Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

Facebook versus Twitter: Cross-Platform Differences in Self-Disclosure and Trait Prediction

Kokil Jaidka, Sharath Chandra Guntuku, Lyle H. Ungar
Computer and Information Science
University of Pennsylvania
{jaidka@sas,sharathg@sas,ungar@cis}.upenn.edu

Abstract

This study compares self-disclosure on Facebook and Twitter through the lens of demographic and psychological traits. Predictive evaluation reveals that language models trained on Facebook posts are more accurate at predicting age, gender, stress, and empathy than those trained on Twitter posts. Qualitative analyses of the underlying linguistic and demographic differences reveal that users are significantly more likely to disclose information about their family, personal concerns, and emotions and provide a more 'honest' self-representation on Facebook. On the other hand, the same users significantly preferred to disclose their needs, drives, and ambitions on Twitter. The higher predictive performance of Facebook is also partly due to the greater volume of language on Facebook than Twitter – Facebook and Twitter are equally good at predicting user traits when the same-sized language samples are used to train language models. We explore the implications of these differences in cross-platform user trait prediction.

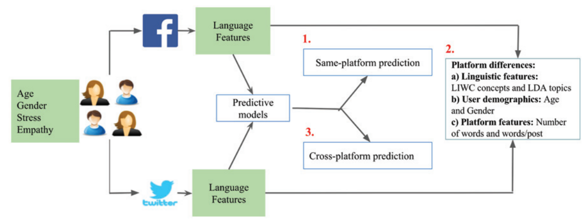


Figure 1: Using a set of users with active accounts on both Facebook and Twitter, we: (1.) compare the predictive performance of both platforms in modeling demographic (age & gender) and psychological traits (stress & empathy); (2.) examine the underlying linguistic, demographic and language sampling differences, and (3.) demonstrate the implications for cross-platform prediction.

Introduction

In the United States, 56% of the adults online use more than one social media platform.[1] Users are increasingly open to the choice of determining where they browse – or post – certain kinds of content. Surveys and studies of social media users in the United States indicate that they use Facebook to connect with friends and family, while they use Twitter to connect with personalities or topics of interest (Joinson 2008; Perrin 2015). In accordance with the platform's 'roles', users also assume different strategies when using or posting information on one of many social networking sites. While the literature on self-presentation in social situations spans decades (Goffman 1956), social media offers a new paradigm to study whether, and how, users disclose less or more about themselves on different social media platforms (Davenport et al. 2014; Lin and Qiu 2013).

This study uses the framework of Tajfel's Social Identity theory (Hogg 2016) to understand the personal and group-level factors that play a role in determining users' self-disclosure on social media platforms. Social media users negotiate between the need to *self-categorize* – for instance, customize user bios to suit particular social media platforms (Zhong et al. 2017) in order to be *socially desirable* (Edwards 1957), as well as the need to *associate* – for instance, self-identify as the member of a certain age group or community (Papacharissi 2002). We hypothesize that the influence of self-categorization and social desirability would be observed in the individual choices to post certain messages on Facebook versus Twitter, and the relationship of user traits with honest vs. positive self-presentation. The influence of group association would be reflected in intra-group similarities and inter-group differences. We are interested in the implications of *linguistic, demographic and platform differences* on self-disclosure.

We focus on Facebook and Twitter – two social media platforms which are used rather differently from one another. The research questions guiding this study are as follows:

- **Predicting user traits:** For the same set of users, how does their Facebook and Twitter language compare at predicting their demographic and psychological traits? (Table 1)
- **Facebook vs. Twitter:**
  - **Linguistic differences:** Which linguistic features predict user's posting preferences? (Tables 2 and 3)
  - **Demographic differences:** Which demographic features predict users' posting preferences? (Table 4 and Figure 2)

[1] http://www.pewinternet.org/2016/11/11/social-media-update-2016/

- **Language sampling differences:** How well do Facebook and Twitter language models perform, when they are trained on the same number of words? (Table 6)
- **Cross-platform prediction:** How well do language models trained on one platform, predict user traits on a different platform? (Table 7)

Our work flow is described in Figure 1 and is motivated by the first research question, wherein we observe that Facebook and Twitter language are not equally accurate at predicting two demographic traits - age and gender - and two psychological traits - empathy (Davis 1983) and stress (Cohen, Kessler, and Gordon 1997).

Background

A body of recent work in computational psychology (Brown 2016; Plonsky et al. 2017) has established that users' social media behavior can be used to reliably predict traits such as age and gender (Sap et al. 2014), personality (Schwartz et al. 2013; Guntuku et al. 2016; Jaidka et al. 2018b; Rieman et al. 2017; Abdul-Mageed et al. 2017), mental health factors such as ADHD (Guntuku et al. 2017b), depression (Guntuku et al. 2017c) and stress (Lin et al. 2014), as well as real-world outcomes such as protest participation (Ahmed, Jaidka, and Cho 2017) and election outcomes (Ahmed, Jaidka, and Skoric 2016; Jaidka et al. 2018a). Positive self-presentation (i.e., posting messages with positive emotions) on Facebook was found to be associated with higher subjective well-being, but honest self-presentation (i.e., posting messages with anger, sadness and anxiety) also had a significant indirect effect on well-being through perceived social support (Kim and Lee 2011). Researchers have recently started exploring the use of multiple social media platforms in profiling user behavior (Manikonda, Meduri, and Kambhampati 2016; Shu et al. 2017), but few studies have explored whether one platform is preferable to another for self-disclosure. The study by Davenport et. al (2014) compared the active usage of Twitter versus Facebook for a college student and an adult population, to suggest that college students on Twitter are more likely to manifest a narcissistic personality through their Twitter activity as compared to their Facebook activity, while no significant differences were reported for the adult population. In related work, Hughes et. al (2012) compared the personality correlates and age- and gender- preferences for the social and informational use (but not language use) of Facebook and Twitter. The authors reported that social media users may prefer to use Facebook for social interactions, but they use Twitter for information consumption.

