ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202

Analysis of the World Cup Pass Network Based on Complex Network Theory

Shengjie Guo[1], Weiqiang Liang[2]

[1]School of Mathematics and Information Technology, Hebei Normal University of Science and Technology, Qinhuangdao 066099, China
[2]School of Mathematics and Information, South China Agricultural University, Guangzhou, 510642, China

Abstract: *Based on the complex network theory, this paper focuses on the passing network of Argentina and France, measures the robustness of the network from multiple static indicators, and analyzes the changes of multiple indicators of the network through the attack on the network. The analysis shows that the Argentina team has more advantages in the effective scale and connectivity of the network, and the French team has stronger network robustness. Both networks are more robust under random attack and less robust under target attack.*

Keywords: *Passing Network, Complex Network, Robustness, Simulation*

1. Introduction

In the final of the 22nd World Cup in 2022, Argentina and France scored 2: 2 in the regular time of the final. This thrilling competition is wonderful. As the two teams competing for the championship, their strength has been controversial. For the comparison of the strength of the two teams, this article will expand the analysis based on their game data.

In recent years, with the development of research on ' complex networks ', there has been a wave of research on real networks using complex network theory. Complex networked theory emerged as the 1960 s [1]. It can not only be used to study the geometric properties of real networks, but also to study the formation mechanism and evolution law of networks. In a football match, the network of passing between players can be used as a typical real network. Based on the analysis of network theory, it can often have a clear description of network structure and various characteristic attributes.

At present, domestic and foreign scholars have a certain degree of research on football pass network. Liu Zeting et al. proposed the method of integrating the matrix representation of the graph with the football network [2]. Bruno Gonçalves et al. predicted the results of the game based on pass data [3]. Fernandez-Navarro et al. characterized the style of different teams ' games based on the characteristics of team data [4]. McHale et al. proposed the application of identifying key players on the pitch for the player passing network [5].

Therefore, this paper will construct the pass network of Argentina players and French players in the regular time of the World Cup final. Set up multiple static indicators to analyze the different characteristics of the network and its robustness. At the same time, the simulation attack is carried out on the pass network, and the change of the network is observed through the change of the index, so as to explore the pass characteristics of the two teams according to the change trend of the network, and analyze the game data to a certain extent.

2. Network construction

In this paper, through the 2022 Qatar World Cup finals, Argentina against France's regular time players ' pass data, the two teams ' pass network is constructed to analyze the characteristics of the two teams. In this paper, the player passing network is expressed as an undirected weighted graph $G = (V, E)$. According to graph theory, V represents the point set and E represents the edge set [6]. In the player passing network, V denotes the collection of players in a team; E represents a set of passing routes in a team's game; the node number is the player number in the team.

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202

It can be seen that in a team pass network, there are $N = |V|$ nodes and $M = |E|$ edges. The weight wij is given to the total number of passes of each two player on the same edge, that is, the weight w_{ij} is given to the edge (v_i, v_j), so as to construct the adjacency matrix $A=(a_{ij})_{n\times n}$, where

$$a_{ij} = \begin{cases} w_{ij} & (v_i, v_j) \in E \\ 0 & else \end{cases} \tag{1}$$

Figure 1 and Figure 2 are the Argentina player passing network and the French player passing network, as follows:

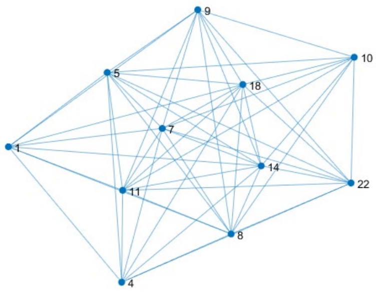


Figure 1: Argentina players pass network

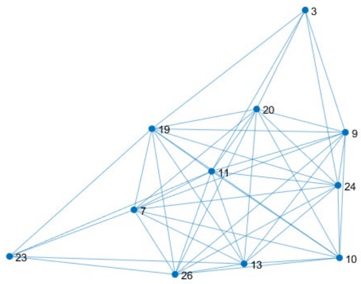


Figure 2: France players pass network

3. Statistical parameters

3.1 Static evaluation methodology

3.1.1 Node average degree and weight distribution difference

In general, the degree k_i of a node v_i in an undirected network is expressed as the number of edges directly connected to the node. The higher the k_i value of the node degree, the higher the network connectivity, the stronger the local bearing capacity of the player as the node. The average degree of all nodes in the network $< k >$ is called the average degree of the network. In the player's pass network, the number of passes on edge represents the weight w_{ij} of the edge, and the point weight H_i with the player as the node represents the total number of passes in the pass network.

