Networks of international football: communities, evolution and globalization of the game

Yang Li (yanglitjroc@gmail.com)
University of Rochester
Gonzalo Mateos
University of Rochester

Research Article

Keywords: Football network, Community structure, Community detection, Network dynamics, Graph similarity, Temporal states

Posted Date: May 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1674354/v1

RESEARCH

Networks of international football: communities, evolution and globalization of the game

Yang Li* and Gonzalo Mateos

*Correspondence:
yanglitjroc@gmail.com
Department of Electrical and Computer Engineering, University of Rochester, Rochester, 14627, NY, US
Full list of author information is available at the end of the article

Abstract

As the most popular sport around the globe, the game of football has recently intrigued much research interest to explore and distill useful and appealing information from the sport. Network science and graph-centric methods have been previously applied to study the importance of football players and teams. In this paper, for the first time we study the macroscopic evolution of the football society from a complex network point of view. Football game records within a time window of over a century were collected and expressed in a graph format, where participant teams are represented by graph nodes and the games between them are the graph edges. We carry out community detection and temporal analysis to reveal the dynamic features and the community structures embedded within the football network, offering the evidence of a continuously expanding football society. Spatio-temporal analysis is also implemented to unveil the temporal states that represent distinct development stages in the football history. Our analysis suggests that the evolution of the game receives considerable impact not only from major sport events, but also from multiple social and political incidents. The game of football and its evolution reflect significant historical transitions and turning points, and can provide a novel perspective for the study of the worldwide globalization process.

Keywords: Football network; Community structure; Community detection; Network dynamics; Graph similarity; Temporal states

Introduction

The game of football and its global social impact

Football (also known as football or association football) is the most popular sport in the world, attracting billions of fans around the globe that regularly practice it and follow professional competitions. Originated in ancient China and subsequently popularized in England (Wikipedia), football has gradually evolved from a local sport enjoyed by only a few nations in the late 19th century, to a nowadays global sport spreading across the world and involving more than 200 countries.

The establishment of the football's international governing body in 1904, the Fédération Internationale de Football Association (FIFA), marks a milestone for the game of football to become an officially recognized sport, and football games either among clubs or between nations have been organized more systematically since then. FIFA is responsible for the organization of major international football tournaments, notably the FIFA World Cup held for the first time in 1930 and the FIFA Women's World Cup which commenced in 1991. It is hierarchically organized in terms of confederations (that can be closely mapped to continental regions), each

of which comprises national association members. Table 1 presents FIFA's current 6 football confederations, along with the year in which each one was founded and the current number of national association members as of the year 2018. Through confederations, FIFA fosters the development of football within continental regions, and also promotes the worldwide progress of international football beyond geographical boundaries.

Table 1 Football confederations in FIFA (as of 2018)

Abbreviation	Name	Year founded	No. of members
CONMEBOL	Confederación Sudamericana de Fútbol	1916	10
UEFA	Union of European Football Associations	1954	55
AFC	Asian Football Confederation	1954	46
CAF	Confédération Africaine de Football	1957	54
CONCACAF	Confederation of North, Central American and Caribbean Association Football	1961	35
OFC	Oceania Football Confederation	1966	11

From a social science vantage point, football serves as a conduit to connect different countries, different continents and most importantly, people with diverse backgrounds all over the world. Recent estimates show that more than half of the world's population consider themselves as association football fans (Sawe, 2017). For example, over 30 billion people ("accumulated" audience) watched the 2006 FIFA World Cup held in Germany. More than one billion fans tuned in to watch the final of the 2014 FIFA World Cup held in Brazil, which is one of the largest television audience of a single match in all sports. Football transcends the boundaries of sport, and brings together people from different parts of the world. As argued in (Foer, 2004), football is much more than a game, or even a way of life. It's a perfect window into the crosscurrents of today's world, with all its joys and sorrows. Changes in the world may affect the global football landscape. For instance, the 1942 and 1946 FIFA World Cups, which Germany and Brazil sought to host, were canceled due to World War II and its aftermath. On the other hand, football itself may also induce socio-economic and political changes into the world. For example, over 20 million fans flocked to Germany during the 2006 World Cup, bringing not only a sporting success but also an economic, political and security boost (DWsports, 2006).

Despite the growing global significance and increasing social impact of football, scientific research on the game is a recent endeavor, mostly aiming at objectively quantifying and identifying variables explaining the value of certain players, or the performance of specific teams. As a result, the scope of said studies is confined to localized entities such as individual (national) teams, clubs, or leagues. On the other hand, the global macro-structure of the football society remains rather unexplored. We contend that its study, facilitated by network analytic methods applied to contemporary and rich datasets, can offer novel insights into the sport, its global connectivity structure, development and evolution.

Previous work on network analysis of football data

Researchers have recently started to integrate network analytic methods into quantitative analyses of football. On the analysis of football players, in the study from (Onody and de Castro, 2004) of the bipartite network (with football players and clubs

as nodes), the degree distribution decays exponentially. In (Sargent and Bedford, 2013), the presence of highly-rated players is demonstrated to provide the most utility within a simulated team network. Small world property is later verified to exist within the networks of interactions (passes and crosses) between professional football players (Gama et al., 2015). On a team-level analysis, a random walk approach is used in (Ribeiro et al., 2010) to model a football league with an emphasis on predicting individual match scores. Their findings indicate that the dynamics of football tournaments can be accurately simulated using a simple probabilistic model. In (Grund, 2012), the concept of network intensity is introduced, which is defined as the passing rate for a football game based on the analysis of team performance of the English Premier League teams. Results therein show that increases in network intensity lead to improved team performance, while increases in network centralization (the degree to which network positions are unequally distributed in a team) have the opposite effect. More recently, a set of network metrics such as density, heterogeneity and centralization is proposed in (Clemente et al., 2015) for offense analysis of football teams with the potential to aid coach decisions.

The previous works mainly focus on localized micro-structures, such as the importance of certain players or overall team performance. So far, to the best of our knowledge, there have been no prior studies on the macro-structure of the international football landscape at a global scale. As graph-centric methods are introducing improvements in many other disciplines, we expect that their applications in the domain of social network analysis, especially in the scenario of football network analysis, would be feasible and beneficial to reveal the evolution path of the game, to reflect the turning points in football history, and to identify critical development stages of the football society for the past decades.

Aim of this study and contributions

The aim of our study is to perform a network-analytic exploration of the macroscopic structure of the men's international football. To that end, we study for the first time a unique dataset including all the official national[1] team football games, ranging all the way from historical clashes at FIFA World Cup finals to friendly games between islands on the Pacific where the game is only played at the amateur level. Nearly 40000 football game records from 1872 to 2016 were collected to support our study. Network graphs were constructed using the aforementioned data with teams as nodes and games as undirected edges. The networks in our study span a global scale, which means they comprehensively capture all football games between national teams happening across the entire world. Moreover, with access to the date of every football game, we are able to construct graphs for arbitrary time horizons. Such rich and dynamic graph data facilitates the analysis of football networks at different temporal resolutions.

We first investigate the existence of community structure in football networks. We test the strength of weak ties theory (Granovetter, 1995) in football networks and validate the community structure linked by weak connections. Community detection algorithms are then applied to unveil various communities in a static graph

[1]Note that by 'national' or 'nation', we are referring to the countries or regions that are recognized by FIFA as individual participants of football games.

including all games from 1872 to 2016, as well as for dynamic networks spanning 11 decades from 1901 to 2010. The evolution of community structures across decades is quite revealing of the development path and the global expansion of football, as only from data in recent decades that we begin to see good correspondence between communities identified and the confederations in Table 1. Consequently, it is intriguing to explore the landscape of football communities in early decades, and to investigate how connectivity (i.e., football game) patterns evolve into the structure we witness nowadays.

Descriptive network statistics of the time-varying networks are also generated, such as the number of games per year and the number of regional games, e.g. games played by European teams only, in each year. Such temporal features would help reveal the rate and directions of the expansion of football, which can be missing from general analyses on globalized markets and societies. With the aim of temporal states extraction, we advocate a graph similarity measure to group the graphs generated for each year from 1901 to 2010 into clusters. Each cluster represents one individual development stage consisting of several years in the football history, namely the temporal states. By referring to various social events in history, we manage to interpret each temporal state and identify the turning points that mark the boundary between states, further verifying the close relationship between the football landscape and the human history.

All in all, our work is the first to examine the macro-structure of football through its network representations. Through this perspective, we offer novel insights on the endogenous and exogenous factors driving the evolution of (community) structure and the globalization of the game. Our approach could also be translated to other domains where evolving patterns over the network are witnessed, with regard to either graph nodal attributes or graph topological connectivity, such as neuroimaging data, traffic data and internet of things (loT).

Data

In this section, we first briefly introduce the data used in our work. Preliminary exploratory analysis is carried out which reveals the micro-structure within the data and indicates the feasibility of network representation. We then provide the details regarding the construction of the football network, laying the foundation of the downstream graph-centric analysis.

Data collection and preprocessing

The data used in this work contains historical football matches between men's national teams. The football match records were parsed from the World Football Elo Ratings website (Eloratings.net; Lasek et al., 2013), ranging from the first recorded and official football match on November 30, 1872 between Scotland and England, to a friendly game played between Martinique and Panama on April 27, 2016. The data set contains 39052 football match records in total, each of which contains necessary details of a football match, including the two participating teams, match venue, and match score, etc. See Table 2 for an example of a football match record.

Besides matches, we also collect information records of all the involved countries. Altogether 238 countries have participated in the game of football, i.e. each one of

Table 2 Example of a football match record

Date	Home Team	Guest Team	GoalsHome	GoalsGuest	Tournament
2014-06-24	Uruguay	Italy	1	0	World Cup
Venue	Home Ranking*	Guest Ranking*	RankHome**	RankGuest**	ThirdPlace***
Brazil	1893	1831	9	15	True

* 'Ranking' refers to the Elorating ranking index on the match date
** 'Rank' refers to the rank position of a team based on Elorating ranking on the match date
*** 'ThirdPlace' indicates if the game is held on neutral ground

them has played at least one football game in history. Each record contains the name of the country, its geographical coordinates (latitude and longitude), continent and football confederation it belongs to. The geographical coordinates are used to mark each country on the map, and the confederation information is used to validate the clustering of countries via community detection (community structures of the football network). Table 3 shows an example record (of the country Austria).

