Modeling Self-Disclosure in Social Networking Sites

Yi-Chia Wang
Language Technologies Institute
Carnegie Mellon University
yichiaw@cmu.edu

Moira Burke
Facebook
mburke@fb.com

Robert Kraut
Human-Computer Interaction Institute
Carnegie Mellon University
robert.kraut@cmu.edu

ABSTRACT

Social networking sites (SNSs) offer users a platform to build and maintain social connections. Understanding when people feel comfortable sharing information about themselves on SNSs is critical to a good user experience, because self-disclosure helps maintain friendships and increase relationship closeness. This observational research develops a machine learning model to measure self-disclosure in SNSs and uses it to understand the contexts where it is higher or lower. Features include emotional valence, social distance between the poster and people mentioned in the post, the language similarity between the post and the community and post topic. To validate the model and advance our understanding about online self-disclosure, we applied it to de-identified, aggregated status updates from Facebook users. Results show that women self-disclose more than men. People with a stronger desire to manage impressions self-disclose less. Network size is negatively associated with self-disclosure, while tie strength and network density are positively associated.

Author Keywords

Social networking sites; Facebook; computer-mediated communication; self-disclosure; personality; audience; context collapse; natural language analysis; applied machine learning.

ACM Classification Keywords

H.5.3. Information Interfaces and Presentation: Group and Organization Interfaces: Evaluation/methodology, Web-based interaction.

INTRODUCTION

When people communicate with others in person or online they share information about themselves that helps others understand who they really are. Self-disclosure is the "act of revealing personal information to others" [3]. Several theories of computer-mediated communication suggest that verbal self-disclosure will be more important and common online than offline because people online are more

anonymous and cannot display thoughts and feelings via non-verbal cues [see Table 1 in 37]. Empirically, people disclose significantly more in computer-mediated communication interactions than in offline ones [28, 51]. However, a recent review suggests that the difference between online and offline self-disclosure is conditioned on a number of factors, including personality, context and the relationship between communication partners [37].

Greater levels of online self-disclosure can be important both for individuals who communicate online and the sites that host their communication. A substantial body of research in both offline and online settings demonstrates the importance of self-disclosure in the formation and maintenance of personal relationships. For example, sharing important parts of our lives improves our relationships [39] and causes others to like us [16]. Experimental research shows that greater self-disclosure leads to greater liking of a conversational partner, feelings of closeness and enjoyment of the conversation [50]. Online self-disclosure is positively associated with intimacy among Facebook friends [40].

Self-disclosure also has implications for the success of social networking sites. Since relationship maintenance is a primary motivation for many people in using social networking sites and because self-disclosure both reflects and enhances social relationships, people are likely to be more satisfied with sites that encourage self-disclosure [49]. Interface elements on these sites influence how much people reveal about themselves. For example, between 2005 and 2014 Facebook increased the number of fields included in users' profiles [1]. They also introduced interface elements such as the privacy dinosaur, which encouraged users to undergo a privacy checkup and become aware of the audiences that could see the information they post [2].

However, online self-disclosure can also have negative effects, most commonly when people share information to a wider audience than they had intended. For example, roughly 37% of companies use social networking sites to research job applicants [27]. Friend networks that encompass multiple social circles can make self-disclosure challenging, as well.

Given the importance of online self-disclosure, an automated measure of self-disclosure in SNSs that can be applied at scale could be very useful for social scientists attempting to understand the conditions that encourage or discourage self-disclosure, for members of social network

sites as the basis for feedback about whether their content is revealing more or less about themselves than they desire, and for service providers to track how changes to the design of their sites influence users' self-disclosure.

We present a novel machine learning model to measure self-disclosure at scale in social network sites. The model is accurate, agreeing with the judgments of trained coders (r=.60). Its features - message length, use of positive and negative emotional vocabulary, mentions of close social ties, use of non-normative language and discussion of more intimate topics - were derived from theory about the nature of personal self-disclosure.

To demonstrate the validity of this measure as well as to advance understanding of online self-disclosure, we applied it to two data sets containing almost nine million de-identified Facebook status updates and show results that are consistent with prior empirical research from laboratory and survey studies on self-disclosure. Status updates exhibited higher self-disclosure if the authors score lower on a self-reported scale measuring the trait of impression management, if they are women rather than men and if their networks of Facebook friends are smaller, denser and of higher average tie strength. All research was observational; no Facebook user's experience was any different from usual as a result of this study.

MEASURING SELF-DISCLOSURE IN LANGUAGE

Self-disclosure in both face-to-face conversations and computer-mediated communication has most often been measured with communicators' retrospective self-reports or by manually coding participants' conversations or posts [e.g., 7, 16, 28, 36, 37]. One goal of our research is to develop an automated approach to assess self-disclosure that can help investigate how people's self-disclosure in online social networking sites differs based on the network structure they experience. Neither participants' retrospective self-reports nor human coding is scalable for examining large archives of members' conversations produced in SNSs. Several studies have demonstrated that it is possible to construct automatic self-disclosure text analyzers [4, 5, 6, 52]. However, the models and approaches proposed in these studies were either domain-specific [6, 52] or provided no ground truth against which to evaluate their accuracy [4]. Although the self-disclosure classifier constructed by Balani and Choudhury [6] had an accuracy of 78%, it was built using over a thousand features. Therefore, it is difficult for researchers to interpret why these features predict self-disclosure. Our research seeks to build a supervised machine learning model that can approximate human judgments about whether people are revealing personal information about themselves in their online posts. To be successful, the model should be accurate, parsimonious (i.e., using a small number of features of the texts), interpretable, and domain-independent.

Building and validating a machine learning model involved three major steps, which we describe in more detail below. Human judges hand-coded the extent of self-disclosure in a sample of 2,000 posts provided by social media users to the researchers with informed consent. Their judgments are both the training data and the "ground truth" for evaluating the accuracy of the machine learning estimates. Second, we represented the posts as a set of linguistic features to be input to the machine learning algorithms. Finally, we constructed statistical machine learning models from the hand-coded data and then evaluated the accuracy of the models.

Data collection and agreement analysis of coded data

In this section, we describe how we operationalized the judgments of self-disclosure and collected 2,000 Facebook status updates with self-disclosure annotations from both the posters (Facebook users recruited from Amazon Mechanical Turk) and from trained judges. After that, we analyzed the agreement between the self-disclosure scores of posters and external judges.

Self-disclosure instrument

Many self-report questionnaires measure self-disclosure, including the Jourard Self-Disclosure Questionnaire [29], the Miller et al. Self-Disclosure Index [36], and Rust's Impression Management scale [45], but most of them conceptualize self-disclosure as a stable personal disposition to reveal personal information. Recently, however, Barak and Gluck-Ofri established a 3-item rating scale to assess self-disclosure in online forum messages, evaluating the degree to which a post exposed the author's personal information, thoughts, and feelings [7].

