Northumbria Research Link

Citation: Zeng, Yifeng, Yu, Shenbao, Pan, Yinghui and Chen, Bilian (2022) Discovering a cohesive football team through players' attributed collaboration networks. Applied Intelligence, 53 (11). pp. 13506-13526. ISSN 0924-669X

Published by: Springer

URL: https://doi.org/10.1007/s10489-022-04199-4 <https://doi.org/10.1007/s10489-022-04199-4>

This version was downloaded from Northumbria Research Link: https://nrl.northumbria.ac.uk/id/eprint/50255/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria University NEWCASTLE

UniversityLibrary

Discovering a cohesive football team through players' attributed collaboration networks

Shenbao Yu[1,2], Yifeng Zeng[3*], Yinghui Pan[4*] and Bilian Chen[1,2]

[1]Department of Automation, Xiamen University, Xiamen, 361005, China.
[2]Xiamen Key Laboratory of Big Data Intelligent Analysis and Decision-making, Xiamen, 361005, China.
[3*]Department of Computer and Information Sciences, Northumbria University, Newcastle, NE1 8ST, UK.
[4]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China.

*Corresponding author(s). E-mail(s): yifeng.zeng@northumbria.ac.uk; panyinghui@szu.edu.cn;
Contributing authors: yushenbao@stu.xmu.edu.cn; blchen@xmu.edu.cn;

Abstract

The process of team composition in multiplayer sports such as football has been a main area of interest within the field of the science of teamwork, which is important for improving competition results and game experience. Recent algorithms for the football team composition problem take into account the skill proficiency of players but not the interactions between players that contribute to winning the championship. To automate the composition of a cohesive team, we consider the internal collaborations among football players. Specifically, we propose a Team Composition based on the Football Players' Attributed Collaboration Network (TC-FPACN) model, aiming to identify a cohesive football team by maximizing football players' capabilities and their collaborations via three network metrics, namely, network ability, network density and network heterogeneity&homogeneity. Solving the optimization problem is NP-hard; we develop an approximation method based on greedy algorithms and then improve the method through pruning strategies given a budget limit. We conduct experiments on two popular football simulation platforms. The experimental results show that our proposed approach can form effective teams that dominate others in the majority of simulated competitions.

Keywords: Football team composition, Attributed collaboration networks, Game analysis, Heterogeneity&homogeneity

1 Introduction

The process of team composition, which aims to discover an appropriate set of individuals with relevant expertise to achieve common goals efficiently, has been a major area of interest in the field of the science of teamwork. As football (also called "soccer" in some countries) requires a high level of teamwork, it is one of the best options for studying the team composition problem since it is characterized by a large amount of communication, interaction, and collaboration between team members. In reality, it is difficult to assess the effectiveness of a football team composition result because it may require a considerable amount of money as well as being labor-intensive. Fortunately, the emergence of a wide variety of football video games, such as Pro Evolution Soccer (PES) [1], Electronic Arts Sports FC (also known as FIFA) [2] and Football Manager [3], offers an opportunity to compose a team based on human preferences and evaluate outcomes efficiently. This opportunity exists not only because gamers can completely redo club designs as well as edit any player in the game but also because the platforms can fully simulate on-pitch football matches. Subsequently, the football team composition task becomes interesting and important on the game platforms.

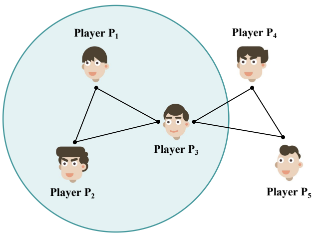
As a multiplayer game, the process of football player selection and team composition is designed to select the most suitable player for a particular playing position and role [1], which is vital for clubs to be able to deliver high sports and financial returns [2]. Such a process is crucial since a poor selection result can affect player loyalty as well as cost a football team millions of dollars [3]. However, the multicriteria complexity and decision-making difficulty make the selection of players a challenging task. Although team managers and coaches use a variety of assessments to choose players by considering many aspects, including player productivity and limited wage budgets, the selection process would be too time-consuming to be realistic, and the accurate evaluation of a player's suitability for a team is also a considerable puzzle. Thus, applying a systematic approach such as the mathematical modeling method is urgent.

Many studies have attempted to address the football team composition problem, but most of them rely on attributes such as players' skills and physical status. For instance, most researchers utilize anthropometric measurements

[1] https://www.konami.com/
[2] https://www.ea.com/games/fifa
[3] https://www.footballmanager.com/

(e.g., age, height, and weight), fitness-related indices (e.g., vertical jump ability and speed), and players' techniques (e.g., short passing and shooting) for the football player selection problem [4]. In addition, the market value and salary of football players are taken into account [5, 6]. Specifically, Zeng et al. [5] considered the players' total salary as a budget constraint and resorted to a submodular function to solve the team composition problem. However, such attributes are not sufficient to measure a football team's competitiveness. Achieving good results depends on not only the high-level players who are involved but also how effectively they collaborate, communicate, and work together as a team.

Assume, for example, a team manager who wants to build a football team consisting of players with distinguished skills in the following areas: {*attacking prowess, ball control, defensive prowess, physical contact, and speed*}. We also assume that there is a network including five football players $\{P_1, P_2, P_3, P_4, P_5\}$ in Figure 1. Each player highlights the corresponding skills, and an edge between two football players indicates that they can collaborate effectively. Such a network is referred to as an *attributed collaboration network* (ACN) [4] [7]. Without considering the connection among players, the manager can select either $\mathbf{C}_1 = \{P_1, P_2, P_3\}$ or $\mathbf{C}_2 = \{P_1, P_4, P_5\}$ - both $\mathbf{C}_1$ and $\mathbf{C}_2$ have the required skill set. However, the candidate set $\mathbf{C}_1$ is the better choice since the network indicates that P_1 cannot work with P_4 and P_5 effectively.



Player	Skills
P_1	defensive prowess
P_2	physical contact, speed
P_3	attacking prowess, ball control
P_4	ball control, speed
P_5	attacking prowess, physical contact

Fig. 1: An example of an ACN with five individual players, each of whom is equipped with several skills.

The existence of an ACN among football players is quite common. In a football league, an obvious type of player collaboration is developed upon whether they are from the same team or country, which is often used to organize players in a club. In this case, the network encodes the fact that football players from the same club or country can communicate more easily and cooperate more harmoniously with each other than those serving in different teams. In addition, it is known that defensive and offensive positions differ in player composition because they are conjunctive and disjunctive tasks respectively [8]. The success of driving off each attack is dependent on completing a joint action. Here, the

[4] We list the abbreviations of major terms throughout this paper in Table A2, Appendix A to ease reading.

weakest defender is detrimental to the team's defensive performance because he or she limits the team's defensive capabilities. In contrast, a team's offensive capacity is determined by the output of the best-performing member. Moreover, the distribution of a team's offensive (defensive) performance can be measured by the network *heterogeneity* (*homogeneity*) [9, 10]. Low heterogeneity (or high homogeneity) indicates that all players share a similar level of interaction through the match, and vice versa. Thus, attacking benefits from heterogeneous players, while homogeneity ensures that there are no weak links among defensive players. This insight facilitates our understanding of the underlying functional mechanism of collaboration and motivates us to develop players' attributed collaboration networks for the football team composition problem.

In this paper, we consider the team composition problem in the context of the Football Players' Attributed Collaboration Network (FPACN). Each node in the network is a football player with certain skills, such as *attacking prowess, ball control, dribbling*, while edges between nodes are constructed based on the clubs they played for and their nationalities, which reflect the affinity between players. After obtaining the attributed collaboration network, given a certain budget, we propose a TC-FPACN model, the acronym for Team Composition based on the Football Players' Attributed Collaboration Network, to identify a set of highly qualified football players and form a remarkably cohesive team. We evaluate the cohesiveness of a football team on the basis of three predefined network metrics, namely, *network ability*, *network density*, and *network heterogeneity&homogeneity*, in the TC-FPACN, whose goal is to discover a football team that maximizes the combination of the three network metrics. As we present the team's properties through the attributed collaboration network, the constrained optimization problem can be converted to finding a maximum density subgraph in a graph, which turns out to be NP-hard [11]. The problem becomes more complicated when players' ability and heterogeneity (or homogeneity) are considered. We propose an approximation algorithm that finds the best team based on greedy algorithms and further improve the algorithm using pruning methods under a budget constraint. We summarize the main contributions of this paper below.

- We propose a Team Composition based on the Football Players' Attributed Collaboration Network (TC-FPACN) model, which incorporates three network metrics (i.e., network ability, network density, and network heterogeneity&homogeneity) to define players' cooperation mechanism.
- We formulate the team composition task as a constrained optimization problem for the TC-FPACN that finds the optimal subgraph based on the network metrics. Since the problem is NP-hard, we propose a greedy algorithm with a pruning technique to solve it.
- We conduct an empirical study on two video game platforms, i.e., Pro Evolution Soccer 2018 (PES2018) and EA SPORTS FIFA 22 (FIFA2022) to evaluate the effectiveness of the proposed model. Simulation results show that our model achieves favorable performance in competition against other teams.

The remainder of the paper is organized as follows. We review related works in Section 2. In Section 3, we first formally introduce the team composition task, then describe the three network metrics of the TC-FPACN and finally, formulate the team composition problem. We propose the new algorithms in Section 4. Section 5 demonstrates the performance of the proposed method. Finally, Section 6 concludes our work and discusses future research directions.

2 Related Work

Since this paper considers finding a cohesive football team based on football players' capabilities and collaborations, we start with a review of football decisions, especially for player selection and team composition, and proceed with related research on the evaluation of personal ability and the retrieval of the team from collaboration networks in general.

2.1 Football player selection and team composition

The process of football player selection and football team composition is a complex problem with conflicting objectives. The traditional solution to this problem is to assess several quantitative factors that are compulsory for coaches and their technical committees to produce the most elite player. These factors include the player's anthropometric measurements [4], fitness-related indices [12], and skills [5, 13]. To name a few, Inan and Cavas [13] analyzed the offensive and defensive characteristics of Turkish Super League football players, such as the long pass accuracy, and developed an artificial neural network model for talent selection. Zeng et al. [5] defined a submodular function that represents the team's skill coverage and used improved greedy algorithms to solve the optimization problem. Given the existence of different duties for football players in the field, many researchers have also considered that the relevant criteria of skills must be assigned according to each player's position [3, 14, 15]. Ozceylan [3], for example, used an analytic hierarchic process to prioritize the criteria for each player based on their position and developed a 0-1 integer linear programming to determine top players in a team.

