



## Research Report

## A study of Facebook behavior: What does it tell about your Neuroticism and Extraversion?



Jianqiang Shen\*, Oliver Brdiczka, Juan Liu

Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA

## ARTICLE INFO

## Keywords:

Personality

Behavior analysis

Social network sites

Facebook

## ABSTRACT

Social Network Sites offer users the opportunity to portray themselves with much freedom. People get an impression about a user based on this user's online profile, posted content and interactions with friends. In this paper, we examine how Neuroticism and Extraversion traits are expressed through such behaviors of a user on Facebook. While previous research relied on self-reported Facebook usage from small samples of college students, we developed a Facebook application to directly retrieve data from 1327 users. This enables us to study fine-grained signals embedded in users' behaviors, such as **writing styles** or **number of "likes"**, and correlate them with users' personality. We present the features found to be significant and show that these features can be used to infer a user's Neuroticism and Extraversion.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Social Network Sites (SNS) facilitate interpersonal interaction and allow for the maintenance of ties that may have otherwise gone dormant (Ellison, Steinfield, & Lampe, 2007). The increased awareness of others' activities could have profound implications for the way we keep connected with others and understand others (Berkovsky, Freyne, & Smith, 2012). People get an impression of a user based on the following SNS behaviors: (1) a **"profile"** created by this user, including **basic demographics, personal interest and a list of friends** this user chose to associate with. (2) Posted content, including **videos, photos and "status updates"** viewable to some or all of this user's friends. Status updates (a.k.a. *posts*) are broadcast messages that are written for others' consumption and usually are not tailored to a particular person. They are a **major means to communicate with friends** on SNS (Kramer, 2010). (3) **Interactions with friends, such as "like" or "comment"** on a post.

People are motivated to be seen as attractive, likeable, competent, and virtuous (Leary, 1996). Though SNS users **strive to project a positive image of themselves** (Barash, Duchenaud, Isaacs, & Bellotti, 2010), their portrayed images could still be quite telling of their underlying characteristics, such as personalities. Personality traits are consistent patterns of thoughts, feelings, or actions that distinguish people from one another (John, Robins, & Pervin, 2010). In a study conducted by Gosling, Gaddis, and Vazire

(2007), 9 undergraduate research assistants rated personality traits of 133 subjects by examining their Facebook profiles. They found some consensus about profile-based personality assessment, with particularly strong consensus for Extraversion. Such personality assessment shows some accuracy, compared with traits reported by the studied subjects and their friends. This suggests that there exist **some specific cues eliciting personality related behaviors** and those cues are actually valid indicators of what someone is like.

The popularity of Social Network Sites (SNS) such as Facebook provides a great opportunity to examine personality inference using significant amounts of data. Indeed, recent SNS offer a wealth of behavioral indicators ranging from profiles to activity statistics that could reflect a user's personality. This behavioral richness could be leveraged to infer the personality of the individual behind an SNS account. However, the problem of modeling personality from Facebook behaviors has achieved very little advance in terms of concrete data analysis. This is largely due to the difficulty in data collection. Researchers typically made their analysis based on the self-reported data from a small sample of students from a single university. There is no guarantee that the reported data is objective and reflects the real behaviors. The homogeneous, small-sized population likely leads to biased conclusion. The contradiction between some research results might be due to the above facts. For example, Ross et al. (2009) showed that Extroversion was found to belong to more Facebook groups and not necessarily be associated with more Facebook friends, while Hamburger and Vinitzky's (2010) demonstrated that Extroversion had a positive effect on the number of friends, but no effect was found with regard to the use of Facebook groups.

\* Corresponding author. Tel.: +1 650 812 4428.

E-mail addresses: [jianqiang.shen@parc.com](mailto:jianqiang.shen@parc.com) (J. Shen), [oliver.brdiczka@parc.com](mailto:oliver.brdiczka@parc.com) (O. Brdiczka), [juan.liu@parc.com](mailto:juan.liu@parc.com) (J. Liu).

In this paper, we use a data-driven approach to personality modeling and prediction for Facebook users. We developed a Facebook application to directly retrieve data from 1327 users. From a user's profile, we extract a rich set of features and perform correlation analysis to discover which features are strongly correlated with the personality. The direct retrieval enables us to study objective, fine-grained signals embedded in users' behaviors, such as writing styles or number of "likes". Privacy concern often is a major hurdle to data collection and needs to be properly addressed. In our work, we have spent extra effort to guard personal information and designed a two-stage approach: (1) careful anonymization in the preprocessing stage to remove personally identifiable information (PII) such as name, address, email address and all numbers, and (2) the innovative design of an activity logger, which processes Facebook activities and retains only aggregated statistics. No raw content is logged in the feature set. Furthermore we assure that the aggregated statistics are sufficiently abstract such that the original content and/or meaning cannot be reconstructed from the feature set. These measures mitigate the privacy concern and are key to our data collection.

