Modeling for Professional Athletes’ Social Networks Based on Statistical Machine Learning

QIAN HUANG, BIN LIU, NINGSHE ZHAO, ZHIYONG ZHANG, QIURONG WANG, XIALI ZHANG, YANG XUN, XIAOYU GE, JIANLAN DING, AND SAI-FU FUNG

1Xi’an Physical Education University, Xi’an 710068, China
2Center of Statistical Research, School of Statistics, Southwest University of Finance and Economics, Chengdu 611130, China
3School of Information Engineering, Xi’an University, Xi’an 710065, China
4Research Institute of Sports Science, Xi’an Physical Education University, Xi’an 710068, China
5Physical Education College, Xi’an Physical Education University, Xi’an 710068, China
6School of Economics and Finance, Xi’an Jiaotong University, Xi’an 710049, China
7Shanghai University of Sport, Shanghai 200444, China
8Department of Sports Economy and Sports Management, Xi’an Physical Education University, Xi’an 710068, China
9Department of Social and Behavioural Sciences, City University of Hong Kong, Hong Kong

Corresponding author: Bin Liu (e-mail: mrlibiin@yeah.net)

This work was supported in part by the National Natural Science Foundation of China under Grant 61440036.

ABSTRACT Professional athletes’ social networks have rarely been explored. In this paper, we propose a framework to analyze athletes’ social networks. Based on the characteristics of athletes’ social network structure and social support theory, the matrix distribution is introduced to describe the network structure. The observation model of the Bayesian network is established, and then the Gaussian process analysis model of sparse matrix is used to investigated the network. We collected real-world data of athletes’ social networks by questionnaires, which contain eight thematic network data. With our method, the interpersonal network of professional athletes is analyzed and the adjacency relationships are predicted. Finally, taking the social subnet of the athlete social network as an example and using the model and algorithm, the node support factor analysis and the complex network community convergence factors are analyzed. We found that professional athletes’ social networks have a stronger small-world characteristic than the general public’s social networks. The proposed model and algorithm provide a new quantitative approach for studying professional athletes’ social networks.

INDEX TERMS Bayesian network, matrix distribution, professional athletes, social network.

I. INTRODUCTION

In recent years, social network has gradually become a hot topic for research. Those studies focused mainly on social network model and social network analysis. The majority of social network models are in line with the actual social network and the international Watts-Strogatz small world network model while domestic scholars, using psychological analysis tools, put forward the social network model of information exchange. In general, the social network model form consists of three types: graph model, algebra model, and block model. The research topics in social network model studies at home and abroad mainly include information dissemination, knowledge transfer, academic topics, and organizational learning. The research on sports social networks has only begun in recent years overseas. There is little research on social networks in China in general and there is even less research on professional athletes’ social networks.

The training of professional athletes in China mainly uses the government-led training system, planned under the city or county-provincial-state guidance. Having limited means of communication and small numbers of total participants, resulting in a relative small size of the social network and a high density, athletes social network research is imperative [1].

Previously, our group described the characteristics of the social network structure of Chinese athletes [2], [3], opened up the domestic under the perspective of complex networks...
on the athletes’ social network and their social support research (based on relational variables of structural analysis). We think that the athletes’ social support network should be constructed from the individual-center social network [3]–[6] and the overall network [1], [3], [7]–[10]. Not only the definition, connotation and classification of individual social support of athletes are expounded [2], but also the impact of individual characteristics on the structure of social support network is studied [5], and some qualitative conclusions are obtained. The overall network characteristics of the athletes’ social support network are analyzed from the two-party relationship and the tripartite relationship [10]. At the same time, with the popularization of mobile internet in recent years, the online sports social forms of organizing sports activities in the form of “enthusiast circle” have gradually increased with mobile location recording software tools. This form provides a better way for sports social data collection, but relatively, in the domestic professional athletes use online sports social way is still relatively small.

