Predicting Weight Loss with Graph Neural Networks Using Social Network Data

Aili Harmon
Department of Computer Science
Vanderbilt University
aili.z.harmon@vanderbilt.edu

Daniel Shu
Department of Computer Science
Vanderbilt University
daniel.m.shu@vanderbilt.edu

Zack Shatsky
Department of Computer Science
Vanderbilt University
zachary.shatsky@vanderbilt.edu

Abstract

The prevalence of obesity in the United States is increasing, and obesity can cause expensive health problems like type 2 diabetes and cancer. We examined data about the features and social activity of users on BOOHEE, a popular Chinese weight loss social media app. We filtered out users with missing feature information or who had a degree of less than five in the friend, mention, or comment network. We used a linear regression neural network and four different graph neural network models to try to predict the final weights of users. None of the GNNs that we tried outperformed our linear regression neural network. Future work in this area could include trying new GNN models or using a different dataset. The average BMI of the users in our filtered dataset was not in the overweight or obese range by American standards, so experiments involving a heavier group of people may have greater public health implications.

Introduction

Over the past few decades, obesity within the United States has increased at an alarming rate. Since 1999, the prevalence of obesity among adults aged 20 years and older has increased from 30.5% to 41.9% (Fryar, Carroll, and J 2020). Child obesity has also increased at a similarly high rate over the past few decades. This trend is problematic because of the devastating consequences that accompany this epidemic. For instance, obesity has been linked with increased rates of heart disease, stroke, type 2 diabetes, and various forms of cancer (Fryar, Carroll, and J 2020). These diseases come with catastrophic physical and emotional reverberations and take a steep financial toll on those they influence. Estimates show that obesity and obesity-linked diseases have led to $173 billion of medical expenditures in 2019 in the U.S., and those inflicted pay an average of $1,861 more per year in health expenses than those of normal weight (Ward et al. 2021). As a result, we have seen the increased popularity of fad diets, nutrition products, and medical tracking devices recently in an attempt to combat this issue.

Among the common innovations to reverse the trends of obesity are mobile applications that track the health or nutrition of users. As mobile smartphone devices become omnipresent in society, applications aid users in recognizing healthy behaviors, motivate users to exhibit sustainable health practices, and discourage unhealthy actions. Such applications are considered satisfactory in their effectiveness of weight loss, particularly with high levels of engagement, but are most beneficial in their convenience and ease of use (Dounavi and Tsoumani 2019). Self-monitoring is one of the most common forms of weight management applications, which gives users the opportunity to track their food intake and/or fitness activity over a period of time. While most applications focus on the individual user, there are some that incorporate social media features into the framework of the platform. Since evidence shows that social support is necessary to maintain a healthy lifestyle, these applications seek to take advantage of the motivational component of social networking that may arise from seeing peers make progress toward their health goals and being able to share their progress with their peers (Jane et al. 2018). WW, Strides, and Peloton are just a few of the platforms that utilize a community-based network to motivate users toward their goals. Throughout the user life-cycle much data is collected, such as demographic information, social activity, and health progression, which can lead to ample opportunity to analyze the influences of social networking on weight loss.

In this paper, we seek to understand the influence of the role of social networking features in weight loss management applications. In doing so, we will attempt to predict the weight loss of individuals based on a variety of factors–both related to their social activity and some of their personal information. Based on our results, we discovered the graph neural networks that are most predictive of weight loss as well as the social aspects of the application that contributes most to an accurate weight loss prediction.

These results are significant for multiple reasons. First, it can be informative for people who seek to lose weight to know what factors may help them most in subconsciously motivating them to lose more weight or stay persistent with their goals. For example, if we determine the friends one has is a better predictor of weight loss, we may suggest a user connect with more friends or connect with those who are better able to reach their own fitness goals. In addition, the results may be informative to the developers of weight loss applications with social media features. If one type of social media data was more predictive of weight loss, future platforms may prioritize it as an essential tool to help users lose weight.

