Zhang, D., Guo, T., Han, S., Vahabli, S., Naseriparsa, M., & Xia, F. (2021). Predicting Mental Health Problems with Personality, Behavior, and Social Networks. *2021 IEEE International Conference on Big Data (Big Data),* 4537–4546.

Available at: https://doi.org/10.1109/BigData52589.2021.9671987

# Predicting Mental Health Problems with Personality, Behavior, and Social Networks

Dongyu Zhang*, Teng Guo*, Shiyu Han*, Sadaf Vahabli†, Mehdi Naseriparsa‡, and Feng Xia‡(✉)
*School of Software, Dalian University of Technology, Dalian 116620, China
†Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran
‡School of Engineering, IT and Physical Sciences, Federation University Australia, Ballarat, VIC 3353, Australia
f.xia@ieee.org

***Abstract*—Mental health is an integral part of human health and well-being. Unhealthy mentality leads to serious consequences such as self-mutilation and suicide, especially for college students. While the literature focused on analysing the relationship between mental health and a single factor such as personality or behavior, accurate prediction is yet to be achieved due to the lack of cross-dimensional analysis and multi-dimensional joint prediction. To this end, this work proposes leveraging multiple factors from three crucial dimensions of mental health: behaviors, personality, and social networks. We recruited 490 college students, and collected their behavioral records from smart cards. In addition, we extracted their psychological traits from questionnaires, and social networks by conducting the survey on the nominating community members. We created a neural network-based model to integrate behavioral, psychological, and social network factors to predict mental health problems. The experimental results verify the efficacy of the proposed model, and demonstrate that the classification model of various factors effectively predicts the students' mental issues.**

***Index Terms*—Mental health, Social networks, Behavior, Personality, Neural networks**

## I. Introduction

Mental health problems have increased dramatically around the globe, especially during the COVID-19 pandemic. These mental health issues are predicted to impose the largest worldwide healthcare cost by 2030 [1]. Maintaining a good mental health is crucial for human well-being. A major challenge for mental health research is to determine the underlying factors that affect individuals' mental health. That's because these factors play an instrumental role in the early prediction, detection, and active treatment of mental illness.

Previous research has suggested that mental health issues arise from a variety of individual and social factors, such as socioeconomic background, health status, personality traits, and personal behaviors [2]. However, the existing studies have focused mainly on the analysis of a single type of factor in the mental conditions of individuals, with a few exceptions [3]. Some researchers argue that the methods which integrate multiple types of factors outperform the methods involving a single factor [3]. As such, using a single type of factor is likely insufficient to assess an individual's mental health. Moreover, no previous studies have leveraged the multiple dimensions of social networks (e.g., friendship, advice, cooperation, and trust) to model mental health problems.

Social networks play an integral part in reflecting individuals' overall mental health condition these days [4]. Their multiple dimensions may allow for a comprehensive study of complex real-world social roles in which the mental health problems are developed. Thus, this is an appropriate approach to understand and analyse mental illness.

In order to better predict and understand the mental illness, this study aims to employ multiple types of factors from three critical mental health dimensions: behaviors, personality traits, and social networks for automatic analysis. To perform the study and conduct the experiment, we recruited 490 college/university students who lived within the same residential area of a large Chinese college. Researchers previously focused on the ability to predict the mental health symptoms of the college students within different cultural settings [5], a passion area on which our research team is focused to address. We used mobile smart cards to collect the online behavioral records, including playing games, social media activity, and watching videos. Moreover, we extracted their "Big Five" personality traits. Finally, we detected mental health problems which are obtained by questionnaires and surveys. The main focus of this study was on the multiple dimensions of social networks and the exploration of 11 corresponding aspects such as friendship, positive attitude, cleverness, study advice, life advice, good news, bad news, cooperation, support, leadership, and trust. As an outcome of this study, we propose a neural network-based model to integrate behavioral, psychological, and social network factors to predict the various mental health problems, including delusion, suicidal intention, inferiority, hostility, and Internet addiction. We believe that the experimental results verify the proposed model's efficacy and demonstrate that the use of multiple factors effectively predicts the mental health problems in college students. Our contributions are as follows:

- We investigate the combined effects of multiple types of factors on mental health, including personality, behaviors, and social networks, while little attention has been received in previous studies.
- We propose a neural network-based model to integrate behavioral, psychological, and social network factors to predict various mental health problems.
- Our study inspires the exploration of multiple dimensions

of mental health mechanisms and other complex factors by using computer science approaches that assist in the early detection and treatment of mental illness.
- We create a unique dataset, which will be released publicly while maintaining confidentiality. This may contribute to the community by allowing other researchers to conduct experiments by using this data.

The rest of this paper is organised as follows: Section II reviews the related work. Section III explains the data that was collected in the experiments. Section IV details the used methodology. Section V presents the experimental results and analysis. Finally, the study outcomes and the importance of the proposed approach are presented in Section VI.

## II. Related Work

### A. Behaviors and Mental Health

There are a number of several relevant studies which focus on the relationship between behaviors and mental health. The results of such studies indicate that the mental health problems (e.g., depression and anxiety) among college students have been increased [6]. In this regard, notable increases were detected in the risk of depression and anxiety among students with associated adverse behavioral characteristics such as frequent drinking, sleep disorders, poor eating behavior, and Internet addiction disorder (IAD) [7]. The latent class analysis was applied to investigate the clustering of six primary risk behaviors (i.e., drinking, smoking, sleep time, lack of exercise, fruit and vegetable intake, and seating time) among 18-year-old Australians [8]. Some researchers employed regression and correlation analysis to study the link between mental health status and two different types of false self-reporting behaviors: *lying behaviors (i.e., status updating and profile details), liking behaviors (i.e., liking posts)*. It shows a significant relationship between anxiety and lying behavior, in which there is a strong relationship between stress, anxiety, depression, and liking behavior [9]. Some studies focused on college students' suicidal behaviors, showing that psychopathology is one of the main reasons for college students' suicidal behaviors [10].

