Using Graph Theory to Analyze Passing Maps of Premier League Teams and Identify Key Players

William Andrian Dharma T 13523006[1,2]
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
[1]13523006@std.stei.itb.ac.id, [2]williamadt123@gmail.com

***Abstract*—Football or soccer in American English, is the most popular sport in the world. Although dubbed the beautiful game and individual brilliance can determine the result of a game, tactics and analysis before the game also holds a key factor in how the game will turn out. This paper will discuss how a passing map analysis is conducted using graph theory and how key players in respect to passing distribution can be identified.**

***Keywords*—Football, Graph Theory, Passing Map, Tactics Analysis**

I. INTRODUCTION

Football, the world's sport, is the most popular sport in the world with around 4 billion fans. Almost every country in the world has football is most popular sport although with notable exceptions including India, the United States, and Canada, to name a few. Of all the football leagues in the world, the most watched and valuable is the English Premier League **Error! Reference source not found.**.

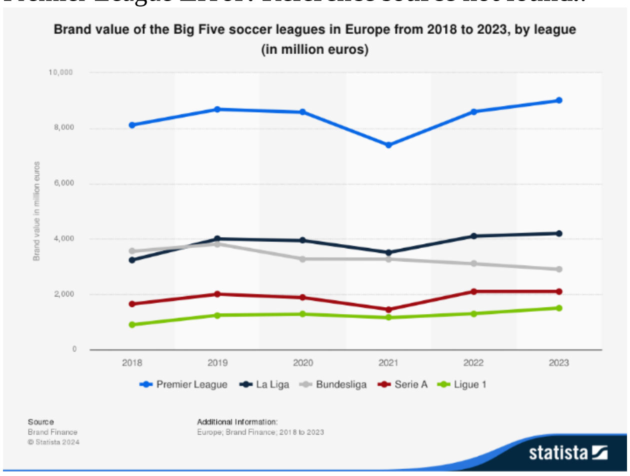


Fig. 1. Brand value of the leading European soccer leagues 2018-2023, by league
Source: https://www.statista.com/statistics/1021424/big-five-leading-football-leagues-brand-value-europe/

Leagues of different countries are generalized to have different styles from one another. For example, the Italian Serie A is thought of a more defensive league, the Spanish La Liga is thought of a more technical league, and the English Premier League is a more physical league. While they are somewhat correct to a degree, the styles of play of the teams within them are still diverse and overlaps happen within the same league or leagues of other nations. For example, Manchester City under the coach Pep Guardiola plays a very possession heavy style of football with having upwards of 60% possession being quite common in their matches. Attacks are done methodically with lots of passes trying to disrupt the structure of the opposition's defense. Moving to their rivals, Manchester United have traditionally been a counterattacking team, utilizing transition phases when the opponent team loses the ball to mount a rapid counterattack usually from the wings, which reached its peak form during the 2007-2008 season with a young Cristiano Ronaldo, Carlos Tevez, and Wayne Rooney being the attacking players capable of quickly delivering fatal counterattacks and led the team to a Premier League win and a UEFA Champions League trophy added to their cabinet. The style of attacking play can be determined by examining the statistics of a team and a big indicator is usually passing. Possession heavy sides would have more passes in comparison to a counterattacking side. Passing maps will also show how teams like to build up their attack.

This paper will restrict the teams analyzed only to mostly being premier league teams from the 2024/2025 season and will mostly use matches from the 2024/2025 season of the Premier League.

The analysis will utilize several methods from graph theory such as weighted graphs, minimum spanning trees, centrality, and tie it in with football related statistics such as xG (Expected Goals), forward passes, key passes, etc.

Applications of such analysis will also be exemplified in this paper by discussing several matches of the 24/25 season and analyzing some of the struggles several teams have during the season.

