The power of dynamic social networks to predict individuals' mental health

Shikang Liu[1], David Hachen[2], Omar Lizardo[3], Christian Poellabauer[1], Aaron Striegel[1],
and Tijana Milenković[1,*]

[1]*Department of Computer Science and Engineering, University of Notre Dame*
[2]*Department of Sociology, University of Notre Dame*
[3]*Department of Sociology, University of California, Los Angeles*
Corresponding author: tmilenko@nd.edu

Precision medicine has received attention both in and outside the clinic. We focus on the latter, by exploiting the relationship between individuals' social interactions and their mental health to predict one's likelihood of being depressed or anxious from rich dynamic social network data. Existing studies differ from our work in at least one aspect: they do not model social interaction data as a network; they do so but analyze static network data; they examine "correlation" between social networks and health but without making any predictions; or they study other individual traits but not mental health. In a comprehensive evaluation, we show that our predictive model that uses dynamic social network data is superior to its static network as well as non-network equivalents when run on the same data. Supplementary material for this work is available at `https://nd.edu/~cone/NetHealth/PSB_SM.pdf`.

Keywords: dynamic social networks, mobile health tracking devices, mental health

1. Introduction

Precision medicine, i.e., giving personalized health-related recommendations to individuals, has received attention both in the clinic via -omics data collection technologies[1,2] and outside the clinic via mobile health tracking devices, online social media behavioral data, or other resources.[3–8] In this study, we focus on the role of individuals' social interactions on their health. Individuals' health-related traits, such as obesity, smoking, depression, physical activities, heart rates, self-perceived health, happiness, or stress can spread through social interactions[9–11] or are correlated with the individuals' positions in their social network.[12–17]

We focus on *mental* health, i.e., depression and anxiety, which are critical public health issues affecting millions of individuals.[18–20] Many such individuals do not recognize that they are depressed or anxious and thus do not seek a timely treatment. To help with this, we focus on developing a machine learning model for predicting individuals' mental health conditions.

Existing studies that propose such models can be divided into non-network ones and network ones. Non-network studies rely on smartphone usage data such as incoming and outgoing call frequency, wearable sensor data such as physical activity (e.g., step count), or online social media behavioral data such as text content.[21–26] These studies typically extract features from such data in its raw form and input those features into off-the-shelf classifiers to make mental health predictions (i.e., the likelihood of an individual being depressed or

anxious). On the other hand, network studies either use the same data as non-network studies but first process the raw data into a network that captures relationships between entities (e.g., by linking two individuals if their physical activity profiles are similar), or they use explicit social network data (e.g., friendships between individuals). Then, they extract network-based features (e.g., an individual's centralities in the network) and input those features into off-the-shelf classifiers.[13,14,17,27] In our previous study,[3] we showed that network analysis of rich social data originating from the NetHealth study[28,29] was more accurate than non-network analysis of *the same data* in the task of predicting individuals' mental health.

In more detail, this previous study of ours[3] is the most comprehensive work to date on predicting individuals' mental health in terms of both data size (the number of considered individuals) and data heterogeneity (the number of considered data types). Namely, we leveraged the rich NetHealth data set containing individuals' social interaction data (SMS communications), health-related behavioral data (physical activity and sleep duration) and a variety of individuals' trait data (personality traits, social status, physical health, and well-being) to predict mental health conditions.[3] We integrated such data into a heterogeneous information network (HIN). Then, we modeled the problem of mental health prediction from our HIN in a novel manner, as applying to the HIN a popular paradigm of a recommender system (RS), which is typically used to predict the preference that an individual would give to an item (e.g., a movie or book). In our case, the items were the individuals' different mental health states. RS gives us an intuitive yet powerful way to predict an individual's mental health state (likelihood of being depressed or anxious) by relying on information about both the individual's and her/his neighbors' behaviors and traits. We found that RS produced more accurate predictions than other types of network methods as well as a fairly comparable *non-network* method using the same data. This confirms the power of our HIN-based predictive framework.

The existing RS methods (and other HIN method types) work on static network data. Extending them to be able to deal with dynamic network data is challenging, as it requires novel methodologies to model individuals' dynamic traits and dynamic social interactions.[30] For these reasons, we constructed our HIN by aggregating dynamic social network data from the considered study time period into a static network. And while even this led to high mental health prediction accuracy, temporal information was lost. Including temporal information could further improve prediction accuracy. So, in this study, we fairly evaluate the power of using dynamic versus static social network data in the task of predicting mental health.

Our study differs from existing studies that associated social networks and individuals' traits in at least one key aspect: 1) whether the study is network-based or not; 2) if yes, whether the study considers a dynamic or static network; 3) whether the study builds machine learning models to predict individuals' traits or "only" examines a potential correlation-like relationship between social networks and traits without making any predictions; 4) whether the study deals with studying mental health or instead it studies other traits (Table 1).