Most approaches mentioned above emphasize the on-pitch sport success. In addition, there are other factors worth considering, such as financial aspects [16, 17] and the future potential of professional football players [18, 19]. For instance, Singh and Lamba [16] resorted to machine learning models including decision tree and gradient boost to identify the factors that affect the financial market values of football players and then used the selected factors to predict the player's market value. In [18], the authors projected a target player's potential by searching the corresponding historical attributes to identify other football players with a similar profile. Zhao et al. [19] defined three attributes, including the potential factor, to evaluate the performance of teams and football players.

Nevertheless, forming a winning football team involves more than having the required mix of skills under the budget limit. Player selection is a difficult decision-making problem that needs to take into account the collaboration mechanisms among football players, which are ignored in the literature.

2.2 Personal ability evaluation

Personal ability is always an important guideline for team composition. Player selection needs to consider quantitative attributes, and the most widely used rating systems for a player are based on performance data. Since there are multiple attributes to consider when assessing a player's ability, algorithms based on multicriteria decision-making (MCDM) are regarded as simple and suitable for developing solutions [20]. As a key component of the MCDM method, the analytic hierarchic process (AHP) is widely used to determine the weights of the selected criteria [21]. Using the AHP methods, each player's attributes are ranked according to their importance in a given position. In parallel, the technique for order of preference by similarity to ideal solution (TOPSIS) – the well-known MCDM method – is applied extensively to rank the alternatives, partly due to its mathematical clarity. A plethora of methods have been developed following this breakthrough, such as TOPSIS-IPA [22] and Fuzzy-TOPSIS [23]. More recently, Sałabun et al. [24] developed a multicriteria model based on the characteristic objects method to evaluate players in team sports.

In addition to MCDM-based models, Liu et al. [25] introduced the text information of postmatch reports written by professional soccer journalists or editors and proposed an affective computing model for the player's performance rating. Furthermore, Pantzalis and Tjortjis [26] conducted an intensive study to define the main attributes that influence a defender's match rating. They found that classic defensive actions such as interceptions and clearances, along with player attributes such as jumping reach and strength, are more suitable for evaluating defenders.

2.3 Collaboration networks for a team formation

A successful team relies on not only individual ability but also communication and collaboration. The study of scientific collaboration aims to compute the fitness level of an expert for collaborating with other experts on a set of skills [27]. Given an expertise collaboration network, Lappas et al. [28] first considered team formation in the presence of a collaboration network and measured effectiveness using communication cost. Furthermore, density-based measurements were proposed [29–31], and the authors generalized the approach [28] by considering the team formation problem as a multiobjective optimization task. For example, Selvarajah et al. [31] aimed to build a more effective team by analyzing various scenarios, such as how frequently team members had worked together in the past. In parallel, Datta et al. [32] proposed a composite mechanism to exploit different elements of individuals and the community given by their expertise and connections. Furthermore, Awal and Bharadwaj [33] quantified and optimized a team's collective ability based on a collective intelligence index, which encodes individuals' knowledge competence and their collaboration competence.

Given that the major limitations of the class of solutions mentioned above are that they fail to capture complex interactions and are computationally intractable, more recent work adopted neural architectures to learn a mapping between the skills and experts' space [34–36]. For instance, Hamidi et al. [36] focused on state-of-the-art neural network methods to learn the dense representations for nodes in the collaboration network and bootstrapped the training process through transfer learning. Similarly, in this paper, we focus on the team formation problem based on the collaboration network and explore an efficient way to find a team. Specifically, we consider a network structure of football players as an attributed collaboration network, where nodes representing players are associated with their skills and the weights attached to edges reflect their degree of affinity.

3 TC-FPACN Model

In this section, we present the TC-FPACN model, which is formed by three network metrics that contribute to determining the cohesiveness of a football team, including network ability, network density, and network heterogeneity&homogeneity. We first formally introduce the team composition task and then detail the network metrics. Finally, we formulate the objective function of TC-FPACN, which is to discover a subnetwork by maximizing the three metrics simultaneously.

3.1 Task formulation

Let $\mathbf{P} = \{\mathrm{P}_n\}$ $(1 \leq n \leq N)$ be a set of football players, and $\mathbf{S} = \{\mathrm{S}_m\}$ $(1 \leq m \leq M)$ be a set of players' skills, where N and M are the number of football players and skills, respectively. Assume that football players are organized in a weighted and undirected graph (i.e., FPACN), denoted as $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. Each node $v_n \in \mathcal{V}$ is associated with a football player P_n equipped with a set of skills [5], while an edge $(i, j) \in \mathcal{E}$ models the relationship between the pair of the players (i.e., P_i and P_j). In addition, for readability, we present the main notations used throughout the paper in Appendix A, Table A2.

In football, it is intuitive that different positions on the pitch highlight different skills, which means that some skills are common (e.g., *body control* and *jump*) while others (e.g., *goalkeeping*) are unique to a particular position (e.g., *goalkeeper*). Thus, we divide football players into three groups - *Forward/Midfielder*, *Backward*, and *Goalkeeper* - according to a player's position in the football field, with the corresponding collaboration network $\mathcal{G} = \mathcal{G}_{\mathrm{F}} \cup \mathcal{G}_{\mathrm{B}} \cup \mathcal{G}_{\mathrm{G}}$, where $\mathcal{G}_{\mathrm{F}}$, $\mathcal{G}_{\mathrm{B}}$, and $\mathcal{G}_{\mathrm{G}}$ are subgraphs for *Forward/Midfielder*, *Backward*, and

[5] In the context of the attributed collaboration network of football players, if not otherwise specified, we use v_n or P_n indiscriminately to represent the same football player.

Goalkeeper respectively. We define the task of football team composition as follows:

Definition 1 *Given an attributed collaboration network of all football players and a limited budget, the goal of our team composition task is to form a cohesive subnetwork (i.e., football team)* $\mathcal{G}'(\mathcal{V}', \mathcal{E}') \subseteq \mathcal{G}(\mathcal{V}, \mathcal{E})$*, where the node set* $\mathcal{V}'$ *represents the selected football players.*

3.2 Three network metrics

The TC-FPACN model considers the cohesiveness of a football team from three aspects: *a*) network ability, *b*) network density, and *c*) network heterogeneity&homogeneity. We now describe the three network metrics in detail.

3.2.1 Network ability

Given a football player $\mathrm{P}_n \in \mathbf{P}$ $(1 \leq n \leq N)$ with a set of skills, each of which is labelled with the corresponding weight and personal level, we first build a model to calculate the personal ability of P_n, denoted ϕ_{P_n}, in Eq. (1).

$$\phi_{\mathrm{P}_n} = \sum_{m=1}^{M} W_{\mathrm{S}_m} L_{\mathrm{P}_n, \mathrm{S}_m}, \tag{1}$$

where W_{S_m} is the weight of skill S_m, and $L_{\mathrm{P}_n, \mathrm{S}_m}$ is the personal level of S_m for player P_n. With the personal ability defined in Eq. (1), we calculate the network ability of $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$ for a football team (i.e., the competency of the whole team), which gives

$$\Phi(\mathcal{G}') = \sum_{n=1}^{|\mathcal{V}'|} \phi_{\mathrm{P}_n}, \tag{2}$$

where $|\mathcal{V}'|$ is the number of selected football players in a team. We can see from Eq. (2) that it is the sum of the personal abilities of the selected players, which means that a higher network ability score contributes to forming a better football team.

3.2.2 Network density

As shown in Eq. (2), a naive scheme for building a football team is to identify suitable players with good skills for each position and then put them together. However, the team's victory depends on not only the number of football stars but also the collaboration of the players, enabling them to function as a cohesive team in the field. Intuitively, good collaboration is commonly built upon players' relationships. To establish relationships among football players, in this paper, we consider whether they come from the same team or country, which is often used for organizing players in a club. Formally, let us consider the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$.

Given any two nodes $v_i, v_j \in \mathcal{V}$ associated with two football players P_i and P_j, if they come from the same country, the same club, or both, we add the edge (i, j) to $\mathcal{E}$, and the relationship is weighted by calculating the *Jaccard* similarity, denoted as $\omega_{i,j}$, in Eq. (3).

$$\omega_{i,j} = \frac{|\mathbf{V}_{\mathrm{P}_i} \cap \mathbf{V}_{\mathrm{P}_j}|}{|\mathbf{V}_{\mathrm{P}_i} \cup \mathbf{V}_{\mathrm{P}_j}|}, \tag{3}$$

where $\mathbf{V}_{\mathrm{P}_i}$ is the vector of player P_i with the elements team name and nationality.

Based on the relationships among football players, we now turn to define the network density for measuring team cohesiveness. Although many methods have been used to define a team's cohesion based on social networks, such as the diameter communication cost [28], density-based measurement [29], and local clustering coefficient [32], the definition of a team's cohesiveness is still an open issue. Different from the existing works, we define the network density to measure the strength of inner-team interaction in the subnetwork $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$ for a football team in Eq. (4).

$$\Psi(\mathcal{G}') = \frac{\sum_{(i,j)\in\mathcal{E}'} \omega_{i,j}}{|\mathcal{E}'|}, \tag{4}$$

where (i, j) is an edge in $\mathcal{E}'$, $\omega_{i,j}$ is the corresponding weight defined in Eq. (3), and $|\mathcal{E}'|$ is the number of edges. If there is no edge between two nodes, we set $\omega_{i,j} = 0$. A larger value of $\Psi(\mathcal{G}')$ suggests that football players are better able to interact with each other, while a smaller value indicates the presence of more ambiguous relationships. To better understand the importance of the network density, we give a toy example below.

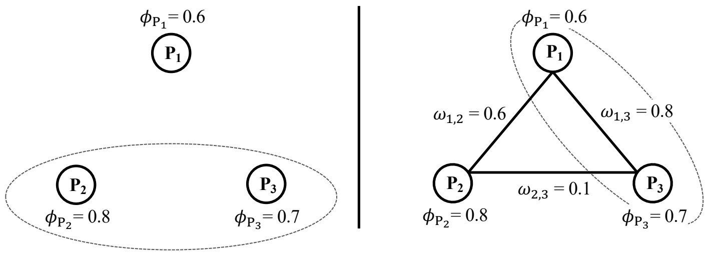


Fig. 2: Two types of networks of three football players. The left-hand side is an edgeless graph, while the graph on the right-hand side shows the connections among players.

