UCL Round of 16 Draw: An Integer Linear Programming Approach

Mohamad Fares El Hajj Chehade

January 2023

1 Introduction

This work attempts to find the fairest draw for the Round of 16 of the UEFA Champions League. If we assume that each team has a certain score, the aim is to minimize the difference in scores between all matches in the round. This ensures no match between two relatively strong teams, which also implies no match between two relatively weak teams.

The two teams facing each other in the Round of 16 must obey the following rules:

- One team must be a group leader, and the other must be a runner-up.
- The teams must belong to different domestic leagues.
- The teams must not belong to the same group.

2 Intuition

Suppose that in some way one can assign a certain score for every team in the Round of 16. This score can be based on their performance in the Champions League and possibly the domestic league. This score may be calculated using a simple formula (like merely the number of points scored in the group stages) or could involve a more sophisticated statistical model. Suppose at this stage that we have this score for every team readily available, regardless of how it was found. In general, the best teams in the league recognized by us, fans, should also have the highest scores, and similarly, the worst teams of the 16 would generally have the lowest scores.

Suppose that the Round of 16 (R16) draw has been performed in the same manner as it has been done over the past years and that it yielded a clash between Real Madrid (with a score, say, of 20) and Liverpool (score: 15). On the other hand, it has placed Benfica (score: 10) with Borussia Dortmund (score:

12). It is obvious that such a draw is not very fair. If we try to sum the scores of the two games separately, we get a score of 35 for the first one and 22 for the second one. The sums are very far from each other. We may conclude that the closer the sums are to each other, the more fair the draw is. In this work, we try to find the combination of matches that yield the closest scores to each other, thus ensuring in some way the fairest draw.

In what follows, the score of the match is used to mean the sum of the scores of the two teams in the match. For the previous example, the score of the first match is 35, and that of the second is 22.

3 Integer Linear Programming Formulation

3.1 Idea

As mentioned in the previous section, we aim to minimize the differences between the scores in the 8 matches. This process of minimization is subject to a set of constraints that include the ones mentioned in the introductory section. This gives rise to an optimization problem that we choose to solve using an Integer Linear Program (ILP). An ILP is convex in nature, and when solvable, would give the optimal solution to the problem.

For a linear objective function, the choice of the difference function is the absolute difference (L1-norm) operator. This operator will use the sum of absolute value errors between the individual scores of each match and the average of these sums. The more this sum of errors is smaller, the closer the individual sums are to the average, and the closer they are to each other. In some sense, we are minimizing a variation of the standard deviation of the scores of each match.

The clear advantage of the sum of absolute errors over the sum of square errors is that the former can be linearized and can consequently be fed into an ILP solver.

3.2 Optimization Problem Formulation

- Let $x_{ij} = \{0,1\}$: A binary variable indicating whether the team that led Group i plays in match j.
- Let $y_{ij} = \{0,1\}$: A binary variable indicating whether the team that was runner-up in Group i plays in match j.
- Let z_j: A positive continuous variable representing the score of match j.
- Let $\bar{z}$: A positive continuous variable representing the average of the scores of the 8 matches.

where $i \in \{1, 2, ..., 8\}$ (number of groups) and $j \in \{1, 2, ...8\}$ (number of matches in R16)

$$\underset{x_{ij}, y_{ij}}{\text{minimize}} \sum_{j=1}^{8} |z_j - \bar{z}| \tag{1}$$

subject to:

- One group leader and one group runner-up in a given match:

$$\forall j \quad \sum_{i=1}^{8} x_{ij} = 1 \tag{2}$$

$$\forall j \quad \sum_{i=1}^{8} y_{ij} = 1 \tag{3}$$

- No two teams from the same group in a given match:

$$\forall j \; \forall i \quad x_{ij} + y_{ij} \leq 1 \tag{4}$$

- No two teams from the same domestic league in a given match:

$$\forall j \quad \forall \, \textit{league} \; l \quad \forall i \in l \quad x_{ij} + y_{ij} \leq 1 \tag{5}$$

- Trivial enough, each team present in one match:

$$\forall i \quad \sum_{j=1}^{8} x_{ij} = 1 \tag{6}$$

$$\forall i \quad \sum_{j=1}^{8} y_{ij} = 1 \tag{7}$$

- The definition of z_j:

$$\forall j \quad z_j = \sum_{i=1}^{8} S_{x_i} x_{ij} + S_{y_i} y_{ij} \tag{8}$$

  where S_{x_i} and S_{y_i} are respectively the scores of the leader and runner-up in Group i.

- The definition of $\bar{z}$:

$$\bar{z} = \frac{1}{8} \sum_{j=1}^{8} z_j \tag{9}$$

In order to linearize the objective function in (1), it is replaced by the following expression:

$$\underset{x_{ij},y_{ij}}{\text{minimize}} \sum_{j=1}^{8} U_j \tag{10}$$

with the addition of the following constraint:

$$\forall j \quad -U_j \leq \bar{z} - z_j \leq U_j \tag{11}$$