2018

Network Analysis of Passing Patterns in Handball

Margrethe Jebsen
Claremont McKenna College

Network Analysis of Passing Patterns in Handball

A Thesis Presented
by
Margrethe Helene Jebsen

To the Keck Science Department
Of Claremont McKenna, Scripps and Pitzer Colleges
In partial fulfillment of
The degree of Bachelor of Arts

Senior Thesis in Physics
April 23, 2018

Acknowledgments

I would like to thank Professor James Wells for being my thesis advisor and for providing me with guidance and feedback throughout this process. I would also like to thank Professor Adam Landsberg for being my secondary reader and giving me valuable advice.

Table of Contents:

Abstract

Network analysis is used in a number of different fields to study systems ranging from molecular interactions to social interactions between humans. With the increase of available data, it is not surprising that network analysis methods are now used to analyze human behavior and interactions in sports. Handball is a popular sport in Europe, and while it has been studied, it has not been studied with network methods. In this work, passing data was collected from two international competitions to study the passing behavior of the Norwegian women's handball team. The data were divided into multiple scenarios to try to explore how the team's and individual players' playing styles depended on whether the team was leading, trailing, or scoring goals. Star players were also compared with their replacements to see if there was any statistical evidence available as to why the star players may be unique. We found differences between the star players and their replacements, some differences between the team's passing behavior when leading and trailing and when scoring goals, and that the team appears to have a specific prescribed playing style for each position.

1 Introduction

1.1 Handball: Rules and Strategies

Handball is a team ball-sport that is one of the most popular sports in Norway. Each team has seven players on the court at any time: a goalkeeper, a mid-back, two side backs, two wings, and a line player. The goalkeeper guards the goal that is at the center of a semicircle with a 6-meter radius; see the solid line in Figure 1 below. No other player is permitted to step within this semicircle while the ball is in play. The defense lies along this semicircle while the offense lies a few meters further out, along the dotted semicircle shown in Figure 1. Unlike other popular ball sports such as soccer, every player is actively involved in both the offense and the defense, i.e. there are no “defenders” or “attackers”.

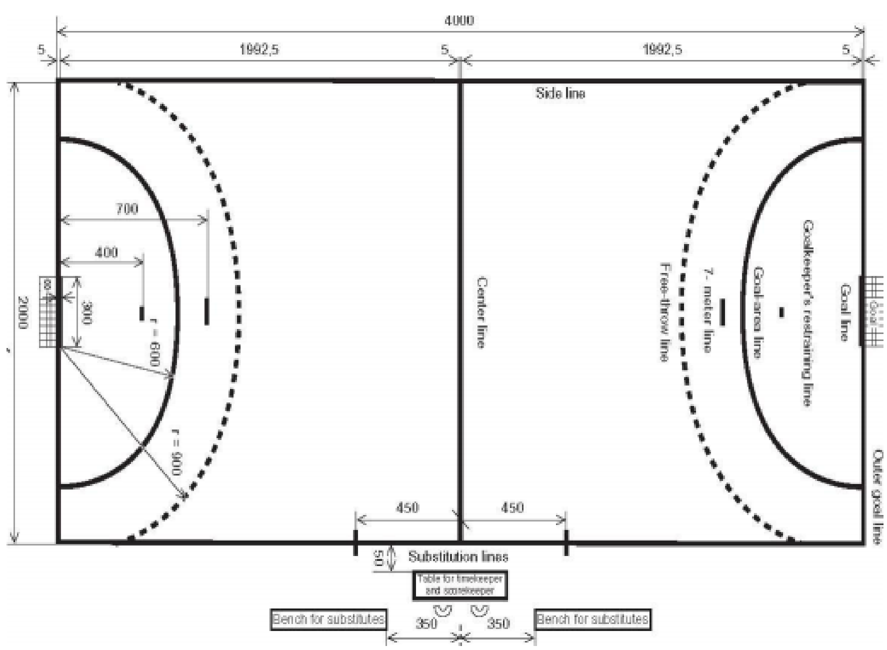

Figure 1: Diagram of a handball court taken from the IHF Rules Handbook [1]

The basic rules of handball are straightforward, so I will not go into much detail about them here. Professional handball is played in two 30-minute halves, where the teams switch sides during half time. To win the game, you must score more goals than the opponent before the time is up. Typically, a team will score between 20 and 30 goals in a game. The average number of goals scored in the 2017 IHF World Women's Handball Championship was 26.1 goals [2]. The game begins at the middle of the court, at the center line depicted in Figure 1. Every member of the team that is in possession of the ball must be on or behind the center line before the referee allows the game to begin. This also happens during a restart after either team scores a goal.

At the start of an offensive play, the mid-back and side backs will generally pass the ball back and forth between them a few times while planning their attack before beginning to pass the ball to the wings. Each individual player may only retain possession of the ball as long as she is dribbling; otherwise she gets three steps or up to three seconds to pass the ball to a teammate or shoot. If she keeps possession longer than this, her team will lose possession of the ball due to traveling. During an offensive play, the line player lies amongst the defense and attempts to secure openings for her teammates or free herself up enough to receive a pass. If the line player does receive a pass while on the line, she will have a very good chance of scoring a goal or securing either a free throw or a penalty. Free throws are awarded to the attacking team if an attacker is fouled by a defender, if a defender steps into the goalkeeper's box, or for other less common infringements. A penalty is awarded if a serious foul is made or if a goal-scoring opportunity is sabotaged due to a foul [1]. Free throws awarded due to fouls occurring within the dotted line concentric to the goalkeeper's box are taken from the 9-meter line. Free throws awarded from fouls made elsewhere are taken from wherever the foul occurred. A penalty is taken from a line 7 meters from the goal; see the short solid line parallel to the top of the goalkeeper's box in Figure 1. Other than the player chosen to take the penalty shot and the defending goalkeeper, no other player, defending or opposing, is permitted to be within the dotted 9-meter semicircle depicted in Figure 1. For a comprehensive review of the rules in handball, see the IHF Rules Handbook listed in the references [1].

Some strategies of handball involve switching positions while in offense. Usually this involves the backs and the wings. The purpose of this strategy may be to confuse or to distract the defense, or to draw the defense to one side of the court or the other in order to create a goal-scoring opportunity. Goal-scoring opportunities include back players finding gaps in the defense, line players opening themselves up to passes, or wings given enough space on either side to jump in at a wide enough angle to score. Therefore, studying passing patterns during offensive plays may give insight about a team's playing style or an individual player's role within a team.

1.2 Background

1.2.1 Network Analysis in Sports

With the increasing amount of data and statistics available about sports, it is not surprising that individuals are using data to analyze team and player behavior in order to, presumably, find the key to managing a successful team. Bekkers and Dabadghao performed extensive research on flow motifs in soccer over four seasons to analyze events such as the transfer of Claudio Bravo to Manchester City [3]. Their analysis was done through the static approach, in which players are only tracked if involved in a ball event. The dynamic approach tracks all players and the ball during the game in real-time. Peña and Touchette studied strategies in soccer using data from the 2010 World Cup [4]. They used network graphs to visualize a team's strategies and potential weaknesses, and also used centrality measures to determine the importance of individual players. Peña and Navarro studied passing behavior of individual players in order to create a tool to measure similarity between players [5].

Clearly, a lot of work has been done in analyzing passing behavior in soccer. While handball is a team ball sport as well, it is very different from soccer. In handball, every player is actively involved in both the offense and the defense, and goal-scoring opportunities may occur several times during one offensive play. Therefore, it is important to explore what kind of research has been done in handball analysis specifically.

1.2.2 Analysis in Handball

While handball is one of the most popular sports in Europe, it is virtually unknown in the United States. Prieto conducted a review of existing scientific literature surrounding handball analysis [6]. His results showed that the field is currently lacking studies focused on women's handball, international competitions, and analyzing defensive behavior. Schrapf and Tilp studied offensive passing behavior in order to identify patterns using clustering techniques [7]. Their study was limited to six matches involving eight teams participating in the European Handball Federation (EHF) European Men's Championship. The same researchers also published a paper on analyzing defensive behavior in handball [8]. They accomplished this by noting the position of the defensive players during goal attempts made by the offensive team. Hassan presents an approach for predicting outcomes of handball matches through the use of modular forward neural networks [9]. The predictions were based off of game details such as number of shots from each position, number of steals and fast break goals, and whether the team won or lost.

The present work is unique because it focuses on women's international handball and because it combines techniques such as flow motifs and comparing individual players, as seen in the above soccer analyses, and applies them to handball.

1.3 Data

1.3.1 Games and Team Selected

This thesis will focus on passes made by the Norwegian women's handball team in the 2017 International Handball Federation (IHF) World Women's Handball Championship and the 2017 Møbelringen Cup.

The IHF Championship is arranged every other year by the IHF and is hosted by one of the competing countries; in 2017 the championship was hosted by Germany. The format of the championship is a single round robin group stage with four groups each comprised of six teams. The group stage is followed by a single-elimination knockout stage involving 16 teams.

The Møbelringen Cup is arranged annually by the Norwegian Handball Federation (NHF) and is sponsored by the Norwegian furniture company Møbelringen. The cup involves just four international teams; Norway and three other countries, and is formatted as a single round robin tournament. In 2017, the other three teams were South Korea, Russia and Hungary. This cup takes place right before the IHF World Women's Handball Championship or the European Women's Handball Championship, depending on the year. Since its inception in 2001, Norway has been the winner 11 times. While this cup is not nearly as prestigious as the IHF World Women's Handball Championship, I chose to include it because it would add more data to my research without changing too many factors such as the players involved, the team's coach, strategies used, and so on.

Historically, the Norwegian women's team has consistently been a favorite. They have won an Olympic medal six times, two of which were gold medals in 2008 and 2012. In the European Women's Handball Championship, the Norwegian women's team has received 7 gold medals and 11 total medals. Since the championship's founding in 1994, the Norwegians have received a medal every year except for in 2000, where they placed 6th. In the IHF World Women's Handball Championship, they have received 10 medals, more than any other country. Three of those medals have been gold and in the 2017 edition, which is studied here, they were the runners up. Thorir Hergeirsson has been the team's head coach since 2009, and was the assistant coach during 2001-2009. Since he became head coach, the Norwegian women's team has missed just one medal in the three above championships: they placed 5th in the 2013 IHF World Championship. Clearly, this is a successful team on an international level.

The players involved in these two tournaments that are often viewed as the star players on the Norwegian women's handball team are Stine Oftedal, Nora Mørk, Veronica Kristiansen, and Katrine Lunde. Oftedal, player number 10, is the current team-captain and was voted MVP of the 2017 IHF World Championships. Mørk, player number 9, has been top-scorer in three international competitions, including in the 2017 World Championships with 66 goals in 8 matches. Kristiansen, player number 4, was a starter for every match observed and had more playing time than any other player. Lunde, player number 16, has been on the All-Star Team five times and tied for top goalkeeper in the 2017 World Championships with a save efficiency of 42% [10] [11] [12].

We will attempt to find out why these players are star players by comparing them to their replacements. "Replacement" in this case simply means a player that plays in the same position as the star player but is not yet considered a star herself. In the matches observed, Mørk's replacement was Amanda Kurtovic, player number 22; Oftedal's was Emilie Christensen, player number 18; and Kristiansen's was Emilie Arntzen, player number 3. We will not compare Lunde to the other goalkeepers due to passing data not being as relevant in determining key behavior and/or skills of a goalkeeper.

1.3.2 Methodology

I collected the passing data myself by watching recordings of the games found in TV2 Sumo’s online archive. This allowed me to pause and re-watch the games as often as needed to record the passes as accurately as possible. However, these recordings came with a disadvantage as well. Often some passes were excluded from the recording in favor of replaying a goal or foul. This was primarily a concern for the very beginning of an offensive play, so was not detrimental to analyzing offensive passing behavior overall. As mentioned in an earlier section, the beginning of an attack usually involves passes between the back players as they plan the attack, so it is not as consequential as the rest of the attack. The plays in which this was an issue were excluded from analysis.

A possession could begin due to a number of events such as the opponent scoring a goal or losing possession of the ball. Figure 2 shows a breakdown of the ways in which possessions started in all the matches observed. Clearly, most possessions began after the opposing team scored a goal. A significant fraction of plays began with free throws and with shots saved by the goalkeeper while in defense.