The remainder of this paper is organized as follows. In the next section, we will discuss the background of the experiment, including related work and our problem statement. Then, we will introduce the dataset by describing some initial data analysis and statistics in Section 3. In Section 4, we will describe our methodology to solve the problem, which includes the graph neural networks and loss we utilized to predict weight loss. In Sections 5 and 6, we will detail the experimental results in more detail and conclude the paper by mentioning potential future work, respectively.

Background

Research related to our paper falls primarily under two categories. The first focuses on identifying the effectiveness of mobile applications in weight loss in experimental trials. The next is research that takes advantage of network data to predict weight loss.

Weight Loss Applications

Because of how prevalent obesity is in society, much research has been conducted on the most effective ways to lose weight, including the efficacy of mobile applications and social media on weight management. Research has shown that mobile applications can help with weight loss (Carter et al. 2013). What was particularly interesting was that people who use a smartphone for self-monitoring weight loss purposes were more likely to adhere to their plan as opposed to people who used a website or paper diary (Carter et al. 2013). Other forms of weight management, such as counting calorie balances (Tsai et al. 2007) and promoting physical activity through the use of sensors on the body (Consolvo et al. 2008), may also be effective. In all cases, the convenience of using a mobile phone to aid in weight management has led to more users adhering to the plan for a longer period of time (Carter et al. 2013) (Consolvo et al. 2008).

Weight Loss Prediction

Less research has been done on the intersection of weight loss and social media in predicting outcomes rather than just effectiveness. Our problem is similar to a classification problem except our loss function is modeled after a regression problem (i.e., we are using an RMSE loss function instead of a cross-entropy loss function). Papers have predicted weight loss based on social media factors (Wang et al. 2017) (Kim, Kim, and Park 2021). In (Kim, Kim, and Park 2021), the authors use an interpretable artificial intelligence algorithm to predict weight loss based on data from Noom, a mobile app that provides the ability to log lifestyle-related activities like food consumption, exercise, and weight, with a percent error of 3.50%. In (Wang et al. 2017), the authors use data from BOOHEE (as in our study) to both predict weight loss of users and identify the individual and social factors most relevant to weight loss by using a random walk based method, feature engineering and user embedding.

Data

The data used in this paper comes from a popular social media app for weight loss in China called BOOHEE. Users who create their account share basic information about themselves, such as their age, gender, and body mass index (BMI) to set up their profiles. Some capabilities of the platform are that users can follow others, they can post their weight loss progress to share with others, and they can comment on the posts of other users. When users post their progress, they can also mention other users. These functionalities can lead to the generation of three graphs: the following/follower relationships (friend network), the mentioner/mentioned user relationships (mention network), and the commenter/poster relationships (comment network). We utilized both the directed and undirected versions of these graphs during our analysis.

Methods

Our goal is to get an accurate prediction of weight loss for each user. In order to do this, however, some preprocessing of the data was required. With nearly 10 million data points, there was originally too much data to work with. After doing some filtering, we were able to reduce the number of nodes to about 30,000. After that, we fed the three graphs into four different GNN networks. Using hyperparameter tuning, we identified the GNN models that best predict weight loss based on the graph data.

Data Preprocessing

Starting with 10 million data points, we first removed data with missing or null values. This reduced the dataset to about 4,000,000 users. In order to reduce the amount of data further, we wanted to ensure we used significant data to create our predictions. This means we would use users who were active in the application and are meaningful nodes within the three graphs (i.e., have a certain degree). Furthermore, we wanted to isolate a common set of users who are present in all three graphs so we could make more direct comparisons between the effectiveness of the interaction types on predicting weight loss. We decided that filtering out nodes with degrees of less than five in any of the graphs would satisfy this notion. In order to do this, we converted the graphs held as Pandas data frames into PyTorch tensors. The nodes were then filtered iteratively based on a mask until each of those user nodes had degree five or greater in all three graphs.