Matrix distribution is a kind of important distribution in statistics, which plays an important role in multivariate statistical analysis in physics, economics, psychology and other fields, and has received high attention in the field of computer science. For social network analysis, by introducing matrix distribution to describe the network structure the internal correlation between nodes can be described, and by introducing non-parametric the nonlinearity of network structure can be described with Gaussian process. As an important branch of social network analysis based on information science, using machine learning as a means to study the relationship between social network entities and the evolution of ideas and behaviors is a new and effective way. Statistical machine learning, especially the algorithms based on Bayesian method, can greatly improve the accuracy of community discovery and prediction of unknown network connections by assuming a certain probability model, which can infer the potential network structure and the underlying factors that form the network from the observation of the network connection relationship. Based on the description of matrix distribution, for social network analysis, the use of machine learning means is helpful to find the implicit relationship of athletes’ social network and to construct and analyze the network in depth.

The rest of this paper is organized as follows. Firstly, in section 2, the present research situation and the research methods used in this paper are introduced. Secondly, Section 3 gives the basic description model of the Professional Athlete social support network, including the network structure description, the node relation and the multi-relationship description between the nodes. Then Section 4 puts forward the observation model and analysis algorithm of professional athlete network based on matrix distribution description. In section 5, the relationship prediction process and algorithm analysis in the deep construction of professional athlete network are given. And section 6 gives the application of the above proposed model in specific networks, including the analysis of the influence factor analysis of athletes’ social support network and the cause of network community division. Finally, in section 7 a retrospective summary of the full work is given.

II. CURRENT STATUS OF RESEARCH AND RELATED WORKS

The traditional social network analysis method pays too much attention to the statistical characteristics of the network, because the social network is sparse in nature (that is, an individual is only connected with a very small number of other individuals in the network) and the nonlinear and local interdependence of individual interaction within the network [11]–[13]. Therefore, traditional social network analysis is insufficient in discovering the structure of social networks and the evolution of dynamic behavior. As an important branch of social network analysis based on information science, using statistical machine learning as a means to study the relationship between social network entities and the evolution of ideas and behaviors is a new and effective idea [12], [14], [15].

In the study of social network, because the rule network model is basically an ideal model of some characteristic performance, the scholars put forward a variety of models in order to better imitate and depict the law of real network. In the late 50s, Erdős and Rényi proposed the basic model of stochastic network, although the average distance of stochastic network is small, but the degree distribution satisfies Poisson distribution, and the agglomeration coefficient is not high, which is not consistent with the characteristics of the actual network. At the last century, Watts and Strogatz constructed a Small World network model, which has a smaller average distance and a higher agglomeration coefficient, similar to the actual social network. Almost at the same time, Barabási and Albert proposed a scale-free network model for online links, which can express the network process of “Matthew Effect”. Although in the above model research, the predecessors have made outstanding contributions, but the existing limitations can not be ignored. For example, the Small World Network model can be used to depict the social network of professional athletes, but because of the static means of network construction makes the network structure is a kind of explicit presentation, this study intends to introduce link prediction into the method of machine learning, so that the construction of network model from the explicit and recessive two parts of the combination to implement.

On the other hand, at present, the academic circles are mostly based on the research of network structure, and less research on network content, especially for the participation attribute of network nodes. Node is the most active factor in social network, and athletes’ social network involves more node attributes, which reflect the actor information [16]. From the point of view of network participation, there are abundant information available for research, such as athletes’ emotional and love networks. The team has also done some
research on the content information fusion of athletes’ network in the follow-up. The social interaction of athletes is limited by its scope, which often presents a social synchronization phenomenon, i.e. attribute convergence. Qi & Zhi proposes an SS model to explain this phenomenon [17], [18]. Kivinen and Tumussen puts forward a kind of more relaxed conditions for social networks to reach consensus by studying the convergence of beliefs about non-Bayesian agents in general social networks [17]. In this small world, the local interaction of athletes also drives the spread of the network [19]–[21]. At the same time, in the social network of athletes, communication will also cause the evolution of network structure [22]. Shan’s team thinks that before the node decides to produce (edge generation), it will experience multiple propagation of the network, that is, it has the effect of cumulative propagation activation.

