IJWR

Social Network Analysis of Football Communications by Finding Motifs

Amir Hossein Ahmadi[a], Babak Teimourpour*[b], Mahtab Mahbood[c]

[a] Master of Information Technology Engineering, Tarbiat Modares University, Tehran, Iran; AmirHossein_Ahmadi@modares.ac.ir
[b] Assistant Professor of Information Technology Engineering, Tarbiat Modares University, Tehran, Iran; B.Teimourpour@modares.ac.ir
[c] Master of Computer Engineering, Amirkabir University of Technology, Tehran, Iran; Mahtab.Mahbood99@aut.ac.ir

ABSTRACT

Statistics, extraction, analysis are vital in sports science. Information technology and data science will significantly increase the quality of research and decisions of sports clubs and organizations. Currently, many coaches and sports institutions use analytics and statistics that are calculated manually. Sports science shows that winning a match depends on different factors.
The purpose of the research is to improve team performance by analyzing social networks, communication networks (such as players' passes and transactions during the match), and analyzing repetitive areas. These results are done by analyzing the data collected from 4 matches of the Persepolis team, including three matches from the first half of the Iranian Premier League in 2018-1399 and a Persepolis match against Al-Sharjah. This research examines the issue from two interconnected aspects: 1- Examining the performance of players individually and as part of a social network. 2- explore the communication network between players and land areas. This analysis uses the innovative method of identifying and classifying motifs.

Keywords: ***Social Network Analysis, Graph Analysis, Motif, Frequent Subgraph, Centrality.***

1. Introduction

Sports disciplines have spread rapidly, but a small percentage of research is related to sports science. Considering this issue, we realize that there is a research hole, or in other words, a need in sports organizations, which can be solved by expanding studies and research in sports science [1].

Information systems are currently used in most affairs. Sport is not a closed system. Sport is considered a mother information system that establishes connections between other subsets of information. Therefore, the use of information systems in these departments increases performance and efficiency and greatly accelerates and facilitates the activities of sports organizations[2].

With the continuous development of sports science, a helpful link between information technology and sports information management science is growing. Research, development, and analysis of sports information will be established as a trend in sports-related science [3]. A lot of information about sports can be obtained from the study of these sciences. Some of these data do not work until valuable information is extracted from them. Examining and identifying practical and valuable information from a large amount of data is called data mining.

Data mining has required data cleaning, integration, conversion, pattern evaluation, and presentation [4].

By examining the results of the analysis of the communication network of the passes, it is possible to identify possible ways to achieve a goal in a team. These reviews will help the technical staff analyze the match and choose the appropriate tactics, player selection, and order in future matches. A large part of the studies that look at the football sport from the point of view of the network theory show the passes as a basis for creating attack situations as a network in which the nodes represent the players, and the connection of the nodes means the pass. between players. The characteristics of players and teams are evaluated quantitatively by analyzing and examining the network of passes [5,6,7].

The relative importance of a node in the network is determined by different criteria of centrality, defined in graph theory and the science of network analysis. Social network analysis became a factor in developing a large part of centrality concepts in network science. Many topics and terms used to determine the indicators of centrality started many types of research in the field of sociological sciences [8].

The network analysis approach in the research showed that one of the most critical criteria of a team's performance is the interaction pattern between the team members.[5] An effective pattern connection can be created by identifying and examining these patterns.

The way players cooperate in team sports is critical. Analyzing the cooperation of players in these sports is very efficient and valuable for competitions. A review of research in this field showed that some studies had used the method of social network analysis to investigate the communication between players in the team [8,9,10]. The higher the coordination of the team, the better the performance and efficiency of the team, and finally, the output obtained from the team will be better. Coordination and appropriateness

http://dx.doi.org/10.22133/ijwr.2022.321880.1110

Citation: A. H. Ahmadi, B. Teimourpour, and M. Mahbood, "Social Network Analysis of Football Communications by Finding Motifs," *International Journal of Web Research*, vol.5, no.2,pp.39-46, 2022, doi: 10.22133/ijwr.2022.321880.1110.

*Coressponding Author
Article History: Received: 25 July 2022; Revised: 16 October 2022 ; Accepted: 19 October 2022

Copyright © 2022 University of Science and Culture. Published by University of Science and Culture. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license(https://creativecommons.org/licenses/by-nc/4.0/). Noncommercial uses of the work are permitted, provided the original work is properly cited.

