Predicting Mood in College Students: Developing a Predictive Model From Multivariate Time Series

Predicting Mood in college students:

Developing a predictive model from multivariate time series

A thesis presented

by

Marianne Aguilar

To

Applied Mathematics

in partial fulfillment of the honors requirement

for the degree of

Bachelor of Arts

Harvard College

Cambridge, Massachusetts

March 29, 2019

Abstract

Sleep diaries often collect useful information regarding students' sleep duration, timing, moods, and relevant daytime activities. The abundance of data provided by these multivariate time series provide a basis by which to carry out predictions for end-of-month results. In particular, end-of-month moods are interesting to predict since they can be indicators of larger health problems, such as depression or anxiety. This paper attempts to model the clusters students fall into based on sleep variables and the time-dependent network that contributes to end-of-month mood ratings in an attempt to find important variables on certain days to target for treatment. It concludes by finding that dependent on the cluster a student falls into, wake time, first event timing, or biological determinants are most important in predicting 28th day moods.

Acknowledgements: This paper could not have been written without the help of a number of people. First and foremost, I'd like to thank Melissa St. Hilaire and Elizabeth Klerman for their help throughout this process. I'm very grateful to Melissa for her constant advice, dedication to helping me learn, and time spent meeting with me or helping me find answers and to Elizabeth for introducing me to the analytical modeling of sleep, welcoming me to do research, and always offering help in any capacity. I'd also like to thank the many friends who offered me emotional support throughout this process. Finally, I'd like to offer special thanks to my parents and sisters for their love and support throughout college. The process of writing this paper was a joy because of the long FaceTime calls with my mom, heartfelt advice from my dad, and hours-long Star Wars speculation and silly conversations with Celi and Clara that interspersed my research.

Contents

1 Introduction

As college students leave the confines of their homes to have more freedom in their schedules and lifestyles, they represent a niche group by which to understand the mechanisms of sleep and its effects, particularly on mental health. College students have been observed to have short sleep duration (only 29.4% obtain the recommended 8 or more hours of sleep per night), low sleep quality (only 34.1% have Pittsburgh Sleep Quality Index [PSQI] levels in the good range) and erratic sleep schedules (20% of college students pull all-nighters at least once a month)[1,2]. In contrast, 64% of high school students receive more than 8 hours of sleep per night and fewer than 20% have sleep onset latency longer than 30 minutes[3]. For college students, this reduction in sleep quantity and quality (PSQI scores greater than 8, as compared to students with lower PSQI scores) produces significantly higher levels of anger, confusion, depression, tension, distress, and exhaustion; higher rates of physical illness, with 3 times as many poor-quality sleepers missing class due to illness; and lower GPA, with a 10-point decrease in Sleep Regularity Index predicting a 0.1 decrease in GPA[1,2]. Moreover, sleep quality and psychiatric symptoms have a significant interaction: sleep duration is significantly related to depressive symptoms. In addition, low sleep quality can worsen alcohol consequences (i.e. risky behaviors such as driving while intoxicated, unplanned sexual activity, and waking up in unexpected places) and contrarily sleep quality can be affected by alcohol consumption when a mental health problem is present[4]. Most importantly, low sleep quality and short sleep duration are known to have an effect on psychiatric symptoms: approximately 40% of individuals with insomnia and 46% of those with hypersomnia have comorbid psychological disorders as compared to 16% of healthy individuals[5]. Based upon this information, the purpose of this paper is to see if it is possible to predict college students' morning mood based on sleep variables collected over time.

In the absence of psychiatric screenings, the best guess for the mental health of a patient is their mental state[7]. Mental state, or mood, is easily accessed via questionnaires. Mathematical models have been developed to simulate the effects of light input on the circadian pacemaker and to project mental state based on lifestyle choices such as sleep duration[7,8]. However, no model has been developed that predicts baseline mood based on other sleep inputs, such as sleep consistency, sleep timing, and sleep quality, and other factors that affect sleep, such as alcohol and caffeine consumption, timing of activities, academic workload, and exercise duration. With the abundance of information collected in daily diaries, it is possible to build a mathematical model for college students' moods based on these other variables. While other models generate predictions for the population (i.e. group-average models), it is hypothesized that baseline mood and sleep phenotype differs across individuals, such that a person who has a certain chronotype or has to wake up early due to activities would be classified differently than another individual. In addition, the unique circumstances of college life suggest that separate models are needed to predict mood in college students. The closest existing model to predict mood is that proposed by Tuarob et al. While it would be an interesting proposition to use this model on sleep diaries collected from college students, their model is designed for prediction from lifestyle choices and daytime activities, such as alcohol intake or weather, and only has one question pertaining to sleep (sleep quality). In addition, while their model, which uses a machine learning algorithm is well suited for large data sets that include missing data points, it cannot be used on time series data, which limits its predictive value since mood, sleep, and activities have time dependence[7]. Similarly, other models have been introduced which classify behavior rather than mood from sleep. For example, Cohen et al. showed that individuals with autism who have sleep deficiencies are more likely to be classified as "low-functioning" individuals (i.e., individuals with behavioral problems). In addition, Sano et al. uses support vector machines that predicted which college students would fall into the top or bottom 20% of happiness scores based on sleep characteristics[9,10]. However, both of these analyses predicted mood categories rather than self-reported values. Moreover, neither Tuarob et al. nor Cohen et al. used multivariate time series as their data input, which means that this data requires different treatment due to its underlying nature, and while Sano et al. does use multivariate time series as input, they ignore sleep and mood phenotypes and treat every person as input to the same classification model[7,9,10]. Thus, there is still a need to model the acute effects of sleep variables over time on mood to better understand the research on sleep and mental health. Nonetheless, these initial models provide a framework from which to start since a machine learning algorithm, although not necessary a regression algorithm, could be well adapted to the noise in our data because it will update parameters based on information from several individuals to yield a usable average.

Thus, we will attempt to build upon the foundations set by previous model attempts to incorporate the variety and amount of information available from diaries while limiting the effects of noise and inter-patient variability. Rather than create a one-size-fits-all neural network model for all individuals that has little accuracy or separate models for each individual that overfits the data, we instead create a model that captures common patterns that exist in multiple clusters of individuals derived from the overall population. Thus, while every individual will provide unique data and patterns, the belief is that once aggregated, individuals will have characteristics in their data that are common among their cluster and can be used to predict other individuals that belong to the same cluster. We propose to develop these models using clustered neural networks and to use the output to generate a graphical model of inter-variable interactions that contribute to prediction of baseline mood. Specifically, this thesis explores appropriate ways to model noisy, multivariate time series over independent samples by testing different clustering methods, testing two different inputs (spectrum-based or standardized) to neural network models for this kind of data to find the most appropriate input, and finally to utilize knowledge of underlying biological, particularly neurological, and psychological relationships to find appropriate underlying models for each kind of data input.

2 Appropriate clustering for multivariate time series

2.1 Data Structure

The data used to build this model has a unique structure that influenced the way this model could be built and the complexity of the question at hand. Two-hundred thirty-eight undergraduate students participated in a 30-day experiment providing 7,202 days of data. Data were collected starting in the Fall of 2013 and lasting until Spring of 2016. In order, start dates included October 28-30 2013, February 12-15 2014, October 3-9 2014, February 12-19 2015, October 6-15 2015, and February 9-11 2016. Thus, students did not start and finish the study on the same day or even on the same day of the week. On each day, participants completed surveys in the morning and evening about academic, extracurricular, and exercise activities, sleep variables, caffeine/drug intake, and self-reported health, mood, alertness, tiredness, and stress. In total, 88 quantitative and categorical variables were collected, along with extraneous comments or IDs to keep data classified. Mood was measured using a visual non-numeric 0-100 scale on five categories (1: 0=sleepy to 100=awake, 2: 0=sad to 100=happy, 3: 0=sluggish to 100=energetic, 4: 0=sick to 100=healthy, 5: 0=stressed to 100=calm). Some of the variables were redundant (e.g., the information these variables encoded included information encoded by another variable) and some variables were added to encode the moving averages of sleep, mood, and caffeine/drug intake. Ultimately, 45 quantitative or categorical variables per day were included in the analysis.

While it may seem unreliable to create predictions based off of sleep diary entries, there is evidence that sleep diary entries are particularly reliable at capturing the true values of sleep variables when collected over a sufficient time period[27]. Specifically, for studies conducted on adolescents, interclass correlation coefficients indicated that five weekday nights of sleep diary entries can give reliable estimates of bedtime, sleep onset latency, and sleep duration, where reliability is defined as the state when a group of days in chronological order have high correlation with subsequent groups of days[27]. In addition, one week of data yielded adequate reliability to estimate bedtime, sleep onset latency, and wake time but not sleep duration or wake time in a sleep diary sample collected from Qatar and the United States[27]. With two weeks of data and 10 non-workdays, it is possible to have excellent reliability (i.e. intraclass correlation coefficient > 0.75) from sleep diaries according to a subsample in the United States[27]. Thus, these studies suggest that 28 days of diary data in our study will have enough values to have an accurate profile of a student's sleep habits.

2.2 Individual clustered heat maps

Using these data, the goal is to create a model that uses 45 variables per day (41 quantitative and 4 categorical) over 27 days to predict mood on the 28th day to determine what combination of variables from a student's prior month history are best able to predict mood at that moment. The first step in this process was to find the Pearson's correlation coefficients across a sample of 41 quantitative variables with the individual values across all days for each student using the *corr* function in Python's Pandas Dataframe. We standardized each variable by subtracting their monthly mean on that variable from the daily value and dividing that by their standard deviation in a month for that variable. Using the *corr* function we found correlations of 0.4 or lower between most variables. These correlations are shown in Figure 1. As expected, the moods, moving average of moods and moving average of sleep variables are highly correlated with the daily mood and sleep variables used to create them. Most other variables have low direct correlation. However, there is a slightly higher correlation between academic work and exercise with each of the 5 moods.

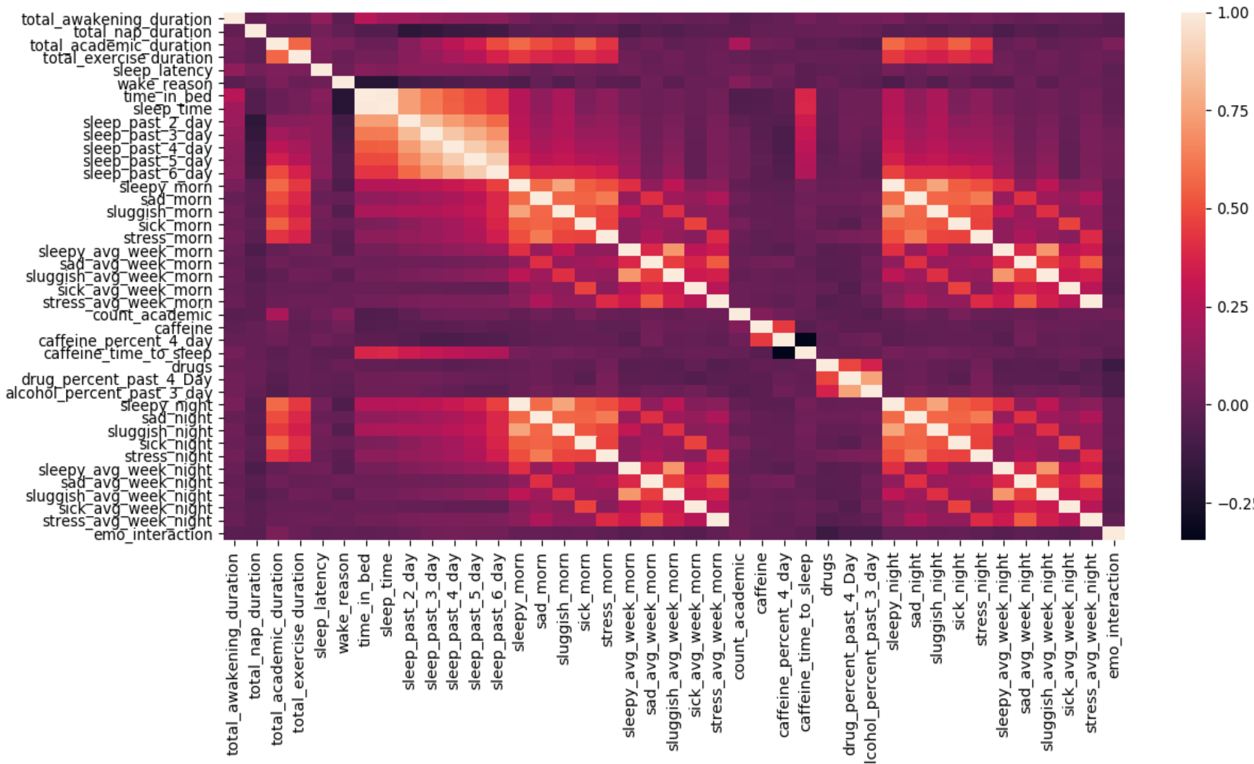

Figure 1: Correlations between 41 quantitative variables for all people

The next step was to understand if students clustered a certain way on all 45 quantitative or categorical variables, which would help clarify whether one model could fit all individuals or if different models are needed for different student phenotypes to maximize accuracy. It is expected that students will cluster differently due to the presence of chronotypes, i.e., an intrinsic sleep-wake pattern, and there is precedent set to cluster students by chronotype to gain a better understanding of what predicts disturbed sleep among different subsections[1]. Chronotypes are an especially important factor when individuals have free schedules, i.e., when they are allowed to wake and sleep when they desire. Chronotypes are highly heritable but can be masked by set schedules. Many of the college students in our study reported a reason that they needed to wake up at a certain time[24]. Nonetheless, even with the presence of activities that force students to wake up at a certain time, we believe that students in our cohort will cluster according to a combination of their chronotype and their reason for waking up[24]. Individual clustered heat maps revealed that several dimensions showed clusters dependent on variability. In particular, for the five mood scales, both those completed in the morning and those completed at night, there were four discernible clusters: a high mood-low variability group, a high mood-high variability group, a low mood-low variability group, and a low mood-high variability group. Examples are shown in Figures 2-6 for each of the 5 mood scales. We can distinguish each of the apparent clusters in the following way: the high mood-low variability group has the same shade of red, orange, or yellow across a row, the high mood-high variability group has a mix of red, orange, and yellow across a row, the low mood-low variability group has the same shade of purple or black across a row, and the low mood-high variability group has a mix of red, purple, and black across a row.