Focusing on the linguistic differences, the analysis by Lin and Qiu (2013) of a set of Twitter posts by 100 Singaporean students and the Facebook posts of a different set of 100 students revealed that the Facebook statuses were more emotional and interpersonal, while the tweets were more casual, explicit and less negative. Our interest is in exploring the implications of these differences in predicting relatively immutable traits such as gender and age, or psychological traits such as empathy and stress, and we focus only on language use.

The technical affordances of social media platforms also influence user disclosure. On Twitter, users are required to express themselves in only 140 characters per post[2], while Facebook's limit on post length is 1000 words. Posts on Twitter are public by default, with an option to switch over to 'protected' mode to restrict readership of all posts ('tweets') to only their Twitter friends. On the other hand, on Facebook there are fine-grained controls per post which lets the user control which subset of people can view and comment on their posts (Debatin et al. 2009). Because of the way our data was collected, we have access to users' public tweets and all of their Facebook posts. We contribute to the previous understanding of self-disclosure on social media platforms, with an analysis of the effect of post length in promoting or hindering self-disclosure.

How well would Twitter language models perform on Facebook language? Researchers often have labeled data from one public platform – such as Twitter or a corpus of blog-posts – but they may want to make predictions based on language from another platform, such as Facebook, for which there may not be labeled data available. In such scenarios, we would like to know how well Twitter-trained models would perform on Facebook language and vice-versa.

Approach

To answer the first research question, we evaluate the performance of models trained with different linguistic features from either social media platform, to predict two demographic traits - age and gender - and two psychological traits - general empathetic concern (empathy) (Davis 1983) and stress (Cohen, Kessler, and Gordon 1997) (Table 1). Although empathy and stress can be modeled as both, states and traits, in this analysis we have used standard psychological scales that consider them as traits.

Next, we compare Facebook and Twitter in terms of the linguistic (Table 3) and demographic (Figure 2) features that create differences between them. We then evaluate the effect of language sampling size, by conducting an ablation analysis of Facebook language models, systematically reducing the language sample according to the number of words, posts and words per post available from Twitter (Table 6).

Finally, we explore the implications of these differences. We evaluate how well models trained on one platform generalize to the other platform (Table 7).

Data collection

The purpose of our data collection was to obtain the Twitter and Facebook language for a set of users who also answered a survey questionnaire. We recruited our subjects – adults in the United States – via Qualtrics [3], a platform for deploying surveys and recruiting participants. Our survey comprised demographic questions and items from standardized psychological scales in randomized order. We obtained participants' consent to access their Facebook status updates and/or Twitter user names, from which we collected their social media posts using the Facebook Graph API and the

[2]At the time of this study, Twitter's 280-character limit had not yet taken force.

[3]www.qualtrics.com/Survey-Software

Twitter API respectively. The study received approval from the Institutional Review Board of the University of Pennsylvania.

We consider those participants who successfully completed the survey, sharing their age, gender, stress as measured using the Cohen's 10-item Stress scale (Cohen, Kessler, and Gordon 1997), general empathetic concern as measured using the Interpersonal Reactivity Index (Davis 1983), and access to their active accounts on both Facebook and Twitter. To choose a threshold for participation, we select the minimum number of words posted by the 500 users who have the most posts on both platforms. This ensures a decent-sized user base and an acceptable language sample per user (Schwartz et al. 2013; Yarkoni 2010). In this manner, we identify 523 users with a minimum language sample of 1700 words per platform. All of the users were from the United States; 229 self-identified as female. The mean age of the sample is 36 (median age is 38), with 79% of the participants identifying as white, 8.3% as African American and 3.3% as Asian. Subjects reported a mean and a median income from all sources of between $20000 - $25000 US dollars. The demographic distribution was similar to the profile of other participants who had provided only a Facebook account.

Data preparation

We collected $647,862$ Facebook posts and $862,807$ Twitter posts, averaging 1238 Facebook and 1649 Twitter posts per subject. We represent language in posts as proportions of: (a) n-grams (words and phrases) (b) theory-driven lexica and (c) topics modeled on n-grams and word embeddings. These approaches and lexica are considered standard in the language analyses of social media posts (Sap et al. 2014; Schwartz et al. 2013; Preoţiuc-Pietro et al. 2015; Guntuku et al. 2017c; Jaidka et al. 2018b).

N-gram distributions: We tokenize Facebook and Twitter posts using the HappierFunTokenizer [4] to produce a total of 1.7 million Facebook tokens and 2.1 million Twitter tokens (uni-grams) which can be further grouped into bi-grams or tri-grams (or referred to more generally as n-grams). We use a bag-of-words representation to reduce each users' posting history to a normalized frequency distribution of their n-grams; i.e., for each person, we divide the count of each n-gram by the total number of n-grams they use. We also follow the same procedure on a balanced dataset of $10,000$ messages sampled from Facebook and Twitter.

Lexica and Part of Speech coverage: We calculate the user-level and message-level proportions of (a) parts of speech, (b) psycho-linguistic conceptual categories and (b) emotional words, using Linguistic Inquiry and Word Count (LIWC) 2015 (Pennebaker, Booth, and Francis 2007), which comprises a number of conceptual, emotional and part of speech categories created by psychologists, which have also been independently evaluated for their correlation with psychological concepts (Pennebaker, Booth, and Francis 2007).

