Personality and Individual Differences 124 (2018) 150–159

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Review

Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis

Danny Azucar, Davide Marengo*, Michele Settanni

Department of Psychology, University of Turin, 10124, Via Verdi 10, Turin, Italy

ARTICLE INFO

Keywords:
Social media
Digital footprint
Big 5 traits
Personality
Data mining
Predictive modeling

ABSTRACT

The growing use of social media among Internet users produces a vast and new source of user generated ecological data, such as textual posts and images, which can be collected for research purposes. The increasing convergence between social and computer sciences has led researchers to develop automated methods to extract and analyze these digital footprints to predict personality traits. These social media-based predictions can then be used for a variety of purposes, including tailoring online services to improve user experience, enhance recommender systems, and as a possible screening and implementation tool for public health. In this paper, we conduct a series of meta-analyses to determine the predictive power of digital footprints collected from social media over Big 5 personality traits. Further, we investigate the impact of different types of digital footprints on prediction accuracy. Results of analyses show that the predictive power of digital footprints over personality traits is in line with the standard "correlational upper-limit" for behavior to predict personality, with correlations ranging from 0.29 (Agreeableness) to 0.40 (Extraversion). Overall, our findings indicate that accuracy of predictions is consistent across Big 5 traits, and that accuracy improves when analyses include demographics and multiple types of digital footprints.

1. Introduction

1.1. Social media and digital footprints

Social media and social network sites have become increasingly popular; currently about 2 billion people worldwide have a Facebook account, and over 1250 million users access Facebook on a daily basis (Statista, 2017). Similarly, Twitter averages about 328 million active users (Statista, 2017), with about 100 million daily users (Aslam, 2017). Social media has revolutionized how people interact with each other, is a virtually unavoidable avenue for social interactions, and a place where users present themselves to the world by creating an online profile. Every day, millions of people express their immediate thoughts, emotions, and beliefs by writing, posting, and sharing content on social media, which is then viewable by the user's online social network. Evidence also suggests that content generated and shared on social media user profiles represents an extension of "one's self" and reflects the actual personality of its individual users rather than project their most desirable traits (Back et al., 2010; Seidman, 2013). Consequently, the interactive nature of social media coupled with its ever-increasing utilization results in a naturally occurring, immense, ecologically valid dataset of online human activity, or *digital footprints*, consisting of information shared by users on their social media profiles - e.g., personal information about age, gender orientation, place of residence, as well shared texts, pictures, and videos (Madden, Fox, Smith, & Vitax, 2007). These digital footprints can be recorded, and have been previously analyzed by researchers from diverse disciplines, including computer science, public health, and social sciences (e.g., De Choudhury, Counts, & Horvitz, 2013; De Choudhury, Counts, Horvitz, & Hoff, 2014; Eichstaedt et al., 2015; Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011; Matz & Netzer, 2017; Padrez et al., 2015; Settanni & Marengo, 2015). In particular, the human migration to social media has steered psychologists toward studying existing relationships between digital footprints and psychological characteristics (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015). The emergence of, and access to, these large user data sets has reshaped the way social science researchers use content analysis to study psychological characteristics and has resulted in the convergence of social and computer sciences. This interdisciplinary work of social and computer sciences has allowed researchers to not only seek to *gain insights* from studying human behaviors on social media, but to also *predict* psychological characteristics and behaviors based on automated data mining and the analysis of digital footprints (Schwartz & Ungar, 2015).

* Corresponding author.
E-mail address: davide.marengo@unito.it (D. Marengo).

https://doi.org/10.1016/j.paid.2017.12.018
Received 1 September 2017; Received in revised form 9 December 2017; Accepted 11 December 2017
Available online 22 December 2017
0191-8869/

1.2. Personality prediction from social media

Personality has been regarded as one of the most important topics in psychological research (Li, Li, Hao, Guan, & Zhu, 2014; Ozer & Benet-Martinez, 2006). Research has shown that personality may be predictive of many aspects of life, including academic success (e.g., Komarraju, Karau, & Schmeck, 2009), job performance (e.g., Judge, Higgins, Thoresen, & Barrick, 1999; Neal, Yeo, Koy, & Xiao, 2012), social status (e.g., Anderson, John, Keltner, & Kring, 2001), health (e.g., Soldz & Vaillant, 1999), success in romantic relationships (e.g., Donnellan, Conger, & Bryant, 2004; Donnellan, Larsen-Rife, & Conger, 2005), political attitudes (e.g., Gerber, Huber, Doherty, Dowling, & Ha, 2010), subjective well-being (e.g., Hayes & Joseph, 2003), and online behaviors (e.g., Wang, 2013). While several models to describe personality exist, one of the most well researched, well regarded, and widely accepted theoretical frameworks of personality is the five-factor (or Big 5) model, comprised of openness to new experiences, conscientiousness, extraversion, agreeableness and neuroticism (McCrae & Costa, 1987; McCrae & John, 1992). Big 5 traits have been shown to be significantly associated with users' behaviors on social media. For example, individuals with high extraversion have been characterized by higher levels of activity on social media (e.g., Blackwell, Leaman, Tramposch, Osborne, & Liss, 2017; Kuss & Griffiths, 2011), and have a greater number of friends (Kosinski, Bachrach, Kohli, Stillwell, & Graepel, 2014) than introverted individuals. Individuals with high neuroticism are more prone to self-disclose hidden aspects of themselves, use social media as a passive way to learn about others (Seidman, 2013), and use more negative words in their posts, or 'status updates' (Schwartz et al., 2013). On the other hand, agreeable individuals tend to use fewer swear words and express positive emotions more frequently in their posts (Schwartz et al., 2013), and are more likely to post pictures expressing a positive mood (Liu, Preotiuc-Pietro, Samani, Moghaddam, & Ungar, 2016). Individuals with high conscientiousness appear to be cautious in managing their social media profiles; they tend to post fewer pictures (Amichai-Hamburger & Vinitzky, 2010), express less "Likes", and engage in less group activity on social media (Kosinski et al., 2014). Furthermore, individuals with high openness tend to have larger networks (Quercia, Lambiotte, Stillwell, Kosinski, & Crowcroft, 2012), and "Like" more content found on social media (Bachrach, Kosinski, Graepel, Kohli, & Stillwell, 2012) than individuals low on the trait. Driven by increasing evidence of the presence of links between personality and online behaviors, researchers have begun exploring the use of digital footprints left by people on social media to infer the Big 5 traits. Researchers in this field have generally employed a common research design consisting of, 1. The administration of self-report questionnaires to assess personality traits of social media users, 2. The collection of digital footprints from users' social media profiles, 3. The processing of these digital footprints to extract single or multiple features to be employed in predictive models, and 4. The evaluation of accuracy of personality predictions based on these features. However, studies vary in terms of type of digital footprints (e.g., text, pictures, Likes, user activity, which may be examined separately or in combination), and social media platforms (e.g., Facebook, Twitter, Instagram, Youtube) examined. For instance, Schwartz et al. (2013) investigated the feasibility of predicting personality traits based on textual features extracted from Facebook status updates using topic-modeling techniques. Similarly, Liu et al. (2016) and Qiu, Lin, Ramsay, and Yang (2012) both analyzed language/text used on Twitter to build predictive models for the Big 5 traits. While Gao et al. (2013), Li et al. (2014), and Wei et al. (2017) inferred the Big 5 traits using samples from the Sina Weibo micro blog albeit using different combinations of digital footprints (activity vs. activity + language vs. activity + language + pictures) in their analysis. Additionally, Kosinski, Stillwell, and Graepel (2013) and Youyou, Kosinski, and Stillwell (2015) explored Big 5 personality predictions based on Facebook Likes. Findings emerging from these studies are heterogeneous with respect to the accuracy of prediction for each personality trait. For instance, using "Likes" data extracted from Facebook, Kosinski et al. (2013) found prediction accuracy to vary significantly across traits, with openness being the easiest to predict. Conversely, Li et al. (2014) analyzed user activity statistics from the Sina Weibo microblog and achieved similar prediction accuracy among all Big 5 Personality traits, and Skowron, Tkalčič, Ferwerda, and Schedl (2016) analyzed language + user features from users of both Twitter and Instagram and found a high prediction accuracy for conscientiousness, but a relatively low prediction accuracy for agreeableness. Even though many studies have been conducted on the subject, this area of psychological research is still quite young, which in part explains the reason for the lack of uniformity in the employed research methods. For example, studies vary largely on sample sizes, type of digital footprints analyzed, and social media platform used for data collection. Given these circumstances with psychological research conducted on social media, there is a need to synthesize and summarize the existing literature in order to evaluate their accuracy, and recommend best methods for personality prediction from social media.

