

MentalNet: Heterogeneous Graph Representation for Early Depression Detection

Ivan Mihov¹ Haiquan Chen¹ Xiao Qin² Wei-Shinn Ku² Da Yan³ Yuhong Liu⁴

¹California State University, Sacramento, USA

²Auburn University, USA

³University of Alabama at Birmingham, USA

⁴Santa Clara University, USA

{ivanmihov, haiquan.chen}@csus.edu, {xqin, weishinn}@auburn.edu, yanda@uab.edu yhliu@scu.edu

Abstract—Depression is one of the leading factors in global disability and a top driver for suicides. Studies have shown that depression has an effect on language usage. In recent years, especially during the COVID pandemic, social media platforms have become the de facto platform for many individuals to self-disclose or discuss mental health issues like depression. This trend presents a unique opportunity for researchers and healthcare professionals to detect potential mental illnesses for early intervention or treatment by taking advantage of the recent advances in machine learning approaches. Existing depression detection methods on social media, however, suffer from two major limitations. First, these solutions heavily rely on the amount, quality, and type of user-posted content. Second, the overlooked social circle impact should be leveraged to enhance the prediction capabilities. In this paper, we propose a depression detection framework, MentalNet, based on heterogeneous graph convolution by capturing users’ interactions (replies, mentions, and quote-tweets) with their friends on social media and differentiating the intimacy of users’ social circles (e.g., family, friends, or acquaintances). Specifically, we formulate the problem of depression detection on social media as a graph classification problem by representing users’ social circles in the format of heterogeneous graphs. MentalNet embraces three modules, (1) extraction of ego-network node features, (2) construction of user interaction graphs, and (3) depression detection based on heterogeneous graph classification. The extensive experiments on Twitter data demonstrate that MentalNet consistently and significantly outperforms the state-of-the-art methods in terms of all the effectiveness metrics. Compared to the baseline methods, MentalNet is able to effectively predict early depression in Twitter users with up to 24% improvement on F1 score.

Index Terms—Representation Learning, Heterogeneous Graph, Depression Detection.

I. INTRODUCTION

Depression is one of the leading factors in global disability and a top reason for suicides. It has been observed that people are increasingly relying on social media platforms such as Twitter to share emotions and personal messages. The content originated by those users and the language used in their posts unveil not only their daily lives but also insightful mental states. In the past, researchers have applied traditional machine learning models to predict depression using social media [1]–[5] with a reasonable accuracy. More recent works have focused on deep learning approaches by utilizing the natural language processing methods [6]–[9]. Though various

approaches exist for depression detection in social media, the current solutions suffer from the following limitations. First, the techniques heavily rely on the amount, quality, and type of user-posted content as training data. Second, the overlooked social circle impact ought to be leveraged to enhance prediction capabilities. We advocate for the exploitation of the relationship (interactions) between users on social media to support accurate and early detection of depression.

Our motivation and contributions. It has been revealed that user interactions on social media provide useful insights on the emotional states of users [10]. In addition, social media not only contributes to depression states, but may be the sole cause of the states [11]. A correlation between depressed users and their ego network has been thoroughly investigated; a user is 93% more likely to be depressed if one of her immediate friends is depressed [12]. Such relational data about users’ ego networks, of course, are non-trivial to capture and analyze. Motivated by the above observations, we propose MentalNet for early detection of depression by capturing the relational data in users’ ego-networks using heterogeneous graph convolution. MentalNet builds on three common types of users’ interactions with their friends on social media, namely, replies, mentions, and quote-tweets. To the best of our knowledge, our work pilots the first concept among its peers to formulate the problem of depression detection on social media as a graph classification problem by representing users’ social circles as heterogeneous graphs. We summarize our contributions as follows:

- We formulate the problem of depression detection to a heterogeneous graph classification problem and propose a novel solution, MentalNet, anchored on heterogeneous graph convolution by capturing users’ various interactions with their friends on social media and differentiating the intimacy of users’ social circles.
- We create a new depression repository PsycheNet-G by injecting user’ various interactions on social media into PsycheNet [13], the first social contagion-based depression repository.
- To capture the linguistic features exhibited in users’ social posts, we design an LSTM autoregressive sequence-to-

sequence autoencoder to extract the latent features in written posts as users' features in their ego-networks.