Gross Happiness score (1 feature): We use LIWC categories to calculate the gross happiness score for each individual in terms of the standardized difference between their use of positive and negative words. According to Kramer (2010), this metric provides a way to compare the relative use of positive and negative expression and generate a metric that is independent of language and dictionary. It is calculated in terms of the average use of positive ($\mu_{user,pos}$) and negative ($\mu_{user,neg}$) emotion words by an individual and the meta-average ((μ)) and meta-sd ((σ)) of positive and negative emotion words over all individuals, as:

$$GH_{user} = \frac{\mu_{user,pos} - \mu_{pos}}{\sigma_{pos}} - \frac{\mu_{user,neg} - \mu_{neg}}{\sigma_{neg}} \quad (1)$$

LDA Topics (2000 features) Data-driven topics are expected to be more representative of social media posts, as compared to theory-based lexica. We use topics mined from a large Facebook corpus to represent every individual as well as all the language on either platform as a normalized frequency distribution of 2000 social-media specific topics. We obtain these topics from the DLATK code repository (Schwartz et al. 2017) and they are based on the Latent Dirichlet Allocation (LDA) of a corpus of approximately 18 million Facebook updates with alpha set to 0.30 to favor fewer topics per document. We represent each user (and platform) in terms of their probability of mentioning a topic, as:

$$usage(topic|user) = \sum_{word' \epsilon topic} p(topic|word) \times p(word, user) \quad (2)$$

Inherently, each topic is realized as a set of words with probabilities. Every individual is thus scored in terms of their probability of mentioning each of the 2000 topics, $(p(topic, user)$, which is derived from their probability of mentioning a word $(p(word|user))$ and the probability of the words being in given topics $(p(topic|word))$.

Word2Vec Topics (100 features): Clustered word embeddings improved the predictive power of language models for user trait prediction. We create 100 clusters of word embeddings, obtained from the skip-gram Word2Vec Twitter corpus (Mikolov et al. 2013) and clustered using the Gensim implementation in Python, and factorized using a word-context PMI matrix (Levy and Goldberg 2014). Finally, we depict each user in terms of their probability of mentioning each of these topics, following equation 2.

Facebook vs. Twitter: Predicting user traits

We evaluate how supervised models trained on Facebook and Twitter language perform on predicting user traits for a held-out set of users, as follows:

- We stratify our set of users into five folds with a uniform distribution of age and gender traits in each fold.
- We conduct a cross-validated weighted linear regression for the real-valued traits (age, empathy and stress), training on the (a) n-grams, (b) LIWC features and (c) LDA and Word2Vec topic features for users in four folds, and testing on the users in the held out fold.
- For predicting age, empathy and stress, we test several regularization methods such as ridge, elastic-net, LASSO and L2 penalized SVMs and obtained the best results using elastic-net regularization.

[4] https://github.com/dlatk/happierfuntokenizing

Platform (N=523)	Models	Training Size	Age (MAE)	Gender (Accuracy)	Empathy (Pearson's R)	Stress (Pearson's R)
Facebook	Baseline	523	10.06	0.54	0	0
	Sap et al. (Sap et al. 2014)	70000	8.82	0.76	-	-
	LIWC	523	7.20	0.87	0.18	0.27
	Word2Vec Topics	523	7.04	0.78	0.18	0.26
	LDA Topics	523	6.78	0.91	0.11	0.26
	N-grams	523	**5.71**	**0.78**	**0.20**	**0.27**
Twitter	Baseline	523	10.06	0.45	0	0
	Sap et al.(Sap et al. 2014)	70000	10.20	0.39	-	-
	LIWC	523	8.59	0.81	0.10	**0.21**
	Word2Vec Topics	523	8.40	0.72	0.07	0.10
	LDA Topics	523	8.58	0.80	-	0.17
	N-grams	523	**8.08**	**0.73**	**0.13**	0.15

Table 1: Prediction performance for age (Mean Absolute Error (MAE) in years), gender (Accuracy %), empathy and stress (Pearson's r) averaged over a five-fold stratified cross-validation across 523 users. We compare our model performance against the baseline results (mean age of the sample, and gender predicted as a random binary outcome)

- For predicting gender, we evaluate gradient boosted classifiers, random forest classifiers and support vector classifiers using the DLATK implementation of Python's sklearn library (Schwartz et al. 2017). The best predictive performance for gender was observed using random forest classifiers with 1000 trees.
- In the case of models trained on n-grams, we perform feature selection by retaining the most frequent 1000 uni-, bi- and tri-grams, and conducting feature selection to retain significantly correlated n-grams ($p < 0.05$) based on a uni-variate regression analysis. Finally, we use randomized principal component analysis (PCA) to avoid overfitting.

We evaluate the performance of language models for predicting age and gender by calculating the average Mean Absolute Error (MAE) and average accuracy (%), and by measuring the average Pearson's r for stress and empathy over held-out data in a five-fold cross-validation setting. Table 1 provides the results, compared against baselines of mean age of the sample for age and a random binary outcome for gender outcome. The Facebook and Twitter models trained with N-gram had the best performance overall. In general, age, empathy and stress are more accurately predicted using Facebook than Twitter language models, while they perform about the same in predicting gender, with Facebook language models showing a slight advantage. The results are comparable to the state-of-the-art age and gender language models provided by Sap et. al (2014). The results for predicting empathy from Facebook N-grams are also comparable to Abdul-Mageed et. al's (2017) prediction using gender and language, with a Pearson's r of 0.25. The results for stress cannot be compared to the findings reported by Lin et. al (2014) because in their model, stress was predicted at the message level as a binary class. In general, the Pearson correlations reported in this paper for language models predicting empathy and stress are weak – however, such results are typical of such studies that correlate language usage with psychological traits (Schwartz et al. 2013).

This preliminary analysis suggests that users' language on Facebook offers more self-disclosure than Twitter about their demographic and psychological traits. The difference is least pronounced in the case of gender, and the most prominent in the case of age. In the following sections, we will explore some of the factors which drive these differences in self-disclosure.

Facebook vs. Twitter: Linguistic differences

We explore the role of linguistic features in determining whether a post is posted on Facebook or Twitter. We formulate a binary classification problem on a balanced random sample of 20,000 posts, labeled according to whether the post was obtained from Facebook or Twitter.