The ability to use digital footprints to accurately predict personality traits may represent a rapid, cost-effective alternative to surveys and reach larger populations, which can be beneficial for academic, health-related, and commercial purposes. With respect to academic research, the development of automated procedures to measure personality would permit to reach larger samples, and obtain measures potentially less prone to social-desirability bias. Furthermore, personality traits have also been shown to act as potential risk and protective factors for many health-related outcomes (Booth-Kewley & Vickers, 1994; Raynor & Levine, 2009; Widiger & Oltmanns, 2017), and to influence beliefs about health (e.g., Hill & Gick, 2011). Therefore, the ability to distinguish online users based on their personality profiles could be leveraged in order to tailor techniques aimed at improving the efficacy of health related messages (Gale, Deary, Wardle, Zaninotto, & Batty, 2015; Lawson, Bundy, & Harvey, 2007; Neeme, Aavik, Aavik, & Punab, 2015; Rimer & Kreuter, 2006) and individual interventions (Chapman, Hampson, & Clarkin, 2014; Franks, Chapman, Duberstein, & Jerant, 2009) directed at online populations, and thus assist in the effective implementation of public health policies (Chapman, Roberts, & Duberstein, 2011; Hengartner, Kawohl, Haker, Rössler, & Ajdacic-Gross, 2016). With respect to commercial applications, knowledge about individuals' personalities can allow for the enhancement and personalization of recommender systems in order to improve user experiences (Bachrach et al., 2012; Farnadi et al., 2016). Also, social media sites, online advertisers, e-commerce retailers, and e-learning websites may be tailored based on individual personality and present information in ways that will be better received by users (Bachrach et al., 2012; Gao et al., 2013; Golbeck, Robles, & Turner, 2011; Kosinski et al., 2013; Markovikj, Gievska, Kosinski, & Stillwell, 2013).

1.3. Aims

The aim of the current study is to conduct a series of meta-analyses to estimate the mean predictive value of digital footprints on each of the Big 5 Personality Traits. Further, we aim to study if the use of different types of digital footprints influence the accuracy of personality prediction, and if data from different social media platforms lead to different results. Lastly, we will check for possible bias in effect size estimates due to study quality.

2. Methods

2.1. Literature search

To identify relevant studies on the relationships between Big 5 personality traits and digital footprints, we followed the literature search strategies proposed by Durlak and Lipsey (1991). We conducted a broad literature search in databases from various disciplines; i.e., Scopus, ISI Web of Science, Pubmed, and Proquest, using multiple groups of keywords. The first group of keywords used referred to social media platforms, namely; *myspace*, *facebook*, *instagram*, *twitter*, *youtube*, *photobucket*, *linkedin*, *social network*, *reddit*, *social media*, *snapchat*, *periscope*, *social networking*, *status updates*, *mypersonality*. A second group of keywords referred to different analytic approaches that have been previously used to analyze digital footprints from social media in association with individual characteristics, which include; *machine learning*, *data mining*, *text analysis*, *language processing*, *closed vocabulary*, *closed dictionary*, *LIWC*, *open vocabulary*, *open dictionary*, *support vector machines*, *text mining*, *topic modeling*, *dictionary*, *latent dirichlet allocation*, *differential language analysis*, *digital footprint*, *differential language*, *computational linguistics*, *content analysis*. These two groups of keywords were each combined with the following keywords referring to personality traits; *personality*, *traits*, *Big-5/Big-Five*, *Five-Factor Model*, *extraversion*, *introversion*, *neuroticism*, *emotional stability*, *openness*, *conscientiousness*, *and agreeableness*. We searched for terms in the following fields: title, abstract and keywords. We then performed Internet searches via www.google.com and Google Scholar to find other available articles, and we performed an additional search by inspecting citations of the included publications from the initial broad database search. Identified papers were then screened by reading the abstracts based on specific inclusion and exclusion criteria. Papers selected based on abstract information were then fully read to ascertain they met criteria for inclusion. The literature search was finalized in May 2017. Flowchart of article selection is reported in Fig. 1.

2.2. Inclusion and exclusion criteria

Papers identified through database searches were screened for the following inclusion criteria - 1. Studies must link digital footprints and Big 5 personality traits at the individual level, 2. Studies must be focused on digital footprints automatically collected from social media, 3. Studies must include a standardized self-report measure to assess Big 5 personality traits (i.e., the Big 5 Inventory; John & Srivastava, 1999; John, Naumann, & Soto, 2008; 10 item Big 5 Inventory; Gosling, Rentfrow, & Swann, 2003; International Personality Item Pool – IPIP, Goldberg et al., 2006), and 4. Studies had to report information about the accuracy of prediction of Big 5 personality traits based on digital footprints. Studies were also excluded from meta-analysis if they reported non-independent data; meaning studies that used overlapping samples for their analysis were excluded (Senn, 2009). In order to resolve this issue, we followed recommendations from previous studies (Hunter, Schmidt, & Jackson, 1982; Sheppard, Hartwick, & Warshaw, 1988), and considered studies as non-independent if they met the following criteria: (1) each effect-size was based on responses from overlapping sample subjects, (2) digital footprints were extracted from the same social media platform, and (3) type of digital footprint used to predict characteristics were the same or partly overlapping. If we found two or more studies to be non-independent based on this criteria, the study with the largest set of digital footprints was included in the analysis. In the case of non-independent studies analyzing the same set of digital footprints, the one with the larger sample size was included in the meta-analysis.

2.3. Research coding

2.3.1. Coding of types of digital footprints

Studies varied considerably in the number and type of investigated digital footprints. Due the heterogeneity in the type of data, research methods, and the fact that many studies did not detail contributions of single digital footprints to overall prediction, studies were coded based on the inclusion (yes/no) of sets of digital footprints, defined based on their content. More in detail, we differentiated between studies including the following types of digital footprints: (1) Utilization of user demographics (e.g., gender, age), (2) Use of Facebook Likes, (3) Utilization of user activity statistics (e.g., number of posts, number of friends or network density, number of received Likes, comments, and user tags), (4) Utilization of language/text features (e.g., tweets from Twitter, status updates and comments from Facebook), (5) Utilization of pictures (e.g., profile pictures, photos from posts), (6) Utilization of multiple vs. single type of digital footprints.

2.3.2. Coding of social media platform

In order to distinguish between the different types of social media platform, we grouped social media sites based on their default privacy settings, differentiating between public (social media platforms whose posts are public domain by default, i.e., Twitter, Sina Weibo, Reddit, and Instagram), and private (social media platform in where posts are visible only to the users' existing network of friends, i.e., Facebook). These factors may play a role in the accuracy of predicting the Big 5 traits.

