Marcello RUSSO

Unveiling English Premier League Team Relationships: A Decade of Network Analysis

Complex Networks

Prof. David Soriano PAÑOS

UNIVERSITAT ROVIRA i VIRGILI

Tarragona, 2025

1. Introduction

1.1. The English Premier League: A Competitive Landscape

The English Premier League (EPL) stands as one of the most prominent and fiercely competitive football leagues globally, captivating audiences with its dynamic play, high stakes, and unpredictable outcomes. Characterized by significant financial investment and a vast talent pool, the EPL serves as an ideal domain for applying advanced analytical techniques to uncover underlying patterns and relationships. Beyond simple league standings, the intricate interactions between teams—their strengths, tactical approaches, and ability to control matches—shape the competitive narrative across seasons. This project aims to delve into these multifaceted relationships, moving beyond traditional statistical aggregation to explore the network structures that emerge from direct team confrontations. Our analysis focuses on a decade of EPL history, specifically the seasons ranging from 2008/09 to 2017/18, providing a robust longitudinal perspective on team dynamics.

1.2. Dataset and Relational Foundations

The foundation of this analysis is a meticulously compiled dataset encompassing match-level statistics for every English Premier League game played between the 2000/01 and 2024/25 seasons. For each match, key performance indicators were collected, including goals scored, goals conceded, and derived metrics related to team aggressiveness and possession control. These raw statistics form the basis for constructing a relational network where teams are represented as nodes. Crucially, the connections (edges) between these teams are established not merely by their participation in a match, but by the *differences* in their performance metrics during direct encounters. This approach allows for the quantification of similarity or dissimilarity between teams based on their head-to-head competitive profiles. For instance, two teams consistently exhibiting small differences in goal margins or control metrics in their matches are considered to have a stronger, more significant connection in the network for that particular attribute. This relational framework transforms raw match data into a structured graph, enabling the application of network science methodologies.

2. Network Construction

2.1. Data Filtering and Aggregation

The initial step in constructing our analytical network involves preparing the raw match data. The ‘create_epl_network‘ function first filters the comprehensive Premier League dataset to isolate matches corresponding to the specified analysis scope, which can be either a single season (e.g., ’2016/17’) or a continuous range

of seasons (e.g., from 2008 to 2018). This ensures that the network is built upon the precise temporal window of interest.
Following the temporal filtering, the function aggregates match-level statistics to capture the cumulative performance between every pair of teams over the defined period. This aggregation is crucial as it sums up all relevant metrics for a given team pair, distinguishing between the roles of 'source' and 'target' team. Specifically, for any two teams A and B, the function accumulates statistics for matches where A was the home team against B, and where A was the away team against B. This bidirectional aggregation results in two distinct entries for each pair: (A, B) representing A's performance against B, and (B, A) representing B's performance against A. This comprehensive summation ensures that all direct confrontations contribute to the overall relational strength and characteristics between teams.

2.2. Definition and Calculation of Performance Metrics

Once aggregated, several key performance metrics are calculated for each directed pair of teams (SourceTeam $\rightarrow$ TargetTeam). These metrics quantify the cumulative performance of the source team when playing against the target team.

- **Goals Scored (goals_scored):** This metric represents the total number of goals scored by the source team against the target team over all their direct encounters within the selected period.

$$\text{goals\_scored}_{(\text{S}\rightarrow\text{T})} = \sum(\text{Goals Scored by S vs T})$$

- **Aggressiveness Committed (aggressiveness_committed):** This metric quantifies the total aggressive actions of the source team against the target team. It is calculated as the sum of fouls, yellow cards, and weighted red cards committed by the source team in matches against the target team. Red cards are typically weighted higher (e.g., by 3) due to their severe impact.

$$\text{aggressiveness\_committed}_{(\text{S}\rightarrow\text{T})} = \sum \text{Fouls}_\text{S} + \sum \text{YC}_\text{S} + 3 \times \sum \text{RC}_\text{S}$$

- **Shot Accuracy (shot_accuracy):** This metric reflects the shooting efficiency of the source team against the target team. It is calculated as the total shots on target by the source team divided by their total shots against the target team.

$$\text{shot\_accuracy}_{(\text{S}\rightarrow\text{T})} = \frac{\sum \text{ShotsOnTarget}_\text{S}}{\sum \text{TotalShots}_\text{S}}$$

(Note: If $\sum \text{TotalShots}_\text{S}$ is zero, the ratio is considered 0.0 to avoid division by zero.)

