DS210 Final Project Report

By SeungJae Hong

Dataset: **Premier League Standings 1993-2024**

1. Introduction

I chose the English Premier League standings dataset because the Premier League is the pinnacle of soccer in Great Britain, and as an avid soccer fan, it's a subject that deeply interests me. My goal is to analyze the dataset to uncover key insights about team performance. I aim to analyzes the English Premier League (EPL) standings from 2000 to 2022 using graph theory, specifically focusing on betweenness centrality. The goal is to quantify each team's influence on the league's structure based on their connections. Betweenness centrality helps identify which teams serve as key "bridges" in the competitive landscape of the EPL. By calculating centrality scores, the project provides insights into the most influential teams and their importance in shaping league outcomes over the years.

2. Dataset

The dataset (**EPL_standings_2000-2022.csv**) includes:

- Season information (years from 2000 to 2022)
- Ranking of the season
- Team names
- Number of games
- Performance metrics such as wins, draws, losses, goal against, goal for, goal difference, and points earned.

3. Project Detail

a. main.rs: This is the entry point of the program. It handles the workflow of the project by:
   i. Loading the Dataset:
      1. Uses graph_construction.rs to construct a graph representation of the dataset.
      2. The dataset contains information about EPL teams, such as their wins, losses, and other metrics.

ii. Computing Betweenness Centrality:
    1. Calls the function from betweenness_centrality.rs to compute centrality scores for each team.
iii. Displaying Results:
    1. Prints the betweenness centrality of each team in a well-formatted output (e.g., rank, team name, and centrality score).
    2. Optionally, displays graph nodes and edges for debugging or visualization.

b. graph_construction.rs: This file is responsible for constructing the graph from the EPL dataset.
    i. Converts the dataset (CSV format) into a graph structure where:
        1. Nodes: Represent teams (e.g., Manchester United, Chelsea).
        2. Edges: Represent relationships or similarities between teams, such as differences in wins, losses, or other metrics.
        3. Edge Weights: Quantify the strength of the connection (e.g., absolute differences in wins).

c. graph_construction.rs: This file is responsible for constructing the graph from the EPL dataset.
    i. Converts the dataset (CSV format) into a graph structure where:
        1. Nodes: Represent teams (e.g., Manchester United, Chelsea).
        2. Edges: Represent relationships or similarities between teams, such as differences in wins, losses, or other metrics.
        3. Edge Weights: Quantify the strength of the connection (e.g., absolute differences in wins).

4. How to Run

a. Dependencies:
    i. Ensure Rust is installed. Use cargo for building and running the project.
    ii. Required crates: petgraph, csv, serde.

b. Steps:
    i. Place the dataset (EPL_standings_2000-2022.csv) in the data folder.
    ii. Run the project:

```
cargo run
```

    iii. View the output in the terminal.

c. Expected Output:
    i. Centrality scores for all teams.
    ii. Graph nodes (team names and indices).

5. Output

a. Betweenness Centrality

```
Betweenness Centrality Scores:
Team: Swansea City                  , Centrality: 21138.4238
Team: Sheffield United              , Centrality: 7519.2636
Team: Blackburn Rovers              , Centrality: 6476.4199
Team: Everton                       , Centrality: 6433.7922
Team: Queens Park Rangers           , Centrality: 5436.0635
Team: Leeds United                  , Centrality: 5160.7032
Team: Bolton Wanderers              , Centrality: 3931.8420
Team: Stoke City                    , Centrality: 3839.8248
Team: Birmingham City               , Centrality: 3753.9228
Team: Southampton                   , Centrality: 3393.1149
Team: Sunderland                    , Centrality: 3155.3199
Team: Aston Villa                   , Centrality: 2812.1733
Team: West Bromwich Albion          , Centrality: 2539.3647
Team: Charlton Athletic             , Centrality: 2410.9198
Team: Burnley                       , Centrality: 2346.2394
Team: Newcastle United              , Centrality: 1745.9440
Team: Watford                       , Centrality: 1636.1888
Team: Crystal Palace                , Centrality: 1589.8461
Team: Fulham                        , Centrality: 1354.9758
Team: Norwich City                  , Centrality: 1325.4285
Team: Portsmouth                    , Centrality: 1140.8551
Team: Leicester City                , Centrality: 1103.0074
Team: Cardiff City                  , Centrality: 1048.7370
Team: Manchester United             , Centrality: 1046.7653
Team: Derby County                  , Centrality: 1034.1439
Team: Brighton & Hove Albion        , Centrality: 782.9818
Team: West Ham United               , Centrality: 729.7027
Team: Blackpool                     , Centrality: 682.1628
Team: Wolverhampton Wanderers       , Centrality: 681.1129
Team: Bournemouth                   , Centrality: 672.7413
Team: Arsenal                       , Centrality: 601.9760
Team: Wigan Athletic                , Centrality: 541.6418
Team: Hull City                     , Centrality: 533.7861
Team: Chelsea                       , Centrality: 505.0953
Team: Brentford                     , Centrality: 447.3592
Team: Middlesbrough                 , Centrality: 439.0549
Team: Huddersfield Town             , Centrality: 397.1668
Team: Tottenham Hotspur             , Centrality: 351.8009
Team: Liverpool                     , Centrality: 325.4974
Team: Manchester City               , Centrality: 290.5077
Team: Reading                       , Centrality: 273.0942
Team: Coventry City                 , Centrality: 252.0937
Team: Ipswich Town                  , Centrality: 82.9083
Team: Bradford City                 , Centrality: 21.9735
```