Some findings are expected and well in alignment with qualitative findings in social science studies. For instance, extroverts engage more actively in Facebook social activities. They share more photos, longer videos, and more status updates. Individuals high in Neuroticism are more likely to post accurate personal information. Our analysis also discovered some findings that we did not originally anticipate. For instance, we originally anticipated that neurotic users are more cautious and thus write less on Internet, but instead our analysis found that neurotic users tend to write longer posts, use more negative sentiment words and strongly subjective words in posts. Furthermore, we build a model to predict a Facebook user's Neuroticism and Extraversion. Our predictor achieves a modest accuracy with correlation  $R > 0.3$ . Giving the noisy nature of the online data and the difficulty in personality analysis, the prediction accuracy is encouraging. In this paper, we also report some preliminary results on other personality traits, with the hope of inspiring future research.

## 2. Related work and theoretical background

**Personality profiling.** Personality traits are consistent patterns of thoughts, feelings, or actions that distinguish people from one another (John et al., 2010). Different theories make different predictions about how mean levels of personality traits change in adulthood (Srivastava, John, Gosling, & Potter, 2003), but it is generally agreed that the personality profile affects our activity (Hogan, Johnson, & Briggs, 1997). First, having a specific personality trait means reacting consistently to the same situation overtime, for example, being agreeable or cooperative means consistently going along with reasonable requests. Second, to respond consistently in the same situation, people must have a capacity to respond to situational cues. Research has shown that it is possible to estimate a stranger's personality solely based on his/her behaviors (Kenny, Horner, Kashy, & Chu, 1992).

Personality profiles are usually constructed through surveys based on proven inventories of questions, e.g. International Personality Item Pool (Goldberg et al., 2006). Personality is readily expressed in specific cues in everyday life. Recent research has shown that when meeting a stranger face-to-face for the first time, it is possible to quickly assess their personality with some accuracy thanks to verbal and non-verbal cues (Kenny et al., 1992). For instance (Kenny et al., 1992), observers generally agree that Extraverted individuals speak louder, with more enthusiasm and energy, and that they are more expressive with gestures. Interestingly, personality cues can also be found in the physical

world: perceivers are able to accurately predict the personality of strangers by looking at their offices and bedrooms (Gosling, Ko, Mannarelli, & Morris, 2002), or even by examining their top ten favorite songs (Rentfrow & Gosling, 2006). There has been some success at predicting personality in a meeting scenario using visual and acoustic indicators (Lepri, Mana, Cappelletti, Pianesi, & Zancanaro, 2009).

**Internet usage & personalities.** Researchers have started exploring the connection between personality traits and general Internet usage. Modeling user personality based on Internet usage could enable better personalization of user interfaces and content (Ehrenberg, Juckes, White, & Walsh, 2008), more efficient collaboration (by forming groups of compatible individuals) in virtual games (Lepri et al., 2009), more precise targeted advertising, or improved learning efficiency by customizing teaching materials and styles (Muldner, Burleson, & VanLehn, 2010), to name just a few possibilities.

Hamburger and Ben-Artzi (2000) demonstrated that on Internet "the poor can get richer", namely, that introverts could compensate themselves for the difficulties they experience in offline social interactions. It was shown that more introverted, less agreeable, and less conscientious students engaged in higher levels of Internet usage (Landers & Lounsbury, 2006). In anonymous forms of online communication such as chat rooms, individuals high on the trait of Neuroticism were more likely to post accurate personal information on their profiles (Hamburger, Wainapel, & Fox, 2002). Guadagno, Okdie, and Eno (2008) conducted a similar study of blogs and found that people high in Openness and high in Neuroticism were more likely to be bloggers.

**Facebook & personalities.** The explosive growth in the number of Facebook users motivated some research investigating personality cues in Facebook profiles. A study conducted by Ross et al. (2009) showed that individuals high on Extraversion were found to belong to more Facebook groups, but not necessarily be associated with more Facebook friends. They also found that Neuroticism was unrelated to the posting of personally-identifying information and those low in Neuroticism preferred posting photos on their Facebook profiles. While Ross et al.'s study relied on self-reports by participants, in a follow-up study Hamburger and Vinitzky (2010) asked a research assistant to hand-code user information on Facebook. Hamburger and Vinitzky's (2010) results are contrary to those of Ross et al.'s (2009) in some aspects. They demonstrated that Extroversion had a positive effect on the number of friends, but no effect was found with regard to the use of Facebook groups, and individuals with high Neuroticism were found to be more inclined to post their photos on Facebook than individuals with low Neuroticism.

