

Document downloaded from:

<http://hdl.handle.net/10251/37048>

This paper must be cited as:

Criado Herrero, R.; Garcia Gonzalez, E.; Pedroche Sánchez, F.; Romance, M. (2013). A new method for comparing rankings through complex networks: Model and analysis of competitiveness of major European soccer leagues. *Chaos*. 23(4):1-10. doi:10.1063/1.4826446.



The final publication is available at

<http://dx.doi.org/10.1063/1.4826446>

Copyright American Institute of Physics (AIP)

# A new method for comparing rankings through complex networks: Model and analysis of competitiveness of major European soccer leagues

Regino Criado,<sup>1,2</sup> Esther García,<sup>1</sup> Francisco Pedroche,<sup>3</sup> and Miguel Romance<sup>1,2</sup>

<sup>1</sup>*Department of Applied Mathematics, Rey Juan Carlos University, Madrid (Spain)*

<sup>2</sup>*Center for Biomedical Technology (CTB), Technical University of Madrid (Spain)*

<sup>3</sup>*Institut de Matemàtica Multidisciplinària, Universitat Politècnica de València, Valencia (Spain)*

In this paper we show a new technique to analyze families of rankings. In particular we focus on sports rankings and, more precisely, on soccer leagues. We consider that two teams compete when they change their relative positions in consecutive rankings. This allows to define a graph by linking teams that compete. We show how to use some structural properties of this *competitiveness graph* to measure to what extent the teams in a league compete. These structural properties are the mean degree, the mean strength and the clustering coefficient. We give a generalization of the Kendall's correlation coefficient to more than two rankings. We also show how to make a dynamic analysis of a league and how to compare different leagues. We apply this technique to analyze the four major European soccer leagues: Bundesliga, Italian Lega, Spanish Liga, and Premier League. We compare our results with the classical analysis of sport ranking based on measures of competitive balance.

**An important feature of a sport competition is the uncertainty about the outcome. Sport industry, governments and followers are interested in having some degree of uncertainty about the competition. In the field of sport rankings the term “competitive balance” attends to measure this degree of uncertainty about the result of a competition. A high competitiveness means that there is high uncertainty about the teams ranking. Classical measures of competitiveness are based on the ratio of wins of each team or other related measures. In this paper we show a new perspective by using techniques from complex networks. We show how to use an ad hoc graph, that we called “competitiveness graph”, to give some measures of the competitiveness of a family of rankings. As an application we make a comparison of the four major European soccer leagues during 2011-12 season and 2012-13 season.**

## I. INTRODUCTION

When studying the ranking of a sport league (e.g., the final rank in a soccer league) one can use usual statistical measures such as the mean and the standard deviation. In fact, this is usually done in sports ranking [1]. We are interested in the dynamic behaviour of a sport league and therefore we need to compare  $r \geq 2$  rankings.

Studies on comparison of  $r$  rankings can be traced back to the seminal paper of Kendall [2], where the Kendall's concordance coefficient is defined. Previous works were focused on the correlation of only two rankings, see, e.g. [3] where the Kendall's correlation coefficient  $\tau$  is defined. We can distinguish three ways of comparing two rankings: a) to use a correlation coefficient (e.g, Spearman's  $\rho$  or Kendall's  $\tau$ , [4]) b) to use a distance between the rankings (such as Spearman's footrule  $D$ , [5], or other

metrics [6]) and c), to use Kendall's concordance coefficient (see [7] for a review). In [8] the authors prove that a) and b) are equivalent. In this paper we use a new technique of comparing  $r$  rankings by using some measures derived from graph theory. In more detail, given  $r$  rankings we use a *competitiveness graph* (CG). Some properties of this CG give us information about the *competitiveness* of the  $r$  rankings. The theoretical definition of the CG and its relation to some well known objects in graph theory are studied in [9]. In this paper we focus on how to use the CG to compare the four major European soccer leagues. In the field of analysis of sport leagues it is commonly used the term *competitive balance* to denote the analysis of sport rankings, see [1] or [10] for historical background. For some results concerning competitive balance in European soccer leagues, see [11], [12], [13]. In section IV we review in detail some classical measures of competitive balance applied to sport rankings.

Our interest in competitiveness comes from the concept of *Competitiveness group*, defined in [14], related with rankings of users in Social Networks based on the Personalized PageRank (PPR); see [15], [16] for details on how to use PPR in the field of Social Networks. In [17] we show a theoretical analysis of PPR, that gives insight about the concept of competitiveness, and we introduced the concept of *effective competitors*. This latter concept motivated the study of the competitors in the frame of complex networks theory.

The structure of the paper is the following. In section II we introduce the basic notation and the main structural properties of the *competitiveness graph* that may be used to analyze the competitiveness of a family of graphs. For each measure introduced we indicate the meaning of *being more competitive*. In section III we analyze the competitiveness of the four major European soccer leagues by using the previous measures. We show how to make an evolutionary analysis of each league, and how to compare the competitiveness of the leagues. In section IV we review some classical measures of com-

petitiveness and we compute them for the four studied leagues. In section V we give some conclusions about the use of the structural properties introduced and we highlight some interesting conclusions derived from the application of our methodology to the four soccer leagues analyzed.

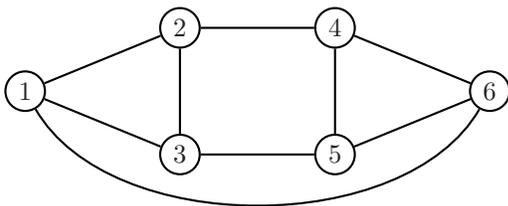
## II. COMPARING RANKINGS AND COMPLEX NETWORK ANALYSIS: A MATHEMATICAL MODEL

Given a set of elements  $\mathcal{N} = \{1, \dots, n\}$  that we will call *nodes* we define a *ranking*  $c$  of  $\mathcal{N}$  as any bijection  $c : \mathcal{N} \rightarrow \mathcal{N}$ . We will write  $i \prec_c j$  when node  $i$  appears first than node  $j$  in the ranking  $c$ . Given a finite family  $\mathcal{R} = \{c_1, c_2, \dots, c_r\}$  of rankings we say that the pair of nodes  $(i, j) \in \mathcal{N}$  *compete* if there exists  $t \in \{1, 2, \dots, r-1\}$  such that  $i$  and  $j$  exchange their relative positions between consecutive rankings  $c_t$  and  $c_{t+1}$ . We define the *competitiveness graph* of the family of rankings  $\mathcal{R}$ , denoted by  $G_c(\mathcal{R}) = (\mathcal{N}, E_{\mathcal{R}})$ , where  $E_{\mathcal{R}}$  denotes the set of edges, as the undirected graph with nodes  $\mathcal{N}$  and edges given by the rule: there is a link between  $i$  and  $j$  if  $(i, j)$  compete.

