
**STI 2018 Conference Proceedings**

*Proceedings of the 23rd International Conference on Science and Technology Indicators*

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

**Chair of the Conference**

Paul Wouters

**Scientific Editors**

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

**Layout**

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

# Gender homophily in citations[1]

Gita Ghiasi[*], Philippe Mongeon [**], Cassidy R. Sugimoto[***], and Vincent Larivière[****]

[*] *gita.ghiasi.hafezi@umontreal.ca*
École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada; The Institute for Data Valorisation (IVADO), Montréal, Quebec, Canada

[**] *p.mongeon@cwts.leidenuniv.nl*
Centre for Science and Technology Studies, Leiden University, P.O. Box 905 2300 AX Leiden, The Netherlands; École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC. H3C 3J7, Canada; and Centre interuniversitaire de recherche sur la science et la technologie C.P. 8888, succ. Centre-ville Montréal (Québec) H3C 3P8

[***] *sugimoto@indiana.edu*
School of Informatics, Computing, and Engineering, Indiana University Bloomington, 901 E 10th St, Bloomington, IN 47405 (USA)

[***] *vincent.lariviere@umontreal.ca*
École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada; and Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC, H3C 3P8, Canada

## Abstract
Citations are one of the key indicators of scientific influence. However, citations are not entirely neutral and are affected by several social factors. This research provides a comprehensive gendered analysis of citation patterns, considering author contribution, research field, and subject similarity. For this purpose, citation data of 7,011,369 articles published in 2008-2016 and indexed in the Web of Science are analyzed. The findings indicate that gender-based homophily in citations exists in all disciplines. While homophily exists outside self-citations, the disproportionate self-citation rates of male authors exacerbate the disparities observed in homophily. This study suggests that men's higher tendency to cite the work of their male-peers could render women's scientific contributions under-recognized. The results of this study inform science policy, by calling into question the neutrality of citation-based metrics and therefore their use as evaluative instruments in the reward system of science.

## Introduction
Citations are one of the key indicators of scientific influence (Håkanson, 2005). They indicate to what extent a piece of research is related to other scholarly publications, and whether knowledge embedded in earlier scientific documents influences new scientific insights. In this context, citation rates are considered as measures of scientific impact, and the author of highly cited papers is thereby perceived as prominent and influential (Maliniak et al., 2013). Therefore, citation measures are increasingly used as a reward currency of science, upon which decisions

[1] This work was supported by the Canada Research Chairs program and the Social Science and Humanities Research Council of Canada.

on resource distribution and salary (Toutkoushian, 1994), hiring, appointment, tenure, and promotion (Holden et al., 2005) are often based.

However, citations are not entirely neutral and are affected by several social factors. One might argue that these factors play an increasingly important role, in a context where citation-based metrics are used for evaluative purposes. One example is the role of familiarity or personal bonds between the citer and the citee, in the sense that researchers are more inclined to cite authors with whom they are personally acquainted (Bornmann and Daniel, 2008; White, 2001).

Moreover, citation measures hinge on the Matthew effect—an allusion to the famous colloquialism "the rich get richer, and the poor get poorer"—in the sense that papers written by highly-cited authors are more likely to garner further citations than those of less-known scientists, leaving less-eminent scientists invisible and unknown (Fowler and Aksnes, 2007; Merton, 1988). This might place women at a disadvantage in fields, as their papers are subject to lower rate of citations compared to those of men after controlling for authorship order (Larivière et al., 2013), journal impact factor and field (Larivière and Sugimoto, 2017), affiliation, tenure status, and method and context (Maliniak et al., 2013).

A deterrent to overcoming gender differences in citations might relate to persistent gender biases in evaluations (Ferber and Brün, 2011), favoring men over women of identical qualifications (Deaux and Taynor, 1973; Ferber and Huber, 1975; Grunspan et al., 2016; Moss-Racusin et al., 2012). Along these lines, self-citations—a prevalent practice in science communication to expand on one's earlier research and an indication of a scholar's dynamic scientific activity— can also present potent effects on inflation of a scholar's citations and strengthen one's position in the scientific community (Costas et al., 2010; Glänzel et al., 2006). Therefore, men's higher proclivity to self-cite their own publications (Ghiasi et al., 2016; King et al., 2015) and women's greater likelihood to receive self-citations from their immediate co-authors (Ghiasi et al., 2016) could thus be consequential to the persistence of the gender disparities in citations.

