

A Study on the Effect of Personality Traits and Social Support on Psychological Health of Chinese College Students Based on Data Mining Technology

Ying Wu (✉ wuguaiguai9896@126.com)

Shanxi Medical University <https://orcid.org/0000-0002-5657-4471>

Huaqing Guo

Shanxi Medical University

Hongxia Li

Shanxi Medical University

Wenxiao Yang

Shanxi Medical University

Research

Keywords: Personality traits, Social support, Psychological health, Data mining

Posted Date: March 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-304899/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

With the continuous development of the society, the pace of life is accelerating, psychological health has become an important factor affecting people's quality of life. Domestic studies have shown that mental diseases have become an important cause of College Students' sick leave, withdrawal, addiction, crime and suicide. Personality traits and social support factors have been considered to be related to psychological health.

Methods

In this study, Symptom Checklist-90 (SCL-90), Eysenck Personality Questionnaire (EPQ) and Perceived Social Support Scale (PSSS) of some freshmen in 2020 were randomly selected from the psychological census database of a university in China as the research data. The decision tree algorithm was used to establish a predictive model for psychological health related personality traits and social support factors of college students. In association analysis, we got the most closely effect factors with psychological health or psychological abnormality by analyzing important association rules. According to the distance, all effect factors were classified into different clusters by system clustering, and the relationship between effect factors and mental health was further analyzed.

Results

Combining the results of the three algorithms, we found that social support and personality traits factors had a certain rule on the effect of college students' psychological health, and social support factors were more important than personality traits factors. Among the social support factors, the most important was family support, followed by friend support and other support. The higher the level of support of the three, the greater the possibility of psychological health. In the personality traits factors, the most important was N, followed by E, and P. The three grades were high or low, may lead to abnormal psychology, and the medium grade was an important condition to maintain psychological health.

Conclusion

The results provided important references for the analysis of the effect factors on psychological health and important theoretical basis for the formulation of psychological intervention measures for college students.

Background

With the continuous development of the society, the pace of life is accelerating, and people's psychological problems are also increasing. Psychological health has become an important factor affecting people's quality of life. College stage is not only the key period of knowledge learning and ideological and moral training, and it is but also an important period of cultivating psychological health, social adaptation and perfecting personality. At present, due to the increasing expectations and concerns of families, schools and society for young people, as well as the increasing social competition, these pressures are increasing day by day, threatening the formation of positive and healthy personality of young people. Domestic studies have shown that mental diseases have become an important cause of College Students' sick leave, withdrawal, addiction, crime and suicide [1–4]. Therefore, we should pay attention to the psychological health education of college students, and deeply analyze their psychological health status and its relationship with personality traits, school, family, society and other influencing factors. Only in this way can we carry out targeted psychological counseling, psychological counseling and health education to help them successfully complete the transition from middle school stage to college stage, and lay a solid foundation for them to adapt to the society in the future.

There are many effect factors of psychological health. Some studies showed that there was a certain association between the personality traits of college students and their psychological status. Some studies showed that the proportion of college students with bad personality traits and mental disorders was increasing in China [5]. Psychological investigation on college students also found that Symptom Checklist-90 (SCL-90) and Eysenck Personality Questionnaire (EPQ) scores were correlated, the more unstable the mood and higher psychoticism of the students, the more likely they were to have psychological problems [6]. Some researchers used the Big Five Inventory (BFI) to investigate college students, the results showed that different gender of college students had its significant difference in the aspect of the openness. The dimensions of the psychological health level had significant negative correlation with the extraversion, agreeableness, sense of responsibility, neuroticism factor, but openness was not significantly related to [7].

Social support refers that individuals who are in a dilemma have its meaningful groups, such as family members, friends, colleagues, relatives and neighbors, etc., providing action, emotional support and information help [8]. Studies suggested that there was a direct and indirect relationship between social support and psychological health. For example, there was a negative correlation relationship between depression and social support, that is, the lower the level of social support, the greater the probability of depression [9]. And good support network can help college students adjust their mood, establish positive attitude, improve their psychological endurance, and have a positive effect on reducing depression and anxiety [10, 11]. In addition, Internet addiction has caused great harm to the mental health of college students, and studies showed that social support was negatively correlated with Internet addiction among college students [12].

