

Manifestations of user personality in website choice and behaviour on online social networks

Michal Kosinski · Yoram Bachrach · Pushmeet Kohli ·
David Stillwell · Thore Graepel

Received: 9 August 2012 / Accepted: 27 August 2013 / Published online: 19 October 2013
© The Author(s) 2013

Abstract Individual differences in personality affect users' online activities as much as they do in the offline world. This work, based on a sample of over a third of a million users, examines how users' behaviour in the online environment, captured by their website choices and Facebook profile features, relates to their personality, as measured by the standard Five Factor Model personality questionnaire. Results show that there are psychologically meaningful links between users' personalities, their website preferences and Facebook profile features. We show how website audiences differ in terms of their personality, present the relationships between personality and Facebook profile features, and show how an individual's personality can be predicted from Facebook profile features. We conclude that predicting a user's personality profile can be applied to personalize content, optimize search results, and improve online advertising.

Keywords Personality · Individual differences · Social networks · Facebook · Personalization · Web search · Online advertising · Machine learning

Editors: Winter Mason, Jennifer Wortman Vaughan, and Hanna Wallach.

M. Kosinski · D. Stillwell
Psychometrics Centre, University of Cambridge, Cambridge, USA

M. Kosinski
e-mail: mk583@cam.ac.uk

D. Stillwell
e-mail: ds617@cam.ac.uk

Y. Bachrach (✉) · P. Kohli · T. Graepel
Microsoft Research, Redmond, USA
e-mail: yobach@microsoft.com

P. Kohli
e-mail: pkohli@microsoft.com

T. Graepel
e-mail: thoreg@microsoft.com

1 Introduction

Decades of psychology research suggest that individuals' behaviour and preferences can be accurately explained by psychological constructs called personality traits (Allport 1962). This is valuable in practice, as it implies that knowledge of an individual's personality enables prediction of both behaviour and preferences across different contexts and environments.

Personality assessment studies have revealed that responses to a relatively short personality questionnaire can predict human behaviour in many different aspects of life—from arriving on time and job performance (Barrick and Mount 1991), to drug use (Roberts et al. 2005) and infidelity (Orzeck and Lung 2005). It is also possible to assess personality by inspecting a person's *behavioural residues*—traces of the individual's actions in the environment. For example, researchers have shown that individuals can identify other people's personality traits by examining their living spaces (Gosling et al. 2002) or music collections (Rentfrow and Gosling 2006). Following the shift of human interactions, socializing, and communication activities towards online platforms, researchers have noted that such behavioural residues are not restricted to the offline environment and showed that personality can be inferred from records of keyboard and mouse use (Khan et al. 2008), contents of personal websites (Marcus et al. 2006; Vazire and Gosling 2004), or Facebook Likes (Kosinski et al. 2013).

This work examines how personality is manifested in users' online behaviour as reflected by the websites they browse and their Facebook activity. As Internet browsing is to a large extent a private activity, relationships between website choices and personality might be unaffected by peer pressure and the tendency to present oneself in a positive manner. Similarly, while the contents of Facebook Status Updates, uploaded Pictures, or the choice of Facebook Likes might carry an element of self-enhancement, the frequencies and distribution of Liking behaviour, number of uploaded Photos, or density of the Friendship network are less likely to be affected by users' conscious attempts to control their image. Thus, website choices and Facebook profile features may offer important and potentially unbiased insights into users' personalities.

The dataset used in this study is relatively large and diverse, consisting of over 350,000 US Facebook users. To the best of our knowledge, this is the largest dataset ever recorded relating psychological traits to web behaviour. Users' personality was measured using a standard International Personality Item Pool questionnaire (Goldberg 1999; Goldberg et al. 2006) representing a widespread Five Factor Model of personality. Users' website preferences were recorded using their website-related Facebook Likes and a questionnaire specifically designed for this study. The Facebook profile features analysed here include: the size and density of the users' Facebook friendship networks, the number of Facebook Groups and Likes that a user has connected with, the number of photos and status updates uploaded by the user, the number of times the user was tagged on photographs uploaded to Facebook, and the number of events attended by the user.

1.1 Five factor model of personality

Individual personality differences have been studied in psychology for a long time. Previous research has shown that personality is correlated with many aspects of life, including job success (Barrick and Mount 1991; Judge et al. 1999; Tett et al. 1991), attractiveness (Byrne et al. 1967), marital satisfaction (Kelly and Conley 1987) and happiness (Ozer and Benet-Martinez 2006).

In our analysis we use the Five Factor Model—the most widespread and generally accepted model of personality (Costa and McCrae 1992; Goldberg 1993; Russell and Karol 1994; Tupes and Christal 1992). The Five Factor Model was shown to subsume most known personality traits and it is claimed to represent the “basic structure” underlying the variety in human behaviour and preferences, providing a nomenclature and a conceptual framework that unifies much of the research findings in the psychology of individual differences.

We now briefly describe the five personality traits (Costa and McCrae 1992; Goldberg 1993; Russell and Karol 1994):

Openness to experience measures a person’s imagination, curiosity, seeking of new experiences and interest in culture, ideas, and aesthetics. It is related to emotional sensitivity, tolerance and political liberalism. People high on Openness tend to have a great appreciation for art, adventure, and new or unusual ideas. Those with low Openness tend to be more conventional, less creative, more authoritarian. They tend to avoid change for its own sake and are usually more conservative and traditional.

Conscientiousness measures the preference for an organized approach to life as opposed to a spontaneous one. People high on Conscientiousness are more likely to be well organized, reliable, and consistent. They plan ahead, seek achievements, and pursue long-term goals. Low Conscientiousness individuals are generally more easy-going, spontaneous, and creative. They tend to be more tolerant and less bound by rules and plans.

Extroversion measures a person’s tendency to seek stimulation in the external world, the company of others, and to express positive emotions. Extroverts tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative, do not mind being at the centre of attention, and make new friends more easily. Introverts are more comfortable in their own company, can be reserved, and tend to seek environments characterized by lower levels of external stimulation.

Agreeableness measures the extent to which a person is focused on maintaining positive social relations. High Agreeableness scorers tend to be friendly and compassionate, but may find it difficult to tell a hard truth. They are more likely to behave in a cooperative way, trust people, and adapt to the needs of others, but consequently they may find it difficult to argue their own opinion.

Neuroticism, often referred to as emotional instability, is the tendency to experience mood swings and negative emotions such as guilt, anger, anxiety, and depression. Highly Neurotic people are more likely to experience stress and nervousness, while those with lower Neuroticism tend to be calmer and more self-confident, but at the extreme they may be emotionally reserved.

1.2 Importance of the results

Our results have three particularly important implications:

Personalization It is valuable for websites, service providers, and brands to know the psycho-demographic profiles of their users. Currently, websites personalize their content, optimize their marketing, and tailor their search results using audience profiles encompassing demographic traits, such as age, gender and income (Hu et al. 2007). If websites and other web services attract audiences with a distinct personality profile, online platforms could greatly expand their understanding of users and thus improve their services and the user experience.

Inferring psychological profiles By finding associations between personality, website preferences, and social network profiles, we provide an alternative avenue for psychological research. In the past, the majority of psychological measurement has relied on self-report questionnaires completed by relatively small numbers of participants. Our approach suggests that personality could be measured automatically based on records of online behaviour; thus enlarging the scope of psychological assessment to an unprecedented scale. It may even improve the quality of results as it considers actual behaviour in the increasingly natural digital environment rather than self-reported test answers. Moreover, it is likely that studying vast samples of digitally recorded behaviour will improve researchers' existing psychological models or suggest new ones.

