

Detecting Personality Traces in Users' Social Activity

Styliani Kleanthous¹(✉), Constantinos Herodotou¹, George Samaras¹,
and Panayiotis Germanakos^{1,2}

¹ Department of Computer Science, University of Cyprus,
CY-1678 Nicosia, Cyprus

{stellak, cherod02, cssamara, pgerman}@cs. ucy. ac. cy

² Suite Engineering UX, Products and Innovation,
SAP SE, 69190 Walldorf, Germany

Abstract. The effect that social media have in our lives nowadays is apparent. Many studies focused on how the differences we hold as people due to our personality, reflect our activities online. In this work we aim to exploit reports of previous work to implicitly build a personality model of Facebook users, based on their Facebook activity. An initial evaluation study shows that using Facebook activity data, we can extract information on user personality and at the same time points in further improvements necessary for more accurate personality prediction.

Keywords: Social networks · Big Five personality model · User modeling

1 Motivation

The behavior of users online has been the subject of many studies in social sciences and computing e.g. [1–4]. Results in cognitive psychology, show that the general personality factors predict very well aspects of internet use [4]. In this line, personality traits can be reflected in the activity and navigation of users online [1, 4].

“Big Five” personality domains are described, as five dimensions that define human personality and predict aspects of human behavior. These five dimensions as formulated by Goldberg [5] are: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. In order for a person to be categorized in one of the five dimensions he/she has to answer the “Big Five” personality questionnaire. Most of the studies performed by social scientists on correlating Facebook activity with personality use this questionnaire as a reference point for user personality prediction, and a second questionnaire for ‘extracting’ the behavior of a user on Facebook e.g. [4, 6].

Similarly, studies coming from the technology perspective consider the “Big Five” questionnaire for extracting the user’s personality, but they automatically extract user activity from Facebook offline (e.g. [7–9]). This can be considered as a more unbiased method for Facebook activity extraction since the user is not directly involved in the process. Furthermore, to correlate Facebook activities to the personality of the user, they employ machine learning and data mining techniques e.g. [3, 7, 9]. However, what

these studies are missing is to feedback the user with the results of his/her personality based on his/her Facebook activity.

Hence, the goal of this work is to take advantage of the results reported on previous theoretical and technical research, particularly in social sciences [1, 4, 6] and computing [2, 7, 8, 10], on which and how, Facebook user activities (e.g. share, like, checkin) relate to the personality of a user. A computational mechanism has been defined for implicitly extracting a user personality model in real time based on the user's activity on Facebook. Within this mechanism a Facebook application has been developed (PersonaWeb app) that allows us to access users' private data and develop a user personality model. The PersonaWeb app is then used to communicate (through visualizations) this information back to the users.

The contribution of this work lays primarily in exploring whether using the data reported on previous research we can: (i) define and implicitly develop a user model in real time, contained of user interaction data and (ii) develop a user personality model, by exploiting the data stored in the user model. Furthermore, through the PersonaWeb app, the user can instantly visually compare the results of his/her personality model extracted, with the results obtained from the "Big Five" personality questionnaire that he/she has answered.

2 Big Five Personality Traits

Before we discuss related technological approaches on extracting personality traits from Facebook it is important to understand how personality traits relate to user, activities and behaviour on Facebook based on the results reported by behavioural and psychology sciences.

Extraversion: People in this dimension have an inherent need to advertise their activities to others and their good mood depends on the feedback they receive from them. People in this category tend to spend more hours in social networking sites [11]. Particularly, in Facebook, they tend to belong to more groups and have more friends [4, 11]. Furthermore, they have the tendency to upload more personal photos than people belonging to other personality dimensions, share more statuses and post more check-ins [7].

Agreeableness: Individuals in this trait are perceived as kind, sympathetic, cooperative, warm and considerate. People who score high on this dimension tend to believe that most people are honest, decent, and trustworthy. The behavior in social networks for people who score high in this dimension, prefer to communicate more with personal messages on Facebook with their friends and be more involved with online games offered in Facebook's API [6]. Furthermore, individuals who fall into this category do not use social networks (i.e. Facebook) for a long time and refrain from making posts on their friends [7, 11].