In order to have people assess the degree of intimacy in their own posts, we adapted definitions and questions from the Barak and Gluck-Ofri Self-Disclosure Rating Scale [7] and the Miller et al. Self-Disclosure Index [36]. We selected and modified questions so that they can be used to measure posters' self-disclosure in a single post. In particular, we conducted pilot studies on Amazon Mechanical Turk (MTurk) to determine the appropriate set of questions and modifications (The detail settings for the MTurk task are described in the next section.) To reduce respondent burden, our goal was to create a reliable scale with only five items. In each round of pilot studies, respondents were asked to enter the text of one of their Facebook status updates and answer several questions using a 7-point Likert scale, ranging from 1 ("not at all") to 7 ("completely") (e.g., "To what extent does this post involve your feelings and emotions, including concerns, frustrations, happiness, sadness, anger, and so on?"). Previous research on self-disclosure has used coarser scales (e.g., a 3-point scale from [7]), but we employed a 7-point scale for greater variance for later model training.) After several rounds of testing with larger sets of items, we created a situational self-disclosure scale based on the five questions listed in Table 1. The composite value of the

To what extent does this post involve	
A.	personal information about yourself [the poster] or people close to you [him/her], such as accomplishments, family, or problems you are [the poster is] having?
B.	personal thoughts on past events, future plans, appearance, health, wishful ideas, etc.?
C.	your [the poster's] feelings and emotions, including concerns, frustrations, happiness, sadness, anger, and so on?
D.	what is important to you [the poster] in life?
E.	your [the poster's] close relationships with other people?

Table 1. Self-disclosure measurement items for Facebook status updates as rated by posters [or research assistants].

	Sample status update	Poster	RA
1	It was so warm out on Saturday... why is there snow everywhere now? :C	1.8	1.6
2	Well, I got into the University of [omitted] MA program. More than 200 applicants, 24 spots. #1 program in the country. so there's my brag and I think I've accomplished enough for this year so can I just play animal crossing or sims for a few weeks, thank you.	5.6	5.2
3	There are few things I cherish more on this Earth more than leftover spaghetti.	5.2	1.7

Table 2. Examples of status updates and average disclosure ratings by posters and research assistants.

answers to these five questions represents the self-disclosure level in a specific update. The scale is reliable, with a Cronbach's alpha of 0.72.

Collecting Self-Disclosure Ratings
In order to construct a dataset of Facebook status updates with hand-coded self-disclosure annotations while at the same time honoring users' privacy and Facebook's terms of service, we recruited active Facebook users from Amazon Mechanical Turk (Turkers) and paid them $0.50 US to contribute and rate their most recent Facebook status update in terms of the degree of self-disclosure it contained. Amazon's Mechanical Turk (https://www.mturk.com) is an online marketplace for crowdsourcing. It allows requesters to post jobs and workers to choose jobs to perform. Jobs are known as Human Intelligence Tasks (HITs).

Our HIT selected only workers from the United States who had 98% or more of their previous submissions accepted. Workers were shown an informed consent document in which they were notified that research assistants would be reading their status updates later. Our university's Institutional Review Board (IRB) approved this task. To ensure that participants were active Facebook users, they were asked to answer questions about their Facebook experience, including "How many days in the past week did you use Facebook?", "How many friends do you have on Facebook?", and "How many photos do you have on Facebook?" Then participants were asked to copy and paste their most recent English pure-text status update. They also rated their post for each of the five questions in Table 1 regarding the degree of intimacy they revealed in it. Table 2 shows some examples of the updates contributed by posters and their composite ratings of self-disclosure. Although many of Turkers' ratings seemed plausible, some seemed highly idiosyncratic. For example, the two individuals who contributed Example 2 and 3 both evaluated them has having substantial self-disclosure (greater than 5 on the 7-point scale), but most experts would consider an announcement about getting into a desired and competitive education program and pride revealed in Example 2 to be

more self-disclosing than the light-hearted statement about leftover spaghetti in Example 3.

Agreement between posters and external judges
Our goal for collecting Facebook status updates and posters' self-disclosure assessments was to build an accurate machine learning model that could be used for examining self-disclosure on social networking sites. Individual differences in Turkers' diligence in attending to the judgment task, the way they interpreted the self-disclosure questions, and the way they used a 7-point scale could lead to noisy training data, which would prevent us from developing an accurate self-disclosure model. This problem of noisy data is compounded because each of the 2,000 status update messages was evaluated by only a single, unique poster.

Furthermore, posters can only describe their intent, but not how an external audience would interpret and evaluate their posts. Indeed, people are poor at judging how others will interpret their online communication [31]. External judges can act as proxies for intended readers or audiences of posts. Thus to reduce noise in the training data and to capture audience judgments, we supplemented posters' judgments of the self-disclosure in their posts with judgments of trained external judges.

We recruited four research assistants (RAs, 1 male and 3 females), with diverse backgrounds from a research-oriented university. They were instructed to rate each update using the same five items in Table 1 from an audience's point of view. That is, they were asked to imagine the poster's intent, by answering questions in which the word "you" was replaced with phrase "the poster." The four RAs initially coded a common set of 50 posts, and met to discuss and resolve their disagreements until reaching a consensus for each of the 50 posts. The average correlation of their ratings was 0.79 before discussion, which increased to 0.82 after discussion. After the training, the four RAs annotated the rest of the 2,000 posts. Each status update was evaluated by at least two RAs. The "outsider's" judgment of a post was then computed by averaging the scores of the RAs who rated the

post. The mean and median of the annotations was 2.52 and 2.12, respectively, and the standard deviation was 1.28.

The last column in Table 2 presents the ratings of RAs for the three examples. Across the 2,000 messages, posters and outsiders agreed moderately on the degree of self-disclosure the posters displayed in their messages (r=.60), which suggested that outsiders or audiences could perceive posters' self-disclosure intent to a reasonable extent. Given this finding and the observation of noisy posters' annotations, we decided to build our machine predictive model based on RAs' annotations to ensure the consistency and quality of the model.

Machine Learning Model of Self-disclosure

To automatically measure self-disclosure we introduced five linguistic features which we believe are key ingredients, including post length, emotional valence, the presence of certain topics, social distance between the poster and a person mentioned in the post, and how well the content of a post fits into social norms. The output was a numerical value representing the degree of self-disclosure in a post. In the following section we explain the rationale and extraction process for each feature.

