New Indicator for Centrality Measurements in Passing-network Analysis of Soccer

Masatoshi Kanbata, Ryohei Orihara, Yuichi Sei, Yasuyuki Tahara and Akihiko Ohsuga
Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan

Keywords: Big Data, Data Mining, Graph Mining, Graph Network, Soccer.

Abstract: A number of fields including business, science, and sports, make use of data analytics. The evaluation of players and teams affect how tactics, training, and scouting are conducted in soccer teams. Data such as the number of shots and goals in match results are often used to evaluate players and teams. However, this is not enough to fully understand the potential of the players and teams. In this paper, we describe a new analysis method using passing-distribution data from soccer games. To evaluate the performance of players and teams, we applied graph mining. We also used an index called centrality, which evaluates individual contributions with an organization. In this research, we propose a new centrality model to improve existing conventional models. In the calculating the centrality of a given player pair, we consider not only the shortest sequence of passing but also longer ones. In this research, we verified the significance of these indicators by applying the data of UEFA EURO 2008, 2012, and 2016. As a result, we found our method to be more consistent with game results than conventional methods.

1 INTRODUCTION

In recent years, data analysis has been widely used, and it has enriched our lives. Among its uses in sports, data analysis is actively performed during games. This analysis has the influence on team management, such as the evaluation of teams and players, team tactics, training of players, and scouting of new players. Another benefit of data analysis is its ability to provide an alternative viewpoint for sports audiences. For example, in baseball, Saber Metrics conducts an objective analysis from a statistical viewpoint, which has an influence on players' evaluations and tactics (Beneventano et al., 2012).

However, conducting an objective analysis in soccer is more complicated. Team play is important in soccer, and ball possession in a soccer game frequently switches between the offensive and the defensive side. Furthermore, players and teams are often evaluated by data such as the number of assists and goals in a game's results, and the players with these statistics in the game tend to be highly evaluated. For this reason, we examine not only the players who scored goals and assists but also the players involved in goals and assists. In addition, by evaluating players and teams through a more appropriate analysis method, this research can contribute to the team's performances using a logical analysis of tactics, evaluation of player's characteristics and abilities.

In this research, we focus on passes that are considered to lead to goals and assists in soccer, and we evaluate the players based on them. In particular, by applying the graph theory, we obtain the sequence of the passes that ended with the shots. We apply the concept of centrality in the graph theory to propose a new evaluation model that improves current conventional models. At the moment, there is no objective method to show if a criterion of an individual player's evaluation is sound. Therefore, we evaluate the soundness of our model by calculating the accuracy of team evaluations based on the model. The accuracy of the team's evaluation can be determined by comparing it to actual game results.

Section 2 introduces conventional research applications of the graph theory to soccer and the conventional method of centrality. In Section 3, we will explain a new centrality model that will improve the conventional method of centrality. In Section 4, we explain the details of the data and the experimental method used to verify the usefulness of the new centrality model. We also present the results of our experiment in this section. Section 5 discusses the results.

Kanbata, M., Orihara, R., Sei, Y., Tahara, Y. and Ohsuga, A.
New Indicator for Centrality Measurements in Passing-network Analysis of Soccer.
DOI: 10.5220/0007377506160623
In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence (ICAART 2019)*, pages 616-623
ISBN: 978-989-758-350-6

2 RELATED WORK

Analysis related to sports have been widely studied, among them, the application of the graph theory to passes in a soccer game has been studied. When conducting this research, data called passing distribution, which records the number of passes as well as the passer and receiver of the pass in a game, are often publicly available.

In soccer, a pass that moves the ball to the opponent's goal is considered to be an important offensive move. Therefore, the analysis to create a graph network based on passing distribution has been done by other researchers. (Goncalves et al., 2017) conducted a tactical analysis that considers the player's positionings and the passes between players using passing distribution. They evaluated the performance of teams participating in international youth games [1] and argued that players and tactical features are different when comparing the two. (Cotta et al., 2013) analyzed the pass tactics of the Spanish national soccer team of the 2010 FIFA World Cup. They investigated the games from the quarter-finals to the final, taking the transition of the number of passes and player's positioning into account.

In addition, (Duch et al., 2010) measured centrality as a method to evaluate the players and the teams in games by creating a graph network from passing distribution. Furthermore, they modified betweenness centrality used for the centrality measurement and proposed flow centrality. This is a centrality measurement using data on only the sequences of passes leading to a shot. They measured flow centrality from the passing-distribution data of EURO 2008. They also evaluated each team by simply averaging players' evaluations. From the results, they argued that the team evaluation based on their method reflects the game's results better than one based on betweenness centrality.

