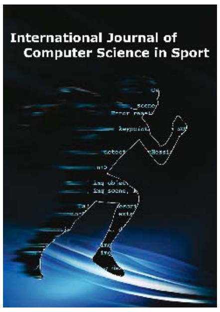
International Journal of Computer Science in Sport

Volume 18, Issue 3, 2019

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2019-0017



Analyzing passing networks in association football based on the difficulty, risk, and potential of passes

Wiig, A.S. [1]*, Håland, E.M.* [1]*, Stålhane, M.* [1]*, and Hvattum, L.M.* [2]

[1]*Department of Industrial Economics and Technology Management, NTNU, Trondheim, Norway*

[2]*Faculty of Logistics, Molde University College, Molde, Norway*

Abstract

This paper investigates the use of network analysis to identify key players on teams, and patterns of passing within teams, in association football. Networks are constructed based on passes made between players, and several centrality measures are investigated in combination with three different methods for evaluating individual passes. Four seasons of data from the Norwegian top division are used to identify key players and analyze matches from a selected team. The networks examined in this work have weights based on three different aspects of the passes made: their probability of being completed, the probability that the team keeps possession after the completed pass, and the probability of the pass being part of a sequence leading to a shot. The results show that using different metrics and network weights leads to the identification of key passers in different phases of play and in different positions on the pitch.

KEYWORDS: NETWORK ANALYSIS, PAGERANK, CENTRALITY, SPORT

Introduction

Network theory is a part of graph theory, and it has several applications, including social networks, computer science, biology, and medicine. Social network analysis is a useful method for examining the relationships and patterns among social entities; it allows researchers to investigate both social structures and individual attributes simultaneously. Different measures of centrality can be computed to identify the importance of an entity, while clustering coefficients can be calculated to study the extent to which the entities in a network cluster together (Boccaletti, Latora, Moreno, Chavez, and Hwang, 2006).

Recently, network analysis has been applied to association football for analyzing passing sequences and identifying key players on teams. For such purposes, teams are commonly viewed as networks where nodes represent players with edges connecting them. The edges can be weighted according to a chosen criterion, with the number of successful passes between two players being the most common choice. This modelling approach can be categorised as a type of social network analysis.

Duch, Waitzman, and Amaral (2010) used social network analysis to determine individual player contributions to teams, while Pena and Touchette (2012) used it to examine team strategies. In the latter paper, the playing style of a team was observed by fixing each node according to the team's tactical strategy. Furthermore, the team's game-play robustness, characterised by the lowest number of intercepted passes required to disturb its natural flow and to isolate a subgroup of its players, was evaluated. Additionally, different centrality measures were calculated to study players' individual contribution to their team in terms of importance and connectedness.

Different centrality measures were also explored by Arriaza-Ardiles, Martín-González, Zuniga, Sánchez-Flores, de Saa, and García-Manso (2018). They computed the players' closeness and betweenness centrality scores to study players' capability to connect with teammates and their ability to make contributions in the play between other players. Furthermore, clustering coefficients were calculated to measure relations between players and to identify the importance of the players who most frequently interacted with each other when ball possession was kept. Another centrality measure, the PageRank centrality, was investigated by Rojas-Mora, Chávez-Bustamante, del Río-Andrade, and Medina-Valdebenito (2017) to find the most important players of a team in matches from the group stage of the Copa America 2015. A combined network for both teams in a match was used, meaning that players on different teams were connected by edges representing miskicks and interceptions.

Gama, Passos, Davids, Relvas, Ribeiro, Vaz, and Dias (2014) used a network-based approach to identify the key players in the attacking phases of play, while Pina, Paulo, and Ara´ujo (2017) investigated the relationship between specific network metrics and teams' outcome of offensive plays, after controlling for the effect of total passes. In the latter paper, a hierarchical logistic regression model was developed using data from the group stage of the UEFA Champions League 2015/2016, and the extent to which network density, clustering coefficients and centralisation could predict a team's performance on offensive plays were examined. Results suggest a negative relationship between the success of offensive plays and network density, which is the only significant predictor variable in the model. This result was supported by Peixoto, Praça, Bredt, and Clemente (2017) who applied social network analysis to study differences across centrality measures in successful and unsuccessful offensive plays.

McHale and Relton (2018) elaborated on the work done by Szczepański and McHale (2016) on assessing players' passing ability and developed a generalised additive mixed model (GAMM) to estimate the likelihood of a pass's success with the purpose of using the results from the model to identify the key passers in a team through a network analysis. The edges in the network were weighted according to the difficulty of the passes between players, given by the estimated likelihoods from the fitted GAMM. By taking pass difficulty into account when weighting the passes, the players' involvements in the passing of the ball can more fairly be compared as it is not only the number of completed passes that is considered. The key players in a team were determined by calculating the exponential centrality measure of each player. Although network analysis within sport seems to be more commonly used in the context of association football, it has also been used in basketball, by Fewell, Armbruster, Ingraham, Petersen, and Waters (2012) and Piette, Anand, and Pham (2011), as well as in volleyball (Clemente, Martins, and Mendes, 2015, Kang, Huh, and Choi, 2015) and cricket (Dey, Ganguly, and Roy, 2017).

This paper contributes to the literature on network analysis for association football in several ways. First, the different centrality measures considered includes the closeness centrality, the betweenness centrality, PageRank centrality, and clustering coefficients. This appears to be the first study where PageRank centrality is calculated alongside other centrality measures. Second, three different network types are considered, where the weights on links are calculated differently. Rather than using the most common choice, with link weights based on the number of successful interactions between two players, the first network is designed by considering the difficulty of passes, as by McHale and Relton (2018). The second and third networks use novel weights, based on the work of Håland,Wiig, Stålhane, and Hvattum (2019). They proposed three different GAMMs, calculating the difficulty of passes, the risk of passes (based on the probability that a pass reaches a teammate and that the next event after the pass is successful), and the potential of passes (the probability that the pass is part of a sequence that leads to a shot by the team in possession). Third, for the PageRank measure, this paper considers centrality from the perspective of both the pass recipient and the passer separately. Fourth, a new data set is used, with matches from four seasons of the Norwegian top division, and a case study is presented giving insights from two selected matches and for players from a selected team. The latter part of the analysis focuses on Rosenborg, the most successful team from Norway in recent years, although the methodology is applicable to any team.

The aim of the work is to investigate how building networks based on different aspects of a pass (difficulty, risk, and potential), and calculating various metrics (closeness, betweenness, two versions of PageRank, and clustering coefficients) can contribute to identify key passers in different positions and roles. Previous work has mainly built networks based on the frequency of passes between players, which is useful to identify the players involved in many passes. Instead using weights based on the potential of passes, players involved in passes leading up to goals can be identified. This information may be more useful when analyzing the attacking patterns of a team, or when scouting opposing teams before matches.

The rest of this paper is structured as follows. Details of the different centrality measures are presented in the next section. The subsequent section then contains a description of the data set used, details of how the different networks and their weights are calculated, and additional information on the calculation of the different centrality measures. Results from applying the metrics to the data set are provided in the penultimate section, including a case study and a comparison of the results to those reported in existing literature. Finally, the last section provides concluding remarks.

Centrality measures

An undirected and unweighted graph $G = (\mathcal{N}, \mathcal{L})$ consists of two sets, $\mathcal{N}$ and $\mathcal{L}$, such that $\mathcal{N} \neq \emptyset$ and $\mathcal{L}$ is a set of unordered pairs of elements in $\mathcal{N}$. The elements of $\mathcal{N}$ are called nodes, and the elements of $\mathcal{L}$ are called edges. Two nodes are adjacent if they are connected by an edge. An edge is defined by two nodes, i and j, and is denoted as (i, j). For a directed graph, direction matters, and the edges are ordered pairs, such that (i, j) and (j, i) are not equal.

An adjacency matrix, A, can be used to represent a graph. For a directed network, it is an $N \times N$ matrix with elements a_{ij} $(i, j = 1, \ldots, N)$. If there is a directed edge from node i to node j, a_{ij} equals one, otherwise the value is zero. Furthermore, the directed edges can be weighted according to the strength of connection between the nodes. A weighted graph G^W adds to an unweighted graph by including an additional set of edge weights, $\mathcal{W}$, with $|\mathcal{L}| = |\mathcal{W}|$. A weighted graph can be described by a weights matrix, W. The elements of an $N \times N$ weights matrix are given by w_{ij}, where w_{ij} represents the strength of the connection between nodes i and j. If two nodes are not directly connected, w_{ij} equals zero. The adjacency matrix for a weighted graph consists of entries such that $a_{ij} = 1$ if $w_{ij} \neq 0$ and $a_{ij} = 0$ if $w_{ij} = 0$.

