IB Extended Essay

Research Question: How can the knowledge of probability and network theory help quantitatively and qualitatively analyse a football team?

Subject: Mathematics

Session: May 2018

Words: 3991 words

Table of contents

INTRODUCTION 2
INTRODUCTION TO NETWORK THEORY 4
DATA USED 11
THE NETWORK 14
CONSTRUCTING THE NETWORK 14
THE FINAL NETWORKS 17
RESULTS 19
CALCULATIONS 20
DATA INTERPRETATION 22
LIMITATIONS OF USING NETWORK THEORY 23
INTRODUCTION TO PROBABILISTIC METHOD 24
POISSON DISTRIBUTION 26
CALCULATIONS 27
CALCULATING Λ 29
LIMITATIONS OF THE PROBABILISTIC MODEL 31
CONCLUSION 32

INTRODUCTION

Football is, arguably, the most famous sport in the world. It is estimated that the 2014 FIFA World Cup Finals were watched by 1 billion people around the world. With such a huge fan-following, it is essential that football teams engage in research that helps them stay ahead of the curve. Traditionally, the sport has lagged behind others, such as baseball and cricket, in the use of statistical data to make decisions. The dynamic nature of football coupled with the constant ball-flow and movement hinders the capability to analyse player and team performance with simple statistical metrics such as goals and assists. In recent years, however, football data companies, such as *Opta* and *Prozone*, have emerged making a significant amount of data available for free and commercial use. This opens avenues for detailed analysis of data in football.

This extended essay looks to seeks to show a proof of concept in how the mathematical theory of networks and probability can be used to analyze data from football matches and provide a qualitative and quantitative measure of the performance of a team. This idea will be applied to two matches of the UEFA Champions League.

A network is a set of objects (or nodes or vertices) that are connected. Networks can be used in the analysis of a multitude of issues, ranging from economical dilemmas to social difficulties to technological innovations. Several mathematical theories have been built around this concept, most notably by Euler, who came up with the formula for a simple closed polygon. Figure 1.1 shows an example of a network – the internet.

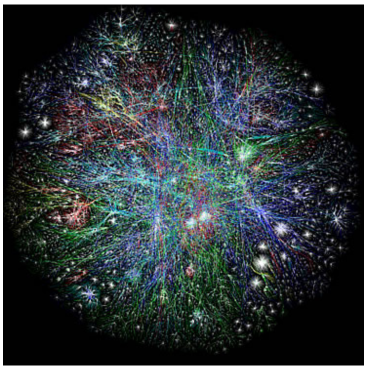


Figure 1.1: A Graphical Representation of the Internet[1]

While these theories have been applied to various fields of study, there has been a limited use of these in sports.

In this essay, using the data that has been made freely available after each match, I will focus on finding a quantifiable depiction of a team's style of play using network theory and probability. All football teams in history have had a distinct style of play and while traditionally it has been observed by football experts, in recent years, this focus has shifted to statisticians who have used complex mathematical techniques, such as probability and graph theory, to quantify the team's style of play. This essay will consist of two parts: the first part focusing on network theory by using the passing distribution and tactical plan of four teams to create a weighted network. The second part of the essay focuses on probability distributions and how it can qualitatively define the defence and attack of a team.

Hence, by the end of the essay, I would have an answer to my research question: How can the knowledge of probability and network theory help quantitatively and qualitatively analyse a football team?

[1] *Graphical Representation of the Internet*. 2011. *Flickr*. Web. 17 Apr. 2017. <https://www.flickr.com/photos/leyinglo/2150307459>.

INTRODUCTION TO NETWORK THEORY

The network theory provides a set of techniques for the analysis of graphs. The fundamental unit of a network is the *node/vertex* while the connections between them are called *edges.*

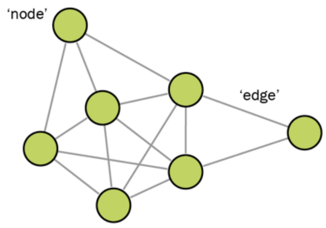


Figure 2.1: A Simple Network[2]

In 1736, Leonard Euler laid the fundamentals of network theory by attempting to solve the *Könisberg Bridge Problem* – which was a situation based on the seven bridges in the city of Könisberg. The problem was to devise a walk through the city that would cross all the bridges only once. Euler converted the problem to a simple graph, and through that, proved that it was impossible.

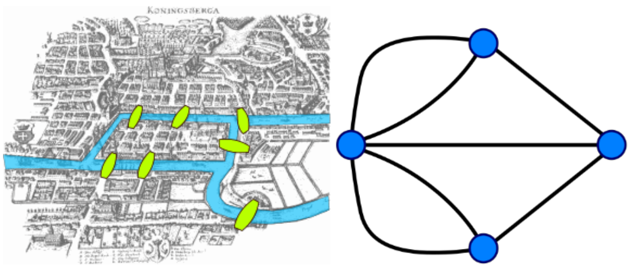


Figure 2.2: Euler's Interpretation of the Könisberg Problem[3]

Euler proved that the possibility to walk through each edge only once depends on the degree of its nodes (which is the number of edges touching it). He argued that the graph should have only two nodes of an odd degree, and hence it was not possible to solve the problem.

As the theory evolved, more concepts were introduced, some of which will be used in order to answer the research question. Key terms need to be defined before the application of this theory is discussed in detail:

---

[2] "Simple Network." *Social Physics*, social-physics.net/wp-content/uploads/2014/01/simplegraph.png.

[3] *Könisberg Graph*. Digital image. *Wikimedia*. Wikipedia, 8 June 2006. Web. 24 Apr. 2017.

1. A *weighted network* is one where the edges have weights. Weights are the strength of connection between two nodes. In some calculations for this essay, weighted networks will be used. The weights will be the number of passes between two players. For example, in figure 2.3, nodes *A* and *B* share an edge with a weight of 4.

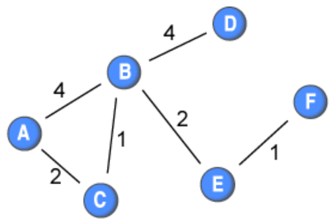


Figure 2.3: A Weighted Network[4]

2. The *geodesic distance* between two nodes, *u* and *v* is the minimum length of paths connecting them.[5]

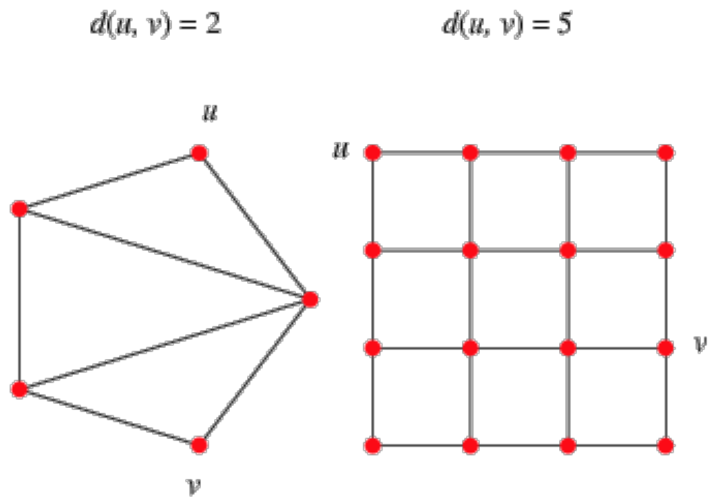


Figure 2.4: Geodesic Distance[6]

3. A *matrix* is a collection of numbers arranged into a fixed number of rows and columns.[7] An example can be seen below:

[4] Opsahl, Tore. “Weighted Network.” *Wikimedia Foundation*, Wikimedia Foundation, 30 Jan. 2007

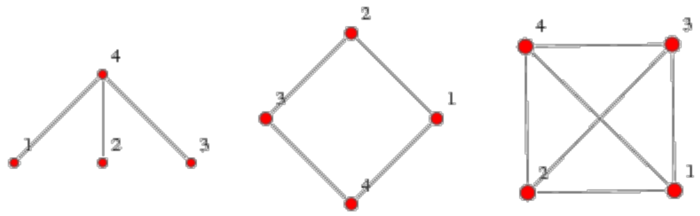
[5] Barile, Margherita. "Graph Distance." From *MathWorld*--A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/GraphDistance.html

[6] Barile, Margherita. "Graph Distance." From *MathWorld*--A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/GraphDistance.html

[7] http://chortle.ccsu.edu/vectorlessons/vmch13/vmch13_2.html

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

4. An *adjacency matrix* is a matrix used to represent a finite graph. The elements of the matrix show whether pairs of vertices are connected (adjacent) or not, in a graph.[8] Adjacency matrices will be used to construct passing matrices for each team. An example can be found here:



$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Figure 2.4: Adjacency Matrix[9]

Methodology

Using the concepts talked about above, a weighted adjacency network will be created using the teams' formations, as they are rough representations of a player's average position in the game. The number of passes between each player will decide the weights between each player, and the thickness of the edges between two players. Essentially, this is an extremely simplified graph as footballers keep moving. However, the weighted edges with varying thicknesses and

[8] Weisstein, Eric W. "Adjacency Matrix." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/AdjacencyMatrix.html

[9] Weisstein, Eric W. "Adjacency Matrix." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/AdjacencyMatrix.html

the calculations made on them can provide an insight into a team's style of play or tactic in the particular match.

