Explaining Tactical Formation and Passing Network in Football with Graph Theory

Rayhan Alghifari Fauzta (13519039)[1]
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
[1]*13519039@std.stei.itb.ac.id*

***Abstract*—As one of the most popular sport in the world, football (or soccer) has evolved periodically in terms of playstyle and formation. A style of play that relies heavily on player's physical attributes have largely been abandoned in favor of technical and tactical prowess. Modern football puts an emphasize on developing the best tactical approach primarily through formation. As players execute the tactics, they will pass the ball to each other and produce a statistical feature known as passing network. Both formations and passing networks can be viewed as graph models in discrete mathematics. This paper will discuss how teams set up their formations and their approach to the game by applying graph theory and its related principles in the process.**

***Keywords*—football, graph, passing network, tactical formation**

I. Introduction

Football is without a doubt the most popular sports in the world, being played professionally by over 128.983 men and women [1] as well as 250 million more players playing it recreationally. Countries with a deeply rooted footballing history have regarded the sport as more than a side activity. Instead, it is part of their identity and pride as a nation. Major international football tournaments such as World Cup and Euros and domestic leagues like Premier League and La Liga have consistently drawn massive number of viewers from all over the world. In terms of money, football is also the most lucrative sports. Barcelona, one of the most successful clubs in the world, generated an annual revenue of €840.8 million in 2020 according to Deloitte [2].

Part of the charm in football is its dynamic and in-depth tactical variations centered around the use of formation. Since first played in late 19th century, playstyle in football has evolved from simple pass-and-shoot to rapid counterattack, intense pressing, *tiki-taka* passing, park-the-bus defense, and many other variations of play. One of the main factors in shaping the playstyle is formation. As stated before, formations have also evolved in shape. The first popular formation in football was 2-8, or 2 defenders at the back with 8 forwards in front. As time goes by, more and more teams realized the need to connect the backline with the forward line with designated midfield players. This resulted in new formations and shape of play, such as 3-5-2, 4-2-3-1, 4-3-3, and the famous 4-4-2 [3].

Graph theory have long been used to study various problems, ranging from transportation and logistical aspects like the Königsberg bridge problem to biology and ecological aspects like the food chain. In football, graph theory can also be applied to analyze the tactical setups used by teams in a game and the statistics attained from the setups. Passing networks and expected goal chains are two examples of statistical features that can be interpreted using graph theory as a reinforcement to another knowledge such as network theory.

This paper aims to explain the principles and reasoning behind formulating a formation in football by applying graph theory in the process. Each player in a football team can be represented by a vertex. The 10 outfield players and 1 goalkeeper can then be positioned accordingly in a way that resembles a graph. Further analysis can then be made by looking at the connection between players and the resulting passing maps from the match using various types of graph such as directed graph and weighted graph.

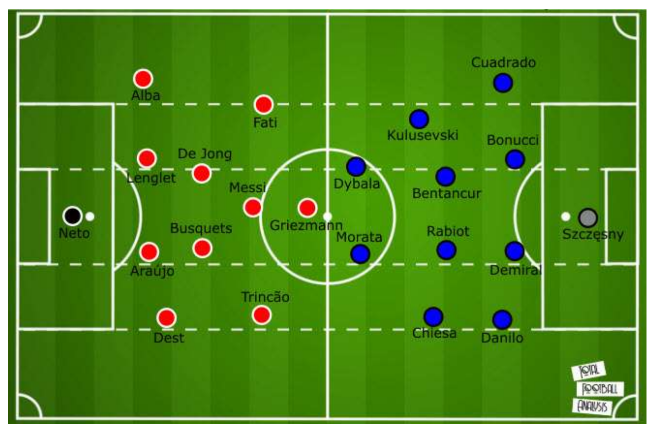

Figure 1. Lineup of FC Barcelona vs Juventus in the UEFA Champions League. Source: https://totalfootballanalysis.com/match-analysis/tactical-preview/uefa-champions-league-2020-21-juventus-vs-barcelona-tactical-preview-analysis-tactics

II. Graph Theory

A. Definition

A graph can be defined as discrete structures consisting of vertices and edges that connect these vertices. A graph of the following:

$$G = (V, E)$$

consists of a nonempty set of vertices (or nodes) V and a set of edges E. Each edge has either one or two vertices associated with it called endpoints [4]. In this paper, vertices will be used to represent players and the edges will represent either action taken by players in a match or interaction between two or more players, both active and passive.

