Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski[a,1], David Stillwell[a], and Thore Graepel[b]

[a]Free School Lane, The Psychometrics Centre, University of Cambridge, Cambridge CB2 3RQ United Kingdom; and [b]Microsoft Research, Cambridge CB1 2FB, United Kingdom

Edited by Kenneth Wachter, University of California, Berkeley, CA, and approved February 12, 2013 (received for review October 29, 2012)

We show that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. The analysis presented is based on a dataset of over 58,000 volunteers who provided their Facebook Likes, detailed demographic profiles, and the results of several psychometric tests. The proposed model uses dimensionality reduction for preprocessing the Likes data, which are then entered into logistic/linear regression to predict individual psychodemographic profiles from Likes. The model correctly discriminates between homosexual and heterosexual men in 88% of cases, African Americans and Caucasian Americans in 95% of cases, and between Democrat and Republican in 85% of cases. For the personality trait "Openness," prediction accuracy is close to the test–retest accuracy of a standard personality test. We give examples of associations between attributes and Likes and discuss implications for online personalization and privacy.

social networks | computational social science | machine learning | big data | data mining | psychological assessment

A growing proportion of human activities, such as social interactions, entertainment, shopping, and gathering information, are now mediated by digital services and devices. Such digitally mediated behaviors can easily be recorded and analyzed, fueling the emergence of computational social science (1) and new services such as personalized search engines, recommender systems (2), and targeted online marketing (3). However, the widespread availability of extensive records of individual behavior, together with the desire to learn more about customers and citizens, presents serious challenges related to privacy and data ownership (4, 5).

We distinguish between data that are actually recorded and information that can be statistically predicted from such records. People may choose not to reveal certain pieces of information about their lives, such as their sexual orientation or age, and yet this information might be predicted in a statistical sense from other aspects of their lives that they do reveal. For example, a major US retail network used customer shopping records to predict pregnancies of its female customers and send them well-timed and well-targeted offers (6). In some contexts, an unexpected flood of vouchers for prenatal vitamins and maternity clothing may be welcome, but it could also lead to a tragic outcome, e.g., by revealing (or incorrectly suggesting) a pregnancy of an unmarried woman to her family in a culture where this is unacceptable (7). As this example shows, predicting personal information to improve products, services, and targeting can also lead to dangerous invasions of privacy.

Predicting individual traits and attributes based on various cues, such as samples of written text (8), answers to a psychometric test (9), or the appearance of spaces people inhabit (10), has a long history. Human migration to digital environment renders it possible to base such predictions on digital records of human behavior. It has been shown that age, gender, occupation, education level, and even personality can be predicted from people's Web site browsing logs (11–15). Similarly, it has been shown that personality can be predicted based on the contents of personal Web sites (16), music collections (17), properties of Facebook or Twitter profiles such as the number of friends or the density of friendship networks (18–21), or language used by their users (22). Furthermore, location within a friendship network at Facebook was shown to be predictive of sexual orientation (23).

This study demonstrates the degree to which relatively basic digital records of human behavior can be used to automatically and accurately estimate a wide range of personal attributes that people would typically assume to be private. The study is based on Facebook Likes, a mechanism used by Facebook users to express their positive association with (or "Like") online content, such as photos, friends' status updates, Facebook pages of products, sports, musicians, books, restaurants, or popular Web sites. Likes represent a very generic class of digital records, similar to Web search queries, Web browsing histories, and credit card purchases. For example, observing users' Likes related to music provides similar information to observing records of songs listened to online, songs and artists searched for using a Web search engine, or subscriptions to related Twitter channels. In contrast to these other sources of information, Facebook Likes are unusual in that they are currently publicly available by default. However, those other digital records are still available to numerous parties (e.g., governments, developers of Web browsers, search engines, or Facebook applications), and, hence, similar predictions are unlikely to be limited to the Facebook environment.