This analysis will be made quantitative by computing:

- the global network variants, characterizing the team as a whole,
- and the local network variants, characterizing each player in a team.

The mathematics behind this is explained below.

A *weighted adjacency matrix* is defined as A_{xy} which represents the number of passes from player x to player y. A higher number of passes would result in a higher weight and a thicker edge.

For some calculations, such as where a binary result is needed (*1* for pass, *0* for no pass), a non-weighted adjacency matrix will be used and will be defined as follows:

$$\varepsilon_{xy} = \begin{cases} 1, & A_{xy} \neq 0 \\ 0, & A_{xy} = 0 \end{cases}$$

To measure individual player performance, the local network invariants will be calculated. To maintain simplicity, they will be segregated into three: *betweenness centrality*, *closeness centrality* and *PageRank centrality*.

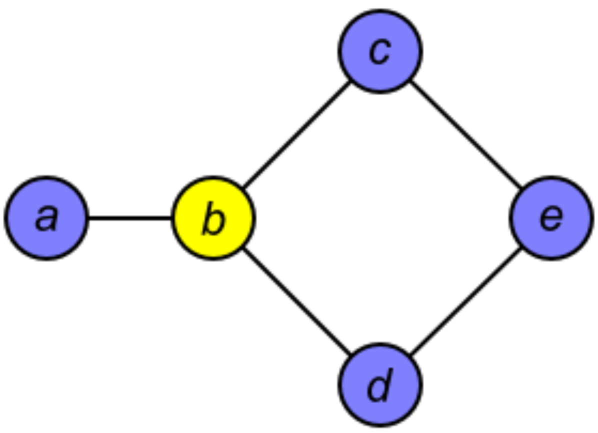


Figure 3.1: Betweenness Centrality

The betweenness centrality is defined as the measure of the extent to which a node lies on paths between other nodes. It is the ratio of the geodesic distances that go through a node. For instance, in *figure 3.1*, node *c* has a betweenness centrality of 1 because it is on the geodesic paths for *a* to *e* and *b* to *e.* Since there are two shortest paths (from *a* to *e* and *b* to *e),* which pass through *c*, the calculation becomes ½ + ½ = 1. This is an oversimplification for the calculations that are up ahead but it provides an insight of what betweenness centrality is. The formula for calculating betweenness centrality is:

$$C'_B\,(x) = \frac{1}{90}\sum_{x \neq y \neq z}\frac{n_{yz}^{x}}{g_{yz}}$$

Where n_{yz}^{x} is the number of geodesic paths from player *y* to player *z* that pass through player *x.* g_{yz} is the total number of geodesic paths between player *y* and player *z.* A normalisation factor of $\frac{1}{90}$ is used. The reason why $\frac{1}{90}$ is taken as the normalization factor is because a normalized betweenness centrality is calculated using this formula, $C'_B = \frac{C_B}{(N-1)(N-2)}$, where *N* is the number of nodes in the entire graph. In case of football, it is 11 nodes (because there are 11 players) and hence (11-1)×(11-2) is equal to 90.

The betweenness will measure how much the flow of the ball between two players, *y* and *z*, is dependent on player *x.* Hence, it provides an indication as to what will be the impact if the player is isolated by the rival team or gets a red card. Essentially, a team would want to have an equal distribution of betweenness centralities as concentrated betweenness would indicate that the team is dependent on a few players. For example, $C'_B\,(x) = 0$ would mean that player *x* is not influential in the game, and removing him won't have much of an effect.

The *closeness centrality* is measure of the inverse of the average geodesic distance of a node in a network.[10] It helps in understanding how easily reachable a player is in the team. It is calculated using the formula:

$$C_x = \frac{10}{\sum_y D_{yx}}$$

Where *N* is the number of nodes in the graph, and D_{yx} represents the geodesic distance between *y* and x.

The *PageRank centrality* measures the importance of a player using the following concept: The idea behind it is that if a player receives passes from an important player, he, too, is important. Mathematically, it is defined through this formula:

$$x_x = p \sum_{x \neq y} \frac{A_{xy}}{L_y^{out}} x_y + q$$

Where p represents the probability that the player *x* will pass the ball away rather than keeping it to himself or shooting, and q represents the popularity awarded to each player. L_y^{out} represents the total number of passes made by player *y*. This roughly assigns a probabilistic value that a player will have the ball after some passes have been made. For this analysis, the values of p and q will be *0.85* and *1* respectively.[11]

For this scenario, PageRank centrality helps in judging the 'importance' of a player in a team as it assigns the probabilistic value of a player having a ball after a certain number of passes have been made.

[10] Peña, Javier López, Dr., and Hugo Touchette, Dr. "A Network Theory Analysis of Football Strategies." Thesis. 2012. A Network Theory Analysis of Football Strategies. 2 July 2012. Web. 17 Apr. 2017.

[11] The values of *p* and *q* are 0.85 and 1 just to show the proof of concept as they are widely used.

Clustering is a measure of the degree to which nodes in a network tend to cluster together.[12] In the context of football, imagine if *x* wants to pass the ball to *z,* but the lane is covered by defenders and hence *x* passes the ball to *y* who then passes it to *z.* If there is a high number of passes from *x* to *y* to *z*, there is a higher clustering score for *y*. The formula for clustering score is:

$$c_z^x = \frac{1}{u_y(u_y - 1)} \sum_{x,z} \frac{\sqrt[3]{A_{xy}A_{yz}A_{xz}}}{\text{maximum } (A)}$$

Where $u_y = \sum_x A_{yx}$, is the *vertex-out degree* that is the passes made by *y,* A_{xy} is the number of passes from *x* to *y*, A_{yz} is the number of passes made by *y* to *z* and A_{xz} is the number of passes from *x* to *z.*

[12] Peña, Javier López, Dr., and Hugo Touchette, Dr. "A Network Theory Analysis of Football Strategies." Thesis. 2012. A Network Theory Analysis of Football Strategies. 2 July 2012. Web. 17 Apr. 2017.

Data Used[13]

The data that has been used is provided by Opta Sports from the *2010/11 UEFA Champions League Final* between Football Club Barcelona and Manchester United Football Club and the *2014/15 UEFA Champions League Final* between Football Club Barcelona and Juventus Football Club. These two matches have been chosen because they are comparable – FC Barcelona won both of them 3-1 – but the style of play in both matches differed immensely.

To avoid confusion regarding substitutions, only the passing data of the 11 players starting the match have been used. Using the data, the following passing tables were created.

The left-hand side column shows the players that the passes are **from**, and the top row shows the players to which the passes are directed **to**.

<u>Table 1:</u> Football Club Barcelona Passing Table vs Manchester United (2010/11 Champions League Final)

From/To	Valdés	Piqué	Mascherano	Alves	Abidal	Xavi	Busquets	Iniesta	Pedro	Messi	Villa
Valdés	0	3	6	3	0	1	3	1	2	0	1
Piqué	4	0	5	1	6	16	6	4	2	1	1
Mascherano	5	7	0	8	4	11	6	1	0	9	2
Alves	1	1	7	0	0	17	9	4	0	17	8
Abidal	2	5	0	0	0	12	7	15	9	4	1
Xavi	2	5	15	22	8	0	12	33	11	23	11
Busquets	0	6	4	6	9	12	0	17	3	15	4
Iniesta	0	4	3	8	7	27	12	0	11	32	2
Pedro	0	0	1	0	9	2	2	15	0	3	0
Messi	0	2	3	11	1	32	10	25	3	0	3
Villa	0	0	1	6	0	4	2	2	1	7	0

[13] The data has been obtained from *Opta Sports*, a world-renowned sports data company.