We report performance in Table 2, in terms of the accuracy in predicting the label for held-out observations in a ten-fold cross-validation using the logistic regression implementation in Python's sklearn package. We use feature selection to first discard those features which are not significantly related in a univariate regression, and then perform randomized principle component analysis to avoid overfitting. Among the N-grams, we find that the use of links, @-mentions and hashtags are highly distinctive of Twitter's vocabulary and contribute to the superiority of the N-gram based model. In Table 3, we identify the weights of some of the top statistically significant predictors in the LIWC-based classifier, as well as example words from the corpus. LIWC categories with a inter-correlation greater than 0.65 amongst each other were discarded before the model was trained. Rows are shaded according to whether the association is significantly linked to Facebook (rows shaded blue) or Twitter (rows shaded red).

Table 3 suggests that users prefer to use Facebook for posts mentioning their personal relationships (family, friends and home), people (male and female), personal concerns (religion and work). Overall, users are more likely to post messages on Facebook that express positive or negative emotion.

Twitter posts reflect more mentions of the self through the use of the first person pronoun, as well as a greater tendency to engage the readers through the use of second person pronouns, than on Facebook. A surprising finding was

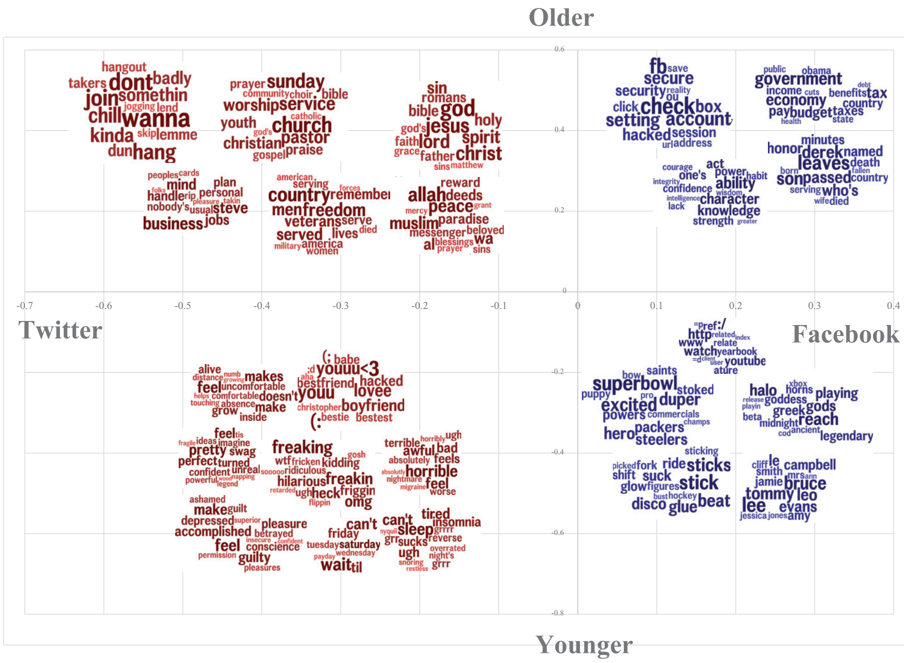


Figure 2: Correlations of words/phrases with Social Media platform (Facebook on the positive x-axis and Twitter on the negative x-axis) and age on the y-axis. Correlations are significant, Simes corrected at p<0.01

Model	Accuracy
Random	0.50
Word2Vec Topics	0.69
LDA Topics	0.73
LIWC	0.75
N-grams	0.83
LIWC + N-grams	**0.84**

Table 2: Accuracy of different logistic regression models in predicting the choice of platform for social media posts on held-out data in a ten-fold cross-validation

that users preferred Twitter to post about their psychological needs and drives – such as their need for achievements and rewards (driven by succeeding, achieving goals and conquering challenges) and their need for affiliation (motivated by close relationships with others) (Jordan and Pennebaker 2015). Their Twitter posts indicate cognitive processes – words used to work things out, for e.g. by the use of comparing and differentiating words, and certain and tentative words. On the whole, users appear to prefer Twitter to 'think aloud' on topics related to their goals and drives.

Facebook vs. Twitter: Demographic differences

We now investigate the demographic differences which lead to different self-disclosure on Facebook and Twitter. First, in Table 4 we identified that a negative correlation with age signifies that younger users in our dataset were significantly more likely to post more on Facebook than on Twitter ($p < 0.05$); this is in line with previous research on the demographics of social media users (Duggan and Brenner 2013).

LIWC Category	Estimate x 10^{-1}	Example words from the data
Family	0.86^{***}	family, baby, mom, son, dad
Male	0.79^{***}	he, his, him, man, son, he's
Female	0.78^{***}	her, she, mom, girl, she's
Negative emotion	0.77^{***}	bad, miss, hate, :(, sick
Positive emotion	0.67^{***}	love, good, happy, :), great, lol
Home	0.60^{***}	home, family, house, bed, room
Religion	0.59^{***}	god, hell, pray, holy, christ
Work	0.41^{***}	work, school, working, class
Friend	0.28^{*}	friends, friend, dear
Risk	-1.02^{***}	bad, stop, wrong, safe, lose
Assent	$-.90^{***}$	awesome, yes, ok, cool
Sexual	-0.87^{***}	fuck
Achieve	-0.76^{***}	work, win, first, best
Netspeak	-0.71^{***}	rt, :), lol, u, gonna, :(, gotta
Reward	-0.64^{***}	good, great, win, excited, won
2nd person pronouns	-0.57^{***}	you, your, u, you're, ya
1st person singular	-0.50^{***}	i, my, me, i'm, im, i've, myself
Affiliation	-0.48^{***}	we, love, our, friend, us, help, share
Differ	-0.44^{***}	not, but, if, or, really
Certainty	-0.42^{***}	all, never, ever, always, every
Discrepancy	-0.32^{***}	want, would, need, could
Compare	-0.21^{***}	like, more, better, best
Tentative	-0.16^{*}	if, some, any, hope, may, trying

Table 3: Summary of the LIWC-based classifier from Table 2, showing the top most predictive LIWC categories for Facebook (blue) and Twitter (red). ***: p¡< 0.001 level; *: p¡< 0.05

Individuals with higher stress are significantly less likely to post on Facebook. Stress has a stronger correlation with negative emotion (r= .21) and (lesser) Gross Happiness (r= -.21) on Facebook than on Twitter (r=.14 for negative emotion and r = -.12 for Gross Happiness; all p <.05).

