An Empirical Study on Symptoms of Heavier Internet Usage among Young Adults

Sai Preethi Vishwanathan*, Levi Malott†, Sriram Chellappan*, P. Murali Doraiswamy†

*Department of Computer Science
Missouri University of Science and Technology
Rolla, Missouri 65409.
{svmc9, lmn93, chellaps}@mst.edu

†Department of Psychiatry and Behavioral Sciences
Duke University
Durham, North Carolina 27708.
murali.doraiswamy@duke.edu

Abstract—Understanding negative consequences of heavy Internet use on mental health is a topic that is gaining significant traction recently. A number of studies have investigated heavy Internet usage, especially among young adults in relation to online games, social media and email. While such studies do provide valuable insights, Internet usage so far has been characterized by means of self-reported surveys only that may suffer from errors and biases. In this paper, we report the findings of a two month empirical study on heavy Internet usage among students conducted at a college campus. The novelty of the study is that it is believed to be the first to use real Internet data that is collected continuously, passively and preserving privacy. A total of 69 Computer Science freshman students were surveyed for symptoms of heavy Internet usage, using the Internet Related Problem Scale, and their campus Internet usage was monitored (after appropriate anonymization procedures to maintain subject privacy). Statistical analysis revealed that several Internet usage features, such as instant messaging, entropy, gaming, web browsing, peer-to-peer usage, remote usage, and email usage exhibit significant correlations with symptoms of Internet addiction like introversion, craving, loss of control and tolerance. Although the study found that Facebook and Twitter usage did not show significant statistical correlations with symptoms of heavier Internet usage, it was found that students tending towards heavier Internet usage used those websites less. We believe that this study provides critical new insights into symptoms of heavier (possibly addictive) Internet usage among young adults, which is now a topic of significant concern to the mental health community today.

Keywords—Addiction, Mental Health, Internet, Privacy

I. INTRODUCTION

Addiction is categorized as continued use of a mood altering substance or behavior despite adverse dependency consequences, or a neurological impairment leading to such behavior [1]. Within this context, behavioral addiction is defined as a compulsion to repeatedly engage in an action to the point where it causes serious negative consequences to various aspects of an individual’s well-being [2]. Recently, one area of research in this realm has been addiction to technology, specifically, the Internet [3], [4], [5], [6], [7], [8].

A. Related Studies on Technology and Internet Addiction

Griffith defined technological addictions as possibly behavioral, due to the lack of chemical substance involved [3]. Subsequent studies by Shotton with computer programmers as subjects introduced the notion of dependents as those who had difficulty controlling their computer use. Also, they tended to be highly educated, had poor social skills, and needed positive intellectual stimulation [5]. With the subsequent wide and pervasive use of the Internet, the issue of addiction to Internet has become a topic of research. In 1997, Young surveyed about 400 adults for problematic Internet use using the adapted DSM-IV criteria for substance abuse [4]. She found that dependent users reported a general loss of control over abilities to restrict their usage, and impairment in certain areas of their daily functioning like academic, relationship, financial etc. More recently though, the issue of young adults affected by excess Internet use has received particular attention. Puekert et. al., reported that up to 3.5% of German teens demonstrate symptoms of excessive Internet use [6]. Konstantinos et al. showed that potential addictive Internet use among young adults in Greece has a prevalence rate of 8.2%, with a majority of males engaging in excessive online gaming [7]. Park et.al., in a 2008 study report that up to 11% of South Korean youth are considered to be at high risk for addictive Internet use [8].

The significance of these studies stems from the fact that excessive Internet has been linked to a variety of negative psychosocial consequences, such as somatization, obsessive-compulsive disorder, depression, anxiety and psychotism [9], [10], [11], which can be particularly dangerous for young people who are at the forefront of technology use. Identifying symptoms of addiction, assessment tools, mental health consequences, and early intervention strategies are all of significant importance to the mental health community today.