- To distinguish between the types of interactions which form the relationships between users and their friends, we construct a graph with heterogeneous edges that represent various types of user interactions on social media, namely, replies, mentions, and quote-tweets. Aiming to denoise the learning process with multiple graph convolution operations, we encode the adjacency matrix of users' each interaction in the format of a doubly stochastic matrix.
- To effectively compare the heterogeneous user interaction graphs involving different numbers of friends, we employ sort pooling [14] to assign friends with similar structural roles in their respective graphs similar positions for subsequent graph convolution operations.
- The extensive experiments using real-world Twitter data verify that MentalNet has a leading edge over the state-of-the-art methods in terms of all the effectiveness measures. Compared to the baseline methods, MentalNet is able to effectively predict early depression in Twitter users with up to 24% improvement on F1 score. The TensorFlow implementation of MentalNet is publicly available¹.

The rest of the paper is organized as follows. In Section 2, we describe PsycheNet-G, the new depression data repository. In Section 3, we elaborate on the MentalNet design. In Section 4, we empirically compare MentalNet with the state-of-the-art methods. We review the related work in Section 5 and conclude the paper in Section 6.

II. AUGMENTATION OF PSYCHENET DATASET

PsycheNet-G. In this study, we extended the PsycheNet [13] dataset by injecting the information about bidirectional replies, mentions, and quotes-tweets between each user and any of her friends. We used Twitter Search API to re-fetch the raw tweets for each user and her friends. All the depression self-diagnostic tweets, as well as their corresponding Twitter timelines, in PsycheNet were further examined by a domain expert for their genuineness to verify the depressed users as ground truth. After that, we excluded those self-diagnostic tweets identified as fake. Table I shows some sample self-diagnostic tweets identified as genuine and fake, respectively. In addition, for every tweet of each user, we examine whether the tweet was in reply to a friend of the user, and whether that friend had any tweets in reply from the user. We summed all of their reply counts to acquire the number of bidirectional replies between each user and her friends. Also, we searched all tweets of each user for mentions of any of her friends, and all friends' tweets for mentions of that user. If both a user and her friend had mentioned each other, we aggregated their mentions as a bidirectional mention count. Lastly, we handled quote-tweets, which are retweets with an added comment and allowed us to capture a more direct communication between users and their

friends. Similarly, we counted the bidirectional quote-tweets between each user and her friends. We named our augmented PsycheNet dataset as PsycheNet-G. Table II reports the basic statistics of PsycheNet-G.

III. PROPOSED METHOD

A. The MentalNet Architecture

MentalNet casts the problem of depression detection on social media into a graph classification problem on heterogeneous graphs. Figure 1 depicts the architecture of MentalNet, which embraces three modules: (1) extraction of ego-network node features, (2) construction of user interaction graphs, and (3) depression detection based on heterogeneous graph classification. First, to capture the linguistic characteristics from user tweets, we train an LSTM autoregressive sequence-to-sequence autoencoder to extract the latent features in written posts as users' node features in their ego-networks. Second, we create for each user an ego-network based on each specific type of user interactions (replies, mentions, and quote-tweets), followed by merging the three resulting ego-networks together to build a heterogeneous graph, where each edge (u, v) is represented as a 3-dimensional vector encoded as the numbers of bidirectional replies, mentions, or quote-tweets between u and v . Third, we feed such heterogeneous ego-network graph to a deep graph convolution network with sort pooling [14] to furnish heterogeneous graph classification.

B. Node Feature Extraction

To capture the linguistic characteristics of users and their friends, we applied an LSTM autoregressive sequence-to-sequence autoencoder [15] to extract the latent features from their posts as their embeddings. In our autoencoder, both encoder and decoder were bidirectional-LSTM networks. Such autoencoder was trained to minimize the reconstruction loss (categorical crossentropy) with the softmax as the final layer and the Adam optimizer to reconstruct the input tweet sequence. The embedding averaged over all the tweets written by each particular user/friend obtained using the encoder-decoder model was then treated as the node features in each user's respective ego-network graph.

