Surf or sleep? Understanding the influence of bedtime patterns on campus

Teng Guo, Linhong Li, Dongyu Zhang
School of Software,
Dalian University of Technology,
Dalian 116620, China
teng.guo@outlook.com;
li.linhong@outlook.com;
zhangdongyu@dlut.edu.cn

Feng Xia*
School of Engineering, IT and Physical Sciences,
Federation University Australia,
Ballarat, VIC 3353, Australia
f.xia@acm.org

ABSTRACT

Poor sleep habits may cause serious problems of mind and body, and it is a commonly observed issue for college students due to study workload as well as peer and social influence. Understanding its impact and identifying students with poor sleep habits matters a lot in educational management. Most of the current researches is either based on self-reports and questionnaires, suffering from small sample size and social desirability bias, or the methods used are not suitable for the education system. In this paper, we develop a general data-driven method for identifying students' sleep patterns according to their Internet access pattern stored in the education management system and explore its influence from various aspects. First, we design a Possion-based probabilistic mixture model to cluster students according to the distribution of bedtime and identify students who are used to stay up late. Second, we profile students from five aspects (including eight dimensions) based on campus-behavior data and build Bayesian networks to explore the relationship between behavioral characteristics and sleeping habits. Finally, we test the predictability of sleeping habits. This paper not only contributes to the understanding of student sleep from a cognitive and behavioral perspective but also presents a new approach that provides an effective framework for various educational institutions to detect the sleeping patterns of students.

1. INTRODUCTION

Sleep matters a lot for health and well-being [38; 44; 4]. Poor sleep habits like insomnia can easily cause serious problems like worsened health-related quality of life and poor health outcomes, especially for college students with psychological immaturity [13]. Research have shown that the majority of college students suffer from sleep disorders like feeling tired, fatigue, or daytime sleepiness and sleeping less than eight hours per night during the school week [6; 27]. 68% of American teenagers report that their sleep time each night is below the recommended eight-ten hours [15] and in East Asia, the sleep time of adolescents on weekdays night is below six hours [29; 14]. Therefore, a significant topic of educational research is identifying college students with bad sleep

---
*Corresponding author

habits timely [40; 3], and understanding its influence. However, detecting sleep patterns accurately faces tremendous challenges. Previous research are mainly based on questionnaires, which is time-consuming and costly making, and it is hard to be scale to a lot of students. In the last decade, scientists propose a variety of techniques for sleep tracking, and some related products are already in use [2; 32]. These techniques can only be used for small-scale analysis due to the high cost of the associated equipment. Some studies have tried to do large-scale sleep-related research through a data-driven approach like [4]. However, such data is difficult to access except for large companies like Microsoft, Google, etc. A data-driven framework for sleep habit detection and analysis that can serve the daily management of universities is urgently needed.

During the past decade, we have witnessed an increased utilization of electronic devices like smartphones and tablets [7]. As they become more lightweight, people may easily use these devices even in bed. Electronic media use in bed after bedtime is more common in younger compared to older age groups [8]. Research show that more than 90% of young people are used to playing mobile phones or tablets before bedtime [21; 19]. Meanwhile, for college students, the campus network is a good choice because of its low price and faster internet speed, which generates tons of data of internet access logs stored in the local education management system. This provides university management divisions a new way to detect the sleep patterns of college students. However, new challenges are introduced as well. First, these data are not easy to obtain due to privacy protection [23; 36]. Second, the volume of data collected in this way is much larger than the traditional method. Targeted methods need to be designed for efficient data mining.

In this paper, we are devoted to identifying students with the habit of staying up late and explore the influence of this habit. First, we use the Internet access pattern to identify the bedtime of each student. Second, we design a Possion-based probabilistic mixture model to cluster students according to the distribution of bedtime and identify students who stay up late. Third, besides sleep habit, we profile students from five aspects in eight dimensions, including interest (reading status, app preference status, and surfing length status), orderliness (breakfast orderliness status and bath orderliness status), finance (financial status), academic performance (academic performance status) and gender. Then we build Bayesian networks to explore the re-

lationship between these characteristics and sleeping habits. Finally, we design an experiment of predictive inference to estimate the predictability of sleep status.

Our contributions could be summarized as follows:

- We design a Possion-based probabilistic mixture model suitable for education management divisions to identify students' sleep patterns.

- We profile students behavior from various aspects and build Bayesian networks to explore the relationship between these characteristics and sleeping habits.

- We conduct comprehensive experiments on a large-scale educational dataset and the extensive results demonstrate the effectiveness.

This paper is organized as follows. In the next section, related work is reviewed. The methods are presented in Section 3. In Section 4, we introduce the experiment setting and analyze the results. In the final section, we present the discussion and conclusion of our work.

2. RELATED WORK

The sleep patterns of college students have attracted the attention of scholars from various fields. As the negative impact of poor sleep quality has been demonstrated, the main concern of scientists is the factors that affect sleep quality. Buboltz et al. [10] carry out a study focused on the sleep disturbance of college students. The results show that most college students suffer from sleep disturbance and women exhibit more sleep disturbance than man do. Moreover, they give some suggestions to reduce the negative influence on academic performance. Tsai et al. [42] also focus on how gender can impact students' sleep. Brown et al. [9] use regression modeling to capture the relationship between sleep hygiene and sleep practices. The results demonstrate that variable sleep schedules, going to bed thirsty, environmental noise, and worrying can all impact the quality of sleep. Tavernier et al. [41] explore how the social tie can impact sleep and the results show that positive social ties can contribute to good sleep quality. Patrick et al. [35] focus on the influence of energy drinks and binge drinks on sleep quality. The results show that drinking the above two drinks can easily lead to poor sleep quality and fatigue the next day. Meanwhile, Schneider et al. [37] focus on adolescents and explore the relationship between sleep characteristics and body-mass index. In this study, they emphasize the importance of gender in related research. Orzech et al. [33] proves that longer digital media use two hours prior to bedtime leads to reduced sleep and later bedtime. They obtain an interesting conclusion that more diverse media uses are associated with longer sleep durations. Moreover, some scholars have focused on sleep patterns and their impact on daily life. Kelley et al. [27] explain adolescents' fatigue and irritability by the sleep problems caused by the disorder of their biological time and social time. Althoff et al. [4] carry out a convincing experiment based on a search engine Bing from Microsoft including 3 million nights of sleep and 75 million interaction tasks. The results show how sleep deprivation can seriously affect daily performance. However, as mentioned before, such type of data is inaccessible for education management divisions.