B. Contributions of this paper

The goal of this study is to further understand heavy Internet usage among young adults and aid towards characterizing potentially harmful usage. While existing studies on this topic (presented above) do provide significant conclusions, they are limited because Internet data collected was by means of self-reported surveys only, which tends to suffer from
human errors, memory limitations, social desirability biases, selection biases, and the inability to capture high dimensional Internet data. To overcome these limitations, we conducted a two month empirical study on heavy Internet usage among college students conducted at Missouri University of Science and Technology (Missouri S&T), which we believe is the first study to use real Internet data that is collected continuously, passively and preserving privacy.

A total of 69 Computer Science freshman students were surveyed for symptoms of heavy Internet usage, using the Internet Related Problem Scale [12], and their Internet usage from the campus network was monitored (after appropriate anonymization procedures to maintain subject privacy). Subsequent statistical analysis revealed that several Internet usage features, such as instant messaging, entropy, gaming, web browsing, peer-to-peer usage, remote usage, and email usage exhibit significant correlations with symptoms of Internet addiction like introversion, craving, loss of control and tolerance. Although the study found that Facebook and Twitter usage did not show significant statistical correlations with symptoms of heavier Internet usage, it was found that students tending towards heavier Internet usage used those websites less.

We believe that this study provides critical new insights into symptoms of heavier (possibly addictive) Internet usage among young adults. While more studies are needed in this realm, there are immediate consequences of the study in this paper. We demonstrate how high dimensional and high volume “big” Internet data when processed appropriately can provide insights into heavy (possibly) addictive Internet usage. The study hence paves the way for deriving markers that can assist in early diagnosis and intervention of addictive Internet use, which is particularly important for teenagers and children today who are amongst the most active Internet users today. By integrating results from this study with related work on how Internet impacts other mental disorders, like depression, stress, anxiety etc., we believe that results of significant value can be made possible to the mental health community on the complex relationships between mental disorders and the Internet today.

II. METHODS

A. Subject Selection

The study was conducted at Missouri S&T in October 2012 by selecting 69 freshmen enrolled in an introductory computer science course. Out of these, 66 were male, and 3 were female. Participants completed the Internet Related Problem Scale (IRPS) [12]. All subjects were 18 years or older and consented to the study. We point out that participants were assigned unique pseudonyms during both surveying and collecting Internet data which were then appropriately linked during analysis to ensure participant non-identifiability.

B. IRPS Scale

In order to assess the degree of heavy Internet usage, we employed the 20-question Internet Related Problem Scale (IRPS) [12]. Adapted from the DSM-IV criteria for substance abuse, the questions in IRPS cover the issue of tolerance, craving, withdrawal, negative life consequences, loss of control, time spent on related Internet activities, and reduction of other activities. Question responses are scored on a Likert scale ranging from 1 (never) to 10 (very frequent). The responses demonstrated good internal consistency with a Cronbach’s alpha of 0.859, which is consistent with previous studies on problematic Internet usage [12], [13].

III. INTERNET DATA PROCESSING

The IT infrastructure of Missouri S&T utilizes Cisco routers to collect and monitor NetFlow data. Internet packets are recorded at these routers and organized into flows. Collected flows are exported to a central location where authorized network administrators can access the data for troubleshooting network connections and policy enforcement. The NetFlow records contain the source/destination IP address of flows, but have no information regarding users or content. To associate specific flows to users, DHCP (Dynamic Host Configuration Protocol) logs provide IP address mappings to specific single sign-on (SSO) user names. DHCP servers issue IP addresses to users for certain periods of time, later becoming available for other users to obtain. To manage the complexity of cross-referencing the DHCP logs, automated scripts were created to generate per-user filters for querying the database. As a result, each user is associated with a specific database identified by their pseudonym. Figure 1 illustrates the collection and processing overview.

The data collection period started on October 1, 2012 and concluded on November 30, 2012. The amount of data contained in a single subjects NetFlow database, after two months of collection, often exceeded a million individual flow records (often more than 500 MB per subject). This necessitated preprocessing data into manageable portions while minimizing the amount of important information lost. To characterize Internet activity, three categories of features were extracted to represent participant Internet usage, namely aggregate, application, and entropy-based traffic features.

