UTRECHT UNIVERSITY

Department of Information and Computing Science

Applied Data Science Master Thesis

Using Network Analysis to Visualize Dominant Pass Lines in Football

Supervisor:
Thijs van der Horst

Candidate:
Umut Evren

First examiner:
Marc van Kreveld

In cooperation with:
Forward Football

Second examiner:
Tim Ophelders

July 5, 2024

Acknowledgement

"If we have to travel from point A to point B, everyone would take the six-lane highway and get there as quickly as possible. Everyone, except Riquelme. He would choose the winding mountain road, that takes six hours, but that fills your eyes with scenes of beautiful landscapes" – Jorge Valdano.

From the time I was a child, watching sports has been my passion. The heroes I've admired, the unforgettable moments of sheer joy, and the heart-wrenching dramas have all shaped my love for the game. I want to thank some iconic figures. Pierre Van Hooijdonk, Miroslav Klose,Thiago Alcantara, David Blatt, Gregg Popovic, Tim Duncan, Manu Ginobili,Dirk Nowitzki, İbrahim Kutluay, Šarūnas Jasikevičius, Vincenzo Nibali, Peter Sagan, Usain Bolt, and Michael Phelps. It was a pleasure watching you

Football... What can I say? Probably the most irrational game ever. It's the joy of having a hero like Alex de Souza, the thrill of an iconic goal celebration, the euphoria of Fergie time, the depth of a moment like The Hand of God, and the chant of Ali İsmail Korkmaz,[1]where the story transcends the game itself. The beauty of football lies in its unpredictability and its ability to be so much more than just a sport. This profound complexity is what has always fueled my undying love for the game.

I dedicate my master thesis to the people who choose the winding mountain and the squad of Barcelona 2011-12, the team playing most spectacularly of all time. Blaugrana, Leo Messi and Pep, gracias.

[1]a 19-year-old Turkish university student, succumbed to his injuries after being attacked by police and counter-protesters during the Gezi Park protests, a wave of civil unrest initially sparked by backlash against urban development plan and flared by forceful eviction of a sit-in at the park protesting the plan. Fenerbahçe fans wove his memory into their chants, a chorus that whispered, 'there is a free world in his dreams,' forever etching Ali İsmail Korkmaz's love for the team and his fight for freedom in their hearts.

Abstract

Within the domain of football analytics, pass-based metrics represent one of the most prevalent methodologies employed to investigate the underlying structure of the game. The act of passing in football, by its very nature, presents a multifaceted action amenable to evaluation from various perspectives. Establishing a framework for passing as a geometrical construct involving numerous actors facilitates a deeper understanding of the game's inherent dynamics. The study employs degree, PageRank and betweenness centrality to detect, interpret and visualize dominant pass lines in football, using the data of Go Ahead Eagles' academy players. In sum, while this particular technique may not yield entirely novel insights for the existing literature, PageRank demonstrates its value in pinpointing pivotal passing hubs within the network. Betweenness centrality, on the other hand, exhibits its potential in representing the concept of shortest paths as a means of establishing a different notion of line-breakers.

Contents

1 Introduction **3**

2 Literature Review **6**

2.1 Network Analysis 6

2.2 Network Analysis in Sports 7

3 Data & Methodology **9**

3.1 Data Collection 9

3.2 Methodology 10

4 Results **15**

4.1 Overview of the results 15

5 Conclusion **20**

Bibliography **25**

1. Introduction

The application of data science has permeated the field of sports analytics, extending its reach beyond the confines of the corporate world. A pivotal example occurred in 2001 when Billy Beane, the manager of the Oakland Athletics, recognized and harnessed the power of big data analysis within the realm of professional baseball. Beane's groundbreaking, data-driven management approach fundamentally transformed the sport, relying solely on statistical analysis to inform decisions. The remarkable impact of this approach is chronicled in the book "Moneyball: The Art of Winning an Unfair Game" [1]. "Moneyball" not only revolutionized baseball but also had a profound influence on the broader landscape of high-performance sports [2]. As the confluence of advanced computing capabilities and augmented data accessibility through better recording and availability has garnered increasing recognition as a catalyst for success, the dissemination of quantitative analyses in the realm of sports has been significantly accelerated. This paradigmatic shift has, in turn, precipitated a widespread adoption of data analytics within the elite sports landscape [3].

Emerging technologies, including wearable devices, multiple-camera player trackers, and drone-based analysis of training sessions, are expanding the methodologies for data collection, thereby creating new opportunities for team sports analysis. Considering the sports sector's exponential growth, there's been an intensified focus to harness an ever-expanding trove of data. Furthermore, the leaps made in information technology have not only facilitated this endeavor but have also enabled specialists in the field to refine and develop metrics that are uniquely suited to the nuances of sports [4]. Over the past ten years, advancements in technology—specifically in areas such as automated tracking systems, video-based motion analysis, and Global Positioning System (GPS) units—have significantly expanded our

capacity to identify crucial performance metrics for both individuals and teams [5], [6],[7]. GPS devices, outfitted with sophisticated sensors such as accelerometers, gyroscopes, and magnetometers, provide a comprehensive suite of data that surpasses basic locational (x–y coordinates) and velocity (distance over time) information. These devices meticulously record details of player movements, encompassing acceleration, deceleration, directional changes, and vertical leaps. Armed with this detailed data, strength and conditioning experts can precisely calibrate training intensities, thereby enhancing tactical decisions related to in-game substitutions and player rotations [8]. Additionally, sports analysts have utilized these insights to construct models that estimate injury risk, enabling coaches to receive timely notifications when player workloads approach unsafe levels [9].