Thanks to the development of technologies, more and more electronic devices are appearing in our daily lives [11] and scientists try to use these advanced technologies (e.g. remote bio-signal monitoring technology) to solve sleep problems. Vroon et al. [43] propose an actuated robotic pillow to improve the quality of sleep through enhancing the interaction with users. De Arriba et al. [16] propose a smartphone-based indicator that can calculate sleep patterns automatically. They try to use this indicator to facilitate the construction of software services in order to improve the efficiency of the learning process. Liu et al. [28] propose a mobile sleep-management system integrated with self-regulated learning strategies and cognitive behavioral therapy. This system can help students to improve their sleep quality by improving their daily schedule and modify strategies to cultivate good learning and healthy lifestyle habits. Zhao et al. [45] design a smart-home system based on the Internet of things, which realizes the function that judges the user's mood and automatically plays corresponding music through facial expression recognition. However, the methods mentioned above are difficult to be used for large-scale sleep detection on campus due to their high cost.

3. METHODS

In this section, we describe the models used in the experiment. First, we introduce a mixed probability model used to cluster students' sleep habits. Then, we introduce the Bayesian network model used to analyze the influence of different sleeping habits.

3.1 Sleep Pattern Recognition

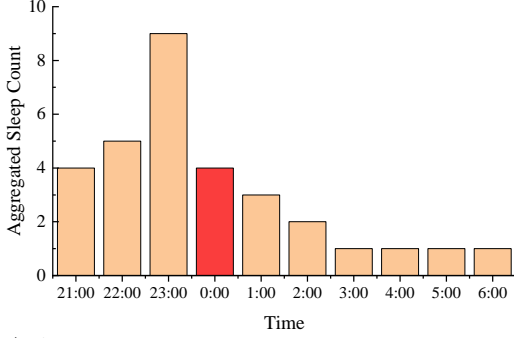#### 3.1.1 Aggregated Sleep Count

For universities that have campus-dedicated networks deployed, web logs can be used to calculate the bedtime of students every night based on the assumption that students will access the Internet through mobile phones or computers before bedtime [5; 21; 19]. In other words, we can use the time of the last network signal every day to quantify the time students go to bed. To mine sleep patterns of different students, we define *aggregated sleep count*, which represents the total number of days a student goes to sleep at a particular time slot. For example, in a certain month, daily bedtime is shown in Figure 1(a). Let's take 0:00 as an example. In this month, there are 4 days that sleep at 0:00 o'clock shown in Figure 1(a) with the red circle, the aggregated sleep count of 0:00 in this month is 4 shown in Figure 1(b) with a red bar. The *aggregated sleep count* of each time period is shown in Figure 1(b).

3.1.2 Mixture Model with Gamma Priors

Inspired by [34], we design a Poisson mixture model to mine the pattern on their aggregated sleep counts in each time period. For student i, we let s_i be the vector of aggregated sleep counts, where $i = 1, 2, 3...N$. The dimensionality D of each vector is the number of time windows ($D = 16$ in this case that we divided the time into 16 segments from 9:00 pm to 4:30 am and the time span of each segment is 30 minutes because: 1, the last class generally is over at 9:00 pm. 2, the sun rises at around 4:30 am). Thus, our data consists of N students each with D-dimensional vectors of aggregated sleep counts.

Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1 21:00	2 21:00	3 23:00	4 23:00	5 23:00	6 23:00
7 0:00	8 21:00	9 21:00	10 0:00	11 1:00	12 1:00	13 23:00
14 0:00	15 22:00	16 22:00	17 22:00	18 2:00	19 22:00	20 6:00
21 23:00	22 3:00	23 4:00	24 2:00	25 0:00	26 23:00	27 1:00
28 23:00	29 23:00	30 5:00	31 22:00			

(a) Bedtime of a student in a month: this is the calendar of a certain month and the number in the middle of the square is the bedtime of the day.

(b) Aggregated Sleep Count of a student in a month: this is the aggregated sleep count of each time period during the month.

Figure 1: Illustration for aggregated sleep count.

To mine the pattern behind this data, we build a probabilistic mixture model with Poisson components. In terms of notation, we let M be the number of components with an index $m = 1, 2, 3...M$. The unobserved latent variable z_i takes values from the set $\{1, 2, ...M\}$ that corresponds to the latent component of student i. Each component consists of a vector of Poisson rate parameterss, $\boldsymbol{\lambda_m} = [\lambda_{m1}, ...\lambda_{md}, ...\lambda_{m16}]$, where d from $[1, 2, 3.....D]$ represents a time window mentioned before. For example, a low value for λ_{md} represents that students rarely sleep during time window d. On the contrary, a high value for λ_{md} represents that students always sleep during time window d.

In this study, the Bayesian approach is used to fit our mixture model. To encourage the model to avoid degenerate solutions, we use Gamma prior distribution for the rate parameter λ_{md} according to the conjugate prior theory. More precisely, we use hyper parameters $\alpha = 1.1$ and $\beta = 0.1$ for the Gamma distribution, which makes it a step-like function that puts zero probability mass at λ_{md} and provides a relatively flat uninformative prior distribution over positive rate parameter values [34].

According to the total probability formula, the probability of each student i under this mixture model is:

$$p(\boldsymbol{s_i}|\boldsymbol{\lambda}) = \sum_{m=1}^{M} p(\boldsymbol{s_i}|z_i = m, \boldsymbol{\lambda_m})p(z_i = m) \quad (1)$$

where $p(z_i = m)$ is the marginal mixing weight for each component. Assuming conditional independence of aggregated sleep count s_{md} given component m, the probability distribution of each component can be decomposed as:

$$p(\boldsymbol{s_i}|z_i = m, \boldsymbol{\lambda_m}) = \prod_{d=1}^{D} p(s_{md}|\lambda_{md}, z_i = m) \quad (2)$$

3.1.3 Parameter Estimation

Expectation-Maximization (EM) algorithm, which is widely used in fitting a mixture model to data[17; 31], is used in this experiment to estimate the parameter λ_m. More specifically, the optimization objective in this experiment is to maximize the product of the data likelihood times the prior probability. This optimization process produces two results. The first one is maximum a posteriori parameter estimates for the Poisson components in the model and the second one is membership weights ω_{im} that is the probability that student i belongs to component m. Each iteration of the EM algorithm consists of two processes, the E (expectation) process and the M (maximization) process. In the E process, the probability of membership is computed for each component $m = 1, 2, ...M$, for each student $i = 1, 2, ...N$.