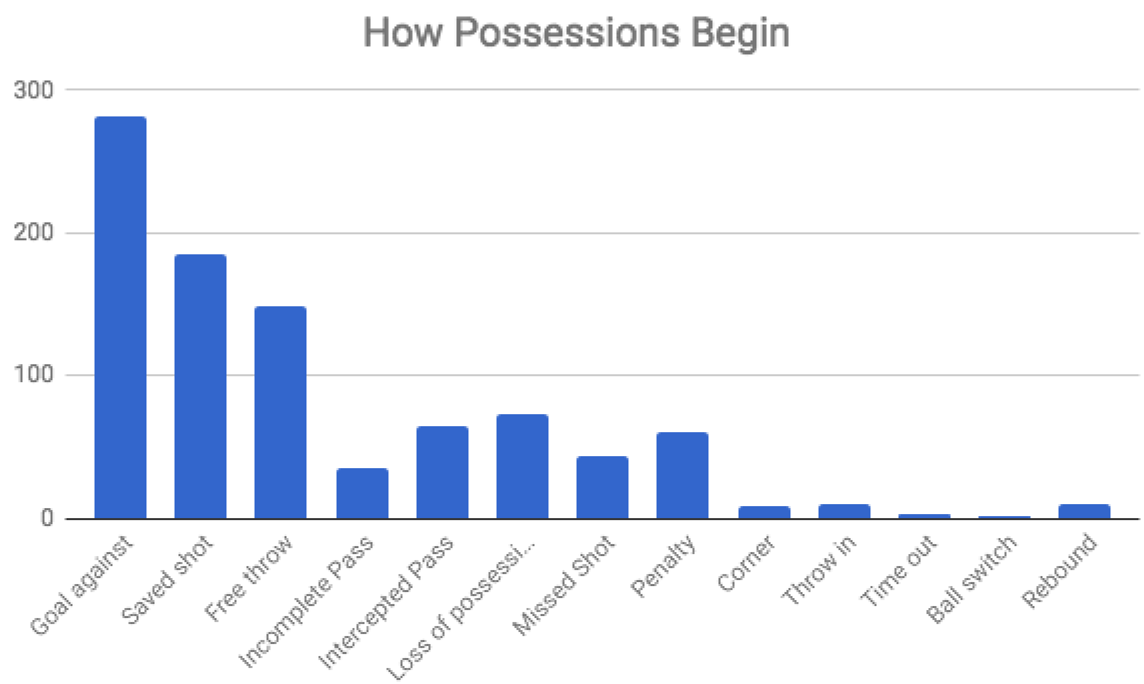

Figure 2: Graph showing the distribution of possible ways for a possession to begin.

Similarly, the Figure 3 shows the breakdown of ways in which the possessions ended. As with the start of a possession, is it clear that possessions end most often due to a goal being scored. Of the 928 possessions recorded in the twelve matches observed, over 350 of them ended with the Norwegian women's team scoring a goal either in open play or as a result of a breakaway, which is a high success rate.

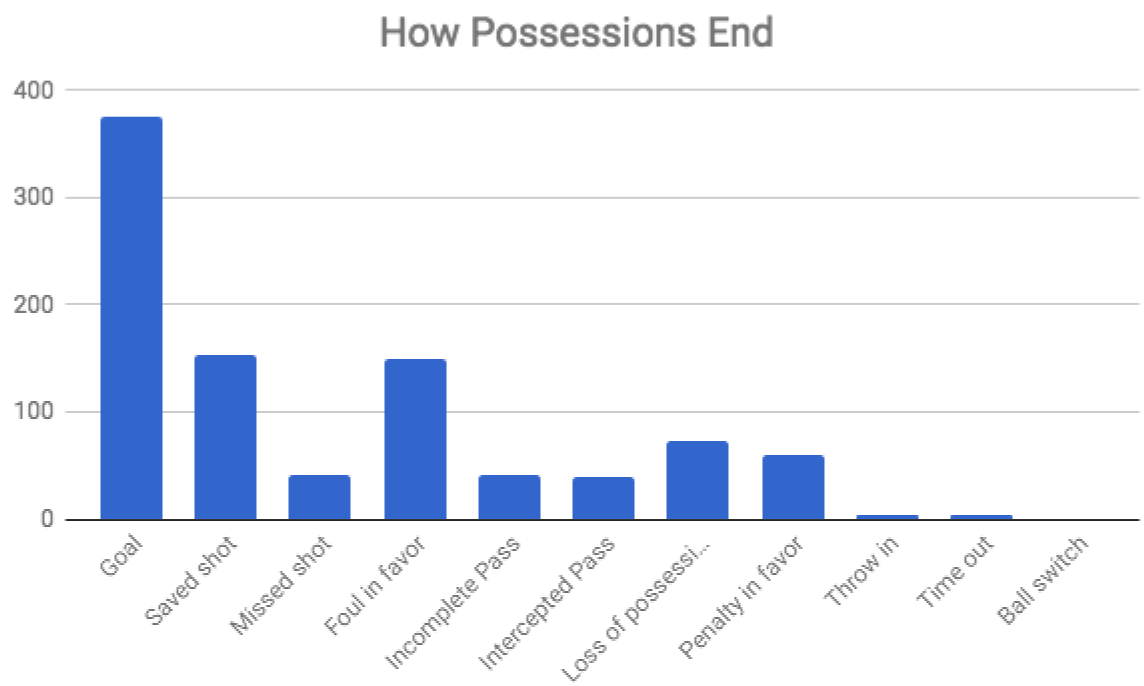

Figure 3: Graph showing the distribution of possible ways for a possession to end.

2 Definitions

2.1 Network Measures

Network measures are used to measure the structure of a network and to measure a given node's importance to the network. Some measures that reflect the structure of the network are the number of edges, number of vertices, number of components, density, and diameter. These will be defined below.

2.1.1 Number of Edges and Vertices

The most fundamental components of a network are the vertices and edges. The vertices represent the individuals or groupings of individuals in the system, while the edges represent the connections between them. In the handball network, the vertices represent individual players and the edges represent directed passes between them. Thus, the number of vertices in the network will simply be equal to the total number of players. The number of edges will be the number of realized connections, i.e. passes, between different players. In some networks, the edges are weighted to reflect the strength of the connection between two nodes. In the handball network, a greater number of passes between two players will result in a more weighted edge between them. On the other hand, if two players never pass to each other, as is the case with two goalkeepers, there will not be an edge connecting their respective vertices in the network.

2.1.2 Density

The density of a graph is the frequency of realized edges over all possible edges [13]. In other words, density is a measure of the total number of realized edges in the network divided by the total number of potential edges. In a directed graph G with V vertices and E edges,

$$den(G) = \frac{E}{V \cdot (V - 1)}$$

where $den(G)$ is the density of the graph [13]. In an undirected graph, we would divide the denominator through by a factor of two to avoid double-counting the total number of possible edges.

2.1.3 Diameter

The diameter of a network is equal to the longest path with no extra steps existing between any two vertices in the network [13]. In other words, diameter represents the shortest path between the most distant vertices, and thus gives us a measure of the size of the network.

2.1.4 Number of Components

The number of components in a graph represents the number of subgraphs that are connected. A connected graph is one in which every vertex is reachable from every other vertex. In the case of the handball network, there is just one component. This means that every player is connected to every other player either directly or indirectly through a passing sequence. Therefore, the handball network is connected and the number of components is trivial.

2.2 Centrality Metrics

Centrality metrics are used to try and determine the relative importance or prominence of vertices in a network. There are a number of different centrality measures used in network analysis, three types will be covered here: vertex degree, closeness centrality, and betweenness centrality.

2.2.1 Vertex Degree

Vertex degree is a measure of the number of edges incident on a given vertex [13]. Consider the undirected toy network shown in Figure 4. The vertex labelled "A" has three incident edges, so it has vertex degree equal to 3. Similarly, vertices B and D also have vertex degree equal to 3. Vertex E only has one edge incident on it, while vertex C has two edges incident on it. Thus, they have vertex degree equal to 1 and 2 respectively. The network shown in Figure 4 is rather unusual because it is very well connected compared to most real networks.

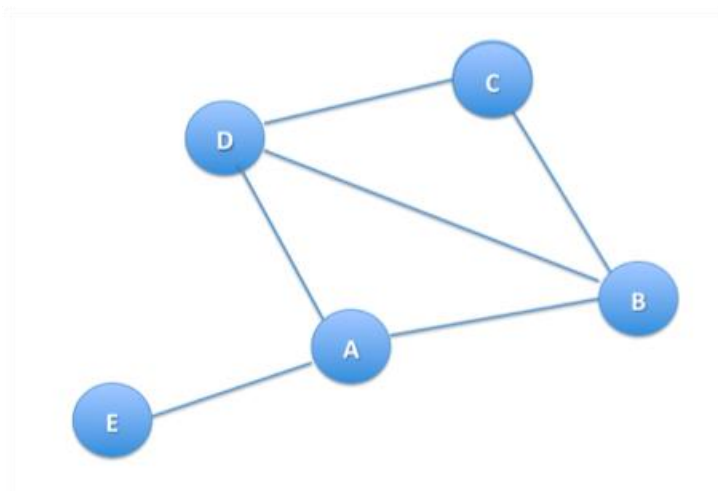

Figure 4: Toy model of a network of five nodes with undirected and unweighted edges.

In a directed graph, it is also possible to think about in- and out-degree of vertices. In the handball network, the in-degree of a vertex is the number of passes from different players received by the player represented by that vertex. Out-degree is the number of different players she passes to. In-degree and out-degree may be used to identify the position of players. For example, we would expect goalkeepers to have higher out-degree than in-degree because it is less common to pass to a goalkeeper than to receive a pass from a goalkeeper.

2.2.2 Closeness Centrality

Closeness centrality measures the extent to which a vertex is "close" to other vertices. The closeness of a vertex is the inverse of the sum of the shortest path distances between it and all the other vertices:

$$C_C(v) = \frac{1}{\sum dist(u,v)}$$

where $dist(u, v)$ is the shortest path distance between the vertex in question and some other vertex [13]. Consider the network shown in Figure 4, let us find the closeness centrality for each vertex. Vertex A is one edge away from vertices B, D and E and two edges away from vertex C. Thus, the sum of the distances between vertex A and the other vertices is 5. Therefore, vertex A has closeness centrality equal to $\frac{1}{5}$ or 0.2. Similarly, vertices B and D also have closeness centrality equal to $\frac{1}{5}$ which means that vertices A, B and D are as "important" to the network with respect to this particular centrality measure. Finally, vertices C and E have closeness centrality equal to $\frac{1}{7}$ and $\frac{1}{8}$ respectively, making them less "close" to the other vertices in the network and thus less central to the network.

In the handball network, we must be careful with how we define distance. The edges in the handball network represent the number of passes between two players, not the distance between them. The larger the number of passes between two players is, the "closer" they ought to be. Thus, in order to use the above equation for closeness centrality, we must take the inverse of each dist(u,v) term to ensure that a greater number of passes between two players translates to greater closeness.

2.2.3 Betweenness Centrality

Betweenness centrality measures the extent to which a vertex is "between" other pairs of vertices [13]. It is given by,

$$C_B(v) = \sum_{s \neq t \neq v} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where the numerator is the total number of shortest paths between the vertices s and t that pass through the vertex v, and the denominator is simply the total number of shortest paths between s and t [13]. Consider Figure 4 once more, we can calculate the betweenness centrality for each vertex by considering its involvement in connections between the other vertices.

Starting with vertex A, let us consider the pair B and C. There exists exactly one shortest path between B and C, but it does not pass through A. Thus, the betweenness centrality term for A is zero so far. The pairs (B, D) and (C, D) also have exactly one shortest path, neither of which passes through A, so A still has zero betweenness. However, the pairs (B, E), (C, E) and (D, E) all have shortest paths that pass through A. Thus, vertex A has betweenness equal to 3.

We repeat this process for the other five vertices and obtain the following table. Clearly, vertex A has the highest betweenness centrality due to the fact that it is the only vertex connected to E. On the other hand, E and C have zero betweenness because the remaining four vertices are connected to one another without the help of either E or C. In this way, betweenness can help determine which individuals are important to the integrity or connectedness of the network and which individuals are not.

Vertex	A	B	C	D	E
Betweenness	3	1	0	1	0

Figure 5: Table showing betweenness centrality for the toy model depicted in Figure 4.

2.3 Three Pass Groups

Three pass groups are the different possible forms of three pass sequences involving two or more players. Figure 6 shows a chain of passes between players A, B, C and D. It contains the passing sequences ABAB, BABC and ABCD. The player in question is labeled as “A”, so the ABCD three pass group represents a passing sequence in which the player in question passed it to her teammate “B”, who passed it to “C”, who then passed it to “D”.

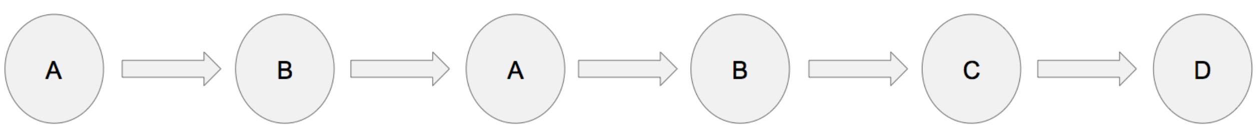

Figure 6: Chain of passes containing the passing sequences ABAB, BABC and ABCD.

Passing groups let us measure playing styles for individual players, thus giving us a tool with which to compare different players’ playing styles. If a player tends to be involved in ABCD sequences, she may not be the type of player that shoots or is awarded penalties. Players with a tendency to be involved in one-two combinations such as ABAB and BABC are likely involved in creating goal-scoring opportunities, play an assisting role in goal-scoring, or play a central position such as mid-back. We can use three pass groups to compare the star players with their replacements to see if they exhibit similar or different playing styles.