Indicators such as betweenness centrality and flow centrality are used for centrality measurements. Flow centrality is a method that measures centrality using data on only the sequences of passes leading to a shot and highly evaluates athletes who participated in the scoring opportunity. However, if it is a small number of sequences in some games, flow centrality may yields inaccurate measurements.

In this research, we propose a new model to alleviate this problem by improving flow centrality. We also evaluate teams based on the model to show the soundness of the model. Let us emphasize that passing distribution only records the number of passes between players and cannot reproduce sequences of passes with three or more players. However, realizing the tendency of the sequence from a large number of passes is possible. Therefore, we will apply and verify this data in our research.

[1]The teams have players under 15 or 17 years of age.

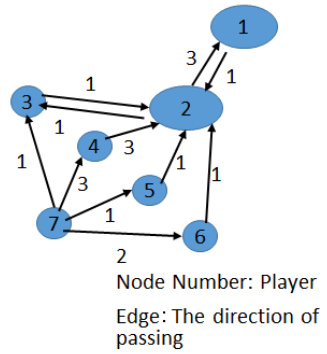

Figure 1: Graph Network Example.

2.1 Betweenness Centrality

Centrality is an index of the degree of influence that any node gives to other nodes in a graph network (Tsugawa and Ohsaki, 2014). This index makes it possible to estimate an important individual in an organization. Because soccer places more emphasis on organizational team play, applying the index would be appropriate.

Creating a graph network using nodes, edges, and edge weights is a feasible endeavor. Figure 1 shows an actual graph network example. When it is applied to a soccer match, the node is the player, the edge is the direction of the pass, and the weight of the edge is the number of passes.

When this concept is applied to the graph theory, it is possible to measure the centrality of the player node by calculating how much the player node involved in the sequence of a specific pass from the graph was created by passing distribution. In soccer, there is a consensus that the more passes a player is involved in, the more contribution to a match she/he makes. We accept this idea and conduct our study based on it.

Betweenness centrality is calculated by counting the number of times a node appears in one or more shortest sequences of passes between nodes. The weights of the sequence are the minimum value of the edge weights. In the example of Figure 1, the shortest sequence from node 7 to node 1 is $7 \rightarrow 3 \rightarrow 2 \rightarrow 1$, $7 \rightarrow 4 \rightarrow 2 \rightarrow 1$, $7 \rightarrow 5 \rightarrow 2 \rightarrow 1$, and $7 \rightarrow 6 \rightarrow 2 \rightarrow 1$. The edge weights of $7 \rightarrow 3 \rightarrow 2 \rightarrow 1$, $7 \rightarrow 5 \rightarrow 2 \rightarrow 1$, and $7 \rightarrow 6 \rightarrow 2 \rightarrow 1$ are one. In addition, $7 \rightarrow 4 \rightarrow 2 \rightarrow 1$ has the edge weights of three, and the number of passes of the shortest sequence is three. Therefore, the number of passes in the shortest sequence from

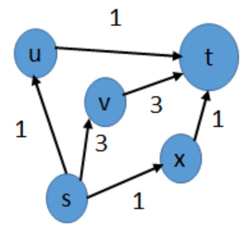

Figure 2: Example of Betweenness Centrality's Graph Network.

node 7 to node 1 is six.

Let σ_{st} be the number of passes in the shortest sequence from node s to node t. Let $\sigma_{st}(v)$ be the number of passes through node v in the shortest sequence from node s to node t. Betweenness centrality representing the centrality of the nodes is defined as follows.

$$betweenness(v) = \sum_{\substack{s,t \in V \\ s \neq v \neq t}} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

Where, V = {1,2, ..., 11 } denotes a set of player nodes (Barthelemy, 2004). For example, the shortest sequence from node s to node t in Figure 2 is $s \to u \to t$, $s \to v \to t$, $s \to x \to t$. In this case, to calculate $\sigma_{st}(v)/\sigma_{st}$, the number of edge weights of $s \to u \to t$ is one, the number of edge weights of $s \to v \to t$ is three, and the number of edge weights of $s \to x \to t$ is one. Therefore, we can write equation (2) as follows:

$$\begin{aligned} \frac{\sigma_{st}(v)}{\sigma_{st}} &= 3/5 \\ betweenness(v) &= \frac{\sigma_{st}(v)}{\sigma_{st}} + \frac{\sigma_{su}(v)}{\sigma_{su}} + \frac{\sigma_{ut}(v)}{\sigma_{ut}} + \frac{\sigma_{sx}(v)}{\sigma_{sx}} \\ &+ \frac{\sigma_{xt}(v)}{\sigma_{xt}} \\ &= 3/5 + 0 + 0 + 0 + 0 = 3/5 \end{aligned} \tag{2}$$

2.2 Flow Centrality

Betweenness centrality is a method to measure centrality from the shortest sequence between nodes, but when you applied it to soccer, the scoring opportunity to measure centrality between the player nodes is unclear.