C. User Heterogeneous Graph Construction

Ego-network creation with multiple edge types. For each user u , we create her/his ego-network graph for each specific interaction type where each edge attribute value is set as the number of interactions in that type between u and a friend. The embeddings obtained from the LSTM autoencoder (as discussed in the previous subsection) are adopted as the node features. We then merge these three ego-network graphs for each user into a heterogeneous graph with each edge being a 3-dimensional vector representing the strengths of different interaction types (replies, mentions, or quote-tweets). Given a user's information, we only keep those friends she/he interacts with. In other words, if a friend is missing all the three interaction types with the associated user, that friend is pruned.

¹<https://github.com/anonymous>

TABLE I: Samples of unedited depression self-diagnostic tweets in PsycheNet-G

Genuine	Umm yes! I have been diagnosed with depression in October and this pandemic ain't helping! I've been diagnosed with depression, and i dont take meds. I've been diagnosed with depression for five years and dealt with it for about ten.
Fake	I've been diagnosed with depression by TikTok Interesting.. you just described me I've been diagnosed with depression but perhaps it's not! Playing bit life and I've been diagnosed with depression at age 3 f*** that

TABLE II: Basic statistics of PsycheNet-G dataset

	Users				Friends per user				Replies per user				Mentions per user				Quote-Tweets per friend				
	Sum	Min	Max	Mean	Sum	Min	Max	Mean	Sum	Min	Max	Mean	Sum	Min	Max	Mean	Sum	Min	Max	Mean	Sum
Depressed group	242	1	518	63	15,321	1	19,414	1,379	333,799	1	17,885	891	215,722	1	461	35	8,470				
Control group	349	1	448	89	31,103	1	132,907	3,119	1,088,470	1	63,532	1,804	629,437	1	22,328	283	98,700				