$$\begin{aligned} \omega_{im} &= p(z_i = m|\boldsymbol{s_i}, \boldsymbol{\lambda}, \alpha, \beta) \\ &\propto p(z_i = m, \boldsymbol{s_i}, \boldsymbol{\lambda_m}|\alpha, \beta) \\ &\propto p(\boldsymbol{s_i}|z_i = m, \boldsymbol{\lambda_m})p(\boldsymbol{\lambda_m}|\alpha, \beta)p(z_i = m) \end{aligned} \quad (3)$$

where ω_{im} is the probability that student i belongs to component m. In the M process, conditioned on the set of membership probabilities ω_{im}, we estimate each parameter via MAP (Maximum Posteriori) estimation:

$$\hat{\boldsymbol{\lambda_m}} = \frac{\sum_i \omega_{im}(\boldsymbol{s_i} + \alpha - 1)}{\sum_i \omega_{im}(\beta + 1)} \quad (4)$$

$$\hat{p}(z_i = m) = \frac{\sum_i^N \omega_{im}}{N} \quad (5)$$

The outputs of the M process provide the input for the next E process, and thus, the cycle of E and M processes continue iteratively.

3.2 Bayesian Network

In this study, Bayesian network (BN) is used to explore the relevance between bedtime patterns and behavioral characteristics [25; 20]. Based on Markov property, the BN uses a set of local distributions to expresses the global joint distribution [1; 12].

3.2.1 Bayesian network

A BN consists of two parts: a directed acyclic graph (DAG) and a set of conditional probability distribution, which can be represented as $\mathcal{B} = \langle G_d, \boldsymbol{\Theta} \rangle$. The graph G_d represents a directed acyclic graph whose vertices correspond to random variables $X_1, X_2, ...X_n$ and whose edges represent direct dependencies between the variables, where n represent the number of variable in $\mathcal{B}$. The $\boldsymbol{\Theta} = (\boldsymbol{\theta_i})_{1 \leq i \leq n}$ denotes the set of all parameters.

The nations we used are introduced as follows:

- n represent the number of variable in $\mathcal{B}$ and each node X_i takes r possible values $x_1, x_2, ...x_r$

- $\pi(X_i)$, denoted the parent nodes of each variable X_i in G_d, and has $q_i = \prod_{m \in \Omega_i} r_m$ possible configurations, where $\Omega_i = \{m; X_m \in \pi(X_i)\}$

- $ch(X_i)$ denotes the children nodes of each variable X_i in G_d

- The conditional distribution of $X_i|\pi(X_i)$ is defined by the matrix of probabilities $\boldsymbol{\theta_i} = (\theta_{ijk})_{1 \leq j \leq q_i, k \in (x_1,...x_r)}$:

$$\theta_{ijk} = P(X_i = k|\pi(X_i) = j) \tag{6}$$

For each X_i, $\boldsymbol{\theta}_{ij} = (\theta_{ijx_1}, \theta_{ijx_2}, ...\theta_{ijx_r})$ is the vector of conditional probabilities defining the distribution of $X_i|\pi(X_i) = j$, where $\Sigma_{k=1}^r \theta_{ijk} = 1$. The joint probability distribution of $(X_1, X_2, ...X_i)$ is given by:

$$P(X_1, X_2, ...X_i) = \prod_i^d P(X_i|\pi(X_i)) \tag{7}$$

3.2.2 Structure Design

To explore the main problem proposed in this paper mentioned above, we first define nine binary variables used in our BN shown in Table 1 (Details are introduced in the following section). We use student data for one semester in this paper. Based on the causal Markov assumption [39], we develop a three-layers raw structure, shown in Figure 2, with two rules: (1), the directed edge can start from one node to another node on the same layer. (2), the directed edge can only point from the lower-layer to the upper level. In other words, we blacklist all outgoing edges from the upper-level nodes to lower level nodes. By construction, no node is allowed to be the parents of gender, and no node is allowed to be the child of academic performance. Such design is consistent with common sense.

3.2.3 Structure Learning

In this research, we use BDeu (Bayesian Dirichlet equivalent uniform) score to evaluate the BN and use the hill-climbing strategy to optimize this score [24]. BDeu score is a variant of the BDeu score, which aims at maximizing the posterior probability of the DAG. We use MLE (maximum likelihood estimation) to fit our BNs [22] in this research.

4. EXPERIMENTS AND RESULTS

4.1 Dataset

The dataset used in this research includes 5,200 students from a Chinese university including freshmen students, sophomores, and juniors and contains a total of about 50 million records of data related to various behavior. For the privacy concern, the students are already pseudonymous in the raw data. Removing the student with incomplete data, we
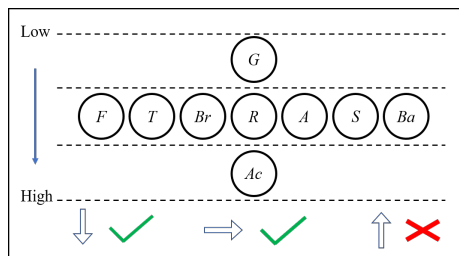

Figure 2: Three-layers raw structure for BN.

Table 2: The number of student in each cluster.

	stay-up	non-stay-up	Total
Junior	664	759	1423
Sophomore	807	773	1580
Freshman	714	532	1246
Total	2185	2064	4249

have 4,249 students left classified into three groups: freshman (1,423 students), sophomore (1,580 students), and junior (1,246 students). This dataset consists of five types of data. The datasets are described as follows:

4.1.1 Internet Access Data

For most students, the campus network is a good choice because of its low price and faster internet speed. The related logs accurately record the student's online information. We use this data to profile students' surfing behavior. The experiment time is from November 2018 to April 2019 including 4,541,247 records data. To ensure the validity of the experimental results, we remove students with fewer related records. In this research, surfing length status T, app preference status A, and sleep status S are calculated based on this data.

4.1.2 Academic Performance Data

Students' academic performance is generally be recorded, which contains the grade, credit of each course for each student. The academic performance data includes 1,048,575 records. In this research, we select the most recent grade data from the period of Internet access data to obtain academic performance data Ac.

