The investigation of relationship between Depressive Behavior and Internet Using among Students of Tehran University

Reza Ahmadi[1], Mohammad Khan ahmadi[2], Maryam Malmir[3], Mahtab Sabzevari[4]

[1:] MA in Exceptional Children, Allameh Tabatabaei University, Iran
[2:] MA in Psychology, Allameh Tabatabaei University, Iran
[3:] PhD Student of Exceptional Children, Islamic Azad University, Iran
[4:] PhD Student of Exceptional Children, Islamic Azad University, Iran
khanahmadimohammad14@yahoo.com

Abstract: In this paper, we report our results on a month long experiment conducted at Tehran University on associating depressive symptoms among university students and Internet using real Internet data collected continuously, unobtrusively and preserving privacy. In our study, 300 undergraduates were researched for depressive symptoms using the LIPE grade. We then collected their on-campus Internet using via Cisco Net Flow records. Subsequent analysis revealed that several Internet using features like average boxes per flow, peer-to-peer (octets, boxes and duration), chat octets, mail (boxes and duration), ftp duration, and remote file octets exhibit a statistically significant correlation with depressive symptoms. Additionally, Mann-Whitney U-tests revealed that average boxes per flow, remote file octets, chat (octets, boxes and duration) and flow duration entropy demonstrate statistically significant differences in the mean values across groups with and without depressive symptoms. To the best of our knowledge, this is the first study that associates depressive symptoms among university students with continuously collected real Internet data.

Keywords: depression, internet, health of mental, Tehran university students

1. Introduction

There are treatments for depression, many victims do not recognize symptoms, and many may be reluctant to seek help. If left untreated, depression can cause appetite loss, sleep disorders, fatigue and anxiety, along with poor academics and higher dropout rates. Detecting depressive symptoms early is hence a critical need today in our university today. In this paper, we report our results on a month long experiment conducted at Tehran University of Science and Technology on associating depressive symptoms among university students with Internet using real campus Internet data collected continuously, unobtrusively and preserving privacy. Depression is a serious health of mental problem affecting a large segment of society today, and particularly university students. In a research by the Centers for Disease Control (CDC) in 2009, 26.1% of students nationwide reported feeling so sad or hopeless almost every day for 2 or more weeks in a row that they stopped doing some usual activities (Joyce and Weibelzahl, 2006). Similar statistics are also reported in health of mental studies by the American University Health Association, and by independent research (Wantland et al., 2004).

Recent studies show that more than 90% of university students in the US actively use the Internet (Anderson and Cuijpers, 2009). While the benefits of Internet for academic learning, research, business and social networking are well known, studies conducted by the Psychological Sciences community have focused on exploring relationships between Internet use and students' health of mental. Studies in (Titov and Dear, 2011) demonstrated that students with depressive symptoms used the Internet much more than those without symptoms. It was also shown that when the Internet was utilized for activities like shopping, depressive symptoms among students increased (Michael and Michael, 2011).

While all of the above studies do provide critical insights into how Internet associates with depressive symptoms among university students, the information they convey is limited. This is because student Internet using in existing studies has been assessed by means of self-reported researchs only. In other words, students themselves reported their volume and type of Internet activity. This method has limitations. First, the volume of collected Internet using data is limited during researching because people's memories fade with time. There may be errors and social desirability bias when students report their own Internet usage. An accurate characterization of Internet using requires representations of significantly higher dimensionality,

and clearly the number of dimensions that can be captured via research is limited.

We conducted a study in 2012 for associating depressive symptoms among university students with their real Internet using data collected continuously, unobtrusively and preserving privacy 1 at Tehran S & T. To the best of our knowledge, this is the first study to do so. The study consisted of the following steps:

Participant Selection and Researching: We recruited 300 students from three undergraduate classes at Tehran S & T in February 2011. The depressive symptoms of participants were quantified using the Center for Epidemiologic Studies Depression (LIPE) grade (Edwards, 2007). In our research, 30% of students met the minimum LIPE criteria for exhibiting depressive symptoms, which compares well with much recent health of mentalresearch.

The Internet using activity of participants was obtained in the form of Cisco Net Flow records collected over the Tehran S & T campus network. For each participant, we derived a number of Internetsusingfeatures divided into three broad categories. The Aggregate category captures raw aggregates of Internet using like flows, boxes, octets, durations etc. The Application using category captures application specific Internet using features like chatting, peer-to-peer, email, ftp, http etc. The Entropy based features captures randomness in Internet using from the perspective of flows, octets, boxes, durations etc.