The final preprocessing step we took was constructing undirected versions of the three graphs because we wanted to experiment with methods dealing with undirected graphs only. This was implemented by duplicating all edges in the networks in the opposite direction, and removing any repeated edges in the new graphs.

Remaining Data

Once we filtered out nodes with a degree of less than 5 in the friend, comment, or mention network, we were left with 33,015 users. The dataset was overwhelmingly female, with 31,827 women and 1,187 men. The average initial weight of these users was 61.4 kg, the average height was 162.9

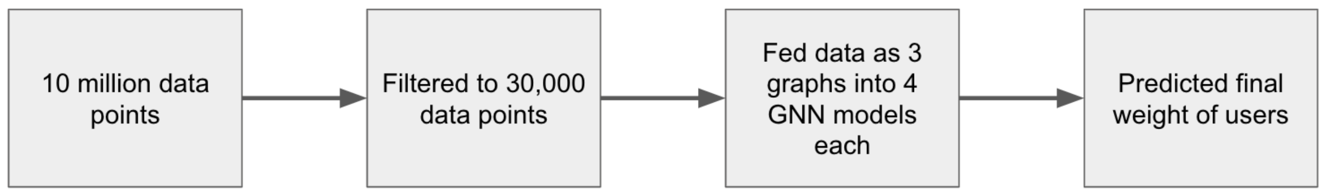

Figure 1: Summary of methodology

cm, the average initial BMI was 23.1, and the average age was 25 years. The average most recently recorded weight of these users was 58.8 kg.

Although the three graphs now share the same set of users, they still display plenty of diversity. We evaluated this by checking the number of edges in each of the directed graphs as well as the sizes of the pair-wise graph intersections. The final friend, mention, and comment networks contain 3,280,543; 1,140,037; and 1,609,929 edges respectively. The sizes of the intersections are listed below:

- Friend-Mention Intersection: 427,986 edges
  - 13.0% of Friend network
  - 37.5% of Mention network
- Friend-Comment Intersection: 870,551 edges
  - 26.5% of Friend network
  - 54.1% of Comment network
- Mention-Comment Intersection: 355,278 edges
  - 31.2% of Mention network
  - 22.1% of Comment network

The highest percentage of one network that is contained in another is 54.1%, so we can reasonably assume that there is enough deviation for the models to show differences in results depending on the relative effectiveness of each interaction type.

Models

The four GNN models we chose to use are the Graph Convolution Network (GCN), Graph Attention Network (GAT), Self-Supervised Graph Attention Network (SuperGAT), and General Convolutional Network. The latter two used the directed versions of the networks, whereas the former used undirected versions of the networks.

GCNs were an advancement upon early versions of graph neural networks in that they deal with non-regular data structures. The GCN model uses an efficient layer-wise propagation based on a first-order approximation of spectral convolutions on graphs (Kipf and Welling 2016). It is called convolutional because filter parameters are usually shared over all locations in a graph. Contrastly, GATs use masked self-attentional layers to assign differing importance to nodes within a neighborhood (Veličković et al. 2017). This allows us to weight the importance of different nodes based on their features. This method is computationally efficient as it does not require knowing the graph structure up front. SuperGAT improves upon GAT in that it uses self-supervision for link prediction based on the graph's average degree and homophily (Kim and Oh 2022). This is particularly useful for noisy graphs. SuperGAT is able to model sparse graphs with a large number of negative samples well. (Kim and Oh 2022). Finally, General Convolution Network is a generalized GNN layer built from a GNN design space and GNN task space (You, Ying, and Leskovec 2020). We then used root mean squared error (RMSE) to determine the performance of the models. A lower RMSE indicates that the model does a better job of predicting the weights of users.

Experiments

We wanted to explore whether the relationships between users in a weight loss social network is predictive of the amount of weight that users will lose. We also wanted to determine which of our GNN models and which set of hyperparameters is the most effective.

