

Personality and Patterns of Facebook Usage

Yoram Bachrach
Microsoft Research
yobach@microsoft.com

Michal Kosinski
University of Cambridge
mk583@cam.ac.uk

Thore Graepel
Microsoft Research
thoreg@microsoft.com

Pushmeet Kohli
Microsoft Research
pkohli@microsoft.com

David Stillwell
University of Cambridge
ds617@cam.ac.uk

ABSTRACT

We show how users' activity on Facebook relates to their personality, as measured by the standard Five Factor Model. Our dataset consists of the personality profiles and Facebook profile data of 180,000 users. We examine correlations between users' personality and the properties of their Facebook profiles such as the size and density of their friendship network, number uploaded photos, number of events attended, number of group memberships, and number of times user has been tagged in photos. Our results show significant relationships between personality traits and various features of Facebook profiles. We then show how multivariate regression allows *prediction* of the personality traits of an individual user given their Facebook profile. The best accuracy of such predictions is achieved for Extraversion and Neuroticism, the lowest accuracy is obtained for Agreeableness, with Openness and Conscientiousness lying in the middle.

Author Keywords

Social Networks, Personality, Big Five Personality Model

ACM Classification Keywords

H.4 Information Systems Applications: Miscellaneous

General Terms

Social Networks, Personality, Algorithms

INTRODUCTION

An individual's success depends largely on the impression made on others. Success on the job market, finding romantic partners, and gaining support and positive attention from one's social background heavily depend on what others think of you. The shift of human interactions, socialization and communication activities towards on-line platforms means that managing the impression of one's on-line presence is increasingly important.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Web Science '12, June 22–24, 2012, Evanston, IL, USA.

One of the most ubiquitous on-line environments, Facebook, is becoming an increasingly natural environment for a growing fraction of the world's population. Currently it facilitates daily interactions of over 800 million users spending more than 40 minutes daily on the platform on average [1]. Facebook profiles became 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 [26], and they are also used when assessing job candidates [10].

Recently, it was shown that people's personality can be successfully judged by the others based on their Facebook profiles [9, 16]. Researchers asked participants to assess the personality traits of the owners of a set of Facebook profiles and showed that they could correctly infer at least some personality traits. Moreover, it was shown that a Facebook profile reflects the actual personality of its owner rather than an idealized projection of desirable traits [3]. This implies that people do not deliberately misrepresent their personalities on their Facebook profiles, or at least do not misrepresent them to a larger extent than in psychometric tests.

The fact that people can judge each other's personality based on Facebook profiles implies two things: an individual's personality is manifested on their Facebook profile, and some aspects of Facebook profiles are used by people to judge others' personalities. However, the overlap between Facebook profile features that contain the actual personality cues and features used by people to form personality judgements does not have to be perfect. It is possible that some of the actual personality cues are ignored or misinterpreted by the people, while some non-relevant features are used in the judgment. Humans are prone to biases and prejudices which may affect the accuracy of their judgements. Also, certain features of a Facebook profile are difficult for humans to grasp. For example, while the number of Facebook friends is clearly displayed on the profile, it is more difficult for a human to determine features such as the network density. Recent work [9] examines which aspects of the Facebook profile humans use to form personality judgements.

The current study focuses on how personality is *manifested* through different features of the Facebook profile. We extract various *high-level features* of a Facebook profile and show how these correlate with its owner's personality, as measured by a standard Five Factor Model personality questionnaire. Examined profile features fall into two broad categories. First, aspects of the profile that depend exclusively

on a user's actions, including: the number of published photos, events and groups the user has uploaded or created and the number of objects the user has "liked". Second, aspects of the profile that depend on the actions of a user and their friends, including the number of times a user has been tagged in photos, and the size and density of their friendship network. The dataset we analyse is relatively large and diverse, consisting of over 180,000 users described by personality score and the records of their Facebook profile features. We show how these features correlate with a user's personality and contrast our findings with the previous work in the area. We also demonstrate how to determine personality based on the Facebook profile rather than on a personality questionnaire.

We continue and expand the work of [2, 11, 15, 21] attempting to overcome some of their limitations, and most of all their rather small (at most a few hundred participants) and biased (mostly student) samples. The small and biased samples make it difficult to reach statistically significant conclusions, or to employ regression techniques to predict users' personalities based on their Facebook profiles. Also, focusing on student populations leads to an unrepresentative sample with reduced variance.

The Big Five Personality Model

We use the Five Factor Model [8, 12, 22], which is currently the most widespread and generally accepted model of personality, whose ability to predict human behaviour has been well-studied. This model was examined in [12, 25] and shown to subsume most known personality traits. It was thus claimed to represent the "basic structure" underlying human personality. This model provides a nomenclature and a conceptual framework that unifies much of the research findings in psychology of individual differences and personality. The five personality traits are described in more detail in the Results section.

Previous research showed that personality is correlated with many aspects of life, including job success [4, 18, 24], attractiveness [6], marital satisfaction [19] and happiness [20]. Existing work [7, 23, 27] has already shown that certain personality traits are correlated with total internet usage and with the propensity of users to use social media and social networking sites. However, these papers focus on the amount of *time* spent using these tools rather than on *how* individuals are using them. This body of work adds value by identifying the personality profiles of heavy internet and Facebook users, but sheds little light on the question how a person's Facebook profile reflects that individual's personality.