Another factor might pertain to "gender homophily" in citations (Potthoff and Zimmermann, 2017)—a term used to describe an effect whereby the citer disproportionately cites references from authors who are of the same gender (for example in Davenport and Snyder (1995); McElhinny et al. (2003); Mitchell et al. (2013)). One of the explanations for men's disproportionate preference for citing men is the Matilda effect (Rossiter, 1993)—a systematic under-recognition of women's contribution to science (Knobloch-Westerwick and Glynn, 2013). Despite these few analyses, the question of homophily in citations has been the focus of few studies. This research intends to fill this gap and map citation patterns of authors of each gender, considering author contribution, research field, and subject similarity.

**Methods**

Data were collected from Clarivate Analytics' Web of Science (WoS), more specifically the Science Citation Index Expanded (SCI-Expanded), the Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (A&HCI) for the years 2008-2016. We focus on articles published from 2008 onward because the full first name of authors (which were used for gender assignation to the lead authors) is indexed in the Web of Science from the year 2008.

First-authors are often associated with researchers whose contributions form the highest share of tasks performed in a paper (Larivière et al., 2016) and a corresponding author is often assigned to the author who is in charge of correspondence and is more likely responsible for the initial conception and supervision of the research project (Mattsson et al., 2011). Therefore,

this paper considers first and corresponding authors as the lead authors and considers both as a proxy for whether the main contributions to the paper was made by men or women.

Gender was assigned to first and corresponding authors, using the Wiki-Gendersort algorithm (Bérubé et al., (in preparation)). This algorithm increases the precision of gender assignation by sifting through the first names of the names covered by Wikipedia and counting the number of masculine and feminine pronouns in the introduction section of the first twenty pages. Gender is assigned to the first name when the same gender was attributed to 75% of Wikipedia pages. No gender is assigned when this threshold is not met. Accordingly, gender is assigned to 68.8% of the 24,960,090 lead authors.

To measure gender homophily in citations, the share of references made to articles led by men for all male-led and female-led papers are compared, both including and excluding self-citations. This paper focus on those articles where (1) gender of all first and corresponding authors of the paper are identified, and (2) all first and corresponding authors are of the same gender. A *given* citation to an article thus falls into four categories: whether a citation is from a male-led authored paper to other paper of the same gender (mm) or the opposite gender (mf), and (2) whether a citation is from a women-led authored paper to other paper of the same gender (ff) or the opposite gender (fm). In this regard, the proportion mm citations ($P_{mm}$) is equal to one minus proportion mf citations ($P_{mf}$) (similarly, $P_{fm}$= 1-$P_{ff}$) and rate of "gender-based homophily" in citations, here, is defined as two equal values: $P_{mm}$ minus $P_{fm}$, or $P_{ff}$ minus $P_{mf}$.

These comparisons are performed with and without self-citations. Self-citations are defined as citations made by one of the lead authors of the paper to other papers on which he/she is listed as an author. Note that a lead citing author is not required to be a lead author in the cited paper for a citation to be considered as a self-citation.

Since the population of citable articles (and therefore the population of citable papers led by men and women) may differ based on research topics, we compared papers led by women with the most similar article led by men that appeared in the same issue of the same journal. We use the cosine similarity of the noun phrases of the title and abstract of the papers to calculate the degree of similarity of articles and to form the nearest neighbour pairs. These pairs thus include only an article for which a similar article led by the opposite gender was published in the same issue of the same journal. Table 1 presents the number of papers led by women and men used in the first part of the analysis, as well as the number of nearest neighbour (similar topic) pairs that are used in the second part of the analysis.