4.1.3 Demographic Data

For all schools, students are required to submit personal information at the time of admission, like hometown, gender, and nation. In this research, we take gender G as a feature.

4.1.4 Financial Data

Campus smart card, in most universities, is used as a recognition tool for his identification. Generally, smart cards can be used for any scene of consumption on campus and thus record tons of data for student consumption behavior, such as bathing and eating. Financial data include 22,176,513 records. In this research, breakfast orderliness status Br, bath orderliness status Ba, and financial status F are obtained based on this data.

4.1.5 Reading Data

Generally, every university has its own library that is the main source of book reading by students. Reading data include 17,209,329 records. Reading status R in this research is obtained from this data.

4.2 Student Clustering

Below, the results of fitting a two-component Poisson mixture model are presented and discussed. Models with more components are explored in this experiment, but the two-components model mainly captures the primary modes of student's sleep patterns and more components models tend to split students into further subgroups without any significant additional insight.

Table 1: Variables definition.

| Variable	Definition	Description
G	Gender	Male or Female
R	Reading status	Classify students according to the number of books borrowed
A	App preference status	Define online preferences based on usage time of each app (Game app or Video app)
T	Surfing length status	Classify students according to their total length of surfing time
Br	Breakfast Orderliness status	Classify students according to the number of times they eat breakfast
Ba	Bath Orderliness status	Classify students according to the regularity of bathing
F	Financial status	Classify students based on daily average spending
Ac	Academic performance status	Classify students based on Grade-Point Average
S	Sleep status	Classify students according to whether they stay up late
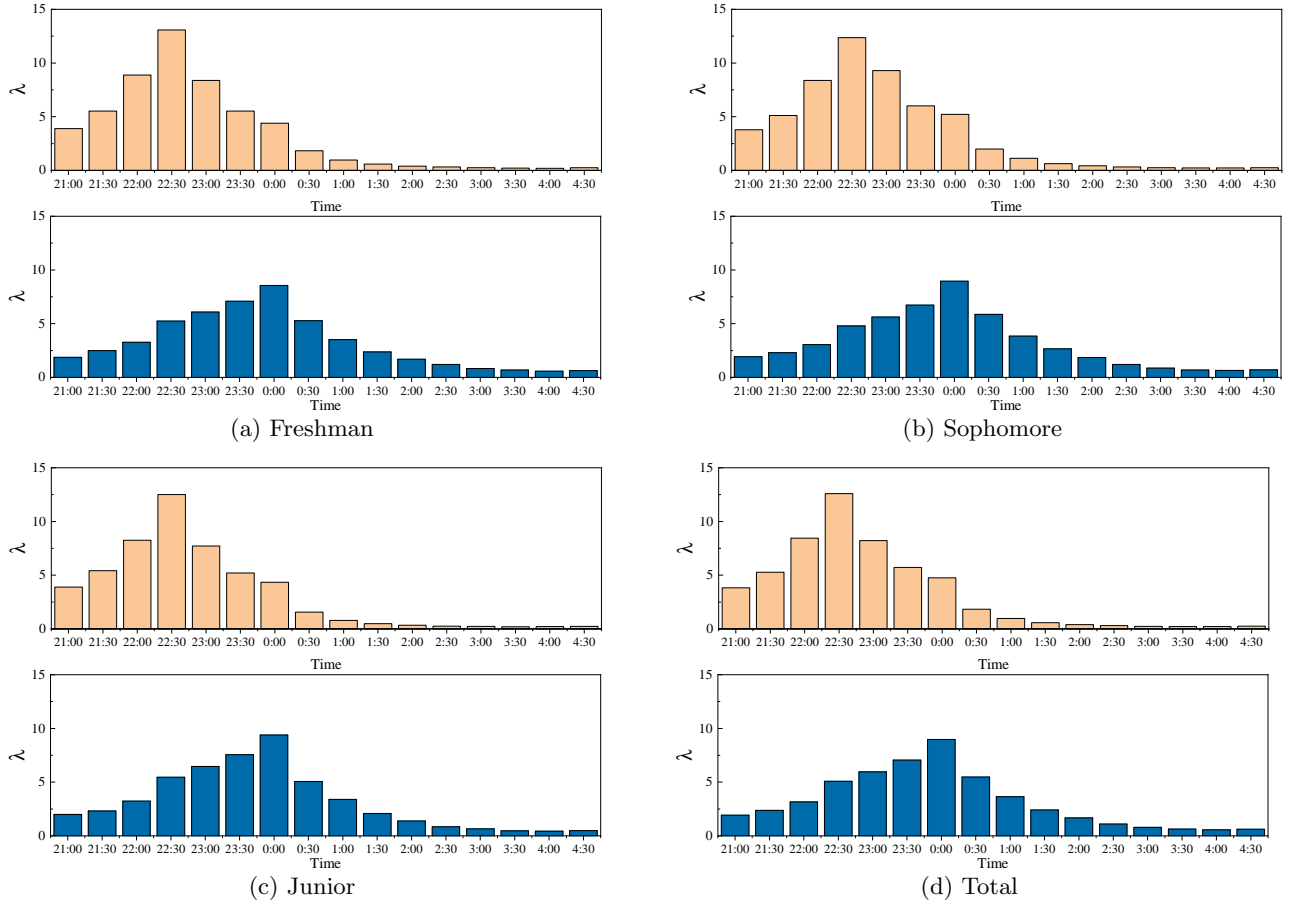
Figure 3: Clustering results of Poisson mixture model with two components.

We set the threshold of component weight at 0.5 to classify student $i = 1, 2, ...N$ into one of the two components, i.e., if $\omega_{i1} \geq 0.5$, then student i is classified into the group of staying up (where $m = 1$ corresponds to the staying up group). We first experiment individually for freshmen, sophomores, and juniors, and then experiment with all students together. Figure 3 shows the λ of each component and Table 2 show the corresponding number of each components. Two distinct sleep patterns can be seen for each group, as the λ of Possion distribution represents the number of times things happen per unit of time. Each pattern is an approximate bell curve. They can be distinguished according to two important characteristics shown as follows:

- First, one of the mixture components has a peak at 10:30 pm (shown as the yellow bar chart) and the other one has a peak at 0:00 am (shown as the blue chart).