3 Results

3.1 Network Measures

3.1.1 All Possessions

This scenario represents every possession belonging to the Norwegian women's handball team in the twelve matches observed. There were 928 possessions, which is an average of 77.33 possessions per game. As we saw earlier, over 350 of these possessions ended in a goal, while the rest ended primarily in missed or saved shots, fouls in favor, or due to a loss of possession. Using the network measures defined in the previous section, we will characterize the handball network for this scenario.

The number of vertices for the handball network representing all possessions is trivial as it is equal to the total number of players involved in both championships, 17. The number of directed edges in the network is 184. We compare the number of realized edges to the number of potential edges to compute the density of the network. In this case, the edges are directed, so the total number of potential edges is equal to the number of vertices V multiplied by one less than the number of vertices, V-1, multiplied by 2. Thus, the number of potential edges is 544 and the density of the network is 0.338.

The diameter of the handball network for all possessions is 2. This means that the most distant vertices in the network are only two edges away from one another. This suggests that the network is very well connected for the all possessions scenario. Figure 7 shows the handball network in this scenario. The color of the nodes corresponds to the position of each player while the size of the nodes is determined by the player's betweenness centrality, which we will explore in more detail in section 3.2.

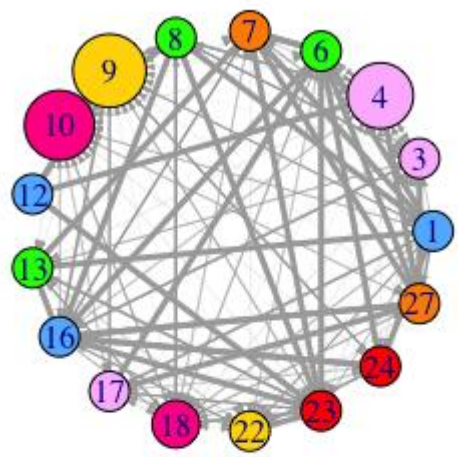

Figure 7: Handball network for the all possessions scenario.

3.1.2 Leading and Trailing

Due to the Norwegian women's handball team success during both tournaments, the amount of data available for trailing possessions is significantly less than for leading possessions. The team was in the lead for 684 out of 928 total possessions, but were only trailing for 142 possessions. In other words, they were leading 73.71% of the time and trailing 15.30% of the time. The remaining possessions occurred while the team was tied. Clearly, there is a large discrepancy between the number of trailing possessions and leading possession. This must be taken into account when making comparisons between the two scenarios.

The number of vertices in both scenarios is 17, as in the all possessions scenario. However, the number of realized edges in the leading scenario is 175 compared to 87 in the trailing scenario. Therefore, the handball network in the leading scenario has a density value of 0.322, which is almost as high as for all possessions, while the trailing scenario has a network density value of 0.160. However, because density is simply a measure of how many potential edges were realized, the sheer number of possessions included will have a large effect because there is more opportunity for players to be more connected.

The diameter of the network in the leading scenario is 2, which is the same as for the all possessions scenario. Thus, despite the team being in the lead, they do not exhibit higher connectedness than in general with respect to this particular network measure. The diameter of the network for the trailing scenario is 3, which is more than for the two previous scenarios. This suggests that the team was less connected when trailing than otherwise.

Figures 8 and 9 below show the handball network for the leading and trailing scenarios respectively. Notice how the leading network appears to have higher connectivity than the trailing network.

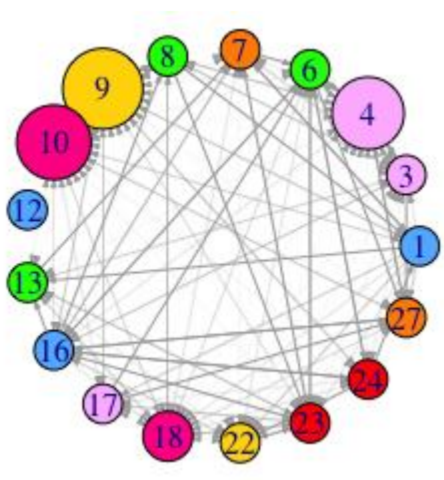

Figure 8: Handball network for leading possessions.

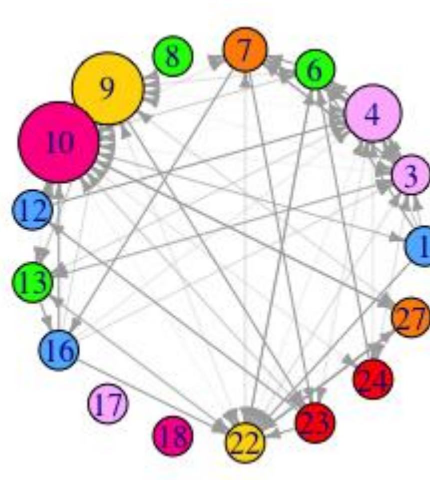

Figure 9: Handball network for trailing possessions.

3.1.3 Goals and Non-Goals

The scenario represents all possessions which ended in a goal. There were 376 possessions that ended in a goal, which is 40.52% of the total number of possessions. In the goals scenario, the number of directed edges is 154. In its complement, i.e. the set of possessions that did not end in a goal, the number of directed edges is 162. Thus, the two scenarios have network density equal to 0.283 and 0.298 respectively. The diameter is 3 for both scenarios. These results suggest that there is not that much of a difference between the two scenarios with respect to these network measures.

Figure 10 shows the handball network for the goals scenario. Notice that it looks more connected than the networks for the previous two scenarios. This may be because every player was involved in a possession than ended in a goal.

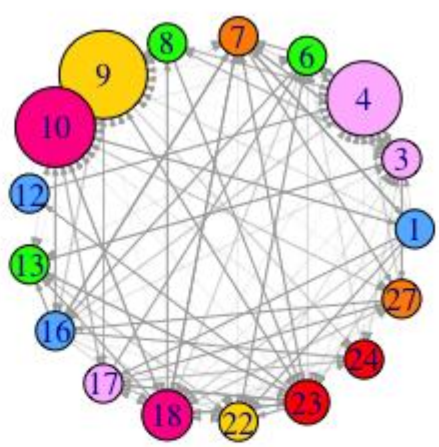

Figure 10: Handball network for possessions ending in goals.

3.2 Centrality Metrics

3.2.1 All Possessions

3.2.1.1 Vertex Degree

Vertex degree is simply the number of edges incident on a given vertex. Figure 11 shows normalized degree, in-degree and out-degree for each individual player in the all possessions scenario. Vertex degree was normalized by dividing the number of incident edges on each vertex by the total number of possible edges, 32. In- and out-degree were normalized by dividing through by the total number of possible "in" and "out" edges, 16.

It is clear from Figure 11 that the "star players" Mørk, player 9, and Oftedal, player 10, have the highest vertex degree values on the team. Star player Kristiansen, player 4, is tied for third-highest vertex degree with Oftedal's replacement, player 18, but has higher vertex degree than her own replacement, player 3. Player 22, Mørk's replacement, also has quite high vertex degree. This suggests that while the star players are "important" to the network in relation to this particular centrality metric, their replacements are about as important.

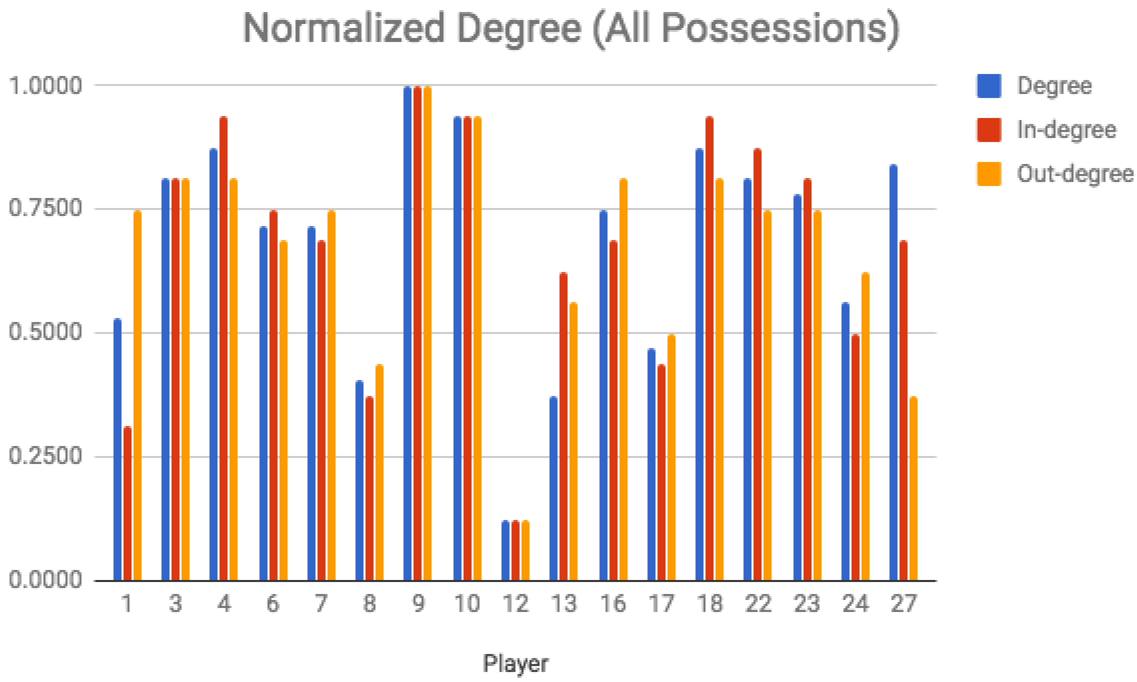

Figure 11: Graph of normalized degree, in-degree and out-degree for each player in all possessions.

3.2.1.2 Closeness Centrality

As defined earlier, closeness centrality measures the extent to which a vertex is "close" to other vertices [13]. Figure 12 shows normalized closeness for each player for the all possessions scenario. It is clear that the players with the highest closeness centrality in this scenario are players 9, 10, 4 and, to some extent, players 3 and 22. The former three players are the star players, while 3 and 22 are the replacements for players 4 and 9 respectively. This suggests that the star players tend to have higher closeness centrality than their replacements, which means that they are more "important" to the network than their replacements.

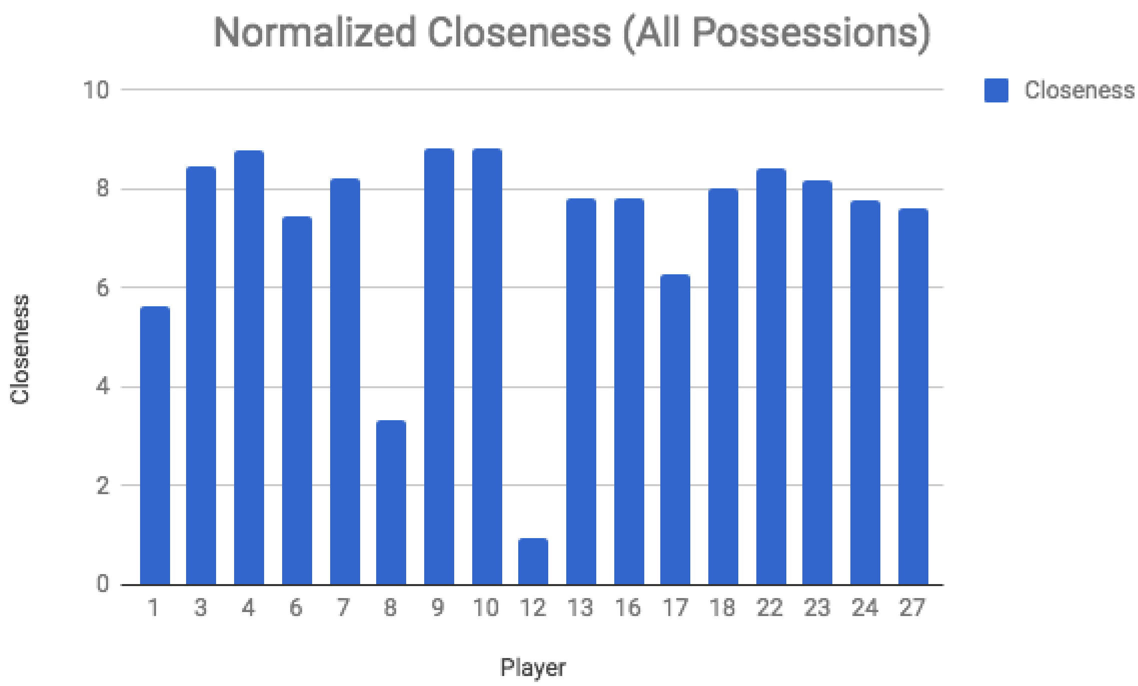

Figure 12: Graph of normalized closeness centrality for all possessions.