2.3.3. Coding of study quality

Due to the relative novelty and multidisciplinary nature of the examined research area, standard methodological procedures for coding study quality have not yet been developed. For this reason, we could not refer to specific guidelines to determine scientific quality of published studies. As an approximation, study quality was assessed by classifying studies based on the rank of the sources they were published in (i.e., peer-reviewed journals and conference proceedings) according to well-known ranking systems of scientific value. More in detail, we used a procedure which differed for peer-reviewed journals and conference proceedings. Concerning articles published in peer reviewed journals, we categorized papers into top, middle and low tiers using the quartile that sources correspond to in the 2016 Scopus CiteScore; quartile 1 was ranked as top tier or high quality, quartile 2 was ranked as middle tier or medium quality, and quartiles 3, 4, and non-indexed studies were ranked as low tier or low quality. In order to assess study quality of proceedings from computer science conferences, we inspected conference ranking as reported in the CORE 2017 and Microsoft Academics databases, which provide rankings of conferences in computer science based on their scientific impact. We considered proceedings as high-quality if at least one of the databases rated the conference with an A (Excellent) score or higher, proceedings with a score of B (Good) were ranked as medium quality, and those with a score of C (ranked conferences meeting minimum standards) and unranked conferences were marked as low quality.

2.4. Strategy of analyses

We collected an effect size for each study, and used Pearson's r to express the accuracy of prediction for the Big 5 personality traits' based on digital footprints. As studies markedly varied in the methods used to study the relationship between digital footprints and personality traits, we employed a twofold approach. The majority of studies (n = 9; Celli, Bruni, & Lepri, 2014; Gao et al., 2013; Kleanthous, Herodotou, Samaras, & Germanakos, 2016; Kosinski et al., 2013; Liu et al., 2016 Study 1 and 2; Skowron et al., 2016; Sumner, Byers, Boochever, & Park, 2012; Wei et al., 2017) tested models using a set of features extracted from digital footprints to predict personality traits. In these cases, we included the overall effect size in the meta-analysis, referring to the predictive power of the model. Some of these studies compared the predictive performance of multiple predictive models based on the same set of features but employing different algorithms (n = 5, Farnadi et al., 2016, Study 1 and 3; Golbeck et al., 2011; Li et al., 2014; Wald, Khoshgoftaar, & Sumner, 2012). For these studies the effect size of the best performing model was included in the analysis. Some other studies (n = 2, Gosling et al., 2011; Qiu et al., 2012) reported multiple effect-sizes (one for each analyzed feature) without furnishing an overall effect size. For instance, Gosling et al. (2011) reported separate effect sizes for

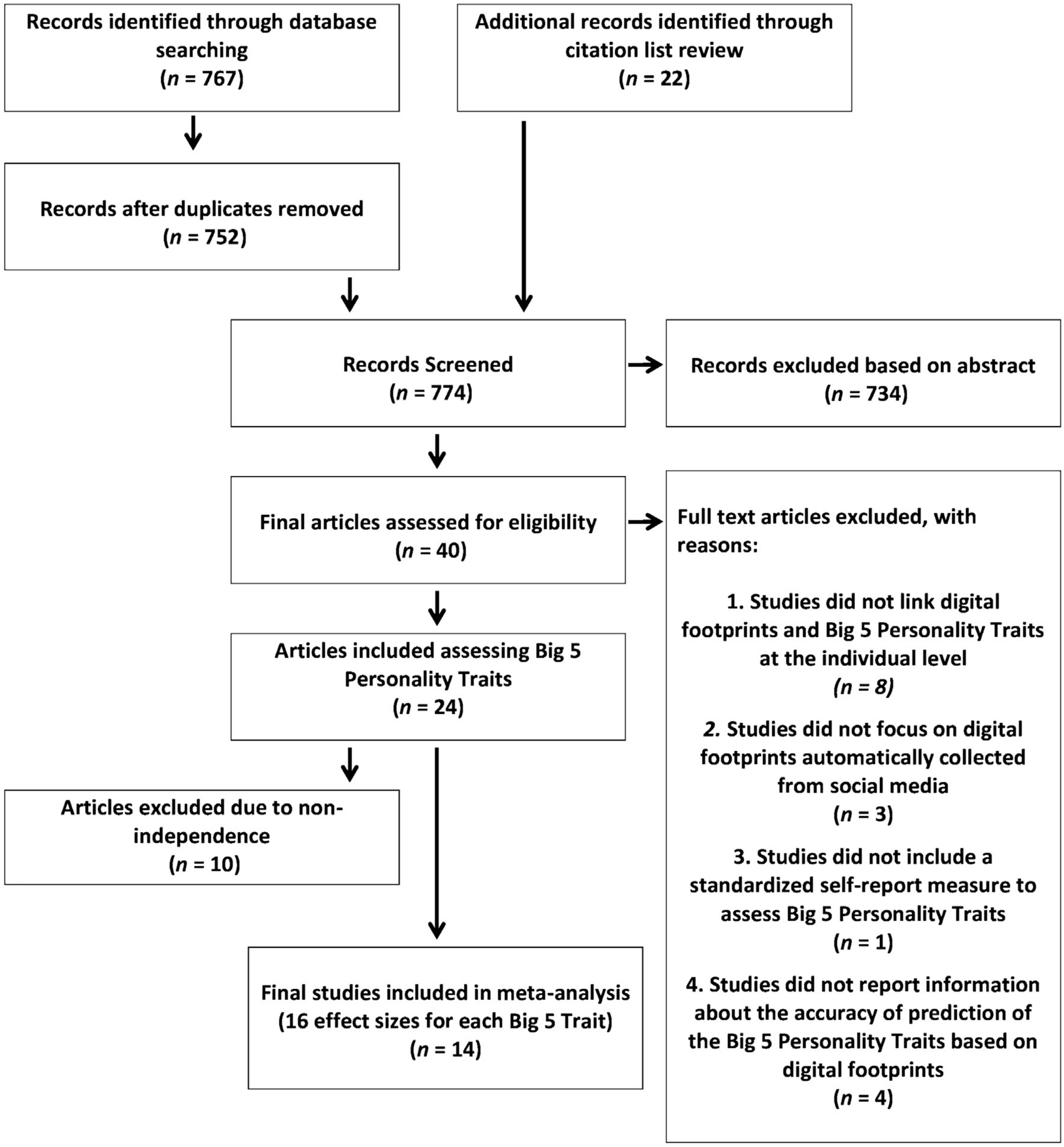

Fig. 1. Flowchart of article selection.

different Facebook activity statistics (e.g., number of friends, number of posts, etc.). In these cases, we included in the meta-analysis the highest effect-size reported, as the best available approximation of the predictive power that would be achieved by a model including the entire set of features as predictors.

Following the indications by Schmidt and Hunter (2014), collected effect sizes were not transformed into Fisher's z scores, since this conversion is not indicated for meta-analytic random-effects models; they yield an upward bias in the estimation of mean correlation, which is normally higher than the bias due to the usage of untransformed correlations. In the event that studies did not report Pearson's r specifically ($n = 4$), the reported effect-sizes were converted to correlations. In case studies reported information about model predictive power using R^2 ($n = 2$), this was converted to correlation by taking the square root of reported value. Area Under the Receiver Operating Characteristic curve (AUROC) statistics ($n = 1$) were first converted to Cohen's d (Ruscio, 2008), and then converted from Cohen's d to r (Rosenthal, Cooper, & Hedges, 1994). When studies provided specificity and sensitivity values ($n = 1$), or positive predicted values (PPV) and negative predicted values (NPV), or when studies provided enough information for computing these statistics, we used this information to compute odds-ratios (Glas, Lijmer, Prins, Bonsel, & Bossuyt, 2003), then converted odds-ratios into Cohen's d (Borenstein, Hedges, Higgins, & Rothstein, 2009), and finally converted Cohen's d into correlations (Rosenthal et al., 1994).