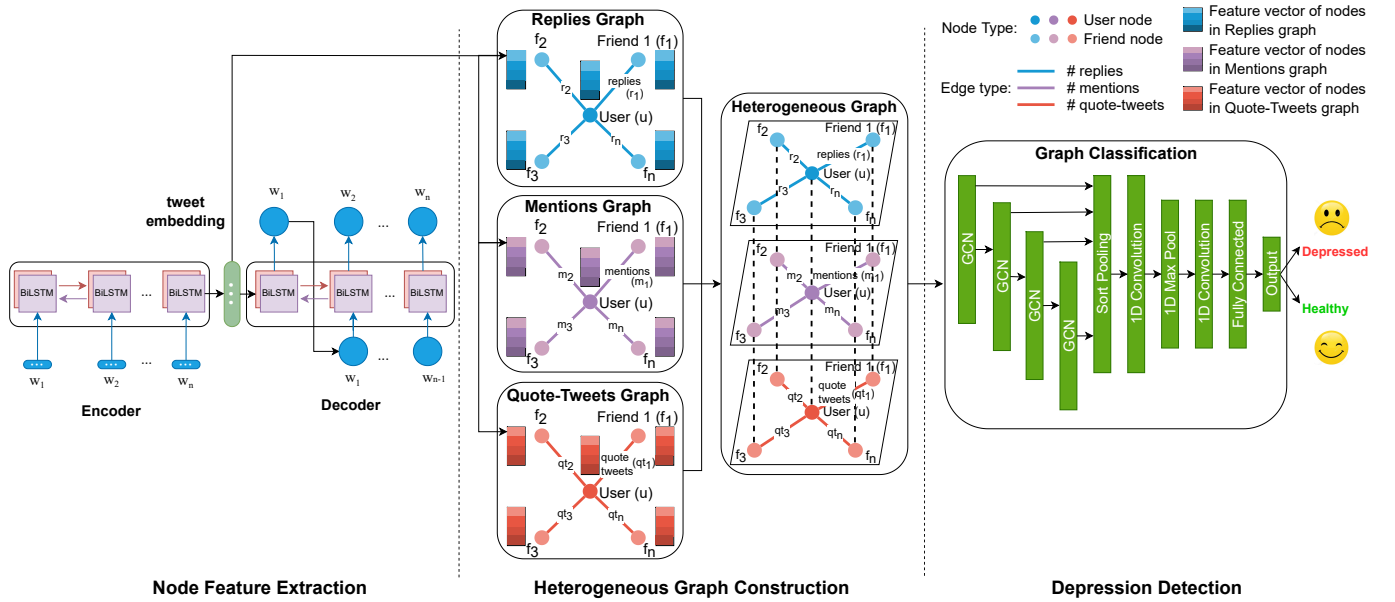


Fig. 1: The architecture of MentalNet, where a heterogeneous ego-network graph is constructed by merging three types of user-friend interactions (replies, mentions, quote-tweets). MentalNet consists of (1) node feature extraction using an LSTM autoregressive sequence-to-sequence autoencoder, (2) construction of heterogeneous user interaction graphs, and (3) depression detection based on heterogeneous graph classification.

Doubly stochastic normalization for heterogeneous graph stabilization. Since the graph convolution is repeated multiple times on constructed heterogeneous graphs in MentalNet, in order to stabilize the subsequent consecutive graph convolutions, we apply doubly stochastic normalization [16] to the constructed heterogeneous graphs before feeding them to the graph convolutional networks. Traditionally, an adjacency matrix used in graph convolution operation [17] is a positive-valued matrix where each element has one-dimensional edge feature, which models the correlation between each node pair in a graph. In our design, given a graph with N nodes, we denote X as an $N \times F$ tensor representation of the node features of the graph and X_{ij} be the value of the j^{th} feature of

the i^{th} node. Also, let E be an $N \times N \times K$ tensor representation of edge features in the graph, E_{ij} be the K -dimensional edge feature vector connecting the i^{th} and the j^{th} nodes, and E_{ijk} be the k^{th} dimension of the edge feature E_{ij} . We convert each matrix E_{**k} , $k = 1, 2, \dots, K$ as a doubly stochastic matrix E'_{**k} as follows:

$$\hat{E}_{ijk} = \frac{E_{ijk}}{\sum_{p=1}^N E_{ipk}} \quad (1)$$

$$E'_{ijk} = \sum_{p=1}^N \frac{\hat{E}_{ipk} \hat{E}_{jp k}}{\sum_{u=1}^N \hat{E}_{upk}}$$

It is observed that normalized matrix E'_{**k} is both left

stochastic and right stochastic (i.e., E'_{**k} is a doubly stochastic matrix). A doubly stochastic matrix is a square matrix of non-negative real numbers, each of whose rows and columns sums to 1. Thus, E'_{**k} meets the following properties:

$$\begin{aligned} E'_{ijk} &\geq 0 \\ \sum_{i=1}^N E'_{ijk} &= \sum_{j=1}^N E'_{ijk} = 1. \end{aligned} \quad (2)$$

In MentalNet, we operate E'_{**k} , $k = 1, 2, \dots, K$ as the adjacency matrix in each graph convolution operation instead of using the edge feature matrix E_{**k} based on the symmetric normalization [14], [17], [18]. Consequently, the resulting Markov chain with such doubly stochastic transition matrix helps to denoise the subsequent convolution learning process by ensuring a uniform stationary distribution [16].

D. Graph Convolution and Sort Pooling

Recently a few works have been proposed to perform graph classification using graph convolution [14], [17], [18]. Here we articulate the graph convolution and sort pooling operations in MentalNet.