In more detail, a group of studies built machine learning models to predict mental health, but they are non-network studies,[21–26,31] unlike our network study. And recall that we already showed that network studies are superior to non-network studies in this task.[3] Among network studies, a majority only explored the existence of a potential "correlation" between individu-

als' social networks and their traits (e.g., whether individuals who have different personality traits tend to have different centrality values). That is, they did not build any predictive models,[9–13,15–17,32–35] which is what we do here. Plus, a majority of these did not focus on mental health, which we do in this study. Of the four network studies that did build predictive models, three used static networks,[3,27,36] and thus, they could not compare the predictive power of dynamic versus static networks, which we do here. Plus, of these three, one study did not even deal with mental health; instead, it dealt with personality traits.[36] The only study that did build predictive models using dynamic social network data did not ask the research question of our study—whether using dynamic network data is more predictive than using static network data.[14] Moreover, while it[14] studied some health-related traits (e.g., happiness, positive attitude) in addition to some non-health-related traits, it did not study mental health (depression and anxiety) as we do. Also, this existing study[14] used as features centralities of nodes in the dynamic network data. This is a typical strategy in many tasks and domains[17,37] and is thus not unique to this existing study.[14] In our study, on top of also using node centralities, we consider additional node features that are based on graphlets.[38–40] Graphlets are subgraphs, i.e., Lego-like basic building blocks of complex networks. They are state-of-the-art features, especially in dynamic networks.[38–40] While these features were proposed by others,[38–40] we are the first of all social-networks-to-health studies to use them, which further distinguishes our work. Note that this existing study[14] was just published, and we became aware of it as we finalized our paper. So, that study[14] and our work are concurrent pieces of work.

Hence, in terms of novelty, we are the first ones to develop predictive models of mental health (depression and anxiety) from dynamic social network data and to study whether using dynamic social network data is more predictive than using static social network data.

Table 1: Summary of existing related work compared to our work proposed in this study. A check mark indicates that the given study satisfies the given criterion.

	Network-based?	Dynamic network?	Predictive model?	Mental health?
Our study	✓	✓	✓	✓
21–26,31			✓	✓
12,15,16,32,34	✓			
33	✓			✓
36	✓		✓	
9,10,17	✓	✓		
3,27	✓		✓	✓
11,13,35	✓	✓		✓
14	✓	✓	✓	

Our contributions are as follows. First, by analyzing our dynamic social network (Section 2.1), we check whether the group of depressed (anxious) individuals occupies different social network positions (i.e., has different dynamic network features) than the group of non-depressed (non-anxious) individuals. Here, the two groups are naturally defined in the data based on one's depression/anxiety trait information. Also, here, we examine *network position* differences in the *trait-based* groups. Second, we check whether individuals who have different

evolving network positions also show depression/anxiety trait differences. Here, we use network clustering to place in the same group those individuals who have similar network positions and in different groups those individuals who have dissimilar network positions. So, here, we examine *trait* differences in the *network-based* groups. In these two analyses, we observe network/trait differences in the trait-based/network-based groups. That is, the network structural information *is* well-associated with the trait information and vice versa. So, it makes sense to develop a predictive model of individuals' mental health traits from their network structural information. Hence, third, we develop such a model, using three features of individuals extracted from the dynamic network: centralities in the dynamic network, dynamic graphlet degree vectors (dynamic GDV),[40] and graphlet orbit transitions (GoT).[38] As a proof-of-concept, we train a logistic regression classifier for each feature to predict individuals as depressed or non-depressed (anxious or non-anxious). To answer our key question—whether using dynamic network data yields more accurate predictions than using static network data—we model the dynamic network as a static network (Section 2.2). We extract two features from the static network: centralities in the static network and static graphlet degree vectors (static GDV).[39] For fairness, these are static counterparts of two of the above dynamic features: centralities in the dynamic network and dynamic GDV, respectively (the third dynamic feature, GoT, has no static counterpart). Also, for fairness, we use the same logistic regression classifier as above on the static network features. We find that using any of the three dynamic network features outperforms using both of the static network features (as well as using a fairly comparable non-network feature). This confirms the superiority of using dynamic network data over using static network data in the task of mental health prediction.

2. Methods

2.1. *Dynamic network*

Data source. Our data come from the NetHealth study, which collected smartphone, wearable sensor (Fitbit), and survey-based trait data about ∼700 undergraduate student participants at the University of Notre Dame from 2015 to 2019.[28,29] For reasons stated in Section 1, of all NetHealth data, we focus on individuals' SMS logs (i.e., social interactions) collected through smartphones and their mental health trait data on depression and anxiety collected through surveys. Of all participants, 615 were iPhone users and the rest were Android users. We focus on the 615 iPhone users because of issues with the Android data consistency.

