

## FOOTBALL NETWORK RESEARCH

**Jiaming Zhang<sup>a</sup>, Linqing Guo<sup>a</sup>, Yukun Liu<sup>b</sup>,  
Shihe Wang<sup>c</sup> and Shuaiqi Liu<sup>a</sup>**

<sup>a</sup>College of Electronic and Information Engineering, Hebei University, Baoding  
Hebei, P. R. China

<sup>b</sup>College of Architecture and Engineering, Hebei University, Baoding Hebei,  
P. R. China

<sup>c</sup>College of Cyberspace Security and Computer, Hebei University, Baoding  
Hebei, P. R. China

---

### Abstract

As societies become more connected, teams are increasingly required to solve complex challenges. We use the football network model as an example to analyze the teamwork strategy. First, based on the acquisition of passing statistics, the establishment of a pass network based on graph theory laid the foundation for its application; second, considering the flow properties of the pass network itself, this paper sets an H index to measure and evaluate the performance of the entire team cooperation, using the Analytic Hierarchy Process evaluates the effectiveness of teamwork; finally, through the construction of the passing network to analyze the impact of key nodes on the game results, the best strategy for teamwork is obtained.

---

\*Corresponding author.

*E-mail address:* ZJM\_HBU@163.com (Jiaming Zhang).

*Keywords:* football, team strategy, performance indicators, passing network.

---

## 1. Introduction

At present, the application of big data technology in social networks, finance, medical care, etc. is constantly developing and maturing, but its application in competitive sports is still in the exploration stage. It is very necessary to optimize the team configuration based on big data analysis, improve the team's combat capabilities, and find the way to "defend the enemy" in the game by analyzing the opponent's technical data. Using big data analysis, the coach can understand the current game situation, the characteristics and performance of each player, the player's defensive range, the opponent team's free zone and other information. Through this information, the coach can more effectively optimize the player's playing time, position, skills and tactics, etc. to improve the overall performance of the player and the team. We mainly discuss football, analyze the network structure formed by the interaction between players, and build an evaluation system that has nothing to do with the results.

This paper selects the NodeXL tool to construct the pass network graph, and on this basis analyzes the pass data between players, fully taps the connotation of pass data, and provides support for player training, tactics formulation, opponent analysis, coach decision making, etc.

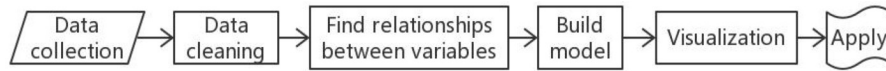
First of all, we fit the data on the basis of more accurate data, use the knowledge of graph theory to construct the pass network, use statistical knowledge to find the adjacency matrix, temporal clustering, spatial clustering analysis of the movement of the ball, the impact of the network. We discussed the weighted form of clustering coefficients to describe the tendency of the team to form a balanced triangle between players and measure local robustness.

Secondly, considering the flow attributes of the passing network itself, this paper sets an H index to measure and evaluate the performance of the entire team cooperation, evaluates the team level by considering four parameters, and uses the Analytic Hierarchy Process to build an evaluation system model to evaluate the effectiveness of team cooperation. The enemy-self analysis of the movement trajectories of various types of balls shows that the strategy is universal and versatile, and has little to do with the opponent's strategy.

Finally, this paper analyzes the impact of key nodes on the results of the game through the construction of the passing network, and obtains the best strategy for teamwork.

## 2. Pass Network Model

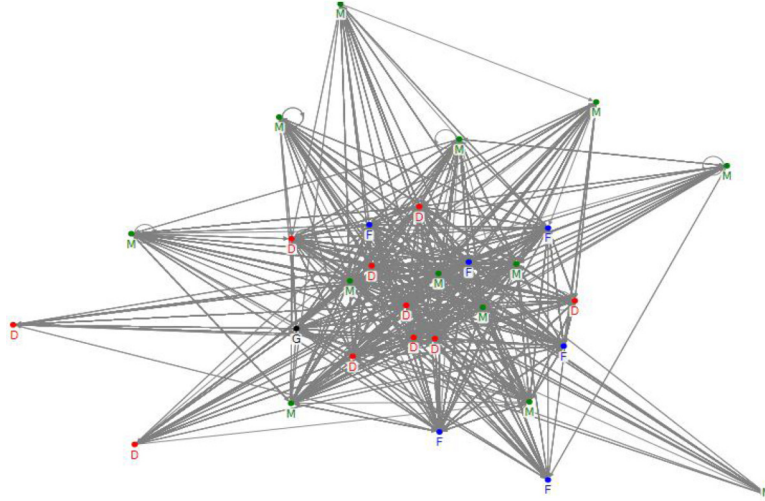
We follow the steps shown in Figure 1 to build a passing network structure framework.



