

The Gender Gap in Scholarly Self-Promotion on Social Media (Supplemental Materials)

Hao Peng, Misha Teplitskiy, Daniel M. Romero, Emőke-Ágnes Horvát

(Dated: May 12, 2025)

I. SUPPLEMENTARY NOTE

A. Data collection

Altmetric. Our data are based on the most complete record of research papers’ online mentions, maintained by Altmetric.com [5]. This service has been tracking the online mentions of research outputs since 2011 on different platforms, including news media and social media such as Twitter and Facebook. We accessed the database (referred to as “Altmetric” hereafter) on October 8, 2019. Altmetric matches online attention to papers based on their unique identifiers, such as the Digital Object Identifier (DOI), PubMed ID, and arXiv ID. Utilizing these identifiers, it also collapses the attention of different versions of each paper into a single unique record [6]. This ensures that the data contain the complete mentions of each paper. We obtain all the mentions of research papers published between 2013 and 2018 in the largest platform, Twitter, which consists of about 74% of all mentions (posts) and 92% of all social media mentions in the entire database [23].

Due to Altmetric’s data license agreement with Twitter, the dataset contains only the tweet ID for each tweet. We thus collect all tweets using the Twitter API. Due to account deletion or changes in the privacy settings by some Twitter users, we have successfully retrieved about 90% of all tweets. The Altmetric data also provide metadata for each paper, such as the DOI, publication year, publication venue, and research topics. The disciplinary catalog uses 26 Scopus Subject Areas, which belong to four broad disciplines: Social Sciences, Life Sciences, Physical Sciences, and Health Sciences [13]. The classification is performed by in-house experts based on the aim and scope of the content a journal publishes [13].

Microsoft Academic Graph. We use the Microsoft Academic Graph (MAG) database [26, 30] (accessed on June 01, 2019) to retrieve other metadata for each paper based on the DOI. We obtain essential author information for each paper, including author’s name and affiliation, affiliation location and rank. We also consider the authors’ previous number of citations and publications up to the publication year of each paper, which reflects authors’ previous productivity when deciding whether to self-promote or not, assuming authors tend to self-promote soon after the publication of a paper.

MAG leverages data mining and artificial intelligence techniques to address author conflation and disambiguation, which ensures that the author’s number of publications is counted accurately [20, 30]. Our dataset at this stage consists of 14,521,585 (paper, author) pairs for 2,983,706 papers after ensuring that we have complete metadata.

B. Gender inference

Gender prediction. We use Ford’s algorithm [14] (accessed on May 1, 2021) to infer the gender of an author based on their first name. The algorithm returns, for a given name, one of 4 categories: Woman, Man, Unisex, Unknown. A name is predicted as “Woman” (“Man”) if it is used for women (men) at least twice as frequently as for men (women), based on data from national statistics institutes [4]. Otherwise,

the name is labeled as “Unisex” (gender-neutral). If the name is not found in the database, it is labeled as “Unknown.”

The percentage of 14.5M observations for Man, Woman, Unknown, and Unisex is 51%, 28%, 16%, 5%, respectively. We exclude “Unisex” and “Unknown” categories from our analytical sample, which consists of 11,396,752 (paper, author) pairs for 2,834,829 papers by 3,503,674 unique authors (58% Man vs. 42% Woman). We treat “Man” as the reference category in all regressions.

Agreement with manually assigned gender labels. We evaluate the performance of this algorithm based on a random sample of 100 authors in our data, for which we manually label their gender. In the labeling process, the author’s gender is determined based on their pronouns and profile pictures displayed on their personal websites, institutional directories, and Wikipedia pages found via Google searches of author names. We use the author’s affiliation to disambiguate multiple authors with the same name. If the gender could not be verified, it is labeled as “Unknown.” There are 19 women, 57 men, and 24 authors with unknown gender in the labeled sample. Based on the set of 76 authors with confirmed gender, the algorithm achieves an accuracy of $F1_m = 0.91$ and $F1_f = 0.88$. A different algorithm, the Genderize API [1, 3, 17, 19] (accessed on June 1, 2021), produces a similar result.

Validation with self-reported gender. We also validate the accuracy of the gender prediction using author self-identified gender labels in the manuscript submission data provided by IOP Publishing (<https://iopublishing.org/>), an academic publishing company specializing in the field of physics.

Each author has self-reported their gender and country of residence, such as China, India, U.S., Canada, Australia, etc. In our evaluation, we focus on authors with self-reported gender as either man or woman and whose names are predicted as either man or woman, as our actual analysis is focused on men and women. There are 432,888 authors, submitting to 62 journals, in this data. Here, we list authors from China (the largest single group) as a separate group because many Chinese names are inherently gender ambiguous and many gender-distinctive Chinese names have the same latin characters [16]. This means that many Chinese authors can have the same English name but they identify with different genders. In our data, there are 475 names associated with both man and woman based on self-reported gender labels, and the vast majority of these names are from Chinese authors. The prediction accuracy for Chinese names is lower than non-Chinese names, as expected, but the overall F1 score is close to 0.9 (Table S18).

Robustness check on non-East Asian names. Since the gender prediction is typically less accurate for East Asian names [25], we repeat our analysis by excluding observations whose author names are predicted to be East Asian ethnicities using the Ethnea API [27] (accessed on Sep 1, 2024). For a specific name, *Ethnea* assigns the ethnicity probabilities based on matched authors in the PubMed database. In the case of two or more predicted ethnicities, we take the one with the highest probability. *Ethnea* predicts 26 individual ethnicities. We categorize seven of them as East Asian, including *Chinese*, *Indonesian*, *Japanese*, *Korean*, *Mongolian*, *Thai*, *Vietnamese*. We obtain consistent results on the subset of the data with non-East Asian names (Table S5).

C. Detecting self-promotion

Each paper has a list of tweets that mention the paper, with each tweet containing the Twitter handle and the screen name of the user (referred to as “tweet names” hereafter). We define self-promotion as the author posting a tweet sharing the unique identifier of their paper, such as the DOI, PubMedID, arXiv ID, etc. Self-promotion on Twitter comes in two different forms, which can manifest in the type of tweet that shares

the research, including (i) an original tweet, or (ii) a retweet. We construct a binary dependent variable indicating whether the author self-promoted the paper or not, based on both retweets and original tweets. The two types of self-promotion differ in how direct the promotion is: the original-tweet-based promotion comes from the authors themselves, whereas the retweet-based promotion originates from others (e.g., an author A retweets a tweet from others sharing A’s paper). Self-promotion in the form of posting an original tweet potentially indicates a stronger intention to advertise one’s paper than does posting a retweet, which is a more indirect promotion.

We determine if an author is among the users who have tweeted the paper based on string matching. There is no perfect method to match author names with tweet names because scholars can use any string as their handle or screen name on Twitter. We thus adopt a simple “containment-matching” approach that searches the author name in tweet names—if either the first name or the last name string is contained in the user handle or screen name (lowercased), we consider the user to be the author of the paper. In case of multiple matches, we use the one with the highest fuzzy matching score [2].

We validate this method using a random sample of 100 papers (each with at least one tweet mentioning it) with manual verification. Due to having multiple authors per paper, there are 521 (paper, author) pairs in the manual labeling process. For each observation, we verify the author against all tweets of the paper to check if the author is among the tweet users. This method achieves an initial F1 score of 0.85 (precision: 0.77 and recall: 0.95).

We refine this method through an iterative process by experimenting with various new heuristics to address false predictions and evaluating its performance based on an independent manual verification process at each iteration. In the final version, we use “containment-matching” only if the tweet names are single-token strings and the author’s first name (or last name) has at least four characters; otherwise, we use “token-matching”: i.e., the first name or the last name should be matched to the tokens of tweet names (split by space or underscore). This heuristic achieves an out-of-sample precision of 0.95 and a recall of 1.00 (F1 score: 0.97).

D. Establishing active status on academic Twitter

The crucial practical challenge in establishing which scholars are active on Twitter is to match them to their Twitter accounts reliably. This matching is difficult because researchers often use Twitter handles unrelated to the names they use for academic papers. Moreover, even if all authors use Twitter handles containing parts of their names, it would still be computationally challenging to match author names to all possible Twitter handles currently in use.

We conduct a thorough literature search to identify the latest methods for linking academics to Twitter users. While no method can perfectly identify whether an author has a Twitter account, the state-of-the-art approach by an independent research team using open datasets produces valuable results [10, 22]. This method utilizes a heuristic to match authors in OpenAlex [24] to a massive dataset of Twitter users who have mentioned a research paper in the Crossref Event data [11]. Thus authors identified in this external dataset are at least somewhat “active” on Twitter. Although the active authors identified in this dataset might still be different from the set of all active authors (e.g., authors whose Twitter usernames are entirely unrelated to their author names), it is a high-precision indicator of which scholars are active on Twitter.

Our data span six years, and authors could have joined Twitter at any point during this timeframe and published multiple papers. With this in mind, we have to carefully code the active status on Twitter for each (paper, author) observation. Specifically, a (paper, author) pair is coded as active on Twitter if two conditions are met: (1) the author is identified as having an active Twitter account in the external dataset [22], and (2) the paper’s publication date is after the author’s earliest tweet time in our Altmetric dataset.

E. Modeling the likelihood of self-promotion

We use a mixed-effects logistic regression [8] to estimate the probability that an author self-promotes a paper as a function of their inferred gender while controlling for the following variables.

Year of publication: Scholars are increasingly promoting their research on Twitter over time (Fig. S1), thus it is important to control for the publication year of a paper.

Journal impact factor: Papers published in high impact journals may be more likely to be shared by their authors (Fig. S2). We obtain the journal impact factor from The Web of Science (2018 version).