Selection of the study time period and pool of individuals. We already produced a collection of manuscripts on various research questions related to various dimensions of the NetHealth data.[3,13] This study is a novel addition to that collection. To be able to draw conclusions across the different dimensions, we match the social interaction and mental health trait data closely between this current study and our previous studies.[3,13] This includes the choices of the study time period and pool of individuals, as follows. During the entire 2015-2019 period, as time went on, more and more students dropped out of the study or became data non-compliant. We wish to consider a time period that is as long as possible and that also includes as many individuals as possible; these two conditions conflict with each other. As a result, we focus on the period from August 2015 to August 2016 because only during this period, the

vast majority of NetHealth participants were actively involved in the study.[3,13] This period covers 31 school weeks (we do not consider 21 break weeks, since these do not have meaningful social network structures).[3,13] We observe that during our study time period, 576 out of the 615 iPhone users actively sent or received SMSs. We use the 576 individuals' SMS logs (i.e., social interactions) to construct a dynamic social network. For each temporal snapshot of the dynamic network, one snapshot per week, nodes are the individuals and there is an edge between two nodes if there is at least one SMS event between the two corresponding individuals during the given week. The 31 resulting snapshots form the dynamic network. Formally, the dynamic network can be represented as a sequence of 31 snapshots $\{G_0, G_1, ..., G_{31}\}$, where each snapshot $G_i = (V_i, E_i)$ is a static graph corresponding to week i; V_i is the set of nodes and E_i is the set of edges during the given week.

Of the 576 individuals, 274 are compliant enough and also have mental health trait data.[3,13] These 274 individuals form the final pool of individuals to be used in our predictive tasks. That is, we use the social interaction data of all 576 individuals' to compute each node's features (needed for our predictive tasks) from the entire network. But when we make mental health predictions and evaluate prediction accuracy, we are able to do so only for the 274 individuals for whom we have the mental health data. Of the 274 individuals, 67 individuals (24.5%) are depressed, and 106 individuals (38.7%) are anxious (Supplementary Section S1).

2.2. *Network analysis methods*

We perform three tasks. Tasks 1 and 2 rely on centralities of nodes, and task 3 relies on centralities plus other network features. We first discuss the considered network centrality measures, followed by tasks 1, 2, and 3, where the latter covers the other considered features.
Measuring positions (centralities) of nodes in the dynamic network.
In each network snapshot, we measure network positions of all nodes with respect to eight popular centrality measures: eccentricity, closeness, betweenness, eigenvector, k-coreness, clustering coefficient, degree, and graphlet degree centrality (Supplementary Section S2).[13,41] Betweenness has the highest time complexity, $O(|V|^3)$, where $|V|$ is the number of nodes in the network. We use the multiple centrality measures because they capture the importance of a node in a network from different perspectives. For each centrality measure, for each node, we compute the given node's centrality value in each snapshot, resulting in 31 centrality values for the 31 snapshots of the dynamic network. Because (i) the network snapshots can have different sizes, (ii) centrality measures can be dependent on the network size, and (iii) we aim to study changes in a given node's network position with time (i.e., across the snapshots), we do not consider the "raw" centrality values. Instead, we convert these into centrality ranks by giving the individual with the lowest "raw" centrality value a rank of 1 (the 1st least central), the individual with the next lowest "raw" centrality value a rank of 2 (the 2nd least central), and so on. We assign the ranks this way so that the intuition remains the same as with the "raw" centrality values: the higher the centrality rank, the more central (i.e., topologically important) the node is. Henceforth, by centrality values, we mean centrality ranks.
Task 1: Do depressed and non-depressed (anxious and non-anxious) individuals occupy different social network positions?
a. Magnitudes of individuals' centralities. Here, we aim to measure whether cen-

tralities of depressed (anxious) individuals are higher or lower (on average over time) than those of non-depressed (anxious) individuals. For each centrality measure and each individual, we average the given individual's centralities over the 31 network snapshots. That is, let C_t be the centrality value of node i at time t. Let vector V_i be $\{C_1, C_2, ..., C_{31}\}$. Then, the magnitude of node i's centralities is $\mu(i) = (\sum_{j=1}^{31} C_j)/31$. Next, we compare the distributions of the average centralities of the depressed and non-depressed (anxious and non-anxious) individuals using the Wilcoxon rank-sum test, whose p-value quantifies the significance of the difference between the two distributions. Since we do eight tests for the eight centrality measures, we adjust the p-values via false discovery rate estimation to correct for the multiple tests.[42] Throughout this paper, we use the adjusted p-value threshold of 0.05.

b. Fluctuations of individuals' centralities. Here, we aim to measure whether centralities of depressed (anxious) individuals vary (increase or decrease) over time more or less than those of non-depressed individuals. For each centrality measure and each individual, we measure the coefficient of variation (CV) of the given individual's centralities in the 31 network snapshots. Given the average value of node i's 31 centrality values μ (as defined above) and the corresponding standard deviation σ, $CV = \frac{\sigma}{\mu}$. CV is widely used for comparing variability between different samples that have different averages.[43] Then, we compare the distributions of CV scores of depressed and non-depressed (anxious and non-anxious) individuals using the Wilcoxon rank-sum test as discussed above.

Task 2: Do individuals who have different evolving network positions show depression/anxiety trait differences?

