Football Market Strategies: Think Locally, Trade Globally

Giulio Rossetti
University of Pisa
Email: giulio.rossetti@di.unipi.it

Vincenzo Caproni
University of Pisa
Email: vincenzo.caproni@gmail.com

***Abstract*—Every year football clubs trade players in order to build competitive rosters able to compete for success, increase the number of their supporters and amplify sponsors and media attention. In the complex system described by the football transfer market can we identify the strategies pursued by successful teams? Where do they search for new talents? Does it pay to constantly change the club roster? In this work we identify archetypal market strategies over 25 years of transfer market as depicted by UEFA professional clubs and study their impact on sportive success. Our analysis underline how, regardless from clubs' available budgets, transfer market strategies deeply impact – on the long run – football sportive performances.**

I. Introduction

Someone might say that football is the most beautiful sport in the world: for sure it is the most popular one. Every day, peoples from all over the world play football, cheer for their preferred teams and players, rejoice for their success and suffer for their defeats. The microcosmos football clubs, players and supporters shape with their daily activities attracts huge attentions from media and sponsors: all the *hype* surrounding this sport plays an important role on how it has evolved, till being transformed – more than other similar realities – into a complex economic system.

During the last years, data science has starting approaching sport analytics so to understand important aspects that such complex systems are able to express: training strategies, player performances, predictive models were proposed, studied and deployed in order to support coaches and societies in improving their results. However, conversely from individual sports, team results are strictly bound to the roster a club posses: a resource that can vary as time goes by. In order to increase their competitiveness clubs heavily rely on transfer markets investing huge budgets in both players and salaries. The majority of sports, due to their limited visibility, are able to attract reduced amount of sponsorships: football, contrarily, gather huge investments that lead to a very active transfer market among clubs. Football clubs, proportionally to the founding their blazon is able to guarantee, trade every year players from all over the world to reach their sportive targets, satisfy their supporters and attract investors.

In this work we propose an approach able to characterize football club market strategies and use them to evaluate their sportive performances. To do so, we automatically identify several archetypal market profiles starting from real world-wide transfer market data and use them in order to cluster clubs showing similar market pattern through time. Once identified clusters of clubs we correlate each one of them with the average rank of the teams they contain in order to ask the following questions: do market profiles correlate with club success? If so, which market strategy "guarantee" optimal results?

The paper is organized as follows: in Section II are discussed relevant works related to both sport data analysis and transfers data analysis; in Section III our dataset is introduced and some of its most relevant characteristics described. Moreover, our analytical model is proposed in Section IV and applied on a real case study in Section V. Finally, Section VI concludes the paper.

II. Related Works

Over the last decade, the analysis of sports data has sparked widespread interest: In [7] teams are modeled as a complex directed networks and analyzed in order to provide interesting informations about the best training patterns. In other works, such as the book Science and Soccer [20], are addressed various aspects of players training and nutrition while in [1] Borrie cares about conditioning and coaching. Hollinger, [11] performed a very criticized study aimed at classifying Basketball players and their abilities; Cycling as well has been object of a study[16] where the performance evolution of professional and amateur cyclists were estimated from the data provided by the tracker Strava[1]; Tennis tactics and patterns were mined in [23]. Other widely popular analysis concerns: tactics extrapolated using the complex network theory [19], the impact of collaboration between team members[4], the importance of single players and the quantification of their contributions to the team [5] as well as sport-dominant geographic regions[9]. Performance evolution analysis, often carried out exploring graph theory models and tools, represent another emerging field as demonstrated in [3], [18]: this kind of analysis has shown to be applicable to define features of a work group, as underlined in Grund[10] and Trequattrini[24], confirming that networks characterized by high intensity and low centralization are associated with better performing teams. Moreover, similar conclusions are stated in [2] where a set of performance metrics are summarized in a single indicator used to study the success of a team.

[1] https://www.strava.com/

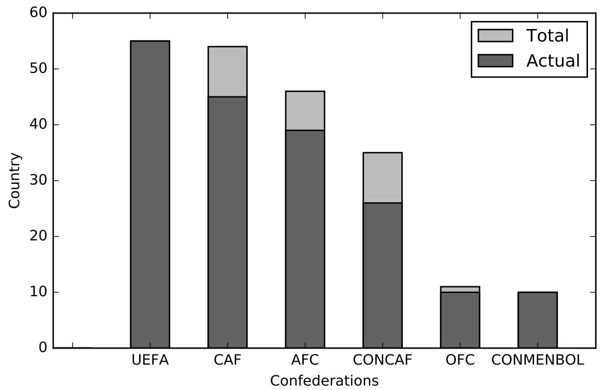


Fig. 1: Transfers representativeness across the FIFA confederations.