The design of the study is presented in Fig. 1. We selected traits and attributes that reveal how accurate and potentially intrusive such a predictive analysis can be, including "sexual orientation," "ethnic origin," "political views," "religion," "personality," "intelligence," "satisfaction with life" (SWL), substance use ("alcohol," "drugs," "cigarettes"), "whether an individual's parents stayed together until the individual was 21 y old," and basic demographic attributes such as "age," "gender," "relationship status," and "size and density of the friendship network." Five Factor Model (9) personality scores (n = 54,373) were established using the International Personality Item Pool (IPIP) questionnaire with 20 items (25). Intelligence (n = 1,350) was measured using Raven's Standard Progressive Matrices (SPM) (26), and SWL (n = 2,340) was measured using the SWL Scale (27). Age (n = 52,700; average, μ = 25.6; SD = 10), gender (n = 57,505; 62% female), relationship status ("single"/"in relationship"; n = 46,027; 49% single), political views ("Liberal"/"Conservative"; n = 9,752;

Author contributions: M.K. and T.G. designed research; M.K. and D.S. performed research; M.K. and T.G. analyzed data; and M.K., D.S., and T.G. wrote the paper.

Conflict of interest statement: D.S. received revenue as owner of the myPersonality Facebook application.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the myPersonality Project database (www.mypersonality.org/wiki).

[1]To whom correspondence should be addressed. E-mail: mk583@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1218772110/-/DCSupplemental.

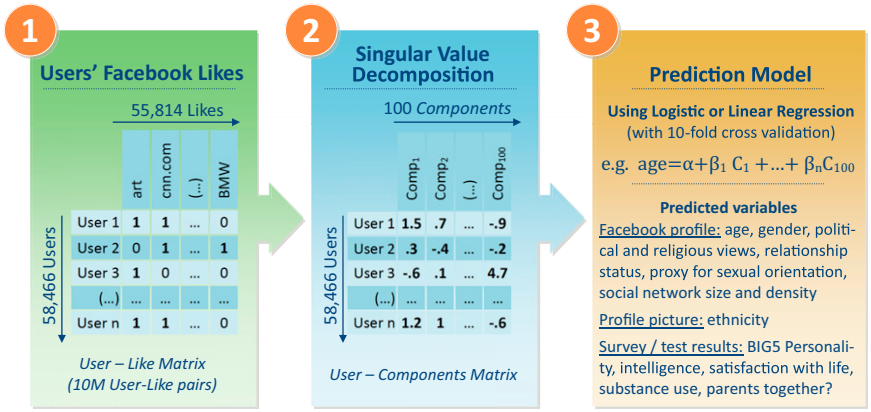

Fig. 1. The study is based on a sample of 58,466 volunteers from the United States, obtained through the myPersonality Facebook application (www.mypersonality.org/wiki), which included their Facebook profile information, a list of their Likes (n = 170 Likes per person on average), psychometric test scores, and survey information. Users and their Likes were represented as a sparse user–Like matrix, the entries of which were set to 1 if there existed an association between a user and a Like and 0 otherwise. The dimensionality of the user–Like matrix was reduced using singular-value decomposition (SVD) (24). Numeric variables such as age or intelligence were predicted using a linear regression model, whereas dichotomous variables such as gender or sexual orientation were predicted using logistic regression. In both cases, we applied 10-fold cross-validation and used the k = 100 top SVD components. For sexual orientation, parents' relationship status, and drug consumption only k = 30 top SVD components were used because of the smaller number of users for which this information was available.

