Self Promotion in US Congressional Tweets

Jun Wang
Independent Researcher
Syracuse, New York
junwang4@gmail.com

Kelly Cui
Leland High School
San Jose, California
kcui930@gmail.com

Bei Yu
School of Information Studies
Syracuse University, New York
byu@syr.edu

Abstract

Prior studies have found that women self-promote less than men due to gender stereotypes. In this study we built a BERT-based NLP model to predict whether a Congressional tweet shows self-promotion or not and then used this model to examine whether a gender gap in self-promotion exists among Congressional tweets. After analyzing 2 million Congressional tweets from July 2017 to March 2021, controlling for a number of factors that include political party, chamber, age, number of terms in Congress, number of daily tweets, and number of followers, we found that women in Congress actually perform more self-promotion on Twitter, indicating a reversal of traditional gender norms where women self-promote less than men.

1 Introduction

Self-promotion is the act of presenting oneself as competent (Jones and Pittman, 1982). It is an important impression management technique in professional communication. Prior studies have found that self-promotion, when combined with other impression management techniques such as ingratiation for likeability, resulted in better interview evaluations (Proost et al., 2010). However, self-promotion was also found to be a risk factor for women—those who self-promoted may have encountered backlash for violating gender stereotypes (Rudman, 1998). This risk was more pronounced in traditionally male-dominated professions such as politicians. Women politicians were faced with the dilemma that, while their job required them to self-promote, doing so may have risked losing likeability and hurt election chances (Okimoto and Brescoll, 2010).

The popularization of social media use in recent years might provide an opportunity for women to escape this dilemma. According to the equalization theory, social media has changed the traditional power structures between politicians and the mass media; as a result, marginalized groups such as women may gain more control in impression management strategies by directly interacting with constituents on social media platforms like Twitter (Seidman, 2013; Vergeer, 2015; Jungherr, 2016; Fountaine, 2017). Thus, politicians' self-promotion behavior on Twitter is worth investigating further. However, there is scant research on content analysis of politicians' self-promotion on Twitter, although prior studies such as (Golbeck et al., 2010) and (Hemphill et al., 2013) have analyzed the topics of Congressional tweets.

In this research, we model Congresspeople's self-promotion on Twitter as an NLP problem. We first manually annotated a corpus of 4,000 tweets as self-promoting or not, and then built a prediction model to identify self-promoting tweets. This model was then used to analyze self-promotion tweets by Congress members from July 2017 to March 2021. We seek answers to the following research questions: (1) To what extent can NLP models identify self-promotion tweets from Congresspeople? (2) Who performed more self-promotion on Twitter, men or women?

2 Related Work

2.1 Theories of self-promotion

In communication theories, self-promotion is considered an important tactic for self-presentation (Goffman, 1959; Giacalone and Rosenfeld, 1986). The self-presentation theory proposed by Jones and Pittman (1982) provided a taxonomy of self-presentation tactics, which defined five strategies with different goals: (1) self-promotion for presenting oneself as competent, (2) exemplification for moral worthiness, (3) ingratiation for likability, (4) intimidation, and (5) supplication for requesting help. In this study we adopted Jones and Pittman's taxonomy, and defined self-promotion as a self-presentation tactic aiming to present oneself

Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 4893–4899
June 6–11, 2021.

as competent.

2.2 Gender gap in self-promotion

The phenomenon of a gender gap in self-evaluation and self-promotion has been well documented in social science research. Exley and Kessler (2019) found that, between equally high-performing men and women, women would self-evaluate more poorly despite evaluating others similarly regardless of gender. Such a gender gap has been observed in traditionally male-dominated professions. For example, many businesswomen were uncomfortable using impression management behaviors (Singh et al., 2002); women MBA graduates were less likely to utilize free-form data fields to promote themselves in their LinkedIn profiles (Altenburger et al., 2017); and women researchers made fewer self-citations than men (King et al., 2017). Women politicians are of particular interest for the gender gap in self-promotion behavior in that their jobs require self promotion, especially during elections and re-elections. Hence, female politicians often face this double bind of likeability vs. competence (Schneider et al., 2010).