In the 2010s, the capability to analyze video data expanded across numerous professional sports. For instance, in the NBA, the SportVU camera system, initially derived from missile-tracking technology, became a mandatory installation, providing detailed tracking for both ball and players, while in baseball, specialized video systems such as PITCHf/x, HITf/x, and FIELDf/x are employed to capture and analyze pitching, hitting, and fielding activities, respectively [8]. The advent of technology has transcended its role from mere observation and analysis, effecting a paradigm shift in the strategic decision-making processes within sports. The insights gleaned from comprehensive data analyses have revolutionized the approach to game planning, influencing not only the tactical choices but also shaping the very nature of gameplay NBA through the number of 3-points [10], [11].The graphical representation of the trajectory and shoot disposition of the ball facilitated an enhanced strategic distribution among the team members during offensive maneuvers. Furthermore, it provided insights into the most advantageous stances for intercepting the ball post-shot, contingent upon the identity of the shooter and the precise location of the shot's origin [12]. Statistical analysis of team sports data is finding practical applications in real-world situations [13]. The evaluation of athletic performance has witnessed a surge in the application of player tracking data [14], particularly within the domain of football [15], [16]. The extraction

and analysis of advanced sports metrics generate valuable insights that aid decision-makers in formulating strategies [17], managing budgets, developing tactics, and enhancing training programs, all aimed at reducing risks and mitigating financial losses [8].

Football's most fascinating feature is its emergent properties [18]. Emergence is described as the formation of new and coherent structures, patterns and properties during the self-organization process in self-complex systems [19]. The complexity of football arises primarily from the numerous interactions between teammates and opponents, but also the context of the game itself [20]. Consequently, a football team represents a complicated, dynamic, nonlinear and adaptable system made up of 11 players, each of whom is also a system. The notion of nonlinearity stems from the idea that the team as whole is not merely as the sum of its players, implying that understanding a football team requires more than just analyzing its players [21]. The complexity of football, characterized by the continuous flow of the ball and typically low scores, renders simple statistics like goals, shots, or assists inadequate for accurately measuring player and team performance [22]. Such statistics hardly reflect a player's true impact on a match's outcome. Instead, a player's actual contribution is often embedded in the team's plays, such as recovering possession or making a decisive pass leading to an assist, indicating that performance metrics should consider these subtler aspects of the game [23].

This paper aims to deploy network analysis to visualize the dominant passing lines on the football pitch, providing some centrality metrics like degree, betweenness and Page Rank to create metrics regarding players'roles in passing network. The inherent abstractness of isolated metrics and the potential for data visualizations to lack contextual richness necessitate a balanced approach that leverages the strengths of both. This research bridges the gap by employing a complementary strategy that integrates these two analytical tools.

2. Literature Review

2.1 Network Analysis

In recent years, there has been a remarkable increase in academic interest directed toward the implementation of networks, encompassing domains such as computer networks, biological networks and social networks. Given that each complex system is supported by a sophisticated network that encapsulates the interactions among its components, the prevalence of networks in science, technology, business and nature is significantly more extensive than what might be perceived at cursory glance [24]. Additionally, it involves mathematical representation of networks, such as stochastic graph models and generative frameworks, pertaining to the theoretical constructs of dynamic phenomena occurring within network structures [25].

Over the years, the intricate interrelations among the agents within social networks have been a focal point in the domain of social network analysis. The concept of centrality has been extensively examined within this analytical framework [26]. Discussions frequently concentrate on undirected, unweighted social networks, adopting a simplified perspective. A consistent definition of an agent's centrality remains elusive even within the confines of these graphs which could be considered as simple [27]. Rather, a diverse array of concepts and interpretations specific to the context of a node's centrality exist, which might arise from varying goals associated with the application of centrality measures. Consequently, a variety of centrality metrics have been formulated, including "betweenness" [28], "eigenvector centrality" [29] and "closeness" [30] among others.

2.2 Network Analysis in Sports

Numerous academic studies have utilized data mining and network science to decipher the intricate success patterns found in both individual and team sports. In a pioneering large-scale data-driven investigation, Cintia and his colleagues scrutinized the workout routines of 30,000 amateur cyclists, sourced from a widely-used fitness social network application [31]. Their findings revealed a consistent pattern in cyclists' training and performance, culminating in the creation of an effective training regimen derived entirely from empirical data. In football, team tactics are shaped by an intricate process derived from a product of a myriad of interrelated variables [32]. Through the lens of network science, one can perceive the structuring of a team as an emergent phenomenon, based on the dynamic interplay among its members, which in turn gives rise to complex passing networks [33].