3.2.1.3 Betweenness Centrality

Betweenness centrality is a measure of the number of connections that must pass through a given vertex. A player in the handball network will have high betweenness if she is important to the connectivity of the network. Figure 13 shows normalized betweenness for each player for this scenario. Only players 4, 9, 10 and 18 have non-zero betweenness in this scenario. Players 10 and 18 are the mid-backs, with 10 being the star player and 18 being her replacement. Player 10 has high betweenness compared with her replacement. Players 4 and 9 are side backs, which is a less central position than the mid-back, but they still have very high betweenness. Their replacements, players 3 and 22 respectively, have zero betweenness in this scenario. Thus, the star players have significantly higher betweenness centrality than their replacements. This means that the star players are more prominent or important to the connectedness of the network than their replacements.

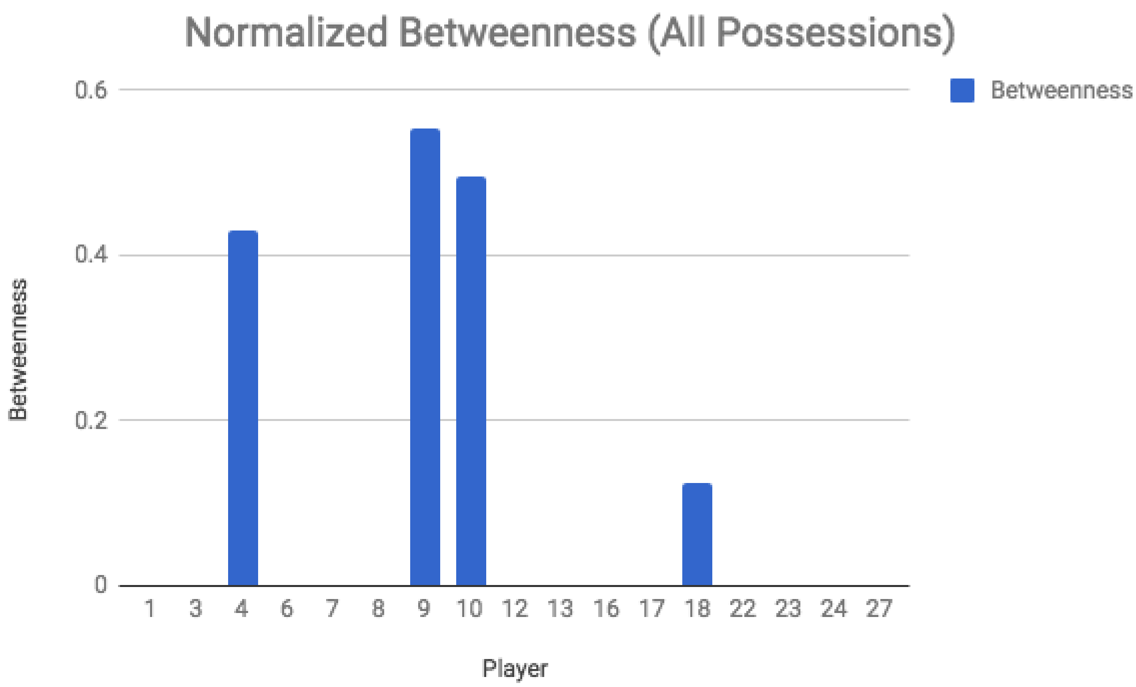

Figure 13: Graph showing normalized betweenness for each player for all possessions.

3.2.2 Leading and Trailing Possessions

3.2.2.1 Vertex Degree

For the leading and trailing scenarios, vertex degree and in- and out-degree may be computed to compare the vertex centrality of each player and see how it changes in the two scenarios.

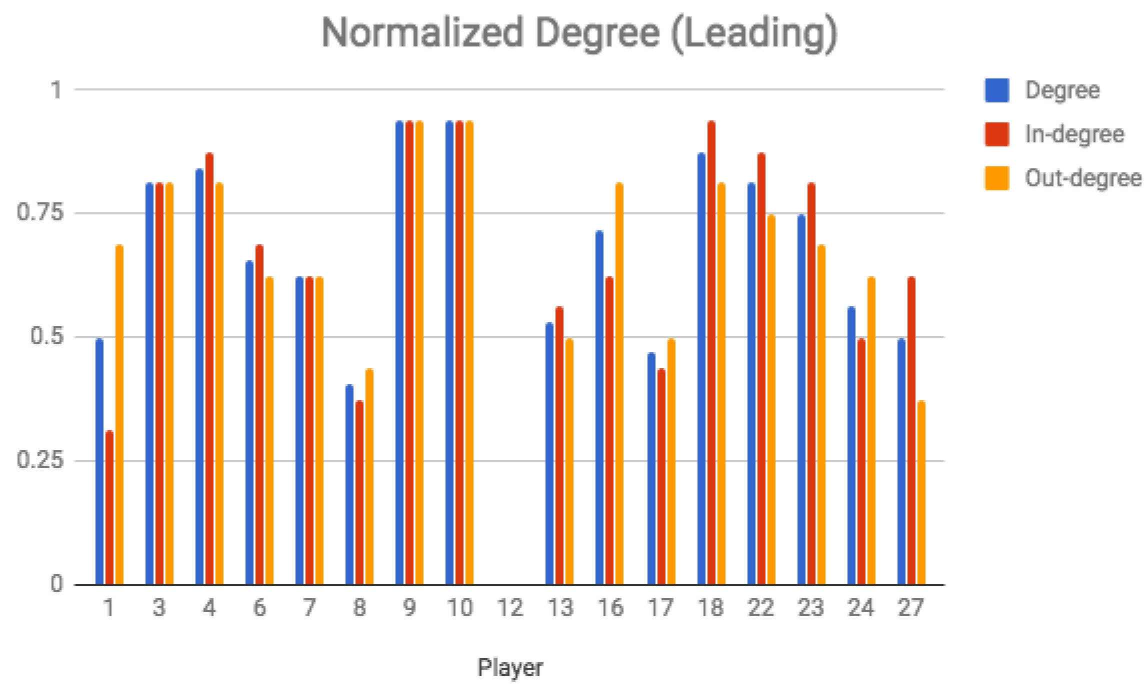

Figure 14: Normalized vertex degree and normalized in- and out-degree for each player for the leading scenario.

Figure 14 shows normalized vertex degree, and normalized in- and out-degree for the leading scenario. Note that player 12 has zero vertex degree, this is because she never played during any leading possessions. This is why there are no players with normalized vertex degree equal to 1. The trend tells us that players 9 and 10 have the highest vertex degree and it tells us that their in-degree is equal to their out-degree. Players 18, 4, 22 and 3 also have high vertex degree, but only player 3 has in-degree equal to out-degree. Players 4, 18 and 22 received passes from a greater number of different players than they themselves passed to, suggesting that they often ended possessions by shooting or by being rewarded a free throw, or that they didn't pass to defensive players such as goalkeepers.

Figure 15 shows normalized vertex degree and normalized in- and out-degree for the trailing scenario. In this scenario, the players 8, 17 and 18 never played, which may explain the significant decrease in vertex degree for players since the leading scenario.

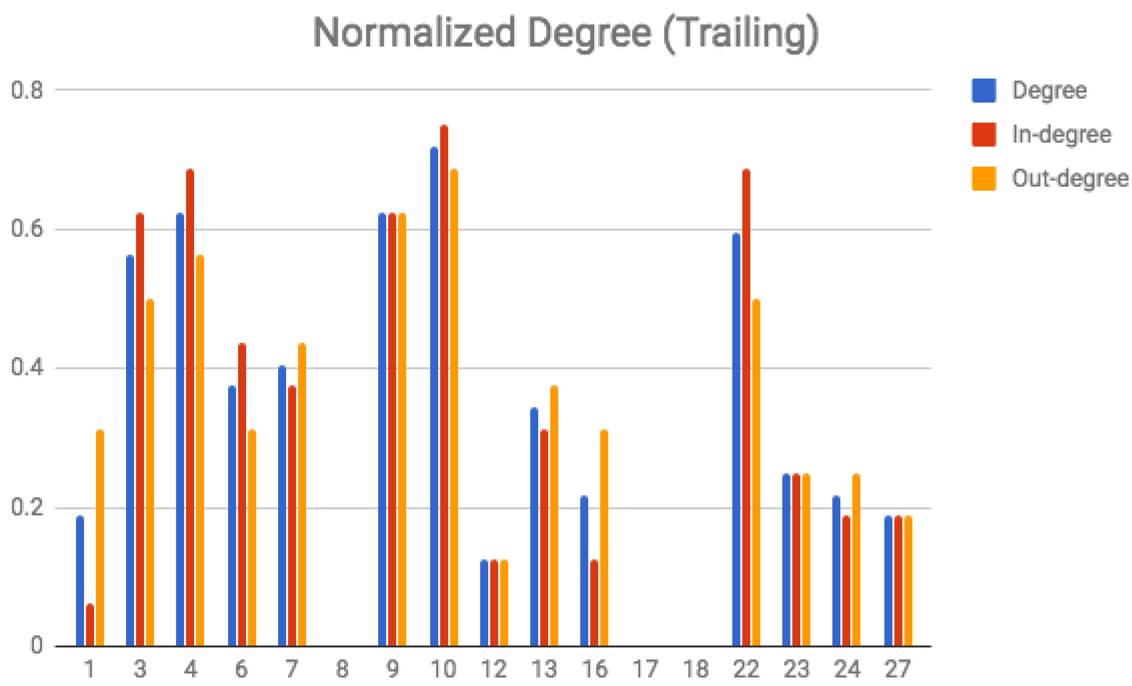

Figure 15: Graph showing normalized vertex degree and normalized in- and out-degree for each player in the trailing scenario.

It is possible that the discrepancies between vertex degree in the leading and trailing scenarios is due to missing players. Therefore, I decided to construct a graph in which vertex degree was normalized by dividing through by the number of possible directed edges for each scenario. In other words, since players 8, 17 and 18 did not play in the trailing scenario, the normalizing factor was not twice 16, but twice 13. Similarly, player 12 did not play during the leading scenario, so the normalizing factor was twice 15 instead of twice 16. This method of normalization produced the graph shown in Figure 16.

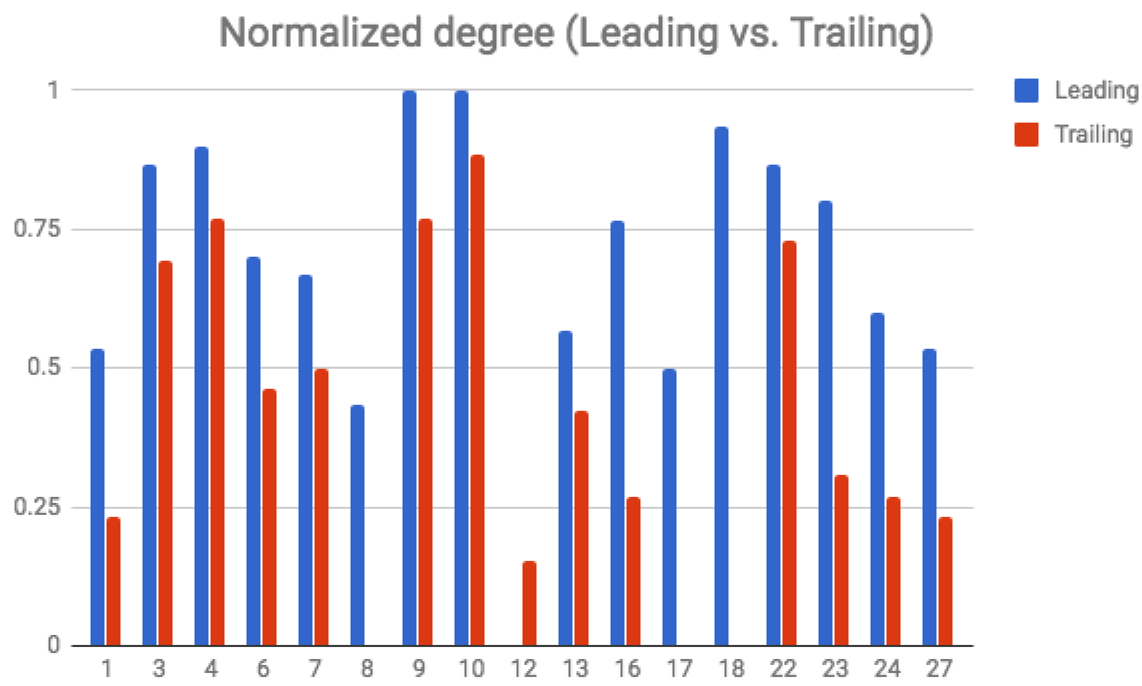

Figure 16: Graph comparing normalized vertex degree of the leading and the trailing scenario. Players that did not play during either of the two scenarios were excluded from the normalization factor.

Despite removing the possibility that missing vertices resulted in discrepancies between vertex degree in the two scenarios, Figure 16 shows that vertex degree in the leading scenario tends to be higher. The only exception is player 12 who did not play in the leading scenario. This result suggests that the team played more connectedly in the leading scenario than in the trailing scenario. It may also suggest that replacement players played fewer minutes when trailing, so there was less opportunity for them to pass to everyone, because the start players had more playing time.