Privacy While it is widely accepted that an individual's personality can be accurately assessed using traditional psychometric tools, such as a personality questionnaires, the ability to automatically infer psychological profiles using digital records of behaviour challenges users' privacy expectations. Such inferences deprive individuals of control over what other parties can learn about them and may breach the trust between users and online service providers.

This paper is a revised and extended version which expands on two preliminary conference papers (Bachrach et al. 2012; Kosinski et al. 2012) presented at the 2012 ACM Web Sciences Conference in Evanston, Illinois. It employs significantly larger datasets and expands the results. We have also broadened the section dealing with earlier work, discussing the similarities and differences of our approach and previous approaches in more detail.

2 Dataset

Our dataset of over 350,000 US Facebook users was acquired from the *myPersonality project*.¹ database, collected using a Facebook application deployed in 2007. The myPersonality application allowed Facebook users to take personality and other psychological tests and obtain feedback on their results. The project database contains over 6 million detailed profiles of Facebook users accompanied by their scores on a wide variety of psychometric measures (Kosinski et al. 2013)

The myPersonality sample is representative of the general Facebook population, with an average age of 24.15 (SD = 6.55), an over-representation of users from the USA (roughly 55 %) and an over-representation of females (58 % of females) which may be attributed to the fact that they spend more time on Facebook and that they are more interested in getting feedback on their personality.² Note that in this study we have chosen participants only from the US in order to avoid biases introduced by cultural differences.

After completing the questionnaire, users could give their opt-in consent to record their Facebook profile information and personality scores for research purposes. This included various Facebook profile features described below, and access to the users' social network bookmarks in the form of their Liked websites. Users were also presented with the opportunity to fill in a Website Preference Questionnaire (WPQ) designed specifically for this study.

¹ Available at: <http://mypersonality.org/>.

² Demographics of Facebook users can be checked on Facebook at <http://newsroom.fb.com/> and at Check-Facebook at www.checkfacebook.com. Facebook is reported to have 20.5 % of its users in the ages 13–17, 26.4 % in the ages 18–25, 26.6 % in the ages 26–34 and 14.8 % in the ages 35–44, similar to the age distribution in our sample.

The WPQ asked users to specify the frequency of their visits to certain websites, providing self-reports regarding users' Internet browsing activity.

We thus had several data sources regarding individual users: their personality trait scores, their Facebook profile features and self-reports regarding their browsing activity.

2.1 IPIP Five Factor Model personality questionnaire

Personality scores used in this study were obtained using the 100 item long International Personality Item Pool questionnaire (Goldberg 1999; Goldberg et al. 2006) measuring Costa and McCrae's Five Factor Model of personality (IPIP FFM questionnaire) (Costa and McCrae 2006).

The quality of the personality scores obtained from myPersonality sample was controlled by examining scale reliability and discriminant validity, as suggested by John and Benet-Martinez (2000). Discriminant validity in our sample (average $r = 0.16$) was better than average discriminant validity reported in a premier empirical journal in personality and social psychology (Journal of Personality and Social Psychology, average $r = 0.20$ in the year 2002; see Gosling et al. 2004 for details). Additionally, the IPIP scale reliabilities in the myPersonality data were on average higher than those reported on the IPIP test publisher's website. This indicates that the quality of the responses in our sample was at least as high as in traditional pencil-and-paper studies.

3 Study 1: personality and website preferences

The goal of this study was to examine how a user's personality is reflected by their Internet web browsing habits and preferences. We start with a review of relevant literature followed by the results.

3.1 Related work

A number of studies have analysed the relationships between online preferences, browsing behaviour and demographic characteristics of website audiences, including age, gender, occupation and education levels, income, and race. Most of these studies (e.g. Baglioni et al. 2003; De Bock and Van Den Poel 2010; Hu et al. 2007; Murray and Durrell 1999; Weber and Jaimes 2011) are based on explicit profile data which is typically collected during the sign up process for an online service. Another approach relies on implicit profile data, or user characteristics that are inferred rather than known. In a typical approach Internet Protocol addresses are used to infer the users' location which, combined with census information, allows inferring characteristics such as education, income, race, etc. (e.g. Weber and Castillo 2010; Weber and Jaimes 2011).

The above studies focused on demographic properties of individuals. To the best of our knowledge, no attempts have been made to relate personality of Internet users to their web browsing behaviour. However, the psychological literature provides some examples of the relationship between personality and other aspects of the users' behaviour in an online setting. For example, Marcus et al. (2006) and Vazire and Gosling (2004) assessed personality using the contents of personal websites, Gill et al. (2006) studied the accuracy of personality judgements based on emails, Back et al. (2008) showed that there is some valid personality related information even in users' email addresses, while Kosinski et al. predicted personality and other psychological traits using Facebook Likes (Kosinski et al. 2013).

3.2 Self-reported website preference data

Self-reported website preferences were collected using the standard approach applied in personality research. In the questionnaire designed for this study, the WPQ, users were asked for the frequency with which they visit 23 websites on a five point scale (from never to regularly). Websites included in the WPQ were selected to be potentially informative about a visitor's personality. For instance, it was assumed (and later confirmed by the results) that songlyrics.com, an online library of song lyrics, would be attractive to outgoing and sociable people or, in other words, people characterized by high levels of extroversion. Moreover, the websites were selected to be neither too popular nor too obscure. Extremely popular websites attract visitors of all personality types and thus are not informative. On the other hand, obscure websites do not attract a reasonable fraction of users and thus are not discriminative.

The WPQ was offered in May 2010 to myPersonality respondents who had previously taken the IPIP FFM questionnaire. We collected completed WPQ questionnaires from 10,897 individual users. On average, respondents reported that they had visited three of the websites in the questionnaire at least rarely ($SD = 1.9$). The maximum number of websites endorsed by a respondent was 13, while around 4 % of the participants did not visit any of the websites included in the questionnaire.

3.3 Liked websites dataset

Users' website preferences were obtained using the Facebook Like feature which allows Facebook users to annotate a website as *Liked*, in order to recommend it to their friends and receive updates or news regarding the website publishers' activities. Users can *Like* a website by clicking the Like button directly on the website (an increasing number of websites offer such functionality) or by joining a website's fan page directly on Facebook.

In contrast to the responses to a questionnaire, the individual records of Liked websites are not influenced by the data collection context, nor limited to a small number of alternatives. However, Liked websites are visible to a user's social circle and thus might be used strategically to convey a desired impression. Also, it should be noted that there is no measure of the degree to which users spend time on each website. Some of the users may like a website simply to promote it to friends, without actually spending much time on it. Conversely, certain websites, such as technical documentation, might be less likely to be promoted with Likes regardless of the time a user spends on them.

We used data recorded between February and March 2011, containing about 153,000 individual US Facebook users resulting in nearly 75,000 unique website-related Likes that were endorsed by at least 20 distinct users. Users that filled in the WPQ questionnaire were removed from this sample to ensure the full independence of the results.

3.4 Aggregated website audience profiles

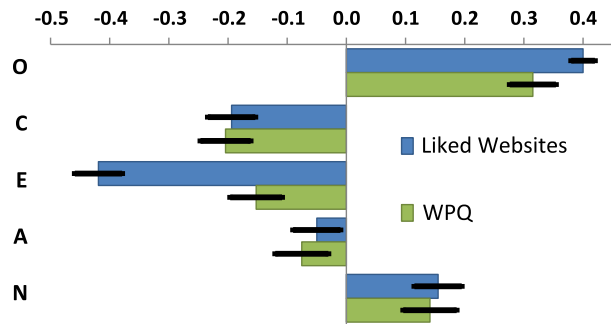
Below we address the question of website audience profiling by presenting the average personality traits observed in audiences of different websites.