Text Processing and Feature Extraction

Post length is the number of words in a post. One component of self-disclosure is the amount of detailed personal information one reveals. Revealing more details about oneself requires writing more text rather than less. Thus, we expected that longer posts would be more revealing than shorter posts.

Positive emotion and negative emotion. According to the self-disclosure instrument in Table 1, revealing emotions and feelings is considered self-disclosing behavior. Moreover, Wang et al. have demonstrated that emotion words can predict emotional self-disclosure [52]. We defined positive emotion and negative emotion features as the frequency of positive and negative tokens in a post. A token was considered positive / negative one if it was found in the positive / negative emotion dictionaries of the Linguistic Inquiry and Word Count program (LIWC) or matched positive / negative emoticons from Wikipedia (http://en.wikipedia.org/wiki/List_of_emoticons). The lists of positive emotion emoticons included smiley (e.g., :-) and :}), laugh (e.g., :D and =D), playful (e.g., :P and xp), and wink (e.g., *) and ;]); the negative ones are sadness (e.g., :(and :c), crying (:'-(and :'( ), angry (e.g., :-‖ and :@), and disgust (e.g., D8 and v.v).

Social distance. Talking about close relationships is a signal of self-disclosure and was included in the self-disclosure scale (see the fifth item in Table 1). Consider the following examples:

a1. My husband can't give up cigarettes.

a2. President Obama can't give up cigarettes.

Both have the same topic (someone's bad habits.) However, it is obvious that (*a1*) discloses more personal information about the author and her circumstances than does (*a2*), since it refers to is the author's husband with whom she presumably has a closer relationship than she does to the president. This example suggests that the social distance between a poster and people mentioned in a post is an important self-disclosure indicator. In addition, prior studies have shown that count of first-person words (e.g., "I," "my," and "myself") can be an effective indicator of self-disclosure in both offline and online communication [18, 28]. Thus, we propose a novel feature measuring the average social distance between posters and all the target(s) they refer to in the post. In contrast to work by Derlaga and Berg [18] and Joinson [28] which only used count of first-person words, we considered all types of person references. The idea is that post authors have an imaginary distance between themselves and each of the people referenced in the post, an estimate of the degree to which they participate in each other's lives.

The feature extraction process involved three steps. The first step was to identify and extract all the people mentioned in a post in our corpus of 2,000 labeled status updates. Person references include singular and plural first-person pronouns (e.g., "I," "me," "our,"), intimate nicknames (e.g., "babe," "darling," "honey"), various types of family (e.g., "husband," "daughter") and friends (e.g., "buddy," "friend"), as well as named entities (e.g., "Harry Potter," "Michael Jackson," "Barack Obama"). Second-person and third-person words were not included because it was impossible to infer the social distance between a post author and second-person or third-person word without knowing its antecedent. Moreover, the person nouns for which they were substituted would have been taken into account when we searched for all people mentions. Except for named entities, all other people words were extracted using a dictionary-based approach, since they comprised a limited set of words. Specifically, we utilized the first-person singular, first-person plural, family, and friend dictionaries in LIWC, and manually created an intimate nickname dictionary.

Several of the steps required baseline text to understand the prevalence of named entities, phrases, and topics across Facebook, and so we selected a random sample of 8,011,980 English Facebook status updates posted between November 2013 and October 2014, a full year to capture all regular events and holidays (the "one-year dataset"). All posts in the dataset were de-identified and analyzed in aggregate on Facebook's servers in accordance with Facebook's data use policy; models were built from counts of terms. No text was viewed by researchers except for the authors' own status updates in order to validate the data processing procedures described below. No Facebook user's experience was changed by this data analysis.

The second step was to identify named entities (NEs) and distinguish private ones from public ones. While a private name was defined as a person whom the author of a post knew, a public name referred to a celebrity, such as a singer or politician. The person-name entity recognizer in the OpenNLP toolkit was applied to find all named entities in status updates. In order to differentiate private names from public ones, we introduced a semi-automatic approach to construct a celebrity name list from the one-year dataset. We first used the person-name recognizer to extract all named entities in the dataset, and then discarded those that occurred fewer than five times. This automatic process resulted in 9,629 unique entities. However, since the name recognizer was not 100% accurate, there were wrongly identified entities in the list, such as "Be Safe," "Merry Christmas," "God Bless." So, we manually pruned the name list, which resulted in a list with 8,434 unique person names. This final list was our celebrity dictionary. Examples include "Robin Williams," "Peter Pan," and "Steve Jobs." A named entity would be categorized as public if it was found in the celebrity dictionary; otherwise, it would be classified as private.

The last step was to calculate a social distance feature for each of the 2000 updates. The feature was the average distance between a poster and each of people referred to in the post. We put people references into one of four categories, and assigned each a relative social distance score of 0, 1, 2, or 3 based on the likelihood the person participated in the poster's life. Those who were more likely to be involved in the poster's life would be assigned a shorter distance score, with 0 representing the poster and 3 representing members of the public. Although we used weights of 0 to 3 to represent social distance, any monotonic coding would produce similar results, as long as psychologically closer people were assigned lower weights. Formally, the social distance of a status update s, $social\_distance(s)$, was defined as following:

$$social\_distance(s) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} distance(p_i), N > 0 \\ 3, N = 0 \end{cases}$$

$$distance(p) = \begin{cases} 0, p \in \{\text{LIWC\_I}\} \\ 1, p \in \{\text{LIWC\_we, LIWC\_family, DIC\_nickname}\} \\ 2, p \in \{\text{LIWC\_friend, NE\_private}\} \\ 3, p \in \{\text{NE\_public}\} \end{cases}$$

where $P = \{p_1, p_2, ..., p_N\}$ denoted the set of people referenced in s; $social\_distance(s)$ was the arithmetic mean of $distance(p_i) \ \forall p_i \in P$ when P was a non-empty set, otherwise it was set to 3. $distance(p)$ was a case function that returned a value indicating the pseudo social distance between the author of s and the people mentioned, p, according to its category. The function returned a distance of 0 when p belonged to the LIWC "I" dictionary, since first-person singular words referred to the author herself / himself. It assigned a distance of 1 when p was family (LIWC_family), or someone close enough to the

author so that she / he used first-person plural words (LIWC_we) to indicate they did something together or used an intimate nickname (DIC_nickname) to refer to the person. Though personal pronouns may indicate other psychological phenomena (such as distancing with the "royal we") [43], we expect these uses to wash out at scale and contribute a small amount of noise. Moving a bit further away from the social circle of the author were people whom the poster knew but was not so familiar with, including general friends (LIWC_friend) and private named entities (NE_private), which got an assignment of 2. The last type of people references was celebrities (NE_public). We assumed most posters do not know celebrities personally, so the function returned a distance of 3 when p was found in the celebrity list.

