New Physics: Sae Mulli,
Vol. 65, No. 4, April 2015, pp. 402∼409

DOI: 10.3938/NPSM.65.402

A Network Approach to the Transfer Market of European Football Leagues

Sangmin Lee · Inho Hong · Woo-Sung Jung*
Department of Physics, Pohang University of Science and Technology, Pohang 790-784, Korea

(Received 24 December 2014 : revised 25 March 2015 : accepted 30 March 2015)

The transfer market for sport players presents an interesting issue both for the people who are directly involved with it and for the many fans. Although attempts have been made to understand it from a perspective of economy and management, an analysis from the view point of a complex system remains in the initial stages. In this research, we analyzed the transfer market of European football leagues as a weighted network in order to understand the detailed transfer patterns. A Google-based standard is used to quantify the value of the 436 transfers that occurred in the summer of 2014. The transfer patterns on a scale of both individual teams and whole leagues validate common sense intuitions about the capitalistic English Premier League. The log-normal distributions for players and teams imply that the network has evolved according to the Yule process. The properties of the network, such as assortativity, strength correlation, and betweenness centrality, provide several significant implications for topology. An assortativity coefficient close to zero represents a randomly-mixed transfer pattern on the league scale, which contradicts the intuitive assumption of disassortativity.

PACS numbers: 89.75.-k, 64.60.aq, 02.50.-r
Keywords: Network, Complex system, Transfer, Football

I. INTRODUCTION

Networks have been studied in order to obtain deep insights into macroscopic aspects of large systems, in cases where the classical approach has reached the limit [1]. The major thrust of recent progress in network science is inspired by the groundbreaking papers of Strogatz [2] and Barabási [3] published at the turn of the century. The world-wide-web [4, 5], social interaction [6, 7], biological systems [8], economic systems [9], and many phenomena have been analyzed using network models. However, the range of possible applications for network models remains enormous, because all interactions between individual agents can be expressed as networks.

The links defined by simple connections in networks can be extended to socioeconomic flows in weighted networks. The flows on networks have been deeply discussed thoroughly investigated for international trades [10–12], energy transmissions [13], and traffic systems [14]. In this manuscript, by adopting existing methods for weighted networks, player transfers between European football teams are analyzed.

Transfer markets in football leagues have attracted the interest of researchers in economy and management. The distinctive features of the football transfer market are a high frequency of transfers and a broad distribution of teams participating in the market. Furthermore, big deals are typically made between teams, not through free agent contracts. These characteristics are suitable for a team-based network analysis, and this can be extended to the league scale as an aggregate of teams. Most previous studies on football transfer markets have focused on statistical analysis [15,16] and the evaluation of players [17–19]. Network analysis has mainly been applied to the matching of games to teams and players [20], and rarely on the transfer of players [21]. The transfer network in reference [21] was treated as an unweighted network based on the number of transferred players.

*E-mail: wsjung@postech.ac.kr

In this study, the European transfer market is analyzed as a weighted network, in order to understand the detailed network structure. We adopt the estimation method using web search results to determine the weight of each link, *i.e.*, the value of each transfer [22, 23]. Meaningful statements on the disjunction between the analysis results and common sense intuitions are extracted from several network properties. The features analyzed, for example, which team traded players a lot and how it will proceed in upcoming seasons, can offer a new statistical aspect for club management.

II. VALUATION OF TRANSFERS BY THE GOOGLE-BASED STANDARD

The network analysis covers the transfers over the summer of 2014, among the three major leagues as ranked by the Union of European Football Association (UEFA): Spanish La Liga, English Premier League (EPL), and German Bundesliga. We also consider trades between the three major leagues and other European leagues; that is, the leagues of Italy, France, Portugal, Russia, Ukraine, and the Netherlands. We have gathered 436 transfer records taking place between 114 teams, from the announcements of clubs and the media [24]. Although the official transfer market of summer 2014 was open from July 1st to September 1st, the transfer data covers a period from February 1st to September 1st, because January 31st is the closure of the transfer market for winter 2013/2014.