65% Liberal), religion ("Muslim"/"Christian"; n = 18,833; 90% Christian), and the Facebook social network information [n = 17,601; median size, $\tilde{X}$ = 204; interquartile range (IQR), 206; median density, $\tilde{X}$ = 0.03; IQR, 0.03] were obtained from users' Facebook profiles. Users' consumption of alcohol (n = 1,196; 50% drink), drugs (n = 856; 21% take drugs), and cigarettes (n = 1211; 30% smoke) and whether a user's parents stayed together until the user was 21 y old (n = 766; 56% stayed together) were recorded using online surveys. Visual inspection of profile pictures was used to assign ethnic origin to a randomly selected subsample of users (n = 7,000; 73% Caucasian; 14% African American; 13% others). Sexual orientation was assigned using the Facebook profile "Interested in" field; users interested only in others of the same sex were labeled as homosexual (4.3% males; 2.4% females), whereas those interested in users of the opposite gender were labeled as heterosexual.

Results

Prediction of Dichotomous Variables. Fig. 2 shows the prediction accuracy of dichotomous variables expressed in terms of the area under the receiver-operating characteristic curve (AUC), which is equivalent to the probability of correctly classifying two randomly selected users one from each class (e.g., male and female). The highest accuracy was achieved for ethnic origin and gender. African Americans and Caucasian Americans were correctly classified in 95% of cases, and males and females were correctly classified in 93% of cases, suggesting that patterns of online behavior as expressed by Likes significantly differ between those groups allowing for nearly perfect classification.

Christians and Muslims were correctly classified in 82% of cases, and similar results were achieved for Democrats and Republicans (85%). Sexual orientation was easier to distinguish among males (88%) than females (75%), which may suggest a wider behavioral divide (as observed from online behavior) between hetero- and homosexual males.

Good prediction accuracy was achieved for relationship status and substance use (between 65% and 73%). The relatively lower accuracy for relationship status may be explained by its temporal variability compared with other dichotomous variables (e.g., gender or sexual orientation).

The model's accuracy was lowest (60%) when inferring whether users' parents stayed together or separated before users were 21 y old. Although it is known that parental divorce does have long-term effects on young adults' well-being (28), it is remarkable that this is detectable through their Facebook Likes. Individuals with parents who separated have a higher probability of liking statements preoccupied with relationships, such as "If I'm with you then I'm with you I don't want anybody else" (Table S1).

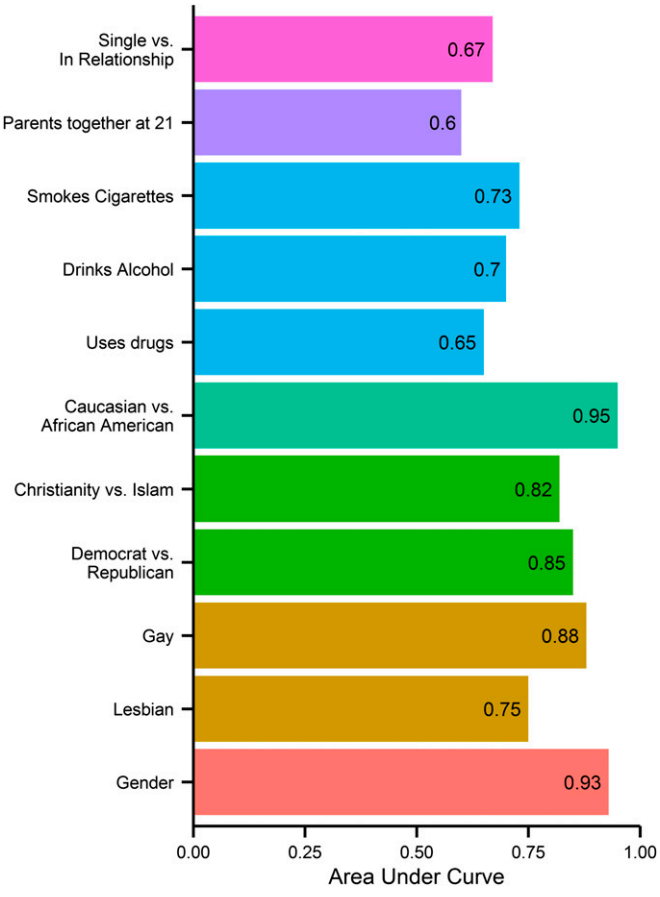

Fig. 2. Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.