We now explore the most distinguishing topics for

User trait	No. of Facebook posts	No. of Twitter posts
Age	$-.10^*$	$-.00$
Gender	$-.02$	$-.01$
Empathy	.02	.02
Stress	$-.10^*$	$-.05$

Table 4: Pearson correlations of user traits with Facebook- and Twitter-posting behavior. $*$: p $<$ 0.05

continuously-valued age on Facebook vs. Twitter. LDA topics offer richer insights than LIWC features. We also found that topic features were better able to represent the language of younger users than LIWC topics. We plot a subset of LDA topic features which are significantly correlated with (a) Age and (b) presence on Twitter or Facebook:

- We conduct a Bonferroni-corrected Pearson correlation of users' topic features on Facebook and their age. This helps us identify which topics are significantly likely to be discussed by younger or older users.
- We conduct a Bonferroni-corrected Pearson correlation of user's topic features, against a binary-valued variable representing whether the source was Facebook or Twitter. This helps us identify which topics are significantly more likely to be found on which platform.
- We plot each of these topics in terms of their correlations with platform use and age on the X- and Y- axis respectively. We depict each topic as a word cloud of its most prevalent words.

Figure 2 illustrates the top few significantly correlated topics (p $<$ 0.05), which are plotted in terms of their Pearson correlation with Facebook (blue topics) and Twitter (red topics), and with the age on the Y-axis. Positions on the plot are adjusted to avoid overlap. On Facebook (bottom right quadrant, -.30$<$ r $<$-.45), younger users talk about gaming, the Superbowl and hockey. Younger users are also more likely to tag their friends in their Facebook posts than older users. On Twitter (bottom left quadrant, -.30$<$ r $<$-.40), younger users are more likely to talk about anxiety, betrayal, mentions of best friends, slang, ideas and accomplishments.

Older users on Facebook (top right quadrant, .12$<$ r $<$.30) discuss governance, voting and philosophy. On Twitter (top left quadrant, .10$<$ r $<$.20), they are more likely to discuss militarization, religion and their jobs. Note that a lot of these results are rather different from Table 3, where posts about religion and jobs likely end up on Facebook. Words depicting slang, emoticons and net-speak was significantly characteristic of younger users on Facebook: a finding that corroborates a language analysis of 70000 Facebook users in the myPersonality dataset (Schwartz et al. 2013).

Facebook vs. Twitter: Language sampling differences

The previous sections identify the different linguistic and demographic differences driving self-disclosure on Facebook and Twitter. Next, we explore whether platform features also affect self-disclosure. In order to assess the importance of sample size, we conduct an ablation analysis by evaluating model performance after randomly reducing the number of posts, words and words-per-post available per user for Facebook, in order to make it match the Twitter language sample. We describe three experiments to test the importance of the language sample size:

- **(WL)** Word-level experiments: To compare the performance of Facebook and Twitter models on the same number of words.
- **(PL)** Post-level experiments: To compare the performance of Facebook and Twitter models on the same number of posts.
- **(PW)** Post-and Word-level experiments: To compare the performance of Facebook and Twitter models on the same number of words per post.
- We experiment with different sample sizes of posts and words: m = 25, 50, 75, 100, 200, etc. until m = 1700, the maximum number of words available for all users.

Table 5 describes our sampling method in more detail. Once these corpora were constructed, we used the same method as described earlier in Section 4 to train and test language models for age, gender, empathy and stress, again using five-fold cross-validation[5].

Results from word- and post-limiting experiments

The results from the post- and post-and-word sampling experiments with different sample lengths are provided in Figure 3a for age and Figure 3b for gender. The X-axes depicts the (log of the) word count threshold which was used to create the post- and post-and-word-level sampling. Performance for age prediction is provided as MAE (lower is better) and for gender prediction is provided as AUC (higher is better). There were no significant differences among the results across the five folds. For age, Figure 3a suggests that if the same number of posts are available from Facebook and Twitter for a user, then predicting age from Facebook language would be significantly more accurate than from Twitter language. However if the same number of words *and* posts are available, then Facebook and Twitter are about at par with each other for predicting age. The results are repeated in the case of gender (Figure 3b), where we see that if the same number of posts are available from Facebook and Twitter, then Facebook language would yield a more accurate prediction of gender than Twitter language; however, for the same-sized language samples Twitter and Facebook are at par for predicting a user's age and gender.

Table 6 presents the performance results only for N = 1700 words for all four traits across the three experiments. In all cases, limiting the number of posts and/or words causes a big drop in the predictive performance of Facebook models as compared to the results reported in Table 1.The first two rows reflect that if the same number of posts by a user are available from both Facebook and Twitter, then in general, Facebook significantly out-performs Twitter in predicting user demographics (p $<$ 0.01). The next pair of rows

[5]The code used to construct the language samples for ablation analysis and to generate the plots is available at https://github.com/kj2013

Strategy	Objective	Description
WL: Word-level sampling	To compare the performance of Facebook and Twitter models on the same number of words	Randomly sample Twitter and Facebook posts without replacement, for a cutoff of cumulative words per user.
PL: Post-level sampling	To compare the performance of Facebook and Twitter models on the same number of posts	Randomly sample Facebook posts without replacement, for the same number of posts per user as in the Twitter WL corpus.
PW: Post-and Word-level sampling	To compare the performance of Facebook and Twitter models on the same number of words per post	Randomly sample without replacement, to retain the same number of n-grams per post as any one post for that user in the Twitter PL corpus.

Table 5: Ablation analysis: the sampling methods used on the Facebook language samples, in the three experimental conditions.