To answer this, we use network clustering to place in the same group (i.e., cluster) those individuals who have similar evolving centrality profiles and in different clusters those individuals who have dissimilar evolving centrality profiles. For reasons described in Supplementary Section S3, we use k-medoids clustering under the Euclidean distance and report results for $k = 4$, i.e., when obtaining four clusters. If the clusters are meaningful, some of them should contain a significant portion of depressed (anxious) individuals while others should contain a significant portion of non-depressed (non-anxious) individuals. We quantify the significance of enrichment of each cluster in depressed and non-depressed (anxious and non-anxious) individuals using the hypergeometric test. Since we test multiple clusters, we adjust the p-values as above.

Task 3: Is using dynamic network data more accurate than using static network data in predicting mental health?

This is our key task. Recall from Section 1 that we create a static version of our dynamic network by connecting two nodes if they have an edge in any temporal snapshot of the dynamic network. Then, we compare node features extracted from the dynamic versus static network under the same classifier (see below). We use the following features of a node in a network.

a. Three dynamic network features. **1.** *Centralities in the dynamic network.* Recall that we consider eight centrality measures, and that for each measure, we obtain a 31-dimensional vector for each node. To hopefully benefit from the different centrality measures, we integrate their eight 31-dimensional feature vectors, resulting in a final $8 \times 31 = 248$-dimensional feature vector for each node. **2.** *Dynamic GDV.*[40] Graphlets, as originally defined in the context of a static network, are small connected non-isomorphic induced sub-

graphs of such a network.[39,44] As an extension of static graphlets to the dynamic network setting, dynamic graphlets were introduced by adding temporal information onto edges of a graphlet, which now become events that appear in a certain temporal order. Dynamic GDV of a node characterizes how the extended network neighborhood of the node evolves by counting for each dynamic graphlet type the number of times the node participates in the given graphlet type. Dynamic GDV's running time complexity is $O(|E| + |E|(\frac{D}{|E|})^{k-1})$, where $|E|$ is the number of edges in the network, D is the number of considered dynamic graphlet types, and k is the maximum number of events over all considered graphlet types. **3.** _GoT_.[38] For a given node, this feature counts how many times in the node's extended network neighborhood each static graphlet type (e.g., a 3-node path) transitions into every other static graphlet type (e.g., a triangle) between every pair of consecutive temporal snapshots. GoT's practical running time is on average 100 times faster than that of dynamic GDV. Dynamic GDV and GoT are complementary dynamic network features.[45]

`b. Two static network features.` **1.** _Centralities in the static network_. In the static network, for each node, for each of the eight centrality measures, we obtain a single centrality value. To benefit from the different measures, we integrate their eight centrality values, resulting in a final 8-dimensional node feature vector. This is the static counterpart of the centralities in the dynamic network above. **2.** _Static GDV_.[39] Static GDV characterizes the structure of a node's extended neighborhood in a static network by counting for each static graphlet type the number of times the node participates in the given graphlet type. Static GDV's running time complexity is $O(|V|d^{j-1})$, where $|V|$ is the number of nodes, d is the maximum node degree, and j is the maximum number of nodes over all considered graphlet types. Static GDV is the static counterpart of dynamic GDV above. There is no static counterpart of GoT.

`c. Raw SMS feature.` For an individual, we count how many SMSs the individual sent or received in each of the 31 weeks, which results in a 31-dimensional node feature vector. This is as similar as possible _non-network_ counterpart of the centralities in the dynamic network.

`d. Feature dimensionality reduction.` When building predictive models, high-dimensional features tend to cause overfitting, meaning that models using such features may fit well on the training data but not predict well on the testing data. Therefore, for each of the features considered above, we generate the corresponding new lower-dimensional feature using principal component analysis (PCA). So, for each of the three dynamic network features, two static network features, and the one non-network feature, we have its pre- and post-PCA versions. Thus, in total, we consider $2 \times (3 + 2 + 1) = 12$ features.

`e. Classification.` For each of the 12 features, we train a logistic regression classifier (we use logistic regression as a proof-of-concept), resulting in 12 classification models for predicting an individual as either depressed or non-depressed (anxious or non-anxious). We consider an additional predictive model – the most accurate HIN-based RS model from our previous work (Section 1),[3] DMF.[46] The above dynamic network-based classification models need to be superior to DMF for it to make sense to incorporate in the future the dynamic network data into the HIN-based framework, per the discussion in Section 1. Note that although DMF could take an HIN as input,[3] in this study, to fairly compare DMF against the above classification models, we use as DMF's input the same (homogeneous) social network from which we extract

the above static network features. In addition, to evaluate the statistical significance of all models' predictive results, we compare them against a random guess model, which works as follows. Recall that in our data, 67 of the individuals are depressed and 207 are non-depressed (106 are anxious and 168 are non-anxious). To make predictions, the random guess model randomly chooses 67 (106) of all individuals and predicts them as depressed (anxious), and it predicts the remaining individuals as non-depressed (non-anxious).

To evaluate the performance of our considered predictive models, we use 5-fold cross-validation (Supplementary Section S4). Given a model's prediction for an individual, taking depression as an example, a true positive (TP) is an individual who is depressed and is also predicted as depressed. A false positive (FP) is an individual who is non-depressed but is predicted as depressed. A false negative (FN) is an individual who is depressed but is predicted as non-depressed. A true negative (TN) is an individual who is non-depressed and is also predicted as non-depressed. Based on these, we compute four popular evaluation measures: precision, recall, F1 score, and accuracy (Supplementary Section S4).