Table 1. Number of male-led articles, female-led articles, and nearest neighbour pairs by discipline

| Discipline | Women-led articles | Men-led articles | Nearest neighbour pairs |
|---|---|---|---|
| **BM** | 773,144 | 1,887,176 | 697,767 |
| **NSE** | 556,080 | 2,458,620 | 498,890 |
| **SSH** | 525,429 | 810,920 | 361,310 |

**Results**

Gender-based homophily in citations exists across all the disciplines (Fig. 1) and becomes less pronounced over time (Fig. 2). It is more prevalent in social sciences and humanities (SSH) than in natural sciences and engineering (NSE) and biomedical sciences (BM) (Fig. 1). However, gender homophily is less conspicuous when self-citations are excluded (Fig. 1 and Fig. 2), which shows that self-citations impose a considerable contribution to gender-based

homophily in citations. Moreover, gender homophily in citations is practiced at the same rate over years when self-citations are excluded (Fig. 2).

Gender homophily propensity decreases when only nearest neighbour pairs are considered, and this decrease is more pronounced in SSH and NSE (Fig. 1), which could correspond to the fragmentation of research topic between genders particularly in these two fields. Our results also reveal that gender homophily in citations is performed almost at the same rate across different disciplines once the topic similarity of papers is controlled for. Exclusion of self-citations does not present a sizable difference in the rate of gender homophily in this case. Overall differences observed are smaller when topic similarity is controlled for, because similar articles are expected to follow a similar referencing pattern.

Figure 1- Gender homophily rate in citations in different disciplines

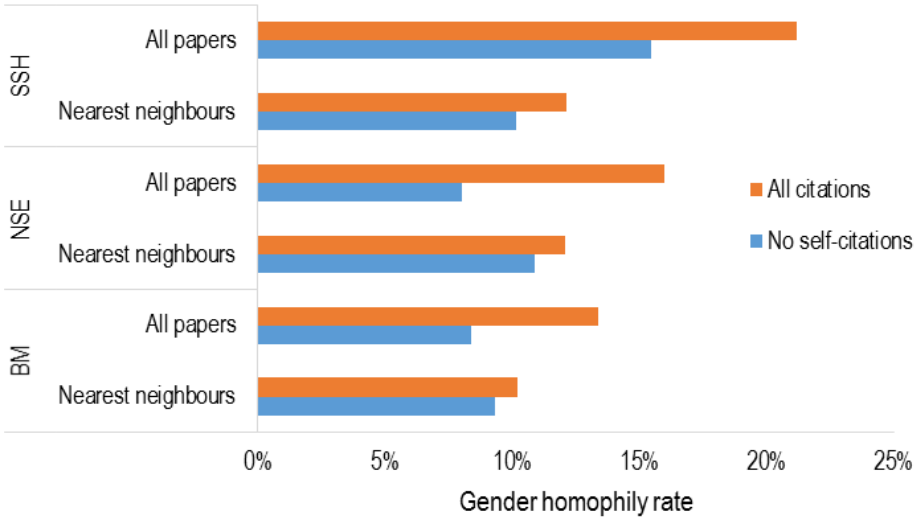
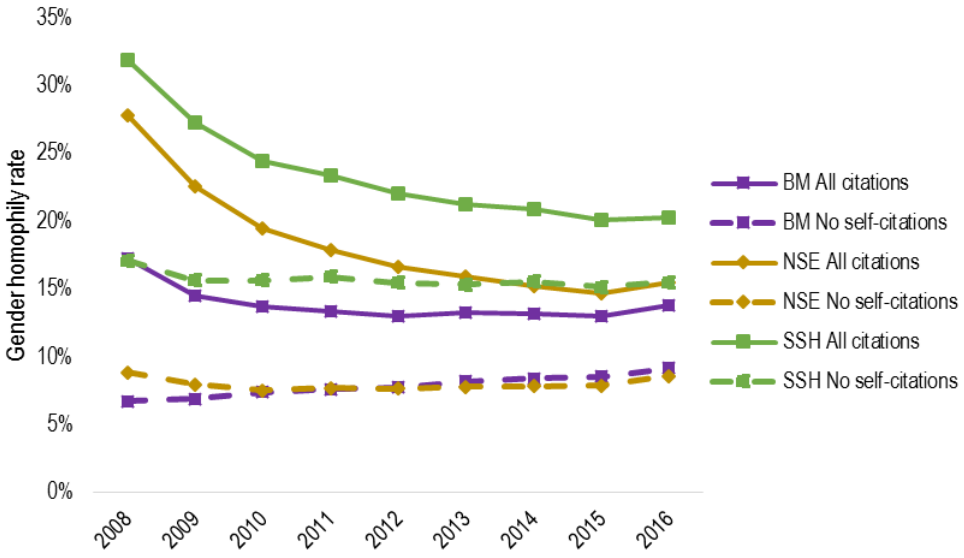