- **Control (control):** This metric serves as an indicator of the source team's ability to dictate offensive play against the target team. It is calculated as the sum of corners obtained and total shots taken by the source team.

$$\text{control}_{(\text{S}\to\text{T})} = \sum \text{Corners}_\text{S} + \sum \text{TotalShots}_\text{S}$$

- **Points Scored (points_scored):** This metric represents the total league points accumulated by the source team from matches played against the target team within the defined period.

$$\text{points\_scored}_{(\text{S}\to\text{T})} = \sum(\text{Points Scored by S vs T})$$

2.3. Directed Graph Construction

The aggregated data and calculated metrics are then used to construct a directed graph (digraph) using NetworkX. Each team present in the filtered dataset becomes a node in the graph. For every unique pair of teams (SourceTeam, TargetTeam) that have played against each other within the specified period, a directed edge is added from SourceTeam to TargetTeam.
Each edge is enriched with a comprehensive set of attributes, including:

- Raw total metrics (e.g., 'goals_scored', 'aggressiveness_committed') representing the source team's cumulative actions against the target team.
- Normalized versions of these metrics (e.g., 'goals_scored_norm', 'aggressiveness_committed_norm'). These are scaled to a range of [0,1] using min-max normalization based on the global minimum and maximum values observed across all teams for that specific metric.
- Other derived metrics (e.g., differences, absolute differences, and their normalized counterparts) which will be discussed in detail in the following section.
- 'matches_played', indicating the number of games between the specific source-target pair.

The decision to initially create a directed graph, despite the later use of undirected graphs for community detection, is strategic. This approach ensures that the valuable directional information (e.g., Team A scoring more goals *against* Team B) is preserved as an edge attribute. While subsequent analyses might leverage undirected relationships (e.g., for general similarity), the underlying directed structure provides a richer foundation for potential future investigations into directed influence, causal relationships, or asymmetric competitive advantages. This foresight allows for maximum flexibility and depth in exploring the complex dynamics of football team interactions.

3. Network Analysis Pipeline

The analytical journey, from raw match data to profound insights into team dynamics, is orchestrated by the 'perform_epl_network_analysis' function. This robust pipeline systematically integrates graph construction, targeted filtering, sophisticated centrality measurements, and advanced community detection, culminating in a correlation analysis that links network positions to real-world performance. Designed for flexibility, the function allows for the analysis of specific seasons or extended periods, across various performance metrics.
The pipeline's core functionality is structured around the following sequential stages:

1. **Initial Graph Construction and Scope Definition:** The process begins by defining the temporal scope of the analysis, whether a single season or a multi-year span. Based on this scope, the 'create_epl_network' function constructs an initial directed graph. This foundational graph encapsulates all aggregated team interactions and the raw and normalized performance metrics derived from direct confrontations, setting the stage for deeper analysis.

2. **League Performance Benchmarking:** In parallel with network creation, the total league points for each team within the selected analytical scope are computed. These points serve as a vital external benchmark against which the network-derived insights, particularly centrality measures, will be evaluated to understand their real-world impact on team success.

3. **Metric-Specific Network Transformation and Analysis:** The core of the pipeline involves an iterative analysis for each selected performance metric (e.g., goals, aggressiveness, control). For each metric:

   - **Targeted Graph Filtering:** A crucial step is the application of the 'filter_graph_by_weight' function. This process transforms the initial comprehensive directed graph into a filtered, undirected graph tailored to the specific metric being analyzed. The filtering relies on the normalized absolute difference ('_diff_norm_abs') between teams for that metric. By applying a configurable threshold and direction (keeping relationships below or above a certain difference), this step effectively isolates teams exhibiting specific levels of similarity or dissimilarity, creating a focused network for subsequent community detection and centrality analysis. The resulting graph is weighted, where edge weights represent the strength of similarity.
   - **Centrality Measurement:** On this filtered, undirected graph, a suite of centrality measures (e.g., Degree, Strength, Betweenness, Closeness, Eigenvector) are computed. These measures quantify the importance and influence of each team within the network's structure for the specific

metric, revealing teams that are highly connected, bridge diverse groups, or exert significant influence through their connections.

- **Community Detection:** Leveraging the same filtered and weighted undirected graph, the Louvain algorithm is employed to identify natural clusters or 'communities' of teams. These communities represent groups of teams that share similar competitive profiles or stylistic tendencies for the given metric. The 'community_resolution' parameter allows for tuning the granularity of these detected clusters.
- **Performance Correlation:** A Pearson correlation analysis is performed to assess the relationship between the calculated network centrality scores and the teams' accumulated league points. This step provides quantitative evidence on how a team's structural role and influence within the network for a given metric correlates with its overall league performance, offering insights into strategic success factors.

4. **Comprehensive Results Aggregation:** Upon completion of the analysis for all specified metrics, the pipeline meticulously aggregates all outputs, including the filtered graphs, centrality scores, and detected communities. This aggregated data forms a comprehensive record of the analysis, ready for detailed interpretation, visualization, and presentation of findings.

This structured approach allows for a systematic and insightful exploration of the complex interdependencies and underlying dynamics within the English Premier League over time.

4. Analysis and Experimental Setup

4.1. Research Questions and Analytical Justification

This project aims to leverage network science to uncover non-trivial patterns and relationships within the English Premier League, moving beyond traditional league tables to explore team dynamics from a novel perspective. Our analysis is driven by several key research questions:

- Can teams be effectively grouped by their underlying strength or distinctive playing style (e.g., attacking prowess, defensive solidity, control-oriented play) using network-based community detection?
- How do significant historical events, such as the 2014/15 season where the Big Six clubs dominated the top league positions, manifest within these network structures? Do they form cohesive communities, and what insights can be derived from their interconnections?