Linear Regression Neural Network

We developed a basic regression neural network to measure the effectiveness of our GNN models against. Our regression neural network has three layers. We tried learning rates of 0.00001, 0.01, and 0.9 and weight decays of 0.0005 and 0.1, for a total of six combinations of hyperparameters. The most effective set of hyperparameters was a weight decay of 0.0005 and a learning rate of 0.01, which produced a validation RMSE of 4.87 and a test RMSE of 4.82. If the edges between users in the friend, comment, and mention networks provide useful information about users' weight loss, we expect at least one of our GNN models to produce a validation and test RMSE of less than 4.82.

Features Dropped	Validation RMSE	Test RMSE
None	4.87	4.82
Weight	5.42	4.92
Weight, BMI	9.28	9.09
Height	4.98	5.60
Height, BMI	4.72	4.80
Age	4.57	4.49
Gender	4.38	5.05
All Except User ID	10.54	10.62

Table 1: Linear Regression with Feature Exclusion

After we discovered the best set of hyperparameters, we ran our linear regression neural network with various features removed from the dataset to see which features were most important for making weight loss predictions. The results of this are shown in Table 1. A person's weight can be calculated using their height and their BMI, so we tried removing weight by itself and removing both weight and BMI. We did the same for height. Removing weight and BMI had the largest impact on the RMSE than removing any other feature. Removing height, age, or gender did not appear to have a large impact of the efficacy of our model. Each trial had a different train-test-validation split, which may have had a slight influence on the RMSEs.

GNN Results

We trained our models over 500 epochs each and ran a grid search on the following sets of hyperparameters:

- Number of Layers: $\{2, 3, 4\}$
- Weight Decay: $\{0.0005, 0.1\}$
- Learning Rate: $\{0.00001, 0.01, 0.9\}$

Our results can be found in Tables 2, 3, 4, 5.

	Friend	Comment	Mention
Best Number of Layers	2	2	2
Best Weight Decay	0.0005	0.0005	0.0005
Best Learning Rate	0.9	0.9	0.9
Validation RMSE	10.24	10.30	10.94
Test RMSE	10.24	10.66	10.74

Table 2: GCN Results

	Friend	Comment	Mention
Best Number of Layers	3	2	2
Best Weight Decay	0.1	0.1	0.1
Best Learning Rate	0.01	0.01	0.01
Validation RMSE	9.68	8.91	8.45
Test RMSE	10.40	8.65	8.54

Table 3: GATConv Results

	Friend	Comment	Mention
Best Number of Layers	2	2	2
Best Weight Decay	0.0005	0.0005	0.1
Best Learning Rate	0.01	0.01	0.01
Validation RMSE	8.74	10.07	8.56

Table 4: SuperGATConv Results
This model was significantly slower than the other three models, so we had to halt our program early. Because of the way our program was designed, we were not able to get the test RMSE without running it to completion.

	Friend	Comment	Mention
Best Number of Layers	2	2	2
Best Weight Decay	0.1	0.0005	0.1
Best Learning Rate	0.01	0.01	0.01
Validation RMSE	4.83	4.69	4.69
Test RMSE	4.75	4.98	4.74

Table 5: GeneralConv Results

Mention	H=16	H=32
Number of Layers	2	2
Best Weight Decay	0.1	0.1
Best Learning Rate	0.05	0.01
Validation RMSE	5.11	4.28
Test RMSE	4.48	4.61

Table 6: Results of Futher Experimentation on GeneralConv

Evaluation

Table 2 shows the results of our GCN model, which was the first undirected GNN that we tried. On the friend, comment, and mention networks, the test RMSEs ranged from 10.24-10.74, which is significantly worse performance than our linear regression neural network. Table 3 shows the results from our GAT model, which was the second undirected GNN that we tried. The test RMSE of the comment network was 8.65 and the test RMSE of the mention network was 8.54, which is better than the GCN model but still not as good as the linear regression neural network.