2.3 Politicians' self-promotion on Twitter

Prior studies have found that self-presentation is a major motivation for social media use (Seidman, 2013). For politicians around the world, Twitter has become a popular social media platform. For example, Jackson and Lilleker (2011) characterized Twitter as a "tool for impression management" among members of the UK Parliament, with self-promotion being the most common among their identified purposes. During the 2014 elections in Belgium and Spain, Coesemans and de Cock (2017) found that Twitter was not only used for professional political communication, but also for personal branding.

Interestingly, recent studies on female politicians' Twitter behavior found patterns that deviated from traditional gender norms. For instance, female House candidates both tweeted more and possessed higher follower counts than their male counterparts in the 2012 election (Evans et al., 2014). It appears that female politicians actively utilize Twitter, perhaps as a way to overcome other systemic obstacles. They also campaigned with more "negative" and "attack-style tweets" than men, which could potentially detract from their image in voters' eyes (Evans and Clark, 2016). However, recent evidence appears to suggest that being seen as ambitious might no longer adversely affect female candidates (Saha and Weeks, 2020). Therefore, it is worthwhile to re-visit the gender gap in self-promotion among politicians on Twitter.

	House			Senate			
	D	R		D	R		total
Female	105	39	(144)	19	10	(29)	173
Male	167	266	(433)	37	55	(92)	525
			(577)			(121)	698

Table 1: Distribution of Congress members of class 115, 116, and 117 across chamber, party, and gender. (D: Democratic; R: Republican; Libertarian and Independent party are excluded).

3 Data and Methods

3.1 Dataset

A data set containing Congress members' tweets from July 1 of 2017 to March 31 of 2021, a total of 45 months' worth of data, was collected from the publicly available repository of Alex Litel's Tweets of Congress project,[1] which includes daily tweets from members of the 115th – 117th Congresses, including both Senate and House.

Besides the tweets, this data set also includes metadata for each Congressperson, such as chamber, party, and a bio ID.[2] Using the bio ID, we were able to link each Congressperson to his/her profile compiled by the @unitedstates project,[3] which includes demographic information such as gender and birthday for members of the US Congress since 1789.

After the data linkage, we obtained about 2 million tweets in total—retweets were excluded—from 698 Congress members. Table 1 provides a summary of the gender, chamber, and party of the Congress members.

Figure 1 shows the median of the number of tweets posted by members of Congress per month from July 2017 to March 2021. Women consistently posted more tweets than men, in accordance with the finding in (Evans et al., 2014). The overall trend for both genders is also consistent with major events that occurred during this period of time, confirming the reliability of the data set; for example: (1) less tweets in August due to Congress recessing

[1] https://github.com/alexlitel/congresstweets

[2] https://github.com/alexlitel/congresstweets-automator/blob/master/data/historical-users-filtered.json

[3] https://github.com/unitedstates/congress-legislators

Common types of self-promotion	Examples
(1) Sharing information about or soliciting participation in events featuring self	"I'm speaking with reporters live at the U.S. Capitol as the House continues its work to put #FamiliesFirst in America's response to the coronavirus pandemic. https://t.co/hDgusBJB1L"
(2) Talking about own work progress and accomplishments, such as introducing or passing bills, demonstrating authority, or acting in leadership positions	"As co-chair of the Medicare for All Congressional Caucus, I fight everyday for every American to access quality healthcare. We need Medicare for All."
	"A patient identifier is a common sense way to reduce medical errors and save lives. Proud the House adopted my amendment this week. https://t.co/6jBfvgUIJc"
(3) Mentioning received recognitions, such as endorsements and awards	"I was honored to join @1SI_Chamber today and receive the "Spirit of Enterprise" Award from the @USChamber https://t.co/FJ199jFm9G"
	"I am honored to have received the endorsement from the BRAFLCIO . Thank you to all the workers, retirees, and their families who truly are the voice and backbone of Florida's labor movement. #aflcio #union #local #fl20 https://t.co/S1hvT1pE6h"

Table 2: Common types of self-promotion tweets used by members of Congress.