$$k_i = \sum_{j=1}^{N} a_{ij} = \sum_{j=1}^{N} a_{ji} \tag{2}$$

$$<k> = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{1}{N}\sum_{i,j=1}^{N} a_{ij} \tag{3}$$

$$H_i = \sum_{j\in N_i} w_{ij} \tag{4}$$

In this paper, the weight distribution difference Y_i of node v_i is used to reflect the dispersion degree of edge weight distribution. If the edge weight of each node is very different, then Y_i tends to 1, which reflects that the weight of an edge plays a major role [7]. At this time, the player 's passing is mostly concentrated on this line, and the interception of any point on the edge will cause a direct interruption in the network, indicating that the weaker the robustness. Therefore, the reciprocal X of the weight distribution difference Y_i is used to represent the local robustness of the network. The larger the X, the stronger the network robustness.

$$Y_i = \sum_{j\in N_i} \left(\frac{w_{ij}}{H_i}\right)^2 \tag{5}$$

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202

$$X = \frac{N}{\sum_{i=1}^{N} Y_i} = \frac{N}{\sum_{i=1}^{N} \sum_{j \in N_i} \left(\frac{w_{ij}}{H_i}\right)^2} \tag{6}$$

3.1.2 Node clustering coefficient and network connectivity

In general, the clustering coefficient C_i of node v_i represents the ratio of the actual number of edges m_i to the possible maximum number of edges between all adjacent nodes of node v_i. The larger the node clustering coefficient C_i, the higher the density of the network node distribution. The average clustering coefficient of all nodes in the network $< C >$is called the average clustering coefficient of the network.

$$C_i = \frac{2m_i}{k_i(k_i - 1)} \tag{7}$$

$$< C > = \frac{1}{N}\sum_{i=1}^{N} C_i = \frac{2}{N}\sum_{i=1}^{N} \frac{m_i}{k_i(k_i - 1)} \tag{8}$$

On this basis, compared with other large networks with more than 1000 nodes and far more edges than nodes, the player passing network can only be a small network [8]. The average clustering coefficient difference $< C >$ in the small network is small, so this paper uses the connectivity γ as the evaluation index of the player passing network. In theory, there is little difference between connectivity and clustering coefficient. Starting from different points, connectivity is not applied to the domain of a node within the network, but the entire network.

$$\gamma = \frac{|M|}{3|N| - 6} \tag{9}$$

3.1.3 Average path length of nodes and network efficiency

In general, the distance d_{ij} between node v_i and v_j is defined as the number of edges on the shortest path connecting two nodes. The average path length L is the average distance between any two nodes, namely

$$L = \frac{1}{\frac{1}{2}N(N-1)}\sum_{i \geq j} d_{ij} \tag{10}$$

In the player passing network, the shorter the average path length, the shorter the passing distance between players, the closer the passing connection in the network, the more stable the network structure. The network diameter D represents the maximum distance, reflecting the size of a network. The standardized network diameter $L_b = L/D$, using 1-L_b to measure the average path length of different scale networks, the larger the value, the stronger the network robustness [7].