When comparing performance of any two predictive models in terms of their five paired runs of 5-fold cross-validation, we evaluate the statistical significance of their performance difference by using the Wilcoxon signed-rank test. Since for each model we compare its performance against the rest of the considered models, we adjust the p-values as described above.

3. Results and discussion

3.1. *Task 1: Depressed and non-depressed (anxious and non-anxious) individuals occupy different social network positions*

We find that the depressed (anxious) individuals have significantly lower magnitudes of centralities and higher fluctuations of centralities than the non-depressed (non-anxious) individuals (Fig. 1 and Supplementary Figs. S1-S6). The former means that the depressed (anxious) individuals are less central, i.e., are more peripheral, in the social network than the non-depressed (non-anxious) individuals. For example, in terms of degree centrality, the depressed (anxious) individuals have fewer social contacts than the non-depressed (non-anxious) individuals. This result obtained from our social interaction (SMS) data collected via smartphones is consistent with existing knowledge that depression and anxiety are associated with having fewer friends based on social interaction data collected from surveys,[35,47,48] which validates that smartphone data may be a good proxy for real-world friendships. The latter means that centralities of the depressed (anxious) individuals vary more with time than centralities of the non-depressed (non-anxious) individuals. For example, in terms of degree centrality, the depressed (anxious) individuals have many friends at some time points but few friends at other time points, while the non-depressed (non-anxious) individuals have more stable friendships. This result indicates the promise of using dynamic network data over using static network data in the task of predicting mental health, since the latter might fail to capture such temporal variation.

Since the depressed and non-depressed (anxious and non-anxious) individuals have different network positions, it makes sense to use supervised learning methods to predict the individuals' mental health based on their network positions, which we do in task 3 below.
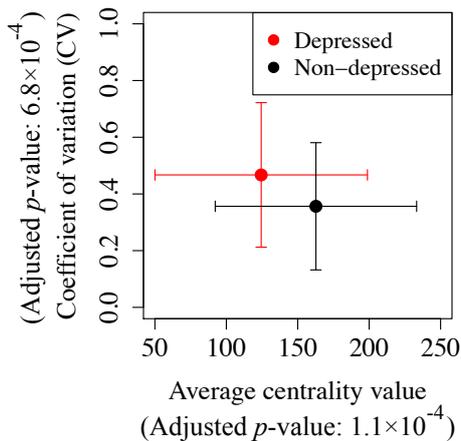
Fig. 1: The difference between centralities of the depressed and non-depressed individuals, in terms of centrality magnitudes (x-axis) and fluctuations (y-axis); see Section 2.2. For each of these two quantities, the average over all depressed (red) or non-depressed (black) individuals is shown, along with the corresponding standard deviation (error bars); the adjusted p-value of the difference is also shown. This figure is for degree centrality and depression. Results are qualitatively similar for the other centralities and anxiety (Supplementary Figs. S1-S6).

3.2. Task 2: Individuals who have different evolving network positions show depression/anxiety trait differences

We find that the individuals in the different network-based clusters (Fig. 2) have different depression (anxiety) traits. Namely, the lower the centrality values (cluster 1 being the least-central), the more likely that individuals are depressed (anxious), and the higher the centrality values (cluster 4 being the most-central), the more likely that individuals are non-depressed (non-anxious) (Fig. 3). While we illustrate these results for degree centrality, results are qualitatively similar for the other centrality measures (not shown due to space constraints).

Thus, network-based features of individuals (their evolving centrality profiles) can distinguish well between depressed (anxious) and non-depressed (non-anxious) individuals. This is an additional confirmation that it makes sense to develop a predictive model of individuals' mental health based on their network structural information, which is what we do next.
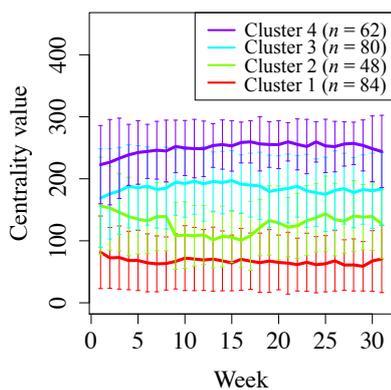

Fig. 2: Centrality values of individuals in the four network-based clusters. For a given cluster, at each time point (week), the average centrality value over all individuals in the cluster is shown, along with the corresponding standard deviation (error bars); also, the number of individuals in the cluster (n) is shown. Clusters are ordered from the one with the lowest centralities (red) to the one with the highest centralities (purple). This figure is for degree centrality-based clusters. Results are qualitatively similar for the other centralities.

