Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)

Are There Gender Differences in Professional Self-Promotion? An Empirical Case Study of LinkedIn Profiles among Recent MBA Graduates

Kristen M. Altenburger
Stanford University (USA)
kaltenb@stanford.edu

Rajlakshmi De
North Carolina State University (USA)
rde2@ncsu.edu

Kaylyn Frazier
Google Inc. (USA)
kfrazier@google.com

Nikolai Avteniev
LinkedIn (USA)
navteniev@linkedin.com

Jim Hamilton
LinkedIn (USA)
jihamilton@linkedin.com

Abstract

Women are more modest than men in expressing accomplishments, referred to as the "feminine modesty effect". Given the importance of highlighting accomplishments and skills for professional advancement, our research revisits the classical question of equal opportunity with a modern dataset to examine how women leverage LinkedIn, a professional social networking site. We first apply propensity score matching methods to identify a subset of similarly qualified female and male U.S. users who recently graduated/will be graduating (2011-2017) from a top-ranked MBA program as indicated on their LinkedIn profile. We then analyze gender differences in online self-promotion choices, an often overlooked aspect of understanding the role of gender in the professional hiring pipeline. Among matched subsets of female and male users, we find that females are less likely relative to males to utilize data fields that require writing in free-form such as the Summary and Job Description fields. However, we find for most universities that females and males are equally likely to include more structured data fields such as Honors and Skills, and for some universities females are more likely to include at least one Skill. This work begins to quantify gender biases in user-provided data and introduces important considerations for how self-presentation choices affect professional opportunity in online hiring platforms.

Introduction

The recent explosion of diverse online social networking platforms provides users an unprecedented opportunity to express themselves, share information, engage each other and communicate directly with organizations (Kietzmann et al. 2011). In the professional setting, self-presentation strategies are critical to career advancement (Rudman 1998). Goffman's seminal work (1959) explores self-presentation and how individuals engage in certain impression management behaviors. There are numerous studies specifically analyzing offline self-presentation in professional settings (Giacalone and Rosenfeld 2013; Roberts 2005), though research is only beginning to unravel how self-presentation translates to online professional settings (Haferkamp et al. 2012; Sievers et al. 2015).

The focus of this work is to isolate possible gender differences in how users present their professional selves. In this paper, we address this nuanced and often overlooked aspect of showcasing one's skills by measuring self-presentation differences between women and men in a matched sample of users on LinkedIn, a premier professional networking site with more than 467 million members. While previous efforts have examined self-promotion in other settings including scholarly self-citations (King et al. 2016), this work evaluates gender differences in self-promotion on a professional social networking site.

We apply traditional statistical matching tools for quasi-experimental design, propensity score matching (Rosenbaum and Rubin 1983), to identify similarly qualified female and male users on the LinkedIn platform and then examine potential gender differences in self-promotion. This paper observes that women are *less* likely relative to men to include profile fields like the Summary which require writing about oneself and identifies that across most universities there are no gender differences in the likelihood to list structured profile fields like Honors or Skills, and finds there are some universities where women are *more* likely relative to men to list a Skill. In the remaining sections, we begin by describing the dataset variables and the statistical matching methodology. This analysis is a small, first step forward in understanding gendered self-promotion in online social networks.

Data: LinkedIn Profiles

We received access to LinkedIn's data as part of the 2015 Economic Graph Challenge. The focus of our analysis is on U.S. users on LinkedIn who state they are recent MBA graduates/will be graduating (2011-2017) from a top 10 MBA program based on the March 2015 rankings from U.S. News & World Report. Recent MBA graduates in particular have a strong impetus to create and/or update their LinkedIn profiles during their MBA studies because it is a period of major career transition and networking. In addition, focusing on recent MBAs at specific universities attenuates the need to control for the entirety of users' job histories because the MBA serves as a baseline similarity feature on top of which we control for recent job history only. We conduct our matching approach at the university-level to account for heterogeneous gender effects due to differences across MBA programs both in terms of resources for developing an online professional presence and support for female students. We then limit our dataset to only U.S. users due to variations in cultural norms in self-promotional behavior.

The current data incorporates gender as the "treatment" condition, matching characteristic categories (i.e. geographic location, current industry, MBA graduation year, and number of years of pre-MBA experience), and the following self-promotion metrics: indicator for including a Summary, Summary length given a Summary is provided, indicator for including a Job Description, indicator for listing Skills, number of Skills given at least one Skill is listed, indicator for including an Honor, and number of Honors listed given at least one Honor is listed. We do not consider endorsements as a self-promotion field because endorsements also depend on others in a user's network to promote a particular skill. The goal of this analysis is to capture self-promotion behavior only.