Indeed, the sportive data nowadays available are not limited to athletes/teams performances and workouts: professional sports design real economic markets that need to be analyzed in order to characterize successful strategies. Still focusing on sport data, it is possible to analyze money flows related to transfers as well as to transfer of players [13].In such kind of studies teams are considered as societies (or clubs) while the players are considered as assets [21]. Studying those data we can highlight two key relevant phenomena: globalization[22] and monetary flows. Globalization capture the patterns of players transfers across different countries and continents while monetary flows, on the other hand, describe the exchange of moneys and players between different clubs: those factors are often analyzed as correlated as in[6]. Globalization and markets offers the chance to deeply analyze different characteristics of a given sport, like the study of national preferences [25], that measures the effect of team nationality composition on fan attendance and overall team quality, as well as the correlation between taxation and international migrations [14]. Moreover, in [15] the analysis of transfer markets related to globalization and monetary fluxes suggest a correlation between the sportive results of clubs and their market activities: such analysis gain particular interest considering previous results on how to improve players piking in drafts [17].

Our work starts from these aspects to investigates the transfer market, searching for correlations between market actions taken by clubs and their sportive results – as in [8] where the authors analyze market volumes on various dimensions like remuneration and transfers.

III. Transfermarkt Data

The dataset we analyzed is collected starting from the website Transfermarkt[2], an online living encyclopedia on football that collects crowdsourced data about teams, players and transfers spanning from professional to juvenile leagues across all the globe. Due to its crowdsourced nature the data present on Transfermarkt appears often to be spurious and of varying quality (players roles partially translated, monetary values express in more than one currency): in order to perform our analysis we thus rely on the UK version of the service that, at the moment of the crawling, appeared to display a more consolidated version of transfer informations. For such reason in the following, whenever we discuss monetary values, we will use GBP as primary currency.

The dataset we collected spans from the early 1900s to 2015 and contains heterogeneous informations regarding the football microcosmos: in this paper we will focus in particular on 25 years of transfer market history (1990-2015) analyzing informations describing *football clubs*, *players*, *transfers* and *clubs results*. In detail, for each of such entities we collect, on yearly basis:

- **Clubs:** name, country, league, roster – 34.327 entities;
- **Players:** name, role, age, market value – 194.237 players;
- **Transfers:** footballer, clubs involved, amount payed, type of transaction (free/payed transfer, loan,...) – 823.321 transactions;
- **Results:** club, league, final rank, points, matches won/lose/draw, goals scored/received.

The dataset is completed by auxiliary data regarding *Stadiums* and *Palmares*, not used in the following analysis. The collected data covers football clubs of countries belonging to all the six FIFA confederations[3]: UEFA (Union des Associations Europeennes de Football), CAF (Confederation Africane de Football), AFC (Asian Football Confederation), CONCAF (Confederation of North, Central America and Caribbean Association Football), OFC (Oceania Football Confederation), CONMENBOL (Confederacion Sudamericana de Futbol).

Figure 1 show the percentage of countries of each confederation having at least a club present in our dataset.

Indeed, not all the confederations are equally represented – UEFA being the more detailed one. In Figure 2(left) we compare their market volume in order to better underline their relative economic power: symmetrically, Figure 2(right) compares the number of traded players per year by each confederation. We can observe that UEFA clubs, along with CONMENBOL ones, constantly monopolized the transfer market during the last decades. Obviously, this results is due to at least two reasons: (i) football is the most popular sport in both Europe and South America, and (ii) Transfermarkt volunteers are prevalently Europeans, thus a data collection bias can be expected.

IV. Club Transfer Strategy

In order to remain competitive year after year, football clubs allocate important budgets to improve their rosters (both covering players' salaries and transfers). Obviously sportive objectives, blazon and economic power deeply affect the expenses that a club is capable to sustain: indeed, teams belonging to the "*gotha*" of the football universe are capable of attract a higher number of supporters and sponsorships than medium/small size realities. In order to focus only on

[2]http://www.transfermarkt.com

[3]http://www.fifa.com/associations

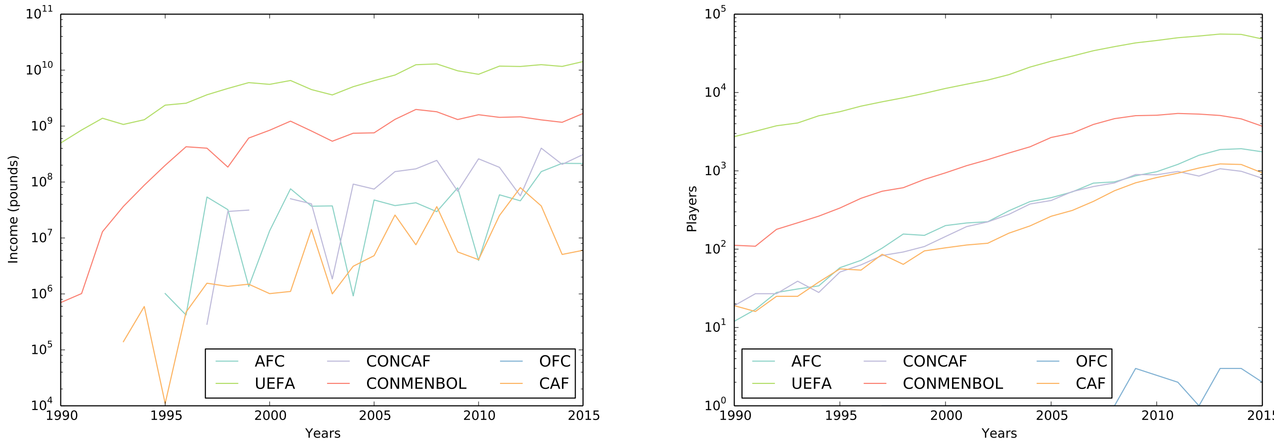


Fig. 2: Transfer Data: *(left)* economic market volumes for each confederation during the observed period; *(right)* number of players traded in each confederation on yearly basis. UEFA and CONMENBOL confederations concur to almost the 85% of trading activities described by the Transfermakt dataset.