To obtain the personality profile of the website preference group, we computed the mean personality scores as well as age and gender of all users who reported to visit (WPQ dataset) or liked (Liked URL dataset) each website. Descriptive statistics of the individual users and audience profiles based on liked URLs are presented in Table 1. The relationship between the number of liked websites and individual personality traits leads to differences between the individual and aggregated values of the average personality trait strengths. For instance,

Table 1 Descriptive statistics; personality, gender and age of individual users and aggregated by website for the Likes dataset. Note that when aggregating by user, the personality traits were first standardized to ensure a zero mean and unit standard deviation

	Aggregated profiles				Individual users			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Gender	.00	1.00	.71	.16	61 % females			
Age	15.71	51.10	21.24	4.14	13.00	65.00	24.36	8.76
Openness	−1.00	1.22	.12	.26	−4.05	2.19	.00	1.00
Conscientiousness	−1.21	.92	−.26	.23	−3.26	1.83	.00	1.00
Extroversion	−1.29	1.02	−.03	.21	−3.13	2.11	.00	1.00
Agreeableness	−1.25	.91	−.10	.20	−3.61	2.24	.00	1.00
Neuroticism	−1.18	1.14	.03	.22	−2.85	2.46	.00	1.00
Number of Users	20	22643	214.69	671.17	–			
Number of Likes	–				1	2,877	104.64	242.54
$N_{websites} = 74,993$					$N_{users} = 153,838$			

Fig. 1 Mean personality predictions for deviantart.com from the two different data sources. The error bars show 95 % confidence intervals



women constitute 61 % of the sample, but as they tend to like more websites than men, on average websites have 71 % of their Likes coming from women. To preserve the clarity of the results' presentation and allow for meaningful comparisons between aggregated profiles, aggregated values were re-scaled within each of the samples to zero mean. For instance, the aggregated values of Openness in the Liked dataset were decreased by its mean value (0.12) as presented in Table 1.

An example of a website audience personality profile, of deviantART.com, is presented in Fig. 1. According to both sources of data, this website attracts an audience that tends to be liberal and artistic rather than conservative and traditional (i.e. with high Openness), spontaneous and flexible rather than well organized (i.e. with low Conscientiousness), shy and reserved rather than outgoing and active (i.e. with low Extroversion), and emotional rather than calm and relaxed (i.e. with high Neuroticism). Both personality theory and common intuition suggest that those results accurately represent the character of deviantART.com users in general—alternative art enthusiasts and artists.

Table 2 provides further evidence of the psychological validity of our results by presenting the six websites with highest and lowest mean scores for each of the personality traits. For example we see that the most liberal, creative, and open to new experience audiences (with high Openness) are especially attracted to (1) modcloth.com, a mod-retro-indie clothing website, (2) boingboing.net, a blog on media, technology and popular cul-

Table 2 Websites with highest and lowest mean personality for each of the five personality traits, estimated on the Likes dataset

Openness	
Liberal, Artistic	Conservative
modcloth.com	gateway.com
senate.gov	newegg.com
boingboing.net	fitnessmagazine.com
astrology-online.com	ourtoolbar.com
gutenberg.org	nhl.com
cafeastrology.com	pier1.com
Conscientiousness	
Organized	Spontaneous
lww.com	candystand.com
ecollege.com	crunchyroll.com
ecnext.com	allthetests.com
exct.net	bestuff.com
education.com	lyricsdepot.com
kodak.com	letmewatchthis.com
Extroversion	
Outgoing, Active	Shy, Reserved
clubzone.com	lyricsty.com
ideeli.com	fanfiction.net
thanksmucho.com	behindthename.com
discoveryeducation.com	newworldencyclopedia.org
list-manage.com	personalitypage.com
trails.com	gaia-online.com
Agreeableness	
Cooperative	Competitive
abebooks.com	localtribune.org
socialsecurity.gov	funnyjunk.com
myrecipes.com	marvel.com
bluemountain.com	allthetests.com
serialssolutions.com	personalitypage.com
ecollege.com	supercheats.com
Neuroticism	
Emotional	Calm, Relaxed
cineplex.com	ncsu.edu
comparedby.us	sheetmusicplus.com
myprofilepimp.com	pitt.edu
barbie.com	highschoolsports.net
yellowpages.ca	myrecipes.com
biglots.com	lww.com

ture, (3) astrology-online.com and cafeastrology.com, astrology websites, (4) gutenber.org, a free e-book repository, (5) failblog.com, containing humorous media content, (6) fin-eartamerica.com, a fine art website, (7) 911tabs.com, a website specializing in guitar tabs, and (8) senate.gov, the website of the United States senate (which at the time of data collection had a majority of Democrat senators).

On the other end of the Openness scale, we see that websites for which the user population is estimated to be most conservative and “conventional” include (1) dealspl.us and newegg.com, shopping deal websites, (2) a variety of health, fitness, recipe and style websites such as fda.gov, mydailymoment.com and fitnessmagazine.com, (3) doctorslounge.com, a website specializing in health and medical jobs, (4) gateway.com, which sells information technology products, (5) nhl.com, the website of the National Ice Hockey League in the United States, and (6) pier1.com which sells furniture and accessories.

3.5 Website categories

To better understand the relationship between personality and website preference, we also aggregated the personality profiles across website categories.

Using classifiers as described by Bennett et al. (2010), we classified each website in the Facebook Like dataset into one of the top two-levels of the Open Directory Project (ODP) document hierarchy (Netscape Communication Corporation 2013). It consists of 219 topical categories such as Arts/Movies, Business/Investing and Sports/Soccer once categories with fewer than 1,000 associated web pages are removed. A logistic regression classifier with L2 regularization was trained using documents tagged with each category in a 2008 crawl of the ODP index. Using these classifiers, we tagged each liked URL in the dataset with the most likely ODP category. We then computed the mean of each of the five personality traits for each ODP category.

Table 3 presents the categories with highest and lowest mean personality score for each of the five personality traits. Again, users of different personalities prefer different website categories and those differences are consistent with personality. For instance, Extroverted users frequent websites related to Music and Internet (the category that contains Facebook and Twitter), while Introverts prefer websites related to Comics, Literature, and Movies. Interestingly, websites related to Mental Health are appeared to be frequented by people with extremely low levels of Agreeableness and Conscientiousness.

3.6 Audience similarity

One of the practical applications of personality profiles of website preference groups might be in personalizing search results and suggesting websites of interest to users. Our next analysis approaches this application by identifying which sets of websites are of interest to similar users, even if the user populations do not overlap. Table 4 shows several websites that appear dissimilar on the surface and do not have much overlap in the audience (in our dataset, the overlap in the audience between any two of the websites in Table 4 is lower than 2 %), but have similar mean psychological profiles. This avenue for personalization would allow identifying other websites to promote to users based on the similarities in personality profile. We see that Tumblr.com (a micro blogging platform), etsy.com (a marketplace of hand-made craft), gaia online.com (advertised as a forum of young open minded people), fanboy.com (marketed as a website for intellectuals with imagination), and rainymood.com (providing sounds of rain to visitors) are frequented by audiences with similar mean personality: liberal, introverted, and rather emotional. Notably, the only website in this group that

Table 3 Categories of websites characterized by the highest and lowest levels of aggregated personality traits. Shown are the top four and bottom four website categories for each personality trait

Openness	
Liberal, Artistic	Conservative
Arts.Animation	Reference.Education
Business.Marketing	Arts.Television
Business.Services	Sports.Soccer
Arts.Photography	Shopping.Children
Conscientiousness	
Organized	Spontaneous
Reference.Education	Health.Mental Health
Shopping.Electronics	Arts.Music
Shopping.Children	Arts.Animation
Reference.Dictionaries	Arts.Literature
Extroversion	
Outgoing, Active	Shy, Reserved
Computers.Internet	Arts.Movies
Reference.Education	Shopping.Children
Science.Environment	Arts.Literature
Arts.Music	Arts.Comic
Agreeableness	
Cooperative	Competitive
Reference.Education	Kids&Teens.Society
Computers.Internet	Health.Mental Health
Business.Logistics	Science.Physics
Health.Diseases	Recreation.Pets
Neuroticism	
Emotional	Calm, Relaxed
Recreation.Pets	Arts.Photography
Recreation.Scouting	Science.Maths
Science.Physics	Business.Marketing
Sports.Hockey	Business.Logistics

attracts a relatively non-spontaneous and well organized users is etsy.com—a market place of hand-made crafts. Apparently, one needs a degree of Conscientiousness in addition to a general arty profile, to trade art.