### B. Psychology, Personality Traits, and Mental Health

Research around the disciplines of psychology and mental health has a long history [11] which includes many areas, such as perception, cognition, emotion, thinking, personality, behavioral habits, computer science, interpersonal relationships, and social relationships [12]. At present, the research on college students' psychological and mental health mainly focuses on the relationship between personality-level characteristics and mental illness. In the past, adolescents' mental health was linked to self-esteem, where there might be a correlation between self-esteem and depression, and between self-esteem and anxiety, according to some studies [13]. The relationships between anxiety, depression, and personality via a linear mixed modelling approach were studied by some researchers. The main focus was to develop the hypothesis based on the previous findings which revealed that anxiety and depression are most prominently associated with neuroticism, as well as extroversion [14]. Researchers have also studied the relationship between over-positive self-evaluation and psychological adaptation by comparing the data of young volunteers five years previous to that of five years later [15]. The results indicated that self-evaluations might be essential for mental health. A study of interest to this research, due to its multi-factor approach, focused on the moderating effects of personality traits in the relationship between religious practices and mental health for university students [16]. Religious, mental health, and personality traits were measured using the religiosity scale of Islam, mental health inventory, and the "Big Five" inventory. The correlation analysis showed a significant relationship between religious belief and behavioral control.

### C. Social Networks and Mental Health

Social networks are one of the most critical factors that affect the college students' mental health and behavior [17], which can be considered as the main controlling element of an individual's general mental health condition [4]. Importantly, social networks can efficiently enable the collection of data to be used to predict relevant characteristics of an individual's mental health [18]. For example, only recently, a variety of behavioral data for social media users has been modelled using a web-based method to diagnose mental health issues [9].

It is extremely challenging to study the impact of social networks on the mental health due to the dynamic and complex nature of these networks [19]. Park and colleagues exploited the potential features of eight common social network characteristics to determine the type of social network with latent profile analysis (LPA). In turn, regression analysis was performed on the identified types based on self-assessed health status and depressive symptoms to explore the health risks posed by group members. The results suggested that isolated nodes in social networks had more physical and mental health risks [20]. In addition, researchers have focused on people with mental health problems and the dynamic process of their social networks, especially when a crisis (i.e., an adverse event in life) comes into the mobility of social network relationships [21]. The results verified that people with mental health problems tend to reduce their activities in the social network when they face a sort of crisis in their lives. Mobile phone sensors have also been used to establish social networks and collect data. The participants' mental health, for instance, has been evaluated by analysing their behavioral patterns [22]. Boonstra and colleagues believe that their work can detect people with mental health problems at early stages, intervene, and treat them promptly by using their detailed techniques.

## III. The Data Set

A total of 490 Chinese college students were recruited from a single college to build this data set. Participants (394 males and 96 females) were freshmen who lived in the same college residential area. They major in software engineering, aged from 18 to 23 (with average age of 19). The data sets contain four types of data: personality traits, behavior, social

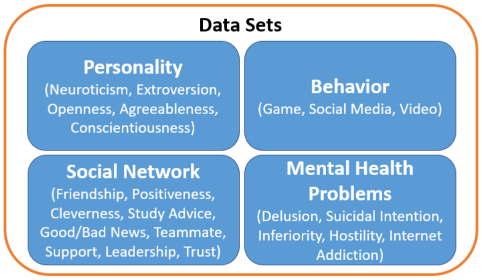



Fig. 1: Four data sets used in this work.

relationships, and mental health problems. We formed a unique data gathering framework which incorporated personality, behavior, social networks, and mental health problems to build and prepare the analytical data set, as presented in Figure 1.

Prior to participation, all participants provided a consent form and consented to take part in the research. They were debriefed after attending the experiments. All participants were informed about the data collected during this research would be coded and kept confidentially by the researcher, with only the researchers having the right to access, all data is securely stored on a password-protected network, all the information is stored with a code, and there is no immediately identifiable information to participants. Part of the research involved taking photographic images. These images are kept secure and stored with no identifying factors, i.e., consent forms and questionnaires.

### A. Personality

To encapsulate the personality factor, the research team formed a consensus to employ the "Big Five" model of personality [23] due to its robustness and relative functionality to this study. Goldberg [23] described this model as a revolution in personality psychology, where five qualities, namely neuroticism, extroversion, openness, agreeableness, and conscientiousness, were detected to cover almost all aspects of personality description. Personality information was gathered by asking participants to complete online questionnaires (80 questions) on the Big Five model. The personality scores were attained by following Goldberg's rating system.

### B. Behavior

The research team collected the online records of the 490 participants from their smart cards over a year to learn about the behavioral aspects of the participants. The record data was updated whenever an individual used a college account to log in to the internal network to use an application. The resulting data contained the date, start time, end time, and the used application. On average, each student has 29277 Internet access records per year. To explore the influence of behavior on mental health, online behaviors were collected, and analysed based on their online dimensions. Finally, the generated data was classified into three behavioral categories: computer games activity, social media activity, and watching videos activity.

### C. Social Networks

Social network information of the $490$ student participants were gathered through online questionnaires, and each participant nominated $5-8$ other people within each question. There were $11$ questions, each one focused on critical dimensions of social networks such as friendship, positiveness, cleverness, study advice, life advice, good news, bad news, teammate, support, leader, and trust. The questions were as follows: (a) list your good friends (friendship); (b) who makes you feel positive and happy? (positiveness); (c) which of your classmates are clever? (cleverness); (d) from whom do you seek advice when you are having trouble with learning? (study advice); (e) who do you seek advice from when you are having trouble in life? (life advice); (f) who do you share the good news with? (good news); (g) who do you share bad news with? (bad news); (h) who will you invite as your teammate? (teammate); (i) Who makes you feel supportive and caring? (support); (j) who has the final decision when you're discussing a choice? (leadership); (k) who do you think is the most trustworthy? (trust). This question set enabled the researchers to construct a detailed understanding of a student participant's social network.

### D. Mental Health Problems

The mental health problems data was achieved by employing the academically accepted scale called *Symptom Checklist-90-R* (SCL-90). This scale enables the participants to self-report measures of psychiatric symptomatology [24]. SCL-90 has 96 items, that covers a wide range of symptoms, and its categories range from feelings, emotions, thinking, consciousness, and behavior to living habits, interpersonal relationships, eating and sleeping, etc. Five of the prominent mental health problems (delusion, suicidal intention, inferiority, hostility, and Internet addiction) were chosen due to their importance in this study. These problems are accounted for their relevance to the study's aims, and for showing high response frequencies to assist in displaying the research aims and data analysis.