Table 2: **Manchester United Football Club Passing Table vs FC Barcelona (2010/11 Champions League Final)**

From/To	Sar	Evra	Ferdinand	Vidic	Fabio	Giggs	Ji-Sung	Valencia	Carrick	Rooney	Hernandez
Sar	**0**	**1**	**6**	**1**	**2**	**1**	**1**	**2**	**1**	**4**	**0**
Evra	**0**	**0**	**1**	**4**	**0**	**4**	**6**	**0**	**2**	**6**	**1**
Ferdinand	**7**	**2**	**0**	**6**	**0**	**1**	**5**	**6**	**5**	**3**	**0**
Vidic	**7**	**4**	**10**	**0**	**0**	**2**	**2**	**0**	**2**	**1**	**0**
Fabio	**0**	**0**	**3**	**0**	**0**	**5**	**0**	**4**	**0**	**3**	**0**
Giggs	**0**	**4**	**2**	**2**	**4**	**0**	**5**	**2**	**1**	**7**	**2**
Ji-Sung	**0**	**4**	**0**	**2**	**1**	**3**	**0**	**1**	**5**	**6**	**1**
Valencia	**1**	**0**	**6**	**1**	**3**	**1**	**0**	**0**	**1**	**3**	**0**
Carrick	**2**	**1**	**1**	**1**	**4**	**6**	**1**	**1**	**0**	**5**	**6**
Rooney	**0**	**2**	**1**	**2**	**1**	**8**	**5**	**5**	**5**	**0**	**4**
Chicharito	**0**	**0**	**0**	**0**	**0**	**3**	**1**	**1**	**3**	**4**	**0**

Table 3: **Juventus Football Club Passing Table vs FC Barcelona (2014/15 Champions League Final)**

From/To	Buffon	Barzagli	Bonucci	Lichtsteiner	Evra	Pogba	Marchisio	Pirlo	Vidal	Tevez	Morata
Buffon	**0**	**3**	**1**	**2**	**1**	**3**	**3**	**4**	**0**	**2**	**0**
Barzagli	**2**	**0**	**6**	**3**	**0**	**4**	**2**	**6**	**3**	**0**	**1**
Bonucci	**3**	**6**	**0**	**2**	**4**	**4**	**1**	**3**	**3**	**0**	**1**
Lichtsteiner	**1**	**1**	**0**	**0**	**0**	**1**	**2**	**3**	**12**	**2**	**1**
Evra	**0**	**0**	**1**	**0**	**0**	**8**	**1**	**2**	**3**	**5**	**0**
Pogba	**0**	**1**	**3**	**1**	**7**	**0**	**2**	**3**	**1**	**4**	**2**
Marchisio	**0**	**0**	**6**	**6**	**0**	**1**	**0**	**6**	**4**	**3**	**3**
Pirlo	**0**	**9**	**3**	**1**	**6**	**5**	**5**	**0**	**3**	**9**	**1**
Vidal	**0**	**5**	**0**	**2**	**2**	**4**	**3**	**8**	**0**	**0**	**7**
Tevez	**0**	**1**	**1**	**0**	**3**	**3**	**5**	**3**	**4**	**0**	**5**
Morata	**0**	**0**	**1**	**1**	**1**	**1**	**2**	**1**	**0**	**2**	**0**

<u>Table 4:</u> Football Club Barcelona Passing Table vs Juventus FC (2014/15 Champions League Final)

From/To	ter Stegen	Piqué	Mascherano	Alba	Alves	Busquets	Iniesta	Rakitic	Messi	Neymar	Suarez
ter Stegen	**0**	**5**	**3**	**5**	**3**	**0**	**2**	**0**	**1**	**0**	**2**
Piqué	**7**	**0**	**5**	**4**	**12**	**0**	**1**	**2**	**2**	**0**	**1**
Mascherano	**1**	**4**	**0**	**10**	**8**	**7**	**2**	**3**	**1**	**3**	**1**
Alba	**1**	**0**	**13**	**0**	**1**	**9**	**14**	**0**	**0**	**21**	**4**
Alves	**0**	**6**	**1**	**4**	**0**	**10**	**1**	**21**	**25**	**0**	**9**
Busquets	**2**	**6**	**6**	**6**	**16**	**0**	**9**	**7**	**9**	**0**	**2**
Iniesta	**0**	**1**	**4**	**14**	**3**	**7**	**0**	**2**	**7**	**10**	**2**
Rakitic	**0**	**3**	**3**	**2**	**14**	**7**	**2**	**0**	**12**	**1**	**1**
Messi	**0**	**2**	**0**	**5**	**10**	**10**	**9**	**7**	**0**	**12**	**6**
Neymar	**0**	**0**	**1**	**8**	**3**	**2**	**7**	**1**	**10**	**0**	**4**
Suarez	**0**	**1**	**0**	**0**	**4**	**1**	**2**	**3**	**7**	**4**	**0**

Using these passing tables, the following adjacency matrices were created:

FC Barcelona Adjacency Matrix vs Manchester United FC (2010/11 Champions League Final)

$$\begin{pmatrix} 0 & 3 & 6 & 3 & 0 & 1 & 3 & 1 & 2 & 0 & 1 \\ 4 & 0 & 5 & 1 & 6 & 16 & 6 & 4 & 2 & 1 & 1 \\ 5 & 7 & 0 & 8 & 4 & 11 & 6 & 1 & 0 & 9 & 2 \\ 1 & 1 & 7 & 0 & 0 & 17 & 9 & 4 & 0 & 17 & 8 \\ 2 & 5 & 0 & 0 & 0 & 12 & 7 & 15 & 9 & 4 & 1 \\ 2 & 5 & 15 & 22 & 8 & 0 & 12 & 33 & 11 & 23 & 11 \\ 0 & 6 & 4 & 6 & 9 & 12 & 0 & 17 & 3 & 15 & 4 \\ 0 & 4 & 3 & 8 & 7 & 27 & 12 & 0 & 11 & 32 & 2 \\ 0 & 0 & 1 & 0 & 9 & 2 & 2 & 15 & 0 & 3 & 0 \\ 0 & 2 & 3 & 11 & 1 & 32 & 10 & 25 & 3 & 0 & 3 \\ 0 & 0 & 1 & 6 & 0 & 4 & 2 & 2 & 1 & 7 & 0 \end{pmatrix}$$

Manchester United FC Adjacency Matrix vs FC Barcelona (2010/11 Champions League Final)

$$\begin{pmatrix} 0 & 1 & 6 & 1 & 2 & 1 & 1 & 2 & 1 & 4 & 0 \\ 0 & 0 & 1 & 4 & 0 & 4 & 6 & 0 & 2 & 6 & 1 \\ 7 & 2 & 0 & 6 & 0 & 1 & 5 & 6 & 5 & 3 & 0 \\ 7 & 4 & 10 & 0 & 0 & 2 & 2 & 0 & 2 & 1 & 0 \\ 0 & 0 & 3 & 0 & 0 & 5 & 0 & 4 & 0 & 3 & 0 \\ 0 & 4 & 2 & 2 & 4 & 0 & 5 & 2 & 1 & 7 & 2 \\ 0 & 4 & 0 & 2 & 1 & 3 & 0 & 1 & 5 & 6 & 1 \\ 1 & 0 & 6 & 1 & 3 & 1 & 0 & 0 & 1 & 3 & 0 \\ 2 & 1 & 1 & 1 & 4 & 6 & 1 & 1 & 0 & 5 & 6 \\ 0 & 2 & 1 & 2 & 1 & 8 & 5 & 5 & 5 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 3 & 1 & 1 & 3 & 4 & 0 \end{pmatrix}$$