Author’s affiliation rank: Authors from higher-ranked institutions may be more likely to self-promote (Fig. S2). We thus consider the rank of their affiliations provided in the MAG. When an author has multiple affiliations in a paper, we use the one with the highest rank. This ranking metric estimates the relative importance of institutions using paper-level features derived from a heterogeneous citation network; while similar to h-index, the method has been shown to produce more fine-grained and robust measurement of impact and prestige [26, 30]. To smooth this variable, we categorize the rank values into ten equally-sized bins (note that a smaller bin indicates a higher rank category).

Author’s affiliation location: We infer the country of author’s affiliation using the latitude and longitude provided in the MAG. We classify countries into two categories: (1) U.S., (2) international (non-U.S.). For authors with multiple affiliations, we only classify them as “U.S.” when at least one of their affiliations is located in the U.S. We use “U.S.” as the baseline to control for the fact that Twitter is an American social media platform that is more likely to be adopted by authors based in the U.S.

Authors’ previous number of publications: Authors’ academic standing or research experience can influence their likelihood to self-promote. To measure this variable, we count each author’s total number of publications before the publication year of the target paper based on MAG. We also categorize this numerical variable into ten equally-sized bins to reduce noise and outliers. Since there is a quadratic relationship between author productivity and self-promotion probability in the data (Fig. S2), we include a second order polynomial term for author productivity in the model.

Authors’ previous number of citations: Authors’ citation counts can impact both their likelihood of self-promotion and the return associated with self-promotion. We count each author’s total number of citations before the publication of the target paper using the MAG database. We log-transform this highly skewed variable to make it more interpretable.

Number of authors: We count the number of authors in each paper. Having more coauthors in a paper may negatively impact an author’s likelihood to promote the paper on their own.

Authorship position: Different authors often play different roles in multi-author projects. For example, authors who play a supportive role in the project may self-promote less frequently (Fig. S1). This variation is often captured by the authorship position in the paper. We thus control for the position of an author using a dummy variable with four categories: (1) first position, (2) middle position, (3) last position, (4) solo author. The “last position” is used as the reference category in the regressions. Note that when “solo author” = 1, the other three positions are all coded as 0.

Research fields: Not all scholars employ Twitter as a channel to share their research, and scientists’ representation on Twitter varies across disciplines [18]. To control for field-specific effects, we use the 26 Scopus Subject Areas. Each subject area is treated as a fixed variable in the regression, and its value is coded as 1 if the paper is assigned that subject (0 otherwise). Note that a paper can belong to multiple subject areas.

Paper random effects: Individual papers have different degrees of newsworthiness (e.g., biomedical papers have much more online coverage than papers from other disciplines [7]). Different papers may vary in the likelihood of being shared on social media by their authors. Gender representation online also varies across disciplines [29]. To capture such paper-level variations, we add random effects for each paper in the model.

F. Robustness check using propensity score matching

As a robustness check, we use propensity score matching to reduce the confounding effects that may be unaccounted for by the logistic regression model. We estimate the propensity score for each observation using a logistic regression with the author’s gender as the dependent variable and all other confounding factors defined previously as controls.

We then match each woman observation to one man observation using the nearest neighbor matching based on the propensity score. In this way, we construct a sample of man observations that is comparable on all observed covariates to the population of woman observations (Table S8). Using the matched sample, we still find that women self-promote their research less often than comparable men (Table S9).

G. Predicting the paper’s number of tweet mentions

We use a negative binomial regression to predict the total number of tweets of a paper. The unit of analysis is still each (paper, author) pair. We focus this analysis on the subset of observations where the author is active on Twitter at the publication date of the paper. The key independent variables are gender, self-promotion status (binary), and their interaction effect. The control variables are the same as defined previously, and include the author’s number of followers on Twitter (log-scale). Due to computational cost and time complexity, the paper random effects are not included in this model.

H. Gender differences in self-promotional language

To compare the textual content of self-promotion posted by men and women, we analyzed the words and sentiment of self-promotional tweets.

First, we used a software called “Vader” [15] to predict the sentiment of each tweet. This tool is specifically developed and attuned to detect sentiment expressed in social media posts and is widely used in past studies [9, 12]. Fig. S5 shows that scholars of both genders tend to share non-negative sentiment in self-promotion, which is not unexpected.

Second, to examine how men and women use different words to convey positive sentiment when sharing their research, we analyzed each of the 25 positive words used in scientific context (collected in [19]) and compared their usage by women and men in self-promotional tweets.

Fig. S6 shows that men use promotional words [21] such as “novel,” “excellent,” “robust,” “remarkable,” and “unprecedented” more often than women, while women use supportive words such as “amazing,” “creative,” “supportive,” and “inspiring” more often than men when promoting their research. Other positive words often used by both men and women in self-promotion include “unique,” “bright,” “prominent,” and “assuring.” The gender differences in the use of positive words suggest that not only do men and women self-promote at different rates, but also that they frame differently the associated language and sentiment.

II. SUPPLEMENTARY FIGURE

Effects of control variables: the raw gender difference. Fig. S1A shows that the average self-promotion probability quadruples from 2013 to 2018 for both genders. First and last authors are more likely than authors in middle positions to self-promote (Fig. S1B), which might be explained by the different roles played by authors in different positions [28]. Across research fields, self-promotion is much more common among social scientists than among physical, health, and life scientists (Fig. S1C). However, regardless of

author roles and research fields, there is a universal gender gap in self-promotion, with men self-advertising between 15% and 64% more often than women (Fig. S1B-C). Fig. S2 shows that other control variables such as the paper's journal impact factor, the author's affiliation rank, and the author's previous publications all correlate with the chance of self-promotion, regardless of gender.



FIG. S1. The average self-promotion probability by gender as a function of publication year, authorship position, and broad discipline in our data. $N = 11,396,752$ in this analysis. **A**, The raw self-promotion probability quadruples from 2013 to 2018 for both genders. **B**, The percentage of self-promotion grouped by authorship position. **C**, the same as **B**, but for the breakdown by papers' discipline. Confidence intervals represent 95% standard errors.

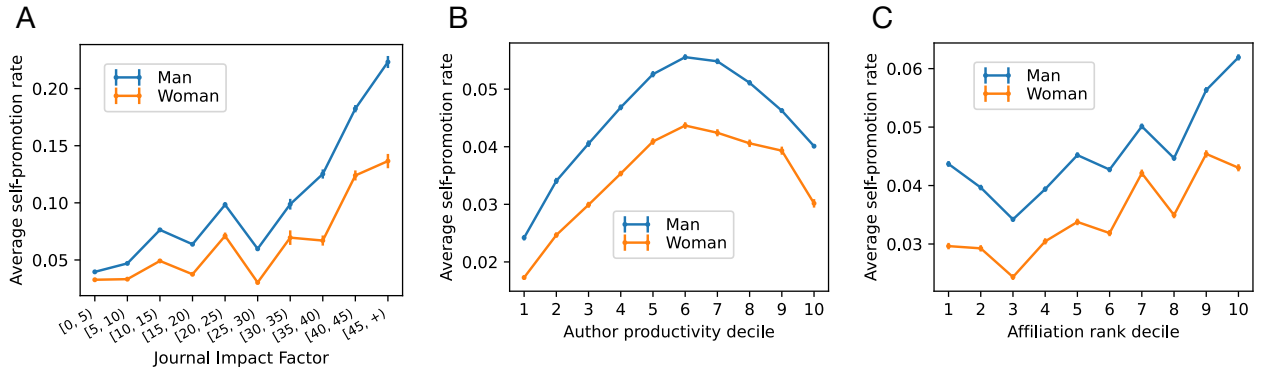


FIG. S2. The percentage of self-promotion cases by gender in our data as a function of **(A)** the paper's journal impact factor, **(B)** the author's productivity decile, measured as the author's previous number of publications, and **(C)** the author's affiliation rank decile. $N = 11,396,752$ in this analysis. Confidence intervals represent 95% standard errors.

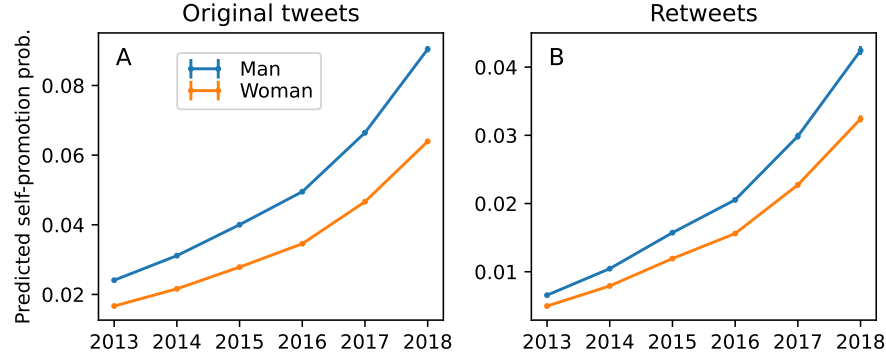


FIG. S3. The predicted probability of self-promotion after controlling for confounding factors in the mixed-effects logistic regression model as in Fig. 2A in the main text. $N = 11,396,752$ in this analysis. The two subplots correspond to two types of tweets (original tweets vs. retweets) based on which the binary dependent variable (self-promotion = True) is coded. Promotion based on original tweets comes directly from the authors, whereas retweet-based promotion originates from other Twitter users. From 2013 to 2018, men self-promote more often than comparable women, across either type of promotional tweets. Error bars indicate 95% bootstrapped confidence intervals.

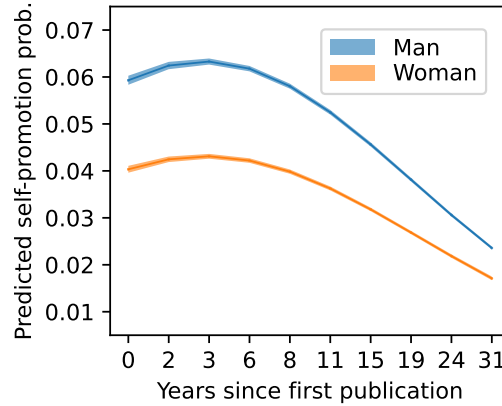


FIG. S4. The likelihood of self-promotion as a function of author career stage, measured as the number of years since first publication. $N = 11,396,752$ in this analysis. The gender gap in self-promotion increases with author's career stage but peaks in the mid-to-late career stage. The prediction is based on our most comprehensive model (Model 5 in SI Table S1) that also adds an interaction term between gender and the career stage variable.