The first directed model we tried used the SuperGAT convolutional layer, and the results are shown in table 4. The results that we got were about the same as the undirected GAT model, and it took significantly longer to run so we halted our experimentation early.

Overall, the GNN with the GeneralConv layer (shown in table 5) was the most effective. This GNN performed about the same as our linear regression neural network, with test RMSEs running from 4.74-4.98. Our linear regression neural network outperformed all of the GNNs in terms of speed, so the linear regression neural network appears to be the best choice overall. Using the edges between nodes in a GNN was not more beneficial than using only the node features in a standard neural network.

In every model, two layers was almost always more effective than three or four layers. The only exception was the GATConv model on the friend network, where the best set of hyperparameters was three layers, a weight decay of 0.1, and a learning rate of 0.01. However, the validation RMSE and test RMSE were higher than the results for the comment and mention networks, where the most effective number of layers was two. Using two layers was usually faster than using three or four, so two layers was the best choice overall. The most effective weight decay and learning rate varied.

For each GNN model, we observe that the three different graphs yielded similar results to each other. This is an indication that the interaction type between users did not provide any significant information that the models could use to better predict weight loss. This may also be a side effect of the GNNs not outperforming the simple linear regression model anyways; since the graph information in general was not useful, it did not matter what specific type of information the graph encapsulated.

Follow-up Experimentation

Due to its relative good performance, some additional experimentation on different hyperparameter sets was performed using the GeneralConv model, the results of which are shown in Table 6. Since there was no significant difference in performance between the different graphs, we chose to use the Mention graph since it is the smallest. Furthermore, we fixed the number of layers to 2 and trained on 300 epochs. A grid search was performed on a more focused set of learning rates and a larger field of weight decay values. We also investigated increasing the hidden dimension size H.

We found no significant improvements over the initial reported RMSEs. It does not appear that the hyperparameters can be further tuned to make the basic models that we used perform better.

Further Discussion on Methods and Efficiency

We have already discussed efficiency briefly in regards the the SuperGatConv model and the Linear Regression model. As another investigation into the efficiency of our methods from a different angle, we trained again on our best hyperparameter sets for the comment and mention graphs for the GCN, GAT, and GeneralConv models to plot the train and test losses. We also did this for the H=32 hyperparameter set for GeneralConv with the mention graph. Some of results are shown in Figures 2, 3, and 4.

We can observe that most of the RMSE values for train and test loss converged quite early and did not require training on the full 500 epochs to reach their lowest values. In general, it seemed that the best models would reach stability by around 300 epochs at the latest, which may have been helpful to learn during our experimental phase as efficiency was a challenge we struggled with.

Conclusion

Overall Discussion

Our GNN models that used data showing the social interactions between BOOHEE users, regardless of the type of social interaction, did not perform significantly better than our linear regression neural network, which only used data about the features of individual users. There are two possible explanations for this. The first possible explanation is that the friend, comment, and mention relationships between users are not predictive of overall weight loss. Maybe the weight loss of BOOHEE users was influenced by real life social interactions or personality factors, and not social interactions on the social media app. The second possible explanation is that the GNN models that we tried were not the best possible models for solving this problem. Future research could help determine which is more likely. We were also unable to find a correlation between weight loss and age, gender, or height.

Challenges and Limitations

We used a small sample of the overall dataset to train and test our models. It's possible that users with less than five friends, comments, or mentions behaved differently than the users that we studied.

Our GNN models took a long time to run on our computers, which limited the amount of hyperparameter combinations that we could try. Using a cloud platform to run our program instead of our personal laptops could've saved time and made it easier for us to run other resource-intensive programs on our computers.

We did not run our models multiple times with the same hyperparameter combinations, so we don't know how variable our results are. It's possible that changing the items in train-validation-test split impacts the RMSEs. To truly determine if our best GNN model was better than our linear regression model, we would have to run each multiple times.