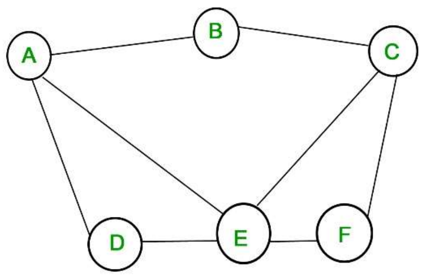

Figure 2. Example of a simple graph. Source: https://www.geeksforgeeks.org/mathematics-graph-theory-basics/

B. Types and Representation

Graph can be classified with a variety of indicators. First is by looking at the orientation of its edges. A graph where the edges do not point to a specific direction or orientation is called an *unoriented graph* or *undirected graph*. Likewise, if each edges of a graph point to a specific direction then the graph is called a *directed graph* or *oriented graph*. In an unoriented graph, the relationship between two vertices is straightforward such that if vertex a is linked to vertex b through an edge, then a is connected to b and b is connected to a. On the other hand, the edges in an oriented graph have a specific direction that goes from one vertex to another. Thus, one of the endpoints acts as a *source* and the other is the *destination*. Oriented graph is often used in football analytics to visualize passing networks in a team, whereas unoriented graph may be used to represent a playing shape like defensive line or offensive line.

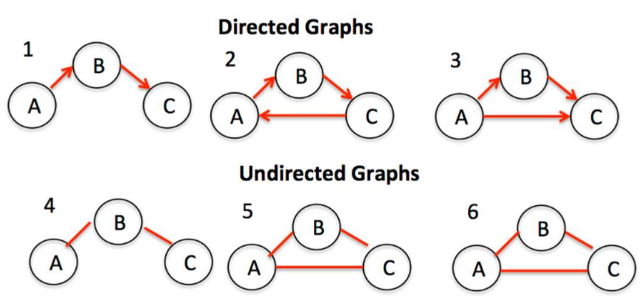

Figure 3. Directed and undirected graph. Source: https://sites.google.com/a/cs.christuniversity.in/discrete-mathematics-lectures/graphs/directed-and-undirected-graph

Another type of graph classification is by looking at its connectivity. A *connected graph* is a graph where every vertex is joined with each other by a path. In other words, every pair of vertices have a path that connect them. If the two vertices are connected by a single edge (the length of the path is 1), then the vertices are adjacent to each other. A graph can also be disconnected. If there exist a pair of vertices where they are not connected with a path, then such graph is referred as a *disconnected graph*. The connection in a connected graph can further be analyzed if the graph is also directed. A directed graph is said to be *weakly connected* if all of its directed edges can produce a connected graph when replaced with undirected edges. A graph is *unilaterally connected* if it contains a directed edge from vertex u to vertex v or a directed path from v to u for every pair of $\{u, v\}$. Lastly, a graph is said to be *strongly connected* if it contains a directed edge from u to v and a directed edge from v to u for every pair of vertices $\{u, v\}$ [5].

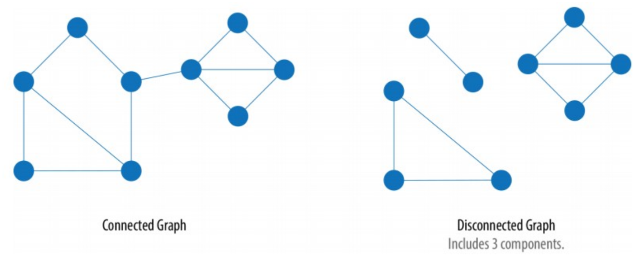

Figure 4. Connected and disconnected graph. Source: https://m.steveclarkapps.com/graphs/

Graph can also be classified by its weight. A graph where each edge has a specific numerical value is called a weighted graph. This weight may correspond to various real-life aspects. In transportation, for example, it can be used to represent the amount of effort needed to travel from one place to another (cost and distance) or the maximum amount of flow that can be moved from one place to another (capacity). One of the most common application of weighted graph is to find the shortest path or the path with minimum weight from an origin vertex to a destination vertex. This task may be carried out using several pathfinding algorithms, such as Dijkstra's algorithm, Jarnik-Prim Minimum Spanning Tree, and Bellman-Ford algorithm.