Table 3 Example record of the country Austria

Name	Latitude	Longitude	Continent	Confederation
Austria	47.52	14.55	Europe	UEFA

The raw data contain all the necessary information about football matches and involved countries, but a few inconsistencies do exist. For example, several countries were split into smaller ones (e.g. collapse of the Soviet Union, East Germany/West Germany, Czechoslovakia, Yugoslavia, etc.). In addition, some countries joined together and participated in football matches as one representative regional team (e.g. Great Britain). In order to maintain data consistency and avoid data redundancy, we locate these anomalies and unify inconsistent data records, for example, assigning the geographical coordinates of England to Great Britain so that it can be located and marked on the map.

Mining frequent football relations among countries

In a football game, two teams play against each other. It is common that these two teams may have played against each other before for multiple times. We call this the frequent football relations. To illustrate this feature and find micro-structures that would add up to construct a complete football network, we applied the Apriori algorithm (Agrawal et al., 1996) to identify the frequent item sets (tuples of teams) in all games from year 1901 to year 2010, i.e. find teams that have played against each other for more than δ times (δ is the threshold, or the so-called minimum support (Agrawal et al., 1996)). We set $\delta = 11$, which is 10% of the total number of years. Table 4 shows some of the frequent relations identified that consist of different number of teams. For each frequent relation, we only list a few examples as illustration.

The largest relation at the threshold of 11 consists of 6 teams. The existence of these frequent relational structures indicates that the whole data set possesses some connectivity patterns. And from the table we can tell that frequent relations mostly exist between countries on the same continent, or countries from the same confederation. This finding suggests that modular structures exist within the football data, thus it is feasible to present the data as a network which could naturally capture the relationship (edges) between teams (nodes).

Table 4 Frequent football relations

Frequent relation	Teams	Occurrence
C_3	England-Scotland-Wales	69
	Denmark-Norway-Sweden	61
	Brazil-Chile-Uruguay	29
	Indonesia-Malaysia-Singapore	21
C_4	England-Scotland-Wales-Northern Ireland	48
	Bahrain-Kuwait-Qatar-Saudi Arabia	16
	Argentina-Brazil-Chile-Uruguay	13
C_5	Bahrain-Kuwait-Qatar-Saudi Arabia-Oman	12
	Chile-Ecuador-Paraguay-Peru-Uruguay	11
C_6	Bahrain-Kuwait-Oman-Qatar-Saudi Arabia-United Arab Emirates	11

* C_n stands for a relation involving n teams

Football network construction

The scientific study of networks, including computer networks, social networks, and biological networks, has received enormous amount of interest in the past decade (Newman, 2010). Networks possess the advantage of transforming complex data into a structural and systematic graph format for presentation and analysis. A network, or graph, is often denoted by $G = \{V, E, W\}$, where V is the set of the nodes and E is the set of edges between the nodes. W is the set of edge weights for weighted graph. For unweighted binary graphs, the edge weights are set to be 1. In this work, the football networks are constructed in the following way.

- A time horizon is specified to delineate the temporal scope of the graph G;
- Nodes in the vertex set V correspond to teams that played at least one game in the prescribed time horizon;
- Undirected edges in E join a pair of nodes in V if the corresponding teams played against each other (at least once); and
- Edge weights in W indicate the number of games played between teams in the prescribed time horizon.

All the constructed networks are undirected, weighted graphs. Fig. 1 shows an example of the football network constructed for the year of 2014. The arcs are the games played, and the endpoints of the edges are the participant countries, marked by their geographical coordinates on the map.

Another way to define the edge weights is to consider the importance of the football match. As indicated by (FIFA), different match type (World Cup, Confederation-level, Friendly, etc.) has different importance. Such importance can be integrated into the edge weights. For example, World Cup matches shall have higher weights than friendly games. In this work, we opt to use the number of games as edge weights, considering the fact that matches with higher importance are fewer in quantity compared with the total number of football matches. While integrating match importance into edge weights might be beneficial, its advantage is not clear to us. Future work shall be devoted to investigate the role of match importance in the construction of football networks.

For an example of the football networks at different timestamps, we plot in Fig. 2 the football networks generated for each World Cup from 1930 to 2014. In each network, edges stand for the games played between participant countries which are located on the map using their geographical coordinates. From Fig. 2, we can clearly

Figure 1 Football network in 2014. The arcs (graph edges) connecting two countries (graph nodes) represent the football games played

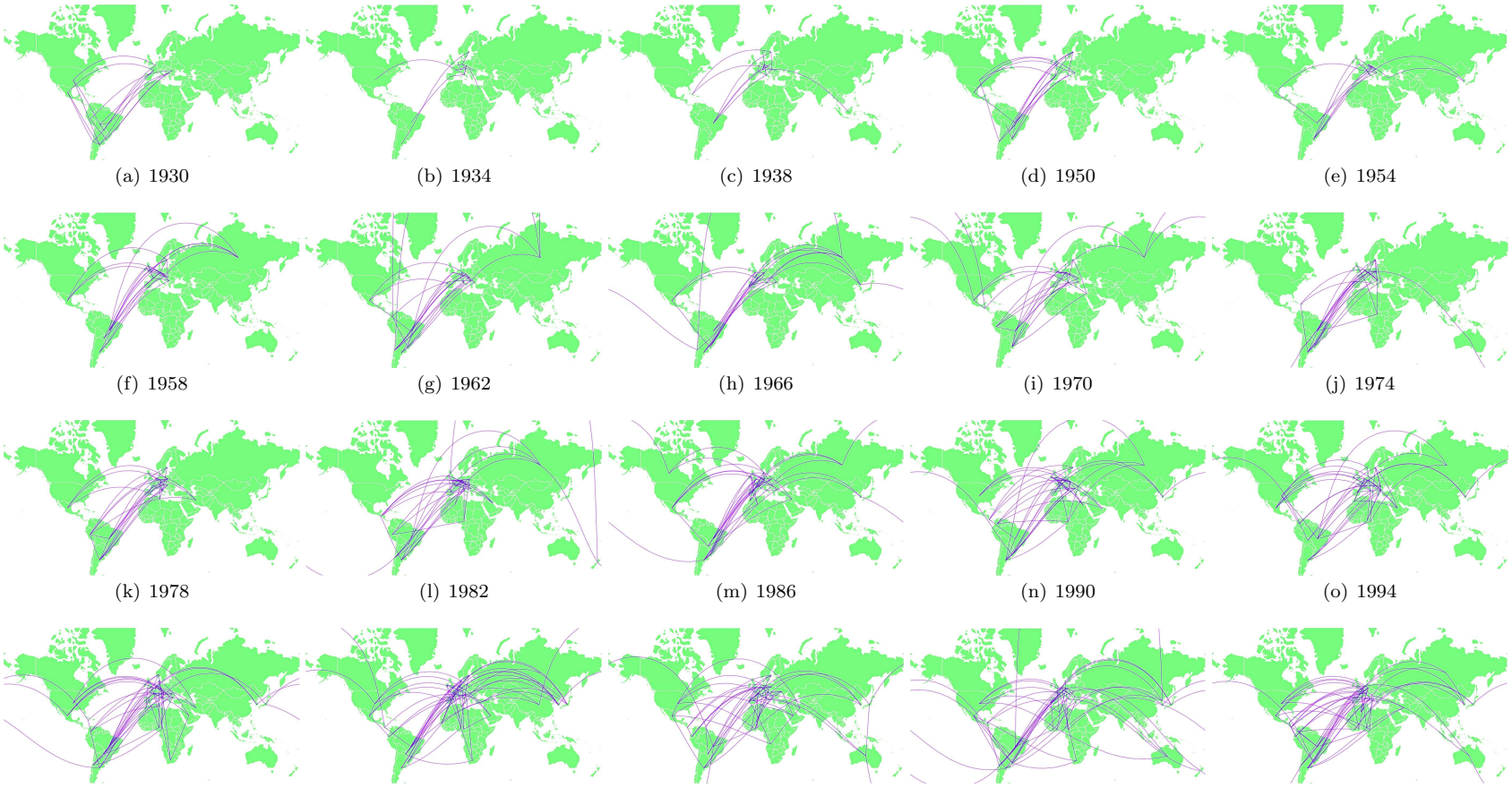
witness the expanding scale of the World Cup with more countries from various continents getting involved, indicating that informative temporal patterns at different timestamps do exist in this data set. These findings motivate us to exploit graph-centric methods to investigate the data, explore the information within the football network of each year, and seek to discover the temporal relationships embedded in the football history represented by a sequential series of football networks.

Analysis and results

In this section, we first briefly review the theoretical background of community detection. Then, the existence of communities within the global football network is verified by checking the extent to which Granovetter's strength of weak ties theory holds in the constructed football graphs. Descriptive graph measures are used to quantify the structural properties of the football networks, which further reveal the dynamics of the network evolution. In the end, we advocate a graph similarity measure that comprehensively integrate various graph properties, further enabling the identification of several temporal states that mark specific development stages of the football history. Via thorough comparison with social history, we manage to determine the great correspondence between the football development stages and significant events occurred in history.

Community detection

One of the most important questions in social network analysis is the identification of "communities", which are loosely defined as collections of individuals who interact unusually frequently (Tantipathananandh et al., 2007). Community detection aims to detect the community structure inside graphs, to identify graph modules and possibly, their hierarchical organization, by using only the information encoded in the graph topology (Fortunato, 2010). Up to now, abundant methods have been proposed for community detection, and most methods can be categorized into traditional methods, modularity-based methods and others; see a thorough review of available algorithms in (Fortunato, 2010).