Juventus FC Adjacency Matrix vs FC Barcelona (2014/15 Champions League Final)

$$\begin{pmatrix} 0 & 3 & 1 & 2 & 1 & 3 & 3 & 4 & 0 & 2 & 0 \\ 2 & 0 & 6 & 3 & 0 & 4 & 2 & 6 & 3 & 0 & 1 \\ 3 & 6 & 0 & 2 & 4 & 4 & 1 & 3 & 3 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 2 & 3 & 12 & 2 & 1 \\ 0 & 0 & 1 & 0 & 0 & 8 & 1 & 2 & 3 & 5 & 0 \\ 0 & 1 & 3 & 1 & 7 & 0 & 2 & 3 & 1 & 4 & 2 \\ 0 & 0 & 6 & 6 & 0 & 1 & 0 & 6 & 4 & 3 & 3 \\ 0 & 9 & 3 & 1 & 6 & 5 & 5 & 0 & 3 & 9 & 1 \\ 0 & 5 & 0 & 2 & 2 & 4 & 3 & 8 & 0 & 0 & 7 \\ 0 & 1 & 1 & 0 & 3 & 3 & 5 & 3 & 4 & 0 & 5 \\ 0 & 0 & 1 & 1 & 1 & 1 & 2 & 1 & 0 & 2 & 0 \end{pmatrix}$$

FC Barcelona Adjacency Matrix vs Juventus FC (2014/15 Champions League Final)

$$\begin{pmatrix} 0 & 5 & 3 & 5 & 3 & 0 & 2 & 0 & 1 & 0 & 2 \\ 7 & 0 & 5 & 4 & 12 & 0 & 1 & 2 & 2 & 0 & 1 \\ 1 & 4 & 0 & 10 & 8 & 7 & 2 & 3 & 1 & 3 & 1 \\ 1 & 0 & 13 & 0 & 1 & 9 & 14 & 0 & 0 & 21 & 4 \\ 0 & 6 & 1 & 4 & 0 & 10 & 1 & 21 & 25 & 0 & 9 \\ 2 & 6 & 6 & 6 & 16 & 0 & 9 & 7 & 9 & 0 & 2 \\ 0 & 1 & 4 & 14 & 3 & 7 & 0 & 2 & 7 & 10 & 2 \\ 0 & 3 & 3 & 2 & 14 & 7 & 2 & 0 & 12 & 1 & 1 \\ 0 & 2 & 0 & 5 & 10 & 10 & 9 & 7 & 0 & 12 & 6 \\ 0 & 0 & 1 & 8 & 3 & 2 & 7 & 1 & 10 & 0 & 4 \\ 0 & 1 & 0 & 0 & 4 & 1 & 2 & 3 & 7 & 4 & 0 \end{pmatrix}$$

The Network

Constructing the Network

To make sense of the matrices, a weighted network was created with nodes (players) fixed in positions corresponding to the team's formation. The weight increased with the number of passes between players. This is an extremely simple way to gauge a team's strategy but can provide a starting point for the analyses of the style of play. Since it is a visual tool, it can show the areas of the field which are used or unused, the passing strategy of the team (short passes or long passes), the involvement of each player, the most important players in the team and whether the team utilized all of its players well or not. No calculations were made while constructing these graphs and they were used as visual tools to help enhance the analysis. These networks were made using *Wolfram's Mathematica*. Here is a snippet of the code used for data in **Table 1**:

Needs

```
Needs["Combinatorica`"]
GraphFromAdjacency[adjm_, namelist_, n_] := Module[{i, j}, Flatten[Table[{namelist[[i]] → namelist[[j]], adjm[[i, j]]}, {i, n}, {j, n}], 1]]
```

Images

```
crestbarca = ;

jerseybarcelona = ;
```

Data

```
posnumbersbarca = {1 → {0, 0}, 3 → {1, 1}, 14 → {-1, 1}, 2 → {2, 2}, 22 → {-2, 2}, 6 → {1, 3}, 16 → {0, 2}, 8 → {-1, 3}, 17 → {2, 4}, 10 → {0, 5}, 7 → {-2, 4}};
numbersbarca = {1, 3, 14, 2, 22, 6, 16, 8, 17, 10, 7};
adjbarca = {{0, 3, 6, 3, 0, 1, 3, 1, 2, 0, 1}, {4, 0, 5, 1, 6, 16, 6, 4, 2, 1, 1}, {5, 7, 0, 8, 4, 11, 6, 1, 0, 9, 2}, {1, 1, 7, 0, 0, 17, 9, 4, 0, 17, 8},
   {2, 5, 0, 0, 0, 12, 7, 15, 9, 4, 1}, {2, 5, 15, 22, 8, 0, 12, 33, 11, 23, 11}, {0, 6, 4, 6, 9, 12, 0, 17, 3, 15, 4}, {0, 4, 3, 8, 7, 27, 12, 0, 11, 32, 2},
   {0, 0, 1, 0, 9, 2, 2, 15, 0, 3, 0}, {0, 2, 3, 11, 1, 32, 10, 25, 3, 0, 3}, {0, 0, 1, 6, 0, 4, 2, 2, 1, 7, 0}};
posbarca = numbersbarca /. posnumbersbarca;
playersbarca = {"VALDES", "PIQUE", "MASCHERANO", "ALVES", "ABIDAL", "XAVI", "BUSQUETS", "INIESTA", "PEDRO", "MESSI", "VILLA"}
graphnumbersbarca = GraphFromAdjacency[adjbarca, numbersbarca, 11]
graphnamesbarca = GraphFromAdjacency[adjbarca, playersbarca, 11]
```

Graphics

```
maxpassbarca = Max[adjbarca]
GraphPlot[graphnumbersbarca,
 EdgeRenderingFunction → ({Blend[{Blue, Black}, #3 / maxpassbarca], AbsoluteThickness[0.2 + #3^2 / 20], Arrowheads[0.01 + #3 / 175], Arrow[#1, 0.45]} &),
VertexLabeling → True,
SelfLoopStyle → None,
VertexCoordinateRules → posnumbersbarca,
Frame → False,
VertexRenderingFunction → (Inset[ImageCompose[jerseybarcelona, Graphics[Text[Style[#2, 25, Bold, White]]]], #2 /. posnumbersbarca, {Center, Center}, 0.75] &),
Epilog → {Inset[crestbarca, {2.25, -0.25}, {Right, Bottom}, 0.7], Inset[Style["Data from Opta", FontSize → 8], {-2.25, -0.25}, {Left, Bottom}]}]
```

Figure 5.1: The code used to create the graph for FC Barcelona vs Manchester United (2010/11)

In the figure above, just for convenience, the code has been divided into distinct sections. The *Needs* and the *Images* section gather the required mathematical and graphical tools for the construction of the network.

The *Data* section contains several variables which are explained below:

1. *posnumbersbarca* is the variable that assigns the coordinates of each player on the graph, with the goalkeeper starting at point (0, 0). Each coordinate has to be individually inputted so that it corresponds to the player's location in the formation. For example, a player with the position coordinate (1, 1) is 1 up and 1 right of the goalkeeper.

2. *numbersbarca* and *playersbarca* are arrays that store the jersey numbers and names of each player.

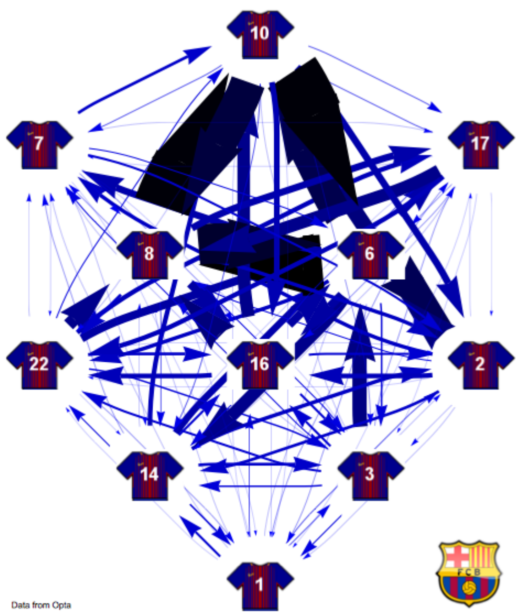
3. *adjbarca* contains the data of the first adjacency matrix on page 13.
4. *graphnumbersbarca* and *graphnamesbarca* relate the adjacency matrix to each respective player.