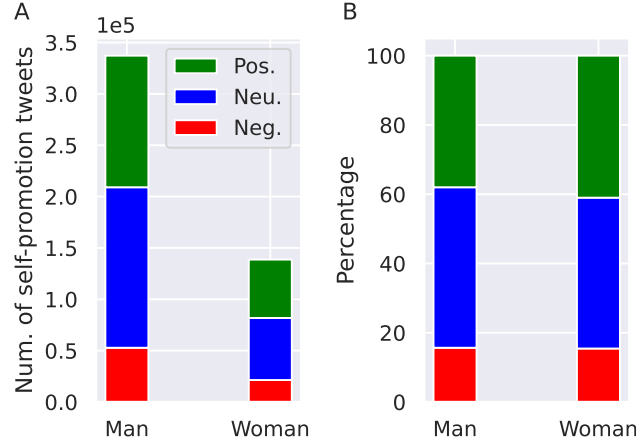


FIG. S5. Both women and men are mostly using non-negative sentiment when self-promoting their research on Twitter. **A**, the number of Positive (Green; 128,127 vs. 56,853), Neutral (Blue; 156,282 vs. 60,396), and Negative (Red; 52,821 vs. 21,417) tweets that involve self-promotion by men and women. **B**, the percentage of three types of tweets by gender. % Positive tweets: 38.0% vs. 41.0%; % Neutral tweets: 46.3% vs. 43.6%; % Negative tweets: 15.7% vs. 15.4%. A Chi-squared test shows that the two distributions are statistically different ($p < 0.000$).

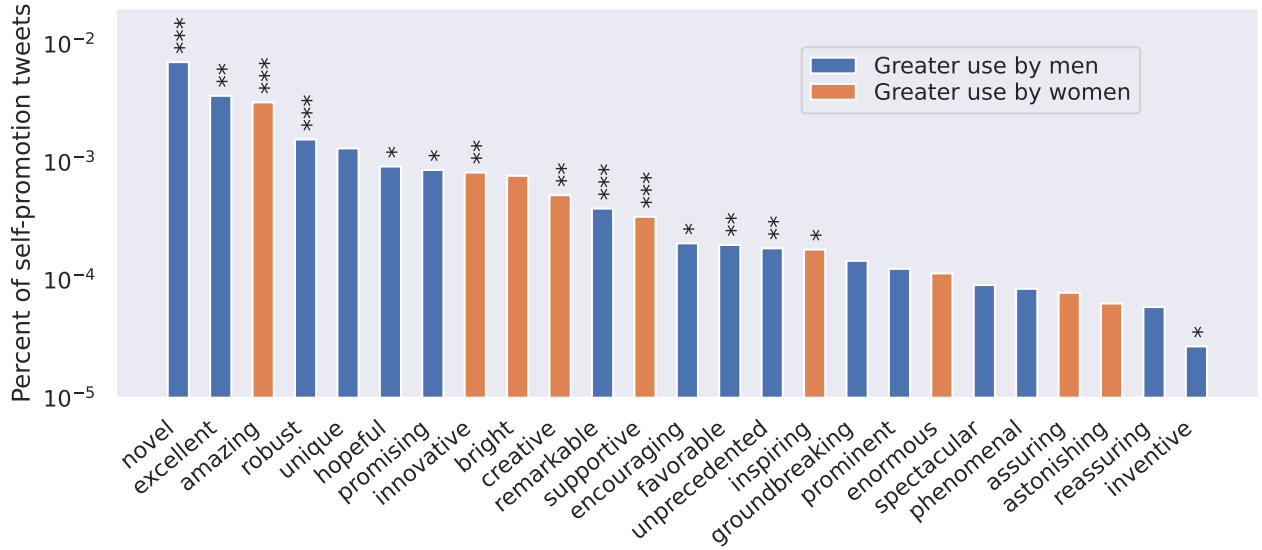


FIG. S6. Gender differences in the usage of 25 positive words [19] in self-promotion tweets. For each word, the variable for each self-promotion observation is set to one if the self-promotion tweet contains that word and to zero if it does not. The y-axis shows the average fraction of self-promotion tweets that have used each positive word. This figure shows that men use promotional words such as “novel,” “excellent,” “robust,” “remarkable,” and “unprecedented” more often than women, whereas women use supportive words such as “amazing,” “supportive,” and “inspiring” more often than men when promoting their research. Stars represent significance levels conducted based on the one-sided Mann–Whitney U test of word frequency distribution of both genders (***: $p < 0.001$, **: $p < 0.01$, and *: $p < 0.05$).

III. SUPPLEMENTARY TABLE

TABLE S1. Results of **five logistic regression models** predicting whether the author self-promoted the paper. All five models are fitted to 11,396,752 (paper, author) observations. Model 1 shows the gender gap without any controls; Model 2 controls for author factors; Model 3 adds the paper’s characteristics; Model 4 controls for research topics; Model 5 further includes paper random effects. The results show that all control variables are correlated with self-promotion and therefore should be included in the final model (Model 5). Note that the squared term for “Author previous num. of publications” is included due to its quadratic relationship with the dependent variable (Fig. S2B). Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable: Self-promotion = True</i>				
	Model 1	Model 2	Model 3	Model 4	Model 5
Woman	−0.298 (<0.001)	−0.265 (<0.001)	−0.284 (<0.001)	−0.311 (<0.001)	−0.348 (<0.001)
Authorship first position		0.337 (<0.001)	0.315 (<0.001)	0.339 (<0.001)	0.361 (<0.001)
Authorship middle position		−0.579 (<0.001)	−0.649 (<0.001)	−0.584 (<0.001)	−0.638 (<0.001)
Authorship solo author		0.894 (<0.001)	0.938 (<0.001)	0.686 (<0.001)	0.828 (<0.001)
Affiliation rank		−0.052 (<0.001)	−0.048 (<0.001)	−0.053 (<0.001)	−0.048 (<0.001)
Affiliation location international		0.046 (<0.001)	0.080 (<0.001)	0.080 (<0.001)	0.052 (<0.001)
Author previous num. of publications		0.266 (<0.001)	0.264 (<0.001)	0.260 (<0.001)	0.286 (<0.001)
Author previous num. of publications (squared)		−0.026 (<0.001)	−0.024 (<0.001)	−0.021 (<0.001)	−0.021 (<0.001)
Author log citations		0.020 (<0.001)	−0.002 (0.014)	−0.013 (<0.001)	−0.031 (<0.001)
Publication year 2014			0.318 (<0.001)	0.320 (<0.001)	0.305 (<0.001)
Publication year 2015			0.653 (<0.001)	0.656 (<0.001)	0.607 (<0.001)
Publication year 2016			0.901 (<0.001)	0.902 (<0.001)	0.846 (<0.001)
Publication year 2017			1.257 (<0.001)	1.252 (<0.001)	1.178 (<0.001)
Publication year 2018			1.598 (<0.001)	1.595 (<0.001)	1.542 (<0.001)
Number of authors			−0.001 (<0.001)	−0.001 (<0.001)	−0.002 (<0.001)
Journal impact factor			0.034 (<0.001)	0.035 (<0.001)	0.033 (<0.001)
Social Sciences				0.843 (<0.001)	0.898 (<0.001)
Materials Science				−0.189 (<0.001)	−0.144 (<0.001)
Engineering				−0.484 (<0.001)	−0.464 (<0.001)
Chemistry				−0.494 (<0.001)	−0.425 (<0.001)
Biochemistry, Genetics & Molecular Biology				−0.156 (<0.001)	−0.167 (<0.001)
Medicine				−0.317 (<0.001)	−0.279 (<0.001)
Nursing				0.248 (<0.001)	0.259 (<0.001)
Agricultural and Biological Sciences				0.482 (<0.001)	0.490 (<0.001)
Pharmacology, Toxicology & Pharmaceutics				−0.628 (<0.001)	−0.568 (<0.001)
Neuroscience				−0.112 (<0.001)	−0.108 (<0.001)
Business, Management and Accounting				−0.200 (<0.001)	−0.181 (<0.001)
Economics, Econometrics and Finance				−0.476 (<0.001)	−0.490 (<0.001)
Chemical Engineering				0.078 (<0.001)	0.067 (<0.001)
Physics and Astronomy				−0.847 (<0.001)	−0.728 (<0.001)
Computer Science				0.107 (<0.001)	0.134 (<0.001)
Decision Sciences				−0.781 (<0.001)	−0.779 (<0.001)
Health Professions				0.942 (<0.001)	0.905 (<0.001)
Psychology				−0.579 (<0.001)	−0.580 (<0.001)
Immunology and Microbiology				−0.170 (<0.001)	−0.175 (<0.001)
Dentistry				−1.390 (<0.001)	−1.263 (<0.001)
Earth and Planetary Sciences				−0.104 (<0.001)	−0.084 (<0.001)
Environmental Science				0.161 (<0.001)	0.168 (<0.001)
Mathematics				−0.211 (<0.001)	−0.242 (<0.001)
Arts and Humanities				−0.226 (<0.001)	−0.229 (<0.001)
Energy				−0.441 (<0.001)	−0.369 (<0.001)
Veterinary				−0.952 (<0.001)	−0.872 (<0.001)
General				0.376 (<0.001)	0.388 (<0.001)
Paper random effects	No	No	No	No	Yes
Intercept	−3.027 (<0.001)	−3.693 (<0.001)	−4.307 (<0.001)	−4.205 (<0.001)	−4.425 (<0.001)
Pseudo- R^2	0.01	0.13	0.16	0.20	0.51
Observations	11,396,752	11,396,752	11,396,752	11,396,752	11,396,752

