Emotional and Linguistic Cues of Depression from Social Media

Nikhita Vedula
Department of Computer Science and Engineering
Ohio State University
vedula.5@osu.edu

Srinivasan Parthasarathy
Department of Computer Science and Engineering
Ohio State University
srini@cse.ohio-state.edu

ABSTRACT

Health outcomes in modern society are often shaped by peer interactions. Increasingly, a significant fraction of such interactions happen online and can have an impact on various mental health and behavioral health outcomes. Guided by appropriate social and psychological research, we conduct an observational study to understand the interactions between clinically depressed users and their ego-network when contrasted with a differential control group of normal users and their ego-network. Specifically, we examine if one can identify relevant linguistic and emotional signals from social media exchanges to detect symptomatic cues of depression. We observe significant deviations in the behavior of depressed users from the control group. Reduced and nocturnal online activity patterns, reduced active and passive network participation, increase in negative sentiment or emotion, distinct linguistic styles (e.g. self-focused pronoun usage), highly clustered and tightly-knit neighborhood structure, and little to no exchange of influence between depressed users and their ego-network over time are some of the observed characteristics. Based on our observations, we then describe an approach to extract relevant features and show that building a classifier to predict depression based on such features can achieve an F-score of 90%.

ACM Reference format:
Nikhita Vedula and Srinivasan Parthasarathy. 2017. Emotional and Linguistic Cues of Depression from Social Media. In *Proceedings of DH '17, London, United Kingdom, July 02-05, 2017,* 10 pages.
https://doi.org/http://dx.doi.org/10.1145/3079452.3079465

1 INTRODUCTION

The health and developmental outcomes in modern society are often shaped by peer interactions. Relationships with family and caregivers during childhood and additionally peers from adolescence to adulthood, are critical to understanding both dimensions of well being and sources of risk during different phases of life [38]. An increasing amount of such interaction happens online via various social media platforms such as Facebook and Twitter. In fact as noted by a recent report from the American Academy of Pediatricians (AAP) [27] and echoed by a recent Pew study [32], social media interactions now represent a key communication modality for the vast majority of US adolescents and young adults and a significant fraction of older adults. As described in the AAP report, benefits of social media include providing "enhanced communication and even technical skills, opportunities for community engagement, collective creativity, diversification of friendships extending beyond the physical neighborhood". However, such online interactions can have a drastic impact on various mental health and behavioral health outcomes such as depression, stress, cyberbullying and even violence [3, 37].

Given the pervasive use of social media, and evidence presented by such studies and reports, a key question then to ask is *whether such use departs significantly from those found in physical (offline) social networks*, as studied by sociologists and psychologists for many decades. We note that it is not our goal here to interject ourselves in the midst of an active debate on whether online social media use displaces (potentially higher quality) offline network interactions and engenders unique risks with online activity (for example, cyberbullying) or alternatively enhances benefits by potentially providing additional sources of social capital beyond existing offline interactions [39]. Having said that, the current study does offer some perspectives on this debate (see key findings below). To be more precise, guided by appropriate social theory, we seek to engage in a simpler question and ask if modern online social media communication exhibits similar patterns of behavior to previously reported studies on offline social engagement. Specifically, our goal in this observational study is to study such effects on a subset of the US population that is active on Twitter, paying particular attention to interactions between clinically depressed individuals and their ego-network. We also seek to examine if one can identify relevant signals from social media activity, engagement and linguistic content to detect symptomatic cues of depression. We believe that studies like this one, which build on and serve to deepen our understanding of previous efforts [7, 13], and represent an important step towards understanding the impact of social media use on mental health and well-being.

To facilitate the analysis of our observational study, we examine network effects related to participation, engagement and ego-neighborhood. We define network participation features to include both passive (tweets a user is exposed to, retweets or mentions a user receives) and active participation (mentions, retweets and conversations made by the user). We define network experience features to include both content (e.g., linguistic cues, emotion) and relational dynamics (e.g., conflict/support, influence) of network embedded interactions. We also examine neighborhood effects and analyze key statistics of the neighborhood such as size, centrality and affinity to form clusters or communities. Our study includes both depressed users and their ego-net(s) as well as normal users (control) and their ego-net(s). Some of our key findings include:

DH '17, July 02-05, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5249-9/17/07…$15.00
https://doi.org/http://dx.doi.org/10.1145/3079452.3079465

- With respect to participatory statistics, users suffering from depression tend to: i) post less frequently as well as later in the evening when compared to their normal counterparts, which agrees with offline studies; ii) have smaller networks with densely clustered pockets; iii) less frequently refer explicitly to their network partners online via retweets and mentions; and iv) have a slightly lower regional entropy for their ego-network (i.e. higher co-location within ego-net) as compared to the non-depressed class of users (see Table 1).
- With respect to engagement and responsiveness, the results largely agree with offline studies, that users suffering from depression are less engaged with their network, and neither influence nor are influenced by their network to any significant extent (see Table 2). However, we do find depressed users receiving reasonable social capital from their online neighbors in terms of reacting to their state of mind via supportive tweets (see Figure 1).
- There is a strong presence of linguistic cues such as self-focused pronoun usage by depressed users online, supporting various offline studies. The differential analysis with respect to the normal class of users is particularly stark and compelling (see Figure 3).
- A majority of depressed users exhibit strong negative emotion in their tweets while their general ego-network tends to be positive (see Figure 5). Moreover, for most depressed users we do not observe much periodicity in their emotional signal with respect to days of the week. On the other hand, non-depressed users tend to be more positive and correlated with their ego-network, showing typical trends of being less positively valenced in their posts during the start of the work week and more positive towards the weekend, agreeing with psychological theory (see Figure 4).

The cues above allow us to build a feature set comprising network, content and user based features which illustrates that in an unsupervised setting, normal users and depressed users are separable. In a supervised setting, we find that using the same feature set one can build a classifier to classify depressed individuals, which achieves an F1-score of 90%.

2 DATA COLLECTION

Before describing our methodology and key questions we seek to study, we discuss our data collection methodology. We emphasize here that our study is observational with no direct engagement with Twitter users [1].

To identify candidate depressed Twitter users for our study, we first collected a set of terms commonly used in conjunction with the word 'depression' from the depression glossary at *www.webmd.com*. We then crawled the Twitter streaming API to extract a sample of tweets mentioning any of these terms, only retaining tweets from regions in the US. After filtering out user accounts providing depression-related medical help, we identified candidate users mentioning such terms frequently. Within this subset we then focused on users who explicitly reported being on anti-depression medication; the names of pharmaceutical drugs typically used to treat clinical depression in the US were obtained from a collaborator. Fifty such users spread across the US were then identified as our ground-truth depressed user class.

[1]OSU's Office of Responsible Research has determined that this study does not meet the US federal definition of human subjects research requiring review and neither IRB review nor exemption review is required. This determination is issued under The Ohio State University's OHRP Federal wide Assurance #00006378.