We identified five ($n = 5$) papers that did not include information about effect-size or did not report enough information to compute correlations (e.g., those who reported only mean absolute error (MAE) and root mean square error (RMSE) statistics, or if results were not fully reported in the study). We then contacted the first or corresponding author of these 5 papers and obtained missing information for one study ($n = 1$). As suggested by previous authors (Bowling & Beehr, 2006; Hershcovis, 2011), papers for which information was not obtained were excluded from the analysis ($n = 4$).

We conducted separate meta-analyses for each Big 5 trait. Meta-analyses were performed using a random-effects model as the true effect size was likely to vary in the individual studies, owing to the variety in data sources, study designs, and analytic approaches. Grubb's test was used to identify outliers. Heterogeneity of the studies' effect-sizes was determined by computing the following statistics: (1) the chi-square Q test of heterogeneity, (2) T^2 estimate of true between-study variance, and (3) the I^2 statistic of proportion of true variation in observed effects. Existence of publication bias was investigated by inspecting funnel plot, and by using Begg and Mazumdar rank correlation test (Begg & Mazumdar, 1994), Egger's intercept test (Sterne & Egger, 2001), Duval and Tweedie's trim and fill procedure (Duval & Tweedie, 2000), and classic fail-safe *N*.

We then analyzed potential moderators using meta-regression models. We measured the possible effects of moderators on study effect-sizes by random-effects univariate meta-regressions using restricted maximum-likelihood estimation. Based on the previous coding procedures for digital footprints, social media platform, and study quality, the authors separately coded all studies for eight potential moderators, which are: 1. Type of social media platform (private vs. public); 2. Utilization of user demographics (yes vs. no); 3. Use of Likes (yes vs.no); 4. Utilization of user activity statistics (yes vs. no); 5. Utilization of language/text features (yes vs. no); 6. Utilization of pictures (yes vs. no); 7. Utilization of multiple vs. single type of digital footprints. 8. Study quality (High, Medium, Low). Overall, coding of moderators required little subjective judgment. Full agreement between coders was reached. In order to conduct moderator analyses, and to acquire acceptably robust coefficient estimates, we followed the suggestion by Fu and colleagues and examined the effect of moderators only if at least 4 studies per group were available (Fu et al., 2014). A critical value of $\boldsymbol{\alpha} = 0.05$ was used in meta-regression analyses. However, given the low number of studies, effects approaching statistical significance ($p < 0.10$) are commented.

All analyses were performed using Comprehensive Meta-analysis software (Version 3.3.070).

3. Results

3.1. Overview of included studies

In total, we identified 24 papers focusing on the analysis of digital footprints extracted from social media and Big 5 Personality traits. Selected papers included 28 studies in which Big 5 personality traits were assessed using versions of the Big 5 Personality Inventory and IPIP measures. 19 studies obtained their samples from Facebook, 5 from Twitter, 3 from the Sina Weibo micro-blogging site, and 1 article used a combined sample from Instagram and Twitter. Twenty studies analyzed a single feature extracted from digital footprints (e.g., user activity, demographics, language, pictures, and Facebook 'Likes'), while 8 studies analyzed a combination of multiple features extracted from digital footprints (e.g., demographics + user activity + language, language + pictures, etc.). For a detailed description of study characteristics refer to Table 1.

Inspection of non-independence led us to exclude a total of 12 studies from the meta-analysis: most of the excluded studies ($n = 11$) were discarded because they used data from the MyPersonality dataset, and analyzed the same type of digital footprints extracted from Facebook. Among studies using MyPersonality data collected on Facebook, we included in the analyses those which examined the most comprehensive set of digital footprints, and in case they examined the same set of digital footprints, we selected those with the largest sample (Farnadi et al., 2016 Study 1; Kosinski et al., 2013). Study 3 by Golbeck (2016) was excluded because it shared the same data with the study by Golbeck et al. (2011). After inspection of studies for non-independence, we selected a subset of 14 papers including 16 independent studies, resulting in 80 independent effect-sizes (16 for each of the Big 5 personality traits). Of the 16 selected studies, 7 were based on data collected from Facebook, 5 from Twitter, 3 from the Sina Weibo micro-blog, and 1 was based on a sample that used combined data from Instagram and Twitter. 9 of these studies were based on analysis including only a single type of digital footprint from social media, while 7 were based on analyses performed on multiple types of digital footprints. Grubb's test failed to identify any outliers, resulting in no further studies being excluded.

3.2. Meta-analyses

3.2.1. Mean effect size

To establish the magnitude of the association between digital footprints and each of the Big 5 personality traits, we conducted five separate meta-analyses analyzing 16 effect-sizes for each trait. Forest plot of effect-sizes included in the meta-analyses are presented in Fig. 2. The estimated meta-analytic correlations were 0.39 (95% CI: 0.30–0.48) for Openness, 0.35 (95% CI: 0.29–0.42) for Conscientiousness, 0.40 (95% CI: 0.33–0.46) for Extraversion, 0.29 (95% CI: 0.21–0.36) for Agreeableness, and 0.33 (95% CI: 0.27–0.39) for Neuroticism. Results of Q test for heterogeneity were significant for each trait (see Table 2). T^2 ranged from 0.01 (neuroticism) to 0.04 (openness), indicating relatively low true heterogeneity between studies. Observed dispersion of effect-sizes was mostly due to true heterogeneity ($I^2 \geq 93.15$).

3.2.2. Publication bias

First, we inspected the funnel plots, plotting the included studies' effect size against its standard error. For each Big 5 trait, the funnel plot was symmetrical, suggesting lack of publication bias. Coherently, Trim-and-fill analyses suggested that no studies were missing on the left side of the mean effect. For each trait, non-significant Begg and Mazumdar test (Openness: $p = 0.39$; Conscientiousness: $p = 0.21$; Extraversion: $p = 0.50$; Agreeableness: $p = 0.24$; Neuroticism: $p = 0.26$) and Egger's test (Openness: $p = 0.31$; Conscientiousness: $p = 0.14$; Extraversion: $p = 0.49$; Agreeableness: $p = 0.44$; Neuroticism: $p = 0.29$) further indicated no significant evidence of publication bias.

For each trait, the fail-safe *N* value was higher than 90 (Openness: $N = 12{,}210$; Conscientiousness: $N = 7688$; Extraversion: $N = 11{,}933$; Agreeableness: $N = 6053$; Neuroticism: $N = 7197$), corresponding to the recommended rule-of-thumb limit of 5 k + 10 (Rosenthal, 1979).

The results of these four tests indicate that it is unlikely that publication bias poses a significant threat to the validity of the findings reported in the current analyses.

3.2.3. Moderator analyses

We examined the following moderating effects: (1) Private vs. public social-media platform, (2) Utilization of user demographics (yes vs. no), (3) Use of Likes (yes vs. no), (4) Utilization of user activity statistics (yes vs. no), (5) Utilization of language/text features (yes vs. no), (6) Utilization of pictures (yes vs. no), (7) Utilization of multiple vs. single type of digital footprints, (8) Study quality (High, Medium, and Low).

Concerning study quality, given the low number of studies marked as low ($n = 2$) and medium ($n = 2$) quality when compared to those marked as high quality ($n = 12$), studies in the low and medium categories were grouped together so as to reach the per-group minimum of 4 studies required for testing the moderator effect. Use of Likes was not tested as a moderator as only one of the included studies used Likes for personality prediction.

Results of univariate regressions showed significant effects for use of multiple types of digital footprints, demographics, and activity statistics. For each trait except agreeableness, results showed an increase in strength of association between digital footprints and personality traits when studies examined multiple types of digital footprints, instead of only one type. However, the effects were statistically significant ($p < 0.05$) only for openness ($\beta = 0.27$, $R^2 = 0.16$), conscientiousness ($\beta = 0.25$, $R^2 = 0.20$), and neuroticism ($\beta = 0.21$, $R^2 = 0.14$). Results of analyses for extraversion suggested a similar trend ($\beta = 0.18$, $R^2 = 0.12$), but the effect did not reach significance ($p = 0.08$).