At present, the college students have been conducted a variety of psychological evaluation and measurement on accumulating a large number of psychological data in many universities. The traditional scale statistics and analysis methods can only obtain the surface information of

psychological data, which was not efficient enough to use the data. As a result, many useful rules and patterns cannot be extracted, and psychological measurement did not fully play its role. Compared with the traditional methods, data mining technology has its unique advantages. It can mine the important information related to college students' psychological problems and effect factors, and find the potential relationship between the factors, etc. So far, a variety of data mining techniques have been applied to the study of college students' psychological problems. Some researchers had discussed and analyzed the feasibility of the decision tree algorithm in solving the psychological problems of vocational college students [13, 14]. Other researchers built a Bayesian network prediction model to analyze the potential relationship between student attributes and psychological test data, and the results showed that the model had good predictive power [15]. In addition, some studies discussed the role of cluster analysis in the prevention of psychological crisis [16]. If a variety of algorithms were introduced into the study of effect factors of college students' psychological health for comprehensive analysis, the importance of various effect factors and their relationships can be discussed more deeply, and more scientific and effective decision support can be provided for psychological health work.

This research adopted the decision tree, association analysis and cluster analysis algorithm to mine college students' personality traits and social support data based on R, found out the key factors that affect psychological health, and analysis the relationship among various effect factors. It will provide a scientific basis for planning, decision-making, enhance the pertinence and effectiveness of psychological health education.

Methods

Data collection and preprocessing

In this study, SCL-90, EPQ and PSSS of some freshmen in 2020 were randomly selected from the psychological census database of a university in China as the research data. The results of SCL-90 were used to determine the psychological health of the students, while EPQ and PSSS were used to evaluate the personality traits and social support of the students. According to the national norm results of SCL-90, the total score was more than 160, or the number of positive items was more than 43, or any factor score was more than 2 points, It was screened as positive [17].

EPQ was a scale compiled by H.J. Eysenck and his wife, who was British psychologist. The effectiveness of EPQ on personality testing has been verified [18, 19], which contains 88 items and 4 scales: Extraversion (E), Psychoticism (P), Neuroticism (N) and Lie (L). Among them, E, N and P are the three dimensions of personality, which are independent of each other. Then, according to the total score (rough score) obtained by the test subjects on each scale. By converting the rough score into the standard score T ($T = 50 + 10 \cdot (x - m) / SD$), we can analyze the personality traits of the subjects. The T score of each scale ranges from 43.3 to 56.7 points as the intermediate type; the T score ranges from 38.5 to 43.3 points or 56.7 to 61.5 points as the tendency type; the T score below 38.5 points or above 61.5 points are typical. According to the T score standard, we divided the data of three personality dimensions into three levels,

that is, $T < 43.3$ is low, marked with "1"; $43.3 \leq T \leq 56.7$, marked with "2"; $T > 56.7$ is high, marked with "3". In this way, the personality dimension can also be divided into 9 personality factors: E1, E2, E3, N1, N2, N3, P1, P2 and P3. This study only investigated three dimensions of personality, so it was not included the L subscale. In order to highlight the authenticity of the three dimensions, the questionnaire containing the L subscale $T > 56.7$ was deleted.

PSSS developed by Zimet was used to measure Social Support [20]. The scale is divided into 3 dimensions, including 4 items of Family Support, 4 items of Friend Support and 4 items of Other Support. The total score of the scale ranged from 12 to 84 points. A score of 12–36 indicates that individuals felt low support; a score of 37–60 indicates that the individual feels moderately supportive; a score of 61–84 indicates that the individual feels a high state of support. Based on this, the data of each dimension were divided into three levels, with 4–12 points as low, marked by "1". 13–20 points was medium, marked by "2"; 21–28 points was high, marked with "3". In this way, the 3 dimensions can also be divided into nine social support factors: Family1, Family2, Family3, Friend1, Friend2, Friend3, Other1, Other2 and Other3.

After screening, a total of 522 valid questionnaires were obtained. 323 were psychological normal questionnaires and 199 were psychological abnormality questionnaires. The discrete personality traits and social support level data and an additional variable of decision result were stored as TXT file in the form of transaction table (Table 1) as the research object of decision tree analysis. The variable "1" of decision result represented psychological abnormality, and "0" represented no abnormality, which was, psychological health. The data of 9 personality traits factors and 9 social support factors were stored in the form of matrix (Table 2) as the research objects of association analysis and cluster analysis. "1" and "0" in the matrix indicated whether the effect factors existed, "1" indicates existence, "0" indicates nonexistence.

Decision tree analysis

Decision tree algorithm is an inductive classification algorithm based on instance category. This algorithm can establish the decision tree model based on the given data set, and extract the intuitive and easy to understand classification rule knowledge from it, which has a good prediction effect, and the result model also has a good interpretation function. Decision tree algorithm is a tree data structure composed of root node, middle node and leaf node. A node can determine the category to which an instance belongs, and its role is to determine how the test case should select the next node by comparing the property values. At present, the commonly used decision Tree algorithms are C4.5 and Classification And Regression Tree (CART). C4.5 is suitable for numeric data And CART is suitable for sub-type data. The CART decision tree algorithm first considers combining the target categories into two super categories according to a certain attribute classification metric. The classification is stopped when one of the subcategories is recursively computed using the same method until the final attribute belongs to a single category.

