Abstract
Social interactions have multifaceted effects on individuals’ mental health statuses, including mood and stress. As a proxy for the social environment, Bluetooth encounters detected by personal mobile devices have been used to improve mental health prediction and have shown preliminary success. In this paper, we propose a vector space model representation of Bluetooth encounters in which we convert encounters into spatiotemporal tokens within a multidimensional feature space. We discuss multiple token designs and feature value schemes and evaluate the predictive power of the resulting features for stress recognition tasks using the StudentLife and Friends & Family datasets. Our findings motivate further discussion and research on bag-of-words approaches for representing raw mobile sensing signals for health outcome inference.

Author Keywords
vector space model; Bluetooth encounters; mental health inference; mobile sensing

ACM Classification Keywords
H.1.2 [Models and Principles]: User/Machine Systems; J.3 [Life and Medical Sciences]: Health; J.4 [Social and Behavioral Sciences]: Psychology
Introduction
Social interaction is a key modulator of mental health [3], although passive detection of social interaction in natural settings remains a challenge. Bluetooth encounters are triggered by physical proximity of personal mobile devices, and these encounters have been used as proxy signals of social interaction [6] within models that infer human social interaction and predict health outcomes [13]. Despite these positive results, feature representations for Bluetooth encounter data are still limited. To address this, we propose a vector space representation of Bluetooth encounter data suitable for mental health inference tasks. Vector space modeling originated in the field of information retrieval [10] as an approach to quantifying the content of textual documents. Each document in such a model is a real-valued vector with its elements representing words in the overall vocabulary. The resulting space of documents supports comparison, similarity measurement, and prediction of outcomes such as thematic category.

The present paper draws an analogy between a word in a vector space model of documents and a Bluetooth encounter between a subject’s device and another device in the natural environment. Over time the subject’s device encounters multiple other devices, forming a collection of Bluetooth encounters that are analogous to the text of a document. In broad terms, the resulting encounters proxy for the narrative of a subject’s social interactions. The aim of the present paper is to study the relationship between this narrative and mental health outcomes. Concretely, we convert Bluetooth encounters (time and device identifier tuples) into spatiotemporal tokens and treat each token as a separate dimension in a feature space. Our approach provides a fine-grained representation of a subject’s Bluetooth encounters, and we hypothesize statistical correlations between locations in the vector space and mental health outcomes.

To the best of our knowledge, this paper is the first to apply vector space feature representations to Bluetooth encounter data for mental health inference. We propose and evaluate (1) Bluetooth encounter token designs including different combinations of temporal and spatial information; and (2) feature value weighting schemes such as binary value, term frequency (TF), and term frequency inverse document frequency (TFIDF) based on prediction performance in stress recognition tasks. We also compare our vector space features with two baseline feature groups: the Bluetooth network features studied in [13] and vector space features of non-Bluetooth-encounter mobile sensing data similar to those constructed in [5]. Finally, we propose future work with topic modeling and word embedding methods on our vector space model to discover meaningful clusters of behavioral patterns. With this study we hope to inspire further discussion and research on bag-of-words approaches to representing mobile sensing signals for more effective mental health inference and management.

Related Work
Automated Mental Health Inference
The need for timely, accurate, and unobtrusive mental health management motivates automated mental health inference and prediction using pervasive mobile sensing data. Researchers have found predictive power in a variety of mobile sensor data such as accelerometer, GPS, phone usage, SMS messages, and phone calls, targeting daily and momentary mental health statuses [8]. Most existing research focuses on characterizing a subject’s intrapersonal behavior (e.g., place visits from GPS data, physical activity levels from accelerometer data) [11] whereas information about one’s interpersonal social interactions has received less
attention despite abundant psychological and sociological studies linking such interactions to mental health [3].