Example 1 *Considering the two undirected, weighted graphs in Figure 2, each node denotes a football player, and the edges reflect the relationship between any two* *players. The values of ϕ_{P_i} and $\omega_{i,j}$ are also shown in the figure. If we ignore the collaborative relationships between football players, it is intuitive that the two players $\{P_2, P_3\}$ are highly scored and shall be selected into a team (see the left-hand side of Figure 2); however, their relationship (the right-hand side of Figure 2) is rather weak. In contrast, the players $\{P_1, P_3\}$ would be the better candidates, as they have the strongest connection, which suggests that the connection strength (network density) among players helps to build and reinforce a cohesive team.*

3.2.3 Network heterogeneity & homogeneity

In this section, we proceed to define the network heterogeneity&homogeneity, which is also an important factor for team cohesiveness in the TC-FPACN. It is well known that heterogeneity and homogeneity are opposites, which means that improving heterogeneity may compromise homogeneity and vice versa. Specifically, heterogeneity highlights the diversity of attributes and behaviors among group members; in contrast, homogeneity emphasizes the within-group similarities regarding these shared attributes.

We adopt the *Gini* coefficient [37] to measure heterogeneity (or homogeneity) for the set of players, denoted Gc. Since the *Gini* coefficient can be calculated in many forms [38–40], we use an approximate calculation method [38] as follows:

$$Gc = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{2N^2u}\sum_{i=1}^{N}\sum_{j=1}^{N}W_{\mathrm{S}_m}|L_{\mathrm{P}_i,\mathrm{S}_m} - L_{\mathrm{P}_j,\mathrm{S}_m}|, \tag{5}$$

where $|L_{\mathrm{P}_i,\mathrm{S}_m} - L_{\mathrm{P}_j,\mathrm{S}_m}|$ measures the difference in the skill level related to S_m between two players P_i and P_j, and u is the average value of skill S_m. In Eq. (5), we see that $Gc = 1$ indicates the maximum heterogeneity, while $Gc = 0$ is the maximum homogeneity, which means that they are interdependent [8].

In the context of football games, the two main tasks are attack and defense, and they require different mechanisms to select players to successfully complete the tasks. Attacks on a goal benefit from players who have different skills and require a set of heterogeneous forward players. However, defense requires homogeneous players since it is expected that most defense players can play in any position in the defense area. Considering that *Forward* and *Midfield* players are involved in the attack and *Backward* players are responsible for the defense, based on the *Gini* coefficient defined in Eq. (5), we measure the network heterogeneity&homogeneity for $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$ for a football team as follows:

$$\Upsilon(\mathcal{G}') = \begin{cases} Gc, & \text{if } v_n \in \mathcal{G}' \cap \mathcal{G}_{\mathrm{F}} \\ \frac{1}{Gc}, & \text{if } v_n \in \mathcal{G}' \cap \mathcal{G}_{\mathrm{B}} \end{cases}, \tag{6}$$

where v_n $(1 \leq n \leq |\mathcal{V}'|)$ represents a football player selected from the two graphs (i.e., $\mathcal{G}_{\mathrm{F}}$ and $\mathcal{G}_{\mathrm{B}}$) simultaneously. Eq. (6) shows that a cohesive team should maximize network heterogeneity for the *Forward/Midfielder* while minimizing it for the *Backward* in the team composition.

3.3 Team composition via three network metrics

As mentioned, we delve into three network metrics of the TC-FPACN model that lay the foundation for building a cohesive football team. Considering all these factors, we introduce the trade-off parameters α and β, where $0 \leq \alpha + \beta \leq 1$, which configures acceptable combinations among network ability, network density, and network heterogeneity&homogeneity. Formally, given the attributed collaboration network of football players $\mathcal{G}(\mathcal{V}, \mathcal{E})$ and a fixed budget (Bu) for recruiting players, we use σ to denote the objective function of the TC-FPACN and then formulate the team composition task as solving the following optimization problem.

$$\begin{aligned} \max_{\mathcal{G}' \subseteq \mathcal{G}} \quad & \sigma(\mathcal{G}') \coloneqq \alpha\, \Phi(\mathcal{G}') + \beta\, \Psi(\mathcal{G}') + (1-\alpha-\beta)\, \Upsilon(\mathcal{G}'), \\ \text{s.t.} \quad & \textstyle\sum_{n=1}^{|\mathcal{V}'|} Cost(\mathrm{P}_n) \leq \mathrm{Bu}, \\ & |\mathcal{V}'| = 11, \end{aligned} \tag{7}$$

where $\sum_{n=1}^{|\mathcal{V}'|} Cost(\mathrm{P}_n)$ denotes the total cost of the football team, in which the function $Cost(\mathrm{P}_n)$ measures the cost of player P_n based on his personal rating, which we will explain in Section 5.1.

As shown in problem (7), the goal of TC-FPACN is to find a subgraph $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$ containing a set of football players that maximize the function considering the three metrics simultaneously. The subgraph $\mathcal{G}'$ contains players for three types of positions in Eq. (8):

$$\mathcal{G}' = \mathcal{G}'_{\mathrm{F}} \cup \mathcal{G}'_{\mathrm{B}} \cup \mathcal{G}'_{\mathrm{G}}, \tag{8}$$

where $\mathcal{G}'_{\mathrm{F}} \subseteq \mathcal{G}_{\mathrm{F}}$, $\mathcal{G}'_{\mathrm{B}} \subseteq \mathcal{G}_{\mathrm{B}}$, and $\mathcal{G}'_{\mathrm{G}} \subseteq \mathcal{G}_{\mathrm{G}}$. Note that we focus on choosing suitable players in the field and neglect bench players, which means that the number of nodes in $\mathcal{G}'$ is 11 (i.e., $|\mathcal{V}'| = 11$), and $\mathcal{G}'_{\mathrm{G}}$ contains one goalkeeper.

4 Optimization method based on greedy algorithm

Given that finding the optimal subgraph based on the optimization function of problem (7) is NP-hard [11], we develop a greedy algorithm to solve the aforementioned team composition problem. We consider a team with a 4-3-3 formation, which is widely-used in international competition. This formation means that there is one goalkeeper, four guards, three midfielders and three forwards on a team. We first leave out the goalkeeper and develop two algorithms to find the best players from *Forward/Midfielder* (i.e., $\mathcal{G}_{\mathrm{F}}$) and *Backward* (i.e., $\mathcal{G}_{\mathrm{B}}$), respectively. Next, we propose a pruning technique to organize the final football team.

Algorithm 1 Finding *Forward/Midfielder* based on a greedy algorithm

Input: $\mathcal{G}_{\mathrm{F}}(\mathcal{V}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}})$, N_{F}, α, β
Output: $\mathcal{G}'_{\mathrm{F}}$
1: Initialize $\mathcal{G}'_{\mathrm{F}} = \emptyset$;
2: **for** $i \leftarrow 1$ to $|\mathcal{V}_{\mathrm{F}}|$ **do**
3:     Record $\mathcal{G}_{\mathrm{F}}^{v_i} \subseteq \mathcal{G}_{\mathrm{F}}$, which consists of v_i and its neighbors;
4: **end for**
5: $v^c \leftarrow \underset{v_n \in \mathcal{V}_{\mathrm{F}}}{\arg\max}\ \alpha\, \phi_{\mathrm{P}_n} + \beta\, \Psi(\mathcal{G}_{\mathrm{F}}^{v_n})$;
6: $\mathcal{G}'_{\mathrm{F}} \leftarrow \mathcal{G}'_{\mathrm{F}} \cup \{v^c\}$;
7: **for** $k \leftarrow 1$ to N_{F} **do**
8:     Find the neighbors $\mathbf{NB}(\mathcal{G}'_{\mathrm{F}})$;
9:     $v^* \leftarrow \underset{v_n \in \mathbf{NB}(\mathcal{G}'_{\mathrm{F}})}{\arg\max}\ \sigma(\mathcal{G}'_{\mathrm{F}} \cup \{v_n\})$;
10:     $\mathcal{G}'_{\mathrm{F}} \leftarrow \mathcal{G}'_{\mathrm{F}} \cup \{v^*\}$;
11:     $\mathcal{V}_{\mathrm{F}} \leftarrow \mathcal{V}_{\mathrm{F}} \setminus v^*$;
12:     $k \leftarrow k + 1$;
13: **end for**
14: **return** $\mathcal{G}'_{\mathrm{F}}$

We show the process to find the best *Forward/Midfielder* players in Algorithm 1. For brevity, we omit the pseudocode for finding the best *Backward* players because the two algorithms differ only in the input: the former selects players from $\mathcal{G}_{\mathrm{F}}$, while the latter chooses players from $\mathcal{G}_{\mathrm{B}}$. As shown in Algorithm 1, we start with an empty graph (line 1), which poses a difficulty to the direct application of the three network metrics; therefore, we need to choose the starting football player. In this paper, we consider a key player with a good trade-off between personal ability and connections to other players. Specifically, for each player, we first extract the subnetwork that consists of the player and the player's neighbors (lines 2-4), and then determine the key player (denoted v^c) that maximizes both personal ability and network density (lines 5-6). The algorithm then proceeds through multiple iterations (lines 7-13). In each loop, the algorithm adds the most suitable player v^* in $\mathcal{G}_{\mathrm{F}}$, who maximizes the value of the objective function of problem (7) (lines 8-10). Note that we remove the player who is selected from $\mathcal{V}_{\mathrm{F}}$ at the end of each iteration, which avoids the same players being selected into the team (line 11). Finally, once the total number of players reaches the size requirement, the algorithm returns the final subgraph $\mathcal{G}'_{\mathrm{F}}$ (line 14).