(a) 1930 (b) 1934 (c) 1938 (d) 1950 (e) 1954

(f) 1958 (g) 1962 (h) 1966 (i) 1970 (j) 1974

(k) 1978 (l) 1982 (m) 1986 (n) 1990 (o) 1994

(p) 1998 (q) 2002 (r) 2006 (s) 2010 (t) 2014

Figure 2 Football network for each World Cup. Each figure depicts the network of a World Cup tournament, with edges representing World Cup matches between participant countries. The first few World Cups mainly involved countries from Europe and South America. With the globalization of international football, more countries, especially Asian and African countries, get involved in the World Cup

The well-known Girvan and Newman method (Girvan and Newman, 2002; Newman and Girvan, 2004) gives a new perspective for community detection by introducing the concept of edge betweenness. Communities are detected by sequentially removing the most important edges in the network. The algorithm also introduces the term of modularity, which serves as a criterion for measuring the quality of the division of networks. The basic idea is to maximize the modularity (Newman, 2006) of the network

$$Q = \frac{1}{2|E|}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2|E|}\right]\frac{s_i s_j + 1}{2}, \tag{1}$$

where $|E|$ is the total number of edges, A_{ij} is the entry of the adjacency matrix on the ith row and jth column that connects node i and j. k_i, k_j are the degree of node i, j respectively, and s_i, s_j are the community assignment for node i and j. When node i and j are in the same community, $s_i s_j = 1$, otherwise $s_i s_j = -1$.

Based on modularity optimization, a whole new set of methods has been proposed. Two advanced approaches were brought up later to speed up the detection process, often referred to as the Fast Newman's algorithm (Newman, 2004) and Louvain algorithm (Blondel et al., 2008). In (Blondel et al., 2008), the algorithm first looks for communities in a local neighborhood of the node. Next, each identified community is aggregated into a new node, adding up to a new network building upon the previous one. Optimize modularity on this secondary network and repeat the steps until a maximum modularity is obtained. This method is among the fastest community detection methods. Consequently, it is implemented in this work for community detection on football networks.

Strength of weak ties in football networks

The natural property of the network structure reflects its capability to bridge the local and the global components. Complex networks often optimize the tie strengths (connection between nodes) to maximize the overall flow in the network (Goh et al., 2001; Maritan et al., 1996). The weak tie hypothesis (Csermely, 2006; Granovetter, 1995) emphasizes the importance of weak ties in connecting communities. Connections with high tie strength are more likely to be structurally-embedded within communities, whereas connections with low tie strength correlate with long-range edges joining communities.

To verify the weak tie hypothesis and identify the intrinsic community structures of the football network, we extract a single graph including all the football games spanning from 1995 to 2015, and use participant teams as nodes and games as edges. In this graph, the numeric tie strength (i.e. edge weight) between two nodes is quantified by the total number of football games played between them. Additionally, follow the definition of the neighborhood overlap of an edge $e_{ij} \in E$ in (Onnela et al., 2007)

$$O_{ij} = \frac{|n(i) \cap n(j)|}{|n(i) \cup n(j)|} = \frac{n_{ij}}{(k_i - 1 + k_j - 1 - n_{ij})}, n(i) := j \in V : (i,j) \in E, \tag{2}$$

where $n(i)$ is the one-hop neighborhood of the node i. n_{ij} is the number of common neighbors shared by node i and node j, and k_i, k_j denote the degrees of node i, j,

respectively. Edges with low overlap are related with two end nodes that do not share many common neighbors, and such edges are more likely to exist between nodes in different communities.

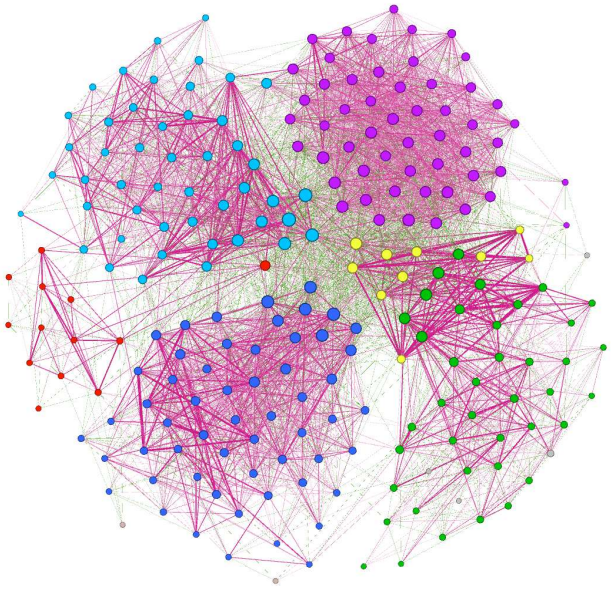
Fig. 3(a) and 3(b) show the network from year 1995 to 2015. The edge colors represent the tie strengths (edge weights) in Fig. 3(a) and edge overlap O_{ij} in Fig. 3(b), respectively. The colors of the nodes correspond to the football confederations they belong to. From the figures, it is clear that edges (in green) with low tie strengths and low overlap are mostly between confederations, while edges (in red) with high tie strengths and high overlap are mostly within confederations. To quantitatively illustrate this property, from all the 5105 edges, edges with the highest 1000 tie strength and edges with the lowest 1000 tie strength are extracted. In each group of 1000 edges, the fraction of edges connecting countries in different confederations is computed. The same procedure is also applied for edge overlap. Table 5 presents the results, which show that edges with high tie strength or high overlap are very unlikely to exist between confederations, while edges with low tie strength or low overlap are more likely to connect countries in different confederations. This result matches the weak tie hypothesis discussed earlier in this section and the visualization in Fig. 3.

Table 5 Fractions of edges between confederations

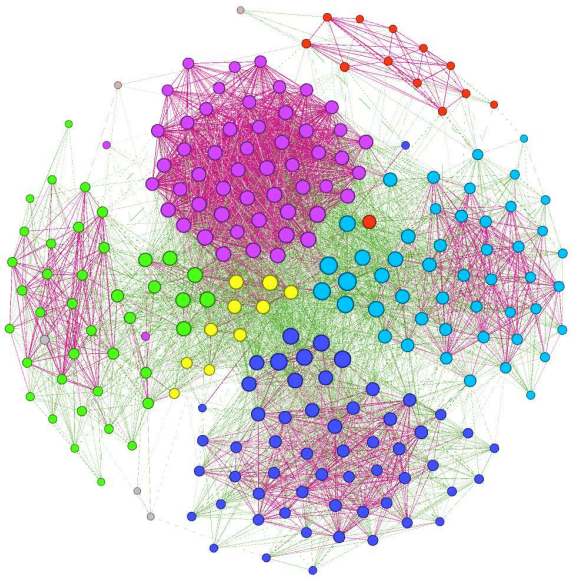
	Highest 1000	Lowest 1000
Tie Strength	2.3%	72.2%
Edge Overlap	6.6%	66.1%

In Fig. 4, we use the tie strength and edge overlap as the (x, y) coordinates and plot all the edges between 1995 and 2015. In Fig. 4(a), edges representing World Cup games are marked as red circles. Most red circles gather in the lower left corner, which indicates that most World Cup games have low tie strength and low overlap. As most World Cup games are played between countries from different continents, this observation again verifies the weak tie hypothesis in the football network where edges with low tie strength and low overlap are most likely between communities. Fig. 4(b) distinguishes games within and between confederations. Red circles that represent games between confederations are located near the origin, validating the existence of weak tie hypothesis in the football network. Fig. 4(b) also provides a visual correspondence to Table 5, showing that most edges with low tie strength and low overlap exist between confederations.

Fig. 3 and Fig. 4 attest the weak tie hypothesis in the football network. As demonstrated in (Granovetter, 1995), most people know about their current jobs from an acquaintance instead of a friend. This fact reveals the role of weak ties in social cohesion and the vital importance of weak ties in message passing within social networks. Similarly, in the football network, edges with low tie strength are between countries with few games played between them. Edges with low overlap indicate that two countries do not share many common neighbors, which means there are few countries that these two countries have both played against. As a result, edges with either low tie strength or low overlap contain vital information about the structure of the football network, and serve as bridges between continents and confederations that contribute to integrate individual football societies into a complete, globalized football network.



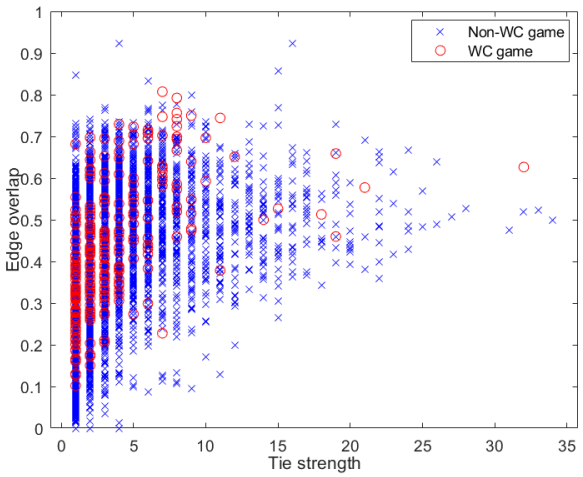
(a) Tie strength



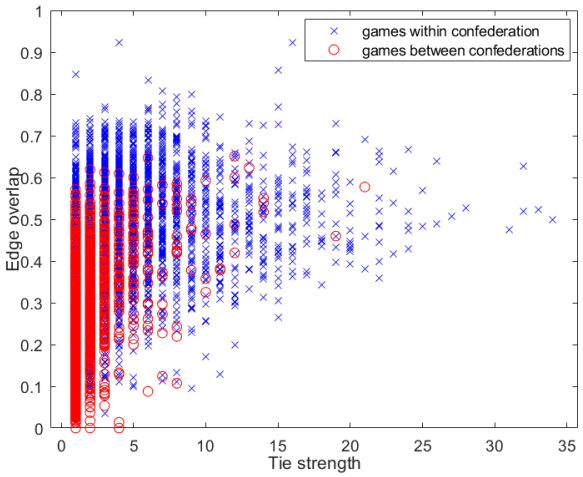
(b) Edge overlap

Figure 3 Football network from 1995 to 2015 visualized via (Bastian et al., 2009). Edges with high tie strengths (total number of football games between two countries) or edge overlap (ratio of common neighbors of the two nodes connected by the edge) are colored in red while edges with low strengths or overlap are depicted in green. Both (a) and (b) show that most green edges are between confederations and red edges are within confederations. Each node color corresponds to one confederation. Purple: UEFA; Blue: CAF; Sky blue: AFC; Red: OFC; Green: CONCACAF; Yellow: CONMEBOL