A sequence of distinct adjacent nodes is defined as a path, and the shortest path refers to the path of minimal distance between two nodes. In an unweighted network, the distance is equivalent to the number of directed edges traversed, whereas for a weighted network the distance is the sum of the weights on the path. To interpret the edge weight between two nodes in a weighted graph as a strength of a connection instead of a cost, Opsahl, Agneessens, and Skvoretz (2010) proposed to inverse the weights before determining the shortest path. Then, the length of the shortest path between nodes i and j can be found by solving:

$$d_{ij}^{w\alpha} = \min\left(\frac{1}{(w_{ih})^{\alpha}} + \cdots + \frac{1}{(w_{hj})^{\alpha}}\right), \tag{1}$$

where α is a positive tuning factor, giving the equivalent to an unweighted network if it has the value of zero.

A directed graph is said to be strongly connected if there is a directed path from node i to node j for every pair of distinct nodes in the network. If there is only an undirected path between the nodes however, the graph is weakly connected. Otherwise, the graph is unconnected.

In the following, four different centrality measures are discussed, together with how they can be interpreted in the context of association football. The four metrics are: closeness centrality, betweenness centrality, PageRank centrality, and clustering coefficients. By constructing a social network of team players, it is possible to identify the influence and importance of players in a team-passing dynamic game. In a passing network, players are represented by nodes and edges represent interactions between the players, such as passes made. The consequences of using different network weights, based on other aspects of passing, are discussed in a later section.

The *closeness centrality* of a node depends on the length of the paths from the node to all other nodes in the network. It provides a measure of a node's independence from other nodes, with higher scores being associated with greater independence (Freeman, 1978). For a weighted network where the weights are considered to be strengths, and both the directed edges going in and out of a node are considered, the closeness centrality of node i is given by (Pena and Touchette, 2012):

$$C_C(i) = \frac{2(N-1)}{\sum_{j \in \mathcal{N} \setminus \{i\}} (d_{ij}^{w\alpha} + d_{ji}^{w\alpha})}, \tag{2}$$

where $d_{ij}^{w\alpha}$ represents the shortest path between nodes i and j given by equation (1), N is the total number of nodes in the graph, and $N - 1$ is a normalisation factor allowing the closeness measure to be comparable across networks of different sizes (Freeman, 1978). Additionally, it lets the measure to be interpreted as the inverse of a node's average distance of the shortest paths to the other nodes in the graph.

In the context of association football, closeness can be interpreted as a measure of the easiness of reaching a particular player within a team. Players achieving higher closeness scores tend to reach more players in fewer passes (Clemente, Martins, and Mendes, 2016).

The *betweenness centrality* of a node is based on the number of shortest paths between two other nodes passing through the node. The idea behind the measure is that a node in a network is central if it is placed on the shortest path between two connecting nodes. Also, it can viewed as a measure of a node's potential to control the flow of information in a graph (Freeman, 1977). The betweenness centrality of node i can be defined as:

$$C_B(i) = \left(\frac{2}{N^2 - 3N + 2}\right) \sum_{\substack{j,k \in \mathcal{N} \\ j \neq k}} \frac{g_{jk}(i)}{g_{jk}}, \tag{3}$$

where g_{jk} is the number of shortest paths from node j to node k, while $g_{jk}(i)$ is the number of shortest paths between nodes j and k passing through node i (Boccaletti et al., 2006). The term to the left of the summation is a normalisation factor that enables comparisons between networks with differing number of nodes, N (Freeman, 1977). The shortest paths can be calculated by a method proposed by Brandes (2001), which is modification of Dijkstra's algorithm (Dijkstra, 1959).

For association football, the betweenness score gives an indication of how the ball-flow between teammates depends on a specific player. Players with high scores play a key role as connecting bridges between teammates, and they contribute to keep ball possession within the team (Gonçalves, Coutinho, Santos, Lago-Penas, Jiménez, and Sampaio, 2017). With the usual approach of using the number of passes between players as weights in a network, a low betweenness score is associated with less involvement in the game, and the effect of removing that player from the game is minimal. From the standpoint of a team, betweenness scores that are evenly distributed among players may indicate a well-balanced passing strategy and less dependence on particular players (Pena and Touchette, 2012).

The *PageRank* algorithm was introduced by Brin and Page (1998) to provide a method of measuring the importance of web pages. The intuition behind the algorithm is that a web page achieves a high PageRank either if many web pages are pointing to it or if some of the web pages pointing to it have a high PageRank themselves. Hence, all web pages' PageRank scores must be calculated simultaneously as the scores depend on each other.

For a network, the PageRank of node i is given by:

$$PR(i) = \frac{1-d}{N} + d \sum_{j \in \mathcal{M}(i)} \frac{PR(j)}{C(j)}, \tag{4}$$

where d is a damping factor representing the probability that a node will join together with other nodes, N is the total number of nodes in the network, $PR(j)$ is the PageRank of node j, $C(j)$ is the number of directed edges departing from node j and $\mathcal{M}(i)$ is the set of nodes that are connected to node i (Fu, Lin, and Tsai, 2006, Page, Brin, Motwani, and Winograd, 1999). This definition deviates from the original definition proposed by Brin and Page (1998) as the first term on the right-hand side in the original equation is divided by the total number of nodes. By doing so, the PageRank scores in a network sum to one.

For a weighted network, the PageRank centrality can be estimated as:

$$PR^w(i) = \frac{1-d}{N} + d \sum_{j \in \mathcal{M}(i)} \frac{w_{ji}}{L_j} PR^w(j), \tag{5}$$

where w_{ji} are elements of the weights matrix, $L_j = \sum_k w_{ij}$ is the sum of the weights on the edges with direction out from node j and the other parameters are as given in equation (4) (Pena and Touchette, 2012). Higher weights on the incoming edges to a node correspond to a higher PageRank for that node.

For team sport analysis, the PageRank centrality is a recursive notion of *popularity* or *importance*. From a recipient's perspective, a player is important when receiving passes from other important players, while from a passer's perspective, a player is important when passing the ball to other important players. Basically, the PageRank centrality assigns to each player the probability that the player will receive or pass the ball after some passes have been made. The damping factor represents the probability that the players in a team will successfully pass the ball to their teammates (Pena and Touchette, 2012).

Clustering coefficients are computed to get a quantification of nodes' tendency to cluster together, and they account for the transitivity of a graph, that is, the probability that adjacent nodes of a node are connected. A high clustering coefficient of node i means that if node i is connected to node j, and node j is connected to node h, then the probability of node i also being connected to node h is high (Barrat, Barthelemy, and Vespignani, 2007). Following the method of Barrat et al. (2007), the weighted local clustering coefficient of node i in an undirected graph is given by:

$$c^w(i) = \frac{1}{s_i(k_i - 1)} \sum_{j,j \in \mathcal{N}} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}, \tag{6}$$

where s_i is the strength of node i, that is, the sum of the node's edge weights of adjacent nodes, k_i is the node degree, which is the number of undirected edges incident to the node, w_{ij} and w_{ih} are weights, a_{ij}, a_{ih}, and a_{jh} are elements of the adjacency matrix, and $s_i(k_i - 1)$ is a normalisation factor ensuresing that $0 \leq c^w(i) \leq 1$. The equation is undefined for nodes with only one connecting node.

For association football, clustering coefficients are computed to get a quantification of players' tendency to cluster together. Such coefficients can be used to assess how close a particular player and his teammates are to become a complete subgraph (Clemente et al., 2016). A high individual clustering score may indicate that a player acts as a middleman for his teammates, and by averaging the players' individual coefficients, a team's clustering coefficient may provide insight into how well-balanced the team is (Pena and Touchette, 2012).

Experimental setup

For the analysis, event-data from Opta Sports's Opta24Feed (Opta Sports, 2018) are used. This data set covers four seasons, 2014–2017, of the Norwegian top division, Eliteserien. For each of

960 matches, the Opta24Feed provides a file describing every event occurring in the match, focusing on on-ball involvements with the addition of some off-ball situations such as bookings and substitutions. All events are recorded manually by three analysts, and include information such as pitch locations and time stamps. Among the on-ball events, a total of 749,859 passes by 831 different players are recorded. This includes passes from open play, headed passes, long passes, and crosses, while excluding free kicks, corners, and throw-ins.

Inputs to build networks are based on passes made between players. However, instead of just using the number of successful passes made, passes are evaluated by one of three different GAMMs, as presented by Håland et al. (2019). Each of these can be stated as follows (Wood, 2006):

$$\eta_i = X_i\beta + Z_i\alpha + f_1(x_1) + \cdots + f_j(x_j), \tag{7}$$

where X_i is a vector of fixed effects with coefficients β, Z_i is a vector of random effects with coefficients α, and f_j are smooth functions of variables x_j. The random effects are assumed to be normally distributed, $\alpha \sim N(0, \Sigma_\sigma)$, with Σ denoting the covariance matrix, parameterised by the coefficient vector σ.

The three GAMMs differ in the definition of the dependent variable. The first model estimates the difficulty of a pass, that is, whether the pass successfully reaches a teammate. The second model measures the risk of passes, with a successful observation indicating that the pass first reached a teammate and then the next event after the pass was also successful. The third model gauges the offensive potential of a pass, identifying the probability that the current sequence of play results in a shot towards the opponent's goal. Each GAMM uses a large set of explanatory variables, listed in Table 1. The fixed-effect variables in each final model were determined using Wald-tests. The resulting models were validated using AUC-ROCs (area under the receiver operating characteristic curves) and AUC-PRs (area under the precision-recall curve) (Fawcett, 2006), with acceptable results. Additional details of the GAMMs and explanations for each variable included, are provided in (Håland et al., 2019).