A. Aggregate Traffic Features

To assess heavy Internet usage, the most straightforward feature is to aggregate each one of the flow attributes individually. Each flow record contains three important spatial quantities: octets, packets, and duration. Octets are equivalent to bytes, which measure how much information was transferred in the flow. Packets contain some number of bytes that constitute an amount of useful information. Duration indicates how long the flow lasted. Four variables are derived from these attributes and their value is the sum of all respective flow entries. Table I outlines the collected features and a short description for each.

B. Application Traffic Features

Raw aggregates while providing useful information mask out fine-grained application features like gaming, email and...
chatting usage that are otherwise very useful to study. To separate out usage of specific applications, destination ports and protocol numbers were used to discriminate application flow records. Identifying applications included referencing the Internet Assigned Numbers Authority [15] and online technical documentation. Additional programs were created to parse the compiled protocol/port-to-application file and tag each individual flow record with a specific application or “unknown”. Social networking usage was determined by matching packet source and destination IP addresses to those owned by Facebook or Twitter. Identified applications were grouped into peer-to-peer (P2P), streaming, chat, remote, HyperText Transfer Protocol (HTTP), mail, file transfer protocol (FTP), Voice-over-IP (VoIP), gaming, and social networking; as shown in Table II. The descriptions in Table II details some examples of applications included in respective groups. Calculated for each group were the aggregate of octets, packets, number of flows, and duration.

C. Entropy based Features

Exploring randomness or unpredictability in Internet usage may provide information on specific habits of participants with heavy usage. Randomness in Internet usage is realized by computing the Shannon Entropy (H) of some variables. Intuitively, entropy estimates the average uncertainty of a series of discrete events. Given a discrete random variable \( x \), Shannon Entropy \( H(x) \) is computed as:

\[
H(x) = - \sum_x p(x) \cdot \log(p(x)),
\]

where \( p(x) \) is the probability of event \( x \) occurring. As \( p(x) \to 1 \) the \( \log(p(x)) \to 0 \), indicating that events with higher probability have lower entropy. The Shannon Entropy for Source IP, Destination IP, Destination Port, Octets, Packets and Flow Duration were calculated in this study.

IV. Statistical Analysis

To obtain a measure of association between Internet usage features and symptoms of heavy Internet usage, tau tests were used to obtain Kendall Tau-b (\( \tau_b \)) correlation coefficients. The Kendall Tau-b coefficient was chosen to determine associations as the corresponding tau test is non-parametric. The captured Internet data can vary widely among individuals making normalizing data difficult, which is required for parametric tests. The tests were performed under the null hypothesis that the dependent variables have no association with the independent variables (\( H_0 : \tau = 0.00 \)). Correlations are presented in Figure 3, where only values significant at the 0.05 level (\( \alpha = 0.05, 2\text{-tailed} \)) are shown. Variables with insufficient
evidence to reject the null hypothesis have been marked with an “x” in Figure 3. More details on the correlations and corresponding discussions are discussed in the next section.

Mann-Whitney U tests were used to determine if students in the higher range of IRPS scores used certain applications different from students scoring lower. The Mann-Whitney U test is a statistical test to determine significant differences in mean values of two populations. In previous studies, the IRPS has not had any specific threshold to separate participants into groups. For analysis purposes in this study, we selected a threshold by observing overall score distributions, shown in Figure 2, and determining the value where there is a noticeable score gap. From Figure 2, it is seen that a gap exists near the total score $TS = 110$ in the IRPS, which enables the separation the participants into normal and high Internet usage groups, denoted by $N$ and $H$ respectively. Group $N$ contained 60 participants while Group $H$ contained 9. The focus of the Mann-Whitney U tests are to determine whether higher scoring participants use specific Internet features less than the lower scoring participants. The main focus of previous studies is determining which Internet applications heavy Internet users are frequenting. Significant information may be gained by testing the which Internet features heavy users are using less than normal users. Consequently, one-tailed Mann-Whitney U tests were performed to identify those features and results are presented next.