To better reveal the community structures of the network using the concepts of tie strength and the edge overlap, edge removal was carried out by removing the strongest edge one by one. The relative size of the giant component (a connected



(a) Tie strength vs. Edge overlap

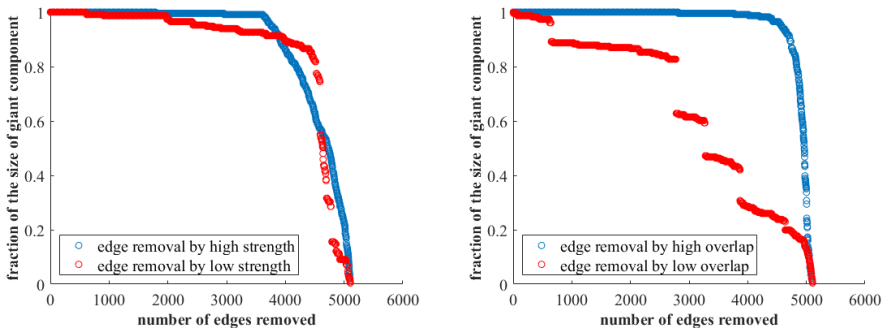


(b) Tie strength vs. Edge overlap

Figure 4 Tie strength vs. Edge overlap. Each point (circle/cross) corresponds to an edge with its tie strength as x-axis value and edge overlap as y-axis value. (a) World Cup (WC) games are marked as red circles and others are marked as blue crosses. Red circles gather in the lower left corner, indicating that most edges of World Cup games have low tie strength and low overlap; (b) Games between countries in different confederations are marked as red circles and games between countries in the same confederation are marked as blue crosses. Red circles gather in the lower left corner, indicating that most edges of games between confederations have low tie strength and low overlap

subset of vertices whose size scales extensively (Newman, 2001)) was computed to check the impact of the removal of each edge. Same procedure is repeated for removal of weakest edges as well. Fig. 5(a) and Fig. 5(b) show the size change of the giant component as edge removal is progressing. From the image, we can see that by progressively removing the edge with either the lowest tie strength or the lowest edge overlap , the size of the giant component shows discontinuity and gaps between points, indicating a sudden disintegration of the network. This means that

removing edges with low strengths or overlap would lead to a breakdown of the original network and the emergence of multiple smaller communities. On the other hand, removing edges with high strengths and overlap gradually shrinks the whole network and does not in fact break the network apart.



(a) Edge removal by tie strength (b) Edge removal by overlap

Figure 5 Change of the size of the giant component with edge removal. In both figures, the blue points correspond to removing the edge with the highest strength or overlap each time. The red points correspond to removing the edge with the lowest strength or overlap each time. The gaps between the red points indicate that removing edges with low strength or overlap will break down the graph and induce smaller communities

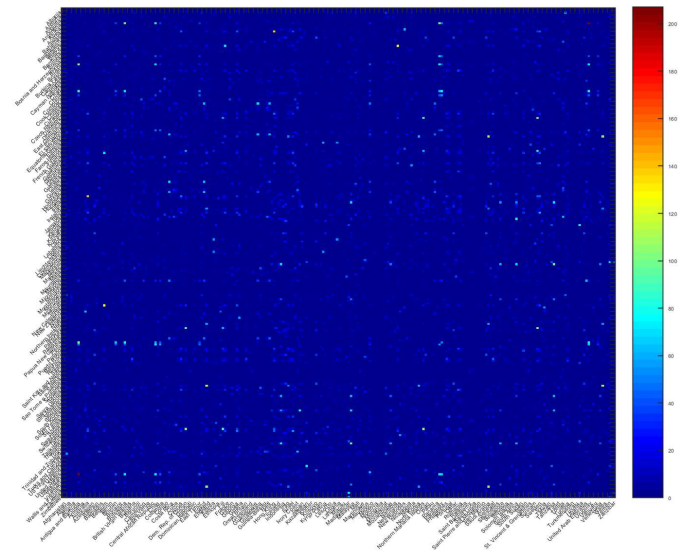
In this section, the weak tie hypothesis has been validated, and the underlying community structures of the football network are revealed by edge removal. The next step is to formally detect these existing communities in the football network via community detection.

Community structure of static football network

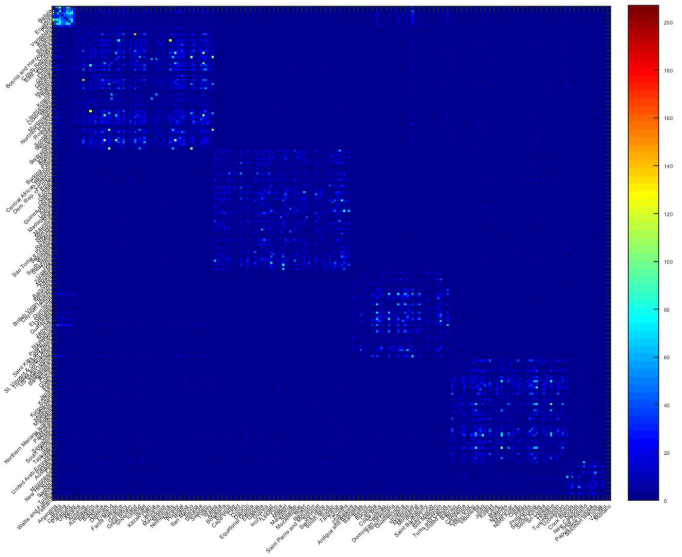
By using the complete data set from year 1872 to year 2016, we constructed a single representative graph of the football network. In this graph, 238 countries involved in football history are included as 238 nodes, and all the 39052 matches are represented by the edges. A 238×238 adjacency matrix shown in Fig. 6(a) was built for this network, whose entries are the number of matches played between two countries.

Applying the Louvain community detection method on this network gives 6 detected communities. We rearrange the rows and columns in Fig. 6(a) based on the community assignment of each node to ensure that nodes of the same community are next to each other in the reformatted adjacency matrix. The new adjacency matrix exhibits a block-wise diagonal format as shown in Fig. 6(b), with each block corresponding to one community. By plotting each community on the world map with different colors in Fig. 7, we see a clear correspondence between the detected communities and the football confederations presented in Table 1.

Although Fig. 7 shows a nice view of the structure of the football network, there exist a few exceptions. For example, Australia joined Asian Football Confederation in 2006 but is still included in the community corresponding to Oceania Football Confederation in Fig. 7. In addition, several countries such as Kazakhstan and Israel left AFC to join UEFA. Such transitions could not be revealed in a single network. Since more and more football matches take place in recent decades, it



(a) Adjacency matrix of the static football network



(b) Adjacency matrix rearranged by node community assignment

Figure 6 Adjacency matrix before and after community detection. (a) the original adjacency matrix of the representative network from 1872 to 2016; (b) the adjacency matrix in which the nodes are rearranged by their community assignments

is reasonable to assume that football networks in earlier years are masked by the crowded networks of recent years where there are much more nodes and edges. The static football network is unable to reveal transitions and changes in the football society. In order to gain a dynamic view of the evolution of the football network, in the next section, we dissect the football history over time and dig into the dynamic properties of the football networks.

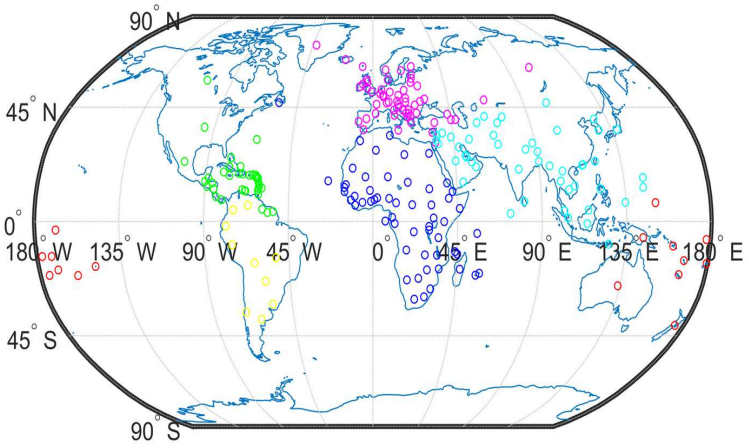


Figure 7 Football community on the world map. Each color corresponds to one community. Based on the geographical locations of the circles on the map, we observe the following relationship between the confederations and the identified communities: purple: UEFA; blue: CAF; sky blue: AFC; red: OFC; green: CONCACAF; yellow: CONMEBOL

Dynamics of the football network

In this section, we focus on the dynamics of football networks with the aim to unveil the evolution and globalization of football society.

Descriptive statistics of dynamic networks

To obtain the temporal dynamics of the football networks, we extract all the football match records from year 1901 to year 2010, and group them into 11 decades to generate one representative football network for each decade. We compute the number of games played either within each football confederation or between confederations per decade. The intention is to distill appealing information, such as which confederations dominate the football world, which confederations experience sudden prosperity or stasis, etc. It would also be interesting to correspond the observations with specific historical events. For example, we would expect to see a significant decrease in the amount of games played in UEFA due to the war, and a sudden spike in the 1950s for CAF as Africa officially enters the football world.