Social normality. Text is less self-revealing when people are saying what everyone else is saying than saying something unique. We quantified social normality as the difference between the language of a status update and the language of the Facebook community as a whole. Specifically, we built a statistical language model representing the linguistic usage of the community, and then calculated the cross-entropy of the update using the Facebook language model. A statistical language model is a probability distribution trained over word sequences (i.e., a corpus) which can be used to assess the probability of an order of words occurring in the corpus [14]. Cross-entropy is a measurement often used in natural language processing applications to evaluate how well a language model predicts a test word sequence. In other words, it can be used to gauge whether one's post fits into a corpus. For instance, Danescu-Niculescu-Mizil and his colleagues compared users' posts in an online community with all the posts in the communities to argue that members of a community adapt to its norms over time [17]. We adopted a similar approach.

In detail, we first constructed a language model estimated from the de-identified year-long corpus. It was a bigram (or word pair) language model with Good-Turing smoothing [23] built using CMU-Cambridge Statistical Language Modeling Toolkit [15]. Refer to [14] for more details about n-gram language models and smoothing techniques.) This language model represented the social norms of the Facebook community, which meant it characterized how the general Facebook community would expect Facebook users to present themselves. Given a status update s, we computed its social normality based on the bigram language model $LM_{Facebook}$ as shown below:

$$social\_normality(s) = -H(s, LM_{Facebook}) = \frac{1}{N}\sum_{i=1}^{N} log P_{LM_{Facebook}}(b_i)$$

where $H(s, LM_{Facebook})$ was the cross-entropy of s under the $LM_{Facebook}$; s was composed of bigrams $(b_1, b_2, ... b_N)$; $P_{LM_{Facebook}}(b_i)$ denoted the probability of the bigram b_i in $LM_{Facebook}$. A status update with a lower social normality value suggested its language looked less similar to the

language on Facebook, which we believed meant it would contain more self-disclosure.

Topic features. Different topics are often associated with different degrees of self-disclosure. Some topics, like physical appearance or work, could potentially contain more personal details than other topics, such as weather and sports. To identify the topics common in status updates, we used Latent Dirichlet Allocation (LDA). LDA is a statistical generative method that can be used to discover hidden topics in documents as well as the words associated with each topic [10]. It analyzes large numbers of unlabeled documents by clustering words that frequently co-occur and have similar meaning into "topics."

Before training our LDA model, we went through several steps to pre-process and clean the data. Our experience suggests that this pre-processing and pruning result in far superior topic models than those from unpruned data. Status updates were segmented into sentences and then tokenized with the Apache OpenNLP library [38], stemmed with the Porter stemmer [44], and lowercased. We removed punctuations and replaced URLs, email addresses, and numbers with tags. Updates were then represented as an unordered set of unigrams (single words) and bi-grams (word pairs).

Across all terms in the de-identified eight million status updates, 83.24% of unigrams appeared only once, and 1000 unigrams accounted for 29.17% of all text. This skew of words is a well-known phenomenon in natural language known as Zipf's law [54]. Therefore, we pruned high- and low-frequency unigrams (those that occurred in more than 0.5% or less than 0.01% of the updates) and bigrams (those that occurred in less than 0.015% of the updates) to reduce noise and vocabulary size. In addition, we excluded all unigrams from a 500-word stopword list (e.g. "the", "and", "in"); bigrams were filtered if both words were stopwords. After pruning, 63.31% of the status updates had fewer than eight n-grams; these documents were too short for successful model training. Therefore, we built topic models from the remaining status updates (N= 2,939,357).

To identify topics in status updates, we built an LDA model treating each status update as a document. The model was set to derive 80 latent topics; this setting produced models that were more interpretable to human judges than models deriving 50, 60, 70, 80, 100, or 120 topics. Topic dictionaries were generated from the 500 terms most strongly associated with each topic, and two experts familiar with SNS content manually named each dictionary. Examples of topics derived from the LDA analysis include Sports (e.g., "football", "player", "score"), Medical (e.g., "doctor", "hospital", "blood"), Food (e.g., "cook", "coffee", "chicken") and Christianity (e.g., "heaven", "christ", "the lord"). See Table 4 for additional examples. Each LDA topical feature calculates the frequency of words in a message matching its corresponding dictionary.

Model Construction and Evaluation

The purpose of our evaluation was to contrast the performance of the machine learning models built using our proposed features with a feature set consisting of unigrams and bigrams, which is frequently used as a baseline for model evaluation. In order to assess the contribution of each proposed feature, we first evaluated them separately and then in combination. Details of our results are below.

Given the input feature representation of a status update, we built machine learning regression models which output a numerical value indicating the degree of self-disclosure in it. The dataset was the 2,000 status updates collected from MTurk workers and annotated by RAs. We used the sequential minimal optimization (SMO) algorithm for support vector machine regression [47] implemented as the SMOreg procedure in Weka [53], a machine learning toolkit, to build the regression models. We used the default linear kernel with all other parameters also set to defaults. The dataset was randomly split into partitions for 10-fold cross-validation. We chose 10-fold cross-validation over leave-one-out cross-validation because they are similar in terms of the size of data points used for training (1,900 versus 1,999), and 10-fold cross-validation is much more time-effective than leave-one-out. We report accuracy in terms of the average Pearson correlation across the 10 folds between the RA-coded ratings and predicted self-disclosure.

Table 3 presents the accuracy results. For a baseline model, we stemmed the raw text, removed stop words, and kept unigrams and bigrams occurring five or more times as features. It had a correlation of .47 (Model 1), but required 814 features. The interesting results were achieved by more parsimonious models using post length (Model 2), positive/negative emotion (Model 3), social distance (Model 4) or the social normality feature (Model 5). Although their performance (as indicated by the correlations of .37, .39, .31 and .17, respectively) were worse than the baseline model, when we built a model using the five features together (Model 6), we got a correlation of .48, which was better than the baseline. One surprising result to note here is that while the social normality feature can predict self-disclosure with a correlation of .17, it is a positive predictor rather than a negative one as we hypothesized. This suggests that a status update using language similar to the Facebook community was considered to be slightly higher in self-disclosure. One explanation for this is that there is a small positive norm of self-disclosure on the site.