For example, consider a set with 6 nodes, labeled from 1 to 6, and the following rankings:

$$\begin{aligned} c_1 &: (1, 2, 3, 4, 5, 6), \\ c_2 &: (1, 3, 4, 2, 5, 6), \\ c_3 &: (1, 2, 5, 3, 4, 6), \\ c_4 &: (3, 2, 6, 1, 5, 4), \end{aligned}$$

where the ranking position is given from left to right; for example, in ranking  $c_4$  the node with highest position is node 3, the node with the second highest position is node 2, etc. The competitiveness graph of this set of nodes and the family of rankings  $\mathcal{R} = \{c_1, c_2, c_3, c_4\}$  is:



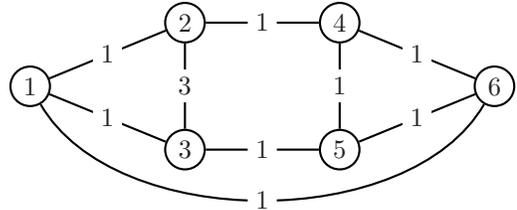
There is a link between nodes 1 and 6 because they exchange their relative positions in rankings  $c_3$  and  $c_4$ ; similarly, there is a link between nodes 1 and 2 because node 1 appears first than node 2 in ranking  $c_3$  but node 2 appears first than node 1 when considering ranking  $c_4$ .

A deeper study of the structural properties of this type of graphs and the interplays with other classic families of graphs can be found in [9].

We say that two nodes  $i, j$  compete  $k$ -times if  $k$  is the maximal number of rankings where  $i$  and  $j$  compete. The *evolutionary competitiveness graph* of  $\mathcal{R}$ , denoted by  $G_c^e(\mathcal{R}) = (\mathcal{N}, E_{\mathcal{R}}^e)$ , will be the weighted undirected

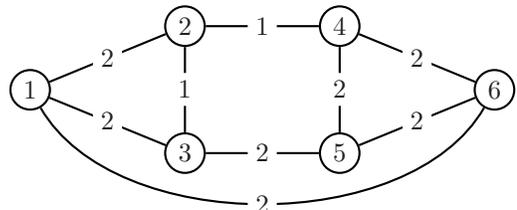
graph with nodes  $\mathcal{N}$  and edges given by the rule: there is an edge between  $i$  and  $j$  labeled with weight  $k$  if  $(i, j)$  compete  $k$  times. Note that the underlying (unweighed) network behind the (weighed) graph  $G_c^e(\mathcal{R})$  is  $G_c(\mathcal{R})$ .

The evolutionary competitiveness graph  $G_c^e(\mathcal{R})$  of the set of nodes  $\{1, 2, 3, 4, 5, 6\}$  and the family of rankings  $\mathcal{R} = \{c_1, c_2, c_3, c_4\}$  is the weighted graph:



The link between nodes 2 and 3 has weight 3 because this pair of nodes compete 3 times: they exchange their relative position between rankings  $c_1$  and  $c_2$ , between  $c_2$  and  $c_3$ , and between  $c_3$  and  $c_4$ . The rest of links are weighted with 1 because their extreme nodes only exchange their relative positions once.

The order of the rankings is fundamental in the calculation of the weights of the evolutionary competitiveness graph, although it does not have influence in the underlying (unweighed) competitiveness graph. For example, if we now consider the same set of nodes  $\{1, 2, 3, 4, 5, 6\}$  and the family of rankings  $\hat{\mathcal{R}} = \{c_1, c_3, c_4, c_2\}$ , we will obtain the following evolutionary competitiveness  $G_c^e(\hat{\mathcal{R}})$  graph:



Notice that the list of node strengths in  $G_c^e(\mathcal{R})$  is  $(3, 5, 5, 3, 3, 3)$  while the list of node strengths for  $G_c^e(\hat{\mathcal{R}})$  is  $(6, 4, 5, 5, 6, 6)$ .

There are several ways to define the *competitiveness* in order to compare two (or more) families of rankings  $\mathcal{R} = \{c_1, c_2, \dots, c_r\}$  and  $\mathcal{S} = \{c_1, c_2, \dots, c_s\}$  possibly coming from different sets of nodes or competitors  $\mathcal{N} = \{1, \dots, n\}$  and  $\mathcal{N}' = \{1, \dots, n'\}$ . In the sequel, let  $G_c^e(\mathcal{R}) = (\mathcal{N}, E_{\mathcal{R}}^e)$  and  $G_c^e(\mathcal{S}) = (\mathcal{N}', E_{\mathcal{S}}^e)$  be two different evolutionary competitiveness graphs. Sometimes we will also refer to the underlying unweighed competitiveness graphs  $G_c(\mathcal{R}) = (\mathcal{N}, E_{\mathcal{R}})$  and  $G_c(\mathcal{S}) = (\mathcal{N}', E_{\mathcal{S}})$ . We will consider, as measures of competitiveness, some different parameters:

**Normalized mean degree.** We define the *normalized mean degree* of a family of rankings  $\mathcal{R}$  as the sum of all the node degrees in the competitiveness graph  $G_c(\mathcal{R})$  divided by the sum over all nodes of their highest possible degrees

$$\text{ND}(\mathcal{R}) = \frac{1}{n(n-1)} \sum_{i \in \mathcal{N}} \text{deg}(i). \quad (1)$$

This parameter gives us a global idea about the relative number of times any two possible competitors have exchanged their respective positions through the corresponding family of rankings. In the previous example, the normalized mean degree  $\text{ND}(\mathcal{R}) = \frac{18}{5}$ .

We say that  $\mathcal{R}$  is more competitive than  $\mathcal{S}$  with respect to the normalized mean degree if  $\text{ND}(\mathcal{R}) > \text{ND}(\mathcal{S})$ .

**Normalized mean strength.** The strength of a node in a weighted graph is the sum of the weights of its incident edges. We define the *normalized mean strength* of a family of rankings  $\mathcal{R}$  as the sum of all edge weights in the evolutive competitiveness graph  $G_c^e(\mathcal{R})$  divided by sum over all possible edges of their highest possible weights:

$$\text{NS}(\mathcal{R}) = \frac{w(E_{\mathcal{R}}^e)}{\binom{n}{2}(r-1)}, \quad (2)$$

where  $w(E_{\mathcal{R}}^e)$  denotes the sum of all weights of the edges of the evolutive competitiveness graph.

We say that  $\mathcal{R}$  is more competitive than  $\mathcal{S}$  with respect to the normalized mean strength if  $\text{NS}(\mathcal{R}) > \text{NS}(\mathcal{S})$ .

In the examples above,  $\hat{\mathcal{R}}$  is more competitive than  $\mathcal{R}$  with respect to the normalized mean strength because  $\text{NS}(\hat{\mathcal{R}}) = \frac{16}{45}$  while  $\text{NS}(\mathcal{R}) = \frac{11}{45}$ .

**Clustering coefficient.** In graph theory, a clique is a set of nodes mutually connected between them. For example, a triangle is a clique formed by three nodes. The clustering coefficient measures how many nodes in a graph tend to cluster together. The clustering coefficient  $C_i$  of a node  $i$  (see, for example [18]) is defined as

$$C_i = \frac{e_i}{\binom{k_i}{2}}, \quad (3)$$

where  $k_i$  is the number of neighbors of node  $i$ ,  $e_i$  is the number of connected pairs between the neighbors of  $i$ , and  $\binom{k_i}{2}$  represents all possible pairs between the neighbors of  $i$ . Given a family of rankings  $\mathcal{R}$ , the *clustering coefficient* of  $\mathcal{R}$  is the average of the clustering coefficients of the nodes of the competitiveness graph  $G_c(\mathcal{R})$ , i.e.,

$$C(\mathcal{R}) = \frac{1}{n} \sum_{i \in \mathcal{N}} C_i. \quad (4)$$

For instance, the clustering coefficient of  $\mathcal{R}$  in the example above is  $\frac{1}{3}$  since the clustering coefficient of each node of  $G_c(\mathcal{R})$  is  $\frac{1}{3}$ .

