Compatibility for Player Substitution Using Pass Distribution Similarity

Heuristic

Two players are considered *substitutable* if their distributions of passes given to and received from all playing positions are sufficiently similar. If these passing patterns align, one player can take the role of another without altering the tactical balance.

Methodology

Step 1: Data Preprocessing

For each player, we extract:

- Most frequently played position

- Total passes **given** to every playing position

- Total passes **received** from every playing position

Step 2: Graph Construction

For each player, we construct a homogeneous directed graph where:

- **Nodes**: All football playing positions

- **Main node**: The player's primary position

- **Outgoing edges**: From the main node to other positions (passes given)

- **Incoming edges**: From other positions to the main node (passes received)

- **Raw edge weights**: Total number of passes given or received

- **Normalized weights**: All edge weights are normalized to eliminate bias due to different match counts or total passes.

$$w_{\text{given}}^{\text{norm}}(i) = \frac{w_{\text{given}}(i)}{\sum_j w_{\text{given}}(j)}$$

$$w_{\text{received}}^{\text{norm}}(i) = \frac{w_{\text{received}}(i)}{\sum_j w_{\text{received}}(j)}$$

where i denotes a particular playing position.

This normalized graph represents the true pass-distribution behaviour of each player.

Step 3: Node Embedding Using Node2Vec

Node2Vec is applied to each graph. The embedding of the player's main node serves as a vector representation of their positional pass tendencies.

Step 4: Similarity Metrics

To measure how closely two players match, we compute:

Cosine Similarity

$$\text{Cosine}(A, B) = \frac{A \cdot B}{\|A\| \, \|B\|}$$

Pearson Correlation

$$\text{Pearson}(A, B) = \frac{\sum_{i=1}^{n}(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^{n}(A_i - \bar{A})^2} \, \sqrt{\sum_{i=1}^{n}(B_i - \bar{B})^2}}$$

Weighted Jaccard Similarity

$$\text{WeightedJaccard}(A, B) = \frac{\sum_{i=1}^{n} \min(A_i, B_i)}{\sum_{i=1}^{n} \max(A_i, B_i)}$$

Two players are considered **substitutable** if at least two of the three scores exceed predefined similarity thresholds.

Step 5: Tkinter Interface

A Tkinter interface is provided where:

- The user inputs two player names

- The system displays their normalized pass-distribution graphs

- The interface outputs whether the two players are substitutable

Examples

Below are example visualizations of the graphs generated for different players:
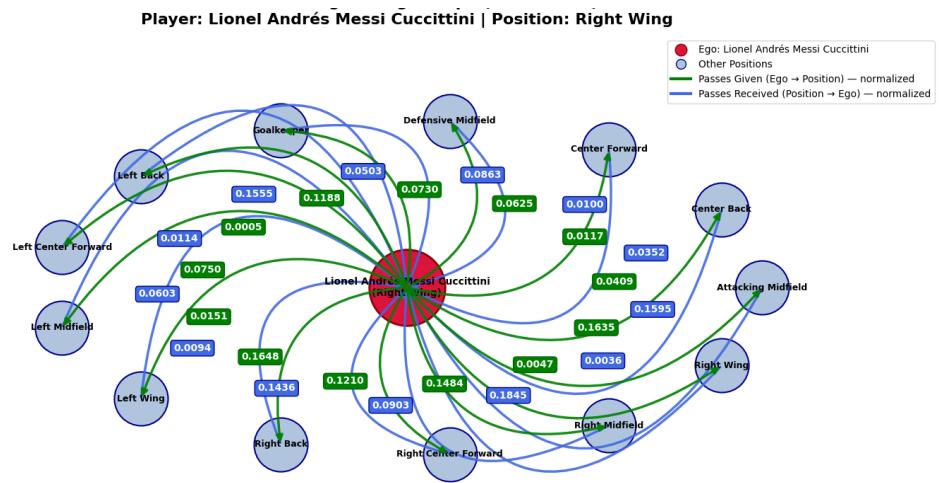
Figure 1: Generated graph
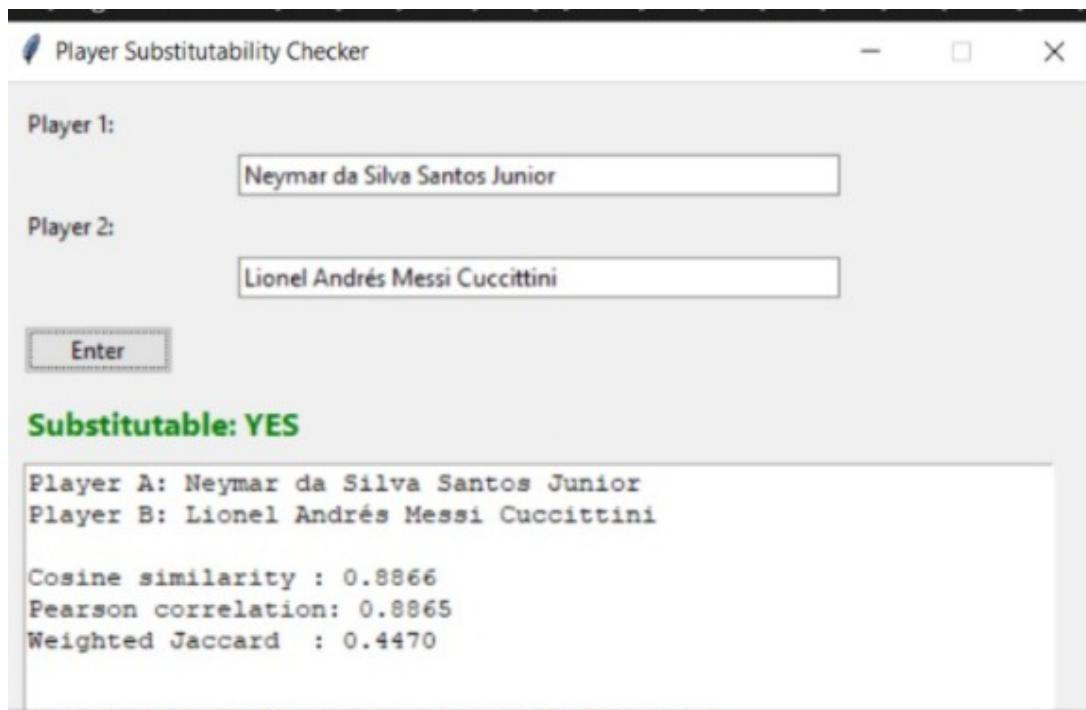
Figure 2: Substitutable
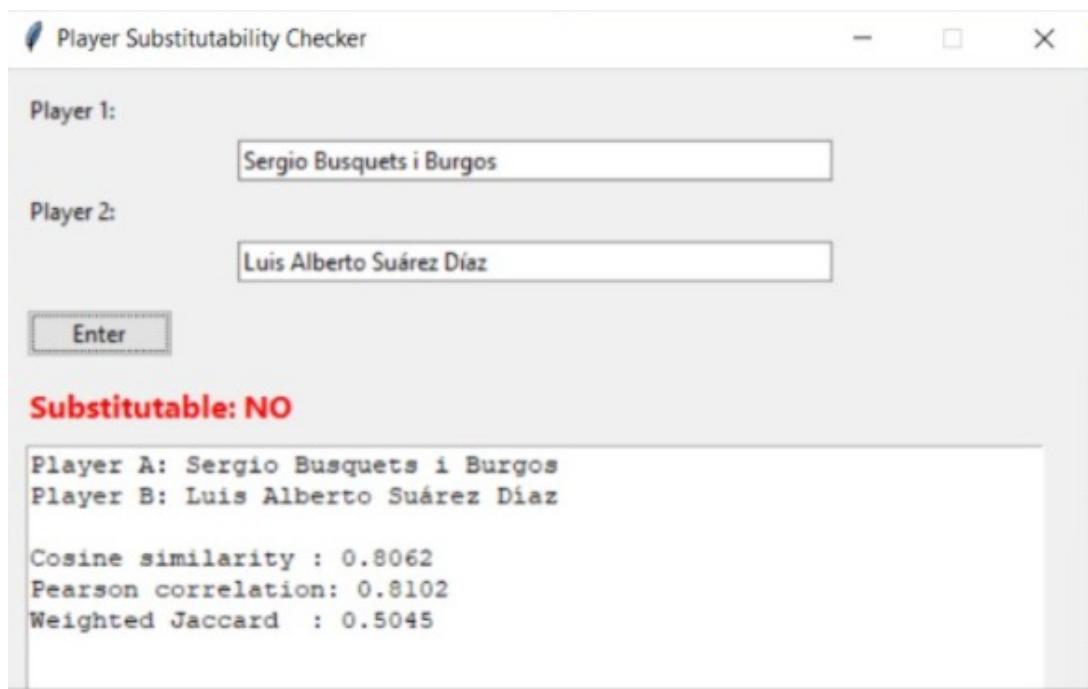
Figure 3: Not Substitutable