Graph convolution for substructure feature extraction. Let G be an undirected graph of $n \in \mathbb{N}$ nodes and an adjacency matrix $\mathbf{A} \in \mathbb{Z}^{n \times n}$. An augmented adjacency matrix of G is defined as $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$. For row-wise normalization, a diagonal degree matrix of $\mathbf{D}$ is defined as $\tilde{\mathbf{D}}_{i,i} = \sum_j \tilde{\mathbf{A}}_{i,j}$. A matrix of learnable parameters is defined as $\mathbf{W} \in \mathbb{R}^{c \times c'}$, $c' \in \mathbb{N}$, where c' is the number of output feature channels. Suppose that the node feature matrix is $\mathbf{X} \in \mathbb{R}^{n \times c}$, $c \in \mathbb{N}$ and a non-linear activation function is denoted by f . The graph convolution operation can be defined as follows:

$$\mathbf{Z} = f\left(\tilde{\mathbf{D}}^{-1} \tilde{\mathbf{A}} \mathbf{X} \mathbf{W}\right), \quad (3)$$

where graph convolution aggregates information about the local substructure by smoothing information about nodes in the local neighborhood. To extract the multi-scale substructure features, in MentalNet, multiple graph convolution layers are stacked, as follows:

$$\mathbf{Z}^{(t+1)} = f\left(\tilde{\mathbf{D}}^{-1} \tilde{\mathbf{A}} \mathbf{Z}^{(t)} \mathbf{W}^{(t)}\right), \quad (4)$$

where $\mathbf{Z}^{(0)} = \mathbf{X}$ and $\mathbf{W}^{(t)} \in \mathbb{R}^{c_t \times c_{t+1}}$.

To extract different substructure features for graph classification, MentalNet consecutively employs Weisfeiler-Lehman subtree kernel [19] during the graph convolution process. Also, in MentalNet, the outputs from consecutive graph convolution layers, $\mathbf{Z}^{(t)}$, $t = 1, \dots, h$, are horizontally concatenated in the form of a tensor $\mathbf{Z}^{(1:h)}$, where h is the number of the graph convolutional layers, before sort pooling.

Sort pooling for effective graph comparison. To ensure that the 1-dimensional convolutional filters are applied to different ego-network graphs consistently, MentalNet orchestrates sort pooling [14] to sort all nodes in a graph based on their embeddings obtained from the graph convolution process. In

the sort pooling layer, the node embeddings are sorted and pooled so that the nodes with similar structural roles are assigned similar positions in their respective graphs before they are fed to 1-dimensional convolution layers for graph comparison. Specifically, the sort pooling takes as input tensor $\mathbf{Z}^{(1:h)}$, where each row is a node's feature descriptor and each column is a feature channel. The input tensor is sorted row-wise according to the last channel of $\mathbf{Z}^{(h)}$ in descending order. The layer's output is a $k \times \sum_{t=1}^h c_t$ tensor, where k is a user-defined parameter. Also, the sort pool layer extends (by using zero-padding) or truncates the output tensors in the first dimension from n to k to make graphs with different sizes (numbers of nodes) to share the same size (k nodes) for effective graph comparison [14].

IV. EXPERIMENTAL VALIDATION

In this section, we empirically show the superiority of MentalNet over the state of the art. To ensure that there was no data leakage, we partitioned all users in the depression and control groups, respectively, into training and test sets with a split of 70% and 30% and removed any friend who is a friend of a user in the training set and also a friend of a user in the test set. All the experiments were conducted with Python v3.7, TensorFlow v2.6.0, and StellarGraph v1.2.1 on Google Colab (with Tesla P100 GPU and 25 GB RAM). The default architecture of MentalNet consists of four graph convolutional layers, a sort pooling layer, two 1-dimensional convolutional layers, two max pooling layers, a fully-connected layer, and an output layer with a binary classification head.

A. Baseline Methods for Comparison

We implemented the following baseline methods for performance comparison:

- **CNNWithMax** [7]. This method extracts word embeddings from users' concatenated tweets, which are then fed to an one-dimensional CNN architecture.
- **CNNWithMax+**. This is an extended version of CNNWithMax, incorporating the tweets from users' friends.
- **CNNFastText** [20]. This methods extracts the word embeddings from users' concatenated tweets, which are then fed to a two-dimensional CNN architecture.
- **CNNFastText+**. This is an extended version of CNNFastText, incorporating the tweets from users' friends.
- **LSTMFastText** [21]. This methods extracts the word embeddings from users' concatenated tweets, which are then fed to three stacked Long Short Term Memory (LSTM) layers.
- **LSTMFastText+**. This is an extended version of LSTMFastText, incorporating the tweets from users' friends.
- **MentalSpot** [13]. The method employs contrastive learning and implements dynamic mean shift pruning to identify the most homogeneous friends. The tweets from each user and her most homogeneous friends are used for depression detection. We retrieved the top-5 most homogeneous friends for each user.