Statistical machine learning, especially Bayesian-based learning algorithms, by observing the network connectivity and the relationships between potential network structures and potential factors in forming networks, can do inference with an assumption of a certain probability model. Their main advantage is the ability to effectively estimate the uncertainty in the network structure. On the other hand, with the introduction of prior probability, it can make full use of the existing background knowledge or do some preliminary analysis through a simple social network. Result [19]–[23]. From the current application of social networks at home and abroad: current social workers do not fully explore the latest research results in the field of social network analysis, so the research methods are too simple, mostly based on simple network statistics [24]–[27], rarely network structure analysis is effectively combined with statistical analysis based on attribute variables [28], [31]. Throughout the major international conferences in the field of social network analysis in recent years (KDD, SocialCom, AAAI, etc.) [32]–[34], machine learning algorithms have become the main method of social network analysis.

At present, social researchers do not fully explore the latest research results in the field of social network analysis, so research methods are too simple, mostly based on simple network statistical characteristics, and rarely combine network structure analysis with statistical analysis based on attribute variables effectively. Although many studies have put forward some good methods and some useful conclusions for the study of athletes’ social network, the research on the social network structure of athletes is mostly static analysis, and there is a lack of research on network dynamic discovery (group relational variable prediction).

III. BASIC MODEL OF SOCIAL NETWORK FOR PROFESSIONAL ATHLETES

As a special network, social network of professional athletes not only need to give the description of the overall network external characteristics, such as network size, but also need to give a description of the internal characteristics of the network, such as node composition. This section mainly expounds the formal representation details in the network construction of professional athletes.

A. EXPRESSION OF THE SOCIAL NETWORK STRUCTURE OF ATHLETES

Graph is a traditional network representation tool, combined with this study, the network representation here is for the athletes’ social network analysis to provide services, so the formal description should take into account the external and internal characteristics of the network. For a network structure for the group of athletes, in order to facilitate the study, adjacency matrixes are used to describe the network topology, and vectors are used to represent the node structure. The athlete network is expressed by figure G (V, e) which represents the collection of nodes in the network, E represents the collection of edges, and the weights of edges represent the strength of the relationship between nodes. The representation of network structure should be able to depict the state of athletes as completely and comprehensively as possible, and we use column vectors to describe the athletes’ nodes. The advantage is that it can not only express macroscopic and mesoscopic network relations, but also express effective network microscopic information, in addition to good scalability and easy expansion.

For the expression of the basic network structure, this study uses the adjacency matrix to represent it. For matrix x, rows and columns both represent identical athlete nodes, and rows and columns are arranged in the same order, and the features in the matrix are two values, representing the relationship between the actors. In such matrices, the elements of the matrix are “1” or “0”, which represent the existence of the relationship, respectively. In the research of this subject, there are eight kinds of network relationship representations involving the athlete group.

B. ATHLETE NODE NETWORK RELATION REPRESENTATION

The analysis and research of athletes’ network relation mainly discusses the influence analysis of the given social relation, and the relationship strength can be expressed by the weight of the middle edge of figure G. In the statistical method, the covariance matrix can be used to represent the correlation nature and degree between node pairs. In order to infer the connection between any two nodes by using the method of machine learning, the synthetic vector form \( [v_i, v_j] \) is used to represent the edges, and the label values are given to them, and the state between the pairs of any nodes is predicted under the support of the training set.

Matrix elements are represented by where they are located. Given network Y (represented as a matrix of \( n \times m \), where \( n \) represents the number of nodes in the network), the elements of row \( i \) and column \( j \), in Matrix Y are recorded as \( Y_{ij} \). Each lattice value in the matrix has its own “label” or location, and you can clearly see the relationship between the rows and columns that are the athletes’ nodes. If rows and columns represent “athlete nodes” from the same collection of nodes, the
elements in the matrix represent the “relationships” between the nodes. If rows and columns represent “athlete nodes” from a collection of two nodes, the elements in the matrix represent a “relationship” between the nodes in the two node collection. If the row represents an athlete node from a collection of nodes, and column represents the event to which the node belongs, the features in the matrix refer to the case where the node belongs to the event. This defines a “membership network.”

The connection between the nodes of professional social network has sparse characteristics, so in the algorithm representation, sparse matrix is used to store it in most places. In addition, there are also the relationship between the correlation and the multivalued relationship, and the matrix is also used for formal description.

C. MULTI-RELATIONSHIP SYNERGY OF ATHLETES’ SOCIAL NETWORKS

Because the athletes’ group interpersonal network has different structural characteristics, its complexity also behaves differently, these network types often present the cooperation and the competition relationship among the athletes, therefore, the concept, the behavior spread between these networks is also worth studying the content. We use multi-relationship synergy algorithm to study how multi-relationship, such as competition and partnership, affects the interpersonal relationship between athletes and the advantages and disadvantages of training results.