Under the *Graphics* subsection, one more variable is defined: *maxpassbarca.* That is the highest number of passes between two players in the entire team. In this case it would be equal to 33, which is the number of passes from Xavi to Iniesta.

The graph is made by using the native *GraphPlot* function. The main purpose of it is to combine the data from the *Data* subsection and the graphics from the *Images* subsection to create a network. The positions of the nodes (players) have already been fixed by the variable *posnumbersbarca* and hence the only aspect left is the arrows in between the players to signify the passes. Since this is going to be a weighted network, the arrow lengths and width must increase and decrease according to the weight assigned to it. Another native function, called the *EdgeRenderingFunction,* is used to achieve this. The darkness and thickness on each arrow depends on the *maxpassbarca* variable. The first statement of the *EdgeRenderingFunction* defines the colour of the arrow – it gets darker as the number of passes increases. The second statement defines the thickness of the arrow and it is defined by the formula shown in the screenshot. The third and fourth statements define the thickness of the arrowhead and the length of the arrow which signify the weights on the edges.

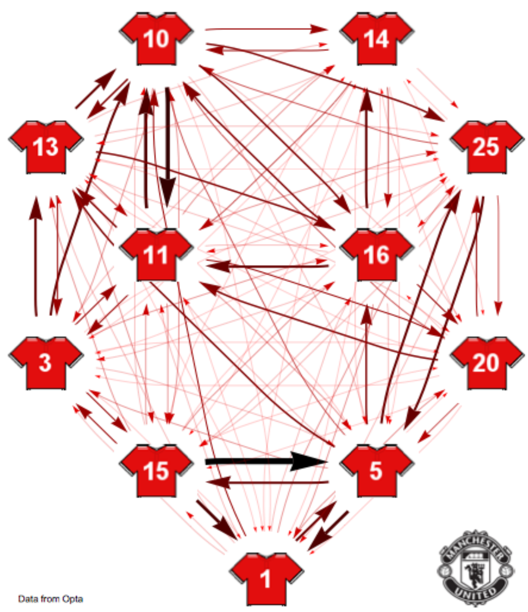
The Final Networks

Figure 6.1: FC Barcelona Passing Network vs Manchester United FC



1: Víctor Valdés, 3: Gerard Piqué, 14: Javier Mascherano, 2: Dani Alves, 22: Eric Abidal, 16: Sergio Busquets, 8: Andrés Iniesta, 6: Xavi Hernandez, 17: Pedro Rodriguez, 7: David Villa, 10: Lionel Messi

Figure 6.2: Manchester United Passing Network vs FC Barcelona



1: Edwin van der Sar, 5: Rio Ferdinand, 15: Nemanja Vidić, 3: Fábio, 20: Patrice Evra, 11: Ryan Giggs, 16: Michael Carrick, 25: Antonio Valencia, 13: Park Ji-Sung, 10: Wayne Rooney, 14: Javier Hernández

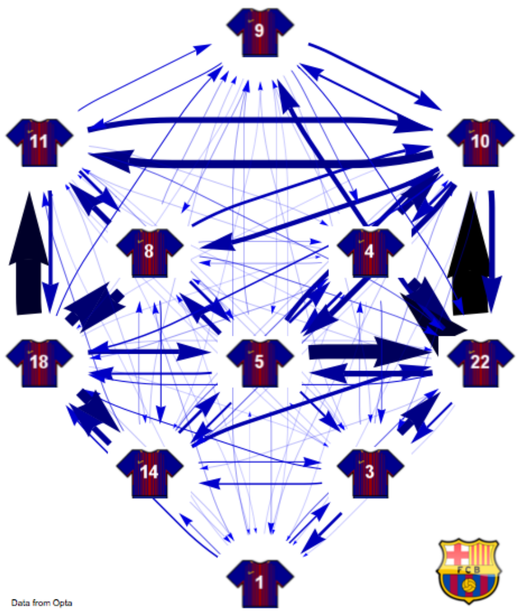
FC Barcelona's starting formation in this match was with a goalkeeper, 4 defenders, 3 midfielders and 3 attackers (a 4-3-3). Manchester United, on the other hand, had a goalkeeper, 4 defenders, 4 midfielders and 2 attackers (a 4-4-2). A pre-quantitative visual analysis can provide an immediate insight into each team's tactics and the narrative of the game. FC Barcelona has dominated the midfield in this game, as shown by the extremely thick and dark arrows between the three midfielders (#8, #6 and #16). The result of this can be seen in Manchester United's passing network where the midfield has been starved of the ball. Another point to be noted in the difference in playing styles is that FC Barcelona has adopted a short-distance passing tactic, with a maximum number of passes between players that are close to each other, whereas Manchester United has adopted a long-distance passing strategy, with a large number of passes going from the central defenders (#15 and #5) to the wingers (#13 and #25).

One area that Manchester United do have an advantage is that they do not rely on two or three players to keep possession of the ball like FC Barcelona do. This means that if Manchester United found a way to isolate those three players (Barcelona's #8, #6 and #10), they would effectively stop Barcelona from playing their game.

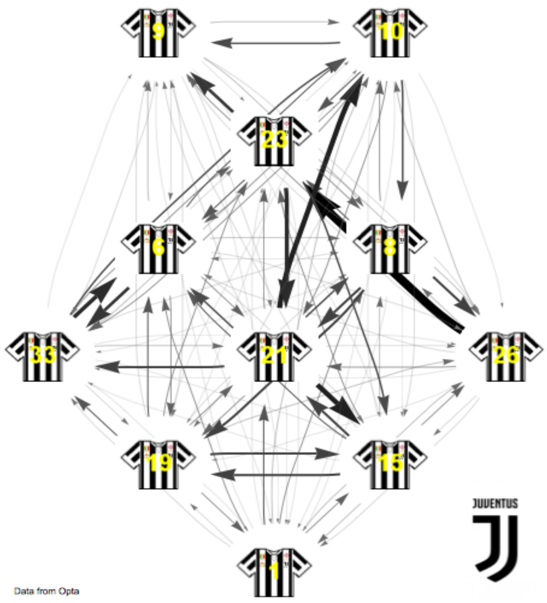
It is amazing to see how much information just a preliminary look at the network graphics can give. However, our understanding will be deepened when we look at the calculations on this network. The process is repeated for Table 3 and 4 to obtain the corresponding networks shown below.

Figure 6.3: FC Barcelona Passing Network vs Juventus FC

Figure 6.4: FC Barcelona Passing Network vs Juventus FC



1: Marc-André ter Stegen, 3: Gerard Piqué, 14: Javier Mascherano,
22: Dani Alves, 18: Jordi Alba, 5: Sergio Busquets, 8: Andrés Iniesta, 4: Ivan Rakitić, 10: Lionel Messi, 11: Neymar da Silva, 9: Luis Suárez



1: Gianluigi Buffon, 15: Andrea Barzagli, 19: Leonardo Bonucci,
33: Patrice Evra, 26: Stephan Lichtsteiner, 21: Andrea Pirlo, 8: Claudio Marchisio, 6: Paul Pogba, 23: Arturo Vidal, 9: Álvaro Morata, 10: Carlos Tevez

Figures 6.3 and 6.4 have been built on the 2014/15 *UEFA Champions League Final* that FC Barcelona won. This match took place about 4 years after the previous one. FC Barcelona has a similar formation to the match against Manchester United whereas Juventus employ a tactic with 4 defenders, 3 midfielders, 1 attacking midfielder and 2 strikers.

Neither team in this match really control the midfield as shown by the dearth of thick arrows there. FC Barcelona rely on the sides of the field to get the ball forward while Juventus FC do not seem to have a clear preference in terms of ball-flow. For FC Barcelona, the wingers (#10 and #11) have become essential as most passes go to them. This is in stark contrast to the FC Barcelona of 2010/11 whose style emphasized on control in the midfield.

In the next section, the results of the calculations are analysed.

