Basic Info

Title: Influence of passing behavior on performance in football

Contributors: Bashir Hamidi and Anoop Kakkireni

Email addresses: bhamidi@clemson.edu, akakkir@clemson.edu

CIDs: bhamidi, akakkir

Project Repository: https://github.com/bashirhamidi/CPSC6030_g11

Background and Motivation

We have selected a project that deals with football passing data for multiple reasons. First, we believe that because most people are familiar with sports to one degree or another, they would be able to relate to and grasp the idea of passing the ball within a team sport contributing to a better team performance. Secondly, football is undoubtedly an international sport with fans across every country including the two collaborators in this project. Both of us grew up playing football and as a result are fans of the sport. Further, we have a hypothesis that passing within football has a significant impact on dynamics of the game. We see similar trends in other sports; for example, the world-renowned basketball team San Antonio Spurs lead by Gregg Popovich have season after season success as a team because of the team dynamics. Many would label the Spurs team dynamics as one in which every player has a role within the team and passes the ball and puts up screens using systematic and well-practiced routines. We believe similar passing habits to contribute to a football team's success across a tournament and maybe even beyond.

Coming across the publication “PassVizor: Toward Better Understanding of the Dynamics of football Passes” our interests were piqued because of the subject matter as the analytical perspective of the publication. The primary questions that the paper was trying to answer aligned with what we were hoping to achieve.

Moreover, there is a misconception that goal scorers are the most important players in a team. From an informed standpoint, this is not true in most cases and there are several players contributing to the success of a team. The visualizations that we hope to achieve will clear this misconception and enable an ill-informed viewer to have a better understanding of the game and players’ importance represented through passing dynamics.

Project Objectives

Football is a passing game. Although the outcome of a game is determined by the number of goals scored, it wouldn't be possible to score goals without passing the ball. There are many statistics to analyse the game of football, but, the considerable increase in the average number of passes over the years calls for an analysis pertaining to the passing dynamics of teams. The primary objective of the project is to understand how various passing patterns and metrics influence the style of play and the outcome of a game. It is also important to acknowledge the fact that goal scoring is a team effort. The contributions of players in different positions has to be accounted for to develop a proper understanding of their impact in a particular game.

Moreover, there's a misconception that goal scorers are the most important players in a team. From an informed standpoint, this is not true in most cases and there are several players contributing to the success of a team. The visualizations that we hope to achieve will clear this misconception and enable an ill-informed viewer to have a better understanding of the game and players' importance represented through passing dynamics.

The questions that we will be trying to answer during the course of the project are:

1. *How does passing behavior affect the outcome of a game?*
2. *What is the impact of different players in the overall performance of the game?*
3. *What are the passing styles and patterns of a particular team?*