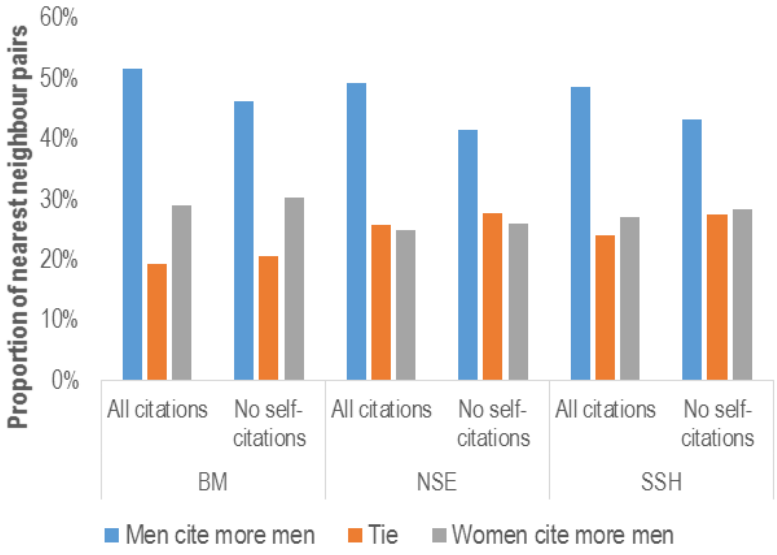

Figure 2- Gender homophily rate of different disciplines over time




The analyses reveal that in around 50% of article pairs of similar topic, male-led articles gave citations to male-led papers more often than the paired female-led paper did (gave citations to male-led papers), while equal gendered practices (where female-led and male-led papers in a

pair cite male-led and female-led papers similarly) are limited to 19%, 26%, and 24% of pairs in BM, NSE, and SSH, respectively (Fig. 3).

Figure 3- Nearest neighbour comparison of men led vs. women led papers' share of citations to male papers




**Discussion and conclusion**

Our results clearly demonstrate that there is gender homophily in referencing patterns; that is, men are more likely than women to cite men, and women cite women more often than men do, even when controlling for fields, self-citation, and topic similarity. Among those, self-citations have a strong effect, and exacerbate the differences observed in homophily, as men have been shown to self-cite themselves more than women (Ghiasi et al., 2016; King et al., 2015). While the gender differences homophily have decreased since 2008, they stabilized in recent years around 15%.

Gender differences in homophily could present serious implications for the reward system of science, especially when combined with women's lower representation in the scientific workforce. Since women only account for less than 30% of authorships (Larivière et al., 2013), men's higher inclination than women to cite the work of men could suggest a perpetual disparity in favor of men. This could render women's scientific contributions under-recognized and contribute to the Matilda effect in science (Rossiter, 1993). Those strong gender biases built into dissemination and use of research suggest that women are at a disadvantage when citation-based indicators of impact are used for evaluative purposes.

**Limitations and future research**

This study only focuses on self-citations (references) made by lead authors. However, all-author self-citations should also be explored, as researchers tend to cite their co-authors more often than other researchers. Therefore, gender homophily in co-authorship practices might also contribute to gender homophily in citations and should be explored in future research. Moreover, this study only focuses on female and male-led papers, does not take into account mixed-gender led articles. Citations received, and references made to (by) mixed-gender papers could potentially influence the gender homophily rate and therefore present an important factor to be incorporated in the future studies.

**References**

Bérubé, N., Sainte-Marie, M., Ghiasi, G., and Larivière, V. ((in preparation)). Assigning gender to first names using Wikipedia.

Bornmann, L., and Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. J. Doc. *64*, 45–80.

Costas, R., Leeuwen, T.N. van, and Bordons, M. (2010). Self-citations at the meso and individual levels: effects of different calculation methods. Scientometrics *82*, 517–537.