The foundational principle of this methodology lies in conceptualizing the teams members as nodes within a network, with the interactions between them –specifically the passes exchanged — being represented as weighted edges. The frequency of passes between any two players is what dictates the weight of these edges [34]. Such a modeling of the team's passing dynamics facilitates the straightforward identification of pivotal players, who are characterized by their extensive connectivity and the higher weights of their associated vertices [35]. By examining the structure of passing networks, one can discern patterns of repeated pass sequences that are indicative of a team's distinctive playing style [36]. These networks, when viewed collectively, reveal a small-world topology characterized by a pronounced clustering coefficient – evidenced by the frequent formation of triangular pass sequences among the three players [37],[38]. This is in sharp contrast to a random null model and is further underscored by the minimal number of steps required to connect any two nodes within the network [39]. Moreover, the detection of motifs – defined by an excessive frequency of specific pass sequences among trios or quartets of players, or within tightly-

knit player communities– remains feasible [40], [41], [42]. In terms of individual players, network motifs function as a tool to map out a player's role within a team and to identify counterparts in other teams who have similar characteristics [43].In a study deploying Temporal Pattern (T-Pattern) detection, recurring sequences of passes within game footage are identified [44]. Another analysis employed clustering techniques on players' on subtrajectories, with the Fréchet distance serving as the metric for similarity [45].

3. Data & Methodology

3.1 Data Collection

The data collection process for this study deployed Local Positioning Measurement (LPM) of Forward Football, which is based on an Ultra Wide Band (UWB) network to precisely track the positions and movements of players and the ball on a football pitch. As for UWB configuration, the system consists of a network of six antennas strategically placed around a full football pitch. These antennas operate at a frequency of 100Hz and are connected to small trackers worn by the players which transmit both heart rate and position data at a rate of 15 times per second. In terms of tracking, each player is equipped with a tracker that continuously transmits their positional data to the network. The algorithm used for determining the position of players and the ball was rigorously tested through the comparisons of the computed positions and against actual ground positions obtained from video recordings. The accuracy of the positioning algorithm was quantified using the Mean Squared Error (MSE) metric, evaluated across various scenarios on the football pitch. The LPM systems facilitated the collection of comprehensive football statistics including tackles, interceptions, passing accuracy, on-target-rates, xT values and so on. This study analyzes entirely successful passes completed by Go Ahead Eagles' academy players. The data was provided by Forward Football. The focus on successfully completed passes is due to the limitations of unsuccessful passes in offering meaningful context for game control analysis.

3.2 Methodology

A network can be formally described as an ordered triple (N, E, W), consisting of a finite set of nodes N, a finite set of edges $E \subseteq N \times N$, and a set of weights W. An ordered pair of vertices, denoted by a function of ψ: $\psi(v_i, v_j) = e_{ij} \subseteq E$, maps the directed connection from the node v_i to node v_j. The function $\omega : \mathcal{E} \to \mathbb{R}$ assigns a weight to each arc, reflecting the strength or cost associated with it. This network can be represented as an adjacency matrix $P = (p_{ij})$, where $p_{ij} = \omega(e_{ij})$ if $e_{ij} \in \mathcal{E}$, and $p_{ij} = 0$ otherwise for $i \neq j$. Self-loops, or arcs connecting a vertex to itself, are not allowed, ensuring $p_{ij} = 0$ when $i = j$.
In a practical application such as sports analytics, the network model can be visualized with players as nodes and passes between them as arcs. In an unweighted network, each pass between players i and j is represented in the adjacency matrix A by $A_{ij} = 1$ if there is a pass, and $A_{ij} = 0$ otherwise. This matrix would typically be of size 11×11 for a soccer team with 11 players, assuming no substitutions. Alternatively, in a weighted network, the strength of each pass is considered, and A_{ij} is assigned a real number corresponding to this strength, allowing for a more detailed analysis of player interactions and performance.

In football analytics and network analysis, the adjacency matrix is a fundamental tool used to represent the relationships and interactions between players on the field. This matrix is a square grid where each row and column corresponds to a player, and the elements within the matrix indicate the presence and strength of interactions between the players. Specifically, in the context of football, the entries in the adjacency matrix typically represent passes between players. A value of zero indicates no pass, while a positive value signifies the number of passes exchanged. The adjacency matrix provides a structured way to analyze passing networks, enabling to quantify the cohesiveness and efficiency of a team's play. By examining the adjacency matrix, key players who act as hubs in the network can be identified, facilitating ball movement and maintaining possession. Furthermore,

advanced metrics such as centrality measures can be derived from the adjacency matrix, offering insights into each player's influence and connectivity within the team.

3.2.1 Degree Centrality

In the domain of graph theory, degree centrality (DC) offers a localized assessment of a vertex's prominence. It measures a vertex's importance exclusively based on the immediate connections it possesses, forgoing consideration of the network's global structure. This intuitive measure essentially reflects the normalized quantity of links incident upon a vertex, as expressed in equation (1).

$$C_D(v_i) = \frac{1}{N-1} \sum_{j=1}^{N} \alpha_{i,j} \tag{1}$$

In this formula, $C_D(v_i)$ is the degree centrality of vertex v_i, N is the total number of vertices in the graph, and $a_{i,j}$ is an element of the adjacency matrix that equals 1 if there is a link between vertices i and j, and 0 otherwise. The sum $\sum_{j=1}^{N} a_{i,j}$ counts the number of links incident on v_i, and the factor $\frac{1}{N-1}$ normalizes this count by the maximum possible degree a vertex can have, which is $N-1$. This formula provides a normalized measure of the number of direct connections that vertex v_i has in the network.Notably, the concept directly aligns with the more general term "degree" used throughout graph theory.