Researchers have employed various sensors such as Bluetooth radio, radio-frequency identification (RFID), and infrared sensors to proxy for social interaction. Among these sensors, Bluetooth is the most widely available within mobile phones and does not require additional hardware beyond the phone itself. The primary use of such signals in related research is for reconstructing social networks within a group of individuals [6]. There is also some research using Bluetooth encounter data for mental health inference. Bogomolov et al. [2] implemented a set of nine Bluetooth encounter features covering encounter counts, entropy, and inter-encounter times in a daily stress estimation problem and found the Bluetooth features to be helpful. Wu et al. [13] investigated Bluetooth encounter network features in a momentary stress estimation task and found that these features improved prediction performance compared to several baselines. To the best of our knowledge, no existing work has proposed a vector space model of Bluetooth encounter signals and examined its utility in mental health inference applications, which is our main goal in this paper.

**Our Approach**

**Bluetooth Encounter Vector Space Model**

Within a Bluetooth data stream, an encounter takes the form of a device identifier with a timestamp. Representing these encounters in a vector space entails treating each distinct type of Bluetooth encounter — a distinct device identifier at a distinct time — as a dimension. The rationale behind this is that each Bluetooth encounter may represent a meaningful real-life proximity or interaction event that has implications for a mental health outcome. Concretely, we codify such information as follows. First, we bin all timestamps by hour (e.g., "2013-03-29 00:02:42" would be binned to "hour00") such that Bluetooth encounters that occurred within the same hour are given the same temporal label. Second, we concatenate the hour block information (e.g., "hour00") with the encountered device identifier to create a Bluetooth encounter token, which takes the format of "[hour block]-[device ID]". Alternatively, one may also choose to keep only the device identifier in the tokens and leave out the time, resulting in tokens with only the device...
Figure 1: Illustration of vector space construction procedure. The upper table shows a sample of raw Bluetooth encounter data. The lower table shows the representation of time periods (analogous to documents) as vectors of time-device token frequencies (analogous to term frequencies).

We assign values to the vector space features of Bluetooth encounters following the conventions of text-based vector space modeling. Specifically, we compute a binary token value, a token frequency (TF) value, and a token frequency inverse period frequency (TFIPF) value for each Bluetooth encounter in the vector space. Given multiple periods of time (e.g., multiple days) during which collections (documents) of Bluetooth encounters are detected by a subject's device, the binary token value is defined as 1 when a token appears in a document regardless of frequency and 0 otherwise; token frequency is defined as the number of times a token appears within a particular period of time; and token frequency inverse period frequency is defined as token frequency multiplied by the natural logarithm of the ratio of the number of collections where a particular token appears to the total number of periods. These three sets of feature values will be used for further analyses and compared.

**Token Augmentation with GPS Data**
The encountered device and the encounter time are two pieces of information native to the Bluetooth data stream. One can also incorporate information from external sensors, when available, to enrich the vector space of Bluetooth encounters. In this paper we explore the augmentation of Bluetooth encounter tokens with GPS data and evaluate the resulting predictive value. Ideally, at the time of any Bluetooth encounter, a subject should have a location that is detectable by a GPS sensor; therefore, one can concatenate the GPS information with the [hour block]-[device ID] tokens to create richer tokens representing a Bluetooth encounter event, with one downside being an enlarged feature space. We encode GPS information through coordinate anonymization, which truncates significant digits of a GPS coordinate to map a raw coordinate (a point) to the rectangular area it belongs to such that no finer geographic information is retained. GPS coordinate anonymization is typically performed to protect user privacy in data transmission. In vector space construction it serves a purpose similar to binning timestamps.

**Bluetooth Encounter Network Features**
Creating a vector space model for Bluetooth encounters does not require any information other than what is available from the subject's device. However, in some cases other sensor data are provided by encountered devices. For example, when we have mobile sensing data from all stu-