Table 1: Participatory Statistics of Users' Ego Network. Measure values are averaged over all users for both the depressed and normal classes, except for the median time of posting.

Type	Measure	Depressed	Normal
Activity	No. of posts (daily)	5.84	7.95
	No. of posts (entire period)	2041.56	3145.88
	Retweet rate (daily)	4.61	7.28
	Retweet rate (entire period)	1366.54	2742.32
	Mention rate (daily)	1.68	4.25
	Mention rate (entire period)	359.78	1048.45
	Median time of posting	11:51 p.m	5:36 p.m.
	Regional entropy of ego-net	3.761	4.483
Specific ego-net properties	Size (1-hop)	1196	3215
	Size (2-hop)	210850	987098
	Density (1-hop)	8.59×10^{-5}	2.67×10^{-5}
	Density (2-hop)	3.44×10^{-7}	1.41×10^{-7}
	User clustering coeff (1-hop)	0.208	0.073
	Eccentricity of user (1-hop)	4.4	2.6

We next used a Twitter Streaming API-based crawler to collect seven months worth of Twitter data, from July 2016 to January 2017, consisting of the tweets of the above identified clinically depressed users, their immediate or one-hop neighbors (a user's followers and followees i.e. friends on Twitter) and their two-hop neighbors (the followers and followees of a user's followers and followees on Twitter). We also used this dataset to expand the depression lexicon we constructed from *www.webmd.com*, to include social media specific terms. For this, we trained a word2vec [23] model on the Twitter dataset, extracted from it the top 500 words most likely to be used in the context of the depression-related set of terms, and added these to the depression lexicon. Further details of the Twitter dataset are presented in Table 1.

For the control group of 'normal' (non-depressed) users, we elected to randomly sample a group of a hundred users based in the US. We explicitly sought to minimize any network interference effects with any of our selected depressed users (i.e. the ego-net of normal users we sampled had negligible overlap with depressed users' ego-nets), by discarding users who did not meet this criterion. The overlap exceptions being highly popular users (such as rock stars, famous sports personalities, major league teams etc.) who may appear on both depressed users' and normal users' ego-nets.

We note that Table 1 shows some interesting differential statistics between normal and depressed users in terms of the size of respective ego-nets and activity levels. We will drill down on some of these aspects in the subsequent sections.

3 METHODOLOGY

Informally, we seek to build a model to predict which users are likely to become victims of clinical depression in the near future based on their behavioral characteristics on social media, and to also identify highly depressed users within a social network. We would additionally like to answer the following question: Is the

emotion of a depressed user on a social network a function of what a user is exposed to (i.e. via a user's immediate or one-hop neighborhood), and a secondary function of what the user's neighborhood is exposed to (i.e. the user's two-hop neighborhood)? As noted previously, we focus on *Twitter* as the social network of choice.

3.1 Network Activity and Participation

Theory: Kawachi et al [19] have shown that depressed users tend to cluster together. Lustberg et al [21] found a strong correlation between depression and insomnia, i.e. depressed users tend to be more active online during the later hours of the night. Also, studies both offline among college students [20] as well as on social networks (Facebook) [24] have shown an increase in social media/internet usage in victims of depression.

Analysis: Table 1 provides basic statistics regarding the online activity of users of both classes on Twitter, as well as various structural features of their ego-network. We find that in line with social and psychological research, the potentially depressed users typically exhibit more of nocturnal behavior than non-depressed class users (see median time of posting). Such users tend to mention and retweet other users in their posts less frequently than their non-depressed counterparts, suggesting a lack of direct interaction with other users. Though the overall activity of the depressed class users is lesser than a non-depressed class user, a smaller percentage of the posts tweeted by depressed users consist of simply retweeting what others have said, as compared to normal users who retweet others much more.

In addition, we computed the regional entropy of the depressed users' ego-network, using the algorithm by Compton et al [5] to extract the location of each tweet. Interestingly, depressed class users on Twitter have a slightly lower tweet location entropy than the non-depressed class of users, suggesting that in terms of their physical geo-location, people in a depressed user's ego network are closer to each other than those in a normal user's ego network.

With respect to ego network characteristics, we observe that the one-hop and two-hop networks of depressed class users are smaller on average when compared to normal users. Their ego networks tend to consist of multiple, tighter-knit clusters (hence the higher density of the network), and the depressed users themselves seem to be connected with a smaller fraction of nodes within their ego network than the normal users. The eccentricity of a depressed user is also significantly larger than that of a normal user within their respective one-hop ego-networks, suggesting that normal users tend to be more centrally located within the ego-network. A somewhat surprising statistic is that in spite of having a lower eccentricity value, the average clustering coefficient of depressed users is quite a bit higher. The *raison d'etre*, as we shall demonstrate shortly, is because of the homophily effect [33] – depressed users tend to be clustered with other potentially depressed users within their ego-net (see Figure 2).

3.2 Network Engagement and Experience

3.2.1 Network Responsiveness to User

Theory: Leading sociologists and psychologists note that victims of clinical depression tend to be socially isolated in an offline setting. In a study, Joiner et al [18] tested whether depressed individuals would be affected by their non-depressed peers in one-on-one interactions, and they found that victims of depression often receive from their peers a negative or unfavorable response or rejection to their constant seeking of reassurance, which in turn exacerbates their depressed state of mind. The fact that depressed persons prefer to associate with others who also tend toward depression (homophily) was also concluded by Rosenblatt et al [33] through an experiment with undergraduate students. They found that depressed people felt worse than earlier after speaking with non-depressed people, but not after speaking with similarly depressed targets. In yet another study on adolescents, Hogue et al [15] conclude that as a result of the 'selection effect', they tend to choose friends possessing similar levels of internal distress.

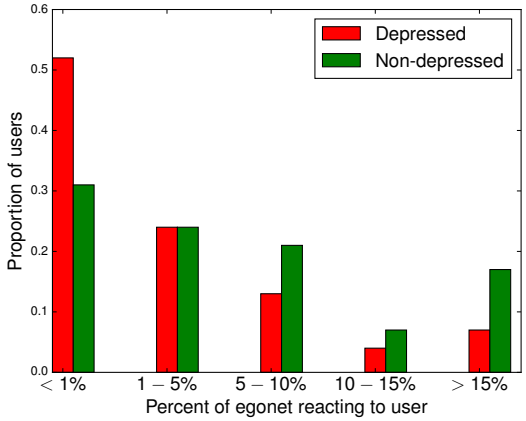

Figure 1: Percentage of ego-network reacting to a user, in the form of mentions, retweets or replies, for both user classes.