In the context of network analysis, the purpose of using these three GAMMs is to derive additional information about each pass, and thereby create network weights that reflect a deeper understanding of passing behavior. Instead of each pass having a weight equal to one, the GAMMs produce a weight for each pass in the interval $[0,1]$, to reflect the difficulty, risk, or potential of the pass. Rather than just looking at which players are central in passing per se, network weights derived from the GAMMs can therefore be used to analyze which players are central in making non-trivial or difficult passes, in making passes that allow the team to keep possession, in receiving and following up on passes that otherwise have a high risk of leading to a loss of possession, and in making passes that have a high probability of generating shots.

Network Setup

To identify the key players on teams in Eliteserien, both directed and undirected weighted passing networks are created to obtain the chosen network metrics. The passing networks consist of all players in a team for a given season who have made or received at least one pass, and the metrics are calculated in three separate cases: pass difficulty, risk, and potential. The networks were constructed in RStudio using the *igraph* package (Csardi and Nepusz, 2006).

Table 1: Summary of explanatory variables used in models by Håland et al. (2019). Fixed-effect variables are denoted by X, random-effects variables by Z, and smooth terms by $f\ (\cdot)$. The types of variables of fixed and random effects are continuous (C), categorical/factor (F), binary (B), and interaction (I).

Variable	Description	Type
$X_{1,s}$	Pass number category in the current sequence of passes ($s = 1,2,3,4$)	F
X_2	Tackle in the previous event	B
X_3	Aerial duel in the previous event	B
X_4	Interception in the previous event	B
$X_{5,i}$	The player making the pass also took part in a tackle ($i = 1$), aerial duel ($i = 2$), or interception ($i = 3$)	B
X_6	Ball recovery due to a loose ball in the previous event	B
X_7	Previous pass was a header	B
X_8	Current pass is a header	B
X_9	Player performing the pass plays for the home team	B
X_{10}	Previous event was a free kick	B
X_{11}	Previous event was a throw-in	B
X_{12}	Corner taken within the past five events	B
$X_{13,i}$	The team attempting a pass just executed a corner ($i = 1$), free kick ($i = 2$), or throw-in ($i = 3$)	B
X_{14}	The match is played on artificial grass	B
$X_{1,2} * X_{10}$	Pass sequence number 2 interacting with free kick in previous event	I
$X_{1,2} * X_{11}$	Pass sequence number 2 interacting with throw-in in previous event	I
$Z_{1,k}$	Player k passing the ball ($k = 1, \dots ,689$)	F
$Z_{2,\mathrm{t}}$	Team t the player is representing ($t = 1, \dots ,19$)	F
$Z_{3,\mathrm{o}}$	Opponent team o to the player passing the ball ($o = 1, \dots ,19$)	F
$f_1(x_0, y_0, x_1, y_1)$	4-D smooth for the start (x_0, y_0) and end coordinates (x_1, y_1) of a pass	C
$f_2(x_2, y_2)$	2-D smooth representing the average position of the player given by coordinates (x_2, y_2)	C
$f_3(x_3)$	1-D smooth representing game time, x_3, in minutes	C
$f_4(x_4)$	1-D smooth for time played by the player passing the ball, x_4	C
$f_5(x_5)$	1-D smooth representing the goal difference, x_5	C
$f_6(x_3, x_5)$	2-D smooth representing the interaction between game time and goal difference	I
$f_7(x_7)$	1-D smooth for the time passed since last occurred event, x_7	C
$f_8(x_8)$	1-D smooth for the Elo rating, x_8, of the opponent team	C
$f_9(x_9) * X_{14}$	1-D smooth functions representing month of play, x_9, interacting with type of grass	I

Defining Weights Matrices

For the case of pass accuracy, the difficulty of the completed passes are used as weights, an approach introduced by McHale and Relton (2018). Both the passer and the recipient are thus seen as better if they pass or receive passes that are more difficult to make, respectively. Following Håland et al. (2019), a GAMM is used to decide the difficulty of a pass by utilising its average predicted likelihood of success, p_1. The average prediction of a pass can be thought of as its easiness, due to it being separated from the skill of the player and the teams. Then, the pass difficulty is given as $1 - p_1$.

When considering the risk of passes, the average predicted probabilities from a second GAMM, p_2, are used to define the weights in the networks. These predictions give the probability that the pass recipients are able to successfully perform the action in the event succeeding a pass. Passers do a better job if they deliver a pass that is easier to follow up, whereas recipients do a better job if they are able to follow up a risky pass, and so the weights are separated for the two cases. Two graphs are considered for each team: one handling the receiving view of a pass and one handling the passing view. The former has weights $1 - p_2$ and the latter has weights given by p_2.

The third GAMM, measuring the potential of passes to end in a shot, is defined in such a way that a different way to define weight matrices is required. This is due to the fact that the pass recipient is not necessarily the player attempting the shot at the end of a passing sequence. Hence, the recipient can not be seen as a good contributor of passing potential just because he or she was able to receive a pass with a low probability of being effective. Therefore, the sum of potential for passes between players is used as weights, to favour players that more often are involved in the offensive play. These weights are used for both passers and recipients.

For each of the three network types, the weights associated with all passes going from one player to another in a season are summed up in a weights matrix, W. As a recipient is needed for all passes, only successful passes are considered. Moreover, for the networks considering the risk and potential of passes, only those passes that are successful according to their defined dependent variables are analysed. The weights matrices are normalised by dividing each entry w_{ij} by the total number of matches in which player i has made a pass to player j during the season considered. The edge weights are defined as strengths, and, consequently, higher weights between players are favourable, and the shortest paths between nodes in a network are decided as defined in equation (1). This way, the highest scores given by the network metrics are assigned to the key players in a team. The three networks with different specifications of weighting are referred to as Network 1 (passing difficulty), Network 2 (passing risk), and Network 3 (passing potential).

Network Metric Specifications

Three centrality measures for a weighted network are considered: closeness, betweenness and PageRank. Additionally, the Barrat clustering coefficient is calculated. The metrics are computed for each player and each season in the data set. If a player has played for different teams in one season, the player is given several scores for that season. For Network 1, all measures are considered, while only the PageRank centrality is considered for Network 2 and Network 3.

The closeness and betweenness measures are calculated by using inverse weights as distances in equations (2) and (3). Dijkstra's algorithm is utilised in the estimation of closeness, while Brandes's algorithm is used for betweenness. When no directed path exists between two nodes, the length of the shortest path between the nodes is considered to be equal to the number of nodes in the network.

The PageRank centrality is considered separately for the pass recipient and the passer. The passer's perspective is obtained by switching the direction of the edges for Network 1 and Network 3. When considering Network 2, two separate graphs are used to calculate each of the PageRanks: one using the recipient graph and one using the passer graph. Equation (5) is used in the calculations, and the damping factor is set to 0.85, which is the commonly used value, such as in the original work of Brin and Page (1998), or by Liu, Liu, Lu, Wang, and Wang (2016) who studied a network of football player transfers. However, the value of the damping factor could also be tuned for a specific application, such as by Lazova and Basnarkov (2015) in an attempt to rank national football teams.

Table 2 summarizes the different measures calculated. All centrality measures are normalised by the maximum score of the measure in the corresponding network. Hence, based on a given centrality measure, the key player in a team gets a score of one. To calculate the Barrat clustering coefficient in equation (6), an undirected graph is needed. This is accomplished by collapsing the corresponding edge weights in the directed graph by averaging them.

Table 2: Summary of measures used.

Abbreviation	Description
$C_C^{w1}(i)$	Closeness centrality based on Network 1
$C_B^{w1}(i)$	Betweenness centrality based on Network 1
$PR_R^{wk}(i)$	PageRank centrality for recipients based on Network $k = 1{,}2{,}3$
$PR_P^{wk}(i)$	PageRank centrality for passing players based on Network $k = 1{,}2{,}3$
$c^{w1}(i)$	Clustering coefficient based on Network 1

Results and discussion

In this section, a validation of the network metrics computed for the players in Eliteserien is performed. Then, the key players on Rosenborg in the 2017 season are presented, together with case studies from two Rosenborg matches. Furthermore, the network approach used here is compared with the ones used in other studies.

Validation

Two approaches are considered to validate the computed network metrics for the players in Eliteserien. The first is to calculate the correlation for each metric across consecutive seasons. Only players appearing for the same team in both seasons are part of the analysis. This can be used to evaluate the stability of the metrics (Franks, D'Amour, Cervone, and Bornn, 2016), that is, whether or not they measure the same properties over time.