Data

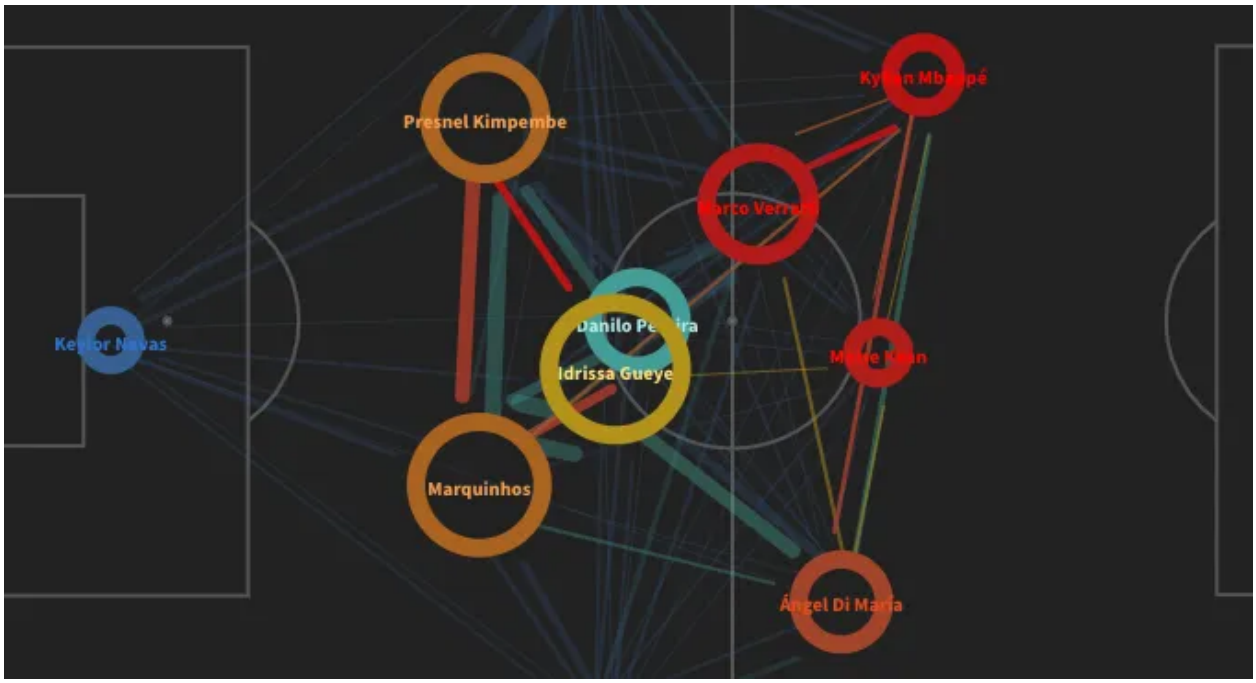
The data has been compiled by Mejorado et al. [1] and published under the Creative Commons Attribution 4.0 International license. This dataset is available for download using the link https://zenodo.org/record/1935382. The data is pre-compiled and packaged into three folders for each FIFA tournament as well as subfolders containing more detailed passing information, player position information, and player performance measures.