Quantifying the influence of each trade or the value of each player is an essential procedure for determining the link weight. The most reasonable criterion for measuring the value of a player would be the wage or the transfer fee. However, only a portion of this data was disclosed as news or gossip by the media, since it is essentially private. We collected 148 transfer fee data items that were disclosed by the various media outlets: The Telegraph [25], Transfermarkt.co.uk [26], and Wikipedia [27]. These do not provide enough information on the whole structure of the network, as only one third of transfers are covered. We propose the Google-based standard as a measure of transfer values, in order to overcome this limitation. The Google-based standard is based on the number of Google search results for a player's name. We confined the search results to Goal.com (*http://www.goal.com*) in order to avoid homonym issues over players's names, and investigated the results from September 1st 2013 to September 1st 2014 to measure the value during this transfer season. Goal.com is the largest on-line football community that posts game data and news, and September 1st is the date of closure of summer transfer markets in both 2013 and 2014. We set the window for the web search results to be one year, because the main transfer market each year is the summer one. Although the web search results may be influenced by personal popularity and media exposure, these external effects can also be regarded as a kind of fame.

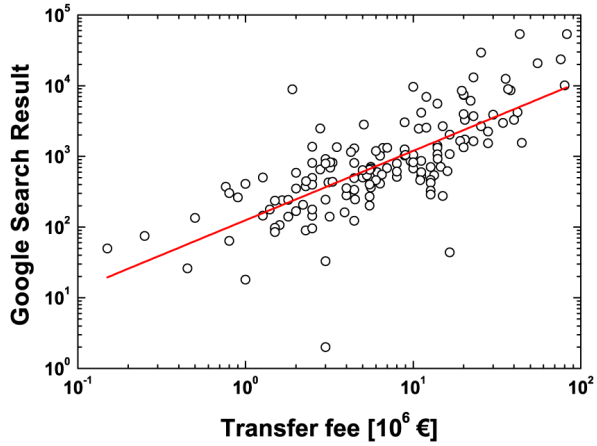

Fig. 1. (Color online) Scatter plot of Google search results and transfer fees for each player. The red line is a linear fit in the log-log scale, and its regression coefficient is 0.98. The Pearson correlation coefficient between two datasets is measured as 0.68.

The spelling of the name of a player should be considered when searching the name. A typical name search is not appropriate for some celebrities, due to their nicknames. For example, not many people mention the full name of Ángel di Maria of Manchester United, with most calling him simply di Maria. In the case of Luis Suárez of FC Barcelona, people generally call him Suarez, as his last name is rather unique. The Google search results actually reflect these phenomena, providing 53,800 search results for Suarez compared with 16,300 for Luis Suárez. In order to correct these dislocations, we searched only for last names for several players who have notably unique last names, and more than 2,000 search results.