RESEARCH QUESTIONS

Several hypotheses have been raised regarding the relation between personality and Facebook profile features. Closest in spirit to our work are [2, 11, 15, 21], briefly described below.

Ross et al. [21] pioneered the study of the relation between personality and patterns of social network use. They hypothesized many relationships between personality and Facebook features, including (1) positive relationship between Extraversion and Facebook use, number of Facebook friends

and associations with Facebook groups; (2) positive relation between Neuroticism and revealing private information on Facebook; (3) positive correlation between Agreeableness and number of Facebook friends; (4) positive correlation between Openness and number of different Facebook features used; (5) negative relationship between Conscientiousness and overall use of Facebook. Unfortunately, this study was based on a relatively small ($n = 97$) and homogeneous sample (mostly female students of the same subject at the same university) which limited the power of their analyses and make it difficult to extrapolate their findings to a general population. Also, this study relied on participants' self-reports of their Facebook profile features, rather than direct observation. Consequently, Ross et al. [21] were only able to present one significant correlation - between Extraversion and group membership, leaving all the remaining hypotheses unverified.

Examining a similar set of hypotheses, Amichai-Hamburger and Vinitzky [2] used the actual Facebook profile information rather than self-reports. They found several significant relationships, however, their sample was still small ($n = 237$) and very homogeneous (Economics and Business Management students of an Israeli university). Moreover, some of their findings were opposite to those of Ross et al. [21]. For example, they find that Extraversion is positively correlated with the number of Facebook friends, but uncorrelated with the number of Facebook groups, whereas Ross et al. [21] find that Extraversion has an effect on group membership, but not on the number of friends. Also, they find that high Neuroticism is positively correlated with users posting their own photo, but negatively correlated with uploading photos in general, while [21] posit that high Neuroticism is negatively correlated with users posting their own photo.

Golbeck et al. [11] attempted to predict personality from Facebook profile information using machine learning algorithms. They use a very rich set of features, including both high-level features, such as the ones we use in this work, and "micro-features" such as 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.

Our work is closest to that of Gosling et al. [15], using both *self-reported* patterns of Facebook usage as well as actual Facebook profile features. However their work was based on a relatively small sample of 157 participants, so while our analysis closely follows theirs, we use a larger sample size.

Studies listed above were based on very limited and often homogeneous samples and lead to some contradictory findings. Our main goal was to use a large and representative sample of Facebook users to settle the question of how personality is expressed in Facebook profiles. We test the following hypotheses: a) Openness and Neuroticism are positively correlated with the number of status updates, photos, groups and "likes" of an individual. b) Conscientiousness is negatively correlated with all aspects of Facebook use: number of friends, likes, photos, etc. c) Extraversion is positively correlated with all aspects of Facebook use d) Agreeableness is positively correlated with the number of friends, groups and

Feature	Details
Friends	number of Facebook friends
Groups	number of associations with groups
Likes	number of Facebook “likes”
Photos	number of photos uploaded by user
Statuses	number of status updates by user
Tags	number of times others “tagged” user in photos

Table 1. Facebook profile features used in this study.

“likes”.

The second major contribution of this work relates to the level of aggregation at which predictions are being made. Most previous work focused on correlating Facebook profile features with personality traits averaged over large *groups*, but were inaccurate on the *individual* level. Similar to [11], we use our large sample to show that by *combining* signals from different Facebook features it is possible to reliably predict personality of *individuals*. Due to the difficulty and cost of testing large samples using a laboratory approach, we used viral marketing to collect personality data using an application within the Facebook environment.

METHODOLOGY

Our dataset of 180,000 users was obtained using *myPersonality*¹, a Facebook application deployed in 2007. The application allows Facebook users to complete a standard Five Factor Model questionnaire [13, 14] and to obtain feedback regarding their personality based on their responses. After filling the questionnaire, users can give their consent to record their Facebook profile information and personality scores for research purposes. The list of features used in this research is listed in Table 1. Many Facebook users had incomplete profile information or security settings preventing us from accessing some parts of their profile, and due to time constraints and bandwidth limitations some of the features were recorded for a fraction of the users. Consequently, not all of features were available for all of the users, but we had at least 15,000 data points per feature and over 50,000 data points for most of the features.

To a large extent our sample was representative of the general Facebook population, with an average age of 24.15 (SD=6.55) and an overrepresentation 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 personality².

CORRELATIONS BETWEEN PERSONALITY TRAITS AND FACEBOOK PROFILE FEATURES

Our first set of results focuses on each of the Facebook features *separately*, and shows their relationship with each of the Big Five personality traits.

¹available at: <https://apps.facebook.com/mypersonality>

²Demographics of Facebook users can be checked on Facebook at <http://newsroom.fb.com/> and at CheckFacebook 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.

Our methodology for correlating Facebook features and personality traits sorts users into deciles according to their score on each Facebook feature. We cluster together users with similar Facebook features, and examine average values of personality trait scores. We sort the n users according to that feature, for example from the user with the smallest number of Facebook friends to the user with the greatest number of friends, to obtain the sorted list u_1, u_2, \dots, u_n . Denote the feature value (e.g. number of friends) of user u_i as c_i . After sorting the users, we partition them into k equal and disjoint sets according to their order, i.e., the set S_1 of $q = \frac{n}{k}$ users with the smallest feature values (smallest number of friends), the following set S_2 of q users with slightly higher feature values and so on until the set S_k of q users of the highest feature values (users with the most friends). This process generates k disjoint sets of q users each, based on a simple “cutoff” criterion. The first set S_1 contains all users with a feature value of at most c_q (i.e. at most c_q friends), the second set S_2 contains all users with feature value of more than c_q but less than c_{2q} , and so on, until the final set S_k which contains all users with a feature value of more than c_{n-q} and at most c_n .