The motifs obtained in this research show the strong connections of the existing nodes (in the sequence) with each other. Each of these motifs is analyzed as a correlation path. Motifs can be used to formulate the strategy of selecting players, to arrange the team, and making the players' solidarity by their capabilities (before the matches). It is also possible to use motifs to analyze the path of solidarity between the players of the rival team (in previous rival matches) and plan and provide solutions to deal with these connections. Motif analysis can also be used during a match.

Investigating the correlation paths formed during a match can reveal part of the team's strategy. The number of repetitions and nodes in each subgraph is essential.

2.6. Transactions Distribution per the Area Graph

Figure 3 shows the communication network relative to the position on the field of play (areas on the field). This graph as to the match between Persepolis and Al-Sharjah, which shows the activity and strategy of the team to rotate the ball or try to penetrate from the middle and in the next dimension of the left-wing.

2.7. Reach Centrality

Each event in this section was assigned a proportional weight, which is explained. First, all occurrences were given a weight of 1. In the next step, the cycles were categorized into successful and unsuccessful. Then the cycles were examined separately, and the weight was assigned to the events within a cycle. In the cycles of the successful category, a positive weight (variable depending on the type of cycle) was given to the last event. From last to first, a fraction of that positive weight was given in descending order to our other events. For unsuccessful cycles, a negative weight (variable depending on the type of cycle) was given to the last event of the cycle, and only the last two events (from the last to the previous) were given as a fraction of the negative weight to the rest of the events in the cycle.

Finally, the total weight of the edges in different cycles is obtained, which indicates the suitability or non-suitability of each path (from the source player to the goal that is successful in reaching the events of the cycle) by considering the influential players.

Finally, a proportional weight is assigned to each edge, according to which the total weight of the edges in different cycles is obtained. The total weight of each cycle shows whether or not each path is suitable. In this way, from the beginning (which is the origin player of the cycle) to the goal (which is to reach the final events of the successful cycle) by considering the effective players (intermediate nodes in the cycle), they are examined.

3. Results

3.1. Average degrees

The results of the analysis of the three criteria of average degrees, weighted average degrees, and graph density show that:

The weighted average score in the match of the Persepolis team vs. Golgohar team was higher than the rest of the matches, which indicates better statistical performance in obtaining access centrality points (good results in scoring cycles and proper use of communications).

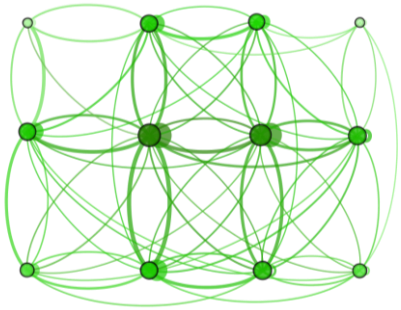

Figure. 3. Frequent subgraph - area correlation

The match between Persepolis and Al-Sharjah teams has the highest average score. This point advantage shows the cooperation of the players in this match compared to the other three matches.

Persepolis match against Al-Sharjah team got the lowest score in this analysis. This issue shows the failure to achieve the desired results in cycles compared to other competitions.

3.2. Motifs - Player Correlation [32]

Meaningful, frequent subgraphs or motifs and correlation paths were obtained in all four matches. A few of them were mentioned. For example, in the match between Persepolis and Paykan, some correlation paths are shown in Figure 4.

The motifs in Figure 4 show the attempt to penetrate from the right side (using players from the right side of the field, for example, player number 17), which is the strategy of the Persepolis team in this match.

The motifs in Figure 5 in the Persepolis-Esteghlal match show the proper relationship between the defense and the midfield. The middle line will not be able to communicate with the offensive line to complete the correlation path.

The motifs in Figure 6 in the Persepolis vs. Al-Sharjah match show the lack of control and improper communication between players 88 and 5. This lack of control leads to the loss of possession of the ball, which repeatedly happened during the match.

Correlation paths with at least three players and three repetitions of the total events of the three matches of the Persepolis Premier League were collected and given in Figure 7. paths were categorized according to whether the beginning and end were in one line (defense, midfield, and attack) (corresponding to the player's position, not the player's current position) or first in one line and end in another. paths that went first on one line, then to another, and finally back to the starting line were considered "other." (For example, in Figure 4, paths 1 to 6 to 17 is regarded as a line of defense.).