Use of demographic statistics was associated with a significant

Table 1
Characteristics of studies included in the meta-analyses.

Study	Self-report	Source (Quality)	Social media	Digital footprints
Bachrach et al., 2012*	IPIP	Proceeding (Low)	Facebook	Activity
Celli et al., 2014 1	Big 5 Inventory - 10	Proceeding (High)	Facebook	Pictures
Farnadi et al., 2016 1*	IPIP	Journal (High)	Facebook	Demographics, Activity, Language
Farnadi et al., 2016 3*	Big 5 Inventory - 10	Journal (High)	Twitter	Demographics, Language
Gao et al., 2013	Big 5 Inventory	Proceeding (Medium)	Sina Weibo	Activity, Language
Golbeck et al., 2011	Big 5 Inventory	Proceeding (High)	Facebook	Demographics, Activity, Language
Golbeck, 2016 1*	IPIP	Journal (Low)	Facebook	Language
Golbeck, 2016 2*	IPIP	Journal (Low)	Facebook	Language
Golbeck, 2016 3	Big 5 Inventory	Journal (Low)	Facebook	Language
Gosling et al., 2011	TIPI	Journal (High)	Facebook	Activity
Kern et al., 2014*	IPIP	Journal (High)	Facebook	Language
Kleanthous et al., 2016	IPIP	Proceeding (Medium)	Facebook	Activity
Kosinski et al., 2013 1*	IPIP	Journal (High)	Facebook	Likes
Kosinski et al., 2014 1*	IPIP	Journal (High)	Facebook	Activity
Li et al., 2014	Big 5 Inventory	Journal (High)	Sina Weibo	Activity
Liu et al., 2016 1	IPIP	Proceeding (High)	Twitter	Language
Liu et al., 2016 2	IPIP	Proceeding (High)	Twitter	Pictures
Markovikj et al., 2013*	IPIP	Proceeding (High)	Facebook	Demographics, Activity, Language
Park et al., 2015*	IPIP	Journal (High)	Facebook	Language
Qiu et al., 2012	Big 5 Inventory	Journal (High)	Twitter	Language
Quercia et al., 2012 1*	IPIP	Proceeding (High)	Facebook	Activity
Schwartz et al., 2013*	IPIP	Journal (High)	Facebook	Language
Skowron et al., 2016	Big 5 Inventory	Proceeding (High)	Twitter, Instagram	Language, Pictures
Sumner et al., 2012	TIPI	Proceeding (Low)	Twitter	Activity, Language
Thilakaratne, Weerasinghe, & Perera, 2016*	IPIP	Proceeding (Medium)	Facebook	Language
Wald et al., 2012	Big 5 Inventory	Proceeding (Low)	Facebook	Demographics, Activity, Language
Wei et al., 2017	Big 5 Inventory	Proceeding (High)	Sina Weibo	Activity, Language, Pictures
Youyou et al., 2015*	IPIP	Journal (High)	Facebook	Likes

Note. Studies included in the meta-analyses are in bold. *Study using MyPersonality datasets.

increase in correlation strength between digital footprints and both agreeableness ($\beta = 0.25$, $R^2 = 0.19$), and neuroticism ($\beta = 0.25$, $p < 0.05$, $R^2 = 0.19$). Results of analyses for openness also revealed a marginally significant ($p = 0.09$) increase in association ($\beta = 0.26$, $R^2 = 0.12$). Similarly, use of activity statistics for prediction purposes was associated with an increase in predictive power over extraversion ($\beta = 0.19$, $R^2 = 0.18$, $p = 0.06$). No other significant moderator effects emerged.

4. Discussion

To our knowledge, this is the first meta-analysis aimed at summarizing findings from studies investigating the predictability of Big 5 personality traits based on digital footprints automatically extracted from social media. Our first aim was to estimate the mean predictive value of digital footprints over each trait. Overall, prediction of Big 5 traits based on the analysis of digital footprints from social media ranged from 0.29 (agreeableness) to 0.40 (extraversion), with no significant differences in effect-size across traits. In general, the emerging relationships between digital footprints and personality seems to be in line with the typical strength of the relationships between personality and behaviors, also known as "personality coefficient" (a Pearson correlation ranging from 0.30 to 0.40; Meyer et al., 2001; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). This indicates that digital records of behaviors on social media may represent a quite reliable source of information for the prediction of individual personality traits. However, some of the studies included in this meta-analysis failed to find significant associations between digital footprints and some of the Big 5 traits (Celli et al., 2014; Farnadi et al., 2016; Gosling et al., 2011; Kleanthous et al., 2016; Liu et al., 2016; Qiu et al., 2012), and a significant effect size heterogeneity emerged between studies. These results support the usefulness of investigating the possible sources of differences in prediction accuracy across studies. Therefore, as a second aim, our study investigated the influence of a set of study characteristics, namely the use of different types of digital footprints, social media platforms, and study quality, on the prediction accuracy of each personality trait. With the exception of agreeableness, our results indicate that prediction accuracy for each trait was stronger when more than one type of digital footprint was analyzed. Concerning the use of specific types of digital footprints, use of demographic data was found to increase prediction accuracy for openness, agreeableness, and neuroticism, while use of activity statistics resulted in an improvement in the accuracy of prediction of extraversion. Also, use of features extracted from texts and pictures posted on social media did not improve prediction accuracy of personality traits over use of other types of digital footprints. These findings appear to be consistent with survey literature indicating the influence of demographic information, such as age and gender, in explaining individual differences on self-reports for Big 5 traits (e.g., Goldberg, Sweeney, Merenda, & Hughes, 1998; Lehmann, Denissen, Allemand, & Penke, 2013; Soto, John, Gosling, & Potter, 2011), as well as the existence of a positive link between extraversion and engagement in social media activities (Blackwell et al., 2017; Kuss & Griffiths, 2011). Furthermore, we found that default privacy settings of social media platforms, namely public vs. private, did not show a significant impact on the accuracy of personality prediction based on social media data. As most of social media platforms provide users with the ability to significantly customize privacy settings, and custom privacy settings are expected to have a stronger influence on users' self-expression on social media than default settings (Waterloo, Baumgartner, Peter, & Valkenburg, 2017), this finding should be taken with caution. Further, default privacy settings are expected to radically mutate over time due to ever-shifting privacy policies of social media platforms (e.g., Barrett, 2016; Warzel, 2014). Future studies exploring the impact of privacy settings on the use digital footprints for personality predictions should consider collecting information about users' actual selected privacy settings.