From Table 3, it is evident that there is a tendency for the same players to be ranked similarly in two consecutive seasons. This is especially true for the cases of the closeness measure, C_C^{w1}, and the PageRank recipient score of Network 1, PR_R^{w1}, as seen from the high correlation across seasons for these measures. As players do develop their abilities across seasons, and their scores depend upon the team's key player for a given season, the correlation for the centrality measures seems to be reasonable. Except for the clustering coefficient, c^{w1}, all metrics have correlation coefficients that are different from 0 given any reasonable level of statistical significance. The clustering coefficient, c^{w1}, has a rather low correlation between seasons, indicating that this metric does not provide consistent scores for players across seasons. In fact, the correlation coefficient for c^{w1} is not significantly different from 0 for the comparison between the 2015 and 2016 seasons, when using a significance level of 5 %. This could be explained by looking into how players receive their clustering scores. If a player is connected to exactly two other players in the graph, the strength of the player (node strength) is equal to the sum of the weights in equation (6), giving a coefficient value of one. However, when the number of connections is increasing, but still covering a low percentage of the total number of players on the team, the score is much lower, but increasing with the number of added connections. Hence, players with a slight increase in involvements from one season to another might have very differing scores for these seasons, leading to a low correlation.

Table 3: Network metric correlation across seasons in Eliteserien for validation. For two consecutive seasons, only players having played on the same team in both seasons are considered.

	2014 vs 2015	2015 vs 2016	2016 vs 2017
$C_C^{w1}(i)$	0.778	0.739	0.842
$C_B^{w1}(i)$	0.494	0.434	0.326
$PR_R^{w1}(i)$	0.798	0.799	0.748
$PR_P^{w1}(i)$	0.550	0.578	0.608
$c^{w1}(i)$	0.152	0.077	0.186
$PR_R^{w2}(i)$	0.612	0.554	0.603
$PR_P^{w2}(i)$	0.670	0.696	0.625
$PR_R^{w3}(i)$	0.626	0.546	0.525
$PR_P^{w3}(i)$	0.423	0.429	0.398

The second validation approach is to look at the correlation between all metrics, based on the data from all seasons, as shown in Figure 1. The idea is to indicate whether the different metrics provide different information (Franks et al., 2016). Here, since the aim is to find key players in a team based on passing, the player rankings should to some extent be similar across the different measures.

The correlation between different network metrics in Figure 1 reveals which of the measures tend to rate players in the same order. The highest positive correlation can be found for the three PageRank scores, for both recipients and passers. Although the closeness measure has been defined such that both the passes received and the passes made are considered, the measure has a higher correlation with the PageRank passer scores compared to the PageRank recipient scores. The Barrat clustering coefficient has a slightly negative correlation to all other metrics. This could be due to the same reason as explained for the correlation across seasons; some players with little involvement are rated to be the best on this measure, but are not in the top ratings for the other measures due to few connections. All the correlation coefficients shown in Figure 1 are significantly different from 0 using any reasonable level of significance, except for the correlation coefficient between PR_R^{w1} and PR_P^{w2} which is not significantly different from 0 using a level of 5 %.

Key Players in Network 1

The key Rosenborg players in the 2017 season in accordance with the network metrics for Network 1 (pass difficulty) are shown in Table 4. Player positions are coded as GK for goalkeeper, CD for central defender, FB for full back, DM for defensive midfielder, CM for central midfielder, WI for winger, and ST for striker. The networks generated contain all players appearing for Rosenborg in the 2017 season. However, to identify key players, only players involved in at least 462 passes are listed. This cut-off is based on the Rosenborg players' average number of passes made and received per match in 2017, multiplied by six, which correspond to having played 20 % of the games in the season. There are two arguments to support such a cut-off. First, that the ratings calculated for players with fewer passes may be noisy, for example as a few lucky assists may overly influence the PageRank passer scores using network 3. Second, that key passers in a team must be expected to perform a certain number of passes.

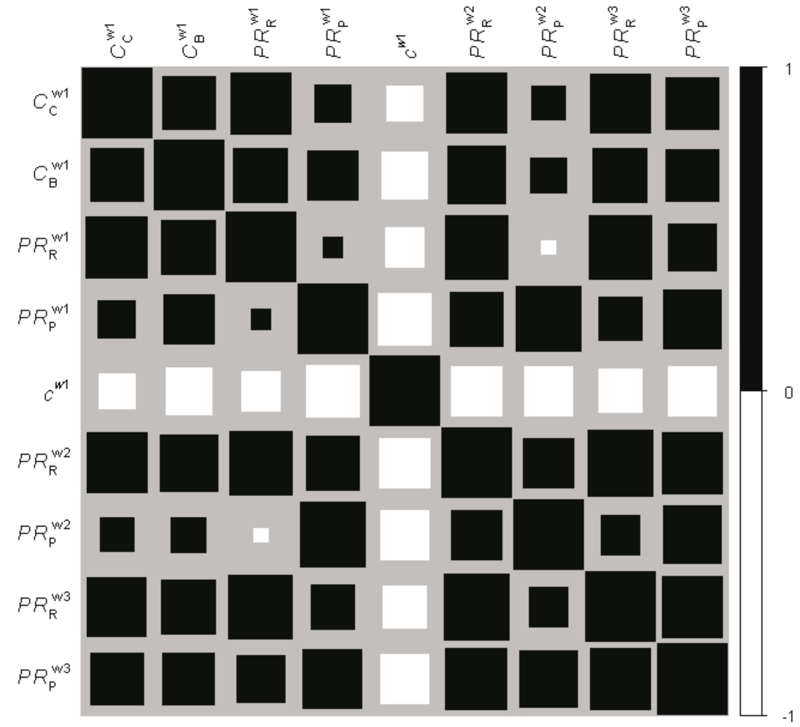


Figure 1: Correlation between the different network metrics. Players from the 2014-2017 seasons of Eliteserien are considered. Larger squares imply a bigger absolute value of the correlation coefficient, with dark squares indicating a positive correlation and white squares indicating a negative correlation.

Closeness is a measure of the easiness of reaching a player. Hence, when using pass difficulties to find the shortest paths between players, players who tend to receive and make more difficult passes will get higher scores. In general, midfielders and attacking players tend to receive high scores, which makes sense as these players are situated in positions on the pitch where passes are given higher weights. For Rosenborg however, their full backs have also received high scores, with two of them, Hedenstad and Meling, being ranked among the top three most important players on the team. This suggests that Hedenstad and Meling are important players in the attacking phase of the team's play, perhaps by making good passes to their teammates on the offensive half.

For the betweenness scores, the same three players as for the closeness measure are among the top three ranked players in Rosenborg. However, there are no clear patterns of which player positions or groups of players that are ranked higher. It seems like players are awarded both for having numerous pass involvements and for performing influential passes as seen from the differing scores. This is true for all teams in Eliteserien.

Even though he is one of the players in Rosenborg involved in the highest number of passes, the betweenness score for Reginiussen is zero, which at first glance seems to be counter-intuitive. A possible explanation for this is that he tends to attempt easier passes than other players and for this reason is not achieving weights in the graph that are large enough to be part of someone's shortest paths. By examining the average difficulty of Reginiussen's pass involvements, the explanation is supported. His average pass difficulty, when considering only passes that are included in the network, was 0.053, while for Hedenstad, for instance, this value was 0.135. When including all n pass involvements as given in the tables, the average difficulty of passes increases for both players. This is a natural development as the extra included passes were unsuccessful and thus potentially more difficult to make. However, Reginiussen's pass involvements still had a considerably lower average difficulty.

Table 4: The key players on Rosenborg in the 2017 season of Eliteserien according to Network 1 are shown. Only players involved in more passes than the equivalent of six matches are considered. The three best players according to each measure are highlighted in bold text. The number of pass involvements, n, is the sum of passes made and received by a player, where passes made also include those that were unsuccessful and thus not part of the network analysis.

Name	Pos	$C_C^{w1}(i)$	$C_B^{w1}(i)$	$PR_R^{w1}(i)$	$PR_P^{w1}(i)$	$c^{w1}(i)$	n
André Hansen	GK	0.319	0.000	0.088	0.693	0.800	1083
Johan Lædre Bjørdal	CD	0.766	0.496	0.205	**0.980**	0.842	1555
Jacob Rasmussen	CD	0.738	0.044	0.255	0.836	0.851	1174
Tore Reginiussen	CD	0.697	0.000	0.203	0.718	0.915	2259
Jørgen Skjelvik	CD	0.797	0.062	0.275	0.901	0.886	1968
Alex Gersbach	FB	0.907	0.035	0.521	0.670	0.903	854
Vegar Hedenstad	FB	**0.999**	**1.000**	0.672	**1.000**	0.849	2420
Birger Meling	FB	**0.992**	**0.646**	0.909	**0.916**	0.821	1747
Anders Konradsen	DM	0.878	0.115	0.604	0.625	**0.932**	1479
Mike Jensen	CM	0.947	0.487	0.788	0.738	0.836	1927
Marius Lundemo	CM	0.735	0.000	0.400	0.596	**0.944**	1257
Fredrik Midtsjø	CM	0.977	0.425	0.751	0.664	0.910	1119
Anders Trondsen	CM	0.910	0.257	0.566	0.771	0.914	633
Pål André Helland	WI	0.943	0.177	0.665	0.562	0.919	920
Milan Jevtovic	WI	**1.000**	**0.681**	**0.952**	0.628	**0.925**	767
Nicklas Bendtner	ST	0.950	0.133	**1.000**	0.584	0.910	1465
M. Vilhjálmsson	ST	0.881	0.310	**0.977**	0.469	0.857	596

The top three most important Rosenborg players by the PageRank recipient score are all offensive players, which is also the trend observed across all teams in Eliteserien. This is not unexpected as these players tend to receive more difficult passes due to their location on the field, which will give higher weights on the edges directed towards them. These players are also popular targets as they usually create more goal-scoring opportunities. For the PageRank passer score, the tendency both overall and for Rosenborg players is that defenders have higher scores. Although they, on average, do not attempt the most difficult passes, the defenders might be seen as important passers due to them completing more passes, many of which are between the defenders themselves so that they might boost each others importance.