- What changes occur in the network dynamics and community structures in the immediate aftermath of an extraordinary event, such as Leicester City's unexpected title win in the 2015/16 season? Does this event cause shifts in typical team groupings or competitive relationships?
- Does a longitudinal analysis across multiple seasons offer a more stable and reliable understanding of team characteristics and relationships compared to single-season snapshots? Are long-term patterns more indicative of inherent team attributes?
- How do network centrality measures, reflecting a team's structural position within the network, correlate with their cumulative league points, offering insights into the relationship between playing style, competitive interaction, and ultimate success?

The application of network analysis to football data is particularly justified by the inherent relational nature of the sport. Unlike individual player statistics, team performance is fundamentally about interactions and comparative strengths. By modeling teams as nodes and their direct competitive differences as weighted edges, we can apply graph theory to identify emergent properties, such as communities of similar playing styles or central teams that act as hubs or bridges.

A significant motivation for our chosen metrics (Goals, Aggressiveness, and Control) and the long-term analysis approach stems from empirical observations. A preliminary long-term correlation analysis across the entire decade (2008-2018) revealed a consistently high Pearson correlation between these three performance variables and accumulated league points. This strong correlation suggests that these metrics are robust indicators of overall team performance and strategic tendencies over extended periods, making them ideal candidates for building our relational network.

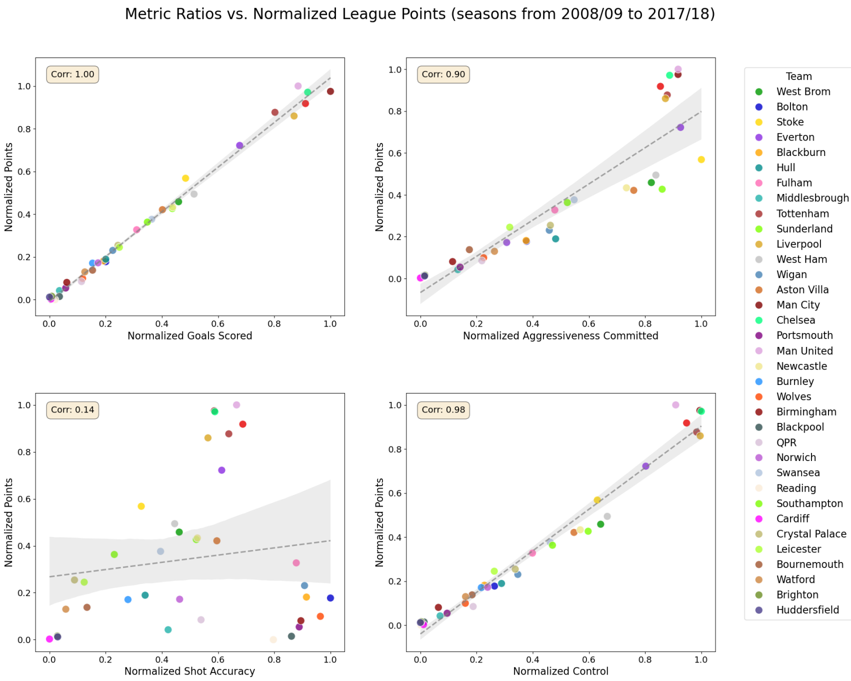

Figura 1: Scatter Plots of Normalized Performance Metrics vs. Total League Points (2008-2018). This figure illustrates the relationships between normalized 'Goals', 'Aggressiveness', 'Control', and 'Shot Accuracy' metrics and the total league points accumulated by teams over the 2008-2018 period. The visual trends support the selection of three of these metrics for network analysis due to their strong correlation with overall team performance('Goals', 'Aggressiveness', and 'Control').

4.2. Experimental Setup: Parameter Configuration for Network Analysis

To address the research questions, we configured the 'perform_epl_network_analysis' pipeline with specific parameters for three distinct scenarios: two single-season analyses and one long-term analysis. The selection of 'keep_above_threshold=False' across all scenarios implies that edges are retained if the absolute difference in the chosen metric between two teams is *less than or equal to* the specified threshold. This configuration prioritizes relationships based on **similarity** in performance. Furthermore, for the single-season analyses, raw absolute differences were used for thresholding, while for the long-term analysis, normalized absolute differences were employed to ensure comparability across diverse seasonal scales.

Scenario 1: 2014/15 Season - Dominance of the Big Six

This scenario investigates a season characterized by the strong performance of the traditional Big Six clubs.

- `analysis_season='2014/15'`: Focuses the analysis solely on the 2014/15 Premier League season.
- `metrics_to_analyze=['goals', 'aggressiveness', 'control']`: The analysis is performed across these three key performance dimensions.
- `thresholds_for_analysis=[1, 5, 6]`:
  - **Goals (threshold=1):** An edge is kept if the absolute difference in total goals scored between two teams in their head-to-head matches during the season is 1 goal or less. This very tight threshold aims to identify teams with highly balanced attacking performances against each other.
  - **Aggressiveness (threshold=5):** An edge is kept if the absolute difference in the calculated aggressiveness metric (fouls + yellow cards + 3×red cards) between two teams is 5 or less. This is a relatively strict threshold, aiming to connect teams with very similar levels of aggressive play.
  - **Control (threshold=6):** An edge is kept if the absolute difference in the control metric (total shots + corners) is 6 or less. This threshold aims to identify teams that exhibited similar levels of offensive pressure and ball dominance when facing each other.
- `keep_above_threshold=False`: Ensures that only highly similar relationships (differences below or equal to the threshold) are considered.
- `use_normalized_abs_diff_for_filter=False`: Uses raw differences for filtering, suitable for single-season analysis where raw counts are directly comparable.