Here (Duch et al., 2010) improved betweenness centrality. They only treat the process of the shortest sequence leading to the shots. Specifically, they newly create the shot node in the graph network and divide it into a shot on goal and a shot wide, and the direction of the player's shot is the edge, and the number of shots is the weight of the edge (Duch et al., 2010). Figure 3 exemplifies this. For this reason, they evaluate the players by considering the opportunity to score directly. They define flow centrality as follows.

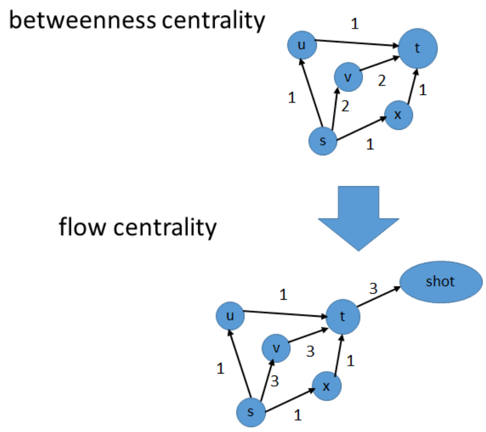

Figure 3: Changes in Graph Network from Betweenness Centrality to Flow Centrality.

$$flow(v) = \sum_{\substack{s \in V \\ t \in U \\ s \neq v}} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3}$$

Where, $U = \{12, 13\}$ is the set of shot nodes and the shot wide node is 12 and the shot on goal node is 13. In Figure 3, to calculate $flow(u)$, we can write equation (4) as follows:

$$\begin{aligned} flow(u) &= \frac{\sigma_{s\ shot}(u)}{\sigma_{s\ shot}} + \frac{\sigma_{t\ shot}(u)}{\sigma_{t\ shot}} + \frac{\sigma_{v\ shot}(u)}{\sigma_{v\ shot}} \\ &+ \frac{\sigma_{x\ shot}(u)}{\sigma_{x\ shot}} \\ &= 1/5 + 0 + 0 + 0 = 1/5 \end{aligned} \tag{4}$$

(Duch et al., 2010), who believed that a player's degree of influence in the game will have the largest evaluation in the sequence directly connected to the goal, arguing that flow centrality can evaluate players more effectively than betweenness centrality.

3 PROPOSED METHOD

The shortest sequence in flow centrality is determined by fixing node t as shot nodes. However, in a soccer game, a team may adopt defensive tactics or maybe unilaterally attacked during a game. As a result, the number of shots would be relatively small in these cases. In applying flow centrality to such games, the number of the shortest sequence to be calculated decreases, and the players involved in the score in the measurement would be limited.

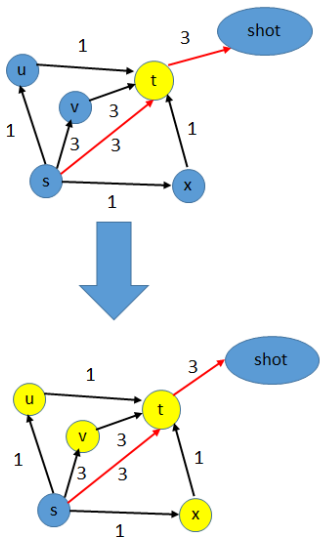

Figure 4: Flow Centrality Problem.

In the example of Figure 4, when flow centrality is applied, the shortest sequence is $s \rightarrow t \rightarrow shot$. In this case, $\sigma_{s\ shot}(t)/\sigma_{s\ shot} = 1$ and $\sigma_{s\ shot}(u)/\sigma_{s\ shot} = \sigma_{s\ shot}(v)/\sigma_{s\ shot} = \sigma_{s\ shot}(x)/\sigma_{s\ shot} = 0$. Therefore, the values of nodes u, v, and x, which are involved in the opportunity of the score, are not evaluated. To evaluate each player's contribution to the score, we need to consider not only node t but also nodes u, v, and x.

Therefore, in this research, in order to consider the players involved in the passes leading to the shot, we increase the number of sequences to be calculated by searching for a sequence whose sequence length is l+1, where l is the length of the shortest sequence.