- Second, the mixture component with 0:00 am peak has a thicker tail compared with the other one with 10:30 pm peak.

Clearly, these two patterns reflect two different types of students: students who stay up later and students who don't stay up late. According to the regulations of the university where we carry out our experiments, all lights in dormitory areas will be turned off at 10:30 to urge students to sleep early, which is consistent with our experimental results, verifying the effectiveness of our proposed method. Moreover, it can be seen in Table 2 that from the freshmen to the juniors, the proportion of students who do not stay up late gradually increased, eventually surpassing the proportion of students staying up late. The possible reason is that they are willing to live a more regular life because of increased academic pressure and forthcoming job-hunting.

4.3 Student Profiling

In this section, besides sleep status (S) defined above through the mixture probabilistic model, we profile students from other five aspects in eight dimensions, including interest $(R, T$ and $A)$, orderliness $(Br$ and $Ba)$, finance(F), Academic performance (Ac) and Gender (G). As shown in Table 1, we sort students according to specific indicators and classify students with 50% as a threshold, resulting in that all variables are binary. The details are introduced in this section. The interest of students is profiled from three aspects: reading habits R, Interest access time T, and internet preferences A. First, for reading habit$R$, we divide students into two categories: those who like to read and those who don't, based on the number of books borrowed during the experiment. Second, for Interest access time T, we divide students into those who like to surf on the Internet and those who don't, based on their average Interest access time. Finally, for app preference status A, apps in our dataset can be classified into two types: games app and video app according to their functions, and compare the time that users spend on both apps to identify their A.

The orderliness of students is profiled from two aspects: breakfast orderliness Br and bath orderliness Ba. For orderliness profiling, the indicator of Br is the number of taking breakfasts during the experiment and the indicator of Ba is the variance of bath date interval during the experiment. Students are classified into two categories according to the corresponding indicators, with a threshold of 50%.

In addition to interest and orderliness, we also profile students from academic performance, financial situation, and gender perspectives. For academic performance, students are divided into two categories based on their Grade-Point Average (GPA). For the financial situation, we divided students into two categories according to their daily spending during the experiment. For gender, students are divided into males and females.

4.4 Bayesian Network Analysis

4.4.1 Bayesian Network

As mentioned above, our dataset includes three types of students: freshmen, sophomores, and juniors. To capture the rules between different groups, we first build BNs for these three groups separately. In this research, we use a hill-climbing strategy to build BNs through optimizing the BDeu score.

1. First, for each group, we learn 200 BNs through a hill-climbing strategy with restarts strategy, and the results are shown in Figure 4. We learn 200 BNs and 300 BNs, respectively, and find that the score of high-quality BNs in these two cases are equal.

2. Second, we select one-third of the BNs with the high score for the null model to build the consensus network [1]. According to the frequency of edge occurring in these BNs, we design null models with bootstrap sampling to filter the invalid edge (shown in Figure 5) [30]. We determine the threshold by keeping the occurring frequency above the mean value plus two times of standard deviation of the random case and in each group, the value is 25, 27, and 25, respectively.

3. Third, we reconstruct the consensus BNs by those edges that occur more frequently than thresholds.

All BNs are shown in Figure 6. Moreover, we synthesize a total network based on three consensus-BNs according to the occurring frequency of each edge shown in Figure 6(d). For the same edge but different directions, we choose the edge with higher occurrence frequency in high score BNs. For example, in three consensus-BNs, the edge from T to A and the edge from A to T have the same occurrence frequency. Because the edge from A to T has a higher occurrence frequency in BNs with a high score, we keep this edge in total BNs.

4.4.2 Structure Analysis

Table 3: Structure analysis for S in the three BNs.

	G	R	A	T	Br	Ba	F	Ac
Freshman		√	√	√	√		√	√
Sophomore	√		√	√	√			√
Junior			√	√				
Times in $\pi(S)$	1		2	2				
Times in $MB(S)$	1	1	3	3	2		1	2
Times in $ch(S)$					2			2

$\sqrt{}$: The variable is in $MB(S)$.
$\sqrt{}$: The variable is in $\pi(S)$.
$\sqrt{}$: The variable is in $ch(S)$.

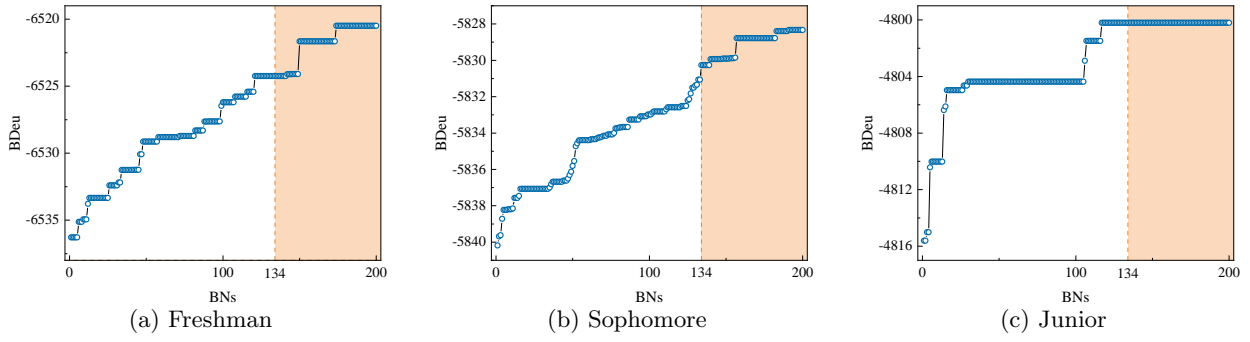(a) Freshman   (b) Sophomore   (c) Junior