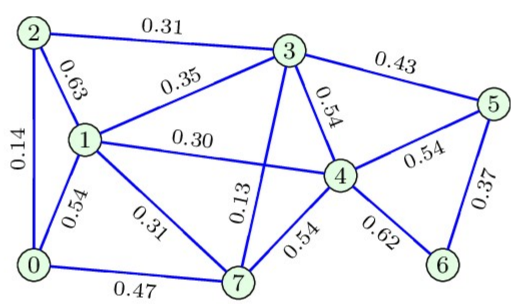

Figure 5. A weighted graph. Source: https://www.researchgate.net/figure/An-example-of-a-weighted-graph_fig1_329350163

Representation of a graph can be done using either adjacency list or adjacency matrix. An *adjacency list* is an array containing the address of all the linked list. In the linked list itself, the first node represents a vertex and the lists connected to the node represents all the vertices connected to the node. An *adjacency matrix* is a two-dimensional array of size v x v, where v is the total number of vertices in a graph. Suppose there is a two-dimensional array:

$$\text{arr}[i][j] = w$$

that means there is an edge from vertex i to vertex j with a weight of w. Both adjacency list and adjacency matrix can be used to represent weighted and unweighted graph.

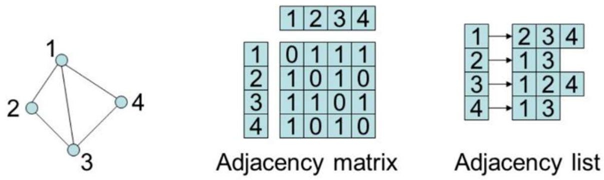

Figure 6. Adjacency list and adjacency matrix. Source: https://bournetocode.com/projects/AQA_A_Theory/pages/graph.html

III. Tactical Formation and Passing Network

A. Tactical Formation

Formation in football explains how a team set up their player's position on the pitch to form a specific shape and in turn influences the team's style of play. A player's position in a formation generally defines whether a player has a predominantly defensive or attacking role, and whether they tend to play towards the wider side of the pitch or centrally.

Formations are typically written by three or four numbers, which represents how many players are in each part of the formation from the most defensive to the most forward. For example, the popular 4–4–2 formation has four defenders, four midfielders, and two forwards. Different formations can be used depending on whether a team wishes to play more attacking or defensive football, and a team may switch formations between or during matches for tactical reasons.

There are two common types of formation based on the players available in the team. The first type is a narrow formation. In this setup, the team will make use of central midfielders and focus their attack through the middle of the pitch. Examples of narrow formations are 4–1–2–1–2 or 4–3–2–1 which allow teams to field up to four or five central midfielders. Narrow formation depends heavily on the full-back (the defenders that occupy the flank) to provide width and to get forward up the pitch as frequently as possible to supplement the attack in wide areas.

Aside from narrow formations, teams may also opt for wide formations if their strength lie on players that operate using the width of the pitch. Wide formations such as 4–2–3–1, 3–5–2 and 4–3–3 commit forwards and wingers high up the pitch. Wide formations allow the attacking team to stretch play and force the defending team to cover more ground. Wide formations are commonly associated with having an attacking style of play. However, determining a particular team's style of play solely on formations can be deceptive. For instance, a team that plays an attacking 4–3–3 formation can quickly revert to a 4–5–1 if two of the three forwards frequently track back in midfield.

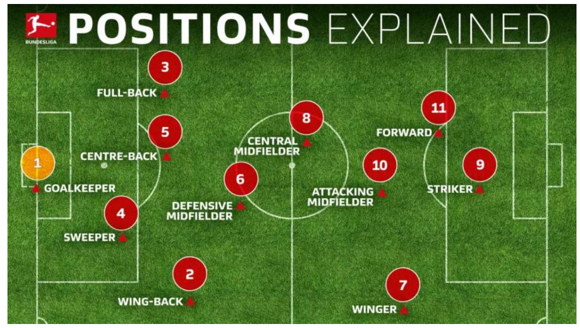

Figure 7. In a match, formations may not be rigid and can change dynamically depending on the flow of the game. Source: https://www.bundesliga.com/en/bundesliga/news/soccer-positions-explained-names-numbers-what-they-do-2579-786

B. Passing Network

Passing network in a football team can be defined as a network with the team's players as nodes and connecting arrows between two players weighted by the successful number of passes completed between them [5]. The weight in a passing network is usually visualized with thickness and hue relative to the number of successful passes between two players.