LinkedIn users are never asked to provide their gender when creating a profile, and we rely on internally predicted gender labels from LinkedIn, dropping users with uncertain gender labels. Users input their geographic location and current industry, such as "Greater New York City Area" for their location and "Banking" for their industry. We use these variables in their existing categorization and only include U.S. users. When a user adds an MBA degree to their profile, they can input start and end years; we utilize the end year as the year of graduation from the MBA program. Finally, we compute a proxy for the number of years of pre-MBA experience by subtracting a user's year of graduation from an undergraduate degree from the start year of their MBA. The following analysis will only apply to users who list an MBA degree on their profile.

For the self-promotion metrics, we consider indicator variables for whether a user has utilized a particular profile section, including Summary, Job Description, Honors and Skills fields, which require additional time and initiative to complete. For the Summary field, among users who complete these fields, we also consider the number of words the user has provided. The Summary field is often one of the first sections on a user's profile, and allows users an open-ended space to write about themselves, whether that is a sentence describing their objective or multiple paragraphs detailing their professional accomplishments. The Honors field allows discrete entries of any accomplishments the user chooses to provide. For both the Honors and Skills fields, we consider whether they exist and if so, the number of entries. If female and male users were equally likely to self-promote, then given a female and male user with similar educational and work histories and similar backgrounds (such as where they're living) we'd expect to see a comparable likelihood to provide a Summary, Job Description, Skills, and Honors.

Methods: Propensity Score Matching

Propensity score matching methods traditionally aim to approximate a randomized experiment for observational studies where a researcher is interested in a causal inquiry. For our purposes, matching methods allow us to create subgroups of female and male users that appear only randomly different on their covariate matching features (i.e. geographic location, current industry, MBA graduation year, and number of years of pre-MBA experience). Due to gender being an "immutable characteristic" (Greiner and Rubin 2011), the interpretation of our results are associative only, and not causal. As explained in more detail in (Greiner and Rubin 2011; Holland 2003), a causal inquiry of the effect of a personal attribute like race or gender would mean randomizing these attributes at conception in order to truly quantify the effect of gender and control for all post-treatment variables. Therefore, causal inquiries are only feasible when randomizing perceived gender via names, and the interpretation of this work is associative only.

Following Rubin's recommendation (2007), we have distinct design and analysis phases to ensure the objectivity of our results. The focus of the design phase is to finalize our propensity score model so that we achieve a subset of users who are reasonably similar along their objective features without yet including any self-promotion data. Then the focus of the analysis phase is to merge in the self-promotion outcome data into our matched samples and compare gender differences across these fields.

The Design Phase: In this set-up, our N units are 2011-2017 MBA professionals at a top MBA program who have a LinkedIn profile at the time we access the data in December 2015. Our matching data fields are denoted by X_i, the "treatment" variable is gender denoted by W_i=1 for female and W_i=0 for male (dropping users with unknown predicted gender labels), and our outcome of interest is promoting in the k*th* data field where, for example, Y_{ik}=1 if the i*th* professional includes ancillary information such as a Summary and 0 otherwise. We temporarily remove our outcome data - all self-promotion fields - while in the Design Phase to ensure the objectivity of our matched sample.

The success of the propensity score approach is dependent on achieving balance in the covariate distributions between the female and male professionals. We explore several matching specifications and ultimately select 1-1 matching with a 0.1 caliper. Note that a caliper (Rosenbaum and Rubin 1985; Althauser and Rubin 1970) is defined as a permissible difference between female and male users to still be considered a match. The order of the female-male matching is done randomly and without replacement. We implement these matching models with the R MatchIt package, version 2.4-21 (Ho et al. 2007). We follow practical recommendations and best practices for matching discussed in (Rubin and Thomas 1996; Rubin 2006).

The Analysis Phase: After finalizing our matched sample, we then assess gender differences in the self-promotion metrics by fitting models that include gender and the matching features for each MBA program except Columbia University which we drop due to a small sample size in the final matched sample. For binary self-promotion data fields such as whether a user includes a Summary, we fit a Logistic regression model and confirm an adequate goodness-of-fit using the Hosmer-Lemeshow test. For count data such as the number of words in a Summary among users who include a