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Device ID 1 (self)</th>
<th>Device ID 2 (other)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-03-29 00:02:42</td>
<td>u08</td>
<td>04:0C:CE:EB:14:7E</td>
</tr>
<tr>
<td>2013-03-29 00:02:42</td>
<td>u08</td>
<td>04:0C:CE:5C:21</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bluetooth encounter tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>h00_04:0C:CE:EB:14:7E</td>
</tr>
<tr>
<td>h00_04:0C:CE:5C:21</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vector space features</th>
</tr>
</thead>
<tbody>
<tr>
<td>u08</td>
</tr>
<tr>
<td>h00_04:0C:CE:EB:14:7E</td>
</tr>
<tr>
<td>h00_04:0C:CE:5C:21</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
dents who enroll in the same class in college, we will usu-
ally observe a Bluetooth encounter network encompassing
multiple individuals as nodes and encounters between them
as edges. This enriched network provides topological at-
tributes of the local encounter network surrounding a sub-
ject as well as information on the social nature of a subject’s
encountered devices (e.g., are they familiar or fortuitous
encounters, how many common encountered devices two
encountering devices share). We previously proposed and
evaluated three groups of Bluetooth encounter network fea-
tures, namely structural, edge, and neighbor attributes (Ta-
bles 3-5 in [13]). These encounter network features will be
used as a baseline for comparison with vector space fea-
tures in the present paper. Intuitively, Bluetooth encounter
network features take advantage of information about a
subject’s encountered “neighbors” in addition to the subject
itself; therefore, we hypothesize these network features will
produce higher performance than the vector space features,
which only utilize the subject's encounter patterns and thus
assume no knowledge of the behaviors of the encountered
devices.

Baseline Vector Space Features
To measure the predictive value of Bluetooth encounter
vector space features, we construct non-Bluetooth tokens
containing only temporal and spatial information as a base-
line feature group. We divide a day into 24 hourly bins, treat
a subject's anonymized GPS locations as places, and cre-
ate distinct hour-place pairs as separate dimensions in a
vector space. For example, if a subject is detected to be at
place A at 2:36pm, then feature/hour-place token [hour 14]-
[place A] will be created and assigned value 1, otherwise
0. We will compare the performance of these spatiotempo-
rnal features alone versus their combination with Bluetooth
encounter features in stress recognition tasks.

Hypotheses
The following hypotheses examine the predictive value of
our Bluetooth encounter vector space features compared to
non-Bluetooth, baseline features (Hypothesis 1) and Blue-
tooth encounter network features (Hypothesis 4) targeting
personal stress outcomes. The hypotheses also investigate
the implications of vector space design (Hypothesis 2) and
feature value weighting scheme (Hypothesis 3) for predic-
tion performance.

Hypothesis 1: Binary, device-only vector space features of
Bluetooth encounters, when combined with baseline vector
space features, predict personal stress levels more accu-
rately than baseline vector space features alone.

Hypothesis 2: As the vector space token design of Blue-
tooth encounters changes from device-only to device-time
to device-time-location, prediction performance targeting
personal stress levels will increase.

Hypothesis 3: As the vector space feature weighting scheme
changes from binary value to TF to TFIPF, prediction perfor-
ance targeting personal stress levels will increase.

Hypothesis 4: Vector space features of Bluetooth encoun-
ters predict personal stress levels less effectively than the
Bluetooth encounter network features [13].

Data
We use two mobile sensing datasets to test our hypothe-
ses: (1) the StudentLife [12] dataset and the Friends &
Family [1] dataset. The StudentLife dataset was collected
from 49 Dartmouth college students who enrolled in the
same class and contains Bluetooth encounters scanned ev-
y 10 minutes among them over two months. The Friends
& Family dataset was collected from young faculty mem-
bers and their spouses totalling 117 people who lived in the
same residential complex at a major research university in North America and contains Bluetooth encounters scanned every 5 minutes among them over 10 months. Bluetooth encounters in both datasets take the format shown in Figure 1. GPS data are available in both datasets although they are affine-transformed in the Friends & Family dataset.

Ground truth data on personal stress level is obtained through mobile phone surveys in both datasets but with different temporal resolutions. In the StudentLife dataset, stress level measurements are obtained through ecological momentary assessment (EMA) surveys deployed on the participants’ smartphones multiple times per day at random times. The survey consists of a question with text “Right now, I am...” and 5 response options “feeling great”, “feeling good”, “a little stressed”, “definitely stressed”, and “stressed out”. We consider the first two as non-stressed and the latter three as stressed. In the Friends & Family dataset, stress surveys are deployed at the end of each day soliciting an assessment of a participant’s perceived stress level on the past day. The question reads “On a scale of 1 to 7, how stressed were you on [day] (with 1 being very unstressed, 4 being neither stressed nor unstressed, and 7 being very stressed)?” and as the question text suggests, participants are asked to choose among 7 options with 5,6,7 being a stressed response and 1,2,3,4 being a non-stressed response.