**Our study & contribution.** In general, existing research agrees that there are associations between Facebook activities and a user's personality. However, the findings are quite controversial, mainly due to the limitation of collected data. This (re-)study seeks to expand the literature by using a larger population and more objective measurements to investigate digital traces related to the personality of Facebook users. Compared with the existing research, our study has the following unique merits. First, unlike the existing research that used a small sample of students from a single university, we recruited 1327 subjects from all over the U.S. Second, instead of relying on self-reports or hand-coding of user profiles, we developed a Facebook application called *iPersonality* to directly retrieve information from Facebook accounts. Third, while the existing research limited features to demographics and high-level usage (spent time, count of friends, groups, albums and photos), direct retrieval enables us to calculate much finer-grained signals that are salient to the audience yet never explored by researchers before, such as written content and interactions with friends.

With the overall goal to resolve some controversy among existing research findings, particularly we attempt to investigate the following 3 research questions:

**RQ1.** *How do users' personality traits correlate with their Facebook activities?*

**RQ2.** *Does there exist any correlation between users' personality traits and their success in gaining social support on Facebook?*

**RQ3.** *To what extent can personality be inferred from the behavior of Facebook users?*

The main contribution of our paper includes the following: (1) we have designed a privacy preserving data logging and feature extraction approach to capture a rich online behavior dataset free of personally identifiable information. (2) Through correlation analysis, our study shows that people's online behaviors often exhibit associations with personality traits. Though such associations have been speculated in various social studies, our work is the first analysis using objective, fine-grained online traces that are not explored before. (3) Over a large dataset collected from a diverse population, we have built predictive models predicting Neuroticism and Extraversion based on online data with a moderate prediction accuracy.

### 3. Research study design

**Privacy-preserving data collection using Facebook app.** Our data collection system was deployed as a web service, so users can easily participate from anywhere. Upon signing up for our study and giving informed consent, each participant first answered an online demographic and psychological survey. Participants rated themselves on 112 psychological items adopted from the International Personality Item Pool (Goldberg et al., 2006), using a Likert scale that ranged from 1 (Very Inaccurate) to 5 (Very Accurate). Based on the survey results, we calculated the ground-truth personality values (ranged from 10 to 50) for each participant.

Participants then authorized our Facebook application, *iPersonality* (Shen, Brdiczka, & Ruan, 2013) to retrieve data from their Facebook accounts through OAuth. OAuth is an open standard for authorization, requiring neither passwords nor user names from users. It allows a user to grant a third party site access to their information stored with another service provider, without sharing their access permissions or the full extent of their data. Our program ensured that users from the same computer only participated once. Using the granted OAuth tokens, our program collected privacy-preserving data by computing a set of high-level, aggregate statistics from each participant's activities within the last 12 months, such as count of sentiment words in posts and count of likes received. Facebook activities were analyzed and extracted in a similar way as described in (Shen, Brdiczka, & Liu, 2013). No raw text content or visual content from the user and/or friends was collected.

In particular, each Facebook text content was heuristically tokenized into a word and punctuation list. We utilized a sentiment dictionary (Wilson, Wiebe, & Hoffmann, 2005) to count positive and negative sentiment words in the written content for each user. Compared with LIWC (Pennebaker, Francis, & Booth, 2001), this dictionary has a much higher coverage and contains human-annotated polarity information on 8221 distinct words. We make the assumption that a set of words (in this case, the positive and negative sentiment words) is defined as having some psychological meaning (in this case, positive or negative emotion), such that a user or group of users who use more words from a certain category are higher in the psychological construct that the category is designed to measure (Kramer, 2010; Pennebaker et al., 2001). Although this assumption may not be "perfect", it is a good enough approximation of underneath emotion and avoids imposing the

extra cost to label posts and train a sentiment classifier (Pang, Lee, & Vaithyanathan, 2002). Such word-count approaches thus have been used extensively in the fields of HCI and psychology (Kramer, 2010). For example, researchers showed that in short blog posts, users known to be angrier show higher incidence of negative words, while more joyful authors use more positive words (Gill, French, Gergle, & Oberlander, 2008). We also counted the number of question and exclamation punctuations, since they usually signal emotion.

**Participant population.** We recruited 1327 participants from all over the U.S. Although we were using emailing to colleagues and posting on Facebook for recruiting, most (> 90%) of our participants were not affiliated with our company or friends with any of the involved researchers. To focus on active Facebook users, we filtered out users that had less than 5 friends on Facebook. Additionally, we filtered out those users whose survey answers had negative inter-item consistency. This left us a diverse population of 917 participants. Table 1 lists various demographic statistics from the valid participants. The distributions of participants' personality scores are roughly Gaussian, with mean 26.1 and standard deviation 8.2 for Neuroticism, and mean 32.0 and standard deviation 7.8 for Extraversion.