II. PRELIMINARIES

A. Graph

A graph in discrete mathematics is defined as a structure consisting of sets of vertices or sometimes called nodes which are connected to one another with edges. A simple graph is a graph where the edges are undirected, unweighted, and contains no loops or multiple edges [2]. A mathematical definition of a simple graph would be:

$$G = (V, E)\ where,$$

- $V\ is\ a\ finite\ set, which\ contains\ the\ vertices\ of\ G,$
- $E\ is\ a\ subset\ of\ \mathcal{P}_2(V)\ (i.e., a\ set\ E\ of\ two-element\ subsets\ of\ V),\ called\ the\ edges\ of\ G$

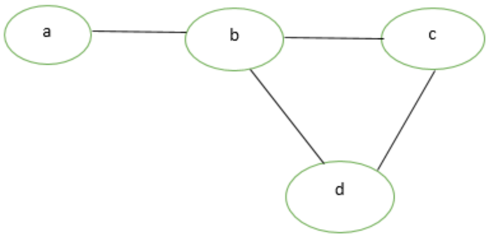


Fig. 2. Example of a simple graph
Source: https://www.geeksforgeeks.org/graph-types-and-applications/

A graph that which has multiple edges or loops are called unsimple graphs. An unsimple graph which has multiple edges is called a multigraph whereas an unsimple graph which contains loops is called a pseudo graph. A mathematical definition of a graph would be as follows [3]:

$$G = (V, E, \phi)\ where,$$

- $V\ is\ a\ finite\ set, which\ contains\ the\ vertices\ of\ G,$
- $E\ is\ a\ finite\ set, which\ contains\ the\ edges\ of\ G,$
- $\phi\ is\ a\ function\ with\ domain\ E\ and\ codomain\ \mathcal{P}_2(V)$

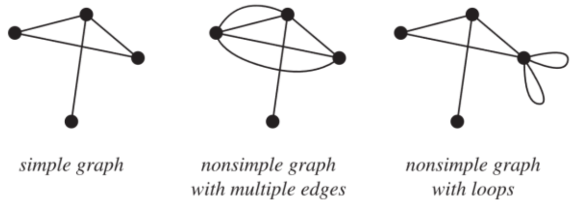


Fig. 3. Examples of simple and unsimple graphs
Source:
https://mathworld.wolfram.com/SimpleGraph.html

B. Directed Graph

A directed graph or digraph is a graph in which edges have directionality. A simple graph where its edges have directions is called a directed simple graph. A directed simple graph can be mathematically defined as [3]:

$$G = (V,\ E)\ where,$$

- $V\ is\ a\ finite\ set, which\ contains\ the\ vertices\ of\ G,$
- $E\ is\ a\ finite\ set, which\ contains\ the\ vertices\ of\ G,$
- $E \subseteq \{(x, y) | (x, y) \in V^2\ and\ x \neq y$

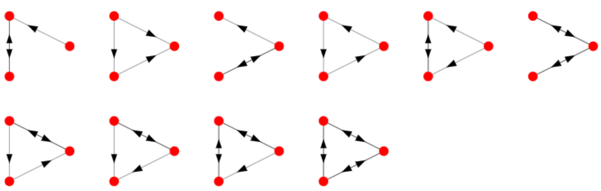


Fig. 4. Examples of directed simple graphs
Source:
https://mathworld.wolfram.com/SimpleDirectedGraph.htm

An unsimple graph where its edges have directions is called a directed multigraph if it has multiple edges and no loops, a directed simple graph permitting loops if it has no multiple edges and has loops, or a directed multigraph permitting loops or a quiver if it has both multiple edges and loops. Such graphs can be mathematically defined as [3]:

$$G = (V,\ E,\ \phi)\ where,$$

- $V\ is\ a\ finite\ set, which\ contains\ the\ vertices\ of\ G,$
- $E\ is\ a\ finite\ set, which\ contains\ the\ vertices\ of\ G,$
- $\phi\ is\ a\ function\ where, \phi : E \rightarrow \{(x, y) | (x, y) \in V^2\}$