Let σ'_{st} be the number of the sequence with a length is l+1 from node s to node t. Let $\sigma'_{st}(v)$ be the number of sequences through node v in the number of the sequence whose length is l+1 from node s to node t. $flow^{+}$ is defined as follows.

$$flow^{+}(v) = \sum_{\substack{s \in V \\ t \in U \\ s \neq v}} \frac{\sigma'_{st}(v)}{\sigma'_{st}} \tag{5}$$

Figure 5 shows an example of a graph network created from sequences whose length is l+1.

From this graph network example, sequences with a length of l+1 include is $s \rightarrow u \rightarrow t \rightarrow shot$, $s \rightarrow v \rightarrow t \rightarrow shot$, and $s \rightarrow x \rightarrow t \rightarrow shot$. Therefore, it is possible to evaluate nodes u, v and x. Furthermore, we can write equations (6) and (7) as follows:

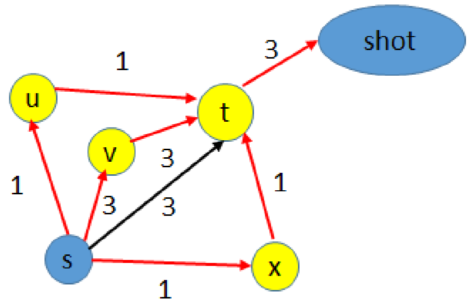

Figure 5: Flow+'s Graph Network Example.

$$\begin{aligned} \frac{\sigma'_{s\ shot}(t)}{\sigma'_{s\ shot}} &= 1 \\ \frac{\sigma'_{s\ shot}(u)}{\sigma'_{s\ shot}} &= \frac{\sigma'_{s\ shot}(x)}{\sigma'_{s\ shot}} = 1/5 \\ \frac{\sigma'_{s\ shot}(v)}{\sigma'_{s\ shot}} &= 3/5 \end{aligned} \tag{6}$$

$$\begin{aligned} flow^{+}(t) &= \frac{\sigma_{s\ shot}(t)}{\sigma_{s\ shot}} + \frac{\sigma_{u\ shot}(t)}{\sigma_{u\ shot}} + \frac{\sigma_{v\ shot}(t)}{\sigma_{v\ shot}} + \frac{\sigma_{x\ shot}(t)}{\sigma_{x\ shot}} \\ &= 1+1+1+1 = 4 \end{aligned} \tag{7}$$

4 EVALUATION EXPERIMENT

4.1 Data Sets

We conducted this experiment using the passing-distribution data of UEFA EURO 2008, 2012, and 2016. Table 1 shows the details of the data. "Matches" represents the number of matches in the tournament, "Teams" the number of teams, and "Evaluated Matches" the number of games excluding draws. "Passes attempted" indicates the number of passes in all games in the tournament, "Passes completed" represents the number of successful passes of all games, and "Ave. Passes" indicates the average number of successful passes. The passing-distribution data were prepared as public data from FIFA and UEFA and it can be obtained directly from the websites (FIFA, 2018) (UEFA, 2018).

Table 1: Data Details.

Data Sets	2008	2012	2016
Teams	16	16	24
Matches	31	31	51
Evaluated Matches	26	24	42
Passes attempted	23756	27517	44490
Passes completed	17679	19893	37947
Ave. Passes	339	414	744

4.2 Experiment

We examined whether the proposed model is an appropriate measure for the players and teams in the match. $P^A_{betweenness}$, P^A_{flow}, and $P^A_{flow^+}$ are arrays of $betweenness$, $flow$, and $flow^+$ of players in the initial lineup of team A, where the arrays are sorted in descending order. We defined the team's evaluations in matches in equations (8), (9), and (10). The players are represented as $n \in \{1,2,...,11\}$.

$$\overline{P}^A_{betweenness}(n) = \frac{1}{n}\sum_{i=1}^{n} P^A_{betweenness}(i)$$
$$(P^A_{betweenness}(1) > P^A_{betweenness}(2) > ...) \quad (8)$$
$$\overline{P}^A_{flow}(n) = \frac{1}{n}\sum_{i=1}^{n} P^A_{flow}(i)$$
$$(P^A_{flow}(1) > P^A_{flow}(2) > ...) \quad (9)$$
$$\overline{P}^A_{flow^+}(n) = \frac{1}{n}\sum_{i=1}^{n} P^A_{flow^+}(i)$$
$$(P^A_{flow^+}(1) > P^A_{flow^+}(2) > ...) \quad (10)$$

(Duch et al., 2010) used team evaluations as the average of the top n players' values from the equations (8), (9) and (10). They calculated this estimate from the team evaluation and by using the bootstrap method. As a result, they showed the usefulness of this method by creating the sample distribution of an estimate and performing a hypothesis test.