Lastly, we found that study quality did not influence the strength of association with personality. Overall, analysis of moderators pointed out that a significant part of the effect size heterogeneity can be traced back to the variety of digital footprints included in the analyses: generally, higher effect sizes have been achieved by studies including multiple types of digital footprints. Further studies will permit to

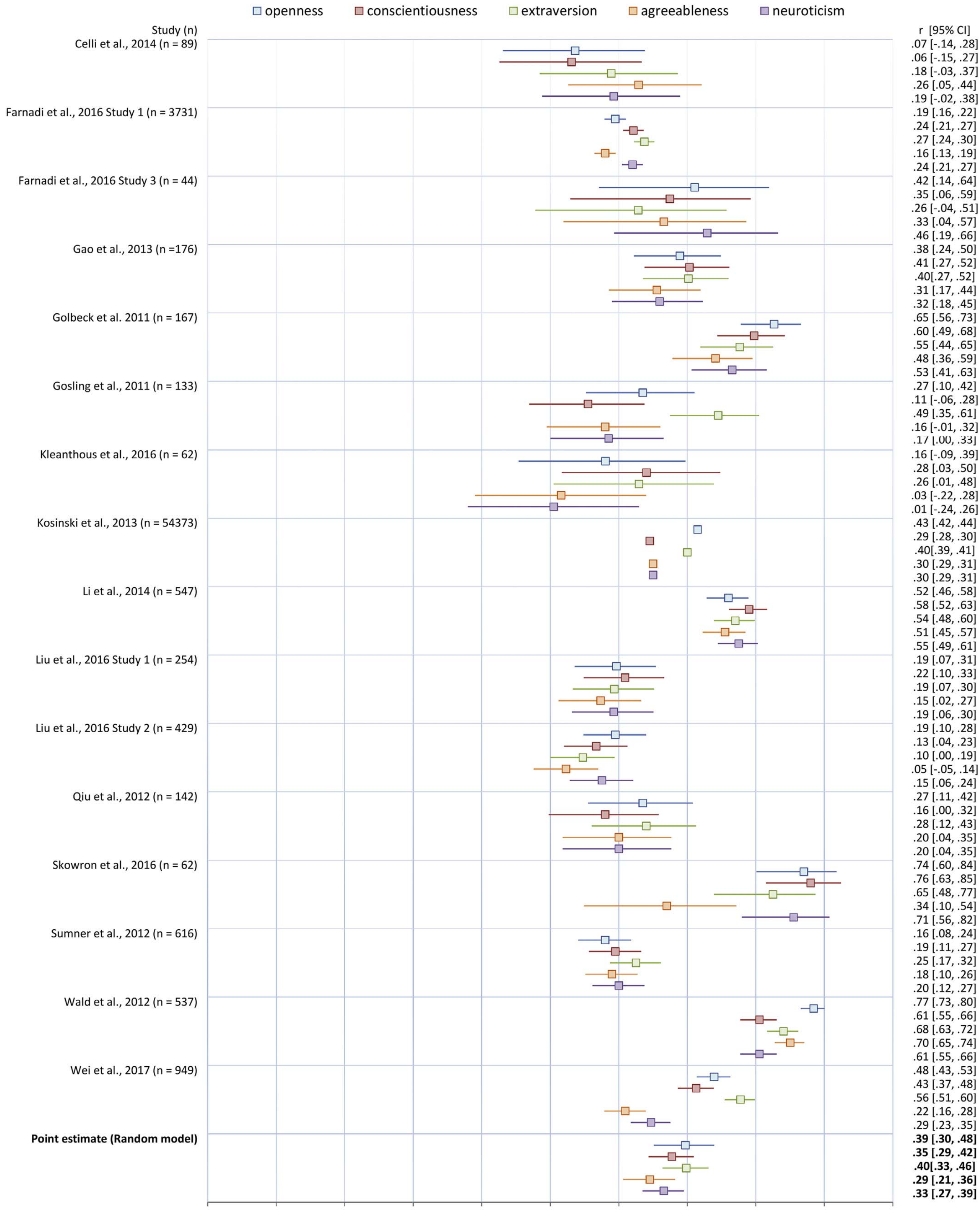

Fig. 2. Forest plot combining effect-sizes and point estimate (random model) for each Big 5 trait.

Table 2
Meta-analytic correlations and heterogeneity statistics for each Big 5 trait.

Trait	Point estimate [95% CI]	Z	Q (df)	I^2	T^2	T
Openness	0.39 [0.30, 0.48]	7.81**	590.00 (15)**	97.46	0.04	0.20
Conscientiousness	0.35 [0.29, 0.42]	9.63**	281.91 (15)**	94.68	0.02	0.14
Extraversion	0.40 [0.33, 0.46]	10.32**	320.911 (15)**	95.33	0.02	0.15
Agreeableness	0.29 [0.21, 0.36]	7.07**	350.77 (15)**	95.72	0.02	0.15
Neuroticism	0.33 [0.27, 0.39]	9.96**	219.014 (15)**	93.15	0.01	0.12

** $p < 0.001$.

confirm this relationship; in order to reach a higher predictive power, scholars should aim to collect and analyze multiple types of digital footprints.

Overall, the predictive power of digital footprints over the individual Big 5 traits, combined with the resemblance in accuracy of predictions across traits, provides encouraging results for researchers who aim to utilize digital footprints from social medial to predict the Big 5 personality traits. Given the relatively recent emergence of personality prediction from social media, and the continuous rapid evolutions that make accessing the large datasets of social media users possible, we expect the accuracy in prediction of the Big 5 traits to improve significantly in the near future. We anticipate an improvement in accuracy due to the ongoing transition from traditional analytic approaches toward a more innovative employment of data mining techniques (e.g., machine learning algorithms) (Kosinski, Wang, Lakkaraju, & Leskovec, 2016), and to the emergence of new techniques to extract essential information from visual data (i.e., image recognition via artificial intelligence) (Guo et al., 2016), which is notably important due to the modern shifts in content sharing on social media from text, to pictures and videos (Statista, 2017).

In light of these considerations, it is worth addressing the ethical issues that may emerge from the development and employment of techniques aimed at assessing individual characteristics on the basis of user data recorded from social media and the internet. The ability to identify people with specific personality profiles, with individual consent, presents an opportunity to customize and enhance online advertising and marketing, improve user's online experience, and inform public health initiatives. On the other hand, possible exploitation, or misuses, of these techniques exist: for example, newspapers recently reported cases which demonstrated the feasibility and adequacy of targeting political propaganda on the basis of information not explicitly disclosed by social media users (Cadwalladr, 2017; Confessore & Hakim, 2017), and reported the use of this information by advertisers to target individuals based on emotional states (Levin, 2017). The dangers associated with the use of these new and emerging techniques to specific areas and subjects should be carefully considered by scholars. It may also prove beneficial to disseminate awareness about these issues among both policymakers and the public audience in order to protect individuals' privacy and prevent possible exploitations of user data.

5. Limitations of the study

The present study is not without limitations. First, given the relatively low number of studies investigating diverse social media platforms and the heterogeneity of both the features analyzed and the analytical approaches employed in the studies included in the analysis, we could not perform a thorough comparison of the accuracy of personality prediction across specific social media platforms. The diverse usage, or activities, users partake in while engaging in specific types of social media platforms might significantly affect the strength of the association between digital footprints and actual personality, improving or hindering the accuracy of predictions. Similarly, the heterogeneity in data extraction and analytic procedures did not permit to compare the contribution of individual features to prediction accuracy. More studies are needed in order to test this hypothesis, as well as to confirm the existence, and establish the strength of moderation effects emerging from the present study.

Second, the present study failed to investigate the impact of cultural differences on the predictability of personality from social media data. Collected data was not sufficient to compare accuracy of personality prediction across different cultural contexts. As most of the included studies either focused on samples of English-speaking users, or explored samples recruited among Chinese users of the Sina Weibo social media platform, there appears to be a need for more studies including non-western populations.

A last limitation concerns the examination of use of visual digital footprints (e.g., pictures, videos) to predict personality. Production and online sharing of visual content is expected to increase dramatically in the next few years (Cisco, 2017), and newer social media platforms focusing on visual content such as Instagram and Snapchat, are now outgrowing older social media platforms (e.g., Facebook, Twitter) in popularity especially among younger people (Richter, 2017). However, only a minority of studies included in the meta-analysis used pictures to predict personality, and none of them included data about videos; further, all examined studies, except for one (Skowron et al., 2016), failed to investigate use of digital footprints collected from highly visual social media platforms such as Instagram and Snapchat. For this reason, results concerning the predictive power of visual data to predict personality are to be taken as preliminary, and further studies focusing on emerging highly-visual social media are needed to establish the relevance of visual digital footprints for the prediction of personality.