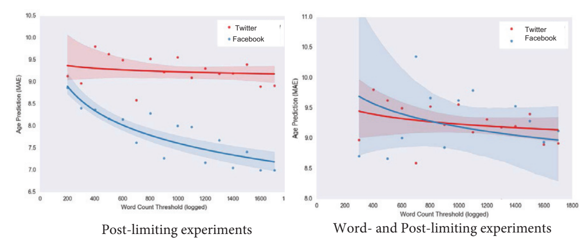
	Platform	**Age**	**Gender**	**Empathy**	**Stress**
Sampling strategy	**Evaluation metric**	**MAE**	**Accuracy**	**Pearson's R**	**Pearson's R**
Post-limiting experiments	Facebook	**7.00***	**0.78***	0.11	**0.11**
	Twitter	8.92	0.69	0.10	0.08
Word-limiting experiments	Facebook	**8.81**	**0.78***	0.11	0.09
	Twitter	8.92	0.69	0.10	0.08
Word- and post-limiting experiments	Facebook	9.12	0.63	0.10	0.09
	Twitter	**8.92**	**0.69***	0.10	0.08

Table 6: Results from the ablation analysis: predictive performance from the post-, word- and post-and-word-limited experiments. * denotes p <0.01. There were no significant differences among the results across the five folds.

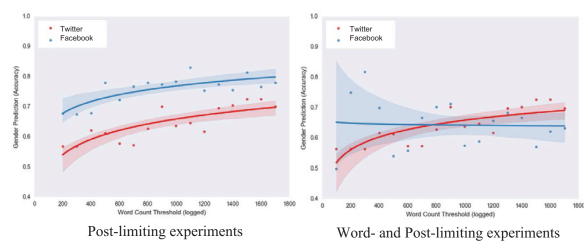
reflects that if the same number of words are available for a user from Facebook and Twitter, Facebook only significantly outperforms Twitter for predicting gender. In the final pair of rows, we see that Twitter outperforms Facebook in predicting age and gender in limited-message situations. In general, the Table suggests that each post in Facebook gives more information about the user's demographic traits than each corresponding post on Twitter. However, if the number of posts in either language sample are the same, then the prediction of psychological traits appears to be unaffected by further manipulations of post-length.

Cross-platform prediction performance

We want to explore if platform differences have implications in the cross-platform prediction of user traits. Table 7 depicts the results from testing on a held out sample of Facebook and Twitter, with the first column indicating performance when no domain transfer is involved, and the second column depicting performance when the model is trained on a different platform. The biggest drop in performance with the Facebook test set is for predicting empathy and stress with a domain-transferred model. On the other hand, Facebook models do reasonably well on a Twitter test set, with the largest drop being a one year increase in the error on age estimates. The results illustrate the challenges in predicting psychological traits from social media, which are further exacerbated when a domain transfer is involved. Depending on the problem being addressed, domain transfer may hinder predictive performance with varying degrees of severity.



(a) Age (Mean Absolute Error)



(b) Gender (Accuracy)

Figure 3: Shown is the decrease in (a) Mean Absolute Error for predicting age, and (b) the increase in accuracy for predicting gender, to the point where Facebook (blue trendlines) and Twitter (red trendlines) language models have comparable performance for the same number of words-per-post. In each graph, the X axis provides the logged limits to the number of 1-grams randomly sampled from each user. The results are provided for (i) Post-limiting sampling (ii) Post- and Word-limiting Sampling. For $N < 200$, the results were often too unstable across the five fold validation, to be included. All images are not to the same scale.

Trained on:	FB	Twitter	Twitter	FB
Tested on:	Facebook		Twitter	
Age (MAE)	**5.71**	7.83	**8.08**	9.03
Gender (Accuracy)	**0.78**	0.79	**0.73**	0.63
Empathy (Pearson's r)	**0.20**	0.07	**0.13**	0.11
Stress (Pearson's r)	**0.27**	0.11	**0.15**	0.13

Table 7: Cross-platform results for age, gender, stress and empathy. The results are from the best performing models of Table 1 in a five-fold cross-validation setting, trained on the source domain and tested on the held-out sample (FB: Facebook).

Discussion

While both Facebook and Twitter are platforms for social networking, they are differently used for self-disclosure and information sharing. Facebook accounts are usually based on real identities, and Facebook friends typically comprises existing friends and acquaintances from users' real lives (Lampe, Ellison, and Steinfield 2008). On Facebook, users identify the things that matter the most to them, such as their home and their emotional connections to their family and friends. They also project positive emotions and excitement. Together, these characteristics suggest that on Facebook, users appear to be driven by a high need for belonging. Need for belonging refers to the fundamental drive to form and maintain relationships (Baumeister and Leary 1995) and is manifested as a pro-social behavior characterized by high positive emotion and high Facebook use (Seidman 2013). Users with higher stress are more likely to express negative emotion on Facebook. In previous work, it has been suggested that users receive more social support on social media when they are able to communicate their needs through self-disclosure, facilitated by honest self-presentation (Kim and Lee 2011). These findings reiterate the previous point that users prefer to use Facebook as a means of emotional and social support.

Twitter does not require any identity information. Previous studies have found that social connections on Twitter tend to be more open and comprise more strangers than those on Facebook (Lin and Qiu 2013). Twitter audiences can thus be thought of as a weak-tied (Granovetter 1973), loosely-knot community (Gruzd, Wellman, and Takhteyev 2011) of people who share the same interests, and discuss information with each other (Bakshy et al. 2012). Because users feel more comfortable sharing their ambitions and goals on Twitter, we infer that they experience a similarity bias in perceiving their Twitter audience as a like-minded community of peers (Holtz and Miller 1985) who share similar backgrounds, attitudes and goals. Their participatory behavior on Twitter resembles a collaborative discussion, with their taking of a stance through the use of words referencing their personal and social identity, engaging others through the use of interpersonal markers and assent words, and working out ideas together through the use of tentative and comparative language (Niculae and Danescu-Niculescu-Mizil 2016). While younger users are more open to discussing negative emotions and anxiety on Twitter, older users are more likely to discuss contentious issues.