Scenario 2: 2015/16 Season - The Leicester City Phenomenon

This scenario examines the season renowned for Leicester City's unexpected title victory, exploring how this anomaly affects the network structure.

- `analysis_season='2015/16'`: Focuses on the 2015/16 Premier League season.
- `metrics_to_analyze=['goals', 'aggressiveness', 'control']`: Same core metrics as Scenario 1.
- `thresholds_for_analysis=[1, 5, 6]`:

- **Goals (threshold=1):** Identical threshold to Scenario 1, seeking very tight goal-scoring similarities.
- **Aggressiveness (threshold=5):** An edge is kept if the absolute difference in the aggressiveness metric is 5 or less, identical to Scenario 1. This consistent threshold allows for a direct comparison of aggressive profiles across seasons.
- **Control (threshold=6):** An edge is kept if the absolute difference in the control metric is 6 or less, identical to Scenario 1. This consistency facilitates a direct comparison of offensive pressure similarities between the two seasons.

- `keep_above_threshold=False`: Retains similar relationships.
- `use_normalized_abs_diff_for_filter=False`: Uses raw differences for filtering.

Scenario 3: Longitudinal Analysis (2008-2018) - Long-Term Stability

This scenario provides a decade-long perspective, aiming to identify more stable and enduring team characteristics and relationships.

- `network_start_year=2008`, `network_end_year=2018`: Encompasses all seasons from 2008/09 up to and including 2017/18.
- `metrics_to_analyze=['goals', 'aggressiveness', 'control']`: Same core metrics.
- `thresholds_for_analysis=[0.18, 0.22, 0.18]`:
  - **Goals (threshold=0.18):** This threshold is applied to the *normalized absolute goal difference*. A value of 0.18 means that teams are considered *similar* in their goal-scoring profile if their normalized absolute goal difference is within the lowest 18 % of all observed differences across the entire decade.
  - **Aggressiveness (threshold=0.22):** Similarly, teams are connected if their normalized absolute aggressiveness difference is within the lowest 22 %.
  - **Control (threshold=0.18):** Teams are connected if their normalized absolute control difference is within the lowest 18 %.

  The use of normalized thresholds (ranging from 0 to 1) is essential for a multi-season analysis, as it accounts for varying raw metric scales and competitive intensities across different years, ensuring that thresholds represent a consistent proportion of observed differences.

- `keep_above_threshold=False`: Retains similar relationships.
- `use_normalized_abs_diff_for_filter=True`: Crucially, this uses normalized absolute differences for filtering, allowing for consistent interpretation of thresholds across a decade where raw values might vary significantly.

These experimental setups are designed to provide a layered understanding of EPL team dynamics, from specific seasonal anomalies to long-term structural trends, by manipulating the granularity and temporal scope of the network analysis.

5. Results and Discussion

5.1. 2014/15 Season: Unpacking the Big Six Dominance

In the 2014/15 season, Chelsea, Manchester City, and Arsenal secured the top three positions, leading the conventional narrative of Big Six dominance. Our network analysis, however, reveals a more nuanced picture regarding how these top teams interacted based on our chosen metrics.

5.1.1. Centrality Measures and Correlations

The centrality analyses for the 2014/15 season reveal specific teams that were most central across different metrics, along with the correlation of these centralities with overall league points.

Cuadro 1: Top Central Teams and Pearson Correlation with League Points (2014/15 Season)

Centrality Measure	Goals	Aggressiveness	Control
Degree	Leicester (-0.11)	Sunderland (-0.37)	Newcastle (-0.52)
Strength	Leicester (-0.13)	Sunderland (-0.37)	Newcastle (-0.53)
Betweenness	Leicester (-0.16)	Southampton (0.08)	Crystal Palace (-0.31)
Closeness	Leicester (-0.06)	Liverpool (0.05)	Crystal Palace (-0.60)
Eigenvector	Leicester (-0.20)	Sunderland (-0.29)	Sunderland (-0.61)

Note: Each cell shows the team with the highest centrality score for that specific measure and metric. The value in parentheses is the Pearson correlation coefficient between that centrality measure (across all teams) and the teams' final league points for the 2014/15 season.

The table above summarizes the top-performing teams in each centrality measure and the overall Pearson correlation with league points for the 2014/15 season. Across all metrics, the correlations between centrality measures and league points were predominantly negative. This suggests that in a season where Big Six clubs dominated the top of the league, being highly similar to many other teams in terms of these specific metrics (i.e., having many closely contested matches with low differences) generally did not correlate with higher league points. For 'goals', Leicester City consistently showed high centrality across all measures, indicating their involvement in a high number of tightly contested matches, though this did not translate to high league standing in this particular season. Similarly, for 'aggressiveness' and 'control', teams from the mid-to-lower table (Sunderland, Newcastle, Crystal Palace, Liverpool, Southampton) often exhibited the highest centralities, further supporting the idea that a high degree of similarity in these aspects of play did not directly lead to top league performance this year.