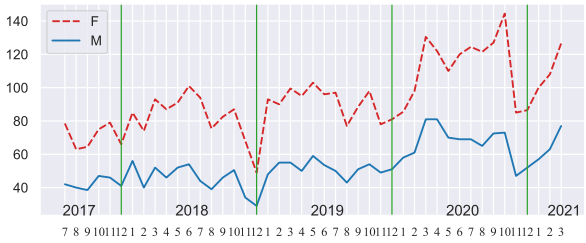

Figure 1: Median number of tweets posted each month by Congress women and men. On average women posted 93 tweets per month and men 54 tweets (retweets not counted).

for the month; (2) less tweets during year-end holidays; (3) a significant decrease right after the 2018 mid-term and the 2020 election; and (4) a significant increase in March 2020 due to the Covid-19 pandemic.

3.2 Annotated corpus

Following Jones and Pittman (1982)'s taxonomy, in this study we define self-promotion as a self-presentation tactic aiming to present oneself as competent. To operationalize the defined concept of self-promotion, two annotators conducted iterative rounds of coding to identify self-promotion content in the tweets. In each round, one hundred tweets were randomly selected and independently coded as either self-promotion or not by the annotators. The disagreements were brought to group discussion. After two rounds of discussion, a sample of 300 tweets was used to conduct an inter-coder agreement test. The result shows an agreement level at 0.80, measured by Cohen's Kappa. The two annotators then each annotated more tweets, resulting in a total of 4,003 annotated tweets, including 914 self-promotion and 3,089 non-self-promotion tweets. We also summarized the three most common types of self-promotion tweets observed during annotation: (1) advertising events featuring self, (2) talking about own work progress or accomplishments, and (3) announcing received recognitions such as awards and endorsements. See Table 2 for tweet examples.

In order to ensure that the training dataset contains a sufficient amount of self-promotion tweets, we over-sampled tweets that contain the word "I" (referred to as *I-tweets*), based on the critical role of self-referencing in self-promotion (Coesemans and de Cock, 2017). We found that over 30% of I-tweets contain self-promotion, while only about 10% of non-I-tweets contain self-promotion. Therefore, although the original ratio of I-tweets vs. non-I-tweets is 0.37 to 1 in the data set, we sampled I-tweets and non-I-tweets by a ratio of 1.7 to 1, resulting in about 2,500 I-tweets and about 1,500 non-I-tweets in the annotated corpus. In addition, to ensure that we have a representative sample, the 4,003 tweets were sampled with each member of Congress contributing at most 10 tweets to the sample.

3.3 Machine learning models

We evaluated two machine learning models on our annotated corpus via 5-fold cross-validation. One is LinearSVM, and the other one is BERT (Devlin et al., 2019). The BERT model [4] achieved a score of macro-F1 at 0.890, and accuracy at 0.923 (see result details in Table 3). In contrast, LinearSVM [5]

[4]Parameter settings for the BERT model: 3 epochs, learning rate=1e-5, max sequence length=128, cased BERT-base pretrained model

[5]Parameter settings for the LinearSVM: scikit-learn LinearSVC, C=1 and tf-idf vectorization with parameters: 1/2/3

achieved a much lower score of macro-F1 at 0.652 and accuracy at 0.868. The BERT model was chosen to be applied to identify all self-promotional tweets in the data set.

Self-promotion	Precision	Recall	F1	Support
no	0.948	0.951	0.950	3089
yes	0.834	0.825	0.829	914
accuracy			0.923	
macro average	0.891	0.888	0.890	

Table 3: Performance of the fine-tuned BERT model.