Subsequent statistical analysis revealed that the following Internet using features correlate with depressive symptoms: average boxes per flow, peer-to-peer (octets, boxes and duration), chat octets, mail (boxes and duration), ftp duration, and remote file octets. Additionally, Mann-Whitney U-tests revealed that average boxes per flow, remote file octets, chat (octets, boxes and duration) and flow duration entropy demonstrate statistically significant differences in their mean values across groups with and without depressive symptoms.

We present preliminary interpretations to our results by integrating them with existing research in Psychological Sciences on associations between depressive symptoms and Internet using among university students.

The remainder of this paper is organized as follows. Section II describes the Data and Methodology. The results are in Section III. Section IV describes conclusion.

2. Material and Methods

In our study, the participant pool consisted of 300 undergraduate students at Tehran from three classes: Psych 48 (General Psychology), CS 246

(Operating Systems) and CS 167 (Data Structures). Psych 48 is taken by students from all departments, while CS 246 and 167 are taken by students from many engineering departments. The research was preceded by a consent form, and there was a minimum age of at least 18 years to participate. The research was conducted in February 2012.

The levels of depressive symptoms among participants were quantified with a one-time research based on the Center for Epidemiologic Studies Depression (LIPE) grade. The LIPEgrade was developed by Lenore Radloff of Utah State University and is used to measure depression levels in the general population (Morahan-Martin and Schumacher, 2003). It consists of 20 questions rated on a 4-point Likertgrade. Possible scores range from 0 to 60, with higher scores indicating greater levels of depressive symptoms. In general, a score of 16 or above on the LIPEgrade is considered indicative of depressive symptoms. The LIPEgrade is widely used and has been extensively tested and validated. It has been shown to be reliable when testing adolescents in high schools and university (Bonettiet al., 2010), although only the LIPE items were scored.

To ensure privacy of participants, appropriate anonymization techniques were enforced during participant selection, researching and collecting Internet using data. The IT department at Tehran S & T provided unique pseudonyms for each participant, and the associations were not disclosed to the research team. Students who completed the LIPEresearch did so using only their pseudonyms, which were tied to their recorded LIPE scores. The IT department remained unaware of the LIPE scores. Additionally, the IT department provided the on-campus Internet using data indexed only by pseudonyms. The only associations available to the researchers were between Internet using data and LIPE scores. In our study, IP addresses were not processed, since the focus was on broad Internet statistics alone 2. Also, the contents of emails, chat and ftp uploads/downloads were not recorded due to privacy considerations.

3. Results

The main source of "Internet Usage" data for this study was Net Flow. Cisco Net Flow technology is a protocol for collecting IP traffic information and is popular. Net Flow data consists of several flows. In our study, Net Flow V5 was used, which contains the following eight fields for each flow after preprocessing:

1) Source IP address,
2) Destination IP address,
3) Source port,
4) Destination port,
5) Protocol,

6) Octets,

7) Boxes and

8) Duration.

The IT department at Tehran S & T collects Net Flow data of all users for troubleshooting network connections and policy enforcement. The Tehran S & T campus has a connection to both the standard commodity Internet and the Internet 2 education research network. Both Internet and Internet 2 traffic pass through the same router where Net Flow statistics recording and exporting are enabled. Every five minutes, these flows are exported from the router to a collector where they are stored for a period of 45 days for analysis purposes before being discarded automatically.

In order to obtain the Net Flow data of participants, the flows pertaining to each participant were identified based on the source IP field, and subsequently filtered and logged to a secure remote server at the end of every month. As the Tehran uses a DHCP (Dynamic Host Configuration Protocol) to provide IP address, the IP address used by a participant at one time could be used by someone else later. Therefore, the extraction process begins by creating a mapping file and associating each user with a set of assigned IP addresses, along with the start and end time stamps. This information is used by a backup daemon to extract user-specific Net Flow information by filtering flows based on the source IP field. The mapping file is created by analyzing DHCP logs that include a participant's user-id, which is that participant's campus email address. Note that this process was executed by the Tehran IT department. This process was completely automated. Subsequently, the Internet using of each participant indexed by appropriate pseudonyms (as discussed in Section II) was delivered to the research team. In this study, the Internet data used was the one collected in February 2012, the month in which the depressive symptoms of participants were researched.