Partitioning the users into groups of equal size but with increasing feature values allows us to examine the relation between the a given Facebook feature and each of the personality traits. We partition users into $k = 10$ large groups (for most features we have several thousands of users in each group). Thus, the average personality trait score of the users in each group is a very accurate estimate of the expected personality trait score of users falling into that group’s cut-off criterion.

We also produce plots presenting the relationship between Facebook features and personality, where horizontal axis represents the average Facebook feature value of the given group (e.g. the average number of Facebook friends of users whose number of friends was within the cut-off range), and vertical axis represents the average personality trait score for this group³. We call such plots “Clustered Scatter Plots”.

Results

We first present the clustered scatter plots for the most significant correlations between Facebook profile features and personality traits. Then we show how accurately personality can be predicted based on the users’ profile features. We present our results regarding each of the Big Five personality traits.

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 high 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 changes and are usually more conservative and closed-minded.

³Note that as the distribution of Facebook features is rarely normal, an average Facebook feature score is usually not exactly in the middle between the top and bottom cut-off values.

Our results, presented on Figures 1, 2, and 3 indicate that Openness is positively correlated with number of users' likes, group associations and status updates. These results are not surprising as all of those features indicate users' greater involvement in seeking new things and ideas and sharing with their friends.

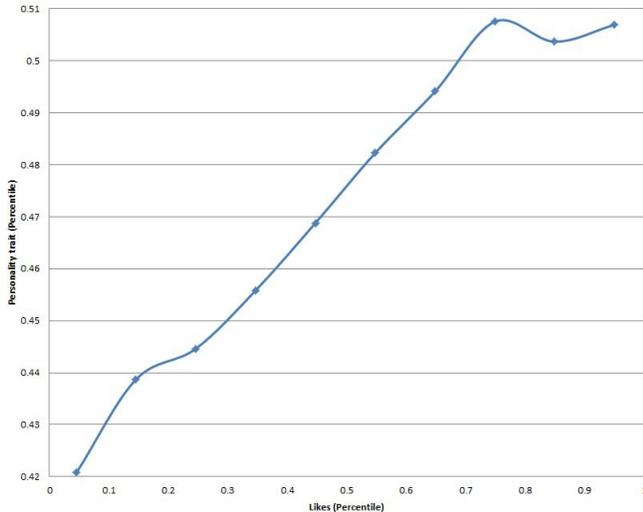


Figure 1. Clustered Scatter Plot showing expected Openness score as a function of the number of users' Facebook likes (see text for the description of plotting technique).

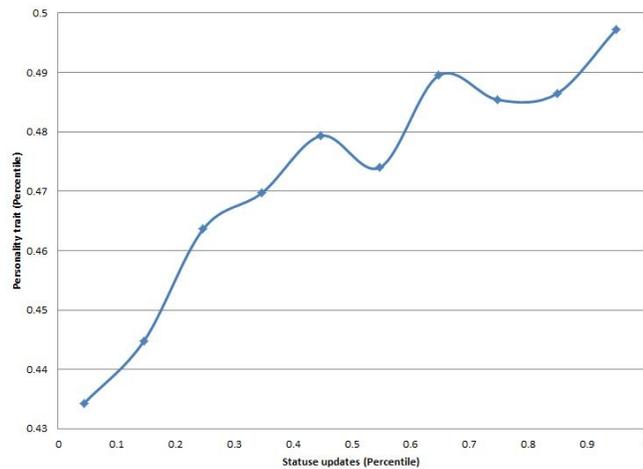


Figure 2. Clustered Scatter Plot showing expected Openness score as a function of the number of users' status updates (see text for the description of plotting technique).

Conscientiousness measures preference for an organized versus spontaneous approach in life. People high on Conscientiousness are more likely to be well organized, reliable, and consistent. They enjoy planning, 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.

As presented on Figures 4, 5, and 6 Conscientiousness is negatively related to the number of likes and group membership, but positively related to the number of uploaded photos.

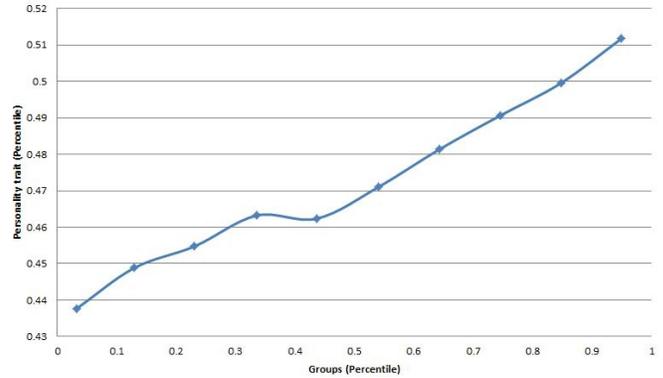


Figure 3. Clustered Scatter Plot showing expected Openness score as a function of the number of associations with Facebook groups (see text for the description of plotting technique).