3.2.2.2 Closeness Centrality

Comparing closeness centrality in the leading and trailing scenarios is tricky due to the difference in the sheer number of possessions occurring in the two scenarios. If we just compute closeness for each scenario without considering the fact that the number of possessions in the leading scenario is 4.8 times that of the trailing scenario, we obtain the graph shown in Figure 17.

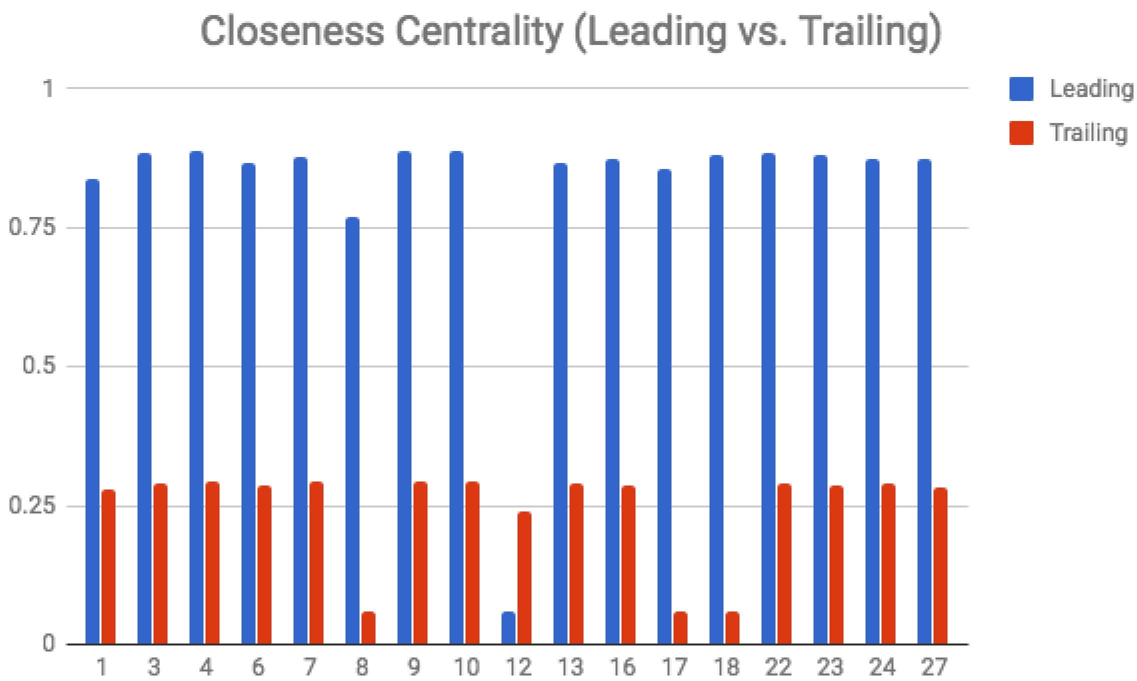

Figure 17: Graph comparing closeness centrality for each player in the leading and trailing scenarios, without taking into account the relative number of possessions in the two scenarios.

While Figure 17 allows us to make conclusions about the general trends existing in the two scenarios, it is difficult to make worthwhile comparisons between the two. In both scenarios, it is clear that the players have very similar closeness centrality, without any real indication of the star players differing from their replacements.

In order to make sensible comparisons between the two scenarios, we divide the adjacency matrices through by the total number of passes made in each scenario before computing closeness centrality. This gives the graph shown in Figure 18.

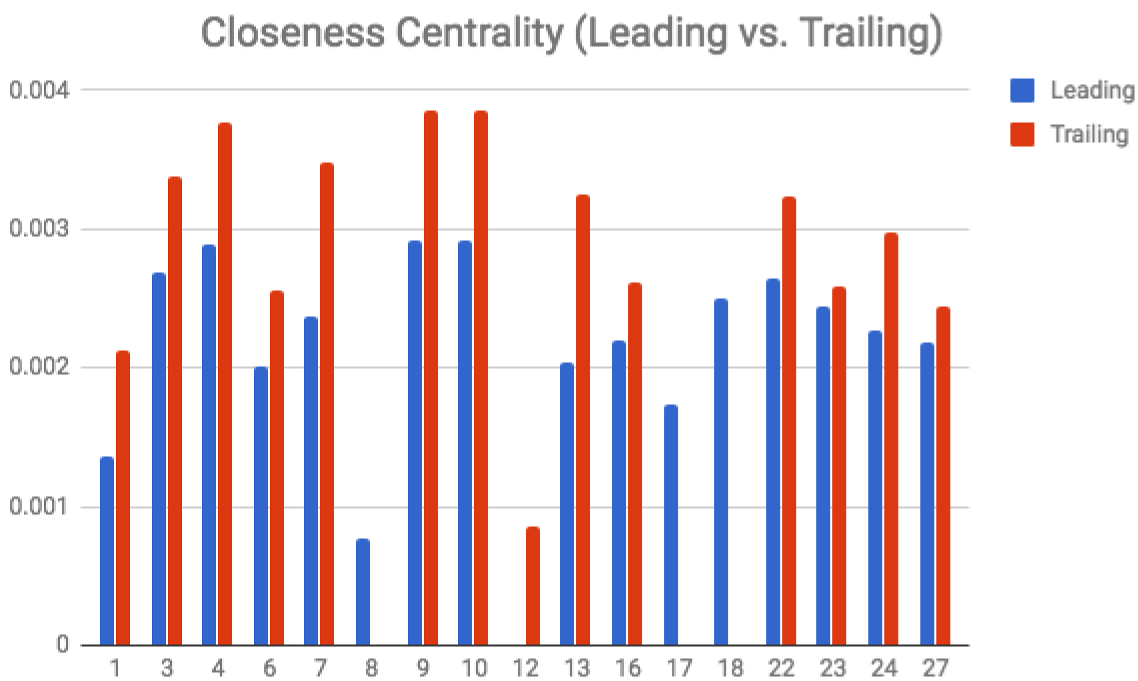

Figure 18: Graph comparing closeness centrality for the leading and trailing scenarios, taking into account difference in the number of possessions occurring during each scenario.

Figure 18 contradicts the result in the previous figure because it suggests that closeness centrality for each player in the trailing scenario tends to be higher than in the leading scenario. The general trend is the same as in the all possessions figure because it shows that in both scenarios players 9, 10, 4 and to some extent players 3 and 22, have the highest closeness centrality.

3.2.2.3 Betweenness Centrality

Finally, let us compare betweenness centrality for each player in the leading and trailing scenarios. Figure 19 shows that the trend is similar to that of the all possessions scenario. The star players 4, 9 and 10 have high betweenness centrality in both scenarios, while player 18 have relatively high betweenness for the leading scenario. Players 10 and 7 have higher betweenness for the trailing scenario suggesting that they are more important and/or prominent to the network when the team is trailing than when the team is leading.

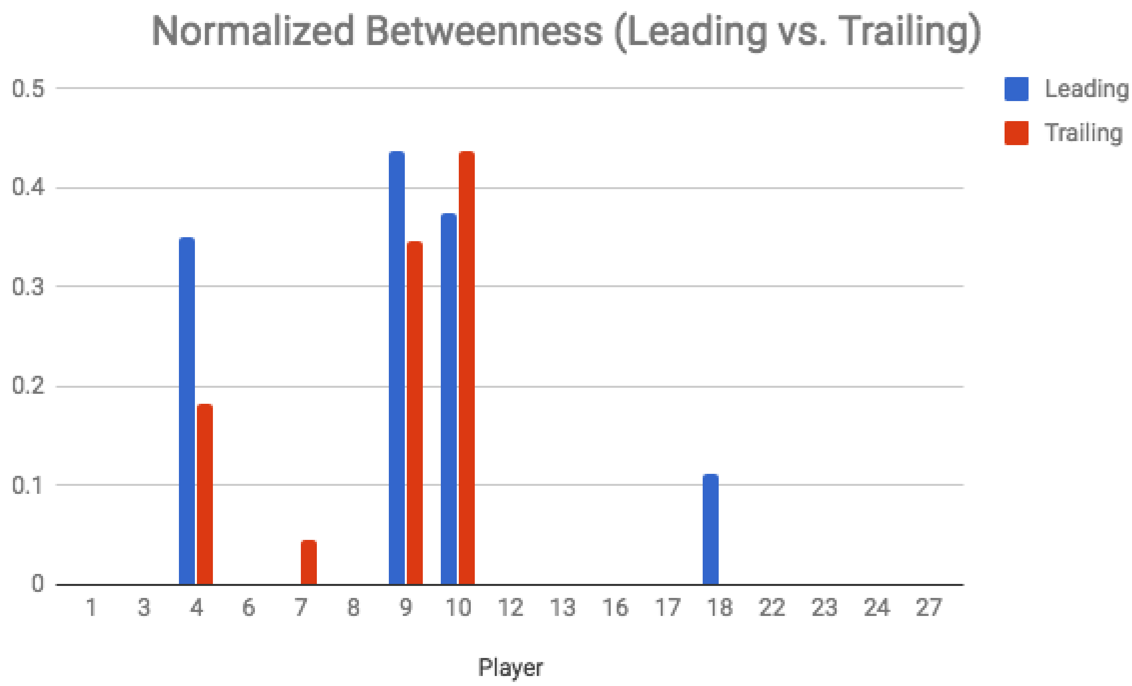

Figure 19: Graph comparing normalized betweenness for each player for the leading and trailing scenarios.

3.2.3 Goals

3.2.3.1 Vertex Degree

In this section we will compare the different centrality metrics for the possessions ending in goals with the possessions not ending in goals.

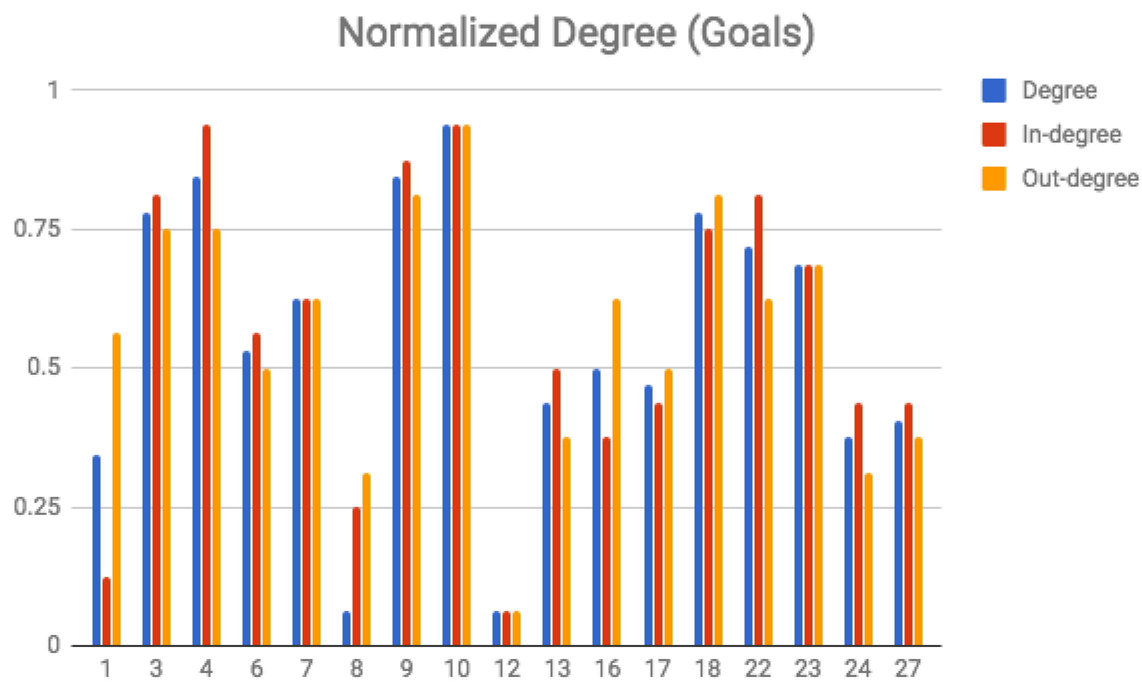

Figure 20: Graph showing normalized vertex degree and in- and out-degree for each player during possessions that ended in goals.

Figure 20 shows normalized vertex degree and normalized in- and out-degree for each player in the possessions ending in goals. It is clear from this figure that our star players, players 4, 9 and 10, have the highest degree in this scenario. This result tells us that the star players are well connected to the rest of the team in possessions that end in goals, which means that they are involved in such possessions along with nearly every other player on the team. In other words, they aren't only involved in scoring goals along with a small subset of the team's players. This suggests that the star players are versatile, flexible and play successfully with most other players on the team.

Figure 21 shows vertex degree and in- and out-degree for each player in the possessions that did not end in goals. Player 9 has higher vertex degree than in the goals scenario, while players 4 and 9 have slightly lower degree. On the other hand, players 18 and 22, who are the replacements for players 10 and 9 respectively, have higher degree than in the goals scenario. These observations suggest that the star players are more connected to the rest of the team during successful possessions, while their replacements are more connected during unsuccessful possessions.