The results from the algorithms above are used as inputs for the final team composition. Since we need to ensure that the total cost of a team does not exceed the budget, we add a pruning strategy to the greedy algorithm. We propose the idea of cost performance, denoted Cp, as a measurement to decide which player must be cut if the total cost exceeds the given budget. Specifically, for a football player P_n, the corresponding cost performance Cp is computed in Eq. (9).

$$Cp(\mathrm{P}_n) = \frac{\phi_{\mathrm{P}_n}}{Cost(\mathrm{P}_n)}. \tag{9}$$

Algorithm 2 Finding the Best Team with Pruning (FBTP)

Input: $\mathcal{G}(\mathcal{V}, \mathcal{E})$, $\mathcal{G}'_{\mathrm{F}}(\mathcal{V}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}})$, $\mathcal{G}'_{\mathrm{B}}(\mathcal{V}_{\mathrm{B}}, \mathcal{E}_{\mathrm{B}})$, α, β, Bu
Output: $\mathcal{G}'$

1: $v^g \leftarrow \arg\max_{v_i \in \mathcal{V}_{\mathrm{G}}} \ \phi_{\mathrm{P}_i}$;
2: $\mathcal{G}'(\mathcal{V}', \mathcal{E}') \leftarrow \mathcal{G}'_{\mathrm{F}} \cup \mathcal{G}'_{\mathrm{B}} \cup \{v^g\}$;
3: **while** $\sum_{i=1}^{|\mathcal{V}'|} Cost(\mathrm{P}_i) > \mathrm{Bu}$ **do**
4: $\quad v^{cut} \leftarrow \arg\min_{v_j \in \mathcal{V}'} \ Cp(\mathrm{P}_j)$;
5: $\quad \mathcal{V}' \leftarrow \mathcal{V}' \setminus v^{cut}$;
6: $\quad$ Find the candidate v^* according to the position of v^{cut};
7: $\quad \mathcal{G}' \leftarrow \mathcal{G}' \cup \{v^*\}$;
8: **end while**
9: **return** $\mathcal{G}'$

We frame the new approach for solving the objective function of the TC-FPACN in problem (7) as the FBTP (Finding the Best Team with Pruning) algorithm presented in Algorithm 2. We first find the best goalkeeper (line 1); and the best team under no budget constraint consists of $\mathcal{G}'_{\mathrm{F}}$, $\mathcal{G}'_{\mathrm{B}}$ and the selected goalkeeper (line 2). The pruning operations are embedded in the greedy algorithm (lines 3-8). Specifically, we use a loop to check whether the total cost of the football team exceeds the budget. If the cost does not satisfy the budget requirement, we perform a pruning strategy that determine the football player v^{cut} with the lowest cost performance (line 4) and remove v^{cut} from the football team $\mathcal{G}'$ (line 5). Next, we choose the other suitable candidate according to the position of v^{cut} (lines 6-7) based on the greedy algorithm. For example, if the position of v^{cut} belongs to *Forward/Midfielder*, we execute the procedures in lines 8-11 of Algorithm 1 to select v^*.

To better illustrate the workflow for constructing a football team based on the algorithms mentioned above, we provide a vivid example in Figure 3, which illustrates the process of finding five football players from *Forward/Midfielder*. We first focus on choosing players without the budget constraint (see the left-hand side of the figure). We start with the key player S and proceed to find the most suitable forward (or midfielder) in each iteration through Algorithm 1. For instance, in step 1, we tend to choose the football player A that maximizes the objective function of problem (7). We return the final selection result (i.e., $\{S, A, B, C, E\}$) in step 4, as the number of players is full. Since the selected players do not consider a proposed budget, on the right-hand side of Figure 3, armed with Algorithm 2, we proceed to conduct the pruning operation by removing the player with the lowest cost performance and then find another

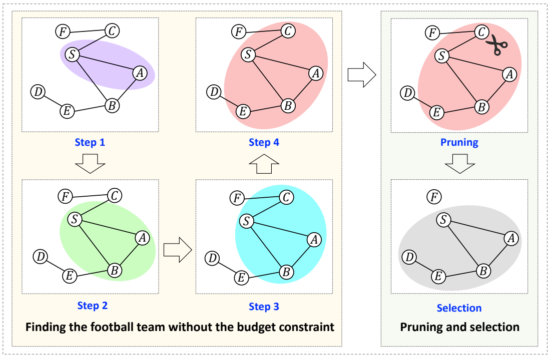


Fig. 3: An example of the process for finding five football players with a budget constraint.

football player, i.e., we remove C and add D. For example, we output the candidate set, $\{S, A, B, E, D\}$ if the total package is no larger than the budget; otherwise, the pruning and selection processes are repeated until the budget requirement is satisfied.

5 Empirical Study: Data Analysis and Team Evaluation

Given the discussions in Section 1, it is difficult to form a series of football teams in the real world to evaluate the performance of the proposed model. Fortunately, football video games provide a convenient and quick way to assess the effectiveness of our model. In this paper, we implement and test our method on the two most popular game platforms (i.e., PES2018 and FIFA2022). Figure 4 shows screenshots of the two platforms; both are classical and full-fledged platforms that not only are equipped with well-simulated football players in real life but also provide hours of entertainment in multiplayer mode, including simulating a football match. We conduct a series of experiments with the quick games of PES2018 and FIFA2022 based on a Windows PC. All the codes are implemented in Python, and the numerical computations are conducted on a server with a 12-core Intel(R) Xeon(R) CPU E5-2620 v3 @2.40 GHz and 16 GB memory. The source code of our method is publicly available at https://github.com/misterbobo/TCFPACN.

5.1 Data analysis

Since the values of many attributes of the team composition are calculated from game data, we first analyze the original data from PES and FIFA and preprocess the data [6]. In PES2018, we retrieve the data that contain $9,563$ football players;

[6]The two datasets we use are publicly available on https://github.com/misterbobo/TCFPACN/tree/main/Data.

(a) PES2018 (b) FIFA2022

Fig. 4: Football game interfaces of the game platforms. The left-hand side is the playing field of PES2018, and the right-hand side shows the user interface of FIFA2022.

we also collect FIFA2022 data, which includes data on 18,278 players from the official website [7]. Table 1 provides a brief overview of the two datasets, both of which list player IDs, positions, and names, as well as descriptions of each player's skills, such as a player's attacking prowess in Table 1a.

Table 1: The structure of the original data in two game platforms

(a) PES2018

Player ID	Player Position	Player Name	Team Name	Nationality	...	Rating	Attacking Prowess	...
1	LWF[1]	C. Ronaldo	MD WHITE	Portugal	...	94	94	...
2	RWF	L. Messi	FC BARCELONA	Argentina	...	94	95	...
3	CF	L. Suárez	FC BARCELONA	Uruguay	...	92	95	...
...	...	...	...	...	...	...	...	...

[1] LWF is short for *Left Wing Forward*. For more detial about the abbreviation of positions, please see https://en.wikipedia.org/wiki/Association_football_positions or Appendix A, Table A3.

(b) FIFA2022

Player ID	Player Name	Nationality	Club	Team Position	...	Overall	Heading Accuracy	...
158023	L. Messi	Argentina	PARIS SAINT-GERMAIN	RW	...	92	70	...
20801	C. Ronaldo	Portugal	MANCHESTER UNITED	ST	...	91	90	...
190871	Neymar JR	Brazil	PARIS SAINT-GERMAIN	LW	...	90	63	...
...	...	...	...	...	...	...	...	...

As seen from Table 1, a player serves in a particular position in a football team. It is also clear that each position has different skill requirements. Consider an example in Table 1a, the skill of *attacking prowess* is crucial for a *Forward* player, while it has no relevance for a goalkeeper. Table 2 shows the assessments of 23 skills for some well-known players in PES2018. The numerical values reflect each player's performance on each skill. As seen from Table 2, it is necessary to link the skills to distinct positions.

For each dataset, we first divide the raw data into three groups (i.e., *Forward/Midfielder*, *Backward*, and *Goalkeeper*) according to each player's position on the pitch. For each group, we rank the skills based on the average

[7] https://sofifa.com/

Table 2: A sample of four players' assessment by 23 criteria in PES2018

Player ID Player Name Player Position		1 C. Ronaldo LWF	4 M. Neuer GK	8 R. Lewandowski CF	17 Sergio Ramos CB
	attacking prowess	94	42	93	65
	ball control	91	68	89	75
	dribbling	86	60	87	66
	low pass	83	65	79	73
	lofted pass	83	69	68	76
	finishing	95	43	92	62
	place kicking	75	65	65	66
	swerve	82	54	70	67
	header	94	70	85	94
	defensive prowess	49	60	53	88
	ball winning	57	41	50	86
Criteria	**kicking power**	94	75	87	75
	speed	89	71	81	78
	explosive power	86	68	81	77
	body control	79	70	79	66
	physical contact	87	88	84	84
	jump	98	83	84	95
	stamina	89	65	79	86
	goalkeeping	40	98	40	40
	catching	40	97	40	40
	clearing	40	98	40	40
	reflexes	40	97	40	40
	coverage	40	96	40	40

values and select the top-10 skill values presented in Table 3. Notably, we ignore the criteria for goalkeepers in Table 3 because both datasets have only a few skills that are relevant to goalkeepers; hence, we include all of them. The weight of each skill is assigned following the principle mentioned in [3], as provided in the last column of Table 3.

The main goal of this paper is to form a cohesive team with a budget constraint (see problem (7)). Therefore, it is necessary to know the salary for each football player. However, there are many football players with missing salaries in both PES and FIFA datasets. It is known that a player's cost is positively correlated with his rating, which is a good indicator. Here, given a football player P_n, we use the fitting function mentioned in [5] to evaluate his cost as follows, which can be used to formulate the total team cost.

$$Cost(\mathrm{P}_n) = \eta\, e^{\theta\, r(\mathrm{P}_n)},$$

where $\eta = 6.375 \times 10^{-4}$, $\theta = 0.1029$. In addition, $r(\mathrm{P}_n)$ denotes the rating (or overall) of P_n (see Table 1).