market strategy, not on the relative importance and economic wealth of clubs, in this section, IV-A, we introduce a synthetic way to profile club market strategy. Moving from our simple definition of club profile, we propose an approach aimed at identifying more general market profiles able to subsume strategies of all the team in our dataset. In IV-B propose a model to capture market profile dynamics through time, thus describing *Club Market Profiles* that allow for a fine grain comparison of club market strategies.

A. Club Profiling

Several informations can be extracted looking to the operations a club makes on the player market in a given year. Indeed, most of the easily identifiable market informations relate to the amount of money spent/cashed: however, such indicators are heavily biased toward wealthy clubs that, following the ubiquitous Pareto distributions, represent less than the 20% of the overall "actors" that are able to move more that the 80% of the money circulating in the entire system (reaching the most expensive – and eligibly – most talented players). In order to avoid such bias we decided to neglect such information and to focus only on the strategy a club follows during a given market season: does the club only sell players? does it only buy them? Are they coming from leagues of the same nation or the club has xenophilic attitudes?

Definition 1 (Transfer)**.** *Let T be the transfer set. $t \in T$ is an individual transfer and can be modeled as a tuple of the form:*

$$t = (p, c^{from}, c^{to}, c^{from}_{country}, c^{to}_{country}, y) \qquad (1)$$

where p is a traded player, y represent the year in which the transfer took place, c^{from} identify the club that sold the player, c^{to} the club that bought the player and $c^{from}_{country}$ / $c^{to}_{country}$ the nationality of the club that participated in the transaction.

Moreover we define:

- $\pi_{c,y}$: all the market transaction in which c is involved (as source or destination) during y;
- $\pi^{from}_{c,y}$ ($\pi^{to}_{c,y}$): all the market transaction in which c is involved as the seller (symmetrically, buyer).

In order to keep simple the club profile we define two families of indicators: *attitude* and *heterogeneity* indicators.

Definition 2 (Market Attitude Indicators)**.** *Let c be a club and y a transfer year, we call* In Transfer *ratio:*

$$in = \frac{|\pi^{to}_{c,y}|}{\pi_{c,y}} \qquad (2)$$

and, symmetrically, Out Transfer *ratio:*

$$out = \frac{|\pi^{from}_{c,y}|}{\pi_{c,y}} \qquad (3)$$

Both the in and out range in [0,1].

Attitude indicators capture the percentage of transfers a club dedicate to acquire as well as get rid of players. It is true that the market value of players play a role in the proportion of such activities (after all, at least the UEFA, is coercing clubs to reach the so called "financial fair-play") however, at least in the last decade we have witnessed to countless transaction alternatives (from segmented payments up to several variants of loans with redemption obligation) that, at least in our opinion make this correlation slightly lower than in the past. However, looking solely to the percentage of incoming and outgoing transfers does not provide a detailed picture of clubs vision. To cope with this limitation we introduce a new class of indicators, namely market *heterogeneity*, that aimed to provide an in-depth specification of the observed market activities:

Definition 3 (Market Heterogeneity Indicators)**.** *Let c be a club, y a transfer year and $X \subset \pi^{to}_{c,y}$ containing all the* *transfers involving players acquired from clubs of the same country of* c. *We thus define:*

$$fsc = \frac{|X|}{|\pi_{c,y}^{from}|} \quad (4)$$

$$foc = \frac{|\pi_{c,y}^{from} - X|}{|\pi_{c,y}^{from}|} \quad (5)$$

where fsc *(*from-same-country*) capture the ratio of players acquired from clubs playing in the same country of* c *and* foc *(*from-other-countries*) the ratio of players acquired from other countries. Symmetrically we define* tsc *and* toc *to model the outgoing market toward the same as well as different countries. All the indicators range in [0,1].*

Market heterogeneity capture xenophilic attitudes of a club in a given year (i.e. to which extent it decides to focus its attentions – or not – on the country market while trading players). Obviously, such indicators are tied to the economic status of the specific club as well as to market trends and more broad financial situation that affect the football microcosms.

Given the proposed indicators we define *Club Profiles*:

Definition 4 (Club Profile)**.** *Fixed a year* y *and a club* c*, we call* $P_y(c)$ *(*Club Profile*) the tuple:*

$$P_y(c) = \{in, out, fsc, foc, tsc, toc\} \quad (6)$$

$P_y(c)$ univocally describe the market strategy of each given club by highlighting the choices its administrators make on how to spend the budget available. Indeed, the proposed profile definition provides a simplification of the complex dynamics expressed by football clubs (for instance, it does not take into account informations regarding the players that compose the roster and are not traded during a given year). However, our main objective is to identify peculiar market patterns across very heterogeneous club realities: in this scenario, the proposed simplification will allow us to focus onto specific indicators, reducing the dimensionality of the problem.