5.1.2. Community Dynamics

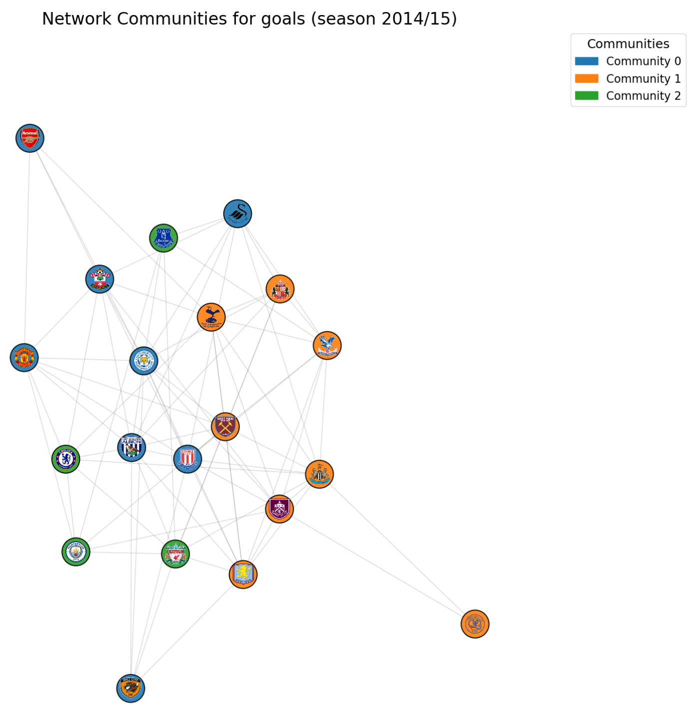

Figura 2: Network Communities for 'Goals' Similarity (2014/15 Season).

Despite their strong league positions, the Big Six consistently failed to form a single, cohesive community across all analyzed metrics. This fragmentation suggests that while they were individually dominant, their underlying playing styles, as captured by goal differentials, aggressiveness levels, or control patterns, were sufficiently diverse to prevent a unified network cluster.

- **Goals:** Top teams are notably dispersed. For instance, Community 0 includes a mix of high-ranking (Arsenal, Man United, Chelsea, Southampton) and mid-to-lower-table teams (Leicester, Stoke, West Brom, Hull). This indicates that their head-to-head matches often resulted in similar, relatively low, goal differences, blurring the lines of 'strength' when viewed through this similarity lens. Notably, Tottenham is isolated in Community 2, suggesting a distinct goal-scoring profile compared to others.
- **Aggressiveness:** The Big Six are similarly fragmented. Community 0, for example, groups Arsenal and Man United with teams like Leicester, Stoke, and West Brom. This broad distribution underlines varied aggressive playing styles, with no single, unified Big Six approach emerging in terms of similar foul rates or card accumulation.
- **Control:** Again, dispersion is evident. Community 0, containing Arsenal, Tottenham, and Man City, also features Stoke and West Brom. This suggests that even among top teams, their approaches to offensive pressure and ball dominance exhibited distinct patterns that prevented a unified cluster.

5.2. 2015/16 Season: The Leicester City Phenomenon Reshaping Dynamics

The 2015/16 season presented a stark contrast, with Leicester City defying all odds to win the Premier League title. This scenario allows us to investigate how such an anomaly is reflected in the network's community structure and centrality dynamics.

5.2.1. Centrality Measures and Correlations

The 2015/16 season presents more varied centrality patterns and intriguing shifts in correlations compared to the previous year.
The analysis of the 2015/16 season's centralities, as summarized in the table above, reveals distinct shifts. For 'goals', teams like West Brom, Swansea, and Bournemouth remain highly central, similar to the previous season, indicating their involvement in a high number of balanced goal difference matches. Leicester City, despite winning the league, does not consistently show the highest centralities in this network, suggesting their success was not primarily driven by high similarity in goal differences with many opponents.
A significant finding emerges in the 'aggressiveness' metric: while Watford and Bournemouth exhibit high degree and strength, Southampton, Leicester, and Watford show high Betweenness and Closeness. Crucially, the Pearson correlations for Betweenness (0.57) and Closeness (0.40) with league points are now positive. This

Cuadro 2: Top Central Teams and Pearson Correlation with League Points (2015/16 Season)

Centrality Measure	**Goals**	**Aggressiveness**	**Control**
Degree	West Brom (-0.07)	Watford (-0.20)	West Ham (-0.35)
Strength	West Brom (-0.08)	Watford (-0.15)	West Ham (-0.36)
Betweenness	Southampton (0.04)	Southampton (0.57)	Everton (-0.42)
Closeness	Norwich (0.19)	Watford (0.40)	Everton (-0.33)
Eigenvector	West Brom (-0.14)	Watford (-0.15)	West Ham (-0.31)

Note: Each cell shows the team with the highest centrality score for that specific measure and metric. The value in parentheses is the Pearson correlation coefficient between that centrality measure (across all teams) and the teams' final league points for the 2015/16 season.

marks a significant shift from 2014/15, implying that in this season, being a bridging team or having short paths to others within the 'aggressiveness' similarity network was highly associated with higher league points. This could indicate that a specific, perhaps pragmatic or adaptable, aggressive style that allowed for consistently competitive (low difference) matches became a viable and even advantageous path to success, which aligns with Leicester's tactical approach.
For 'control', West Ham, Watford, and Everton display the highest centralities. However, similar to 2014/15, the correlations with league points largely remain negative. This suggests that, even in Leicester's unique season, a strategy based on similarity in control patterns (often indicating less dominant possession but perhaps more balanced matches) did not directly translate to higher league points for most teams.