In this study, personality traits and social support data were divided into three levels, so CART was used to construct the data prediction model. The decision tree model was implemented through the "rpart"

package of R, Table 1 was used as training set, 6 personality and social support dimensions as input variables, and output variables as decision results. Gini coefficient was used for heterogeneity index, and no specific parameters were set for pruning.

Apriori association analysis

Association analysis is an important way to reveal the intrinsic structural characteristics of data. The purpose of association analysis is to find simple association relation or sequence relation among things based on existing data. Association rule mining was one of the most active research methods in data mining. The original motivation was proposed for the problem of shopping basket analysis, with the purpose of discovering the relation rules between different items in the transaction database [21]. These rules describe the buying behavior of customers and can be used to guide businesses to scientifically arrange the purchase, inventory and shelf design.

Apriori algorithm is a classical algorithm of association rules, which is the core of all association mining algorithms. Apriori algorithm involves three important parameters which are Support, Confidence and Lift. Support measures the universality of the application of association rules. The higher the Support, the more common the rule is. Confidence reflects the accuracy of the association rules. The higher the Confidence is, the greater the chance of the latter term appears under the condition that the former term of the rule exists. Lift reflects the practicability of association rules, and only the rules with Lift greater than 1 can be useful.

In this study, Apriori algorithm was used to conduct association analysis on 18 factors related to personality traits and social support, which was realized by "arules" package of R. The matrix of personality traits and social support factors were introduced into the transaction table, and a decision result factor was added to the transaction table. "Yes" indicated psychological abnormality, and "No" indicated psychological health. In order to control the number of rules and screen out the most effective rules, the minimum Support was set to 0.1, the minimum Confidence was 0.4, and the right item of association rules was "Yes". Then set the minimum Support to 0.2, the minimum Confidence to 0.5, and the right item of the association rule to "No". In this way, the association rules of psychological abnormality and psychological normality can be mined respectively. The mining process consisted of two main stages: the first stage was to identify all frequent itemsets from the data set, and the second stage was to generate association rules from these high-frequency itemsets.

Clustering analysis

In many data mining algorithms, clustering analysis is a certain algorithm which is divided the data set into the nature of the similar data points of several classes, so as to reveal the internal relation among data and different methods, and has been successfully applied to many fields, including pattern recognition, image processing, data analysis and market research and so on.

System cluster, also known as hierarchical clustering, designs algorithms from the perspective of distance and connectivity. The observation points close to each other in space are one class, and the

clustering results obtained are generally deterministic and hierarchical. In the process of systematic clustering, all observation points are gradually merged into small classes, and small classes are merged into middle classes and even large classes.

The systematic clustering analysis in this study was implemented through the package of "hclust" in R, and the analysis object was the matrix of personality traits and social support factors (Table 2). The measure method was set as "ward", namely the sum of squares of dispersion.

Results

The results of Decision tree

The final decision tree grew from top to bottom, including 1 root node, 4 internal nodes, and 6 leaf nodes (Fig. 1). Each rounded rectangle represents a node, which contained three numbers. The number in the first row represents the output decision result, and the number in the second row from left to right represents the number of people who made the decision for psychological health and abnormality. The decision tree contained four factors. In terms of decision order, the first level was Family Support; the second level was Friend Support and N; the third level was N and E.

The reasoning rules of psychological abnormality corresponding to 3 leaf nodes were: Family Support level was 1 or 2 and N level was 3, the confidence was 0.767, there were 33 people; Family Support level was 1 or 2, N level was 1 or 2, and E level was 3, the confidence was 0.576, with 19 people; Family Support level was 3, Friend Support level was 1 or 2 and N level was 3, the confidence was 0.786, with 11 people.

The other 3 leaf nodes corresponded to the reasoning rules of psychological health: Family Support level was 1 or 2, N level was 1 or 2, and E level was 1 or 2, the confidence was 0.623, there were 48 people; Family Support level was 3, Friend Support level was 1 or 2, and N level was 1 or 2, the confidence was 0.596, with 31 people; Family Support level was 3, Friend Support was 3, the confidence was 0.716, with 217 people.

Association analysis results

After calculation, we obtained 6 association rules for psychological abnormality with Lift greater than 1 (Fig. 2). In the figure, the larger the circle, the greater the Support, and the darker the color, the greater the Lift was. The association rule with maximum Support was {Other2} => {Yes}, with the Support was 0.170; the association rule with maximum Lift was {N3} => {Yes}, with the Support was 1.37.