In order to study the multi-relationship synergy of the network, we analyze the connections and causes between the associations, such as for some of the existing Community Division \( \pi = \{C_1, C_2, \cdots, C_n\} \), for any two communities \( C_i, C_j \), the connection edge set is \( E_{ij} = \{e_1, e_2, \cdots, e_k\} \). We analyze the influence of the division of \( \pi \) on the edge of the set \( E_{ij} \), and through the derivation of the classification, we can get the vector \( s = (\omega_1, \omega_2, \cdots, \omega_k) \) about the influence score of the \( E_{ij} \), according to the specified threshold \( u > 0, u \in \mathbb{R}^+ \), for \( |\omega_k| > u \), we can get a subset of the \( E_{ij} \) and \( E_s \). \( E_s \) is the key link we need to focus on. According to the set \( E_s \), we can learn the main influencing factors of forming a community among athletes according to some kind of network relationship.

IV. MODEL AND ALGORITHM OF ATHLETE NETWORK BASED ON MATRIX DISTRIBUTION DESCRIPTION

For a given athlete network \( Y \) (represented as a \( n \times m \) matrix, \( n \) represents the number of nodes in the network), we describe \( Y = X + E \), \( X = U^T W V \), here \( U \) and \( V \) are both matrix (respectively \( n \times d \) and \( m \times d \)), which can be seen as a representation of nodes in the network in hidden space, \( M \) can be regarded as the number of node categories. Each column of \( u \) represents the clustering information for the node, which we call \( U \) the Membership Matrix, and \( W \) is a \( m \times m \) matrix that represents the interaction between categories. \( E \) is noise in the network.

A. NETWORK OBSERVATION MODEL OF ATHLETES BASED ON MATRIX DISTRIBUTION

The connection relationship in the network may be noisy, however the t-distribution has robust characteristics to effectively remove noise, so this study will assume that \( W \) obeys a matrix \( T \) distribution (Matrix-variate t-distribution) with a degree of freedom of \( R \), that is, \( W \sim mt (r, 0, I, L) \), here \( I \) is the unit matrix. We’re going to get \( X \) which obeies a matrix \( T \) distribution of degrees of freedom \( r \), and its covariance matrices \( K \) and \( G \) contain information on the lines and columns of a given network matrix \( Y \), that is \( X \sim mt (r, 0, K, G) \).

For symmetric networks (such as athlete-athlete networks), we assume that \( K \) equals \( G \), and for asymmetric networks (such as athlete-coach networks), it can be assumed that \( K \) is not equal to \( G \), therefore the introduction of covariance matrix can describe the mutual dependence (interdependence) between nodes. On the other hand, if \( K \) and \( G \) are nonlinear covariance matrices, such as covariance matrices defined by Gaussian functions, then the interactive relationship of nodes in the network is a nonlinear relationship. As shown in Figure 1, covariance functions \( K \) and \( G \) describe the association relationships on rows and columns, respectively.

Figure 1 uses row and column covariance functions to describe the similarities between different types of individuals in asymmetric networks, such as athlete-coach networks. Nonlinear covariance functions can be used to describe the nonlinear interaction between individuals.

B. SPARSE MATRIX GAUSSIAN PROCESS ANALYSIS MODEL

Using the attribute information of the edges of the network [11], we can establish a relational model such as the bottom and between attributes:

\[
\begin{align*}
x_{ij} = \beta^T r_{ij} + m_{ij}
\end{align*}
\]

In the above formula, the first part is a linear regression model, \( \beta \) is a regression weight, and the second part is a hidden interactive relational variable \( m_{ij} \) is the element(i,j) in Matrix \( M \). Where matrix \( M \) is determined by a hidden matrix \( U \), while \( x_{ij} \) is a hidden variable corresponding to the edges, and unlike the variable \( y_{ij} \) of the edges, \( x_{ij} \) is a continuous variable.