This may indicate that conscientious people are less eager to show their appreciation for an object or a group. Additionally, it may be the case that conscientious, better organized and less spontaneous individuals consider using Facebook to be a waste of time and a distraction from other activities such as work and thus tend to like fewer objects and join fewer groups. However, Figure 6 shows that more conscientious people tend to upload more images to Facebook, so in this aspects they tend to be more active than other users. It is possible that the diligence and good organization of highly conscientious people predisposes them to focus more on uploading and organizing their pictures using the tools Facebook offers.

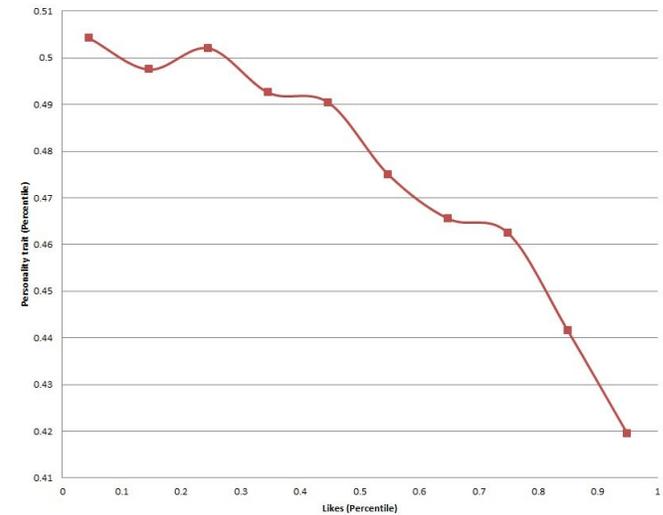


Figure 4. Clustered Scatter Plot showing expected Conscientiousness score as a function of the number of users' likes (see text for the description of plotting technique).

Extraversion measures a person's tendency to seek stimulation in the external world, company of others, and express positive emotions. Extraverts tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative, do not mind being the centre of attention, and make new friends more easily. Introverts are more likely to be solitary or reserved and seek environments characterized by lower levels of external stimulation.

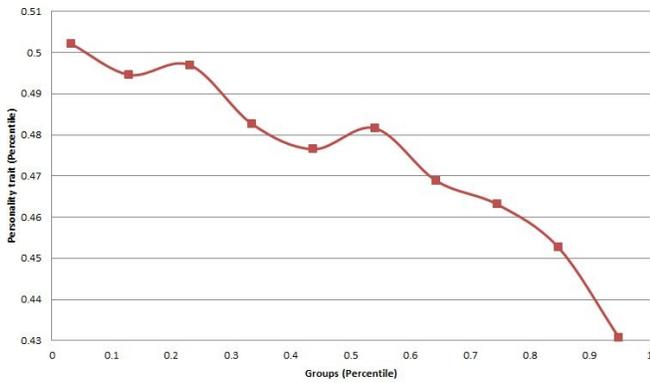


Figure 5. Clustered Scatter Plot showing expected Conscientiousness score as a function of the number of associations with Facebook groups (see text for the description of plotting technique).

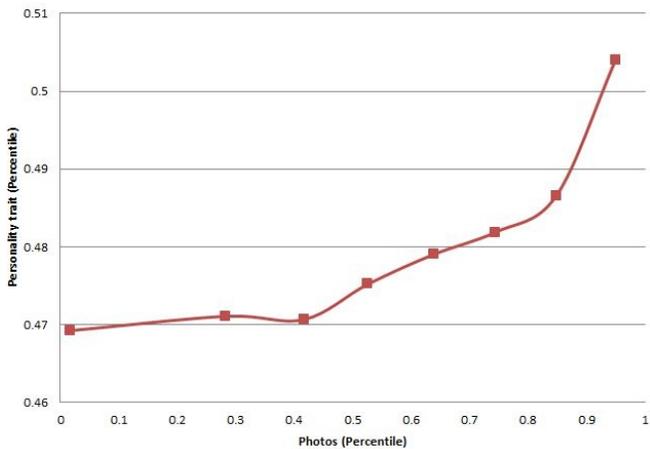


Figure 6. Clustered Scatter Plot showing expected Conscientiousness score as a function of the number of photos uploaded by the user.

Our results show that Extraverts are more likely to reach out and interact with other people on Facebook. They are more actively sharing what is going on in their lives or their feelings with other people (and let other people respond to these) using status updates (Figure 7). Extraverts seem to be more likely to engage with a content and objects shared by their friends by “liking” them, indicating their appreciation or sympathy (Figure 8). Also, they tend to interact more with other users using Facebook groups, which allow exchanging information and interacting with a wider set of people than the direct friends of the user (Figure 9).

Finally, Extraversion relates to the number of Facebook friends, as depicted by Figure 10. The correlation is generally positive, but the increase in average Extraversion occurs only for groups with the average number of friends above 50th percentile. Also, those with extremely few friends (around 30) are on average more Extraverted than those with average and below average numbers of friends. This can possibly be explained by the fact that Facebook friends accrue with time - users discover new friends and add new ones, but rarely delete them. Additionally, some users sign up for Facebook but never actively use it. Therefore, it is likely that users who have very few friends are not especially introverted but sim-

ply abandoned their accounts or joined Facebook relatively recently. Consequently, the relationship between Extraversion and number of friends becomes pronounced for those users who have more than average number of friends.