3.7 Data validation

To mitigate the risk of biases in our data and to minimize the risk of random effects, we evaluate the consistency of the findings between the two sources of website preference data.

Table 4 Similarities between personality profiles of art-related websites, estimated using the Liked URL dataset. The columns labelled O through N represent the five personality traits, *freq* indicates the number of distinct users who liked each website. The column labelled *SEM* is the standard error of the mean, which was of similar magnitude for all of the five personality traits and is hence presented in a single column

Domain	O	C	E	A	N	freq	SEM	Pearson correlation					
								1	2	3	4	5	6
(1) deviantART.com	.40	-.19	-.42	-.05	.16	3,154	.01-.02	1					
(2) Tumblr.com	.23	-.23	-.16	-.10	.22	639	.03	.89	1				
(3) Etsy.com	.41	.14	-.26	.07	.10	612	.03	.88	.59	1			
(4) Gaia online.com	.33	-.23	-.40	.00	.19	2,076	.02	.99	.91	.82	1		
(5) Fanboy.com	.36	-.27	-.44	-.04	.22	128	.07-.09	.99	.93	.81	1	1	
(6) RainyMood.com	.36	-.22	-.35	.03	.12	236	.06-.07	.99	.88	.84	.99	.98	1

Table 5 Pearson's correlation between personality estimated using both WPQ and Likes datasets

Personality Trait	
Openness	0.77
Conscientiousness	0.83
Extroversion	0.57
Agreeableness	0.78
Neuroticism	0.83
# common websites	23

Table 6 Correlation between personality profiles estimated using our datasets. Correlation coefficients were averaged using Fisher's z transformation

domain	WPQ to Likes
deviantart.com	0.98
tumblr.com	0.78
etsy.com	0.98
nba.com	0.94
snagajob.com	0.85
job.com	0.77
gamefaqs.com	0.59
digg.com	0.86
ancestry.com	0.64
myxer.com	0.92
pandora.com	0.84
barnesandnoble.com	0.34
foodnetwork.com	0.83
Average correlation	0.83

First, we estimate Pearson's product-moment correlation between the aggregated personality traits across the two datasets, as shown in Table 5. For instance, the value of 0.83 for Conscientiousness between Likes dataset and WPQ indicates that average Conscientiousness of the website preference group in the Facebook Likes dataset correlates highly ($r = 0.83$) with the average Conscientiousness estimated using the WPQ dataset. The correlation coefficients presented in Table 5 indicate high consistency between aggregated personality profiles established using different sources of data collected from different individuals.

Second, to examine the consistency of the entire personality profiles, we correlated the five personality estimates between datasets (Table 6). There were 14 websites for which the data was available in both samples (note, that there were only 23 websites in the WPQ sample). The average Pearson product-moment correlation between personality profiles estimated using two different samples ranged from 0.78 to 0.83. This indicates that aggregated personality profiles were stable across the datasets. For instance, the profile of deviantART.com presented in Fig. 1 is very similar across all of the datasets.

The high level of consistency observed across the samples provides strong evidence supporting the validity of our findings and methods.

4 Study 2: personality and facebook profile features

Facebook profiles have become an important source of information used to form impressions about others. For example, people examine other people's Facebook profiles when trying to decide whether to start dating them (Zhao et al. 2008), and when assessing job candidates (Finder 2006).

Study 2 explores the relationship between personality and the features of the Facebook profiles. We continue and expand the work of Amichai-Hamburger and Vinitzky (2010), Golbeck et al. (2011), Gosling et al. (2011), and Ross et al. (2009) regarding personality and social network profiles, attempting to overcome some of the limitations of those studies, in particular their relatively small (at most a few hundred participants) and biased (mostly student) samples. The large sample used in this study is more representative of the general online population and enables us to make more statistically significant conclusions. We also employ regression techniques to predict users' personalities based on their Facebook profiles.

This section starts with a description of the previous work relevant to this study followed by the results. The correlations between profile features and personality reported in the results are compared with those reported by Amichai-Hamburger and Vinitzky (2010), Ross et al. (2009).

4.1 Related work

Existing work (Correa et al. 2010; Ryan and Xenos 2011; Zhong et al. 2011) has shown that certain personality traits are correlated with total internet usage and with the propensity of individuals to use social media and social networking sites. However, these papers focus on the amount of *time* spent using these tools rather than the specific features individuals engage and interact with. These papers add value by identifying the personality profiles of heavy internet and Facebook users, but shed little light on the issue of how a person's Facebook profile reflects their personality.

The existing research has additionally shown that Facebook profiles reflect the actual personality of their owners rather than an idealized projection of desirable traits (Back et al. 2010). Researchers asked participants to assess the personality of the owners of a set of Facebook profiles and revealed that they could correctly infer at least some personality traits. This implies that they do not deliberately misrepresent their personalities on their Facebook profiles, or at least do not misrepresent them to a larger extent than they do in psychometric tests.

Despite the fact that people can judge other people's personalities based on their Facebook profiles or web browsing history, it is possible that some of the personality cues are ignored or misinterpreted. As humans we are prone to biases and prejudices which may affect the accuracy of our judgements. Recent work (Evans et al. 2008), examining what aspects of the Facebook profile individuals use to form personality judgements, shows that certain features are difficult to grasp for people. For example, while the number of Facebook friends is clearly displayed on the profile, people cannot easily determine features such as the network density (whether a user's friends know each other).

Several earlier papers investigate the relationships between personality traits and Facebook profile features. We briefly describe below a selection of studies closest in spirit to our work.

Golbeck et al. (2011) attempted to predict personality from Facebook profile information using machine learning algorithms. They used a very rich set of features, including both

Facebook profile features, such as the ones we use in this work, but also the words used in status updates. However, their sample ($n = 167$) was very small, especially given the number of features used in prediction ($m = 74$), which limits the reliability and generalizability of their results.

Gosling et al. (2011) revealed several connections between personality and self-reported Facebook features. For example, they showed the positive relationship between Extroversion and frequency of Facebook usage and engagement in the site. As in offline contexts, Extroverts seek out virtual social engagement, leaving behind a behavioural residue such as friendship connections or picture postings. However their work was based on a relatively small sample of 157 participants, again limiting the reliability and generalizability of their results.

Quercia et al. (2012) studied the relationship between Facebook popularity (number of contacts) and personality traits, showing that Extroversion predicts the number of Facebook contacts. They also found no statistical evidence for the relationship between popularity and self-monitoring—a personality trait describing an ability adapt to new forms of communication, present oneself in likeable ways, and maintain superficial relationships.