3.2.2 Betweenness

Betweenness centrality quantifies the degree to which a node is positioned on the shortest paths between other nodes in a graph. Nodes with high betweenness centrality wield significant influence within a network due to their role in controlling the flow of information between other nodes. The algorithm determines the shortest paths between all pairs of nodes in the graph, assigning a score to each node based on the number of these shortest paths that traverse it. Consequently, nodes that frequently appear on the shortest paths between others receive higher betweenness centrality scores. These nodes are crucial in maintaining network communication; their removal would significantly disrupt information flow, as they are situated on numerous paths that messages typically follow. Therefore, if a node with high betweenness is removed, all messages that previously passed through it must be rerouted via alternative paths (Newman, 2010)

Consider a directed graph $G = \langle V, E \rangle$, where V denotes the set of vertices and E denotes the set of edges. Let $\sigma(x, y)$ represent the number of shortest paths between vertices x and y, and let $\sigma(x, y \mid v)$ represent the number of shortest paths between vertices x and y that pass through vertex v.

$$C_B(v_i) = \sum_{x,y \in V} \frac{\sigma(x, y \mid v_i}{\sigma(x, y)} \tag{2}$$

where $C_B(v_i)$ represents the betweenness centrality of vertex v_i. The term $\sigma(x, y)$ denotes the number of shortest paths between vertices x and y, while $\sigma(x, y \mid v_i)$ represents the number of those shortest paths that pass through vertex v_i. Notably, $\sigma(x, y \mid v_i) = 0$ if v_i is one of the vertices x or y, i.e., $v_i \in \{x, y\}$.This formula calculates the average fraction of shortest paths between all pairs of nodes that pass through v.

- If $x = y$, then $\sigma(x, y) = 1$.
- If $v \in \{x, y\}$, then $\sigma(x, y \mid v_i) = 0$.

The number of shortest paths $\sigma(x, y)$ can be calculated recursively:

$$\sigma(x, y) = \sum_{u \in \text{Pred}(y)} \sigma(x, u) \tag{3}$$

where $\text{Pred}(y)$ is the set of predecessors of y on the shortest path from x:

$$\text{Pred}(t) = \{u : (u, y) \in E, d(x, y) = d(x, u) + 1\} \tag{4}$$

Here, $d(x, u)$ represents the distance between nodes x and u.

3.2.3 Page Rank Centrality

PageRank, developed by Google's founders,is another centrality method specifically designed for web search which measures the importance of webpages based on the structure of hyperlinks on the web. The core idea is to assign a score of importance to each webpage, assuming that a page is more important if it has many incoming links from other important pages. PageRank also considers the significance of the pages providing links, positing that links from more important pages carry greater weight and thus enhance the ranking of the linked pages. Consequently, the importance of a page is determined by the aggregate value of the votes from its incoming links. This mechanism ensures that links from highly ranked pages contribute more substantially to the ranking of other pages, thereby refining the overall ranking process.

Let $A = (a_{i,j})$ be the adjacency matrix representing a directed graph. The PageRank centrality x_i of a node i can be expressed as:

$$x_i = \alpha \sum_k \frac{a_{k,i}}{d_k} x_k + \beta \tag{5}$$

In this formula, α and β are constants, while d_k denotes the out-degree of node k. If node k has no outgoing links, d_k is defined to be 1. When represented in matrix form, the equation becomes:

$$[x = \alpha x D^{-1} A + \beta \tag{6}$$

Here, β is a vector where each element is a constant positive value, and D^{-1} is a diagonal matrix with the i-th diagonal element being $1/d_i$.PageRank includes an endogenous component, which depends on the network's topology, and an exogenous component, which is independent of the network structure. Thus, x can be computed as:

$$x = \beta(I - \alpha D^{-1} A)^{-1} \tag{7}$$

Algorithm 1 PageRank Calculation

function PAGERANK(M, d = 0.85)
    # Parameters:
    # M: adjacency matrix
    # d: damping factor, default is 0.85
    $N \leftarrow$ number of columns in M
    $w \leftarrow$ initialize a vector of size N with each element equal to $\frac{1}{N}$
    $M\_hat \leftarrow d \times M$
    $v \leftarrow M\_hat \times w + (1 - d)$
    while $\sqrt{\sum (w - v)^2} \geq 1 \times 10^{-10}$ **do**
        $w \leftarrow v$
        $v \leftarrow M\_hat \times w + (1 - d)$
    end while
    return v
end function

4. Results

4.1 Overview of the results

Given that the initial form of the match data does not constitute network data, an adjacency matrix was generated from the pass data using a custom Python function. Since the rows represent the passed player and the columns represent the receiver, the adjacency matrix (Figure 4.1) does not have symmetrical structure

Algorithm 2 Create Adjacency Matrix

```
procedure CREATE_ADJACENCY_MATRIX(passes, player_ids)
    matrix ← DataFrame initialized with zeros, index and columns as player_ids
    for all row in passes do
        matrix[passedPlayerId][receivedPlayerId] ← matrix[passedPlayerId][receivedPlayerId] + 1
    end for
    return matrix
end procedure
```