Furthermore, the model built with 80 topic features (Model 7) achieved a correlation of .57, substantially better than the baseline. These moderate accuracy correlations are convincing evidence for our assumption that the concepts of post length, positive/negative emotion, social distance, social normality, and topics are essential components and indicators of self-disclosure. To understand the topics most

	Feature set	Num. of features	Corr.
1	Baseline (unigrams + bigrams)	814	0.47
2	Post length	1	0.37
3	Positive/negative emotion	2	0.39
4	Social distance	1	0.31
5	Social normality	1	0.17
6	Post length + positive/negative emotion + social distance + social normality	5	0.48
7	Topics	80	0.57
8	Topics + social distance + social normality	82	0.59
9	Post length + positive/negative emotion + topics + social distance + social normality	85	0.60

Table 3. Evaluation results with alternative feature sets.

Topic	Sample vocabulary	Regression weight
Christianity	shall, christ, spirit, the lord, of god	-0.70
Birthday	love you, happy birthday, my baby	0.51
Family Relationship	husband, wife, my mom, marry, my dad, the best, my daughter, in law	0.50
Politics	country, nation, american, govern	0.33
Deep Thoughts	the world, human, earth, create, key, purpose, soul, inspire, life	-0.32
School	student, write, teacher, test, grade	0.29
Weekend plan	wait for, n't wait, relax, spent, time with, so excited, this weekend, yay	0.26
Memorial	miss, angel, rip, heaven, pass away	0.24
Names	mary, smith, jack, jame, johnson	0.23
Medical	doctor, hospital, blood, leg, surgery	0.23

Table 4. Top 10 ranked topic features and their corresponding sample vocabulary in the model trained with 80 topic features.

relevant to self-disclosure, we further examined the top 10 ranked topics in Model 7. The results are shown in Table 4. We found that, for example, topics like Politics and Memorial were positive indicators of self-disclosure, while Christianity and Deep Thoughts were negative signals. It was noteworthy that there were two high-ranked topics (Family Relationships and Names) that overlapped with the information used in the social distance feature. This might be the reason why adding social distance and social normality features with the topic 80 features resulted in only a small gain in accuracy over the topic model by itself (Model 8). Lastly, we built a model combining all the proposed features (Model 9), which achieved the highest correlation among all the experiments (.60). Given the adequate validity of the last model, we then applied it to detect self-disclosure for all the status updates in the later analyses.

FACTORS RELATED TO SELF-DISCLOSURE

To test the validity of this automated measurement as well as reexamine and advance our understanding about online self-disclosure, we used it to replicate empirical patterns found in previous experimental and survey research on self-disclosure or suggested by network structure theory. We focused on individual differences among the posters and audience factors that might affect self-disclosure.

Poster Characteristics Influencing Self-disclosure

Personality: Impression Management

Goffman says in *The Presentation of Self in Everyday Life*, "When an individual appears in the presence of others, there will usually be some reason for him to mobilize his activity so that it will convey an impression to others which it is in his interests to convey" [22]. This phenomenon is known as *self-presentation* and sometimes called *impression management* [22, 46], which refers to the process through which people try to control the images others form about them. Impression management is generally thought of as the inverse of self-disclosure, by controlling the personal information one reveals. Researchers have developed self-report impression management scales to measure this concept as a stable personality trait, such as the Self-Monitoring scale [48], the

Balanced Inventory of Desirable Responding [42], and Rust's Impression Management scale [45]. These scales were used to assess individuals' desire for managing the impressions they make on others and appearing socially acceptable. Example items in these scales include "There are times when it's not sensible to tell the truth," "Like most people I sometimes blame my mistakes on others," and "I have some pretty awful habits" (reversed). These items suggest that people with a stronger desire to manage impressions would tend to hide the truth about themselves from others if they believe it hurts their images. We expect this tendency would influence how much they are willing to disclose to others, especially in a wide-audience environment such as Facebook status updates. The impression management model proposed by Leary and Kowalski [32] suggests that self-disclosure can endanger people's impressions of the discloser, since it involves the revelation of one's internal world, which usually consists of personal information or emotions that are socially awkward or morally questionable. Thus, we hypothesize below:

H1: Individuals with a stronger desire for impression management will self-disclose less.

Gender

It is both a cultural stereotype in the United States and an empirical reality that women self-disclose more than men. A meta-analysis involving over 23,000 people in 205 studies found that women on average were more self-revealing than men (d=.18) [19]. Women were self-disclosed more when demands for positive self-presentations were lower, including when talking to other women (d=.35) rather than men (d=.00) and when talking to friends (d=.28), spouses (d=.22) or parents (d=.25) rather than strangers (d=.07). In this study, we reexamine the following hypothesis:

H2: Women will self-disclose more than men.

Audience Factors Influencing Self-disclosure

Audience structure can affect language usage during social interactions [25]. SNS offers users a unified platform to

build and maintain social connections [33, 41], which can serve as a lens for understanding and comparing how individuals adjust their self-disclosure according to various audience factors.

Public vs. Private Communication

Communication through SNSs can be distinguished based on how directed and public the interaction is [9, 12]. Whereas directedness measures whether the target of the communication is a particular friend, publicness measures the possibility that an individual's behavior might be seen by others and the number of others who might see it. Facebook status updates are undirected, since they are typically published to the entire social network of a poster and not targeted at any specific person. However, the degree of publicness of individuals' status updates depends on the number of friends they have. Having more friends implies that updates are more "public." Since self-disclosure involves revelation of personal details which are considered private and people have less control of who sees their status updates when they have more friends, we hypothesize that:

H3: Network size will be negatively correlated with self-disclosure.

Closeness to Communication Targets

Empirical studies of dyadic relationships show that people reveal less personal details to acquaintances than to close friends [16]. We expect to find a similar result when considering one's social network as a whole. People with a greater number of strong ties in their networks should be more comfortable disclosing:

H4: Average tie strength will be positively correlated with self-disclosure.

Context Collapse on Social Networking Sites

Much of what we know about self-disclosure comes from studies of dyads [e.g., 16, 39]; we know less about when people self-disclose to wider audiences of multiple friends, such as on social network sites. These sites allow people to share with others from many parts of their life at once, a phenomenon known as *context collapse* [11, 33]. Context collapse may cause people to self-disclose less, because they would feel uncomfortable sharing intimate information appropriate for family and friends with relative strangers in their networks. That is, they might self-censor and only present information appropriate to the lowest common denominator [26].

Context collapse online makes impression management challenging [33]. People have to meet the expectations and interests of many different audiences. Given that people's networks are comprised of both weak and strong ties, they may self-disclose less as their networks become more diverse. Network density is the interconnections among the ties in one's social network. Network density is likely to signal the degree of context collapse. Higher network density suggests that friends are more connected and thus

have fewer disconnected clusters. Therefore we hypothesize that:

H5: Network density will be positively correlated with self-disclosure.

