Search For A Complete And Transitive Ranking Of Football Leagues

Arseniy Stolyarov, Gleb Vasiliev
Higher School of Economics, New Economic School, Toros Research

JEL classification: Z20

Keywords: data science, sports research, economics of sports, football, graph theory

The position of force in football world has changed a lot for the last 10 years (the appearance of China and MLS in the global field, the relative decline of Serie A, etc), consequently, it becomes more and more important to invent a tool to compare different leagues across time. The current availability of data together with the increased intensity of transfers provide more opportunities for such research than ever before. Such research has several practical applications. First of all, it helps to understand the general flaw of power and resource in the modern football, which may be very useful for potential long-term investors that may better understand the changing environment. Moreover, it shows similar leagues in terms of average attractiveness for players (without the description of main influencing factors), which may be very helpful for players that make a transfer choice. Furthermore, several analytical packages now provide statistical data on almost all players in the world. However, comparison of players made solely on the comparison of these numbers is not always a good idea because football is a collective game and, thus, statistical data does not only depend on the talent and motivation of a player but also on the same characteristics of his opponent. Thus, the introduced ranking will at least partially solve the problem and will include comparisons across leagues into the pricing model. Also, such ranking may help both spectators and mass media to choose what to watch and what to show. For advertisement companies it may also be important both currently and strategically (as it does not simply show the current situation, but might be also used for the projection). Lastly, it provides several valid and reliable methods for the analysis of sportive leagues as elements of graph and, thus, applying main methods of graph theory to both classification of leagues and understanding of idiosyncratic cluster effects. The second section of the paper is devoted to the analysis of the existing attempts and methods to resolve the stated classification problem. The third section describes two possibilities (*number-preference* and *fee-preference*) of pairwise comparisons across leagues based solely on the information set of transfers between the two compared leagues. This methods are good in terms of IIA [1], but both of them cannot solve the problems with transitivity of preferences across leagues. The fourth section solves this problem by the use of three standard graph metrics (PageRank, closeness centrality, betweenness centrality) and introduction of a new metric that unites in it the betweenness of the node and the "directness" of the node in terms of the weight of the incoming transfers in the total market value of all transfers corresponding to that league. The fifth section focuses on the various weights that can be given to the calculated parameters and, consequently, to the properties of the obtained ranking.

1 Related Work

Several recent papers use network approach to study football transfer markets.

Lee et al. [6] considers transfers among the clubs from the English Premier League, the Bundesliga and Italian Serie A in the transfer window of summer 2014. This article assigns weights to these transfers based on the number of web search results corresponding to the involved players. The validity of this approach is later proved by the existence of positive correlation between the defined weights and the transfer fees. The authors also investigate topological characteristics of the network and find that the network follows a randomly-mixed pattern of transfers, i.e., inter-league transfers doesn't dominate intra-league ones and vice versa.

Liu et al. [7] use the dataset of player transfers among nearly 400 clubs from 24 leagues from 2010 to 2015 to explore the relationship between different network properties and the clubs' strategy and behaviour in the global transfer market. The authors classify leagues according to the relationship between match performance and transfer profitability of the participating clubs and study correlations between club network characteristics in different team categories.

Naidenova and Parshakov [8] create a transfer network with vertices corresponding to European football clubs and combine vertex features with team attributes from FIFA football game in order to build a linear model for predicting clubs' average points per match. The overall goal of the paper is very different from the one used here, but the metric (node centrality) is the same.