The clustering coefficients are in general high for all players in Eliteserien, with players having few pass involvements being awarded with a coefficient of one. As seen from Table 4, none of the Rosenborg players who received a score of one made the cut-off, which was also the case for all other teams. With almost all players being connected on a team, the observation of consistently high scores is reasonable.

Key Players in Network 2 and Network 3

Moving on to Network 2 and Network 3, the cut-offs used for pass involvements are 308 and 42 respectively, and the computed centrality measures for Rosenborg players in 2017 are presented in Table 5. Interestingly, although not being among the most important players and by far not making the cut-off for Network 1 or Network 2, Samuel Adegbenro seems to be an important offensive contributor to the team as he makes the cut-off for Network 3. Adegbenro joined Rosenborg mid-season and has thus fewer connections with his teammates, which is the probable reason for why he is not recognised by the other networks.

The PageRank recipient scores for Network 2 identify the full backs Meling and Hedenstad and the strikers Vilhj´almsson and Bendtner as the most important players. Strikers are not surprisingly scoring well here due to their position on the pitch. In general, passes they receive are more difficult to make, and thus also more challenging to follow up. For the more defensive players, high scores on this measure might indicate that they are tactically good contributors to their teams. Across the teams in Eliteserien, the player positions among the highest rated players are differing.

As defenders tend to make easy passes between each other, it is not surprising that they also obtain the highest scores for PageRank in terms of delivering passes that are easy to follow up. Passes that are easier to make are often also easier to follow up, such that defenders will have high weights on the outgoing edges due to both higher values of p_2 and many pass attempts. In a way, this PageRank measure is thus not a good indicator of who are the best players to spot opportunities and to make tactically good passes.

Although offensive players might be thought of as being more effective, and thus should be captured by the PageRank score for effectiveness in Network 3, this is not always the case for all teams. This is probably due to the way the weights are defined. By counting involvements and not accounting for difficulty in any way, players with more pass involvements will be considered as more important. Attacking players have fewer pass involvements, but are important in the offensive play through other involvements than those that are considered in the network. Shots are mostly attempted by the attackers, meaning that they might not have been part of the sequence leading up to the attempt itself other than having received the final pass. Hence, they could receive high PageRank recipient scores, but could in principal get low rankings for the passer's score. The PageRank scores for effectiveness are thus a way of identifying the most important players in terms of frequency of offensive pass involvements, and not a way of finding the overall offensive contributor in a team. In Eliteserien, offensive players tend to dominate the top lists in terms of being recipients, while for the PageRank passer score, the player positions vary.

For Rosenborg's case, the top three most important recipients according to Network 3 represent the main outfield player positions: defender, midfielder and attacker. Bendtner, the 2017 league top scorer, is found to be the most important player. When looking at the PageRank passer scores instead, offensive players, wingers included, are ranked lower. The high scores for Meling and Hedenstad support the rationale behind their closeness score; they seem to be important players in Rosenborg's offensive play.

Case Studies

Two case studies involving Rosenborg matches from the 2017 season are presented below. In Case I, Sandefjord played against Rosenborg, with Rosenborg having a season-high ball possession of 70 %. Case II is a match between Viking and Rosenborg in which Rosenborg had a season-low ball possession of 40 % (Verdens Gang AS, 2018). Rosenborg's average ball possession in 2017 was 54.4 % (WhoScored.com, 2018).

For each case, network metrics are calculated for each player on both teams. A graphical representation of their passing networks based on Network 1 is also provided. In the networks, nodes represent the players on each team by their shirt number, and they are ordered by the starting formation of the team. The directed edges between nodes are weighted, and thicker lines indicate a stronger relationship between two players, with a stronger relationship being more passes between the players or, in general, higher difficulty on the passes observed between them.

Table 5: The key Rosenborg players in terms of Network 2 and Network 3 in the 2017 season of Eliteserien. Only players involved in more passes than the equivalent of six matches according to each of the pass perspectives are considered. The three best players according to each measure are highlighted in bold text. The number of pass involvements, n, is the sum of passes made and received by a player and that was included in the corresponding GAMM. This count also includes passes that were unsuccessful and thus not part of the network analysis.

		Risk			Potential		
Name	**Pos**	$PR_R^{w2}(i)$	$PR_P^{w2}(i)$	n	$PR_R^{w3}(i)$	$PR_P^{w3}(i)$	n
André Hansen	GK	0.398	0.402	659	0.179	0.465	50
Johan Lædre Bjørdal	CD	0.499	**0.909**	1260	0.520	0.723	106
Jacob Rasmussen	CD	0.624	**1.000**	982	0.479	0.795	70
Tore Reginiussen	CD	0.500	**0.844**	1801	0.710	0.849	154
Jørgen Skjelvik	CD	0.542	0.809	1536	0.551	0.696	125
Alex Gersbach	FB	0.527	0.542	550	0.483	0.504	54
Vegar Hedenstad	FB	**0.770**	0.641	1699	0.745	**0.994**	234
Birger Meling	FB	**1.000**	0.677	1259	**0.938**	**1.000**	149
Anders Konradsen	DM	0.660	0.559	1039	0.883	0.718	151
Mike Jensen	CM	0.715	0.486	1138	0.617	**0.871**	211
Marius Lundemo	CM	0.557	0.615	918	0.578	0.570	94
Fredrik Midtsjø	CM	0.659	0.361	694	**0.944**	0.822	171
Anders Trondsen	CM	0.666	0.497	466	0.751	0.586	55
Samuel Adegbenro				449	0.585	0.486	46
Pål André Helland	WI	0.471	0.273	432	0.786	0.623	144
Milan Jevtovic	WI	0.520	0.297	886	0.662	0.551	104
Nicklas Bendtner	ST	0.738	0.365	380	**1.000**	0.708	198
M. Vilhjálmsson	ST	**0.744**	0.319	659	0.779	0.599	95

Case I: Sandefjord versus Rosenborg

The match between Sandefjord and Rosenborg was played on the 5th of April 2017, with Rosenborg winning with three goals against zero. Graphical representations of the teams' passing networks are depicted in Figure 2, while the computed network metrics for players on both teams are tabulated in Table 6 and Table 7.

The much higher percentage of ball possession of Rosenborg is evident from the thicker directed edges between their players for Network 1 as shown in Figure 2. In general, the connections between Rosenborg players are stronger than that of Sandefjord players. For Rosenborg, there are no clear pattern in the passing network, while for Sandefjord, strong connections exist between offensive players on the left-hand side.

For Network 1, Rosenborg players' closeness scores are on average higher than the closeness scores of Sandefjord players, which is intuitive when taking the differences in ball possession into account. With high closeness scores, players are more easily reached and the ball is more readily kept within the team. Also, the betweenness scores of Rosenborg players are more evenly distributed across the team compared to the case of Sandefjord, which may indicate that Sandefjord is more dependent on certain players on the team to keep possession of the ball. These observations are somewhat supported by the teams' passing networks.

Considering the PageRank recipient scores for Network 1, the tendency is that offensive players achieve the highest scores, which is reasonable. Different patterns are seen for the PageRank passer scores for the two teams. In the case of Rosenborg, defensive players are achieving the highest scores, while central midfielders have the highest scores for Sandefjord. These differences could be due to the fact that both teams have a set playing style. In a 4-3-3 formation, Rosenborg wants to play out from the back and put pressure on the opponent team on the offensive half when the opponent team is established offensively and defensively respectively. For Sandefjord, the central midfielders are crucial players both in the attacking and defensive phases of the play in their 3-5-2 formation (Sandefjord Fotball, 2017).

For both teams, the clustering coefficients are relatively high for all players. Rosenborg's average clustering coefficient for the match (0.780) is slightly higher than the team coefficient of Sandefjord (0.779) when considering only players who played at least 20% of the regular game time. These numbers are lower than the corresponding seasonal average team coefficients given in Table 8. This might be due to the fact that the match was played at the very beginning of the season. At the start of a season, it has been a long period of time since most teams have played proper matches. Hence, the players might need some time to adjust to match situations and are thus not playing their best together just yet.