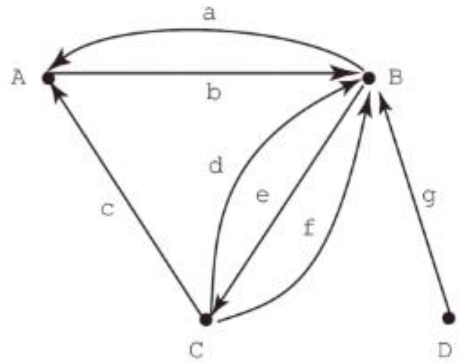
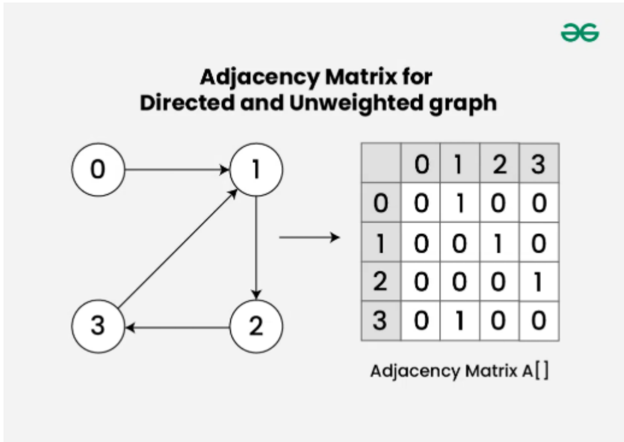


Fig. 5. Example of directed multigraph [3]

C. Graph Terminologies

1. Adjacent

Two nodes are adjacent to each other if they are directly connected by one or more edges. In a directed graph, a node X is adjacent to Y if there is an edge *a* leaving from X to Y. On the contrary, that means that node Y is adjacent from X. The adjacent relations between nodes in a graph can be represented in an adjacency matrix which maps the adjacent to relation between nodes.



	0	1	2	3
0	0	1	0	0
1	0	0	1	0
2	0	0	0	1
3	0	1	0	0

Fig. 6. Adjacency matrix for a directed an unweighted graph
Source:
https://www.geeksforgeeks.org/adjacency-matrix-of-directed-graph/

2. Incident

An edge is incident to the two nodes they connect. Vice versa, a node is incident to all edges it is connected to. For a directed graph, a node X is incident to an edge *a* if *a* is an edge leaving from node X. On the contrary, a node X is incident from an edge *a* if *a* is an edge entering node X. The incident relations between nodes and edges can be represented as an incidence matrix. For a directed graph, an incident to relation is mapped to 1 and an incident from relation is mapped to -1.

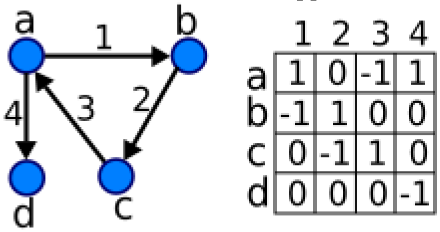


Fig. 7. Incidence matrix of a directed graph
Source: https://commons.wikimedia.org/wiki/File:Incidence_matrix_-_directed_graph.svg

3. Degree
   The degree of a node is the number of edges that are incident to it. For a directed graph they can be separated to in-degree and out-degree based on the direction of the incident edges. For example, looking at Fig. 5., the node B has an in-degree of 4 and an out-degree of 2.

D. Weighted Graph

A weighted graph is an undirected or directed graph in which vertices, edges, or both are assigned numerical values called weights [4].