Since we only have the data of people who chose to participate, we want to understand the potential bias introduced by our sample. We compare the demographics of our participants with the demographic information published by Facebook in March 2012 (BusinessAndIndustryPortal, 2013). We found that the demographics are somehow consistent between our participants and the overall Facebook population. For example, it was reported that 57% Facebook users were female and 57% were college- or university-educated. While we cannot claim that any of our findings must generalize to the entire population of Facebook users, we want to

**Table 1**  
Summary of demographic profiles.

Parameter		Value/percentage
Age		18–71 years (Mean 29 years, SD 8 years)
Gender	Female	57%
	Male	43%
Marital status	Single/divorced	53%
	Engaged/married/partnered	47%
Children	None	64%
	Some that live at home	33%
	None that live at home	3%
Education	High school or less	38%
	Currently in university	20%
	Completed university	31%
	Graduate or professional degree	11%
Job status	Full-time worker	40%
	Part-time worker/student	19%
	Full-time student	17%
	Unemployed	13%
	Home-maker	10%
	Retired	1%
Job category	Computer/IT	11%
	Education/training	8%
	Retail	7%
	Medical/health	6%
	Other	68%
Native English speaker?	Yes	96%
	No	4%

emphasize that our participants are from a diverse background and represent a large population of Facebook users.

**Study plan.** We are interested in inferring personality from the behaviors of a Facebook user. Thus we focus on cues visible to the audience and ignore behaviors that the general audience cannot observe directly, such as private messages and time spent on Facebook. A user's created profiles and shared posts are obviously two such visible cues. In addition, an individual's personality affects the motivation and capability to seek social interactions (John et al., 2010). On Facebook, such interactions are reflected in comments and "likes" from friends. We thus examine such previously unexplored cues.

Two personality traits particularly attract our attention – *Extraversion* and *Neuroticism*. A personality trait is an internal characteristic that corresponds to an extreme position on a behavioral dimension (John et al., 2010). They are basic tendencies that remain stable across the life span. *Extraversion* implies an energetic approach to the social world and includes traits such as sociability and positive emotionality. *Neuroticism* contrasts emotional stability and is sometimes called emotional stability. It is generally agreed that personality traits *Neuroticism* and *Extraversion* are significantly related to online activities (Hamburger & Ben-Artzi, 2000; Hamburger & Vinitzky, 2010; Hamburger et al., 2002; Kraut et al., 2002; Landers & Lounsbury, 2006). We thus focus on those two personality traits in the paper. To inspire future research, we also present some preliminary results on several other personality traits.

#### 4. Results

From each participant's Facebook account, we calculated 154 features, covering demographics, profile descriptions, produced content, and interactions with friends. It is important to mention that we are not claiming that we have extracted all possible variables for analysis in this paper, but rather, that we have extracted a meaningful and manageable subset of cues that cover a wide range of behaviors on Facebook. To answer RQ1, a description of derived variables, along with their means and standard deviations are presented in Table 2. For brevity, we excluded some similar features that were not significantly correlated with *Neuroticism* and *Extraversion*.

**Neuroticism** relates to *emotional stability* (John et al., 2010). High scorers tend to be *nervous, sensitive, vulnerable*; low scorers tend to be *calm, relaxed, secure, confident*. Gender is the most significant cues in our data: our female participants score much higher in *Neuroticism*. Older and married participants tend to be slightly more calm and relaxed, but the correlations are very weak.

Contrary to Ross et al.'s results (Ross et al., 2009) and consistent with Hamburger and Vinitzky's (2010), we find that *high scorers in Neuroticism* are inclined to share personally identifying information – they post *more personal interests and create more albums*. This style of posting information is likely to intentionally bolster psychological support (Mantovani, 2001) that could otherwise be missing for these psychologically vulnerable individuals. This finding also supports Hamburger et al.'s theory (Hamburger et al., 2002) that individuals high in *Neuroticism* are more likely to post accurate personal information. From our data, we also have the following interesting findings regarding *Neuroticism*:

*Finding 1: neurotic users tend to write longer posts, use more negative sentiment words and strongly subjective words in posts.*

*Finding 2: high scorers in Neuroticism are more successful in gaining social support: their posts get significantly more comments from friends.*

**Extraversion** relates to sociability and positive emotionality (John et al., 2010). High scorers tend to be sociable, friendly,

talkative; low scorers tend to be reserved, shy, quiet. Male, young and unmarried participants were likely to be more extroverted, though not significantly.