TABLE III: Performance comparison by varying the number of friends per user (50 tweets per friend).

	5 Friends						9 Friends						All Friends					
	Healthy (0)			Depressed (1)			Healthy (0)			Depressed (1)			Healthy (0)			Depressed (1)		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CNNWithMax	0.78	0.79	0.78	0.75	0.72	0.73	0.78	0.79	0.78	0.75	0.72	0.73	0.78	0.79	0.78	0.75	0.72	0.73
CNNWithMax+	0.79	0.75	0.77	0.75	0.71	0.73	0.77	0.65	0.70	0.70	0.64	0.67	0.75	0.76	0.75	0.71	0.70	0.70
CNNFastText	0.78	0.74	0.76	0.76	0.68	0.72	0.78	0.74	0.76	0.76	0.68	0.72	0.78	0.74	0.76	0.76	0.68	0.72
CNNFastText+	0.75	0.71	0.73	0.71	0.67	0.69	0.71	0.71	0.71	0.70	0.68	0.69	0.79	0.69	0.74	0.72	0.70	0.71
LSTMFastText	0.76	0.74	0.75	0.73	0.69	0.71	0.76	0.74	0.75	0.73	0.69	0.71	0.76	0.74	0.75	0.73	0.69	0.71
LSTMFastText+	0.77	0.74	0.75	0.73	0.69	0.71	0.76	0.67	0.71	0.68	0.65	0.66	0.73	0.70	0.71	0.67	0.58	0.62
MentalSpot	0.80	0.79	0.80	0.77	0.75	0.76	0.82	0.80	0.81	0.78	0.74	0.76	0.75	0.71	0.73	0.71	0.65	0.66
MentalNet	0.81	0.80	0.80	0.74	0.70	0.72	0.86	0.80	0.83	0.81	0.77	0.79	0.89	0.88	0.89	0.87	0.84	0.86

TABLE IV: Performance comparison by varying the number of tweets per user/friend (all friends considered per user).

	5 Tweets						50 Tweets						100 Tweets					
	Healthy (0)			Depressed (1)			Healthy (0)			Depressed (1)			Healthy (0)			Depressed (1)		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CNNWithMax	0.75	0.64	0.69	0.72	0.62	0.67	0.78	0.79	0.78	0.75	0.72	0.73	0.77	0.75	0.76	0.74	0.71	0.72
CNNWithMax+	0.75	0.62	0.68	0.72	0.63	0.67	0.75	0.76	0.75	0.71	0.70	0.70	0.73	0.75	0.74	0.70	0.68	0.69
CNNFastText	0.68	0.64	0.66	0.66	0.65	0.65	0.78	0.74	0.76	0.76	0.68	0.72	0.78	0.73	0.75	0.74	0.67	0.72
CNNFastText+	0.70	0.65	0.67	0.63	0.64	0.63	0.79	0.69	0.74	0.72	0.70	0.71	0.73	0.69	0.71	0.73	0.68	0.70
LSTMFastText	0.72	0.68	0.70	0.70	0.66	0.68	0.76	0.74	0.75	0.73	0.69	0.71	0.77	0.70	0.73	0.73	0.68	0.70
LSTMFastText+	0.69	0.68	0.68	0.67	0.56	0.61	0.73	0.70	0.71	0.67	0.58	0.62	0.70	0.66	0.68	0.67	0.56	0.61
MentalSpot	0.71	0.65	0.68	0.65	0.62	0.64	0.75	0.71	0.73	0.71	0.65	0.66	0.71	0.68	0.69	0.68	0.60	0.64
MentalNet	0.79	0.80	0.79	0.76	0.75	0.75	0.89	0.88	0.89	0.87	0.84	0.86	0.78	0.78	0.78	0.75	0.74	0.74