PREDICTING SELF-DISCLOSURE IN STATUS UPDATES FROM PERSONAL AND AUDIENCE CHARACTERISTICS

In this section, we examined the relationships of self-disclosure with poster characteristics and audience factors. Unlike prior research studying these relationships at the dyadic or message level, we examined them at the personal network level by, for instance, averaging self-disclosure of all of a person's status updates and considering the average tie strength that person had with all of her / his Facebook friends. We did this because we were interested in Facebook status updates, which are not targeted at any specific person and potentially visible to all Facebook friends of a user.

Poster Characteristics and Self-disclosure

To test the relationship of the self-reported trait of impression management (H1) and gender (H2) with self-disclosure, we utilized the dataset from the myPersonality project (http://mypersonality.org/). The myPersonality project, founded by Kosinski et al. [30], uses a Facebook app to collect anonymized data from Facebook users, such as their profile information and social network statistics, and combines them with personality scores measured by questionnaires. Specifically, the dataset contains users' status updates as well as their demographic information and self-report impression management scores, based on Rust's Impression Management scale [45]. We applied our machine learning model to measure self-disclosure in the users' updates, computed an average self-disclosure value for each user based on all her / his updates, and compared the average values with the corresponding self-report impression management scores. Analyzing the data from 2,878 users, we found a correlation of -0.19 (n=2,878, p<0.0001), which showed a negative relationship between self-reported desire for impression management and self-disclosure, and thus confirmed Hypothesis 1. We also calculated the correlation of users' gender (1 for male and 0 for female) with their average self-disclosure values, and it was -0.23 (n=153,726, p<0.0001), confirming Hypothesis 2, that women self-disclose more than men.

Audience Factors and Self-disclosure

The automatic self-disclosure model was applied to a new dataset of all of the posts written by a random sample of 412,470 English language Facebook active users for approximately one month in late 2014. All data was again de-identified and analyzed in aggregate on Facebook's servers; no text was viewed by researchers and no user's experience on the site was changed. We collected demographic information as control variables, including gender, age and the number of days they logged into Facebook in the past month. While gender was a binary variable with one (1) indicating male and zero (0) for

female, the other two were continuous, numeric variables. We also included a snapshot of their social network size and structure at the beginning of the data collection period.

Dependent Variable
- Self-disclosure: We computed a self-disclosure score for each user by averaging the machine-coded self-disclosure values of all their status updates.

Independent Variables
- Social network size: The number of friends a user had in the beginning of the data collection period.

- Social tie strength: We estimated tie strength between each user in the sample and all of his or her friends, using counts of communication frequency and other dyad-level variables, substantively identical to the techniques described in Gilbert and Karahalios [21] and Burke and Kraut [13].

- Social network density: This variable was the number of friendship connections among a user's friends. We normalized this measure by the total number of possible links among friends a user had, so that it corresponded to the portion of the possible connections within a user's friend network that were actually connected.

Except for the binary variable Male, all the numerical control and independent variables were standardized and centered, with a mean of zero and standard deviation of one. Additionally, we took the log of the variable Network size before it was standardized, since it had a skewed distribution. Table 5 reports the descriptive statistics for the variables used in regression models before standardization.

	Mean	Median	S.D.	Min	Max
Age	35.65	32	14.24	14	114
Number of logins	26.40	28	4.26	0	28
Network size	492.76	329	558.19	0	4,968
Tie strength	0.32	0.31	0.05	0.07	1
Network density	0.09	0.07	0.06	0	1
Self-disclosure	2.50	2.31	.83	1	7

Table 5. Descriptive statistics for the variables in the regression analyses.

When do people self-disclose more?
Table 6 presents five linear regression models predicting self-disclosure. Model 1 reports the effects of the control variables. In the rest of the models, we tested hypotheses regarding social network features and self-disclosure. Because network size is correlated with network density (r=-0.32) and average tie strength (r=-0.53), we first tested the effects of the three network variables separately in Models 2, 3, and 4. We then analyzed their effects together in a single model (Model 5). The intercept in the models represents a woman with all numerical variables at their means, who would disclose at a level of 2.595 on a 1 to 7 scale. Betas represent the effect on self-disclosure from a binary variable having a value of 1, or a one standard deviation increase in continuous independent variables. We also reported R-squared values in Table 6. Although the values are small, the outcome we were predicting (i.e., self-disclosure in one's language) is relatively subtle.

Model 1 shows that males revealed significantly less about themselves in their status updates than females (2.319 versus 2.595). Older posters disclosed more than younger people. However, the significant negative beta for number of logins suggests that the more active someone is on Facebook, the less he or she self-discloses. In Model 2, we found that when controlling for demographic information and activity level of these users, their social network size negatively predicted their self-disclosing behavior. The self-disclosure level decreased 0.01 point for users who had one standard deviation more friends, which confirmed Hypothesis 3. We investigated the effect of average social tie strength in Model 3. The result demonstrates that the closer individuals were to their friends, the more they self-disclosed in status updates. This finding confirmed Hypothesis 4. Model 4 tests Hypothesis 5 and shows a positive correlation between social network density and self-disclosure. That is, in one's social network, when there were more friends who were also friends with each other, that person would be more willing to share her / his personal details. In the last model, we examined the simultaneous effects of the three network variables on self-disclosure. While the effects of average tie strength and

DV: Self-disclosure	Model 1		Model 2		Model 3		Model 4		Model 5	
Explanatory Variable	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.
Male	-.276***	.003	-.275***	.003	-.267***	.003	-.276***	.003	-.267***	.003
Age[1]	.100***	.001	.097***	.001	.091***	.001	.100***	.001	.093***	.001
Number of logins[1]	-.043***	.001	-.042***	.001	-.045***	.001	-.043***	.001	-.046***	.001
Network size[2]	-	-	-.010***	.001	-	-	-	-	.007***	.002
Average tie strength[1]	-	-	-	-	.030***	.001	-	-	.033***	.002
Network density[1]	-	-	-	-	-	-	.004**	.001	.003*	.001
(Intercept)	2.595***	.002	2.595***	.002	2.592***	.002	2.596***	.002	2.592***	.002
R²	0.0429		0.0430		0.0440		0.0429		0.0441	
Number of observations					412,398					

1: standardized and centered. 2: Logged (base 10), standardized, centered.
* p<0.05, **p<0.01, ***p<0.001

Table 6. Results of the regression analyses.

network density were similar to those in Model 3 and 4, the effect direction of network size changed from negative to positive. Possible explanations for this surprising result will be discussed in the next section.