In contrast, correlation paths with at least three players and three repetitions of the total events of the Persepolis vs. Al-Sharjah match were obtained, which were given for comparison in Figure 8. It is clear that the communication between the defensive and midfield lines and the attacking line players is weak, and these players are not present in the process of correlation paths of the players. This weakness is exacerbated by the lack of strong communication between

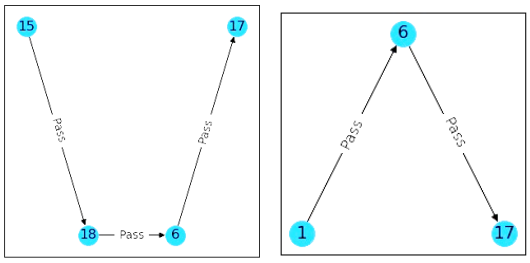

Figure. 4. Motifs - Node Destination 17

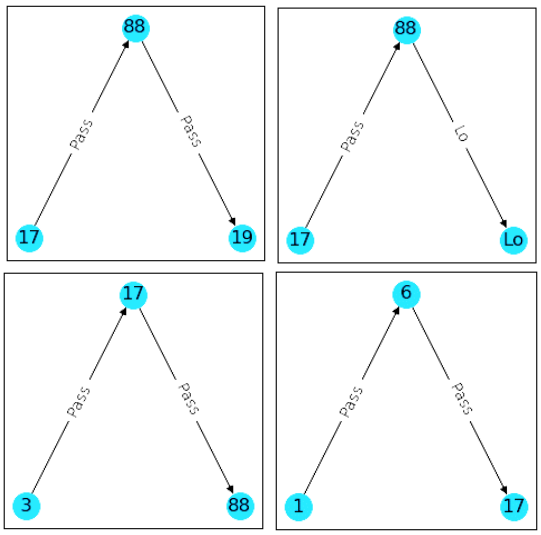

Figure. 5. Motifs - Weakness in attack

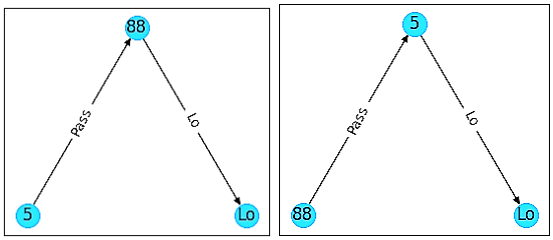

Figure. 6. Motifs - Leaking the ball

attacking players and the lack of correlation paths on the offensive line. Comparing this statistic from the Premier League matches with the sample of this statistic in the match against Al-Sharjah, results that the mentioned weakness persists.

3.3. Motif - Area Correlation

By extracting the correlation of the areas from repetitive transactions and obtaining the patterns of Persepolis vs. Al-Sharjah match and classifying them, the following results are determined:

Figure 9 shows the correlation paths with the criterion of dividing the field of play into Persepolis vs. Al-Sharjah fields and the frequent rotation of the ball in these areas.

Figure 10 shows the correlation paths with the criterion of dividing the playing field into defensive-midfield-offensive lines and frequent rotation of the ball in these areas.

Figure 11 shows the correlation paths with the criterion of dividing the playing field into left-right-middle wings and frequent rotation of the ball in these areas.