We say that  $\mathcal{R}$  is more competitive than  $\mathcal{S}$  with respect to the clustering coefficient  $\mathcal{C}$  if  $C(\mathcal{R}) > C(\mathcal{S})$ .

Similarly, we can consider other graph parameters such as the normalized size of the maximal clique (i.e., the number of nodes of the maximal clique contained in the graph divided by the number of nodes of the graph) the normalized size of the largest connected component, etc. For each of these parameters, a family of rankings is more competitive than another family if this parameter in the (evolutive) competitiveness graph is bigger, see [19]. There are other graph parameters that work the other

way round: the smaller they are, the more competitive a family of rankings is. Examples of such parameters are the number of connected components and the Kendall's coefficient  $\tau$  that we will treat in the sequel.

**Generalized Kendall's  $\tau$  correlation coefficient.** It is well known that if we take two rankings  $c_1$  and  $c_2$  of a set  $\mathcal{N}$  of  $n$  elements, then the *Kendall's  $\tau$  correlation coefficient* [3] is defined as

$$\tau(c_1, c_2) = \frac{\tilde{K}(c_1, c_2) - K(c_1, c_2)}{\binom{n}{2}}, \quad (5)$$

where  $\tilde{K}(c_1, c_2)$  denotes the number of pairs  $(i, j)$  that do not compete with respect to  $\mathcal{R} = \{c_1, c_2\}$ , and  $K(c_1, c_2)$  denotes the number of pairs  $(i, j)$  that compete. Notice that  $\binom{n}{2}$  is the number of all possible pairs of nodes  $(i, j)$ . If we consider the competitiveness graph  $G_c(\mathcal{R})$  with respect to  $\mathcal{R} = \{c_1, c_2\}$  then  $K(c_1, c_2) = |E_{\mathcal{R}}|$ , the number of edges of  $G_c(\mathcal{R})$ , and  $\tilde{K}(c_1, c_2) = \binom{n}{2} - |E_{\mathcal{R}}|$ , so

$$\tau(c_1, c_2) = 1 - \frac{2|E_{\mathcal{R}}|}{\binom{n}{2}} = 1 - \frac{4|E_{\mathcal{R}}|}{n(n-1)}. \quad (6)$$

The number of edges  $E_{\mathcal{R}}$  of the competitiveness graph  $G_c(\mathcal{R})$  of a family  $\mathcal{R}$  of rankings is related with the Kendall's correlation coefficient of two rankings: if we denote by  $E(c_1, c_2)$  the edges of the competitiveness graph of the family  $\{c_1, c_2\}$ , we have that

$$\begin{aligned} |E_{\mathcal{R}}| &\geq \max_{c_1, c_2 \in \mathcal{R}} |E(c_1, c_2)| \\ &\geq \frac{n(n-1)}{4} (1 - \min \tau(c_1, c_2)) \end{aligned} \quad (7)$$

and this inequality is in fact an equality when  $r = 2$ . Similarly,  $E_{\mathcal{R}} = \bigcup_{c_1, c_2 \in \mathcal{R}} E(c_1, c_2)$  so

$$\begin{aligned} |E_{\mathcal{R}}| &\leq \sum_{c_1, c_2 \in \mathcal{R}} |E(c_1, c_2)| \\ &= \frac{n(n-1)}{4} \left( \binom{r}{2} - \sum_{c_1, c_2 \in \mathcal{R}} \tau(c_1, c_2) \right). \end{aligned} \quad (8)$$

This inequality turns to be an equality when  $r = 2$ .

We can define a *generalized Kendall's correlation coefficient*  $\tau(\mathcal{R})$  of a family  $\mathcal{R}$  of  $r \geq 2$  rankings: following the original definition (number of pairs that do not compete  $\tilde{K}(\mathcal{R})$  minus number of pairs that compete  $K(\mathcal{R})$ , divided by the number of all possible pairs  $\binom{n}{2}$ ), we set

$$\begin{aligned} \tau(\mathcal{R}) &= \frac{\tilde{K}(\mathcal{R}) - K(\mathcal{R})}{\binom{n}{2}} = 1 - \frac{2|E_{\mathcal{R}}|}{\binom{n}{2}} \\ &= 1 - \frac{4|E_{\mathcal{R}}|}{n(n-1)}. \end{aligned} \quad (9)$$

We can also construct an *evolutive Kendall's correlation coefficient*  $\tau(\mathcal{R})_e$  if we take into account the number

of times each pair of nodes compete. In this sense, we define

$$\tau(\mathcal{R})_e = 1 - \frac{2 w(E_{\mathcal{R}}^e)}{\binom{n}{2}(r-1)}, \quad (10)$$

where  $w(E_{\mathcal{R}}^e)$  denotes the sum of all weights of the edges of the evolutive competitiveness graph. The denominator  $\binom{n}{2}(r-1)$  represents the sum over all possible edges of their highest possible weights.

The evolutive Kendall's coefficient of a family of rankings  $\mathcal{R}$  is directly related to the normalized strength of the family

$$\tau_e(\mathcal{R}) = 1 - 2 \text{NS}(\mathcal{R}). \quad (11)$$

We say that  $\mathcal{R}$  is more competitive than  $\mathcal{S}$  with respect to the Kendall's coefficient if  $\tau_e(\mathcal{R}) < \tau_e(\mathcal{S})$ . Notice that the smaller the Kendall's coefficient  $\tau_e(\mathcal{R})$  is, the more competitive  $\mathcal{R}$  is.

By using these parameters we can compare the competitiveness of the major European soccer leagues in 2011-12 and 2012-13. This is the main goal of next section.

### III. AN APPLICATION TO THE COMPETITIVENESS OF THE MAJOR EUROPEAN SOCCER LEAGUES IN 2011-12

In this section we will illustrate the method introduced in section II by analyzing the competitiveness of four major European soccer leagues during 2011-12 season and 2012-13 season. From a competitiveness point of view, a season of a (soccer) league of  $n$  teams is a family of  $2n-2$  rankings  $\mathcal{R} = \{c_1, \dots, c_{2n-2}\}$  such that each  $c_k$  corresponds to the classification of the teams at matchday  $k$ . Note that the ranking are temporarily ordered and they aggregate all the information about all the matches and their results from the beginning of the season until the corresponding matchday.

We have chosen the classifications of the following major European soccer leagues along 2011-12 and 2012-13:

- The German **Fußball-Bundesliga** [20] (also called, in short, the *Bundesliga*) which is the top league at the German soccer league system. During season 2011-12 and 2012-13 it is contested by 18 teams and operates on a system of promotion and relegation with the *2nd. Bundesliga*.
- The Italian **Lega Serie A** [21] (also called *Serie A TIM* or the *Calcio*, until 2010), which is at the top of the Italian soccer league system and it is contested by 20 Italian teams.
- The Spanish **Liga BBVA** [22] (also called *La Liga*), which is the First Division of the Spanish Professional Soccer National Association (LFP). It is contested by 20 teams, with the three lowest placed teams relegated to the *Segunda División* and

replaced by the top two teams in that division plus the winner of a play-off.