TABLE S2. Coefficients of all variables in Model 5 (Table S1) that predicts **whether the author is active on Twitter when the paper is published**. The model is fitted to 11,396,752 observations. The negative woman coefficient indicates that women are indeed less likely to be active on Twitter (or have a Twitter account) than comparable men, which explains part of the overall gender gap in self-promotion rates, shown in Table S1. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Active on Twitter = True
Woman	-0.134 (<0.001)
Authorship first position	0.295 (<0.001)
Authorship middle position	-0.241 (<0.001)
Authorship solo author	0.650 (<0.001)
Affiliation rank	-0.071 (<0.001)
Affiliation location international	0.117 (<0.001)
Author previous num. of publications	0.670 (<0.001)
Author previous num. of publications (squared)	-0.039 (<0.001)
Author log citations	-0.031 (<0.001)
Publication year 2014	0.498 (<0.001)
Publication year 2015	0.891 (<0.001)
Publication year 2016	1.197 (<0.001)
Publication year 2017	1.511 (<0.001)
Publication year 2018	1.832 (<0.001)
Number of authors	-0.001 (<0.001)
Journal impact factor	0.011 (<0.001)
Social Sciences	0.863 (<0.001)
Materials Science	-0.156 (<0.001)
Engineering	-0.412 (<0.001)
Chemistry	-0.451 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.154 (<0.001)
Medicine	-0.084 (<0.001)
Nursing	0.294 (<0.001)
Agricultural and Biological Sciences	0.364 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.423 (<0.001)
Neuroscience	-0.104 (<0.001)
Business, Management and Accounting	-0.135 (<0.001)
Economics, Econometrics and Finance	-0.470 (<0.001)
Chemical Engineering	-0.009 (0.522)
Physics and Astronomy	-0.708 (<0.001)
Computer Science	0.178 (<0.001)
Decision Sciences	-0.630 (<0.001)
Health Professions	0.706 (<0.001)
Psychology	-0.448 (<0.001)
Immunology and Microbiology	-0.197 (<0.001)
Dentistry	-1.041 (<0.001)
Earth and Planetary Sciences	0.008 (0.364)
Environmental Science	0.228 (<0.001)
Mathematics	-0.262 (<0.001)
Arts and Humanities	-0.251 (<0.001)
Energy	-0.369 (<0.001)
Veterinary	-0.595 (<0.001)
General	0.100 (<0.001)
Intercept	-5.842 (<0.001)
Pseudo- R^2	0.42
Observations	11,396,752

TABLE S3. Coefficients of all variables in Model 5 (Table S1) fitted to 618,742 observations for which **the author is active on Twitter at the publication date of the paper**. It shows that the overall gender gap in self-promotion (Table S1) is not mainly driven by the gender difference in Twitter presence (Table S2), as the disparity exists even among authors active on Twitter. Note that the quadratic term for *Author previous num. of publications* is dropped because it is linearly correlated with the dependent variable in this subset (including the quadratic term produces similar results). Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	-0.189 (<0.001)
Authorship first position	0.333 (<0.001)
Authorship middle position	-0.741 (<0.001)
Authorship solo author	0.375 (<0.001)
Affiliation rank	0.005 (<0.001)
Affiliation location international	0.243 (<0.001)
Author previous num. of publications	-0.109 (<0.001)
Author log citations	-0.022 (<0.001)
Publication year 2014	0.080 (<0.001)
Publication year 2015	0.171 (<0.001)
Publication year 2016	0.226 (<0.001)
Publication year 2017	0.438 (<0.001)
Publication year 2018	0.650 (<0.001)
Number of authors	-0.002 (<0.001)
Journal impact factor	0.022 (<0.001)
Social Sciences	0.322 (<0.001)
Materials Science	0.057 (0.039)
Engineering	-0.099 (<0.001)
Chemistry	-0.038 (0.064)
Biochemistry, Genetics & Molecular Biology	-0.095 (<0.001)
Medicine	-0.396 (<0.001)
Nursing	-0.026 (0.200)
Agricultural and Biological Sciences	0.371 (<0.001)
Pharmacology, Toxicology & Pharmaceutics	-0.268 (<0.001)
Neuroscience	-0.030 (0.054)
Business, Management and Accounting	-0.088 (0.002)
Economics, Econometrics and Finance	-0.167 (<0.001)
Chemical Engineering	0.122 (<0.001)
Physics and Astronomy	-0.235 (<0.001)
Computer Science	-0.065 (0.003)
Decision Sciences	-0.312 (<0.001)
Health Professions	0.519 (<0.001)
Psychology	-0.349 (<0.001)
Immunology and Microbiology	0.031 (0.080)
Dentistry	-0.698 (<0.001)
Earth and Planetary Sciences	-0.093 (<0.001)
Environmental Science	0.122 (<0.001)
Mathematics	0.133 (<0.001)
Arts and Humanities	0.008 (0.008)
Energy	-0.031 (0.412)
Veterinary	-0.616 (<0.001)
General	0.507 (<0.001)
Intercept	0.312 (<0.001)
Pseudo- R^2	0.32
Observations	618,742

TABLE S4. The same as Model 5 in Table S1, but the model is fitted to 12,205,072 observations including **Woman, Man, and Gender-neutral names**. Authors with gender-neutral names are less likely than those with masculine and feminine names to self-promote their papers. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	-0.348 (<0.001)
Unisex	-0.753 (<0.001)
Authorship first position	0.364 (<0.001)
Authorship middle position	-0.640 (<0.001)
Authorship solo author	0.844 (<0.001)
Affiliation rank	-0.048 (<0.001)
Affiliation location international	0.049 (<0.001)
Author previous num. of publications	0.286 (<0.001)
Author previous num. of publications (squared)	-0.021 (<0.001)
Author log citations	-0.031 (<0.001)
Publication year 2014	0.305 (<0.001)
Publication year 2015	0.604 (<0.001)
Publication year 2016	0.845 (<0.001)
Publication year 2017	1.176 (<0.001)
Publication year 2018	1.542 (<0.001)
Number of authors	-0.002 (<0.001)
Journal impact factor	0.034 (<0.001)
Social Sciences	0.907 (<0.001)
Materials Science	-0.147 (<0.001)
Engineering	-0.469 (<0.001)
Chemistry	-0.429 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.171 (<0.001)
Medicine	-0.274 (<0.001)
Nursing	0.263 (<0.001)
Agricultural and Biological Sciences	0.494 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.571 (<0.001)
Neuroscience	-0.104 (<0.001)
Business, Management and Accounting	-0.178 (<0.001)
Economics, Econometrics and Finance	-0.493 (<0.001)
Chemical Engineering	0.062 (<0.001)
Physics and Astronomy	-0.723 (<0.001)
Computer Science	0.131 (<0.001)
Decision Sciences	-0.780 (<0.001)
Health Professions	0.904 (<0.001)
Psychology	-0.577 (<0.001)
Immunology and Microbiology	-0.177 (<0.001)
Dentistry	-1.254 (<0.001)
Earth and Planetary Sciences	-0.075 (<0.001)
Environmental Science	0.169 (<0.001)
Mathematics	-0.238 (<0.001)
Arts and Humanities	-0.228 (<0.001)
Energy	-0.374 (<0.001)
Veterinary	-0.858 (<0.001)
General	0.392 (<0.001)
Intercept	-4.442 (<0.001)
Pseudo- R^2	0.52
Observations	12,205,072

TABLE S5. The same as Model 5 in Table S1 fitted to 9,793,077 observations. The data exclude authors whose names are predicted as **East Asian ethnicities** based on *Ethnea* [27]. The result shows that our finding is robust for authors whose names display clear gender signals. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	−0.332 (<0.001)
Authorship first position	0.361 (<0.001)
Authorship middle position	−0.623 (<0.001)
Authorship solo author	0.766 (<0.001)
Affiliation rank	−0.044 (<0.001)
Affiliation location international	0.096 (<0.001)
Author previous num. of publications	0.280 (<0.001)
Author previous num. of publications (squared)	−0.020 (<0.001)
Author log citations	−0.038 (<0.001)
Publication year 2014	0.315 (<0.001)
Publication year 2015	0.626 (<0.001)
Publication year 2016	0.868 (<0.001)
Publication year 2017	1.210 (<0.001)
Publication year 2018	1.578 (<0.001)
Number of authors	−0.002 (<0.001)
Journal impact factor	0.032 (<0.001)
Social Sciences	0.815 (<0.001)
Materials Science	−0.078 (<0.001)
Engineering	−0.440 (<0.001)
Chemistry	−0.415 (<0.001)
Biochemistry, Genetics & Molecular Biology	−0.137 (<0.001)
Medicine	−0.298 (<0.001)
Nursing	0.228 (<0.001)
Agricultural and Biological Sciences	0.457 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	−0.562 (<0.001)
Neuroscience	−0.117 (<0.001)
Business, Management and Accounting	−0.184 (<0.001)
Economics, Econometrics and Finance	−0.484 (<0.001)
Chemical Engineering	0.079 (<0.001)
Physics and Astronomy	−0.759 (<0.001)
Computer Science	0.144 (<0.001)
Decision Sciences	−0.697 (<0.001)
Health Professions	0.870 (<0.001)
Psychology	−0.581 (<0.001)
Immunology and Microbiology	−0.180 (<0.001)
Dentistry	−1.268 (<0.001)
Earth and Planetary Sciences	−0.119 (<0.001)
Environmental Science	0.164 (<0.001)
Mathematics	−0.292 (<0.001)
Arts and Humanities	−0.244 (<0.001)
Energy	−0.350 (<0.001)
Veterinary	−0.914 (<0.001)
General	0.382 (<0.001)
Intercept	−4.299 (<0.001)
Pseudo- R^2	0.50
Observations	9,793,077