Different football confederations have different development paths. Some entered the football world early while others had a late start. Fig. 8 shows the number of football games played within and between confederations. In Fig. 8(a), it is clear that the dominant confederation is UEFA shown as the red line with the most games played. It suffered a severe drop down in the number of games in the decade of 1941-1950 due to the second World War. CONMEBOL shown as the sky blue diamond, as the first established confederation, does not experience much interruption and shows a steady growth. AFC and CAF do not have many football games until the 1950s, the decade in which both confederations were established. The first Asian Cup and the first African Cup were also held in that decade.

In addition, it is worth noticing that the 1990s witnesses great increases in number of games played in multiple confederations. Such increase makes sense if we look into football history for reference. The 1998 World Cup grew from 24 teams to 32 teams

and allowed more teams from Africa, Asian and North America to participate. This change could significantly increase the eagerness of countries in these areas to join football and also enhance the competition. More friendly games and qualification games would be played within confederations.

Fig. 8(b) shows the interaction between confederations. The communication between confederations is basically growing as more games are played in recent decades. Exception such as the number of games between AFC and OFC, shown as the sky blue line with diamonds, can be explained as Australia left OFC and joined AFC in 2006. Consequently, the original between-confederation edges between Asian countries and Australia now belong to the within-confederation edges of AFC. The line with the blue star shows a significant increase of football games between AFC and UEFA starting from the 1990s. This change is strongly related to the ambition of Asian countries in developing their football and the growing economy of Asia where money are spent to invite European teams for friendly games.

Other graph measures can also be useful to capture the dynamics of networks. In this work, we explore the measure of global efficiency (Latora and Marchiori, 2001) which is the average of inverse shortest path length. The average efficiency of a network G is defined as:

$$E(G) = \frac{1}{n(n-1)} \sum_{i<j \in G} \frac{1}{d_{ij}},$$

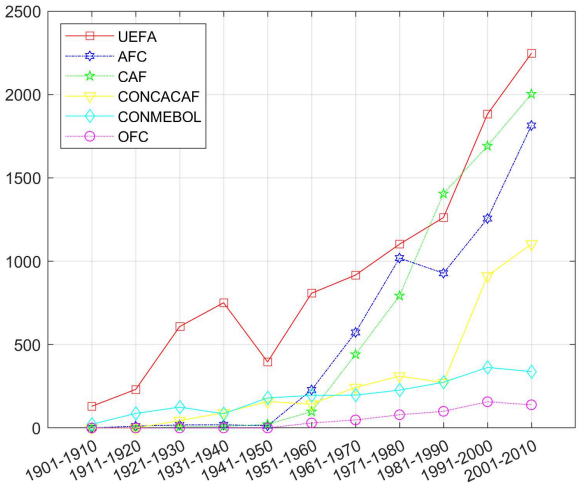
where n denotes the total number of nodes and d_{ij} denotes the length of the shortest path between node i and node j. The global efficiency is defined as:

$$E_{global}(G) = \frac{E(G)}{E(G^{ideal})}$$

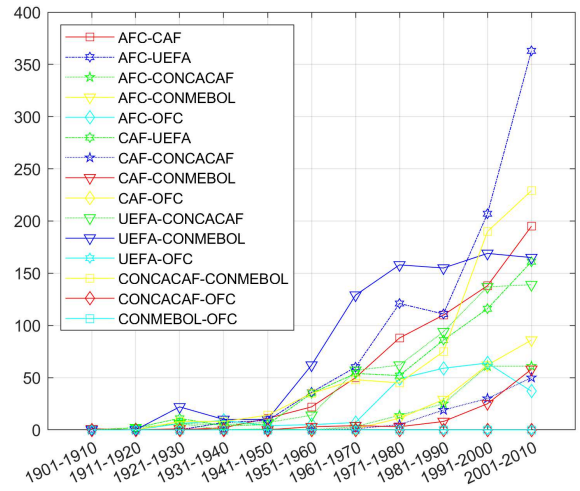
where G^{ideal} is the graph with n nodes and all possible edges are present. Global efficiency serves as a quantitative measure of the average distance it takes for a node to reach another node. Networks with high global efficiency should have more edges thus connections between nodes are efficient.

Fig. 9 shows the computed global efficiency for the football networks constructed for each decade. An obvious valley appears in the 1940s when the WWII broke out. This finding matches the results in Fig. 8. During the wartime, there were fewer football games thus most connections between the nodes were cut off, and the global efficiency of the network is severely compromised.

So far, we have looked at several descriptive statistics of the dynamic in the football networks. In Fig. 7, we observe the correspondence between football communities and real-world football confederations. However, football networks do not always appear in such form as different confederation were established at different times in history. In addition, it is practical to assume that communities in early decades are more localized in certain regions, while in recent decades communities are much more spread out and nodes in the same community may have huge geographical distance in between. Thus, it is reasonable to assume that football networks maintain different community structures in different decades.



(a) Number of games within confederations



(b) Number of games between confederations

Figure 8 Number of edges within/between confederations per decade. (a) total number of football games played within each confederation for each decade; (b) total number of football games played between any two confederations for each decade

Community structure of dynamic networks

To reveal the dynamics of the community structure of football networks, we applied the Louvain community detection algorithm on the 11 networks for the 11 decades from year 1901 to year 2010. Fig. 10 and Fig. 11 give the adjacency matrices and visualization of the networks for 4 example decades. Both figures show a clear trend of the community evolution of the football network. Starting from early decades, the communities are quite local, and the correspondence between the identified communities and confederations was not clear. As more countries join the football society, the communities start to grow with more nodes. The boundaries between communities also become more apparent as shown in Fig. 10(c) and Fig. 11(a). In these

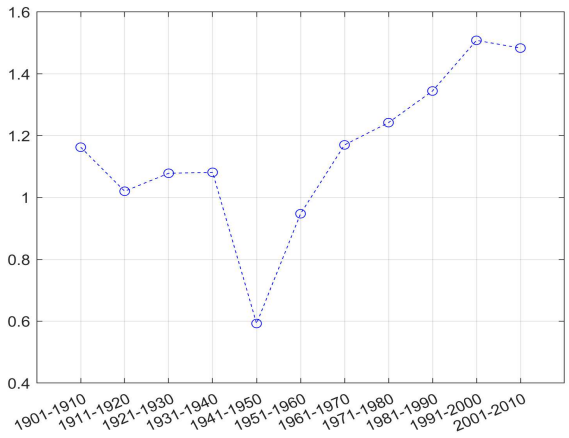


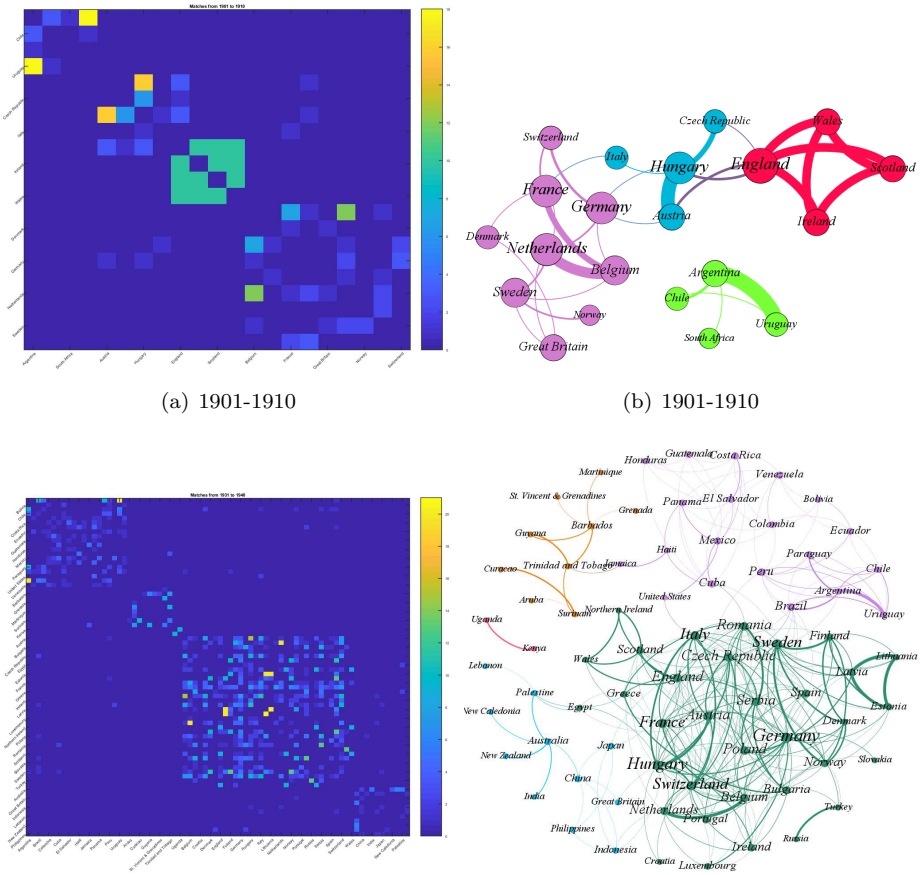
Figure 9 Global efficiency of the football networks per decade. An ascending trend is witnessed, with the exceptions due to the world wars

decades, the communities are slowly transforming their structures, sharing more and more similarity with the actual football confederations. Fig. 11(c) and 11(d) show the football network of the last decade, in which the community assignment for each node is basically the same with the actual confederation affiliation of each country. From these figures, a clear evolution path of the football network is unveiled, and the temporal features beneath such evolving community structures definitely worth more in-depth research.

Temporal states extraction

In the previous section, we brought up the assumption that different stages exist during the evolution of football. It would be helpful to identify individual states in football history which represent different evolution stages. As stated in multiple literatures (Koutra et al., 2013; Papadimitriou et al., 2010; Zager and Verghese, 2008), graphs belonging to the same state or same cluster shall exhibit high similarity. As a result, a reasonable way to find temporal states in football history is to first calculate the similarity between football networks per year, and gather the graphs with high similarities into one group.