Large networks are often disconnected and often contain many small connected network sub graphs. If two nodes belong to different small connected sub graphs, the distance between them will be infinite because there is no edge connection. Therefore, the global network efficiency E is introduced, that is,

$$E = \frac{1}{L} = \frac{2}{N(N-1)}\sum_{i \geq j} \frac{1}{d_{ij}} \tag{11}$$

3.1.4 Assortative coefficient of nodes

If there is an edge between two nodes in the network is independent of the degree of the two nodes, that is, the degree of the two endpoints of a random edge in the network is random, then there is

$$e_{jk} = q_j q_k \quad \forall j, k \tag{12}$$

The network is said to have no degree correlation, or the network is neutral. Conversely, the network is said to have degree correlation. For the degree-related network, if the nodes with large degree tend to

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202

be connected to the nodes with large degree, then the network is said to be positively correlated, or the network is said to be assortative. Conversely, the network is called mismatched. In general, to characterize whether a network is assortative or disassortative, it is often measured by the value of the assortative coefficient *R*, that is,

$$R = \frac{1}{\sigma_q^2} \sum_{j,k} jk\left(e_{jk} - q_j q_k\right) \tag{13}$$

$$\sigma_q^2 = \sum_k k^2 q_k^2 - \left[\sum_k k q_k\right]^2 \tag{14}$$

If $R > 0$, the network is assortative; if $R < 0$, the network is mismatched. The size of $|R|$ reflects the strength of network assortativity. Assortative matching means that nodes with similar attributes are easy to approach each other. In the player passing network, players are more inclined to pass the ball to players in the same position or in the front and back layers. Few players pass the ball to players with large differences in location attributes, which greatly reduces the risk of network connectivity being interrupted, making network connectivity more stable and network robustness stronger.

3.1.5 Betweenness of nodes

In general, the betweenness BC_i of node v_i is defined as the number of the shortest paths of the node. In the player passing network, from the perspective of football passing transmission, the higher the betweenness, the greater the importance of player nodes, and the more significant the impact of removing these nodes on the connectivity of the entire network.

$$BC_i = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}} \tag{15}$$

Among them, g_{st} is the number of the shortest paths from node v_s to node v_t, and n^i_{st} is the number of shortest paths through node v_i in g_{st} shortest paths from node v_s to node v_t.

3.1.6 Cycles and robustness

In general, the robustness of a network often reflects the stability and robustness of the network. In this paper, the robustness of the player's passing network is: when the player's passing edge encounters a deliberate attack or a random attack, a team can have other lines to ensure the ability of the passing network to continue to connect. For describing this ability, this paper uses the number of cycles μ defined by Berge.

$$\mu = |M| - |N| + 1 \tag{16}$$

However, in theory, the number of circles often increases with the increase of network size, because the increase of network nodes makes the probability of network encountering attack events increased. Therefore, when we evaluate the robustness of a network, we should consider reducing the impact of network size, and select μ^T as a further evaluation index of robustness.

$$\mu^T = \frac{\mu}{|N|} \tag{17}$$

3.2 Dynamic evaluation methodology

3.2.1 Relative size of maximum connected subgraph

In general, after some nodes in the network fail, the network is still connected, indicating that the node failure fault is robust. This change in network performance after node failure is usually described by the relative size *S* of the largest connected subgraph. The larger the value, the more complete the network composed of the remaining nodes and the smaller the loss of network performance.

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202

$$C_r(v_i) = \frac{N_i^R}{N} \tag{18}$$

N_i^R represents the number of nodes in the largest connected subgraph after deleting node v_i.

3.2.2 Attack strategy of network

Considering that it is impossible to delete nodes in the network in a football match, in other words, the attack should not act on the players in the player's passing network, but each attack should be aimed at deleting the passing edge, which is also in line with the practical significance of interception in football matches.

For the random attack strategy, each attack randomly deletes an edge in the network until the network is decomposed. For the target attack strategy, each attack targets the node with the largest betweenness BC_i, and then deletes one of the largest paths in the number of all passing edges of the node until the network is decomposed.

4. Network robustness analysis

4.1 Static index analysis

Through the static parameters proposed above, the parameters are obtained according to the pass data of Argentina and France players, as shown in table 1:

Table 1: Statistical parameters of pass network between Argentina and France players

name	*<k>*	*<C>*	1-L_b	*R*	*X*	γ	μ^T	*E*
argentina	8.364	7.3159	0.6764	-0.1236	5.5533	16.4815	39.5455	10.6293
france	9.273	5.3816	0.7562	-0.0584	5.8420	13.4074	32	8.6225

First of all, through the calculation, although the node N is 11 the pass side number M is different. The Argentina team pass side number is 46, the French team is 51. At the same time, the total number of passes between Argentina and France in the game is also different, which is 445 and 362 respectively, so the size of the network is also different.