- The British **Barclays Premier League** [23] (also simply called the *Premier League*) which is the professional league at the top of the English men's association soccer clubs. It is contested by 20 clubs and it operates on a system of promotion and relegation with the *Football League*.

Since each National League of  $n$  teams along a season can be understood as a (ordered) family  $\mathcal{R}$  of  $2n-2$  rankings, we can consider its (evolutive) competitiveness graph  $G_c^e(\mathcal{R})$ , that we will call the (evolutive) competitiveness graph of the league in the corresponding season. Figure 1 shows the evolutive competitiveness graphs of the four European Leagues considered during 2012-13 (see [19] for the competitiveness graphs of the Leagues considered during 2011-12). This Figure shows, e.g, that the FC Bayern-München is not connected with any team: it occupied the first position during all the season 2012-13. The same happens with FC Barcelona in the Spanish league.

Now we can obtain information about the competitiveness in each league from structural properties of the corresponding (evolutive) competitiveness graph  $G_c^e(\mathcal{R})$ , following the ideas introduced in section II. Table I shows several structural parameters for the four Leagues during 2011-12 and 2012-13 season and we can see that there are some clear differences in the competitiveness among the leagues. Actually, some leagues are more competitive with respect to some parameters, while the ordering changes if we consider other structural measures. For example, the Bundesliga is more competitive than the Lega Serie A during 2011-12 season with respect to the Kendall's coefficient  $\tau$ , but the Lega Serie A is more competitive than the Bundesliga in the same season with respect to the evolutive Kendall's coefficient  $\tau_e$ . In general, the results derived from the weight distributions of the (evolutive) competitiveness graph (such as the normalized mean strength or the evolutive Kendall's coefficient  $\tau_e$ ) are more accurate than those coming from the (unweighed) competitiveness graph (such as the normalized mean degree). Following this idea, we get that during 2011-12 season the Italian Lega Serie A was the most competitive one, followed by the Spanish Liga BBVA, the German Bundesliga and finally the British Premier League. Similarly, during 2012-13 season, the most competitive league was the German Bundesliga, followed by the Spanish Liga BBVA, the Italian Lega Serie A and finally the British Premier League, despite the fact that either in the German and the Spanish Leagues the champion of each league (The Bayern München and F.C.Barcelona) occupied the first position along the whole season.

In addition to the *static* analysis of certain structural parameters presented before, we can consider the evolution of these parameters along the season in order to measure the fluctuations of competitiveness. In fact we can

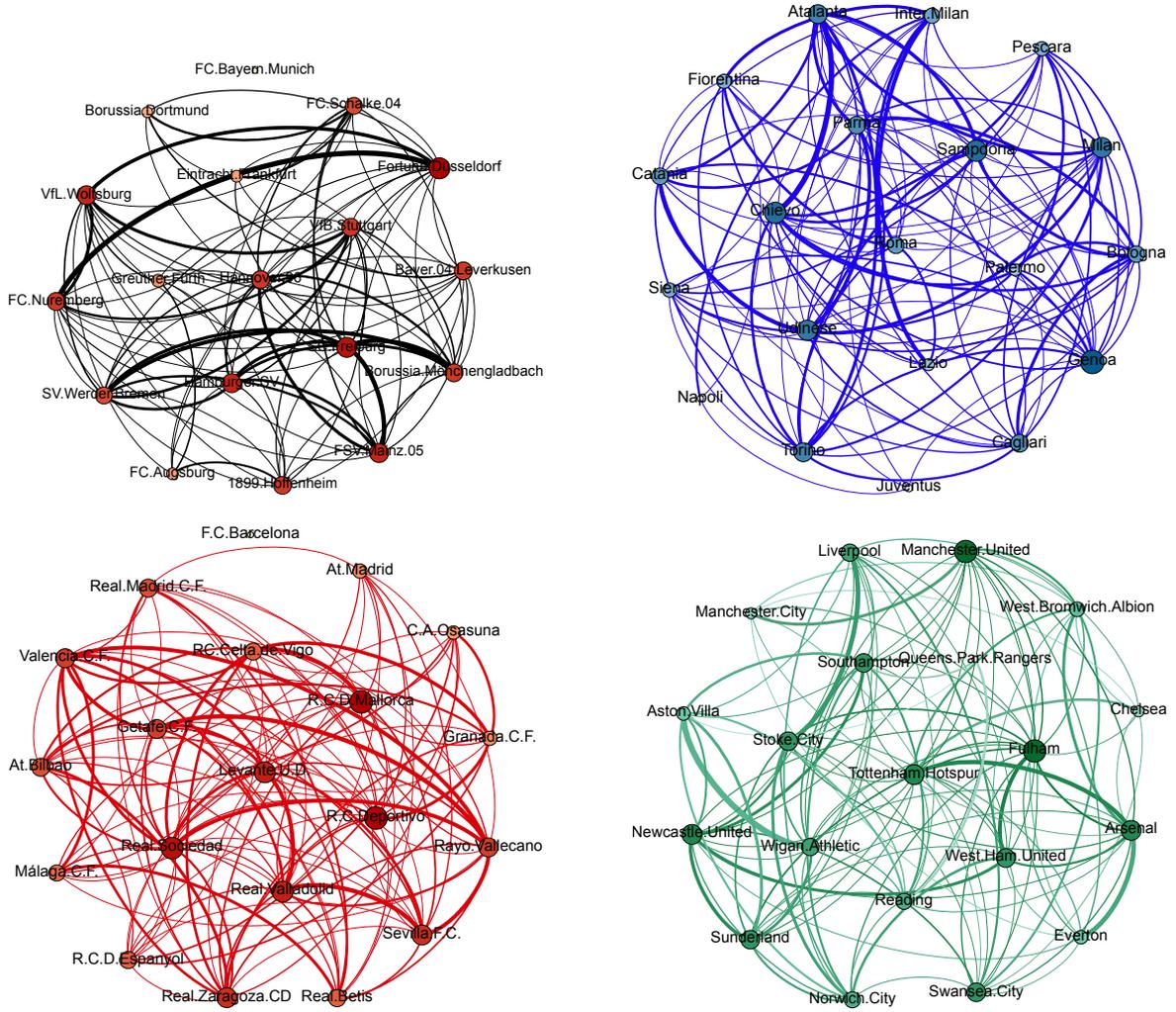


FIG. 1. The evolutive competitiveness graph for the German *Bundesliga* (on black), the Italian *Lega Serie A* (on blue), the Spanish *Liga BBVA* (on red) and the British *Premier* (on green) during season 2012-13.