## IV. METHODOLOGY

### A. Problem Formulation

The research team investigated the relationship between personality ($\boldsymbol{P}$), behavior ($\boldsymbol{B}$), social relationship ($\boldsymbol{S}$), and mental health problems ($\boldsymbol{M}$). For each participant $i$, their personality features $< p_1, p_2, p_3, p_4, p_5 >$ were constructed based on five values from Big Five model and three categorised behaviors $< b_1, b_2, b_3 >$ based on the time spent on three kinds of activities (game, social media, and video). We analyzed eleven social networks $< s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11} >$ for eleven social relationships. In addition, the model was designed by using personality and behavior features and the network structure information to predict the five mental health problems $< m_1, m_2, m_3, m_4, m_5 >$.

### B. Exponential Random Graph Modelling

To further explore the characteristics of personality and behavioral attributes on social networks, an exponential random graph model (ERGM) is employed [25]. ERGM is a statistical model which is often used in social network research and can support graph learning [26]. It focuses on the edges of the network to explore the impact of various factors on generation and maintenance of the edges, including the network's structure and attributes of the nodes.

ERGM models the collected network information and assumes that the observed network is a specific implementation of a set of all possible networks with similarly essential characteristics such as the same number of nodes in the network. In addition, reasonable assumptions can be made about this stochastic process and network configurations of interest which can be studied such as reciprocated ties, transitive triads or homogeneity in the observed network. Some crucial terms should be defined to understand the ERGM model [27].

- **Statistic**: Network statistic is the count of a particular configuration of different types in the network, such as the number of the reciprocated ties.
- **Parameter**: A parameter measures the importance of statistics, similar to the weight of each variable in the regression model. ERGM assigns a probability to a graph by a sum of statistics weighted by parameters. A positive parameter denotes a higher probability to those graphs with relevant configurations. For example, if a transitive triads parameter is large and positive, graphs with many transitive triads are more probable in the graph distribution for that model. Also, the observed network has more transitivity than a random graph when the model simulates the entire network well.
- **Estimation**: It is the estimate of the parameter. ERGM simulates the distribution of random graphs from a set of initial parameters, determines the parameter values by comparing them with the observed graphs, and repeats the process until the parameter estimates are obtained.
- **Standard error (SE)**: SE is a measure of the precision of parameter estimates. The lower SE indicates the certainty of the estimation. SE provides a measure of the significance of the estimates. For instance, we can determine the parameter is significant with an absolute estimate value more than twice with SE.

The general form of ERGM is presented as follows:

$$Pr(\mathbf{Y} = y) = (1/\kappa)exp\{\sum_{A} \eta_A g_A(y)\} \quad (1)$$

where $y$ is a specific implementation of the network, $A$ is all possible configurations, and $\eta_A$ is the parameter of corresponding configuration $A$. $g_A(y) = \prod_{y_{ij} \in A} y_{ij}$ denotes the network statistic corresponding to the configuration $A$ which equals to 1, if the configuration is observed in the network $y$, otherwise equals to 0. $\kappa$ is a normalizing quantity for ensuring the equation is a proper probability distribution. This general form shows the probability of observing a particular network $y$ and it depends on $g_A(y)$ and $\eta_A$ of all configurations $A$. Refer to [27] for more details about the ERGM.

In this study, each participant was defined and treated as a node, with a focus on their social ties which is presented within a network. Also, the analysis focused on their personality and behavior as being individual node attributes to study the characteristic of these two factors on the ties in the network. ERGM is implemented by R language and $ergm$ package.

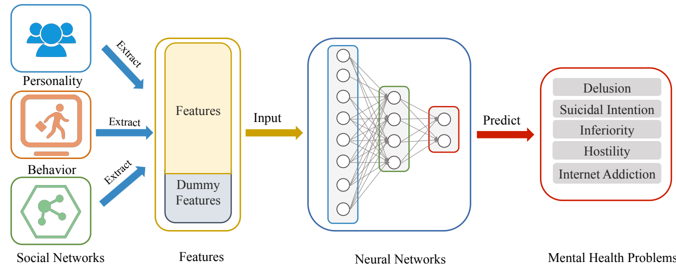


Fig. 2: Overall framework of the prediction model.

### C. Prediction

This section proposes a prediction model based on neural networks, and explains the constructed model. The overall framework of the model is presented in Figure 2. To begin the modelling, three factors (personality, behavior, and social networks) were extracted. Because negative samples significantly outnumbered positive samples, dummy features were added based on a unique strategy to balance the samples. Following this, we will elaborate the model's input features (personality, behavior, and social networks) and the output (the probability of suffering one mental health problem).

*1) Features:* Three factors of personality, behavior, and social networks were considered to predict mental health problems, which are used to extract features for each sample.

**Personality:** This feature included the personality indicators introduced in *previous Section*. These indicators were organised as 5-dimensional features, with each feature representing one corresponding personality indicator.

**Behavior:** To explore the influence of behavior on mental health, students' online activity time was categorised into three classes: video, game, and social media activities. Through the time spent by students, 3-dimensional online features were collected, each of which represents one corresponding class.

**Social networks:** In addition to the two above factors, the authors believe that mental health problems can be inferred by social activity which is usually reflected in social networks. However, determining how to analyse these networks is a challenging task. In this study, Google's PageRank algorithm was employed due to having an effective method for identifying the importance of nodes and recognising the structural identities of the nodes [28]. We will explain the calculation of *PR* node values through the use of $PageRank$ next.

Let $G$ denote a network, and $N = \{v_1, v_2, ...., v_{|N|}\}$, where $|N|$ is the number of the nodes to denote the set of the nodes in the network. Then, we calculate *PR* values through *PageRank* in the following iterative manner:

$$PR(i) = d * \sum_{j \in I_i} \frac{PR(j)}{|O_j|} + \frac{1-d}{|N|} \quad (2)$$

where $I_i$ is the set of $v_i$'s neighbors linking to it, and $O_j$ is the set of $v_j$'s neighbors to be linked. $d$ is a damping factor, and is set to $0.85$. Initially, *PR* values of all nodes are set to $1/|N|$. We stop the iteration when the maximum absolute error was less than $10^{-6}$.

In this study, 11 kinds of social networks were constructed, as presented in *previous Section*. In the social networks, one node represents one student, and the directed edge between the two nodes represents the asymmetric relationship of two corresponding students. It is noted that the 11 networks have identical indexed nodes because they are constructed according to identical students. With *PageRank* operating on the 11 networks, each node has 11 distinct $PR$ values. These values were connected to show 11-dimensional features.

To integrate all studied features from personality, behavior, and social networks, all obtained features described above were combined as one. After that, these features are used as an input for the prediction model which is presented next.