5.2.2. Community Dynamics

Leicester City's unique performance translates into distinct community affiliations, often separating them from traditional top clubs. The fragmentation of the Big Six largely persists.

- **Goals:** Crucially, the champions, Leicester City, are part of Community 0 alongside Chelsea, Man United, Man City, and other diverse teams like Bournemouth and Stoke. This broader community for goals, compared to the previous season, suggests a wider range of teams engaging in matches with similar tight goal differentials, potentially reflecting a more unpredictable season. The Big Six continue to be fragmented, with Arsenal, Tottenham, and Liverpool appearing in Community 1. This reinforces the idea that even Leicester's success did not force the Big Six into a singular goal-based cluster.

- **Aggressiveness:** Leicester is found in Community 2 alongside Bournemouth, Norwich, Arsenal, and West Ham, indicating a shared aggressiveness profile with these teams. This suggests their aggressive play (or lack thereof, if low differences are with less aggressive teams) was part of a broader tactical approach shared by a mix of clubs. The Big Six again show dispersion, highlighting varied approaches to aggressive play within the league.

- **Control:** Leicester prominently features in Community 3 with Norwich, Aston Villa, Swansea, and Crystal Palace. This emphasizes a shared control profile that diverged significantly from the possession-heavy styles often associated with traditional top clubs. Their success with this distinct control strategy suggests a departure from the typical pathways to dominance. The Big Six remain distributed across various communities (e.g., Arsenal, Man City, Liverpool in Community 0; Chelsea, Tottenham in Community 1), reinforcing their varied tactical approaches to ball possession and control.

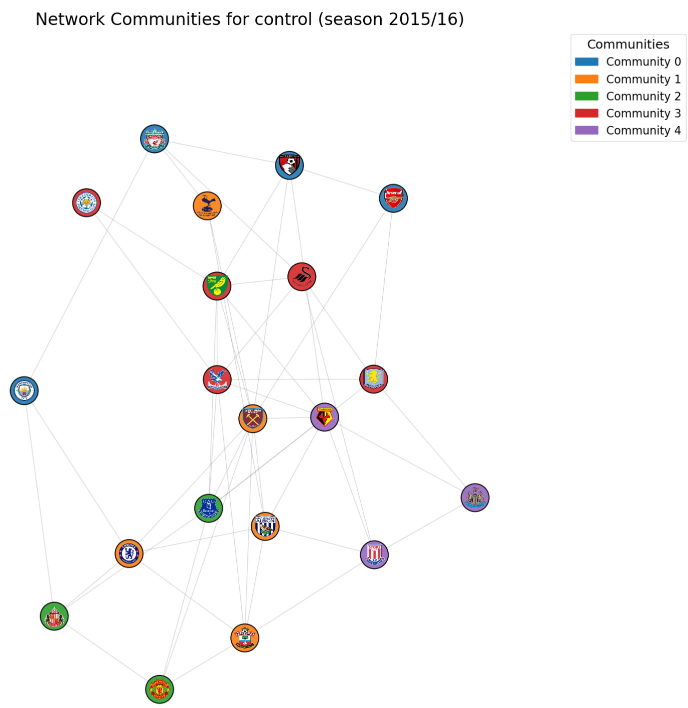

Figura 3: Network Communities for 'Control' Similarity (2015/16 Season).

The 2015/16 analysis highlights that Leicester City's triumph was not necessarily predicted by conventional network centralities across all metrics. Their success is perhaps better understood through their unique community affiliations and the shifting landscape of correlations, particularly the newfound positive relationship between aggressiveness centrality and league points. This suggests that the 'winning formula' could shift, and a style allowing for consistently close matches in aggression became a viable path to success, or at least a characteristic of successful teams. The persistent fragmentation of the 'Big Six' in communities underscores their diverse tactical identities despite their reputation.

5.3. Long-Term Perspective: 2008/09 - 2017/18 Seasons

Extending the analysis to a ten-season period (2008/09 - 2017/18) allows for a more robust examination of enduring network dynamics, mitigating the impact of single-season anomalies like Leicester's title. This aggregated view helps to identify stable patterns in how similarity in goals, aggressiveness, and control relates to sustained performance and the structure of the league. Over this period, traditional powerhouses like Manchester United, Manchester City, and Chelsea consistently accumulated the most points, setting the stage for evaluating their network behavior in a long-term context.

