

Community Detection and Player Influence Within Football Matches

Mitchell vom Scheidt, Willem Strydom

Washington University in St. Louis, St. Louis, 63130

November 9, 2023

Abstract

This project aims to investigate and analyze data collected from numerous football matches across the 2018/2019 Champions League competition. Specifically, we will construct a complex network for each team within a given match with the goal of determining which player, or group of players were the most important for each team. The method for constructing the complex networks is still to be determined. However, one approach is to construct a node for each player in the starting lineup, where two nodes are connected by an edge if a pass was successfully played from one player to the other. These edges will be weighted, where the weight of an edge is the number of successful passes one player played to another throughout the game.

We are interested in performing this analysis as the results of this sort of network analysis could have beneficial applications to football clubs. Performing this analysis on your own team could give insights into how successfully a club's strategy is being implemented, and which players have the most impact. On the other hand, performing this analysis on your opponents in preparation for a match could be beneficial. A club could use it to help determine the best strategy to use during a game to help increase their chances of victory.

The way in which this project will perform its analysis is by implementing various algorithms to determine the most important player, or group of players for a specific team. These algorithms will include calculating numerous centrality measures, such as betweenness, closeness and eigenvector centrality. In addition, we will implement community detection algorithm's to investigate whether there was a group, or groups of players which had a significant impact on the game. The preliminary choice of community detection algorithm's are the greedy modularity-based algorithm, which finds communities using greedy modularity maximization and the Louvain community detection algorithm, which is a heuristic method based on modularity optimization.

Our source of data for this investigation can be found at this GitHub repository: <https://github.com/statsbomb/open-data>. StatsBomb provides open-source data which we are able to use for our analysis. Additionally, they have a python package, statsbombpy, which we will utilize to import the necessary data so that we can manipulate and analyze it. The data is in a JSON format, which initially we are not able to construct our network from. We will use the data to create our own dataframe and pick out the relevant columns to construct our own network.

Keywords— Node Centrality, Community Detection, Complex Network Construction