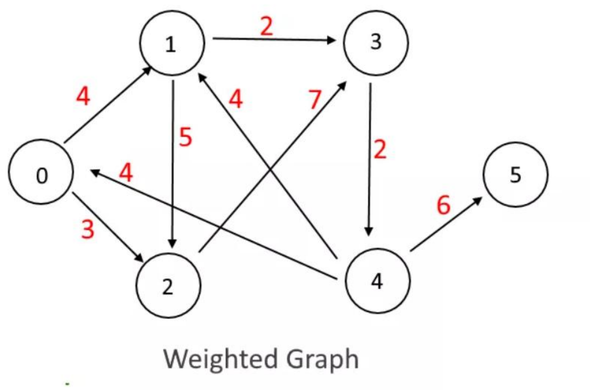


Fig. 8. Example of a directed weighted graph
Source: https://github.com/erenkeskin/directed-weighted-graph?tab=readme-ov-file

A weighted graph will have several changes from an unweighted graph for its adjacency and incidence matrix. A weighted graph's adjacency matrix will contain the weight of the edge and if it is directed, a non-adjacent node will be indicated with infinity. A weighted graph's incidence matrix will be similar to a unweighted graph's incidence matrix but will multiply the value with the weights of the edges.

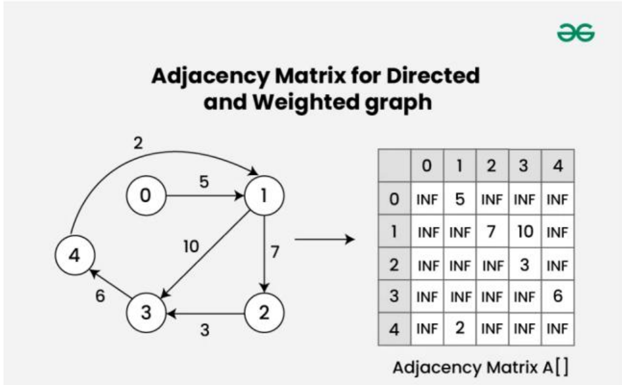


Fig. 9. Example of an adjacency matrix for a directed and weighted graph
Source: https://www.geeksforgeeks.org/adjacency-matrix-of-directed-graph/

This paper will use the representation of vertices as players and edges as passes between them, with the weight being the reciprocal of the number of passes.

E. Closeness Centrality

Closeness Centrality is a measure of centrality of a node in a weighted graph. It is defined as the reciprocal of the sum of the weights of the shortest paths between a node and every other node in the graph [5] which has the formula:

$$C(x) = \frac{1}{\sum_y d(y,x)}$$

where $d(y,x)$ is the shortest path between nodes x and y.

A normalized version of closeness centrality is given by multiplying the previous formula with $N-1$ where N is the number of nodes in a graph.

$$C(x) = \frac{N-1}{\sum_y d(y,x)}$$

Although this is useful for comparing graphs of different sizes, since the size of a football team playing in a match is always the same, ignoring red cards, this paper will use the unnormalized version.

F. Betweenness Centrality

Betweenness Centrality is a measure of centrality of a node in a weighted graph which could show how important is a node within a graph. It is defined as:

$$C_B(y) = \sum_{x \neq y \neq z} \frac{\sigma_{xz}(y)}{\sigma_{xz}}$$

where $\sigma_{xz}(y)$ is the number of shortest paths that pass through node y and σ_{xz} being the total number of shortest paths between node x and z. The sum is taken from all pairs of nodes (x,z), with $x \neq y \neq z$.

G. Eigenvector Centrality

Eigenvector Centrality is a metric which measures how influential is a node to a graph. A high score is obtained by having connections to other high-scoring nodes. It is defined as [7]:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

Where x_i is the eigenvector centrality of node i, λ is the eigenvalue of the corresponding vector, and A_{ij} is the element of the adjacency matrix A. It can be rewritten in vector form as:

$$\lambda x = \mathrm{A} \cdot x$$

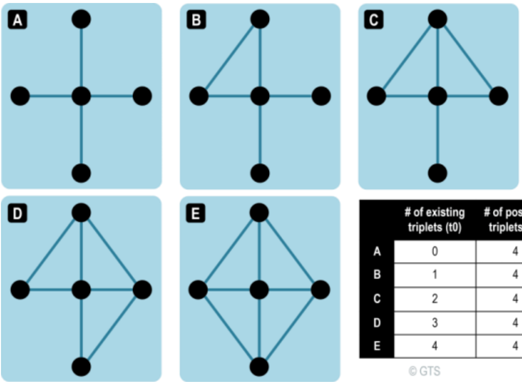
Then, to define a dominant eigenvector, an eigenvector algorithm such as power iteration is used to get the score that will be used [8]. Eigenvector centrality is commonly used for network analysis across domains and a variation called PageRank was famously used in the past by Google to rank search engine results [10].