Results

Calculations

The calculations for the measures defined in the methodology (i.e. Betweenness Centrality, Closeness Centrality, PageRank Centrality and Local Clustering Coefficient) were also done using Wolfram Mathematica. There was no need to input the formulae as there are pre-defined functions in the software. Here is a snapshot of the code used:

Calculations

```
betweennessbarca = BetweennessCentrality[WeightedAdjacencyGraph[adjbarca /. {0 → ∞}]]
Thread[{numbersbarca, playersbarca, betweennessbarca}] // TableForm

closenessbarca = ClosenessCentrality[WeightedAdjacencyGraph[adjbarca /. {0 → ∞}]]
Thread[{numbersbarca, playersbarca, closenessbarca}] // TableForm

rankbarca = PageRankCentrality[WeightedAdjacencyGraph[adjbarca /. {0 → ∞}]]
Thread[{numbersbarca, playersbarca, rankbarca}] // TableForm

FindClique[WeightedAdjacencyGraph[adjbarca /. {0 → ∞}]]
FindClique[WeightedAdjacencyGraph[adjbarca /. {0 → ∞}]][[1]]

MeanClusteringCoefficient[WeightedAdjacencyGraph[adjbarca]]
localbarca = LocalClusteringCoefficient[WeightedAdjacencyGraph[adjbarca /. {0 → ∞}]]
Thread[{numbersbarca, playersbarca, localbarca}] // TableForm

AverageBetweenness = Mean[betweennessbarca]

Thread[{numbersbarca, playersbarca, betweennessbarca, closenessbarca, rankbarca, localbarca}]
```

Table 2.1: Result Table for FC Barcelona 2010/11

NUMBER	NAME	BETWEENESS CENTRALITY	CLOSENESS CENTRALITY	PAGERANK CENTRALITY	LOCAL CLUSTERING COEFFICIENT
1	VALDES	0.536	0.400	0.0568	0.861
3	PIQUÉ	1.59	**0.500**	0.0849	0.833
14	MASCHERANO	**2.59**	0.244	0.0976	0.890
2	ALVES	1.23	**0.435**	0.0857	0.893
22	ABIDAL	1.39	0.333	0.0805	**0.940**
6	XAVI	**3.13**	0.227	**0.106**	0.811
16	BUSQUETS	1.86	0.189	**0.105**	0.827
8	INIESTA	1.86	0.233	**0.105**	0.827
17	PEDRO	0.583	0.313	0.0850	**0.930**
10	MESSI	1.57	0.345	0.0999	0.861
7	VILLA	0.661	0.286	0.0922	0.877

Table 2.2: Result Table for Manchester United 2010/11

NUMBER	NAME	BETWEENESS CENTRALITY	CLOSENESS CENTRALITY	PAGERANK CENTRALITY	LOCAL CLUSTERING COEFFICIENT
1	DER SAR	1.10	**0.667**	0.0570	0.813
3	EVRA	0.967	0.435	0.0835	0.814
5	FERDINAND	3.00	0.357	0.0883	**0.820**
15	VIDIC	1.5	0.417	0.0923	**0.857**
20	FABIO	0.167	0.217	0.0728	**0.857**
11	GIGGS	**5.18**	0.455	**0.121**	0.679
13	JI-SUNG	2.23	0.500	0.0947	0.724
25	VALENCIA	**3.32**	**0.556**	0.097	0.76
16	CARRICK	**5.18**	**0.667**	**0.107**	0.716
10	ROONEY	**5.18**	0.370	**0.121**	0.679
14	HERNANDEZ	0.167	0.476	0.0657	0.0904

Table 2.3: Result Table for FC Barcelona 2014/15

NUMBER	NAME	BETWEENESS CENTRALITY	CLOSENESS CENTRALITY	PAGERANK CENTRALITY	LOCAL CLUSTERING COEFFICIENT
1	TER STEGEN	0.310	0.323	0.0509	0.800
3	PIQUÉ	2.20	0.370	0.0895	0.807
14	MASCHERANO	**2.88**	0.417	0.0878	0.764
18	ALBA	2.57	0.370	0.0941	0.807
22	ALVES	1.55	0.417	**0.107**	0.819
5	BUSQUETS	**2.99**	0.244	0.0938	**0.862**
8	INIESTA	2.25	0.333	**0.108**	0.802
4	RAKITIC	1.24	**0.476**	0.092	0.859
10	MESSI	1.46	0.227	0.0967	0.800
11	NEYMAR	0.762	**0.455**	0.0745	**0.881**
9	SUAREZ	1.80	0.323	0.106	0.825

Table 2.4: Result Table for Juventus FC 2014/15

NUMBER	NAME	BETWEENESS CENTRALITY	CLOSENESS CENTRALITY	PAGERANK CENTRALITY	LOCAL CLUSTERING COEFFICIENT
1	BUFFON	0.786	0.435	0.0407	0.762
15	BARZAGLI	2.39	0.476	0.0787	0.780
19	BONUCCI	**4.99**	0.455	0.0932	0.754
26	LICHTSTEINER	**3.42**	**0.625**	0.0932	0.789
33	EVRA	0.744	0.400	0.0805	0.784
6	POGBA	2.77	0.476	**0.113**	0.741
8	MARCHISIO	1.706	0.357	**0.110**	0.794
21	PIRLO	2.77	0.477	**0.113**	0.741
23	VIDAL	1.376	0.333	0.0959	**0.820**
10	TEVEZ	1.62	0.410	0.0856	0.804
9	MORATA	1.43	**0.667**	0.0962	**0.880**

Data Interpretation

The tables show the betweenness, closeness, page rank and clustering scores of the four teams and depict the way the teams carried out their tactics.

The four measures bring out aspects of the game that are difficult to be seen while watching the game. For the first match, FC Barcelona versus Manchester United, there is a higher score for betweenness, PageRank and closeness centralities for the Manchester United team. The two highest betweenness scores belong to four Manchester United players: Michael Carrick, Ryan Giggs, Antonio Valencia and Wayne Rooney. Since betweenness is a measure of the importance of each player in a team, it is evident that these three are the backbone of the team. There are five players in the team with an extremely high betweenness centrality (Antonio Valencia, Ryan Giggs, Rio Ferdinand, Michael Carrick and Wayne Rooney). This is a weakness in their team as it means that if any one of those five players are isolated, the passing network could be disconnected. FC Barcelona has a low and consistent betweenness score which implies that there is less reliance on individual skill and an emphasis on team play. The low betweenness scores also mean that, unlike Manchester United, if a player is isolated from the network, there won't be much of an effect to the overall passing network as the reliance on individual talent is extremely low. The PageRank centrality scores echo the same. The highest two PageRank centralities also belong to three of the four players with the highest betweenness scores: Giggs, Carrick and Rooney. Again, this exhibits the importance of these players in the team. While these players are phenomenal in every sense, the Manchester United team over-relies on them and that can be exploited. FC Barcelona's two highest PageRank scores belong to the three midfielders – Xavi, Iniesta and Busquets. The style of playing, called *Juego de Posicion* (or Positional Play), emphasises on control in the midfield, and the PageRank scores reiterate the same. The one measure where Manchester United has the upper hand in is the closeness centrality. Most of the players have high closeness scores which correspond to those players being well connected in the network and having a short geodesic passing distance. FC Barcelona, on the other hand, do not have such high closeness scores which indicate that the team isn't extremely well connected or that most passes were long-distance ones. Two FC Barcelona players have the two-highest local clustering coefficient scores – Eric Abidal and Pedro Rodriguez – indicating that the players were involved in a lot of interplay. The measures generally point towards the fact that FC Barcelona dominated the game, which is true, as the game ended in a 3-1 win for the Catalans.

The second match was played in May of 2015, 4 years after the first one and there is a huge difference in the style of play. The key elements of FC Barcelona's game remain the same. There is an emphasis on team play over individual skill as the betweenness score is lower than that of Juventus'. The two highest betweenness scores belong to Leonardo Bonucci and Stephan Lichsteiner – two defenders which indicate that Juventus was on the backfoot the entire match. The biggest shift in style for FC Barcelona is shown through the PageRank centrality. In the match from 2011, the three FC Barcelona midfielders had the highest PageRank Centrality scores, however in this match, it is the three Juventus FC midfielders – Pirlo, Marchisio and Pogba – that have the two highest PageRank scores. This signals a shift from the traditional FC Barcelona style of control in the midfield to playing through the wings. Mathematically speaking, the weight of the edges between the three midfielders have dramatically decreased and the weight of the edges surrounding the wingers have increased. The two-highest local clustering coefficient scores are of Neymar and Morata – two attackers. It is unusual to see two attackers being involved in the passing network so heavily but that seemed to be the tactic employed by the manager in those matches.