6. Conclusions

Overall, the present meta-analysis demonstrates that Big 5 personality traits can be inferred using digital footprints extracted from social media with remarkable accuracy. The ability to make distinct but similarly accurate predictions of Big 5 traits allows for the identification of social media users with different personality profiles. This information is of utmost relevance since it can be beneficial for research, commercial, and public health purposes. First, the ability to assess personality in an unobtrusive way via the analysis of social media data would allow researchers to reach larger samples and obtain measures, which are potentially less biased than traditional self-reports. Next, accurate predictions of the Big 5 traits could be usefully applied to online marketing and advertising by making it possible to profile individuals, and tailor advertisements automatically displayed in individual users profiles based on personality (Bachrach et al., 2012). Furthermore, areas of human-computer interactions (HCI) may use this information to create adaptive and personalized systems in order to provide rich and best possible user experiences (Farnadi et al., 2016), and recommendation systems may also capitalize on this information by including personality dimensions to their current user models and present information in ways that will be most attractive to users (Golbeck et al., 2011; Nass & Lee, 2000). Finally, at the public health level, the ability to tailor online messages based on social media user's personality information could be used to improve the implementation of public health programs by increasing the efficacy of targeted health campaigns, screening programs, and interventions directed at online populations (Chapman et al., 2014; Franks et al., 2009).

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality.

Computers in Human Behavior, 26(6), 1289–1295.
Anderson, C., John, O. P., Keltner, D., & Kring, A. M. (2001). Who attains social status? Effects of personality and physical attractiveness in social groups. *Journal of Personality and Social Psychology, 81*(1), 116.
Aslam, S. (2017). *Twitter by the numbers: Stats, demographics & fun facts.* OMNICORE. Retrieved from: https://www.omnicoreagency.com/twitter-statistics/.
Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of Facebook usage. *Proceedings of the 4th annual ACM web science conference* (pp. 24–32). ACM. http://dx.doi.org/10.1145/2380718.2380722.
Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*(3), 372–374.
Barrett, B. (2016). You should go check Facebook's new privacy settings. *Wired*. Retrieved from https://www.wired.com/2016/06/go-check-facebooks-new-privacy-settings/.
Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics,* 1088–1101.
Blackwell, D., Leaman, C., Tramposch, R., Osborne, C., & Liss, M. (2017). Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences, 116*, 69–72.
Booth-Kewley, S., & Vickers, R. R. (1994). Associations between major domains of personality and health behavior. *Journal of Personality, 62*(3), 281–298.
Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). Meta-regression. *Introduction to meta-analysis* (pp. 187–203). .
Bowling, N. A., & Beehr, T. A. (2006). *Workplace harassment from the victim's perspective: A theoretical model and meta-analysis*.
Cadwalladr, C. (2017). The great British Brexit robbery: How our democracy was hijacked. *The Guardian*. Retrieved from www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy.
Celli, F., Bruni, E., & Lepri, B. (2014). Automatic personality and interaction style recognition from Facebook profile pictures. *Proceedings of the 22nd ACM international conference on multimedia* (pp. 1101–1104). ACM. http://dx.doi.org/10.1145/2647868.2654977.
Chapman, B. P., Hampson, S., & Clarkin, J. (2014). Personality-informed interventions for healthy aging: Conclusions from a National Institute on Aging work group. *Developmental Psychology, 50*(5), 1426.
Chapman, B. P., Roberts, B., & Duberstein, P. (2011). Personality and longevity: Knowns, unknowns, and implications for public health and personalized medicine. *Journal of Aging Research* 759170.
Cisco (2017). *Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021.* Cisco. Retrieved from https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.
Confessore, N., & Hakim, D. (2017). Data firm says 'secret sauce' aided trump: Many scoff. *The New York Times*. Retrieved from www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html?_r=0.
De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social media as a measurement tool of depression in populations. *Proceedings of the 5th annual ACM web science conference* (pp. 47–56). ACM.
De Choudhury, M., Counts, S., Horvitz, E. J., & Hoff, A. (2014). Characterizing and predicting postpartum depression from shared Facebook data. *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing* (pp. 626–638). ACM.
Donnellan, M. B., Conger, R. D., & Bryant, C. M. (2004). The Big Five and enduring marriages. *Journal of Research in Personality, 38*(5), 481–504.
Donnellan, M. B., Larsen-Rife, D., & Conger, R. D. (2005). Personality, family history, and competence in early adult romantic relationships. *Journal of Personality and Social Psychology, 88*(3), 562.
Durlak, J. A., & Lipsey, M. W. (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology, 19*(3), 291–332.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... Weeg, C. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science, 26*(2), 159–169.
Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., ... De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction, 26*(2–3), 109–142. http://dx.doi.org/10.1007/s11257-016-9171-0.
Franks, P., Chapman, B., Duberstein, P., & Jerant, A. (2009). Five factor model personality factors moderated the effects of an intervention to enhance chronic disease management self-efficacy. *British Journal of Health Psychology, 14*(3), 473–487.
Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., ... Santaguida, P. (2014). Conducting quantitative synthesis when comparing medical interventions. *Methods guide for effectiveness and comparative effectiveness reviews*. *254*.
Gale, C. R., Deary, I. J., Wardle, J., Zaninotto, P., & Batty, G. D. (2015). Cognitive ability and personality as predictors of participation in a national colorectal cancer screening programme: The English Longitudinal Study of Ageing. *Journal of Epidemiology and Community Health, 69*(6), 530–535.
Gao, R., Hao, B., Bai, S., Li, L., Li, A., & Zhu, T. (2013). Improving user profile with personality traits predicted from social media content. *Proceedings of the 7th ACM conference on recommender systems* (pp. 355–358). ACM. http://dx.doi.org/10.1145/2507157.2507219.
Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., & Ha, S. E. (2010). Personality and political attitudes: Relationships across issue domains and political contexts. *American Political Science Review, 104*(1), 111–133.
Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. (2003). The diagnostic odds ratio: A single indicator of test performance. *Journal of Clinical Epidemiology, 56*(11), 1129–1135. http://dx.doi.org/10.1016/S0895-4356(03)00177-X.
Golbeck, J. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research, 2*(1), 2.
Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *CHI'11 extended abstracts on human factors in computing systems* (pp. 253–262). ACM. http://dx.doi.org/10.1145/1979742.1979614.
Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public domain personality measures. *Journal of Research in Personality, 40*, 84–96. http://dx.doi.org/10.1016/j.jrp.2005.08.007.
Goldberg, L. R., Sweeney, D., Merenda, P. F., & Hughes, J. E. (1998). Demographic variables and personality: The effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes. *Personality and Individual Differences, 24*(3), 393–403.
Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking, 14*(9), 483–488. http://dx.doi.org/10.1089/cyber.2010.0087.
Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528.
Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing, 187*, 27–48.
Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual Differences, 34*(4), 723–727.
Hengartner, M. P., Kawohl, W., Haker, H., Rössler, W., & Ajdacic-Gross, V. (2016). Big Five personality traits may inform public health policy and preventive medicine: Evidence from a cross-sectional and a prospective longitudinal epidemiologic study in a Swiss community. *Journal of Psychosomatic Research, 84*, 44–51.
Hershcovis, M. S. (2011). "Incivility, social undermining, bullying… oh my!": A call to reconcile constructs within workplace aggression research. *Journal of Organizational Behavior, 32*(3), 499–519.
Hill, E. M., & Gick, M. L. (2011). The big five and cervical screening barriers: Evidence for the influence of conscientiousness, extraversion and openness. *Personality and Individual Differences, 50*(5), 662–667.
Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies. Vol. 4*Sage Publications Inc.http://dx.doi.org/10.1037//0021-9010.75.3.334.
John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research. 3. Handbook of personality: Theory and research* (pp. 114–158).
John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research. 2(199). Handbook of personality: Theory and research* (pp. 102–138).
Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621–652.
Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., ... Seligman, M. E. (2014). The online social self: An open vocabulary approach to personality. *Assessment, 21*(2), 158–169. http://dx.doi.org/10.1177/1073191113514104.
Kleanthous, S., Herodotou, C., Samaras, G., & Germanakos, P. (2016). Detecting personality traces in users' social activity. *International conference on social computing and social media* (pp. 287–297). Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-39910-2_27.
Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences, 19*(1), 47–52.
Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning, 95*(3), 357–380. http://dx.doi.org/10.1007/s10994-013-5415-y.
Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist, 70*(6), 543.
Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 110*(15), 5802–5805. http://dx.doi.org/10.1073/pnas.1218772110.
Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods, 21*(4), 493.
Kuss, D. J., & Griffiths, M. D. (2011). Online social networking and addiction—A review of the psychological literature. *International Journal of Environmental Research and Public Health, 8*(9), 3528–3552. http://dx.doi.org/10.3390/ijerph8093528.
Lawson, V. L., Bundy, C., & Harvey, J. N. (2007). The influence of health threat communication and personality traits on personal models of diabetes in newly diagnosed diabetic patients. *Diabetic Medicine, 24*(8), 883–891.
Lehmann, R., Denissen, J. J., Allemand, M., & Penke, L. (2013). Age and gender differences in motivational manifestations of the Big Five from age 16 to 60. *Developmental Psychology, 49*(2), 365.
Levin, S. (2017). Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless'. *The Guardian*. Retrieved from www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens.
Li, L., Li, A., Hao, B., Guan, Z., & Zhu, T. (2014). Predicting active users' personality based on micro-blogging behaviors. *PLoS One, 9*(1), e84997. http://dx.doi.org/10.1371/journal.pone.0084997.