\[\text{FIGURE 1.} \quad \text{Interactive relationships between nodes.}\]
For $M$, after a given hidden variable $U$ the expression is as follows:

$$
p(M|U) = GP_{d,n}(M; 0, K, G)
= (2\pi)^{-\frac{n^2}{2}} |K|^{-\frac{n}{2}} |G|^{-\frac{n}{2}}
\times \exp \left\{ -\frac{1}{2} Tr \left( K^{-1}MG^{-1}M^TF \right) \right\}
$$

(2)

where, $K_{ij} = k_1(u_i, u_j)$, $G_{ij} = k_2(u_i, u_j)$ $\cdot|\cdot$ is to calculate the matrix determinant operation, $k_1$, $k_2$ are nonlinear kernel functions, specifically defined as $k_1(u_i, u_j) = \varnothing (u_i) \varnothing (u_j)$, here, the $u_i$ is a $d \times 1$ vector, $d$ is the number of hidden communities, $\varnothing (\cdot)$ is a nonlinear mapping function. In this way we can define the relationship between matrix $M$ and $U$ as follows,

$$
M = U^TW
$$

(3)

here the matrix $W$ is a known quantity that can be given a matrix Gaussian priori,

$$
W \sim N_{d,d}(W; 0_{row}, \Omega_{col})
$$

(4)

here, $\Omega_{row}$ and $\Omega_{col}$ are line covariance matrices and column covariance matrices respectively.

1) PROBIT MODEL

The Probit model is depicted that the hidden variables $x_{ij}$ of the edges establish a connection with the variable $y_{ij}$ of the edges by an intermediate variable $z_{ij}$,

$$
p(y_{ij}|z_{ij}, x_{ij}) = \{ \delta (y_{ij} = 1) \delta (z_{ij} > 0) \}
+ \{ \delta (y_{ij} = 0) \delta (z_{ij} > 0 \leq 0) \}
$$

(5)

$$
p(z_{ij}|x_{ij}) = N(z_{ij}|x_{ij}, 1)
$$

(6)

2) U’S PRIORI

For variable $u$, due to the sparse nature of the network, we can give the relational matrix $U$ a Laplace priori as follows,

$$
p(U) = \prod_i \exp (-\lambda U_{ij})
$$

(7)

So far, we can obtain the joint distribution of all variables according to Bayesian theory, that is, the sparse matrix Gaussian process model of this paper,

$$
p(Y, Z, M, \beta, U) = \prod_{1 \leq i, j \leq n} p(y_{ij}|z_{ij}, x_{ij}) p(z_{ij}|x_{ij}) p(\beta)
\times p(M|U) p(U)
$$

(8)

The graph model of the joint distribution indicates that, as shown in Figure 2,

**V. RELATIONSHIP PREDICTION IN THE DEEP CONSTRUCTION OF INDIVIDUAL NETWORK OF ATHLETES**

In the expression of the individual relationship (link) of the Athlete network, on the basis of the static adjacency matrix, the Bayesian network analysis model based on matrix distribution is used, and the network of professional athletes’ interpersonal relationship is analyzed according to the eight thematic networks supported by the questionnaire, and the adjacent relations are predicted. Thus, it is possible to infer the implicit relationship between individuals and realize the deep construction of the whole network.

**A. PROBLEM DESCRIPTION AND ALGORITHM**

The problem can be described as follows: for the 144 valid questionnaires recovered from this study, there are eight kinds of network relationships that need to be analyzed, such as daily help, mood talk, social interaction, income discussion, management discussion, marriage and love discussion, employment help and achievement help, and the relational data matrix can be regarded as a collection of training sets and test sets. The event set is randomly divided by five-weight cross-verification, different training sets and test sets are obtained, and two nodes are selected in all nodes, that is, one of the 20,736 events is taken, and the Bayesian network based on matrix distribution is used to predict the link of nodes.

Suppose the Membership matrix $U$ obeys a sparse distribution, such as the Laplace distribution (Laplace distribution), which allows $u$ to become a sparse matrix. On the other hand, for the attribute characteristics of network nodes, some background knowledge of the network, as well as the network statistical characteristics obtained through complex network analysis, can be used as prior knowledge to ensure the existence of rows or columns in the variance matrix, so as to maximize the use of existing knowledge and thus more effective modeling of social networks. The Bayesian learning algorithm based on matrix distribution is shown in Figure 3.