5.2 Performance metrics and parameter settings

To understand game results intuitively, we use *goal difference* (GD) and *team points* (Tps), which are the general rules in international competitions, as the metrics to evaluate team performance. Specifically, given a competition set $\Delta = \{\Delta_1, \Delta_2, \cdots, \Delta_Z\}$, where Δ_z $(1 \le z \le Z)$ represents a match and Z is the total number of matches, the value of GD for a football team is calculated as

Table 3: The criteria for Forward/Midfielder and Backward

Platform	Group	Criteria	Average Value	Weight
PES2018	**Forward/Midfielder**	explosive power	74.2812	8
		speed	74.0443	10
		stamina	73.6505	6
		kicking power	72.2510	6
		ball control	72.1408	10
		dribbling	71.4847	8
		low pass	70.5880	6
		physical contact	70.5880	4
		body control	70.2905	8
		attacking prowess	70.1156	10
	Backward	physical contact	74.2202	10
		stamina	73.9685	6
		jump	72.6409	10
		speed	72.5572	4
		ball winning	71.8546	10
		defensive prowess	71.5585	10
		explosive power	71.4511	6
		header	70.5864	10
		kicking power	69.8391	6
		low pass	67.6840	6
FIFA2022	**Forward/Midfielder**	movement agility	69.7926	8
		movement acceleration	69.6549	8
		movement balance	69.2812	6
		movement sprint speed	69.2538	8
		power stamina	66.6179	6
		skill ball control	66.2300	10
		skill dribbling	65.1919	10
		attacking short passing	64.4739	10
		power jumping	64.0808	4
		power shot power	64.0674	6
	Backward	power strength	70.6532	10
		power jumping	69.3325	4
		power stamina	67.7168	8
		defending standing tackle	66.5593	10
		mentality aggression	65.4470	4
		defending sliding tackle	64.8977	10
		movement sprint speed	64.8364	8
		defending marking	63.7573	6
		movement acceleration	63.7514	6
		mentality interceptions	63.7312	10

the number of goals scored in all matches minus the number of goals conceded, which gives

$$\mathrm{GD} = \sum_{z=1}^{Z} \delta_s(\Delta_z) - \delta_c(\Delta_z),$$

where $\delta_s(\Delta_z)$ and $\delta_c(\Delta_z)$ are the number of goals scored and conceded in one match, respectively. Tps denotes the total match scores of a team, as shown in Eq. (10).

$$\mathrm{Tps} = \sum_{z=1}^{Z} \mathrm{Tp}(\Delta_z), \tag{10}$$

where $\mathrm{Tp}(\Delta_z)$ is a team point for one match, which gives

$$\mathrm{Tp}(\Delta_z) = \begin{cases} 3, & \text{if win} \\ 0, & \text{if draw} \\ -1, & \text{if lose} \end{cases}.$$

Unless stated otherwise, we set the number of matches $Z = 30$ and set $\mathrm{Bu} = 100$ to simulate the unconstrained budget case. In addition, for PES2018, we use the FBTP algorithm with the settings $\alpha = 0.6$ and $\beta = 0.2$ in the TC-FPACN to solve the optimization problem (denoted TC-FPACN+FBTP); similarly, we set $\alpha = 0.4$ and $\beta = 0.4$ for FIFA2022. We further present a sensitivity analysis of parameters α and β based on our new evaluation strategy in Section 5.5.

5.3 Simulation results

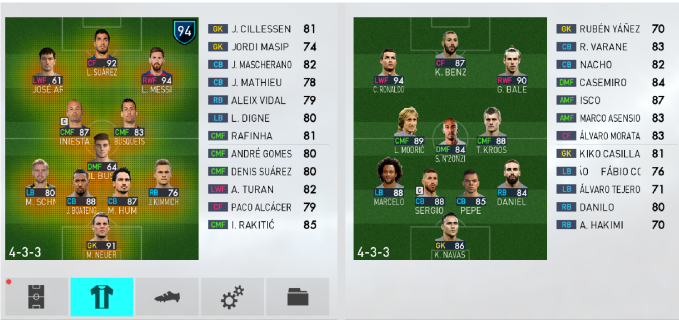
As the team budget has a large impact on team composition, we investigate the capability of the TC-FPACN+FBTP to deal with different team composition scenarios (i.e., with or without the budget constraint).

Table 4: Selected football players in PES2018 and FIFA2022 with $\mathrm{Bu} = 100$

Platform	Group	Player Name	Player Position	Cost
PES2018	**Forward/Midfielder**	Lionel Messi	RWF	52.48
		Luis Suárez	CF	
		Iniesta	CMF	
		Sergio Busquets	DMF	
		Oriol Busquets	DMF	
		José Arnáiz	LWF	
	Backward	Jérôme Boateng	CB	
		Mats Hummels	CB	
		Joshua Walter Kimmich	RB	
		Marcel Schmelzer	LB	
	Goalkeeper	Manuel Peter Neuer	GK	
FIFA2022	**Forward/Midfielder**	Raheem Sterling	LW	34.95
		Gabriel Jesus	ST	
		Bernardo Silva	RW	
		Fernando Luiz Rosa	DMF	
		Felix Nmecha	CAM	
		Philip Foden	CM	
	Backward	Kyle Walker	RB	
		Luke Shaw	LB	
		Fikayo Tomori	CB	
		Jamaal Lascelles	CB	
	Goalkeeper	Ederson Santana de Moraes	GK	

5.3.1 Team performance without a budget constraint

In this subsection, we conduct experiments to show the effectiveness of the team generated by the TC-FPACN+FBTP that ignores the budget constraint. We show our team formation results in Table 4. Based on the recommended players, we compose our DREAM TEAM in PES2018 and FIFA2022, denoted DT-PES and DT-FIFA, respectively (see the left-hand side of Figure 5a and Figure 5b). To conduct a performance comparison and ensure the fairness of competitions, we select a team in PES2018 with a cost approximately equal to DT-PES, namely, MD WHITE [8] (the right-hand side of Figure 5a), which is one of the most competitive teams in the game. In FIFA2022, we choose MANCHESTER UNITED [9] as the competitor (pictured on the right in Figure 5b), which not only has a similar cost to DT-FIFA but also has the leading record in its football league.



(a) The players of DT-PES (left) and MD WHITE (right).

(b) The players of DT-FIFA (left) and MANCHESTER UNITED (right).

Fig. 5: Recommended players to compose DT-PES *v.s.* MD WHITE in PES2018 and DT-FIFA *v.s.* MANCHESTER UNITED in FIFA2022.

[8] https://www.realmadrid.com/
[9] https://www.manutd.com/

Table 5 shows the battle results on the two game platforms, including the scoreline of each match, the total cost, Tps, and GD. A close inspection of the match results in the table shows that DT-PES wins more matches than MD WHITE in PES2018, and DT-FIFA achieves good performance than MANCHESTER UNITED in FIFA2022. Moreover, the cost of our team is slightly smaller than that of MD WHITE (or MANCHESTER UNITED). It is clear that whichever platform we use, our team dominates through the 30-race series, which highlights the effectiveness of the proposed model.

Table 5: Match results for DT-PES *v.s.* MD WHITE and DT-FIFA *v.s.* MANCHESTER UNITED without a budget constraint

Platform	Battle	Game Results										Win	Draw	Lose	Cost	Tps	GD
PES2018	**DT-PES**	3:3	1:3	5:5	1:0	2:0	0:2	1:2	4:1	8:2	0:0	16	7	7	52.48	41	29
	v.s.	2:0	3:1	4:0	6:2	3:1	4:1	2:2	2:3	4:2	5:3	:	:	:	:	:	:
	MD WHITE	2:2	6:4	2:0	2:0	2:4	2:4	1:1	3:5	3:1	0:0	7	7	16	58.58	5	-29
FIFA2022	**DT-FIFA**	2:1	1:0	0:2	1:1	2:2	2:2	2:2	2:0	0:0	2:0	12	13	5	34.95	31	11
	v.s.	1:1	2:2	2:1	2:1	2:0	0:2	0:1	0:0	2:1	2:1	:	:	:	:	:	:
	MANCHESTER UNITED	2:2	3:0	0:1	1:1	2:0	0:0	3:1	1:1	0:2	1:1	5	13	12	37.01	3	-11

To demonstrate the strength of our team, we simulate matches in which random teams battle with MD WHITE and MANCHESTER UNITED, respectively. There are two ways to generate a random team. Given the total cost of MD WHITE (or MANCHESTER UNITED) as the budget constraint, one way is to pick a player for each position randomly based on the average budget, while the other way is first to pick a few players that consume most of the budget and then select other players based on the remaining budget. We name the resulting teams RAND 1 and RAND 2, respectively in PES2018, and RAND 3 and RAND 4, respectively in FIFA2022. The simulated results are shown in Table 6. From the perspective of Tps and GD, we find that our teams perform better than all the random teams when competing against MD WHITE in PES2018 and MANCHESTER UNITED in FIFA2022.

Table 6: Match results for random teams against MD WHITE in PES2018 and MANCHESTER UNITED in FIFA2022

Platform	Battle	Game Results										Win	Draw	Lose	Cost	Tps	GD
PES2018	**RAND 1**	4:1	6:1	4:2	5:6	2:0	2:0	0:3	4:5	1:0	2:0	12	2	16	50.46	20	-6
	v.s.	1:1	1:3	0:2	1:2	0:2	0:2	0:4	1:7	2:0	6:0	:	:	:	:	:	:
	MD WHITE	2:2	3:2	1:0	3:0	1:3	2:3	0:1	1:2	1:4	1:5	16	2	12	58.58	36	6
	RAND 2	2:0	4:0	0:0	1:3	1:0	1:3	1:0	5:0	0:1	0:4	8	6	16	52.44	8	-19
	v.s.	0:3	0:5	1:0	1:2	2:4	5:5	2:2	4:5	0:0	0:4	:	:	:	:	:	:
	MD WHITE	1:1	1:1	0:1	1:2	1:3	2:3	1:0	3:0	1:3	1:6	16	6	8	58.58	40	19
FIFA2022	**RAND 3**	1:3	2:2	0:3	2:2	2:3	1:2	1:2	0:1	1:3	3:1	3	7	20	33.98	-11	-28
	v.s.	1:2	1:3	1:2	1:2	1:3	2:2	2:2	1:2	0:3	2:2	:	:	:	:	:	:
	MANCHESTER UNITED	2:3	2:1	1:2	2:3	1:1	0:5	2:1	2:2	2:3	2:3	20	7	3	37.01	57	28
	RAND 4	0:4	0:1	2:2	1:2	2:4	1:3	0:3	1:3	1:2	2:1	3	6	21	34.72	-12	-38
	v.s.	1:2	3:2	1:4	2:4	1:2	0:1	1:2	1:3	2:3	0:3	:	:	:	:	:	:
	MANCHESTER UNITED	2:2	1:3	1:1	2:2	2:4	0:4	1:4	4:2	2:2	1:1	21	6	3	37.01	60	38

5.3.2 Team performance considering different budget constraints

It is common for football player recruitment to be constrained by a budget crunch. In this subsection, we discuss the performance of the TC-FPACN+FBTP by adjusting the budget constraint. In PES2018, since MD WHITE is one of the best teams with the highest cost burden, we use its cost as the budget limit (denoted as Bu_{hi}), and set the budget change from Bu_{lo} to Bu_{hi}, where $\mathrm{Bu}_{lo} = 10$ and $\mathrm{Bu}_{hi} = 60$. Similarly, in FIFA2022, we set $\mathrm{Bu}_{hi} = 40$, whose value is close to the cost of MANCHESTER UNITED, and $\mathrm{Bu}_{lo} = 0$. We define the budget levels in Table 7.