**Figure 1.** Structural framework.

Visually analyze the data, portray all players based on graph theory, and create a pass network graph  $G$ . Graph  $G$  refers to a binary  $(V(G), E(G))$ , where  $V(G) = \{v_1, v_2, \dots, v_k\}$ : represents the collection of all players in the passing network, each player is regarded as a node  $v_i (1 \leq i \leq k)$ ,  $k$  is the order of the graph  $G$ , that is the number of players, and  $E(G) = \{ \langle p_i \rightarrow p_j \rangle | p_i \in V, p_j \in V \}$  represents the passing relationship between any vertices (players). The collection, for example,  $p_i \rightarrow p_j$  means player  $p_i$  pass to  $p_j$ . The network reflects the relationship between team players' various actions and the ball movement as much as possible. Because different players have different positions and influences in the passing network, we set different vertices to distinguish different players in the passing network.

The pass network is a directed graph, as shown in Figure 2 below.



**Figure 2.** Passing network.

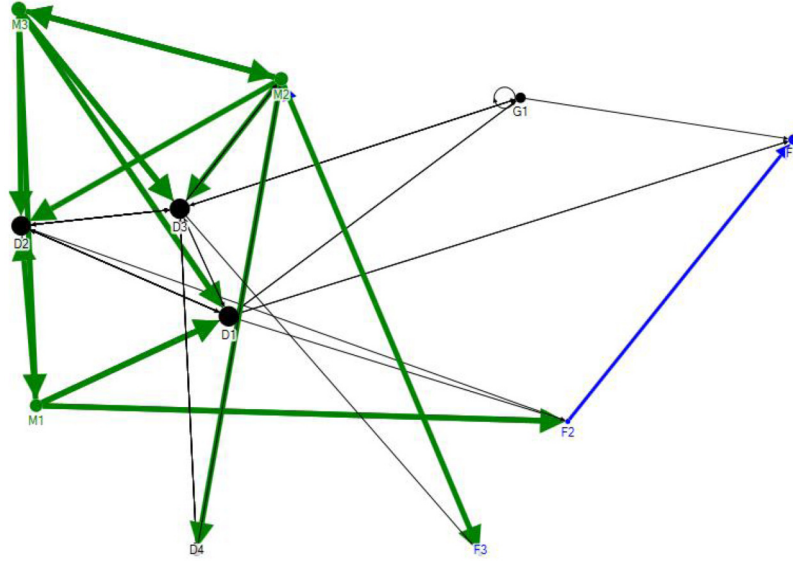
Degree centrality (CD) describes the number of connected edges of a node, and characterizes the ability of a node to communicate directly with other nodes from the local characteristics of the network. The formal description is:

$$CD(\nu) = \frac{\deg(\nu)}{N - 1}, \quad (1)$$

$$\deg(\nu) = \sum_{j=1}^N x_{\nu j} (\nu \neq j). \quad (2)$$

Equation (2) represents the number of edges connected to a node  $\nu$  in the network.  $\sum_{j=1}^N x_{\nu j}$  is used to calculate the number of direct edges between node  $\nu$  and other  $N - 1$  nodes. If  $\nu$  is associated with  $j$ , and  $\nu \neq j$ , then  $x_{\nu j} = 1$ , otherwise  $x_{\nu j} = 0$ . The larger the value  $CD(\nu)$ , the stronger the communication ability of the node has. The more passes between players, the more important it is in the network. Figure 3 is an example of the network obtained, taken from the first game. Each node is marked with

the player's name. Only members of the team within the specific time considered are shown here. The size of the node depends on the degree center, and different shades correspond to the players with different responsibilities.