Similarly, we obtained 12 association rules for psychological health with Lift greater than 1 (Fig. 3). Among them, the association rule with the maximum Support was {Family3} => {No}, with the Support was 0.480. The association rule with maximum Lift was {Family3, Friend3} => {No}, with the Lift was 1.16.

System clustering results

The clustering process integrated personality traits and social support factors into larger categories based on distance from close to far from small categories, and the similarity within categories gradually decreases (Fig. 4). The 18 factors were divided into two clusters with a distance of 30, and the two clusters did not gather into one cluster until the end of the clustering process, with obvious differences. Cluster 1 included Family3, Friend3, Other3, P2, E2, and N2. Cluster 2 contained 12 factors including Other1, Family1, Friend1, E3, P1, N1, E1, P3, N3, Family2, Friend2 and Other2.

Discussions

In this study, the decision tree algorithm was used to establish a predictive model for psychological health related personality traits and social support factors of college students. The model includes four factors: Family Support, Friend Support, N and E. From the perspective of decision-making order, Family Support was the most important factor affecting psychological health, while the other three factors also play an important role. From the perspective of Social Support, the higher the level of Family Support and Friend Support, the greater of psychological health was, and vice versa. From the perspective of personality traits, a low or medium N and E grades were associated with psychological health, while a high N and E grades were associated with psychological abnormality. In other words, college students who were lack of support from family and friends and have unstable extroversion were prone to psychological abnormality and should be given priority. These factors should also be taken into account when conducting psychological intervention on college students.

Secondly, in the association analysis, we obtained 6 association rules of psychological abnormality with Lift greater than 1 and 12 association rules of psychological normality. In general, the former and the latter were most closely related, and once the former appeared, the latter were more likely to occur simultaneously. The association rules with the maximum Support for psychological abnormality and psychological normality were {Other2} => {Yes} and {Family3} => {No}, respectively. It suggested that students with general support were more likely to be psychologically abnormal, while students with good family support were more likely to be psychologically normal. The association rules with the greatest Lift were {N3} => {Yes} and {Family3, Friend3} => {No}, which meant that students with unstable emotions had psychological abnormalities or students with good support from family and friends were psychologically normal. These two association rules were of more practical significance. It can be concluded from the results of association rules that the factors most closely related to psychological health were high family support and high friend support. Students with emotional instability or other general support need special attention, they may have psychological abnormality.

Finally, from the results of system clustering, the 18 effect factors were obviously divided into two clusters. The first cluster contained 6 factors, including the highest level of all social support factors and the medium level of all personality traits factors. Combining the results of decision tree and association analysis, it can be concluded that such factors were closely related to psychologically normality,

suggesting that students with good social support and moderate personality traits were more likely to be psychological health.

The second cluster contained the remaining 12 factors, many of which play an important role in the decision-making and association analysis of abnormal psychology, indicating that such factors were closely related to psychological abnormality. This cluster suggested that poor social support and extreme personality traits may lead to psychological abnormality.

Combining the results of the three algorithms, we found that social support and personality traits factors had a certain rule on the effect of college students' psychological health, and social support factors were more important than personality traits factors. Among the social support factors, the most important was family support, followed by friend support and other support. The higher the level of support of the three, the greater the possibility of psychological health. Of the personality traits factors, the most important was N, followed by E, and P. The three grades were high or low, may lead to abnormal psychology, and the medium grade was an important condition to maintain psychological health.

Of course, this study also had some limitations. The data was from a single source, and only one college freshman's psychometric data was selected. Because the sample size was small, the test set was not set in the decision tree modeling, which may have a certain impact on the result.

Conclusion

In this study, three data mining techniques, namely decision tree analysis, association analysis and cluster analysis, were applied to analyze the relationship between psychological health, personality traits and social support of college students, and to dig out the important effect factors of psychological health as well as the association characteristics and rules. The results provided important references for the analysis of the effect factors on psychological health and important theoretical basis for the formulation of psychological intervention measures for college students.

Abbreviations

SCL-90

Symptom Checklist-90; EPQ:Eysenck Personality Questionnaire; BFI:Big Five Inventory; PSSS:Perceived Social Support Scale; CART:Classification And Regression Tree

Declarations

Acknowledgements

The authors would like to express gratitude to the participants for their time and willingness to participate in this study.