Data Processing

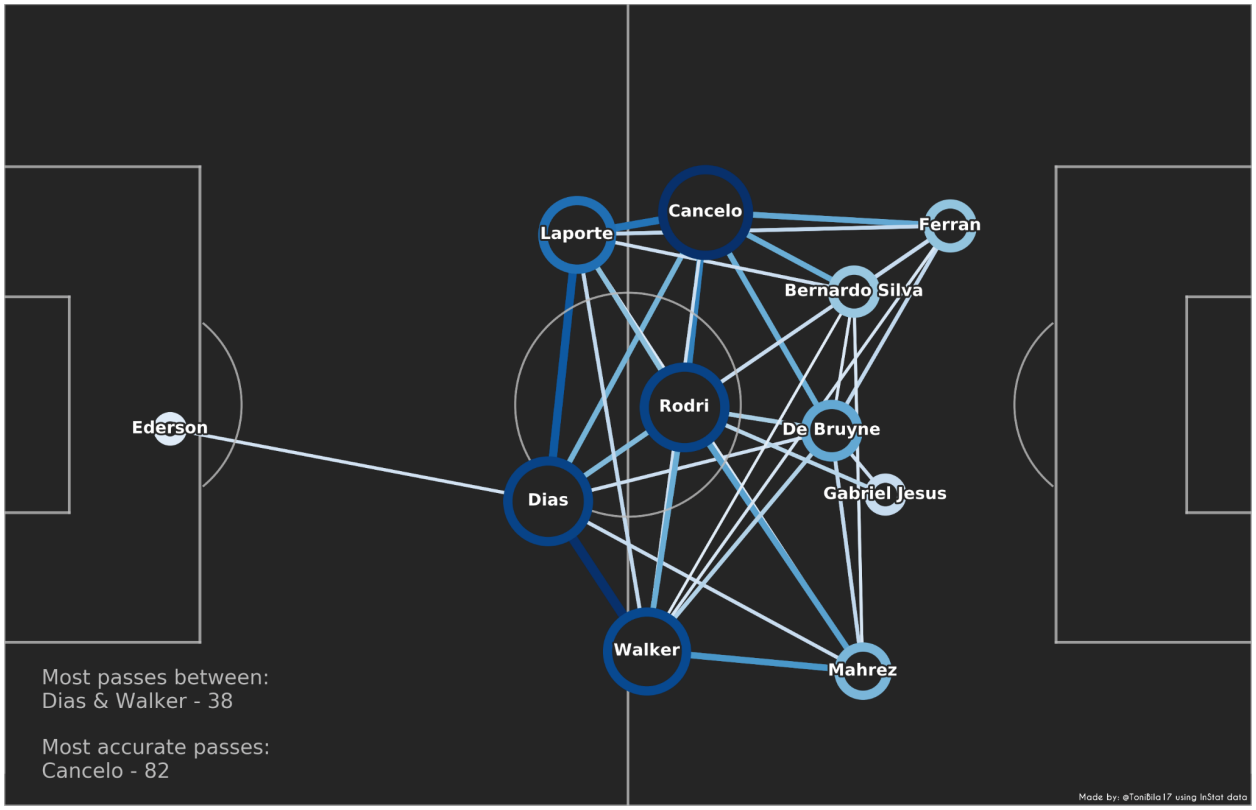
We have a dataset that consists of match data from three World Cup tournaments: the 2010 tournament hosted by South Africa, the 2014 tournament hosted by Brazil, and lastly, the 2018 tournament hosted by Russia. For each of these tournaments, we have respective spreadsheets that contain general match information such as match number, teams, player names, player position, number of minutes played, number of passes attempted and completed, number of passes received, goals scored, shots made, fouls committed, and fouls suffered. In separate CSV files, for each match, we also have a distribution of the passes by each player within that match. The latter files are organized by country names and require a fair amount of manipulation in order to make sense of the data.
We plan to do the data processing using R and Excel depending on the task underway. If we decide to utilize the passing distribution spreadsheets, we will need to write a script to process the data as there are 385 individual CSVs that correspond to the matches and contain the passing distribution of each player. Because this data is not well-labeled, it may take additional manual processing steps that would significantly add to the time needed to process the data.

Visualization Design

Sample 1:



https://i2.wp.com/statsbomb.com/wp-content/uploads/2021/04/Paris-Saint-Germain-vs-Lyon-2021-03-21-Pass-Map.png?fit=679%2C367&ssl=1