Analysis: In order to examine whether the above ideas extend to online social networks, we study basic network engagement effects of the depressed and non-depressed users in our study. Figure 1 displays the percentage of the users' ego-network reacting to the user in the form of mentions, retweets or replies. This gives an indication of the extent to which a depressed class user is able to influence his ego-network. We observe that nearly half of the depressed users have no or minimal impact on their network (network response towards them is $< 1\%$), while for roughly the other half, $5 - 15\%$ of their network reacts to them in the form of replying to them, mentioning them or retweeting their tweets. This implies that the depressed user in these cases is engaged and able to exert some influence on his ego-network. For normal users the level of engagement is distributed between $0 - 15\%$. Certain depressed users receive a significant amount of support from their ego network ($> 15\%$). The results here suggest that in the online setting while some users tend to be socially isolated (agreeing with some of the above offline studies), some of them are adequately engaged with their ego-networks both in terms of participation (Table 1) and engagement (Figure 1).

We next examine the relative position of a user with respect to his/her network structure. We aggregate all the tweets made by the depressed users and the users belonging to their ego-network,

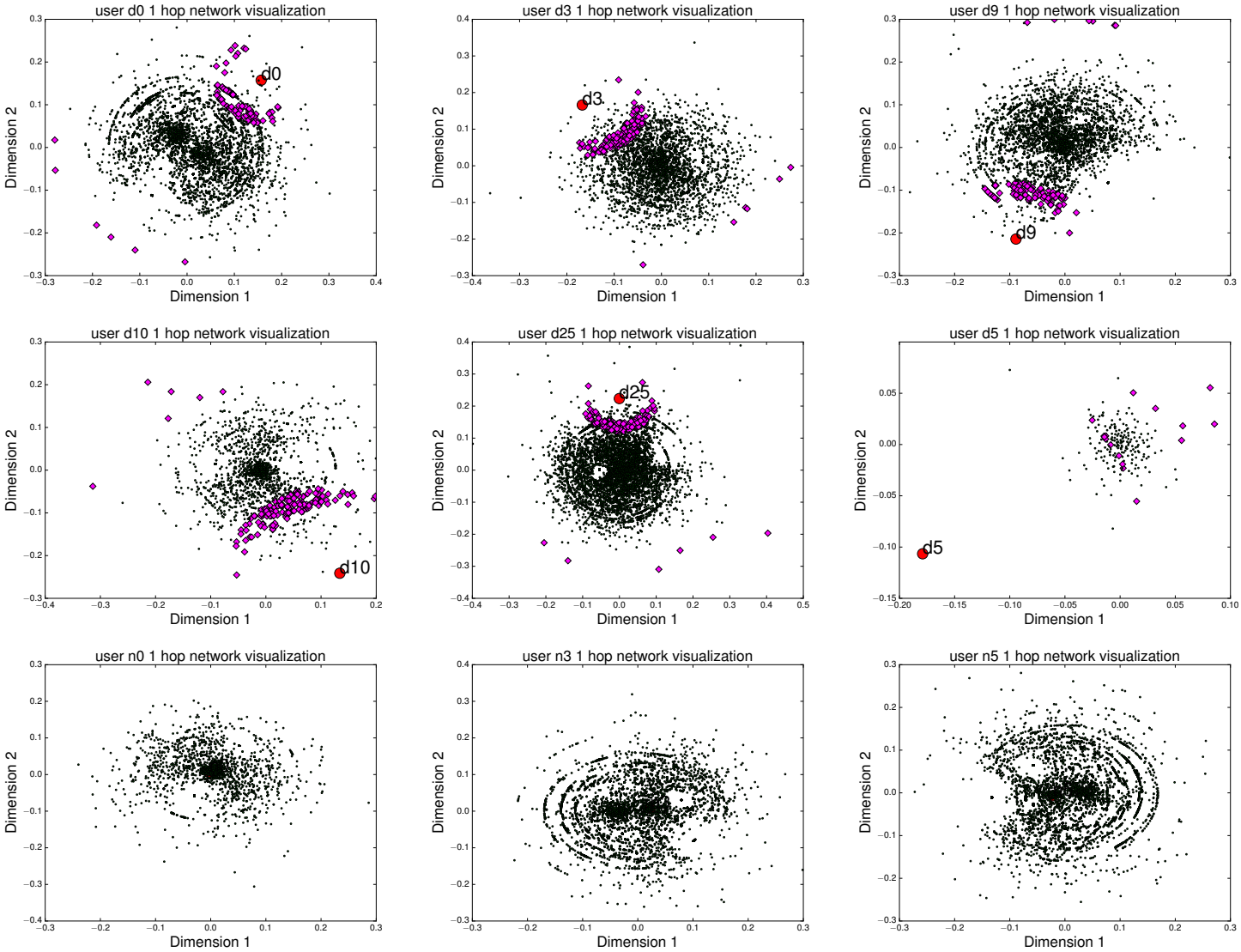

Figure 2: Visualization of selected users belonging to the depressed and normal class within their ego-net. The x and y axes represent the 2 dimensions obtained from multi-dimensional scaling. The green points represent the ego-net while the red point represents the user. The pink points represent the users in the ego-net of the depressed user, who have also been predicted as depressed. The first two rows of plots belong to depressed users and the third row belongs to the normal class of users.

and obtain vector representations of each word in the tweets from our pre-trained word2vec model (described in Section 2). For the purpose of this evaluation, only original content posted by a user or ego-net partner is levered; retweets are not considered in our aggregation to avoid bias effects. After accumulating the vector representation of each word for each user, we perform a dimension-wise average of all the vectors to get a single high-dimensional vector representation for each user. To visualize this, we use Multi-dimensional Scaling (MDS) [22] with cosine similarity as a similarity metric, to scale down the users' feature vectors to two dimensions. We depict this in Figure 2 for a subset of users belonging to both classes. The x and y axes represent the two dimensions obtained from MDS. For each user plot, the green points represent the users belonging to their ego-network while the large red point represents the users themselves. We observe that the depressed users are like outliers in their networks, at a significant distance away from the core of their network. Multiple potentially depressed users tend to cluster together (the pink points – more on this in Section 4). The non-depressed class users in the last row of Figure 2 are more centrally located and are similar to the other users in their network.