The average degree can measure the robustness of the network. From the above data, the average degree of the French team's network is greater than that of the Argentine team. Therefore, the French team's player passing network is more robust and stable. However, according to the analysis of the static parameters of Feng Chun [7], the size of a network's robustness cannot be measured only by the average degree of the network. Some data show that the less the number of nodes and edges, the more it can highlight the robustness of a network. For this reason, this paper introduces the network weight distribution difference X_i to further compare the robustness of the two teams. Argentina team network weight distribution difference X = 5.5533, French team weight Distribution difference X = 5.8420. Through comparative analysis, it is found that the difference in weight distribution of the French team is greater than that of the Argentine team. Therefore, it can be seen that the overall distribution of the French team 's players ' passing route is uniform, with more paths to replace and similar path lengths. The network structure is more stable and the robustness is stronger.

In the numerical performance of $< C >$ and γ, starting from the data : the aggregation effect of the Argentine team is much larger than that of the French team, indicating that the Argentine team pays more attention to the passing between players, and the scale of network connectivity will be further enlarged. However, since R is negative, it indicates that the player 's passing network is generally mismatched, reflecting that the player is more inclined to pass to players with large differences in position attributes. The performance of most Argentina teams is that although the scale of passing between players is large, the average number of passes received by players is large, but it is more for long passes, with the risk of network connectivity being interrupted. However, due to the small value of R, to a certain extent, it shows that the probability of interruption risk is small. Therefore, although the number of interruptions will exist under the network scale with a large number of connections, the reflection of probability and frequency is also small. It can also explain the performance and fact that the Argentine team player network efficiency E = 10.6293 is also much larger than the French team E = 8.6225.

Finally, according to the 1-L_b two teams were 0.6764 and 0.7562 and the number of μ^T, two teams were 39.5455 and 32 respectively. It is further proved that the French team network is more robust than the Argentine team, but the overall difference is not significant.

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202

In general, it can be concluded that the Argentine team has a greater advantage in the effective scale and connectivity of the network than the French team, but the French team's network is more robust. That is, in the face of attacks, the effective organization path has some advantages in reconnecting the network, but there is little difference between the two teams as a whole, that is, both tend to be stable. The Argentine team has a larger base of receiving and passing, more times of controlling the ball, more network connectivity and paths. Therefore, the Argentine team has a more obvious distinction than the French team under the overall network operation efficiency.

4.2 Dynamic simulation analysis

According to the two attack methods mentioned above: target attack and random attack to the Argentina team passing network and the French team passing network. Among them, the horizontal axis represents the proportion of the number of attacks to the total number of attacks caused by the overall decomposition of the network through the index *f*. The vertical axis represents the average path length, weight distribution difference, average clustering coefficient, assortative coefficient, relative size of the largest connected subgraph, and network efficiency through indicators *L*, *X*, *C*, *R*, *S*, and *E*, respectively.

4.2.1 Analysis of dynamic parameters of Argentina players' passing network

From the Figure7, it can be seen that the initial maximum connected subgraph *S* of the Argentina players ' passing network is 1. Under the action of the two attack modes, *S* can remain stable at a certain stage. But the division of stages is obviously different. Under the target attack, when $f < 0.906$, the network as a whole can maintain the stability under connectivity; under random attacks, when $f = 0.760$, the stage of maintaining connectivity of the whole network ends. In the subsequent two attack modes, it can also be found that the number of attacks on the network in the random attack mode to maintain the same maximum connected subgraph size as in the target attack is less than the number of attacks on the target attack. This shows that the passing network is less robust under random attack, and the network structure is more stable under target attack.

It can be seen from the Figure 4 that with the increase of the number of attacks, the difference of network weight distribution is obviously different in the two attack modes. In the target attack mode, the weight distribution difference maintains an upward trend, indicating that the network has more paths to replace and the path length is similar. The more stable the network structure is, the stronger the robustness is. Correspondingly, under random attacks, the network becomes more vulnerable and on the verge of collapse. It further proves that the network has poor robustness under random attack and good characteristics under target attack.