B. Performance Comparison

Impact of the number of friends per user. Table III compares the performance of MentalNet with the baseline methods in terms of precision, recall, and F1 scores by varying the number of friends per user (we fixed the number of tweets per friend to 50). As shown in Table III, MentalNet outperformed all the baseline methods in most cases. When the number of friends was 5, MentalNet showed similar performance with MentalSpot since the same group of friends were taken into account for each user. For the depressed group, the performance scores were 0.74, 0.70, and 0.72, respectively, for MentalNet and 0.77, 0.75, and 0.76, respectively, for MentalSpot. When n was increased to 9, MentalNet consistently outperformed the baseline methods with precision, recall, and F1 scores of 0.86, 0.80, and 0.83, respectively, for the healthy group, and 0.81, 0.77, and 0.79 for the depressed group. When we utilized all friends for each user, MentalNet yielded precision, recall, and F1 scores of 0.89, 0.88, 0.89, respectively, for the healthy group, and 0.87, 0.84, and 0.86, respectively, for the depressed group, which was a significant improvement over all baseline methods. This can be explained by the fact that MentalNet utilizes graph neural networks to capture users' interactions with their friends on social media and therefore is able to differentiate the intimacy of users' different social circles. On the other hand, simply increasing the number of friends for all baseline methods will sometimes deteriorate the prediction results since not every friend equally contributes to a user's mental status.

Impact of the number of tweets per user/friend. Table IV compares the performance of MentalNet with the baseline methods in terms of precision, recall, and F1 scores by varying the number of tweets per friend (we used all the friends for each user). As shown in Table IV, MentalNet consistently outperformed all the baseline methods in all cases. When m was set to 5, MentalNet yielded precision, recall, and F1 scores of 0.79, 0.80, and 0.79, respectively, for the healthy group, and 0.76, 0.75, and 0.75, respectively, for the depressed group. With m set to 50, MentalNet significantly outperformed the baseline methods with the precision, recall, and F1 score of 0.89, 0.88, and 0.89, respectively, for the healthy group, and 0.87, 0.84, and 0.86, respectively, for the depressed group. The best performance of MentalNet was achieved for both the healthy group and depressed group when the number of tweets was set to 50. The reason could be that using 50 most recent tweets of all friends until the self-diagnosis point is a more accurate estimate for a user's mental status due to the dynamic nature of depression over time, since a tweet published a long time ago might not indicate the user's current depression status.

V. RELATED WORK

Since traditional machine learning approaches [1]–[5] for depression detection using social media data heavily rely on feature engineering, a handful of deep learning-based approaches to detecting depression have been proposed. Yates *et al.* advocated for a convolutional network model to detect and identify depressed users based on textual information [6].

In [7], the CBOW and skip-gram models were used to capture users' linguistic features, which were then fed into different convolutional and BiLSTM network architectures. Gui *et al.* piloted a multi-agent model to predict depression in users by analyzing both text and relevant images, where a GRU network facilitates feature extraction from text and VGG-Net was applied to extract visual features from images [8]. In [9], a convolutional network was used to extract linguistic features from users' tweets while a GRU-based network was employed to deal with the temporal and semantic modalities. In [22], a multi-modal approach was proposed to leverage text, image, and behaviour features to originate an aggregated depression score for each post on Instagram, where a convolutional network was created for images, a recurrent neural network was applied for text, and a random forest model was utilized as the behavior classifier. Lin *et al.* [23] designed a multi-modal approach by using BERT to extract text features and convolutional networks to extract visual features from images. Unfortunately, all the aforementioned methods suffer from two major limitations. First, these solutions heavily depend on the amount, quality, and type of user posted content, which requires a significant amount of user generated content (tweets). Second, the existing techniques ignore interactions between users on social media, which ought to be leveraged to enhance the prediction capabilities to support accurate and early detection of depression. Pirayesh *et al.* [13] proposed MentalSpot, which is primed to identify the homogeneous friends and to explore social posts from homogeneous friends to detect depression. However, MentalSpot fails to capture users' various interactions with their friends on social media, thereby being unable to differentiate the intimacy of users' multiple social circles.