Figure 4: Scores of 200 BNs from restart strategy. We select one-third of the BNs with the high score (shown in the yellow area) for the null model and the consensus network.
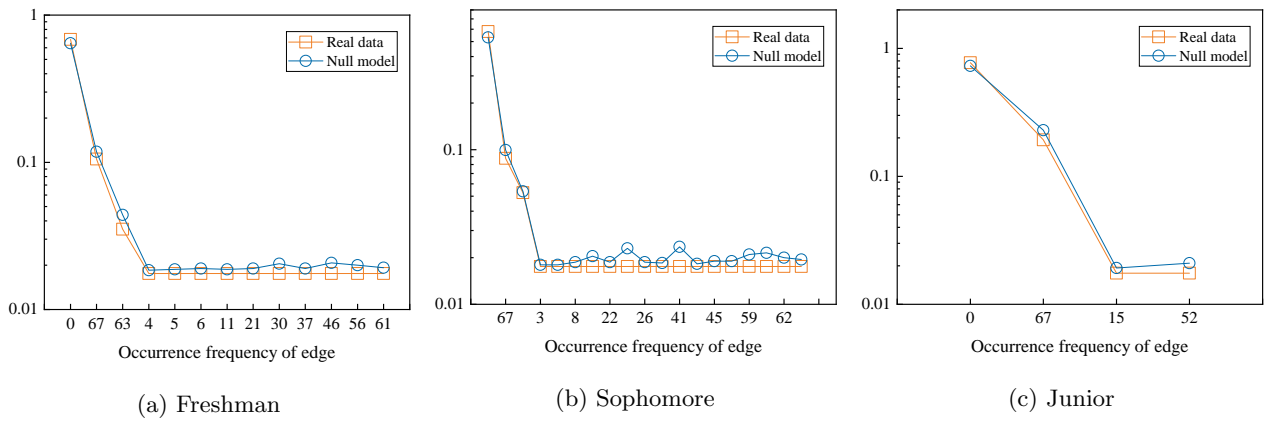

(a) Freshman   (b) Sophomore   (c) Junior

Figure 5: Null model for each group.

(a) Freshman   (b) Sophomore   (c) Junior   (d) Total
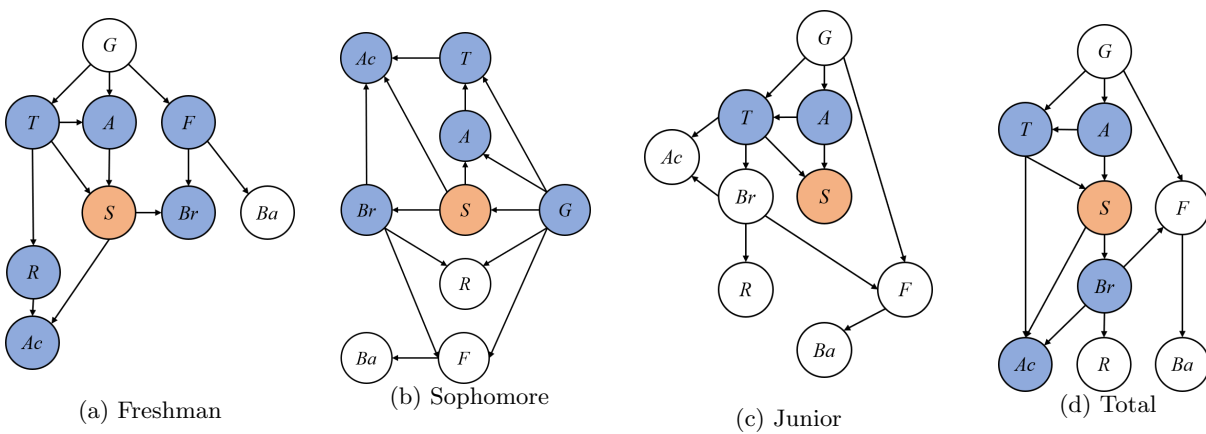
Figure 6: Consensus BN for each group. The orange node is sleep status and its Markov Blanket is represented by blue node.
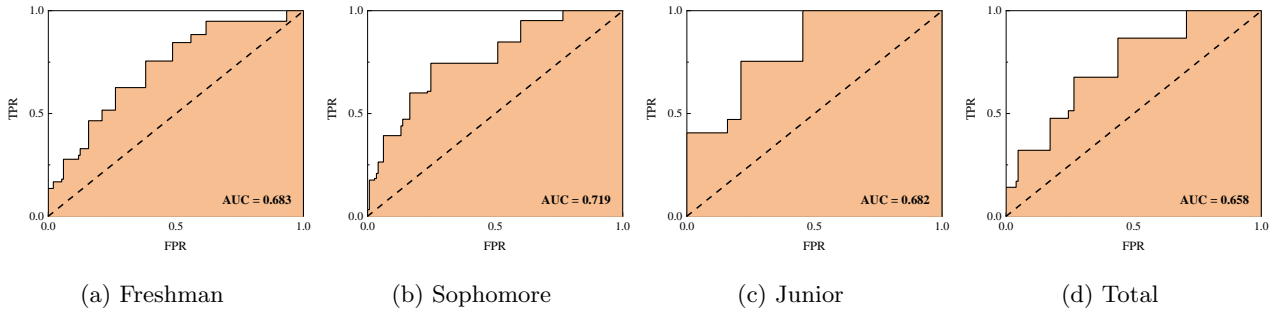
Figure 7: Prediction performance of each BNs. We use ROC to measure the prediction performance. AUC is indicated by yellow and diagonal dotted lines indicate 0.5 in random cases.

The three networks shown in Figure 6 represent three types of students: freshmen, sophomores, and juniors. It is obvious that the behaviors of different groups are different. According to the Markov property of the Bayesian network, the parents node $\pi(X_i)$ are the most related to the output of the targeted node X_i. In this case, we summary those BNs in Figure 6 and show the results in Table 3.

First, both A and T are frequent in $\pi(S)$, representing that a strong connection exists between sleep habit and surfing habit. Specifically, from the perspective of App preference, the probabilities of staying up late for people who love watching videos are 0.75, 0.63, and 0.70, in freshmen, sophomores, and juniors, respectively. For people who love playing games, the probabilities are 0.5, 0.39, and 0.47. Video lovers prefer to stay up late than game fans. Then, we summarize the most popular games among college students, and the top three are Glory of Kings, PUBG (Playerunknown's Battlegrounds), and LOL (League of Legends), which account for more than 80 percent (This result is consistent with the findings of previous studies that people who primarily engage in group play have superior adherence to people who primarily play alone [26]). A common feature of these games is that they are all web-based battle games. Compared to previous single-player games, these games bring all players together to collaborate or compete with each other through an online platform. Players play with real people instead of computers, which is closer to social behavior. In this case, if the user's enemies or teammates are sleeping, then this game naturally loses fun, which explains why game fans don't like staying up late. This phenomenon can also be analyzed from another perspective. According to the theory mentioned in [18], 'spectator experience' and a sense of social presence are very important for players of online games, which can also explain the finding of our experiment. In contrast, watching a video is a behavior that a student can do well on their own, even in the middle of the night. Note that the above results show that people who love to watch videos are more likely to stay up late rather than that videos are to keep people stay up late compared with games, due to the differences in the number of each type of students.