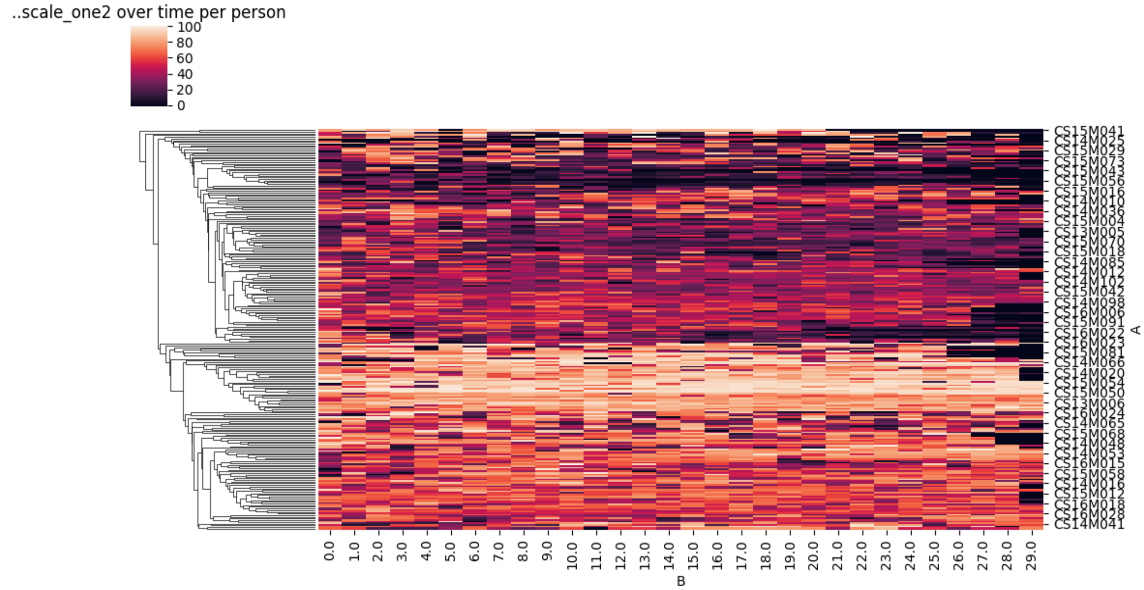

Figure 2: Clustered individuals by sleepiness

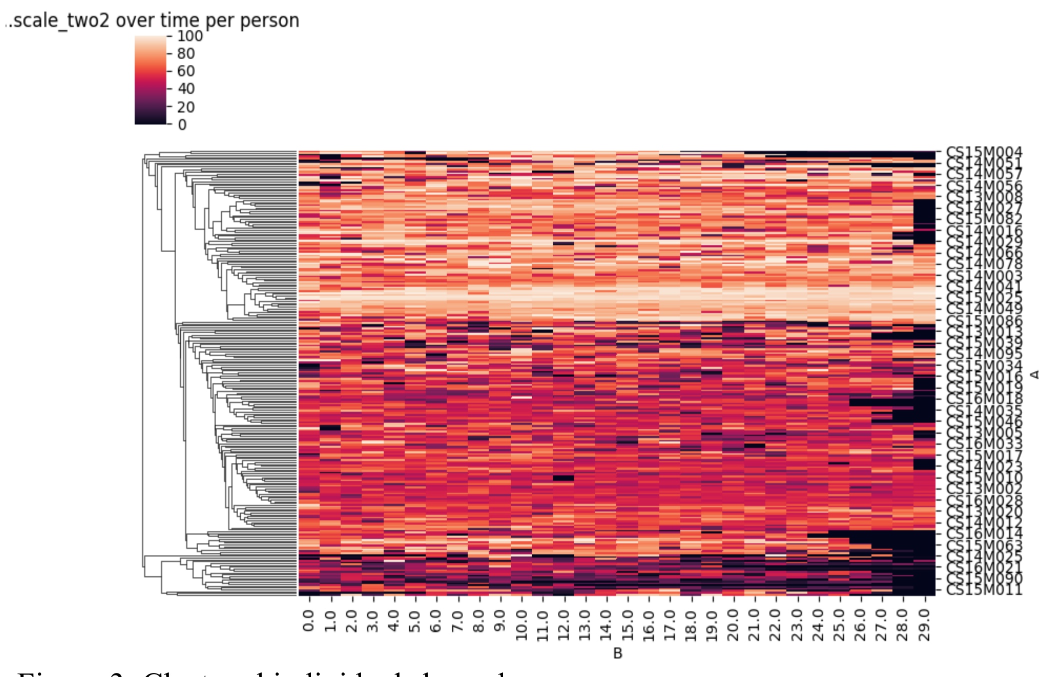

Figure 3: Clustered individuals by sadness

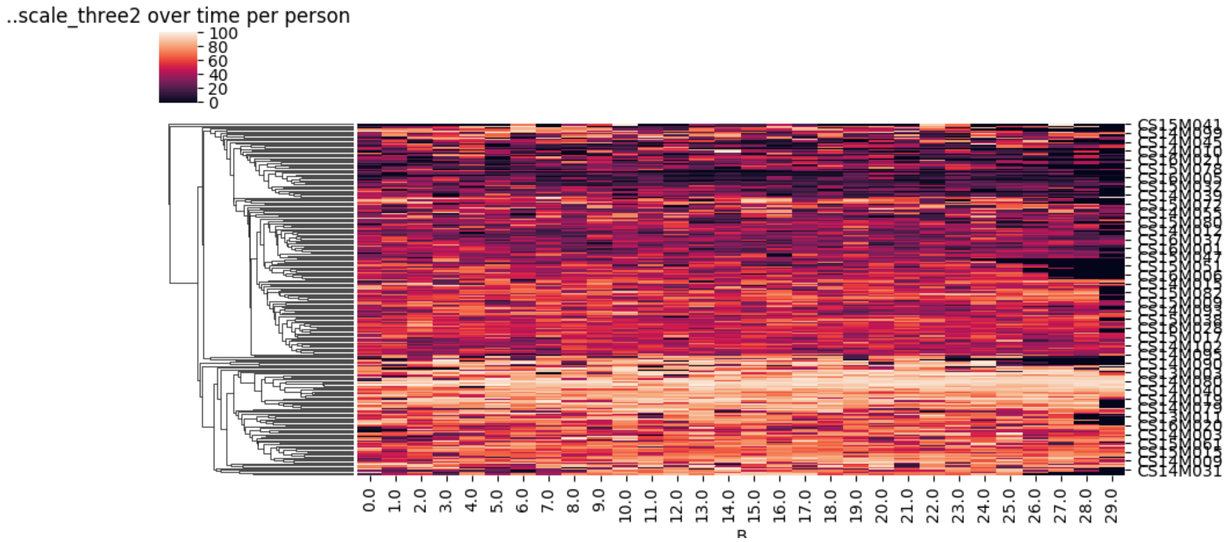

Figure 4: Clustered individuals by sluggishness

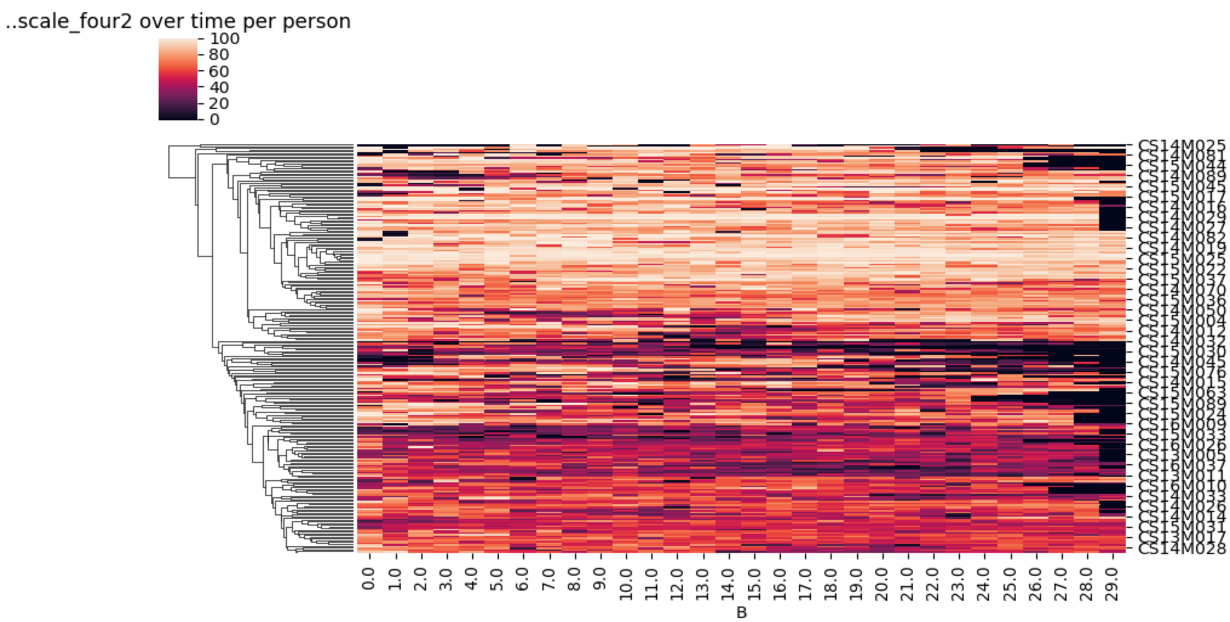

Figure 5: Clustered individuals by health

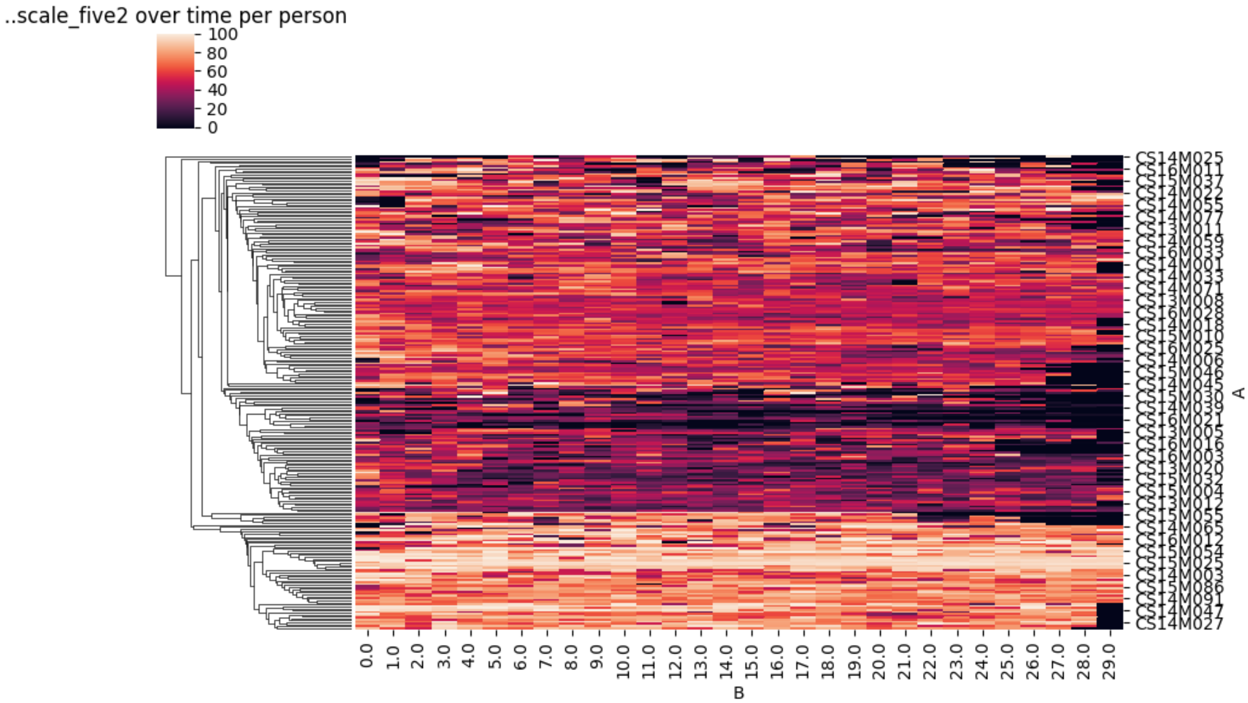

Figure 6: Clustered individuals by stress

This finding prompted an interest in exploring a more comprehensive clustering mechanism that could capture the distance between time series while taking into account these 45 dimensions. Currently, clustering the individual time series creates a disparate picture of how individuals themselves - who are understood by the composite of all the measured variables over time - group, since on one measure an individual could be grouped with individuals that he or she is not grouped with on another measure. Thus, an individual did not belong to one set group because not all the variables are accounted for at the same time using this clustering method. Once individuals are placed appropriately in clusters, however, a model could be built for each cluster to account for the variability in individuals' time series data while still attempting to build a larger model that can predict across different individuals.

2.3 Spectrum-based quasi-distances

Having done a preliminary visual scan of the clusters that arise on the mood dimensions for these time series, the next step was to use a quantitative method that could combine the multiple dimensions being measured over time to yield a better understanding of the number of clusters and the placement of each student into a cluster. The time series in our dataset do not all start on the same day across students. Furthermore, the data were collected across different seasons. Therefore, it may be counterintuitive to compare values on the same day, despite the fact that such a comparison takes into account the possibility that starting a new routine of filling out sleep diaries could in some way affect sleep patterns. Thus, the first attempt to cluster these time series accurately was to use a quasi-distance[11].

Ravishanker et al. propose that when attempting to find a distance metric between multivariate time series to then cluster based on those distances, one should use a quasi-distance based on a likelihood ratio test and the periodograms of the time series on each dimension. The motivating idea is that the spectra comprising a time series are more informative than the original time series itself, and since the periodogram itself is an estimate of the spectral density, it can serve to help create one unifying distance measure between two 45-dimensional time series. An example of this initial periodogram for a frequency is shown below in Equation 1.

$$\text{Equation 1: } I_{x,jk}(\omega_l) = \frac{1}{2\pi T}\left(\sum_{t=1}^{T} x_{tj} e^{-itw_l}\right)\left(\sum_{t=1}^{T} x_{tk} e^{itw_l}\right)$$

Starting from the smoothed periodograms, shown below of each time series for each student, each frequency is represented by an n x m matrix where diagonal entries are estimates of the direct spectrum for that frequency and the ij-th entry is an estimate of the cross spectrum of the i-th and j-th component for that frequency.