To help us understand the linguistic cues used in self-promoting tweets, we conducted an analysis with LIME - a machine learning interpretation tool (Ribeiro et al., 2016). We first sampled 5000 tweets, then for each tweet, we ran LIME (paired with the above fine-tuned BERT model) to find the most salient words (specifically, top 7 words with the highest weights). This resulted in the following list of content words related to expressing self-promotion: *bill, legislation, Tune, introduced, Act, proud, honored, bipartisan, joining, live*, etc. When adding the context in which these words occur, we found such phrases as:

1. *I am proud to introduce / cosponsor / support / vote for a [bipartisan] bill / legislation*
2. *Be sure to tune in / I'm live now / I'm hosting a virtual town hall*
3. *I'm honored to have received / earned / be recognized by*

We also examined a sample of prediction errors to identify areas for future improvement. The prediction model missed some self-promotion tweets that are implicit without direct attribution, such as the Case 1 in Table 4. An implicit self-promotion tweet may also attribute the credit to a group instead of oneself, or self-promote through someone else's words, such as a direct quote from a voter. The prediction model also mistook some non-self-promotion tweets as self-promotion, due to linguistic similarity. For example, in the Case 2 in Table 4, a Congress member attended a social event to demonstrate their moral worthiness rather than their competency. The error analysis shows that more clarifying training examples may further improve the prediction model.

ngrams and min_df=3

Case 1. *"I am always willing to stand up for what I believe in, but I will always do it as respectfully as possible and with a goal toward building the greatest power. This strategy is working and US-Progressives have more power than ever before."* (Note: self-promotion, false negative prediction)

Case 2. *"Yesterday, on the steps of the State Capitol in Sacramento, I joined hundreds in rallying against the state's latest water grab in the San Joaquin Valley."* (Note: not self-promotion, false positive prediction)

Table 4: Examples of the BERT model prediction error.

4 Results

Applying the above trained BERT model to the 2 million tweets posted by the Congress members, we found that 16.7% of the tweets contained self-promotion.

To examine gender difference in self-promotion, we adopted a generalized linear mixed-effects regression framework, in which (1) the fixed-effects factors are gender (`F/M`), political party (`D/R`), chamber (`house/senate`), age, number of terms served in Congress, number of daily tweets (representing tweet frequency), and number of followers (preprocessed with log transformation due to its highly skewed distribution), (2) the random-effects factors are the author and the date of a tweet; and (3) the dependent variable is whether a tweet contains self-promotion, of which the value comes from the BERT prediction result.

We fed the 2 million observations of tweets into the mixed-effects model, using the `glmer()` function of the R package `lme4` (Bates et al., 2014)—see Appendix A for the detailed regression formula and Appendix C for the distribution of the four numerical factors. Table 5 shows a significant gender difference when controlling for other factors: women in Congress are more likely to self-promote in their tweets than their men colleagues.

We are also interested in further examining whether this gender difference has been consistent over the time. To answer this question, for each month from July 2017 to March 2021, we fit the monthly data to the mixed-effects model, and then from each monthly model we calculated the estimated marginal means or expected means.Specifically, we used the `ggemmeans()`

	Coef	Std Err	P-value
gender [F]	**0.110**	0.044	0.013 *
party [R]	0.071	0.039	0.065 ·
chamber [senate]	**0.224**	0.057	0.000 ***
age	**−0.006**	0.002	0.003 **
num_terms	0.008	0.005	0.108
daily_tweets	**−0.019**	0.000	0.000 ***
followers_log	**−0.136**	0.015	0.000 ***
AIC	1695986		
Num. obs.	1981428		

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Table 5: Coefficients of the mixed-effects regression model in predicting if a tweet shows self-promotion. Note that `gender[M]`, `party[D]`, and `chamber[house]` are used as the reference levels of their corresponding factors. Also see Appendix B for the odds ratios of the factors.

