EPSN (Extraction of Particular Sports as Networks): A Sports API for Network Analysis

Collin J. Kovacs

23 April 2022

Summary

In recent years, various sports and its data have been analyzed more to find more efficient techniques and patterns to improve the coaches and players abilities. But, data extracted from sports stats and events are given and shown in relational or tabular format. `EPSN` strives to revolutionize the way users analyze sports data as whole. While allowing users to choose which level they would like to analyze the desired sport, Team or Player, the user is given a response in Graph Markup Language (GML) format so it can be saved and used for later or parsed, cached, and analyzed in Python using `NetworkX` (Hagberg, Swart, and S Chult 2008).

Statement of need

When using network data formats like `.gml` or a `.csv` that explicitly states what the source and target nodes are, there is no room for customization and variability. Often times, finding the right network data is hard, time consuming, and sometimes impossible. This is where `EPSN` comes into play. `EPSN` is a wrapper of `Sportsipy` (Clark 2019) developed for the sole purpose of producing sports affiliations as networks. This project was developed due to my original inability to find quality datasets that represented networks in a clean, novel and unoverused way. `EPSN` provides flexible queries to a user who is interested in researching anything about sports but at a network-based level. This API provides two levels, Team and Player, to allow the user to study interactions between these associations.

One example of a problem that could be analyzed is studying the network of MLB players from 1971 to understand how the interactions of players led to the Pittsburgh Pirates winning the World Series. Another example would be

to determine schedules NFL teams by predicting links given the data. The possibilities are endless and they are up to the user to make use of them!

References

Clark, Robert. 2019. "Sportsipy: A Free Sports API Written for Python." *GitHub Repository*. GitHub. https://github.com/roclark/sportsipy/.

Hagberg, Aric, Pieter Swart, and Daniel S Chult. 2008. "Exploring Network Structure, Dynamics, and Function Using NetworkX." Los Alamos National Lab.(LANL), Los Alamos, NM (United States).