TABLE I. Computation of the normalized mean degree ( $ND(\mathcal{R})$ ), the normalized mean strength ( $NS(\mathcal{R})$ ), the clustering coefficient ( $C(\mathcal{R})$ ), the generalized Kendall's  $\tau$  coefficient ( $\tau(\mathcal{R})$ ), the evolutive Kendall's coefficient ( $\tau_e(\mathcal{R})$ ), the number of connected components ( $NC(\mathcal{R})$ ), the normalized size of the maximal connected component ( $SC(\mathcal{R})$ ) and the normalized size of the maximal clique ( $SQ(\mathcal{R})$ ) of the evolutive competitiveness graph  $G_v^e(\mathcal{R})$  of the considered leagues in 2011-12 and 2012-13.

| League and Season      | $ND(\mathcal{R})$ | $NS(\mathcal{R})$ | $C(\mathcal{R})$ | $\tau(\mathcal{R})$ | $\tau_e(\mathcal{R})$ | $NC(\mathcal{R})$ | $SC(\mathcal{R})$ | $SQ(\mathcal{R})$ |
|------------------------|-------------------|-------------------|------------------|---------------------|-----------------------|-------------------|-------------------|-------------------|
| Bundesliga 2011-12     | 0.7255            | 0.0604            | 0.8126           | -0.4510             | 0.8792                | 1                 | 1                 | 0.5556            |
| Lega Serie A 2011-12   | 0.6895            | 0.0690            | 0.8300           | -0.3789             | 0.8620                | 1                 | 1                 | 0.5500            |
| Liga BBVA 2011-12      | 0.6368            | 0.0656            | 0.8273           | -0.2737             | 0.8688                | 1                 | 1                 | 0.5500            |
| Premier League 2011-12 | 0.5684            | 0.0572            | 0.7887           | -0.1368             | 0.8856                | 2                 | 0.9000            | 0.6000            |
| Bundesliga 2012-13     | 0.6078            | 0.0636            | 0.7721           | -0.2157             | 0.8728                | 2                 | 0.9444            | 0.5000            |
| Lega Serie A 2012-13   | 0.6158            | 0.0576            | 0.7801           | -0.2316             | 0.8848                | 1                 | 1                 | 0.5000            |
| Liga BBVA 2012-13      | 0.6737            | 0.0619            | 0.8143           | -0.3474             | 0.8762                | 2                 | 0.9500            | 0.5500            |
| Premier League 2012-13 | 0.6053            | 0.0536            | 0.7710           | -0.2105             | 0.8927                | 1                 | 1                 | 0.4000            |

associate a temporal (weighted) network  $G_c^e(\mathcal{R}, t)$  to each league of  $n$  teams as follows. For every  $1 \leq t \leq 2n - 2$ ,  $G_c^e(\mathcal{R}, t) = G_c^e(\mathcal{R}_t)$ , where  $\mathcal{R}_t = \{c_1, \dots, c_t\}$  (i.e.  $\mathcal{R}_t$  is the set of rankings corresponding to the  $t$  first matchdays). Hence, the structural evolution of some structural parameters of this time-varying topology network  $G_c^e(\mathcal{R}, t)$  is useful for analyzing the fluctuations of the competitiveness along the season. Furthermore, this evolution can help spotting the matchday that gave a boost in the competitiveness of the league. In order to avoid spurious fluctuations coming from the first matchdays, we only analyze the evolution of the temporal network  $G_c^e(\mathcal{R}, t)$  with  $4 \leq t \leq 2n - 2$ .

Following this idea, Figure 2 shows the evolution of the generalized Kendall's coefficient  $\tau$  (see panels (a) and (d)), the evolutive Kendall's coefficient  $\tau_e$  (panels (b) and (e)) and the clustering coefficient (panels (c) and (f)) for the four major European leagues considered during 2011-12 and 2012-13. This Figure shows that in both seasons the *competitiveness rankings* among the four leagues are different if we consider generalized Kendall's coefficient  $\tau$  or the evolutive Kendall's coefficient  $\tau_e$ . Furthermore, if we consider other structural parameter (the clustering coefficient), the corresponding ranking coincides with the ranking coming from the evolutive Kendall's coefficient  $\tau_e$  in 2011-12 season (where the Lega Serie A was the most competitive, followed by the Liga BBVA, the Bundesliga and finally the Premier League) and it coincides with the ranking coming from the generalized Kendall's coefficient  $\tau$  in 2012-13 season (where the most competitive league was the Spanish Liga BBVA, followed by Lega Serie A, the Bundesliga and the Premier League).

The analysis of the evolution of these structural parameters is also useful for studying the differences in the evolution of the competitiveness during 2011-12 and 2012-13. In 2011-12 season there were no significant fluctuations in the competitiveness rankings along the seasons (i.e. the competitiveness ranking among the four Leagues does not change too much along the season), while in 2012-13 season there were some changes in the competitiveness ranking specially during the second half of the season. For example, panel (e) shows that, despite the fact that the competitiveness of the Bundesliga was quite low at the beginning of the season, it got better, by overcoming the BBVA Liga about matchday 20, which makes the Bundesliga more competitive since then. Similar analysis can be performed by considering other structural parameters (such as the normalized mean strength, the normalized size of the maximal clique and other), as it is illustrated in [19].

Similarly, we can fix a league and compare the competitiveness of this league along different seasons by analyzing the evolution of some structural parameters of the temporal network  $G_c^e(\mathcal{R}, t)$  along different seasons. Figure 3 shows a comparison of the normalized mean degree (panels from (a) to (d)) and the normalized mean strength (panels from (e) to (h)) for the German Bundesliga (panels (a) and (e)), the Italian Lega Serie A

(panels (b) and (f)), the Spanish Liga BBVA (panels (c) and (g)) and British Premier League (panels (d) and (h)) during 2011-12 season (lines in red in all the panels) and 2012-13 (lines in black in all the panels). This figure illustrates that the conclusions obtained strongly depend on the kind of structural parameter that we are considering. In general, the parameters that take into account the weights of the evolutive competitiveness graph are more sensible and therefore they give a better knowledge of the competitiveness of the soccer leagues.

In any case, the behavior of each of the four soccer leagues along 2011-12 and 2012-13 are quite different. For example, if we analyze the evolution of the normalized mean strength (panels from (e) to (h) in Figure 3) we can see four different situations. The Bundesliga (panel (e)) was less competitive during the first half of 2011-12 than during the same period of 2012-13, but its competitiveness improved significantly in the second half of 2012-13 and it was more competitive during this period than in the previous season. The Lega Serie A (panel (b)) was definitively less competitive during 2012-13 than during 2011-12. The Liga BBVA exhibits a behavior that it is the opposite to the Bundesliga's one: it was much more competitive during the first half of 2012-13 season, but in the second half of the season the competitiveness was reduced and actually it was more competitive in the second half of 2011-12 than in 2012-13. Finally, the Premier League kept its competitiveness at the same level during 2011-12 and 2012-13. Similar analysis can be performed by considering other structural parameters (such as the normalized mean strength, the normalized size of the maximal clique and other), as it is illustrated in [19].