Facebook is mainly utilized to sustain existing offline relationships or support offline connections, as opposed to meeting new people (Ellison et al., 2007). Contrary to Ross et al.'s results (Ross et al., 2009) and consistent with Hamburger and Vinitzky's (2010), *extroverts have significantly more Facebook friends* than introverts in our dataset. Our data also leads to the following findings regarding *Extraversion*:

*Finding 3: extroverts engage more actively in Facebook social activities. They share more photos, longer videos, and more status updates.*

*Finding 4: extroverts get more comments and "likes" from friends, while the average number of comments or "likes" per one post or photo is not significantly different from introverts. When extroverts post a status update or photo, the percentage of friends commenting or liking it is much lower than introverts.*

**Discussion.** Different theories make different predictions about how mean levels of personality traits change in adulthood (Srivastava et al., 2003). The biological view of the Five-factor theory proposes the plaster hypothesis: all personality traits stop changing by age 30. In contrast, contextualist perspectives propose that changes should be more varied and should persist throughout adulthood. Our data suggests that there exists an age effect on personality traits: though not significantly, older participants tend to be slightly more calm and relaxed, and they are less likely to be extroverted.

Neurotic users tend to write longer posts, use more negative sentiment words and strongly subjective words in posts, which could be explained in those ways: (1) neurotic users have a strong interest in using the Internet for communication (Kraut et al., 2002), and are more willing to share their bad experience in detail in order to solicit social support, (2) as *Neuroticism* reflects a person's tendency to experience psychological distress, high levels of the trait are associated with a sensitivity to threat (John et al., 2010).

Neurotic users get more comments from friends and are more successful in gaining social support. Though status updates entail no obligation to reply, they may elicit supportive response (Burke, Kraut, & Marlow, 2011). Neurotic users are more successful probably because they tend to share more dramatic stories and thus attract more empathy. Although neurotic users tend to have fewer friends on Facebook, their online friends keep a tight relationship with them. Higher percentages of their online friends respond to their Facebook activities and give comments or "likes" to their posts.

Extroverts share more photos, longer videos, and more status updates. They make more comments on others' posts. It is interesting that extroverts tend to use more exclamation marks, perhaps to exaggerate things a little and attract some attention. As *Extraversion* reflects a person's tendency to experience positive emotionality, high scorers use few negative words in their posts.

Although extroverts engage more actively in Facebook social activities, their average number of comments or "likes" per one post or photo is not significantly different from introverts. Moreover, the percentage of friends commenting or liking their status update or photos is much lower than introverts. Thus, besides supporting the "rich get richer" theory that extroverts gain more from their Internet use than introverts (Kraut et al., 2002), our results also suggest that extroverts maintain diverse and heterogeneous relationships with friends, and many connections are weak. Research found that friend count was a robust predictor of bridging social capital, the form of social ties that loosely link people across a cleft that may divide society (Burke et al., 2011). Those extroverts with more friends and smaller percentage of interactive friends may explain the reason.



**Table 2**Means, SD, and correlations of features. Correlations in bold are  $p < .05$ .

Feature			Mean(SD)	Neurot.	Extrav.
Demographics	1.	Gender (0: male, 1: female)	1:57.2%	<b>0.20</b>	−0.04
	2.	Age	28.6(8.0)	−0.02	−0.03
	3.	Relationship (0: unmarried, 1: otherwise)	1:47.2%	−0.01	−0.03
Profiles	4.	# of friends	341.5(345.7)	− <b>0.15</b>	<b>0.23</b>
	5.	# of groups	8.33(12.54)	0.03	0.04
	6.	# of close groups	2.67(4.22)	0.01	0.03
	7.	# of listed educations	0.92(1.31)	0.00	0.00
	8.	# of listed employments	0.59(1.25)	−0.05	<b>0.07</b>
	9.	# of interests	2.81(5.62)	<b>0.11</b>	−0.03
Content	10.	Percent of videos with titles	0.38(0.44)	−0.02	<b>0.08</b>
	11.	Average video length (s)	18.67(39.06)	0.00	<b>0.07</b>
	12.	# of albums	9.33(8.27)	<b>0.08</b>	<b>0.07</b>
	13.	# of photos	293.08(469.10)	0.05	<b>0.10</b>
	14.	Percent of photos with titles	0.13(0.15)	0.04	− <b>0.06</b>
	15.	# of posts	264.20(410.22)	0.01	<b>0.08</b>
	16.	# of liking someone's status	9.52 (25.24)	0.01	0.01
	17.	# of comments	78.08(133.10)	0.05	<b>0.06</b>
	18.	# of question marks per 100 words in posts	0.99(3.70)	0.04	−0.01
	19.	# of exclamation marks per 100 words in posts	3.44(6.48)	−0.04	<b>0.07</b>
	20.	# of negative words per 100 words in posts	2.91(2.64)	<b>0.10</b>	−0.04
	21.	# of positive words per 100 words in posts	5.84(4.26)	0.04	0.01
	22.	# of strong subjectives per 100 words in posts	6.38(4.60)	<b>0.08</b>	−0.02
	23.	Avg # of words in a post	4.46(7.56)	<b>0.10</b>	−0.02
Interactions	24.	# of likes in photos	143.55(193.21)	−0.01	<b>0.14</b>
	25.	Avg # of likes in a photo	0.58(1.40)	−0.02	−0.02
	26.	Avg percent of friends giving like in a photo	0.60%(1.73%)	0.03	− <b>0.14</b>
	27.	# of comments in photos	116.39(153.40)	0.03	<b>0.12</b>
	28.	Avg # of comments in a photo	0.53(1.53)	0.01	−0.03
	29.	Avg percent of friends commenting in a photo	0.71%(3.55%)	0.03	− <b>0.07</b>
	30.	# of tags in photos	44.39(96.75)	0.01	<b>0.14</b>
	31.	Average percent of friends tagged in a photo	0.06%(0.19%)	0.03	− <b>0.08</b>
	32.	# of likes in posts	117.19(216.82)	0.02	<b>0.11</b>
	33.	Avg # of likes in a post	0.49(1.01)	0.01	0.05
	34.	Avg percent of friends giving likes in a post	0.24%(0.43%)	<b>0.12</b>	− <b>0.11</b>
	35.	# of comments in posts	112.01(241.58)	<b>0.07</b>	0.04
	36.	Avg # of comments in a post	0.45(0.89)	<b>0.09</b>	−0.03
	37.	Avg percent of friends commenting in a post	0.32%(1.33%)	<b>0.06</b>	− <b>0.08</b>
	38.	# of likes of his/her comments	18.44(35.94)	<b>0.08</b>	0.05
	39.	Avg # of likes in a comment	0.21(0.17)	0.01	0.00
	40.	Avg percent of friends giving likes in a comment	0.14%(0.29%)	<b>0.11</b>	− <b>0.21</b>
	41.	# of friends giving at least 1 comment to posts	41.69(43.05)	0.00	<b>0.13</b>
	42.	Percent of friends giving at least 1 comment	19.4%(23.2%)	<b>0.12</b>	− <b>0.11</b>