Ross et al. (2009) pioneered the study of the association between personality and patterns of social network usage. The study proposes a number of hypotheses but reports only one significant correlation—between Extroversion and group membership. A relatively small ($n = 97$) and homogeneous sample (mostly female students studying the same subject at a single university), and a potentially unreliable approach to collecting data (participants' self-reports of their Facebook profile features, rather than direct observation) may have prevented the authors from finding more significant connections and make it difficult to extrapolate findings to a general population.

In a similar study, Amichai-Hamburger and Vinitzky (2010) used actual Facebook profile information rather than self-reports, although their sample was still small ($n = 237$) and homogeneous (Economics and Business Management students of an Israeli university). They found several significant correlations, but some of their findings were in contradiction to those of Ross et al. (2009). For example, they found that Extroversion was positively correlated with the number of Facebook friends, but uncorrelated with the number of Facebook groups, whereas Ross et al. (2009) found that Extroversion had an effect on group membership, but not on the number of friends. Additionally, they found that high Neuroticism was positively correlated with users posting their own photo, but negatively correlated with uploading photos in general, while Ross et al. (2009) argued that high Neuroticism is negatively correlated with users posting their own photo.

Following the work of Ross et al. (2009) and Amichai-Hamburger and Vinitzky (2010) the present study focused on the relationship between personality and Facebook use, but based on a much larger sample (350,000 versus 97 and 237 users in Ross et al. (2009) and Amichai-Hamburger and Vinitzky (2010) respectively). Similar to Amichai-Hamburger and Vinitzky (2010) we have recorded actual features of the Facebook profiles instead of relying on potentially unreliable self-reports such as those used in Ross et al. (2009).

4.2 Facebook profile features

Facebook profile features were obtained for more than 354,000 US Facebook users, who had used Facebook for at least 24 months before the data was recorded.

The total number of friends, events, status updates, photos, photo tags and membership in groups accrue on user profiles over time. In order to account for this process, before analysis, those features were divided by the number of months since the user was active on Facebook,

Table 7 Summary of Facebook features used in this study, including their labels, number of users for which data on given feature was available, median value, and thresholds of the first and third quartiles

Label	Details	n	\tilde{x}	Q_1	Q_3
Friends	Average number of friends	62,780	295	182	488
Network Density	Density of the friendship network	28,835	.03	.014	.05
Groups	Groups joined	97,622	15	7	32
Likes	Number of Facebook Likes	109,837	82	33	196
Photos	Number of photos uploaded	110,540	10	4	29
Statuses	Number of status updates posted	71,912	99	38	195
Photo Tags	Number of times others tagged user in photos	315,169	20	4	83
Events	Number of events attended	8,977	10	4	32
Age	User's age	198,959	23	20	28

estimated by looking for an earliest sign of users' activity in our records—e.g. users' first status update, photo tag, uploaded photo, or attended event (we did not have access to the date on which given Facebook account was created). Interestingly, the number of Likes did not significantly depend on the amount of time since joining Facebook and hence we used the total number here.

The density of the friendship network largely relates to its size, which is a well known property of social networks. Therefore, a simple linear regression model was built explaining log-transformed density by the log-transformed network size, and it was used to remove the effect of the network size on density. The residual present in this model, which is an equivalent of the density not explained by the sheer size of the network, was used as a measure of an individual user's network density.

Many Facebook users had incomplete profile information or their privacy settings did not allow for accessing some parts of their profile. Consequently, not all of the features were available for all of the users, but we had at least 9,000 data points per feature and over 100,000 data points for the majority of the features. The frequencies of Facebook features used in this study are presented in Table 7.

4.3 Correlating personality with Facebook profile features

We began by correlating personality with Facebook profile features using Spearman's rank correlation, which is appropriate for variables characterized by a long tailed distribution. Table 8 summarizes the correlations found. We have tested the statistical significance of these results using a t -distribution test and all reported correlations were significant at the $p < .01$ level. We carried out an additional statistical significance test, and compared the top and bottom thirds of the population in terms of various Facebook features (for example, the third of the population with the fewest, and the most friends). We used a Mann-Whitney-Wilcoxon test (MWW-test, also known as the a Mann-Whitney U test or the Wilcoxon rank-sum test) to determine whether the top and bottom thirds of the population differ significantly in terms of their mean personality score (for various different traits). Again, the test showed all relations are significant at the $p < .01$ level.

Selected correlations are also represented on Figs. 2 to 7. The horizontal axis of those figures represents the standardized psychological trait score while the vertical axis represents the median of a given Facebook feature (e.g. the median number of Facebook friends for users characterized by a given level of Extroversion). In order to increase the clarity of

Table 8 Statistically significant correlations between personality and Facebook profile features. All reported correlation coefficients are significant at $p < .01$ level

Psychological Trait	Profile Feature	Spearman Rank Correlation
Openness	Likes	.09
	Statuses	.07
	Groups	.07
Conscientiousness	Likes	−.11
	Groups	−.09
Extroversion	Statuses	.12
	Friends	.18
	Events	.08
	Groups	.08
Neuroticism	Likes	.12
	Statuses	.09

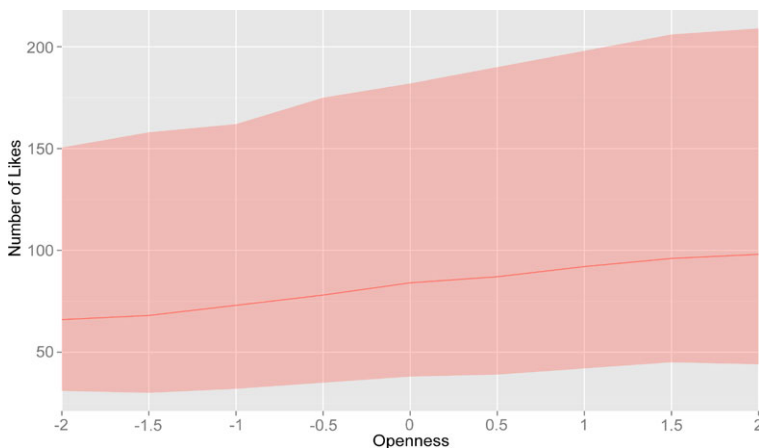


Fig. 2 Median number of Likes for users characterized by different levels of Openness. Ribbon represents the interquartile range, or the middle 50 percentiles of the number of users' Likes

the plots, users were grouped by their standardized trait scores rounded to the nearest half integer (e.g. users with standardized Openness score between 1.75 and 2.24 were grouped together and represented by a score of 2). The shaded ribbon represents the interquartile range, IQR (also referred to as “middle fifty”) for the Facebook feature. The IQR is the range of values between the 25 % percentile of the population to the 75 % percentile of the population.

Table 8 presents significant Spearman rank correlations between Facebook profile features and psychological traits. It is clear that the correlations, while psychologically meaningful, are relatively low. However, inspecting the plots which represent the same relation graphically offers some valuable insights.