Table 7: The budget levels and the corresponding range of values

Budget Level	The Range of Budget Value	
	PES2018	FIFA2022
Level I	(10, 20)	(0, 8)
Level II	[20, 30)	[8, 16)
Level III	[30, 40)	[16, 24)
Level IV	[40, 50)	[24, 32)
Level V	[50, 60]	[32, 40]

Table 8: Match results under different budget constraints

(a) PES2018

Battle	Game Results										Win	Draw	Lose	Cost	Tps	GD	BudgeLevel
DT-PES	2:0	1:1	2:0	4:0	1:0	2:0	0:2	1:1	1:0	2:0	14	11	5	13.32	37	18	
v.s.	0:0	0:1	1:0	1:0	1:0	0:0	0:1	3:0	0:0	0:0	:	:	:	:	:	:	LevelI
AS RED WHITE	0:0	0:0	0:0	0:0	0:0	0:1	2:0	1:0	0:1	1:0	5	11	14	13.87	1	-18	
DT-PES	0:4	1:0	3:1	1:1	3:0	2:0	3:1	2:0	1:0	0:0	21	5	4	22.63	59	27	
v.s.	3:1	0:2	1:0	1:0	1:0	1:0	1:0	1:0	0:0	0:0	:	:	:	:	:	:	LevelII
VALENCIA	3:0	2:0	4:1	1:0	0:1	0:0	1:0	4:0	4:1	0:4	4	5	21	23.00	-9	-27	
DT-PES	5:0	4:0	4:2	3:0	8:0	3:0	2:0	5:0	2:1	4:0	20	8	2	39.34	58	53	
v.s.	0:0	0:0	1:1	1:0	3:0	2:0	1:1	0:0	1:0	2:0	:	:	:	:	:	:	LevelIII
LONDON FC	0:0	0:0	0:1	0:0	3:0	0:1	1:0	1:0	2:1	3:0	2	8	20	42.01	-14	-53	
DT-PES	3:0	3:0	2:0	2:0	1:1	2:0	0:0	1:0	1:1	2:0	16	12	2	43.14	46	24	
v.s.	2:0	1:0	2:1	1:0	0:0	1:0	0:0	2:0	1:1	2:0	:	:	:	:	:	:	LevelIV
PM BLACK WHITE	0:0	0:0	0:0	0:0	0:0	0:2	1:0	0:0	1:0	0:1	2	12	16	49.39	-10	-24	

(b) FIFA2022

Battle	Game Results										Win	Draw	Lose	Cost	Tps	GD	BudgeLevel
DT-FIFA	1:1	1:1	2:2	0:1	1:1	3:0	1:2	1:2	1:1	2:0	11	11	8	7.89	25	10	
v.s.	0:1	0:0	1:0	1:2	1:1	1:0	0:1	3:1	2:2	1:0	:	:	:	:	:	:	Level I
CD TONDELA	1:1	2:2	2:0	2:1	2:0	0:2	1:2	3:0	1:1	2:1	8	11	11	7.95	13	-10	
DT-FIFA	2:0	3:1	1:0	2:2	1:3	1:1	3:0	1:2	2:2	1:2	14	7	9	15.81	33	7	
v.s.	2:1	2:1	0:1	2:0	1:0	1:2	0:2	1:0	2:1	1:2	:	:	:	:	:	:	Level II
FC NANTES	2:1	2:1	2:3	1:1	0:2	2:1	2:1	1:1	2:2	1:1	9	7	14	15.82	13	-7	
DT-FIFA	1:3	2:1	2:3	2:1	0:2	1:0	1:2	2:2	1:2	1:0	12	7	11	22.72	25	-2	
v.s.	0:2	3:2	2:0	2:3	1:3	4:1	2:4	0:2	0:4	2:1	:	:	:	:	:	:	Level III
REAL SOCIEDAD	2:2	2:1	1:1	3:1	1:1	2:1	3:0	2:2	2:2	2:2	11	7	12	23.82	21	2	
DT-FIFA	0:2	1:2	1:1	2:0	2:2	2:2	0:1	1:2	2:1	1:2	13	7	10	28.92	29	10	
v.s.	2:0	1:1	1:2	1:1	2:2	3:1	1:0	2:1	1:2	1:0	:	:	:	:	:	:	Level IV
AC MILAN	2:3	3:1	0:2	2:2	2:0	3:0	3:0	1:2	3:2	2:1	10	7	13	27.73	17	-10	

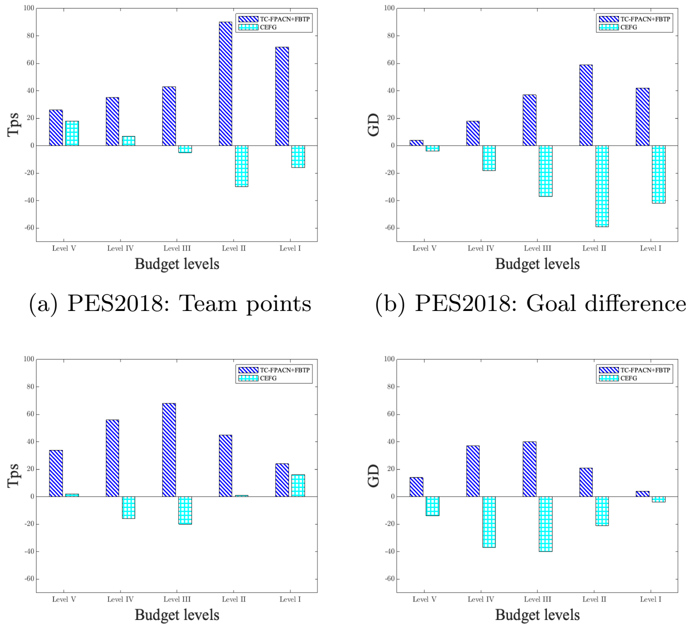
Since Table 5 shows the outcomes of the simulated matches against MD WHITE and MANCHESTER UNITED, both of which have a cost of Level V, we select only four typical teams on each game platform whose costs fall within Level I to Level IV. Specifically, in PES2018, we choose AS RED WHITE, VALENCIA, LONDON FC, and PM BLACK WHITE; in FIFA2022, the four teams are CD TONDELA, FC NANTES, REAL SOCIEDAD, and AC MILAN. For each competitor, we use the corresponding budget level as the constraint to select football players who constitute the DREAM TEAM based on the TC-FPACN+FBTP. We show the match results in Table 8. As shown in the table, all eight teams generated by the TC-FPACN+FBTP are more successful at winning events in terms of Tps. In addition, except for losing two goals when playing a 30-game series against REAL SOCIEDAD in FIFA2022, the remaining teams formed with our method still win the series with the superior goal difference. The match results suggest that the proposed method can assemble a team that wins nearly all the competitions given a budget level.

5.4 Method comparisons

In this subsection, we compare the TC-FPACN+FBTP with other approaches from two aspects. We first compare the TC-FPACN+FBTP with the other football team composition method, namely, CEFG (Cost-Effective Forward selection Greedy) [5]. Next, we discuss the performance of the search strategy based on the random walk algorithm (RW) [41], which is widely used in many areas (e.g., recommender systems [42, 43], community detection [44, 45], and sampling algorithms [46]) for solving the constrained optimization problem (7), denoted TC-FPACN+RW.

5.4.1 Comparison with the CEFG method

We first compare the team composition quality of our method with the CEFG. We again use the PES2018 and FIFA2022 game platforms and focus on the Tps and GD of the two methods for different budget levels. For a given budget constraint, we first generate two football teams on a platform with the TC-FPACN+FBTP and CEFG and then simulate 30 matches between the two teams. Figure 6 compares the simulation results, from which we can conclude that the team generated by the TC-FPACN+FBTP dominates the play on the football pitch. In addition, the data in all four figures shows that the Tps (or GD) increases first and then decreases with the increase in the budget level, and the numerical value reaches a peak at Level II in PES2018 and Level III in FIFA2022. Interestingly, a closer observation reveals that the cost of our team at Level II in PES2018 is approximately equivalent to that at Level III in FIFA2022. A possible reason for the disappointing performance of CEFG is that the team recommended by the CEFG tends to contain a few superstars, and the remaining players may lack competitiveness, especially at a small budget level, thereby leading to poor match results. However, the TC-FPACN+FBTP is more efficient for building a cohesive team that balances the ability in each position and facilitates collaboration among players. Thus, the results suggest that the proposed method generates reliable and promising performance and is not constrained by the choice of game platform.



(a) PES2018: Team points (b) PES2018: Goal difference

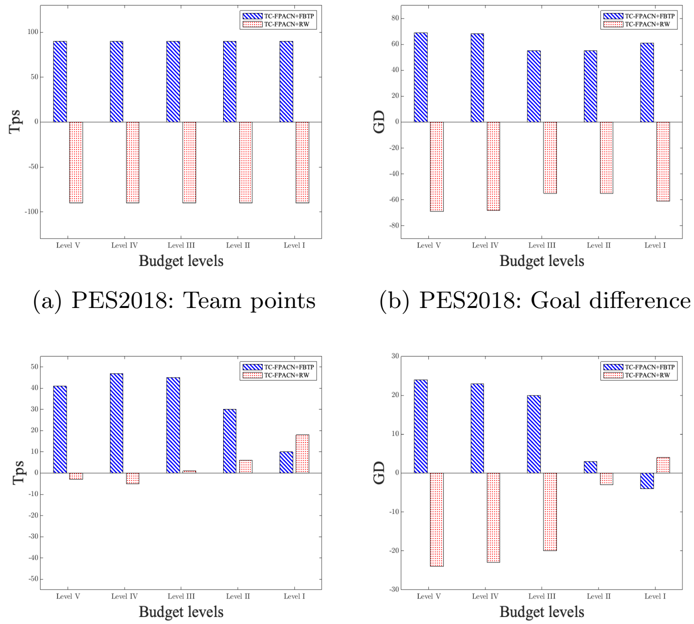
(c) FIFA2022: Team points (d) FIFA2022: Goal difference

Fig. 6: The performance of the TC-FPACN+FBTP and CEFG under different budget levels in terms of Tps and GD.

5.4.2 Comparison with the random walk strategy

In this subsection, we compare the performance of the TC-FPACN+FBTP with the TC-FPACN+RW. Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be the attributed collaboration network of football players. The TC-FPACN+RW begins at a node v_i randomly, and at each step, it moves to another node v_j with a probability proportional to the weight of edge (i, j). We consider the probability (or weight) from v_i to v_j based on the objective function value σ that includes v_j in problem (7), which means that a higher value of σ results in a greater probability of choosing node v_j. The searching process stops if the required number of football players is met, and all the nodes selected in this way form the final football team.