3.3. Task 3: The dynamic network has more power to predict individuals' mental health than the static network

Recall from Section 2.2 that we develop a predictive classification model for each of the three dynamic and two static network features, plus the dynamic raw SMS data (non-network) feature. For each feature, we focus on the best of its pre- and post-PCA versions; we have found that using the post-PCA version is more accurate than using the pre-PCA version for
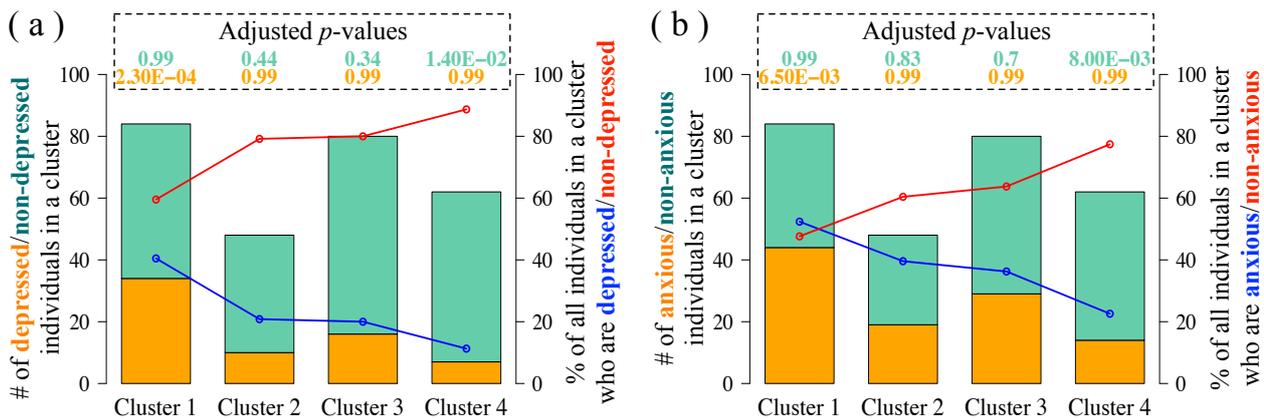
Fig. 3: The enrichment of the four network-based clusters in **(a)** depressed and non-depressed individuals, and **(b)** anxious and non-anxious individuals. In a given panel, the left y-axis shows the number of depressed and non-depressed (anxious and non-anxious) individuals in a given cluster. For example, in panel (a), in cluster 1 containing 84 individuals, there are 34 depressed and 50 non-depressed individuals. The right y-axis shows the percentage of depressed and non-depressed (anxious and non-anxious) individuals in a given cluster. For example, in panel (a), of the individuals in cluster 1, 40% are depressed and 60% are non-depressed (the two percentages add up to 100%). On top of a given bar (corresponding to a given cluster), the adjusted p-values are shown indicating the statistical significance of the enrichment of the cluster in depressed and non-depressed (anxious and non-anxious) individuals. For example, in panel (a), cluster 1 is significantly enriched (adjusted p-value<0.05) in depressed individuals, and cluster 4 is significantly enriched in non-depressed individuals; no other cluster is significantly enriched in either depressed or non-depressed individuals (this is directly related to the fact that the blue values corresponding to the enrichment in depressed individuals decrease going from cluster 1 to cluster 4, and because the red values corresponding to the enrichment in non-depressed individuals decrease going from cluster 4 to cluster 1). These results are for degree centrality-based clusters. Results are qualitatively similar for the other centralities.

each considered feature except one, namely centralities in the static network. Also, we evaluate the static RS model, DMF, and the random guess model. We evaluate the models with respect to four evaluation measures: precision, recall, F1 score, and accuracy.

Our findings are as follows, and they hold for both depression and anxiety, as well as for all four evaluation measures. All three dynamic network feature-based classification models are significantly more accurate (adjusted p-value<0.05) than (i) both of the static network feature-based classification models, (ii) the DMF model, (iii) the non-network classification model, and (iv) the random guess model (Fig. 4 and Supplementary Figs. S7-S8). This confirms the hypothesis of our paper that the dynamic network features have more predictive power than any one of the other features. The three dynamic network feature-based classification models perform similarly, with none of them having perfect performance. This indicates their likely complementarity and a potential promise of developing in the future an ensemble learning approach that would integrate the different dynamic network features.
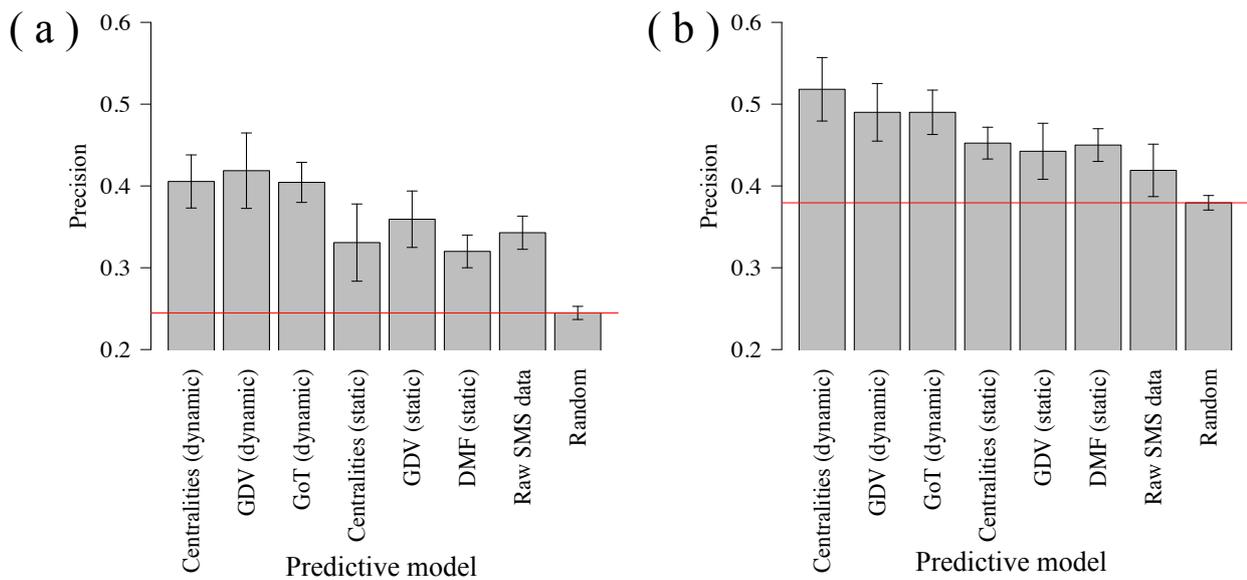
Fig. 4: The performance of the considered predictive models for (a) depression and (b) anxiety. For each model, the performance averaged over the five runs of 5-fold cross-validation is shown, along with the corresponding standard deviation. In parentheses, we show whether the feature in a given model is extracted from the dynamic or static network. The red line corresponds to the performance of the random guess model. These results are for precision. Results are qualitatively similar for the other evaluation measures (Supplementary Figs. S7-S8).