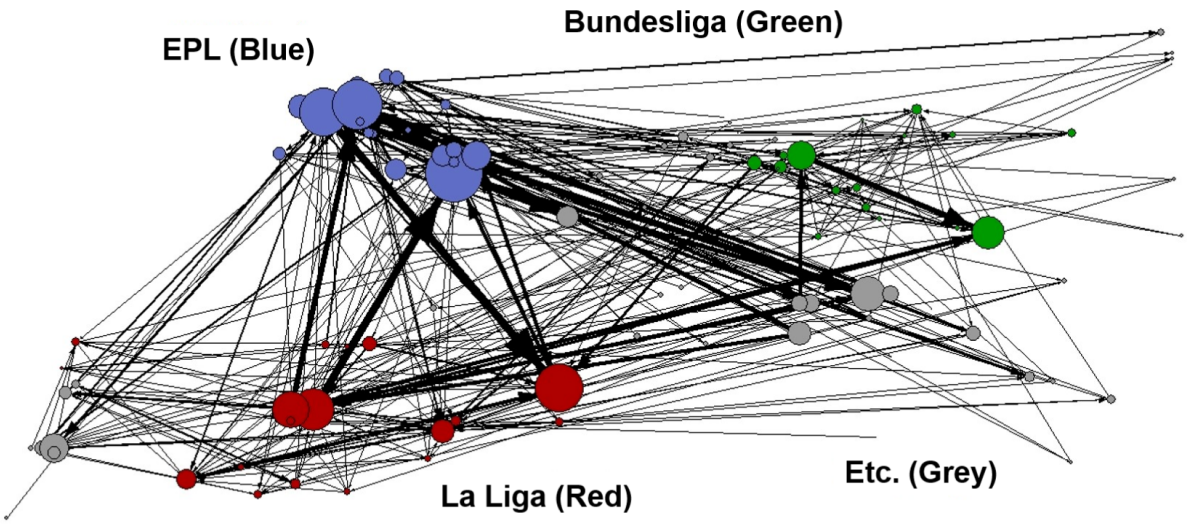

Fig. 2. (Color online) Visualization of European football transfers in the summer of 2014. A total of 436 transfers (arrows) among 114 teams (circles) are represented. The size of a circle and the width of an arrow are proportional to the strength of a team and the weight of a transfer, respectively. The location of a team refers to the geographical position of its hometown. As several teams have the same hometown (*e.g.*, teams in London and Madrid), their locations are slightly shifted in the figure.

The proposed Google-based standard should be validated before being applied to an analysis. Correlation with the exact standard, the transfer fee, can verify the validity of the suggested standard. The monetary units of a transfer fee are Euros and Pounds, and their exchange rate is fixed to that of September 1, 2014 (1.27%/£1) for simplicity. The Pearson correlation coefficient of transfer fees and Google search results is measured as 0.68, which proves the validity of the Google-based standard. The scatter plot of transfer fees and Google search results is presented in Fig. 1. As the regression coefficient close to 1 reflects linear proportionality, the number of Google search results can be directly employed as an alternative to the transfer fee, without modeling their relations.

III. NETWORK CONSTRUCTION AND TRANSFERS BETWEEN LEAGUES

We construct the weighted network by connecting the teams involved in trades. The transfer weight of a player, $w_{ij}(p)$, is defined by the number of search results for the player p transferred from a team i to a team j. The link weight between two teams is equal to the total weight of all traded players between them. The participation of a team in the transfer market is expressed as the node strength. The total strength $s(i)$ of a team i can be obtained by a summation of the in- and out-transfer weights following Eq. (1). Because strength is used as a measure of the total quantity of trades, we treat the network as undirected in the calculation of strength.

$$s(i) = \sum_p \sum_j [w_{ij}(p) + w_{ji}(p)]. \quad (1)$$

Figure 2 succinctly presents how transfers occurred last summer. The EPL teams have a larger size than the others, on average. On the other hand, the La Liga and Bundesliga teams are smaller except for two notably large teams: Real Madrid and FC Barcelona. They participated in the transfers of two well-known players, in James Rodriguez and Luis Suárez. Figure 2 shows that the EPL is more affluent and commercially active than the other leagues. That highest net inflow is for the EPL, shown in Table 1, also supports that claim. These outcomes coincide with the intuition that the EPL is prominently capitalistic.

These results are also consistent with youth promotion and salary data. First, only 31.8% of active EPL players were domestically born players in the 2013/14 season, whereas for the La Liga and Bundesliga the figures were 59% and 50%, respectively [29]. This means

Table 1. Table representing the size of league-to-league transfers. The diagonals show the assortative behavior of each league.

From/To	EPL	La Liga	Bundes-liga	*Etc.*	Total
EPL	99,029	71,637	10,899	85,391	266,956
La Liga	121,140	31,294	13,408	25,780	191,622
Bundes-liga	2,060	23,632	39,375	2,791	67,858
Etc.	85,770	29,113	12,954	0	127,837
Total	307,999	155,676	76,636	113,962	654,273

that the EPL has tried to assemble squads using capital, rather than with youth promotion. A capital-based organization accompanies the recruiting of foreign players, and this affects the decrease of domestically born players. Second, the average salary of EPL players is estimated as £2,273,277 annually, which is far larger than those of La Liga and Bundesliga players, at £1,213,024 and £1,456,565, respectively [30]. The high salaries of EPL players also supports the hypothesis that the EPL is a capital-based league.