5.3.1. Centrality Measures and Correlations

The long-term analysis of centrality measures provides insights into which teams consistently acted as key players in these similarity networks over a decade, and how these roles correlated with cumulative league points. The use of normalized absolute differences is crucial here, as it allows for meaningful comparison of team-vs-team differences over many seasons, irrespective of overall league trends in scoring or aggression.

Cuadro 3: Top Central Teams and Pearson Correlation with Cumulative League Points (2008/09 - 2017/18)

Centrality Measure	**Goals**	**Aggressiveness**	**Control**
Degree	Middlesbrough (-0.57)	Man City (0.39)	West Brom (-0.52)
Strength	Middlesbrough (-0.59)	Hull (0.35)	West Brom (-0.54)
Betweenness	Brighton (-0.34)	Arsenal (0.29)	Huddersfield (-0.26)
Closeness	Brighton (-0.66)	Newcastle (0.15)	Huddersfield (-0.67)
Eigenvector	Stoke (-0.58)	Hull (0.25)	West Brom (-0.53)

Note: Each cell shows the team with the highest centrality score for that specific measure and metric over the 10-season period. The value in parentheses is the Pearson correlation coefficient between that centrality measure (across all teams) and the teams' cumulative league points from 2008/09 to 2017/18.

The table above summarizes the long-term centrality analysis. A notable trend for the aggregated period is the consistent negative correlation between most centrality measures and league points for both Goals and Control metrics. For Goals, teams like Middlesbrough, West Ham, and Stoke consistently show high centralities, indicating their frequent involvement in matches with tightly contested goal differences over the decade. However, these strong similarities correlate negatively with cumulative points (e.g., Closeness vs. Points: -0.66), suggesting that true long-term dominance in goals is achieved by creating significant differences, not by consistently matching opponents. Similarly, for Control, teams such as West Brom, West Ham, and Newcastle are consistently highly central, yet correlations remain strongly negative (e.g., Closeness vs. Points: -0.67). This reinforces that imposing control and creating large differences in possession or shot attempts, rather than engaging in balanced matches, is characteristic of consistently high-performing teams over time. In contrast, a significant shift occurs for Aggressiveness: correlations are now predominantly positive (e.g., Degree vs. Points: 0.39, Strength vs. Points: 0.35, Betweenness vs. Points: 0.29). This indicates that over a decade, a team's consistent involvement in matches with similar aggression levels (low absolute differences in fouls or cards) positively correlates with higher cumulative league points. Top teams like Manchester City and Arsenal, which demonstrate high centralities here, appear to maintain a certain 'aggressive equilibrium' across many games, suggesting that a disciplined or tactically consistent level of aggression contributes to sustained success, differentiating them from teams with more extreme aggressive profiles.

5.3.2. Community Dynamics

The long-term analysis of community dynamics reveals more pronounced and conceptually intuitive groupings when considering team strength and historical performance. The thresholds used for this multi-season analysis ('0.18' for goals and control, '0.22' for aggressiveness) are normalized, allowing for consistent comparison across different seasons and thus reflecting enduring similarities.

- **Goals:** For the goals metric, a clearer division emerges. Community 0 predominantly comprises Big Six teams (Arsenal, Chelsea, Man United, Liverpool, Tottenham, Man City) along with other consistent Premier League teams that often finished in the upper half of the table (Everton, Swansea, Southampton, Crystal Palace, Leicester, Bournemouth, Watford, Brighton, Huddersfield). This suggests that over a decade, these teams generally exhibited relatively balanced goal differentials in their head-to-head matches, creating a large, cohesive group based on this aspect of play. Community 1 largely consists of teams that frequently yo-yoed between the Premier League and Championship or were typically in the lower half of the table, indicating a distinct long-term goal-scoring similarity profile among them.

- **Aggressiveness:** The communities for aggressiveness also show a tendency to group teams based on their league standing over time. Community 0, for

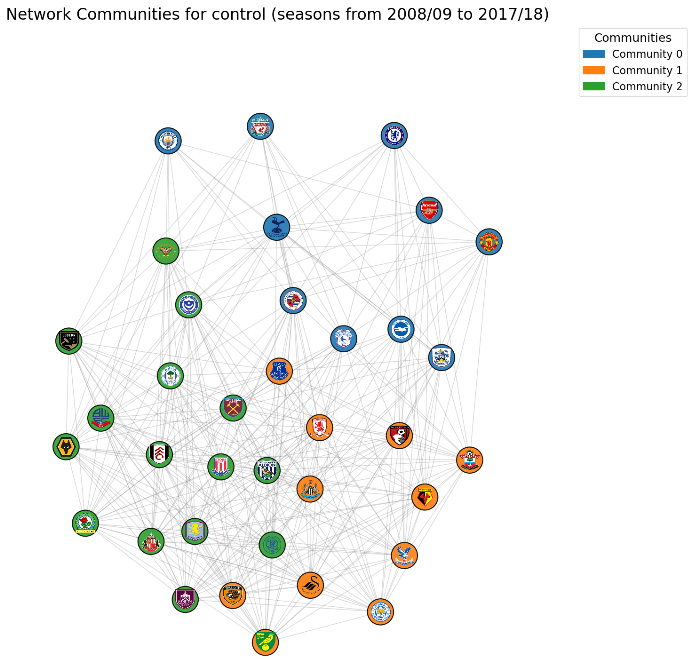

Figura 4: Network Communities for 'Control' Similarity (2008/09 - 2017/2018 Seasons).

instance, includes several Big Six teams (Tottenham) alongside teams that consistently maintained a certain level of pragmatic play (Leicester, Southampton, Swansea, Watford, Brighton, Huddersfield). This might indicate a shared, perhaps less overtly aggressive or more disciplined, approach to play amongst these teams over the long run. The fragmentation, however, still exists, reflecting that even within the Big Six, aggressive styles can vary significantly.