TABLE S6. The same as Model 5 in Table S1, but the dependent variable is **whether the author is the first to self-promote the paper**. The model is fitted to 11,396,752 observations. This analysis shows that the gender gap is not driven by potential self-promotion dynamics in mixed-gender teams. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Is the first self-promoter?
Woman	-0.323 (<0.001)
Authorship first position	0.352 (<0.001)
Authorship middle position	-0.757 (<0.001)
Authorship solo author	0.930 (<0.001)
Affiliation rank	-0.049 (<0.001)
Affiliation location international	0.053 (<0.001)
Author previous num. of publications	0.299 (<0.001)
Author previous num. of publications (squared)	-0.023 (<0.001)
Author log citations	-0.013 (<0.001)
Publication year 2014	0.287 (<0.001)
Publication year 2015	0.562 (<0.001)
Publication year 2016	0.775 (<0.001)
Publication year 2017	1.051 (<0.001)
Publication year 2018	1.348 (<0.001)
Number of authors	-0.015 (<0.001)
Journal impact factor	0.022 (<0.001)
Social Sciences	0.804 (<0.001)
Materials Science	-0.112 (<0.001)
Engineering	-0.410 (<0.001)
Chemistry	-0.349 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.139 (<0.001)
Medicine	-0.253 (<0.001)
Nursing	0.213 (<0.001)
Agricultural and Biological Sciences	0.408 (<0.001)
Pharmacology, Toxicology & Pharmaceutics	-0.510 (<0.001)
Neuroscience	-0.095 (<0.001)
Business, Management and Accounting	-0.140 (<0.001)
Economics, Econometrics and Finance	-0.440 (<0.001)
Chemical Engineering	0.073 (<0.001)
Physics and Astronomy	-0.658 (<0.001)
Computer Science	0.113 (<0.001)
Decision Sciences	-0.703 (<0.001)
Health Professions	0.683 (<0.001)
Psychology	-0.523 (<0.001)
Immunology and Microbiology	-0.142 (<0.001)
Dentistry	-1.198 (<0.001)
Earth and Planetary Sciences	-0.074 (<0.001)
Environmental Science	0.144 (<0.001)
Mathematics	-0.263 (<0.001)
Arts and Humanities	-0.207 (<0.001)
Energy	-0.279 (<0.001)
Veterinary	-0.821 (<0.001)
General	0.268 (<0.001)
Intercept	-4.336 (<0.001)
Pseudo- R^2	0.86
Observations	11,396,752

TABLE S7. The same as Model 5 in Table S1. The model is fitted to 176,097 observations for which **the paper is solo-authored**. This analysis shows that our finding is robust for single-authored papers. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	−0.092 (<0.001)
Affiliation rank	−0.025 (<0.001)
Affiliation location international	0.169 (<0.001)
Author previous num. of publications	0.267 (<0.001)
Author previous num. of publications (squared)	−0.020 (<0.001)
Author log citations	−0.098 (<0.001)
Publication year 2014	0.417 (<0.001)
Publication year 2015	0.697 (<0.001)
Publication year 2016	0.898 (<0.001)
Publication year 2017	1.214 (<0.001)
Publication year 2018	1.471 (<0.001)
Journal impact factor	0.025 (<0.001)
Social Sciences	0.497 (<0.001)
Materials Science	0.156 (0.120)
Engineering	−0.119 (0.033)
Chemistry	−0.537 (<0.001)
Biochemistry, Genetics & Molecular Biology	−0.067 (0.038)
Medicine	−0.373 (<0.001)
Nursing	0.059 (0.288)
Agricultural and Biological Sciences	0.237 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	−0.315 (<0.001)
Neuroscience	0.036 (0.498)
Business, Management and Accounting	−0.196 (<0.001)
Economics, Econometrics and Finance	−0.461 (<0.001)
Chemical Engineering	0.227 (0.045)
Physics and Astronomy	−1.124 (<0.001)
Computer Science	0.166 (<0.001)
Decision Sciences	−0.418 (<0.001)
Health Professions	0.527 (<0.001)
Psychology	−0.414 (<0.001)
Immunology and Microbiology	−0.017 (0.779)
Dentistry	−1.490 (<0.001)
Earth and Planetary Sciences	−0.152 (0.002)
Environmental Science	0.091 (0.016)
Mathematics	−1.112 (<0.001)
Arts and Humanities	−0.329 (<0.001)
Energy	−0.166 (0.076)
Veterinary	−1.433 (<0.001)
General	−0.015 (0.802)
Intercept	−2.958 (<0.001)
Pseudo- R^2	0.15
Observations	176,097

TABLE S8. The mean of all control variables by gender in the full data vs. the matched samples in the propensity score matching analysis shown in Table S9.

Feature	Before matching		After matching	
	Woman	Man	Woman	Man
Authorship first position	0.19	0.15	0.19	0.19
Authorship middle position	0.66	0.64	0.66	0.66
Authorship solo author	0.01	0.02	0.01	0.01
Affiliation rank	4.35	4.28	4.35	4.38
Affiliation location international	0.65	0.65	0.65	0.65
Author previous num. of publications	4.08	5.41	4.08	4.04
Author log citations	5.93	7.84	5.93	5.85
Publication year 2014	0.15	0.15	0.15	0.15
Publication year 2015	0.17	0.17	0.17	0.17
Publication year 2016	0.18	0.18	0.18	0.18
Publication year 2017	0.19	0.18	0.19	0.19
Publication year 2018	0.18	0.19	0.18	0.18
Number of authors	30.23	59.76	30.23	27.18
Journal impact factor	5.23	5.59	5.23	5.25
Social Sciences	0.09	0.06	0.09	0.08
Materials Science	0.03	0.04	0.03	0.03
Engineering	0.03	0.05	0.03	0.03
Chemistry	0.05	0.06	0.05	0.05
Biochemistry, Genetics & Molecular Biology	0.25	0.22	0.25	0.25
Medicine	0.46	0.40	0.46	0.45
Nursing	0.03	0.01	0.03	0.03
Agricultural and Biological Sciences	0.11	0.11	0.11	0.12
Pharmacology, Toxicology & Pharmaceutics	0.04	0.03	0.04	0.04
Neuroscience	0.05	0.05	0.05	0.05
Business, Management and Accounting	0.01	0.01	0.01	0.01
Economics, Econometrics and Finance	0.01	0.01	0.01	0.01
Chemical Engineering	0.02	0.02	0.02	0.02
Physics and Astronomy	0.04	0.09	0.04	0.04
Computer Science	0.01	0.03	0.01	0.02
Decision Sciences	0.00	0.00	0.00	0.00
Health Professions	0.01	0.01	0.01	0.01
Psychology	0.04	0.02	0.04	0.04
Immunology and Microbiology	0.06	0.05	0.06	0.06
Dentistry	0.00	0.00	0.00	0.00
Earth and Planetary Sciences	0.02	0.04	0.02	0.02
Environmental Science	0.04	0.05	0.04	0.05
Mathematics	0.01	0.02	0.01	0.01
Arts and Humanities	0.00	0.00	0.00	0.00
Energy	0.01	0.01	0.01	0.01
Veterinary	0.01	0.01	0.01	0.01
General	0.01	0.02	0.01	0.01

TABLE S9. Logistic regression model that predicts whether the author has self-promoted the paper or not. The model is fitted to 8,051,300 **matched observations based on the propensity score** using the nearest neighbor matching (one man observation is matched with one woman observation). Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	-0.313 (<0.001)
Authorship first position	0.338 (<0.001)
Authorship middle position	-0.619 (<0.001)
Authorship solo author	0.804 (<0.001)
Affiliation rank	-0.057 (<0.001)
Affiliation location international	0.107 (<0.001)
Author previous num. of publications	0.266 (<0.001)
Author previous num. of publications (squared)	-0.020 (<0.001)
Author log citations	-0.023 (<0.001)
Publication year 2014	0.324 (<0.001)
Publication year 2015	0.657 (<0.001)
Publication year 2016	0.897 (<0.001)
Publication year 2017	1.260 (<0.001)
Publication year 2018	1.603 (<0.001)
Number of authors	-0.001 (<0.001)
Journal impact factor	0.036 (<0.001)
Social Sciences	0.834 (<0.001)
Materials Science	-0.333 (<0.001)
Engineering	-0.441 (<0.001)
Chemistry	-0.602 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.224 (<0.001)
Medicine	-0.312 (<0.001)
Nursing	0.209 (<0.001)
Agricultural and Biological Sciences	0.522 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.667 (<0.001)
Neuroscience	-0.137 (<0.001)
Business, Management and Accounting	-0.231 (<0.001)
Economics, Econometrics and Finance	-0.448 (<0.001)
Chemical Engineering	0.063 (<0.001)
Physics and Astronomy	-0.845 (<0.001)
Computer Science	0.118 (<0.001)
Decision Sciences	-0.816 (<0.001)
Health Professions	0.861 (<0.001)
Psychology	-0.582 (<0.001)
Immunology and Microbiology	-0.190 (<0.001)
Dentistry	-1.421 (<0.001)
Earth and Planetary Sciences	-0.073 (<0.001)
Environmental Science	0.148 (<0.001)
Mathematics	-0.267 (<0.001)
Arts and Humanities	-0.222 (<0.001)
Energy	-0.414 (<0.001)
Veterinary	-0.851 (<0.001)
General	0.341 (<0.001)
Intercept	-4.174 (<0.001)
Pseudo- R^2	0.10
Observations	8,051,300

TABLE S10. The woman coefficient in a Model 5 (Table S1) fitted to all observations in each of the **four broad disciplines**. Note that a paper can belong to more than one discipline. The result shows that women self-promote less than men across difference research fields. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