**Figure 3.** The distinction between player roles.

### 3. Cluster Analysis

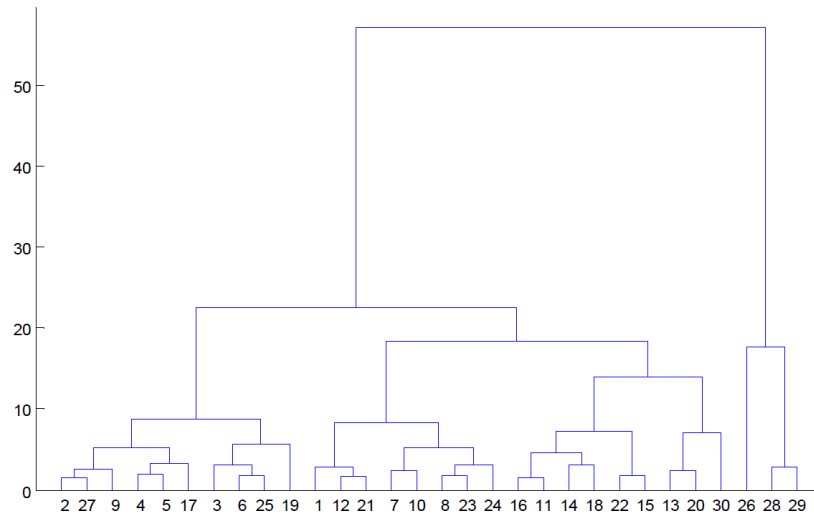
#### 3.1. Basic idea

We use a clustering model to classify the passing network, and use the local clustering coefficient of the node  $i$  as the percentage of the nodes directly connected to it. Between them, the degree can be averaged along  $N$  nodes of the network to obtain the average clustering. We use weighted clustering coefficients  $C_w(i)$  to measure the likelihood of a given player  $i$  linking to those around him.

$$C_w(i) = \frac{\sum_{j,k} w_{ij} w_{jk} w_{ik}}{\sum_{j,k} w_{ij} w_{ik}}, \quad (3)$$

where  $j$  and  $k$  are any two players of the team,  $w_{ij}$  and  $w_{ik}$  are the number of passes between the third player  $i$  and the first two players. We average all participants' clustering coefficients  $C_w(i)$ , according to a formula, which is:  $C = \frac{1}{N} \sum_{i=1}^N C_w(i)$  to obtain the clustering coefficient of the entire passing network. Here, the weighted form of the clustering coefficient describes the tendency of the team to form a balanced triangle between players, which is a method of measuring local robustness.

The idea of cluster analysis believes that there are similarities and intimacy between the pass data and indicators we studied. The entire classification system, as shown in Figure 4, represents the affinity between all indicators. The Brown's passing network has the characteristics of high density and short distance. Most players have had direct contact with each other, and the Browns' passing trend is obvious.



**Figure 4.** Cluster analysis diagram.

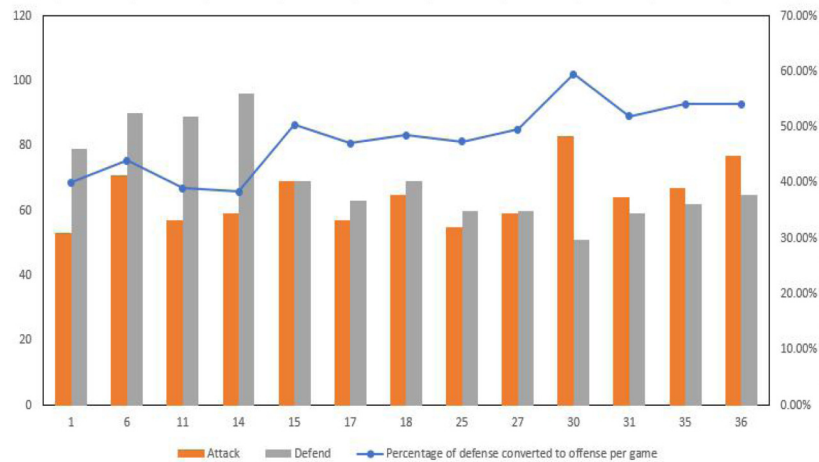
#### 4. The Main Parameters

Analytic Hierarchy Process is used to evaluate the index system in the passing network. Using the main evaluation model, coupled with qualitative analysis of team-level processes, conclusions about the structure, configuration, and dynamics of teamwork are obtained. We drew up four parameters to create the model, extracted a set of performance indicators based on passing, and summarized it up as H indicators to measure and evaluate the cooperation performance of the entire team.