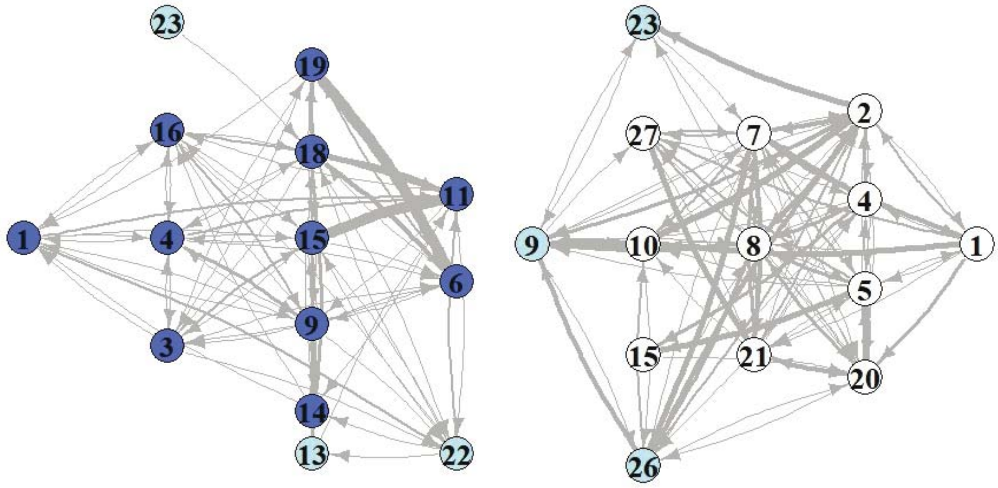
Moving on to the PageRank recipient and passer scores for Network 2, defensive players are recognised with the highest scores for Rosenborg, while central midfielders are the highest rated players in the case of Sandefjord. The importance of the players in these specific player positions seems to be related to the teams' playing styles as for the case of the PageRank passer measure for Network 1. For Network 3, the distribution of PageRank scores indicates well which players are more involved when shots are attempted as players that have not been involved in sequences leading to shots do not receive a score. The results are intuitive and the passer scores also support the findings that defenders and midfielders seem to be important players in the attacking play for Rosenborg and Sandefjord respectively.

Table 6: The overall key figures according to Network 1 for both teams in Case I. The first column for each team indicate the players' jersey numbers.

		Sandefjord							Rosenborg				
No.	**Name**	C_C^{w1}	C_B^{w1}	PR_R^{w1}	PR_P^{w1}	c^{w1}	**No.**	**Name**	C_C^{w1}	C_B^{w1}	PR_R^{w1}	PR_P^{w1}	c^{w1}
1	I. Jónsson	0.220	0.000	0.091	0.356	0.765	1	A. Hansen	0.219	0.000	0.084	0.557	0.836
3	A. Seck	0.552	0.000	0.151	0.373	0.782	2	V. Hedenstad	0.961	0.769	0.688	0.679	0.667
4	C. Hansen	0.602	0.000	0.102	0.497	0.801	4	T. Reginiussen	0.618	0.308	0.123	0.760	0.665
6	P. Morer	0.799	0.177	0.701	0.349	0.773	5	J. Rasmussen	0.875	0.872	0.260	1.000	0.636
9	H. Storbæk	0.987	0.710	0.833	0.939	0.787	7	M. Jensen	0.955	0.436	0.901	0.393	0.742
11	F. Kastrati	0.991	0.774	1.000	0.380	0.812	8	A. Konradsen	1.000	0.923	0.713	0.648	0.803
13	Naglestad	0.712	0.000	0.222	0.089	1.000	9	N. Bendtner	0.774	0.051	0.688	0.191	0.712
14	E. Kebbie	0.857	0.032	0.466	0.426	0.971	10	Vilhjálmsson	0.795	0.154	1.000	0.161	0.795
15	E. Vallés	1.000	0.516	0.526	1.000	0.728	15	E. Rashani	0.666	0.026	0.148	0.115	1.000
16	E. Kane	0.868	0.210	0.560	0.525	0.942	20	A. Gersbach	0.934	1.000	0.666	0.642	0.664
18	W. Kurtovic	0.993	1.000	0.725	0.835	0.475	21	F. Midtsjø	0.936	0.359	0.747	0.286	0.874
19	V. Bindia	0.785	0.000	0.521	0.303	0.763	23	P. Helland	0.732	0.051	0.227	0.113	0.936
22	A. Sødlund	0.814	0.016	0.383	0.245	0.743	26	M. Jevtovic	0.944	0.692	0.841	0.323	0.894
23	M. Holt	0.394	0.000	0.068	0.109	NA	27	M. Bakenga	0.855	0.000	0.587	0.180	0.915

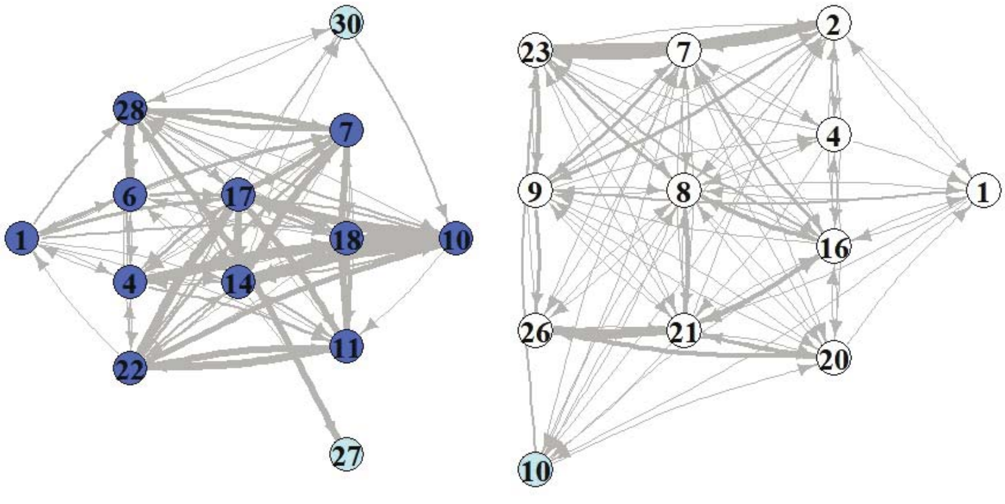
Table 7: The PageRank scores for Network 2 and Network 3 for players on both teams in Case I. The first column for each team indicates the players' jersey numbers. Players not participating in a passing sequence leading to a shot do not have PageRank scores for Network 3.

		Sandefjord						Rosenborg			
No.	**Name**	PR_R^{w2}	PR_P^{w2}	PR_R^{w3}	PR_P^{w3}	**No.**	**Name**	PR_R^{w2}	PR_P^{w2}	PR_R^{w3}	PR_P^{w3}
1	I. Jónsson	0.303	0.327			1	A. Hansen	0.295	0.257		
3	A. Seck	0.213	0.313			2	V. Hedenstad	1.000	0.541	1.000	1.000
4	C. Hansen	0.320	0.497	0.389	0.826	4	T. Reginiussen	0.459	0.685	0.395	0.653
6	P. Morer	0.562	0.340	1.000	0.684	5	J. Rasmussen	0.756	1.000	0.196	0.254
9	H. Storbæk	0.861	0.768	0.144	0.536	7	M. Jensen	0.643	0.233	0.631	0.580
11	F. Kastrati	0.446	0.137	0.061	0.464	8	A. Konradsen	0.966	0.639	0.668	0.798
13	Naglestad	0.294	0.062	0.093	0.236	9	N. Bendtner	0.660	0.150	0.730	0.421
14	E. Kebbie	0.355	0.094			10	Vilhjálmsson	0.665	0.155	0.715	0.525
15	E. Vallés	0.852	1.000	0.061	1.000	15	E. Rashani	0.119	0.050	0.157	0.077
16	E. Kane	0.395	0.390	0.348	0.704	20	A. Gersbach	0.443	0.421		
18	W. Kurtovic	1.000	0.521	0.112	0.899	21	F. Midtsjø	0.479	0.167	0.460	0.642
19	V. Bindia	0.549	0.412	0.911	0.527	23	P. Helland	0.278	0.088	0.302	0.166
22	A. Sødlund	0.518	0.356			26	M. Jevtovic	0.513	0.138	0.577	0.386
23	M. Holt	0.090	0.086			27	M. Bakenga	0.345	0.124	0.452	0.627



Panel A: Sandefjord (3-5-2) Panel B: Rosenborg (4-3-3)

Figure 2: Graphical representations of the passing networks for the teams in Case I. Nodes are placed with respect to the starting eleven for each team, coloured based on the team's jersey colour and given names based on the players' jersey numbers. Substitutes are depicted with a separate colour (light blue). The directed edges are weighted based on the difficulty of passes between players.



Panel A: Viking (4-2-3-1) Panel B: Rosenborg (4-3-3)

Figure 3: Graphical representations of the passing networks for the teams in Case II. Nodes are placed with respect to the starting eleven for each team, coloured based on the team's jersey colour and given names based on the players' jersey numbers. Substitutes are depicted with a separate colour (light blue). The directed edges are weighted based on the difficulty of passes between players.

Case II: Viking versus Rosenborg

The match between Viking and Rosenborg was played on April 17th 2017, and Rosenborg won with the final score 0-1. In Figure 3, a graphical representation of each team's passing network for Network 1 is shown, and in Table 9 and Table 10, the calculated network metrics are presented.