[1] Independence of irrelevant alternatives

2 Methods

The six different types of rankings are provided:

1. **Number-preference** We say that League A is ***number-preferred*** to League B during the period T if the number of transfers from League B to League A exceeds number of transfers per team from League A to League B during the period T. This method does not depend on the measurement error of the actual transfer fee of the individual.
2. **Fee-preference**We say that League A is ***fee-preferred*** to League B during the period T if the total market fee of transfers (measured by Transfermarkt.de) from League B to League A exceeds the total market fee of transfers (measured by Transfermarkt.de) from League A to League B during the period T. This method weights transfers in terms of importance by using transfer-fee as a proxy
3. **Normalized PageRank**The idea is taken from the standard Google PageRank, in which links were substituted by transfers.

$$P_{\sigma_i} = \frac{PageRank}{AvgClubs},$$

where $AvgClubs$ is the average number of clubs in a league for the 2010-2016 period.
4. **The betweennes centrality** of a node (league) here is calculated as follows:

$$C_v = \sum_{s \neq v \neq t} \frac{\sigma_s t(v)}{\sigma_s t}$$

5. **The closeness centrality** is in some sense the average distance between the point and all other points in the graph. In this article we calculate it in a traditional manner as:

$$BC_j = \frac{N}{\sum_{i=1} N d_{ij}}$$

6. As far as the graph theory suggests several methods of evaluation the most important node or vertices, all such methods do not take into account the binary relation between the leagues. The method provided in this section unites the two approaches into one. Here is the formula for the method:

$$R_i = BW_i \times \frac{\sum_{j=1}^{n} fee_j \times \mathbb{1}(v_i = 1)}{\sum_{j=1}^{n} fee_j}$$

where, BW_i is the betweenness centrality of the country node in the undirected graph, fee_j is the fee of a transfer, v_i is a boolean function that is equal to 1 for the transfers to the country and 0 otherwise.

3 Conclusions

This article has several important results:

1. It is impossible to construct a complete and transitive binary relations and, consequently, ranking using only transfers between the pairs of leagues (informational parsimony) due to the high share of intransitive triples, some of which survive even under strict filters.
2. Three different parameters of leagues (nodes) were calculated and may be later used as a proxy of the attractiveness of a league by players. Unless, we cannot state yet that these ratings perfectly deal with the general attractiveness of a league, all of them (or even some linear combination of them can be used as an instrumental variable to struggle with endogeneity of regressors from the overall attractiveness of a league by players.
3. Four versions of Borda count ranking were derived from the parameters of nodes, all of which may be used as an attempt to rank leagues by their attractiveness depending on a researchers understanding of attractiveness.
4. A directed transfer network for the 28 major leagues was calculated, which may be extremely useful to both agents and players.
5. The negative sample rank Spearman correlation of a Normalized PageRank of a league with its centrality is discovered, which is a striking feature and should be explained via some model.

References

[1] Andrew B. Adelsman, Rony M.; Whinston. Sophisticated Voting with Information for Two Voting Functions. *Journal of Economic Theory*, 15(1):145–159, June 1977.

[2] Chris Anderson and David Sally. *The Numbers Game*. Why Everything You Know About Football is Wrong. Penguin UK, May 2013.

[3] Pierre-Louis Giscard, Nils Kriege, and Richard C Wilson. A general purpose algorithm for counting simple cycles and simple paths of any length. *arXiv.org*, December 2016.

[4] Donald B Johnson. Finding All the Elementary Circuits of a Directed Graph. *SIAM Journal on Computing*, 4(1):77–84, March 1975.

[5] K.J.Arrow. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, 58(4):328–346, August 1950.

[6] Sangmin Lee, Inho Hong, and Woo-Sung Jung. A Network Approach to the Transfer Market of European Football Leagues. *New Physics Sae Mulli*, 65(4):402–409, April 2015.

[7] Xiao Fan Liu, Yu-Liang Liu, Xin-Hang Lu, Qi-Xuan Wang, and Tong-Xing Wang. The Anatomy of the Global Football Player Transfer Network: Club Functionalities versus Network Properties. *PLoS ONE*, 11(6):e0156504, February 2016.

[8] Iuliia Naidenova and Petr Parshakov. Transfer Networks In European Football and the Clubs' Performance. In *XVIII April International Academic Conference on Economic and Social Development*. HSE - Perm, April 2017.

[9] H. P. Young. Condorcet's Theory of Voting. *American Political Science Review*, 82(4):1231–1244, March 1988.