$$H = \{ \text{Ball right, Flexibility, Liquidity, Key nodes} \}. \quad (4)$$

##### 4.1. Ball right

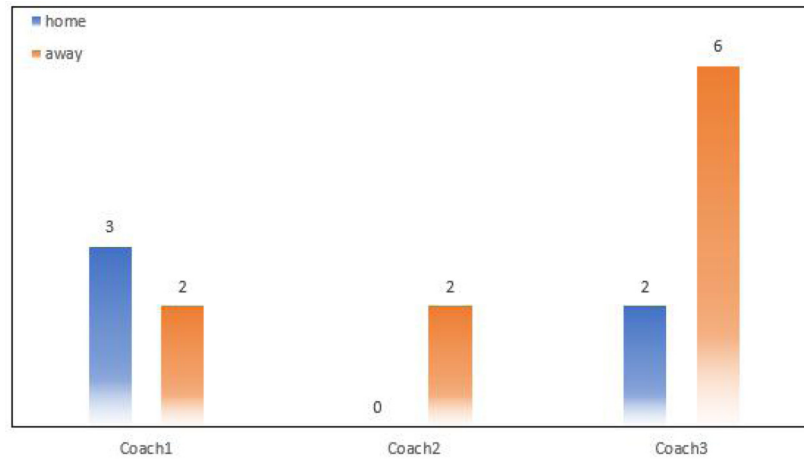
Defense is the key to football games, but if you can get the right to the ball and mobilize the offensive, you can make the team have more offensive initiative. We take the ratio of the team's total number of defense-to-offense to the total number of defense-to-defense as the proportion of defense-to-turnoff. As shown in Figure 5, personal data is a manifestation of single-point defense ability, and overall team data is a manifestation of team style.



**Figure 5.** Percent turn defense per game.

## 4.2. Flexibility

We realized from Figure 6 that the passing strategy frequently used by the team can disrupt the opponent's game mode. We constantly change the strategy based on the actual situation, and flexibly use tactics according to the nature of the "stream" in the network.

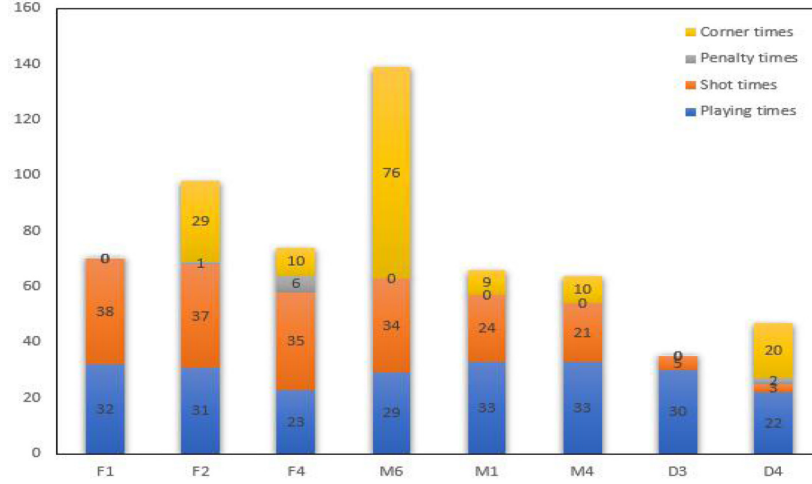


**Figure 6.** Impact of different strategies on results.

## 4.3. Liquidity

Players in different positions have different tasks and have different performance in player interaction. As shown in Figure 7, mobility shows the speed at which players pass the ball from the back to the front.