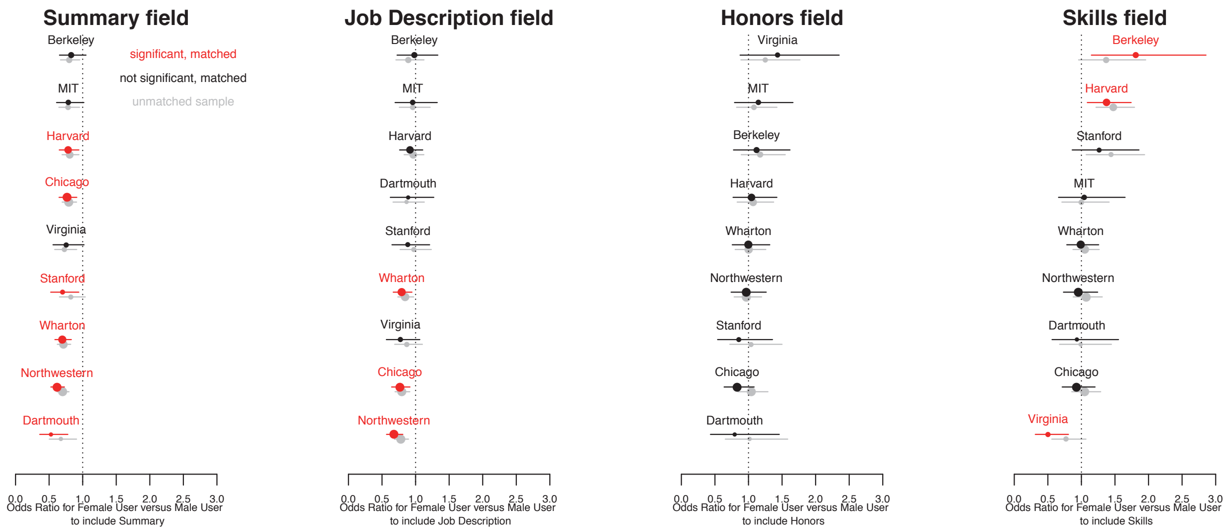

Figure 1: Odds Ratio Coefficient Estimates and 95% Confidence Intervals for Self-Promotion Fields.

Summary field on their profile, we assess gender differences via a randomization test (Ernst and others 2004). We first compute the observed absolute mean difference in counts between females and males, and then compute a null distribution of absolute mean differences by permuting gender labels. To determine statistical significance, we then compute the proportion of null absolute differences that are greater than the observed one, and determine statistical significance if this proportion is less than 0.05, based on 5,000 iterations.

For the binary self-promotion fields where we fit a Logistic regression, we show our results in Figure 1. Also, for comparison of the results relative to an unmatched subset of users, we also report in gray on Figure 1 the naive estimates of gender differences without controlling for other features. We report the odds ratio coefficient estimate for gender for each of the models by taking $e^{\beta_{Female}}$, where the β_{Female} coefficient is from the Logistic regression model fit. Therefore, in Figure 1, the interpretation of the vertical line at 1 indicates no gender differences, while <1 indicates female users are less likely to include the specified data field and >1 indicates female users are more likely to include the specified data field. We report 95% confidence intervals as well and significant gender differences at the 0.05-level are displayed in red. We now describe the analysis for each of the four fields (i.e. Summary, Job Description, Honors, and Skills):

Summary field: We evaluate gender differences in the likelihood to include a Summary as part of the profile. As shown in Figure 1, we observe women are overall less likely than their male counterparts to complete the Summary field. To illustrate how to interpret the meaning of these effect sizes, for Dartmouth we see the odds of a female user including a Summary, controlling for the matching variables, is 0.53 times as large as the odds of a male including a Summary. The size of the points on the figure are proportional to the number of users in a university in the resulting matched sample. Among those users who do include a profile, we do not find any statistically significant gender differences in Summary length via the randomization test.

Job Description field: Next, we evaluate differences in the likelihood to include a Job Description and discover varying odds of a female user including one compared to the odds for male users. Similar to the Summary figure, the Description figure displays the odds ratio coefficient interpretation from the Logistic regression along with the 95% confidence intervals. We observe that for several universities women are significantly less likely than men to include a Description.

Honors field: We measure differences in the likelihood to include at least one Honor and find no gender differences across universities. Then, based on the randomization test, we also do not find any statistical gender differences in the total number of Honors listed among users who do include at least one Honor.

Skills field: Finally, we analyze gender differences in the likelihood to include at least one Skill on the profile. We observe a few universities where women are more likely to include a Skill and one where women are less likely to include a Skill. Then among users who do include at least one Skill, we observe statistically significant gender differences in female versus male Skill counts respectively for Berkeley (19.8 vs. 21.9), Chicago (17.5 vs. 19.6), Harvard (16.3 vs. 17.4), MIT (18.8 vs. 20.8), Northwestern (18.8 vs. 20.5), and Wharton (16.6 vs. 19.0).

Conclusion

We discover that female users typically have a lower odds of including Summary and Job Description fields in their profile relative to male users. This portion of the profile, unlike listing Honors or Skills, requires the user to take an extra step in providing information about themselves. While this work is unable to speculate on the causal impact of these differences, we encourage future work that investigates this question on how self-presentation decisions affect online professional opportunity, similar to (Chiang and Suen 2015). This work also has implications for how recruiters should interpret such self-presentations. For example, if recruiters do not currently prioritize reading the Summary field, then such self-presentation differences should not matter. Yet, listing more Skills may increase the likelihood of a user being found by a recruiter.