H. Transitivity

The transitivity of a graph is 3 multiplied by the number of triangles in a graph divided by the number of triples [9]. A triangle consists of three nodes that are all connected whereas a triple consists of three nodes that not all pairs have to be connected. The multiplier 3 comes from the fact that each triangle also contributes to three other connected triples. The formula is defined as:

$$T = \frac{3 \times number\ of\ triangles}{number\ of\ connected\ triples}$$

Transitivity can explain how tightly connected a graph is and could show clusters.



	# of existing triplets (t0)	# of possible triplets (t1)	Transitivity (t0/t1)
A	0	4	0
B	1	4	0.25
C	2	4	0.5
D	3	4	0.75
E	4	4	1.0

© GTS

Fig. 10. Transitivity of several graphs
Source: https://transportgeography.org/contents/methods/graph-theory-measures-indices/transitivity-graph/

I. Minimum Spanning Arborescence

A Minimum Spanning Arborescence (MSA) is the directed analog of the minimum spanning tree in an undirected graph. It can be calculated with several algorithms with one being the Chu-Liu/Edmond's algorithm [11][12].

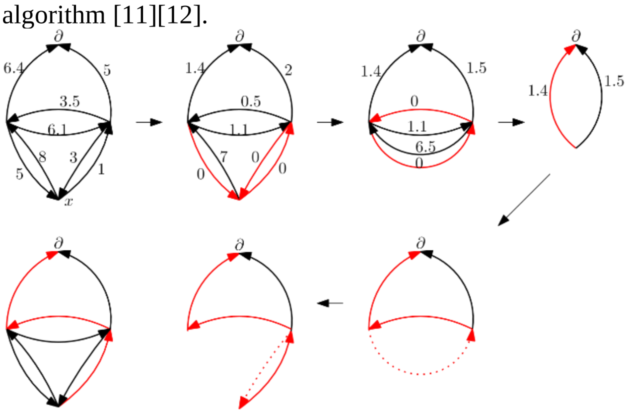


Fig. 11. Example of the Chu-Liu/Edmond's algorithm
Source: https://arxiv.org/html/2401.13238v1

III. Methodology

A. Experiment Dataset

This paper uses data available online on theanalyst.com which is provided by Opta Analyst. The website provides lots of match statistics and passing maps which will be the basis for our analysis.

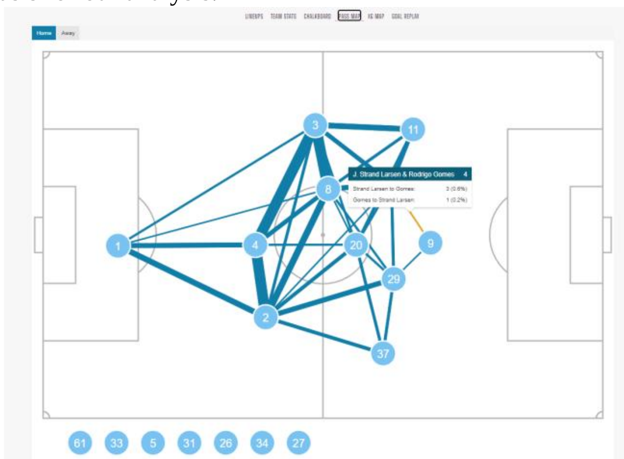


Fig. 12. Pass map of Wolverhampton Wanderers FC vs Nottingham Forest FC on 6 January 2025
Source: https://theanalyst.com/2023/07/opta-football-match-centre/?competitionId=2kwbbcootiqqgmrzs6o5inle5&seasonId=9n12waklv005j8r32sfjj2eqc&matchId=c65kd5w22z083sj10frkt7f9w

Any additional statistics will use fotmob.com especially individual statistics of a match.