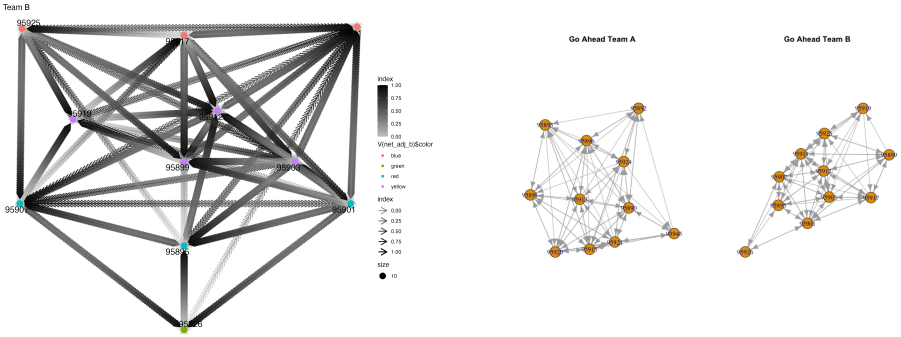
Adjacency Pass Matrix

Team A

Index	Player	ID 95892	ID 95913	ID 95896	ID 95890	ID 95915	ID 95891	ID 95945	ID 95921	ID 95920	ID 95893	ID 95924	Total Successful Passes	Total Number of Passes
95892	Adam Benbouker	0	0	6	2	1	0	0	0	0	2	4	16	19
95913	Enes Gundogdu	2	0	3	6	6	2	0	4	2	1	2	28	38
95896	Eser Gurbuz	2	2	0	2	1	3	0	0	1	1	9	22	49
95890	Figo Ausems	3	7	2	0	5	2	9	13	2	0	12	60	73
95915	Jermaine Yayo Hoogeveen	1	7	0	6	0	4	1	17	5	1	0	45	53
95891	Kaid Phiri Becker	1	3	4	0	1	0	0	3	1	1	0	15	20
95945	Lucas Luisman	2	0	0	11	2	0	0	4	1	0	0	20	25
95921	Mikai Ozcan	0	6	4	18	20	1	2	0	4	1	2	58	67
95920	Sadiq Moennoe	0	4	3	4	1	2	0	2	0	0	2	19	24
95893	Sven ten Bokum	0	1	0	2	2	1	1	0	2	0	1	11	14
95924	Yari Westerbeek	3	4	7	5	1	1	1	6	1	1	0	33	39

Figure 4.1: Team A's adjacency matrix

The initial visualizations employ the ggraph and igraph packages' basic plotting functionalities. However, as depicted in Figure 4.2, these ap-

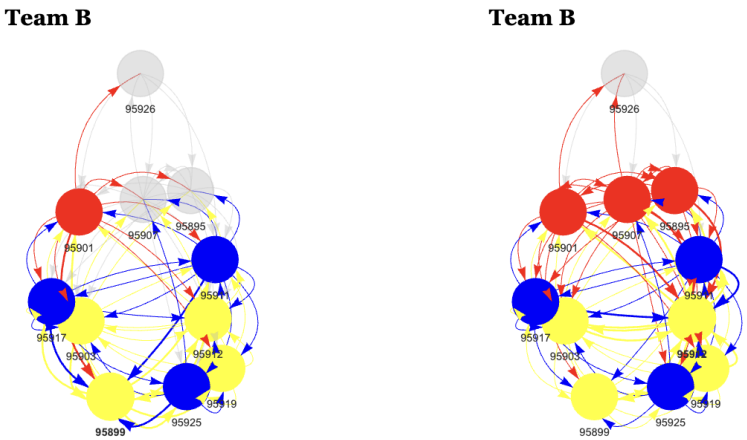


(a) ggraph implementation **(b)** igraph plot

Figure 4.2: Comparison of two different times

proaches demonstrate a limited capacity for effectively conveying interpretable results pertaining to dominant pass lines.

In contrast, the second stage adopts a distinct approach implemented through visNetwork library, which is a R package using JavaScript components to visualize networks, [46]. Figure 4.3 illustrates the networks of two contrasting players: one with an exceptionally high value and the other with an exceptionally low value. The network on the left, representing the player with the lowest degree centrality, exhibits a limited number of connections. Conversely, the network on the right, representing the player with the highest degree centrality, demonstrates a dominant passing relationship with a multitude of other players.



(a) The Representation of Player 95899 **(b)** The Representation of Player 95912

Figure 4.3: Comparison of how players at the opposing ends diverge.

Desaturated colors for minimally involved players and the use of directional arrows significantly enhance the clarity of this visualization. This approach minimizes visual clutter, allowing users to easily focus on key player interactions and recurring patterns. The arrows intuitively communicate both the direction and frequency of passes through their direction and density.

Team A (PR)

Node	PageRank
95890	0.15599392
95921	0.14179249
95915	0.11449622
95924	0.11203056
95913	0.10668841
95896	0.10145158
95920	0.06442495
95891	0.06113434
95892	0.05521228
95945	0.04825527
95893	0.03851999

(a) Page Rank for Team A

Team B (PR)

Node	PageRank
95912	0.16494595
95911	0.12396670
95925	0.11089680
95903	0.10501199
95907	0.09803606
95895	0.09472388
95901	0.08409635
95917	0.07907170
95899	0.05614033
95919	0.05560319
95926	0.02750705

(b) Page Rank for Team B

Figure 4.4: Comparison of two different times

Figure 4.4 depicts a comparative ranking of both teams' PageRank centrality, encoding color differentiation to represent centrality levels. Players identified by IDs 95890, 95921, and 95912 emerge as prominent pass hubs, playing a decisive role in facilitating pass traffic, given that the influence a node gains from being linked to a central node diminishes proportionally if the central node distributes its "links" generously across a large number of other nodes[47].[1]

[1]Therefore, this element functions as a critical junction within the network, analogous to a station. Unlike a typical node, it's not solely defined by the number of connections or their distribution but acts as a central hub for channeling key interactions.