3.2.2 *Linguistic Content (Pronoun) Analysis*

Theory: Various studies have analyzed the linguistic style and content associated with the text and/or speech of depressed individuals [2, 16, 34]. These attest to the fact that depressed individuals

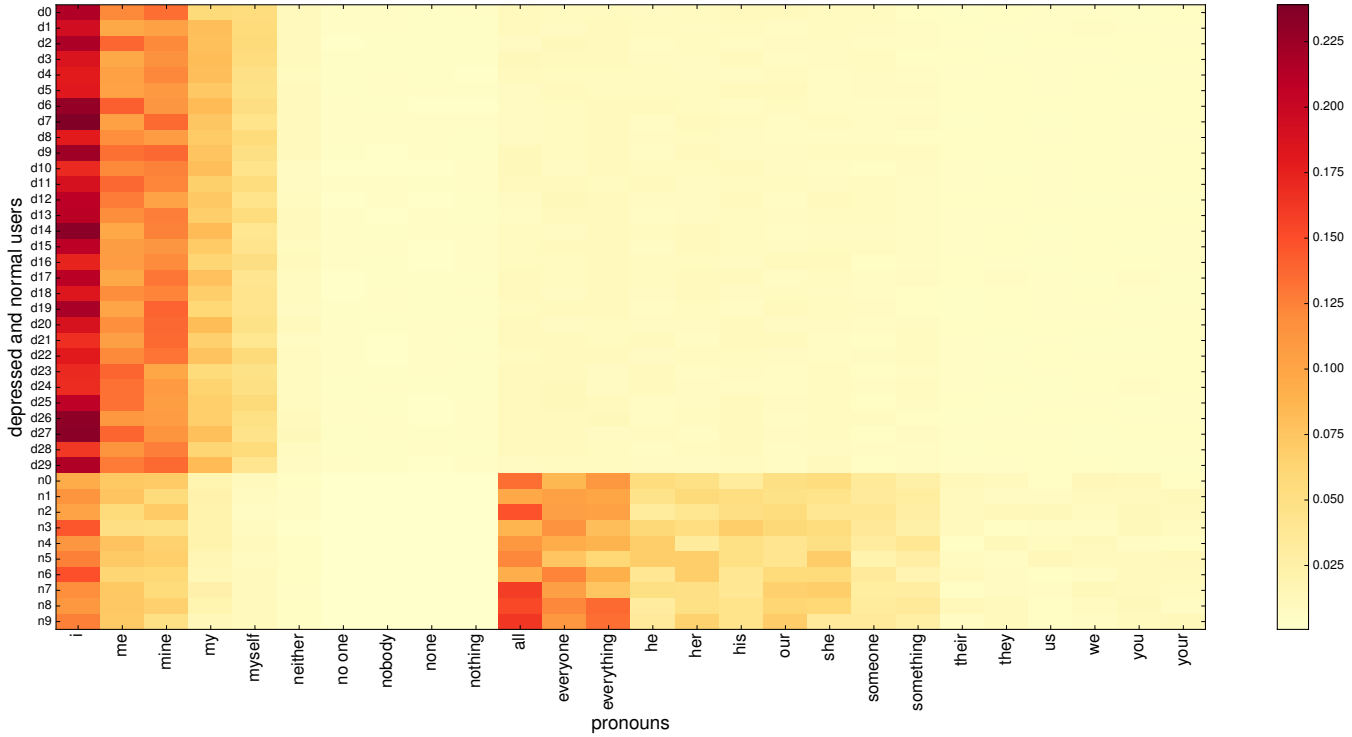

Figure 3: Heatmap distinguishing linguistic pronoun usage of depressed users from normal users. Self-focused (e.g. 'I', 'me', 'my', 'mine') and group connotation pronouns (e.g. 'our', 'we') have the highest differential capability between the two classes. In the interests of space, we present a representative sample of non-depressed and depressed users in this figure.

have a higher propensity towards self-focus, which translates to an increased linguistic usage of personal pronouns associated with the self such as 'I' and its derivatives, and a reduced use of third-person pronouns or those exhibiting collective connotation.

Analysis: In connection with the above findings, we explore the linguistic style patterns of the depressed and non-depressed class users in terms of the pronoun usage in their tweets. We first identify the top pronouns that are most frequently used by the users of both classes in their tweets. Figure 3 shows a heat map in which the colorbar represents the frequency of the pronouns with respect to the total number of unique words in the tweet vocabulary. In this figure and subsequently, the usernames beginning with the letter 'n' represent normal class users and those beginning with the letter 'd' represent depressed class users. We observe perfect separation of the users into two classes. We observe that the self and negative connotation pronouns (the first ten pronouns on the x-axis) are used relatively heavily by the users belonging to the depressed class, and second-person pronouns such as 'you' or those that denote group connotations such as 'we', 'our', 'they' etc are hardly used by this class of individuals. This indicates that these users are more inclined to talking about themselves in an isolated manner, without including themselves with other people.

3.2.3 Content-based Emotion Analysis

Theory: Multiple analyses related to the spread of depressive symptoms were performed on a densely interconnected social network of 12067 people as part of the Framingham Heart Study [10]. Depressive symptoms were evaluated using CES-D (Center for Epidemiological Studies Depression Scale) scores, and results confirmed that both low and high CES-D scores (i.e. absence and presence of depression in an individual) in a given period were strongly correlated with the CES-D scores in the individual's friends and neighbors, extending up to three degrees of separation (one's friends' friends' friends i.e. the three-hop neighborhood of an individual). Interestingly, this study also confirmed that while positive emotions such as happiness seem to spread across a social network (conditioned on geographical location), negative emotions such as sorrow, anxiety, distress or depression do not possess the same contagion effect i.e. they do not spread across a social network. Studies in the literature have also analyzed the affective valence associated with the text and/or speech of depressed individuals, and have found the presence of negative emotional affect to a much higher degree in the language of victims of clinical depression [16, 26].

Analysis: We first aggregate all the tweets that have been collected for each user and their ego-network on a per-day basis. For the purpose of quantifying the emotional strength within the textual content expressed by users in their tweets, we use a tool called SentiStrength [36] that is customized to detect positive or negative emotion within short, informal texts characteristic to social media.

Figure 4 displays the distribution of average positive and negative emotion over the seven days of the week, for selected users