IV. WEIGHT AND STRENGTH DISTRIBUTIONS

The probability distributions of players having weight w, and of teams having strength s, are examined as a fundamental analysis for understanding the statistical characteristics. Since these parameters are extracted from web search results, the accumulation mechanism for them can explain the origin of the distributions. The most commonly observed distributions in complex systems are power-law and exponential distributions. A power-law distribution appears in a scale-free network, and an exponential distribution results in a growing network with limitations. The number of citations in publications provides an example of a power-law distribution.

In contrast with citations, the strength distribution of players and the weight distribution of teams shows log-normal behavior, not corresponding to either the power-law or the exponential distribution. Such log-normal behavior occurs in a process (Yule process) that evolves in proportional to its current size, *i.e.*, preferential attachment [31]. A log-normal distribution is obtained as a transient-state solution of preferential attachment, whereas a power-law distribution constitutes a steady-state solution. When an event related to a famous player occurs, the news will be widely reported and reproduced in comparison with the same news for an unknown player. The occurrence of a log-normal distribution due to preferential attachment is considered natural, since reproduced interest can result in positive enhancement of a player's reputation. The search data for the last transfer season can be regarded as a snapshot in the evolutionary process of reputation, compared with the life span of a player. Therefore, these distributions can be understood as the result of evolution through the Yule process [32].

V. RANDOMLY-MIXED PATTERN IN TRANSFERS

Several network properties, including assortativity, strength correlation with neighbors, and betweenness centrality, are evaluated in order to investigate the transfer patterns. Assortativity indicates what portion of all transfers have occurred between the homogeneous groups. Equation (2) provides the definition of the assortativity coefficient, r. In Eq. (2), E_{ij} describes the number of total transfer weights between two leagues i and j as described in Table 1, and $\mathbf{E}$ is the matrix with elements E_{ij}. $\|\mathbf{E}\|$ denotes the summation of all of the components of $\mathbf{E}$ [1,28].

$$r = \frac{\mathrm{Tr}\frac{\mathbf{E}}{\|\mathbf{E}\|} - \|(\frac{\mathbf{E}}{\|\mathbf{E}\|})^2\|}{1 - \|(\frac{\mathbf{E}}{\|\mathbf{E}\|})^2\|}. \tag{2}$$

An assortativity coefficient close to 1 or -1 corresponds to assortative or disassortative mixing, where a coefficient close to zero represents random mixing. The coefficient tells us which transfers dominate the transfer patterns: intra-league transfers or inter-league transfers. The calculated assortativity coefficient of -0.07 describes a randomly-mixed pattern of transfers.