Limitations of Using Network Theory

As with any exploration, there are limitations to this approach of analysing football matches. Since football is a dynamic game, the position of the players is constantly changing. This aspect has not been taken into account while constructing the networks used for the visual analysis. Manual work also implies that there is scope for manual errors. While all the data used in this essay has come directly from Opta Sports, a world-renowned sports data company, the manual typing of the number of passes to and fro each player could lead to some errors. To reduce the chance of such errors, each dataset was reviewed three times before being used.

Introduction to Probabilistic Method

This method of finding the strengths and weaknesses of a football team focuses only on two aspects of the game: attacking and defending. Using this parameterized probabilistic model, one can quantify a team's strength in defence and attack. Complex variations of this model are used extensively by betting companies while setting the odds.

The *Parameterized probabilistic model* divides the 90 minutes of the game into 6 parts of 15 minutes. In each part, there is a probability, p, that team A scores. Assuming that team A scores a maximum of one goal in each period, the number of goals scored would follow the Bernoulli distribution.[14]

The Bernoulli distribution is the probability distribution of a random variable which takes the value 1 with probability p and a value 0 with a probability $q = 1 - p$. It is the probability

[14] Taïeb, Franck. "A Model of Football Games." Science4All, 31 Jan. 2016, www.science4all.org/article/a-model-of-football-games/.

distribution of any single experiment that has a Boolean-valued answer. It is a special case of the binomial distribution where only a single experiment is conducted (n = 1). [15]

$$f(k;p) = \begin{cases} q = (1-p), & for\ k = 0 \\ p, & for\ k = 1 \end{cases};$$

It can also be expressed as:

$$f(k;p) = p^k (1-p)^{1-k}\ for \in \{0,1\};$$

Let *p* be the probability of scoring in a period. The goals that team *A* scores is the sum total of goals in all 6 periods. Hence, it is the sum of 6 independent and identically distributed Bernoulli variables which have a value of 1 with a probability, *p.*

If *Team A* scores in the 3rd and 6th periods only, the probability is:

$$\boldsymbol{Probability\ of\ A\ Scoring = (1-p)(1-p)p(1-p)(1-p)p = p^2(1-p)^4}$$

However, this is the probability of scoring twice in any of the 6 time periods. To compute the probability of *Team A* scoring twice, the probabilities of all configurations must be added up. This is done by using the binomial coefficient: $\binom{6}{2}$ or *6 choose 2*. In this scenario, there are 15 ways in which *Team A* can score 2 goals.

To generalize it, if there are *n* periods, then the number of configurations when *team A* scores *k* goals equals to $\binom{n}{k}$. The generalized equation then becomes:

[15] "Bernoulli Distribution." Wikipedia, Wikimedia Foundation, 16 Nov. 2017, en.wikipedia.org/wiki/Bernoulli_distribution.

$$P(k\ Goals) = \binom{n}{k} p^k (1-p)^{n-k}$$

This is the binomial law.

Poisson Distribution

However, this has one limitation – it assumes that only one goal can be scored per 15 minutes. This is not the case in a football match. A way to overcome this is by increasing the number of periods to infinity. The binomial law is now converged to a Poisson distribution. It depends on a single parameter, λ, and denotes the expected number of a goals.

To define the expected number of goals, λ, two new variables need to be introduced. One for *Strength in Attack* and one for *Weakness in Defense.* So, for example, if *Team A* faces *Team B*, the average number of goals scored by *A* against *B* would be denoted by λ.

$$Expected\ Number\ of\ Goals\ Scored\ by\ Team\ A\ against\ Team\ B =$$
$$\boldsymbol{Strength\ in\ Attack\ of\ Team\ A} \times \boldsymbol{Weakness\ in\ Defense\ of\ Team\ B}$$

To calculate the strength and weaknesses in attack and defense, one must look at the mean number of goals scored and conceded per game. Teams with a higher than average number of goals scored per game have a higher attacking 'strength' and teams with higher than average number of goals conceded have a higher defensive 'weakness'. This technique is demonstrated below.

The formula for attack strength is:

$$\frac{Average\ Goals\ Scored\ by\ Team\ A - Average\ Goals\ Scored\ in\ Tournament}{Average\ Goals\ Scored\ in\ Tournament} + 1$$

It is essentially an indicator how much better the team has performed than the average, goal-scoring wise. So, a score of 1.40 would indicate that the team has scored 40% more goals than average.

The formula for defensive weakness is:

$$\frac{Average\ Goals\ Conceded\ by\ Team\ A}{Average\ Goals\ Conceded\ in\ Tournament}$$

This is an indicator of how much worse a team has performed compared to the average, goal-conceding wise. So, a score of 0.736 would indicate that the team has conceded 26.4% less goals than average.

Calculations

Using the statistics available on the UEFA website (uefa.com), the following table of values was obtained. The average goals scored in this tournament was 1.31 while the average goals conceded was 1.59:

Table 4.1: Attack and Defence Metrics for UEFA Champions League 2010/11 Semi-Finalists

Name of Team	Average Goals Scored per game	Average Goals Conceded	**Attack 'Strength'**	**Defensive 'Weakness'**
Schalke 04	1.83	1.17	$\frac{1.83-1.31}{1.31}+1=\mathbf{1.40}$	$\frac{1.17}{1.59}=\mathbf{0.736}$
Real Madrid	2.08	0.5	$\frac{2.08-1.31}{1.31}+1=\mathbf{1.59}$	$\frac{0.5}{1.59}=\mathbf{0.314}$
FC Barcelona	2.31	0.69	$\frac{2.31-1.31}{1.31}+1=\mathbf{1.76}$	$\frac{0.69}{1.59}=\mathbf{0.434}$
Manchester United	1.46	0.54	$\frac{1.46-1.31}{1.31}+1=\mathbf{1.11}$	$\frac{0.54}{1.59}=\mathbf{0.340}$

The UEFA Champions League has 4 stages – the qualification stage, the group stage, the knockout stage and the final. The data for the calculations was obtained from the group stage onwards. This data included 32 teams.

A similar table was constructed for the semi-finalists of the UEFA Champions League 2014/15

Table 4.2: Attack and Defence Metrics for UEFA Champions League 2014/15 Semi-Finalists

Name of Team	Average Goals Scored per Game	Average Goals Conceded	**Attack 'Strength'**	**Defensive 'Weakness'**
Bayern Munich	2.75	1.08	$\frac{2.75-1.33}{1.33}+1=\mathbf{2.07}$	$\frac{1.08}{1.58}=\mathbf{0.684}$
FC Barcelona	2.38	0.85	$\frac{2.38-1.33}{1.33}+1=\mathbf{1.79}$	$\frac{0.85}{1.58}=\mathbf{0.538}$
Juventus FC	1.31	0.77	$\frac{1.31-1.33}{1.33}+1=\mathbf{0.984}$	$\frac{0.77}{1.58}=\mathbf{0.487}$
Real Madrid	2	0.75	$\frac{2-1.33}{1.33}+1=\mathbf{1.50}$	$\frac{0.75}{1.58}=\mathbf{0.475}$

In this case, the average goals conceded throughout the tournament were 1.58 while the average goals scored were 1.33.

Calculating λ

The finals of the Champions League in 2010/11 were between FC Barcelona and Manchester United. Using the data above, a rough prediction of the outcome of the match can be made:

Goals Scored by FC Barcelona = Average Goals Scored per Match × Attack Strength of FC Barcelona × Defensive Weakness of Manchester United

$$\lambda_{Barça} = \boldsymbol{Goals\ Scored\ by\ FC\ Barcelona = 1.31 \times 1.76 \times 0.340 = 0.784}$$

Goals Scored by Manchester United = Average Goals Scored per Match × Attack Strength of Manchester United×Defensive Weakness of FC Barcelona

$$\lambda_{ManU} = \boldsymbol{Goals\ Scored\ by\ Manchester\ United = 1.31 \times 1.11 \times 0.434}$$
$$\boldsymbol{= 0.631}$$

If these λ are taken in isolation, it would seem as if the final score-line would be 0.784 – 0.631.