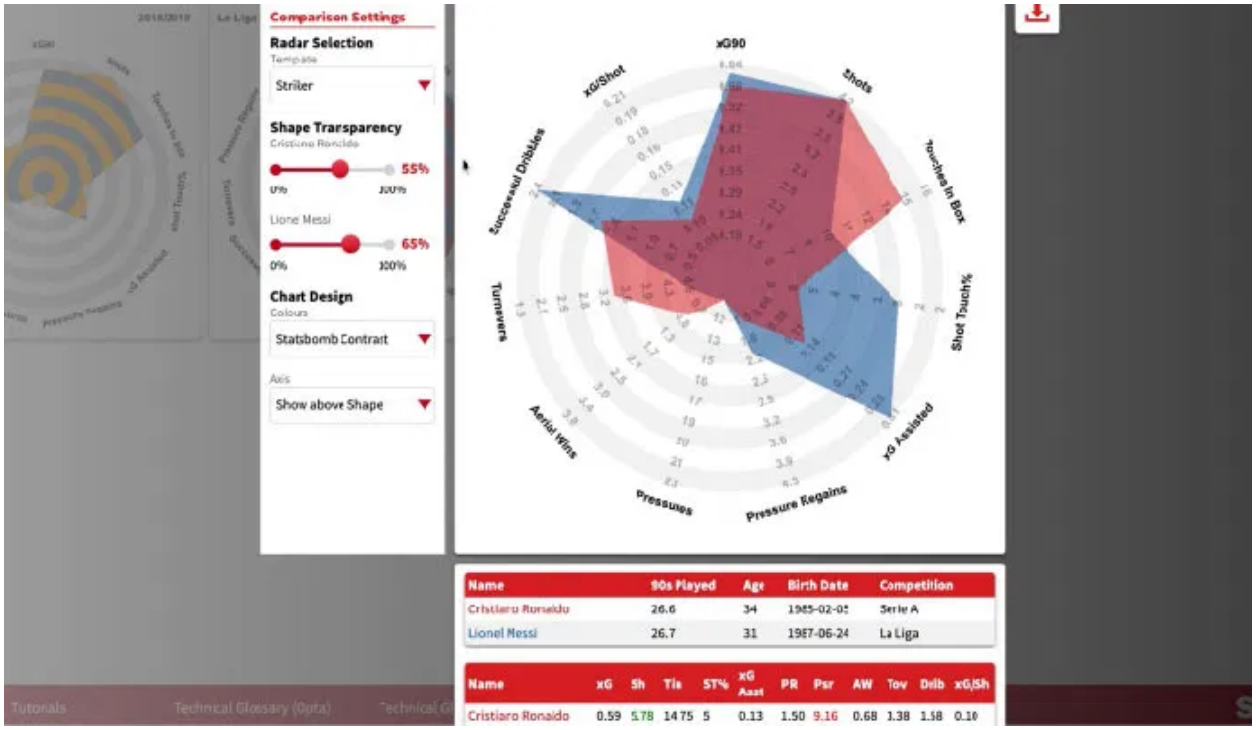


https://totalfootballanalysis.com/wp-content/uploads/2020/11/City-passing-network.png

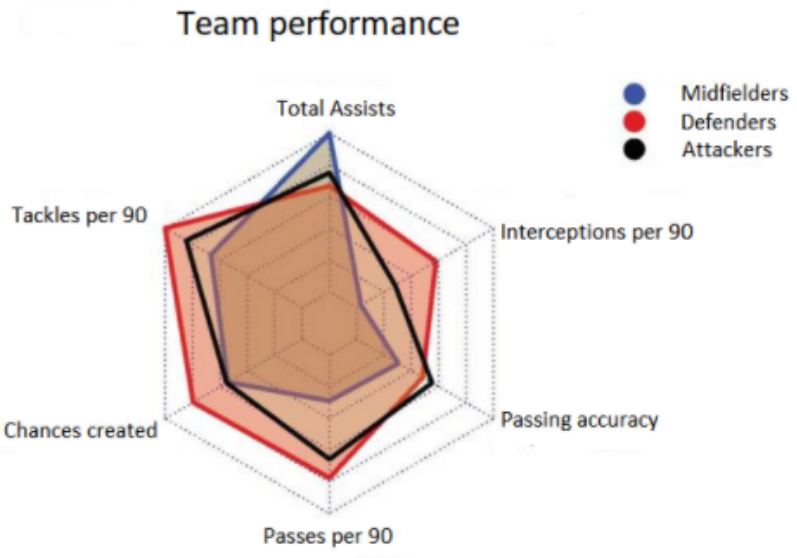
We will try to develop a visualisation as the ones mentioned above to create a network of passes between all the players in a team. This will give us an idea of the passing style and the play style of a particular team. It will also highlight the key players responsible for creating chances throughout the game.

The network graph has the ability to convey the patterns in passing behaviour because it actually depicts the location of players on the field. It is easy for the viewer to understand the visualisation based on his/her understanding of the game. The thickness of the lines or the saturation of the lines used to connect different players can be used to portray the number of passes between them.