V. RESULTS AND DISCUSSION

A. Statistical Correlations

1) Total Score: The total score of participants showed a positive correlation with remote_flows ($\tau_b = 0.17$) and duration_total ($\tau_b = 0.16$). As expected, there is a statistically significant association between the total IRPS scores and amount of time spent online. This is in agreement with the result of many current studies [3], [4], [6], [8]. The positive correlation of remote_flows is not immediately clear, although it is noted that universities host network shares for individual users and a variety of remote-enabled computers for student use. Computer science students tend to be at the forefront in

![Fig. 2. Histogram of Participant Internet Related Problem Scale Scores](image)

![Fig. 3. Results of Kendall Tau-b ($\tau_b$) Tests Between Internet Usage and IRPS Responses](image)
using such services for a variety of purposes including academics, content downloading, content sharing etc. Additional studies are needed to explain this correlation.

2) Introversions: Introversions was measured by the student’s scaled response to if they feel more comfortable with objects than people. The following features showed positive correlations with introversion: HTTP ($\tau_{b,avg} = 0.22$), gaming ($\tau_{b,avg} = 0.20$), remote ($\tau_{b,avg} = 0.22$), aggregate totals ($\tau_{b,avg} = 0.26$), packets-per-flow entropy ($\tau_b = 0.21$) and duration_entropy ($\tau_b = 0.18$). Nearly every application group had at least one feature that significantly correlated with introversion, along with every aggregate total feature and entropy. FTP, media streaming, and mail were the only application groups not showing significant correlations.

There are a number of studies that have investigated introversion, sometimes described as loneliness, with respect to increased Internet usage from the perspective of increased downloading music, playing games and email usage [16], [17]. While these studies explain many of the correlations, very few studies have been conducted to determine correlates between heavy Internet usage and specific forms of online communication. The derived features contain three different forms of communication: email, instant messaging, and VoIP. Email was the only communication application that did not support any correlations with introversion in this study.

3) Craving: The primary survey criteria for craving symptoms are related to staying online for longer than intended. Showing significant positive correlations with craving include HTTP ($\tau_{b,avg} = 0.19$), chatting ($\tau_{b,avg} = 0.20$), mail_octets ($\tau_b = 0.17$), FTP ($\tau_{b,avg} = 0.17$), game_flows ($\tau_b = 0.16$), duration_entropy ($\tau_b = 0.19$), packets_entropy ($\tau_b = 0.20$), and aggregate totals ($\tau_{b,avg} = 0.21$). The issue of excess chatting, email usage and online gaming correlating with unhealthy Internet usage has been documented in prior studies [18]. Correlations between duration_entropy and packets_entropy with symptoms are craving are quite revealing. They indicate that frequent multi-tasking or switching between applications that demonstrates randomized behavior (and hence increasing entropy) tend to create a feeling of staying online longer than intended.

4) Withdrawal: Withdrawal questions consisted of scaled responses for constantly pondering on what is happening on the Internet, or an increased anxiety to connect to the Internet after being away from it. There was insufficient evidence to support any correlations between withdrawal and Internet usage.

5) Negative Effects: Negative effects include scaled questions about sleep pattern disruption and possible tardiness. Variables correlating with negative effects include chatting ($\tau_{b,avg} = 0.24$), ftp_packets ($\tau_b = 0.18$), and ftp_octets ($\tau_b = 0.19$). As previously mentioned, FTP usage typically indicates content downloads that could lead to late-night usage or distraction from academics. Online chatting has also been associated with these patterns in prior studies.

6) Related Activities: This symptom assessed how much time students felt like they involved in activities related to the Internet, such as reading Internet magazines, reading e-books, etc. There was insufficient evidence to support any correlations between related activities and Internet usage.

7) Loss of Control: Loss of control aims to measure the participants attempts to recognize and unsuccessfully reduce their amount of Internet usage. The only correlating variable included was game_packets ($\tau_b = -0.19$). One possible explanation for this is that students scoring higher on this category may have logically recognized online gaming as a potential harmful area to allocate online time.