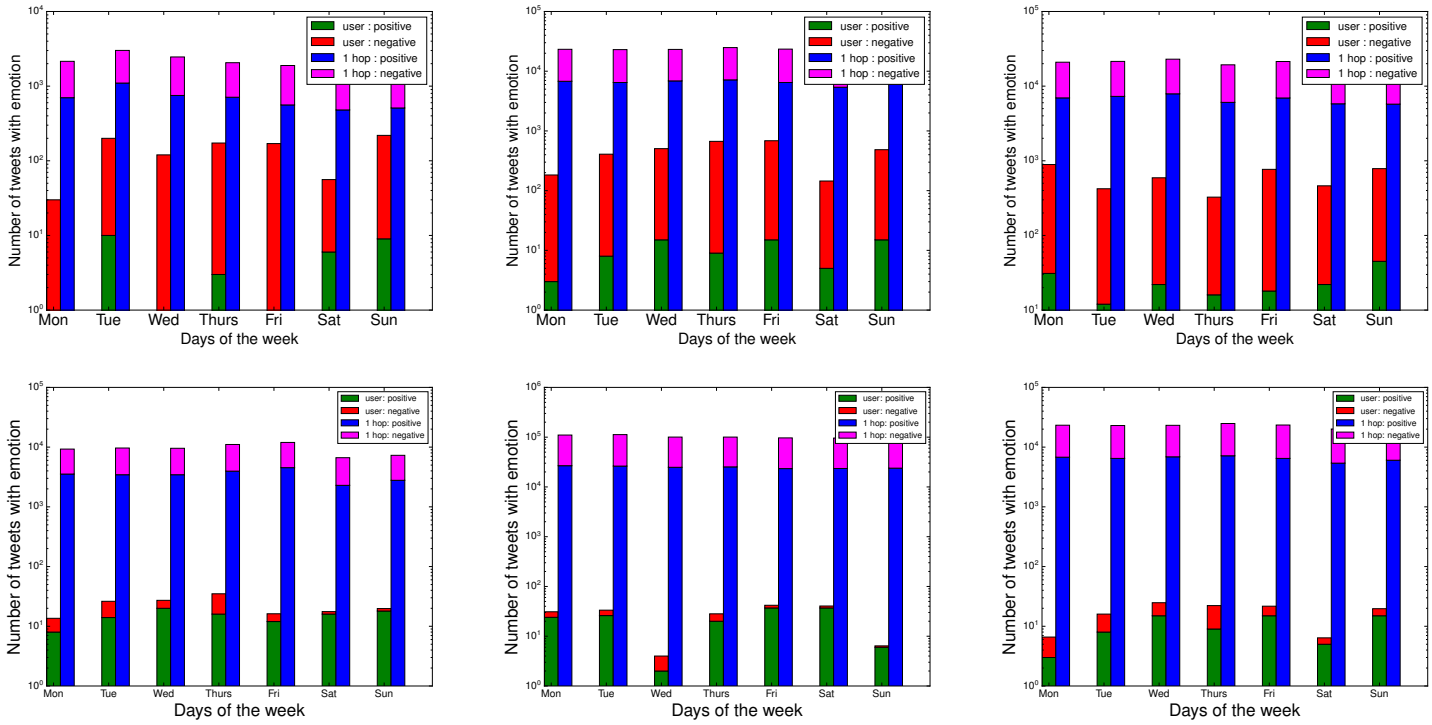

Figure 4: Number of tweets with positive and negative emotions of users and their one-hop and two-hop networks, over days of the week. The first row of figures is for the depressed class of users and the second is for the non-depressed class.

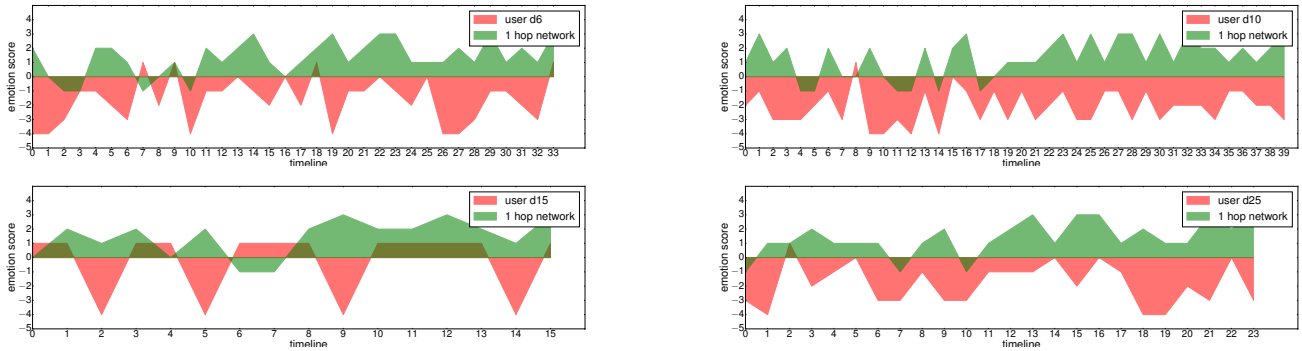

Figure 5: Emotion scores of selected depressed class users and their one-hop networks over time (red represents the user's emotion and green represents the average emotion of the one-hop network). The brown regions show overlap between the emotion of depressed users and their network. One unit of time on the x-axis corresponds to three actual days of user activity.

belonging to the depressed and non-depressed classes. The tweets of the user as well as their network were first aggregated according to the day of the week on which they were posted and then segregated into positive and negative after averaging the predominant emotion expressed in their content. The x-axis represents the day of the week and the y-axis shows the number of tweets of the user or their network expressing positive or negative emotion. We can see from the stacked bar chart that for users belonging to the depression class, the tweets expressing negative emotion heavily outnumber the tweets expressing positive emotion. We do not observe any significant relationship between the predominant emotion of a user and the day of the week, which largely agrees with social and psychological research. On the other hand for individuals belonging to the normal class, positive emotion is dominant and these users seem to exhibit more negative emotion in their tweets on the initial working days of the week i.e. Mondays and Tuesdays, and as they reach the end of the week their tweets grow more positive. The average emotion expressed by users belonging to the one-hop network as well as the two-hop network (not shown in the figure for the sake of brevity) of both classes is largely positive.

We next analyze the temporal distribution of the overall emotion expressed by the potentially depressed users and their ego-network

Table 2: Cross correlation analysis of a selected representative sample of depressed class and non-depressed class users' emotion distribution over time with the users of their one-hop and two-hop network.

User [neg, pos]	1-hop [neg, pos]	Corr 1-hop (Lag/Lead)	Time 1-hop (Lag/Lead)	Corr 2-hop (Lag/Lead)	Time 2-hop (Lag/Lead)
d1 [-2, 1]	[-1, 3.5]	0.118	1	0.106	1
d5 [-3, 1]	[-1, 3]	0.209	-1	0.061	-1
d8 [-3, 1]	[-2, 3]	0.23	1	0.065	-1
d10 [-4, 1]	[-1.5, 3]	0.282	1	0.132	-1
d12 [-2, 1]	[-1.5, 3.5]	0.116	1	0.11	0
d14 [-2, 1]	[-1, 3.5]	0.078	-1	0.06	1
d15 [-3, 1]	[-2, 3]	0.213	1	0.128	0
d16 [-2, 1]	[-1.5, 3.5]	0.036	1	0.012	0
d21 [-2, 2]	[-1, 4]	0.017	1	0.031	-1
d22 [-3, 2]	[-1.5, 4]	0.025	-1	0.019	1
d24 [-2, 1]	[-2, 3.5]	0.186	1	0.177	-1
d25 [-3.5, 1]	[-2, 3]	0.277	1	0.218	-1
d29 [-4, 1]	[-1.5, 3]	0.301	1	0.237	-1
n0 [-2, 3]	[-1, 5]	0.516	-1	0.467	-1
n1 [-1, 4]	[-1, 5]	0.628	-1	0.818	-1
n3 [-1, 3]	[-2, 4.5]	0.65	-1	0.868	-1
n5 [-2, 3]	[-1, 4]	0.51	0	0.579	-1
n6 [-1.5, 3]	[-1, 4]	0.78	-1	0.76	-1
n8 [-1, 3]	[-1, 3.5]	0.517	0	0.511	-1

(Figure 5). For this, we aggregate all the tweets made by a given user as well as their ego-network over three-day intervals, not considering the days when the user does not tweet anything. As earlier, we eliminate from consideration the retweets of a user. The x-axis represents the timeline, where one unit on the x-axis corresponds to three days, and the y-axis represents the average emotion score for that duration. The red and green plots represent the emotion distribution of the user and his ego network respectively. As expected, the emotion expressed by the depressed users is predominantly negative with some regions of positivity, whereas the overall emotion of the users' ego-networks is positive. The overlap in emotion of the depressed users with their ego-net over time (the brownish colored portions in the plots) is low. This confirms that a depressed class user is not very likely to get influenced by the emotion prevalent in their neighborhood, and tends to remain socially isolated, in line with social and psychological studies.