Funding

The authors gratefully acknowledge the support for this study from “Analysis on Scientific Collaboration and Trends Prediction of Research Fronts in Psychiatry Field (No. 71503152)”, awarded by National Natural Science Foundation of China. This study is also supported by Program for the Outstanding Innovative Teams of Higher Learning Institutions of Shanxi (OIT) and Program for the Philosophy and Social Sciences Key Research Base of Higher Education Institutions of Shanxi (PSRB) under Grant No. (20190114).

Availability of data and materials

The data used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The research data were collected from the university psychological census database. All participants gave written consent after being informed to the aim of the survey and their rights to refuse to participate.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Humanities and Social Sciences, Shanxi Medical University, No. 56 South Xinjian Road, Taiyuan 030001, China.

References

1. Chen JW, Wang T. A preliminary study on the relationship between self-esteem, self integration and psychological health of College Students. *Chinese Journal of Clinical Psychology*. 2004;2:53-54.
2. Zhu HY, Zhang F, Shen MW. The construction of Internet related social psychological health concept for College Students. *Chinese Journal of Clinical Psychology*. 2005;13(1):4-8.
3. Yang L, Liu Y, Wu F. Assessment of College Students' Mental Health and Psychological Intervention's Effect. *Chinese Journal of Clinical Psychology*. 2005;13(1):102-104.
4. Shao YK, Zheng HY, Yu QN. An analysis of the reasons for the students of Panjin vocational and technical college to quit from 2006 to 2009. *Preventive Medicine Tribune*. 2009;05:83+85.

5. Wu HL, Sun YS. The Relationship between Mental Health Status and Personality Characteristics of Medical Students. *China Journal of Health Psychology*. 2013;21(1):129-131.
6. Liu CM, Zhao W. The Relation between Personality Characteristics and Mental Health of College Students. *Journal of Hebei North University (Social Science Edition)*. 2008;1:64-67.
7. Li KS, Zhou SJ. Study on the personality and psychological health of College Students. *Journal of Xiangtan Normal University (Social Science Edition)*. 2008;30(1):157-159.
8. Hefner J, Eisenberg D. Social Support and Mental Health Among College Students. *American Journal of Orthopsychiatry*. 2009;79(4):491-499.
9. Bener A, Alsulaiman R, Doodson L, et al. Depression, Hopelessness and Social Support among Breast Cancer Patients: in Highly Endogamous Population. *Asian Pacific Journal of Cancer Prevention*. 2017;18(7):1889-1896.
10. Prashant T. The Moderating Effect of Perceived Social Support on Stress and Depression among University Students. *Online Journal of Health and Allied Sciences*. 2016;15(3).
11. Sun CY, Liu HQ, Yuan J. The relationship between social support, anxiety and depression of College Students. *Chinese Journal of Public Health*. 2007;23(12):1518-1519.
12. Mo PKH, Chan VWY, Chan SW, Lau JTF. The role of social support on emotion dysregulation and internet addiction among chinese adolescents: a structural equation model. *Addictive Behaviors*. 2018;82:86-93.
13. Lu T. Research and practice of data mining technology in Higher Vocational Students' information system. *China CIO News*. 2015;05:151.
14. Chen ZF, Wu HH. The application of data mining in Higher Vocational Students' Psychological Education. *Practical Electronics*. 2013;22:170-170,112.
15. Zhang H, Huang ZJ. Bayesian network and its application in psychological field. *Modern Preventive Medicine*. 2007;034(007):1284-1285, 1287.
16. Li Z, Zhang YX, Hu H. The application of cluster analysis in the management of College students' mental health. *Journal of Hubei Engineering University*. 2014;34(6):53-57.
17. Yang YM, Li J. The application of cluster analysis algorithm in the analysis of College Students' mental health. *Journal of the Chinese Society of Education*. 2015;S1: 27-29.
18. Cox BJ, Enns MW, Clara IP. The multidimensional structure of perfectionism in clinically distressed and college student samples. *Psychological Assessment*. 2002;14(3):365-373.
19. Henderson JW, Donatelle RJ, Acock AC. Confirmatory Factor Analysis of the Cancer Locus of Control Scale. *Educational and Psychological Measurement*. 2002;62(6):995-1005.
20. Blumenthal JA, Burg MM, Barefoot J, et al. Social support, type A behavior, and coronary artery disease. *Psychosomatic Medicine*. 1987;49(4):331-340.
21. Humayun M. Data Mining Framework for Generating Sales Decision Making Information Using Association Rules. *International Journal of Advanced Computer Science and Applications* 2016;7(5):378-385.