Liu, L., Preotiuc-Pietro, D., Samani, Z. R., Moghaddam, M. E., & Ungar, L. H. (2016). Analyzing personality through social media profile picture choice. *ICWSM* (pp. 211–220). .

Madden, M., Fox, S., Smith, A., & Vitax, J. (2007). *Digital footprints.* Pew Research Center. Retrieved from http://www.pewinternet.org/2007/12/16/digital-footprints/.

Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013). Mining Facebook data for predictive personality modeling. *Proceedings of the 7th international AAAI conference on weblogs and social media (ICWSM 2013), Boston, MA, USA* (pp. 23–26). .

Matz, S. C., & Netzer, O. (2017). Using big data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences, 18*, 7–12.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., ... Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*(2), 128.

Nass, C., & Lee, K. M. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 329–336). ACM.

Neal, A., Yeo, G., Koy, A., & Xiao, T. (2012). Predicting the form and direction of work role performance from the Big 5 model of personality traits. *Journal of Organizational Behavior, 33*(2), 175–192.

Neeme, M., Aavik, A., Aavik, T., & Punab, M. (2015). Personality and utilization of prostate cancer testing: Evidence for the influence of neuroticism and conscientiousness. *SAGE Open, 5*(3) (2158244015593324).

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401–421.

Padrez, K. A., Ungar, L., Schwartz, H. A., Smith, R. J., Hill, S., Antanavicius, T., ... Merchant, R. M. (2015). Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ Quality and Safety, 25*(6), 414–423.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology, 108*(6), 934. http://dx.doi.org/10.1037/pspp0000020.

Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality, 46*(6), 710–718. http://dx.doi.org/10.1016/j.jrp.2012.08.008.

Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., & Crowcroft, J. (2012). The personality of popular Facebook users. *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 955–964). ACM. http://dx.doi.org/10.1145/2145204.2145346.

Raynor, D. A., & Levine, H. (2009). Associations between the five-factor model of personality and health behaviors among college students. *Journal of American College Health, 58*(1), 73–82.

Richter, F. (2017). Snapchat crowned number 1 by American teens. *Statista*. Retrieved from https://www.statista.com/chart/4823/teenagers-favorite-social-networks/.

Rimer, B. K., & Kreuter, M. W. (2006). Advancing tailored health communication: A persuasion and message effects perspective. *Journal of Communication, 56*(s1).

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*(4), 313–345.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638.

Rosenthal, R., Cooper, H., & Hedges, L. V. (1994). Parametric measures of effect size. *The handbook of research synthesis,* 231–244.

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*(1), 19. http://dx.doi.org/10.1037/1082-989X.13.1.19.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings.* Sage Publicationshttp://dx.doi.org/10.4135/9781483398105.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One, 8*(9), e73791. http://dx.doi.org/10.1371/journal.pone.0073791.

Schwartz, H. A., & Ungar, L. H. (2015). Data-driven content analysis of social media: A systematic overview of automated methods. *The Annals of the American Academy of Political and Social Science, 659*(1), 78–94.

Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences, 54*(3), 402–407.

Senn, S. J. (2009). Overstating the evidence–double counting in meta-analysis and related problems. *BMC Medical Research Methodology, 9*(1), 10.

Settanni, M., & Marengo, D. (2015). Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology, 6*.

Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research, 15*(3), 325–343. http://dx.doi.org/10.1086/209170.

Skowron, M., Tkalčič, M., Ferwerda, B., & Schedl, M. (2016). Fusing social media cues: Personality prediction from Twitter and Instagram. *Proceedings of the 25th international conference companion on World Wide Web* (pp. 107–108). International World Wide Web Conferences Steering Committee. http://dx.doi.org/10.1145/2872518.2889368.

Soldz, S., & Vaillant, G. E. (1999). The Big Five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality, 33*(2), 208–232.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100*(2), 330.

Statista (2017). The most popular mobile social networking apps in the United States as of July 2017, by monthly users. *Statista*. Retrieved from www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*(10), 1046–1055.

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. *Machine learning and applications (ICMLA), 2012 11th international conference on*. *Vol. 2*. *Machine learning and applications (ICMLA), 2012 11th international conference on* (pp. 386–393). IEEE. http://dx.doi.org/10.1109/icmla.2012.218.

Thilakaratne, M., Weerasinghe, R., & Perera, S. (2017). Knowledge-driven approach to predict personality traits by leveraging social media data. *Proceedings - 2016 IEEE/WIC/ACM international conference on Web intelligence, WI 2016 7817065* (pp. 288–295). . http://dx.doi.org/10.1109/WI.2016.0048.

Wald, R., Khoshgoftaar, T., & Sumner, C. (2012). Machine prediction of personality from Facebook profiles. *Proceedings of the 2012 IEEE 13th International Conference on Information Reuse and Integration, IRI 2012 6302998* (pp. 109–115). . http://dx.doi.org/10.1109/IRI.2012.6302998.

Wang, S. S. (2013). "I share, therefore I am": Personality traits, life satisfaction, and Facebook check-ins. *Cyberpsychology, Behavior, and Social Networking, 16*(12), 870–877.

Warzel, C. (2014). Facebook makes a major change to its privacy policies. *BUZZFEED*. Retrieved from: http://www.buzzfeed.com/charliewarzel/facebook-makes-a-huge-change-to-itsprivacy-policies#.wd6ogxwa5.

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2017). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society, 1461444817707349*.

Wei, H., Zhang, F., Yuan, N. J., Cao, C., Fu, H., Xie, X., ... Ma, W. Y. (2017). Beyond the words: predicting user personality from heterogeneous information. *Proceedings of the tenth ACM international conference on web search and data mining* (pp. 305–314). ACM. http://dx.doi.org/10.1145/3018661.3018717.

Widiger, T. A., & Oltmanns, J. R. (2017). Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry, 16*(2), 144–145.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, 112*(4), 1036–1040. http://dx.doi.org/10.1073/pnas.1418680112.