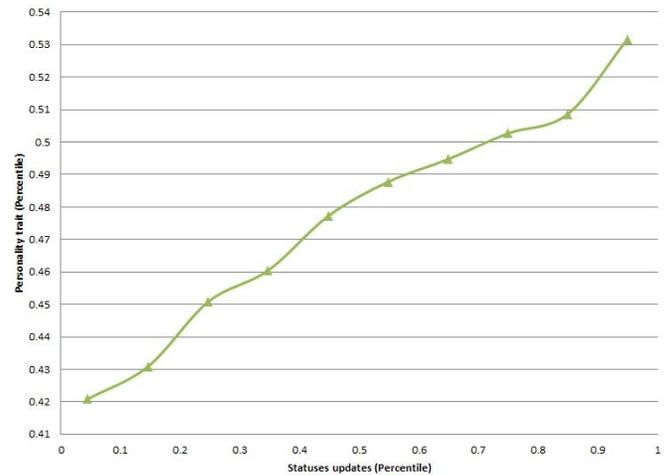


Figure 7. Clustered Scatter Plot showing expected Extraversion score as a function of the number of users' status updates (see text for the description of plotting technique).

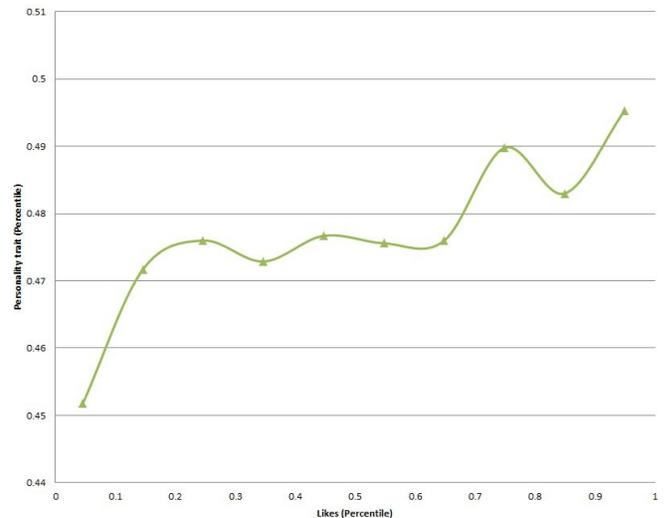


Figure 8. Clustered Scatter Plot showing expected Extraversion score as a function of the number of users' likes (see text for the description of plotting technique).

Agreeableness measures the extent to which a person is focused on maintaining positive social relations. High Agreeableness people tend to be friendly and compassionate, rather than cold or suspicious. They are more likely to behave in a cooperative way, trust other people, and adapt to their needs. Unsurprisingly, such likable people more often appear in pictures with other users, as expressed by higher number of tags (Figure 11). Note however, that the effect is visible only for a relatively large number of tags, with no significant correlation for users with fewer than 50 tags.

Those low on Agreeableness are focused on themselves, less likely to compromise, and may be less gullible. They also

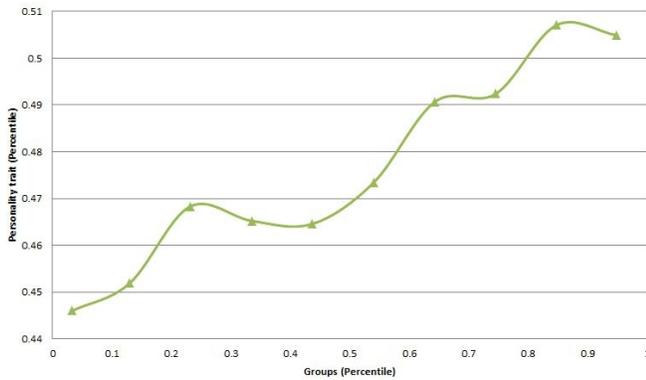


Figure 9. Clustered Scatter Plot showing expected Extraversion score as a function of the number of associations with Facebook groups (see text for the description of plotting technique).

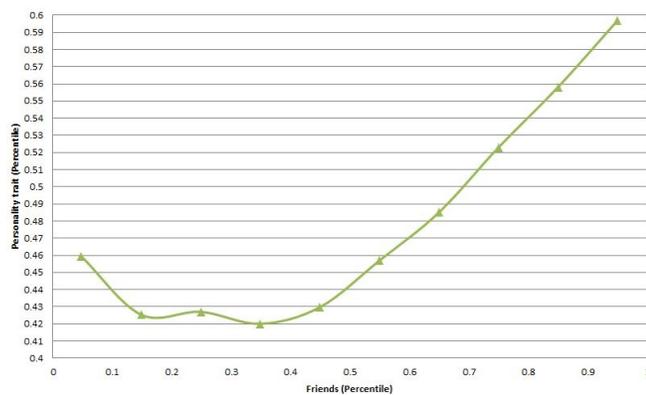


Figure 10. Clustered Scatter Plot showing expected Extraversion score as a function of the number of Facebook friends (see text for the description of plotting technique).