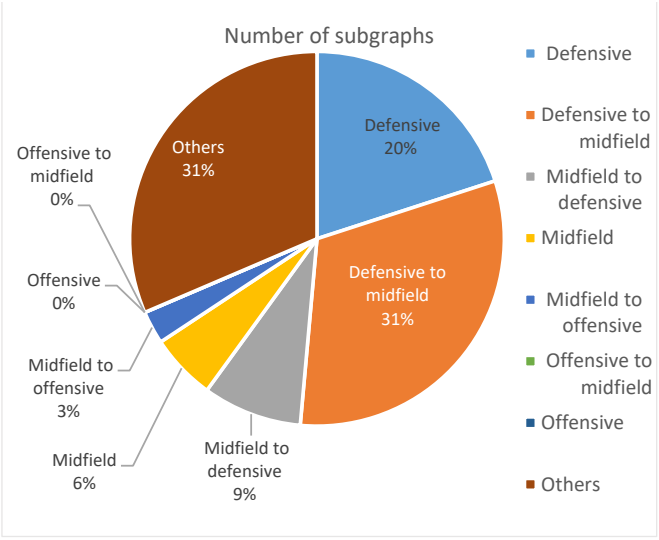

Figure. 7. Categorizing Persepolis motifs

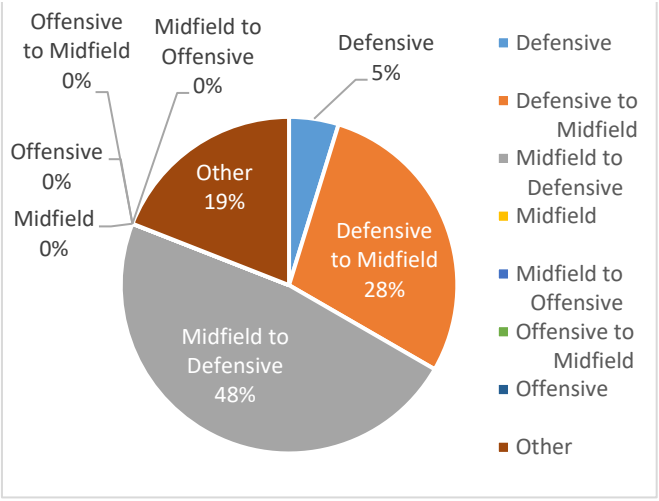

Figure. 8. Categorizing Persepolis repetitive motifs

Figure 11 shows the correlation paths with the criterion of dividing the playing field into left-right-middle wings and frequent rotation of the ball in these areas.

As a result, the frequent correlation of different areas shows that Persepolis (48% of the cases inside the opponent's field) is playing against Al-Sharjah and the strategy of direct and right-side penetration (39%).

3.4. Centrality

Reach centrality is calculated and illustrated with Gephi visualization software. The Figures 12-15 show the players based on dark blue, light blue, and cream colors from high to low degree of centrality.

In Figures 12-15, graphs are displayed where the color of the players' nodes indicates the result of the hits they hit the ball and their impact on the match process. The darker the color of the knot, it means that he sent valuable balls (shots, crosses, assists, etc.) and was an effective player in the match. Finally, one of the strategies according to this analysis is to deliver the ball to these players.