4. Conclusions

In this paper, we develop a predictive model of mental health that uses rich dynamic social network data. We demonstrate that using the dynamic network data has advantage over using its static network equivalent as well as its non-network equivalent in this predictive task. So, our previous study that is the most comprehensive work to date on predicting individuals' mental health in terms of heterogeneity of the considered data,[3] which yielded high prediction accuracy despite using static social network data (which had to be used due to current methodological challenges in the field of heterogeneous network analysis), could be further improved by incorporating the data dynamics. This challenging direction that requires novel algorithmic thinking is the subject of our future work. Other challenges that may warrant future work are as follows. First, none of the considered models are even close to perfect. For example, the highest precision over all models is 42% for depression and 52% for anxiety. So, predictive models need to be further improved before they can be translated to clinical practice. Second, as studies of this type becomes more popular, lots of privacy-related challenges will arise that will need to be addressed in order to protect privacy of individuals whose health-related information is used, especially if the data are to be released for public use in scientific research. Third, in our study, dynamic GDV is the best model in depression prediction. One challenge is to improve its scalability, i.e., the running time of counting dynamic graphlets in a network. Similar was the case with static graphlets when they first appeared over a decade ago. Yet, faster approaches have been developed for static graphlets.[49,50] So, we expect similar

advances to happen for dynamic graphlets, which appeared only several years ago.[40] Fourth, although our study is the largest among all studies of this type in terms of the number of participants, its number of participants is still relatively small. Also, we focus on a population of college students, and hence, it is unclear whether our results are generalizable to other populations. So, an opportunity and a challenge is to collect data from a much larger number of participants and from a more diverse population.