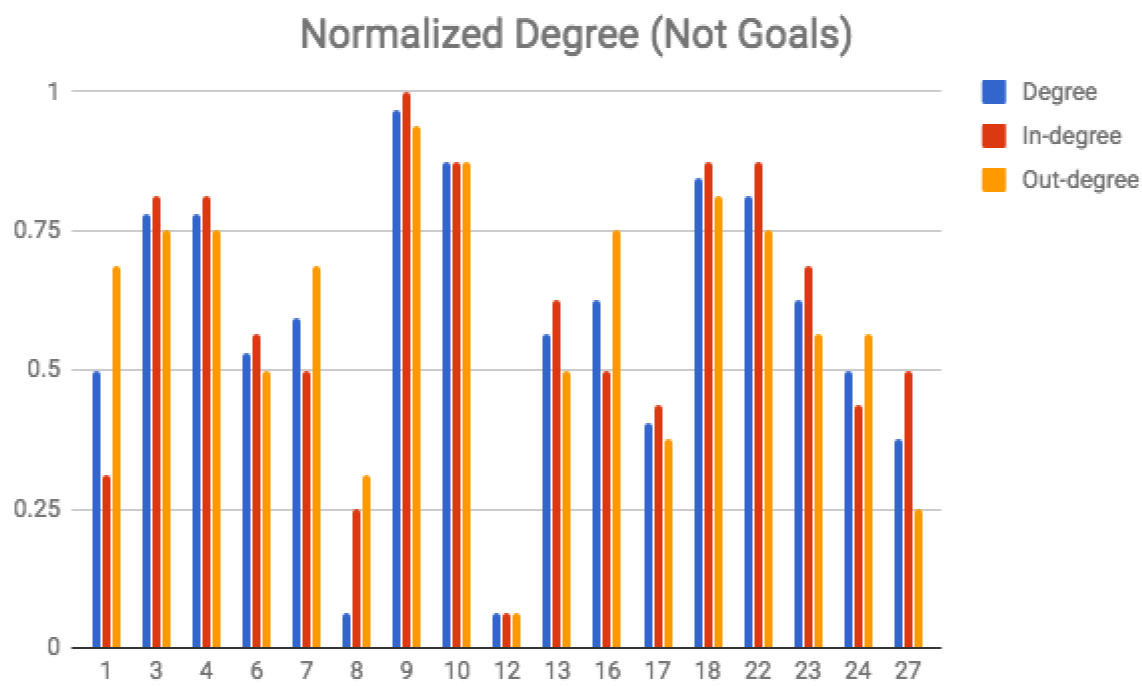

Figure 21: Graph showing normalized vertex degree and in- and out-degree for each player during possessions that did **not** end in goals.

3.2.3.2 Closeness Centrality

Figure 22 compares the normalized closeness centrality for each player in both scenarios. As with the leading and trailing scenarios, there exists a significant difference between the number of goals possessions and the number of possessions not ending in goals. This may explain why the closeness centrality for each player is so much higher for the possessions not ending in goals than for possessions that do end in goals, as shown in Figure 22.

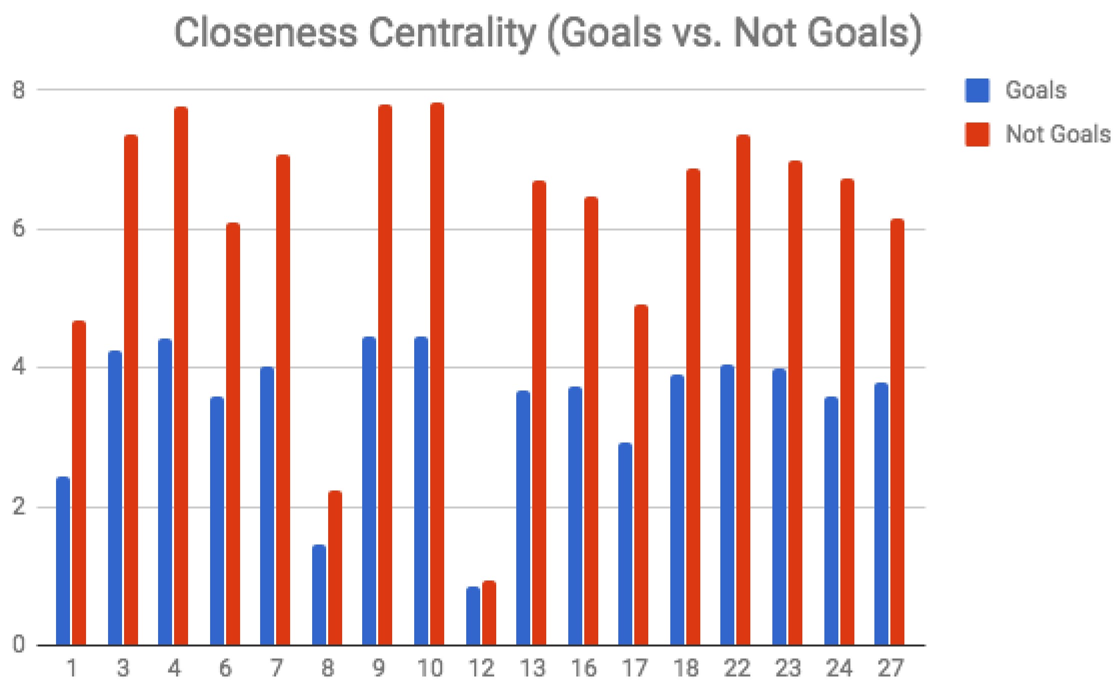

Figure 22: Graph comparing closeness centrality for possessions ending in goals with that for possessions not ending in goals.

In order to make comparisons between the two scenarios easier, we again divided the adjacency matrices through by the total number of passes in each scenario before computing closeness centrality. This resulted in the following graph shown in Figure 23. Now, it is clear that closeness centrality tends to be higher in the goals scenario. However, the general trend in both cases is the same: players 9, 10, 4 and 3 have the highest closeness centrality. Other than closeness being higher in general for possessions ending in goals, there does not seem to be much difference in relative closeness centrality between players in either scenario.

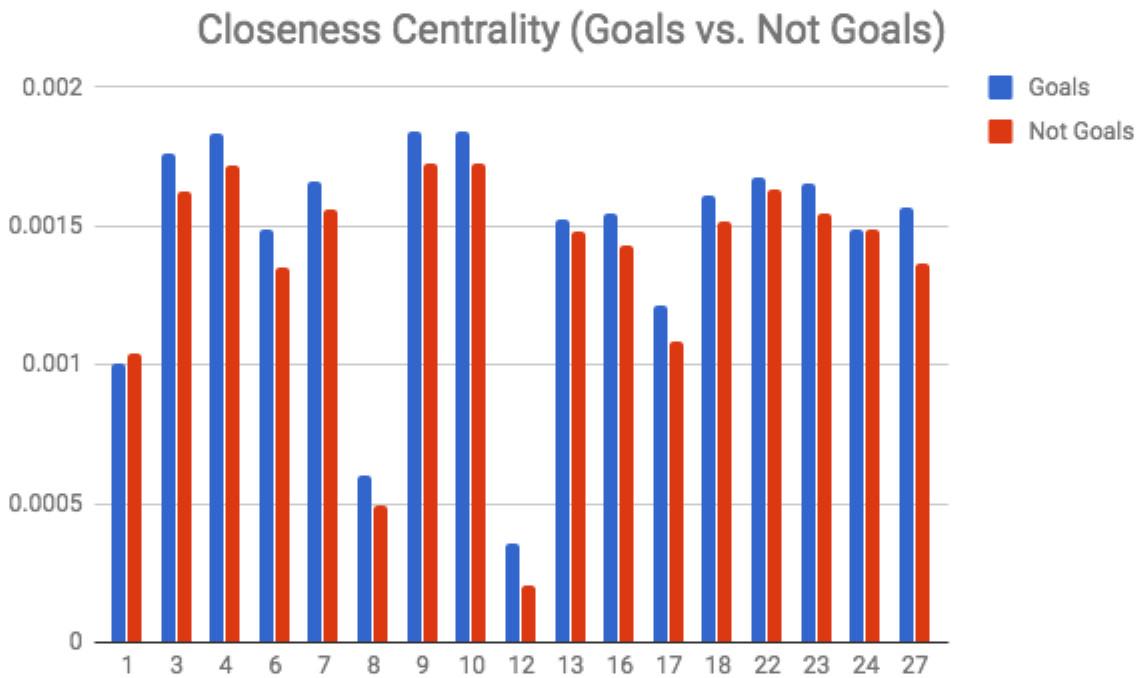

Figure 23: Graph comparing closeness centrality of the goals scenario with that of the non-goals scenario, taking into account the difference in the number of possessions in each scenario.

3.2.3.3 Betweenness Centrality

Finally, let us compare betweenness centrality for the two scenarios. Figure 24 shows normalized betweenness centrality for each player for the two scenarios. The results are similar to the previous two betweenness centrality graphs in that players 4, 9, 10 have the highest betweenness, while player 18 has relatively high betweenness. However, the results in the goals scenario differ because player 23 also has relatively high betweenness. This is actually not too surprising. In the IHF World Championship, player 23 scored more goals than any other wing player and scored more fastbreak goals than anyone else on her team [11]. In other words, her involvement in possessions ending in goals may explain why her betweenness in this scenario is high compared to the other wing players.

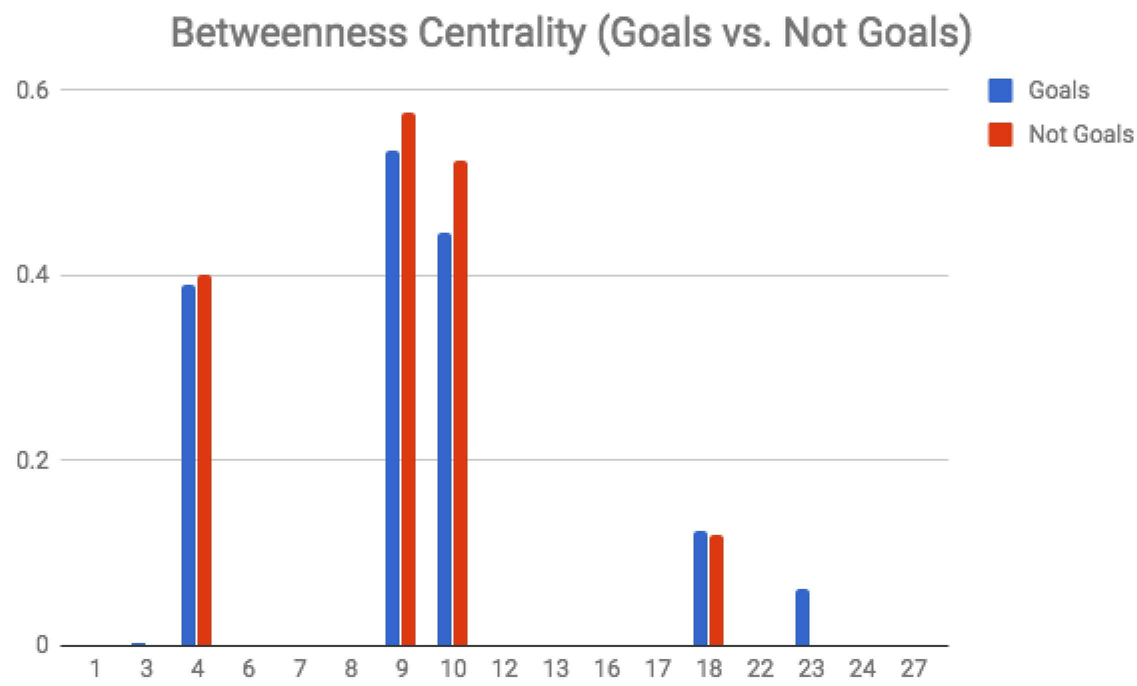

Figure 24: Graph comparing betweenness centrality for the goals and not goals scenario.

3.3 Three Pass Groups

3.3.1 All Possessions

Figure 25 shows the three pass group distributions for each player in the all possessions scenario. Blue corresponds to similarity while red corresponds to dissimilarity. Each row and column represents a player. The diagonal contains comparisons between a player and herself, which is a trivial comparison in this case since the graph in Figure 25 compares playing styles for just one scenario.