There are three main types of passing networks based what the nodes represent. First is player passing networks where the nodes are composed of players in the team. Next is pitch passing networks where the nodes are made of specific regions of the pitch that are occupied by players performing the passes. The third type is a combination of pitch-player passing networks where the nodes are a combination of players and their position on the pitch when they made the pass [7]. When visualizing a passing network, pitch-player passing network is preferred because it can also explain the average position of players in a match on top of showing the passing attempts of the team.

Basic visual analysis can be made by computing the weighted adjacency matrix A, with A_{ij} being the number of passes from player i to player j. The weight or number of passes will be used as a measure of the strength of an arrow in the network and to define the distance between players. Distance d_{ij} is defined as the distance given by the length of the shortest path connecting the nodes i and j, where the length of a path is obtained by adding the lengths l_{ij} of the arrows according to the following conditions:

$$l_{ij} = \begin{cases} 0, & i = j \\ \dfrac{1}{A_{ij}}, & i \neq j \end{cases}$$

The length of an arrow between two players is considered infinite if they do not pass the ball to each other. It is worth noting that definition of distance used in this context is not necessarily correlated with the physical distance between players on the pitch [6].

Although arguments can be made that passing network is an oversimplification of a game of football, it does provide an insight into a team's tactical setup and style of play. Passing network can be used to determine areas of the pitch that are favored or neglected by a team. It can also tell a story about the types of passing game preferred, whether they play a possession-based football with short range passes or take the route one with long range passes straight to the forwards. The network can also be used by a team to detect under-performing players, detect potential problems between teammates who are not passing the ball as often as their position dictates, identify player positioning problems where two or more players occupy the same area on the pitch, as well as to detect weaknesses in rivals.

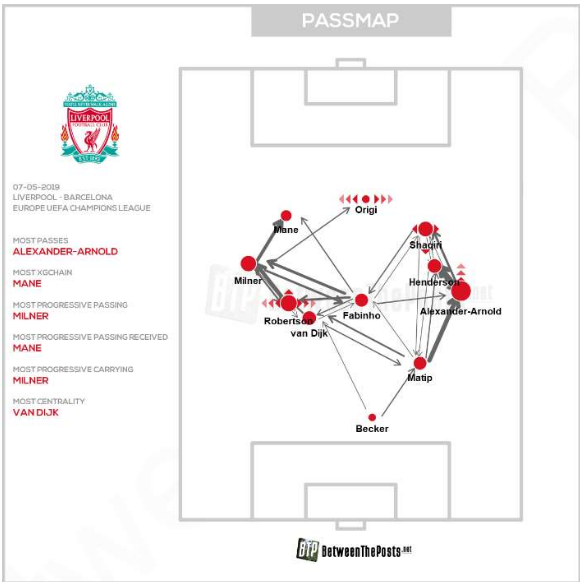

Figure 8. Liverpool pass map vs Barcelona in May 2019. Source: https://betweentheposts.net/liverpool-barcelona-4-0-klopp-tactics-clever-corner-night-to-remember-anfield/

IV. Analysis of Tactical Formation and Passing Network

To analyze the formational aspect of the game, we will take a look at how Manchester United, hereafter referred as United, set up their tactical approach for their match against RB Leipzig in the UEFA Champions League 2020/2021 Group Stage. United would go on and lose the game with a scoreline of 2-3 to the German side. This is how United lined up for the game:

Figure 9. Manchester United starting eleven versus RB Leipzig. Source: UEFA Champions League

Based on the lineup, initial assumption can be made that United intended to play a 4-2-3-1 formation with four defenders at the back, two defensive midfielders, one attacking midfielder, two wingers, and one striker. However, player position and roles must be taken into account here. Alex Telles, the player on the left wing, is actually a left back rather than a left winger. The original left back, Shaw, was assigned to the centre back role accompanying Maguire and Lindelöf. To keep the shape balanced, Greenwood moved up top to partner Rashford as a striker. These changes resulted in a 5-3-2 (or 3-5-2 depending on how the full back operates) formation which can be represented using graph as follows:

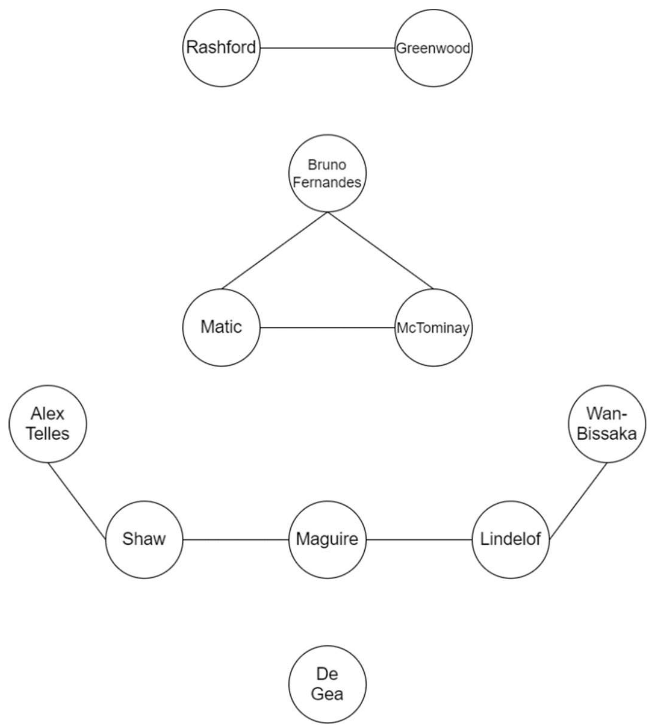

Figure 10. Manchester United's 5-3-2 formation represented with graph.

From the disconnected graph above, we can see that there are four subgraphs representing three playing units. Starting from the defensive unit, De Gea acts as the goalkeeper with protection from a trio of Shaw, Maguire, and Lindelöf tasked with keeping the central regions safe, followed by Alex Telles on the left and Wan-Bissaka on the right to guard the flanks. Moving to the midfield, Matic and McTominay will provide support to the defense as well as distributing the ball safely from the back to the front. Attacking midfielder Bruno Fernandes will be the creative powerhouse of the team who will find ways to create goalscoring opportunities with his movement and passing. United's goalscoring threats in front are Rashford and Greenwood, both playing as advanced forwards tasked with breaking down the opposition defense and eventually finding the back of the net.

Five defenders at the back implies that United wanted to play a defensive and disciplined football to frustrate the opposition. They would then break out with fast counterattack to find the goals needed to win the game. However, this plan would fumble and exploited by Leipzig. To understand how, take a look at the following diagram of Manchester United's average player positions:

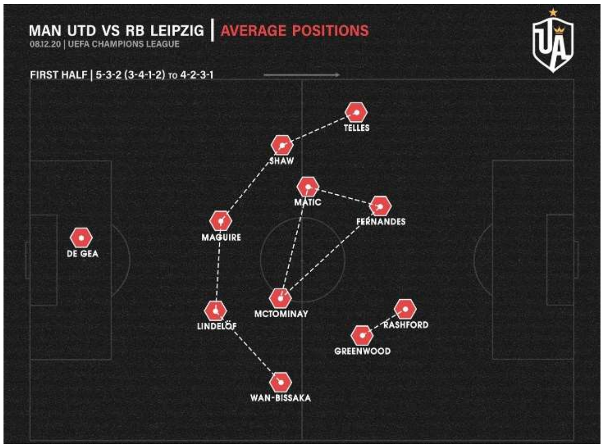

Figure 11. United's average positions in the first half. Source: United Arena

The average positions diagram above shows three anomalies surrounding United's playstyle. First, the left flank is too crowded and unbalanced with Shaw playing further forward but Alex Telles does not operate higher on the pitch. Moving to the right flank, the situation became quite different from the left in that the player's positioning are too loose. Wan-Bissaka, Lindelöf, and McTominay are too far from each other. This space in defense would then be exploited by Leipzig for their first and second goal. In the forward department, playing a defensive football means that United's attack during the first half was nonexistent. Rashford and Greenwood have to cover the right side of the pitch due to constant bombardment from Leipzig left back. Consequently, Bruno Fernandes was left alone in the center. Attacking movement from all three sides of the pitch would be easily nullified by Leipzig's defense. United would end up 2-0 down in the first half, with both goals coming from the loosely defended right side of the pitch.