Openness Liberal and open to experience individuals tend to Like more items on Facebook (Fig. 2), post more status updates and join more groups, which is consistent with the definition of this personality trait. Highly open users do not only choose different Likes and

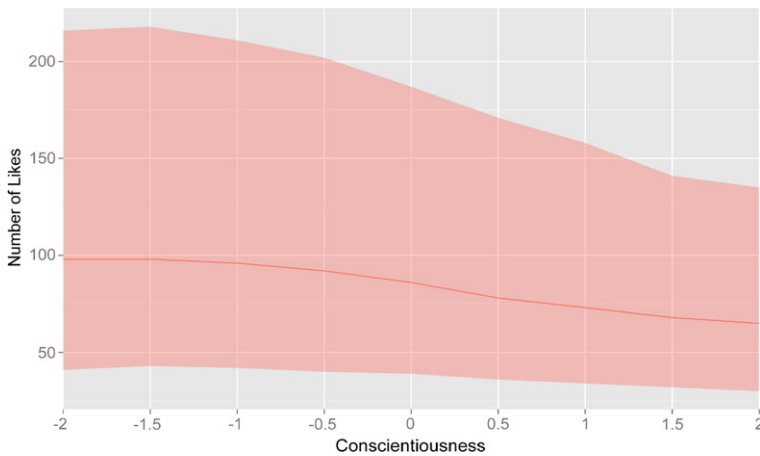


Fig. 3 Median number of Likes for users characterized by different levels of Conscientiousness. Ribbon represents the interquartile range, or the middle 50 percentiles of the number of users' Likes

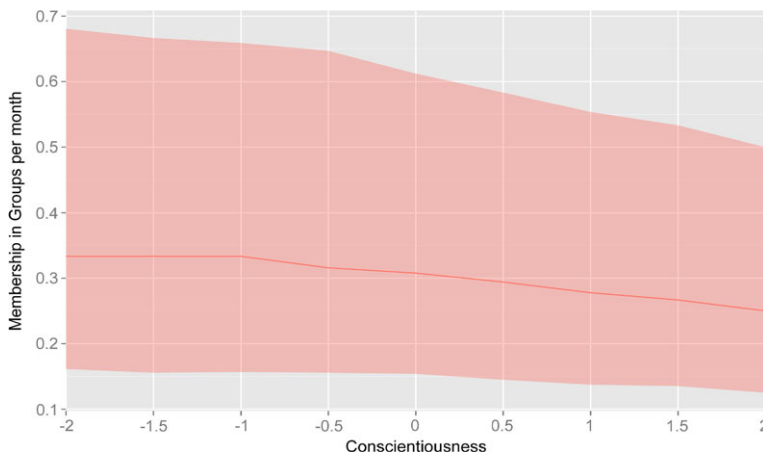


Fig. 4 Median number of groups joined per month by users characterized by different levels of Conscientiousness. Ribbon represents the interquartile range, or the middle 50 percentiles of the number of groups joined per month

Groups than conservative ones (as shown by Kosinski et al. 2013) but are also willing accept a wider range of objects.

This results confirm the hypotheses of Ross et al. (2009) and results presented in Amichai-Hamburger and Vinitzky (2010) suggesting that individuals high on Openness are more willing to use Facebook as a communication tool and to use a greater number of features.

Conscientiousness As presented in Figs. 3 and 4, spontaneous (low on Conscientiousness) individuals tend to join more groups and Like more things. Interestingly, conscientious individuals do not only join less groups and use Like feature less frequently, but also are more homogeneous in doing so, as indicated by a significant drop in the interquartile ranges.

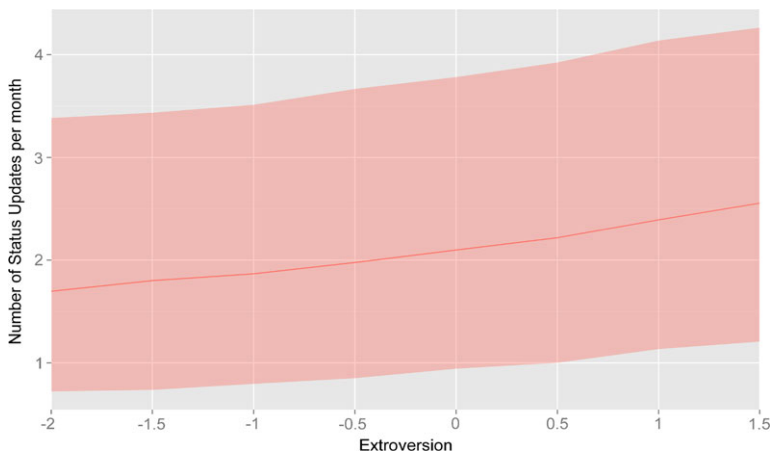


Fig. 5 Median of status updates posted per month by users characterized by different levels of Extroversion. Ribbon represents the interquartile range, or the middle 50 percentiles of the number of status updates posted per month

Figure 3 shows that the median number of Likes among highly conscientious individuals is higher by 40 Likes from the most spontaneous ones. Also, while 25 % of spontaneous users have more than 210 Likes, the same value for conscientious users is lower by a third (140 Likes).

These results confirm the hypothesis of Ross et al. (2009) that Conscientiousness is negatively related to engaging in Facebook activities (importantly, Ross et al. (2009) were not able to confirm this hypothesis). Furthermore, the results do not support the hypothesis and results presented in Amichai-Hamburger and Vinitzky (2010) who found that Conscientiousness was positively related to number of friends.

Extroversion Our results show that Extroverts are generally more likely to reach out and interact with others on Facebook. They more actively share what is going on in their lives or their feelings with other people (and allow other people respond to these) using status updates (Fig. 5), they attend more events, and interact more with other individuals using Facebook groups, which allows them to exchange information and connect with individuals outside their immediate friendship circle. Finally, Extroversion relates to the number of Facebook friends, as showed by Fig. 6.

Previous results of Ross et al. (2009) show a positive link between Extroversion and group membership but no relationship with the number of friends, while Amichai-Hamburger and Vinitzky (2010) showed a positive link between Extroversion and number of friends but no effect in regards to the use of Facebook groups. The current work suggests that such conflicting results may have stemmed from the relatively small sample sizes limiting the ability to establish significant relationships.

Agreeableness Agreeableness does not appear to be significantly correlated with any of the Facebook profile features studied in this paper. This result suggests that the relationship between Agreeableness and the number of friends hypothesized but not confirmed by both Ross et al. (2009) and Amichai-Hamburger and Vinitzky (2010) may in fact be non-existent.

Neuroticism Figure 7 show that Neuroticism is positively correlated with the number of Facebook Likes, indicating that more emotional users tend to use the Like function more

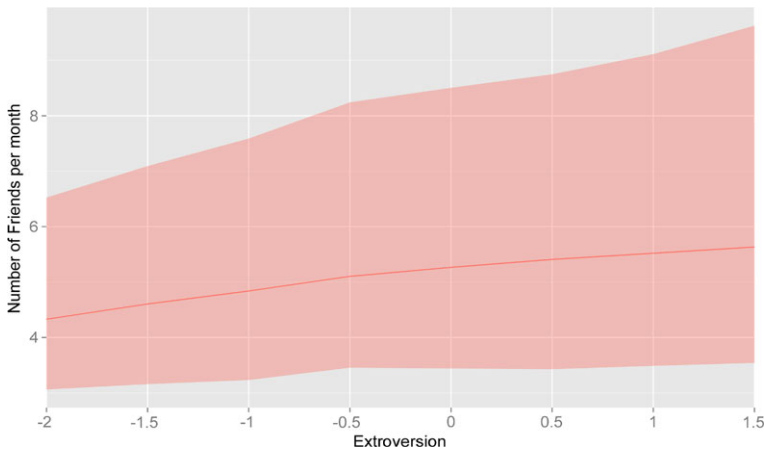


Fig. 6 Median of friends added per month by users characterized by different levels of Extroversion. Ribbon represents the interquartile range, or the middle 50 percentiles of the number of friends added per month