Multiple literatures (Bunke, 2000; Bunke et al., 2007; Macindoe and Richards, 2010; Wilson and Zhu, 2008) have discussed the calculation of graph similarity. However, each method only consider one aspect of the graph features and may lose other useful information. In (Li et al., 2011), the authors included 20 features regarding graph properties to generate a feature vector per graph. After normalizing the feature vectors, they fed them into a SVM for graph classification. The mean of the node degrees and the mean of node clustering coefficients are combined with other global measures such as the global efficiency in the final feature vectors. This procedure ignores the node correspondence between graphs, and may cause information loss regarding the different degree distribution between graphs. Also, combining all the measures into a singe vector makes it vague to determine the contribution from each measure. In order to preserve the node correspondence and take advantage of the graph-level measures, we bring up the following framework



(a) 1901-1910 (b) 1901-1910

(c) 1931-1940 (d) 1931-1940

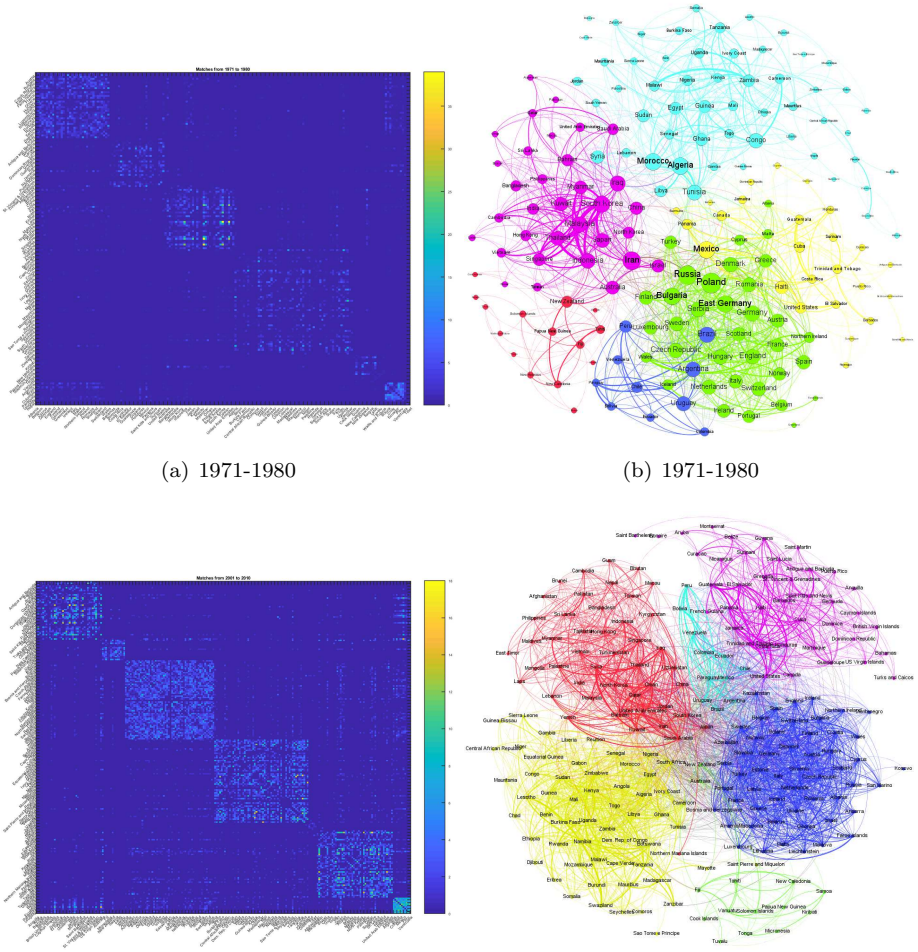
Figure 10 Adjacency matrices and visualization for decade 1901-1910 and 1931-1940. Left: adjacency matrices. Right: network visualization. In (b) and (d), each color stands for one community. In early decades, boundaries between communities in the adjacency matrices as shown in (a) and (c) were not clear, and the correspondence between community assignment and confederation as shown in (b) and (d) was poor

to combine different categories of graph measures for the computation of graph similarities. We continue to use the match records from year 1901 to year 2010, and construct 110 football networks in total, one for each year. The graph measures are categorized as the following 3 types.

(1) Node-level: degree, clustering coefficient (Watts and Strogatz, 1998), closeness (Freeman, 1978), local efficiency (Latora and Marchiori, 2001).

- For each graph with N nodes, compute each of the above measures and generate a $N \times 1$ vector for each node-level measure;
- Concatenate the vectors for all the 110 networks into a $110 \times N$ matrix;
- Calculate the correlation coefficient between matrix rows;
- Transform the correlation coefficient c_{ij} into similarity measure to constrain the value into [0,1];

$$s_{ij} = \frac{c_{ij} + 1}{2}, i, j = 1, ..., 110$$



(a) 1971-1980 (b) 1971-1980

(c) 2001-2010 (d) 2001-2010

Figure 11 Adjacency matrices and visualization for decade 1971-1980 and 2001-2010. Left: adjacency matrices. Right: network visualization. Networks get more crowded and complex with more nodes and edges in recent decades, and the community structure is more similar with the modern football landscape represented by the six confederations

- Generate a 110×110 similarity matrix for each node-level measure.

(2) Graph-level: number of nodes, number of edges, average path length, global efficiency (Latora and Marchiori, 2001), diameter, radius, graph energy (Gutman, 1978), link density (Black, 2008) and transitivity (Newman, 2003).

- Calculate the above measures and construct a 9×1 feature vector per graph;
- For all the 110 graphs, construct a 110×9 feature matrix;
- Normalize the columns with z-normalization;
- Generate a 110×110 similarity matrix based on the similarity (defined above) between rows.

(3) Structure-level: vertex-edge-overlap (VEO) (Papadimitriou et al., 2010). The vertex-edge-overlap similarity is defined as

$$sim_{VEO}(G_1, G_2) = 2 \times \frac{E_1 \cap E_2 + V_1 \cap V_2}{E_1 + E_2 + V_1 + V_2}$$

where $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$. The vertex-edge-overlap measures the structural similarity between graphs. The VEO is computed between every pair of graphs of the 110 graphs, and a 110×110 similarity matrix is generated.

After the three similarity matrices are obtained, we further calculate the average of them as the final similarity matrix. Fig. 12 presents the exact pipeline for the above procedure to identify temporal states in football history. This framework takes into consideration the node-level graph measures to preserve the node correspondence between graphs, the global graph measures to consider overall graph properties, and the structural properties (nodes and edges) to quantify the topological similarity between graphs. In Fig. 12, a clear dissection of the original complete network in the top left corner is presented. Three levels of similarity calculation are carried out and the results shall be discussed soon in later sections.

Fig. 13 presents the graph similarity matrices obtained based on node-level graph measures (degree, closeness, clustering coefficient, local efficiency). In Fig. 13, we can see a clear boundary before and after the 1940s due to the second World War, and along the diagonal several blocks could be identified leading to potential temporal states. Fig. 14(a) shows the graph similarity matrix based on global measures. From the image, we can roughly identify two clusters with a blurry boundary around 1950 to 1960. These years could be identified as a transition stage from earlier stage when the football society just started to grow, to current modern stage with established confederations. Fig. 14(b) shows the vertex-edge-overlap similarity matrix, which exhibits a similar pattern with Fig. 13.

Fig. 15 presents the ultimate graph similarity matrix as the average of all the similarity matrices obtained above. This matrix combines all three levels of similarity, taking features of the network from all aspects into account. We can see clear partitions of the 110 years in football history. If we go along the diagonal line, several individual states could be visually identified. We further carry out a community detection procedure on the similarity matrix of Fig. 15, and partition all the 110 years into community of years, enabling the following temporal states to emerge.

- *1901-1908*: Start of the history
  In early years, only a few countries were playing football, and they played mostly against each other. Football networks in these years share high similarity (see the bright red block on the upper left corner) as they consists of similar nodes and roughly identical edges.
- *1909-1913*: Embryonic form of globalization
  In 1908, the first official football tournament was played at the Summer Olympics in London. Most participant countries were from Europe, yet it was the first time that football appeared as an international sport.
- *1914-1918*: WWI
  The first World War broke out in July 1914 and ended in November 1918. During the war, football in Europe was severely impacted. Fewer international football games were played and some football players even joined the army at that time. Meanwhile, other areas such as South America was less affected. CONMEBOL was founded at this time, and the first Copa América was held in 1916. This explains the high similarity between football networks in these years as shown in Fig. 15.