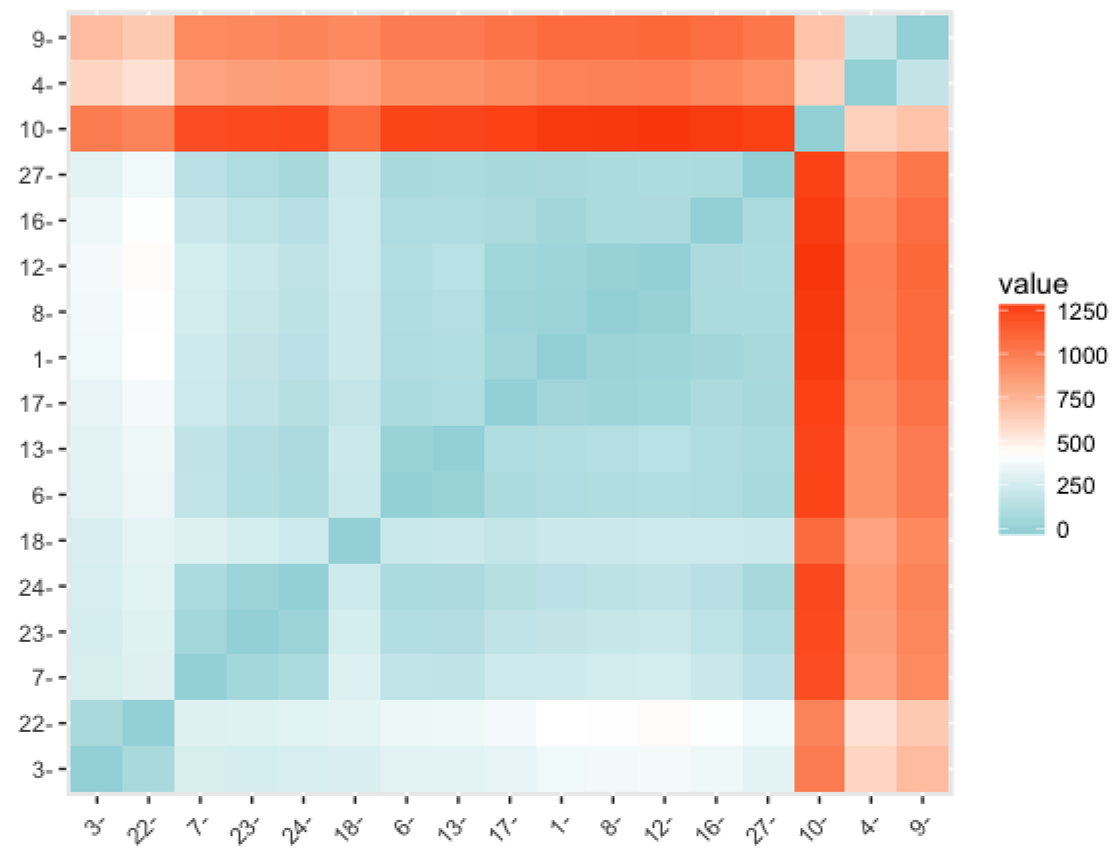

Figure 25: Graph comparing three pass group distributions for each player in the all possessions scenario.

It is clear that players 4, 9 and 10 are the most dissimilar from the rest of the team as their intersections are mostly red. Players 4 and 9 are similar to one another, but they are both different from player 10. Players 3 and 22 exhibit somewhat unique playing styles, but they are also quite different from the star players that they replace, i.e. players 4 and 9 respectively. Playing time may be a significant factor because the star players play more and therefore have more opportunity to participate in a greater number of different three pass sequences. In the following section, we normalize the three pass groups distribution and show that the star players are actually very similar to their replacements if playing time is accounted for.

If we ignore the star players, we can make observations about the rest of the team. The line players, players 6 and 13, are very similar to one another, as are the left wing players,

players 23 and 24. This is a neat result because it shows that position may have a lot to do with one's playing style.

3.3.2 Leading and Trailing

In order to remove the effect of each player's playing time, we normalized the three pass groups distributions for the leading and trailing scenarios. This was done by dividing the three pass groups distributions through by the total number of three pass groups belonging to a given player. For example, if some player's three pass groups distribution included 10 of ABAB, 5 of ABCD and 15 of BABA, normalizing her distribution would result in ⅓ of ABAB, ⅙ of ABCD and ½ of BABA. Figure 26 shows the normalized three pass groups for the leading scenario, with red corresponding to similarity and blue to dissimilarity. The players are ordered by similarity, so they're placed next to the players they are most similar to.

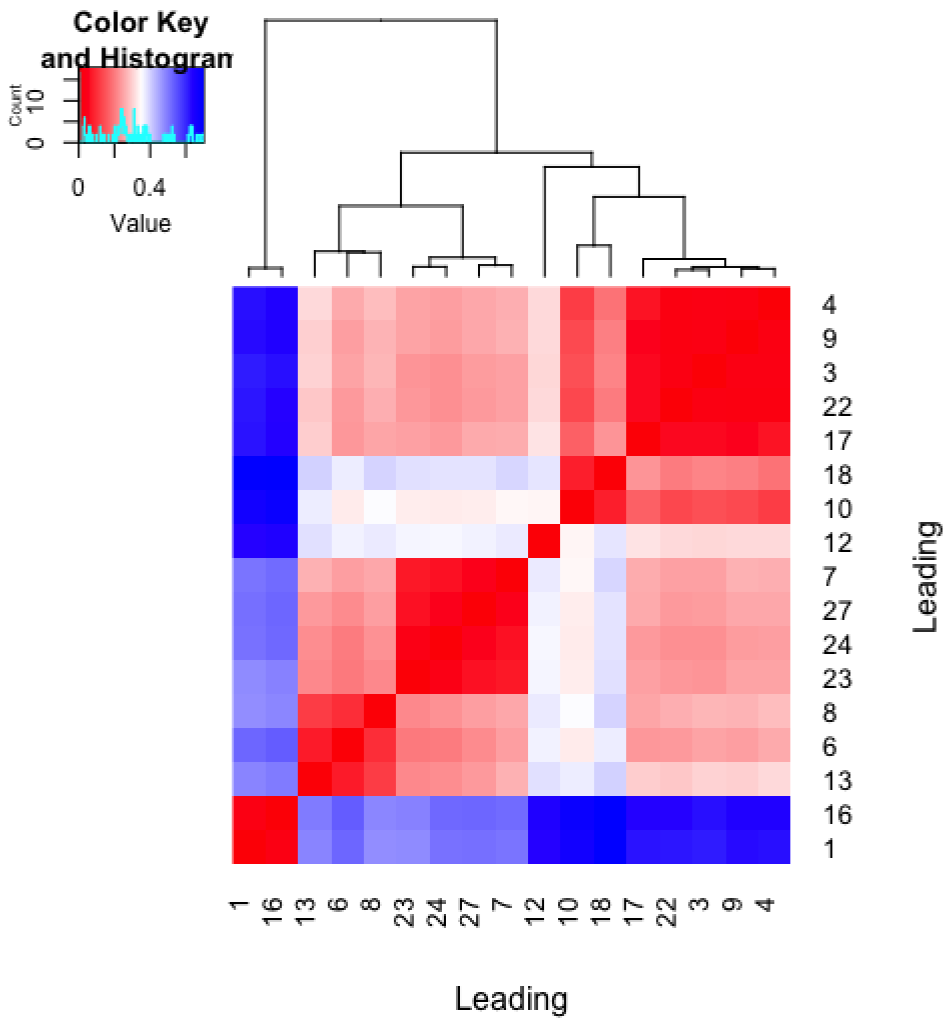

Figure 26: Graph showing normalized three pass group distribution in the leading scenario. The dendrogram shows how similar the players are to one another.

With the exception of player 12, who didn't play at all during the leading scenario, Figure 26 shows that players are ordered almost perfectly by position. This suggests that each position has a certain playing style that the players adhere to.

Players 4, 9, 3, 22, and 17 are the side backs. It is clear from Figure 26 that they are almost identical to one another in terms of their normalized three pass groups distributions. Amongst them, players 4 and 9 are more similar to one another than their replacements, players 3 and 22 respectively, but not by much. Similarly, players 10 and 18, the mid-backs, are almost identical to one another. Continuing down the line, we can see the players 7, 27, 23 and 24 are also almost identical to one another. These are the wing players. 7 and 27 are slightly more similar to one another, which is neat as they are the two right wings, while the left wings, players 23 and 24, are slightly more similar to one another as well. The line players, players 6, 8 and 13, are also remarkably similar, with 6 and 13 being almost identical. Finally, the two goalkeepers that actually played during the leading scenario, players 1 and 16, are completely different from everyone else, but almost identical to one another. This makes sense since goalkeepers play more defensively than the rest of the team, so you would expect them to only be involved in the beginning of a passing sequence.

The result in the trailing scenario is similar. Players 8, 17 and 18 did not play during the trailing scenario, but we can still make observations about the rest of the team. Figure 27 shows normalized three pass groups for the trailing scenario.

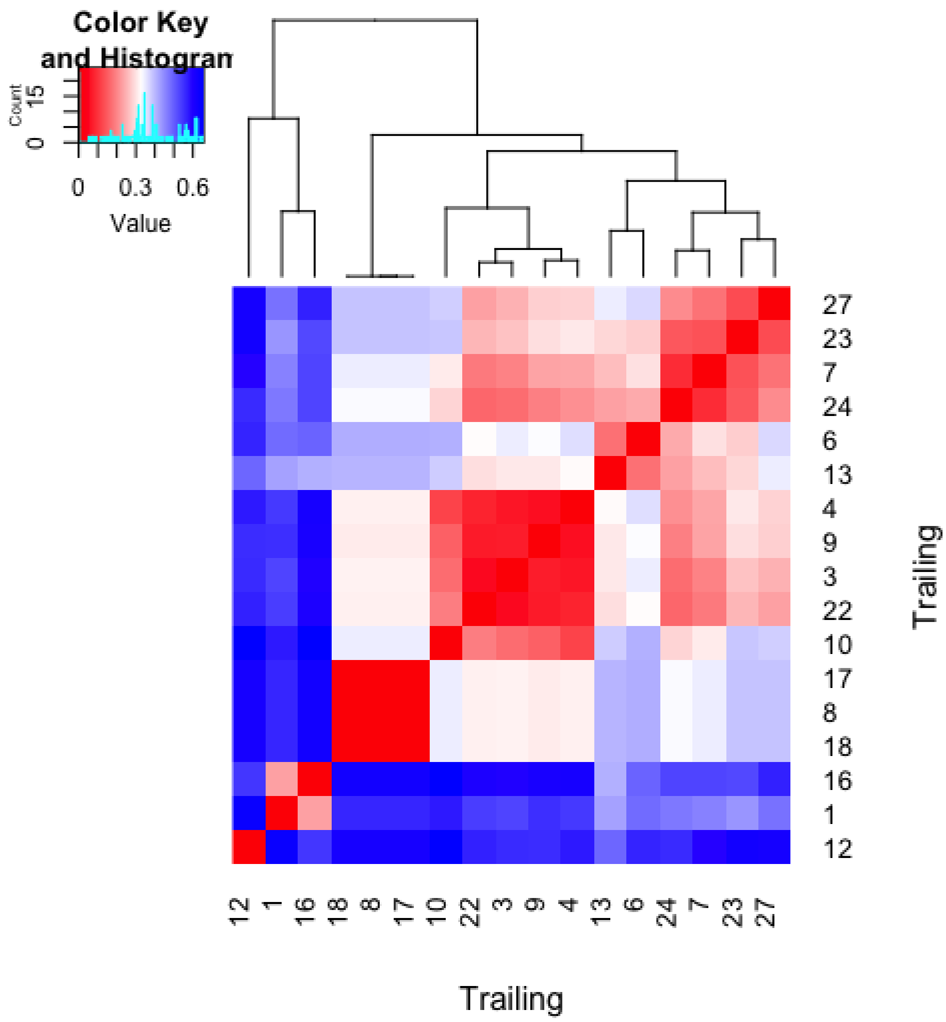

Figure 27: Graph showing normalized three pass group distribution in the trailing scenario. Again, the dendrogram shows how similar the players are to one another.

As with the leading scenario, the players are ordered almost perfectly by position with the exception of the players that did not play. The wing players are next to each other, but are not ordered by right and left wing this time. The line players are most similar to one another, but less so than in the leading scenario. The mid-back, player 10, is lumped in with the side backs, but is about as similar to them in this scenario as she was in the leading scenario. The side backs appear to be almost identical in this scenario as well. The goalkeepers, 1 and 16, are not as similar in this scenario, but they are more similar to one another than to anyone else.

These results suggest that the Norwegian women's handball team has an ideal playing profile for each position that its players try to adhere to. Evidently, their attempts are successful because the individual players' playing styles were sorted nearly perfectly by what position they played. It also suggests that there is no difference in playing style between the star players and their replacements other than the fact that the star players get more playing time.

3.3.3 Goals

Finally, we will compare normalized three pass groups distributions for possessions ending in goals with that of possessions not ending in goals. Figure 28 below shows normalized three pass groups for the goals scenario. Immediately, we can observe that there are fewer overall differences between players than in the leading or trailing scenarios. This may be because involvement in goals requires being a goalscorer, providing someone else with assists, and so on. However, the players are still sorted almost perfectly by position in this scenario. This suggests that in possessions ending in goals, there still exists a playing profile for each position that the players exhibit, but the players became more similar to one another overall.

The graph for the non-goals scenario is shown in Figure 29. This graph is more similar to the leading and trailing scenarios because there are clearer differences between positions than in the goals scenario. Again, the players are sorted nearly perfectly by position.