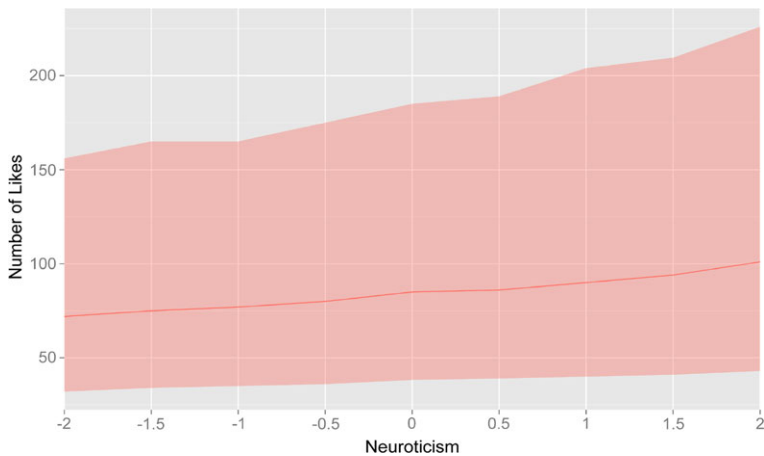


Fig. 7 Median number of Likes for users characterized by different levels of Neuroticism. Ribbon represents the interquartile range, or the middle 50 percentiles of the number of users' Likes

frequently. While 75 % of the most stable users like fewer than approximately 150 Likes, 75 % of the most emotional users like more than 220 Likes.

Those results are in agreement with the hypothesis proposed in Ross et al. (2009) and Amichai-Hamburger and Vinitzky (2010) suggesting that neurotic individuals would be more willing to share personal information on Facebook. While we did not find any significant relationships between Neuroticism and number of photos uploaded by the users, we found positive relationships between Neuroticism and number of likes, and status updates—serving similar function.

4.4 Predicting personality

So far we have examined the relationship between personality traits and Facebook profile features. We now discuss *predicting* personality based on multiple profile features. We

Table 9 Predicting personality, Satisfaction with Life, Intelligence, and age based on multiple profile features using a multivariate linear regression with 10-fold cross validation. Table presents prediction accuracy expressed by the Pearson correlation coefficient, sample size, and Facebook features used in the prediction

Trait	Accuracy (r)	n	Features used in the prediction
Openness	.11	18,720	Friends, Groups, Likes, Network Density, Photo Tags, Photos
Conscientiousness	.16	18,720	Friends, Groups, Likes, Network Density, Photo Tags, Photos
Extroversion	.31	16,900	Friends, Groups, Likes, Network Density, Photo Tags, Statuses
Agreeableness	.05	45,565	Friends, Likes, Photo Tags
Neuroticism	.23	9,515	Friends, Likes, Photos, Statuses
Satisfaction with Life	.33	311	Events, Likes, Network Density, Photo Tags
Intelligence	.20	395	Events, Likes, Photo Tags, Photos, Statuses
Age	.50	3,826	Events, Friends, Groups, Likes, Network Density, Photo Tags, Photos

used a simple prediction method, multivariate linear regression, and examined our results using 10-fold cross validation. The Facebook profile features used in this analysis were log-transformed in order to normalize their distribution. As a measure of the goodness of fit, we used the Pearson correlation coefficient between the predicted and actual personality values.

We first performed a bi-directional stepwise variable selection based on Akaike Information Criterion (Burnham and Anderson 2002) to select the best model from a set of models by minimizing the Kullback-Leibler divergence between the model and the truth. This greedy procedure starts with all predictive variables and keeps removing the variable which, when removed, most improves the quality of the model until no further improvement is possible. Next, it repeatedly adds the variable that most significantly improves the quality of the model. Effectively, each of the personality traits is being predicted with a different subset of Facebook profile features. As the overlap between the data related to different Facebook features is not complete, each of the personality traits is predicted using a sample of a different size.

The results, presented in Table 9, indicate that Extroversion is most highly expressed by Facebook features, followed by Neuroticism, Conscientiousness, and Openness. Agreeableness is the hardest trait to predict using our Facebook profile features and the simple model used in this study.

As a comparison we present the accuracy achieved while using the same data to predict age and two additional psychological traits: Intelligence and Satisfaction with Life (see Kosinski et al. (2013) for details on how those traits were measured). It is apparent that all of the psychological traits, apart from Agreeableness, are manifested with similar strength in Facebook profile features. Age can be predicted with significantly higher accuracy ($r = .5$). Note, however, that while age is relatively easy to estimate, personality scores are estimated with a significant degree of error. For instance, the accuracy of the Extroversion scale used in this study, expressed by its test-retest reliability (correlation between the scores of the same person taking the test on two different occasions), equals $r = .75$, constituting an upper limit for the prediction accuracy for this trait.

It is worth comparing the prediction accuracies achieved in this study with those achieved in the same environment but using different types of signal. In our previous study we explored the predictive power of Likes associated with a given Facebook account and achieved the accuracy of $r = .3$ for Conscientiousness, Agreeableness, and Neuroticism, $r = .43$ for Openness, $r = .4$ for Extroversion, and $r = .75$ for age (Kosinski et al. 2013). The accuracies achieved in the current paper are consistently lower indicating that individual selection of Likes is more informative in terms of personality. This effect is especially strong for Openness, which is relatively well manifested in the selection of Likes, but very weakly represented in the aggregate features of Facebook behaviour.

We note that multiple linear regression is one of the simplest predictive methods. However, inspecting the relationship between log-transformed Facebook features and personality (that are relatively weak but predominantly linear), and the similar accuracy achieved using other prediction methods, indicate that the results presented in Table 9 do indeed capture the ability to predict personality using the Facebook profile features used in this study.³

5 Limitations

Studies measuring personality are often limited to the lab environment and rely on a small or moderate population of volunteers to self-report their personal behaviours and preferences under certain situations. While in this study we have used a very large and hence more representative population of respondents, we have also restricted it to US-based Facebook users. Although this avoided cultural biases in our results, it limits the generalizability of our findings. In addition, our volunteers came from a typically western, educated, industrialized, rich and democratic (WEIRD) (Henrich et al. 2010) society. Moreover, our observations are limited to volunteers who opted-in to participate in the research. While the demographic structure of the population used in this study matches the general Facebook population, it is possible that our volunteers differed from the general population in some other way, for example in their psychological traits. We hope this self-selection effect is partly mitigated by the scale of our experiments.

Another issue that plagues traditional psychometric studies is that participants may lie when taking tests. To some extent, we believe that information regarding Liked websites on Facebook is less prone to lying and misrepresentation, as people provide this data in a natural environment rather than in a test situation and had to opt-in for their data to be recorded. That said, Facebook users may be selective in which websites they Like, promoting the impression that they have a particular personality that is perhaps different from their actual personality (for further discussion on this issue, see Back et al. 2010).

6 Conclusions and future work

We studied how a user's personality manifests itself in their use of online social networks, as reflected by features of their Facebook profile, and in their preference for websites. Potential

³We applied several additional machine learning methods for predicting traits, which may allow exploiting various forms of non-linear relations between variables. These include tree based rule-sets M5 Rules (Holmes et al. 1999), support vector machines (SVM) with polynomial and RBF kernels, and decision stumps (for details on these methods see Bishop 2006; Holmes et al. 1994). The results using these approaches are similar to the ones we obtained using linear regression, giving some support to our conjecture that the errors stem mostly from a high variance in personality traits even among people with similar Facebook profile features.

applications for this work are online advertising and recommender systems. By analysing information from online social networks it would be possible to “profile” individuals, automatically dividing users into different segments, and tailor advertisements to each segment based on their personality. Similarly, one can imagine building recommender systems based on personality profiles.