Does the seemingly rational behavior of game lovers contradict the popular view of game addiction? We analyze the relation between app preference A and total surfing length T due to that A appears frequently in the parent node of T. As mentioned before, according to the total length of surfing time, we divide students into long-time internet users and short-time internet users. The probabilities of game fans

surfing the Internet for a long time are 0.57, 0.60, and 0.57 in three groups, which is higher than 0.25, 0.22, and 0.22 for video lovers. This is in line with the current popular view of game addiction. At the same time, the above mentioned can also explain why the short-time internet user prefers to stay up late. (Note that this finding does not conflict with the conclusions in [33] that a longer duration of digital media use was associated with reduced total sleep time and later bedtime because they focus on the digital media two hours before bedtime instead of all day).

Second, for the child nodes of S, the variables of breakfast status Br and academic performance Ac appear frequently according to Table 3. In the matter of having breakfast, the probability of students staying up late is 0.2 lower than normal students, which is consistent with common sense that students who like to stay up late usually get up late to miss breakfast. For the academic performance, students who go to bed early are more likely to achieve a good grade, which is consistent with previous findings, both in biology [27] and cognitive science [4].

4.4.3 Inference

The Markov property of the Bayesian network implies that conditioned on the Markov blanket of a node (shown in Figure 6), the probability distribution of the node is independent of the rest of the network shown as the following equation:

$$X \perp \{U - MB(X) - \{X\}\}|MB(X) \qquad (8)$$

where U is the set of all random variables in BNs and $MB(X)$ represents the Markov blanket of variable X, which is a set of nodes that consists of the parents of the node, the children of the node, and any other parents of the children of that node.

In our research, we design an experiment of predictive inference to estimate the predictability of sleep status. In other words, we want to test whether the sleep state S can be predicted rather than pursue prediction performance. So, we use MLE to fit our datasets and use Area Under The Curve (AUC) to evaluate the prediction performance. For AUC, if the AUC is bigger than 0.5, representing that the prediction result is better than the random model. In other words, the S is predictive. From Figure 7, the AUC value for each group is 0.683, 0.719 and 0.682, which reflects the predictability of sleep pattern.

5. DISCUSSION AND CONCLUSION

In this paper, we design a Possion-based probabilistic mixture model to identify students who are used to stay up late based on the Internet access patterns. We apply this model to a real-world dataset and classify students into two groups: students who are used to stay up late and students who sleep on time. We profile students from five aspects in eight dimensions, including interest (reading status, app preference status, and surfing length status), orderliness (breakfast orderliness status and bath orderliness status), finance (financial status), academic performance (academic performance status) and gender. Then we build Bayesian networks to explore the relationship between these characteristics and sleeping habits and find that surfing habits have a big impact on sleep habits. Finally, we test the predictability of sleeping habits based on campus behavior features.

The assumption of our experiment that students will access the Internet through mobile phones or computers before bedtime is reasonable because the Internet pervades every aspect of our lives, including entertainment, study, social contact, and so on. However, the underlying assumptions made in this study raise a couple of limitations. First, we identify students staying up late based on a hypothesis that students use mobile phones or computers to access the Internet before going to bed. So, it is hard to detect the people who don't have this habit. Second, more data of life details need to be collected for drawing a valid and solid conclusion. There are multiple avenues for future work. First, we intend to expand our dataset and investigate this issue from more aspects. Second, we only check if the sleep status can be predicted rather than pursuing a precise prediction. Next, we plan to design a prediction model with a good performance and integrate this model into the modern educational management system and apply real-time data to detect the sleep status of students.

6. REFERENCES

[1] R. Agrahari, A. Foroushani, T. R. Docking, L. Chang, G. Duns, M. Hudoba, A. Karsan, and H. Zare. Applications of bayesian network models in predicting types of hematological malignancies. *Scientific reports*, 8(1):6951, 2018.

[2] W. Al-Salman, Y. Li, and P. Wen. Detecting sleep spindles in eegs using wavelet fourier analysis and statistical features. *Biomedical Signal Processing and Control*, 48:80–92, 2019.

[3] Z. Alimoradi, C.-Y. Lin, A. Broström, P. H. Búlow, Z. Bajalan, M. D. Griffiths, M. M. Ohayon, and A. H. Pakpour. Internet addiction and sleep problems: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 47:51–61, 2019.

[4] T. Althoff, E. Horvitz, R. W. White, and J. Zeitzer. Harnessing the web for population-scale physiological sensing: A case study of sleep and performance. In *WWW '17 Proceedings of the 26th International Conference on World Wide Web*, pages 113–122, 2017.

[5] K. Bartel, R. Scheeren, and M. Gradisar. Altering adolescents' pre-bedtime phone use to achieve better sleep health. *Health Communication*, 34(4):456–462, 2019.

[6] M. B. Becerra, B. S. Bol, R. Granados, and C. Hassija. Sleepless in school: The role of social determinants of sleep health among college students. *Journal of American College Health*, (1):1–7, 2018.

[7] H. D. Bedru, S. Yu, X. Xiao, D. Zhang, L. Wan, H. Guo, and F. Xia. Big networks: A survey. *Computer Science Review*, 37:100247, 2020.

[8] B. Bjorvatn, J. Mrdalj, I. W. Saxvig, T. Aasnæs, S. Pallesen, and S. Waage. Age and sex differences in bedroom habits and bedroom preferences. *Sleep medicine*, 32:157–161, 2017.

[9] F. C. Brown, W. C. Buboltz Jr, and B. Soper. Relationship of sleep hygiene awareness, sleep hygiene practices, and sleep quality in university students. *Behavioral medicine*, 28(1):33–38, 2002.

[10] W. C. Buboltz Jr, F. Brown, and B. Soper. Sleep habits and patterns of college students: a preliminary study. *Journal of American college health*, 50(3):131–135, 2001.

[11] C. Burr, J. Morley, M. Taddeo, and L. Floridi. Digital psychiatry: Risks and opportunities for public health and wellbeing. *IEEE Transactions on Technology and Society*, 1(1):21–33, 2020.