Team A (Betweenness)

Node	Betweenness
95893	25.495238
95920	15.059524
95915	14.266667
95891	8.169048
95945	4.700000
95892	2.726190
95896	2.233333
95890	2.000000
95924	1.069048
95913	0.750000
95921	0.000000

Team B (Betweenness)

Node	Betweenness
95911	26.8166667
95907	11.4666667
95899	11.3750000
95925	11.0666667
95917	9.1750000
95919	6.7083333
95901	5.1916667
95895	1.0000000
95903	0.8333333
95926	0.6666667
95912	0.0000000

Figure 4.5: Comparison of Betweenness

Betweenness centrality reveals a striking contrast compared to the findings from the previous method. Notably, the player ranked lowest in PageRank centrality for Team A emerges as the team's leader in betweenness centrality. While the disparity may not be as pronounced for all players, a similar trend is observed for players 95920 and 95907.The exceptional performance of player 95911 (Cem Eroglu) in both PageRank and betweenness centrality suggests a potential outlier. This unique scenario warrants further investigation into the influence of a "roaming" role. Such a role might involve the player receiving the ball in deeper positions and then facilitating the ball movement forward. In simpler terms, by actively covering specific areas of the field, this player functions as a key passing station and potentially possesses the shortest path lengths for ball distribution.

While centrality metrics visualized through intricate network graphs remain valuable tools for network analysis, chord diagrams offer a complementary approach to data storytelling. By creating a radial layout, chord diagrams effectively communicate the interrelatedness between data points, fostering a clearer understanding of the underlying relationships.

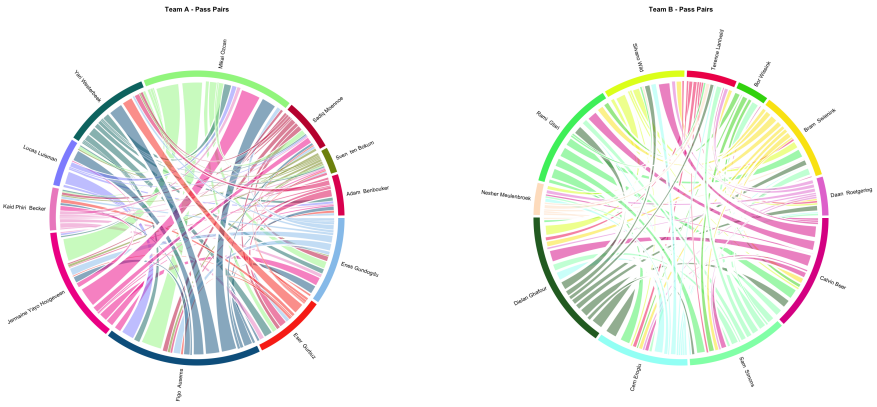


Figure 4.6: Comparison passing pairs via chord diagram

Figure 4.6 employs a chord diagram to depict pass pairs. The values, which represent number of passes betweeen each pairs, of for each team are normalized internally, allowing for a clear comparison of pass frequencies within each team. This normalization reveals Team A's dominance in terms of pass pairs. Since Team A has more dominant pairs, its links representing passes are relatively thicker, whereas Team B does not have dominant duos like Figu Ausems-Yari Westerbeek or Hoogeveen-Mikai Ozcan.

The implementations are open to public: github.com/umutevren

5. Conclusion

This research underscores the value of applying network analysis to football. It allows for both interpretation and visualization of dominant pass lines, offering a comprehensive approach to understanding passing patterns. Traditionally, player passing impact has been evaluated solely through degree centrality, which simply reflects the number of passes made. This paper proposes an alternative method by deploying PageRank and betweenness centrality. PageRank plays a critical role in pinpointing critical pass hubs through iterative calculations, while betweenness centrality focuses on identifying "shortest paths" across the pitch. In football terminology, a player with high betweenness centrality can be likened to a *regista-esque* weaving short connections and facilitating ball movement between teammates. Overall, this study concludes that degree centrality in passing can be regarded as a network representation of the mainstream approach to passes, as it primarily considers the number of nodes. Conversely, PageRank is useful for identifying critical passing lanes and hubs. Finally, betweenness centrality has the potential to provide a novel framework for line-breaking passes [2]by combining the shortest path with other metric routes.

This study acknowledges several limitations that present opportunities for future research. First, the analysis considers only successful passes. However, a received pass may be quickly lost, rendering the initial pass inconsequential. Consequently, this study does not account for the actions that follow the pass. In the real world of football, the action following a pass significantly impacts its evaluation. For example, a successful pass into the opponent's half, which consists of smaller spaces and requires a higher level

[2]a line-breaking pass is one that penetrates the opposition's defensive formation and advances the ball significantly forward. For more detail

of skill to execute[3], followed by a successful dribble, can pose a significant threat to the opposition. Due to limitations in data consistency between passing data and other action datasets (such as dribbles, interceptions, etc.), this study does not incorporate any follow-up actions.

[3]For a more in-depth exploration of crucial position-specific actions, check notion of Juego posicional

Bibliography

[1] M. Lewis, *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton, 2004.

[2] B. Gerrard, "Is the moneyball approach transferable to complex invasion team sports?" *International journal of sport finance*, vol. 2, pp. 214–230, Nov. 2007.