In this research, we set $n = 3, 11$ based on (Duch et al., 2010) and verified whether the values of $\overline{P}^A(n)$ and $\overline{P}^B(n)$ are directly linked to the outcome of a match of A versus B. We calculate the accuracy of our method as the ratio of agreement between the values of $\overline{P}^A(n)$ and $\overline{P}^B(n)$ and a match's outcome agree (Duch et al., 2010). We did not conduct the evaluation in the case of a draw.

4.3 Result

The results of our experiments on the data from EURO 2008, 2012, and 2016 will be described below.

Table 2: Accuracy of Match Results.

EURO 2008	$betweenness$	$flow$	$flow^+$
$\overline{P}(11)$	0.35	0.42	0.46
$\overline{P}(3)$	0.35	0.50	0.65
EURO 2012	$betweenness$	$flow$	$flow^+$
$\overline{P}(11)$	0.42	0.46	0.46
$\overline{P}(3)$	0.38	0.42	0.42
EURO 2016	$betweenness$	$flow$	$flow^+$
$\overline{P}(11)$	0.59	0.35	0.49
$\overline{P}(3)$	0.41	0.35	0.41

Table 2 shows the accuracy of betweenness centrality, flow centrality, and flow+ centrality as a ratio of the level of agreement between the experiment's array values and match outcomes. Furthermore, Table 3, 4, and Table 5 depict $\overline{P}^A_{flow}(3)$ and $\overline{P}^B_{flow}(3)$ as well as $\overline{P}^A_{flow^+}(3)$ and $\overline{P}^B_{flow^+}(3)$. The country's name is expressed in FIFA code, and when the match results and values do not match, the values are followed by the letter F (FIFA, 2010).

5 DISCUSSION

Table 2 shows that $flow^+$ of $\overline{P}$ shows higher values than $betweenness$ and $flow$. Because $flow^+$ calculates sequences whose length is l+1, more sequences leading to the shot are calculated than the conventional model. From this, the player involved in the sequence that ended with a shot was evaluated. Therefore, this is considered to be more accurate than the conventional method. In this research, we verified $flow^+$'s soundness with sequences with a length of l+1. From now on, it is necessary to estimate the appropriate length of the sequence. It is also necessary to verify a new index in which both the sequences with a length of l+1 and the sequences with a length of l are applied.

Furthermore, in EURO 2012 and EURO 2016, the accuracy was low overall. The value of $flow^+$ in $\overline{P}$ did not exceed $flow$ in EURO 2012 and $betweenness$ in EURO 2016. From this result, in order to verify the cause of mismatch of $flow^+$ against the outcomes, we examined the distribution of the value of $flow^+$. As an example of a mismatch of $\overline{P}$, Figure 6 shows the distribution of $flow^+$ for the match between Italy and the Republic of Ireland in EURO 2012. Furthermore, Figure 7 depicts the distribution of $flow^+$ for the match between Croatia and Spain.

Figure 6 shows an example where Italy won but $\overline{P}$ in $flow^+$ is higher for the Republic of Ireland than for Italy. The variance of the value of $flow^+$ for Italy is 4.68, whereas the Republic of Ireland is 15.1, which

Table 3: Team Evaluation of Each League Game in EURO 2016.

Gr.A	$P_{flow}(3)$	$P_{flow^+}(3)$
FRA-ROM	**3.92-5.82,F**	**7.05-8.92,F**
ALB-SUI	**6.33-4.55,F**	**8.60-7.06,F**
FRA-ALB	**4.52-6.54,F**	**7.55-8.70,F**
ROM-ALB	**5.28-3.62,F**	**6.84-6.67,F**
Gr.B	$P_{flow}(3)$	$P_{flow^+}(3)$
WAL-SVK	**5.11-6.33,F**	**8.46-8.69,F**
RUS-SVK	6.4-4.44	**8.42-7.32,F**
ENG-WAL	**4.24-7,F**	**7.84-9.98,F**
RUS-WAL	4.03-4.38	**6.77-6.51,F**
Gr.C	$P_{flow}(3)$	$P_{flow^+}(3)$
POL-NIR	**4.24-4.33,F**	7.53-4.91
GER-UKR	**3.79-6.44,F**	**6.04-8.18,F**
UKR-NIR	3.45-4.33	5.62-6.64
UKR-POL	3.33-4.72	6.18-7.60
NIR-GER	**4.79-3.2,F**	5.28-5.77
Gr.D	$P_{flow}(3)$	$P_{flow^+}(3)$
TUR-CRO	**4.67-2.27,F**	**8.22-5.09,F**
ESP-CZE	**3.40-5.81,F**	**5.12-8.30,F**
ESP-TUR	5-2.67	7.09-4.31
CZE-TUR	4.07-5.61	6.93-7.75
CRO-ESP	**3.72-4.61,F**	**5.91-6.99,F**
Gr.E	$P_{flow}(3)$	$P_{flow^+}(3)$
BEL-ITA	4.34-6	7.06-8.45
ITA-SWE	6.22-3	7.42-4.59
BEL-IRL	4.01-6.83	5.94-8.53
ITA-IRL	2.67-4.18	4.39-7.63
SWE-BEL	**4.89-4.11,F**	**8.33-6.60,F**
Gr.F	$P_{flow}(3)$	$P_{flow^+}(3)$
AUT-HUN	3.94-2.12	6.63-5.96
ISL-AUT	**3.83-4.39,F**	7.12-6.42