This analysis of online self-presentation contributes to the rapidly developing literature documenting gender bias in the professional and social marketplaces such as differences in callback rates when randomizing the gender of a name on a resume (Bertrand and Mullainathan 2004), the effect of perceived gender on social media influence (Nilizadeh et al. 2016), differences in self-disclosure rates on Facebook (Wang, Burke, and Kraut 2016), and an analysis of gender bias in computer science faculty hiring (Way, Larremore, and Clauset 2016). The global desire to promote gender diversity in professional settings creates tailwinds for information-gathering projects like ours. For example, what are the economic effects of gender differences in online self-promotion and what might be the implications? Are there differences in how recruiters perceive self-promoting women compared to self-promoting men with the same objective qualifications? As companies become accountable for structural equality, recruiters need to recognize gender differences in order to hire diverse employees. Providing the right information to both women and talent identifiers is the first step towards a shift in workplace demographics.

Acknowledgments

We thank those at LinkedIn who made the 2015 Economic Graph Challenge possible, and also thank Edo Airoldi, Nikola Andric, Fabrizia Mealli, and Johan Ugander for helpful comments. Supported in part by an National Defense Science and Engineering Graduate Fellowship.

References

Althauser, R. P., and Rubin, D. 1970. The computerized construction of a matched sample. *American Journal of Sociology* 76(2):325–346.

Bertrand, M., and Mullainathan, S. 2004. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review* 94(4):991–1013.

Chiang, J. K.-H., and Suen, H.-Y. 2015. Self-presentation and hiring recommendations in online communities: Lessons from linkedin. *Computers in Human Behavior* 48:516–524.

Ernst, M. D., et al. 2004. Permutation methods: a basis for exact inference. *Statistical Science* 19(4):676–685.

Giacalone, R. A., and Rosenfeld, P. 2013. *Impression management in the organization*. Psychology Press.

Goffman, E. 1959. The presentation of self in everyday life. *Garden City, NY: Anchor*.

Greiner, D. J., and Rubin, D. B. 2011. Causal effects of perceived immutable characteristics. *Review of Economics and Statistics* 93(3):775–785.

Haferkamp, N.; Eimler, S. C.; Papadakis, A.-M.; and Kruck, J. V. 2012. Men are from mars, women are from venus? examining gender differences in self-presentation on social networking sites. *Cyberpsychology, Behavior, and Social Networking* 15(2):91–98.

Ho, D.; Imai, K.; King, G.; and Stuart, E. 2007. Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software* http://gking.harvard.edu/matchit.

Holland, P. W. 2003. Causation and race. *ETS Research Report Series* 2003(1):i–21.

Kietzmann, J. H.; Hermkens, K.; McCarthy, I. P.; and Silvestre, B. S. 2011. Social media? get serious! understanding the functional building blocks of social media. *Business Horizons* 54(3):241–251.

King, M. M.; Bergstrom, C. T.; Correll, S. J.; Jacquet, J.; and West, J. D. 2016. Men set their own cites high: Gender and self-citation across fields and over time. *arXiv preprint arXiv:1607.00376*.

Nilizadeh, S.; Groggel, A.; Lista, P.; Das, S.; Ahn, Y.-Y.; Kapadia, A.; and Rojas, F. 2016. Twitter's glass ceiling: The effect of perceived gender on online visibility. In *Tenth International AAAI Conference on Web and Social Media*.

Roberts, L. M. 2005. Changing faces: Professional image construction in diverse organizational settings. *Academy of Management Review* 30(4):685–711.

Rosenbaum, P. R., and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Rosenbaum, P. R., and Rubin, D. B. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39(1):33–38.

Rubin, D. B., and Thomas, N. 1996. Matching using estimated propensity scores: relating theory to practice. *Biometrics* 52(1):249–264.

Rubin, D. B. 2006. *Matched sampling for causal effects*. Cambridge University Press.

Rubin, D. B. 2007. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine* 26(1):20–36.

Rudman, L. A. 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.

Sievers, K.; Wodzicki, K.; Aberle, I.; Keckeisen, M.; and Cress, U. 2015. Self-presentation in professional networks: More than just window dressing. *Computers in Human Behavior* 50:25–30.

Wang, Y.-C.; Burke, M.; and Kraut, R. 2016. Modeling self-disclosure in social networking sites. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 74–85. ACM.

Way, S. F.; Larremore, D. B.; and Clauset, A. 2016. Gender, productivity, and prestige in computer science faculty hiring networks. In *Proceedings of the 25th International Conference on World Wide Web*, 1169–1179.