[3] I. G. McHale and S. D. Relton, "Identifying key players in soccer teams using network analysis and pass difficulty," *Eur. J. Oper. Res.*, vol. 268, pp. 339–347, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:3572927.

[4] L. Gyarmati and M. Hefeeda, "Analyzing in-game movements of soccer players at scale," *ArXiv*, vol. abs/1603.05583, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:18229111.

[5] M. Herold, F. Goes-Smit, S. Nopp, P. Bauer, C. Thompson, and T. Meyer, "Machine learning in men's professional football: Current applications and future directions for improving attacking play," *International Journal of Sports Science Coaching*, vol. 14, Oct. 2019. DOI: 10.1177/1747954119879350.

[6] A. Rossi, L. Pappalardo, P. Cintia, F. M. Iaia, J. Fernàndez, and D. Medina, "Effective injury forecasting in soccer with gps training data and machine learning," *PloS one*, vol. 13, no. 7, e0201264, 2018. DOI: 10.1371/journal.pone.0201264. [Online]. Available: https://doi.org/10.1371/journal.pone.0201264.

[7] C. Carling, R. Thomas, and A. Williams, "Performance assessment in field sports," *Journal of Sports Science and Medicine*, vol. 8, Mar. 2009.

[8] E. Morgulev, O. Azar, and R. Lidor, "Sports analytics and the big-data era," *International Journal of Data Science and Analytics*, vol. 5, Jun. 2018. DOI: 10.1007/s41060-017-0093-7.

[9] G. Fried and C. Mumcu, Eds., *Sport Analytics: A data-driven approach to sport business and management*. Routledge, 2017, ISBN: 9781138667136.

[10] Z. Lowe, "Lights, cameras, revolution," *Grantland*, Mar. 2013. [Online]. Available: https://grantland.com/features/the-toronto-raptors-sportvu-cameras-nba-analytical-revolution/.

[11] T. Zając, K. Mikołajec, P. Chmura, M. Konefał, M. Krzysztofik, and P. Makar, "Long-term trends in shooting performance in the nba: An analysis of two- and three-point shooting across 40 consecutive seasons," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, p. 1924, 2023, Published 2023 Jan 20. DOI: 10.

3390/ijerph20031924. [Online]. Available: https://doi.org/10.3390/ijerph20031924.

[12] K. Goldsberry, "Courtvision : New visual and spatial analytics for the nba," 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:4205086.

[13] J. Albert, J. Bennett, and J. J. Cochran, *Anthology of Statistics in Sports* (ASA-SIAM Series on Statistics and Applied Probability). SIAM, 2005, vol. 16, p. 327, ISBN: 0898718384.

[14] B. Drust, "Performance analysis research: Meeting the challenge," *Journal of Sports Sciences*, vol. 28, no. 9, pp. 921–922, Jul. 2010. DOI: 10.1080/02640411003740769.

[15] V. Di Salvo, F. Pigozzi, C. González-Haro, M. S. Laughlin, and J. K. De Witt, "Match performance comparison in top english soccer leagues," *International Journal of Sports Medicine*, vol. 34, no. 6, pp. 526–532, Jun. 2013, Epub 2012 Nov 26. DOI: 10.1055/s-0032-1327660.

[16] M. Bush, C. Barnes, D. T. Archer, B. Hogg, and P. S. Bradley, "Evolution of match performance parameters for various playing positions in the english premier league," *Human Movement Science*, vol. 39, pp. 1–11, Feb. 2015, Epub 2014 Nov 18. DOI: 10.1016/j.humov.2014.10.003.

[17] A. Adhikari, A. Majumdar, G. Gupta, and A. Bisi, "An innovative super-efficiency data envelopment analysis, semi-variance, and shannon-entropy-based methodology for player selection: Evidence from cricket," *Annals of Operations Research*, vol. 284, Jan. 2020. DOI: 10.1007/s10479-018-3088-4.

[18] Y. Yamamoto and K. Yokoyama, "Common and unique network dynamics in football games," *PLoS ONE*, vol. 6, no. 12, e29638, Dec. 2011. DOI: 10.1371/journal.pone.0029638. [Online]. Available: https://doi.org/10.1371/journal.pone.0029638.

[19] J. Goldstein, "Emergence as a construct: History and issues," *Emergence*, vol. 1, pp. 49–72, Mar. 1999. DOI: 10.1207/s15327000em0101_4.

[20] A. Ric, S. Robertson, and D. Sumpter, Eds., *Football Analytics 2021: The Role of Context in Transferring Analytics to the Pitch*. Barcelona, Spain: Barça Innovation Hub, 2021. [Online]. Available: https://barcainnovationhub.com.

[21] C. Willy, E. A. Neugebauer, and H. Gerngroß, "The concept of nonlinearity in complex systems," *European Journal of Trauma and Emergency Surgery*, vol. 29, pp. 11–22, 2003. DOI: 10.1007/s00068-003-1248-x.

[22] J. Peña and H. Touchette, "A network theory analysis of football strategies," *In C. Clanet (ed.), Sports Physics: Proc. 2012 Euromech Physics of Sports Conference, p. 517-528, 'Editions de l'École Polytechnique, Palaiseau, 2013. (ISBN 978-2-7302-1615-9)*, 2012. arXiv: 1206.6904 [math.CO].