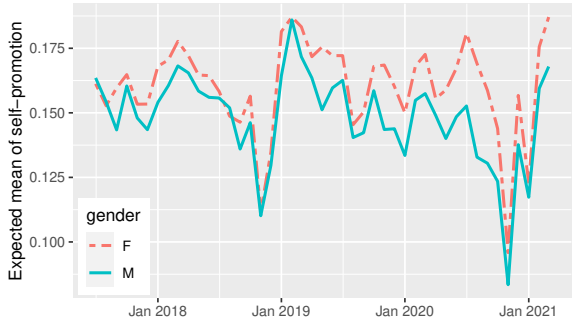

Figure 2: Women consistently performed more self-promotion than men over the time.

function[6] in the R package `ggeffects` to do the calculation (Ludecke, 2018). As shown in Fig. 2, we can see that women consistently exhibited more self-promotion than men.

In addition to the gender effect, Table 5 also shows other significant factors for self-promotion: (1) Senators are more likely to send self-promotion tweets than House Representatives; (2) young people self-promote more than old people; and (3) Congress members with fewer followers or those who tweet less frequently are more likely to do self-promotion.

While more research is needed for causal interpretations, these findings seem to be consistent with common sense knowledge. As mentioned in Table 2 (common types of self-promotion tweets), most self-promotion tweets were advertising events or touting accomplishments, endorsements, and awards. Since Senators represent the entire states, while members of the House represent individual districts, Senators are in general more politically powerful, and might be involved in more activities that they can use for self-promotion. It is probably not surprising that younger members do more self-promotion on Twitter as they are more social media savvy. The negative correlation between self-promotion and the tweet frequency (daily tweets) or number of followers indicates that for the members who are less active on Twitter or have fewer followers, self-promotion accounts for a larger proportion of their tweets, suggesting that their Twitter use is somewhat more focused on self-promotion.

5 Conclusion

Contribution. We built an annotated corpus of self-promotion tweets posted by Congress members, and trained a BERT-based prediction model with 0.89 macro-F1 score. To the best of our knowledge, this is the first NLP model for predicting self-promotion in political tweets. Applying this model to 2 million Congressional tweets from July 2017 to March 2021, we found that 16.7% of Congressional tweets contained self-promotion. After controlling for a number of factors we found women in Congress perform significantly more self-promotion on Twitter than their male colleagues. This indicates a reversal of traditional gender norms where women self-promote less than men.

Limitations. Although the data set we used is large and spans almost 4 years, more data are needed to evaluate whether the self-promotion prediction model is generalizable to politicians outside of the US Congress, such as those in other government branches (e.g. executive and judicial) and levels (e.g. states and counties), and other countries. Based on our manual annotations, we would speculate that the model should be generalizable to some extent in that self-promotion content shares some common terms such as describing leadership roles and sharing news on awards and endorsement. However, some self-promotional content may be domain-specific, e.g. accomplishment on introducing and passing bills is only applicable to legislators.

[6] For categorical factors such as party, `ggemmeans()` averages over their categories; for numerical factors such as age, their mean values (e.g., age=60) are used.

• The annotated corpus, related data and code are available at https://github.com/junwang4/self-promotion-in-congress-tweets.

Acknowledgement

We thank the reviewers for very helpful comments, and Albert Wang for help with annotation.

References

Kristen M Altenburger, Rajlakshmi De, Kaylyn Frazier, Nikolai Avteniev, and Jim Hamilton. 2017. Are there gender differences in professional self-promotion? An empirical case study of linkedin profiles among recent mba graduates. In *ICWSM'2017*.

Douglas Bates, Martin Machler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67:1–48.

Roel Coesemans and Barbara de Cock. 2017. Self-reference by politicians on twitter: Strategies to adapt to 140 characters. *Journal of Pragmatics*, 116:37–50.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT'2019*, pages 4171–4186.