Future Work

Our simple GeneralConv GNN model could be refined further in the future. We could explore the effects of more complex features such as dropout. There were some other parameters for the GeneralConv layer that we didn't change that we could try tweaking in the future. There may be a style of convolutional layer that we didn't test in our project that may perform better than the GeneralConv layer. Machine learning is a rapidly advancing field, so there may be a new type of GNN developed in the next few years that works better on this dataset.

In our program, we used the user's initial weight, age, height, gender, and bmi to predict the user's most recent updated weight. Some of the users in our dataset updated their weight multiple times, but we only used the first and last weight data point. We also didn't look at the timestamp of the most recent weight update. In the future we could develop a program that looks at the timestamps and weights listed on all of the users' weight updates, and uses that information to predict when the user will post another update and what their next weight will be.

It might be interesting to try similar experiments on a different dataset. There may be cultural differences between people in China and people in other countries. However, finding another network with publicly accessible data that has both information about users' weights and users social activity may be difficult. Researchers could get participants on an app like Twitter to agree to take a survey and share their recent weight fluctuations, then look at data from the users' social media account.

On average, the users in the dataset that we used to test our models had a BMI of 23.1. In order to be classified as overweight in the United States, an individual must have a BMI of at least 25. To be classified as obese, an individual must have a BMI of 30 or more (Ward et al. 2021). This means that many of the most active users of BOOHEE are not overweight or obese. Studying people with higher BMIs in future research may have greater implications for public health. Patterns of weight loss may be different for individuals with higher BMIs because they have more fat that they

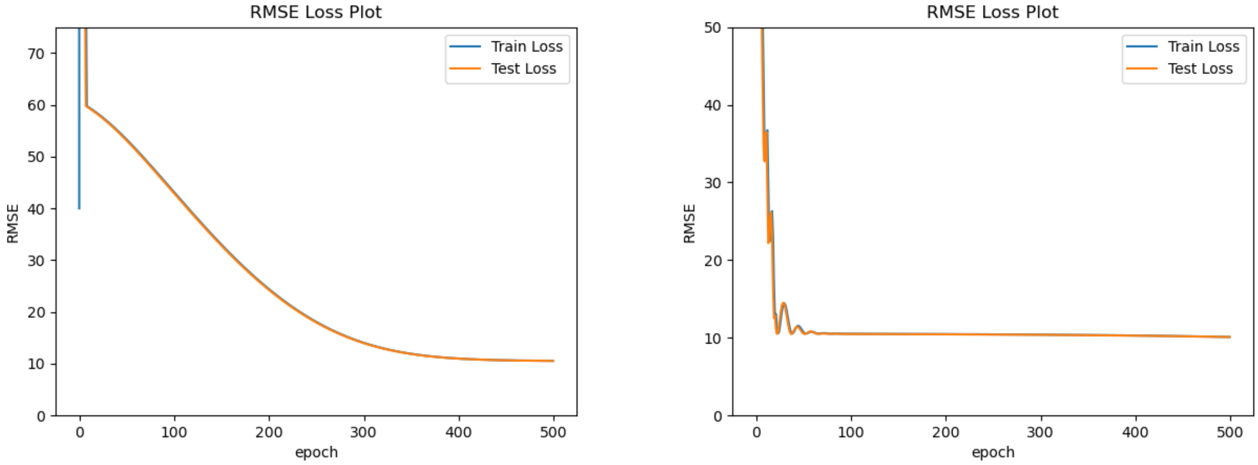

Figure 2: Mention graph GCN Losses (left) and GAT Losses (right)