After the questionnaire processing, the preliminary statistical characteristics of the network were obtained. From the social network modeling process, all the prior knowledge of the social network were acquired. In addition to the statistical attributes of the network, we also obtained the social network system of the individual and network topology, network nodes of the explicit clustering relationship, network noise. A network observation model based on matrix distribution was established. Based on the observation model, the potential characteristics of the network and the convergence factors of the network community are obtained.
B. DEFAULT LINK PREDICTION

According to the supporting data of eight networks of recy- cling questionnaire, the default connection in the relational matrix of individual network of athletes is predicted by sev- eral parallel experiments. Link prediction AUC statistical results are shown in Figure 4:

where the X-axis scale represents the dimension of the hidden relationship U. We test the performance of the algorithm on link prediction when the hidden variable u takes different results. A to H distribution represents eight thematic networks.

C. PREDICTIVE ALGORITHM EVALUATION

As shown in Figure 4, AUC represents classification accuracy, which is larger, indicating the better the model’s analysis (classification) of such data. In this article, the larger the AUC, the more accurate the prediction of the relationship between the nodes on behalf of our model. This prediction includes predicting the existence of relationships between two nodes and the non-existence of relationships between the two nodes. The statistical results of AUC are based on the correct prediction of these two events. In the prediction of node relationship based on eight kinds of network data, AUC results reach 0.99, which shows that the accuracy of the algorithm is ideal.

On the other hand, as shown in table 1 is the result of the Prediction analysis algorithm, based on the support data of the recovery questionnaire, through a number of parallel experiments, the average value as the estimation of the true value is worth the accuracy to 0.98, representing the size of the system error of the algorithm itself. In the results of this report, from the confusion matrix representation, TN = 67,FP = 0,FN = 3,TP = 67, the positive example coverage (true positive Rate) can reach 96%, representing the algorithm to judge the true correct rate; negative case coverage (True Negative Rate) is 100%, which indicates that the algorithm can determine all negative cases, and the false alarm rate (false positive Rate) is 0%, which indicates that the algorithm predicts the relationship of all instances, and the false alarm rate (false Negative Rate) is 4%. It represents the positive sample proportion of the algorithm which is predicted to be negative in the example prediction.

F1 score is a statistical index used in this study to measure the accuracy of algorithm classification, taking into account the accuracy and recall rate of the classification model. F1 scores can be seen as a harmonic average of model accuracy and recall rates, with a maximum value of 1 and a minimum value of 0.

\[
F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}
\]
TABLE 2. Family network factors analysis.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Desc</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>economy</td>
<td>Financial situation of the family</td>
<td>0</td>
</tr>
<tr>
<td>education</td>
<td>Level of education</td>
<td>1</td>
</tr>
<tr>
<td>hometwon</td>
<td>Family sources</td>
<td>2</td>
</tr>
<tr>
<td>sex</td>
<td>Athlete Sex</td>
<td>3</td>
</tr>
<tr>
<td>social_network</td>
<td>Who do you usually communicate with</td>
<td>4-9</td>
</tr>
<tr>
<td>sports</td>
<td>Sports in which the project is carried out</td>
<td>10</td>
</tr>
<tr>
<td>year</td>
<td>Athlete years</td>
<td>11</td>
</tr>
<tr>
<td>SCL</td>
<td>10 Factors of SCL90</td>
<td>12-21</td>
</tr>
</tbody>
</table>

In the work of this algorithm, F1-Score = 0.98, which shows that the classification accuracy can reach 98%. It worked pretty well.

VI. APPLICATION OF ATHLETES’ NETWORK MODEL AND ALGORITHM BASED ON MATRIX DISTRIBUTION

For social communication networks, we extract a total of 21 factors and classify them.

A. ANALYSIS ON SOCIAL NETWORK NODE OF PROFESSIONAL ATHLETES

Taking the athletes’ social network as an example, among these factors, as shown in table 2, it covers factors and basic information about the impact and contribution of professional athletes in social communication networks. Including family economic status, family origin, gender, who will turn to, exercise projects, sports years, SCL90 eight factors, of which the 10 factors of SCL90 are mainly for the athletes’ current psychological state assessment, including hostility, anxiety, psychosis, terror, paranoia, other, compulsive symptoms, somatization, Interpersonal sensitivity, depression 10 psychological state.