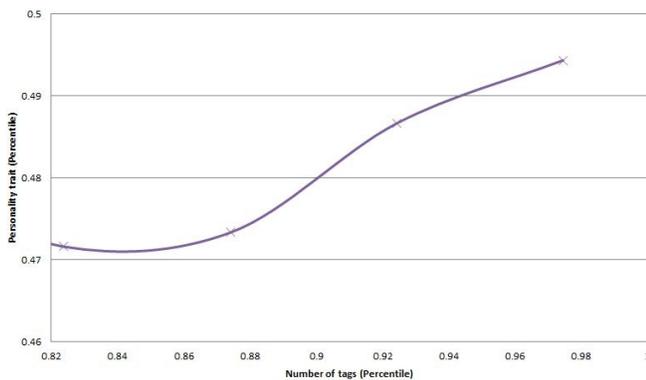


Figure 11. Clustered Scatter Plot showing expected Agreeableness score as a function of the number of times a user was tagged in photos.

tend to be less bound by social expectations and conventions, and be more assertive. That might explain why Agreeableness is somewhat negatively correlated with the number of likes (Figure 12). It is likely that users characterized by low Agreeableness are less concerned with what others may think about them and thus are liking different objects more freely, while more Agreeable people may be afraid that liking things can put them in the opposition to their friends. For exam-

ple, liking Christianity may offend one's Muslim or atheist friends.

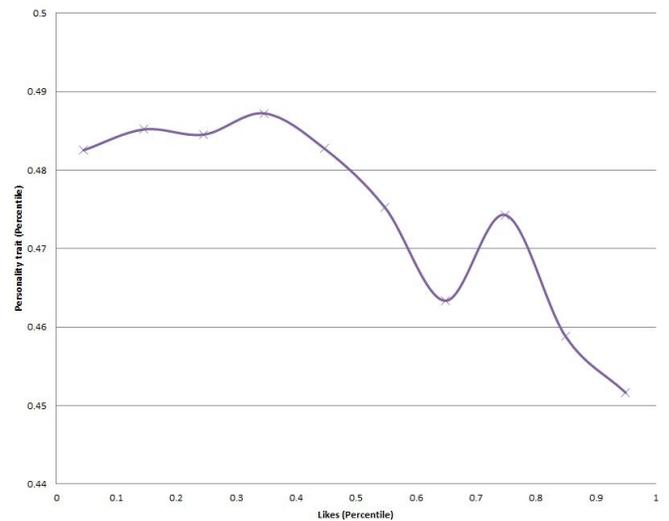


Figure 12. Clustered Scatter Plot showing expected Agreeableness score as a function of the number of users' likes (see text for the description of plotting technique).

In general, however, Agreeableness appears to be less correlated with high-level Facebook features than the other four of the Big Five personality traits. Most of the high-level Facebook features, such as the number of groups, likes or friends show no significant monotone correlation with Agreeableness.

Neuroticism, often referred to as emotional instability, is a 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 self-confident.

Figures 13 and 14 show that Neuroticism is positively correlated with the number of Facebook likes and slightly negatively correlated with number of groups. This effect is somewhat similar to the correlation between the number of Facebook likes and Openness to experience, shown in Figure 1. However, especially in the case of likes, the effect for Neuroticism is moderate for the lower levels of likes and stronger for the users with many likes, while the effect for Openness is strong for lower numbers of likes and saturates as the number of likes increases.

One possible explanation for the correlation between Neuroticism scores and the number of likes and groups is that more Neurotic users often tend to feel negative emotions such as anxiety, anger, or depression. One way to help alleviate these is to seek support from friends. Thus Neurotic users may seek support through activity in Facebook groups or hope to get support by liking other users' updates, hoping they would reciprocate by supporting them.

Figure 15 shows an interesting relation between number of friends and Neuroticism. It seems that average Neuroticism increases with number of friends, until reaching peak levels

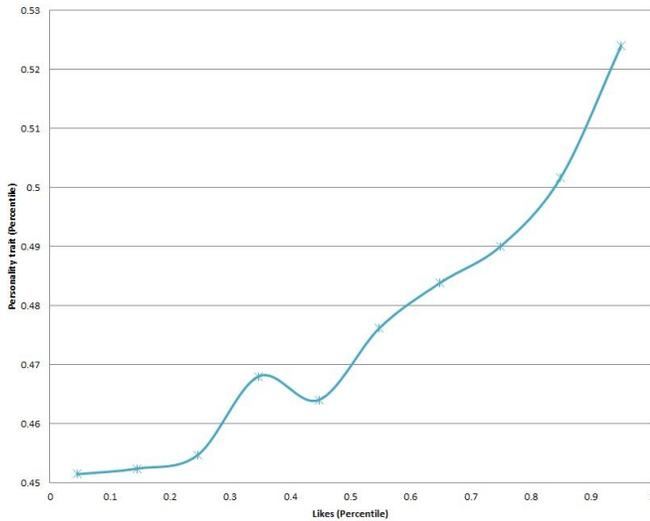


Figure 13. Clustered Scatter Plot showing expected Neuroticism score as a function of the number of users' likes (see text for the description of plotting technique).

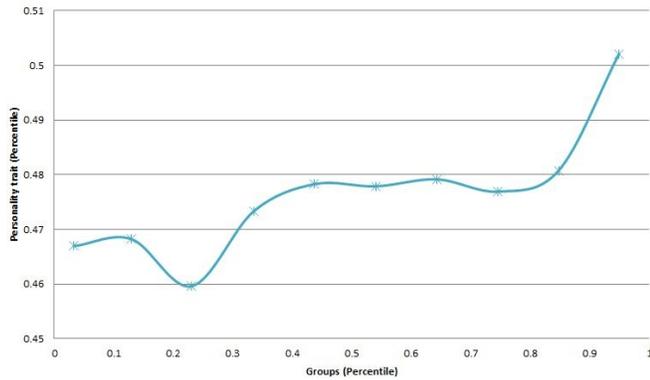


Figure 14. Clustered Scatter Plot showing expected Neuroticism score as a function of the number of associations with Facebook groups (see text for the description of plotting technique).

for roughly 200 friends. Beyond this peak level, Neuroticism becomes negatively associated with number of friends. One possible way to explain this effect is that Neurotic users need more support to alleviate negative feelings, which they can only get from a relatively small number of close friends. Neurotic However, beyond a certain number of friends, additional friendships are typically very superficial, providing little additional support. Thus, very Neurotic people may tend to have fewer friends, but maintain closer relations, providing more support.