Table 8: The average team clustering coefficient for each team in the 2017 season of Eliteserien. Only players with pass involvements above the equivalent of six matches are considered for each team. For the entire league, this number corresponds to 364 pass involvements.

Team	$\bar{c}_i^{w1}$
Aalesund	0.906
Brann	0.906
Odd	0.886
Rosenborg	0.883
Sarpsborg 08	0.864
Haugesund	0.860
Lillestrøm	0.854
Strømsgodset	0.847
V°alerenga	0.844
Tromsø	0.842
Molde	0.833
Sandefjord	0.824
Viking	0.807
Stabæk	0.805
Kristiansund	0.797
Sogndal	0.761

As expected, when taking the difference in ball possession into account, the strength of the connections between Viking players is stronger than for the case of Rosenborg players in the passing networks. Interestingly, the connections between Rosenborg's full back, central midfielder and winger on the right-hand side stand out to be strong. Thus, it seems like Rosenborg is more dependent on the players on this side when their ball possession is low, compared to the situation when they are dominating the play, as they did in Case I.

Considering the closeness scores of Rosenborg players in Network 1, many of the same top rated players as for Case I are recognised. Thus, these players seem to be central for Rosenborg independent of the game development and the distribution of ball possession between the playing teams. Moreover, the players achieving the highest closeness scores are also achieving high betweenness scores. Thus, these players are easy to reach and have a high involvement in the game. Compared to Case I, more Rosenborg players have a betweenness score of zero. It seems like higher ball possession is related to having fewer betweenness scores of zero, which intuitively makes sense as a low betweenness score is associated with less involvement in the match.

Similar to Case I, Rosenborg's offensive players dominate the highest ratings according to the PageRank recipient scores for Network 1, while more defensive players achieve the highest PageRank passer scores. For Viking however, the PageRank recipient scores tend to be higher for players playing on the left-hand side of the pitch, while the PageRank passer scores are higher for players playing in positions central on the field. Viking's clustering coefficient of 0.849 is higher than their average coefficient for the entire season which is tabulated in Table 8, but it is slightly lower than Rosenborg's coefficient from the match (0.852). Although Viking performed poorly in 2017 and were relegated from Eliteserien, they apparently played a good match against Rosenborg, which might explain the difference in Viking's clustering coefficients.

Table 9: The overall key figures for both teams in Case II. The first column for each team indicate the players' jersey numbers.

Viking							Rosenborg						
No.	**Name**	C_C^{w1}	C_B^{w1}	PR_R^{w1}	PR_P^{w1}	c^{w1}	**No.**	**Name**	C_C^{w1}	C_B^{w1}	PR_R^{w1}	PR_P^{w1}	c^{w1}
1	I. Austbø	0.264	0.000	0.085	0.644	0.884	1	A. Hansen	0.208	0.000	0.081	0.405	0.915
4	M. Ledger	0.604	0.233	0.146	0.829	0.818	2	V. Hedenstad	1.000	1.000	0.417	1.000	0.881
6	K. Mets	0.667	0.000	0.138	0.736	0.982	4	T. Reginiussen	0.766	0.000	0.140	0.709	0.886
7	S. Adegbenro	0.958	0.488	0.645	0.510	0.913	7	M. Jensen	0.989	0.857	0.642	0.811	0.843
10	K. Appiah	0.912	0.442	0.635	0.441	0.897	8	A. Konradsen	0.943	0.393	0.503	0.932	0.842
11	Z. Bytyqi	0.971	0.744	0.606	0.656	0.951	9	N. Bendtner	0.901	0.357	0.896	0.317	0.763
14	A. Danielsen	0.918	0.093	0.440	1.000	0.742	10	M. Vilhjálmsson	0.681	0.000	0.143	0.284	0.823
17	S. Ernemann	1.000	1.000	1.000	0.665	0.827	16	J. Skjelvik	0.835	0.000	0.201	0.986	0.918
18	J. Ryerson	0.751	0.070	0.273	0.752	0.847	20	A. Gersbach	0.874	0.107	0.590	0.453	0.780
22	C. Kronberg	0.900	0.140	0.454	0.740	0.873	21	F. Midtsjø	0.997	0.571	0.688	0.672	0.777
27	M. Bringaker	0.472	0.000	0.291	0.091	1.000	23	P. Helland	0.963	0.357	1.000	0.349	0.858
28	K. Haugen	0.924	0.512	0.649	0.647	0.610	26	M. Jevtovic	0.894	0.214	0.957	0.270	0.936
30	S. Michalsen	0.481	0.000	0.125	0.208	1.000							

Table 10: The PageRank scores for Network 2 and Network 3 for players on both teams in Case II. The first column for each team indicate the players' jersey numbers. Players not participating in a passing sequence leading to a shot do not have PageRank scores for Network 3.

Viking						Rosenborg					
No.	**Name**	PR_R^{w2}	PR_P^{w2}	PR_R^{w3}	PR_P^{w3}	**No.**	**Name**	PR_R^{w2}	PR_P^{w2}	PR_R^{w3}	PR_P^{w3}
1	I. Austbø	0.241	0.440			1	A. Hansen	0.377	0.310		
4	M. Ledger	0.351	0.529	0.432	0.473	2	V. Hedenstad	0.929	0.626	0.380	0.751
6	K. Mets	0.352	0.834	0.265	0.320	4	T. Reginiussen	0.520	0.620		
7	S. Adegbenro	0.527	0.327	0.426	0.374	7	M. Jensen	1.000	0.469	0.547	0.531
10	K. Appiah	0.673	0.329	0.374	0.301	8	A. Konradsen	0.870	0.777	0.337	0.446
11	Z. Bytyqi	0.673	0.352	0.145	1.000	9	N. Bendtner	0.730	0.477	0.587	1.000
14	A. Danielsen	0.754	1.000	0.081	0.675	10	M. Vilhjálmsson	0.294	0.202	0.484	0.953
17	S. Ernemann	1.000	0.734	0.603	0.320	16	J. Skjelvik	0.771	1.000		
18	J. Ryerson	0.533	0.777	0.081	0.265	20	A. Gersbach	0.663	0.489	0.341	0.596
22	C. Kronberg	0.650	0.523	0.143	0.828	21	F. Midtsjø	0.828	0.503	0.724	0.841
27	M. Bringaker	0.152	0.076	0.388	0.138	23	P. Helland	0.897	0.312	1.000	0.089
28	K. Haugen	0.540	0.554	1.000	0.643	26	M. Jevtovic	0.688	0.263	0.425	0.107
30	S. Michalsen	0.149	0.112	0.365	0.339						

The importance of the players on Rosenborg's right-hand side is seemingly captured by the PageRank recipient scores for Network 2. For Viking, two midfielders are achieving the highest ratings for this measure. Considering the PageRank passer scores for Network 2, the same patterns as seen for Network 1 are present. In general, the highest PageRank scores for Rosenborg players in Network 3 are dominated by offensive players, while more defensive players have high scores for Viking. Compared to Case I, where more defensive players on Rosenborg received higher PageRank passer scores, it seems like the defenders have had a lower offensive contribution in this match. As these players appear to be important for Rosenborg, this might explain the team's lower ball possession in the game.

It is also interesting to see the difference in scores of the two strikers for the measures of Networks 2 and 3. The Rosenborg striker has high values, which is in line with the season averages of all strikers in Eliteserien, while the striker of Viking has uncharacteristically low scores. This may indicate that Viking had trouble involving him in the game, both as a passer and as a recivever, perhaps because the central defenders of Rosenborg were doing a good job of keeping him out of the game. Looking at Figure 3 we can see that not a single successfull pass was made from either winger of Viking and to their striker.

Comparison with Existing Literature

In this work, the key players on teams in Eliteserien are found through network analyses with the majority of the edge weights being based on the results from the predicted probabilities of passes' success. The idea of using pass difficulties as weights instead of the number of successful passes between players was introduced by McHale and Relton (2018). They only consider the accuracy of passes however, whereas two more aspects explaining the success of passes are investigated in separate networks here. They also considered different types of network metrics, and used a different data set. While player scores across leagues cannot be compared directly, due to different teams, players, and playing styles, general trends seen for player positions can give some insights. McHale and Relton (2018) identified mostly attacking players and midfielders in the top lists for teams playing in the English Premier League when calculating the exponential centrality, which is also the case for the closeness scores of Eliteserien players in Network 1.

Using the number of passes between players as weights, Pena and Touchette (2012) evaluated players' individual contributions to teams by calculating the closeness, betweenness, and PageRank centrality measures as well as the Onnela clustering coefficient. In general, high clustering coefficients are observed across the teams, an observation that is supported by the results for Network 1. Also, the betweenness scores seem to vary a lot, with no clear patterns apparent for the different player positions. Other than this, no obvious similarities are observed, which could be due to the fact that the scores given in (Pena and Touchette, 2012) are based on players' performance from one single match, giving a poor basis for comparisons between the two studies.