5. Acknowledgements

References

1. N. Malod-Dognin, J. Petschnigg and N. Pržulj, *Current Opinion in Systems Biology* **7**, 1 (2018).
2. V. Gligorijević, N. Malod-Dognin and N. Pržulj, *Proteomics* **16**, 741 (2016).
3. S. Liu, F. Vahedian, D. Hachen, O. Lizardo, C. Poellabauer, A. Striegel and T. Milenković, *arXiv preprint arXiv:1906.04346* (2019).
4. C. A. Latkin and A. R. Knowlton, *Behavioral Medicine* **41**, 90 (2015).
5. J. M. Perkins, S. Subramanian and N. A. Christakis, *Social Science & Medicine* **125**, 60 (2015).
6. R. Pastor-Satorras, C. Castellano, P. Van Mieghem and A. Vespignani, *Reviews of Modern Physics* **87**, p. 925 (2015).
7. T. W. Valente and S. R. Pitts, *Annual Review of Public Health* **38**, 103 (2017).
8. N. K. Cobb, A. L. Graham and D. B. Abrams, *American Journal of Public Health* **100**, 1282 (2010).
9. N. A. Christakis and J. H. Fowler, *New England Journal of Medicine* **2007**, 370 (2007).
10. N. A. Christakis and J. H. Fowler, *New England Journal of Medicine* **358**, 2249 (2008).
11. J. N. Rosenquist, J. H. Fowler and N. A. Christakis, *Molecular Psychiatry* **16**, p. 273 (2011).
12. Y. Youm, E. O. Laumann, K. F. Ferraro, L. J. Waite, H. C. Kim, Y.-R. Park, S. H. Chu, W.-T. Joo and J. A. Lee, *BMC Geriatrics* **14**, p. 102 (2014).
13. S. Liu, D. Hachen, O. Lizardo, C. Poellabauer, A. Striegel and T. Milenković, *Applied Network Science* **3**, p. 45 (2018).
14. S. Lin, L. Faust, P. Robles-Granda, T. Kajdanowicz and N. V. Chawla, *PLOS ONE* **14**, p. e0217264 (2019).
15. M. H. Schafer, *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* **66**, 795 (2011).
16. S. A. Haas, D. R. Schaefer and O. Kornienko, *Journal of Health and Social Behavior* **51**, 424 (2010).
17. L. Meng, Y. Hulovatyy, A. Striegel and T. Milenković, *IEEE Transactions on Network Science and Engineering* **3**, 32 (2016).
18. M. Berk, A. Brnabic, S. Dodd, K. Kelin, M. Tohen, G. S. Malhi, L. Berk, P. Conus and P. D. McGorry, *Bipolar Disorders* **13**, 87 (2011).
19. M. M. Aldarwish and H. F. Ahmad, in *IEEE 13th International Symposium on Autonomous Decentralized System*, 2017.
20. P. D. McGorry, *Journal of Nervous and Mental Disease* **203**, p. 310 (2015).
21. D. C. Mohr, M. Zhang and S. M. Schueller, *Annual Review of Clinical Psychology* **13**, 23 (2017).
22. T. Glenn and S. Monteith, *Current Psychiatry Reports* **16**, p. 523 (2014).
23. S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar and J. C. Eichstaedt, *Current Opinion in Behavioral Sciences* **18**, 43 (2017).
24. A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi and A. S. Pentland, in *ACM International Conference on Multimedia*, 2014.
25. A. Wongkoblap, M. A. Vadillo and V. Curcin, *Journal of Medical Internet Research* **19** (2017).
26. M. De Choudhury, M. Gamon, S. Counts and E. Horvitz, *Seventh International AAAI Conference on Weblogs and Social Media* **13**, 1 (2013).
27. X. Wang, C. Zhang and L. Sun, in *2013 IEEE 13th International Conference on Data Mining Workshops*, 2013.
28. R. Purta, S. Mattingly, L. Song, O. Lizardo, D. Hachen, C. Poellabauer and A. Striegel, in *ACM International Symposium on Wearable Computers*, 2016.
29. L. Faust, R. Purta, D. Hachen, A. Striegel, C. Poellabauer, O. Lizardo and N. V. Chawla, in *International Workshop on Social Sensing*, 2017.
30. Y. Koren, in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009.
31. A. Sano, A. J. Phillips, Z. Y. Amy, A. W. McHill, S. Taylor, N. Jaques, C. A. Czeisler, E. B. Klerman and R. W. Picard, in *IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks*, 2015.
32. J. Bollen, H. Gonçalves, G. Ruan and H. Mao, *Artificial Life* **17**, 237 (2011).
33. M. L. Hatzenbuehler, K. A. McLaughlin and Z. Xuan, *Social Science & Medicine* **75**, 1184 (2012).
34. J. Krause, R. James and D. Croft, *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 4099 (2010).
35. D. R. Schaefer, O. Kornienko and A. M. Fox, *American Sociological Review* **76**, 764 (2011).
36. J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe and A. Pentland, in *ACM Conference on Ubiquitous Computing*, 2012.
37. F. E. Faisal and T. Milenković, *Bioinformatics* **30**, 1721 (2014).
38. D. Aparício, P. Ribeiro and F. Silva, *PLOS ONE* **13**, p. e0205497 (2018).
39. T. Milenković and N. Pržulj, *Cancer Informatics* **6**, CIN (2008).
40. Y. Hulovatyy, H. Chen and T. Milenković, *Bioinformatics* **31**, i171 (2015).
41. T. Milenković, V. Memišević, A. Bonato and N. Pržulj, *PLOS ONE* **6**, p. e23016 (2011).
42. W. S. Noble, *Nature Biotechnology* **27**, p. 1135 (2009).
43. H. Abdi, *Encyclopedia of Research Design* **1**, 169 (2010).
44. N. Pržulj, D. G. Corneil and I. Jurisica, *Bioinformatics* **20**, 3508 (2004).
45. D. Aparício, P. Ribeiro, T. Milenković and F. Silva, *Bioinformatics, doi: 10.1093/bioinformatics/btz119* (2019).
46. L. R. Drumond, E. Diaz-Aviles, L. Schmidt-Thieme and W. Nejdl, in *ACM International Conference on Conference on Information and Knowledge Management*, 2014.
47. F. M. Alpass and S. Neville, *Aging & Mental Health* **7**, 212 (2003).
48. L. K. George, D. G. Blazer, D. C. Hughes and N. Fowler, *British Journal of Psychiatry* **154**, 478 (1989).
49. T. Hocevar and J. Demsar, *Bioinformatics* **30**, 559 (2014).
50. M. Rahman, M. A. Bhuiyan and M. A. Hasan, *IEEE Transactions on Knowledge and Data Engineering* **26**, 2466 (2014).