Sample Idea 2:



https://i2.wp.com/statsbomb.com/wp-content/uploads/2019/05/player-comparison.jpg?fit=692%2C400&ssl=1

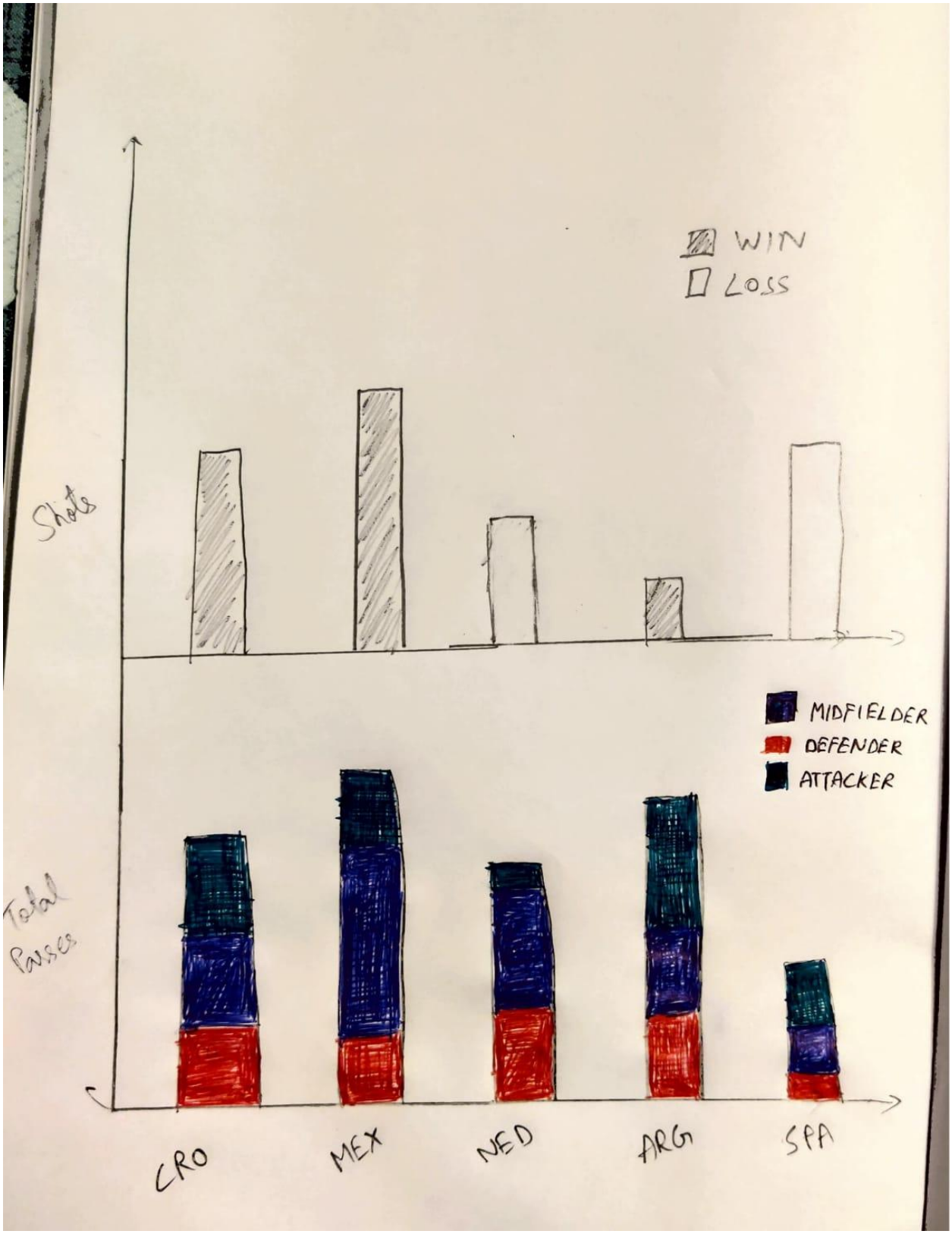


One of the other ideas is to create a radial graph similar to the ones mentioned above to showcase the skill level of various players. All the metrics here will be normalized and hence will give a better picture of the contributions of different players. The intention behind using a radial graph

is to estimate the competency of a player in each of the attributes revealing the weaknesses and strengths.
The same idea can be extended by aggregating the data based on the position of the players namely, midfielders, attackers and defenders.