[23] J. Duch, J. S. Waitzman, and L. A. N. Amaral, "Quantifying the performance of individual players in a team activity," *PLOS ONE*, vol. 5, no. 6, e10937, 2010. DOI: 10.1371/journal.pone.0010937.

[24] A.-L. Barabasi, *Network Science*. Cambridge, UK: Cambridge University Press, 2016, ISBN: 978-1107076266.

[25] M. Newman, *Networks: An Introduction*. Oxford: Oxford University Press, 2010, ISBN: 9780199206650. DOI: 10.1093/acprof:oso/9780199206650.001.0001. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780199206650.001.0001.

[26] D. Lusher, G. Robins, and P. Kremer, "The application of social network analysis to team sports," *Measurement in Physical Education and Exercise Science*, vol. 14, no. 4, pp. 211–224, 2010. DOI: 10.1080/1091367X.2010.495559.

[27] S. P. Borgatti and M. G. Everett, "A graph-theoretic perspective on centrality," *Social networks*, vol. 28, no. 4, pp. 466–484, 2006.

[28] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978, ISSN: 0378-8733. DOI: https://doi.org/10.1016/0378-8733(78)90021-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0378873378900217.

[29] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *The Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972. DOI: 10.1080/0022250X.1972.9989806.

[30] M. A. Beauchamp, "An improved index of centrality," *Behavioral Science*, vol. 10, no. 2, pp. 161–163, 1965. DOI: 10.1002/bs.3830100205.

[31] P. Cintia, L. Pappalardo, and D. Pedreschi, ""engine matters": A first large scale data driven study on cyclists' performance," Dec. 2013. DOI: 10.1109/ICDMW.2013.41.

[32] M. Kempe, D. Memmert, S. Nopp, and M. Vogelbein, "Possession vs. direct play: Evaluating tactical behavior in elite soccer," *International Journal of Sport Science*, vol. 4, pp. 35–41, Nov. 2014. DOI: 10.5923/s.sports.201401.05.

[33] J. Buldu, J. Busquets, J. Martínez, *et al.*, "Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game," *Frontiers in Psychology*, vol. 9, p. 1900, Oct. 2018. DOI: 10.3389/fpsyg.2018.01900.

[34] P. Passos, K. Davids, D. Araujo, N. Paz, J. Minguéns, and J. F. Mendes, "Network as a novel tool for studying team ball sports as complex social system," *Journal of science and medicine in sport / Sports Medicine Australia*, vol. 14, pp. 170–6, Dec. 2010. DOI: 10.1016/j.jsams.2010.10.459.

[35] J. Gama, P. Passos, K. Davids, *et al.*, "Network analysis and intra-team activity in attacking phases of professional football," *Interna-*

tional Journal of Performance Analysis in Sport, vol. 14, Dec. 2014. DOI: 10.1080/24748668.2014.11868752.

[36] L. Gyarmati and X. Anguera, "Automatic extraction of the passing strategies of soccer teams," *arXiv preprint arXiv:1508.02171*, 2015. DOI: 10.48550/arXiv.1508.02171. [Online]. Available: https://arxiv.org/abs/1508.02171.

[37] T. Narizuka, K. Yamamoto, and Y. Yamazaki, "Statistical properties of position-dependent ball-passing networks in football games," *Physica A: Statistical Mechanics and its Applications*, vol. 412, Nov. 2013. DOI: 10.1016/j.physa.2014.06.037.

[38] C. Cotta, A. Mora, C. Molina, and J. Merelo Guervós, "Fifa world cup 2010: A network analysis of the champion team play," *CoRR*, vol. abs/1108.0261, Aug. 2011. DOI: 10.1007/s11424-013-2291-2.

[39] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998. DOI: 10.1038/30918.

[40] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science (New York, N.Y.)*, vol. 298, pp. 824–7, Nov. 2002. DOI: 10.1126/science.298.5594.824.

[41] L. Gyarmati, H. Kwak, and P. Rodriguez, "Searching for a unique style in soccer," *arXiv preprint arXiv:1409.0308*, 2014. DOI: 10.48550/arXiv.1409.0308. [Online]. Available: https://arxiv.org/abs/1409.0308.

[42] F. M. Clemente, M. S. Couceiro, F. M. L. Martins, and R. S. Mendes, "Using network metrics in soccer: A macro-analysis," *Journal of human kinetics*, vol. 45, pp. 123–134, 2015.

[43] J. Peña and R. Navarro, "Who can replace xavi? a passing motif analysis of football players," Jun. 2015.

[44] A. Borrie, G. Jonsson, and M. Magnusson, "Temporal pattern analysis and its applicability in sport: An explanation and exemplar data," *Journal of sports sciences*, vol. 20, pp. 845–52, Nov. 2002. DOI: 10.1080/026404102320675675.

[45] J. Gudmundsson and T. Wolle, "Towards automated football analysis: Algorithms and data structures," 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:15369923.

[46] A. B.V., Contributors, and B. Thieurmel, *Visnetwork: Network visualization using 'vis.js' library*, R package version 2.1.2, 2022. DOI: 10.32614/CRAN.package.visNetwork. [Online]. Available: https://CRAN.R-project.org/package=visNetwork.

[47] M. Franceschet, "Pagerank: Standing on the shoulders of giants," *Communications of the ACM*, vol. 54, no. 6, pp. 92–101, 2011. DOI: 10.1145/1953122.1953146.