In order to test whether adding the three audience factors as predictor variables results in a statistically significant improvement in the fit of Model 1, we conducted the likelihood ratio test to evaluate the difference between Model 1 and each of the four audience models. The tests show that Model 2 ($p<0.0001$), Model 3 ($p<0.0001$), Model 4 ($p<0.001$), and Model 5 ($p<0.0001$) all fit significantly better to the data than Model 1.

DISCUSSION
In this paper, we developed a supervised machine learning model to detect the degree of self-disclosure in status updates and then used the model to replicate patterns from previous empirical work and theory. Through the process of building the model, we demonstrated that message length, emotional valence, the presence of certain topics, social distance between a poster and people mentioned in a post, and how well the content of a post fits into social norms were important constituents of self-disclosure. We showed that women self-disclose more than men, and people who more strongly desire to manage the impressions they make on others self-disclose less. We then demonstrated that social network size was negatively associated with self-disclosure, while network density and average tie strength had positive correlations with self-disclosure. Most of the results are consistent with those found or suggested by prior literature, which validate the effectiveness of the machine learning model we proposed.

One unexpected result in our analyses is that the estimates of the effects of network size in Model 2 and 5 have different signs. Although network size correlates with tie strength and network density, we confirmed that multi-collinearity is not a problem, with the all variance inflation factors less than 1.7. The result may be substantive, rather than methodological, challenging our assumptions about the meaning of the network variables and how they affect self-disclosure. While we hypothesized that a larger network size would lead to less self-disclosure because it makes communication more public, it may be that people believe that posting to larger networks exposes messages to weaker ties. Even though network size was designed to measure publicness, it grows by disproportionately adding weaker ties into the network [35]. This size and average tie strength are intrinsically lined. As a result, when average tie strength is held constant in Model 5, adding more people to the network seems to lead to an increase in self-disclosure.

This study not only replicates empirical patterns found in previous research but also extends the existing literature in both social sciences and linguistics. It advances our knowledge of how people self-disclose and maintain relationships in SNS by utilizing machine learning to analyze a large archive of online communication text. Most

early research on self-presentation or self-disclosure in online environments focuses on dyadic contexts and online dating sites. For example, some scholars have investigated how online dating participants manage their profile presentations to draw the attention of potential dates [20, 24]. In recent years, there has been an increasing number of studies exploring how people perform to their entire social network, not just potential dates [e.g., 8, 9, 34, 40]. Self-presentation to one's social network differs from self-presentation to potential dates. Online dating services target the development of romantic relationships, typically among dyads of roughly the same age. In contrast, online social networking services support people as they present themselves to a variety of partners with various types of social relationships [41].

Moreover, our research may have better generalizability than past research on online self-disclosure [e.g., 9, 40, 49], since it was based on a diverse, large sample of online communication. As a sensitivity test, we replicated the analyses reported here on de-identified, aggregated posts from Facebook users in Australia and Singapore and discovered similar results. Second, with the automatic self-disclosure model introduced in this paper, we will be able to develop and test more theories regarding online self-disclosure in the future.

The findings in this study also have practical implications for improving user experiences in the social web. If designers of these sites know how users of social networking sites navigate multiple audiences to manage impressions, they can improve their services by providing better affordances to users. For example, when network size and diversity become large enough that a person might not feel comfortable sharing personal news with friends, the site might nudge that person to share to a smaller group or a custom list of friends.

Limitations and Future Directions
Our current findings are based on a static view of the relationship between audience network structure and self-disclosure. We can only make correlational claims, not causal ones. One possible future direction is to perform a controlled experiment in a lab, making participants' online network size and diversity more or less salient and examining how that affects their willingness to self-disclose. Another future direction is to analyze audiences' responses to posters, so that we will have a better understanding of how audiences perceive and react to self-disclosure and whether they interpret the self-presenter's messages in the same way that the self-presenter intended. Are posts that are higher in self-disclosure perceived as higher quality by friends of the poster? Or are other post features more important? The answers to these questions would help site designers understand the degree to which context collapse affects the quality of post inventory.

Selection bias in the Amazon Mechanical Turk sample may also weaken the model. We know little about workers who

chose not to participate in the study or how representative our sample is. By virtue of their online employment, these workers may be more technologically savvy or spend more time on Facebook, and thus their self-disclosure behaviors and perceptions may be different from people who use Facebook less often. Furthermore, workers were asked to select their most recent post but may not have. Future work should gather ratings from a more representative sample.

Although our self-disclosure machine learning model performs reasonably well, there is still room for improvement, given that the average annotation correlation among RAs is 0.7, which can be considered the upper bound for model performance. Our current approach utilizes a linear kernel to train the model, which assumes features are independent. So, one potential next step is to consider combinations or interaction among features. Moreover, as we pointed out earlier, there were some topic features capturing concepts or information similar to those in the social distance feature. Thus, another possibility of improving the model is to remove redundant features or disentangle the relationships among features.