We also consider another type of analysis of the competitiveness of the European soccer leagues that take into account the distribution of the weights of the evolutive competitiveness graph. If we have two different leagues with the same strength at each node, the competitiveness should be different if these strengths are concentrated in some couples of nodes than if they are distributed in a big set of pairs of *loosely competitor* nodes. In other words, we should measure how deeply the couples of teams compete in order to get a sharp idea of the competitiveness in each league. It should be interesting to take the evolutive competitiveness graph of a league of a given season and only considering the competitiveness between nodes that flip their positions a significant number of times. In order to do this, we can fix a threshold  $0 \leq \omega_o \in \mathbb{Z}$  and study the structural properties of the *filtered* evolutive competitiveness graph  $G_c^e(\mathcal{R})|_{\geq \omega_o}$ , which is the graph whose (weighted) adjacency matrix  $\tilde{A} = (\tilde{a}_{ij})$  is given by

$$\tilde{a}_{ij} = \begin{cases} a_{ij}, & \text{if } a_{ij} \geq \omega_o, \\ 0, & \text{otherwise,} \end{cases} \quad (12)$$

where  $A = (a_{ij})$  is the (weighted) adjacency matrix of the evolutive competitiveness graph  $G_c^e(\mathcal{R})$ . Note that for every  $0 \leq \omega_o \in \mathbb{Z}$ , the *filtered* evolutive competitiveness graph  $G_c^e(\mathcal{R})|_{\geq \omega_o}$  only takes into account the competitiveness

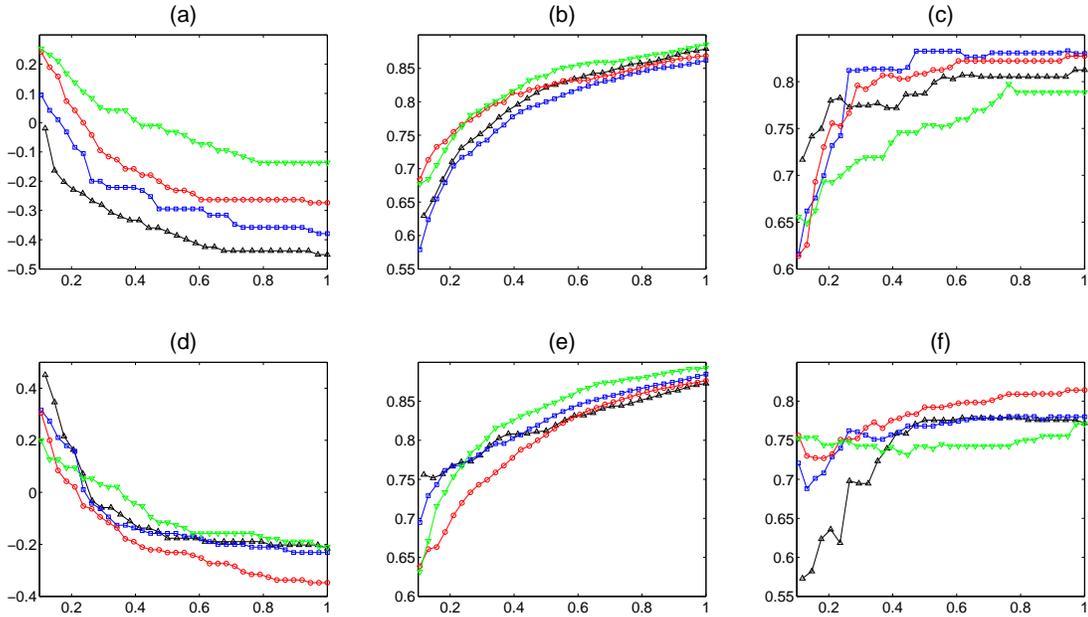


FIG. 2. A comparison of the generalized Kendall's coefficient  $\tau$  (panels (a) and (d)), the evolutive Kendall's coefficient  $\tau_e$  (panels (b) and (e)) and the clustering coefficient (panels (c) and (f)) for the German Bundesliga (in black), the Italian Lega Serie A (in blue), the Spanish Liga BBVA (in red) and British Premier League (in green) during 2011-12 (first row, i.e. panels (a), (b) and (c)) and 2012-13 (second row, i.e. panels (d), (e) and (f)).

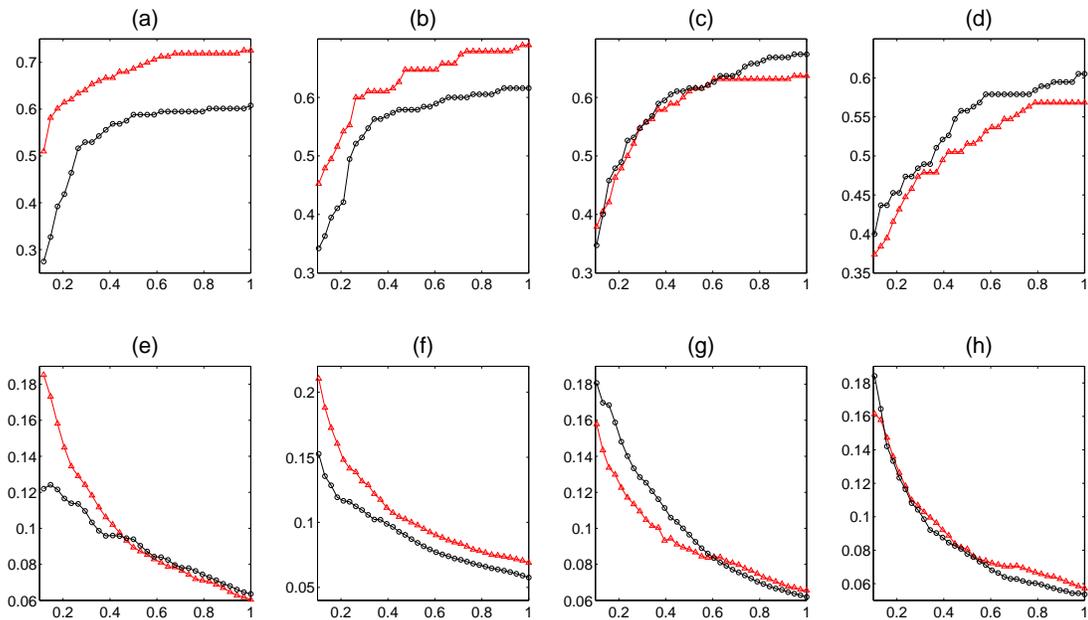


FIG. 3. A comparison of the normalized mean degree (panels from (a) to (d)) and the normalized mean strength (panels from (e) to (h)) for the German Bundesliga (panels (a) and (e)), the Italian Lega Serie A (panels (b) and (f)), the Spanish Liga BBVA (panels (c) and (g)) and British Premier League (panels (d) and (h)) during 2011-12 (lines in red in all the panels) and 2012-13 (lines in black in all the panels).

between the nodes such that compete at least  $w_o$  times. By using this new tool, we can study how deeply the couples of teams compete in a league by measuring several properties of  $G_c^e(\mathcal{R})|_{\geq w_o}$  and how they change when we consider different values of  $0 \leq w_o \in \mathbb{Z}$ .