Equation 2: $I_x^*(\omega_l) = \frac{\sum_{|k| \leq m} I_x(\omega_{l+k})}{2m+1}$ where m is a pre-determined integer.

For each frequency, the time series are then compared using a likelihood ratio test rather than Euclidean distance. This likelihood ratio for each frequency uses the null hypothesis that the two n x m spectral matrices are equal and tests that hypothesis using the ratio of the determinant of the n x m matrices over the sum of the two. These are then use to make the quasi-distance per frequency.

Equation 3: $H_o : \lambda_x(\omega) = \lambda_y(\omega)$

Equation 4:

$$L\left(\lambda_x(\omega), \lambda_y(\omega)\right) = \frac{|I_x^*(\omega_l) I_y^*(\omega_l)|^{2m+1-p} |\lambda_x(\omega)\lambda(\omega)|^{-(2m+1)} e^{-tr(\lambda_x^{-1}(\omega) I_x^*(\omega) + \lambda_y^{-1}(\omega) I_y^*(\omega))}}{(\Gamma_p(2m+1))^2}$$

Equation 5: $Q(\omega) = \frac{L_0(\tilde{\lambda}_x(w)\tilde{\lambda}_y(w))}{L_0(\tilde{\lambda}_x(w)) L_0(\tilde{\lambda}_y(w))} = 2^{2p(2m+1)} \frac{|I_x^*(\omega_l)|^{2m+1} |I_y^*(\omega_l)|^{2m+1}}{|I_x^*(\omega_l) + I_y^*(\omega_l)|^{4m+2}}$

These measures over all the frequencies sampled are combined to form a quasi-distance as their average is taken to give one measure that gives the average deviance based on the null hypothesis that two multivariate time series are equal. It can then be used to perform hierarchical clustering.

Equation 6: $Q^* = \overline{Q(w)}$

However, for our samples, this measure did not work well, often resulting in non-existent values or 0 due to the order of magnitude of the computed values from our signals. This could be due to smaller values in our periodograms than in the sample data used in the Ravishanker paper. Regardless, even when logarithmic values were used, the values did not yield a reasonable clustering output. This led to the conclusion that perhaps rather than having issues with the periodograms themselves, the problem arises from the use of hierarchical clustering based on a likelihood ratio test. If there is enough noise in the data, then the null hypothesis that two individuals have distinct spectral frequencies may fail to be rejected, such that it is optimal when carrying out a hierarchical clustering to cluster all students together because the difference in quasi-distance between different pairs is so small. The Ravishanker paper motivated the idea that while the quasi-distance may not be the way to approach this multivariate time series, basing clustering on periodograms to capture the frequencies rather than exact values may be one way to proceed with clustering.

2.4 Feature agglomeration

The next step in attempting to cluster the data was to use a feature agglomeration algorithm instead of the quasi-distances to cluster students who could be seen as belonging to one feature group based on their time series. Using the same method as that suggested for clustering 8 x 8 pixel photos, this algorithm works similarly to Principle Component Analysis (PCA) but is adapted for our purposes by making it consider groups of students' time series as similar features rather than to yield features throughout time series. Thus, because inputs were vectors of day-variable pairings in this case, to consider an individual as the same feature as another individual is most analogous to considering them to exhibit the same phenotype, rather than to try to cluster different types of phenotypes in one student's time series. To see if clusters were intuitive, an analysis was conducted to determine the variables that defined each cluster. Without mood input or wake times, clusters mostly exist due to sleep times, and when there are only two clusters the only significant difference is in average sleep over the past five days. Significance was tested by using a two-sample t-test to see whether the mean of the averages of a group's data differed for each quantitative measure of interest.

Specifically, we began by inputting the variables in Table 1 over 27 days, which amount to 1215 inputs for each day-variable combination per student.

Table 1: 45 Day-variable inputs	
Total night awakenings duration	$\alpha_1,\ldots,\alpha_{27}$
Total nap duration	$\beta_1, \ldots,\beta_{27}$
Total academic work duration	$\psi_1, \ldots,\psi_{27}$
Total exercise duration	$\delta_1,\ldots,\delta_{27}$
Bedtime	$\varepsilon_1,\ldots,\varepsilon_{27}$
Sleep latency	$\varphi_1,\ldots,\varphi_{27}$
Reason for waking	$\gamma_1, \ldots,\gamma_{27}$
Wake time	$\eta_1,\ldots,\eta_{27}$
Total time in bed	$\iota_1,\ldots,\iota_{27}$
Total sleep duration	$\xi_1, \ldots,\xi_{27}$
Average sleep in the past 2 days	$\kappa_1, \ldots,\kappa_{27}$
Average sleep in the past 3 days	$\lambda_1,\ldots,\lambda_{27}$
Average sleep in the past 4 days	$\mu_1,\ldots,\mu_{27}$
Average sleep in the past 5 days	$\nu_1,\ldots,\nu_{27}$
Average sleep in the past 6 days	$o_1,\ldots,o_{27}$
First event time	$\pi_1,\ldots,\pi_{27}$
Sleepiness rating in the morning	$\rho_1,\ldots,\rho_{27}$

Health rating in the morning	$\theta_1, \dots, \theta_{27}$
Stress rating in the morning	$\omega_1, \dots, \omega_{27}$
Average sleepiness in the past 6 mornings	$\varsigma_1, \dots, \varsigma_{27}$
Average sadness in the past 6 mornings	$\chi_1, \dots, \chi_{27}$
Average sluggishness in the past 6 mornings	$\upsilon_1, \dots, \upsilon_{27}$
Average health in the past 6 mornings	$\zeta_1, \dots, \zeta_{27}$
Average stress in the past 6 mornings	$A_1, \dots, A_{27}$
Number of academic activities that day	$B_1, \dots, B_{27}$
Whether caffeine was taken	$\Psi_1, \dots, \Psi_{27}$
Percent of past 4 days in which caffeine was taken	$\Delta_1, \dots, \Delta_{27}$
Time between previous night's bedtime and previous day's caffeine intake	$E_1, \dots, E_{27}$
Clock time of caffeine intake	$\Phi_1, \dots, \Phi_{27}$
Whether drugs/alcohol were taken	$\Gamma_1, \dots, \Gamma_{27}$
Percent of the past 4 days that drugs were taken	$H_1, \dots, H_{27}$
Percent of the past 3 days that alcohol was taken	$I_1, \dots, I_{27}$
Sleepiness rating in the evening	$\Xi_1, \dots, \Xi_{27}$
Sadness rating in the evening	$\Lambda_1, \dots, \Lambda_{27}$
Sluggishness rating in the evening	$M_1, \dots, M_{27}$
Health rating in the evening	$N_1, \dots, N_{27}$

Stress rating in the evening	$O_1, \dots, O_{27}$
Average sleepiness in the past 6 evenings	$\Pi_1, \dots, \Pi_{27}$
Average sadness in the past 6 evenings	$P_1, \dots, P_{27}$
Average sluggishness in the past 6 evenings	$\Sigma_1, \dots, \Sigma_{27}$
Average health in the past 6 evenings	$T_1, \dots, T_{27}$
Average stress in the past 6 evenings	$\Theta_1, \dots, \Theta_{27}$
Type of emotional interactions in the day	$\Omega_1, \dots, \Omega_{27}$

For the periodogram trial with 4 clusters, we found that groups were significantly different on reason for waking, which we considered interesting but not informative with regards to sleep since this could divide students into groups of people who wake naturally and those who often have morning responsibilities. Taking out reason for waking, we found that clusters would form along the lines of total amount of academic work, total sleep time, and average sleep time over the past 5 days. By eventually taking out mood variables as well, the differences between clusters occurred due to time in bed, total amount of academic work, and average sleep over the past 5 days. Clustering along school work, sleep time or time in bed, and average sleep over the past 5 or 6 days held across 3 and 4 clusters and whether standardized data or periodograms were used. This was not the case for 2 clusters when either standardized or periodogram data was used, indicating that perhaps it is better to divide data into more clusters to be able to cluster according to sleep behavior. The reason is that only allowing for 2 clusters can divide individuals along categorical data, such as wake reason or caffeine users, as was found in this case when a higher cutoff of 0.2 was used for significance to find how periodogram data divided into 2 groups. With no significant difference between groups when two clusters were used, we took this as an indication that if one wishes to create models for students who fall under certain sleep profiles, then it is best to use 3-4 clusters based on sleep parameters rather than mood level and variability as described in section 2.1. This is due to the fact that clustering reveals that groups arise from sleep data rather than from mood data.

3 Relationships between variables

3.1 Background for modeling Scale 2: Sadness

To build the model, one can begin to incorporate the biology underlying the systems. The most obvious relationship to begin examining is that of sleep and sadness. While no causal relationship has ever been established, sleep has a u-shaped relationship with depression, where too little sleep (anywhere less than 6 hours) is linked to a higher risk of depression, too much sleep (anywhere greater than 10 hours) is linked to depression as well due to excessive sleep being a side effect of depression, and there exists an optimal zone of sleep where risk of depression is at its lowest[12]. Of course, different individuals have different optimal sleep minima that could yield the lowest risk of depression based on their sleep phenotype, but this model attempts to address that by the clustering based on chronotype before building the model. However, as a general class, a starting point to model the relationship between sleep and sadness is to use a concave, u-shaped function.

With other variables, such as caffeine intake or exercise, the relationship to sadness, stress, or other moods is not as clear. However, these variables often mediate or affect sleep, which is often reflected in their association to sleep. The first of these relationships is that of caffeine intake relative to sleep time. Interestingly, caffeine is part of a feedback loop with sleep where caffeine doses and timing initially disrupts sleep, leading to greater sleepiness the following day, which in turn often leads to greater caffeine consumption[13]. This feedback loop is particularly interesting to model as a variable contributing to sleep that can be worked in to the u-shaped relationship between sleep duration and sadness[12]. In addition to an existing feedback loop, the timing of caffeine matters as well, with timing of caffeine about 3 hours before bedtime producing the maximum number and length of awakenings through the night and lowest sleep quality[14].

Exercise provides another interesting attribute to investigate since it affects both the sleep cycle itself and mood through the release of endorphins, providing two avenues by which to affect sadness in a model[15,16,17,18,19]. Exercise's effect on sadness has been shown to help release endorphins and is linked to higher emotional affect and an increase in the amount and vigor of exercise decreases emotional sensitivity to sadness[15,16]. This means that when an individual would normally respond strongly to a negative emotional event, exercise helps regulate that response to decrease the effect, perhaps by having an inverse relationship with previous evening moods[16]. In addition to that, exercise also affects sleep in a bidirectional relationship[17]. Specifically, circadian phase delay has a linear relationship with timing of exercise, which means that exercise could have a linear relationship with sleep onset, which has a linear relationship with sleep duration that should be considered[19].

Thirdly, napping has a significant relationship with depression when napping occurs in high doses, but can also produce happier mood after a short dose. Napping has a u-shaped relationship with sleep quality, a measure we try to capture in number of awakenings[20]. This relationship seems to arise from a combination of a negative linear relationship with sleep during the week and a positive linear relationship with sleep during the weekend, such that too few naps can be indicative of poor sleep quality that exists throughout the day and too many naps can be indicative of poor sleep quality that is addressed through sleep periods in the day[20].

While the data used in the previously mentioned studies often refer to the relationship between raw/standardized data and mood, these relationships can easily be extended to the relationship between periodograms and mood. This is because the frequency-based estimate yields an estimate of the spectral matrix, which means it yields an estimate for the spectrum of frequencies that make up a signal. This is informative since an individual whose sleep is characterized by a high frequency for both sleep duration and sadness, for example, could perhaps be best estimated to have a value "x" if past individuals with that composition of frequencies had values close to "x" on the day we estimate for. Thus, since we know about the u-shaped relationship between sleep duration and sadness, an individual whose sleep duration and moods have fluctuated very often in the past could be best estimated with an average of mood values. This could better capture the effect of the day of the week on estimates because an individual could have one frequency better represented than another due to the presence of data from an extra Friday at the end of the month rather than an extra Sunday or Monday in the beginning of the month. It could also better capture an overall estimate of a person's mood since the frequencies give a profile of an individual on more measures as opposed to standardized data. However, the use of data from the previous days may still be needed for an accurate estimate since it could combine information about the frequencies and whether we are falling on the upper or lower part of a signal. Thus, we included both spectral frequencies and some previous days' standardized data to obtain what we hope is the best estimate for the 28th day mood.

3.2 Background for modeling Scale 1 and Scale 4: Sleepiness and Health

In the same stream of logic, the relationships between our variables of interest and perceived health can provide information on the relationship between these other mood scales and past standardized or spectral variables.

Sleepiness goes up as sleep duration and quality decrease with an approximately equal rate of change among individuals who are heavy drinkers, with a slightly higher increase when alcohol consequences are added on to sleepiness and late bedtime[23]. While this applies to a subsection of the general population who drinks heavily, Lund et al.'s study of college students reveals that as PSQI scores increase, sleep duration decreases and we see a 28% increase in feelings of fatigue once individuals leave the optimal PSQI zone (PSQI<6) and enter the borderline PSQI zone (PSQI=6-7) and an even greater change in the magnitude of fatigue ratings when one goes from borderline to the poor PSQI zone (PSQI>7)[1]. This finding suggests a relationship between sleep quantity variables that increases the rate of change in sleepiness rating as sleep quantity decreases.

For the heavy drinker population, perceived health has the lowest mean score when respondents have a combination of sleepiness, late bedtime, sleep disturbances, and alcohol consequences than when they only have sleepiness, late bedtime, and consequences, sleepiness and late bedtime, or only sleepiness[23]. Moreover, the change in mean perceived health in the DeMartini study that occurs when adding the variable sleep disturbances, which is a sign of poor sleep quality, to the combination of sleepiness, late bedtime, and alcohol consequences is a 14% change and about 6x greater than when late bedtime is added to sleepiness and about 36x greater than when consequences are added to the combination of sleepiness and late bedtime[23]. In a similar matter, Lund et al. studied the percentage of students who skipped class in the past two months due to illness, which is a reflection of the perceived health scale measured in our sample since it takes a low perceived health score to skip class, irrespective of actual health. They found that the percentage of students who skip class due to low perceived health in the borderline PSQI zone is 12.5% higher than those who have optimal sleep quality and quantity and 37.5% higher for those in the poor PSQI zone[1]. As in the sleepiness case, this finding suggests a relationship between sleep quality and quantity variables in which the rate of change in sleepiness rating increases as sleep quality and quantity decrease.