Discipline	Number of Observations	Woman Coefficient (P value)
Social Sciences	804,121	-0.195 (<0.001)
Life Sciences	4,123,459	-0.466 (<0.001)
Physical Sciences	2,561,568	-0.284 (<0.001)
Health Sciences	4,952,083	-0.340 (<0.001)

TABLE S11. The same as Table S10, but each model is fitted to observations where the author is active on Twitter. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

Discipline	Number of Observations	Woman Coefficient (P value)
Social Sciences	77,120	-0.143 (<0.001)
Life Sciences	193,077	-0.218 (<0.001)
Physical Sciences	109,732	-0.169 (<0.001)
Health Sciences	257,918	-0.167 (<0.001)

TABLE S12. Coefficients of the logit model that predicts whether the author self-promoted the paper. The model includes an interaction term between gender and journal impact factor and is fitted to 11,396,752 (paper, author) observations. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	-0.275 (<0.001)
Journal impact factor	0.036 (<0.001)
Woman x Journal impact factor	-0.011 (<0.001)
Authorship first position	0.359 (<0.001)
Authorship middle position	-0.638 (<0.001)
Authorship solo author	0.827 (<0.001)
Affiliation rank	-0.048 (<0.001)
Affiliation location international	0.051 (<0.001)
Author previous num. of publications	0.288 (<0.001)
Author previous num. of publications (squared)	-0.021 (<0.001)
Author log citations	-0.031 (<0.001)
Publication year 2014	0.305 (<0.001)
Publication year 2015	0.606 (<0.001)
Publication year 2016	0.846 (<0.001)
Publication year 2017	1.178 (<0.001)
Publication year 2018	1.541 (<0.001)
Number of authors	-0.002 (<0.001)
Social Sciences	0.893 (<0.001)
Materials Science	-0.145 (<0.001)
Engineering	-0.462 (<0.001)
Chemistry	-0.425 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.167 (<0.001)
Medicine	-0.280 (<0.001)
Nursing	0.252 (<0.001)
Agricultural and Biological Sciences	0.491 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.569 (<0.001)
Neuroscience	-0.108 (<0.001)
Business, Management and Accounting	-0.177 (<0.001)
Economics, Econometrics and Finance	-0.483 (<0.001)
Chemical Engineering	0.067 (<0.001)
Physics and Astronomy	-0.725 (<0.001)
Computer Science	0.138 (<0.001)
Decision Sciences	-0.774 (<0.001)
Health Professions	0.903 (<0.001)
Psychology	-0.581 (<0.001)
Immunology and Microbiology	-0.175 (<0.001)
Dentistry	-1.261 (<0.001)
Earth and Planetary Sciences	-0.082 (<0.001)
Environmental Science	0.169 (<0.001)
Mathematics	-0.238 (<0.001)
Arts and Humanities	-0.227 (<0.001)
Energy	-0.370 (<0.001)
Veterinary	-0.875 (<0.001)
General	0.379 (<0.001)
Intercept	-4.447 (<0.001)
Pseudo- R^2	0.51
Observations	11,396,752

TABLE S13. Coefficients of the logit model that predicts whether the author self-promoted the paper. The model includes an interaction term between gender and number of publications, and is fitted to 11,396,752 (paper, author) observations. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	-0.410 (<0.001)
Author previous num. of publications	0.277 (<0.001)
Author previous num. of publications (squared)	-0.020 (<0.001)
Woman x Author previous num. of publications	0.021 (<0.001)
Woman x Author previous num. of publications (squared)	-0.001 (0.027)
Authorship first position	0.362 (<0.001)
Authorship middle position	-0.638 (<0.001)
Authorship solo author	0.828 (<0.001)
Affiliation rank	-0.047 (<0.001)
Affiliation location international	0.052 (<0.001)
Author log citations	-0.031 (<0.001)
Publication year 2014	0.305 (<0.001)
Publication year 2015	0.606 (<0.001)
Publication year 2016	0.846 (<0.001)
Publication year 2017	1.178 (<0.001)
Publication year 2018	1.541 (<0.001)
Journal impact factor	0.033 (<0.001)
Number of authors	-0.002 (<0.001)
Social Sciences	0.898 (<0.001)
Materials Science	-0.144 (<0.001)
Engineering	-0.464 (<0.001)
Chemistry	-0.425 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.167 (<0.001)
Medicine	-0.280 (<0.001)
Nursing	0.259 (<0.001)
Agricultural and Biological Sciences	0.491 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.568 (<0.001)
Neuroscience	-0.108 (<0.001)
Business, Management and Accounting	-0.181 (<0.001)
Economics, Econometrics and Finance	-0.490 (<0.001)
Chemical Engineering	0.067 (<0.001)
Physics and Astronomy	-0.728 (<0.001)
Computer Science	0.134 (<0.001)
Decision Sciences	-0.779 (<0.001)
Health Professions	0.905 (<0.001)
Psychology	-0.580 (<0.001)
Immunology and Microbiology	-0.175 (<0.001)
Dentistry	-1.263 (<0.001)
Earth and Planetary Sciences	-0.084 (<0.001)
Environmental Science	0.168 (<0.001)
Mathematics	-0.242 (<0.001)
Arts and Humanities	-0.230 (<0.001)
Energy	-0.370 (<0.001)
Veterinary	-0.872 (<0.001)
General	0.388 (<0.001)
Intercept	-4.399 (<0.001)
Pseudo- R^2	0.51
Observations	11,396,752

TABLE S14. Coefficients of the logit model that predicts whether the author self-promoted the paper. The model includes an interaction term between gender and affiliation rank and is fitted to 11,396,752 (paper, author) observations. Note that we inverted the affiliation rank bins in Fig. 3D in the main text so that a larger bin indicates a higher rank decile in the plot. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Self-promotion = True
Woman	-0.357 (<0.001)
Affiliation rank	-0.048 (<0.001)
Woman x Affiliation rank	0.002 (0.02)
Affiliation location international	0.052 (<0.001)
Authorship first position	0.361 (<0.001)
Authorship middle position	-0.638 (<0.001)
Authorship solo author	0.828 (<0.001)
Author previous num. of publications	0.286 (<0.001)
Author previous num. of publications (squared)	-0.021 (<0.001)
Author log citations	-0.031 (<0.001)
Publication year 2014	0.305 (<0.001)
Publication year 2015	0.607 (<0.001)
Publication year 2016	0.846 (<0.001)
Publication year 2017	1.178 (<0.001)
Publication year 2018	1.542 (<0.001)
Journal impact factor	0.033 (<0.001)
Number of authors	-0.002 (<0.001)
Social Sciences	0.898 (<0.001)
Materials Science	-0.144 (<0.001)
Engineering	-0.464 (<0.001)
Chemistry	-0.425 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.167 (<0.001)
Medicine	-0.279 (<0.001)
Nursing	0.259 (<0.001)
Agricultural and Biological Sciences	0.490 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.568 (<0.001)
Neuroscience	-0.108 (<0.001)
Business, Management and Accounting	-0.181 (<0.001)
Economics, Econometrics and Finance	-0.490 (<0.001)
Chemical Engineering	0.067 (<0.001)
Physics and Astronomy	-0.728 (<0.001)
Computer Science	0.134 (<0.001)
Decision Sciences	-0.779 (<0.001)
Health Professions	0.905 (<0.001)
Psychology	-0.580 (<0.001)
Immunology and Microbiology	-0.175 (<0.001)
Dentistry	-1.263 (<0.001)
Earth and Planetary Sciences	-0.084 (<0.001)
Environmental Science	0.168 (<0.001)
Mathematics	-0.242 (<0.001)
Arts and Humanities	-0.229 (<0.001)
Energy	-0.369 (<0.001)
Veterinary	-0.872 (<0.001)
General	0.387 (<0.001)
Intercept	-4.423 (<0.001)
Pseudo- R^2	0.51
Observations	11,396,752

TABLE S15. Negative binomial regression model that predicts **the total number of tweets of a paper**. The model controls for the author's follower count on Twitter and includes an interaction term between gender and self-promotion. The model is fitted to 618,742 (paper, author) observations where the author is active on Twitter at the paper's publication date. Note that the squared term for *Author previous num. of publications* is dropped because it is linearly correlated with the dependent variable. Paper random effects are not included in the model due to computational infeasibility. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Number of tweets for paper
Woman	0.148 (<0.001)
Self-promotion = True	1.235 (<0.001)
Woman x (Self-promotion = True)	-0.127 (<0.001)
Authorship first position	0.110 (<0.001)
Authorship middle position	0.290 (<0.001)
Authorship solo author	-0.206 (<0.001)
Affiliation rank	-0.008 (<0.001)
Affiliation location international	0.069 (<0.001)
Author previous num. of publications	-0.102 (<0.001)
Author log citations	0.050 (<0.001)
Author log follower count	0.095 (<0.001)
Publication year 2014	0.239 (<0.001)
Publication year 2015	0.399 (<0.001)
Publication year 2016	0.531 (<0.001)
Publication year 2017	0.797 (<0.001)
Publication year 2018	0.924 (<0.001)
Number of authors	0.001 (<0.001)
Journal impact factor	0.065 (<0.001)
Social Sciences	-0.116 (<0.001)
Materials Science	-0.532 (<0.001)
Engineering	-0.264 (<0.001)
Chemistry	-0.971 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.143 (<0.001)
Medicine	-0.035 (<0.001)
Nursing	0.212 (<0.001)
Agricultural and Biological Sciences	0.332 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.420 (<0.001)
Neuroscience	-0.008 (0.349)
Business, Management and Accounting	-0.408 (<0.001)
Economics, Econometrics and Finance	-0.129 (<0.001)
Chemical Engineering	0.091 (<0.001)
Physics and Astronomy	-0.583 (<0.001)
Computer Science	-0.161 (<0.001)
Decision Sciences	-0.438 (<0.001)
Health Professions	0.672 (<0.001)
Psychology	-0.101 (<0.001)
Immunology and Microbiology	-0.007 (0.441)
Dentistry	-0.677 (<0.001)
Earth and Planetary Sciences	-0.416 (<0.001)
Environmental Science	-0.168 (<0.001)
Mathematics	0.478 (<0.001)
Arts and Humanities	-0.151 (<0.001)
Energy	-0.588 (<0.001)
Veterinary	-0.661 (<0.001)
General	0.449 (<0.001)
Intercept	0.364 (<0.001)
Observations	618,742