B. Generalizing Club Profiles

So far we described a simple model able to capture quantitative informations regarding the players traded during a given year by a given football club. Indeed, each *Club Profile* is able to summarize the activity of a single and well defined club: can we use the same set of informations to describe more general profiles, thus identifying market trends shared by clubs all over the world?

In order to answer such question, we approach the identification of *Market Profiles* as an unsupervised learning task:

Definition 5 (Market Profiles)**.** *Fixed a year* y *and a set of* Club Profiles $\mathcal{P}_y$ *we call* $\mathcal{M}_y$ *the set of* Market Profiles *during* y *and identify it as:*

$$\mathcal{M}_y = Kmeans(\mathcal{P}_y, k) \quad (7)$$

where the function $Kmeans$[4] *implements the homonym clustering algorithm. The parameter,* k *– e.g. the number of clusters – is usually identified by optimizing the sum of squared errors (SSE).*

A market profile describes a trading strategy applied by several clubs in a given year. As we will discuss in our case study, described in Section V, each market profile $m_i \in \mathcal{M}_y$ can be interpreted and labeled analyzing its medoid, i.e. the *Club Profile* among the ones grouped by m_i that minimizes the euclidean distances among all the other entities belonging to the cluster itself.

Market Profiles provide a succinct description of the trading strategies pursued by clubs world-wide during a specific year. However, clubs rarely maintain the same strategy for a long period: club objectives changes, rosters always needs to be improved, unpredictable events (game injuries, players' suspension…) can deeply affect the clubs plans: indeed, *no battle plan survives contact with the enemy*. All these reasons, along with the highly dynamism of the transfer market often due to the aging of the "goods" traded, make clear the main limitation of the proposed profiles: they can only picture static snapshots of club trading activities.

1) Profile Dynamics: In order to cope with the evolution of club market strategies we decided to extend the proposed profile exploiting one essential information: the temporal dimension. We defined both *Club* and *Market Profiles* to be parametric w.r.t. a time window, thus focusing our attentions on the activities that took place during a single year at a time. In order to extend such static profiles to a more dynamic model we pursue the following strategy:

i) we extract, for each club and year a profile, as in Def. 4;
ii) we identify the set of *Market Profiles*, as in Def. 5
iii) we then build, for each club, a time series in which for each year is identified the *Market Profiles* assumed the club.

The proposed workflow can be implemented following several modeling choices that deeply affect the produced results. Indeed the first step is trivial to accomplish and its output univocally identified by the Def. 4. Conversely, the second step can be implemented in at least three different ways:

- **Complete knowledge scenario:** *Market Profiles* are computed starting from all the *Club Profiles* disregarding the year, thus allowing multiple profiles to be considered for each team while applying Def. 5;
- **Partial knowledge scenario:** *Market Profiles* are computed on yearly basis: each club is represented by a single profile and all the profiles considered are computed on the same time window. This scenario can be further divided in two:
  - **Time-varying Profiles:** *Market Profiles* are independently computed for each time window (i.e., every year);

[4] In the following we will use the Kmeans++ implementation provided by the python library *sklearn*.

– **Fixed Profiles:** *Market Profiles* are computed once during a bootstrap phase, i.e. during the first year, and are inherited for the subsequent time windows.

Obviously each of the proposed alternatives has its own strength as well as drawbacks, as shown in Table I.

In the following analysis we decided to take advantages of the *Partial knowledge – fixed profiles* scenario. Such choice was made in order to simplify the identification and characterization of the market strategies. Indeed, working with a preset number of profiles (whose interpretation is fixed) may make impossible to recognize novel emerging trends: however, due to the high level definition of the indicators we used to build clubs profiles we can make the assumption that such trends, if not captured in the bootstrap phase, are not particularly interesting for our investigation, or equivalently that we are interested to observe only the evolution of club's strategies within certain specific trading categories.

Once computed both *Club* and *Market Profiles* we approach the last step of the proposed workflow by assigning each $P_y(c) \in \mathcal{P}$ to the more similar $m_i \in \mathcal{M}$. We solve the mapping problem through a *k-nearest neighbors* (KNN) approach, building $(P_y(c), m_i)$ pairs that minimize the euclidean distance among the *Club Profile* and the selected *Market* one. This step allow us to build for each club a time series capturing the evolution in time of its market strategy starting from a precomputed set of *Market Profiles* (our classes) and all its yearly individual profiles. More formally, at the end of this process we get a set of *Club Market Profiles*:

Definition 6 (Club Market Profiles)**.** *Let c be a club and $\mathcal{P}(c)$ its ordered set of* Club Profiles *defined as $\mathcal{P}(c) = \{P_0(c), \ldots, P_n(c)\}$. Given the set of the* Market Profiles $\mathcal{M}$ *and a function $\theta(P_i(c), \mathcal{M})$ assigning a given* Club Profile *to the most similar* Market Profile*, we call* Club Market Profile *of c*

$$CM(c) = \{\theta(P_0(c), \mathcal{M}), \ldots, \theta(P_n(c), \mathcal{M})\} \tag{8}$$

namely, the time series of Market Profiles *describing the market strategy of c.*

Strategy	Advantages	Drawbacks
Complete Knowledge	(i) Fixed number and interpretations of *Market Profiles*, (ii) stability through time	All the yearly profiles need to be known beforehand: in case of new observations *Market Profiles* need to be recomputed from scratch.
Time-varying Profiles	Capture the yearly specificity of the global transfers market.	(i) Varying number of *Market Profiles*, (ii) every year the obtained profiles needs to be interpreted from scratch.
Fixed Profiles	(i) Fixed number and interpretations of *Market Profiles*, (ii) allows incremental construction of time series when a new observation appears.	As time goes by some of the identified profiles may be scarcely populated and novel ones not represented at all.