shows that there are variations in the value of $flow^+$ for the Republic of Ireland.

Although the relative evaluation of each player does not change depending on $flow$ and $flow^+$, we can observe a phenomenon in which particular players get extremely high evaluations with $flow^+$, resulting in a high team evaluation.

In addition, the Croatia versus Spain game in Figure 7 shows that Spain won, but it is the mismatch of $flow^+$ against the outcomes. In this game, the overall evaluation of the Spanish team is low.

From this, we found that there are mainly two patterns in the trend of $flow^+$ in cases of inconsistency. From this point on, we analyzed these two patterns further and examined the current problem and solution of $flow^+$.

The reason why the value of a specific player, such as Keith Andrews in the Republic of Ireland in Figure 6, becomes so large is that the number of shots in the

Table 4: Team Evaluation of Each League Game in EURO 2012.

Gr.A	$P_{flow}(3)$	$P_{flow^+}(3)$
RUS-CZE	3.75-2.72	7.33-5.41
GRE-CZE	**5.42-5.33,F**	7.07-8.49
CZE-POL	3.3-2.64	5.55-5.11
GRE-RUS	5.67-1.67	7.93-4.65
Gr.B	$P_{flow}(3)$	$P_{flow^+}(3)$
NED-DEN	**2.33-1.71,F**	**5.49-3.32,F**
GER-POR	5.29-4.36	7.59-6.22
DEN-POR	**5.25-4.46,F**	**8.16-7.20,F**
NED-GER	3.82-4.2	**7.34-5.07,F**
DEN-GER	**4.22-4.10,F**	**6.71-5.81,F**
POR-NED	3.60-2.88	6.97-5.67
Gr.C	$P_{flow}(3)$	$P_{flow^+}(3)$
IRE-CRO	**4.34-3.84,F**	**5.49-3.32,F**
ESP-IRE	**3.11-6.43,F**	**5.12-6.43,F**
CRO-ESP	**6.33-4.52,F**	**8.76-6.51,F**
ITA-IRE	**2.54-6.17,F**	**5.94-8.36,F**
Gr.D	$P_{flow}(3)$	$P_{flow^+}(3)$
UKR-SWE	**3.56-4.21,F**	**6.32-6.87,F**
UKR-FRA	**5.12-3.67,F**	**7.99-6.52,F**
SWE-ENG	**3.34-2.75,F**	**6.96-5.54,F**
ENG-UKR	4.12-2.86	7.12-6.27
SWE-FRA	4.44-2.44	7.12-6.1

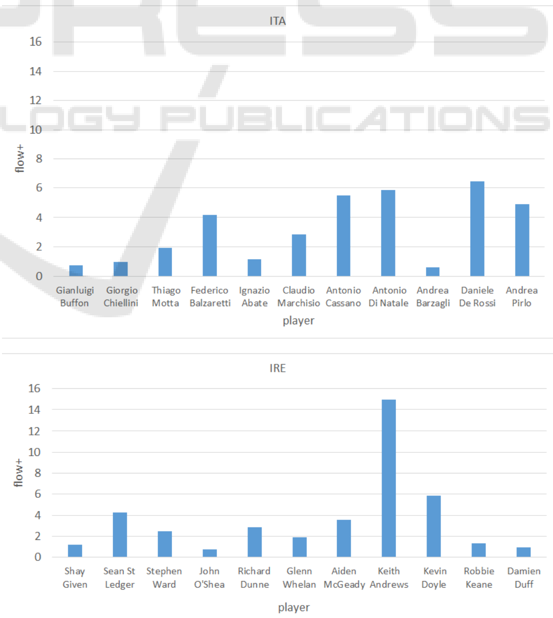

Figure 6: Flow+ of Italy and the Republic of Ireland.

game is extremely small and the tactics used to specify a particular player as the destination of passes are carried out. In the future, when the value of the index of a specific player becomes large, it would be necessary to discount it. Furthermore, Table 6 shows the breakdown of the number of shots in a EURO 2008

Table 5: Team Evaluation of Each League Game in EURO 2008.