It can be seen from Figure 3 that due to the weight of the attack on the edge, the overall edge of the network remains connected, and the average path length of the network basically decreases proportionally with the increase of attacks.

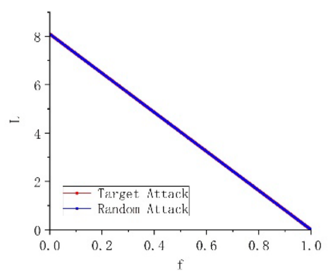


Figure 3: Argentina's L changes

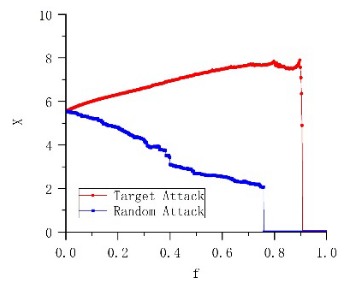


Figure 4: Argentina's X changes

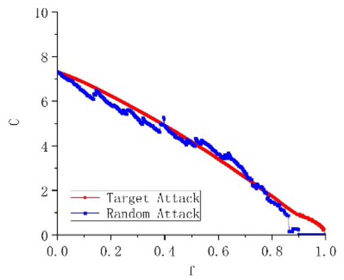


Figure 5: Argentina's C changes

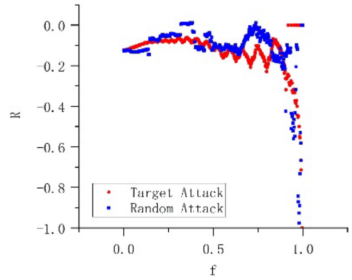
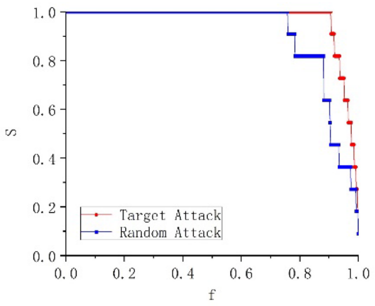


Figure 6: Argentina's R changes

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202



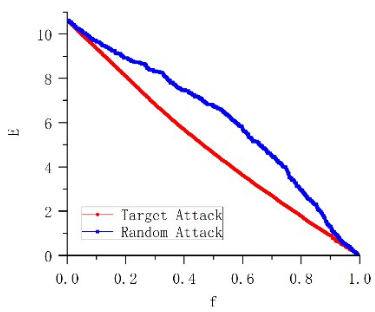


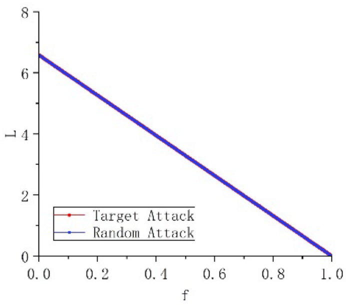
Figure 7: Argentina's S changes *Figure 8: Argentina's E changes*

From the Figure6, it can be seen that R of the network belongs to the disassortative network as a whole with the increase of the number of attacks, but the local change is very unstable, with the more f tends to 1, it can be seen that the greater the fluctuation. In $f \in [0.6404 , 0.8202]$, the assortative coefficient under random attack is higher than that under target attack, indicating that in the middle and late stages of the number of attacks, the target attack against the network has a more differentiated degree of attack damage to the network edge, which greatly increases the risk of network interruption.

It can be seen from the Figure5 that with the increase of the number of damaged edges, the network connectivity shows a downward trend as a whole. However, the difference is that the degree of network agglomeration effect under target attack is obviously different from that under random attack, which is a very stable decline; random attack leads to a small drop in connectivity. It is worth noting that when $f \in [0.5146 , 0.7191]$, the connectivity of the network under the random attack is slightly higher than that of the target attack, and the overall interval is closer to the assortative coefficient. It also further shows that the target attack has a significant effect on the network connectivity in this interval, and the risk of network interruption is greatly increased.