3.3 Background for bidirectional relationships

As mentioned in section 3.1, there are a number of feedback loops that involve three variables, such as the feedback loop between caffeine, sleep, and mood. In addition, there are a number of more complicated relationships among some of the variables contributing to 28th day baseline mood. For example, given a person with an underlying psychiatric problem, which in the case of depression could be reflected in low energy, happiness, and calmness in the previous month, poor sleep quality may be associated with more alcohol-related consequences, specifically when both a psychiatric problem and low sleep duration and quality are present[4]. Alcohol consequences in turn have effects on future days, and therefore the alcohol measure we include in our analysis is not easily related to sleep or mood on its own[4]. Moreover, psychiatric problems in general have a bidirectional relationship with sleep disturbances, which means their effects can compound each other over time[5]. In addition, another interesting relationship that can have an indirect effect on mood is that between day of the week and sleep. Specifically, the typical Monday through Friday college course schedule can often restrict sleep opportunity. However, weekends are usually free from scheduled events, and often yield different sleep duration and timing, which means that chronotype and other biological determinants for sleep are often exposed under these conditions[24]. For college students, however, weekends are often also associated with greater alcohol intake, which relates back to the previous discussion of complex interactions among these variables.

3.4 Background for behavioral changes

In addition to this, one must ask the question of how to extend sleep duration or improve sleep quality if that is the desired goal. Improving sleep quality has been previously discussed, as caffeine intake taken further from one's bedtime, napping and its u-shaped relationship with sleep quality, and alcohol intake can all be regulated. However, sleep quantity is something that can also be addressed via bedtimes and wake times. As students undergo puberty, many of them have their ideal bedtime shift to a later time, such that moving bedtime earlier is an option but may not prove particularly effective[13]. However, extending wake time, which would also mean changing the timing of the first event of the day - which for many students is school start time, consistently results in longer sleep times and is associated with motivation, declined depressive mood reports, fewer health center visits, and fewer reports of exhaustion likely as a result of this extended sleep time[13].

To extend upon this idea, there are a number of relationships for which our diaries do not have direct measurements but that can have bearing upon our models nonetheless. An important one concerns the relationship between smartphone use, sleep quality, and depression or anxiety. For instance, while we do not collect the time spent on a digital media device before sleep, it has been shown that computer work is associated with less total sleep time, less overall time in bed, and a later bedtime while surfing the internet is associated with disrupted sleep[26]. This is compounded by the fact that high levels of smartphone use are significantly related to higher levels of depression[25]. Therefore, students' variety and prevalence of digital media device use before bedtime could affect sleep and indirectly affect mood.

4 Finding the best method to build a model for multivariate time series

4.1 Feed-forward neural networks

A number of options are available to build models for time series data that can predict based on past observations and other variables, including auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA) models, which use moving averages to generate predictions. One of these models could have been implemented to predict 28th day day-variable values, including mood, based on a cluster's previous 27 days of day-variable inputs from our dataset. However, rather than create a model that tracks previous inputs and the means of the population separately, we hoped to acknowledge the ways in which day-variables contribute to mood together since variables are often related to each other over time. To still allow for the moving average to capture the fluctuations in data over time, we calculated each day's average sleep duration over 2, 3, 4, 5, and 6 days, average weekly morning sleepiness, sadness, sluggishness, health, and stress, average weekly evening sleepiness, sadness, sluggishness, health, and stress, average number of days in which caffeine was taken over 4 days, and average number of days in which drugs were taken over 3 days. We decided to use the neural network as the basis for our model since through its summation of inputs via weights, it can account for the degree to which variables occur together as one predicts 28th day mood.

We began using the neural network regressive model in the following way:

1. Begin by inputting the variables in Table 1.
2. Sum together each variable in different combinations to each of "n" hidden neurons.
3. At each hidden neuron, transform using a standard function (linear, logistic sigmoid or hyperbolic tangent).
4. If applicable, sum the output of each hidden neuron in different combinations to the next set of hidden neurons and transform once more using a standard function (linear, logistic sigmoid or hyperbolic tangent).
5. Sum final layer of hidden neurons with different weights to yield a numeric output.

6. Compare estimate to actual value to adjust weights for next iteration and adjust weights in each level to obtain a greater degree of accuracy.

Python's Sklearn package includes a neural network component that will train the model to yield weights that minimize error. Specifically, we use the Adam optimizer to minimize the mean squared error[29]. This optimizer does so using a stochastic gradient descent, which uses the gradient of the objective function to approach a minimal point by going according to the negation of the gradient and uses stochasticity by randomizing the point chosen to update the parameters being estimated[29]. Adam takes this a step further by changing the learning rate of the parameters using the first and second moments of the gradient and dividing the learning rate by these for recent gradients[29]. This method works particularly well for artificial neural networks like the ones being used in this problem[29].

Before determining the correct number of hidden layers, it was first necessary to conduct some sensitivity analysis to determine the correct number of neurons for the hidden layers. Comparing the effect of using 5, 10, or 100 hidden neurons, we get the results in Figure 7.

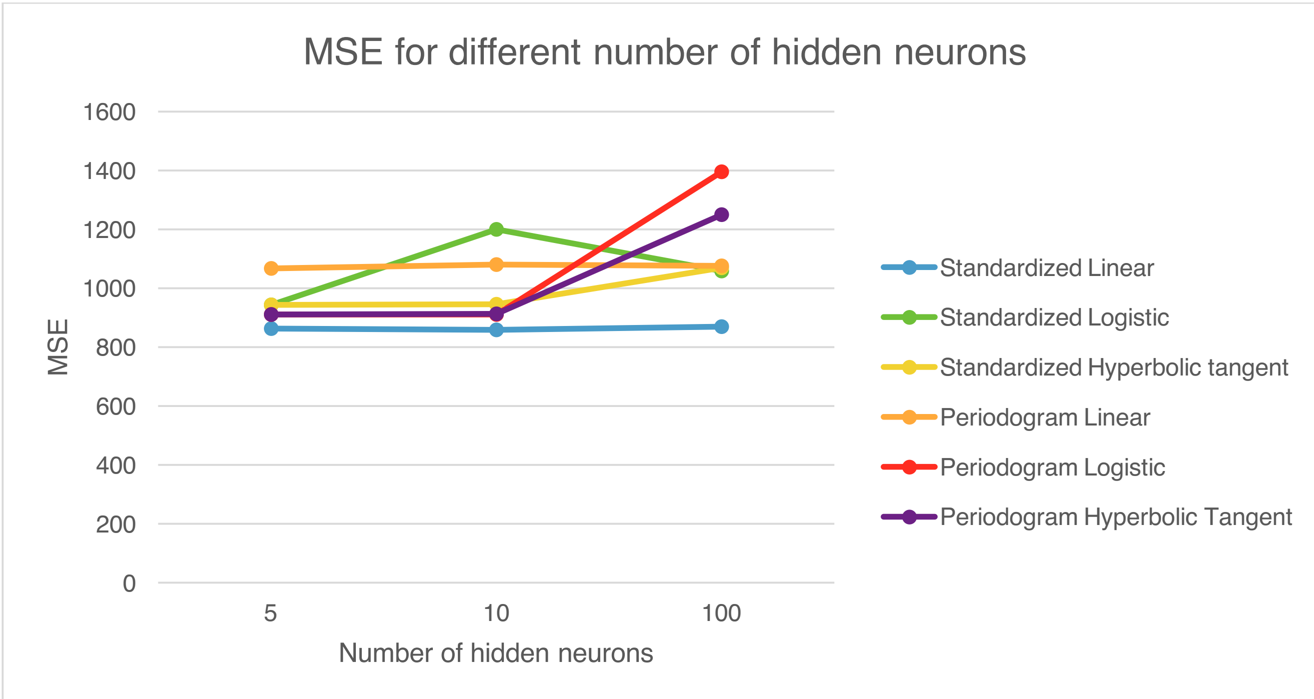

Figure 7: MSE across number of hidden neurons for two hidden layers and 4 clusters

This seems to indicate that across the different inputs and activation functions, it is generally the case that 10 hidden neurons yields the best results. This holds across 2 or 3 clusters as well, so we decided to use 10 hidden neurons.

4.2 Proposed form and output for neural networks with 1 hidden layer

In the basic form proposed above, a neural network could be used for any sort of blind prediction and tuned with parameters until the best prediction is found. However, given the information known about sleep and its relationship to mood and daytime activities, we propose hyperbolic tangent functions can be combined to approximate the intended shape. This idea is guided by a deeper understanding of how neural networks work with regards to nonlinear, continuous data as shown by Muller et al., where they propose that nonlinear functions can be approximated through a combination of localized functions, and by the proof written by Cybenko where he states that a summation of sigmoidal functions can be discriminatory and with the right parameters approximate a continuous function[21,22]. Assuming that there is some underlying continuous function in the space $\mathbb{R}^{1215}$ that predicts 28th day morning mood based on 27 entries over time on 45 scales, this idea could guide the logic that some combination of functions must be used to approximate a complicated function of this kind.

Muller et al. give the following example with a one-dimensional variable and linear function f(x). The function $g(x) = f(x) - f(x - c)$ creates a bump in the graph at $x = \frac{c}{2}$ and this bump can be moved by transforming the function to the form shown in Equation 7[21]. These bumps are meant to approximate the shape of the underlying graph in n-dimensional space by representing the minima and maxima that define the graph.

$$Equation\ 7: g(x) = \alpha[f(\alpha x - c_1) - f(\alpha x - c_2)].$$

This already creates one bump, which can be an approximation to a function such as $y = x^2$. To add more than one bump, it would be necessary to add another hidden neuron. To mirror the actions of a neural network, one sums the two function to obtain a function of the form in Equation 8.

$$\text{Equation 8: } g_1(x) + g_2(y) = \alpha[f(\alpha x - c_1) - f(\alpha x - c_2) + f(\alpha y - c_3) - f(\alpha y - c_4)]$$

Where $g_1(x)$ and $g_2(y)$ intersect, one reaches the highest value, such that if the activation threshold is chosen between α and the maximum in this continuous space, all other parts are suppressed with the exception of the maximal bump, which can be summarized in a two-dimensional form in Equation 9.

$$\text{Equation 9: } g(x, y) = f(\alpha[f(\alpha x - c_1) - f(\alpha x - c_2) + f(\alpha y - c_3) - f(\alpha y - c_4) - 1.5])$$

Cybenko complicates this further by stating that for some function f(x), there is some sum G(x) that is the best approximation to f(x) in continuous space[22]. By then proving the sigmoidal functions are discriminatory, he shows these functions are the best approximation of continuous space when summed to form G(x)[22].

Based off this intuition, we propose that the noisy, multivariate time series we hope to use for prediction of mood at the end of the month can be adapted to fit the forms proposed by past data through the sum of activation functions in hidden layers. We begin by proposing the following form for one hidden layer with 10 hidden neurons as shown in Equation 10.

Equation 10:

$$g\left(\begin{matrix} \alpha_1,\ldots,\alpha_{27},\beta_1,\ldots,\beta_{27},\psi_1,\ldots,\psi_{27},\delta_1,\ldots,\delta_{27},\varepsilon_1,\ldots,\varepsilon_{27},\varphi_1,\ldots,\varphi_{27},\gamma_1,\ldots,\gamma_{27},\eta_1,\ldots,\eta_{27},\iota_1,\ldots,\iota_{27}, \\ \xi_1,\ldots,\xi_{27},\kappa_1,\ldots,\kappa_{27},\lambda_1,\ldots,\lambda_{27},\mu_1,\ldots,\mu_{27},\nu_1,\ldots,\nu_{27},o_1,\ldots,o_{27},\pi_1,\ldots,\pi_{27},\rho_1,\ldots,\rho_{27}, \\ \sigma_1,\ldots,\sigma_{27},\tau_1,\ldots,\tau_{27},\theta_1,\ldots,\theta_{27},\omega_1,\ldots,\omega_{27},\varsigma_1,\ldots,\varsigma_{27},A_1,\ldots,A_{27},B_1,\ldots,B_{27},\Psi_1,\ldots,\Psi_{27}, \\ \Delta_1,\ldots,\Delta_{27},E_1,\ldots,E_{27},\Phi_1,\ldots,\Phi_{27},\Gamma_1,\ldots,\Gamma_{27},H_1,\ldots,H_{27},I_1,\ldots,I_{27},\Xi_1,\ldots,\Xi_{27},\Lambda_1,\ldots,\Lambda_{27}, \\ M_1,\ldots,M_{27},N_1,\ldots,N_{27},O_1,\ldots,O_{27},\Pi_1,\ldots,\Pi_{27},P_1,\ldots,P_{27},\Sigma_1,\ldots,\Sigma_{27},T_1,\ldots,T_{27},\Theta_1,\ldots,\Theta_{27}, \\ \Omega_1,\ldots,\Omega_{27} \end{matrix}\right) =$$

$$\sum c_i \tanh\left(\sum \begin{matrix} \alpha_1,\ldots,\alpha_{27},\beta_1,\ldots,\beta_{27},\psi_1,\ldots,\psi_{27},\delta_1,\ldots,\delta_{27},\varepsilon_1,\ldots,\varepsilon_{27},\varphi_1,\ldots,\varphi_{27},\gamma_1,\ldots,\gamma_{27},\eta_1,\ldots,\eta_{27},\iota_1,\ldots,\iota_{27}, \\ \xi_1,\ldots,\xi_{27},\kappa_1,\ldots,\kappa_{27},\lambda_1,\ldots,\lambda_{27},\mu_1,\ldots,\mu_{27},\nu_1,\ldots,\nu_{27},o_1,\ldots,o_{27},\pi_1,\ldots,\pi_{27},\rho_1,\ldots,\rho_{27}, \\ \sigma_1,\ldots,\sigma_{27},\tau_1,\ldots,\tau_{27},\theta_1,\ldots,\theta_{27},\omega_1,\ldots,\omega_{27},\varsigma_1,\ldots,\varsigma_{27},A_1,\ldots,A_{27},B_1,\ldots,B_{27},\Psi_1,\ldots,\Psi_{27}, \\ \Delta_1,\ldots,\Delta_{27},E_1,\ldots,E_{27},\Phi_1,\ldots,\Phi_{27},\Gamma_1,\ldots,\Gamma_{27},H_1,\ldots,H_{27},I_1,\ldots,I_{27},\Xi_1,\ldots,\Xi_{27},\Lambda_1,\ldots,\Lambda_{27}, \\ M_1,\ldots,M_{27},N_1,\ldots,N_{27},O_1,\ldots,O_{27},\Pi_1,\ldots,\Pi_{27},P_1,\ldots,P_{27},\Sigma_1,\ldots,\Sigma_{27},T_1,\ldots,T_{27},\Theta_1,\ldots,\Theta_{27}, \\ \Omega_1,\ldots,\Omega_{27} \end{matrix}\right)$$

where entries are found in Table 1.