Experiments
To test Hypotheses 1-4, we perform predictive modeling targeting a categorical stressed/non-stressed response variable as discussed in Section Data using vector space features constructed following the descriptions in Section Our Approach. For the StudentLife dataset, we treat Bluetooth encounters detected within the 18-hour window leading up to each stress level self-report as a document. For the Friends & Family dataset we treat the entire day corresponding to each end-of-day self-report as the time period associated with each document. We experiment with naive Bayes, support vector machine, and random forest classifiers. Random forest yielded the best results and we will only report performance by random forest in the next section. To evaluate the prediction results we adopt a subject specific leave-one-out cross validation setup where area under the ROC curve (AUC) is computed for each subject in each dataset.

Results
As shown in Table 1, incorporating Bluetooth encounter vector space features (binary valued, device-only tokens) improved prediction performance compared to spatiotemporal baseline features alone for both the momentary and the daily stress recognition tasks. This result supports Hypothesis 1 and our proposed vector space representation of Bluetooth encounter data for mental health inference.

Listed in Table 2 are the average AUC scores achieved by the three vector space token designs and the three feature value weighting schemes with the two datasets. The best performing group is the device-time-GPS tokens with TFIPF feature values for the momentary recognition task using the StudentLife dataset (AUC = 0.714) whereas the best performing group for the daily recognition task using Friends & Family dataset is the simpler binary valued device-only token features (AUC = 0.664). Our results do not support Hypothesis 2 or 3 as more sophisticated token design and feature value scheme did not result in better performance.

Comparing the best performance achieved across the three token designs, we found that device-time tokens performed poorer in both stress recognition tasks than device-only and device-time-GPS tokens. Our results were inconclusive re-
Table 1: Binary valued, device-only vector space features of Bluetooth encounters enhance stress recognition prediction performance (AUC).

<table>
<thead>
<tr>
<th>Features</th>
<th>StudentLife (momentary)</th>
<th>Friends&amp;Family (daily)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hour-GPS</td>
<td>0.688</td>
<td>0.649</td>
</tr>
<tr>
<td>Hour-GPS + Bluetooth</td>
<td>0.714</td>
<td>0.662</td>
</tr>
</tbody>
</table>

Regarding the relative performance of the latter two designs. This suggests that the identity of one’s Bluetooth encounters drives the predictive power and the timestamp of the encounters introduces more noise to predictive modeling than the additional information it provides. In other words, whether one is in proximity of a particular device may matter more than when such proximity events occur, as far as the effect on one’s stress level is concerned. This result also indicates the value of geographic information in mental health inference, as the addition of GPS information in token design improved the performance of device-time tokens.

Comparing the best performance achieved with the three feature value weighting schemes, we found that TF is consistently the worst choice in both tasks with each token design. We expected that TFIPF would perform better than TF as in many text-based problems; however, TF did not outperform binary values in our tasks. Between results obtained with the two datasets, a shared pattern is that device-time tokens with TF feature values appear to be the worst choice when modeling personal stress outcomes, regardless of its temporal scale (momentary versus daily).

Table 2: Predictive performance (AUC) with each vector space token design, feature value scheme, and dataset; the best value for each token design is bolded.

<table>
<thead>
<tr>
<th>Token design</th>
<th>Weighting</th>
<th>StudentLife (momentary)</th>
<th>Friends&amp;Family (daily)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device-only</td>
<td>Binary</td>
<td>0.689</td>
<td>0.664</td>
</tr>
<tr>
<td></td>
<td>TF</td>
<td>0.705</td>
<td>0.642</td>
</tr>
<tr>
<td></td>
<td>TFIPF</td>
<td>0.711</td>
<td>0.654</td>
</tr>
<tr>
<td>Device-time</td>
<td>Binary</td>
<td>0.683</td>
<td>0.634</td>
</tr>
<tr>
<td></td>
<td>TF</td>
<td>0.688</td>
<td>0.636</td>
</tr>
<tr>
<td></td>
<td>TFIPF</td>
<td>0.709</td>
<td>0.643</td>
</tr>
<tr>
<td>Device-time-GPS</td>
<td>Binary</td>
<td>0.709</td>
<td>0.654</td>
</tr>
<tr>
<td></td>
<td>TF</td>
<td>0.683</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>TFIPF</td>
<td>0.714</td>
<td>0.649</td>
</tr>
</tbody>
</table>