TABLE S16. The coefficients of Woman and “Woman x (Self promotion = True)” in the negative binomial regression model shown in Table S15. Separate regressions are fitted to data on papers published in each year. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

Publication year	<i>Dependent variable:</i> Number of tweets for paper	
	Woman coeff.	Woman x (Self promotion = True) coeff.
2013	0.272 (<0.001)	-0.170 (<0.001)
2014	0.169 (<0.001)	-0.064 (0.039)
2015	0.158 (<0.001)	-0.131 (<0.001)
2016	0.179 (<0.001)	-0.191 (<0.001)
2017	0.130 (<0.001)	-0.095 (<0.001)
2018	0.110 (<0.001)	-0.066 (<0.001)

TABLE S17. The same as Table S15, but is fitted to 20,216 observations for which the paper is solo-authored and the author is active on Twitter. The “Woman x (Self-promotion = True)” coefficient is still negative but non-significant, possibly due to a small number of observations. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Number of tweets for paper
Woman	0.144 (<0.001)
Self-promotion = True	1.451 (<0.001)
Woman x (Self-promotion = True)	-0.053 (0.181)
Affiliation rank	-0.014 (<0.001)
Affiliation location international	0.004 (0.827)
Author previous num. of publications	-0.077 (<0.001)
Author log citations	0.064 (<0.001)
Author log follower count	0.154 (<0.001)
Publication year 2014	0.211 (<0.001)
Publication year 2015	0.458 (<0.001)
Publication year 2016	0.526 (<0.001)
Publication year 2017	0.790 (<0.001)
Publication year 2018	0.964 (<0.001)
Journal impact factor	0.035 (<0.001)
Social Sciences	-0.039 (0.098)
Materials Science	-0.754 (<0.001)
Engineering	-0.273 (<0.001)
Chemistry	-0.868 (<0.001)
Biochemistry, Genetics & Molecular Biology	0.072 (0.059)
Medicine	-0.028 (0.275)
Nursing	0.195 (0.003)
Agricultural and Biological Sciences	0.250 (<0.001)
Pharmacology, Toxicology & Pharmaceutics	0.188 (0.070)
Neuroscience	0.311 (<0.001)
Business, Management and Accounting	-0.174 (<0.001)
Economics, Econometrics and Finance	0.166 (<0.001)
Chemical Engineering	1.121 (<0.001)
Physics and Astronomy	-0.103 (0.216)
Computer Science	0.400 (<0.001)
Decision Sciences	-0.353 (0.005)
Health Professions	1.103 (<0.001)
Psychology	0.139 (0.003)
Immunology and Microbiology	0.330 (<0.001)
Dentistry	-0.473 (0.257)
Earth and Planetary Sciences	-0.070 (0.245)
Environmental Science	-0.031 (0.458)
Mathematics	0.003 (0.968)
Arts and Humanities	-0.318 (<0.001)
Energy	-0.413 (<0.001)
Veterinary	-0.337 (0.175)
General	0.343 (<0.001)
Intercept	-0.843 (<0.001)
Observations	20,216

TABLE S18. The same as Table S15, but the self-promotion variable is defined as an author advertising his or her paper within one day after the paper’s publication. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Number of tweets for paper
Woman	0.093 (<0.001)
Self-promotion = True	0.744 (<0.001)
Woman x (Self-promotion = True)	−0.089 (<0.001)
Authorship first position	0.180 (<0.001)
Authorship middle position	0.172 (<0.001)
Authorship solo author	−0.099 (<0.001)
Affiliation rank	−0.007 (<0.001)
Affiliation location international	0.097 (<0.001)
Author previous num. of publications	−0.109 (<0.001)
Author log citations	0.039 (<0.001)
Author log follower count	0.135 (<0.001)
Publication year 2014	0.212 (<0.001)
Publication year 2015	0.351 (<0.001)
Publication year 2016	0.487 (<0.001)
Publication year 2017	0.776 (<0.001)
Publication year 2018	0.904 (<0.001)
Number of authors	0.001 (<0.001)
Journal impact factor	0.066 (<0.001)
Social Sciences	−0.096 (<0.001)
Materials Science	−0.490 (<0.001)
Engineering	−0.345 (<0.001)
Chemistry	−0.990 (<0.001)
Biochemistry, Genetics & Molecular Biology	−0.116 (<0.001)
Medicine	−0.103 (<0.001)
Nursing	0.155 (<0.001)
Agricultural and Biological Sciences	0.429 (<0.001)
Pharmacology, Toxicology & Pharmaceutics	−0.480 (<0.001)
Neuroscience	−0.008 (0.343)
Business, Management and Accounting	−0.389 (<0.001)
Economics, Econometrics and Finance	−0.080 (<0.001)
Chemical Engineering	0.121 (<0.001)
Physics and Astronomy	−0.613 (<0.001)
Computer Science	−0.147 (<0.001)
Decision Sciences	−0.485 (<0.001)
Health Professions	0.726 (<0.001)
Psychology	−0.191 (<0.001)
Immunology and Microbiology	−0.014 (0.164)
Dentistry	−0.691 (<0.001)
Earth and Planetary Sciences	−0.484 (<0.001)
Environmental Science	−0.251 (<0.001)
Mathematics	0.557 (<0.001)
Arts and Humanities	−0.147 (<0.001)
Energy	−0.367 (<0.001)
Veterinary	−0.801 (<0.001)
General	0.478 (<0.001)
Intercept	0.599 (<0.001)
Observations	618,742

TABLE S19. The same as Table S15, but the dependent variable is the number of scientists (“researcher”) who have mentioned each paper. Note that the types of Twitter audiences are categorized by in-house experts from Altmetric. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Number of scientists for paper
Woman	0.136 (<0.001)
Self-promotion = True	1.272 (<0.001)
Woman x (Self-promotion = True)	−0.088 (<0.001)
Authorship first position	0.062 (<0.001)
Authorship middle position	0.304 (<0.001)
Authorship solo author	−0.262 (<0.001)
Affiliation rank	−0.020 (<0.001)
Affiliation location international	0.045 (<0.001)
Author previous num. of publications	−0.189 (<0.001)
Author log citations	0.098 (<0.001)
Author log follower count	0.096 (<0.001)
Publication year 2014	0.308 (<0.001)
Publication year 2015	0.456 (<0.001)
Publication year 2016	0.623 (<0.001)
Publication year 2017	0.925 (<0.001)
Publication year 2018	1.099 (<0.001)
Number of authors	0.001 (<0.001)
Journal impact factor	0.062 (<0.001)
Social Sciences	−0.091 (<0.001)
Materials Science	−0.487 (<0.001)
Engineering	−0.453 (<0.001)
Chemistry	−0.696 (<0.001)
Biochemistry, Genetics & Molecular Biology	0.177 (<0.001)
Medicine	−0.423 (<0.001)
Nursing	0.268 (<0.001)
Agricultural and Biological Sciences	0.401 (<0.001)
Pharmacology, Toxicology & Pharmaceutics	−0.568 (<0.001)
Neuroscience	−0.036 (<0.001)
Business, Management and Accounting	−0.380 (<0.001)
Economics, Econometrics and Finance	−0.200 (<0.001)
Chemical Engineering	0.122 (<0.001)
Physics and Astronomy	−0.254 (<0.001)
Computer Science	−0.101 (<0.001)
Decision Sciences	−0.149 (<0.001)
Health Professions	0.879 (<0.001)
Psychology	−0.113 (<0.001)
Immunology and Microbiology	0.107 (<0.001)
Dentistry	−1.440 (<0.001)
Earth and Planetary Sciences	−0.385 (<0.001)
Environmental Science	−0.130 (<0.001)
Mathematics	0.394 (<0.001)
Arts and Humanities	−0.230 (<0.001)
Energy	−0.904 (<0.001)
Veterinary	−1.053 (<0.001)
General	0.572 (<0.001)
Intercept	−1.086 (<0.001)
Observations	618,742

TABLE S20. The same as Table S15, but the dependent variable is the number of non-scientists (including “member of the public,” “practitioner,” and “science communicator”) who have mentioned each paper. Note that the types of Twitter audiences are categorized by in-house experts from Altmetric. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	<i>Dependent variable:</i>
	Number of non-scientists for paper
Woman	0.145 (<0.001)
Self-promotion = True	1.121 (<0.001)
Woman x (Self-promotion = True)	-0.125 (<0.001)
Authorship first position	0.126 (<0.001)
Authorship middle position	0.278 (<0.001)
Authorship solo author	-0.182 (<0.001)
Affiliation rank	-0.010 (<0.001)
Affiliation location international	0.066 (<0.001)
Author previous num. of publications	-0.079 (<0.001)
Author log citations	0.040 (<0.001)
Author log follower count	0.092 (<0.001)
Publication year 2014	0.228 (<0.001)
Publication year 2015	0.378 (<0.001)
Publication year 2016	0.495 (<0.001)
Publication year 2017	0.734 (<0.001)
Publication year 2018	0.864 (<0.001)
Number of authors	0.001 (<0.001)
Journal impact factor	0.064 (<0.001)
Social Sciences	-0.140 (<0.001)
Materials Science	-0.621 (<0.001)
Engineering	-0.188 (<0.001)
Chemistry	-1.176 (<0.001)
Biochemistry, Genetics & Molecular Biology	-0.220 (<0.001)
Medicine	0.046 (<0.001)
Nursing	0.293 (<0.001)
Agricultural and Biological Sciences	0.338 (<0.001)
Pharmacology, Toxicology & Pharmaceuticals	-0.344 (<0.001)
Neuroscience	0.068 (<0.001)
Business, Management and Accounting	-0.459 (<0.001)
Economics, Econometrics and Finance	-0.019 (0.289)
Chemical Engineering	0.071 (<0.001)
Physics and Astronomy	-0.577 (<0.001)
Computer Science	-0.213 (<0.001)
Decision Sciences	-0.402 (<0.001)
Health Professions	0.646 (<0.001)
Psychology	-0.008 (0.499)
Immunology and Microbiology	-0.076 (<0.001)
Dentistry	-0.753 (<0.001)
Earth and Planetary Sciences	-0.386 (<0.001)
Environmental Science	-0.192 (<0.001)
Mathematics	0.285 (<0.001)
Arts and Humanities	-0.116 (<0.001)
Energy	-0.660 (<0.001)
Veterinary	-0.518 (<0.001)
General	0.505 (<0.001)
Intercept	0.036 (0.019)
Observations	618,742

