Gender's Influence on Academic Collaboration in a University-Wide Network

Logan McNichols, Gabriel Medina-Kim, Viet Lien Nguyen, Christian Rapp, and Theresa Migler(✉)

California Polytechnic State University, San Luis Obispo, CA 93405, USA
tmigler@calpoly.edu

Abstract. We present a collaboration network of the faculty of a large polytechnic state university in the United States. In our network vertices represent researchers: faculty members at the university and collaborators of these faculty members. Edges between two researchers in this network represent some sort of collaborative experience: conference publication, journal publication, or grant application. In this paper we present a study of this network and various subnetworks with respect to gender.

Keywords: Collaboration network · Interdisciplinary collaboration · Gender

1 Introduction

Co-authorship networks, in which vertices represent authors and two authors are connected if they have co-authored a paper together, are fundamental to our understanding of dynamics among groups of researchers. In particular, there is a strong drive for academics to become *interdisciplinary*, to combine two or more academic disciplines into one activity. For example, the European Union framework program Horizon 2020 has policy actors that state that interdisciplinary research will be the "key to future scientific breakthroughs" [2]. It is predicted that the future of research will become increasingly interdisciplinary [13]. In order to study interdisciplinary collaborations it is necessary to study networks that include as many disciplines as possible. The university being studied has six colleges and 49 distinct departments. We present a large network of 20,822 vertices, 1,855 of these vertices represent the university's faculty and 18,967 of these vertices represent external collaborators. Two researchers are connected if they have ever shared a collaborative experience: a co-publication of a conference or journal paper or a book or a grant application. We constructed this network with a view to understand the behavior of collaborations across disciplines.

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2019, SCI 882, pp. 94–104, 2020.
https://doi.org/10.1007/978-3-030-36683-4_8

2 Related Work

Collaboration networks are among the most studied networks in network science. Collaboration networks range from the popular "film actor network", where vertices represent film actors and two film actors are connected if they have ever appeared in a film together [18], to the most often studied co-authorship networks of academics, where vertices represent researchers and two researchers are connected if they have ever co-authored a paper [4,6,7,15,16]. Most of these collaboration networks are built with respect to a certain field, for example vertices represent mathematicians and two mathematicians are connected if they have published a paper [10]. There have been studies of collaborative networks across several related departments within a university. In 1978 Friedkin studied the collaborative networks of 128 faculty members across six physical science departments [9]. However, to the best of the authors' knowledge there has been no large-scale university-wide collaboration network across many disciplines constructed or studied.

The analysis of gender with respect to collaborations has been widely studied. Ductor, Goyal, and Prummer showed that women have fewer collaborators, collaborate more often with the same co-authors, and a higher fraction of their co-authors are co-authors of each other [8]. Holman and Morandin found that researchers preferentially co-publish with colleagues of the same gender, and show that this'gender homophily' is slightly stronger today than it was 10 years ago [12]. Using a dataset with more than 270,000 scientists, Araújo, Araújo, Moreira, Herrmann, and Andrade show that while men are more likely to collaborate with other men, women are more egalitarian regardless of how many collaborators each scientist has, with the exception of engineering where the bias disappears with increasing number of collaborators [3]. Abramo, D'Angelo, and Murgia study the scientific work of Italian academics and found that women researchers register a greater capacity to collaborate, with the exception of international collaboration, where there is still a gap in comparison to male colleagues [1]. And West, Jacquet, King, Correll, and Bergstrom showed that on average women publish fewer papers than men [19]. In what follows we will confirm most of these observations for our network.

3 The Network

The university that we are building this collaboration network for has approximately 1,473 current faculty members, both full and part-time. The university is mainly an undergraduate university with 21,037 undergraduate students and 775 graduate students. The university has six colleges: the College of Agriculture, Food and Environmental Sciences, the College of Architecture and Environmental Design, the College of Business, the College of Engineering, the College of Liberal Arts, and the College of Science and Mathematics. Among these six colleges there are 49 distinct departments. We refer to researchers at this university as *university researchers*. Collaborators of university researchers are *external researchers* (although most external researchers work at a different university).