Summary of Findings and Statistical Significance

The clustered scatter plots provide a convenient visualization of the relationships between personality traits and Facebook profile features, but do not measure the strength and significance of such relationships. To test statistical significance we applied two tests. First, we have tested the statistical significance of the correlations (against the null hypothesis of no correlation) using a t-distribution test. Except for the relation between Agreeableness and the number of tags which is

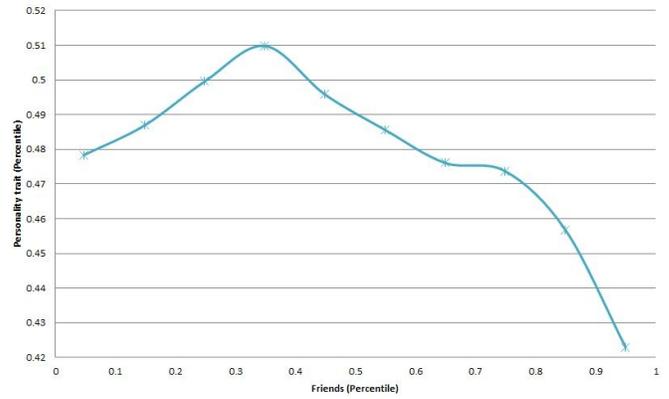


Figure 15. Clustered Scatter Plot showing expected Neuroticism score as a function of the number of Facebook friends.

Personality Trait	Profile Feature	Pearson Correlation
Openness	Likes	0.102
	Statuses	0.062
	Groups	0.077
Conscientiousness	Likes	-0.088
	Groups	-0.0697
	Photos	0.0330
Extraversion	Statuses	0.117
	Likes	0.034
	Groups	0.069
	Friends	0.177
Agreeableness	Likes	-0.036
Neuroticism	Likes	0.075
	Friends	-0.059

Table 2. Statistically significant correlations between personality traits Facebook profile features (at a significance level of $p < 1\%$).

further discussed below, all correlations were found to be significant at the $p < 1\%$ level. This provides strong evidence for the above claims regarding correlations⁴. We carried 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 friends and with 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 has shown all relations are significant at the $p < 1\%$ level. Table 2 summarizes the correlations found.

We note that Neuroticism has a generally significant negative correlation with the number of friends, but Figure 15 reveals that the relation roughly follows an inverse U curve. Similarly, the correlation between Extraversion and the number of friends is positive, but Figure 10 reveals the strong positive correlation holds mostly for high numbers of friends. Though

⁴Note the relation between Neuroticism and the number of friends indicate a non-linear relation. Still, the general linear negative trend for this relation was found significant at the $p < 1\%$ level.

Trait	R^2	RMSE
Openness	0.11	0.29
Conscientiousness	0.17	0.28
Extraversion	0.33	0.27
Agreeableness	0.01	0.29
Neuroticism	0.26	0.28

Table 3. Predicting personality traits using Facebook features through multivariate linear regression

the relation between Agreeableness and the number of Facebook tags was weaker, an MWW-test shows that the top 10% of the population in the number of Facebook tags has significantly higher Agreeableness scores than the bottom 10%, at the $p < 5\%$ level. This provides some evidence that Agreeableness is positively correlated with the number of tags, but this relation is weaker than the other relations in Table 2.

PREDICTING PERSONALITY

Previous sections examined the correlations between each of the Big Five personality traits and Facebook profile features. We now turn to making predictions about an individual's personality based on multiple profile features. We focus on a subset of 5,000 individuals for whom all of the Facebook features listed in Table 1 were available.

We first used multivariate linear regression with 10-fold cross validation, attempting to predict each of the traits using the available profile information. As a measure of the goodness of fit, we used the coefficient of determination, R^2 . For our regression, we have converted raw scores of both features and personality traits to percentiles. Thus, rather than saying that a user has 100 friends, we say she is in the 20th percentile in terms of the number of friends. Similarly, rather than saying a user has an Extraversion score of 4, we say she is in the 80th percentile of Extraversion scores. This improves the quality of the regression, and allows easy comparison between traits. Table 3 lists the R^2 and root mean squared error (RMSE) for each of the personality traits, predicted using the Facebook profile information.

As the above table indicates, some personality traits can be predicted with reasonable accuracy using Facebook features, whereas other traits are more difficult to predict using the high-level Facebook features we have used. For Extraversion we obtained the model with the best fit, with an R^2 value of 0.33, indicating quite accurate a prediction. Predictions regarding Neuroticism are also reasonably accurate, with R^2 of 0.26. The model for Conscientiousness has a lower fit, and the model for Openness is even less accurate. It seems that Agreeableness is the hardest trait to predict using our Facebook profile features, and our model has a poor fit indeed.