Since hyperbolic tangent is not limited to outputting in the range of 0 to 1, this form should capture both the maxima and minima around specific input vectors that discriminates from other vectors using 1215 cutoffs created by neural networks for certain inputs around a "bump". Moreover, the sigmoidal function is still captured in the hyperbolic tangent, so it should sum to be closer to a nonlinear, continuous function than a linear activation function. Finally, the sigmoidal, and therefore hyperbolic tangent, function can likely capture the feedback loops often found between inputs into this network (as discussed in sections 3.1 and 3.2) since inputs are often either transformed to a value of 1 or a value of -1, mirroring the additive nature of a positive feedback loop and detractor nature of a negative feedback loop.

We began by running the model with the first variation for the multivariate time series, which takes periodogram data as well as recent data as inputs. While we hypothesized the form shown above in Equation 10, we tested with the linear, hyperbolic tangent, and logistic activation functions to see how their performance compared. The individuals were clustered using Feature Agglomeration into 2, 3, or 4 sleep and academic work phenotype groups with MSE values as shown below in Figure 8. The same comparisons run with only the standardized time series as inputs are also shown below in Figure 8. Across the different data inputs, activation functions, and number of clusters we can see that the test of the six models with 1 hidden layer is to use a larger number of clusters, standardized data, and the linear function.

Figure 8: MSE for 2, 3, and 4 clusters made by periodogram and standardized data with different activation functions and one hidden layer

4.3 Proposed form and output for neural networks with 2 hidden layers

While it was informative to run the model with one hidden layer, Muller et al. point out that a greater number of hidden layers can result in a greater number of bumps in higher dimensions since multiple maxima-producing functions are now combined several times. Thus, with this intuition, we imagine that two hidden layers will give a more accurate prediction since it can capture the many feedback loops more accurately, which are likely maxima or minima themselves in n-dimensional space.

The comparisons carried out in 3.4 were repeated once more with 2 hidden layers with an activation equation shown in Equation 11.

Equation 11:

$$\sum C_j \sum c_i \tanh \left(\sum \begin{array}{c} \alpha_1, \ldots, \alpha_{27}, \beta_1, \ldots, \beta_{27}, \psi_1, \ldots, \psi_{27}, \delta_1, \ldots, \delta_{27}, \varepsilon_1, \ldots, \varepsilon_{27}, \varphi_1, \ldots, \varphi_{27}, \gamma_1, \ldots, \gamma_{27}, \eta_1, \ldots, \eta_{27}, \iota_1, \ldots, \iota_{27}, \\ \xi_1, \ldots, \xi_{27}, \kappa_1, \ldots, \kappa_{27}, \lambda_1, \ldots, \lambda_{27}, \mu_1, \ldots, \mu_{27}, \nu_1, \ldots, \nu_{27}, o_1, \ldots, o_{27}, \pi_1, \ldots, \pi_{27}, \rho_1, \ldots, \rho_{27}, \\ \sigma_1, \ldots, \sigma_{27}, \tau_1, \ldots, \tau_{27}, \theta_1, \ldots, \theta_{27}, \omega_1, \ldots, \omega_{27}, \varsigma_1, \ldots, \varsigma_{27}, A_1, \ldots, A_{27}, B_1, \ldots, B_{27}, \Psi_1, \ldots, \Psi_{27}, \\ \Delta_1, \ldots, \Delta_{27}, E_1, \ldots, E_{27}, \Phi_1, \ldots, \Phi_{27}, \Gamma_1, \ldots, \Gamma_{27}, H_1, \ldots, H_{27}, I_1, \ldots, I_{27}, \Xi_1, \ldots, \Xi_{27}, \Lambda_1, \ldots, \Lambda_{27}, \\ M_1, \ldots, M_{27}, N_1, \ldots, N_{27}, O_1, \ldots, O_{27}, \Pi_1, \ldots, \Pi_{27}, P_1, \ldots, P_{27}, \Sigma_1, \ldots, \Sigma_{27}, T_1, \ldots, T_{27}, \Theta_1, \ldots, \Theta_{27}, \\ \Omega_1, \ldots, \Omega_{27} \end{array} \right)$$

Figure 9 shows how MSE changes when two hidden layers, each with 10 neurons, are used for standardized and periodogram data inputs. Over all the tested models, standardized linear still gives the lowest MSE and that improvement is not drastic even when the number of hidden layers is increased.

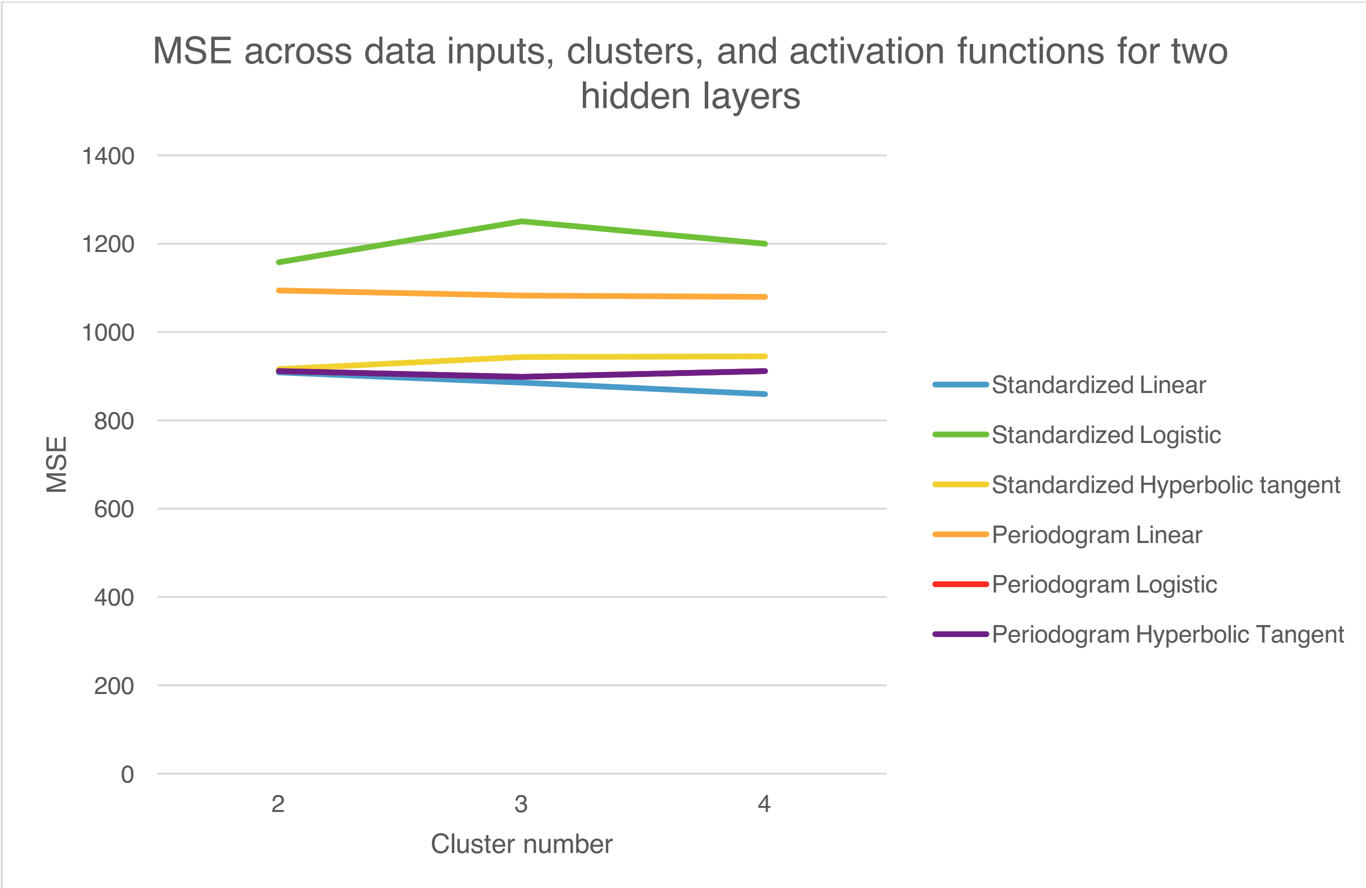

Figure 9: MSE for 2, 3, and 4 clusters made by periodogram and standardized data with different activation functions and two hidden layers

4.4 Proposed form and output for neural networks with 3 hidden layers

With logic similar to that proposed in 3.5, we assumed that three hidden layers would be most accurate in capturing the relationships underlying the data that would then be used to predict 28th day mood. The comparisons carried out in 3.4 were repeated once more. With 3 hidden layers, the greater complexity should be able to capture the nonlinear interactions in the data set more accurately. This is shown in Equation 12.

Equation 12:

$$\sum D_k \sum C_j \sum c_i \tanh \left(\sum \begin{matrix} \alpha_1, \ldots, \alpha_{27}, \beta_1, \ldots, \beta_{27}, \psi_1, \ldots, \psi_{27}, \delta_1, \ldots, \delta_{27}, \varepsilon_1, \ldots, \varepsilon_{27}, \varphi_1, \ldots, \varphi_{27}, \gamma_1, \ldots, \gamma_{27}, \eta_1, \ldots, \eta_{27}, \iota_1, \ldots, \iota_{27}, \\ \xi_1, \ldots, \xi_{27}, \kappa_1, \ldots, \kappa_{27}, \lambda_1, \ldots, \lambda_{27}, \mu_1, \ldots, \mu_{27}, \nu_1, \ldots, \nu_{27}, o_1, \ldots, o_{27}, \pi_1, \ldots, \pi_{27}, \rho_1, \ldots, \rho_{27}, \\ \sigma_1, \ldots, \sigma_{27}, \tau_1, \ldots, \tau_{27}, \theta_1, \ldots, \theta_{27}, \omega_1, \ldots, \omega_{27}, \varsigma_1, \ldots, \varsigma_{27}, A_1, \ldots, A_{27}, B_1, \ldots, B_{27}, \Psi_1, \ldots, \Psi_{27}, \\ \Delta_1, \ldots, \Delta_{27}, E_1, \ldots, E_{27}, \Phi_1, \ldots, \Phi_{27}, \Gamma_1, \ldots, \Gamma_{27}, H_1, \ldots, H_{27}, I_1, \ldots, I_{27}, \Xi_1, \ldots, \Xi_{27}, \Lambda_1, \ldots, \Lambda_{27}, \\ M_1, \ldots, M_{27}, N_1, \ldots, N_{27}, O_1, \ldots, O_{27}, \Pi_1, \ldots, \Pi_{27}, P_1, \ldots, P_{27}, \Sigma_1, \ldots, \Sigma_{27}, T_1, \ldots, T_{27}, \Theta_1, \ldots, \Theta_{27}, \\ \Omega_1, \ldots, \Omega_{27} \end{matrix} \right)$$

For the periodogram data, MSE behaves in a similar way to previously seen models in that nonlinear activation functions are best at predictions. However, for standardized inputs, the linear activation function still works best, albeit with higher MSE scores than with fewer hidden layers. Overall, there is evidence that the standardized linear form still gives the lowest MSE and that three hidden layers do not improve predictions for this dataset.

4.5 Network visualization of the model

Because it is difficult to capture the number of relationships between the variables, we propose an alternative to understanding the influence of sleep variables on mood by visualizing the network contributing to 28th day mood. Rather than simply use correlations, which were often weak due to comparisons that only capture the pairwise relationship between two variables, we propose using the contributions to hidden layers on each day to capture both the time series nature of the data as well as the number of variables that are all contributing to predict 28th day mood. Thus, the network to be imagined is one that shows the relationship between variables that have nonzero contributions to 28th day mood to understand how changing one of these variables would affect the whole network and in turn affect the value predicted on the 28th day. We propose that for the population of college students, there is some 1215x10 weight matrix that predicts the 28th day mood with the same entries as extracted from the diary with lowest possible MSE. For this population weight matrix, we are interested in conducting the calculations below to create networks that capture the relationship between variables summarized over 27 days that contribute to 28th day mood. We decided not to use the 1215x10 weight matrix itself to understand which day-variable entries are most important because while weights are mostly non-zero, they are miniscule due to the fact that day-variables from earlier days often affect later days, such that an optimal way to assign weights is to weigh earlier days more heavily and allow those entries to later propagate throughout the network in smaller values. This was confirmed by finding that the variables with the highest sum of the squared sum of weights over 10 layers are the first variables that feed into the network. Thus, it makes more sense to find central nodes that affect other nodes in the network rather than simply the highest weighed nodes.

For each of the 1215 day-variable entries, we had 10 weights assigned to that value to feed into each of 10 hidden neurons. For each pair of day-variable entries, we would do the calculation shown in Equation 13 to yield a 1215x1215 summary variable matrix.

$$Equation\ 13: \omega_{ij} = \frac{\sum_{k=0}^{9} w_{ik} w_{jk}}{\sum_{k=0}^{9} |w_{ik}||w_{jk}|}\ \forall\ i \in 0, \ldots, 1214, j \in 0, \ldots, 1214$$

This 1215x1215 matrix is of interest since it captures the strength of association between these day-variable entries as a proportion of the maximal association we could have. The intention is not to calculate correlation, but rather association across entries in two non-random vectors and the skewness towards opposition or support for one another. This matrix also allows comparison between entries since certain variables could be more opposed or in favor of one another depending on the value yielded since a higher value means that overall higher same direction interactions are happening relative to the maximum interaction that could occur.