Conscientiousness: Is a characteristic that defines a person who is being thorough, careful, or vigilant. It is recorded that because people with high conscientiousness are more committed to goals, their activity on Facebook will be decreased in relation with

other users of the social network [6]. This implies that they will spend less time in Facebook; they will have fewer friends and they will publish fewer photographs and statuses [10]. In addition, people who fall into this category will rarely like any posts or belong to groups [3].

Neuroticism: This dimension describes people with the tendency to experience strongly negative emotions, such as anger, anxiety, or depression. People characterized by neuroticism, tend to be more frequent users of Facebook since they want to control the information about themselves and their environment [6]. Thus, the most frequent activity they practice is to disseminate information or statements that they approve. In contrast they avoid publishing photos of themselves [11]. Furthermore, neurotics tend to have fewer friends on Facebook, but at the same time, use often the like function in posts of these friends [3, 7].

Openness: Describes people with a general appreciation for art, emotion, adventure, unusual ideas, imagination, curiosity, and variety of experience. People who belong in this dimension are more likely to hold unconventional beliefs. People who are distinguished for their openness to experience, due to their receptivity available for new experiences, will tend to use the social network more to inquire new experiences [7]. They will publish frequently [12] their statuses and they tend to use Facebook's like function in anything that intrigues them [6].

3 Extracting Personality Traits from Facebook

Due to the rapid development of Facebook, compared to other social networks, and due to the enormous amount of information available for most users, many research groups have tried to acquire and exploit the log data in order to draw conclusions in relation to personality. Two main techniques are used and discussed below.

Semi-automated data mining approaches utilize algorithms to extract information from public profiles on Facebook [9, 13]. In any case, the users involved in the study have to complete a personality questionnaire in order for the researchers to get an indication on the users' personality. Data mining algorithms and machine learning [7] are followed in analysing and correlating the activity of users to personality traits. User's replies to the personality questionnaire are used for evaluating the models developed. These studies showed that textual elements [3] and demographic profile information of users can provide indication of user's personality and that indeed personality is closely related to social networks usage. Although data mining algorithms and machine learning approaches are un-obstructive methods for the user, and predict user personality with high accuracy, publicly available information are getting fewer [7] as time passes due to Facebook's new privacy policies and settings. Consequently, the information one can get using this method is not rich and similarly to the previous discussion the user is not directly getting anything back.

A different approach is followed in automated processes. A Facebook application is created that requests user approval to extract their personal activity data hence, richer data for research purposes. In this case, however, to encourage the user to provide access to his data, the application needs to provide some feedback to the user [8].

Although creating Facebook applications for research purposes is becoming a trend among the HCI community [8, 9, 14], here we will focus only on the most relevant approaches to ours. The pioneers of this methodology were the MyPersonality team [9]. The project runs since 2007 (latest reports refer to 7.5 million users to have accessed the MyPersonality Facebook application). The purpose was to extract several Facebook activity and demographic features and correlate patterns of behaviour to personality traits. What the user was getting back through the MyPersonality app was their scores on the psychometric tests they took and nothing related to his Facebook activity.

In [8], classification trees employed in predicting Alternative Five Model personality features. Users required to answer the ZKPQ-60-cc personality questionnaire and in return, the application presented the users with (i) the results of their personality test, (ii) information on similar users who have used the application, (iii) the choice to compare their results with the results of their friends (if they have completed the test) and (iv) the ability to invite their friends to use the application. Although the results show an accurate prediction of 70 % for all traits, the user data extracted were limited to the number of posts in a user's wall, the number of user's friends and the number of months the user used Facebook.

In contrast to previous work, in this paper we are exploring a different approach of exploiting user data extracted from Facebook. Our aim in this work is not primarily to explore the accuracy that can be achieved in predicting personality traits from Facebook activity data (these has been done already in previous work), but to introduce a different approach on how the collected data can be exploited and to also be presented as useful information to the user in real time - all of the approaches mentioned above analyzed the collected data off-line.