[12] C.-F. Chien, S.-L. Chen, and Y.-S. Lin. Using bayesian network for fault location on distribution feeder. *IEEE Transactions on Power Delivery*, 17(3):785–793, 2002.

[13] D. Combs, J. L. Goodwin, S. F. Quan, W. J. Morgan, S. Shetty, and S. Parthasarathy. Insomnia, health-related quality of life and health outcomes in children: a seven year longitudinal cohort. *Scientific reports*, 6:27921, 2016.

[14] J. N. Cousins, K. Sasmita, and M. W. L. Chee. Memory encoding is impaired after multiple nights of partial sleep restriction. *Journal of Sleep Research*, 27(1):138–145, 2018.

[15] J. N. Cousins, E. van Rijn, J. L. Ong, K. F. Wong, and M. W. L. Chee. Does splitting sleep improve long-term memory in chronically sleep deprived adolescents? *npj Science of Learning*, 4(1):8, 2019.

[16] F. de Arriba Pérez, J. M. S. Gago, and M. C. Rodríguez. Calculation of sleep indicators in students using smartphones and wearables. In *New Advances in Information Systems and Technologies*, pages 169–178. Springer, 2016.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[18] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore. Alone together? exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 407–416, 2006.

[19] I. N. Fossum, L. T. Nordnes, S. S. Storemark, B. Bjorvatn, and S. Pallesen. The association between use of electronic media in bed before going to sleep and insomnia symptoms, daytime sleepiness, morningness, and chronotype. *Behavioral sleep medicine*, 12(5):343–357, 2014.

[20] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[21] J. Grønli, I. K. Byrkjedal, B. Bjorvatn, Ø. Nødtvedt, B. Hamre, and S. Pallesen. Reading from an ipad or from a book in bed: the impact on human sleep. a randomized controlled crossover trial. *Sleep medicine*, 21:86–92, 2016.

[22] D. Grossman and P. Domingos. Learning bayesian network classifiers by maximizing conditional likelihood. *Twenty-first international conference on Machine learning - ICML '04*, page 46, 01 2004.

[23] T. Guo, F. Xia, S. Zhen, X. Bai, D. Zhang, Z. Liu, and J. Tang. Graduate employment prediction with bias. In *Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press, 2020.

[24] D. Heckerman, D. Geiger, and D. M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.

[25] F. V. Jensen et al. *An introduction to Bayesian networks*, volume 210. UCL press London, 1996.

[26] M. D. Kaos, R. E. Rhodes, P. Hämäläinen, and T. N. Graham. Social play in an exergame: how the need to belong predicts adherence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[27] P. Kelley, S. W. Lockley, R. G. Foster, and J. Kelley. Synchronizing education to adolescent biology:'let teens sleep, start school later'. *Learning, Media and Technology*, 40(2):210–226, 2015.

[28] Y.-M. Liu, L. Wang, H.-C. Chu, and C. Yang. Development of a mobile sleep-management system for improving students' lifestyles based on a self-regulated learning strategy. In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 623–628. IEEE, 2017.

[29] J. C. Lo, J. L. Ong, R. L. Leong, J. J. Gooley, and M. W. Chee. Cognitive performance, sleepiness, and mood in partially sleep deprived adolescents: The need for sleep study. *Sleep*, 39(3):687–698, 2016.

[30] M. J. McGeachie, H.-H. Chang, and S. T. Weiss. Cgbayesnets: Conditional gaussian bayesian network learning and inference with mixed discrete and continuous data. *PLOS Computational Biology*, 10(6):e1003676, 2014.

[31] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.

[32] K. Okano, J. R. Kaczmarzyk, N. Dave, J. D. Gabrieli, and J. C. Grossman. Sleep quality, duration, and consistency are associated with better academic performance in college students. *NPJ science of learning*, 4(1):1–5, 2019.

[33] K. M. Orzech, M. A. Grandner, B. M. Roane, and M. A. Carskadon. Digital media use in the 2 h before bedtime is associated with sleep variables in university students. *Computers in human behavior*, 55:43–50, 2016.

[34] J. Park, R. Yu, F. Rodriguez, R. Baker, P. Smyth, and M. Warschauer. Understanding student procrastination via mixture models. *International Educational Data Mining Society*, 2018.

[35] M. E. Patrick, J. Griffin, E. D. Huntley, and J. L. Maggs. Energy drinks and binge drinking predict college students sleep quantity, quality, and tiredness. *Behavioral sleep medicine*, 16(1):92–105, 2018.

[36] D. Peters, K. Vold, D. Robinson, and R. A. Calvo. Responsible ai-two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, 1(1):34–47, 2020.

[37] A. C. Schneider, D. Zhang, and Q. Xiao. Adolescent sleep characteristics and body-mass index in the family life, activity, sun, health, and eating (flashe) study. *Scientific Reports*, 10(1):1–10, 2020.

[38] E. B. Simon and M. P. Walker. Sleep loss causes social withdrawal and loneliness. *Nature communications*, 9(1):1–9, 2018.

[39] P. Spirtes, C. Glymour, and R. Scheines. *Causation, prediction, and search*. 1993.

[40] R. Tavernier and T. Willoughby. Sleep problems: predictor or outcome of media use among emerging adults at university? *Journal of Sleep Research*, 23(4):389–396, 2014.

[41] R. Tavernier and T. Willoughby. A longitudinal examination of the bidirectional association between sleep problems and social ties at university: The mediating role of emotion regulation. *Journal of youth and adolescence*, 44(2):317–330, 2015.

[42] L.-L. Tsai and S.-P. Li. Sleep patterns in college students: Gender and grade differences. *Journal of psychosomatic research*, 56(2):231–237, 2004.

[43] J. Vroon, C. Zaga, D. Davison, J. Kolkmeier, and J. Linssen. Snoozle–a robotic pillow that helps you go to sleep: Hri 2017 student design competition. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 399–400. ACM, 2017.

[44] H. Zhao, W. Gui, H. Huang, Y. Liu, H. Ding, W. Fan, S. Huang, W. Yang, X. Wang, and G. Chen. Association of long-term sleep habits and hypertension: a cross-sectional study in chinese adults. *Journal of Human Hypertension*, pages 1–10, 2019.

[45] X. Zhao, J. Li, W. Liu, J. Zhang, and Y. Li. Design of the sleeping aid system based on face recognition. *Ad Hoc Networks*, 99:102070, 2020.