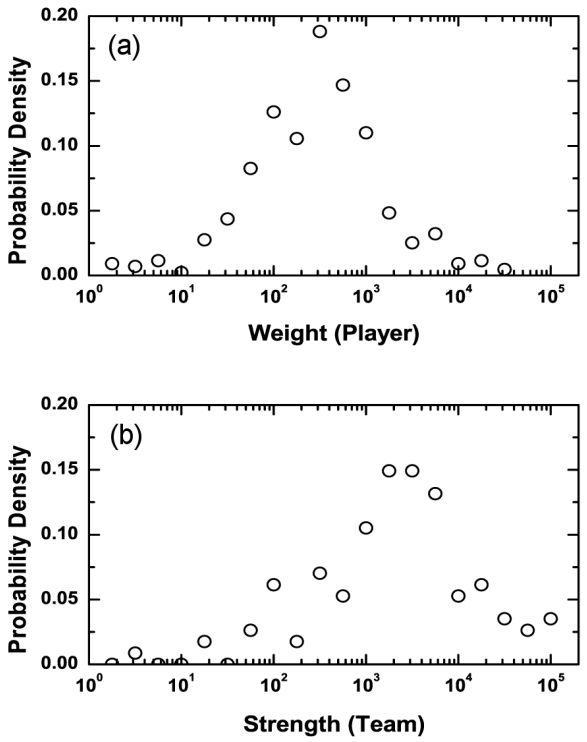

Fig. 3. (a) Weight distribution for transferred players. (b) Strength distribution for participating teams. The probability density is obtained by log-binning in the x-axis range of $10^{0.25}$. The extent to which players and teams have attracted interest on Goal.com webpages is shown in (a) and (b), respectively. Both distributions obviously seem to be log-normal. (a) is clearer, because it is the metadata for (b).

According to popular belief, intra-league transfers are expected to be avoided, as they can strengthen a buyer who is directly competing with the seller. This intuition is shown to be groundless, because the assortativity would come out as -1 in such a situation. The observed randomly-mixed pattern is also exceptional compared with the characteristics of common social networks, and other networks. Common social networks behave assortatively, and others disassortatively [1].

Strength correlation of a node i with its neighbors $nn(i)$ is another effective measure for homophilic clustering of nodes. It can be determined by comparing the node's strength $s(i)$ with the mean strength of the neighbors $\langle s_{nn}(i)\rangle$ as defined in Eq. (3). The comparison shows how nodes are connected with other nodes having similar strength, which describes the homophily of nodes [33]. The set of neighbors $nn(i)$ indicates the trade opponents of a team i, and $k(i)$ is the number of neighbors

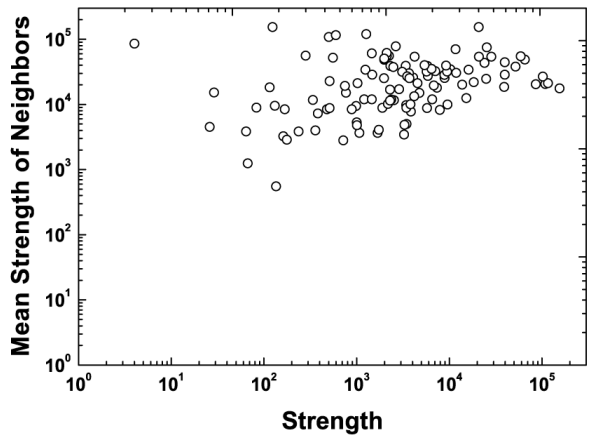

Fig. 4. Scatter plot of strength of a team and the mean strength of neighboring teams. The Pearson correlation coefficient is measured as 0.037, and the trend shows no special correlation.

of the node i.

$$\langle s_{nn}(i)\rangle = \frac{\sum_{nn(i)} s(nn)}{k(i)}. \tag{3}$$

No strength correlation in Fig. 4 implies no clustering of similar strength nodes, which is different from commonly observed clustering in social networks. The lack of homophily in assortativity and strength correlation reflects that the value of players, rather than competition inside a league or the clustering of actively trading teams, dominates the transfers in the three major leagues.

Betweenness centrality is another parameter for estimating the connectivity. The betweenness centrality of a node is proportional to the number of shortest paths from all nodes to all others that pass through that node [35]. Equation (4) defines the betweenness centrality $g(i)$ of a node i, where σ_{ab} is the total number of the shortest paths from a node a to b, and $\sigma_{ab}(i)$ is the number of those paths that pass through i. The shortest paths for all-to-all connections are evaluated using the Floyd-Warshall algorithm and the betweenness centrality is identified by tagging the paths [36]. The calculation of betweenness centrality is conducted for the unweighted network to see only the topological effect.

$$g(i) = \sum_{a\neq b\neq i} \frac{\sigma_{ab}(i)}{\sigma_{ab}}. \tag{4}$$

Betweenness centrality is usually compared with degree or strength, in order to determine whether a topologically central node has many or heavy connections. In