VI. CONCLUSION

In this study, we propose a novel depression detection framework, MentalNet, anchored on heterogeneous graph convolution by capturing users' interactions on social media and differentiating the intimacy of users' social circles. We formulate the problem of depression detection as a graph classification problem by representing each user as her heterogeneous graph with different interaction types with her friends. Extensive experiments on the Twitter data demonstrate that MentalNet significantly outperforms the state-of-the-art methods in terms of all the effectiveness metrics.

ACKNOWLEDGEMENT

We thank Dr. Nuo Yang (University of Rochester Medical Center, Rochester, NY) for evaluating and verifying depression self-diagnostic tweets in PsycheNet-G.

REFERENCES

- [1] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 858–862.
- [2] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," *Icwsn*, vol. 13, pp. 1–10, 2013.
- [3] D. Preotjuc-Pietro, J. Eichstaedt, G. Park, M. Sap, L. Smith, V. Tobolsky, H. A. Schwartz, and L. Ungar, "The role of personality, age, and gender in tweeting about mental illness," in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, Jun. 5 2015.
- [4] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.
- [5] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in Twitter," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Jun. 2014.
- [6] A. Yates, A. Cohan, and N. Goharian, "Depression and self-harm risk assessment in online forums," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Sep. 2017.
- [7] A. Hussein Orabi, P. Buddhitha, M. Hussein Orabi, and D. Inkpen, "Deep learning for depression detection of Twitter users," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, New Orleans, LA, Jun. 2018.
- [8] T. Gui, L. Zhu, Q. Zhang, M. Peng, X. Zhou, K. Ding, and Z. Chen, "Cooperative multimodal approach to depression detection in twitter," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 110–117, Jul. 2019.
- [9] H. Zogan, X. Wang, S. Jameel, and G. Xu, "Depression detection with multi-modalities using a hybrid deep learning model on social media," *CoRR*, vol. abs/2007.02847.
- [10] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Echo chambers: Emotional contagion and group polarization on facebook," *Scientific reports*, vol. 6, p. 37825, 2016.
- [11] M. Hunt, J. Young, R. Marx, and C. Lipson, "No more fomo: Limiting social media decreases loneliness and depression," *Journal of Social and Clinical Psychology*, vol. 37, pp. 751–768, 11 2018.
- [12] J. N. Rosenquist, J. H. Fowler, and N. A. Christakis, "Social network determinants of depression," *Molecular psychiatry*, vol. 16, no. 3, pp. 273–281, 2011.
- [13] J. Pirayesh, H. Chen, X. Qin, W. Ku, and D. Yan, "MentalSpot: Effective early screening for depression based on social contagion," in *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*. ACM, 2021, pp. 1437–1446.
- [14] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, "An end-to-end deep learning architecture for graph classification," in *AAAI*, 2018.
- [15] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems, December 8-13 2014, Montreal, Quebec, Canada*, 2014, pp. 3104–3112.
- [16] B. Wang, A. Pourshafeie, M. Zitnik, J. Zhu, C. D. Bustamante, S. Batzoglou, and J. Leskovec, "Network enhancement as a general method to denoise weighted biological networks," *Nature Communications*, vol. 9, Aug 2018.
- [17] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017.
- [18] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," 2019.
- [19] N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *Journal of Machine Learning Research*, vol. 12, no. 77, pp. 2539–2561, 2011.
- [20] M. Troczek, S. Koitka, and C. Friedrich, "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, pp. 588–601, 04 2018.
- [21] J. Wolohan, "Estimating the effect of COVID-19 on mental health: Linguistic indicators of depression during a global pandemic," in *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Jul. 2020.
- [22] C. Chiu, H.-Y. Lane, J.-L. Koh, and A. Chen, "Multimodal depression detection on instagram considering time interval of posts," *Journal of Intelligent Information Systems*, vol. 56, 02 2021.
- [23] C. Lin, P. Hu, H. Su, S. Li, J. Mei, J. Zhou, and H. Leung, "Sensemood: Depression detection on social media," in *Proceedings of the 2020 International Conference on Multimedia Retrieval*, ser. ICMR '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 407–411.