In terms of the implications for machine learning, our findings suggest that the size of the language sample can greatly impact the ability of trained models to predict self-disclosure. In general, language models trained on Twitter may not be easily transferable to Facebook. The same users wield a qualitatively and emotionally richer vocabulary on Facebook than Twitter in terms of grammar, diversity of linguistic concepts and emotion words. However, most studies which model language on social media have disregarded platform differences when using a pre-trained model on a new data source.

Conclusion

The focus of this analysis is on comparing self-disclosure on Facebook vs. Twitter, looking beneath the many methods to predict user traits on language and into the more fundamental exploration of the linguistic, demographic and language sampling differences in a within-group comparison. We collect and process social media data for the same users for both platforms, which enables us to perform a comparative analysis under a proper scientific setup. We find that our set of users do prefer to self-disclose more on Facebook than on Twitter, but the reasons for this are both linguistic and quantitative. Consequently, it is challenging to generalize findings of user behavior from one social media platform to all others: a caveat that is worth noting for scholars planning other tasks which use crowd-sourcing platforms for recruiting subjects and annotators.

Social media users with multiple social media accounts may strategize their self-disclosure on each platform to suit their self- and social identity, to the extent that their language on one platform does not cohere with their language on another. As per our analysis, these users are not different from an average social media user with a single social media account: a paired t-test between the average proportion of LIWC features of our subjects and the excluded participants who shared access to only a Facebook account, revealed no significant differences in their language (p = 0.93). Users in our dataset are more honest with negative emotion expression (Kim and Lee 2011; Edwards 1957), and more open about their need for belonging (Baumeister and Leary 1995) on Facebook than on Twitter. On the other hand, they appear to consider Twitter a community of their peers with similar ambitions and problems (Holtz and Miller 1985), with older people are more likely to use Twitter to mention contentious issues.

A comparison of sample differences suggests that platform affordances do play a role in determining users' self-disclosure behavior. The results suggest that a major factor underlying Facebook's superiority over Twitter is the lack of restrictions on the lengths of posts, which give people more space to express themselves.

We recommend that cross-platform differences can be overcome by using domain adaptation methods to adjust pre-trained models on new social media platforms. Although we have focused on language – in future work, it would be interesting to conduct a cross-platform comparison that also considers the photographs (Guntuku et al. 2017a), (Guntuku, Roy, and Weisi 2015), list of interests (Kosinski, Stillwell, and Graepel 2013) and other cues shared by users as a part of their social media profiles. We did not have information about the devices used to post messages; however, since we found no difference in the diurnal posting behavior on Facebook and Twitter, it is reasonable to expect that users can access either platform on their device of choice.

Our study contributes to the literature with a new understanding of the implications of the choice of social media platform and the size of language samples used for training and testing predictive models. An underlying assumption is that the survey participants were honest while answering the surveys. At the very least, the results suggest that socially desirable behavior observed through surveys is the most similar to the socially desirable behavior manifested by the same users on Facebook than on Twitter. A self-selection bias implies that the users surveyed are not representative of the general American population; however we have at least validated that the sub-sample that was analyzed here is similar to the overall population of survey participants in our study. We conclude with a few observations and recommendations:

- For researchers in computational social science who are looking to work across platforms, the results suggest that one-size-fits-all NLP approaches should be used with caution.
- Social media platforms have significantly different language characteristics. For predictive tasks, supervised domain adaptation is recommended, especially when a language model is trained on Twitter, and for predicting behavioral traits such as empathy and stress (Mejova and Srinivasan 2012).
- The language sample per user should comprise at least 200 words per user in order to have stable predictive performance.
- Sufficiently-sized language samples from Twitter can defuse the performance gap between Facebook- and Twitter-based predictive models. Accordingly, we recommend focusing Twitter language analyses on 'heavy users' of Twitter (Krishnamurthy, Gill, and Arlitt 2008).

Acknowledgements: The authors acknowledge the support from Templeton Religion Trust, grant TRT-0048.

References

Abdul-Mageed, M.; Buffone, A.; Peng, H.; Eichstaedt, J. C.; and Ungar, L. H. 2017. Recognizing pathogenic empathy in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM, 448–451.

Ahmed, S.; Jaidka, K.; and Cho, J. 2017. Tweeting indias nirbhaya protest: a study of emotional dynamics in an online social movement. *Social Movement Studies* 16(4):447–465.

Ahmed, S.; Jaidka, K.; and Skoric, M. M. 2016. Tweets and votes: A four-country comparison of volumetric and sentiment analysis approaches. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM, 507–510.

Bakshy, E.; Rosenn, I.; Marlow, C.; and Adamic, L. 2012. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, 519–528. ACM.

Baumeister, R. F., and Leary, M. R. 1995. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin* 117(3):497.

Brown, S. M. 2016. Machine learning methods for computational psychology. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI, 4291–4292. AAAI Press.

Cohen, S.; Kessler, R. C.; and Gordon, L. U. 1997. *Measuring stress: A guide for health and social scientists*. Oxford University Press.

Davenport, S. W.; Bergman, S. M.; Bergman, J. Z.; and Fearrington, M. E. 2014. Twitter versus facebook: Exploring the role of narcissism in the motives and usage of different social media platforms. *Computers in Human Behavior* 32:212–220.

Davis, M. H. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology* 44(1):113–126.

Debatin, B.; Lovejoy, J. P.; Horn, A.-K.; and Hughes, B. N. 2009. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication* 15(1):83–108.

Duggan, M., and Brenner, J. 2013. *The demographics of social media users, 2012*, volume 14. Pew Research Center's Internet & American Life Project Washington, DC.

Edwards, A. L. 1957. The social desirability variable in personality assessment and research.

Goffman, E. 1956. The presentation of self. *Life as theater: A dramaturgical sourcebook*.

Granovetter, M. S. 1973. The strength of weak ties. *American journal of sociology* 78(6):1360–1380.

Gruzd, A.; Wellman, B.; and Takhteyev, Y. 2011. Imagining twitter as an imagined community. *American Behavioral Scientist* 55(10):1294–1318.