4 Computational Framework

A computational framework has been developed following the general framework of adaptive systems proposed by Jameson at [15] and consists of two phases: Data Extraction and Processing, for building the User Model; and User Model Application, for extracting the personality model.

4.1 Facebook Data Extraction for User Modeling

The extraction of Facebook activity data has been done using a Facebook application (PersonaWeb app), which allowed us to get users' permissions for accessing their personal data as input to the framework. The data extracted include publically available information about a user and also private activity data (e.g. friends of a user, posts liked, shares, types of posts liked, checkins, checkins that a user was tagged in, events attended and created etc.).

Additional features have been defined by the authors (e.g. active friends), that can be considered to be a list of friends of a user with whom the user 'regularly' interacts with. In order for a user to be considered as an active friend of a given user, he/she had

to publish at least four posts directly on that user's wall, or appear in a Facebook activity together, during a period of a year. The reason for the four posts threshold is for excluding birthday and name-day wishes.

In every user model we keep a vector (*nuv*) that consists of arithmetic normalized values of the aggregated data collected based on thresholds defined. The thresholds' values defined based on a sample of Facebook users who participated in the study presented in a following section and thus excluded from the overall evaluation sample. After considering reports of Facebook user activity, we defined the activity of 'light' to 'heavy' Facebook usage for each element in Table 1. Elements in *nuv* can take values from 1 to 5, to simulate the scores of answers in a "Big Five" personality questionnaire. This is used in the extraction, of the user personality model, and the similarities between users in our system.

Table 1. Vector *nuv* consists of aggregated arithmetic values of user activity on Facebook

Values in <i>nuv</i>
of likes
of pages a user is following
of friends
of active friends
of self tags created by a user
of tags created by other users for a specific user
of events attended
of events created
of checkins created by others and mention a user
of checkins created by a user
of status posts made
of links posted by a user
percentage of user's profile photos with respect to all the photos uploaded
percentage of user's favourite page type with respect to all pages the user likes

4.2 Deriving the User Personality Model

The most important application of the user model in this work is the extraction of the personality of a user. Based on studies mentioned on previous work [1–3, 6–9], we identified Facebook activity that relates positively or negatively and with varied importance to each personality trait (Table 2). In addition, to the positive or negative relevance of an activity to a personality trait, we assigned weights of importance that an activity has, to a personality trait, and can take values from 0 to 1. The process of defining the weights is an initial attempt to experiment with this concept and thus, have been defined based on reports in related work on the importance of Facebook activities for a personality trait [1, 6, 7, 10].

The calculation of the value of each personality trait for a user a is done using Eq. 1 (if activity is positively related) and Eq. 2 (if activity is negatively related). In Eqs. 1 and 2, where ptv_a can be any of the five personality traits as defined in the “Big Five” model; act_weight is the weight value assigned to an activity (e.g. like, check-in, share); $activity_{ia}$ is the aggregated value of an activity in Table 2 as stored in nuv , for a user.

$$ptv_a = ptv_a + (act_weight * activity_{ia}) \tag{1}$$

$$ptv_a = ptv_a + (act_weight * (5 - activity_{ia} + 1)) \tag{2}$$

The extracted personality model for each user is presented to him/her through the PersonaWeb Facebook app, in a graphical way (Fig. 3) along with the results of the Big Five personality test they took.

Table 2. Facebook activity that relates positively or negatively to each personality trait

Personality trait	Positive relevance	Negative relevance
Extraversion	Likes, friends, status, user_checkins, user_events, self_tagprofile_pic	
Agreeableness	Tags, links	Likes, checkins, active_fr
Conscientiousness	page_type	active_fr, status, likes, events, tags, links, pages
Neuroticism	Likes, pages	Friends, profile_pic
Openness	Status, likes, events, pages	page_type

5 PersonaWeb Facebook Application

The purpose of developing the PersonaWeb Facebook App is twofold. Firstly, to be able to get user permission to extract his activity data; secondly, the application allowed us to visualize information kept in the user model and information regarding user results on the personality questionnaire. Initially the user is logging in to the application for the first time and a dialogue box is presented to him asking for his permission to release his data to the application.