B. Analysis Tools

This paper will use python as the programming language used for analysis. We will use libraries such as networkx, numpy, matplotlib, etc. The weighted directed graph will be manually created based on the pass map. The goalkeeper will be ignored.

C. Analysis Metrics

An analysis of a passing map will result in several metrics such as:

1. Closeness Centrality for every player
2. Betweenness Centrality for every player
3. Eigenvector Centrality for every player
4. Transitivity for the team
5. Minimum Spanning Arborescence of the passing map

D. Analysis Interpretation and Results

The metrics will have an interpretation as follows:

1. A high Closeness Centrality would mean a player is highly connected to other players.
2. A high Betweenness Centrality would mean a player is highly important in distributing the ball to other players.
3. A high Eigenvector Centrality would mean a player has a large influence on the whole passing network.

4. A high Transitivity would mean that a team's passing is tightly connected.
5. The Minimum Spanning Arborescence would give us a simplified passing map which could discover the passing and build up tendencies of the team.

All of those will be used to analyze and discuss the tactics of a team.

IV. Applications

A. Match Analysis

To demonstrate the methods in this paper, let us do a match analysis of the match between Liverpool and Manchester United on 5 January 2025 which ended 2 – 2.

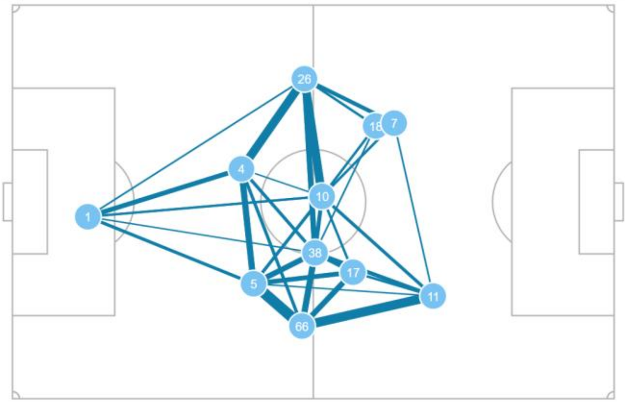


Fig. 13. Pass map of Liverpool
Source: https://theanalyst.com/2023/07/opta-football-match-centre/?competitionId=2kwbbcootiqqgmrzs6o5inle5&seasonId=9n12waklv005j8r32sfjj2eqc&matchId=c4ke9hxyesq5m0zfx31p3s6c4

Player	Closeness	Betweenness	*Eigenvector*
26	3.8595	0.3611	*0.24*
4	3.2102	0.1389	*0.2982*
38	3.7324	0.1944	*0.3882*
10	3.592	0.1111	*0.5003*
18	2.2705	0.0	*0.2497*
7	2.7012	0.0	*0.3304*
66	3.7097	0.1389	*0.1493*
5	3.432	0.1389	*0.2471*
17	3.1337	0.0	*0.3118*
11	3.3409	0.0	*0.3168*

Transitivity (Reciprocal of Passes): 0.6184

Table 1. Metrics for Liverpool players

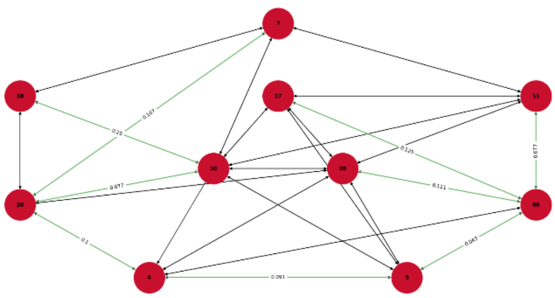


Fig. 14. Minimum Spanning Arborescence of Liverpool (marked in green)