On one hand, we have shown that personality can be, to some extent, inferred from a user’s Facebook profile, which gives rise to important privacy concerns, especially since many users may not be aware of how revealing such information can be. Whereas, our analysis indicates that preferences for online content and websites reflect the personality of users and that aggregate statistics regarding an *audience* personality can be reliably collected in a privacy preserving manner and be used to improve the quality of Internet services.

Many interesting directions are left open for future research. First, we have already pointed out that we only used very specific features and that we believe that a wide variety of other cues warrants additional study, especially regarding “micro-level” features such as the specific groups a user is a member of, or the specific items they Like (see Kosinski et al. 2013 for further discussion of this).

Second, we have also noted that this study, similarly to other online studies, suffers from user sampling biases caused by self-selection and sparsity. More sophisticated approaches could be used to overcome such biases and provide accurate uncertainty estimates by using meaningful priors.

Third, our analysis examines actual behaviour of users in their natural online environment, rather than simply using self-reports. We hope that such approaches would allow increasing the scale as well as the quality of psychological assessment. By observing that personality is reflected in online behaviour, our approach enables further studies of personality and its relation with other aspects of online behaviour. Given a personality profile of a sufficiently large set of websites, larger scale studies of personality based on browsing behaviour are likely possible, allowing personality to be correlated with any other observable information about users.

Finally, another important direction for future work would be the study of privacy preserving mechanisms such as differential privacy (Dwork 2006) for processing and aggregating online behavioural data, to provide even stronger guarantees that users’ privacy is respected in such studies.

References

- Allport, G. W. (1962). The general and the unique in psychological science. *Journal of Personality*, 30, 405–422.
- Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, 26(6), 1289–1295.
- Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of facebook usage. In *Proceedings of the ACM web science conference*.
- Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality*, 42(4), 1116–1122.
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21(3), 372.
- Baglioni, M., Ferrara, U., Romei, A., Ruggieri, S., & Turini, F. (2003). Preprocessing and mining web log data for web personalization. In *Proceedings of the congress of the Italian association for artificial intelligence*.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44(1), 1–26.
- Bennett, P. N., Svore, K., & Dumais, S. T. (2010). Classification-enhanced ranking. In *International world wide web conference*.

- Bishop, C. M., & SpringerLink (Service en ligne) (2006). *Pattern recognition and machine learning* (Vol. 4). New York: Springer.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Berlin: Springer.
- Byrne, D., Griffitt, W., & Stefaniak, D. (1967). Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology*, 5(1), 82.
- Correa, T., Hinsley, A. W., & De Zuniga, H. G. (2010). Who interacts on the web? The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2), 247–253.
- Costa, P. T. Jr., & McCrae, R. R. (1992). Neo personality inventory–revised (neo-pi-r) and neo five-factor inventory (neo-ffi) professional manual. In *Psychological assessment resources*, Odessa, FL.
- Costa, P. T., & McCrae, R. R. (2006). *Revised NEO personality inventory (NEO PI-R): manual*. Boston: Hogrefe.
- De Bock, K., & Van Den Poel, D. (2010). Predicting website audience demographics for web advertising targeting using multi-website clickstream data. *Fundamenta Informaticae*, 98(1), 49–70.
- Dwork, C. (2006). Differential privacy. In *International colloquium on automata, languages and programming*.
- Evans, D. C., Gosling, S. D., & Carroll, A. (2008). What elements of an online social networking profile predict target-rater agreement in personality impressions. In *Proceedings of the international conference on weblogs and social media* (pp. 1–6).
- Finder, A. (2006). For some, online persona undermines a résumé. *New York Times*, 11.
- Gill, A. J., Oberlander, J., & Austin, E. (2006). The perception of e-mail personality at zero-acquaintance. *Personality and Individual Differences*, 40, 497–507.
- Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *ACM SIGCHI conference on human factors in computing systems* (pp. 253–262).
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *The American Psychologist*, 48(1), 26.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe*, 7, 7–28.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84–96.
- Gosling, S. D., Ko, S., 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.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *The American Psychologist*, 59(2), 93–104.
- Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, 14(9), 483–488. doi:[10.1089/cyber.2010.0087](https://doi.org/10.1089/cyber.2010.0087).
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not weird. *Nature*, 466(7302), 29.
- Holmes, G., Donkin, A., & Witten, I. H. (1994). Weka: a machine learning workbench. In *Proceedings of the 1994 second Australian and New Zealand conference on intelligent information systems*, 1994 (pp. 357–361). New York: IEEE Press.
- Holmes, G., Hall, M., & Prank, E. (1999). *Generating rule sets from model trees*. Berlin: Springer.
- Hu, J., Zeng, H.-J., Li, H., Niu, C., & Chen, Z. (2007). Demographic prediction based on user's browsing behavior. In *The world wide web conference* (pp. 151–160).
- John, O. P., & Benet-Martinez, V. (2000). Measurement, scale construction, and reliability. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 339–369). New York: Cambridge University Press.
- Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52(3), 621–652.
- Kelly, E. L., & Conley, J. J. (1987). Personality and compatibility: a prospective analysis of marital stability and marital satisfaction. *Journal of Personality and Social Psychology*, 52(1), 27.
- Khan, I. A., Brinkman, W.-P., Fine, N., & Hierons, R. M. (2008). Measuring personality from keyboard and mouse use. In *Proceedings of the 15th European conference on cognitive ergonomics: the ergonomics of cool interaction*.
- Kosinski, M., Stillwell, D., Kohli, P., Bachrach, Y., & Graepel, T. (2012). Personality and website choice. In *Proceedings of the ACM web science conference*.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. In *Proceedings of the National Academy of Sciences*.
- Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology*, 90(6), 1014–1031.

- Murray, D., & Durrell, K. (1999). Inferring demographics attributes of anonymous Internet users. In *ACM SIGKDD conference on knowledge discovery and data mining*.
- Netscape Communication Corporation (2013). Open directory project. <http://dmoz.org/>.
- Orzeck, T., & Lung, E. (2005). Big-five personality differences of cheaters and non-cheaters. *Current Psychology*, 24, 274–287.
- Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57, 401–421.
- Quercia, D., Lambiotte, R., Kosinski, M., Stillwell, D., & Crowcroft, J. (2012). The personality of popular facebook users. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*.
- 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.
- Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: an empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, 58(1), 103–139.
- Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with facebook use. *Computers in Human Behavior*, 25(2), 578–586.
- Russell, M. T., Karol, D. L., & Institute for Personality, and Ability Testing (1994). *The 16PF fifth edition administrator's manual*. Champaign: Institute for Personality and Ability Testing.
- Ryan, T., & Xenos, S. (2011). Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage. *Computers in Human Behavior*, 27(5), 1658–1664. doi:[10.1016/j.chb.2011.02.004](https://doi.org/10.1016/j.chb.2011.02.004).
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel Psychology*, 44(4), 703–742.
- Tupes, E. C., & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, 60(2), 225–251.
- Vazire, S., & Gosling, S. D. (2004). E-perceptions: personality impressions based on personal websites. *Journal of Personality and Social Psychology*, 87, 123–132.
- Weber, I., & Castillo, C. (2010). The demographics of web search. In *ACM SIGIR special interest group on information retrieval*.
- Weber, I., & Jaimes, A. (2011). Who uses web search for what: and how. In *International conference on web search and data mining* (pp. 15–24).
- Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on facebook: digital empowerment in anchored relationships. *Computers in Human Behavior*, 24(5), 1816–1836.
- Zhong, B., Hardin, M., & Sun, T. (2011). Less effortful thinking leads to more social networking? the associations between the use of social network sites and personality traits. *Computers in Human Behavior*, 27(3), 1265–1271. doi:[10.1016/j.chb.2011.01.008](https://doi.org/10.1016/j.chb.2011.01.008).