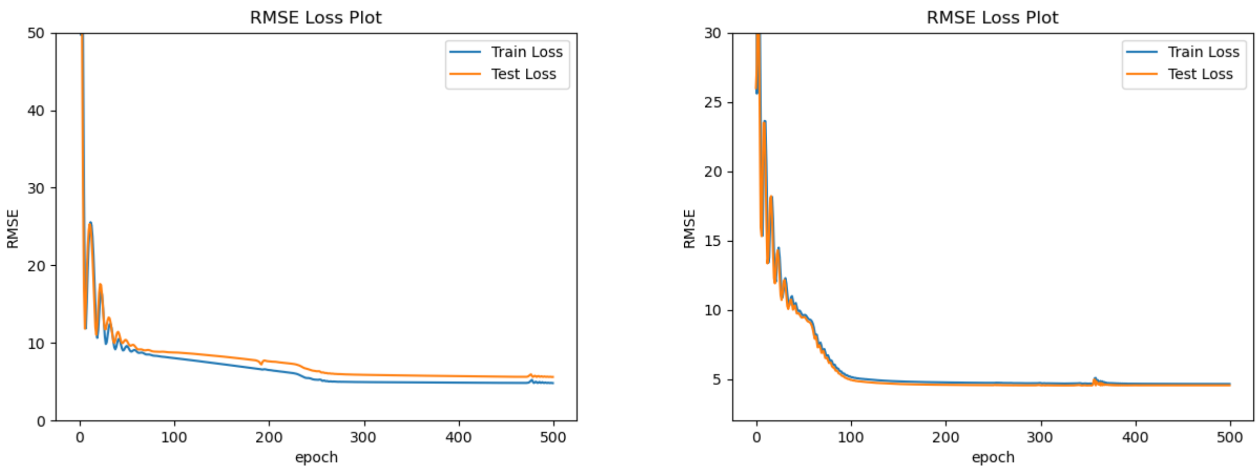

Figure 3: Mention graph GeneralConv Losses for H=16 (left) and H=32 (right)

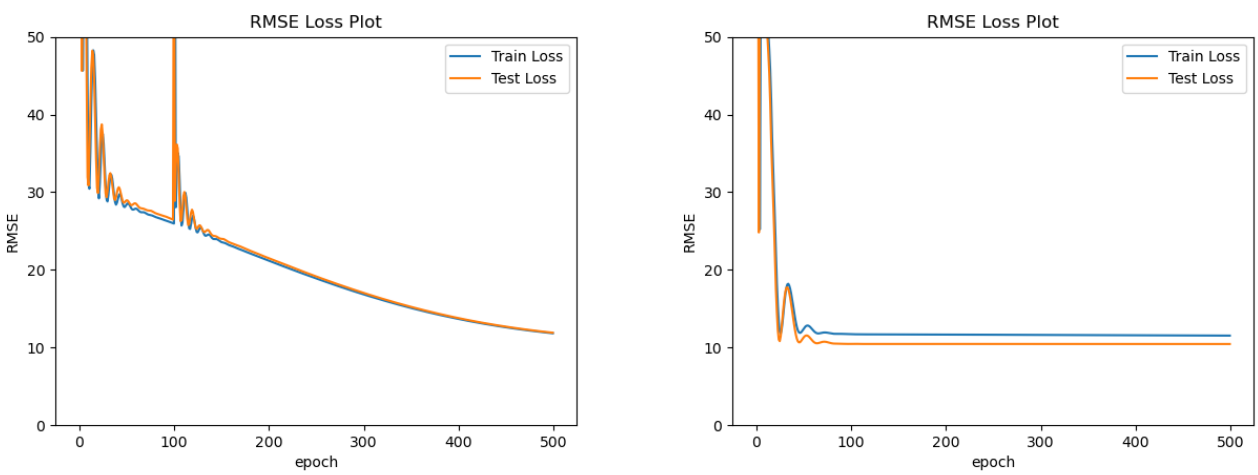

Figure 4: Comment graph GCN Losses (left) and GAT Losses (right)

could potentially lose. Losing weight has more health benefits for people with BMIs in the obese range than for people with BMIs in the normal range, like many BOOHEE users.