We further strengthen our claim of the depressed sentiment of a user not affecting or being affected by his/her neighborhood to a significant extent. For this, we inspect the cross correlation between the temporal emotion distribution of the users belonging to the depressed and non-depressed classes with that of their one-hop and two-hop networks (see Table 2). In order to compute this, as earlier, we aggregated the daily tweets of each user and their network, eliminated retweets, and the days when the user did not post any tweets. We then computed the cross correlation values over time with respect to the daily average emotion scores between the user and those of his one-hop network (third column of Table 2), and the user and those of his two-hop network (fifth column of Table 2). We investigated if we could observe a temporal lag (indicating that emotion permeates from the user's ego-network to the user) or lead (an emotion contagion from the user to his ego-network) associated with the highest cross correlation value (fourth and sixth columns of Table 2). A positive value in the fourth and sixth columns represents a lead while a negative value represents a lag. The parentheses in the first two columns contain the range of the emotion score for the users and their ego-net respectively, averaged over all their tweets.

The best correlation values of the depressed users' aggregated average emotion with that of their one-hop or two-hop network are quite low (≤ 0.3), while they are much higher for the normal class users (> 0.5 in most cases). We observe that for users belonging to the depressed class, the average emotion scores range between -2 and -4, while for the ego-networks as well as normal users, they are largely positive with slight negativity. Some normal class users such as $n0$ and $n5$ are slightly more negative than others. We do not find any significant evidence of a lag or lead in time between the emotion of a user belonging to the depressed class and his ego-network, indicating that the depressed users seem to be largely isolated from and unaffected by their neighbors and/or network. Normal users predominantly tend to lag behind their network i.e. appear to be influenced by the emotion of their immediate neighbors (average correlation value of > 0.5) within a day.

Finally, in order to evaluate the differences between the two groups of users with respect to the various behavioral attributes detailed above, we perform a test of statistical significance, shown in Table 3. We find that the difference between the two groups is statistically significant with respect to most attributes.

4 PREDICTIVE MODEL FOR DEPRESSION

Differential Analysis: Based on our investigation thus far, we now conduct a differential analysis to identify features that can be used to distinguish between the behavior of users that potentially suffer from clinical depression and those that do not.

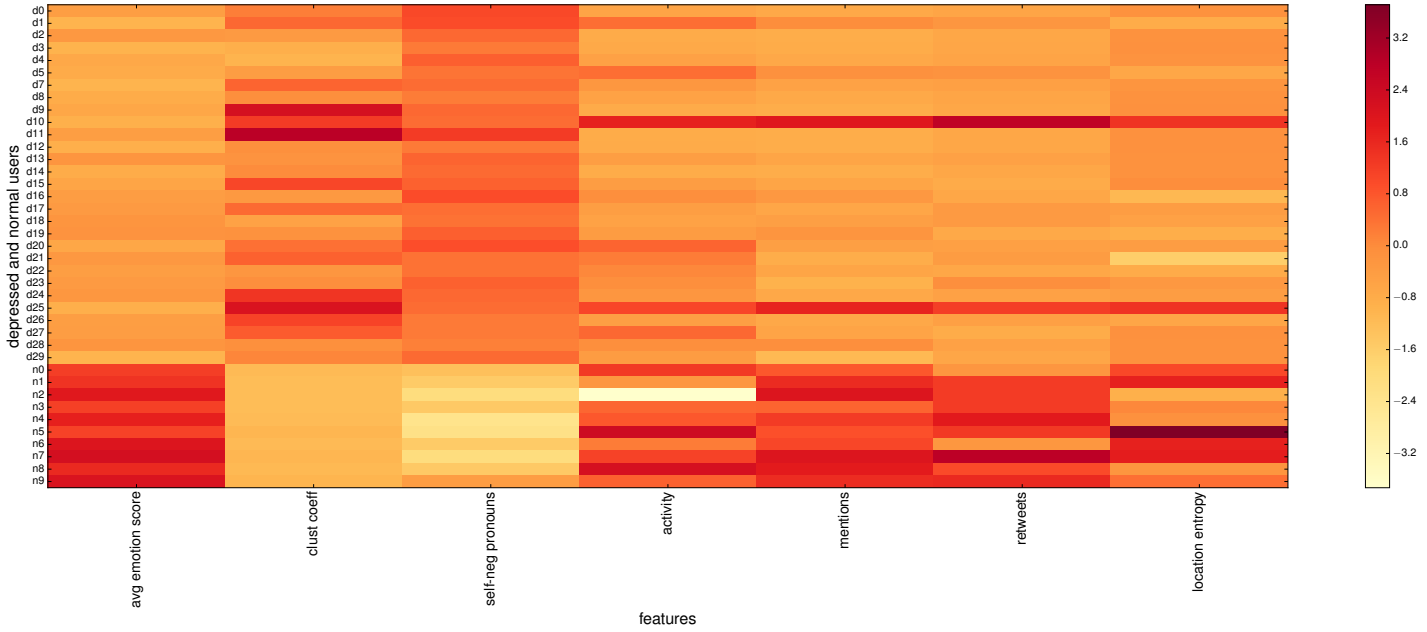

Figure 6: Heatmap distinguishing behavioral features of depressed users from non-depressed users. In the interests of space, we present results over a representative of non-depressed and depressed users (the same sample as Figure 3).