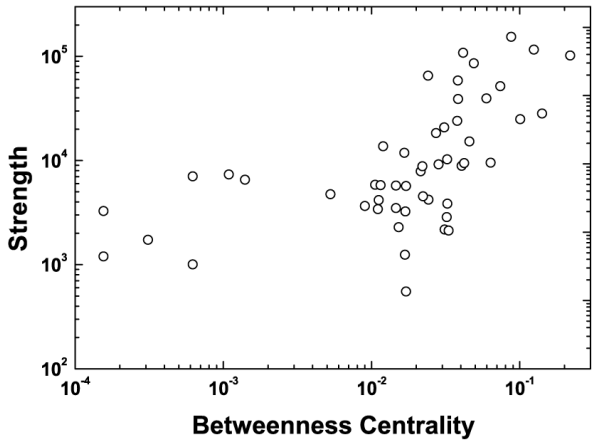

Fig. 5. Scatter plot of betweenness centrality and strength for each team. The teams on no shortest path have a betweenness centrality of 0; they are not depicted in this figure. The Pearson correlation coefficient of strength and betweenness centrality is 0.64.

general, betweenness centrality has a strong positive correlation with degree or strength, because passing through a central node is effective for arriving at other nodes [34]. In the transfer network, the correlation of betweenness centrality and strength can indicate whether a central team of the transfer network is actually a heavy trader.

Figure 5 depicts a positive correlation between strength and betweenness centrality for most teams. Several teams that show a weak correlation are exceptional. These teams trade few valuable players; the strength is not small, but the betweenness centrality is extremely small. In the entire region, the proportional trend between strength and betweenness centrality is observed as being generally in line with other networks [34]. This proportionality means that the teams located at the center of the transfer network are actually the heavy traders, in terms of trading volume.

VI. CONCLUSION

In summary, the European football transfer market in the summer of 2014 has been analyzed as a weighted network, in order to understand the topological and quantitative aspects of the transfer market. We collected 436 transfers that included at least one EPL, La Liga, or Bundesliga team, and these involved 114 teams. We quantified the value of each transfer based on web search results, in order to illuminate the whole structure of the transfers. Intuitions about the capitalistic EPL are explained through the visualization of team-to-team transfers and a table of league-to-league transfers. The resulting log-normal distributions of player weight and team strength are understood by the growth mechanism with preferential attachment.

Transfer patterns and topological characteristics have been investigated through an analysis of network parameters. An assortativity coefficient close zero shows that the transfer market exhibits a randomly-mixed pattern; players are transferred to each league indifferently. This invalidates the intuitive assumption of disassortativity between teams in the same league. An absence of strength correlation between a team and its neighbors implies no clustering of similarly sized nodes. Through a positive correlation between strength and betweenness centrality, we have been able to conclude that the central team in the topology actually behaves as a heavy trader.

In this manuscript, the most significant results are the validation of the valuation method using web search results, and the observation of randomly-mixed transfer patterns. Measuring the value of a person or an abstract object is a very difficult task, because of the extensive variety of variables to be considered. The valuation method using web search results has great advantages because of its simplicity, when the searching range is defined in a way that appropriately rejects linguistic errors and dummies, such as homonyms and spam. The positive correlation of Google search results and transfer fees proved the validity of this method as a reference index. Applications to the other objects is also expected, under a well-defined searching range.

The randomly mixed transfer pattern that we observed is a unique characteristic of the European football transfer network. The teams involved in the three major leagues can be clearly classified into three communities. The assortativity coefficient close to zero implies that transfers within the three major leagues are not affected by what community that the trade opponent belongs to. It appears that the three major leagues can be classified at an equal level, at least in terms of player transfers. An absence of strength correlation between neighboring teams also implies a similar characteristic at the team

scale. This implies that activeness in the transfer market does not have an influence on transfer patterns. A lack of homophily on both team and league scales is a distinctive feature of the European football transfer network, whereas in general, social networks are assortative, and other networks are disassortative.

Our valuation method and network analysis results are expected to contribute to the management of sports teams. The decision maker of a sports team could integrate these analyses into their management strategy. For example, a team can choose either to find famous players through transfer markets, or to grow talented youths. If the records for each team are collected and analyzed annually, then a long-term decision could be made through a network approach.