- **Control:** In the control metric, a strong Big Six cluster emerges in Community 0 (Arsenal, Chelsea, Man United, Liverpool, Tottenham, Man City) along with Reading, Cardiff, Brighton, and Huddersfield. This is a significant finding, suggesting that over a decade, these dominant teams consistently demonstrated similar levels of offensive pressure and ball control, forming a more unified network based on this metric. Community 1 groups more mid-table and often relegated teams, further highlighting a separation based on long-term control profiles.

The multi-season aggregation appears to stabilize community detection, leading to more discernible patterns that often align with perceived team strength and long-term league presence, especially for the control metric where a Big Six dominated community is more evident.

5.4. Conclusions

The long-term analysis provides a more robust answer to our initial questions:

- **Grouping by Strength:** While the Big Six don't always form monolithic groups, especially in single-season goals and aggressiveness similarity networks, the long-term view for control demonstrates a stronger tendency for dominant teams to cluster. This suggests that sustained control and offensive pressure are more defining characteristics of long-term strength that lead to network similarity among top clubs, compared to transient goal differences or aggressive outbursts.

- **Impact of Anomalies (Leicester):** While Leicester's single-season triumph in 2015/16 highlighted a unique path, the long-term data contextualizes it. Over a decade, consistent point accumulation still largely belongs to the Big Six who, through their differing styles, nevertheless exhibit similarities that define distinct community structures. Leicester's success was an outlier, not a shift in the fundamental long-term network dynamics for the entire league.

- **Long-Term Improvement in Correlations:** A key observation is the shift in correlations, particularly for aggressiveness. In the short-term 2014/15 season, being similar in aggression was detrimental. In 2015/16, it became beneficial (likely due to Leicester's style). Over the long term (2008/09 - 2017/18), the consistently positive correlations for aggressiveness centralities indicate that a certain level of competitive balance in aggressive play, leading to low differences, is actually a characteristic of teams that achieve sustained success. This suggests that mature, successful clubs either avoid excessive aggression (reducing differences with less aggressive teams) or maintain a tactical, consistent level of aggression that is well-matched by many opponents, allowing them to consistently collect points. This is a significant finding: over time, avoiding highly asymmetric aggressive encounters seems to be part of a winning formula.

In essence, while single seasons can reveal surprising tactical variations (like Leicester's unique profile), the decade-long analysis offers a picture of underlying stability where consistent control defines the top tier, and a balanced approach to aggressiveness emerges as a surprisingly positive correlator with long-term success. The negative correlations for goals and control similarity centralities consistently highlight that true dominance comes from creating *differences*, not similarities, in these metrics.

6. Limitations and Future Work

6.1. Limitations of the Current Study

This network analysis provides valuable insights into EPL team dynamics, yet it's important to acknowledge its inherent limitations. A primary constraint lies in the **granularity of the data**, as using aggregated season-level metrics necessarily smooths out the finer, week-to-week tactical adjustments and individual match nuances that are crucial in football. Furthermore, the **definition of similarity** through fixed thresholds, while consistent, might not capture all forms of relevant relationships between teams, and different thresholding strategies could yield alternative network structures. The **scope of selected metrics** also presents a limitation; while 'goals', 'aggressiveness', and 'control' offer a broad overview, they don't encompass every critical facet of team performance, such as intricate defensive schemes, creative attacking plays, or granular player-level contributions. Lastly, the study's **static analysis** approach, treating networks as fixed snapshots, inherently overlooks the dynamic and evolving nature of football relationships, and the **identified correlations,** though insightful, do not directly imply causality.

6.2. Future Work and Extensions

Building upon this foundational research, several exciting avenues for future work could significantly enhance our understanding of football team dynamics. A crucial step involves transitioning to **dynamic network modeling**, which would allow for the capture and analysis of evolving team relationships and community structures over time, perhaps on a match-by-match or monthly basis. Integrating **multi-layer networks** could provide a more holistic view by incorporating diverse data sources such as player transfers, coaching changes, or even financial metrics, revealing complex interdependencies. Furthermore, the development of **predictive models** leveraging network features like centrality scores or community memberships could move the analysis beyond description to forecasting future team performance. Exploring more **advanced football analytics** for defining similarity would offer deeper, more nuanced insights into tactical profiles. Finally, applying this comprehensive methodology to **other global football leagues** could facilitate comparative studies, identifying universal patterns of team interaction versus those unique to specific football cultures. These extensions promise to offer a richer, more granular, and ultimately more predictive understanding of the beautiful game's intricate dynamics.