Following this idea, Figure 4 shows the fluctuation of the evolutive Kendall's coefficient  $\tau_e$  of the *filtered* evolutive competitiveness graph  $G_c^e(\mathcal{R})|_{\geq w}$  as a function of the threshold  $0 \leq w \in \mathbb{Z}$ . In each panel we have considered the *filtered* evolutive competitiveness graph  $G_c^e(\mathcal{R})|_{\geq w}$  for a fixed league (during 2011-12 and 2012-13) and we have computed its evolutive Kendall's coefficient  $\tau_e$  as a function of the threshold  $0 \leq w \in \mathbb{Z}$ . As before, we have considered the German Bundesliga (panels (a) and (e)), the Italian Lega Serie A (panels (b) and (f)), the Spanish Liga BBVA (panels (c) and (g)) and British Premier League (panels (d) and (h)) during 2011-12 (lines in red in all the panels) and 2012-13 (lines in black in all the panels) in Figure 4. In order to get more information, we have computed the original evolutive competitiveness graph  $G_c^e(\mathcal{R})$  at two instances in order to get a more detailed information about the competitiveness: at the end of the first part of the season (panels from (a) to (d)) and at the end of the season (panels from (e) to (h)). Since in section II we have said that a league  $L_1$  is more competitive than  $L_2$  (with respect to the evolutive Kendall's coefficient) if the evolutive Kendall's coefficient of the evolutive competitiveness graph of  $L_1$  is less than the corresponding value for the evolutive competitiveness graph of  $L_2$ , the later the increase of the fluctuation of the evolutive Kendall's coefficient occurs, more competitive a league is. Hence, Figure 4 shows that the four leagues exhibit different behaviors. The competitiveness of the Bundesliga was more competitive in 2012-13 than in 2011-12, either after the first half of the season (panel (a)) and at the end of the season (panel (e)). The competitiveness of the Lega Serie A was less competitive in 2012-13 than in 2011-12, either after the first half of the season (panel (b)) and at the end of the season (panel (f)), i.e. this league has the opposite behavior than the Bundesliga. The competitiveness of the Liga BBVA was higher in 2012-13 than in 2011-12 after the first half of the season (panel (c)), but it was weaker in 2012-13 than in 2011-12 at the end of the season (panel (g)), so the Spanish league lost competitiveness from one season to the following. Finally, the Premier League kept the strength of the competitiveness in a similar level during seasons 2011-12 and 2012-13, either after the first half of the season (panel (d)) and at the end of the season (panel (h)). Similar analysis can be performed by considering other structural parameters (such as the normalized mean strength, the relative size of the maximal clique and other), as it is illustrated in [19]. In next section we apply some classical measures of competitive balance to the leagues just considered.

#### IV. COMPARISON WITH SOME RESULTS OBTAINED BY USING STANDARD MEASURES OF COMPETITIVE BALANCE

The most common measure of competitive balance is the standard deviation of the proportion of wins  $w_i$  each team has achieved in a season. This standard deviation can be defined as

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( w_i - \frac{1}{2} \right)^2}, \quad (13)$$

where  $n$  is the number of teams. In general, the higher is  $\sigma$  the lower is the *competitive balance* (and the higher is the *competitive inequality*). In a league where each team has the same probability of win a match one will have  $\sigma = 0$ , and therefore an increase in  $\sigma$  means a decrease in the *competitive balance*. It is said that a league (or, in general,  $r$  rankings) has increased its competitiveness when it has decreased its  $\sigma$ . Some other measures are variations of this one, see [24], [25]. In particular, to use as a comparison with our measures, we shall compute the coefficient called NAMSIS (National Measure of Seasonal Imbalance, [13]):

$$\text{NAMSIS} = \sqrt{\frac{\sum_{i=1}^n \left( w_i - \frac{1}{2} \right)^2}{\sum_{i=1}^n \left( w_{i,max} - \frac{1}{2} \right)^2}}, \quad (14)$$

which is a normalized  $\sigma$ . The term  $w_{i,max}$  is the win ratio of team  $i$  when there is complete predictability: team 1 wins all the matches, team 2 wins all except two matches (home and away), ..., team  $n$  loses all the matches. In MATLAB notation, the vector with components  $w_{i,max}$  is computed as  $[n-1 : -1 : 0]/(n-1)$ .

Another usual measure of competitive balance is the Herfindahl-Hirschman Index (see [26]). A related index, called HICB ( $H$ -index of competitive balance), see [11], is defined as

$$\text{HICB} = 100 n \sum_{i=1}^n s_i^2, \quad (15)$$

where  $s_i$  is the ratio of points scored in a season by team  $i$ . An increase in the HICB coefficient means a decrease in the competitive balance [11], [1].

A different measure of competitive balance, which is more related with the mathematical literature on rankings, is the *Churn* [27]

$$C_t = \frac{1}{n} \sum_{i=1}^n |c_{i,t} - c_{i,t-1}|, \quad (16)$$

where  $c_{i,t}$  is the ranking of team  $i$  in season  $t$ . Then the *Adjusted Churn* is defined as

$$C_{ad} = \frac{C_t}{C_{t,max}}, \quad (17)$$

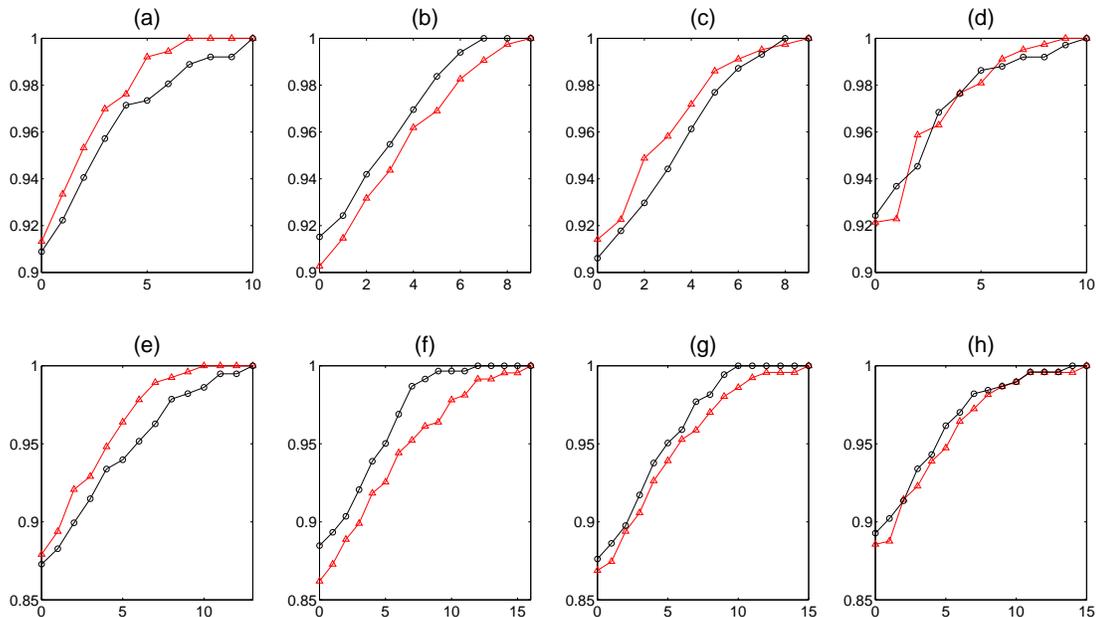


FIG. 4. A comparison of the evolutive Kendall's coefficient  $\tau_e$  of the evolutive competitiveness graph  $G_c^e(\mathcal{R})$  in terms of the threshold applied for the German Bundesliga (panels (a) and (e)), the Italian Lega Serie A (panels (b) and (f)), the Spanish Liga BBVA (panels (c) and (g)) and British Premier League (panels (d) and (h)) during 2011-12 (lines in red in all the panels) and 2012-13 (lines in black in all the panels). The evolutive competitiveness graphs have been considered at the end of the first part of the season (panels from (a) to (d)) and at the end of the season (panels from (e) to (h)).

where  $C_{t,max}$  is the maximum value of  $C_t$  given a league of  $n$  teams. This happens e.g., when  $c_t = [1, 2, \dots, n]$  and  $c_{t+1} = [n, n-1, \dots, 2, 1]$ . This index indicates competitive balance between seasons. If two rankings are the same then  $C_t = 0$ . It is considered that high values of  $C_t$  imply high values of competitive balance [27], [1]. Note that the Adjusted Churn is a normalized Spearman's footrule  $D(c_t, c_{t+1})$ . In fact,  $D(c_t, c_{t+1}) = nC_t$ . Standard tools such as the Spearman's footrule and Kendall's  $\tau$  (see equation (5)) have also been used in sports ranking (see [10], [1], [25]). Kendall's  $\tau$  allows for a dynamic study of the leagues, in allowing the comparison of two seasons. Following [10] a competition is perfectly balanced in a dynamic sense when the ranking is random, that is, any ranking is equally probable, and rankings in consecutive seasons are independent of each other. It is assumed that low values of  $\tau$  correspond to high competitive balance.