ACKNOWLEDGEMENTS

We are grateful to all of the members of Complexity in Social System Laboratory and the Department of Physics at POSTECH. This work was supported by Mid-career Researcher Program through the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2013-R1A2A2A04017095) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2010-0021987).

REFERENCES

[1] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[2] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[3] A. L. Barabási, R. Albert and H. Jeong, Physica A **28**, 69 (2000).

[4] R. Albert, H. Jeong and A. L. Barabási, Nature **401**, 130 (1999).

[5] A. Broder, R. Kumar, F. Maghoul, P. Raghavan and S. Rajagopalan *et al.*, Comput. Networks **33**, 309 (2000).

[6] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz and A. Schubert *et al.*, Physica A **311**, 590 (2002).

[7] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).

[8] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabási, Nature **407**, 651 (2000).

[9] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).

[10] L. De Benedictis and L. Tajoli, World Econ. **34**, 1417 (2011).

[11] M. Á. Serrano and M. Boguñá, Phys. Rev. E **68**, 015101(R) (2003).

[12] K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski and S. S. Manna, J. Stat. Mech. **2008**, P02002 (2008).

[13] L. A. N. Amaral, A. Scala, M. Barthélémy and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).

[14] V. Latora and M. Marchiori, Physica A **314**, 109 (2002).

[15] B. Frick, Scot. J. Polit. Econ. **54**, 422 (2007).

[16] E. Amir and G. Livne, J. Bus. Finan. Account. **32**, 549 (2005).

[17] F. Carmichael, D. Forrest and R. Simmons, Bull. Econ. Res. **51**, 0307 (1999).

[18] B. Reilly and R. Witt, Appl. Econ. Lett. **2**, 220 (1995).

[19] E. Feessa and G. Muehlheusser, Eur. Econ. Rev. **47**, 645 (2003).

[20] R. N. Onondy and P. A. de Castro, Phys. Rev. E **70**, 037103 (2004).

[21] K. Kapanova, *Football Transfers Looked From a Social Network Analysis Perspective*, http://www.blankchapters.com. (accessed Mar. 10, 2015).

[22] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer and M. S. Smolinski *et al.*, Nature **457**, 1012 (2009).

[23] S. H. Lee, P.-J. Kim, Y.-Y. Ahn and H. Jeong, PLoS ONE **5**, e11233 (2010).

[24] Category: Football transfers summer 2014, Wikipedia, http://en.wikipedia.org (accessed Oct. 3, 2014).

[25] Transfer Rumours: Premier League gossip and news, The Telegraph, http://www.telegraph.co.uk/sport/football/football-transfers/10826604/Transfer-rumours-Premier-League-gossip-and-news.html (accessed Oct. 3, 2014).

[26] Transfers & Rumors, Transfermarkt.co.uk, http://www.transfermarkt.co.uk/statistik/transferrekorde (accessed Oct. 3, 2014).

[27] List of German football transfers summer 2014, Wikipedia, http://en.wikipedia.org (accessed Oct., 3, 2014).

[28] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).

[29] T. Kim, *[EPL Focus] Illusion of Youth System: Foreign Players in Golden Age.*, http://www.footballist.co.kr (accessed Nov. 23, 2014).

[30] J. Jang, *EPL Average Salary is the Largest: 2.27M £*, http://sports.hankooki.com (accessed Nov. 23, 2014).

[31] M. Mitzenmacher, Int. Math. **1**, 226 (2004).

[32] M. Y. Choi, H. Choi, J. Y. Fortin and J. Choi, Europhys. Lett. **85**, 30009 (2009).

[33] R. P. Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[34] M. Barthélemy, Phys. Rep. **499**, 1 (2011).

[35] L. C. Freeman, Sociometry **40**, 35 (1977).

[36] M. J. Atallah, *Algorithms and Theory of Computation Handbook* (CRC Press, Boca Raton, 1998), Chap. 6.