If the user clicks accept then a pop-up window appears prompting the user to ‘share’ his derived personality model on Facebook. At the right-top corner the user can find a button that leads him to the PersonaWeb project web site, where he can find a “Big Five” personality questionnaire. A graphical representation of his Like activity on the social network (Fig. 2) follows. A second graph gives information on pages and their type that the user liked (Fig. 1). Furthermore, the user is becoming aware of (i) the person/page that the most posts he liked come from, (ii) his friend who tagged him most in posts and (iii) the four most recent events he attended.

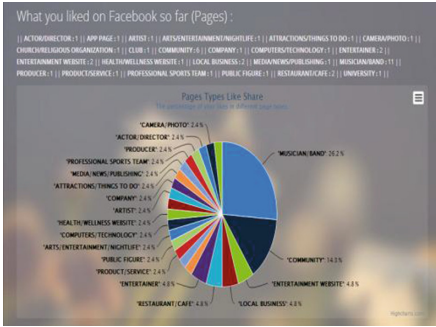


Fig. 1. Graphical statistical analysis of the types of pages a user likes

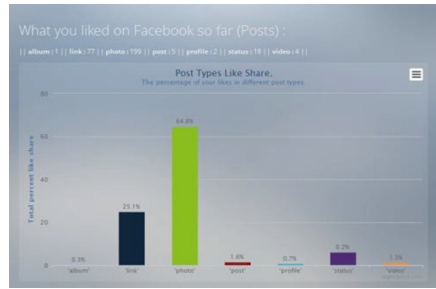


Fig. 2. Summary of user like activity on different Facebook categories

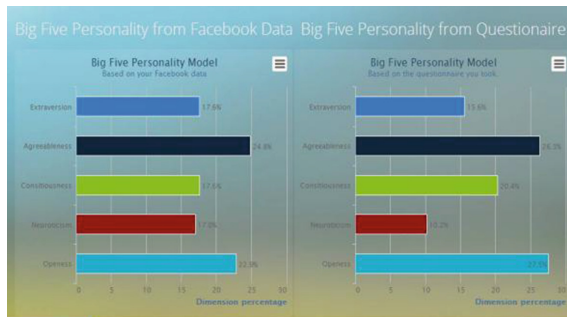


Fig. 3. Personality models of the user as derived from: the questionnaire (right); and Facebook based on activity data (Left)

The second part of the application provides a graphical representation of the user's personality model. It demonstrates the percentage scores of the user on each personality trait. These percentages reflect the user's personality based on his activity and interaction with other users in the social network as discussed above. If the user has completed the personality questionnaire, he can also see a second personality graph based on the results of the questionnaire; in that case he can visually compare the two graphs (see Fig. 3). In different case a message appears followed by a link prompting the user to complete the questionnaire. Below the two personality graphs the user can find an explanation of each Big Five personality trait.

6 Evaluation Study

6.1 Sampling and Procedure

The methodology followed was to perform an initial evaluation study using real Facebook users, who will be willing to fill in the Big Five personality test online and

release their data for us to use. Thus, a call for participation was distributed on mailing lists for recruiting volunteers that have active Facebook accounts. This approach allowed us to pull Facebook users of different ages and demographic orientation. The message that was sent, was explaining the reason for the study and the steps that the users had to follow. An additional method for attracting participants was the friend-of-a-friend approach where users shared their personality model results on their Facebook wall through the PersonaWeb app and as a result their friends became aware of the app and joined the study. We allowed eight days for people to access the PersonaWeb app and to complete the “Big Five” personality questionnaire.

Our dataset consisted of 62 active users of Facebook, 38 men and 24 women, with ages between 17 and 59, and average age of 32.08 (Std. Dev = 9.903). Participants were asked to click on a link to the PersonaWeb app, and provide consent for us to extract their activity data. The users were explicitly asked to click on the link provided to complete the “Big Five” personality questionnaire. After the completion of the questionnaire the user was redirected back to the app where he/she could compare both personality models (Fig. 3).

6.2 Data Analysis and Results

The main purpose of this evaluation study was to examine (i) whether the information extracted in the user model reflected the users’ activities; and (ii) whether we can utilize results reported in existing literature that correlated personality traits and Facebook activity to predict personality traits in real time.

To approach the first point of this study, we requested feedback from the participants in the form of casual written conversation [16]. We contacted participants through email requesting their comments on whether the information they received through the PersonaWeb Facebook app (e.g. likes, page types, most tagged from, most posts you like come from, events attended, recommendations) were representative of their actions in the social network and represented the current situation of their Facebook usage and their opinion on the recommendation they received. This technique of validating a user model is in line with methods followed in evaluating user modeling and adaptive systems.

25 out of 62 participants replied to our request. The general comments we had were mostly in favor of the information users received but we collected also some constructive criticism as well. Users thought that our system extracted very accurately the “image” of their ‘like’ activity and they thought that what was presented to them as a decomposition of the ‘page types’ they are following was indeed and exact useful. Some comments focused in the fact that there was also reflection of their past activity in the extracted information and that this may interfere with inferences done based on the user model. Furthermore, with respect to the last three events attended, the users thought that this was just a reminder of events they attended long time ago and not a feature that added to their experience using this application. On the other hand 5 users reported that this was not a useful feature to have and it should be eliminated. The users appreciated the ‘similar users’ information communicated to them and mentioned that

they were curious to further explore their similarities to these people, especially since the profile pictures and user names provided were clickable links to that user's profile.

With respect to the second goal of this evaluation, similarly to previous studies mentioned, the results of the "Big Five" personality questionnaire that each user completed were used to compare the two personality models obtained for each user. A Pearson's product-moment correlation employed.

According to the results of the correlation analysis (Table 3), it appears to be a weak positive correlation for the *Extraversion* trait between the two personality models with $r = 0.259$ at 0.05 significance level. The results in *Agreeableness* trait show a positive but not significant correlation between the two models $r = 0.032$. Similar to extraversion, *Conscientiousness* trait appears to give a weak positive correlation between the two models with $r = 0.281$ at the 0.05 significance level. In contrast *Neuroticism* and *Openness* appear to negatively correlate between the two models with $r = -0.010$ and $r = -0.161$ respectively.

Table 3. Pearson correlation results between the Big Five personality models extracted based on the questionnaire and the computational model introduced.

Big Five personality trait	Pearson correlation results
Extraversion	0.259
Agreeableness	0.032
Conscientiousness	0.281
Openness	-0.161
Neuroticism	-0.010

The initial results show that the activities considered in our work as important for modeling the personality of the users for the traits of Extraversion, Agreeableness and Conscientiousness reflect in a minimal extend the personality of a Facebook user in our sample. In the case of Openness and Neuroticism the results show that the activities employed are not sufficient and further refinement of the model is needed. These results are not surprising to us since research particularly on social sciences and psychology still report contradictory results on how and which Facebook activities are important for each personality trait in the Big Five model e.g. [1, 10, 14]. Our model is strongly depended on reports of previous research; consequently, our results reflect this contradiction and call for a more refined model definition.

Given the above results we further explored the correlation of activity features we extracted to the personality traits of *Neuroticism* and *Openness*, in an attempt to explain the negative correlations. Initial results show strong positive correlation of Neuroticism with the number of *active friends* a user has and also with the number of *checkins a user is tagged in*. Additionally, consistent with [14] and inconsistent with [10] Neuroticism appears to correlate positively with *like and share* activity on Facebook. A negative relation appears with the number of *events* the user attended. For openness the number of *links* a user shared on Facebook appears negatively related with this trait. This information had not reported on existing literature and hence was not included in

our initial personality model. We are currently working on this study to correlate Facebook activity features to other personality traits for improving our models.

7 Conclusion

The motivation behind this work was to implicitly extract a user personality model based on information reported on previous work on activity correlation to personality traits. Compared to previous work, we have used much richer private user data to build a user's personality model and we exploited previous and existing literature from psychology and computing in an innovative way (e.g. real-time computation of the user personality model and instant visualization of the personality prediction results to the user through the Facebook app). However, at this stage we cannot claim statistical comparison of our results with studies in the area of machine learning e.g. [3, 7, 8] primarily due to the limited size of users participated in our evaluation study and due to the work-in-progress state of our work.

In addition to correlating Facebook activities to personality traits we are looking into exploring further the potentials of our approach. For example the assignment of weights of importance to Facebook activities needs to be explored further through several studies in order to see how this can affect the accuracy of predicting the personality model of a user. Although the idea of defining active friends in the user model is in-line with theory and was similar to previous work [8], we believe there is more to 'who can be considered as an active friend' e.g. (textual analysis). Finally, a larger study with more users will provide more accurate results and outlook on the benefits of this approach.

References

1. Amichai-Hamburger, Y., Vinitzky, G.: Social network use and personality. *Comp. Hum. Behav.* **26**(6), 1289–1295 (2010)
2. Caci, B., Cardaci, M., Tabacchi, M.E., Scrima, F.: Personality variables as predictors of Facebook usage. *Psychol. Rep.* **114**(2), 528–539 (2014)
3. Golbeck, J., Robles, C., Turner, K.: Predicting personality with social media. In: Proceedings of CHI 2011 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2011), pp. 253–262 (2011). <http://dx.doi.org/10.1145/1979742.1979614>
4. McElroy, J.C., Hendrickson, A.R., Townsend, A.M., DeMarie, S.M.: Dispositional factors in internet use: personality versus cognitive style. *MIS Q.* **31**(4), 809–820 (2007)
5. Goldberg, L.R.: An alternative description of personality: the Big-Five factor structure. *J. Pers. Soc. Psychol.* **59**(6), 1216–1229 (1990)
6. Moore, K., McElroy, J.C.: The influence of personality on Facebook usage, wall postings, and regret. *Comput. Hum. Behav.* **28**(1), 267–274 (2012)
7. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., Stillwell, D.: Personality and patterns of Facebook usage. In: Proceedings of the 4th Annual ACM Web Science Conference (WebSci 2012), pp. 24–32 (2012). <http://dx.doi.org/10.1145/2380718.2380722>
8. Ortigosa, A., Carro, R.M., Quiroga, J.I.: Predicting user personality by mining social interactions in Facebook. *J. Comput. Syst. Sci.* **80**(1), 57–71 (2014)

9. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. (PNAS)* **110**(15), 5802–5805 (2013)
10. Ross, C., Orr, E.S., Sisc, M., Arseneault, J.M., Simmering, M.G., Robert Orr, R.: Personality and motivations associated with Facebook use. *Comput. Hum. Behav.* **25**(2), 578–586 (2009)
11. Seidman, G.: Self-presentation and belonging on Facebook: how personality influences social media use and motivations. *Pers. Individ. Differ.* **54**(3), 402–407 (2013)
12. Markovikj, D., Gievaska, S., Kosinski, M., Stillwell, D.: Mining Facebook data for predictive personality modeling. In: *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM 2013)*, Boston, MA, USA (2013)
13. Wald, R., Khoshgoftaar, T., Sumner, C.: Machine prediction of personality from facebook profiles. In: *2012 IEEE 13th International Conference on Information Reuse & Integration (Iri)* (2012)
14. Shen, J., Brdiczka, O., Liu, J.: A study of Facebook behavior: what does it tell about your neuroticism and extraversion? *Comput. Hum. Behav.* **45**, 32–38 (2015)
15. Jameson, A.: Adaptive interfaces and agents. In: Jacko, J.A., Sears, A. (eds.) *Human Computer Interaction Handbook*, pp. 105–127. CRC Press, Taylor & Francis Group, Boca Raton, NY (2003)
16. Heritage, J., Atkinson, J.M.: *Structures of Social Action: Studies in Conversation Analysis*. Cambridge University Press, Cambridge (1984)