In table II we show the NAMSIS coefficient given by equation (14), and the HICB coefficient given by (15). We have computed these coefficients using the data from [20], [21], [22], and [23]. According both the NAMSIS coefficient and the HICB in the season 2011-12 the most competitive league was Lega Serie A. This is in accordance with our conclusions using  $\tau_e$  in section III. Regarding the season 2012-13 the most competitive league was the Spanish Liga BBVA (by using NAMSIS coefficient) or Lega Serie A (by using HICB). Note that the values of these coefficients are very similar for these two

| League         | Season 2011-12 |        | Season 2012-13 |        |
|----------------|----------------|--------|----------------|--------|
|                | NAMSIS         | HICB   | NAMSIS         | HICB   |
| Lega Serie A   | 0.6410         | 107.82 | 0.6495         | 110.40 |
| Liga BBVA      | 0.6581         | 109.74 | 0.6419         | 110.73 |
| Premier League | 0.6669         | 110.54 | 0.7135         | 111.33 |
| Bundesliga     | 0.6788         | 109.58 | 0.6784         | 111.15 |

TABLE II. NAMSIS and HICB coefficients for the four major European soccer leagues.

leagues in that season. We have obtained, using  $\tau_e$  that the most competitive league in this season is Bundesliga. Therefore, in this case our measures give a different conclusion.

In table III we show the values of the Adjusted Churn  $C_{ad}$ , computed by using equation (17), and the Kendall's  $\tau$  for the four analyzed leagues. Both coefficients  $C_{ad}$  show that the Bundesliga is the one that has increased more its competitiveness from season 2011-12 to season 2012-13, since this league has the highest value of  $C_{ad}$  and the lowest value of  $\tau$ . This is in accordance with our conclusions obtained in section III from panel e) of Figure 3, and from panel e) in Figure 4. A similar analysis could be done using another classical measures of competitive balance. As we have noted before, depending on whether measure is used one can obtain different conclusions. In

| League         | Season 2011-12 to season 2012-13 |                  |
|----------------|----------------------------------|------------------|
|                | Adjusted churn                   | Kendall's $\tau$ |
| Lega Serie A   | 0.28                             | 0.63             |
| Liga BBVA      | 0.34                             | 0.54             |
| Premier League | 0.30                             | 0.59             |
| Bundesliga     | 0.37                             | 0.49             |

TABLE III. Adjusted Churn coefficient,  $C_{ad}$ , and Kendall's  $\tau$  for the four major European soccer leagues.

section III we have exploited some of the possibilities of the structural measures introduced in this paper. Some additional results can be obtained from [19].

## V. CONCLUSIONS

We have shown that the use of a graph (*competitiveness graph*) associated to a family of rankings can be useful to analyze the competitiveness associated to the rankings. We have introduced some parameters (mean degree, mean strength, clustering coefficient) that can be used to measure the competitive balance of a sport ranking. We have related the *competitiveness graph* with classical measures of rankings, such as the Kendall's correlation coefficient, and we have given a generalized Kendall's correlation coefficient for  $r$  rankings that can be computed

easily from the competitiveness graph. To show the applicability of our measures we have analyzed the four major European soccer leagues, obtaining among others the following conclusions:

- In 2011-12 season the most competitive league was the Italian Lega Serie A, if we take into account the evolutive Kendall's coefficient, the normalized mean strength and the clustering coefficient (see Table I).
- In 2012-13 season the most competitive league was the German Bundesliga, if we take into account the evolutive Kendall's coefficient, the normalized mean degree, and the normalized mean strength, see Table I).
- From season 2011-12 to season 2012-13 the league that has increased more its competitiveness was the Bundesliga.

## ACKNOWLEDGEMENTS

This paper was partially supported by Spanish MICINN Funds and FEDER Funds MTM2009-13848, MTM2010-16153 and MTM2010-18674, and Junta de Andalucía Funds FQM-264.

- 
- [1] S. Dobson and J. Goddard, *The economics of football (2nd ed.)*, Cambridge Univ. Press. (2011).
- [2] M.G. Kendall and B.B. Smith, *Ann. Math. Statist.* **10**(3), 275 (1939).
- [3] M. G. Kendall, *Biometrika* **30**(1-2), 81 (1938).
- [4] P. Diaconis and R.L. Graham, *J. Roy. Stat. Soc. (Series B)* **39**(2), 262 (1977).
- [5] P.K. Sen, I.A. Salama and D. Quade, *Chilean J. Stat.* **2**(1), 3 (2011).
- [6] R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar and E. Vee, *SIAM J. Discrete Math.* **20**(3), 628 (2006).
- [7] P. Legendre, *J. Agricult. Biol. Environ. Stat.* **10**(2), 226 (2005).
- [8] E.J. Emond and D.W. Mason, *J. Multi-Crit. Decis. Anal.* **11**: 17.28 (2002).
- [9] R. Criado, E. García, F. Pedroche and M. Romance, Preprint (2013).
- [10] L. Groot, *Economics, uncertainty and European football : trends in competitive balance*. Edward Elgar cop. Northampton, USA. (2008).
- [11] T. Pawlowski, Ch. Breuer and A. Hovemann, *J. Sport Economics* **11**(29), 186 (2010).
- [12] A. Feddersen and W. Maennig, *Trends in competitive balance: Is there evidence for growing imbalance in professional sport leagues?*, Hamburg contemporary economic discussions, No. 01/2005, University of Hamburg. 2005.
- [13] K. Goossens, *Rivista di diritto ed economia dello sport* **2**(2), 77 (2006).
- [14] F. Pedroche *Math. Compt. Modelling* **52**, 1052 (2010).
- [15] F. Pedroche, *Int. J. Bif. Chaos* **22**(7), art. no. 1250162. (2012)
- [16] F. Pedroche, F. Moreno, A. González and A.A. Valencia *Math. Compt. Modelling* **57** 1891 (2013).
- [17] E. García, F. Pedroche and M. Romance *Linear Algebra and its Applications* **439**, 640 (2013).
- [18] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, *Physics Reports* **424**, 175 (2006).
- [19] Supplementary material.
- [20] <http://www.bundesliga.com/>
- [21] <http://www.legaseriea.it/>
- [22] <http://www.ligabbva.com/>
- [23] <http://www.premierleague.com/>
- [24] B.R. Humphreys, *J. Sports Economics* **3**(2), 133 (2002).
- [25] M. Kringstad, *Competitive Balance in Complex Professional Sports Leagues*, Doctoral Thesis, The University of Leeds. Leeds University Business School, 2008.
- [26] P.D. Owen, M. Ryan and C.R. Weatherston, *Rev. Ind. Organ* **31**, 289 (2007).
- [27] D. Mizak, J. Neral and A. Stair, *Economics Bull.* **26**(3), 1 (2007).