**Figure 7.** Number of events for different players in a game.

#### 4.4. Key nodes

The PageRank algorithm maps the importance of nodes to a specific number, which can calculate the player's importance in the passing network. Drawing on the PageRank algorithm, this paper proposes an algorithm that calculates the importance of players in the passing network, PlayerRank, which is calculated as

$$PR(p) = (1 - \alpha) + \alpha \sum_{i=1}^k \frac{PR(T_i)}{C(T_i)}, \quad (5)$$

$PR(p)$  indicates the importance of a player  $p$  in the passing network,  $PR(p) \in [0, 10]$ . The larger its value, the higher its importance. Parameter  $\alpha \in [0, 1]$  can adjust the weight proportional relationship between the two attributes. Let  $n$  denote the number of players (vertices) in the passing network;  $T_i (i = 1, 2, \dots, n)$  represents the set of players passing to the player;  $C(T_i)$  is the number of players  $T_i$  passing out;  $\frac{PR(T_i)}{C(T_i)}$  represents the value of  $PR$  given to player  $p$  by player  $T_i$

passing to player  $p$ . In the PageRank algorithm, a passing network is linked by another passing network with a high  $PR$  value, indicating that the passing network itself is of good quality and its  $PR$  value will also increase.

## 5. Best Strategies

Based on the pass network, we use principal component analysis and spatio-temporal analysis to bring important games to victory, and find that the number of shots and positive shots contributes a lot to the victory, showing a team's desire to attack and ability to reach the opponent's key defense. The number of shots reflects the team's offensive strength, and the number of goals is better than the opponent determines the direction of the game.

The performance of a team depends on the core players. The Brown team accounted for 87% of all team passes in 38 games. Obviously, short-range passing has a primary value as a defensive strategy because it deprives the opponent of possession of the ball, which is the factor that determines the outcome of the game.

## 6. Conclusion

We analyzed our proposed method, compared with other algorithms, the pass network we built based on graph theory is simple and clear, and can fully reflect the situation of all players. Our model is simple and flexible and easy to adapt, with low characteristics and evaluation complexity, which can be extended to applications in other fields.

## References

- [1] L. C. Freeman, A set of measures of centrality based on betweenness, *Sociometry* 40(1) (1977), 35-41.  
DOI: <https://doi.org/10.2307/3033543>
- [2] Yue-Jin Tan, Jun Wu and Hong-Zhong Deng, Evaluation method for node importance based on node contraction in complex networks, *Systems Engineering Theory and Practice* 26(11) (2006), 79-83.
- [3] Jing Chen and Linfu Sun, Evaluation of node importance in complex networks, *Journal of Southwest Jiaotong University* 3 (2009), 132-135.
- [4] Zhang Tingping, *Research on the Algorithm of Node Importance Evaluation in Complex Networks*, Southwest University, 2015.
- [5] S. Wasserman and K. Faust, *Social Network Analysis*, Cambridge University Press, Cambridge, UK, 1994.  
DOI: <https://doi.org/10.1017/CBO9780511815478>
- [6] Li Bo and Wang Lei, Feasibility analysis of social network analysis on passing performance in football match, *Journal of Beijing Sport University* 08 (2017), 117-124.
- [7] Zhang Tao, Yu Jiong, Liao Bin, Guo Binglei, Bian Chen, Wang Yuefei and Liu Yan, The construction and analysis of pass network graph based on GraphX, *Journal of Computer Research and Development* 53(12) (2016), 2729-2752.  
DOI: <https://doi.org/10.7544/issn1000-1239.2016.20160568>
- [8] Xu Xiaofei, *Study on the Classic Offensive Tactics System of American Basketball*, Beijing: Beijing Sport University, 2016, 6.
- [9] Li Zhigao and Liu Jun, Establishment of team-based performance evaluation index system and its application, *Business Research* 13 (2005), 30-33.
- [10] Cao Weihua, Analysis of transmission control characteristics of Tiki-Taka tactical play in Spain based on social network analysis, *Journal of Chengdu Sport University* 4 (2019), 65-72.
- [11] Gu Yingwei and Li Juan, Research on key performance indicator (KPI) system, *Modern Management Science* 6 (2007), 79-80.