*2) Prediction Model:* Numerous models have been used to predict tasks, such as logistic regression, and support vector machine models (SVM) [29], most of which assumed inputs were linear with outputs. However, we hypothesised that the relationship between features and mental health problems might be nonlinear in this work. Thus, a prediction model based on neural networks was implemented for its great nonlinear design [30]. $X$ denotes the inputs where the $i_{th}$ row $X_i$ are features of $i_{th}$ sample, and $Z$ denotes the output. Based on neural networks, the model is defined as follows:

$$Z = softmax(ReLU(XW^0 + B^0)W^1 + B^1) \tag{3}$$

where $W^0$ and $W^1$ are learned weights, and $B^0$ and $B^1$ are learned biases. Prediction problems were converted to binary classifications and assisted in defining a model for each mental health problem to be predicted. The $Z$ is an $n$-by-2 matrix, where $n$ is the number of samples. $Z_{i0}$ is the probability that sample $i$ is ordinary, and $Z_{i1}$ is the probability of sample $i$ to be abnormal. The $Softmax$ function renders $Z_{i0}+Z_{i1} = 1$. In the designed prediction model, sample $i$ is predicted to suffer the mental health problems if $Z_{i1} > Z_{i0}$.

In our data set, some of the data have corresponding labels while the other part of the data does not. To train the prediction model, we employed the semi-supervised learning. Then, we evaluated the cross-entropy error over all labeled samples:

$$\mathcal{L} = -\sum_{l \in Y_L} y \ln Z_{l0} + (1-y) \ln Z_{l1} \tag{4}$$

where $Y_L$ is set of sample indices with labels, $y$ is indicator that is set to 1 for abnormal samples, and 0 for ordinary ones.

Additionally, regularisation was imposed on all learned weights and biases to prevent over-fitting. Therefore, the final loss function is defined as follows:

$$\mathcal{L}_{loss} = \mathcal{L} + \alpha \mathcal{L}_{reg} \tag{5}$$

where $\alpha$ is the parameter that balances cross-entropy error and regularisation loss that is set to $0.05$. We set hidden layer ($W^0$) with 8 nodes, output layer ($W^1$) with 2 nodes, $learning\ rate$ to $0.005$, and employ $dropout$ [31] with $keeping\ rate = 0.7$. Also, we employ *Adam* to optimize the loss function [32], and stop training when the epoch reaches 200.

TABLE I: Summary of the five mental health problems.

| Mental Health Problems | Abnormal samples | Ordinaries | Totals |
|---|---|---|---|
| Delusion | 68 | 422 | 490 |
| Suicidal Intention | 36 | 454 | 490 |
| Inferiority | 14 | 476 | 490 |
| Hostility | 11 | 479 | 490 |
| Internet Addiction | 16 | 474 | 490 |

*3) Optimization:* Before training the prediction model, training and test sets were firstly generated, which contained positive and negative samples. In this paper, abnormal students were regarded as positive samples, and ordinary ones were coined as negative samples. A challenging issue is maintaining the balance between positive and negative samples since ordinaries outnumber abnormals significantly in our dataset, as indicated in Table I. A strategy was developed to address this issue, whereby balanced training and test samples were generated. This strategy includes the following steps: First, divide the abnormal samples equally into two parts, both seen as positive samples. One part was used for training, and the other was for testing. Second, randomly select 200 samples, from ordinaries, as negative samples used for training. Then, negative examples used for testing, which had the same number as the positive samples used for testing, were randomly selected from the remaining ordinaries. Third, employ Synthetic Minority Oversampling Technique - SMOTE to generate some positive dummy samples. In this article, due to the small number of students with psychological abnormalities, the label are unbalanced in the data. Therefore, we use the SMOTE algorithm to generate minority data in the training set to improve the generalization performance of the algorithm. SMOTE is an oversampling method, and the basic idea is to analyse the minority samples, then, add new samples to the dataset based on minority samples. Given that the features of a positive sample were used for the test, we chose from remaining sample features with the minimum Euclidean distance $x$. The dummy feature $\hat{x}$ is calculated as follows:

$$\hat{x} = x + RANK(0,1)(\tilde{x} - x) \tag{6}$$

where $x$ and $\tilde{x}$ represent two samples with the same label. $RANK(0, 1)$ is a function that produces a random value from 0 to 1, the dummy feature $\hat{x}$ was added to positive samples. Then, *SMOTE* was applied until the number of positive dummy samples reached to 200.

## V. Results

In this section, we describe the distribution of the collected data, and briefly analyze each of the three mental health factors. These factors are personality, behavior, and social network. We analyze the correlations between these factors,

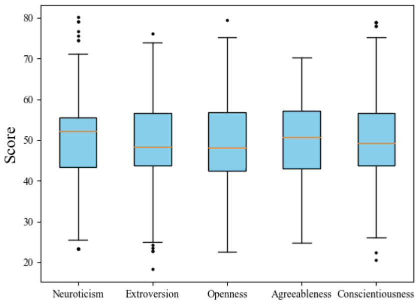



Fig. 3: Box plot of personality. Black dots represent extreme value.

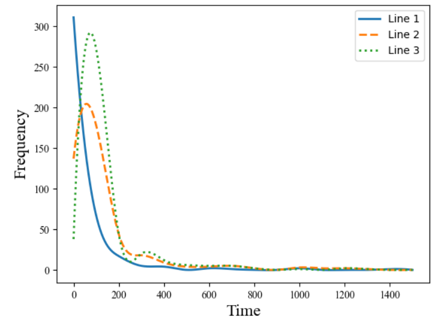



Fig. 4: Behavioral data distribution fit curve. Line 1 - Game Time, Line 2 - Social Media Time, and Line 3 - Video Time.

and outline the overall performance of the proposed model for prediction of the mental health problems.

### A. Analysis of Multiple Types of Factors

*1) Personality:* The data distributions for five personality dimensions (neuroticism, extroversion, openness, agreeableness, and conscientiousness) are presented in Figure 3. Neuroticism has the highest average score above 50, while openness has the lowest score. This indicates that participants are more prone to unstable emotions such as anxiety and depression. At the level of dispersion, the scores for openness and agreeableness were spread out further from their average value than others, as their span and inter-quartile range are much longer than others. There are almost no extreme values for agreeableness and openness, and yet neuroticism has more extreme values. Finally, there are different degrees of skew in the data of Neuroticism, Extroversion and Openness.