However, this can never be the case as the goals scored are integers. These values of λ are the 'expected' or the average values – if the game was played over and over again, this would be the average score. The distribution expresses the probability of a number of events occurring in a given time period, if the average rate of the occurrence is known and the events are independent. [16]

The formula for Poisson Distribution is as follows:

---

[16] Spiegelhalter, David, and Yin Lam Ng. "Understanding Uncertainty: Football Crazy."Understanding Uncertainty: Football Crazy, 5 Sept. 2009, plus.maths.org/content/understanding-uncertainty-football-crazy.

$$P(X; \lambda) = \frac{\lambda^X e^{-\lambda}}{X!} \ ;$$

Using this and a Graphic Display Calculator, we get the distribution below:

Table 4.3: Poisson Distribution for the Number of Goals Scored for FC Barcelona and Manchester United in UEFA Champions League Final 2010/11

Team Name	Zero Goals Scored (X = 0)	One Goal Scored (X = 1)	Two Goals Scored (X = 2)	Three Goals Scored (X = 3)	Four Goals Scored (X = 4)
FC Barcelona ($\lambda = 0.784$)	0.457	0.358	0.140	0.0367	0.00719
Manchester United ($\lambda = 0.631$)	0.532	0.336	0.106	0.0223	0.00351

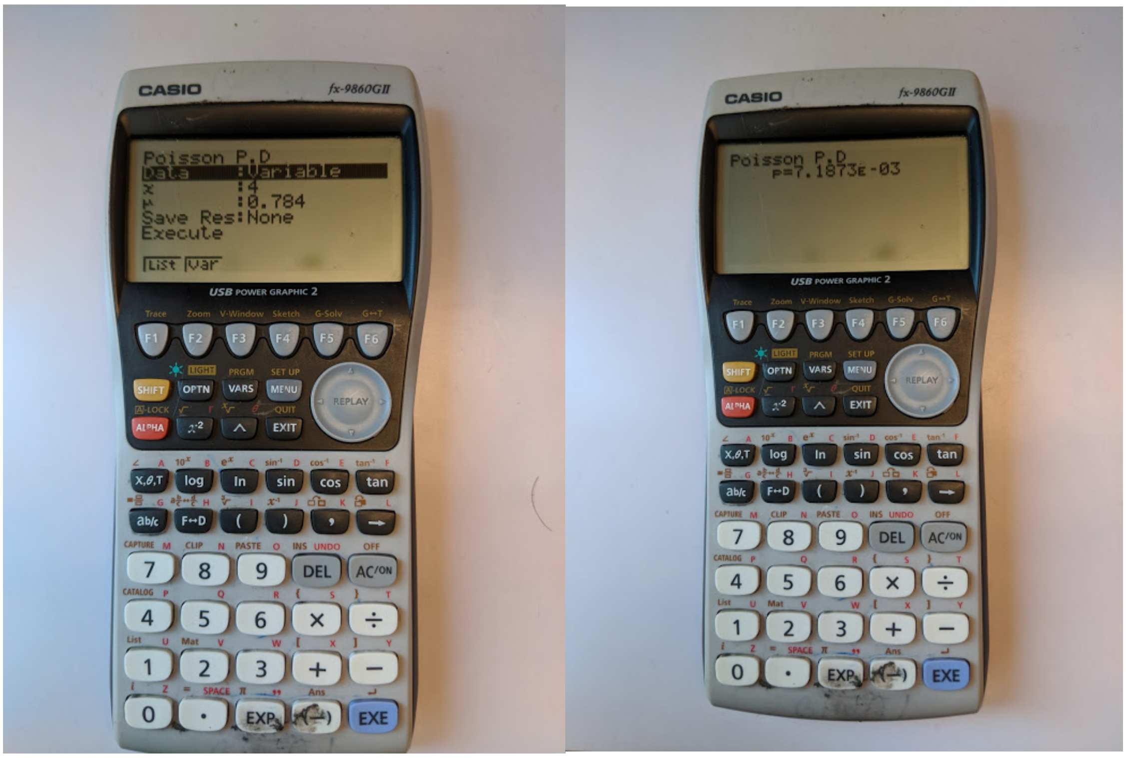


Image 4.1: Calculations for the Poisson Distribution

The final result of this match was 3-1 in the favour of FC Barcelona. Looking at the table, the probability of such a score-line is extremely low –

$$0.0367 \times 0.336 = 0.0123$$

The probability of this occurring is 1.23%. This is an extremely low probability for such a result. However, the limitations of this model explain this.

Table 4.3: Poisson Distribution for the Number of Goals Scored for FC Barcelona and Juventus FC in UEFA Champions League Final 2014/15

Team Name	**Zero Goals Scored (X = 0)**	**One Goal Scored (X = 1)**	**Two Goals Scored (X = 2)**	**Three Goals Scored (X = 3)**	**Four Goals Scored (X = 4)**
FC Barcelona ($\lambda = 1.16$)	0.313	0.364	0.211	0.0816	0.0237
Juventus FC ($\lambda = 0.704$)	0.495	0.348	0.123	0.0288	0.00506

The 2014/15 UEFA Champions League Final was held between FC Barcelona and Juventus FC. It ended 3-1 in Barcelona's favour again. The probability of that scoreline was 2.84%.

$$0.0816 \times 0.348 = 0.0284$$

In some cases, this model is accurate. In others, the result depends on a lot of external factors discussed in the limitations. In cases where the model does not fit the score, the network theory analysis can be used to qualitatively assess a team's performance, and where it went right or wrong.

Limitations of the Probabilistic Model

The limitations of this model are extremely evident. The model does not take into account external factors. The two matches analysed were finals of the tournaments and hence the motivation and pressure has not been taken into account. Furthermore, the value of λ depends on past results in the cup. However, that is not sufficient to predict the score between two teams because the opponents they faced in the past are different.

Conclusion

Whereas neither methods on their own provide a complete analysis of a team's strength and weaknesses, when both the methods are combined, they provide a comprehensive analysis. The probabilistic model gives a quantitative analysis of a team's defence and attack strength while the network theory model explains why that is so, qualitatively. These two must be used in unison for any data to make sense. Football coaches can predict the scoreline of a match in the future based on past events but can alter the actual scoreline by looking at the strengths and weaknesses of the opponent using network theory. This, along with a lot of other data analysis used currently in football, can be used during a match to provide a qualitative analysis to both coaches in a match. So, network theory can provide a stepping stone to indicating how the team's style of play is, but more work needs to be done in this field for it to ultimately be proved successful.

References

1. Baez, John. “Network Theory.” 2015. *University of California, Riverside* . <http://math.ucr.edu/home/baez/econ.pdf>.

2. Beardon, Alan. *Euler's Formula and Topology*. February 2011. 17 April 2017. <https://nrich.maths.org/1384>.

3. Brin, Sergey, and Lawrence Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *The Anatomy of a Search Engine*. N.p., n.d. Web. 24 Apr. 2017. <http://infolab.stanford.edu/~backrub/google.html>.

4. D. R. Brillinger. A potential function approach to the flow of play in soccer. *J. Quant. Analys*. Sports, 3(1), 2007.

5. “One Game to Play!” One Game to Play! | Understanding Uncertainty, 17 Jan. 2001, understandinguncertainty.org/node/228.

6. Peña, Javier López, Dr., and Hugo Touchette, Dr. "A Network Theory Analysis of Football Strategies." Thesis. 2012. *A Network Theory Analysis of Football Strategies*. 2 July 2012. Web. 17 Apr. 2017.

7. M. Hughes and I. Franks. Analysis of passing sequences, shots and goals in soccer. *J. Sports Sci.*, 23(5):509–14, 2005.

8. Maksim Tsvetovat Follow. “3 Centrality.” *LinkedIn SlideShare*, 19 Feb. 2009, www.slideshare.net/maksim2042/3-centrality.

9. Soramaki, Kimmo. “Applications of Network Theory in Finance and Production.” 17 May 2015. *Slide Share.* 17 April 2017. https://www.slideshare.net/soramaki/applications-of-network-theory-in-finance-and-production

10. Spiegelhalter, David, and Yin Lam Ng. “Understanding Uncertainty: Football Crazy.”Understanding Uncertainty: Football Crazy, 5 Sept. 2009, plus.maths.org/content/understanding-uncertainty-football-crazy.

11. Taïeb, Franck. “A Model of Football Games.” Science4All, 31 Jan. 2016, www.science4all.org/article/a-model-of-football-games/.