To repair the problems, United made several changes in the second half by seizing the initiatives to attack. They made several substitutes and switched to a 4-2-3-1 diamond attacking formation. The midfield that was overrun in the first half was strengthened by bringing in one extra midfielder in van de Beek. Matic was also substituted with Pogba to provide more pace and fresh legs to go forward. Greenwood played deeper in midfielder which resulted in Rashford acting as the lone striker. The attacking burden was lifted from the flanks to ensure that Williams, who replaces Shaw in the left, and Wan-Bissaka are fully focused on the defensive aspects, although they occasionally make forward runs especially nearing the end of the match. These changes would not immediately take effect, however, and United were unfortunate to let in a third goal due to defensive lapses before being able to score themselves. Average positions of United players during the second half is as follows:

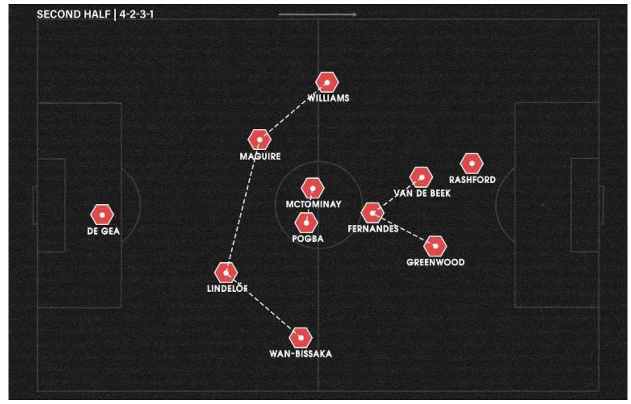

Figure 12. United's average positions in the second half. Source: United Arena

Analyzing the passing networks gathered from the match can then be done after the formational aspects have been understood. Passing networks offer more insights into what went right and what went wrong in terms of positioning as well as determining if the attacking pattern is effective or not. Due to recency of the match and technical limitations at the time of writing, only the first half pass map will be shown. Here is United's passing network during the first half:

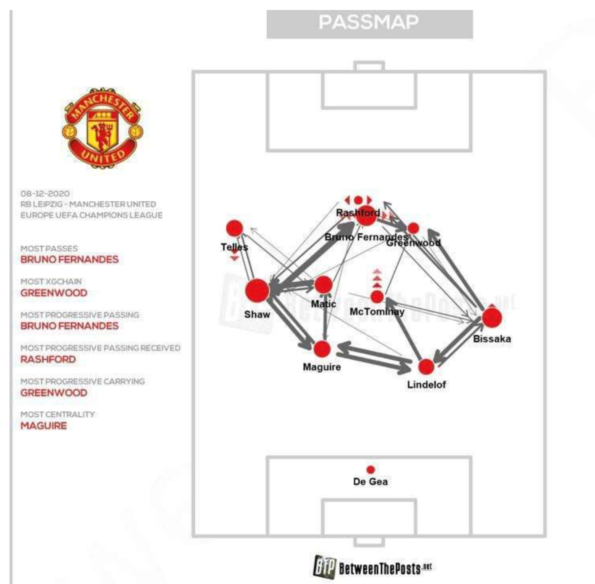

Figure 13. Pass map of Manchester United against RB Leipzig. Source: https://betweentheposts.net/rasenballsport-leipzig-manchester-united-early-leipzig-blitz-proves-too-much-for-uniteds-comeback-attempt-3-2/