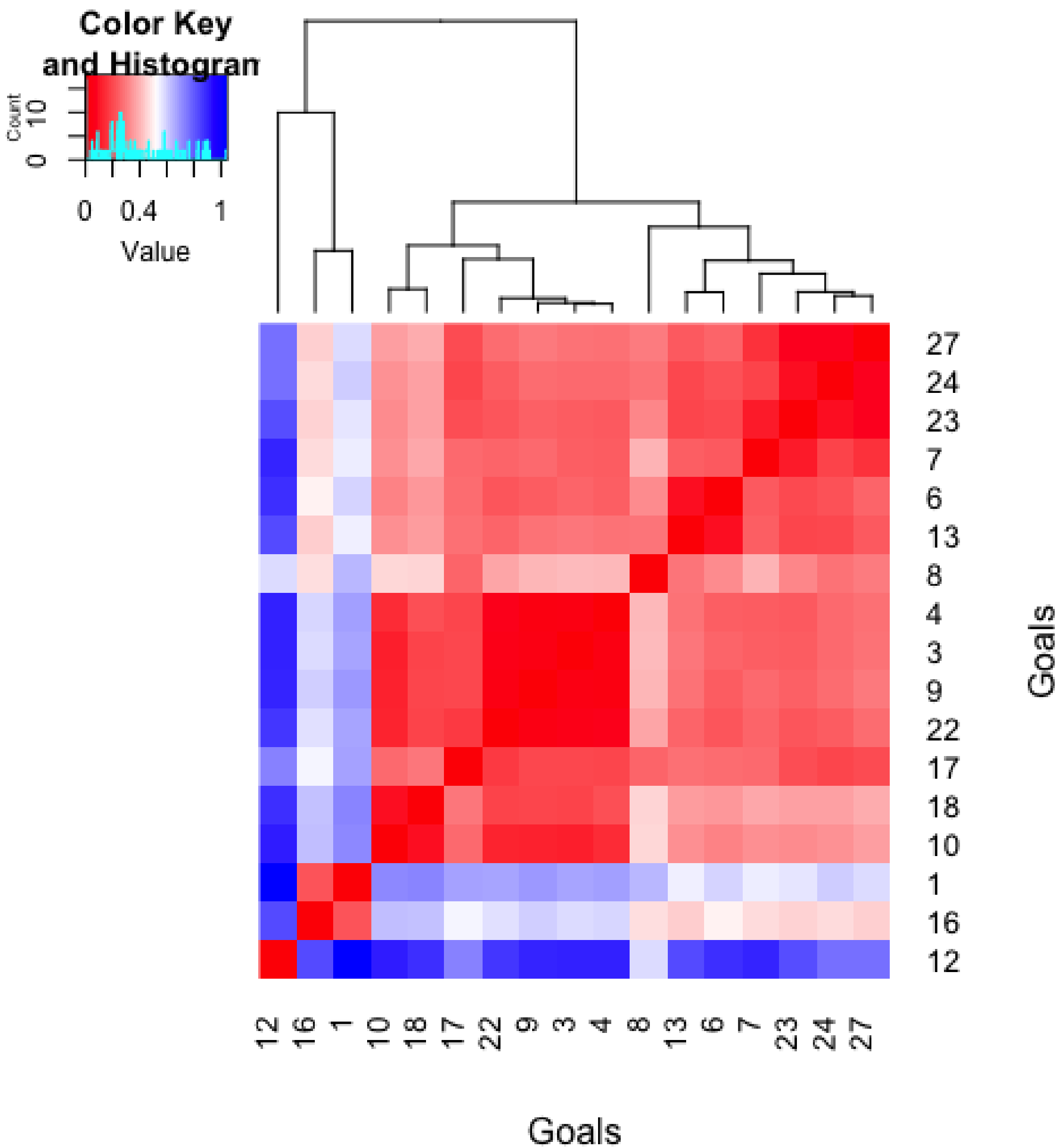

Figure 28: Graph showing normalized three pass group distribution in the goals scenario.

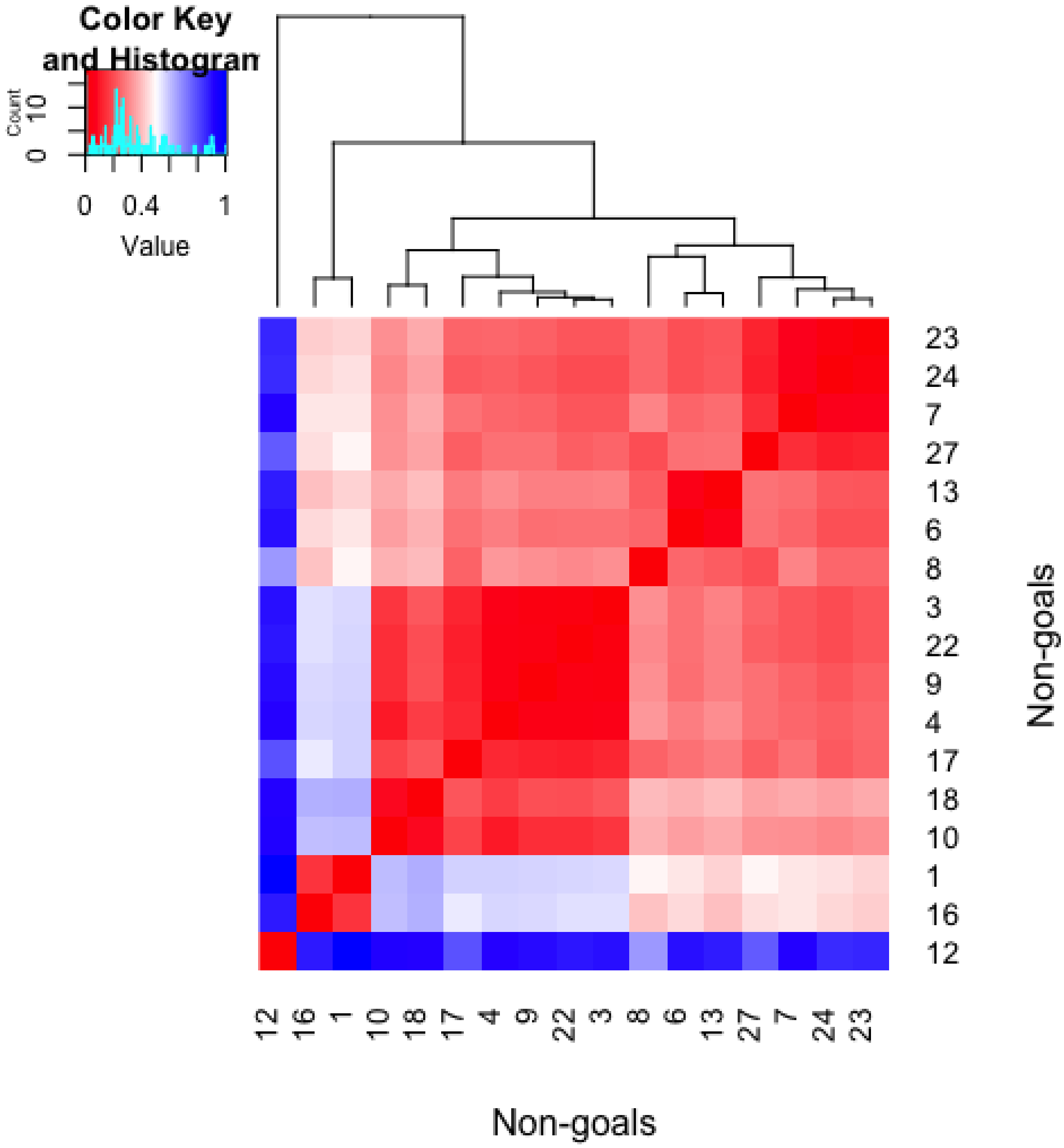

Figure 29: Graph showing normalized three pass group distribution in the non-goals scenario.

4 Conclusions

The questions explored in this thesis attempt to identify differences between the star players and their replacements; differences in the team's and individual player's playing styles when leading and when trailing; and differences in the team's and individual player's playing styles when scoring goals. We approached these questions by looking at various network measures and centrality metrics, and by using three pass groups to build a tool with which we could compare players' playing styles.

We found that in terms of centrality metrics, there were significant differences between the star players and their replacements. Star players tended to have higher centrality than their replacements, this was most apparent with the betweenness centrality metric. We also found that the star players had different three pass groups distributions than their replacements in the unnormalized scenario. However, in the normalized case it was clear that the star players had almost identical playing styles to their replacements in every scenario. These results suggest that each position has a prescribed playing profile that every player appears to adhere to. This is a neat result that would be fascinating to compare with other women's teams or the Norwegian men's team.

We found that players tended to have higher vertex degree in the leading scenario than in the trailing scenario, suggesting that the connectedness of the team may be correlated to their success. The three pass groups graphs also provided some evidence that players adhere to their respective positions' prescribed playing styles better in the leading scenario than in the trailing scenario. This suggests that the team's ability to adhere to this playing style may affect their success.

Comparing possessions that ended in goals with possessions that did not end in goals produced some interesting results as well. We found that normalized closeness centrality tended to be higher for the goals scenario, but that this was not the case for either vertex degree or betweenness centrality. This discrepancy may be due to the fact that there were more possessions that did not end in goals than possessions that did. However, it could also be due to the team not possessing significantly different centrality metrics in the two scenarios. On the other hand, the three pass groups graphs showed that the there was more similarity in players' playing styles in the goals scenario than in the non-goals scenario. The non-goals scenario was similar to the leading scenario in that every player was sorted by their position, and there appeared to be distinct and significant differences between playing styles for different positions. These results suggest that instead of adhering to their position's prescribed playing style, players all took on similar playing styles during possessions that ended in goals. This may be due to a certain set of three pass groups being more suitable for goals, which naturally would create more similarity between players in this scenario.

Thus, each question motivated interesting results that are significant. The team played differently when leading than when trailing, and the players played more similarly overall during possessions than ended in goals. The star players were different from the replacements in terms of centrality and because of reasons relating to playing time. The star players had higher centrality than their replacements, but their playing styles were nearly identical to that of their replacements. This suggests that while each position may call for a certain playing style, the star players are still more important to the team because of their importance to the passing network. High centrality means that they provide connections between many different players, which in handball translates to them not being limited to playing alongside a certain subset of the team's players. Being a “star player” may have more to do with one’s ability to play well with others than with one’s playing style or ability to score goals.

Further research in network analysis of passing patterns in handball could involve analyzing three pass groups distributions of the Norwegian men’s team to explore whether or not this prescribed playing style that I have discussed exists in men’s handball as well. It would also be interesting to compare the Norwegian women’s team to other international favorites such as France or the Netherlands.

5 References

[1] IHF “Rules of the Game”, Edition 1 July, 2016. Available at: http://www.ihf.info/upload/pdf-download/rules_english.pdf

[2] IHF Fixtures and Results, Available at: http://www.ihf.info/en-us/ihfcompetitions/worldchampionships/womensworldchampionships/ihfwomen%E2%80%99sworldchampionshipingermany2017/fixturesandresults.aspx

[3] Bekkers, Joris, and Shaunak Dabadghao. "Flow Motifs in Soccer: What can passing behavior tell us?." (2017).

[4] Pena, Javier López, and Hugo Touchette. "A network theory analysis of football strategies." *arXiv preprint arXiv:1206.6904* (2012).

[5] Peña, Javier López, and Raúl Sánchez Navarro. "Who can replace Xavi? A passing motif analysis of football players." *arXiv preprint arXiv:1506.07768* (2015).

[6] Prieto, Jaime, Miguel-Ángel Gómez, and Jaime Sampaio. "From a static to a dynamic perspective in handball match analysis: a systematic review." *The Open Sports Sciences Journal* 8, no. 1 (2015).

[7] Schrapf, Norbert, and Marcus Tilp. "Action sequence analysis in team handball." (2013).

[8] Tilp, M., and N. Schrapf. "Analysis of tactical defensive behavior in team handball by means of artificial neural networks." IFAC-PapersOnline 1 (2015): 48.

[9] Hassan, Amr. "The Use of Modular Feed Forward Neural Networks in Anticipating the Results of Handball Championship 2015." *American Journal of Sports Science* 3, no. 5 (2015): 93-97.

[10] IHF Women’s World Championship 2017, Statistics. Available at: http://www.ihf.info/en-us/ihfcompetitions/worldchampionships/womensworldchampionships/ihfwomen%E2%80%99sworldchampionshipingermany2017/statistics.aspx

[11] IHF Women’s World Championship 2017, Team Info, Roster and Player Statistics. Available at http://www.ihf.info/en-us/ihfcompetitions/worldchampionships/womensworldchampionships/ihfwomen%E2%80%99sworldchampionshipingermany2017/teaminfo.aspx

[12] IHF Women’s World Championship 2017, Fixtures and Results. Available at http://www.ihf.info/en-us/ihfcompetitions/worldchampionships/womensworldchampionships/ihfwomen%E2%80%99sworldchampionshipingermany2017/fixturesandresults.aspx

[13] Kolaczyk, Eric D., and Gábor Csárdi. *Statistical analysis of network data with R*. Vol. 65. New York: Springer, 2014.