Guntuku, S. C.; Zhou, J. T.; Roy, S.; Weisi, L.; and Tsang, I. W. 2016. Who likes what, and why? insights into personality modeling based on imagelikes'. *IEEE Transactions on Affective Computing*.

Guntuku, S. C.; Lin, W.; Carpenter, J.; Ng, W. K.; Ungar, L. H.; and Preoţiuc-Pietro, D. 2017a. Studying personality through the content of posted and liked images on twitter. In *Proceedings of the 2017 ACM on web science conference*, 223–227. ACM.

Guntuku, S. C.; Ramsay, J. R.; Merchant, R. M.; and Ungar, L. H. 2017b. Language of adhd in adults on social media. *Journal of attention disorders* 1087054717738083.

Guntuku, S. C.; Yaden, D. B.; Kern, M. L.; Ungar, L. H.; and Eichstaedt, J. C. 2017c. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* 18:43–49.

Guntuku, S. C.; Roy, S.; and Weisi, L. 2015. Evaluating visual and textual features for predicting user likes. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, 1–6. IEEE.

Hogg, M. A. 2016. Social identity theory. In *Understanding Peace and Conflict Through Social Identity Theory*. Springer. 3–17.

Holtz, R., and Miller, N. 1985. Assumed similarity and opinion certainty. *Journal of Personality and Social Psychology* 48(4):890.

Hughes, D. J.; Rowe, M.; Batey, M.; and Lee, A. 2012. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. *Computers in Human Behavior* 28(2):561–569.

Jaidka, K.; Ahmed, S.; Skoric, M.; and Hilbert, M. 2018a. Predicting elections from social media: A three-country, three-method comparative study. *Asian Journal of Communication* (Online first).

Jaidka, K.; Buffone, A.; Eichstaedt, J.; Rouhizadeh, M.; and Ungar, L. 2018b. Modeling and visualizing locus of control with facebook language. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Joinson, A. N. 2008. Looking at, looking up or keeping up with people?: motives and use of facebook. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1027–1036. ACM.

Jordan, K. N., and Pennebaker, J. W. 2015. Seeking power, finding friends, and making the grade: Drive orientations of the primary frontrunners. [Online; posted 16-November-2015].

Kim, J., and Lee, J.-E. R. 2011. The facebook paths to happiness: Effects of the number of facebook friends and self-presentation on subjective well-being. *CyberPsychology, behavior, and social networking* 14(6):359–364.

Kosinski, M.; Stillwell, D.; and 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.

Kramer, A. D. 2010. An unobtrusive behavioral model of gross national happiness. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 287–290. ACM.

Krishnamurthy, B.; Gill, P.; and Arlitt, M. 2008. A few chirps about twitter. In *Proceedings of the first workshop on Online social networks*, 19–24. ACM.

Lampe, C.; Ellison, N. B.; and Steinfield, C. 2008. Changes in use and perception of facebook. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 721–730. ACM.

Levy, O., and Goldberg, Y. 2014. Dependency-based word embeddings. In *ACL (2)*, 302–308.

Lin, H., and Qiu, L. 2013. Two sites, two voices: Linguistic differences between facebook status updates and tweets. In *International Conference on Cross-Cultural Design*, 432–440. Springer.

Lin, H.; Jia, J.; Guo, Q.; Xue, Y.; Li, Q.; Huang, J.; Cai, L.; and Feng, L. 2014. User-level psychological stress detection from social media using deep neural network. In *ACM MM*.

Manikonda, L.; Meduri, V. V.; and Kambhampati, S. 2016. Tweeting the mind and instagramming the heart: Exploring differentiated content sharing on social media.

Mejova, Y., and Srinivasan, P. 2012. Crossing media streams with sentiment: Domain adaptation in blogs, reviews and twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Niculae, V., and Danescu-Niculescu-Mizil, C. 2016. Conversational markers of constructive discussions. In *Proceedings of NAACL-HLT*, 568–578.

Papacharissi, Z. 2002. The presentation of self in virtual life: Characteristics of personal home pages. *Journalism & Mass Communication Quarterly* 79(3):643–660.

Pennebaker, J. W.; Booth, R. J.; and Francis, M. E. 2007. Linguistic inquiry and word count: Liwc [computer software]. *Austin, TX: liwc. net*.

Perrin, A. 2015. Social media usage. *Pew Research Center*.

Plonsky, O.; Erev, I.; Hazan, T.; and Tennenholtz, M. 2017. Psychological forest: Predicting human behavior. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, AAAI, 656–662. AAAI Press.

Preoţiuc-Pietro, D.; Volkova, S.; Lampos, V.; Bachrach, Y.; and Aletras, N. 2015. Studying user income through language, behaviour and affect in social media. *PloS one* 10(9):e0138717.

Rieman, D.; Jaidka, K.; Schwartz, H. A.; and Ungar, L. 2017. Domain adaptation from user-level facebook models to county-level twitter predictions. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, 764–773.

Sap, M.; Park, G.; Eichstaedt, J. C.; Kern, M. L.; Stillwell, D.; Kosinski, M.; Ungar, L. H.; and Schwartz, H. A. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 11461151. Citeseer.

Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Dziurzynski, L.; Ramones, S. M.; Agrawal, M.; Shah, A.; Kosinski, M.; Stillwell, D.; Seligman, M. E.; et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8(9):e73791.

Schwartz, H. A.; Giorgi, S.; Sap, M.; Crutchley, P.; Ungar, L.; and Eichstaedt, J. 2017. Dlatk: Differential language analysis toolkit. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 55–60.

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.

Shu, K.; Wang, S.; Tang, J.; Zafarani, R.; and Liu, H. 2017. User identity linkage across online social networks: A review. *ACM SIGKDD Explorations Newsletter* 18(2):5–17.

Yarkoni, T. 2010. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality* 44(3):363–373.

Zhong, C.; Chan, H.-w.; Karamshu, D.; Lee, D.; and Sastry, N. 2017. Wearing many (social) hats: How different are your different social network personae?