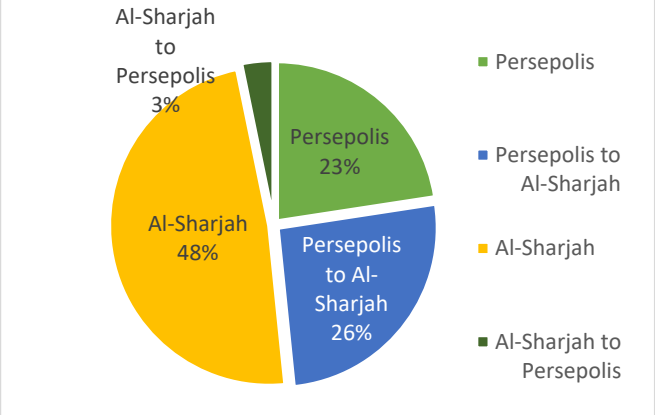

Figure. 9. Categorizing the ball rotation in repetitive patterns in ground

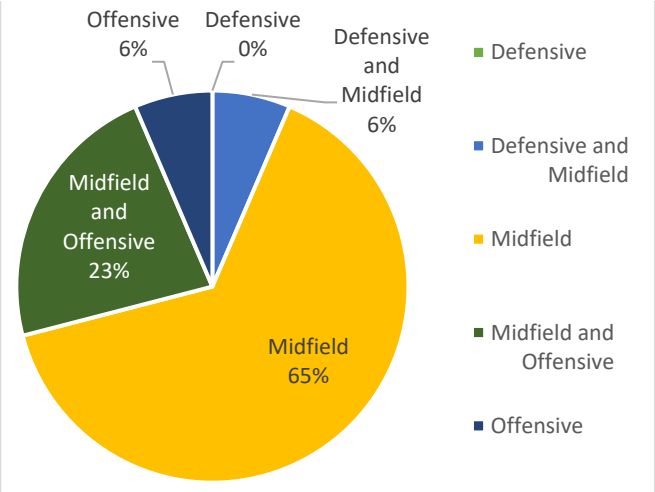

Figure. 10. Categorizing the ball rotation in repetitive patterns in lines

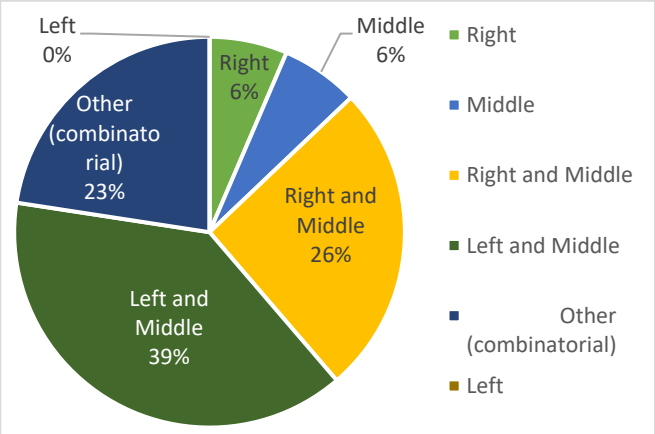

Figure. 11. Categorizing the ball rotation in repetitive patterns in sides

3.5. N-pass cycle

During the competition, the cycles obtained different results with different points, which were obtained from the sum of the points of the routes in the access center. Each cycle was labeled based on the total number of passes in that cycle. Cycles were categorized based on labels. (1P: cycle with pass length 1).

The diagram in Figure 16 shows the ratio of the total points obtained in each type of cycle to the number of cycles of that type. The diagram in Figure 16 is divided into two parts,

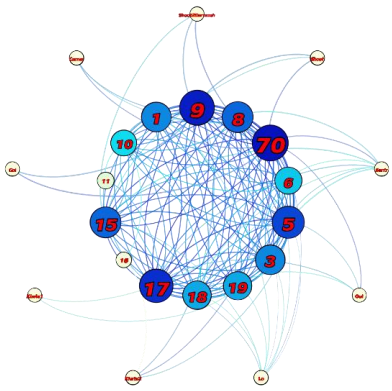

Figure. 12. Displaying the performance of Persepolis players vs. Gol Gohar in terms of points gained in the match

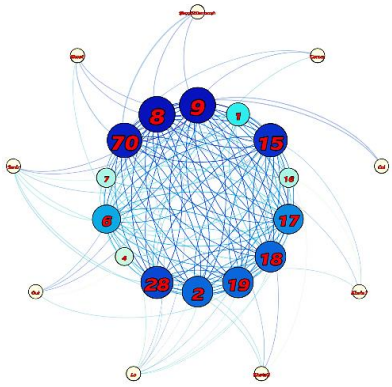

Figure. 13. Displaying the performance of Persepolis players vs. Paykan in terms of points gained in the match

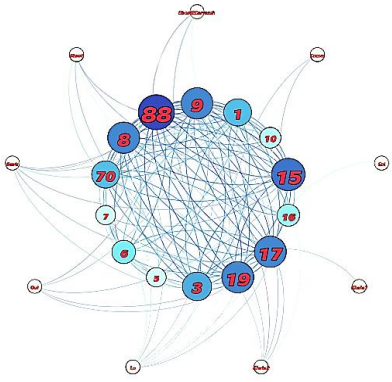

Figure. 14. Displaying the performance of Persepolis players vs. Esteghlal in terms of points gained in the match

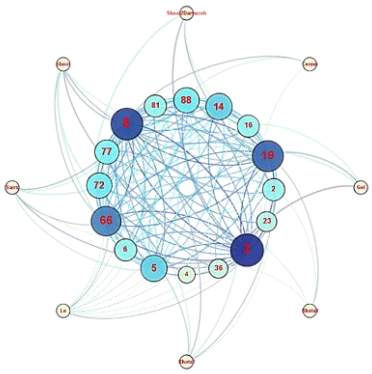

Figure. 15. Displaying the performance of Persepolis players vs. Al-Sharjah in terms of points gained in the match

including the three matches of the Persepolis team in the Premier League of the Persian Gulf and the Asian match of the Persepolis team against the Al-Sharjah team. The results show better results in the 7-pass and 8-pass cycles for the Persepolis team. Therefore, this team has performed better (relative to the total number of cycles) in long cycles. In contrast, this team shows poor performance in short pass cycles. This analysis shows the strategy of Persepolis in the matches to take possession of the ball and possessive play and create opportunities from the opponent's closed playing style.