As we can see from Table 1. And Fig. 14., several key observations can be made. In terms of Closeness, A. Robertson with the number 26 is top of the list followed by R. Gravenberch with the number 28 and T. Alexander-Arnold with the number 66. This indicates that Liverpool likes to use their fullbacks, which are A. Robertson and T. Alexander-Arnold, in distributing the ball and building up the attack. Then, the third highest closeness goes to R. Gravenberch with A. Mac Allister with the number 10 following close behind. As they are both the central midfielders in Liverpool's 4-2-3-1 formation, it makes sense that they have high Closeness since it is a central midfielder's role to control play and distribute the ball.

Next on Betweenness, A. Robertson again leads the list this time significantly. This shows that Liverpool likes to pass around the ball with different players on the left wing with A. Robertson as the main controller. On the other wing, T. Alexander-Arnold has a lower Betweenness that could probably be due to his more direct attacking style of play, which is shown by his preference in passing to M. Salah with the number 11 rather than sideways to the midfielders. Finally, the leader of the Eigenvector metric is A. Mac Allister with the number 10 with a score of 0.5003, he is indicated as the key player in Liverpool's passing and distribution. As we can see, he is the only midfielder that has successful passes to all the 4 attacking players which are the 3 attacking midfielders and the striker. He had an assist in that game which further strengthened the point.

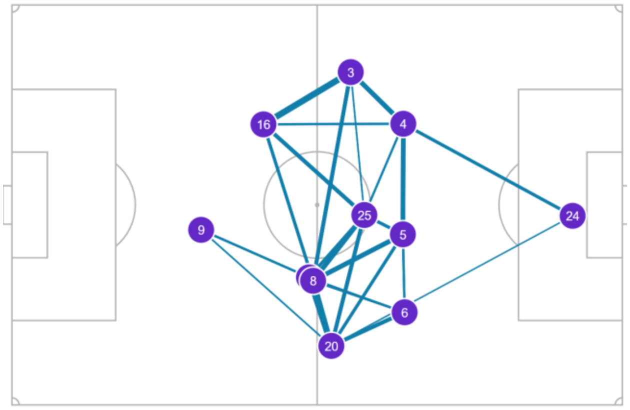


Fig. 15. Pass map of Manchester United
Source: https://theanalyst.com/2023/07/opta-football-match-centre/?competitionId=2kwbbcootiqqgmrzs6o5inle5&seasonId=9n12waklv005j8r32sfjj2eqc&matchId=c4ke9hxyesq5m0zfx31p3s6c4

Player	Closeness	Betweenness	*Eigenvector*
6	2.1579	0.0278	*0.3852*
20	2.8546	0.1806	*0.2557*
8	3.642	0.375	*0.4292*
5	3.4467	0.0833	*0.4002*
37	3.4468	0.2153	*0.2483*
25	2.5472	0.1389	*0.4259*
4	2.7739	0.0278	*0.1738*
16	2.6741	0.0833	*0.2588*
3	3.4357	0.1944	*0.2026*
9	1.7609	0.0	*0.246*