*2) Behavior:* This online behavior is separated into three categories: game, social media, and video behaviors. As the survey shows, time activity of students' behaviors was logged by considering these three types of behaviors. The frequency distribution fitting curve is presented in Figure 4. Three different colour curves represent game, social media, and video activities respectively. Through the fitting curve, we observe that the vast majority of participants spent less than $500$ minutes on these three types of behavior, and most people spent less than $200$ minutes on social media. Intuitively, the network time that the video category takes up is the longest, followed by social media. Also, the shortest is the game activity as lots of students don't play online games.

*3) Social Network:* Eleven individual networks were defined by different types of relationships and analysed in the survey. This study focused on the in-degree results in given networks as the out-degree is fixed from $5$ to $8$. As presented in Figure 5, the distributions of in-degrees of $11$ networks look like left-skewed bell-shaped curves. Most of the average network in-degree is about $4.50$, except for the good friend network with the highest average in-degree of $4.97$, which indicates friendship may be the most important social relationship in campus life. That probably accounts for the need for more friends than leaders. Most students prefer to take decisions into their own hands when completing a task, corresponding to the lowest in-degree in the leader network. In addition, about $75\%$ of the network in-degrees are between $2$ and $6$, with the most frequent in-degrees to be $3$ for all networks. A few people have larger in-degrees that occupy a central location in social networks. Due to space limitations, the other two networks (e.g., life advice and bad news) are not placed, and their distributions are almost identical to others.

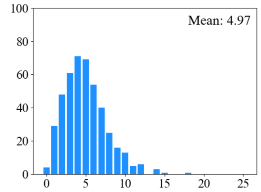
*4) Mental Health Problems:* As illustrated in Table I, the total number of people who participate in the survey is $490$. The vast majority of these people have no psychological problems. The proportion of students who have delusion is the highest, about $14\%$, and the most negligible proportion of the patients belongs to hostility, which is about $2\%$.

### B. Correlation Analysis

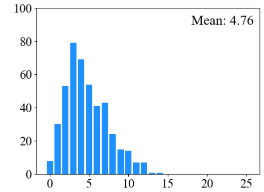
The internal correlation of each factor and external correlation between terms from different factors are explored in this part. More attention needs to be paid to the correlation between mental health problems and the three characteristics (personality, behavior and social network).

*1) Internal Correlation:* The internal correlation explains the relationship between different items under the same factor. The results are too simple to list and are explained briefly here. We analysed the in-degree correlation between various networks and found that some students are at the center of social relations. More specifically, if a person is in one social network center with high in-degree, he is also in a central position in other networks. People with strong social skills are active in a variety of social relationships. In terms of various online behaviors, there was a positive but weak correlation between video and social media. Regarding personality, neuroticism has a robust negative correlation with the other four personality types. The correlation between extroversion, openness, agreeableness, and conscientiousness is positive and significant ($p < .005$). For mental health problems, hostility and Internet addiction have a significantly positive correlation.

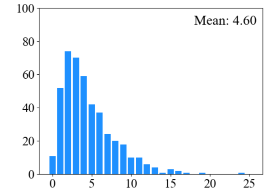
*2) External Correlation:* The correlations of mental health problems with personality and behavior are illustrated in Table II. Personality and mental health problems have significantly weak correlations where mental health problems refer to abnormalities in psychological characteristics. Delusion has a significantly weak correlation with all personalities and is only



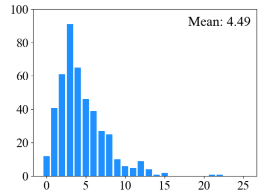

(a) Friendship



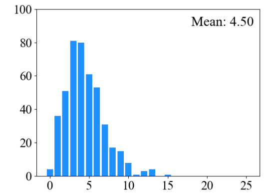

(b) Positiveness



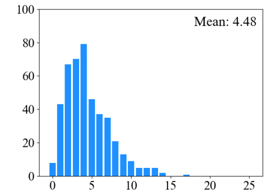

(c) Cleverness



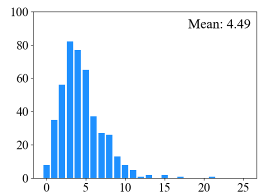

(d) Study Advice



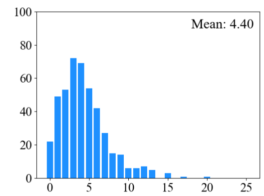

(e) Good News



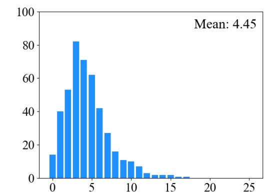

(f) Teammate




(g) Support




(h) Leadership




(i) Trust

Fig. 5: Histograms of in-degree for various social networks. The ordinate is the frequency and the abscissa is the in-degree.

positively related to neuroticism. This denotes the difference of patients with delusions, that is evident in psychological characteristics. Inferiority is related obviously to personality, while hostility is not associated with any terms. Thus, the more negative and paranoid personality, the more likely to suffer from delusion and inferiority. For the correlation between mental problems and online behavior, a significant correlation is shown between game time and inferiority (0.14, $p < 0.01$). Therefore, schools and teachers should pay attention to the students who spend a lot of time on online games. They may suffer from inferiority because of long-term lack of social communication. Also, they become isolated in the game environment while knowing that behavior has no significant correlation with personality.

### *C. Network Analysis*

We firstly study the correlation between social network in-degree and three factors (personality, behavior, and psychological problems). The study demonstrates that the relationship between online behavior and social network is weak. In terms of personality, extroversion is significantly related to all social network in-degree ($p < 0.01$). This proves that an extroverted person is better at socialising. Also, openness has a significant correlation with the cleverness (0.12, $p < 0.05$) and the teammate (0.11, $p < 0.05$) networks, which indicates that creative people are more likely to be considered clever and more likely to rely on others when making decisions. Agreeableness is related to the good news network (0.12, $p < 0.05$), which verifies that we tend to share the good news with those who can understand our happy emotions. People with high conscientiousness are more likely to be considered clever and take center positions in the study of advice, good news, bad news, teammate, leader, and trust networks ($p < 0.05$), where we need a reliable person to guide us sometimes. The weak relation between social network in-degree and mental problems suggests that the shallow feature of the network, such as in-degree, cannot explain the mental state of the nodes.