Gr.A	$P_{flow}(3)$	$P_{flow^+}(3)$
SUI-CZE	4.37-7.83	6.64-8.71
POR-TUR	4.66-3	7.32-4.55
CZE-POR	4.86-4.89	7.4-7.66
SUI-TUR	3.19-4.67	6.88-7.09
SUI-POR	**5-5.31,F**	8.17-7.94
TUR-CZE	**3.9-4.05,F**	6.6-6.49
Gr.B	$P_{flow}(3)$	$P_{flow^+}(3)$
AUT-CRO	4.83-5.12	7.23-7.7
GER-POL	5.67-3.87	8.86-6.98
CRO-GER	**3.23-4.22,F**	**6.38-6.87,F**
POL-CRO	**4.07-3.51,F**	6.45-6.87
AUT-GER	**4.72-2.71,F**	**6.92-5.24,F**
Gr.C	$P_{flow}(3)$	$P_{flow^+}(3)$
NED-ITA	**3.27-3.33,F**	6.04-5.19
NED-FRA	4.58-3.26	7.24-6.24
NED-ROU	**3.68-5,F**	**6.69-7.18,F**
FRA-ITA	**5.33-3.57,F**	**8.33-5.69,F**
Gr.D	$P_{flow}(3)$	$P_{flow^+}(3)$
ESP-RUS	**3.57-4.61,F**	**6.79-7.13,F**
GRE-SWE	**3.72-3.12,F**	6.07-6.25
SWE-ESP	4.33-4.5	5.91-7.26
GRE-RUS	**4.97-4.43,F**	6.89-7.53
GRE-ESP	4.58-4.73	6.26-8.29
RUS-SWE	**2.86-3.2,F**	**5.95-6.34,F**

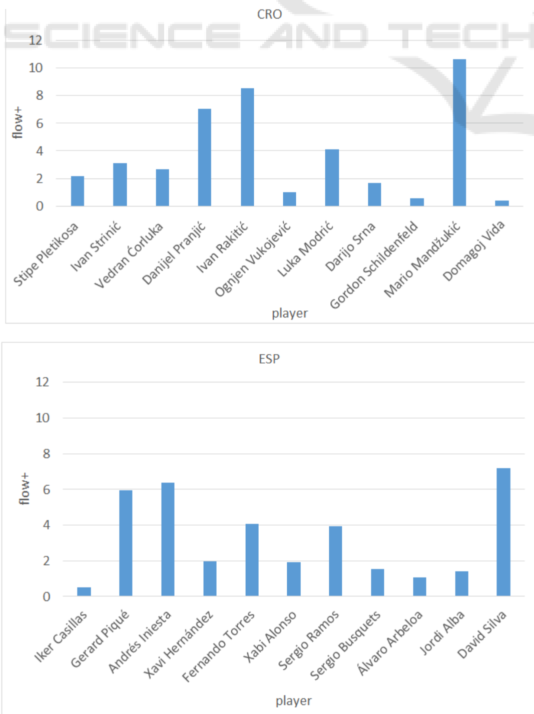

Figure 7: Flow+ of Croatia and Spain.

Table 6: Breakdown of the Number of Shots in a EURO 2008 Game.

	Winning Team	Losing Team
Total Attempts	337	287
On-Target	175	121
Off-Target	162	166
Ave. On-Target	6.73	4.64

Table 7: Breakdown of the Number of Shots in a EURO 2012 Game.

	Winning Team	Losing Team
Total Attempts	347	292
On-Target	195	135
Off-Target	152	157
Ave. On-Target	8.13	5.63

game, and Table 7 shows the breakdown of the number of shots in a EURO 2012 game. The number of shots is “Total Attempts”, the number of shots on goal is “On-Target”, the number of the shots wide is “Off-Target”, and the number of the average shots on goal is “Ave. On-Target”.

From Table 6 and Table 7, it can be presumed that the winning teams are superior in the number of shots, the number of shots on goal, and the number of the average shots on goal. From this assumption, it can be inferred that the number of shots is larger in the winning team than the defeated team, and the shot accuracy tends to be higher in the winning teams. In this research, we calculated $flow^+$ by treating team A and team B equally. In the future, changing the edge weight by shots out of the frame or shots in the frame is a necessary step.