In order to derive meaningful statistics, we have to preprocess Net Flow data for each participant into an N-dimensional feature vector. Also, as the number of rows associated with a participant approaches millions when aggregated over a month, preprocessing also compresses the data into manageable proportions. As the space of all possible feature vectors is large, care must be taken to extract features that are likely to associate with depressive symptoms. Inspired by related research in the Psychological Sciences community (as discussed in Section I), we derived three broad features of Internet using presented below.

The simplest feature is a representation of overall aggregate traffic statistics, such as total boxes, flows and octets. Although the granularity is low,

these features can be used to answer questions like: "Does more Internet using associate with increased depressive symptoms"? In our study, aggregate flow statistics were derived using the flow-report in the flow-tools suite. Additionally, bash scripting was used to extract the data and convert it into a feature vector, one per participant.

Traffic aggregation alone has low granularity. For example, an aggregate of high email and low chatting may appear similar to an aggregate of low email and high chatting. Application-level statistics capture more information by sub-categorizing aggregate traffic features by application. In other words, traffic features such as flows, octets, boxes and duration are derived per application such as http, email, peer-to-peer (p2p), chat, etc.

A total of 61 applications were identified by filtering flows based on set combinations of destination port and destination protocol fields, as allocated by IANA (Internet Assigned Numbers Authority) (Kim et al., 2009). Since Net Flow data was only logged for on-campus Internet usage, some application categories like socks, squid, and blubster, showed little or no activity. Universities tend to block such services due to security and copyright issues, and students also tend to limit such activities on campuses. In our study, 25 applications were hence retained. C. Entropy Based features.

Difficulty concentrating or making clear decisions is a symptom of depression among university students. We capture randomness in Internet using via Shannon Entropy (H). Intuitively, entropy estimates the average uncertainty of a series of discrete events. Given a discrete random variable X, Shannon entropy H (X) is:

Statistical analysis was performed to correlate the Internet using data collected, with LIPE scores (both collected in February 2012). For each feature derived, Pearson's, Spearman Rho, and Kendall tau-b correlation coefficients were determined. Additionally, T-tests were attempted to identify Internet using features that significantly differentiated participants exhibiting depressive symptoms from those that did not. The T-test assumes a normal data distribution and homogeneity of variance. Normality was verified by observing P-P plots, while Levene's test was used to assess the equality of variance. If the data deviated from a normal distribution, the non-parametric Mann-Whitney U-test was used.

In this section, we present some practical interpretations to our results by integrating them with existing research in the Psychological Sciences community on associations between depressive symptoms among university students and Internet usage.

Average boxes per flow: The average box per flow is high when a large number of boxes are generated per flow. Larger number of boxes per flow is typical under Internet streaming and downloading, which is common when watching videos and gaming. This is intuitive, as gaming and video watching are common symptoms of Internet addiction that have been shown to associate with depressive symptoms (Boals, 2010).

Peer-to-Peer usage: The correlation observed between peer-to-peer using and depressive symptoms is intuitive. Sharing files like music, movies, photos etc. are primary reasons for using peer-to-peer services. Students are prone to be addicted to such kinds of content, which may explain this trend.

Chatting: Excess online chatting can affect the psychology of young people in terms of causing social isolation and loneliness in the real world, potentially leading to depressive symptoms. People with depression are also known to join "Depression Chat Rooms" to overcome feelings of isolation. This may explain Chat octets being significantly high for students with depressive symptoms.

Email: Excessive email using identified in our study as statistically correlating with depressive symptoms is supported by studies in. Frequent email checking may relate with high levels of anxiety, which in-turn correlates with depressive symptoms. It is also theorized that email addiction is a form of impulsive-compulsive disorder in the sense that victims (especially young people) suffer from a compulsive and irresistible need to check messages (often even in the middle of the night).

Flow Duration Entropy: As discussed before, difficulty concentrating or making clear decisions are symptoms of depressive behavior among students. When Flow Durations have high entropy, it is likely a result of frequent switching among multiple Internet applications, which is likely to result in highly variable flow durations, and hence high Entropy. Frequent switching may also reflect an attempt to elevate feelings in the face of Anhedonia, when there is desperation to find something - an interesting article, an e-mail, a pleasing video, etc., to derive a momentary spark of pleasure and elevate mood.

Ftp and Remote File usage: It is not completely clear why ftp duration and remote file octets correlate with depressive symptoms. One interpretation could be that since excess ftp using and remote file octets are indicative of excess file transfers, this could indicate addiction to certain types of files that may associate with depressive symptoms. In our study, we do not access the content of files exchanged, and hence we are limited in the nature of conclusions derived here. Interestingly though, ftp boxes and ftp octets did not show statistically significant correlations; only the ftp duration did. Our on-going studies attempt to further explain these trends based on more discussions with counselors, clinical psychologists and educators, and with more experiments with larger subject sizes.