The ERGM analysis of the three factors (personality, behavior, and mental health problem) within 11 social networks is shown in Table III. Insignificant parameters are not displayed. From the result of ERGM, participants tend to nominate those students with high extroversion, and people with high agreeableness are more appreciative of the people around them. They are more likely to nominate others in all social relationships. Students who suffer from delusion are not likely to nominate people similar to themselves and have less social relations. Additionally, 1) neurotic people may be more likely to be nominated in the friendship, positiveness, cleverness, study advice, and support networks; 2) people with high openness prefer to choose creative and imaginative people in cleverness, advice, news, teammate, leader, and trust networks; 3) in the cleverness network, students are likely to nominate people with different extroversion degrees, and people with low agreeableness and high conscientiousness receive more nominations; 4) in the study advice network, people prefer to

TABLE II: Correlations between mental health problems and other factors. We use the student test to test the significance of the correlation coefficient. **:$p$ <.01, *:$p$ <.1

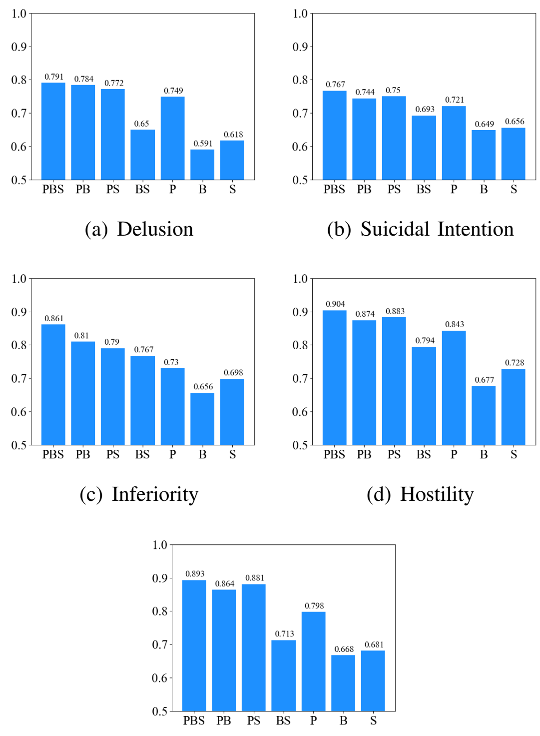
| | Neuroticism | Extroversion | Openness | Agreeableness | Conscientiousness | Game Time | Social Media Time | Video Time |
|---|---|---|---|---|---|---|---|---|
| Delusion | .25** | -.24** | -.14** | -.21** | -.20** | .01 | -.02 | .06 |
| Duicidal Intention | .13** | -.07 | .00 | -.12* | -.11* | -.01 | -.02 | -.04 |
| Dnferiority | .10* | -.04 | -.13** | -.09* | -.10* | .14** | .03 | -.02 |
| Hostility | .08 | .07 | -.06 | .04 | .00 | .07 | .06 | -.02 |
| Internet Addiction | -.01 | 0.08* | -.08* | .02 | -.07 | .01 | -.04 | -.01 |

choose those with low openness and high conscientiousness. Perhaps the study method of creative students is not suitable for most people; 5) agreeableness plays an integral part in cleverness, life advice, and good news networks. Also, conscientiousness is vital in study advice and teammate relationships; 6) in support and leader networks, people tend to choose those who are different from themselves in neuroticism.

### *D. Prediction Performance*

Correlation analysis discloses the relationship between mental problems and the three factors - personality, behavior, and social relationship. In this section, the performance of different factors is compared with predicted features, and our model's efficacy is evaluated. In addition, we claim that the proposed model is robust enough to cover the individual differences between demographical data like academic performance, gender, and ethnicity. According to the statistical analysis, the proposed model outperforms traditional analysis methods. All presented experiments were run 50 times. This indicates that all of the reported results were averaged.

*1) Accuracy:* In this section of the paper, the efficacy of the proposed model for predicting the five kinds of mental illnesses is evaluated. In addition, the performance of different factors and their combinations are compared. For the brevity of the experimental results, personality, behavior, and social networks are abbreviated as P, B, and S, respectively. PBS is used to denote the three factors, PB to denote the two factors of personality and behavior, and P to denote the single factor of personality. Other abbreviations are used similarly.




(a) Delusion (b) Suicidal Intention (c) Inferiority (d) Hostility (e) Internet Addiction

Fig. 6: Accuracy of the prediction model. P, B, and S represent Personality, Behavior, and Social Networks, respectively.

Figure 6 illustrates the prediction accuracy of different factors and their combinations. As can be seen from this figure, no matter which illness is predicted, the prediction model performs at its best when all three factors are included. PBS outperforms the second-best combination, PB, by up to 5.1% for predicting inferiority. Therefore, these results indicate the necessity to consider all three factors when attempting to predict mental illnesses. The findings are listed below:

- When a single factor is used for prediction, the implementation of P always performs better than B or S. This verifies that P is better reflected and has a greater correlation with mental conditions than B or S.
- The single implementation of the factors B or S achieves impressive performance, even accuracy of 0.728. This result is captivating because only a few studies investigated the correlation between the two factors and mental illnesses. This work further demonstrates that more studies should explore their correlation. It is also observed that S leads to more beneficial outcomes than B. This result could be explained because the feature set obtained through S is 11-dimensional which is more comprehensive than B which is 3-dimensional. Thus, more information is extracted from S.
- Using the combination of two factors consistently outperforms the methods which use them separately. The implementation of combined three factors provides the best performance. The three factors stem from three distinct aspects, capturing different dimensions of information, which is beneficial in predicting the mental illnesses. The combination of more factors integrates more diverse and valuable information to achieve better performance.

*2) Robust:* To effectively demonstrate the robustness of our prediction model, and to reflect individual differences in social