Tables

Table 1 Partial training set of decision tree analysis

Decision	E	P	N	Family	Friend	Other
1	1	2	2	3	3	3
1	1	3	3	3	3	3
1	3	2	1	3	3	3
1	1	2	2	3	3	3
1	2	2	2	3	3	3
1	1	1	2	3	3	3
1	2	2	2	2	2	2
1	2	2	2	3	3	3
0	2	2	1	2	2	2
0	1	2	2	3	3	3
0	2	1	1	2	2	2
0	3	1	1	3	2	2
0	2	3	2	3	3	3
0	2	2	1	3	2	3
0	1	2	2	3	3	3

Table 2 Partial matrix of personality traits and social support factors

	1	2	3	4	5	6	7	8
E1	1	1	0	1	0	1	0	0
P1	0	0	0	0	0	1	0	0
N1	0	0	1	0	0	0	0	0
E2	0	0	0	0	1	0	1	1
P2	1	0	1	1	1	0	1	1
N2	1	0	0	1	1	1	1	1
E3	0	0	1	0	0	0	0	0
P3	0	1	0	0	0	0	0	0
N3	0	1	0	0	0	0	0	0
Family1	0	0	0	0	0	0	0	0
Friend1	0	0	0	0	0	0	0	0
Other1	0	0	0	0	0	0	0	0
Family2	0	0	0	0	0	0	1	0
Friend2	0	0	0	0	0	0	1	0
Other2	0	0	0	0	0	0	1	0
Family3	1	1	1	1	1	1	0	1
Friend3	1	1	1	1	1	1	0	1
Other3	1	1	1	1	1	1	0	1