4. Conclusion

Team success depends on many factors. This study analyzed the communication patterns of players in previous matches to analyze social networks of passes and in-game transactions. The strong interaction of team members and the cooperation of these people as a complete network makes the team dynamic and improves performance. This research has looked at sports from an innovative and networked perspective, which has provided results, claims, and solutions.

A database of matches was created to perform the analyzes performed in this research. The database included three matches of the Persepolis team in the Premier League of the Persian Gulf and one match in the Asian League from this team. The analyzes were done from two perspectives:

• Detailed view at the level of nodes: In this view, players were considered as part of a social network, and factors such as types of node degrees, points of each node, etc., were investigated.

• Network generalities: In this view, the players' communication network analysis from a macro perspective and the investigation of network characteristics were done. The network was illustrated as a graph, and the following factors: correlation path, n-pass cycle, critical path, etc., were calculated and analyzed.

In the first part, degree types, total degree-score of nodes, and graph density were calculated. On average, each player (node) was connected with 6 or 7 other nodes. Also, the density of the communication graph in the match with Al-Sharjah has increased unexpectedly compared to the average grades (both weighted and weightless).

In the second part, the correlation paths of players and areas of the field were identified and categorized. Most of the players' paths of solidarity among the players were with the position of the defensive line or the connection of the defensive and midfield line. In contrast, these routes did not exist in the offensive line and midfield to attack. On the other hand, for the game between Persepolis and Al-Sharjah, the correlation paths of more areas in the middle and the left regions inside the opponent's field were obtained, which showed that the team played and played forward in this match. Reach centrality, critical path, and n-pass cycles were calculated, which showed that the 8-pass cycles were effective for Persepolis. Finally, due to the lack of communication and strong correlation with the offensive line and in the offensive line, the existence of a problem within this line or effective communication with it was concluded in this team. This problem was seen both in the Premier League matches and in the Asian game of the Persepolis team. However, the team's victory in the tournament despite this problem can depend on the following two

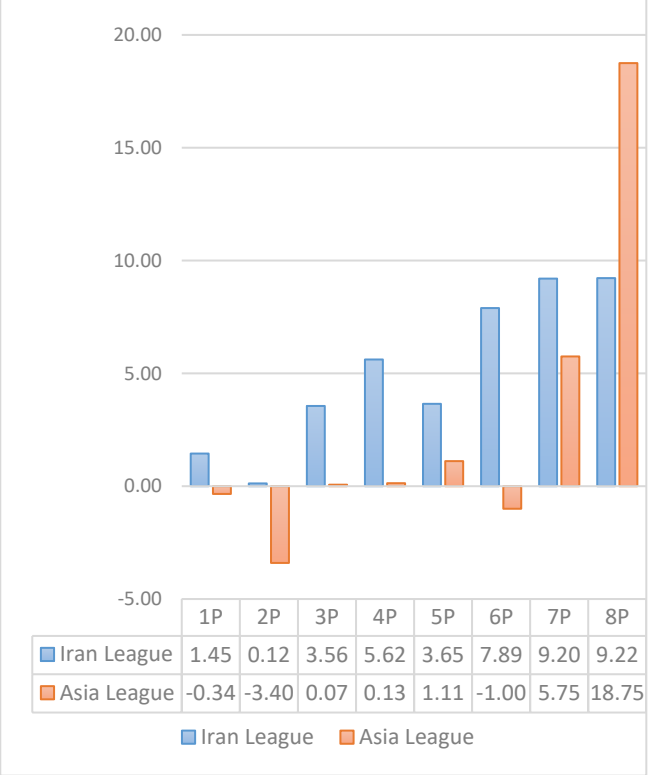

Figure. 16. Displaying the ratio of the total points earned per cycle to the number of cycles

influential factors:

1. The skill of individual players to play as nodes outside the communication network.

2. Support and replace players from other lines (except the offensive line) to achieve the desired result during matches.