Similar to the process of the simulation match mentioned in Section 5.4.1, we use the team formed by the TC-FPACN+FBTP to compete against the team set up by the TC-FPACN+RW in PES2018 and FIFA2022. The simulation results are compared in Figure 7. Figure 7a and Figure 7b show that the team generated by the proposed method wins all matches against the team produced by the TC-FPACN+RW on the PES2018 platform. In addition, Figure 7c and Figure 7d show that the team formed by the TC-FPACN+FBTP also shows enough dominance to win matches under four budget constraints (i.e., from Level II to Level V). A possible explanation for the results might be that the RW strategy focuses only on neighbors of the current node in the players' network in each searching step, which is easily trapped in a local optimum, thereby compromising the discovery of the most suitable players. Note that at Level I, the value of Tps and GD of the team built via our method is smaller than the team produced by the TC-FPACN+RW (see the rightmost bars in Figure 7c and Figure 7d), which means our team lost most of the matches. This result is likely to be related to the very low budget, which fails to recruit even one competitive football player. Nevertheless, the overall results show the effectiveness of the proposed FBTP searching algorithm.



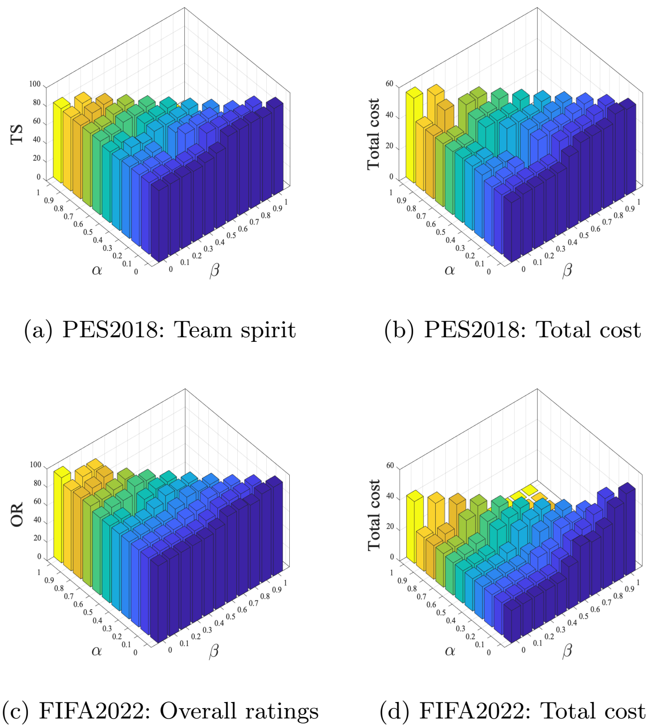
(a) PES2018: Team points (b) PES2018: Goal difference

(c) FIFA2022: Team points (d) FIFA2022: Goal difference

Fig. 7: The performance of the TC-FPACN+FBTP and TC-FPACN+RW under different budget levels in terms of Tps and GD.

5.5 Sensitivity analysis of the parameters

In this subsection, we discuss the parameter sensitivity of the TC-FPACN model, which includes α and β, under no budget constraint. We again use the FBTP algorithm to choose football players. Since it is time-consuming to simulate all matches for different parameter settings, we seek another indicator to evaluate team performance efficiently instead of using Tps and GD. In PES2018, we observe that there is an eye-catching number, namely, *Team Spirit* [10] (TS), when we complete the configuration of a football team (e.g., the upper right corner of the left-hand side in Figure 5a). In fact, TS indicates how good the relationship is on the pitch, and a high TS value could occur in a player who has an affinity for the manager's team instructions, which naturally leads to better teamwork. In FIFA2022, due to the lack of a similar concept to TS, we use the overall rating (OR), which is calculated by first summing the ratings of all football players on a team and then computing the average (e.g., see the player's rating on the left-hand side of Figure 5b). We assume that a higher value of OR indicates better team performance.



(a) PES2018: Team spirit (b) PES2018: Total cost

(c) FIFA2022: Overall ratings (d) FIFA2022: Total cost

Fig. 8: The values of TS (for PES2018) and OR (for FIFA2022) and the corresponding total costs under different settings of α and β.

Armed with TS and OR, as well as the total team cost, we set up the tests to loop through all values of α and β, and the increment of α and β in each iteration

[10] https://www.konami.com/wepes/2018/manual/ps4/en-us/myclub.html

is 0.1. If we select a smaller increment, the evaluation becomes more labor-intensive, and the recommended players do not change much. Figure 8 exhibits the results when tuning α and β. If $\alpha = 0$ and $\beta = 0$, the objective function of problem (7) maximizes the heterogeneity&homogeneity of a team, which results in both poor TS and OR values. Similarly, if $\alpha = 1$ and $\beta = 0$, the function considers only the network ability, which not only leads to a degradation in the TS or OR value but also increases the cost burden. Additionally, there is a slight incline in the values of TS and OR when increasing β. This observation suggests that the network density is an important factor that noticeably benefits the team spirit, and it also demonstrates that football is a team sport. Given the results in Figure 8a and Figure 8b, we can choose appropriate settings for the parameters $\alpha = 0.6$ and $\beta = 0.2$ for PES2018 because they achieve the highest team spirit value while incurring a relatively low cost. For FIFA2022, Figure 8c and Figure 8d show that at the grid point $(\alpha, \beta) = (0.4, 0.4)$, we obtain a good balance of a relatively high OR value and a low total cost; hence, we use this pair of parameters as the tuning result.

6 Conclusions

In this paper, we study the problem of optimizing football team composition in the context of the attributed collaboration network of football players. Since the team's success requires full cooperation between football players, we propose a team scoring function that considers three network metrics, namely, network ability, network density, and network heterogeneity&homogeneity. We then convert the constrained team composition task into the problem of finding an optimal subgraph in the attributed collaboration network. To tackle this problem, we present a novel approach that searches a subgraph by using a greedy algorithm with pruning techniques. We conduct an empirical study of the proposed techniques on two simulated game platforms (PES2018 and FIFA2022). The experimental results show that our method can build a competitive team.

Despite achieving good performance, we have barely scratched the surface of football players' cooperation mechanisms. In particular, the search strategy tends to be trapped in a local optimum in our study. Further work needs to be conducted to investigate sophisticated social factors and delve into how they interact, as well as to explore other search optimization algorithms based on a given budget constraint. In addition, although our new approach aims to determine a football team composition, the investigation of the cooperation factors in this paper can be generalized to solve the team cohesion problem. We will investigate such a generalization and its applications in other problem domains.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61836005 and 62176225 and the Youth Innovation Fund of Xiamen under Grant No. 3502Z20206049.

References

[1] Rajesh, P., Alam, M., Tahernezhadi, M., *et al.*: A data science approach to football team player selection. In: 2020 IEEE International Conference on Electro Information Technology (EIT), pp. 175–183 (2020). https://doi.org/10.1109/EIT48999.2020.9208331

[2] Salles, S.A.F., Hora, H.R.M.d., Erthal, M., Santos, A.C.d.S.G.d., Shimoya, A.: Operations research contributions for football teams formation: a systematic review. Pesquisa Operacional **39**, 277–293 (2019). https://doi.org/10.1590/0101-7438.2019.039.02.0277

[3] Ozceylan, E.: A mathematical model using ahp priorities for soccer player selection: a case study. South African Journal of Industrial Engineering **27**(2), 190–205 (2016). https://doi.org/10.7166/27-2-1265

[4] Abidin, D.: A case study on player selection and team formation in football with machine learning. Turkish Journal of Electrical Engineering and Computer Sciences **29**(3), 1672–1691 (2021). https://doi.org/10.3906/elk-2005-27

[5] Zeng, Y., Shen, G., Chen, B., Tang, J.: Team composition in pes2018 using submodular function optimization. IEEE Access (2019). https://doi.org/10.1109/ACCESS.2019.2919447

[6] Payyappalli, V.M., Zhuang, J.: A data-driven integer programming model for soccer clubs' decision making on player transfers. Environment Systems and Decisions **39**(4), 466–481 (2019). https://doi.org/10.1007/s10669-019-09721-7

[7] Wang, W., Liu, J., Tang, T., Tuarob, S., Xia, F., Gong, Z., King, I.: Attributed collaboration network embedding for academic relationship mining. ACM Transactions on the Web (TWEB) **15**(1), 1–20 (2020). https://doi.org/10.1145/3409736

[8] Gelade, G.A.: The influence of team composition on attacking and defending in football. Journal of Sports Economics **19**(8), 1174–1190 (2018). https://doi.org/10.1177/1527002517716974

[9] Zepp, C., Kleinert, J.: Homogeneity of prototypical attributes in soccer teams. Sage Open **5**(3), 1–10 (2015). https://doi.org/10.1177/2158244015602517

[10] Ingersoll, K., Malesky, E., Saiegh, S.M.: Heterogeneity and team performance: evaluating the effect of cultural diversity in the world's top soccer league. Journal of Sports Analytics **3**(2), 67–92 (2017). https://doi.org/10.3233/JSA-170052

[11] Khuller, S., Saha, B.: On finding dense subgraphs. In: International Colloquium on Automata, Languages, and Programming, pp. 597–608 (2009). https://doi.org/10.1007/978-3-642-02927-1_50

[12] Damian, P., Cristian, P., Dragoş Florin, T.: Considerations regarding the selection in the football game. Ovidius University Annals, Series Physical Education & Sport/Science, Movement & Health **21** (2021)

[13] Inan, T., Cavas, L.: Estimation of market values of football players through artificial neural network: a model study from the turkish super league. Applied Artificial Intelligence **35**(13), 1022–1042 (2021). https://doi.org/10.1080/08839514.2021.1966884

[14] Anamisa, D., Kustiyahningsih, Y., Yusuf, M., Rochman, E., Putro, S., Syakur, M., Bakti, A.: A selection system for the position ideal of football players based on the ahp and topsis methods. In: IOP Conference Series: Materials Science and Engineering, vol. 1125, p. 012044 (2021). https://doi.org/10.1088/1757-899X/1125/1/012044

[15] Nasiri, M.M., Ranjbar, M., Tavana, M., Santos Arteaga, F.J., Yazdanparast, R.: A novel hybrid method for selecting soccer players during the transfer season. Expert Systems **36**(1), 12342 (2019). https://doi.org/10.1111/exsy.12342