F1 Score	China	non-China	All Countries
Woman	0.61	0.90	0.83
Man	0.80	0.98	0.95

TABLE S21. The gender prediction accuracy based on authors' self-reported gender in the IOP Publishing data. There are 71,869 authors from China and 361,019 authors from outside of China. Here we list authors from China as a separate group because Chinese names typically do not encode a clear gender signal when written in English characters [16], and China is the largest single group in this data. As expected, the prediction accuracy for Chinese names is lower than for non-Chinese names, but the overall F1 score is close to 0.9.

TABLE S22. Coefficients of all variables in our Model 5 that predicts **whether the author self-promoted the paper**. A separate model is fitted to the first authors in Social Sciences and the last authors in Natural Sciences. Coefficients for 26 Scopus Subject Areas are included in the model but omitted here for clarity. Significance test is based on the two-sided Wald test. P values are shown in parentheses.

	Social Sci. (First Auth.)	Natural Sci. (Last Auth.)
Woman	-0.209 (<0.001)	-0.241 (<0.001)
Author previous num. of publications	0.395 (<0.001)	0.291 (<0.001)
Author previous num. of publications (squared)	-0.025 (<0.001)	-0.030 (<0.001)
Affiliation rank	-0.032 (<0.001)	-0.066 (<0.001)
Affiliation location international	0.250 (<0.001)	0.073 (<0.001)
Number of authors	-0.041 (<0.001)	-0.004 (<0.001)
Journal impact factor	0.090 (<0.001)	0.033 (<0.001)
Author log citations	-0.093 (<0.001)	0.037 (<0.001)
Publication year 2014	0.312 (<0.001)	0.354 (<0.001)
Publication year 2015	0.672 (<0.001)	0.693 (<0.001)
Publication year 2016	0.937 (<0.001)	0.965 (<0.001)
Publication year 2017	1.199 (<0.001)	1.344 (<0.001)
Publication year 2018	1.561 (<0.001)	1.706 (<0.001)
Intercept	-3.173 (<0.001)	-4.489 (<0.001)
Observations	205,398	1,751,444

TABLE S23. The “woman” coefficient in a Model 5 in Table S1 that predicts **whether the author self-promoted the paper**. A separate model is fitted for each of the 26 Scopus Subject Areas. Women self-promote less often than men in all subfields except in “Dentistry” and “Earth and Planetary Sciences”. Significance test is based on the two-sided Wald test. P values are shown in the last column.

Subfield	Women’s coefficient	Std. Dev.	P value
1 Social Sciences	-0.19	0.01	<0.001
2 Materials Science	-0.44	0.03	<0.001
3 Engineering	-0.34	0.03	<0.001
4 Chemistry	-0.46	0.02	<0.001
5 Biochemistry, Genetics and Molecular Biology	-0.59	0.01	<0.001
6 Medicine	-0.34	0.01	<0.001
7 Nursing	-0.22	0.02	<0.001
8 Agricultural and Biological Sciences	-0.34	0.01	<0.001
9 Pharmacology, Toxicology and Pharmaceuticals	-0.50	0.03	<0.001
10 Neuroscience	-0.44	0.02	<0.001
11 Business, Management and Accounting	-0.22	0.03	<0.001
12 Economics, Econometrics and Finance	-0.20	0.03	<0.001
13 Chemical Engineering	-0.44	0.03	<0.001
14 Physics and Astronomy	-0.39	0.03	<0.001
15 Computer Science	-0.27	0.03	<0.001
16 Decision Sciences	-0.31	0.06	<0.001
17 Health Professions	-0.53	0.03	<0.001
18 Psychology	-0.28	0.02	<0.001
19 Immunology and Microbiology	-0.53	0.02	<0.001
20 Dentistry	-0.21	0.11	0.067
21 Earth and Planetary Sciences	-0.02	0.02	0.297
22 Environmental Science	-0.20	0.01	<0.001
23 Mathematics	-0.44	0.04	<0.001
24 Arts and Humanities	-0.14	0.04	<0.001
25 Energy	-0.14	0.04	0.002
26 Veterinary	-0.21	0.06	<0.001

-
- [1] Gender detection. <http://codingnews.info/post/gender-detection.html>, 2015.
 - [2] fuzzywuzzy. <https://github.com/seatgeek/fuzzywuzzy>, 2021.
 - [3] Genderize. <https://genderize.io/>, 2021.
 - [4] gendercomputer. <https://github.com/tue-mdse/genderComputer>, 2022.
 - [5] Altmetric.com. Our sources. <https://www.altmetric.com/about-us/our-data/our-sources/>, 2023.
 - [6] Altmetric.com. <https://www.altmetric.com/about-us/our-data/how-does-it-work/>, 2023.
 - [7] Sumit Kumar Banshal, Vivek Kumar Singh, Pranab K Muhuri, and Philipp Mayr. Disciplinary variations in altmetric coverage of scholarly articles. *arXiv:1910.04205*, 2019.
 - [8] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 2015.
 - [9] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1217–1230, 2017.
 - [10] Rodrigo Costas, Philippe Mongeon, Márcia R Ferreira, Jeroen van Honk, and Thomas Franssen. Large-scale identification and characterization of scholars on Twitter. *Quantitative Science Studies*, 1(2):771–791, 2020.
 - [11] Crossref. Crossref data. <https://www.crossref.org/services/event-data/>, 2023.
 - [12] Thomas Davidson, Dana Warmley, Michael Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, pages 512–515, 2017.
 - [13] Elsevier. All science journal classification codes. <http://bartoc.org/en/node/20290>, 2021.
 - [14] Denae Ford, Alisse Harkins, and Chris Parnin. Someone like me: How does peer parity influence participation of women on stack overflow? In *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 239–243. IEEE, 2017.
 - [15] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 216–225, 2014.
 - [16] Jizheng Jia and Qiyang Zhao. Gender prediction based on chinese name. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 676–683. Springer, 2019.
 - [17] Fariba Karimi, Claudia Wagner, Florian Lemmerich, Mohsen Jadidi, and Markus Strohmaier. Inferring gender from names on the web: A comparative evaluation of gender detection methods. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 53–54, 2016.
 - [18] Qing Ke, Yong-Yeol Ahn, and Cassidy R Sugimoto. A systematic identification and analysis of scientists on twitter. *PLOS ONE*, 12(4):e0175368, 2017.
 - [19] Marc J Lerchenmueller, Olav Sorenson, and Anupam B Jena. Gender differences in how scientists present the importance of their research: observational study. *British Medical Journal*, 367, 2019.
 - [20] Microsoft. How microsoft academic uses knowledge to address the problem of conflation and disambiguation. <https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-uses-knowledge-address-problem-conflation-disambiguation/>, 2018.
 - [21] Neil Millar, Bojan Batalo, and Brian Budgell. Trends in the use of promotional language (hype) in abstracts of successful national institutes of health grant applications, 1985-2020. *JAMA Network Open*, 5(8):e2228676–e2228676, 2022.
 - [22] Philippe Mongeon, Timothy D Bowman, and Rodrigo Costas. An open data set of scholars on Twitter. *Quantitative Science Studies*, 4(2):314–324, 2023.
 - [23] Hao Peng, Daniel M Romero, and Emőke-Ágnes Horvát. Dynamics of cross-platform attention to retracted papers. *Proceedings of the National Academy of Sciences*, 119(25):e2119086119, 2022.
 - [24] Jason Priem, Heather Piwowar, and Richard Orr. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv:2205.01833*, 2022.
 - [25] Lucía Santamaría and Helena Mihaljević. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, 4:e156, 2018.

- [26] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang. An overview of microsoft academic service (mas) and applications. In *WWW*, 2015.
- [27] Vetle I Torvik and Sneha Agarwal. Ethnea—an instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database. *International Symposium on Science of Science*, 2016.
- [28] Nicole A Vasilevsky, Mohammad Hosseini, Samantha Teplitzky, Violeta Ilik, Ehsan Mohammadi, Juliane Schneider, Barbara Kern, Julien Colomb, Scott C Edmunds, Karen Gutzman, et al. Is authorship sufficient for today’s collaborative research? a call for contributor roles. *Accountability in Research*, 28(1):23–43, 2021.
- [29] Orsolya Vásárhelyi, Igor Zakhlebin, Staša Milojević, and E.-Á. Horvát. Gender inequities in the online dissemination of scholars’ work. *Proceedings of the National Academy of Sciences*, 118(39):e2102945118, 2021.
- [30] Kuansan Wang, Zhihong Shen, Chi-Yuan Huang, Chieh-Han Wu, Darrin Eide, Yuxiao Dong, Junjie Qian, Anshul Kanakia, Alvin Chen, and Richard Rogahn. A review of microsoft academic services for science of science studies. *Frontiers in Big Data*, 2:45, 2019.