Declarationss

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

AA: Study design, acquisition of data, statistical analysis, interpretation of the results, drafting the manuscript, revision of the manuscript; BT: Study design, Supervision, revision of the manuscript; MM: Study design, acquisition of data, drafting the manuscript;

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] T. Oktavia, F. L. Gaol, T. Hosoda, and A. Syahir, “Sport Science Model to Support the Professional Sports Organization Decision Making,” in 2020 International Conference on Information Management and Technology (ICIMTech), Aug. 2020, pp. 599–604, doi: 10.1109/ICIMTech50083.2020.9211238.

[2] J. Miočić, L. Zekanović-Korona, and B. Bosančić, “Information Systems in Sports Organizations: Case Study of the Sports Association of the City of Zadar,” in 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), May 2019, pp. 1362–1367, doi: 10.23919/MIPRO.2019.8756923.

[3] D. Jinguo, "The Theoretical Construction and Application System Development Study of Sports Information Management," in 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Jul. 2020, pp. 526–529, doi: 10.1109/ICPICS50287.2020.9202135.

[4] G. Kaur and G. Jagdev, "Analyzing and Exploring the Impact of Big Data Analytics in Sports Science," in 2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), Feb. 2020, pp. 218–224, doi: 10.1109/Indo-TaiwanICAN48429.2020.9181320.

[5] Grund, T. U., "Network structure and team performance: The case of English premier league soccer team", Social Networks, 34(4), 682–690, 2012.

[6] Duch, J., Waitzman, J. S., & Amaral, L. A., "Quantifying the performance of individual players in a team activity", PLoS One, 5(6), e10937, 2010.

[7] Clemente, F. M., Couceiro, M. S., Martins, F. M. L., & Mendes, R. S., "Using network metrics in soccer: A macro-analysis", Journal of Human Kinetics, 45, 123–134, 2015.

[8] Al Falahi, Kanna, Nikolaos Mavridis, and Yacine Atif., "Social networks and recommender systems: a world of current and future synergies." In Computational Social Networks, pp. 445-465. Springer, London, 2012.

[9] Duch, J., Waitzman, J. S., & Amaral, L. A., "Quantifying the performance of individual players in a team activity", Plos One, 5(6), 109-19, 2010.

[10] Bourbousson, J., Poizat, G., Saury, J., & Seve, C., "Team coordination in basketball: Description of the cognitive connections among teammates", Journal of Applied Sport Psychology, 22(2), 150-66, 2010.

[11] Fewell, J. H., Armbruster, D., Ingraham, J., Petersen, A., & Waters, J. S., "Basketball teams as strategic networks", Plos One, 7(11), 474-85, 2012.

[12] Kawasaki, T., Sakaue, K., Matsubara, R., & Ishizaki, S., "Football pass network based on the measurement of player position by using network theory and clustering", *International Journal of Performance Analysis in Sport*, *19*(3), 381-392, 2019.

[13] Clemente FM , Martins FML , Kalamaras D , Wong DP , Mendes RS ., "General net- work analysis of national soccer teams in Fifa World Cup 2014". Int J Perform Anal Sport,15(1), 2015.

[14] Duch J , Waitzman JS , Amaral LA., "Quantifying the performance of individual players in a team activity", PLoS One,5(6),e10937, 2010.

[15] Peña JL , Touchette H ., "A network theory analysis of football strategies", In: Clanet C, editor. Sports physics: proceuromech physics of sports con- ference. Palaiseau, France: Editions de l' \ 'Ecole Polytechnique, Palaiseau, p. 517–28, 2012 .

[16] Clemente FM , Martins FML , Kalamaras D , Wong DP , Mendes RS ., "Midfielder as the prominent participant in the building attack : a network analysis of na- tional teams in Fifa World Cup 2014", Int J Perform Anal Sport, 704–22, 2015.

[17] R. Marcelino, J. Sampaio, G. Amichay, B. Gonçalves, I. D. Couzin, and M. Nagy, "Collective movement analysis reveals coordination tactics of team players in football matches," Chaos Solitons Fractals, vol. 138, p. 109831, Sep. 2020, doi: 10.1016/j.chaos.2020.109831.

[18] Lago-Ballesteros, J., Lago-Peñas, C., and Rey, E., "The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team", Journal of Sports Sciences, 30(14), 1455-1461, 2012.