Figures

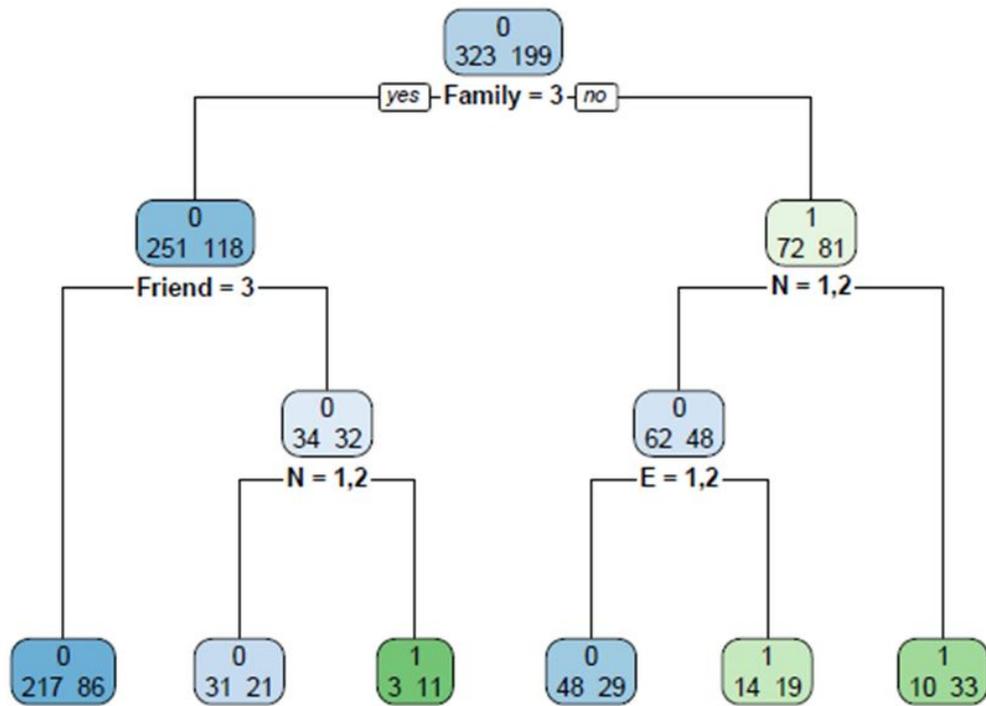


Figure 1

The decision tree diagram of psychological health

Grouped Matrix for 6 Rules

Size: support
Color: lift

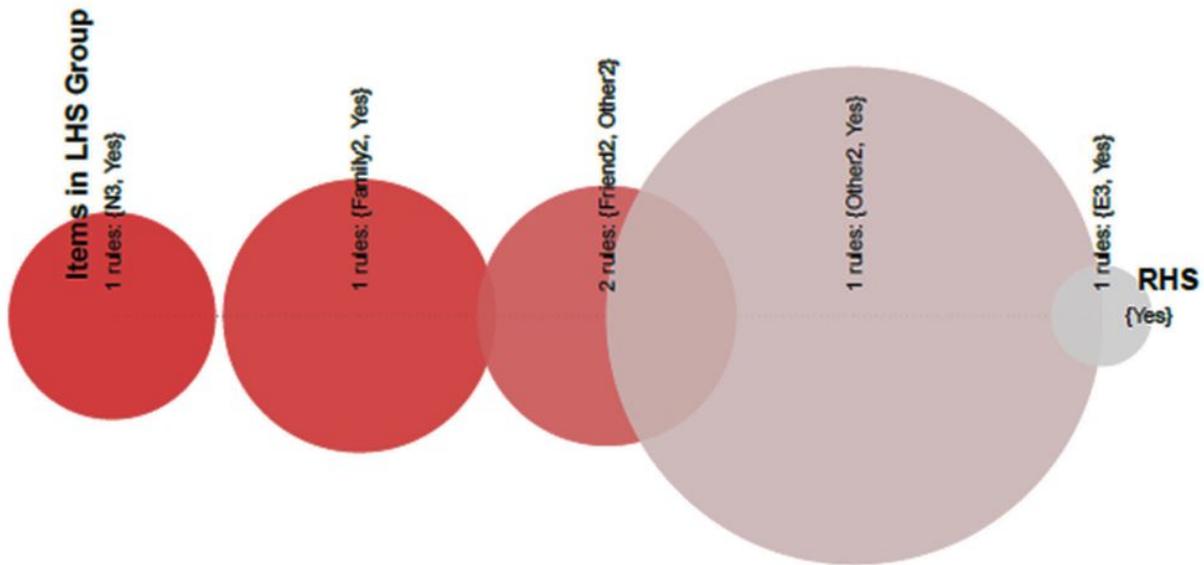


Figure 2

The diagram of 6 association rules for psychological abnormality