We wish to depict the attributes ‘distance covered with possession of the ball’ , ‘passes attempted’, ‘passes completed’, ‘passing accuracy’, and the attributes corresponding to the percentage of time spent on different parts of the field to estimate the impact and usefulness of each player. All this data pertains to the national team of Brazil in the 2014 World Cup.

Sample Idea 3:


WIN
LOSS
Shots
MIDFIELDER
DEFENDER
ATTACKER
Total Passes
CRO
MEX
NED
ARG
SPA

To understand the direct relationship between the passing behavior and the outcome of the game in terms of the no.of chances created and the end result itself, we can use a stacked bar graph. The bar graph is a simple way of understanding how the change in passing behaviour from game to game influenced the performance of the national team of Brazil in the 2014 world cup.

Must-Have Features

We would like to have a clear-cut way of showing teams' performance being driven from player passing statistics. We would like to have this performance data shown to users in a way that's very clear and understandable. At this point, we imagine this to be a feature that displays the influence of passing on match outcome in a binary way, maybe coloring those matches that were predicted by passing in green vs those that were not in gray.

Optional Features

We would like to implement a filter feature for users to select use of data from specific tournaments. This would allow users to explore the possibility of passing performance impact over multiple tournament years.
Further, we would like to implement a feature in which users are able to filter out specific countries and explore the matching performance of one team as opposed to the aggregate.
Lastly, another nice-to-have feature would be mouseover interactive graphics that allow users to select a particular country and have the figure filtered to that country's data only.

Project Schedule

Week	**Delegates, Anoop (A) and Bashir (B)**	**Tasks, Responsibilities, and Deliverables (bolded)**
October 18	A & B	Ironing out the skeleton for the website and key svg elements. Begin implementing skeleton for site. Meet 2x/week or as necessary to collaborate and discuss progress.

October 25	A & B	Have a fully implemented skeleton by the start of the week and begin working on the svg elements. Split the work of the svg element between A & B to tackle them quickly. Meet 3x/week or as necessary to collaborate and discuss progress.
November 1	A & B	Continue to implement svg elements. Adjust the skeleton as necessary. Meet 3x/week or as necessary to troubleshoot, collaborate, and discuss progress. **Assignment Project Milestone 3 - Project prototype due on Sun Nov 7, 2021**
November 8	A & B	Continue to work on the prototype. Implement changes as necessary. Discuss upcoming deliverables. **Assignment Project Milestone 4 - Peer evaluation due on Sun Nov 14, 2021**
November 15	A & B	Create an outline of the final project report, split the tasks, and begin working on the report. Create an outline of the screen-cast and split the work. Create a script of necessary topics to be covered and how to present them.
November 22	A & B	Split the work of presentation, come up with an outline, practice the presentation and provide critical and helpful feedback to each other.
November 29	A & B	**Assignment Project Milestone 5 - Presentation due on Tue Nov 30, 2021** Make final changes to the project report. **Assignment Project Milestone 6 - Final submission due on Sun Dec 5, 2021**

References

1. Martinez Mejorado, Denisse Aneth, & Ramirez Marquez, Jose Emmanuel. (2018). FIFA World Cups 2010, 2014 and 2018 matches' information datasets (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1935382
2. X. Xie, *et al*.,"PassVizor: Toward Better Understanding of the Dynamics of Soccer Passes" in *IEEE Transactions on Visualization & Computer Graphics*, vol. 27, no. 02, pp. 1322-1331, 2021. doi: 10.1109/TVCG.2020.3030359