[19] Tenga, A., Holme, L., Ronglan, L.T., and Bahr, R., "*Effect of playing tactics on goal scoring in Norwegian professional soccer*", Journal of Sports Sciences, 28(3), 237-244, 2010.

[20] Reep, C., Pollard, R., & Benjamin, B., "*Skill and chance in ball games*", Journal of the Royal Statistical Society, A, 134, 623-629, 1971.

[21] Mike Hughes & Ian Franks., "*Analysis of passing sequences, shots and goals in soccer*", Journal of Sports Sciences, 23:5, 509-514, 2005.

[22] Paixão, Paulo, Jaime Sampaio, Carlos H. Almeida, and Ricardo Duarte., "How does match status affects the passing sequences of top-level European soccer teams?." International Journal of Performance Analysis in Sport 15, no. 1, 229-240, 2015.

[23] Clemente, Filipe Manuel, Hugo Sarmento, and Rodrigo Aquino., "Player position relationships with centrality in the passing network of world cup soccer teams: Win/loss match comparisons", Chaos, Solitons & Fractals 133,109625, 2020.

[24] Clemente, F. M., Martins, F. M. L., Kalamaras, D., Wong, P. D., & Mendes, R. S., "General network analysis of national soccer teams in FIFA World Cup 2014", International Journal of Performance Analysis in Sport, 15(1), 80–96, 2015.

[25] Clemente, Filipe Manuel, Fernando Manuel Lourenço Martins, Dimitris Kalamaras, Joana Oliveira, Patrícia Oliveira, and Rui Sousa Mendes. "The social network analysis of Switzerland football team on FIFA World Cup 2014." Journal of Physical Education and Sport 15, no. 1 : 136, 2014.

[26] Clemente, Filipe Manuel, Micael Santos Couceiro, Fernando Manuel Lourenço Martins, and Rui Sousa Mendes. "Using network metrics to investigate football team players' connections: A pilot study." Motriz: Revista de Educação Física 20, no. 3 : 262-271, 2014.

[27] Malta, Pedro, and Bruno Travassos. "Caraterização da transição defesa-ataque de uma equipa de Futebol." Motricidade 10, no. 1: 27-37, 2014.

[28] Y. Li, R. Ma, B. Gonçalves, B. Gong, Y. Cui, and Y. Shen, "Data-driven team ranking and match performance analysis in Chinese Football Super League," Chaos Solitons Fractals, vol. 141, p. 110330, Dec. 2020, doi: 10.1016/j.chaos.2020.110330.

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

[30] Chouhan, A. et al. (2021) 'Shotifier: A Binary Shot Conversion Classifier Pipeline for Football Forwards', in"2021 IEEE International Conference on Big Data and Smart Computing (BigComp) ", Jan.2021, pp. 156–163. doi: 10.1109/BigComp51126.2021.00038.

[31] B. Aalbers and J. Van Haaren, "Distinguishing between roles of football players in play-by-play match event data," in Proceedings of International Workshop on Machine Learning and Data Mining for Sports Analytics, pp. 31–41, 2018.

[32] A. H. Ahmadi, A. Noori, and B. Teimourpour, "Social Network Analysis of Passes and Communication Graph in Football by mining Frequent Subgraphs," in 2020 6th International Conference on Web Research (ICWR), Apr. 2020, pp. 1–7, doi: 10.1109/ICWR49608.2020.9122303.

AmirHossein Ahmadi received his B.Sc degree in Buin Zahra Technical University, Qazvin, Iran in 2019 and Currently, he is an M.Sc. student in Information Technology at Tarbiat Modares University. His research interests include data mining, machine learning, network science and programming.

Babak Teimourpour received his B.Sc degree in Industrial Engineering, Sharif University, Tehran, Iran in1996 and received his M.Sc. degree in Socio-Economic Systems Engineering, Institute for Research on Planning and Developement, Tehran, Iran in 1998. Also he received his Ph.D. in Industrial Engineering at Tarbiat Modares University, Tehran, Iran in 2010. His research interests include data mining, social network analysis.

Mahtab Mahbood received her B.Sc degree in Buin Zahra Technical University, Qazvin, Iran in 2020 and Currently, She is an M.Sc. student in Computer Engineering at Amirkabir University of Technology (Tehran Polytechnic) . Her research interests include network analysis and machine learning.