8) Reduced Activities: Reduced activity questions included a measure of how students felt about degradation of their productivity and also any perceived reduction in social or leisure time because of time spent online. No significant correlations were obtained between this symptom and Internet usage features derived.

9) Escape from Other Problems: Survey criteria for escape from other problems are situations involving the use of the Internet to avoid other pressing issues, or using the Internet to elevate mood. remote_flows ($\tau_b = 0.20$), and duration_total ($\tau_b = 0.17$) were the only variables to associate with escape from other problems. While correlations with duration_total is understandable, correlations with remote_flows needs further studies.

10) Tolerance: The two survey criteria corresponding to tolerance to Internet usage were a feeling of never having enough information from the Internet, and an increased online presence over the last twelve months. There were significant correlations between Tolerance and total_octets ($\tau_b = 0.18$) and total_packets ($\tau_b = 0.19$), both of which are indicators of high volume of Internet usage. Also, there were significant correlations between streaming_flows ($\tau_b = 0.20$) and using VoIP ($\tau_{b,avg} = 0.23$) applications with increased tolerance to Internet usage.

B. Mean Differences in Groups

Recall that using a IRPS total score threshold value of 110, participants were separated into groups of normal Internet activity (Group N, n = 60) and heavy Internet activity (Group H, n = 9). Statistical tests were constructed to determine which Internet features Group H used less than Group N. Under the null and alternate hypotheses

$$H_0 : \mu_N \leq \mu_H$$

$$H_A : \mu_N > \mu_H$$

Mann-Whitney U tests revealed that social_packets, social_octets, and social_flows values were statistically different between Group N and Group H. Group H tended to use Facebook and Twitter much less than those in Group N, as the results show in Figure 4. For each feature, Group N accumulate over twice the bytes or total. From these results, it seems as though heavier Internet users prefer other forms on online communication rather than social networking. One possible explanation is the use of Internet for entertainment versus social interactions. We point out that there have been prior studies suggesting that individuals engaging in Internet use for entertainment purposes may be more problematic than those seeking social interactions [16]. Even though instant messaging, gaming, and VoIP applications are instances of social mediums, they maintain a sense of pseudo-identity through the use of usernames. Social networking sites, like Facebook and Twitter, are promoted in a very different way;
personal profiles are made and there is a focus on maintaining real-life relationships via online mechanisms. It may be the case that these attributes could be key determinants in assessing both degree and symptoms of heavy Internet usage among young adults.

### TABLE III. COMPARISON AND RESULTS OF MANN-WHITNEY U TESTS (ONE-TAILED)

<table>
<thead>
<tr>
<th>Features</th>
<th>One-Tailed P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>social_packets (total)</td>
<td>0.049</td>
</tr>
<tr>
<td>social_octets (bytes)</td>
<td>0.046</td>
</tr>
<tr>
<td>social_flows (total)</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Fig. 4. Values of Statistically Different Mean Internet Feature Values

VI. CONCLUSIONS

In this paper, the results of a two month experiment conducted at the Missouri S&T campus on associating symptoms of heavy Internet usage with collected Internet data were analyzed. A number of fine grained Internet usage features that correlate with symptoms of heavy Internet usage were identified: like tolerance, craving, negative effects, introversion and escape from other problems. While most findings agree with previous studies, we also identify several interesting new findings that deserve more research. To the best of our knowledge this is the first empirical study on heavier (possibly addictive) Internet usage that uses real Internet usage data collected continuously, unobtrusively and preserving privacy. It also enables novel applications of “big” Internet data after appropriate processing in the realm of human behavior and mental health. While this study focused purely on Internet statistics, future studies could also explore correlations between symptoms of heavy Internet usage and web content. Understanding known (and possibly emerging) correlations between Internet usage symptoms of mental disorders like depression, anxiety, stress etc., and positioning them with results of this study are also topics of future investigation.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation (NSF) under Grant No. 1254117 and 1205695. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES