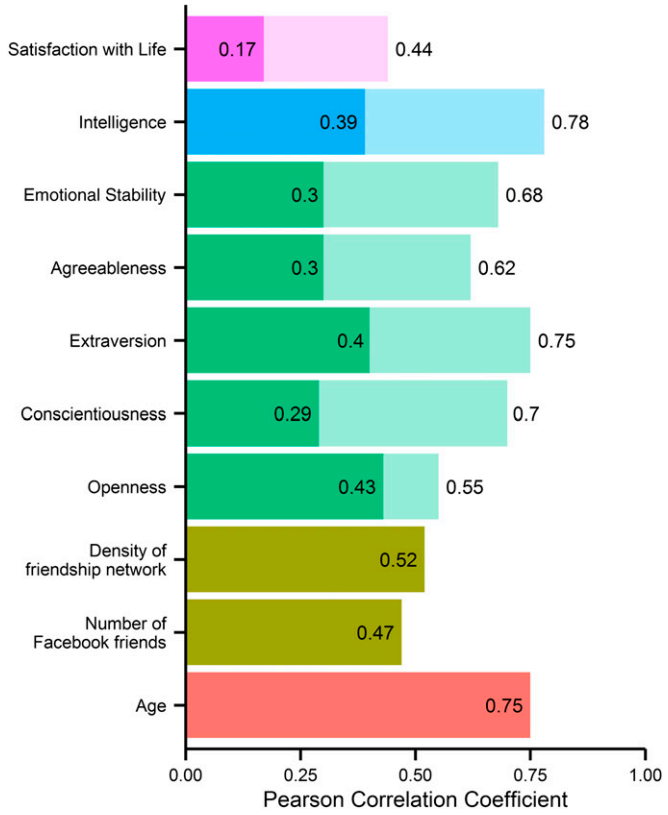

Fig. 3. Prediction accuracy of regression for numeric attributes and traits expressed by the Pearson correlation coefficient between predicted and actual attribute values; all correlations are significant at the $P < 0.001$ level. The transparent bars indicate the questionnaire's baseline accuracy, expressed in terms of test–retest reliability.

Prediction of Numeric Variables. Fig. 3 presents the accuracy of predicting numeric variables as expressed by the Pearson product–moment correlation coefficient between the actual and predicted values. The highest correlation was obtained for age ($r = 0.75$), followed by density ($r = 0.52$) and size ($r = 0.47$) of the Facebook friendship network. Closely following were the personality traits of "Openness" ($r = 0.43$), "Extraversion" ($r = 0.40$), and "Intelligence" ($r = 0.39$). The remaining personality traits and SWL were predicted with somewhat lower accuracy ($r = 0.17$ to 0.30).

Psychological traits are examples of latent traits (i.e., traits that cannot be measured directly). As a consequence, their values can only be measured approximately, for example, by evaluating responses to questionnaires. The transparent bars presented in Fig. 3 indicate the accuracy of the questionnaires used as expressed by their test-retest reliabilities (Pearson product–moment correlation between the questionnaire scores obtained by the same respondent at two points in time). The correlation between the predicted and actual Openness score ($r = 0.43$) was very close to the test–retest reliability for Openness ($r = 0.50$). This indicates that for the Openness trait, observation of the user's Likes is roughly as informative as using their personality test score itself. For the remaining traits, prediction accuracies correspond to roughly half the questionnaire's test-retest reliabilities.

The relatively lower prediction accuracy for SWL ($r = 0.17$) may be attributable to the difficulty of separating long-term happiness (29) from mood swings, which vary over time. Thus, although the SWL score includes variability attributable to mood, users' Likes accrue over a longer period and, so, may be suitable only for predicting long-term happiness.

Amount of Data Available and Prediction Accuracy. The results presented so far rely on individuals for which between one and 700 Likes were available. The median number of Likes was 68 per individual (IQR, 152). Therefore, what is the expected accuracy given a random individual and how does prediction accuracy change with the number of observed Likes? Using a subsample ($n = 500$) of users for whom at least 300 Likes were available, we ran predictive models based on randomly selected subsets of $n = 1, 2, \ldots, 300$ Likes. The results presented in Fig. 4 show that even knowing a single random Like for a given user can result in nonnegligible prediction accuracy. Knowing further Likes increases the accuracy but with diminishing returns from each additional piece of information.

Predictive Power of Likes. Individual traits and attributes can be predicted to a high degree of accuracy based on records of users' Likes. Table S1 presents a sample of highly predictive Likes related to each of the attributes. For example, the best predictors of high intelligence include "Thunderstorms," "The Colbert Report," "Science," and "Curly Fries," whereas low intelligence was indicated by "Sephora," "I Love Being A Mom," "Harley Davidson," and "Lady Antebellum." Good predictors of male homosexuality included "No H8 Campaign," "Mac Cosmetics," and "Wicked The Musical," whereas strong predictors of male heterosexuality included "Wu-Tang Clan," "Shaq," and "Being Confused After Waking Up From Naps." Although some of the Likes clearly relate to their predicted attribute, as in the case of No H8 Campaign and homosexuality, other pairs are more elusive; there is no obvious connection between Curly Fries and high intelligence.

Moreover, note that few users were associated with Likes explicitly revealing their attributes. For example, less than 5% of users labeled as gay were connected with explicitly gay groups, such as No H8 Campaign, "Being Gay," "Gay Marriage," "I love Being

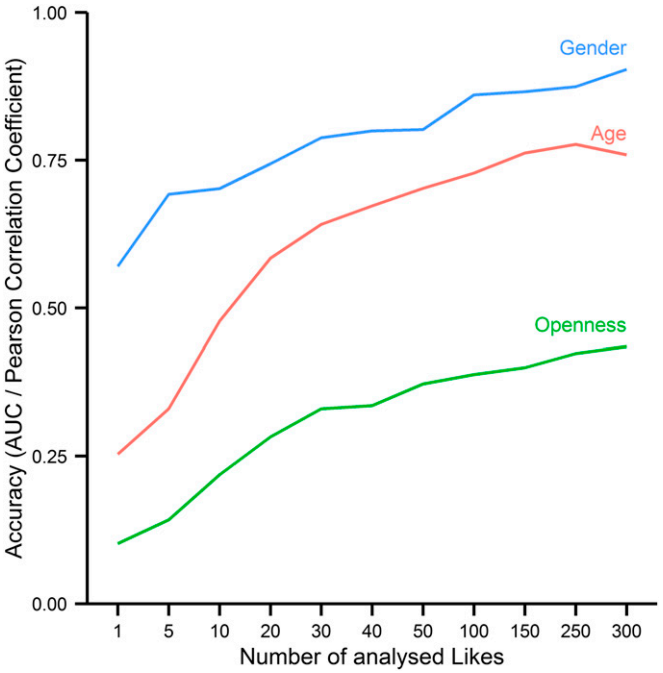

Fig. 4. Accuracy of selected predictions as a function of the number of available Likes. Accuracy is expressed as AUC (gender) and Pearson's correlation coefficient (age and Openness). About 50% of users in this sample had at least 100 Likes and about 20% had at least 250 Likes. Note, that for gender (dichotomous variable) the random guessing baseline corresponds to an AUC = 0.50.

Gay," "We Didn't Choose To Be Gay We Were Chosen." Consequently, predictions rely on less informative but more popular Likes, such as "Britney Spears" or "Desperate Housewives" (both moderately indicative of being gay).

This is further illustrated in Fig. S1, which shows the average levels of personality traits and age for several popular Likes. Each Like attracts users with a different average personality and demographic profile and, thus, can be used to predict those attributes. For example, users who liked the "Hello Kitty" brand tended to be high on Openness and low on "Conscientiousness," "Agreeableness," and "Emotional Stability." They were also more likely to have Democratic political views and to be of African-American origin, predominantly Christian, and slightly below average age. The same Likes were used to create Fig. S2, presenting their relative popularity in four groups: Democrats, Christians, Homosexuals, and African-American individuals. For example, although liking "Barack Obama" is clearly related to being a Democrat, it is also relatively popular among Christians, African Americans, and Homosexual individuals.

Conclusions

We show that a wide variety of people's personal attributes, ranging from sexual orientation to intelligence, can be automatically and accurately inferred using their Facebook Likes. Similarity between Facebook Likes and other widespread kinds of digital records, such as browsing histories, search queries, or purchase histories suggests that the potential to reveal users' attributes is unlikely to be limited to Likes. Moreover, the wide variety of attributes predicted in this study indicates that, given appropriate training data, it may be possible to reveal other attributes as well.

Predicting users' individual attributes and preferences can be used to improve numerous products and services. For instance, digital systems and devices (such as online stores or cars) could be designed to adjust their behavior to best fit each user's inferred profile (30). Also, the relevance of marketing and product recommendations could be improved by adding psychological dimensions to current user models. For example, online insurance advertisements might emphasize security when facing emotionally unstable (neurotic) users but stress potential threats when dealing with emotionally stable ones. Moreover, digital records of behavior may provide a convenient and reliable way to measure psychological traits. Automated assessment based on large samples of behavior may not only be more accurate and less prone to cheating and misrepresentation but may also permit assessment across time to detect trends. Moreover, inference based on observations of digitally recorded behavior may open new doors for research in human psychology.

On the other hand, the predictability of individual attributes from digital records of behavior may have considerable negative implications, because it can easily be applied to large numbers of people without obtaining their individual consent and without them noticing. Commercial companies, governmental institutions, or even one's Facebook friends could use software to infer attributes such as intelligence, sexual orientation, or political views that an individual may not have intended to share. One can imagine situations in which such predictions, even if incorrect, could pose a threat to an individual's well-being, freedom, or even life. Importantly, given the ever-increasing amount of digital traces people leave behind, it becomes difficult for individuals to control which of their attributes are being revealed. For example, merely avoiding explicitly homosexual content may be insufficient to prevent others from discovering one's sexual orientation.

There is a risk that the growing awareness of digital exposure may negatively affect people's experience of digital technologies, decrease their trust in online services, or even completely deter them from using digital technology. It is our hope, however, that the trust and goodwill among parties interacting in the digital environment can be maintained by providing users with transparency and control over their information, leading to an individually controlled balance between the promises and perils of the Digital Age.

ACKNOWLEDGMENTS. We thank Yoram Bachrach, Alan Blackwell, George Danezis, Stephen Emmott, David Good, Peter Key, Emre Kiciman, Pushmeet Kohli, Drew Purves, Jason Rentfrow, John Rust, and Duncan Watts for discussions about the topic of this study, as well as for comments on the manuscript. Demonstration of personality prediction based on individuals' Likes is available at http://www.youarewhatyoulike.com. M.K. received funding from Boeing Corporation and from Microsoft Research.

1. Lazer D, et al. (2009) Computational social science. *Science* 323(5915):721–723.
2. Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.
3. Chen Y, Pavlov D, Canny JF (2009) Large-scale behavioral targeting. *International Conference on Knowledge Discovery and Data Mining*, pp 209–218.
4. Butler D (2007) Data sharing threatens privacy. *Nature* 449(7163):644–645.
5. Narayanan A, Shmatikov V (2008) Robust de-anonymization of large sparse datasets. *IEEE Symposium on Security and Privacy*, pp 111–125.
6. Duhigg C (2012) *The Power of Habit: Why We Do What We Do in Life and Business* (Random House, New York).
7. İnce HO, Yarali A, Özsel D (2009) Customary killings in Turkey and Turkish modernization. *Middle East Stud* 45(4):537–551.
8. Fast LA, Funder DC (2008) Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *J Pers Soc Psychol* 94(2):334–346.
9. Costa PT, McCrae RR (1992) *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual* (Psychological Assessment Resources, Odessa, FL).
10. Gosling SD, Ko SJ, Mannarelli T, Morris ME (2002) A room with a cue: Personality judgments based on offices and bedrooms. *J Pers Soc Psychol* 82(3):379–398.
11. Hu J, Zeng H-J, Li H, Niu C, Chen Z (2007) Demographic prediction based on user's browsing behavior. *International World Wide Web Conference*, pp 151–160.
12. Murray D, Durrell K (1999) Inferring demographic attributes of anonymous Internet users. *Revised Papers from the International Workshop on Web Usage Analysis and User Profiling*, eds Masand BM, Spiliopoulou M (Springer, London), pp 7–20.
13. 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.
14. Goel S, Hofman JM, Sirer MI (2012) Who does what on the Web: Studying Web browsing behavior at scale. *International Conference on Weblogs and Social Media*, pp 130–137.
15. Kosinski M, Kohli P, Stillwell DJ, Bachrach Y, Graepel T (2012) Personality and website choice. *ACM Web Science Conference*, pp 251–254.
16. Marcus B, Machilek F, Schütz A (2006) Personality in cyberspace: Personal Web sites as media for personality expressions and impressions. *J Pers Soc Psychol* 90(6):1014–1031.
17. Rentfrow PJ, Gosling SD (2003) The do re mi's of everyday life: The structure and personality correlates of music preferences. *J Pers Soc Psychol* 84(6):1236–1256.
18. Quercia D, Lambiotte R, Kosinski M, Stillwell D, Crowcroft J (2012) The Personality of popular Facebook users. *ACM Conference on Computer Supported Cooperative Work*. Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp 955–964.
19. Bachrach Y, Kohli P, Graepel T, Stillwell DJ, Kosinski M (2012) Personality and patterns of Facebook usage. *ACM Web Science Conference*. Proceedings of the ACM Web Science Conference, pp 36–44.
20. Quercia D, Kosinski M, Stillwell DJ, Crowcroft J (2011) Our Twitter profiles, our selves: Predicting personality with Twitter. *IEEE International Conference on Social Computing*. Proceedings of the 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, pp 180–185.
21. Golbeck J, Robles C, Edmondson M, Turner K (2011) Predicting personality from Twitter. *IEEE International Conference on Social Computing*, pp 149–156.
22. Golbeck J, Robles C, Turner K (2011) Predicting personality with social media. *Conference on Human Factors in Computing Systems*, pp 253–262.
23. Jernigan C, Mistree BF (2009) Gaydar: Facebook friendships expose sexual orientation. *First Monday* 14(10).
24. Golub GH, Kahan W (1965) Calculating the singular values and pseudo-inverse of a matrix. *J Soc Ind Appl Math* 2(2):205–224.
25. Goldberg LR, et al. (2006) The international personality item pool and the future of public-domain personality measures. *J Res Pers* 40(1):84–96.
26. Raven JC (2000) The Raven's progressive matrices: Change and stability over culture and time. *Cognit Psychol* 41(1):1–48.
27. Diener E, Emmons RA, Larsen RJ, Griffin S (1985) The satisfaction with life scale. *J Pers Assess* 49(1):71–75.
28. Musick K, Meier A (2010) Are both parents always better than one? Parental conflict and young adult well-being. *Soc Sci Res* 39(5):814–830.
29. Schimmack U, Diener E, Oishi S (2002) Life-satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. *J Pers* 70(3):345–384.
30. Nass C, Lee KM (2000) Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *J Exp Psychol* 7(3):171–181.