A Network Tour of Football Transfer

Group 23

December 2019

1   Introduction

Soccer is a very popular sport in Europe. At the same time, this attractive event has triggered a series of economic effects each year. During the transfer window, a lot of soccer players transfer from original club to another, which affect both the performance and economic condition of different clubs. According to statistics, the European football market is now estimated to be worth 25.5 billion. Upon this, we wonder the connection of this large amount of trade and the features that can influence the transfers.

   The goal of the project is to exploit the connection of the soccer player transfers among clubs in different European leagues. The analysis will be implemented in the network approach. During the project, we will create a graph based on the transfer of soccer players and exploit the data with tools like clustering, dimensionality reduction, classification, etc.

2   Dataset

The dataset that we choose to use are **Player Transfer Data of nine majors European football leagues from 1992 to 2018** :

1. English Premier League

2. English Championship

3. dutch eredivisie

4. french ligue 1

5. german bundesliga

6. italian serie A

7. portugese liga nos

8. spanish primera division

9. russian premier liga

Each dataset include features like:**club name, player, name, age, position, club involved name, fee transfer movement, fee cleaned league name, year,**

We may merge these data into a single dataset and proceed analysis in **European scale**, or pick a single league and proceed more fine grained analysis.

Besides, the **FIFA score data** will also probably be used.

3   Network

Two possibilities:

Players based network: Each player is a node and the link could be "if two players played in a same team before".

Clubs based network: Each club represents a node and the link would represent a transfer(transfers) between two clubs and the weight could be for example the transfer fee amount or the number of transfers.

The tools that we choose to use: **clustering (spectral clustering, k-means) dimensionality reduction (PCA, MDS, LLE, ISOMAP, Laplacian eigenmaps, t-SNE)**

4   Research Questions

In terms of the standard data exploration, some interesting questions would be:

- In the European context, which clubs mainly serve other clubs with players, and which ones get transferences from smaller clubs?

- Which countries have clubs that receive more players from abroad and which countries more often use their internal market?

In terms of the analysis of the structure of the network, we would like to investigate:

- Which clubs occupy the center of the European football transfer market, and which of them are more peripheric?

- Are there clusters of clubs that trade mainly with each other (in terms of number of players / fees)?

We would like to performe some classification task, such as for example (need to rewrite / more ideas):

- Is it possible to infer the wealth of a club by analyzing the patterns of transfers between it and other clubs