Analysis of team passing networks considering expected threat

Authors:

Lihong Ji

Joe McMullen

Tutors:

Javier Martin Buldu

1. Introduction

Football is an invasion sport in which each team has 11 players (10 outfield and 1 goalkeeper) and aim to score by moving the ball into the opposing team’s goal. Given that each team consists of multiple agents working together to overcome the other, complexity arises. We can therefore consider each team as a complex system in which the behaviour of the team cannot be determined by considering the individual players alone.

As a result, analysis and prediction of football performance can use the approaches developed in the field of complex sciences. These approaches were developed by considering the interactions, structures and behaviours of complex systems such as the internet and organism biology.

Network science is one branch of complexity science which considers a system as nodes which represent entities and edges which represent a link between the nodes. This link can be a distance, communication, or other interaction. In the case of football, individual players can be considered as nodes and the edges between them can represent some action between the players.

Given that the goal of football is to move the ball to a desired area, passing between teammates is important, not just for progress of the ball but to prevent the opposing team from claiming possession. Directed networks in which the nodes represents the players of a team and the edges represent the volume of passes from one player to another have become popular as a method of performance analysis. Particularly when visualised on a pitch with average location of each player as the node location. This gives an idea of team structure and pattern of play.

Football is also a game with sparse rewards, the average number of goals in each La Liga game in he 2022/23 season was 2.51. Over a game of 90 minutes, and due to the large number of variables, and stochasticity, we would like to understand better which team deserved to win.

This has led to the development of additional metrics including expected goals (xG) which can be defined as the probability that a shot results in a goal and can be considered as chance creation. Many of these xG models use classifiers such as decision trees or XGBoost trained on large datasets of shots.

However, xG can only assign value to shots taken and again these are sparse in the context of football. Additionally, they only tell us about the performance of the shooter and do not give context on how the other 10 players on the field contributed to the chance creation. To understand individual contribution to chance creation, we want to assign a value to all actions on the pitch in terms of their contribution to the likelihood of scoring a goal.

Expected threat is a metric that provides the likelihood of scoring at the end of a possession based on location on the pitch. Areas of the pitch are considered as nodes in a Markov chain and actions are the station. Using this, we can calculate how in possession actions increase or decrease the likelihood of actions and analyse the contribution of players to the threat of a team over a match or season.

1. Mission Statement:

Our aim to bring together xT and passing networks. We will do this by considering the edges between players of the same team as the accumulation of xT due to passes between players. We will compare this to the traditional approach of edges representing pass volume to determine what additional insight we can learn through the combination of network science and advanced football metrics.

1. Data
   1. Data Description

The data used in this study consists of match information and event data for 192 games in the first half of the 2023/24 La Liga season provided by Opta. For each match, we have 4 files:

* Team

Contains team id and name for both teams in each match, the teams which featured can be seen below:



* All Events

Contains data for all events that occurred during a match, this includes information on the identifier of the match in which the event occurred, half of the game, the minute and the second in which it occured, the outcome of the event, the position of the event, the identifier, name, position and team from the player who originated the event, the team, identifier and name of the goalkeeper that receives the shot, the expected goal probability and some other details about the shots that are not used

* Pass

Additional detailed information on pass events not found in all events including receiving player information and location, the sequence and possession number that each pass was part of and whether the pass was offside.

* Player

Summation of player information for each game, including identifier, name and shirt number, what position they started the game in and how many minutes they played