Transitivity: 0.5935

Table 2. Metrics for Manchester United players

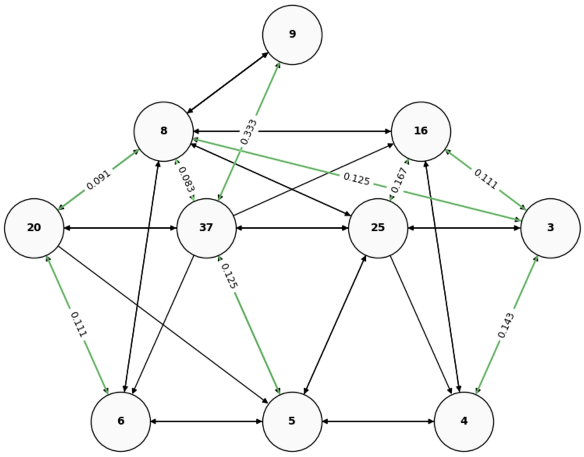


Fig. 16. Minimum Spanning Arborescence of Manchester United

As we can see from Table 2. And Fig. 16., several key points can be made. In terms of Closeness, B. Fernandes with the number 8 is top of the list followed by K. Mainoo with the number 37, H. Maguire with the number 5, and N. Mazaroui with the number 3 in very similar numbers. This indicates that B. Fernandes is at the center of play for Manchester United and often works together with the other 3 which explains the very close similarity of the scores and by Fig. 16. in which B. Fernandes has a high degree in. Now for Betweenness, B. Fernandes again significantly leads, which further cements his position as a key player in distribution and build up for Manchester United. Finally for Eigenvector, B. Fernandes is once again top of the list followed closely by M. Ugarte with the number 25. By this analysis, it is safe to say that B. Fernandes is a crucial player for Manchester United, this is further cemented by the fact that he had an assist that game and won man of the match. This is also confirmed by the fact that he has also been called the most crucial player in that Manchester United team for a few seasons.

Both teams have similar transitivity which shows that both teams are quite tight in their passing although Liverpool seems to prefer a higher number of passes than Manchester United.

V. Conclusion

With the example demonstrated in this paper, it has been shown that this method of football analysis supplemented with graph theory has its merits. The example correctly points out B. Fernandes as the key player in Manchester United and deeper analysis would extract more information. Further extensions and metrics added to this method would probably provide more information about the tactics a football team uses and add robustness to this method.

VI. Acknowledgment

I would like to give thanks to my family who supported me in my studies and my lecturer Ir. Rila Mandala, M.Eng., Ph.D., which has explained the material in detail and clarity.

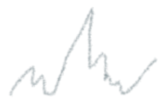
References

[1] Brand Finance Plc. 2023. *Brand Finance Football 50 2023*
[2] Gibbons, A. *Algorithmic Graph Theory*. Cambridge, England: Cambridge University Press, 1985. p.2.
[3] *Bender, Edward A.; Williamson, S. Gill (2010). Lists, Decisions and Graphs. With an Introduction to Probability*.
[4] Weisstein, Eric W. "Weighted Graph." From *MathWorld*--A Wolfram Web Resource. https://mathworld.wolfram.com/WeightedGraph.html
[5] Sabidussi, G (1966). "The centrality index of a graph". *Psychometrika*. **31** (4): 581–603. doi:10.1007/bf02289527. hdl:10338.dmlcz/101401. PMID 5232444. S2CID 119981743.
[6] Brandes, Ulrik (2001). "A faster algorithm for betweenness centrality" (PDF). *Journal of Mathematical Sociology*. **25** (2): 163–177.
[7] M. E. J. Newman. "The mathematics of networks"
[8] Richard von Mises and H. Pollaczek-Geiringer, *Praktische Verfahren der Gleichungsauflösung*, ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik 9, 152-164 (1929).
[9] Nykamp DQ, "Definition of the transitivity of a graph." From *Math Insight*. http://mathinsight.org/definition/transitivity_graph
[10] Sullivan, Danny (2007-04-26). "What Is Google PageRank? A Guide For Searchers & Webmasters". *Search Engine Land*. Archived from the original on 2016-07-03.
[11] *Edmonds, J. (1967), "Optimum Branchings", Journal of Research of the National Bureau of Standards Section B, **71B** (4): 233–240, doi:10.6028/jres.071b.032*
[12] Chu, Yeong-Jin; Liu, Tseng-Hong (1965), "On the Shortest Arborescence of a Directed Graph" (PDF), Scientia Sinica, XIV (10): 1396–1400

Pernyataan

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 8 Januari 2025


ttd

William Andrian Dharma T - 13523006