Previous research suggests that playing online games is like being “alone together” – surrounded by other players, but not necessarily actively interacting with them (Ducheneaut, Yee, Nickell, & Moore, 2006). Game players, instead of playing with other people, rely on them as an audience for their in-game performances, as an entertaining spectacle, and as a diffuse and easily accessible source of information and chitchat. We observe the same situation on Facebook: when a Facebook user posts a status update, averagely only 0.24% of the friends give “likes” and only 0.32% of the friends give comments. Moreover, averagely only 19.4% of friends comment a user's status updates at least once. This suggests alternative design strategies for social networking where encouraging and supporting direct interactions might be less important than designing for the “spectator experience” and a sense of social presence (Reeves, Benford, O'Malley, & Fraser, 2005).

Regarding RQ2, we found neurotic and introverted users were more successful in gaining social support. Although they tend to have fewer friends on Facebook, higher percentages of their online friends respond to their Facebook activities and give comments or “likes” to their posts. The appeal of being “alone together” in virtual worlds can usually be explained by users' desire for an audience, a sense of social presence, and a spectacle (Ducheneaut et al., 2006). As neurotic and introverted users care less about such things, their online friends are more likely to be their real friends

than “audience”. Their online friends are thus more active in engaging direct interactions with them.

**Personality inference.** To examine how well Neuroticism and Extraversion can be inferred from the behaviors of Facebook users, we normalized our features and conducted forward step-wise regressions, which start with no variables in the model, test the addition of each variable using correlation coefficients, add the variable (if any) that improves the model the most, and repeat this process until none improves the model. Note that our goal here is to understand whether meaningful statistic models can be constructed, not to exhaust different machine learning algorithms to achieve the highest accuracy. The final models with selected features and weights are presented in Tables 3 and 4. Both regressions were significant at  $p < .001$ . According to Cohen (1988), an  $R$  of 0.30 is a medium effect size, and thus our regression models had around medium effect sizes. Effect size estimates the strength of a phenomenon or a sample-based magnitude of the effect under the alternate hypothesis. Regarding RQ3, our result suggests that virtual cues constructed from behaviors on Facebook can be used to provide statistically significant models of a Facebook user's personality.

Those regression models are simple and fit our intuition. For example, in the Neuroticism model, gender and number of friends carry the largest weights, followed by number of negative words;

**Table 3**Regression on Neuroticism. For significance level: \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

	Coefficient	Sig. level
Intercept	21.73	***
1. Gender	4.93	***
4. # of friends	−1.82	***
5. # of groups	0.60	**
9. # of interests	0.33	*
20. Avg # of negatives in posts	0.71	*
22. Avg # of strong subs in posts	−0.43	
23. # of words in a post	0.12	
26. % of friends liking a photo	−0.15	
37. % of friends commenting a post	0.04	
Correlation R	0.32	

**Table 4**Regression on Extraversion. For significance level: \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ .

	Coefficient	Sig. level
Intercept	31.38	***
4. # of friends	1.52	***
6. # of close groups	−0.50	**
8. # of listed employments	0.21	
24. # of Likes in photos	3.37	*
26. % of friends liking a photo	−0.15	
30. # of tags in photos	0.34	**
31. % of friends tagged in a photo	−0.29	**
37. % of friends commenting a post	−0.33	**
Correlation R	0.36	

for Extraversion, number of “likes” in photos and number of friends have the largest weights.