Table 3: Predictive performance (AUC) comparison of vector space and network features of Bluetooth encounters.

<table>
<thead>
<tr>
<th>Features</th>
<th>StudentLife (momentary)</th>
<th>Friends&amp;Family (daily)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Vector Space</td>
<td>0.714</td>
<td>0.664</td>
</tr>
<tr>
<td>Network</td>
<td>0.735</td>
<td>0.631</td>
</tr>
<tr>
<td>Network + Best Vector Space</td>
<td>0.760</td>
<td>0.654</td>
</tr>
</tbody>
</table>

As well as the two groups combined. For the momentary problem, Bluetooth network features performed better than vector space features, supporting Hypothesis 4 that information about behaviors of the devices (and by extension their users) encountered by a subject carries information regarding the subject’s mental health status. When the two feature spaces are combined, they achieved better performance than each alone. For the daily task, the comparison between vector space and network features is reversed. We suspect that the differences in these comparisons results
from the different degree of closeness of the social relationships between study participants in the two datasets. In the StudentLife study, participants are undergraduate students enrolled in the same class who likely have integrated social lives. We observed many non-participant Bluetooth encounters shared by these subjects. In contrast, in the Friends & Family study, the tie between participants is that they live in the same apartment complex, which might not include daily socialization events. Although Bluetooth encounter features can be computed regardless of the closeness of the network, the characteristics of a subject’s Bluetooth encounter network are likely more indicative of health outcomes when the Bluetooth encounter network reflects proximity events with a higher proportion of a subject’s social contacts. Testing this hypothesis is beyond the scope of this paper; however, this finding (1) validates the utility of vector space representation of Bluetooth signals especially when we do not have access to proximity network data of a subject’s close social contacts (as in the case of Friends & Family dataset); and (2) motivates smart health researchers and practitioners to incorporate Bluetooth encounter network features with vector space features proposed in this paper when data are available for a relatively close-knit group.

Concluding Remarks
This work proposes a vector space representation of Bluetooth encounter data for mental health inference and measures predictive utility in stress recognition tasks with two public datasets. Our results support Hypothesis 1 regarding the value of vector space representations of Bluetooth encounter data. Our results do not support Hypothesis 2 or 3, revealing implications of token design and feature value scheme for prediction performance. Lastly, our results partially support Hypothesis 4, indicating that closeness of social relationships could be a factor in Bluetooth encounter network features’ predictive advantage over vector space features.

We envision several directions of future research, as a vector space representation of Bluetooth encounters opens the door to several additional methods. Topic modeling might be used to extract clusters of proximity events that convey social insights. In particular, we anticipate the applicability of structural topic modeling (STM) [9], which allows the effect of a document-level covariate on the topic composition to be modeled. A fluctuating mental health status associated with a period of time from which a collection of Bluetooth encounters took place naturally serves as a document-level covariate. How such covariates (e.g., high or low stress levels) correlate with topical composition of Bluetooth encounters would be an interesting avenue of exploration. Moreover, word embedding methods [7] might be used to compute a weight vector for each token in a vector space, with higher degree of similarity between weight vectors indicating higher likelihood of token co-occurrence. Such weights may provide a useful representation for understanding human behavior and predicting health outcomes. Finally, the applicability of vector space representation extends to many mobile sensors beyond Bluetooth encounters (e.g., accelerometer and audio). Future discussion and research on bag-of-words approaches to representing raw mobile sensing signals for intelligent health inference would further validate and extend our preliminary results.

REFERENCES

2. Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Sandy Pentland. 2014. Daily stress recognition from mobile phone data, weather conditions