We note that multiple linear regression is one of the simplest statistics/machine learning methods. We also applied several more sophisticated machine learning methods for predicting traits, including tree based rule-sets, support vector machines, and decision stumps (for details on these methods see [5, 17]). However, for all of the personality traits, both the R^2 values and RMSE change very little when using more sophisticated machine learning methods (changes are mostly in the third

digit after the decimal point).

CONCLUSIONS AND LIMITATIONS

In this study we show that personality traits are correlated with patterns of social network use, as reflected by features of Facebook profile, using a sample which is greater in size than that used in previous work by several orders of magnitude. Table 2 summarizes the key significant correlations, settling open hypotheses and contradictory findings in earlier work (such as [2, 21]). Further, the study shows that by combining several features, we can make relatively accurate predictions regarding a *individual's* personality, with Extraversion being most easy to predict and Agreeableness being most elusive. One potential application for our work is online advertising and recommender systems. By analysing information from social networks it would be possible to "profile" individuals, automatically dividing users into different segments, and tailor advertisements to each segment based on personality. Similarly, one can imagine building recommender systems based on personality profiles.

The approach presented here has several limitations. First, the data used may suffer from a self-selection bias, as we only have data for users who are active on Facebook and who have decided to use our personality analysis application. Further, users were able to control the information stored regarding their profile, so we only had data for users who chose to let us access this information. Also, the Facebook features we used are high-level aggregate features. For example, we used the number of Facebook "likes", rather than examining which objects were liked, or we counted the number of status updates rather than considering the words used in the status updates. It remains an open research question to see whether such fine-grained information can be used to predict personality, and to see whether personality can also be predicted using other potentially observable online behaviour such as a user's internet browsing or web search history.

ACKNOWLEDGEMENTS

We would like to thank Sam Gosling for very valuable comments on this work.

REFERENCES

1. Facebook f8: Redesigning and hitting 800 million users. *LA Times*, September 2011.
2. Y. Amichai-Hamburger and G. Vinitzky. Social network use and personality. *Computers in Human Behavior*, 26(6):1289–1295, 2010.
3. M.D. Back, J.M. Stopfer, S. Vazire, S. Gaddis, S.C. Schmukle, B. Egloff, and S.D. Gosling. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21(3):372, 2010.
4. M.R. Barrick and M.K. Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26, 1991.
5. C.M. Bishop and SpringerLink (Service en ligne). *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

6. D. Byrne, W. Griffitt, and D. Stefaniak. Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology*, 5(1):82, 1967.
7. T. Correa, A.W. Hinsley, and H.G. De Zuniga. Who interacts on the web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2):247–253, 2010.
8. P.T. Costa Jr and R.R. McCrae. Neo personality inventory–revised (neo-pi-r) and neo five-factor inventory (neo-ffi) professional manual. *Odessa, FL: Psychological Assessment Resources*, 1992.
9. D.C. Evans, S.D. Gosling, and A. Carroll. 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*, pages 1–6, 2008.
10. A. Finder. For some, online persona undermines a résumé. *New York Times*, 11, 2006.
11. J. Golbeck, C. Robles, and K. Turner. Predicting personality with social media. In *CHI*, pages 253–262, 2011.
12. L.R. Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1):26, 1993.
13. L.R. Goldberg. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, 7:7–28, 1999.
14. L.R. Goldberg, J.A. Johnson, H.W. Eber, R. Hogan, M.C. Ashton, C.R. Cloninger, and H.G. Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96, 2006.
15. S.D. Gosling, A.A. Augustine, S. Vazire, N. Holtzman, and S. Gaddis. Manifestations of personality in online social networks: Self-reported facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, 2011.
16. S.D. Gosling, S. Gaddis, and S. Vazire. Personality impressions based on facebook profiles. 2007.
17. G. Holmes, A. Donkin, and I.H. Witten. Weka: A machine learning workbench. In *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, pages 357–361. Ieee, 1994.
18. T.A. Judge, C.A. Higgins, C.J. Thoresen, and M.R. Barrick. The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, 52(3):621–652, 1999.
19. E.L. Kelly and J.J. Conley. Personality and compatibility: A prospective analysis of marital stability and marital satisfaction. *Journal of Personality and Social Psychology*, 52(1):27, 1987.
20. D.J. Ozer and V. Benet-Martinez. Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.*, 57:401–421, 2006.
21. C. Ross, E.S. Orr, M. Sisic, J.M. Arseneault, M.G. Simmering, and R.R. Orr. Personality and motivations associated with facebook use. *Computers in Human Behavior*, 25(2):578–586, 2009.
22. M.T. Russell, D.L. Karol, Institute for Personality, and Ability Testing. *The 16PF fifth edition administrator's manual*. Institute for Personality and Ability Testing Champaign, IL, 1994.
23. T. Ryan and S. Xenos. Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage. *Computers in Human Behavior*, 2011.
24. R.P. Tett, D.N. Jackson, and M. Rothstein. Personality measures as predictors of job performance: a meta-analytic review. *Personnel psychology*, 44(4):703–742, 1991.
25. E.C. Tupes and R.E. Christal. Recurrent personality factors based on trait ratings. *Journal of Personality*, 60(2):225–251, 1992.
26. S. Zhao, S. Grasmuck, and J. Martin. Identity construction on facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, 24(5):1816–1836, 2008.
27. B. Zhong, M. Hardin, and T. Sun. Less effortful thinking leads to more social networking? the associations between the use of social network sites and personality traits. *Computers in Human Behavior*, 2011.