Being able to infer a user's personality from online cues has direct impact on HCI research, given the field's long-standing interest in interface personalization and system customization (Yee, Ducheneaut, Nelson, & Likarish, 2011). Fine-grained modeling of user preferences and personality can help computer systems further adapt to different users. Modeling user personality based on online cues and behaviors will enable better personalization, collaboration and targeted advertising. Our results show that it is feasible to build a modestly reliable model to predict an SNS user's personality. The simple models presented in Tables 3 and 4 can be directly utilized to customize content delivery based on estimated personality profiles, so Facebook users can have more appealing user experience.

**A peek at other personality traits.** Using the collected data, we also did some preliminary analysis on other personality traits – Agreeableness, Openness and Conscientiousness (John et al., 2010). Note we are claiming that the results are comprehensive or conclusive. Rather, such preliminary results serve as an exploring of possible social networking cues on those personality traits, with the hope of promoting further concrete research on this topic.

*Agreeableness* is a tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others. High scorers tend to be friendly, caring, cooperative; low scorers tend to be critical, rude, suspicious. We found that Agreeableness is well correlated with the number of friends (Correlation = 0.15, with  $p < .05$ ). Given the factor that persons with high Agreeableness are more likely to be friendly, it is not surprising that they tend to get more friends on Facebook.

*Openness* describes the breadth, depth, originality, and complexity of an individual's mental life. High scorers tend to be original, curious, complex; low scorers tend to be conventional, narrow interests, uncreative. The two most promising indicators we found are age (Correlation = −0.21, with  $p < .05$ ) and number of sports (Correlation = 0.13, with  $p < .05$ ). Our data shows that older persons are less likely to be interested in new and exciting things,

which again suggests there exists an age effect on personality traits. Because of their curiosity nature, those users with high Openness tend to like more sport activities on Facebook.

*Conscientiousness* describes socially prescribed impulse control that facilitates task and goal-directed behavior. High scorers tend to be reliable, well-organized, self-disciplined; low scorers tend to be disorganized, undependable, negligent. We found that high scorers tend to be older (Correlation of age = 0.10, with  $p < .05$ ) and post fewer interests on Facebook (Correlation of # of interest = −0.07, with  $p < .05$ ).

Our results suggest that in general there are some strong indicators of personality traits on Facebook. In the future, we plan to extract and build more activity cues and correlate them with different personality traits.

## 5. Conclusions

Our effort is inscribed in a broader scope of research seeking to establish correspondence between people's real-world context and online social behaviors. Through careful design of direct retrieval of users' Facebook activities, we were able to collect a large dataset from participants with diverse background. The purpose of this (re-)study was to use this large user sample to examine the nature of Facebook use, explore the connections between personality traits and behaviors of Facebook users, and resolve some controversy among existing research findings. By rigid analysis of this large, unbiased dataset, we establish a data-driven correlation model between people's online behaviors and their personality traits.

We present personality cues constructed from demographics, profiles, content and user interactions, along with their predictive power. Direct retrieval of data from Facebook accounts enables us to study fine-grained signals that have never been explored before, including produced content and interactions with friends. We confirmed that neurotic users are inclined to share personally identifying information and extroverts have significantly more Facebook friends than introverts (Hamburger & Vinitzky, 2010; Ross et al., 2009). We have some interesting discoveries, for instance, neurotic and introverted persons are more successful in gaining social supports and had interactions with many of their Facebook friends. The age of users has some effect on their personality traits. Step-wise regressions show that we can construct statistically significant models to infer a Facebook user's Neuroticism and Extraversion.

Personality study has long been of interest to the user modeling community (Kraut et al., 2002). Information needs vary based on a user's personality (Lepri et al., 2009; Yee et al., 2011), and accurate personality modeling could enable better personalization of user interfaces and more precisely targeted advertising, to name but a few possibilities. We have identified informative features and built models to profile Facebook users with a moderate accuracy. Our analysis has verified the existence of correlations between people's Facebook behaviors and their personality traits. We hope that the method presented in this paper can be extended to other online study problems as well. We plan to look into other social dynamics, such as influence propagation, dynamics between Friends, and roles people take in their online social groups. Through such effort we hope to achieve a better understanding of the online social space, so as to make online platforms more helpful to real-world users.