From the Figure 8, the initial network efficiency is 10.6293. Under the attack, the change of network efficiency generally shows a downward trend, but then the attack makes the speed of network efficiency change from slow to fast, and the speed change from fast to slow under the target attack. When $f = 0.5$, the efficiency under random attack is 6.738, and the decrease proportion is 36.61 %. Under the target attack, the efficiency is 4.621, and the decline proportion is 56.53 %, which indicates that the network has been seriously damaged.

4.2.2 Analysis of dynamic parameters of French players ' passing network



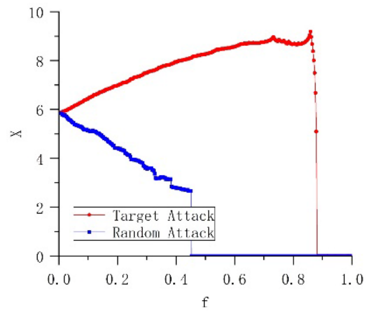
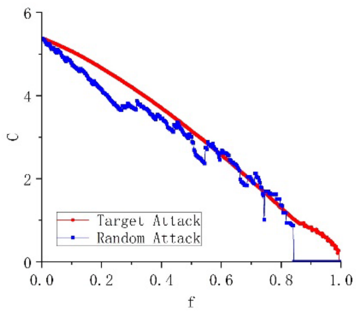


Figure 9: French's L changes *Figure 10: French's X changes*



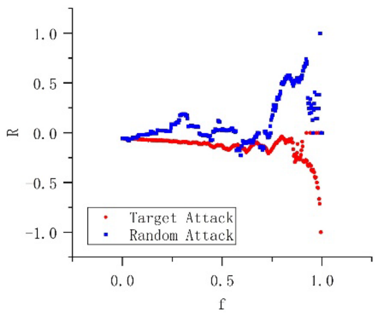
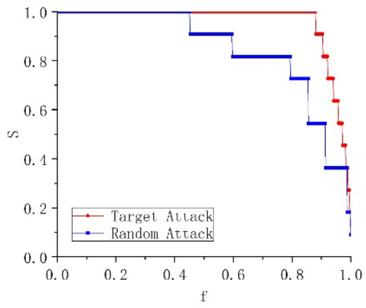


Figure 11: French's C changes *Figure 12: French's R changes*

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202



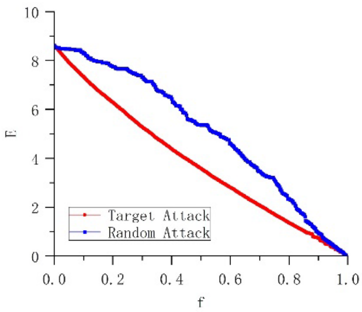


Figure 13: French's S changes *Figure 14: French's E changes*

It can be seen from Figure13 that *S* of the network can maintain a certain value under different attack times, but in the two attack modes, the threshold of the number of attacks that each maintains is significantly different. When the network maintains the initial maximum connected subgraph, it can withstand 319 attacks under the target attack. At this time, $f = 0.8785$, indicating that most attacks cannot cause disruption to the network. Under random attacks, when $f = 0.4502$, that is 164 attacks, the network cannot maintain the previous complete connectivity. Under the subsequent attacks of the two modes, it can be found that the number of attacks that the target attack can carry is always greater than or equal to the random attack mode. The overall results show that the French players' pass network has poor robustness under random attack, while the network structure is more stable under target attack. The conclusion of Figure 9 is the same as the Argentina team above.

In Figure 10, with the increasing number of attacks, the difference of network weight distribution is obviously different in the two attack modes. In the target attack mode, when the difference of weight distribution is $f \leq 0.8591$, the overall upward trend is maintained, indicating that the network has more paths to replace and the path length is similar under the attack. The more stable the network structure is, the stronger the robustness gradually becomes. However, the difference of weight distribution in $f \in [0.8591\ ,\ 0.8812]$ decreases sharply, X changes from 9.1879 to 0. Compared with it, the difference of network weight distribution under random attack decreases from 5.842 to 2.6625 under 45 % before attack. A subsequent attack directly makes *X* become 0 until the network is decomposed. It shows that the network becomes more fragile and on the verge of collapse more quickly under the random attack. The data analysis is consistent with the conclusion of the maximum connected subgraph size, which further proves the above characteristics.