ACKNOWLEDGMENTS

REFERENCES
1. Alessandro Acquisti, Laura Brandimarte, & George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science, 347*(6221), 509-514.
2. Reed Albertgotti. 2014, May 22. Facebook's Privacy Dinosaur Is Back: New Members' Posts Aren't Automatically 'Public' Anymore. *Wall Street Journal*
3. Richard L Archer. 1980. Self-disclosure *The self in social psychology* (pp. 183-205). Oxford: Oxford University Press.
4. JinYeong Bak, Suin Kim, & Alice Oh. 2012. Self-disclosure and relationship strength in Twitter conversations. In *ACM ACL '12*, Jeju Island, Korea.
5. JinYeong Bak, Chin-Yew Lin, & Alice Oh. 2014. Self-disclosure topic model for classifying and analyzing Twitter conversations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar.
6. Sairam Balani, & Munmun De Choudhury. 2015. Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, Seoul, Korea.
7. Azy Barak, & Orit Gluck-Ofri. 2007. Degree and reciprocity of self-disclosure in online forums. *Cyberpsychol Behav, 10*(3), 407-417.
8. Vladimir Barash, Nicolas Ducheneaut, Ellen Isaacs, & Victoria Bellotti. 2010. Faceplant: Impression (Mis)management in Facebook Status Updates. In *ICWSM*.
9. Natalya N. Bazarova, Jessie G. Taft, Yoon Hyung Choi, & Dan Cosley. 2012. Managing Impressions and Relationships on Facebook: Self-Presentational and Relational Concerns Revealed Through the Analysis of Language Style. *Journal of Language and Social Psychology*.
10. David M. Blei, Andrew Y. Ng, & Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res., 3*, 993-1022.
11. danah michele boyd. 2008. *Taken Out of Context: American Teen Sociality in Networked Publics*: University of California, Berkeley.
12. Moira Burke, Cameron Marlow, & Thomas Lento. 2010. Social network activity and social well-being. In *ACM CHI 2010*, Atlanta, Georgia, USA.
13. Moira Burke, & Robert E. Kraut. 2014. Growing closer on facebook: changes in tie strength through social network site use. In *ACM CHI 2014*, Toronto, Ontario, Canada.
14. Stanley F. Chen, & Joshua Goodman. 1996. An empirical study of smoothing techniques for language modeling. In *ACM ACL '96*, Santa Cruz, California.
15. Philip Clarkson, & Roni Rosenfeld. 1997. Statistical Language Modeling Using the CMU-Cambridge Toolkit. In *Proceedings ESCA Eurospeech*. http://www.speech.cs.cmu.edu/SLM/toolkit.html
16. Nancy L. Collins, & Lynn C. Miller. 1994. Self-disclosure and liking: a meta-analytic review. *Psychol Bull, 116*(3), 457-475.
17. Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, & Christopher Potts. 2013. No country for old members: user lifecycle and linguistic change in online communities. In *WWW '13*, Rio de Janeiro, Brazil.
18. Valerian J. Derlaga, & John H. Berg. 1987. *Self-Disclosure: Theory, Research and Therapy*: Springer.
19. Kathryn Dindia, & Mike Allen. 1992. Sex differences in self-disclosure: a meta-analysis. *Psychol Bull, 112*(1), 106-124.
20. Nicole Ellison, Rebecca Heino, & Jennifer Gibbs. 2006. Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment. *Journal of Computer-Mediated Communication, 11*(2), 415-441.
21. Eric Gilbert, & Karrie Karahalios. 2009. Predicting tie strength with social media. In *ACM CHI '2009*, Boston, MA, USA.
22. Erving Goffman. 1959. *The Presentation of Self in Everyday Life*: Doubleday.
23. I. J. Good. 1953. The population frequencies of species and the estimation of population parameters. *Biometrika, 40*(3-4), 237-264.
24. Jeffrey T. Hancock, Catalina Toma, & Nicole Ellison. 2007. The truth about lying in online dating profiles. In *ACM CHI 2007*, San Jose, California, USA.

25. Susan C. Herring. 2007. A Faceted Classification Scheme for Computer-Mediated Discourse. *Language@Internet, 4*(1).

26. Bernie Hogan. 2010. The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. *Bulletin of Science, Technology & Society, 30*(6), 377-386.

27. Huffington Post. 2012, Apr 20. 37 Percent Of Employers Use Facebook To Pre-Screen Applicants, New Study Says. Retrieved May 15, 2015, from http://www.huffingtonpost.com/2012/04/20/employers-use-facebook-to-pre-screen-applicants_n_1441289.html

28. Adam N. Joinson. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology, 31*(2), 177-192.

29. Sidney M. Jourard, & Paul Lasakow. 1958. Some factors in self-disclosure. *J Abnorm Psychol, 56*(1), 91-98.

30. Michal Kosinski, David Stillwell, & Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 110*(15), 5802-5805.

31. Justin Kruger, Nicholas Epley, Jason Parker, & Zhi-Wen Ng. 2005. Egocentrism over e-mail: Can we communicate as well as we think? *Journal of Personality and Social Psychology, 89*(6), 925-936.

32. Mark R. Leary, & Robin M. Kowalski. 1990. Impression management: A literature review and two-component model. *Psychol Bull, 107*(1), 34.

33. Alice E. Marwick, & danah michele boyd. 2010. I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society*.

34. Soraya Mehdizadeh. 2010. Self-presentation 2.0: narcissism and self-esteem on Facebook. *Cyberpsychol Behav Soc Netw, 13*(4), 357-364.

35. Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, & Alessandro Provetti. 2014. On Facebook, most ties are weak. *Commun. ACM, 57*(11), 78-84.

36. Lynn C. Miller, John H. Berg, & Richard L. Archer. 1983. Openers: Individuals who elicit intimate self-disclosure. *Journal of Personality and Social Psychology, 44*(6), 1234-1244.

37. Melanie Nguyen, Yu Sun Bin, & Andrew Campbell. 2012. Comparing online and offline self-disclosure: a systematic review. *Cyberpsychol Behav Soc Netw, 15*(2), 103-111.

38. OpenSource. 2010. The Apache OpenNLP library. from https://opennlp.apache.org/

39. Debra L. Oswald, Eddie M. Clark, & Cheryl M. Kelly. 2004. Friendship Maintenance: An Analysis of Individual and Dyad Behaviors. *Journal of Social and Clinical Psychology, 23*(3), 413-441.

40. Namkee Park, Borae Jin, & Seung-A Annie Jin. 2011. Effects of self-disclosure on relational intimacy in Facebook. *Computers in Human Behavior, 27*(5), 1974-1983.

41. Malcolm Parks. 2010. Who are Facebook friends? Exploring the composition of Facebook friend networks. In *Proceedings of the Annual Meeting of the International Communication Association*, Singapore.

42. Delroy L. Paulhus. 1991. Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA, US: Academic Press.

43. James W. Pennebaker, Matthias R. Mehl, & Kate G. Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology, 54*(1), 547-577.

44. Martin Porter. 2006. Porter stemmer. from http://tartarus.org/martin/PorterStemmer/

45. John Rust, & Susan Golombok. 2009. Psychometric assessment of personality in occupational settings *Modern Psychometric: The Science of Psychological Assessment* (Third ed., pp. 165-182). New York, NY: Routledge.

46. Barry R. Schlenker. 1980. *Impression Management: The Self-concept, Social Identity, and Interpersonal Relations*: Brooks/Cole Publishing Company.

47. S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, & K. R. K. Murthy. 2000. Improvements to the SMO algorithm for SVM regression. *Neural Networks, IEEE Transactions on, 11*(5), 1188-1193.

48. Mark Snyder. 1974. Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*(4), 526-537.

49. Whitney P. Special, & Kirsten T. Li-Barber. 2012. Self-disclosure and student satisfaction with Facebook. *Computers in Human Behavior, 28*(2), 624-630.

50. Susan Sprecher, Stanislav Treger, & Joshua D Wondra. 2013. Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships, 30*(4), 497-514.

51. Lisa Collins Tidwell, & Joseph B. Walther. 2002. Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time. *Human Communication Research, 28*(3), 317-348.

52. Yi-Chia Wang, Robert E. Kraut, & John M. Levine. 2015. Eliciting and Receiving Online Support: Using Computer-Aided Content Analysis to Examine the Dynamics of Online Social Support. *J Med Internet Res, 17*(4), e99.

53. Ian H. Witten, Eibe Frank, & Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*: Morgan Kaufmann Publishers Inc.

54. George Kingsley Zipf. 1949. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.