b. Graph Construction (node)

```
Graph Nodes:
Node Index: 0  , Team: Manchester United
Node Index: 1  , Team: Arsenal
Node Index: 2  , Team: Liverpool
Node Index: 3  , Team: Leeds United
Node Index: 4  , Team: Ipswich Town
Node Index: 5  , Team: Chelsea
Node Index: 6  , Team: Sunderland
Node Index: 7  , Team: Aston Villa
Node Index: 8  , Team: Charlton Athletic
Node Index: 9  , Team: Southampton
Node Index: 10 , Team: Newcastle United
Node Index: 11 , Team: Tottenham Hotspur
Node Index: 12 , Team: Leicester City
Node Index: 13 , Team: Middlesbrough
Node Index: 14 , Team: West Ham United
Node Index: 15 , Team: Everton
Node Index: 16 , Team: Derby County
Node Index: 17 , Team: Manchester City
Node Index: 18 , Team: Coventry City
Node Index: 19 , Team: Bradford City
Node Index: 20 , Team: Blackburn Rovers
Node Index: 21 , Team: Fulham
Node Index: 22 , Team: Bolton Wanderers
Node Index: 23 , Team: Birmingham City
Node Index: 24 , Team: West Bromwich Albion
Node Index: 25 , Team: Portsmouth
Node Index: 26 , Team: Wolverhampton Wanderers
Node Index: 27 , Team: Crystal Palace
Node Index: 28 , Team: Norwich City
Node Index: 29 , Team: Wigan Athletic
Node Index: 30 , Team: Reading
Node Index: 31 , Team: Sheffield United
Node Index: 32 , Team: Watford
Node Index: 33 , Team: Stoke City
Node Index: 34 , Team: Hull City
Node Index: 35 , Team: Burnley
Node Index: 36 , Team: Blackpool
Node Index: 37 , Team: Swansea City
Node Index: 38 , Team: Queens Park Rangers
Node Index: 39 , Team: Cardiff City
Node Index: 40 , Team: Bournemouth
Node Index: 41 , Team: Brighton & Hove Albion
Node Index: 42 , Team: Huddersfield Town
Node Index: 43 , Team: Brentford
```

6. Insights

a. High Betweenness Centrality Teams:
   i. Teams like Swansea City, Sheffield United, Blackburn Rovers, and Everton often act as "bridges" in the league, influencing other teams' standings.
   ii. These teams may represent key turning points in competition dynamics.

b. Low Betweenness Centrality Teams:
   i. Teams with lower scores, such as Manchester United, Ipswich Town, and Bradford City have less influence on the network. This could happen if the team is consistently at the top or bottom of the standings, with less variability in their interactions.

c. League Dynamics:
   i. The results highlight how certain teams consistently impact the competitive structure.

7. Test

```
   Compiling epl_betweenness_project v0.1.0 (/Users/hongseungjae/ds210_fp_seungjae_hong)
    Finished `test` profile [unoptimized + debuginfo] target(s) in 1.95s
     Running unittests src/main.rs (target/debug/deps/epl_betweenness_project-94409beea80b1f03)

running 1 test
test tests::test_graph_construction_and_centrality ... ok

test result: ok. 1 passed; 0 failed; 0 ignored; 0 measured; 0 filtered out; finished in 0.00s
```

8. Conclusion

This project demonstrates the application of graph theory to sports analytics. By analyzing betweenness centrality, it highlights the key roles certain teams play in the EPL's competitive structure. These insights can be valuable for stakeholders, such as analysts, coaches, and fans, to understand league dynamics better.