Grouped Matrix for 12 Rules

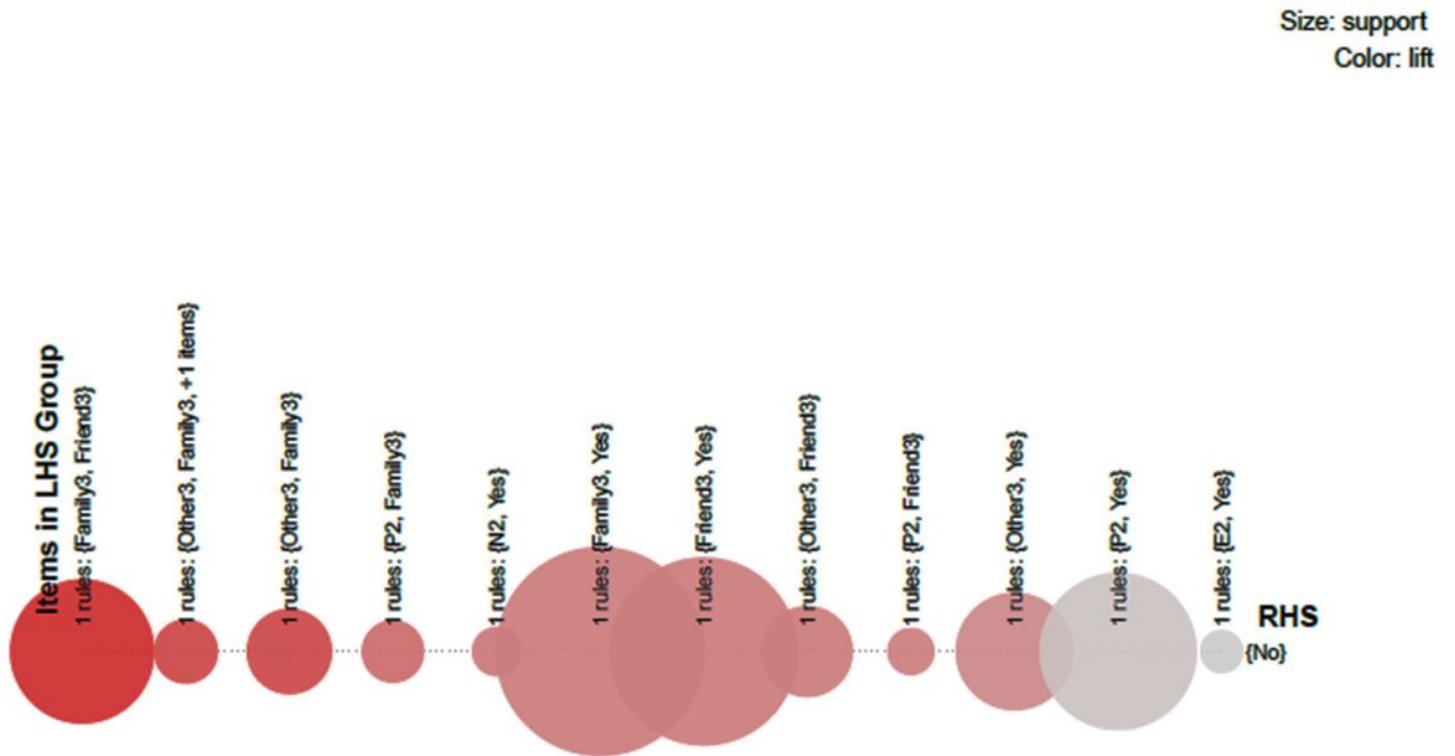


Figure 3

The diagram of 12 association rules for psychological health

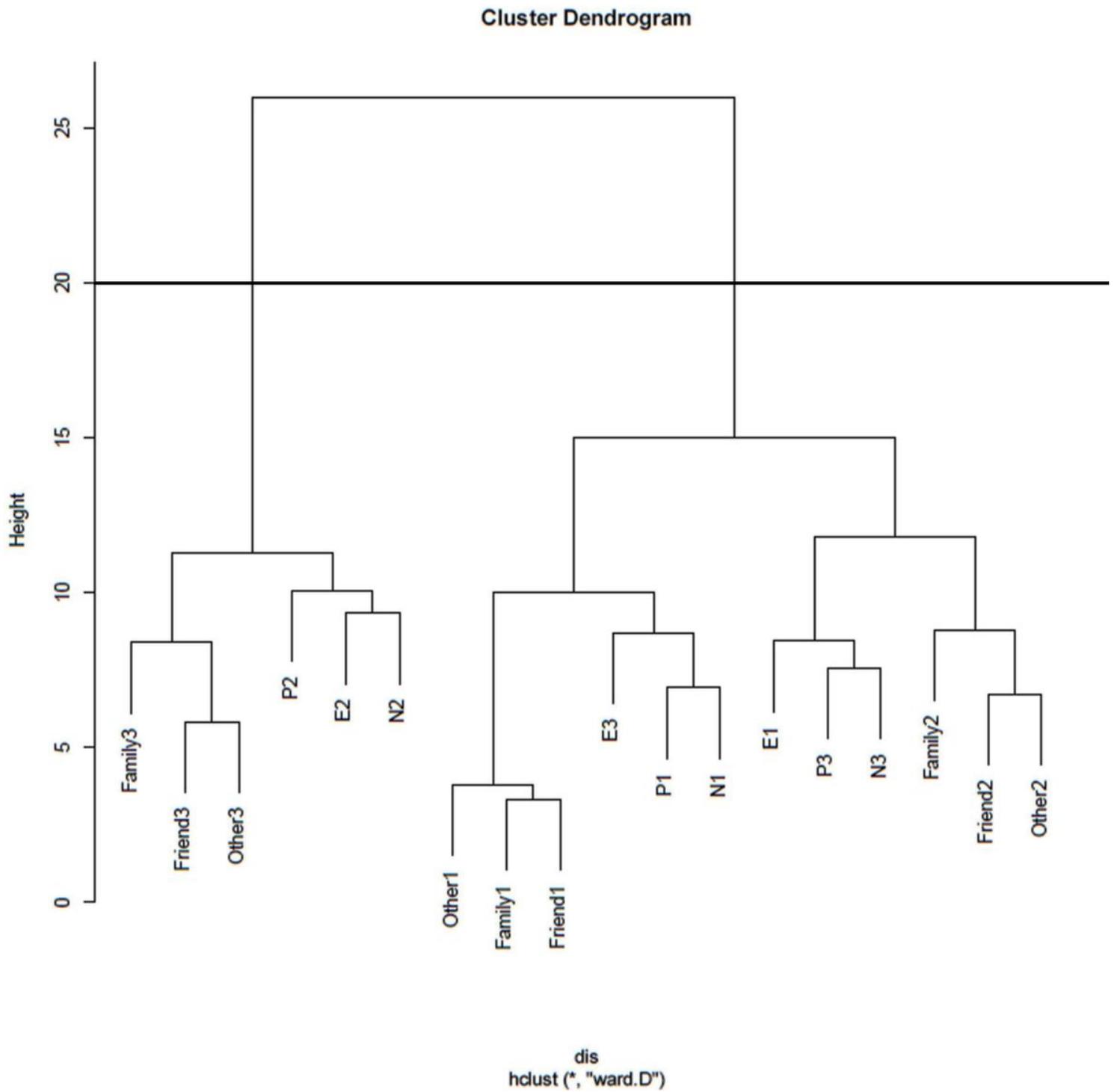


Figure 4

The cluster dendrogram of 18 personality traits and social support factors