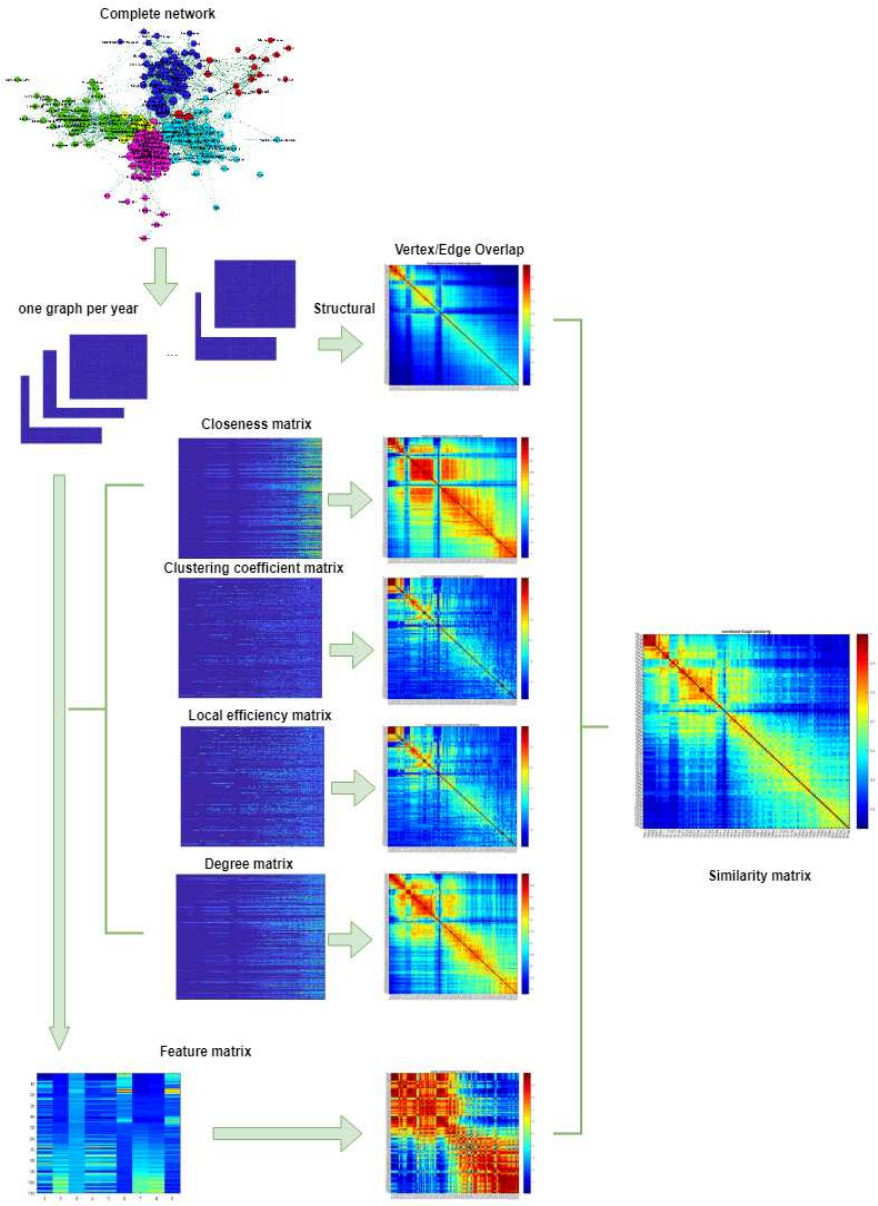
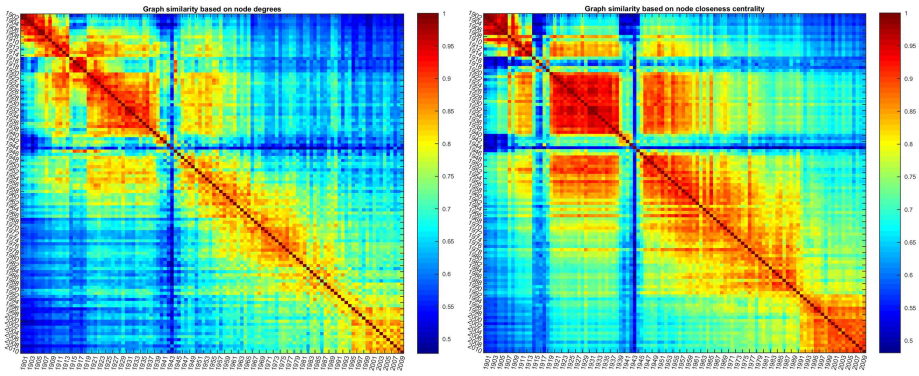
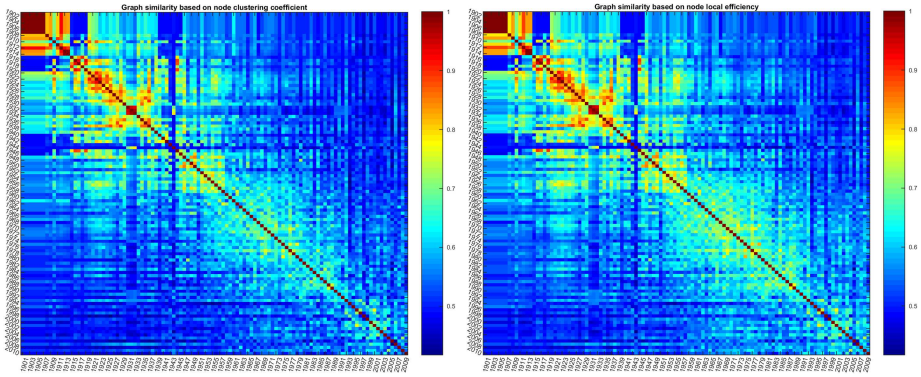


Figure 12 Pipeline for identifying temporal states in football history. 110 networks are generated for each year from 1901 to 2010. Node-level, graph-level, structure-level similarity matrices are computed and the final similarity matrix is given on the right

- *1919-1938*: Rebuilding
  After the WWI, peace again returned Europe and football, as the most popular sport there, got prosperous one more time in Europe. Football networks in this stage have high similarities, showing a steady and healthy growth of the game of football. In 1930, Uruguay held the first World Cup with 13 teams. Another 2 World Cups were held in 1934 and 1938 in Italy and France, respectively. Football in this stage showed an increase in the total number of games played, total number of nations participated, and the diversity of the participant nations. Although most nations playing football were from Eu-

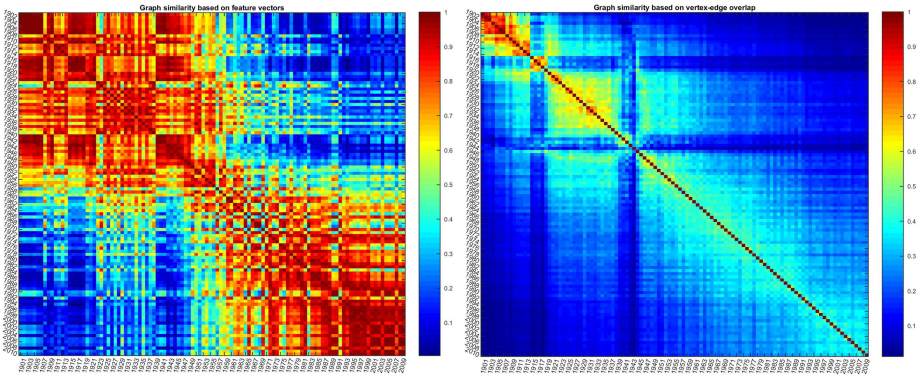


(a) degree (b) closeness



(c) clustering coefficient (d) local efficiency

Figure 13 Graph similarity matrices based on node-level graph measures. Similarity matrix based on (a) node degree (b) node closeness (c) node clustering coefficient (d) node local efficiency



(a) global measures (b) vertex-edge-overlap

Figure 14 Graph similarity matrices via global graph measures (a) and structural similarity (b)

rope and South America, several Asian and African countries, such as China, Egypt, Palestine, also joined the football world in this period.

- *1939-1945*: WWII
  The World War II, from 1939 to 1945, was a disaster to the whole world. In

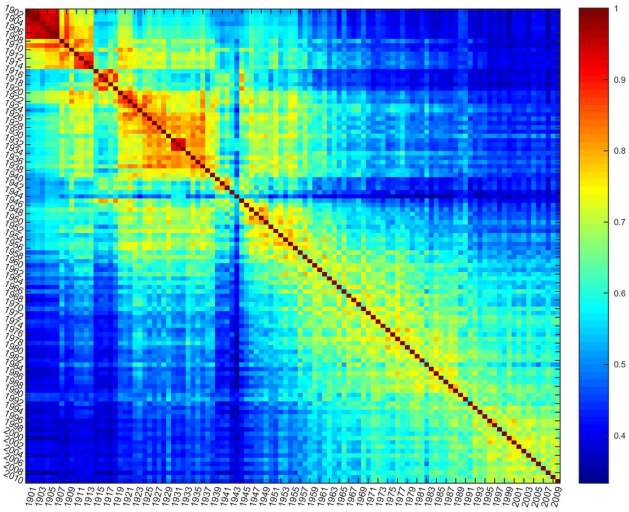


Figure 15 Ultimate graph similarity matrix. The matrix is the average of matrices shown in Fig. 13 and Fig. 14. The matrix indicates that several states existing in football history.

Fig. 15, we can see a blue cross in this time period as the football networks in these years are not at all similar with the rest of the networks. World Cup was forced to stop due to the war. Countries focused merely on how to survive instead of playing football. However, football was not completely forgotten, as in some neutral nations football was still quite popular. Moreover, football was very popular among soldiers and even inside prisoner-of-war camps.

- *1946-1959*: Post-war recovery
  After the second World War, everything began to recover, including football as well. In 1950, Brazil held the 4th World Cup which had been cut off for 12 years. The whole football world started to heal itself. In this period, besides the international football events such as the World Cup and the Olympics, regional football also embraced a fast growth, including
  - UEFA and AFC founded in 1954
  - The first Asian Cup was held in 1956, won by South Korea
  - CAF is founded and the first African Cup of Nations was held in Sudan

  It is also interesting that graphs in this stage share much similarity with the graphs in the stage before the second World War. This exactly shows the attempt of the football society to get recovered to the status as the one before war.
- *1960-1990*: Prospering
  In the previous stage, most regional football confederations were established with the exception of CONCACAF which was founded in 1961. The structure of the modern football world was more or less built up. In this stage, the football world experienced a peaceful growing stage for 30 years. There were not much significant events happening to change the overall landscape. In Fig. 15, we can see a large block from year 1960 to 1990, with a more or less uniform similarity between the networks.

- *1991-2010*: The tremendous change
  In 1991, the Soviet Union dissolved. This historical and political event had its own impact on the football world. In Fig. 15, a new block appears starting from year 1991, indicating that from this moment the network enters a new stage and has less similarity with previous networks. The changes in the new networks are immense as the Soviet Union, which originally was a single and important node in the network, is dissolved into multiple new nodes. With these newly emerged nodes, more football games have been played thus the edges are also significantly influenced. We investigate the political history and generate Table 6. The nations in the table were either emerged as new nations after the cold war, or got independent from Soviet Union. The table also includes the first football game played by new countries, and the last game before cold war and first game after cold war for other countries. 20 new nodes emerged or re-emerged after Soviet Union dissolved. This political change brings significant alternation of the overall structure of the football network, especially due to the fact that most of the new countries joined UEFA, the confederation with the most impact to the whole football world.