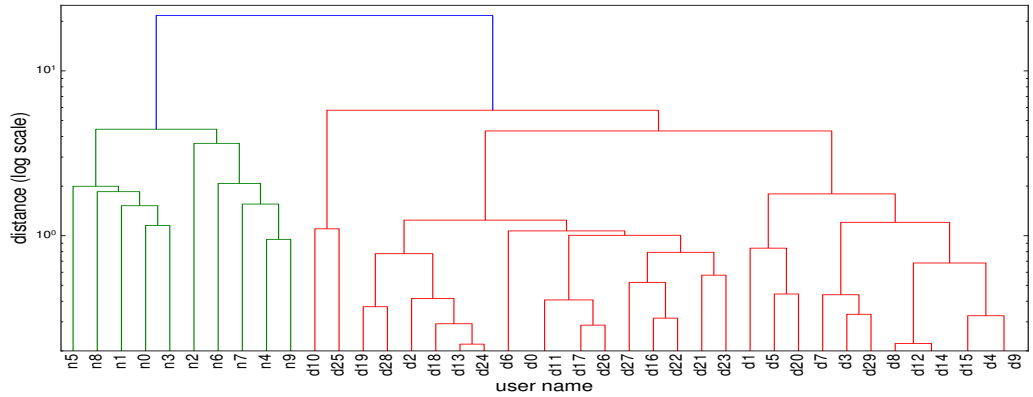

Figure 7: Dendrogram of depressed and normal users based on hierarchical clustering of user features. In the interests of space, we present results over a representative of non-depressed and depressed users (the same sample as Figure 3).

In Figure 6, we plot a heat map distinguishing online behavioral features of depressed class users from non-depressed class users. The y-axis shows a selected sample of users belonging to the depressed and normal classes. We explore the following contextual, linguistic and structural and network engagement based features representative to their behavioral identity on social media (displayed on the x-axis): average textual emotion expressed, user clustering coefficient with respect to their one-hop network, linguistic style (the proportion of self and negativity related pronoun usage in their text), average user activity (number of posts), average number of mentions and retweets received by the user and the regional entropy of their ego-network. All feature values have been normalized using their z-scores, denoted by the colorbar.

A differential analysis reveals similar feature trends among the depressed class users, that depart from the trends observed for the normal users. Depressed class users exhibit significantly negative emotion in their tweets. They have a higher clustering coefficient than their non-depressed counterparts in most cases. The proportion of self and negative pronoun usage is notably higher in case of the depressed class users. Depressed class users seem to be less active overall than regular users. Further, the number of retweets and mentions received by most depressed class users is clearly lower than those received by the non-depressed class users, showing that the depressed users don't seem to have much of an effect on their ego network. The ego-networks of depressed class users seem more geographically co-located than those of the normal class users.

Table 3: Two-sample t-test of significance comparing both user classes (significant differences in bold)

Feature	p-value
Average user emotion	**0.000142**
Clustering coefficient	**0.05**
Pronoun usage	**0.004**
User Activity	0.112
No. of mentions	**0.00243**
No. of retweets	**0.00121**
Location entropy of egonet	0.078
% of reaction obtained from egonet (Figure 1)	**0.043**
Correlation of user emotion with 1-hop network (Table 2)	**0.00713**
Correlation of user emotion with 2-hop network (Table 2)	**0.00904**

There are however some users (for example, $d10$ and $d25$ in Figure 6) belonging to the potentially depressed class who are quite different from the rest of their class members. They seem to be more similar to the normal class of users in features such as activity, mentions, retweets and network location entropy. They are more active online and receive positive reinforcement from their network, in the form of an increased number of mentions and retweets.

Building Predictive Models: Since the features identified above discriminate quite well between the two classes of users, we utilize them in order to predict depression in our two user classes. For this, we first inspect how well these features are able to separate the users into 2 classes using Ward's method of agglomerative hierarchical clustering [22]. Figure 7 shows this result for a sample of 40 users in the form of a dendrogram, where the x-axis represents the users and the y axis represents the distance between clusters. We notice that users $d10$ and $d25$ who were identified as somewhat distinct from the rest of their class in Figure 6, appear to be more similar to the normal class of individuals here as well.

We subsequently construct a binary classifier using the above identified feature set. We use the Gradient Boosted Decision Trees classifier [11] along with 5-fold cross validation on the full dataset of 150 users. The features distinguishing between the users of the two classes in decreasing order of importance are: words expressing negative emotion, self and group-focused pronouns, user clustering coefficient, activity, retweets, mentions and location entropy. We measure the performance of the classifier using the standard metrics of accuracy of both classes, micro-F1 score and macro F1-score. We achieve *an accuracy of 0.9 for the depressed class users, a slightly lower accuracy of 0.87 for the normal/non-depressed class of users, a micro-F1 score of 0.9 and a macro-F1 score of 0.8901.* Drilling down on the results, we find that only 5 depressed users are misclassified as normal. Of the misclassified normal users, we find that while their tweets do not express depression, they still exhibit an increasing use of words with negative emotions such as violence or anger.

We additionally use this classification model to predict whether the ego-net of depressed users consists of other similarly depressed users. This would further endorse the theory of clinically depressed users tending to assemble together as a group. While we lack ground truth, we show the users that have been classified as depressed within the one-hop network by the pink colored points in Figure 2. The originally depressed and the predicted depressed individuals tend to cluster together with an average clustering coefficient of 0.153, exhibiting some degree of homophily. Many of them are also connected to each other based on their Twitter follower-followee relationship. We validated that these users actually show signs of depression by manually looking at their Twitter feed. We find our results to correlate quite well with some of the social and psychological research described earlier. In addition, some other network users who did not cluster together with the original depressed class user are also predicted as depressed. As expected, these users are quite far from the core of the network. Some cases (such as user $d5$ in Figure 2) do exist who do not cluster together with other similarly depressed users in their network.

5 RELATED WORK

As detailed in Section 3, several experiments have been conducted and theories posited in the fields of social science, psychology, psychiatry, medical science and linguistics in conjunction with the onset and spread of clinical depression and its symptoms in individuals [2, 15, 16, 18–21, 24, 26, 34]. While the importance and utility of such empirical research cannot be underestimated, a key challenge associated with it is the difficulty in obtaining data pertaining to specific individuals, as well as monitoring them for long periods of time. Therefore in recent years, researchers have been employing social networking websites in order to collect data as well as study behavioral characteristics of people related to various aspects of mental and psychological health. Social media has been used to study dissemination of health information [14, 35], as well as to gain key insights related to the spread of diseases and their symptoms [6, 30]. Prior work [4, 6–9, 13] has also highlighted the usefulness of social media in various issues concerning mental health. Jelenchick et al [17] and Moreno et al [24] analyzed the phenomenon of escalating signs of 'Facebook Depression' among users due to rising use of the social network *Facebook*. Changes in mood and emotional state of individuals is reflected on their social media profiles, according to multiple studies conducted on Twitter data [1, 12]. Park et al [29] observed that people make posts online regarding their depression and even treatment received. Analyzing textual content of individuals has also proved to be helpful in identifying signs of various mental disorders among them [2, 16, 25, 28, 31, 34, 40]. DeChoudhury et al in [7] use social media as a tool to study postpartum depression in pregnant women. In another work [8], they leverage social media analysis to estimate an individual's risk of having Major Depressive Disorder (MDD). Our work builds on these efforts, drilling down on emotional, linguistic (self-focused pronouns) and location-based cues in addition to standard activity patterns and features of the ego-network. We study emotional contagion and temporal relationships between depressed individuals and their ego network, as well as their orientation in their network topology. Furthermore, we show how to realize a practical and accurate classifier for potentially classifying users who may be suffering from depressive tendencies by focusing on seven high level features.