Heather K. Evans and Jennifer Hayes Clark. 2016. "You tweet like a girl!": How female candidates campaign on Twitter. *American Politics Research*, 44(2):326–352.

Heather K. Evans, Victoria Cordova, and Savannah Sipole. 2014. Twitter Style: An Analysis of How House Candidates Used Twitter in Their 2012 Campaigns. *Political Science & Politics*, 47(2):454–462.

Christine L Exley and Judd B Kessler. 2019. The gender gap in self-promotion. Working Paper 26345, National Bureau of Economic Research.

Susan Fountaine. 2017. Whats not to like?: A qualitative study of young women politicians self-framing on twitter. *Journal of Public Relations Research*, 29(5):219–237.

Robert A Giacalone and Paul Rosenfeld. 1986. Self-presentation and self-promotion in an organizational setting. *The Journal of social psychology*, 126(3):321–326.

Erving Goffman. 1959. *The Presentation of Self in Everyday Life*. Anchor Books.

Jennifer Golbeck, Justin M Grimes, and Anthony Rogers. 2010. Twitter use by the US Congress. *Journal of the American Society for Information Science and Technology*, 61(8):1612–1621.

Libby Hemphill, Jahna Otterbacher, and Matthew Shapiro. 2013. What's congress doing on twitter? In *CSCW'2013*, pages 877–886.

Nigel Jackson and Darren G. Lilleker. 2011. Microblogging, Constituency Service and Impression Management: UK MPs and the Use of Twitter. *The Journal of Legislative Studies*, 17:86–105.

Edward E Jones and Thane S Pittman. 1982. *Toward a general theory of strategic self-presentation*, volume 1, chapter 9. Lawrence Erlbaum Associates.

Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics*, 13(1):72–91.

Molly M King, Carl T Bergstrom, Shelley J Correll, Jennifer Jacquet, and Jevin D West. 2017. Men set their own cites high: Gender and self-citation across fields and over time. *Socius*, 3.

Daniel Ludecke. 2018. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, 3(26):772.

Tyler G Okimoto and Victoria L Brescoll. 2010. The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, 36(7):923–936.

K. Proost, B. Schreurs, K. D. Witte, and Eva Derous. 2010. Ingratiation and self-promotion in the selection interview: The effects of using single tactics or a combination of tactics on interviewer judgments. *Journal of Applied Social Psychology*, 40:2155–2169.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *KDD'2016*, page 11351144.

Laurie A Rudman. 1998. Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, 74(3):629–645.

Sparsha Saha and Ana Catalano Weeks. 2020. Ambitious women: Gender and voter perceptions of candidate ambition. *Political Behavior*, pages 1–27.

Andrea Kupfer Schneider, Catherine H Tinsley, Sandra Cheldelin, and Emily T Amanatullah. 2010. Likeability v. competence: The impossible choice faced by female politicians, attenuated by lawyers. *Duke Journal of Gender Law & Policy*, 17(2):363–384.

Gwendolyn Seidman. 2013. Self-presentation and belonging on facebook: How personality influences social media use and motivations. *Personality and individual differences*, 54(3):402–407.

Val Singh, Savita Kumra, and Susan Vinnicombe. 2002. Gender and impression management: Playing the promotion game. *Journal of Business Ethics*, 37(1):77–89.

Maurice Vergeer. 2015. Twitter and political campaigning. *Sociology compass*, 9(9):745–760.

Appendix

Appendix A

Table 6 gives the formula that we used to construct the regression model with the `glmer()` function in the R package `lme4` (Bates et al., 2014).

self_promotion ∼	
gender	`// M, F`
+ party	`// D, R`
+ chamber	`// house, senate`
+ age	`// numerical`
+ num_terms	`// numerical`
+ daily_tweets	`// numerical`
+ followers_log	`// numerical`
+ (1 | tweet_author)	`// random effect`
+ (1 | tweet_date)	`// random effect`

Table 6: Formula of the logistic linear mixed-effects regression model.