It can be seen from Figure 12 that *R* of the network becomes more obvious with the increase of the attacks. Under random attack, the network as a whole belongs to the assortative network; under the target attack, the whole network belongs to the mismatch network. Before $f = 0.6547$, the coefficient values in the two modes are not much different, but after this, as the number of attacks enters the middle and late stages, the difference in the coefficient value becomes more obvious. Then the attack assortativity coefficient gradually tends to 0.74, then decreases slightly and finally stabilizes around 0.3. Under the attack of the target, the value of the mismatch coefficient decreases and stabilizes in the trend of $f < 0.497$, and then until $f = 0.93$, the value of the mismatch coefficient is unstable, but the overall trend is down and down. Finally, as the attack enters the later stage of network decomposition, the degree of network mismatch becomes more obvious.

From Figure 11, with the increase of the number of damaged edges in the network, the network connectivity shows a downward trend as a whole, but the difference is that the degree of network agglomeration effect under target attack is obviously different from that under random attack, which is a stable decline: random attack leads to a small drop in connectivity.

From Figure 14, the initial *E* is 8.6225. Under the attack, the change of network efficiency generally shows a downward trend, but then the attack makes the speed of network efficiency change from slow to fast, and the speed change from fast to slow under the target attack. Under the target attack, the efficiency is 3.5798, and the decline proportion is 58.48 %, which indicates that the network has been seriously damaged.

5. Conclusions

In the final of the 22nd World Cup, Argentina scored 2: 2 in the main game against France. As the last two champions, the strength of the two teams is relatively stable, so there is also the significance of

ISSN 2616-5805 Vol. 4, Issue 2: 10-18, DOI: 10.25236/AJMS.2023.040202

analyzing the pass data of the two teams.

Through the analysis of the static indicators and dynamic simulation of the two players' passing network above, it can be found that the Argentina team has stronger connectivity and the French team has more advantageous robustness. In the face of cyber attacks, the French team has more alternative paths to reorganize the connection of the pass, and the network matching coefficient after the attack is more obvious. After sacrificing some of the pass routes, the network class structure is clearer, and the risk of the ball being interrupted is greatly reduced, and the remaining network is more stable. Compared with the French team, the Argentine team network shows different characteristics. In the face of attacks, the network can carry more attacks to maintain the connectivity of the overall network, maintain the whole team together, have more pass route choices, and organize the team's effective layout and defensive attack.

References

[1] Liu Jianxiang. Complex networks and their research progress in China [J]. Journal of Systems Science, 2009, 17 (04): 31-37.
[2] Liu Zeting, Ge Huanmin, Shen Yanfei. Teaching Exploration and Practice on the Integration of Graph Theory and Passing Network [J]. College Mathematics, 2022, 38 (01): 20-25.
[3] Gonçalves Bruno, Coutinho Diogo, Santos Sara, Lago-Penas Carlos, Jiménez Sergio, Sampaio Jaime. Exploring Team Passing Networks and Player Movement Dynamics in Youth Association Football [J]. PloS one, 2017, 12(1).
[4] Fernandez-Navarro Javier, Fradua Luis, Zubillaga Asier, Ford Paul R, McRobert Allistair P. Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams [J]. Journal of sports sciences, 2016, 34(24).
[5] Ian G. McHale, Samuel D. Relton. Identifying key players in soccer teams using network analysis and pass difficulty [J]. European Journal of Operational Research, 2018, 268(1).
[6] Wang L , Lin K , Zhang W , et al. Invulnerability Analysis of Power Distribution Network Based on Topology Structure[C]// 2019 6th International Conference on Systems and Informatics (ICSAI). 2019.
[7] Feng Chun, Zhu Qian, Yu Bao. Urban rail transit network robustness simulation [J]. Computer simulation, 2018, 35 (10): 182-186 + 461.
[8] Lai Liping. Research on the characteristics of subway network based on complex network theory [J]. Journal of Natural Science, Harbin Normal University, 2016, 32 (06): 30-33.