6 CONCLUSION

We perform an empirical study on Twitter to understand the online behavior of potentially depressed users against a differential control group of normal users. After building a lexicon of words

regularly used in conjunction with clinical depression, we examine a wide range of social media related signals such as linguistic style, emotional signals, user engagement, geo-location and network topology to detect symptomatic cues of depression online. We notice significant deviations in the behavior of depressed users from the control group in the form of reduced and nocturnal online activity patterns, reduced active and passive network participation, increase in textual negative emotion, distinct linguistic styles (e.g. self-focused pronoun usage), highly clustered and tightly-knit neighborhood topology, a slightly higher geo-location proximity among ego-network members and little to no exchange of influence among depressed users and their ego-network over time. Based on these observations, we extract relevant features and build a classifier to predict depression among individuals. It achieves an F-score of 90%. Most of our empirical findings corroborate quite well with theoretical literature from the social sciences, medicine and psychology, suggesting that social media interactions may offer a crucial diagnostic tool for clinicians.

Competing Interests Statement: The authors have no competing interests to declare.

7 ACKNOWLEDGMENTS

The authors would like to thank Dr. Christopher Browning for his useful insights and suggestions on an early version of this work. This work was partially supported by the following grants: NSF-EAR-1520870 and NIH-1R01 HD088545-01A1. Any opinions, findings, and conclusions in this material are those of the author(s) and may not reflect the views of the respective funding agencies.

REFERENCES

[1] Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *ICWSM* (2011).
[2] Wilma Bucci and Norbert Freedman. 1981. The language of depression. *Bulletin of the Menninger Clinic* (1981).
[3] John T Cacioppo, James H Fowler, and Nicholas A Christakis. 2009. Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of Personality and Social Psychology* (2009).
[4] Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016. Quantifying and Predicting Mental Illness Severity in Online Pro-Eating Disorder Communities. In *ACM CSCW*.
[5] Ryan Compton, David Jurgens, and David Allen. 2014. Geotagging one hundred million twitter accounts with total variation minimization. In *IEEE Big Data*.
[6] Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. *NAACL HLT* (2015).
[7] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *ACM SIGCHI*.
[8] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media.. In *ICWSM*.
[9] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *ACM SIGCHI*.
[10] James H Fowler and Nicholas A Christakis. 2008. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal* (2008).
[11] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* (2002).
[12] Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* (2011).
[13] Glen Coppersmith Mark Dredze Craig Harman. 2014. Quantifying mental health signals in Twitter. *ACL* (2014).
[14] Carleen Hawn. 2009. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. *Health Affairs* (2009).
[15] Aaron Hogue and Laurence Steinberg. 1995. Homophily of internalized distress in adolescent peer groups. *Developmental Psychology* (1995).
[16] Rick E Ingram, Debra Cruet, Brenda R Johnson, and Kathleen S Wisnicki. 1988. Self-focused attention, gender, gender role, and vulnerability to negative affect. *Journal of Personality and Social Psychology* (1988).
[17] Lauren A Jelenchick, Jens C Eickhoff, and Megan A Moreno. 2013. "Facebook depression?" Social networking site use and depression in older adolescents. *Journal of Adolescent Health* (2013).
[18] Thomas E Joiner, Mark S Alfano, and Gerald I Metalsky. 1992. When depression breeds contempt: Reassurance seeking, self-esteem, and rejection of depressed college students by their roommates. *Journal of Abnormal Psychology* (1992).
[19] Ichiro Kawachi and Lisa F Berkman. 2001. Social ties and mental health. *Journal of Urban Health* (2001).
[20] Raghavendra Kotikalapudi, S Chellappan, F Montgomery, D Wunsch, and K Lutzen. 2012. Associating depressive symptoms in college students with internet usage using real Internet data. *IEEE Technology and Society Magazine* (2012).
[21] Lisa Lustberg and Charles F Reynolds. 2000. Depression and insomnia: questions of cause and effect. *Sleep Medicine Reviews* (2000).
[22] Oded Maimon and Lior Rokach. 2005. *Data Mining and Knowledge Discovery Handbook*. Springer.
[23] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*.
[24] Megan A Moreno, Lauren A Jelenchick, Katie G Egan, Elizabeth Cox, Henry Young, Kerry E Gannon, and Tara Becker. 2011. Feeling bad on Facebook: depression disclosures by college students on a social networking site. *Depression and Anxiety* (2011).
[25] Yair Neuman, Yohai Cohen, Dan Assaf, and Gabbi Kedma. 2012. Proactive screening for depression through metaphorical and automatic text analysis. *Artificial Intelligence in Medicine* (2012).
[26] Jean Nunn, Andrew Mathews, and Peter Trower. 1997. Selective processing of concern-related information in depression. *British Journal of Clinical Psychology* (1997).
[27] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, and others. 2011. The impact of social media on children, adolescents, and families. *Pediatrics* (2011).
[28] Thomas E Oxman, Stanley D Rosenberg, and Gary J Tucker. 1982. The language of paranoia. *American Journal of Psychiatry* (1982).
[29] Minsu Park, Chiyoung Cha, and Meeyoung Cha. 2012. Depressive moods of users portrayed in Twitter. In *ACM SIGKDD Workshop on Healthcare Informatics*.
[30] Michael J Paul and Mark Dredze. 2011. You are what you Tweet: Analyzing Twitter for public health. *ICWSM* (2011).
[31] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use. *Annual Review of Psychology* (2003).
[32] Andrew Perrin. 2015. Social media usage. *Pew Research Center* (2015).
[33] Abram Rosenblatt and Jeff Greenberg. 1991. Examining the world of the depressed: Do depressed people prefer others who are depressed? *Journal of Personality and Social Psychology* (1991).
[34] Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion* (2004).
[35] Daniel Scanfeld, Vanessa Scanfeld, and Elaine L Larson. 2010. Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control* (2010).
[36] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *JASIST* (2012).
[37] Koji Ueno. 2005. The effects of friendship networks on adolescent depressive symptoms. *Social Science Research* (2005).
[38] Thomas W Valente. 2010. *Social networks and health: Models, methods, and applications*.
[39] Patti M Valkenburg and Jochen Peter. 2007. Online communication and adolescent well-being: Testing the stimulation versus the displacement hypothesis. *Journal of Computer-Mediated Communication* (2007).
[40] Walter Weintraub. 1981. *Verbal behavior: Adaptation and psychopathology*.