Future research could specifically focus on obese individuals who have weight loss surgery. A healthy diet post surgery is necessary for weight loss surgery to be effective (Papalazarou et al. 2012). If social networks can positively influence the lifestyle habits of people who have had weight loss surgery, weight loss apps could be life-changing.

Another research question that could be studied is whether existing recommendation systems for apps like TikTok can learn how much a user weighs based on how they interact with the app. It's possible that certain types of posts or videos are shown more often to people who weigh a certain amount or have recently lost or gained weight. A research experiment could examine data about the posts that are recommended to users and the users' weights, and see if there is a correlation.

Source Repository

`https://github.com/danielshu7/WeightLoss-Network-Analysis`

References

[Carter et al. 2013] Carter, M. C.; Burley, V. J.; Nykjaer, C.; and Cade, J. E. 2013. Adherence to a smartphone application for weight loss compared to website and paper diary: Pilot randomized controlled trial. *Journal of Medical Internet Research* 15(4).

[Consolvo et al. 2008] Consolvo, S.; McDonald, D.; Toscos, T.; Chen, M.; Froehlich, J.; Harrison, B.; Klasnja, P.; LaMarca, A.; LeGrand, L.; Libby, R.; Smith, I.; and Landay, J. 2008. Activity sensing in the wild: A field trial of ubifit garden. 1797–1806.

[Dounavi and Tsoumani 2019] Dounavi, K., and Tsoumani, O. 2019. Mobile health applications in weight management: A systematic literature review. *American Journal of Preventive Medicine* 56(6):894–903.

[Fryar, Carroll, and J 2020] Fryar, C.; Carroll, M.; and J, A. 2020. Prevalence of overweight, obesity, and severe obesity among children and adolescents aged 2–19 years: United states, 1963–1965 through 2017–2018. NCHS Health E-Stats.

[Jane et al. 2018] Jane, M.; Hagger, M.; Foster, J.; Ho, S.; and Pal, S. 2018. Social media for health promotion and weight management: A critical debate. *BMC Public Health* 18(1).

[Kim and Oh 2022] Kim, D., and Oh, A. 2022. How to find your friendly neighborhood: Graph attention design with self-supervision.

[Kim, Kim, and Park 2021] Kim, H. H.; Kim, Y.; and Park, Y. R. 2021. Interpretable conditional recurrent neural network for weight change prediction: Algorithm development and validation study. *JMIR Mhealth Uhealth* 9(3):e22183.

[Kipf and Welling 2016] Kipf, T. N., and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *CoRR* abs/1609.02907.

[Papalazarou et al. 2012] Papalazarou, A.; Yannakoulia, M.; Kavouras, S. A.; Komesidou, V.; Dimitriadis, G.; Papakonstantinou, A.; and S. Sidossis, L. 2012. Lifestyle intervention favorably affects weight loss and maintenance following obesity surgery. *Obesity*.

[Tsai et al. 2007] Tsai, C. C.; Lee, G.; Raab, F.; Norman, G. J.; Sohn, T.; Griswold, W. G.; and Patrick, K. 2007. Usability and feasibility of pmeb: A mobile phone application for monitoring real time caloric balance. *Mobile Networks and Applications* 12(2-3):173–184.

[Veličković et al. 2017] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2017. Graph attention networks.

[Wang et al. 2017] Wang, Z.; Derr, T.; Yin, D.; and Tang, J. 2017. Understanding and predicting weight loss with mobile social networking data. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, 1269–1278. New York, NY, USA: Association for Computing Machinery.

[Ward et al. 2021] Ward, Z. J.; Bleich, S. N.; Long, M. W.; and Gortmaker, S. L. 2021. Association of body mass index with health care expenditures in the united states by age and sex. *PLOS ONE* 16(3):1–10.

[You, Ying, and Leskovec 2020] You, J.; Ying, R.; and Leskovec, J. 2020. Design space for graph neural networks.