Table 6 Network nodes emerged after cold war

Country	Before cold war	After cold war
Slovenia	-	1991
Moldova	-	1991
Ukraine	-	1992
Belarus	-	1992
Uzbekistan	-	1992
Kyrgyzstan	-	1992
Turkmenistan	-	1992
Tajikistan	-	1992
Kazakhstan	-	1992
Bosnia and Herzegovina	-	1993
Macedonia	-	1993
Kosovo	-	1993
Croatia	1956	1990
Lithuania	1940	1990
Georgia	1935	1991
Latvia	1942	1991
Estonia	1942	1991
Armenia	1935	1992
Azerbaijan	1935	1992
Slovakia	1944	1992

The above analysis lists several stages in the football history from year 1901 to year 2010. Over a century of football history is partitioned into 8 states. The interesting fact about these stages is that, although people would assume that the changes of the football networks would be mainly related with the changes within the football world, such as the expansion of World Cup or the commencement of new football tournaments, it is also significantly related with historical and political events, such as wars and political incidents. This shows a perfect evidence of the social impact of football and its ability of offering an insight into the changes in the world. Changes in the world would affect football and football on the other hand, could reflect changes in the world.

Discussions and Conclusions

This paper analyzes the evolution of football society from a macroscopic point of view. The focus is on the complete football history instead of certain teams, players

or leagues as in previous works. Network science disciplines are applied to mine the temporal dynamics and community structures of football networks. Our findings show the existence of community structures in football society, and the identified temporal dynamics demonstrate the continuous growth of the football society with correspondence to globalization process.

Community detection method was later applied on the networks to show the expanding size of communities in each decade. Football communities formed in early years with only a few countries close in geographical distance, and then developed into large ones with nearly all the countries in each continent. The scales of communities are also expanding from local regions to a whole continent. The convergence of the communities and the confederations were also significantly improved from early decades to recent decades.

In this paper we proposed a framework for calculating graph similarity for graph clustering. The framework integrates multiple graph metrics including three levels (node-level, graph-level and structure-level) and considers graph properties from all aspects. Based on the graph similarity matrix, 8 temporal states, each representing one distinct development stage in football history, were found. The temporal states possess great correspondence with both the changes in football society and social events in the world history.

For the first time, our research analyzes the big picture of the football society and offers a new perspective of the research on football. The method and the framework used in this paper can also be applied on the network analysis on data from other domains. Future research may target at continuous analysis of football data by including more football game records. It would also be a promising aspect to analyze the football network structure using the spectral graph theory and graph signal processing techniques.

Declarations

Availability of data and materials

The datasets supporting the conclusions of this article are available on Eloratings.net. More information of the data can be found at http://www.eloratings.net/. The datasets can be accessed via https://github.com/YangLiyli131/footballNetworkData.

Competing interests

The authors declare that they have no competing interests.

Funding

Work in this paper was supported by the NSF award CCF-1750428.

Author's contributions

YL processed and analyzed the data and conducted the research scheme in this work. YL was responsible for generating the plots and tables in this work. GM provided supervision of the whole research plan and contributed to the polishing of this work. All authors read and approved the final manuscript.

Authors' information

YL received a B.Sc. degree in Electrical Engineering (Automation) from Tianjin University, Tianjin, China, in 2013, and a M.Sc. degree in Electrical and Computer Engineering from the University of Rochester in 2016. He joined the Ph.D. program at the University of Rochester in January 2017.

GM is an Associate Professor with the Dept. of Electrical and Computer Engineering, University of Rochester as well as an Asaro Biggar Family Fellow in Data Science. He is also affiliated with the Goergen Institute for Data Science.

Acknowledgements

Not applicable.

References

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I., *et al.*: Fast discovery of association rules. Advances in knowledge discovery and data mining **12**(1), 307–328 (1996)

Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 3, pp. 361–362 (2009)

Black, P.E.: Sparse graph. Dictionary of Algorithms and Data Structures, National Institute of Standards and Technology (NIST) (2008)

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), 10008 (2008)

Bunke, H.: Graph matching: Theoretical foundations, algorithms, and applications. In: Proc. Vision Interface, vol. 2000, pp. 82–88 (2000)

Bunke, H., Dickinson, P.J., Kraetzl, M., Wallis, W.D.: A Graph-theoretic Approach to Enterprise Network Dynamics vol. 24. Springer, Berlin (2007)

Clemente, F.M., Couceiro, M.S., Martins, F.M.L., Mendes, R.S.: Using network metrics in soccer: a macro-analysis. Journal of human kinetics **45**(1), 123–134 (2015)

Csermely, P.: Weak links: Stabilizers of complex systems from proteins to social networks. Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks, by Peter Csermely. 2006 XX, 408 p. 37 illus. 3-540-31151-3. Berlin: Springer, 2006., 37 (2006)

DWsports: Germany's World Cup Report Hails Economic, Social Success (2006). http://www.dw.com/en/germanys-world-cup-report-hails-economic-social-success/a-2263053 Accessed 2 Oct 2018

Eloratings.net: Eloratings. http://www.eloratings.net/ Accessed 27 Apr 2016

FIFA: Men's Ranking Procedure. https://www.fifa.com/fifa-world-ranking/procedure-men Accessed 11 Jun 2021

Foer, F.: How Soccer Explains the World. HarperCollins Publishers, New York (2004)

Fortunato, S.: Community detection in graphs. Physics reports **486**(3-5), 75–174 (2010)

Freeman, L.C.: Centrality in social networks conceptual clarification. Social networks **1**(3), 215–239 (1978)

Gama, J., Couceiro, M., Dias, G., Vaz, V.: Small-world networks in professional football: conceptual model and data. European Journal of Human Movement **35**, 85–113 (2015)

Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the national academy of sciences **99**(12), 7821–7826 (2002)

Goh, K.-I., Kahng, B., Kim, D.: Universal behavior of load distribution in scale-free networks. Physical Review Letters **87**(27), 278701 (2001)

Granovetter, M.: Getting a Job: A Study of Contacts and Careers. University of Chicago press, Chicago (1995)

Grund, T.U.: Network structure and team performance: The case of english premier league soccer teams. Social Networks **34**(4), 682–690 (2012)

Gutman, I.: The energy of a graph. 10. steiermrkisches mathematisches symposium (stift rein, graz, 1978). Ber. Math.-Statist. Sekt. Forsch. Graz (100-105) (1978)

Koutra, D., Vogelstein, J.T., Faloutsos, C.: Deltacon: A principled massive-graph similarity function. In: Proceedings of the 2013 SIAM International Conference on Data Mining, pp. 162–170 (2013)

Lasek, J., Szlávik, Z., Bhulai, S.: The predictive power of ranking systems in association football. International Journal of Applied Pattern Recognition **1**(1), 27–46 (2013)

Latora, V., Marchiori, M.: Efficient behavior of small-world networks. Physical review letters **87**(19), 198701 (2001)

Li, G., Semerci, M., Yener, B., Zaki, M.J.: Graph classification via topological and label attributes. In: Proceedings of the 9th International Workshop on Mining and Learning with Graphs (MLG), San Diego, USA, vol. 2 (2011)

Macindoe, O., Richards, W.: Graph comparison using fine structure analysis. In: Social Computing (SocialCom), 2010 IEEE Second International Conference On, pp. 193–200 (2010)

Maritan, A., Colaiori, F., Flammini, A., Cieplak, M., Banavar, J.R.: Universality classes of optimal channel networks. Science **272**(5264), 984–986 (1996)

Newman, M.: Networks: an Introduction. Oxford university press, Oxford (2010)

Newman, M.E.: Scientific collaboration networks. i. network construction and fundamental results. Physical review E **64**(1), 016131 (2001)

Newman, M.E.: Ego-centered networks and the ripple effect. Social Networks **25**(1), 83–95 (2003)

Newman, M.E.: Fast algorithm for detecting community structure in networks. Physical review E **69**(6), 066133 (2004)

Newman, M.E.: Modularity and community structure in networks. Proceedings of the national academy of sciences **103**(23), 8577–8582 (2006)

Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. Physical review E **69**(2), 026113 (2004)

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.-L.: Structure and tie strengths in mobile communication networks. Proceedings of the national academy of sciences **104**(18), 7332–7336 (2007)

Onody, R.N., de Castro, P.A.: Complex network study of brazilian soccer players. Physical Review E **70**(3), 037103 (2004)

Papadimitriou, P., Dasdan, A., Garcia-Molina, H.: Web graph similarity for anomaly detection. Journal of Internet Services and Applications **1**(1), 19–30 (2010)

Ribeiro, H., Mendes, R., Malacarne, L., Picoli, S., Santoro, P.: Dynamics of tournaments: the soccer case. The European Physical Journal B **75**(3), 327–334 (2010)

Sargent, J., Bedford, A.: Evaluating australian football league player contributions using interactive network simulation. Journal of sports science & medicine **12**(1), 116 (2013)

Sawe, B.E.: What Are The Most Popular Sports In The World? (2017).

https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html Accessed 2 Oct 2018

Tantipathananandh, C., Berger-Wolf, T., Kempe, D.: A framework for community identification in dynamic social networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 717–726 (2007)

Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world'networks. nature **393**(6684), 440 (1998)

Wikipedia: Football. https://en.wikipedia.org/wiki/Football Accessed 18 May 2022

Wilson, R.C., Zhu, P.: A study of graph spectra for comparing graphs and trees. Pattern Recognition **41**(9), 2833–2841 (2008)

Zager, L.A., Verghese, G.C.: Graph similarity scoring and matching. Applied mathematics letters **21**(1), 86–94 (2008)