## Acknowledgments

The authors gratefully acknowledge support for this work from DARPA through the ADAMS (Anomaly Detection At Multiple Scales) program funded project GLAD-PC (Graph Learning for

Anomaly Detection using Psychological Context). Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the government funding agencies. We thank the anonymous reviewers for their constructive comments.

## References

- Barash, V., Ducheneaut, N., Isaacs, E., & Bellotti, V. (2010). Faceplant: Impression (mis) management in Facebook status updates. In *Proc. of ICWSM-10*.
- Berkovsky, S., Freyne, J., & Smith, G. (2012). Personalized network updates: Increasing social interactions and contributions in social networks. In *Proc. of UMAP-12* (pp. 1–13).
- Burke, M., Kraut, R., & Marlow, C. (2011). Social capital on Facebook: Differentiating uses and users. In *CHI-11* (pp. 571–580).
- BusinessAndIndustryPortal. (2013). Who uses Facebook? <<http://www.business.qld.gov.au/business/running/marketing/online-marketing/using-facebook-to-market-your-business/who-uses-facebook>>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum.
- Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). Alone together?: Exploring the social dynamics of massively multiplayer online games. In *Proc. of CHI-06* (pp. 407–416).
- Ehrenberg, A. L., Juckes, S. C., White, K. M., & Walsh, S. P. (2008). Personality and self-esteem as predictors of young people's technology use. *Cyberpsychology & Behavior*, 11(6), 739–741.
- Ellison, N., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), 1143–1168.
- Gill, A. J., French, R. M., Gergle, D., & Oberlander, J. (2008). The language of emotion in short blog texts. In *Proc. of CSCW-08* (pp. 299–302).
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96.
- Gosling, S., Gaddis, S., & Vazire, S. (2007). Personality impressions based on Facebook profiles. In *Proc. of ICWSM-07* (pp. 26–28).
- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82(3), 379–398.
- Guadagno, R., Okdie, B., & Eno, C. (2008). Who blogs? Personality predictors of blogging. *Computers in Human Behavior*, 24(5), 1993–2004.
- Hamburger, Y. A., & Ben-Artzi, E. (2000). The relationship between extraversion and neuroticism and the different uses of the internet. *Computers in Human Behavior*, 6(4).
- Hamburger, Y. A., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, 26(6), 1289–1295.
- Hamburger, Y. A., Wainapel, G., & Fox, S. (2002). "On the internet no one knows i'm an introvert": Extroversion, neuroticism, and internet interaction. *CyberPsychology & Behavior*, 5(2), 125–128.
- Hogan, R., Johnson, J., & Briggs, S. (1997). *Handbook of personality psychology* (3rd ed.). Academic Press.
- John, O. P., Robins, R. W., & Pervin, L. A. (2010). *Handbook of personality: Theory and research* (3rd ed.). The Guilford Press.
- Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L.-C. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, 88–97.
- Kramer, A. (2010). An unobtrusive behavioral model of gross national happiness. In *Proc. of CHI-10* (pp. 287–290).
- Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues*, 58(1), 49–74.
- Landers, R., & Lounsbury, J. (2006). An investigation of big five and narrow personality traits in relation to internet usage. *Computers in Human Behavior*, 22(2), 283–293.
- Leary, M. (1996). *Self-presentation: Impression management and interpersonal behavior*. Westview Press.
- Lepri, B., Mana, N., Cappelletti, A., Pianesi, F., & Zancanaro, M. (2009). Modeling the personality of participants during group interactions. In *Proc. of UMAP-09* (pp. 114–125).
- Mantovani, G. (2001). The psychological construction of the internet: From information foraging to social gathering to cultural mediation. *CyberPsychology & Behavior*, 4(1), 47–56.
- Muldner, K., Bursleson, W., & VanLehn, K. (2010). Yes!: Using tutor and sensor data to predict moments of delight during instructional activities. In *Proc of UMAP-10* (pp. 159–170).
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proc. of EMNLP '02* (pp. 79–86).
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and word count (LIWC2001)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Reeves, S., Benford, S., O'Malley, C., & Fraser, M. (2005). Designing the spectator experience. In *Proc. of CHI-05* (pp. 741–750).
- Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. *Psychological Science*, 17(3), 236–242.
- Ross, C., Orr, E., Sisic, M., Arseneault, J., Simmering, M., & Orr, R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, 25(2), 578–586.
- Shen, J., Brdiczka, O., & Ruan, Y. (2013). A comparison study of user behavior on Facebook and Gmail. *Computers in Human Behavior*, 29(6), 2650–2655.
- Shen, J., Brdiczka, O., & Liu, J. (2013). Understanding email writers: Personality prediction from email messages. In *Proc. of UMAP-2013*.
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, 84(5), 1041–1053.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT-EMNLP* (pp. 347–354).
- Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011). Introverted elves & conscientious gnomes: The expression of personality in world of warcraft. In *Proc. of CHI-2011* (pp. 753–762).