TABLE I: Market Profiles construction strategies: advantages and drawbacks.

In the following Section we will discuss how *Club Market Profiles* can be fruitfully used to evaluate time-aware market similarities among football clubs and, thus, to draw some conclusions on the influence trading strategies have on sportive results.

V. Case Study: UEFA 1990-2015

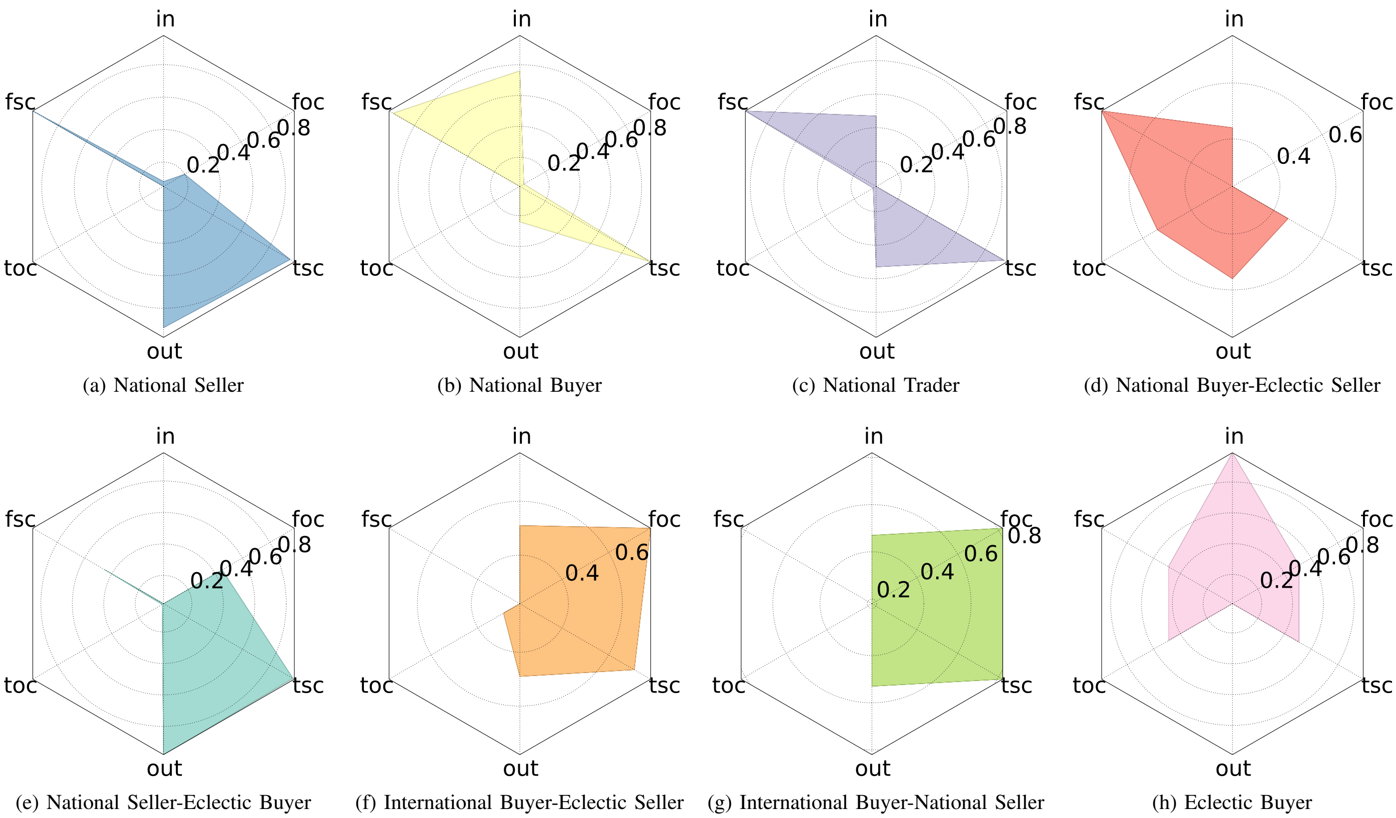
In order to better bound the performed analysis we decided to focus on a specific case study: we analyze the trades of clubs militating in leagues of countries affiliated to the UEFA confederation. In V-A, we use such filtered dataset to build *Club Market Profiles* and correlate them with teams' success. Finally, in V-B we briefly discuss our findings, highlight their implication and discuss the effect roster stability has on clubs' performances.

A. Club Market Strategy Profiling

We have seen in Section III that UEFA's data can be considered the most complete ones among the six FIFA confederations. Starting from this observation we focused our attention on the transfers involving clubs belonging to such confederation in a very specific time period: 1990-2015. Once applied such filtering we get a dataset composed by 1,756 clubs of 47 countries, for a total of 43,900 *Club Profiles*.

Moving from the definitions provided in Section IV as a first step we extracted *Market Profiles* from the *Club Profiles* computed for the year 1990. Figure 3 synthesize, using radar charts, the medoids of the clusters we obtain. The obtained profiles – whose numbers and definitions have appeared to be surprisingly stable across all the years selected for our analysis – can be described as follows:

- **National Seller**. This category, composed on average by 8% of the clubs on yearly basis. National sellers focus on nation-wide outgoing transfers – Fig. 3(a).
- **National Buyer**. Clubs having such profile are 12% of total: as opposite to the previous profile, National Buyers only acquire players – still on the national market – limiting as possible incoming transfers, Fig. 3(b).
- **National Trader**. This is the most common market profile describing on average the 28% of the clubs that balance incoming and outgoing transfers focusing almost exclusively on the national market – Fig. 3(c).
- **National Buyer-Eclectic Seller**. This strategy is followed by 5% of the clubs: differently from *National Buyer*, the clubs within this cluster compensate the incoming national market strategy with an outgoing eclectic one, thus selling to both national and international buyers – Fig. 3(d).
- **National Seller-Eclectic Buyer**. Almost 18% of the football companies fall in this profile. Symmetrically to the previous profile, in this cluster fall all those clubs that acquire players independently from the country of the team they are affiliated while sell only on the national market – Fig. 3(e).
- **International Buyer-Eclectic Seller**. This cluster groups 7% of clubs that acquire players almost exclusively on the



(a) National Seller (b) National Buyer (c) National Trader (d) National Buyer-Eclectic Seller

(e) National Seller-Eclectic Buyer (f) International Buyer-Eclectic Seller (g) International Buyer-National Seller (h) Eclectic Buyer

Fig. 3: UEFA Market Profiles.

international market while sell them both to clubs of the same country as well as to clubs of different ones – Fig. 3(f).

- **International Buyer-National Seller**. As the previous profile it describes clubs particularly active on the international markets (even though only for player acquisition). However, the clubs complying with this profile (almost 7% of the total) sell their players only within the national market – Fig. 3(g).
- **Eclectic Buyer**. Finally, the remaining 17% of the analyzed clubs have limited and balanced outgoing market and a particularly prominent incoming one. Eclectic Buyers acquire players disregarding the nationality of the clubs in which they currently perform – Fig. 3(h).

After having identified our *Market Profiles* we built, for each club, a *Club Market Profile* spanning from 1990 to 2015. We end up with a set of time series each one of them describing the evolution of market strategies of a given football club playing in a country affiliated to the UEFA confederation.

Once obtained the temporal annotated description of club profiles we exploited such informations to identify clusters having similar *Club Market Profiles*. To do so we applied kmodes [12], an unsupervised learning approach able to cluster data described by categorical features. Kmodes works following the same rationale of kmeans: it identifies k clusters by iteratively computing distances between the objects passed as input (i.e., the computed *Club Market Profiles*) and the medoids identified during the previous iteration. Each object is assigned to the nearest cluster and the algorithm complete its computation when a stable solution is reached. We identified k=5 as the optimal number of clusters while minimizing the SSE[5]. The identified clusters describes different market strategies: analyzing their centroids we get the following categorization:

- **Cluster 0**: The clubs in this cluster alternate the adoption of *National Seller* and *National Seller-Eclectic Buyer* strategies;
- **Cluster 1**: All the *Club Market Profiles* associated to this group alternate *National Seller-Eclectic Buyer* to the *National Trader* strategy;
- **Cluster 2**: Differently from the previous cluster the clubs here collected are identified by the adoption of *National Buyer-Eclectic Seller* strategie and, partially by the *National Trader*;
- **Cluster 3**: Like Cluster 2 the profile of clubs here represented are characterized by *National Buyer-Eclectic Seller* profiles. However, instead of focusing in the national market exclusively for selling players they alternate their principal strategy with a *National Buyer* one;
- **Cluster 4**: Finally, in this cluster are grouped clubs constantly characterized by a single profile: *National*

[5]Sum of squared errors

Cluster	Coverage	Avg. Rank	Club Example
0	(13%)	[1-4]	Barcelona, Porto, Monaco, Bayern Munich
1	(13%)	[3-6]	Juventus, Chelsea, Olympique Marseille, Atletico Madrid
2	(18%)	[5-10]	Arsenal, Inter Milan, Bayer Leverkusen, AJ Auxerre
3	(7%)	[7-14]	Everton, Tenerife, Genoa, Stuttgart
4	(49%)	[12-18]	Sheffield, Napoli, Auxerre, Leicester City

TABLE II: *Club Market Profiles* clusters. For each cluster is reported the percentage of clubs covered as well as their average national league rank. Moreover, for each cluster are shown some examples.

Buyer.

In order to evaluate if clubs pursuing similar market strategies for protracted periods are also experiencing comparable success in their respective national leagues, we cross-referenced the identified clusters with the yearly national league rankings of the teams they group together.

To do so we computed, for each football team, the mode of its yearly league final ranks in the period 1991-2015: then, we calculated for each cluster the expected team position as the mean of its teams' ranks mode (approximated by their standard deviation), thus obtaining the results reported in Table II.

Our findings show that the overlap among average club rankings within the identified clusters is very low: this suggest that the *Club Market Profiles* represented by each cluster can provide insights on the success a club is expected to reach in its national league. Clubs that sell players almost exclusively on the national market but acquire them eclectically (Clusters 0 and, partially, 1) are the ones that are able to constantly reach good placements: among them we can find Barcellona, Bayern Munich, Juventus and Chelsea. Conversely, clubs that focus their attention almost prevalently on the national market while acquiring players (Clusters 2, 3 and 4) are not able to take the leap and reach high placements or, if they do, they are not able to sustain them for a long period. Examples are: Arsenal, Inter Milan, Stuttgart, Napoli.

Obviously our results are strongly tied to the period selected for the analysis that span over 25 years of market strategies and football results: indeed, varying the time window we can identify slightly different team assignation to clusters, even though we observed that the optimal number of groups and their medoids remain surprisingly stable across several timespan (ranging from 25 down to 5 years).

B. Rosters, Stability and Performances

Analyzing transfer data and cross-referencing them with clubs' sportive results we obtained a quite clear picture: choose and sustain a specific market strategy through time affect the club success. Moreover, it seems that in order to perform better than national competitor it is mandatory for a club to nurture an *international* profile, searching and acquiring talents on the international market, not only on the national one. Even though this result seems obvious due to the increasingly globalized world we live in, it is not so. In order to avoid distortions and bias due to the different economic power of clubs we decided to focus only on market strategies – considering top clubs and third category ones alike. Considering such choice, our findings suggest that at each level of football professionalism it is mandatory to scout talents not only at the country level: the transfers can be loans, exchanges as well as onerous ones but they must show openness of the club toward different national football realities.

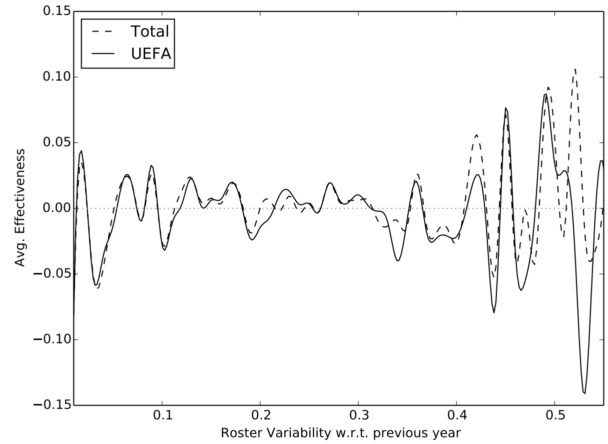


Fig. 4: Roster Variability vs. Team avgerage Effectiveness. Roster stability deeply affect team performances: on average, market strategies that change roster for less than 30% w.r.t. the previous year guarantee continuity in the results.

Indeed, trade players worldwide seems to be the only strategy able to assure club competitiveness on the long run. However, one question still remains unanswered: how much a club needs to change its roster in order to observe a significant impact on its performances? Given a club c and its rosters R_0, R_1 observed during two consecutive years we call roster variability:

$$\epsilon = 1 - \frac{|R_0 \cap R_1|}{|R_0|} \tag{9}$$

ϵ identify how much, in percentage, the roster of c changed after a year of market activity. Symmetrically, given the points p_0, p_1 scored during two subsequent years by c in its league we define its effectiveness as:

$$\kappa = \frac{p_1 - p_0}{p_0} \tag{10}$$

In Figure 4 we report, for the whole dataset as well as for the UEFA case study, the average impact of roster variability on club effectiveness. We can notice that both trends follows a very similar shape that can be qualitatively decomposed in three sub-regimes:

- **High stability** ($\epsilon \in$ [0,0.1]): the roster changes are below the 10% and the average performances w.r.t. the previous year appears to fluctuate. It is likely that the traded players are only second liners, not fundamental ones;
- **Average stability** ($\epsilon \in$ [0.1,0.3)): the roster starts changing in some of its key player. The performances are quite stable and comparable with the ones of the previous year.

- **Low stability** ($\epsilon \geq 0.3$): the roster is heavily changed (probably due to the end of a cycle and/or the change of coach/societal vertices). This scenario is the one in which the performances variates the most: a good part of the team first line is changed and the lack of togetherness is likely to cause performance worsening.

In the end it seems that, at least on average, conservative market choices involving no more than 30% of the whole team are the safest choice for clubs that need to maintain national competitiveness as well as to improve their actual roster with some alternatives (second liners) or few highly talented first line players. Conversely, apply a higher turnaround represent a very risky choice to made: it can lead to a high effectiveness as well as to dramatic negative peaks since it deeply affect all at once the foundation of the roster itself.