B. Applications of our Results

There are a number of applications of our methodology and results from this paper. We present three below.

Investigating associations between other health of mental disorders and Internet usage: Our methodology is general and can be used to study associations between Internet using and other health of mental disorders like anorexia, bulimia, ADHD, schizophrenia etc. We could also investigate associations between other Internet features like visits to social networking sites, late night Internet use, and randomness in Internet using times etc. with depressive symptoms.

Proactively discovering depressive symptoms from passive and unobtrusive Internet using monitoring: Using the correlating Internet using features derived in our study, we are currently investigating algorithmic techniques to proactively discover depressive symptoms among students by passive, unobtrusive and run-time monitoring of Internet usage. To do so, we are planning to conduct largergrade studies. However, there are practical concerns in terms of false positives and negatives, along with concerns on ethics and privacy of subjects in the realm of detection. While we believe that the techniques developed can assist in early, personalized and (possibly) in-home health of mental care, we believe that a number of stakeholders from multiple disciplines and organizations need to be involved prior to their practical deployment.

Designing Internet (or Computer) based interventions for depression: There are many recent studies exploring Internet based intervention strategies for alleviating depression. Our results in this paper could yield new insights on designing and administering effective Internet based interventions for mental disorders. Our work can also enable run-time adaptation of intervention strategies based on severity of symptoms for a subject. Furthermore, our results will impact the evaluation of Internet based intervention strategies. With our results, one could easily test the efficacy of Internet based intervention strategies by verifying corresponding changes in correlating Internet using features identified in this study. This, we believe will positively impact the design of effective Internet based interventions in the future.

4. Discussions

In this paper, we report results from a study conducted at Tehran S & T on associating depressive symptoms among university students with their Internet usage. We have identified that average boxes per flow, peer-to-peer (octets, boxes and duration), chat octets, mail (boxes and duration), ftp duration, and remote file octets show statistically significant correlations with depressive symptoms. Additionally, Mann-Whitney U-tests revealed that average boxes per flow, remote file octets, chat (octets, boxes and duration) and flow duration entropy demonstrate statistically significant differences in the mean values across groups with and without depressive symptoms. To the best of our knowledge, this is the first study associating Internet using with depressive symptoms among university students using real Internet data collected in an continuously, unobtrusively and preserving privacy.

Corresponding Author:
Mohammad Khan ahmadi
MA in Psychology, Allameh Tabatabaei University, Iran
khanahmadimohammad14@yahoo.com

References

1. A. Boals. Correlates of Video Game Addiction. PhD thesis, University of North Texas 2010.
2. D. Joyce and S. Weibelzahl. Text messaging as a means to lowering barriers to help-seeking in students with depression, Proceedings of IADIS International Conference e-Society, Dublin, Ireland, 2006.
3. D.J. Wantland, C.J. Portillo, W.L. Holzemer, et.al. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. Journal of Medical Internet Research, 2004.
4. G. Anderson and P. Cuijpers. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cognitive Behaviour Therapy, 2009.
5. J. Kim, R. La Rose, and W. Peng. Loneliness as the cause and the effect of problematic internet use: The relationship between internet use and psychological well-being. Cyber Psychology & Behavior, 12(4):451–455, 2009.
6. J. Morahan-Martin and P. Schumacher. Loneliness and social uses of the Internet. Computers in Human Behavior, 19(6):659–671, 2003.
7. L. Bonetti, M.A. Campbell, and L. Gilmore. The relationship of loneliness and social anxiety with children's and adolescents' online communication. Cyber Psychology, Behavior, and Social Networking, 13(3):279–285, 2010.
8. M.G. Michael and K. Michael. The Fall-Out from Emerging Technologies: on Matters of Surveillance, Social Networks and Suicide. IEEE Technology and Society Magazine 30.3 (2011): 15-18.
9. N. Titov and B. Dear. Trans-diagnostic internet treatment for anxiety and depression: A randomized controlled trial. Behavior Research and Therapy, 2011.
10. S Edwards, L Tinning, JSL Brown, Reluctance to seek help and the perception of anxiety and depression in the United Kingdom: a pilot vignette study. The Journal of Nervous and Mental Disease, 2007.

12/25/2018