These variables would then be further summarized over all days to obtain a 45x45 matrix summarizing whether variables tend to occur together with the same or opposite signs over all 27 days. This is shown in Equation 14.

$$Equation\ 14:$$

$$W_{mn} = \det\left(\begin{bmatrix} \omega_{27m+1,27n+1} & \cdots & \omega_{27m+1,27n+27} \\ \vdots & \ddots & \vdots \\ \omega_{27m+27,27n+1} & \cdots & \omega_{27m+27,27n+27} \end{bmatrix}\right) \forall\, m \in 0, \dots, 44, n \in 0, \dots, 44$$

However, since we only have one sample of college students, we decided to bootstrap samples to obtain an average matrix to use for network creation. This is shown below in Equations 15-17 for 100 samplings from the 238 student sample.

$$Equation\ 15: \omega^{*}_{ij,h} = \frac{\sum_{k=0}^{9} w_{ik,h} w_{jk,h}}{\sum_{k=0}^{9} |w_{ik,h}||w_{jk,h}|}\ \forall\, i \in 0, \dots, 1214, j \in 0, \dots\ , 1214, h \in 1, \dots, 100$$

$$Equation\ 16: \widetilde{\omega_{ij}} = \sum_{h=1}^{100} \omega^{*}_{ij,h}\ \forall\, i \in 0, \dots, 1214, j \in 0, \dots\ , 1214$$

$$Equation\ 17:$$

$$\widetilde{W_{mn}} = \det\left(\begin{bmatrix} \widetilde{\omega_{27m+1,27n+1}} & \cdots & \widetilde{\omega_{27m+1,27n+27}} \\ \vdots & \ddots & \vdots \\ \widetilde{\omega_{27m+27,27n+1}} & \cdots & \widetilde{\omega_{27m+27,27n+27}} \end{bmatrix}\right) \forall\, m \in 0, \dots, 44, n \in 0, \dots, 44$$

To produce the graphs below, we used the highest and lowest 200 values of the $\widetilde{W}$ matrix to be able to see only the most significant interactions. These networks are shown in Figures 10-13 when created with the hidden layer contributions from the model that uses standardized data inputs, four clusters, and two hidden layers since it had the lowest MSE.

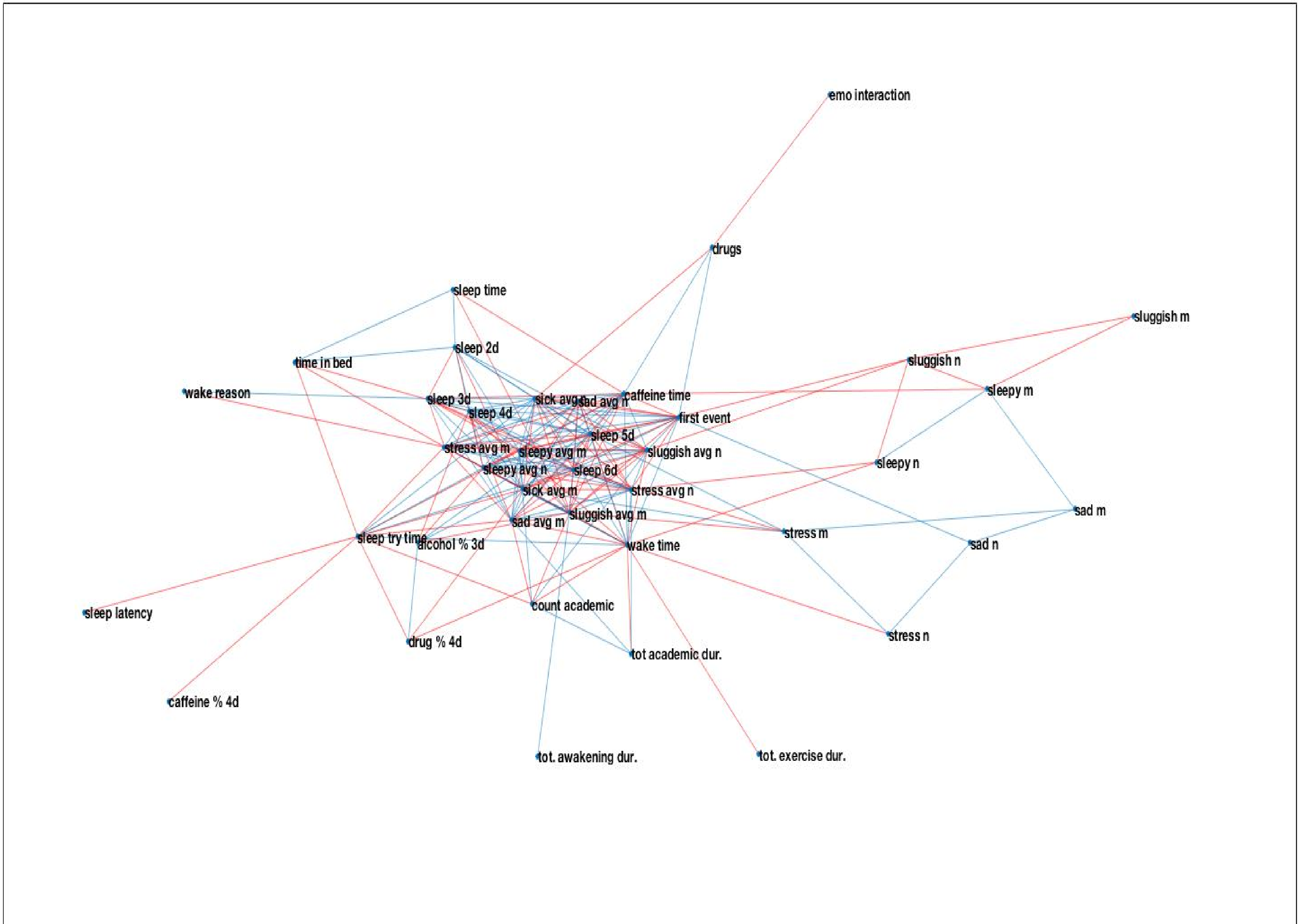

Figure 10: Network for variables contributing to calculating 28^{th} day mood influencing each other negatively(red) or positively(blue) in Cluster 1

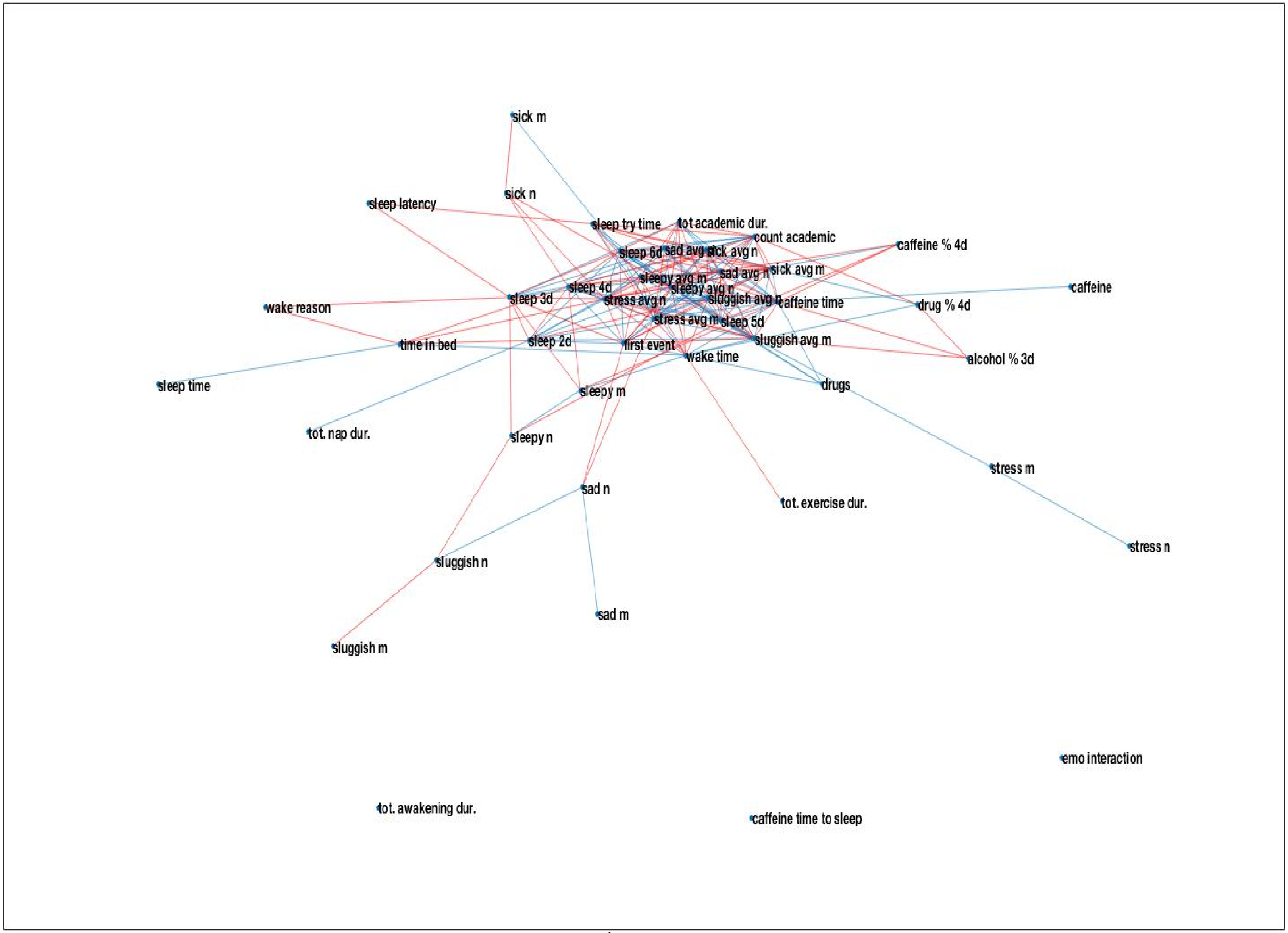

Figure 11: Network for variables contributing to calculating 28^{th} day mood influencing each other negatively(red) or positively(blue) in Cluster 2

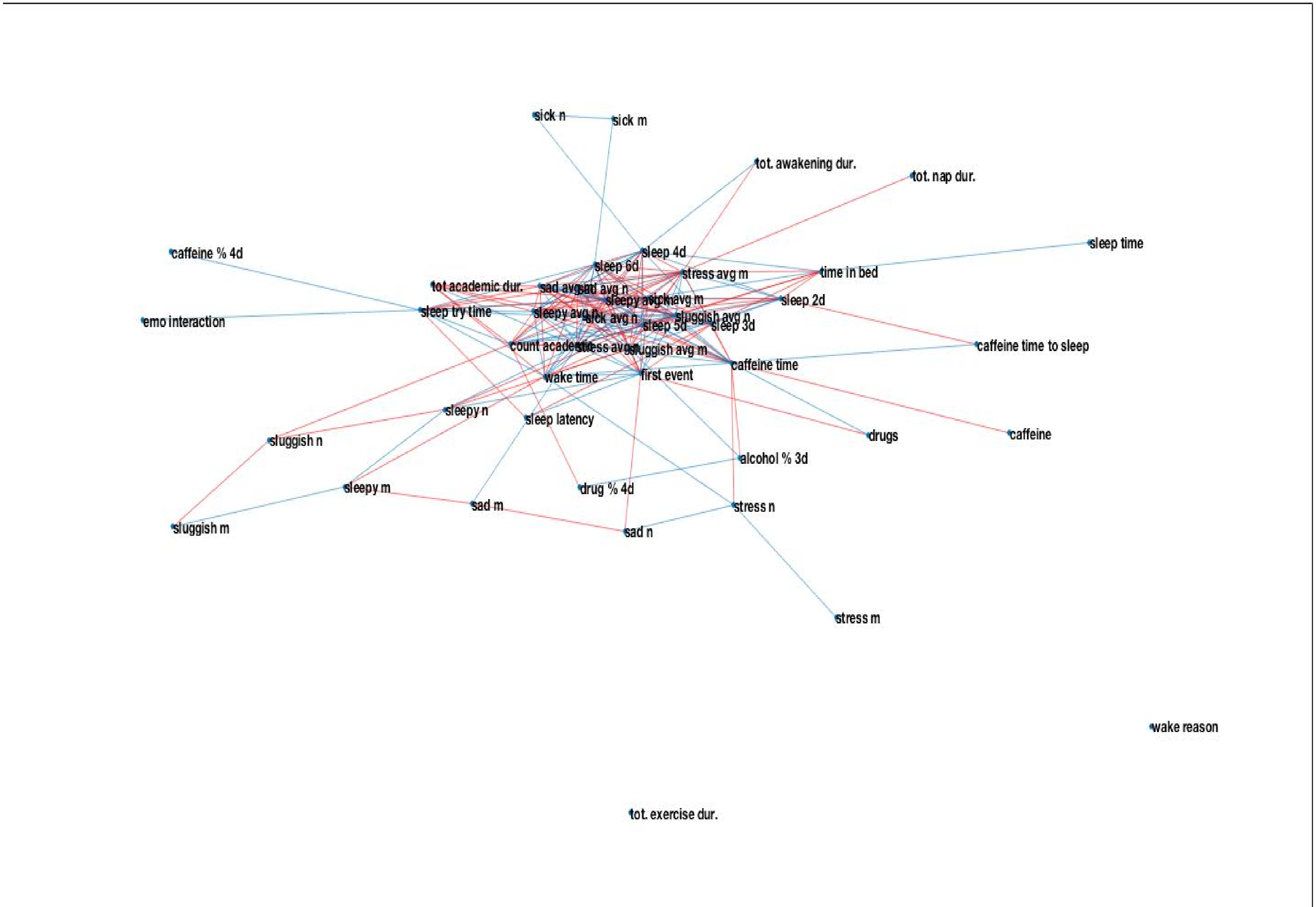

Figure 12: Network for variables contributing to calculating 28th day mood influencing each other negatively(red) or positively(blue) in Cluster 3