VI. Conclusion

In this work we analyzed 25 years of football related data covering the trading activities of clubs within the UEFA confederation. We define *Market Profiles* in order to categorize the market choices yearly made by clubs in our dataset: to do so we disregard informations capturing the economic power of clubs and focus our attention on more high-level *strategies*. From such profiles we build up *Club Market Profiles*, a 25 year-long representation of club market strategies: finally, we clustered them so to identify clubs following the same market profiles over time and correlate the identified groups with their sportive success. Our analysis show that club success over the years is affected by the market strategies pursued and by roster stability: in the end, if you want to win your local league you have to trade globally!

As future work we plan to extend our analysis to the other FIFA confederation, to take into account other valuable information in the market profile definition (such as the average year of players traded and their role) and to analyze the correlations among the wealth of clubs and their market strategy. Indeed, another interesting aspect to study will be the impact of the identified of market strategies have on the international success of clubs: does a local optimal market strategy make a club a strong candidate to win the UEFA Champions League?

References

[1] Andrew Borrie, Gudberg K Jonsson, and Magnus S Magnusson. Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. *Journal of sports sciences*, 20(10):845–852, 2002.

[2] Paolo Cintia, Fosca Giannotti, Luca Pappalardo, Dino Pedreschi, and Marco Malvaldi. The harsh rule of the goals: data-driven performance indicators for football teams. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1–10. IEEE, 2015.

[3] Paolo Cintia, Salvatore Rinzivillo, and Luca Pappalardo. A network-based approach to evaluate the performance of football teams. In *Machine Learning and Data Mining for Sports Analytics Workshop, Porto, Portugal*, 2015.

[4] Carlos Cotta, Antonio M Mora, Cecilia Merelo-Molina, and Juan Julián Merelo. Fifa world cup 2010: A network analysis of the champion team play. *arXiv preprint arXiv:1108.0261*, 2011.

[5] Jordi Duch, Joshua S Waitzman, and Luís A Nunes Amaral. Quantifying the performance of individual players in a team activity. *PloS one*, 5(6):e10937, 2010.

[6] Annika Eliasson. The european football market, globalization and mobility among players. *Soccer & Society*, 10(3-4):386–397, 2009.

[7] Giorgio Fagiolo. Clustering in complex directed networks. *Physical Review E*, 76(2):026107, 2007.

[8] Bernd Frick. The football players'labor market: Empirical evidence from the major european leagues. *Scottish Journal of Political Economy*, 54(3):422–446, 2007.

[9] Akira Fujimura and Kokichi Sugihara. Geometric analysis and quantitative evaluation of sport teamwork. *Systems and Computers in Japan*, 36(6):49–58, 2005.

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

[11] John Hollinger. The player efficiency rating, 2009.

[12] Zhexue Huang. Clustering large data sets with mixed numeric and categorical values. In *Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining,(PAKDD)*, pages 21–34. Singapore, 1997.

[13] D Jones, R Parkes, and A Houlihan. Football money league: Changing of the guard. *Manchester: Deloitte Sports Business Group, http://www2.deloitte.com/content/dam/Deloitte/global/Documents/Audit/gx-deloitte-football-money-league-2006.pdf, accessed*, 9:2010, 2006.

[14] Henrik Jacobsen Kleven, Camille Landais, and Emmanuel Saez. Taxation and international migration of superstars: Evidence from the european football market. *The American Economic Review*, 103(5), 2013.

[15] Xiao Fan Liu, Yu-Liang Liu, Xin-Hang Lu, Qi-Xuan Wang, and Tong-Xing Wang. The anatomy of the global football player transfer network: Club functionalities versus network properties. *PloS one*, 11(6):e0156504, 2016.

[16] Paolo Cintia Luca Pappalardo, Dino Pedreschi. Engine matters: A first large scale data driven study on cyclists' performance. *ICDM 2013 - DMCS Workshop*, 2013.

[17] Cade Massey and Richard H. Thaler. The loser's curse: Decision making and market efficiency in the national football league draft. *Management Science*, 59(7):1479–1495, 2013.

[18] Pedro Passos, Keith Davids, Duarte Araújo, N Paz, J Minguéns, and Jose Mendes. Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*, 14(2):170–176, 2011.

[19] Javier López Pena and Hugo Touchette. A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904*, 2012.

[20] Thomas Reilly. *Science and soccer*. Routledge, 2003.

[21] PeterJ. Sloane. The labour market in professional football*. *British Journal of Industrial Relations*, 7(2):181–199, 7 1969.

[22] Matthew Taylor. Global players? football, migration and globalization, c. 1930-2000. *Historical Social Research / Historische Sozialforschung*, 31(1 (115)):7–30, 2006.

[23] Antonio Terroba, Walter A Kosters, and Jonathan K Vis. Tactical analysis modeling through data mining. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*. Citeseer, 2010.

[24] Raffaele Trequattrini, Rosa Lombardi, and Mirella Battista. Network analysis and football team performance: a first application. *Team Performance Management: An International Journal*, 21(1/2):85–110, 2015.

[25] Dennis P. Wilson and Yung-Hsiang Ying. Nationality preferences for labour in the international football industry. *Applied Economics*, 35(14):1551–1559, 2003.