Davenport, E., and Snyder, H. (1995). Who cites women? whom do women cite?: an exploration of gender and scholarly citation in sociology. J. Doc. *51*, 404–410.

Deaux, K., and Taynor, J. (1973). Evaluation of Male and Female Ability: Bias Works Two Ways. Psychol. Rep. *32*, 261–262.

Ferber, M.A., and Brün, M. (2011). The Gender Gap in Citations: Does It Persist? Fem. Econ. *17*, 151–158.

Ferber, M.A., and Huber, J.A. (1975). Sex of Student and Instructor: A Study of Student Bias. Am. J. Sociol. *80*, 949–963.

Fowler, J.H., and Aksnes, D.W. (2007). Does self-citation pay? Scientometrics *72*, 427–437.

Ghiasi, G., Larivière, V., and Sugimoto, C.R. (2016). Gender differences in synchronous and diachronous self-citations. In Proceedings of the 21st International Conference on Science and Technology Indicators, Valencia, Spain, p.

Glänzel, W., Debackere, K., Thijs, B., and Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. Scientometrics *67*, 263–277.

Grunspan, D.Z., Eddy, S.L., Brownell, S.E., Wiggins, B.L., Crowe, A.J., and Goodreau, S.M. (2016). Males Under-Estimate Academic Performance of Their Female Peers in Undergraduate Biology Classrooms. PLOS ONE *11*, e0148405.

Håkanson, M. (2005). The impact of gender on citations: An analysis of college & research libraries, journal of academic librarianship, and library quarterly. Coll. Res. Libr. *66*, 312–322.

Holden, G., Rosenberg, G., and Barker, K. (2005). Bibliometrics: A potential decision making aid in hiring, reappointment, tenure and promotion decisions. Soc. Work Health Care *41*, 67–92.

King, M.M., Correll, S.J., Jacquet, J., Bergstrom, C.T., and West, J.D. (2015). Men set their own cites high: Gender and self-citation across fields and over time (Working paper. Retrieved from http://www. eigenfactor. org/gender/self-citation/SelfCitation. pdf).

Knobloch-Westerwick, S., and Glynn, C.J. (2013). The Matilda Effect-Role Congruity Effects on Scholarly Communication: A Citation Analysis of Communication Research and Journal of Communication Articles. Commun. Res. *40*, 3–26.

Larivière, V., and Sugimoto, C.R. (2017). The end of gender disparities in science? If only it were true...

Larivière, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C.R. (2013). Bibliometrics: Global gender disparities in science. Nature *504*, 211–213.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., and Sugimoto, C.R. (2016). Contributorship and division of labor in knowledge production. Soc. Stud. Sci. *46*, 417–435.

Maliniak, D., Powers, R., and Walter, B.F. (2013). The gender citation gap in international relations. Int. Organ. *67*, 889–922.

Mattsson, P., Sundberg, C.J., and Laget, P. (2011). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. Scientometrics *87*, 99–105.

McElhinny, B., Hols, M., Holtzkener, J., Unger, S., and Hicks, C. (2003). Gender, publication and citation in sociolinguistics and linguistic anthropology: The construction of a scholarly canon. Lang. Soc. *32*, 299–328.

Merton, R.K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. Isis 606–623.

Mitchell, S.M., Lange, S., and Brus, H. (2013). Gendered citation patterns in international relations journals. Int. Stud. Perspect. *14*, 485–492.

Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., Graham, M.J., and Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. Proc. Natl. Acad. Sci. *109*, 16474–16479.

Potthoff, M., and Zimmermann, F. (2017). Is there a gender-based fragmentation of communication science? An investigation of the reasons for the apparent gender homophily in citations. Scientometrics *112*, 1047–1063.

Rossiter, M.W. (1993). The Matthew Matilda effect in science. Soc. Stud. Sci. *23*, 325–341.

Toutkoushian, R.K. (1994). Using Citations to Measure Sex Discrimination in Faculty Salaries. Rev. High. Educ. *18*, 61–82.

White, H.D. (2001). Authors as citers over time. J. Assoc. Inf. Sci. Technol. *52*, 87–108.