Rojas-Mora et al. (2017) investigated three matches from the group stage of Copa America and calculated the PageRank scores for all players on the field. Although the comparison is made on limited data, the PageRank recipient scores for Network 1 and the PageRank scores for players in Copa America both indicate that players playing in more offensive positions on the pitch are more important to their teams.

By only considering offensive sequences that ended with shots in their network analysis, Peixoto et al. (2017) had a similar approach to Network 3. In-degree and out-degree centrality measures, which are linked to passes received and passes made, respectively, were calculated and revealed that strikers and midfielders scored highest on the respective measures. The PageRank scores for Network 3 are likewise linked to either the passer or the recipient of a pass. The PageRank recipient scores support the finding that offensive players are more important, whereas the PageRank passer scores did not have a consistent pattern of player positions among the key players.

The network analyses shown in this work indicate that the formation of a team can reveal information about which of the team's players the play is centred around. By using such information in the pre-game analyses, a team can accommodate their game plan according to the strengths and weaknesses of the opponent team. As the key players found for the teams are not only based on the number of passes between players, but also the difficulty of the passes, the opponent players that make the smartest passing alternatives may be identified and actions to stop them can be taken.

Concluding remarks

In this paper we have used network analysis to attempt to identify key players on a team. Each team is modelled as a network where the players are represented by nodes and the edges connecting them represent the interactions between the players. The edges are weighted based on three different criteria: the difficulty of passes performed between players, the risk of the passes, and the potential of the passes to result in shots.

The main finding is that setting network weights according to different criteria lead to the identification of different players as key passers and/or pass recipients. For instance, some players are more central in terms of passing the ball so that the sequence of play ends with a shot, whereas other players may be central in terms of being able to make passes that avoid losing possession. It is possible to use the network analysis over a full season of matches, but also in individual matches. For the latter, it may be possible to determine that a team uses a different pattern of passing depending on the type of opposition.

In addition to using different network weights, four different centrality measures were considered for the different networks. These measures can, for instance, provide information about which players are more involved in the play, which players tend to be most important playmakers, and how well-balanced a team is. Such information may be beneficial for a team when deciding upon their game plan. Results also showed that the positions of the key players vary among the measures. Offensive players tend to be ranked higher on the closeness centrality and the PageRank measure for recipients, while defenders, with many easy passes between each other, tend to be ranked higher on the PageRank measure for passers.

The methodology examined in this paper may provide a good starting point for analyzing contributions to passing from players within a professional team. It is possible to apply the network analysis to any league where sufficient event-level data are available. Based on comparisons with existing studies from the literature, some general findings seem to carry over to other leagues, such as the fact that players from certain positions are favoured by some of the metrics.

The current approach is not without limitations, though. One of the measures examined proved to give unreasonable results when attempting to find key players in teams: The PageRank passer scores for the passing risk network favoured defenders and would not recognise the true tactical pass makers from the Norwegian top division. Hence, the weights used for this network should be reconsidered to get reliable results for the passing aspect that is explored.

Furthermore, it remains unclear whether the player scores calculated can be used to assess player transfers, as scores are relative to the current set of teammates. One direction for future work could be to investigate how the player scores can be used to predict the performance of players after moving to new teams, even in the case that this involves moving to a different league system. It is also unknown how the scores of a player may change depending on the role that the player has within a team. To provide an example, defenders and midfielders may have quite different roles in the passing network, but some players can be used in either position. Such players should perhaps be evaluated by separate scores for each position.

This study has used three different generalised additive mixed models (GAMMs) to evaluate each pass made. While these GAMMs are state-of-the-art in terms of evaluating proper ties of the passes made, their effectiveness could influence the results of the network analysis. As the GAMMs used, derived from (H°aland et al., 2019), do not utilize information about the locations of opposing players, it may be conjectured that more precise evaluations of passes can be obtained by using player tracking data, as in (McHale and Relton, 2018). However, it remains an open question whether this increased accuracy in evaluations will have a substantial effect on the analysis of passing networks.

Acknowledgements

The authors wish to thank two anonymous reviewers for their useful comments, which helped to improve this paper. Opta Sports provided access to relevant data from Eliteserien, without which this work could not have been performed. Association football coach Hugo Pereira is thanked for his input and feedback.

References

Arriaza-Ardiles, E., Martín-González, J., Zuniga, M., Sánchez-Flores, J., de Saa, Y., & García-Manso, J. (2018). Applying graphs and complex networks to football metric interpretation. *Human Movement Science*, 57, 236–243.

Barrat, A., Barthelemy, M., & Vespignani, A. (2007). The architecture of complex weighted networks: Measurements and models. In: Caldarelli, G., & Vespignani, A., eds., Large Scale Structure And Dynamics Of Complex Networks: From Information Technology to Finance and Natural Science, World Scientific, 67–92.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424, 175–308.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25, 163–177.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30, 107–117.

Clemente, F., Martins, F., & Mendes, R. (2015). There are differences between centrality levels of volleyball players in different competitive levels? *Journal of Physical Education and Sport*, 15, 272.

Clemente, F., Martins, F., & Mendes, R. (2016). Social network analysis applied to team sports analysis, Netherlands: Springer International Publishing.

Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695, 1–9.

Dey, P., Ganguly, M., & Roy, S. (2017). Network centrality based team formation: A case study on T-20 cricket. *Applied Computing and Informatics*, 13, 161–168.

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, 269–271.

Duch, J., Waitzman, J., & Amaral, L. (2010). Quantifying the performance of individual players in a team activity. *PloS One*, 5, e10937.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874.

Fewell, J., Armbruster, D., Ingraham, J., Petersen, A., & Waters, J. (2012). Basketball teams as strategic networks. *PloS One*, 7, e47445.

Franks, A., D'Amour, A., Cervone, D., & Bornn, L. (2016). Meta-analytics: tools for understanding the statistical properties of sports metrics. *Journal of Quantitative Analysis in Sports*, 12, 151–165.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35–41.

Freeman, L. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1, 215–239.

Fu, H.-H., Lin, D., & Tsai, H.-T. (2006). Damping factor in Google page ranking. *Applied Stochastic Models in Business and Industry*, 22, 431–444.

Gama, J., Passos, P., Davids, K., Relvas, H., Ribeiro, J., Vaz, V., & Dias, G. (2014). Network analysis and intra-team activity in attacking phases of professional football. *International Journal of Performance Analysis in Sport*, 14, 692–708.

Gonçalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jiménez, S., & Sampaio, J. (2017). Exploring team passing networks and player movement dynamics in youth association football. *PloS One*, 12, e0171156.

Håland, E., Wiig, A., Stålhane, M., & Hvattum, L. (2019). Evaluating passing ability in association football. *IMA Journal of Management Mathematics*, forthcoming.

Kang, B., Huh, M., & Choi, S. (2015). Performance analysis of volleyball games using the social network and text mining techniques. *Journal of the Korean Data and Information Science Society*, 26, 619–630.

Lazova, V. & Basnarkov, L. (2015). PageRank approach to ranking national football teams. arXiv preprint arXiv:1503.01331.

Liu, X.F., Liu, Y.-L., Lu, X.-H., Wang, Q.-X., & Wang, T.-X. (2016). The Anatomy of the Global Football Player Transfer Network: Club Functionalities versus Network Properties. *PLoS ONE* 11: e0156504.

McHale, I. & Relton, S. (2018). Identifying key players in soccer teams using network analysis and pass difficulty. *European Journal of Operational Research*, 268, 339–347.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32, 245–251.

Opta Sports (2018). World leaders in sports data. https://www.optasports.com/, accessed on 13/4/2018.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Peixoto, D., Praça, G., Bredt, S., & Clemente, F (2017). Comparison of network processes between successful and unsuccessful offensive sequences in elite soccer. *Human Movement*, 18, 48–54.

Pena, J. & Touchette, H. (2012). A network theory analysis of football strategies. arXiv preprint arXiv:1206.6904.

Piette, J., Anand, S., & Pham, L. (2011). Evaluating basketball player performance via statistical network modeling. In: MIT Sloan Sports Analytics Conference.

Pina, T., Paulo, A., & Araújo, D. (2017). Network characteristics of successful performance in association football. A study on the UEFA champions league. *Frontiers in Psychology*, 8, 1173.

Rojas-Mora, J., Chávez-Bustamante, F., del Río-Andrade, J., & Medina-Valdebenito, N. (2017). A methodology for the analysis of soccer matches based on pagerank centrality. In: Peris-Ortiz, M., Álvarez-García, J., & Del Río Rama, M., eds., Sports Management as an Emerging Economic Activity, Springer, 257–272.

Sandefjord Fotball (2017): "Sportsplan," https://drive.google.com/file/d/0B9wYsNKQFBUFMkRpejFDaFM3OFk/, (accessed on 10/04/2018).

Szczepański, Ł. & McHale, I. (2016). Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179, 513–533.

Verdens Gang AS (2018): "VG LIVE," URL https://vglive.no/.

WhoScored.com (2018): "Whoscored.com," URL https://www.whoscored.com/.

Wood, S. (2006): Generalized additive models: an introduction with R, Boca Raton, Florida: Chapman and Hall/CRC.