TABLE III: Significant estimates of various parameters for eleven social networks. The estimates have been multiplied by 1000. N: neuroticism, E: extroversion, O: openness, A: agreeableness, C: conscientiousness, b1: game time, b2: social media time, b3: video time, m1: delusion, m2: suicidal intention, m3: inferiority, m4: hostility, m5: Internet addiction. receiver: the specific attribute value of receiver in tie, sender: the specific attribute value of sender in tie. absdiff: the absolute value of the difference between the specific attribute values of receiver and sender. match: the case that receiver and sender have same value in specific attribute. ***:$p < .001$, **:$p < .01$, *:$p < .05$

| | Friendship | Positiveness | Cleverness | Study Advice | Life Aadvice | Good News | Bad News | Teammate | Support | Leadership | Trust |
|---|---|---|---|---|---|---|---|---|---|---|---|
| absdiff.N | - | - | - | - | - | - | - | - | -6.815* | -6.071* | - |
| absdiff.E | - | - | 5.794* | - | - | - | - | - | - | - | - |
| absdiff.O | - | - | -9.387** | -7.775* | -7.133* | -6.701* | -5.958* | -8.371** | - | -10.44*** | -6.598* |
| receiver.N | 8.161** | 6.578* | 8.209** | 7.063* | - | - | - | - | 6.226* | - | - |
| receiver.E | 17.69*** | 19.72*** | 13.50*** | 13.15*** | 16.50*** | 15.23*** | 16.71*** | 16.63*** | 15.98*** | 18.88*** | 13.39*** |
| receiver.O | - | - | - | -7.245* | - | - | - | - | - | - | - |
| receiver.A | - | - | -6.649* | - | 7.059* | 6.559* | - | - | - | - | - |
| receiver.C | - | - | 21.73*** | 15.70*** | - | - | - | 8.667* | - | - | - |
| sender.A | 20*** | 20.63*** | 21.87*** | 19.99*** | 21.34*** | 21.10*** | 19.88*** | 21.48*** | 21.13*** | 19.94*** | 21.74*** |
| match.m1 | -140.2** | -119.5* | -110.3* | -190.3*** | -198.8*** | -184.2*** | -149.2** | -192.6*** | -232.8*** | -110.9* | -173.4** |
| match.m3 | - | - | - | - | - | - | - | - | - | - | -208.2* |

TABLE IV: Accuracies of PBSA, PBSG, and PBSE. The numbers in parentheses are the percentage changes of accuracy after the corresponding feature is added to PBS.

| Mental Health Problem | PBSA | PBSG | PBSE |
|---|---|---|---|
| Delusion | 0.792(+0.1%) | 0.787(-0.4%) | 0.792(+0.1%) |
| Suicidal Intention | 0.764(-0.3%) | 0.764(-0.3%) | 0.765(-0.2%) |
| Inferiority | 0.857(-0.4%) | 0.864(+0.3%) | 0.868(+0.7%) |
| Hostility | 0.900(-0.4%) | 0.900(-0.4%) | 0.908(+0.4%) |
| Internet Addiction | 0.897(+0.4%) | 0.891(-0.2%) | 0.897(+0.4%) |

background and academic performance, the GPA (Grade Point Average), gender, and ethnicity of all students were collected and were regarded as three distinct features within this paper. The following abbreviations have been used to understand the analysis: A as GPA, G as gender, and E as ethnicity. The accuracy of PBSA, PBSG, and PBSE is presented in Table IV, where the numbers in parentheses are the percentage changes of accuracy after the corresponding feature is added to PBS.

From Table IV, it is evident that all percentage changes are less than $0.5\%$, with the exception of PBSE when predicting inferiority. Therefore, this result reveals that our proposed prediction model is robust enough to reflect individual differences by using the targeted demographics of GPA, gender, and ethnicity. Although considering individual differences leads to slight changes in the overall accuracy, the underlying reason is worth exploring. However, it is beyond the scope of this paper and is left as an avenue for future research.

*3) Comparison:* To compare our proposed prediction method with traditional approaches, LR, SVM, KNN, Decision Tree, and Nave Bayes were chosen as they are well-supported and appropriate methods for statistical analysis. Table V presents the prediction accuracies of our proposed prediction method and the traditional methods. The percentages of a setback that the conventional methods suffer are highlighted within parentheses of the table. The data in Table V verifies that our proposed prediction model significantly outperforms the five traditional methods. The largest gain achieved by our model is $28.7\%$, whose complementary approach is Naive Bayes. For the other methods, the least gain is up to $6.1\%$, a significant improvement. LR and SVM assume the dependency between features and tasks by linear functions. KNN, Decision Tree, and Naive Bayes hypothesize the level of independence between features. Experimental results suggest that these traditional methods are insufficient to predict mental health problems. Neural networks capture the nonlinear relation between features and work through activation functions and mine correlation between features through multiple stacks. Indeed, the proposed model is based on neural networks; thus, it predicts the mental illness more effectively. Furthermore, these results verify that such nonlinearity and correlation are worth considering for future mental health research.

## VI. Conclusions

This paper investigates the relation of multiple types of factors, including personality, behaviors, and social networks, with mental health. We provided a comprehensive analysis for residential college freshmen students based on the data collected from real-world sources. The achievements are as follows: 1. mental problems correlate strongly with personality and behaviors; 2. students who play online games for a long time may suffer from inferiority; 3. the more negative, closed, and paranoid the personality, the more likely they may suffer from inferiority and delusion. Meanwhile, we propose a neural network-based model to integrate these factors to predict various mental health problems based on their correlation. Google's PageRank captured the social relationship factor to consider the structural information of a network. The experimental results verified that our proposed model predicted the students' mental health with high accuracy and outperformed traditional statistical methods. All three factors can significantly improve the prediction efficiency, where personality was the most compelling feature. In the future, we will collect more diverse student data to conduct experiments to obtain more effective experimental results.

TABLE V: Accuracy comparison of our prediction model with traditional models. The numbers in parentheses show the traditional models difference with ours. LR, SVM and KNN are the abbreviations of Logistic Regression, Support Vector Machine and K-Nearest Neighbor.

| Mental health problem | Our Model | LR | SVM | KNN | Decision Tree | Naive Bayes |
|---|---|---|---|---|---|---|
| Delusion | 0.791 | 0.761(-3.0%) | 0.744(-4.7%) | 0.711(-8.0%) | 0.697(-9.4%) | 0.714(-7.7%) |
| Suicidal Intention | 0.767 | 0.703(-6.4%) | 0.727(-4.0%) | 0.732(-3.5%) | 0.718(-4.9%) | 0.626(-14.1%) |
| Inferiority | 0.861 | 0.774(-8.7%) | 0.800(-6.1%) | 0.824(-3.7%) | 0.764(-9.7%) | 0.764(-9.7%) |
| Hostility | 0.904 | 0.879(-2.5%) | 0.883(-2.1%) | 0.875(-2.9%) | 0.883(-2.1%) | 0.800(-10.4%) |
| Internet Addiction | 0.893 | 0.797(-9.6%) | 0.866(-2.7%) | 0.848(-4.5%) | 0.790(-10.3%) | 0.606(-28.7%) |


## Acknowledgements

This work is partially supported by the National Natural Science Foundation of China under Grants No. 62076051. The authors would like to thank Shengjun Xu, Yue Hu, Nan Shi, Lei Wang, Lulu Li, and Grant Meredith for help with the first draft of the paper, as well as all participants of the experiments.