Furthermore, we analyzed the cause of the tendency of values like Spain in Figure 7. $flow^+$ is an index that quantifies how much is involved in the sequence of passes that leads to a scoring opportunity. Figure 8 shows a scatterplot of the number of shots and the number of passes of the winning team of EURO 2012, and Figure 9 shows a scatterplot of the defeated team's number of shots and number of passes in EURO 2012.

In Figure 8, Spain's the number of successful passes is over four hundred and the number of shots increases accordingly. In fact, the correlation coefficient between the number of shots and the number of passes of the winning team is 0.45, whereas it is 0.63 for the defeated team. From this, we can say there is a correlation. Based on this, it is thought that the more the number of successful passes increases, the more the number of scoring opportunities increases as well. In Spain, the number of successful passes is much larger than that of other teams, therefore the value of $flow^+$ will most likely not become too large

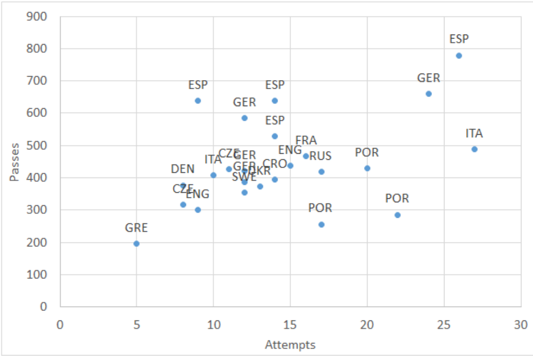

Figure 8: Scatterplot of the Number of Shots and the Number of Passes of EURO 2012's Winning Teams.

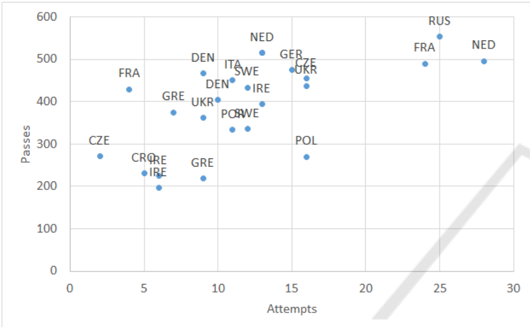

Figure 9: Scatterplot of the Number of Shots and the Number of Passes of EURO 2012's Defeated Teams.

even if the number of shots increases. This can be improved by limiting the number of successful passes to a certain range.

6 CONCLUSIONS

In this research, we analyzed the contribution of a soccer player in the matches by utilizing graph theory's centrality concept from passing-distribution data. As a result, we found that the proposed model sometimes has better accuracy than the conventional model. We were able to verify the soundness of the proposed model. By improving $flow^{+}$, new indicators may emerge in future studies. In addition, an objective method is necessary to show whether the individual player's evaluation is sound.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP16K00419, JP16K12411, JP17H04705, JP18H03229, JP18H03340, JP18K19835. We acknowledge the professor Shinichi Honiden of the Waseda University for providing opportunities to study and the members of our laboratory.

REFERENCES

Barthelemy, M. (2004). Betweenness centrality in large complex networks. In *The European Physical Journal B pp.163–168*.

Beneventano, P., Berger, P., and Weinberg, B. (2012). Predicting run production and run prevention in baseball: The impact of sabermetrics. In *International Journal of Business, Humanities and Technology Vol. 2 No. 4 pp.67-75*.

Cotta, C., Mora, A., Merelo, J. J., and Merelo-Molina, C. (2013). A network analysis of the 2010 fifa world cup champion team play. In *JOURNAL OF SYSTEMS SCIENCE and COMPLEXITY(2013) 26: pp.21–42*.

Duch, J., Waitzman, J., and Amaral, L. (2010). Quantifying the performance of individual players in a team activity. In *PLoS ONE 5(6): e10937. https://doi.org/10.1371/journal.pone.0010937*.

FIFA (2010). Fifa's 208 member associations. http://www.fifa.com/mm/document/fifafacts/organisation/52/00/10/fs-120_01a_ma.pdf.

FIFA (2018). https://www.fifa.com/.

Goncalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jimenez, S., and Sampaio, J. (2017). Exploring team passing networks and player movement dynamics in youth association football. In *PLoS ONE 12(1): e0171156. doi:10.1371/journal*.

Tsugawa, S. and Ohsaki, H. (2014). A study on the robustness of centrality measures in network analysis. In *Dai 57 kai zidou seigyo rengou kouenkai ronbunsyu pp.1072-1077*.

UEFA (2018). https://www.uefa.com/insideuefa/mediaservices/presskits/.