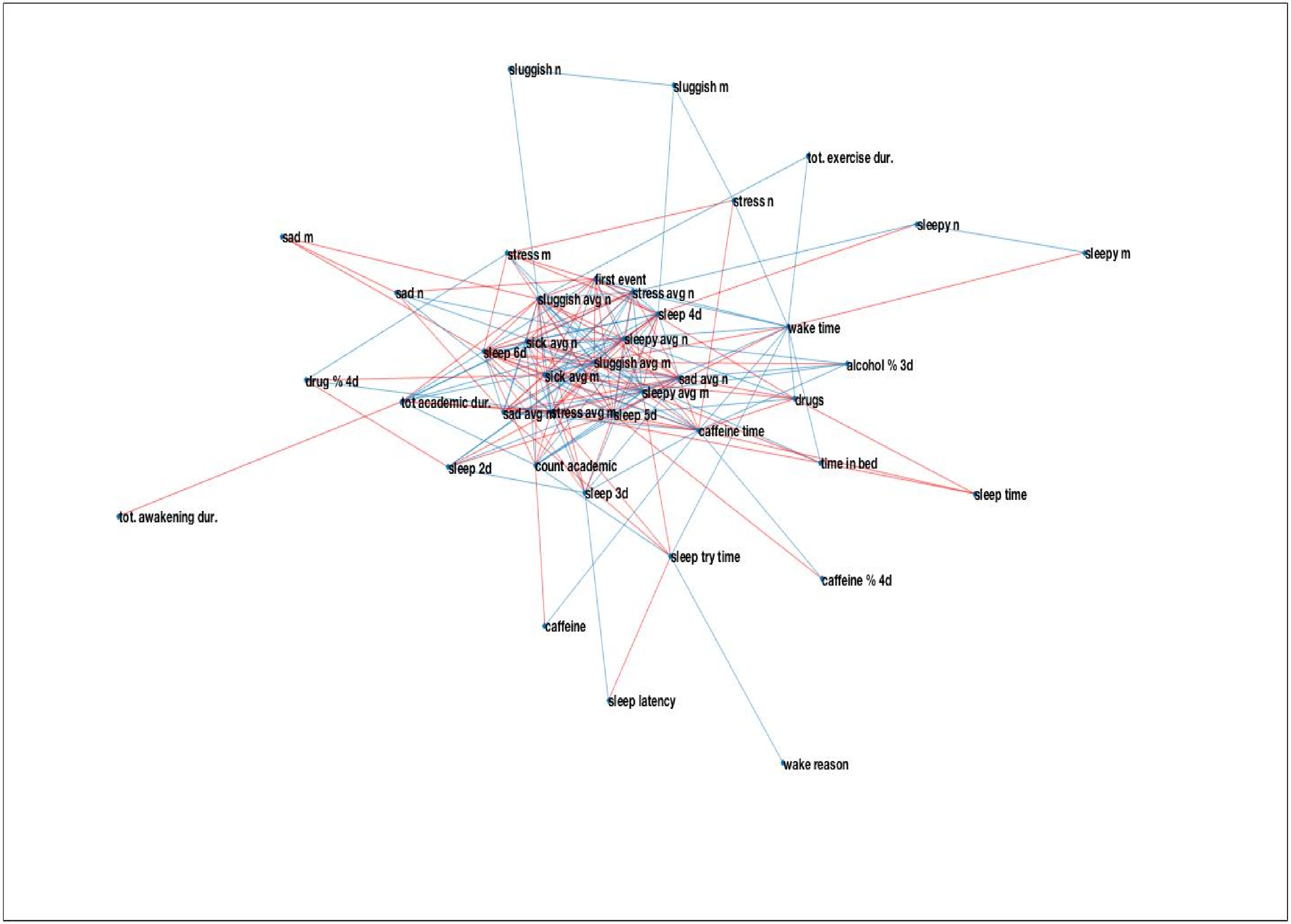

Figure 13: Network for variables contributing to calculating 28th day mood influencing each other negatively(red) or positively(blue) in Cluster 4

4.6 Finding central nodes and propagating effects of change

While the graphs above are interesting for summarization, they gloss over the importance of temporality in predicting mood. To gain a better understanding of temporality, we used the $\widetilde{\omega}$ matrix to find the most important day-variable pairings to consider in the prediction of 28th day mood. The goal of this analysis was to identify potential parameters that could be adjusted at strategic times to have a positive influence on future mood. Unlike the $\widetilde{W}$ matrix, the edges represented in $\widetilde{\omega}$ have time-directionality incorporated since a day-variable on an earlier day affects a day-variable on a later day but not vice versa. To determine the importance of certain nodes in $\widetilde{\omega}$, which is a 1215x1215 matrix, we first selected for the most extreme 2% of edge weights; only these weights were chosen to prevent oversaturating centrality measures with edges that correspond to a slight tendency to vary in the same way in the neural network weight matrix. This yields relationships in which day A-variable B almost always contributes to the neural network in the same or in the opposing direction as day C-variable D. In other words, knowing variable B on day A will allow us to predict its relationship to day C-variable D. Furthermore, this predicted relationship can be defined as an effect on variable D on day C since variable B on day A precedes it. We then measured centrality via an average Matlab_R2018b's *degree* and *betweenness* functions per node to account for the fact that while we are interested in the number of first degree relationships between day-variables, it is also important to account for when day-variables mediate other relationships.

We considered the top five nodes as measured above in each cluster to be central hubs to be targeted. To better understand these hubs' effects on other variables in the network, we looked to the shortest path to every other day-variable via Matlab's shortest path function, and used these results to understand days and variables that are important to consider as ultimate predictors of 28^{th} day mood.

From our summary matrix of the relationship between day-variable weights in the weight matrices of the bootstrapped neural networks, we learned that the nodes with the strongest relationships (top 2% magnitude of weights between edges), highest degree, and highest centrality overall in each cluster are often very similar. For cluster 1, the most important nodes included wake time on day 11 (on average a Monday or Tuesday), wake time on day 12 (on average a Tuesday or Wednesday), wake time on day 15 (on average a Friday or Saturday), wake time on day 19 (on average a Wednesday or Thursday), and time of caffeine intake on day 17 (on average a Sunday or Monday). For cluster 2, the most important nodes included wake time on day 12 (on average a Monday or Tuesday), first event time on day 11 (on average a Sunday or Monday), first event time on day 15 (on average a Thursday or Friday), first event time on day 17 (on average a Saturday or Sunday), and time of caffeine intake on day 7 (on average a Wednesday or Thursday). For cluster 3, the most important nodes included wake time on day 12 (on average a Sunday or Monday), first event time on day 10 (on average a Friday or Saturday), first event time on day 11 (on average a Saturday or Sunday), average weekly morning health on day 13 (on average a Monday or Tuesday), and time of caffeine intake on day 7 (on average a Tuesday or Wednesday). For cluster 4, the most important nodes included wake time on day 12 (on average a Tuesday or Wednesday), wake time on day 13 (on average a Wednesday or Thursday), average weekly morning health on day 13 (on average a Wednesday or Thursday), average weekly morning stress on day 11 (on average a Monday or Tuesday), and time of caffeine intake on day 7 (on average a Thursday or Friday). It is interesting to note that for cluster 1, central hubs tend to occur in a window that covers mostly weekdays. For cluster 2, central hubs are also mostly occurring in a window containing weekdays, with the exception of first event time on day 17. For cluster 3, the same is true except that the weekend-exclusive hub is first event time on day 11. Finally, for cluster 4, all nodes occur within windows containing weekdays. While these do not give conclusive results, given that the day of the week is based on the average day of the week that day x would fall on for that cluster, they do seem to indicate the behavior and sleep constraints that occur during the week have more importance than those from the weekend, even when the central hub is related to average mood scales. It is important to frame projected effects of changing variables and recommended behavioral changes with this knowledge in mind.

While this does not mean that these are direct predictors of 28th day mood, it does mean that based on the phenotypes of students, certain sleep variables on a specific day can have effects that trickle down over time and compound its contribution to predictions of 28th day mood because these variables are related over time and every variable included in this network had non-zero entries into the weight matrix of the neural network. Therefore, as a predictor of 28th day mood, the aforementioned day-variables are central hubs that are important to consider as one tries to predict mood based on monthly data but also as one tries to change behavior with the expectation of seeing future changes. They prove to be very interesting in the context of sleep because the shortest path to any other connected node in network is at most 6 edges; these central hubs are often connected to the subsequent day's sleep duration, sleep quality, and mood measures.

For cluster 1, via one, two or three edges, every central hub positively affects subsequent total awakening duration, total nap duration, amount of time spent on schoolwork, amount of time spent exercising, bedtimes, sleep latency, almost every day's average sleep, almost every day's morning mood, and even caffeine or drug intake. Similar results are found in clusters 2, 3, and 4, with additional effects on subsequent days' bedtimes, wake times, average morning or evening moods, and categorization of emotional interactions. This finding is noteworthy, because while these students have different central hubs, these hubs are ultimately affecting the same variables that are then inputs into the neural network. Thus, it would seem that to produce the same effect in 28th day mood for each group of students, the time and variable to address would depend on the sleep phenotype of the student. In this case, students in cluster 1 seem to be more responsive to the time they wake, suggesting perhaps they may need extended sleep in the morning. Students in cluster 2 seem to be more responsive to timing of the first event earlier into the month-long study. This finding may not mean that their chronotype differs significantly from students in cluster 1, since first event timing determines wake time to some degree, but rather that their schedules may be more influential to their mood ratings than those in cluster 1. In cluster 3, students have some of the same central hubs with the addition of an average weekly mood - health - earlier in the month, which may mean these students are less dependent on sleep to determine mood because their schedules are more accommodating to their ideal sleep timing. Finally, in cluster 4, students have the greatest reliance on average weekly moods, with some wake time reliance as well, suggesting that their ideal sleep schedule may already be accommodated by their schedule. Interestingly, caffeine intake early in the month is a hub for all the students, which may suggest that caffeine intake habits as displayed early in one's data collection are indicative of future habits and therefore future sleep outcomes.

While these effects varied between being positive and negative, they show that often, it may be useful to change a specific behavior to see changes about two weeks later in relation to mood. However, this leaves the question of how to change certain inputs to predict certain effects on mood. While it would seem to be a matter of plugging in different values into the vector input for the neural network in each cluster, this would not capture the number of changes per variable that could be optimal for increasing mood ratings. Thus, perhaps it is of better use to observe the net effects of changing inputs into the network of day-variable pairs. We carried this out in cluster 1 to see the net effect that an increase of 1 standard deviation in a central hub measure would have on an input vector where every item was originally at its mean. We used a depletion factor of 0.9 per edge crossed. Multiplying this across the neural network's weight matrices, one can see the changes shown in the table below.

Table 2: Cluster 1 projected mood change			
	28^{th} day alertness	28^{th} day happiness	28^{th} day health
Wake time on day 11	14.00	4.27	21.49
Wake time on day 12	8.93	7.41	-4.24
Wake time on day 15	13.87	10.80	16.73
Wake time on day 19	-0.81	-1.81	-1.46
Time of caffeine intake on day 17	-2.66	-5.30	-11.09

While the above changes may seem arbitrary, they reveal the importance of temporality as well as behavioral changes that are needed to evoke a desired mood change. It is not as simple as changing one variable on one day; however, changing these variables could produce positive effects when changed at the most appropriate times. For this group of students, delaying wake time, thereby increasing sleep duration, on days 11-15 seems to have the greatest positive effect on alertness, happiness, and health on day 28, with the exception of the effect of wake time on day 12 on 28^{th} day health. Moreover, avoiding the intake of caffeine too late on day 17, which is about a week and a half before day 28, would seem to consistently prevent a decrease in 28^{th} day alertness, happiness, and health. If these students have a chronotype that favors a later bedtime and wake time, then it would make sense that wake onset should be delayed early on since it would produce positive effects, such as increased time in bed, sleep duration, morning moods, and delayed bedtime according to a personally preferred schedule, that would compound over time to yield a net positive effect. It is also informative that even if these students are waking at later times, caffeine intake too close to bedtime is still detrimental to long-term mood because of the disruptions it causes in sleep over time.

For cluster 2, this same simulation of network-wide change and neural network predictions was carried out and the results are shown below in Table 3.

Table 3: Cluster 2 projected mood change			
	28^{th} day alertness	28^{th} day happiness	28^{th} day health
Wake time on day 12	12.99	11.37	5.85
1st event time-day 11	19.46	13.81	13.40
1st event time-day 15	28.83	7.96	11.08
1st event time-day 17	20.14	4.49	10.22
Time of caffeine intake on day 7	-5.81	0.44	-6.73

For this group of students, increasing wake time on day 12 also seems to increase alertness, happiness, and health, but more interestingly, delaying the time of the first event in the day on days 11-17 seems to be more important, as increases of one standard deviation in the timing of the first event increases mood ratings across the board for these students. The aforementioned idea that these students may also have a chronotype that favors later bedtime and wake time is consistent here since delaying the first time event allows more time to sleep in the morning, which would improve these students' moods over time. In addition, while caffeine does not have as strong of a negative effect on 28^{th} day happiness, intake of caffeine at later times does indicate risk of lowering alertness and health on the 28^{th} day for these students.

For cluster 3, this effect is shown in Table 4.

Table 4: Cluster 3 projected mood change			
	28th day alertness	28th day happiness	28th day health
Wake time on day 12	-0.09	4.60	0.83
1st event time-day 10	23.64	30.42	16.89
1st event time-day 11	12.06	13.98	16.57
Avg. weekly morning health on day 13	16.53	16.86	30.83
Time of caffeine intake on day 7	3.35	-16.81	-16.23

For this group of students, first event time about two weeks before day 28 also stands out as having a strong positive effect on eventual alertness, happiness, and health when increased by one standard deviation, and time of caffeine intake also continues to have a strong negative effect on eventual happiness and health when increased. However, this group is interesting in that future mood begins to have a strong reliance on average weekly morning health, such that an increase in average weekly morning health on day 13 would predict an increase in morning alertness, happiness, and health two weeks later. It is logical that improving mood earlier on would improve mood in the future, but this is not a behavioral change that can be implemented the way extending sleep time via wake time or first event time is. Rather, this seems to be an indication that these students do not have the same level of a behavioral problem that decreases sleep quality or quantity, such that mood can begin to be predicted simply from previous data. This is the same cluster that had a significantly higher sleep latency than cluster 1 and 2, significantly lower time in bed than cluster 2, significantly lower average sleep over 2 days and significantly higher average sleep over 4 days than cluster 2. This may seem to be contradictory and seems to be some extension of Simpson's Paradox where the average over more time is higher than the average over shorter periods of time. However, the takeaway from this is that the reliance on feelings of health may indicate a health problem that interferes with sleep along with behavior.