[16] Singh, P., Lamba, P.S.: Influence of crowdsourcing, popularity and previous year statistics in market value estimation of football players. Journal of Discrete Mathematical Sciences and Cryptography **22**(2), 113–126 (2019). https://doi.org/10.1080/09720529.2019.1576333

[17] Arrul, V.S., Subramanian, P., Mafas, R.: Predicting the football players' market value using neural network model: a data-driven approach. In: 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), pp. 1–7 (2022). https://doi.org/10.1109/ICDCECE53908.2022.9792681

[18] Vroonen, R., Decroos, T., Van Haaren, J., Davis, J.: Predicting the potential of professional soccer players. In: Proceedings of the 4th Workshop on Machine Learning and Data Mining for Sports Analytics, vol. 1971, pp. 1–10 (2017)

[19] Zhao, H., Chen, H., Yu, S., Chen, B.: Multi-objective optimization for football team member selection. IEEE Access (2021). https://doi.org/10.1109/ACCESS.2021.3091185

[20] Dadelo, S., Turskis, Z., Zavadskas, E.K., Dadeliene, R.: Multi-criteria assessment and ranking system of sport team formation based on objective-measured values of criteria set. Expert Systems with Applications **41**(14),

6106–6113 (2014). https://doi.org/10.1016/j.eswa.2014.03.036

[21] Syaifudin, Y.W., Puspitaningayu, P.: Predicting winner of football match using analytical hierarchy process: an analysis based on previous matches data. In: 2021 International Conference on Data Analytics for Business and Industry (ICDABI), pp. 47–52 (2021). https://doi.org/10.1109/ICDABI53623.2021.9655836

[22] WANG, J.: A novel rugby team player selection method integrating the topsis and ipa methods. International Journal of Sport Psychology **52**, 137–158 (2021). https://doi.org/10.7352/IJSP.2021.52.137

[23] Baharin, N.H., Rashidi, N.F., Mahad, N.F.: Manager selection using fuzzy topsis method. In: Journal of Physics: Conference Series, vol. 1988, p. 012057 (2021). https://doi.org/10.1088/1742-6596/1988/1/012057

[24] Sałabun, W., Shekhovtsov, A., Pamučar, D., Wątróbski, J., Kizielewicz, B., Więckowski, J., Bozanić, D., Urbaniak, K., Nyczaj, B.: A fuzzy inference system for players evaluation in multi-player sports: the football study case. Symmetry **12**(12), 2029 (2020). https://doi.org/10.3390/sym12122029

[25] Liu, W., Xie, X., Ma, S., Wang, Y.: An improved evaluation method for soccer player performance using affective computing. In: 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 324–329 (2020). https://doi.org/10.1109/ICAIBD49809.2020.9137435

[26] Pantzalis, V.C., Tjortjis, C.: Sports analytics for football league table and player performance prediction. In: 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA), pp. 1–8 (2020). https://doi.org/10.1109/IISA50023.2020.9284352

[27] Ghasemian, F., Zamanifar, K., Ghasem-Aghaee, N.: An evolutionary non-linear ranking algorithm for ranking scientific collaborations. Applied Intelligence **48**(2), 465–481 (2018). https://doi.org/10.1007/s10489-017-0990-4

[28] Lappas, T., Liu, K., Terzi, E.: Finding a team of experts in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 467–476 (2009). https://doi.org/10.1145/1557019.1557074

[29] Li, C.-T., Shan, M.-K.: Team formation for generalized tasks in expertise social networks. In: 2010 IEEE Second International Conference on Social Computing, pp. 9–16 (2010). https://doi.org/10.1109/SocialCom.2010.12

[30] Juárez, J., Brizuela, C.A.: A multi-objective formulation of the team formation problem in social networks: preliminary results. In: Proceedings

of the Genetic and Evolutionary Computation Conference, pp. 261–268 (2018). https://doi.org/10.1145/3205455.3205634

[31] Selvarajah, K., Zadeh, P.M., Kobti, Z., Palanichamy, Y., Kargar, M.: A unified framework for effective team formation in social networks. Expert Systems with Applications **177**, 114886 (2021). https://doi.org/10.1016/j.eswa.2021.114886

[32] Datta, A., Tan Teck Yong, J., Ventresque, A.: T-recs: team recommendation system through expertise and cohesiveness. In: Proceedings of the 20th International Conference Companion on World Wide Web, pp. 201–204 (2011). https://doi.org/10.1145/1963192.1963289

[33] Awal, G.K., Bharadwaj, K.K.: Team formation in social networks based on collective intelligence–an evolutionary approach. Applied Intelligence **41**(2), 627–648 (2014). https://doi.org/10.1007/s10489-014-0528-y

[34] Sapienza, A., Goyal, P., Ferrara, E.: Deep neural networks for optimal team composition. Frontiers in Big Data **2**, 14 (2019). https://doi.org/10.3389/fdata.2019.00014

[35] Hamidi Rad, R., Fani, H., Kargar, M., Szlichta, J., Bagheri, E.: Learning to form skill-based teams of experts. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 2049–2052 (2020). https://doi.org/10.1145/3340531.3412140

[36] Hamidi Rad, R., Bagheri, E., Kargar, M., Srivastava, D., Szlichta, J.: Retrieving skill-based teams from collaboration networks. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2015–2019 (2021). https://doi.org/10.1145/3404835.3463105

[37] Gini, C.: Measurement of inequality of incomes. The Economic Journal **31**(121), 124–126 (1921). https://doi.org/10.2307/2223319

[38] Alvaredo, F.: A note on the relationship between top income shares and the gini coefficient. Economics Letters **110**(3), 0–277 (2011). https://doi.org/10.1016/j.econlet.2010.10.008

[39] Deltas, G.: The small-sample bias of the gini coefficient: results and implications for empirical research. Review of Economics and Statistics **85**(1), 226–234 (2003). https://doi.org/10.1162/rest.2003.85.1.226

[40] Sueyoshi, T., Qu, J., Li, A., Liu, X.: A new approach for evaluating technology inequality and diffusion barriers: the concept of efficiency gini coefficient and its application in chinese provinces. Energy **235**, 121256 (2021). https://doi.org/10.1016/j.energy.2021.121256

[41] Lovász, L.: Random walks on graphs. Combinatorics **2**(1-46), 4 (1993). https://doi.org/10.1007/BFb0077189

[42] Feng, S., Zhang, H., Cao, J., Yao, Y.: Merging user social network into the random walk model for better group recommendation. Applied Intelligence **49**(6), 2046–2058 (2019). https://doi.org/10.1007/s10489-018-1375-z

[43] Pradhan, T., Pal, S.: A multi-level fusion based decision support system for academic collaborator recommendation. Knowledge-Based Systems **197**, 105784 (2020). https://doi.org/10.1016/j.knosys.2020.105784

[44] Bahadori, S., Moradi, P., Zare, H.: An improved limited random walk approach for identification of overlapping communities in complex networks. Applied Intelligence **51**(6), 3561–3580 (2021). https://doi.org/10.1007/s10489-020-01999-4

[45] Guo, K., Wang, Q., Lin, J., Wu, L., Guo, W., Chao, K.-M.: Network representation learning based on community-aware and adaptive random walk for overlapping community detection. Applied Intelligence, 1–19 (2022). https://doi.org/10.1007/s10489-021-02999-8

[46] Shao, Y., Huang, S., Miao, X., Cui, B., Chen, L.: Memory-aware framework for efficient second-order random walk on large graphs. In: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, pp. 1797–1812 (2020). https://doi.org/10.1145/3318464.3380562

Appendix A

Table A1: Summary of abbreviations

Abbreviations	Full Name or Descriptions
PES2018	Pro Evolution Soccer 2018
FIFA2022	EA Sports FIFA 22
ACN	Attributed Collaboration Network
FPACN	Football Players' Attributed Collaboration Network
TC-FPACN	Team Composition based on Football Players' Attributed Collaboration Network
MCDM	Multi-Criteria Decision-Making
AHP	Analytic Hierarchic Process
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
Bu	The fixed Budget
DT-PES	DREAM TEAM generated by the proposed method in PES2018
DT-FIFA	DREAM TEAM generated by the proposed method in FIFA2022
FBTP	Finding the Best Team with Pruning
Tps	Team Points
GD	Goal Difference
CEFG	Cost-Effective Forward selection Greedy
RW	Random Walk Algorithm
TS	Team Spirit
OR	Overall Ratings

Table A2: Notations adapted in the paper

Symbol	Description
P_n	a football player
S_m	a skill of football players
$\mathcal{G}(\mathcal{V}, \mathcal{E})$	a graph of football players with a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$
$\mathcal{G}_{\mathrm{F}}$	a graph of the *Forward/Midfielder*

$\mathcal{G}_{\mathrm{B}}$	a graph of the *Backward*
$\mathcal{G}_{\mathrm{G}}$	a graph of the *Goalkeeper*
ϕ_{P_n}	the personal ability of P_n
$\Phi(\mathcal{G}')$	the network ability of $\mathcal{G}'$
$\Psi(\mathcal{G}')$	the network density of $\mathcal{G}'$
$\Upsilon(\mathcal{G}')$	the network heterogeneity&homogeneity of $\mathcal{G}'$
$Cp(\mathrm{P}_n)$	the cost performance of football player P_n

Table A3: The most common positions used in association football

Platform	**Position Abbreviations**	**Full Name**
	SS	Second Striker
	CF	Center Forward
	LWF	Left Wing Forward
	RWF	Right Wing Forward
	AMF	Attacking Midfiedler
PES2018	CMF	Center Midfielde
	DMF	Defensive Midfielder
	LMF	Left Midfielder
	RMF	Right Midfielder
	CB	Center Back
	LB	Left Back
	RB	Right Back
	GK	Goalkeeper
	LS	Left Striker
	LF	Left Forward
	CF	Center Forward
	RF	Right Forward
	RS	Right Striker
	ST	Striker
	LW	Left Winger
	RW	Right Winger
	LAM	Left Attacking Midfielder
	CAM	Center Attacking Midfielder
	RAM	Right Attacking Midfielder
FIFA2022		

LM	Left Midfielder
LCM	Left Central Midfielder
CM	Central Midfielder
RCM	Right Central Midfielder
RM	Right Midfielder
LDM	Left Defensive Midfielder
CDM	Central Defensive Midfielder
RDM	Right Defensive Midfielder
LWB	Left Wing Back
RWB	Right Wing Back
LB	Left Back
LCB	Left Central Back
CB	Center Back
RCB	Right Central Back
RB	Right Back
GK	Goalkeeper