## References


[1] T. Kushner and A. Sharma, "Bursts of activity: Temporal patterns of help-seeking and support in online mental health forums," in *Proceedings of The Web Conference 2020*, 2020, pp. 2906–2912.

[2] D. Zhang, N. Shi, C. Peng, A. Aziz, W. Zhao, and F. Xia, "Mam: A metaphor-based approach for mental illness detection," in *International Conference on Computational Science (ICCS)*, 2021.

[3] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, and A. S. Pentland, "Daily stress recognition from mobile phone data, weather conditions and individual traits," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 477–486.

[4] L. Meng, Y. Hulovatyy, A. Striegel, and T. Milenković, "On the interplay between individuals evolving interaction patterns and traits in dynamic multiplex social networks," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 1, pp. 32–43, 2016.

[5] W. E. Copeland, E. McGinnis, Y. Bai, Z. Adams, H. Nardone, V. Devadanam, J. Rettew, and J. J. Hudziak, "Impact of covid-19 pandemic on college student mental health and wellness," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 60, no. 1, pp. 134–141, 2021.

[6] D. Eisenberg, J. Hunt, and N. Speer, "Mental health in american colleges and universities: variation across student subgroups and across campuses," *The Journal of nervous and mental disease*, vol. 201, no. 1, pp. 60–67, 2013.

[7] Y.-l. Ye, P.-g. Wang, G.-c. Qu, S. Yuan, P. Phongsavan, and Q.-q. He, "Associations between multiple health risk behaviors and mental health among chinese college students," *Psychology, health & medicine*, vol. 21, no. 3, pp. 377–385, 2016.

[8] K. E. Champion, M. Mather, B. Spring, F. Kay-Lambkin, M. Teesson, and N. C. Newton, "Clustering of multiple risk behaviors among a sample of 18-year-old australians and associations with mental health outcomes: a latent class analysis," *Frontiers in public health*, vol. 6, p. 135, 2018.

[9] E. J. Wright, K. M. White, and P. L. Obst, "Facebook false self-presentation behaviors and negative mental health," *Cyberpsychology, behavior, and social networking*, vol. 21, no. 1, pp. 40–49, 2018.

[10] S. P. Becker, A. S. Holdaway, and A. M. Luebbe, "Suicidal behaviors in college students: frequency, sex differences, and mental health correlates including sluggish cognitive tempo," *Journal of Adolescent Health*, vol. 63, no. 2, pp. 181–188, 2018.

[11] H. H. Clark and E. V. Clark, *Psychology and Language: An Introduction to Psycholinguistics*. Harcourt College Pub, 1977.

[12] J. Liu, J. Tian, X. Kong, I. Lee, and F. Xia, "Two decades of information systems: a bibliometric review," *Scientometrics*, vol. 118, no. 2, pp. 617–643, 2019.

[13] L. Keane and M. Loades, "Low self-esteem and internalizing disorders in young people–a systematic review," *Child and Adolescent Mental Health*, vol. 22, no. 1, pp. 4–15, 2017.

[14] M. H. De Moor, A. Beem, J. H. Stubbe, D. I. Boomsma, and E. J. De Geus, "Regular exercise, anxiety, depression and personality: a population-based study," *Preventive medicine*, vol. 42, no. 4, pp. 273–279, 2006.

[15] C. R. Colvin, J. Block, and D. C. Funder, "Overly positive self-evaluations and personality: Negative implications for mental health." *Journal of personality and social psychology*, vol. 68, no. 6, p. 1152, 1995.

[16] S. Sultan, F. Kanwal, and I. Hussain, "Moderating effects of personality traits in relationship between religious practices and mental health of university students," *Journal of religion and health*, pp. 1–11, 2019.

[17] J. Wang, B. Lloyd-Evans, D. Giacco, R. Forsyth, C. Nebo, F. Mann, and S. Johnson, "Social isolation in mental health: a conceptual and methodological review," *Social psychiatry and psychiatric epidemiology*, vol. 52, no. 12, pp. 1451–1461, 2017.

[18] A. Hosseini, T. Chen, W. Wu, Y. Sun, and M. Sarrafzadeh, "Heteromed: Heterogeneous information network for medical diagnosis," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018, pp. 763–772.

[19] F. Xia, S. Yu, C. Liu, J. Li, and I. Lee, "Chief: Clustering with higher-order motifs in big networks," *IEEE Transactions on Network Science and Engineering*, 2021.

[20] N. Park, Y. Jang, B. Lee, D. Chiriboga, S. Chang, and S. Kim, "Associations of a social network typology with physical and mental health risks among older adults in south korea," *Aging & mental health*, vol. 22, no. 5, pp. 631–638, 2018.

[21] S. Walker, A. Kennedy, I. Vassilev, and A. Rogers, "How do people with long-term mental health problems negotiate relationships with network members at times of crisis?" *Health Expectations*, vol. 21, no. 1, pp. 336–346, 2018.

[22] T. W. Boonstra, J. Nicholas, Q. J. Wong, F. Shaw, S. Townsend, and H. Christensen, "Using mobile phone sensor technology for mental health research: Integrated analysis to identify hidden challenges and potential solutions," *Journal of medical Internet research*, vol. 20, no. 7, p. e10131, 2018.

[23] L. R. Goldberg, "The development of markers for the big-five factor structure." *Psychological assessment*, vol. 4, no. 1, p. 26, 1992.

[24] L. R. Derogatis, K. Rickels, and A. F. Rock, "The scl-90 and the mmpi: A step in the validation of a new self-report scale," *The British Journal of Psychiatry*, vol. 128, no. 3, pp. 280–289, 1976.

[25] A. Stivala, G. Robins, and A. Lomi, "Exponential random graph model parameter estimation for very large directed networks," *PloS one*, vol. 15, no. 1, p. e0227804, 2020.

[26] F. Xia, K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, and H. Liu, "Graph learning: A survey," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 109–127, 2021.

[27] D. Lusher, J. Koskinen, and G. Robins, *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press, 2013.

[28] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[29] T. Guo, F. Xia, S. Zhen, X. Bai, D. Zhang, Z. Liu, and J. Tang, "Graduate employment prediction with bias," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 670–677.

[30] L. Aitchison, "Why bigger is not always better: on finite and infinite neural networks," in *International Conference on Machine Learning*. PMLR, 2020, pp. 156–164.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[32] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.