For cluster 4, this effect is shown in Table 5.

Table 5: Cluster 4 projected mood change			
	28th day alertness	28th day happiness	28th day health
Wake time-day 12	0.97	-3.72	-1.91
Wake time-day 13	1.54	3.79	4.67
Avg. weekly morn. health on day 13	-9.15	2.58	-13.57
Avg. weekly morn. calm on day 11	-5.29	-0.003	-13.90
Time of caffeine intake on day 7	0.60	-0.50	10.37

For this group of students, wake time increases on day 13 would also indicate some improvement in eventual alertness, happiness, and health. However, increasing average weekly morning health and calmness on days 13 and 11 respectively has a contradictory effect to the ideas motivated by cluster 3's outcomes since it would decrease eventual alertness and health. In addition, it indicates an even greater reliance on underlying biological determinants of mood rather than behavior. This group was originally clustered in a way that made them significantly different from other clusters in that their time in bed is significantly higher than those in cluster 3, their sleep duration is significantly higher than those in cluster 1 and cluster 2, and their average sleep over 2 days is significantly lower than those in cluster 2. This seems to be another example of Simpson's paradox, but what this result may indicate is that although individuals in cluster 4 are below their two-day average more than those in cluster 2, they may have a higher overall average as a result of individual days' higher values that are not smoothed as they are in a two-day moving average. To tie together these differences, the predicted change in moods that are contrary to expectations, and the biological contributions to mood, their overall better sleep would seem to indicate that wake time or first event timing is not the behavior to change for them. Rather, their previous existing mood may predict a decreased mood if improved upon because doing so would be outside of the normal range of moods seen for this cluster and would interfere with the biological determinants of mood. With no behavior to change and the impossibility of changing average moods, this may indicate that these students are most stable as they are.

5 Discussion

We began this research by asking whether it is possible to mathematically model college students' morning mood based on sleep and sleep-related variables in the previous 27 days. Because individuals often exhibit sleep phenotypes and chronotypes that determine their ideal wake and bed time as well as the time spent sleeping, we were interested in creating a clustered model such that the same class of model is built for each cluster of students, but the weights differ per cluster. We attempted to cluster the data using two methods: a spectrum-based, likelihood ratio-based quasi-distance to carry out hierarchical clustering and feature-based agglomeration that returns which groups of students appear as if they were all one feature. We decided to proceed with feature-based agglomeration with two types of inputs: periodograms that approximate the spectral frequencies and standardized scores from the past 27 days versus only standardized scores from the past 27 days. We tested the difference between using 2, 3, or 4 clusters in creating clustered models. To build the model within each cluster, we used neural networks since we felt they would most accurately capture the time dependency of each variable entered into the activation function. We used the weight matrix produced by the neural network to build a network that describes the relationship between variables contributing to the 28th day mood scales we produce as output. To extend this to have a greater understanding beyond our sample, we conducted a bootstrap simulation such that the weight matrix is an average of the quasi-distances between different day-variables.

With regards to the best data input, while periodogram-based inputs were an attempt to capture the frequencies at which the data varied, only inputting standardized values ultimately lowered mean-squared error regardless of cluster number and activation function. This may be because the standardized data may still capture fluctuations over time due as values that dip below the mean may correspond with a certain level of rating one's mood for that class of students. This ties in to the fact that increasing the number of clusters consistently lowered mean-squared error as well since these clusters, which separated students along lines of school work, sleep time or time in bed, and average sleep over the past 5 or 6 days, may have therefore been capturing fluctuations in these values across groups of students with different behaviors and therefore had weights in the neural network weight matrix that were tuned to these behaviors.

Comparisons across activation functions and number of layers proved to be informative as well. While the differences across the same number of clusters between different activation functions proved to be minimal, the 4-cluster, 2-layer linear condition proved to be most accurate. The weights assigned to day-variables as they fed into the layers proved to be more interesting from a network-wide perspective since the nonlinearities captured by the 2 layers could be extended to create weights between nodes in a network without changing the values of the weights because they ultimately fed into a linear activation function.

This adds to the already existing pool of knowledge in a number of ways. Drastically changing future perception of mood is not as simple as changing a single variable on a single day. However, mood improvement can mean targeting certain behaviors, as predicted by previous research, that can slowly aggregate to produce certain results. While the improvement of mood about two weeks after changes in central nodes may be an artifact of a changed behavior being kept over time, it is important to note that this is evidence that behavior interacts with biology to produce changes. This is especially true with regard to chronotype, which must be taken into account to create a treatment plan. For example, as noted above, for a student who not only wants to increase mood ratings, but may be trying to do so to combat the risk of depression or anxiety, falling into a cluster such as cluster 1 would suggest a re-evaluation of wake timing about two weeks before the intended improvement and attempting to extend sleep as much as possible as a treatment plan while falling into cluster 2 involves a re-evaluation of one's overall schedule as well to see results in following weeks. Moreover, if one falls into cluster 3 or 4, the decreased reliance on sleep variables as central hubs indicates that treatment may not be as related to behavior as it is to addressing the biological causes of sadness, depression, stress, or anxiety. This is important as one considers that existing research on depression, anxiety, and sleep have shown that low sleep and high sleep increase risk of depression even as depression, which is due to uncertain causes but likely due to triggers and biological predisposition to develop it, causes longer sleep hours[12]. Thus, this indicates that different clusters of students may have different sleep-depression graphs since clusters 1 and 2 exhibit the expected relationship where increasing sleep would decrease depression-related moods while clusters 3 and 4 may be less radically improved as sleep improves since there are other factors that are decreasing the rate of change in the sleep-depression curve[12]. It also stands with information from the Owens study that for students, wake time and delayed morning activities often produce more sleep duration and therefore lower depressive mood symptoms and visits to the health clinic[13]. Likewise, this information adds nuance to the conversation about caffeine since timing is shown to be critical to long-term effects. The fact that caffeine timing in the middle of the month is so important could be due to a confirmation in the middle of the month that a certain time is typical for caffeine intake and the combined effects from the past having effects that multiply in the future until we reach 28th day, such that a later caffeine timing in the middle of the month combines the effects of past caffeine intake and subsequent effects on sleep that affect 28th day mood[13,14]. Moreover, the timing has increasing effects on sleep quality, which may explain why sleep duration would be beneficial to extend across all clusters if caffeine is not addressed since it may help make up for the number of awakenings in the night[14]. These findings could prove to be useful as universities consider the timing of classes or attempt to advise students. Knowing more information about their sleep history could help categorize the type of student and the best schedule for them. Moreover, shifting classes to start at later times across the board could help students in every cluster.

There are a number of improvements that could have been implemented in this research. The failure to implement clustering based on periodograms and a likelihood ratio test is one salient example. Perhaps one solution was to not rely on Python's hierarchical clustering algorithm and instead attempt to optimize the distance between individuals in the same cluster to be as small as possible by writing the script oneself. This optimization would try to limit the sum of distances between individuals in a cluster by taking the sum of quasi-distances between individuals when placed in the same cluster to yield optimal clusters. However, this may fall upon the same pitfall as the hierarchical clustering, which means that noise in the data may render quasi-distances so small that it is optimal to have one large cluster when quasi-distances are based on a likelihood ratio test with a null hypothesis that two individuals have the same distribution of spectral frequencies. There also could have been different methods used to establish causality between variables rather than having used a network perspective of the interactions contributing to 28th day mood. Namely, the use of Hill's criterion to establish causation would aid in determining risk factors to address as a society rather than only using network connections and predictive capabilities of certain variables as we did in this study.

6 References

1. Lund H.G., Reider B.D., Whiting A.B., & Prichard J.R. Sleep Patterns and Predictors of Disturbed Sleep in a Large Population of College Students. *Journal of Adolescent Health.* 46. 124-132. doi:10.1016/j.jadohealth.2009.06.016. (2010)

2. Phillips A.J.K., Clerx W.M., O'Brien C.S., Sano A., Barger L.K., Picard R.W., Lockley S.W., Klerman E.B., & Czeisler C.A. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Nature*. 7: 3216. 1-13. doi:10.1038/s41598-017-03171-4. (2017)

3. Short M.A., Gradisar M., Gill J., & Camfferman D. Identifying Adolescent Sleep Problems. *PLOS.* 8: 9. 1-6. (2013)

4. Miller M.B., Van Reen E., Barker D.H., Roane B.M., Borasari B., McGeary J.E., Seifer R., & Carskadon M.A. The Impact of Sleep and Psychiatric Symptoms on Alcohol Consequences among Young Adults. *Addict Behavior.* 66. 138-144. doi:10.1016/j.addbeh.2016.11.023. (2017)

5. Winokur A. The Relationship Between Sleep Disturbances and Psychiatric Disorders. *Psychiatric Clinic North America.* 38. 603-614. (2015)

6. Ojio Y., Nishida A., Shimodera S., Togo F., & Sasaki T. Sleep Duration Associated with the Lowest Risk of Depression/Anxiety in Adolescents. *Sleep and Psychiatric Disorders.* 39: 8. 1555-1563. http://dx.doi.org/10.5665/sleep.6020. (2016)

7. Tuarob S., Tucker C.S., Kumara S., Giles C.L., Pincus A.L., Conroy D.E., & Ram N. How Are You Feeling?: A Personalized Methodology for Predicting Mental States from Temporally Observable Physical and Behavioral Information. *Journal of Biomedical Information.* 68. 1-19. doi:10.1016/j.jbi.2017.02.010. (2017)

8. Klerman E.B., Dijk D., Kronauer R.E., & Czeisler C.A. Simulations of light effects on the human circadian pacemaker: implications for assessment of intrinsic period. *The American Physiological Society.* R272-R283. (1996)

9. Cohen S., Fulcher B.D., Rajaratnam S.M.W., Conduit R., Sullivan J.P., St. Hilaire M.A., Phillips A.J., Loddenkemper T., Kothare S.V., McConnell K., Ahearn W., Braga-Kenyon P., Shlesinger A., Potter J., Bird F., Cornish K.M., Lockley S.W. Behaviorally-determined sleep phenotypes are robustly associated with adaptive functioning in individuals with low functioning autism. *Scientific Reports*. 7:14228. 1-8. (2017)

10. Sano A., Yu A., McHill A.W., Phillips A.J.K., Taylor S., Jaques N., Klerman E.B., Picard R.W. Prediction of Happy-Sad Mood from Daily Behaviors and Previous Sleep History. (2015)

11. Ravishanker N., Hosking J.R.M., Mukhopadhyay J. Spectrum-based Comparison of Stationary Multivariate Time Series. *Springer Science+Business Media, LLC.* 12. 749-762. (2010)

12. Ojio Y., Nishida A., Shimodera S., Togo F., Sasaki T. Sleep Duration Associated with the Lowest Risk of Depression/Anxiety in Adolescents. *Sleep*. 39: 8. 1555-1562. (2016)

13. Owens J. Insufficient Sleep in Adolescents and Young Adults: An Update on Causes and Consequences. *American Academy of Pediatrics*. 134: 3. e921- e932. (2014)

14. Drake C., Roehrs T., Shambroom J., Roth T. Caffeine Effects on Sleep Taken 0, 3, or 6 Hours before Going to Bed. *Journal of Clinical Sleep Medicine.* 9: 11. 1195-1200. (2013)

15. Bernstein E.E., McNally R.J. Acute aerobic exercise helps overcome emotion regulation deficits. *Cognition and Emotion*. 31: 4. 834-843. (2017)

16. Edwards M.K., Rhodes R.E., Mann J.R., Loprinzi P.D. Effects of acute aerobic exercise or meditation on emotional regulation. *Physiology & Behavior*. 186. 16-24. (2018)

17. Uchida S. Shioda K., Morita Y., Kubota C., Ganeko M., Takeda N. Exercise effects on sleep physiology. *Frontiers in neurology*. 3: 48. 1-5. (2012)

18. Kline C.E. The bidirectional relationship between exercise and sleep: Implications for exercise adherence and sleep improvement. American Journal of Lifestyle Medicine. 8: 6. 375-379. (2014)

19. Barger L.K., Wright K.P., Hughes R.J., Czeisler C.A. Daily exercise facilitates phase delays of circadian melatonin rhythm in very dim light. AJP-Regul Integr Comp Physiol. 286. (2004)

20. Ye L., Johnson S.H., Keane K., Manasia M., Gregas M. Napping in College Students and Its Relationship With Nighttime Sleep. Journal of American College Health. 63: 2. 88-97. (2015)

21. Muller B., Reinhardt J., Strickland M.T. Neural Networks: An Introduction. *Springer*. (1995)

22. Cybenko G. Approximations by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems*. 2. 303-314. (1989)

23. DeMartini K.S., Fucito L.M. Variation in Sleep Characteristics and Sleep-Related Impairment in At-Risk College Drinkers: A Latent Profile Analysis. *Health Psychology.* 33: 10. 1164-1173. (2014)

24. Inderkum A. P, Tarokh L. High Heritability of Adolescent Sleep-Wake Behavior on Free, but not School Days: A Long-Term Twin Study. *Oxford University Press*. (2018)

25. Demirci K., Akgonul M., Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. *Journal of Behavioral Addictions*. 4: 2. 85-92. (2015)

26. Orzech K. M., Grandner M. A., Roane B. M., Carskadon M. A. Digital media use in the 2 h before bedtime is associated with sleep variables in university students. *Comput Human Behavior*. 55: A. 43-50. (2016)

27. Short M.A., Arora T., Gradisar M., Taheri S. Carskadon M.A. How Many Sleep Diary Entries Are Needed to Reliably Estimate Adolescent Sleep?. *Sleep*. 40: 3. 1-10. (2017)

28. Wong M.L., Lau E.Y.Y., Wan J.H.Y., Cheung S.F., Hui C.H., & Ying D.S. The interplay between sleep and mood in predicting academic functioning, physical health and psychological health: A longitudinal study. *Journal of Psychosomatic Research.* 74: 4. 271-277. (2013)

29. Brownlee, J. Gentle introduction to the Adam optimization algorithm for deep learning. (2017).https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/