Appendix B

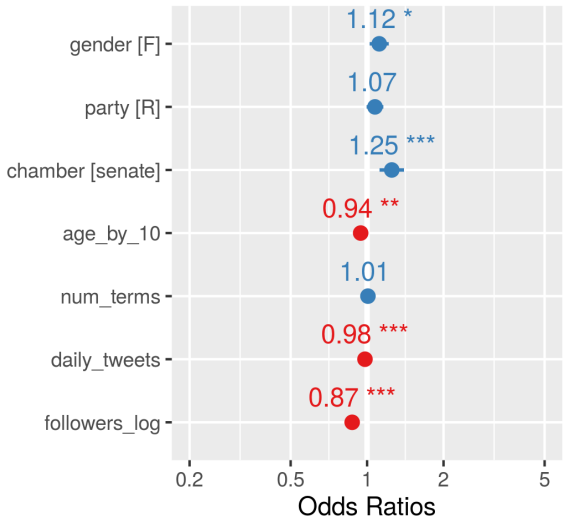

Figure 3: Odds ratios of the above independent variables obtained by fitting the regression model with 2 million Congressional tweets. For *age*, here we present its odds ratio in terms of $age/10$ (without scaling, the ratio is 0.99).

Appendix C

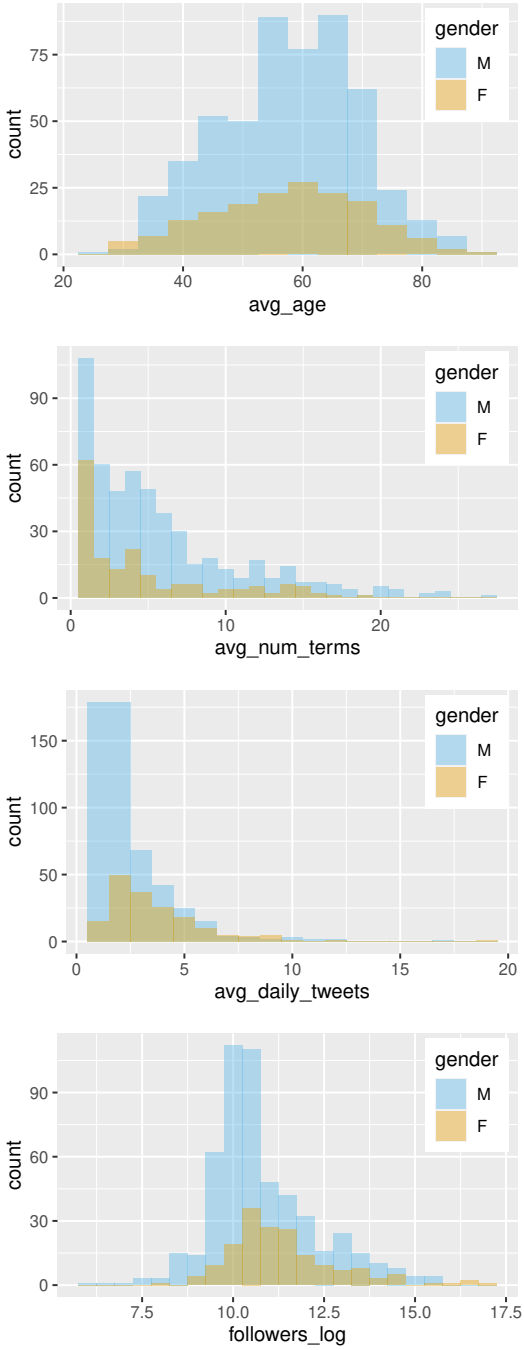

Figure 4: Distribution of the four numerical factors used in the regression model: *age*, *num_terms*, *daily_tweets*, and *followers_log*. Our data span about 4 years, so the value of age or num_terms changes as the tweet date varies; as a result, we show each person's average values (in terms of *median*). Similarly, for variable daily_tweets, we also show a person's average here.