At its most basic form, passing network (or pass map) can be interpreted as a directed graph. the passing networks gathered from the match can then be done after the formational aspects have been understood. Passing networks offer more insights into what went right and what went wrong in terms of positioning as well as determining if the attacking pattern is effective or not. Looking at the pass map above, it is clear that United focused their attack through the center. Player with most centrality is Harry Maguire, meaning that he is the most influential player on the pitch for United. However, this does not necessarily indicate that he is the most important player for the team. Centrality refers to the one most responsible for directing the ball and controlling the tempo in general. Maguire, playing as a centre back, receives the ball regularly from goalkeeper De Gea as well as his two partners in defense, Shaw and Lindelöf. He would then distribute the ball to Shaw who acts as the connecting link to the attackers, particularly Bruno Fernandes. Fernandes is arguably the most important player here. He holds the title for most passes as well as most progressive passes. When Fernandes has the ball, he would fulfill his duty by performing defense-splitting passes to either Rashford or Greenwood. In this case, Rashford is shown to be the player who received the most progressive passes. Greenwood primarily acts as the dribbler and distractor, performing forward runs when he has the ball in order to beat his guard or disrupt Leipzig's defense in general.

Moving on from the general flow of play, the pass map tells a clear story about where United's weaknesses lie. First major weak link in the team is surprisingly the midfield. There is no connection between Matic and McTominay. They operate almost independently without coordination, let alone pass the ball to each other. It is also clear that McTominay operates too far in the middle from Wan-Bissaka and Lindelöf. This creates a catastrophic situation where Wan-Bissaka is forced to protect the right side alone while still providing attacking threat, a duty even the best wing back in the world would struggle to accomplish. On the left wing, the situation is reversed but somewhat similar from the right. Telles as a left winger is simply ineffective. He rarely made forward runs and track back too frequently to join the defense. Attacking part of the team is also not blameless. In terms of positioning, there is no clear distinction as to where Rashford, Fernandes, and Greenwood should operate. Rashford and Greenwood are too deep in midfield and too far in distance between them. This puts unnecessary difficulties on Fernandes to determine where he should pass. All of those problems were eventually capitalized by Leipzig. They exploited the right flank by bombing down the area continuously, resulting in two goals in the first half. Changes made by United in the second half were unfortunately not enough, as they succumbed to a 2-3 loss and eventually dropping out of the UEFA Champions League for the 2020/2021 season.

V. Conclusion

Mathematical principles can be applied to virtually every aspects of human life, including sports and specifically football. Knowledge of graph theory can be utilized to understand the reasoning behind how teams prepare their formation for a game of football. Furthermore, execution of said formations can be reviewed by looking at the passing networks generated from matches. This paper is only a dip at the surface by sampling one team and one match using one core principle. Deeper exploration can be done by applying more complex and rigorous principles like network theory, clustering, and/or pagerank centrality as well as acquiring bigger datasets to be analyzed.

VI. Acknowledgment

The author would like to express utmost gratitude to God, families, and friends who provided unconditional supports through thick and thin especially during a period of difficult times when this paper was written. Author also thanked Mr. Rinaldi Munir as head lecturer for IF2120 Discrete Mathematics course and Mrs. Fariska Zakhralativa Ruskanda as class lecturer for their spirit and determination to keep delivering knowledge despite major teaching difficulties. Lastly, the author would like to thank every football analyst in the world for their collective effort to enhance the beautiful game and taking it to higher grounds through the power of mathematics and science.

References

[1] FIFA. 2019. *Professional Football Report 2019*. Zürich: FIFA.

[2] Deloitte Sports Business Group. 2020. *Eye on the prize - Football Money League*. Manchester: Deloitte.

[3] Wilson, Jonathan. 2008. *Inverting the Pyramid: The History of Football Tactics*. London: Orion Books.

[4] Rosen, K. H. 2007. *Discrete Mathematics and Its.Application (7th ed.).* New York City: McGraw-Hill Higher Education.

[5] Bang-Jensen, J., & Gregory Z. Gutin. 2008. *Digraphs: Theory, Algorithms and Applications (2nd. ed.)*. Springer Publishing Company, Incorporated.

[6] Peña, J. L., & Touchette, H. 2012. "A network theory analysis of football strategies". ArXiv. http://arxiv.org/abs/1206.6904, retrieved December 9th 2020.

[7] Buldu, J. M., et. al. 2018. "Using network science to analyze football passing networks: Dynamics, space, time and the multilayer nature of the game". ArXiv. https://arxiv.org/ftp/arxiv/papers/1807/1807.00534.pdf, retrieved December 9th 2020.

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.

Jakarta, 10 Desember 2020

Rayhan Alghifari Fauzta (13519039)