From the results of factor analysis, as shown in Figure 5, we can see that the farther the weight value deviates from the center, the greater the influence of the representative factor on the relationship between the athletes’ social networks in the achievement network, and conversely, if the weight is closer to the center axis, the less influence the factor has on the connection in the achievement network.

From the results in the figure 5, we can see that factor 0 family economic status, factor 7 usually and teammate communication, as well as factor 10 engaged in sports projects will have a greater positive impact on social communication networks, and education level also has a greater contribution to social communication networks but a negative impact. At the same time, the interpersonal relationship factor in SCL90 will also have a greater negative impact.

B. ANALYSIS OF COMMUNITY RELATIONS OF PROFESSIONAL ATHLETES’ GROUP NETWORK

From the existing research results, it is found that the social relationship of professional athletes is usually obvious in the small world characteristics, it is necessary to further study the community phenomenon, according to the previous establishment of the athletes network sparse matrix Gaussian process model, we can use it for a certain analysis of the causes of the community. The formation of the community is influenced by some key factors, and we use machine learning as a means of Community relationship analysis for a typical group relationship network social network, as shown in Figure 6 as a community division of social networks.

In the social networking network, as shown in Figure 6, according to the clustering relationship, the network is divided into four communities according to the connection, excluding small communities with fewer than 20 nodes. And according to the number of nodes descending to each community in turn numbered 1, 2 and .... The sparse matrix Gaussian process analysis of the remaining 3 larger communities is calculated according to the weight between 22, and the results are as shown in Figure 7:

From Figure 7, it is found that the group 1,2 and subgroup 2,3,10 factors are the most influential factor in the formation of the differences between the two subgroups. Looking for the questionnaire, we found that factor 10th corresponds to the questionnaire 401 question: whether the team...
emergency treatment is appropriate. In the different understanding of this problem, it leads to the difference between nodes, which results in different communities. As you can learn, the impact of factor 10th makes it easier for nodes to converge in 2 communities compared to 2 communities and 3 communities; 1 communities compared to 2 communities, the impact of factor 10th makes it easier for nodes to converge in 2 communities than in 1 communities. In social networking relationships, the convergence of nodes for a particular community also reflects a certain trend of cooperation and competition.

VII. CONCLUSION

Based on matrix distribution, this study describes the social network structure of professional athlete groups, establishes the construction and analysis model of athletes’ social network, and applies the model and algorithm through questionnaire dataset.

In the construction of professional athletes’ social network, according to the questionnaire data, eight networks of professional athletes are predicted, and the model-based prediction algorithm shows good predictive performance, which reflects the ideal performance of the algorithm from the statistics of classifier. In the process of network node factor analysis, This machine learning method can quickly and accurately obtain the influence degree of node factor on various relational networks, and is suitable for the analysis of supporting attribute factors of professional athletes’ nodes. In the community analysis study of the complex network of professional athletes, it can be seen that the community convergence of professional athletes is influenced by certain factors. From the analysis results, the algorithm can get the cooperation trend of athletes in network relations, in addition, the community is mostly small-scale groups, the characteristics of the small world is obvious. Therefore, the proposed model and algorithm have good adaptability and provides a good working basis for future expansion research.

REFERENCES

XIAOYU GE received the master’s degree in education from the Xian Physical Education University, in 2016. He is currently pursuing the Ph.D. degree with the Shanghai University of Sport. His work has been published several article in different journals. His research interests include management science and sports science.

SAI-FU FUNG received the Ph.D. degree from the University of Warwick, Coventry, U.K. He is currently a Teacher with the Department of Social and Behavioral Sciences, City University of Hong Kong. He published in various SSCI and SCI listed journals, such as the IEEE Transactions on Neural Networks and Learning Systems, Health and Quality of Life Outcomes, and Psychology Research and Behavior Management. His current research interests include artificial intelligence, sociology of sport, political sociology, and psychometrics.

JIANLAN DING is currently pursuing the Ph.D. degree with Shaanxi Normal University. She is currently an Associate Professor with the Department of Sports Economy and Sports Management, Xi’an Physical Education University, Xi’an, China. She has published many high-level articles in public. Her research interests include sports management, social networking, neural marketing, and data mining.