A vertex in the network represents a researcher who is either a university researcher or an external researcher who collaborates with a university researcher. We initially populated the vertex set by scraping each of the 49 departments' webpages for faculty lists. Then as we found collaborative experiences for these university researchers with external researchers, we added these external researchers to our vertex set.

The edges in our network represent *collaborative experiences* between two researchers which represent a scholarly work (a publication or grant application) where a university researcher is included as an author. We populated the collaboration data from five sources:

- Grants applications from the university's grants office
- Publication records from Google Scholar, Microsoft Academic, and *arxiv*
- In the case of the Computer Science Department, personal webpages and curricula vitae for faculty members

Based on many findings that Google Scholar provides a comprehensive coverage that meets or exceeds that of similar bibliographic databases [11], we chose to use collaborations from Google Scholar as our "backbone" set of collaborations. One criticism of Google Scholar is that there may be duplicate entries [11]. We chose not to use services such as ResearchGate because there has been much criticism that such services give equal weight to legitimate journals as well as predatory or hijacked journals [14].

Note that a single collaborative experience with i authors will produce a clique in the graph of size i.

4 Properties of the Networks

We consider three nested networks:

- **The Computer Science Network:** Vertices represent faculty in the Computer Science Department and their collaborators.
  - This network has 54 Computer Science Department university researcher vertices and 1,314 external researcher vertices (1,368 total vertices), and 6,682 edges.
- **The College of Engineering Network:** Vertices represent faculty in the College of Engineering and their collaborators.
  - This network has 207 College of Engineering university researcher vertices and 4,569 external researcher vertices (4,776 total vertices), and 25,438 edges.
- **The University-Wide Network:** Vertices represent faculty at the university and their collaborators.
  - This network has 1,855 university researcher vertices (this includes emeritus faculty) and 18,967 external researcher vertices (20,822 total vertices), and 106,728 edges.

In these three networks edges represent collaborative experiences between two researchers in the network. We exclude collaborative experiences with more than seven researchers as we feel that we cannot draw collaborative information from such large groups. If there are twenty authors we can't expect that every pair of the twenty authors has ever even met the other.

We note that the at the university being studied, the Computer Science Department is housed in the College of Engineering, so these networks are truly nested.

The Computer Science Network is the only network that has been human verified. For each computer science university researcher in this network, we verified that all of the collaborations listed on this researcher's curriculum vitae were in the network, and further that there weren't any duplicate entries.

5 Gender and the Networks

In order to analyze our networks with respect to gender we used *Gender API*, a gender inference service. Gender API currently supports 178 countries and provides confidence parameters *samples*, the number of database records matching the request, and *accuracy*, the reliability of the assignment. When comparing with other gender inference tools, Gender API is the best performer in terms of fraction of inaccuracies [17]. Unfortunately, many external researchers in the network first initial to represent their first name. In such cases we mark the gender as *unknown*. Further there are cases where Gender API cannot conclusively identify a name as male or female, we also mark these genders as unknown.

5.1 Basic Gender Statistics

In the University-Wide network there are 5,216 total female vertices, 691 of which are female university vertices, 13,000 total male vertices, and 1,150 of which are male university vertices, 2,606 total unknown vertices, 14 of which are unknown university vertices.

In the College of Engineering Network there are 811 total female vertices, 31 of which are female university vertices, 3,274 total male vertices, 173 of which are male university vertices, and 691 total unknown vertices, 3 of which are unknown university vertices.

In the Computer Science Network there are 282 total female vertices, 11 female university vertices, 966 total male vertices, 43 male university vertices, and 120 total unknown vertices, 0 of which are unknown university vertices. See Fig. 2 for a summary of these counts.

In the following we use the term *internal* to mean: for the University-Wide Network, simply university researchers, for the College of Engineering Network, university vertices who are further College of Engineering faculty members, and for the Computer Science Network university vertices who are further Computer Science faculty members. See Fig. 1 for the Computer Science network with only internal edges and vertices.

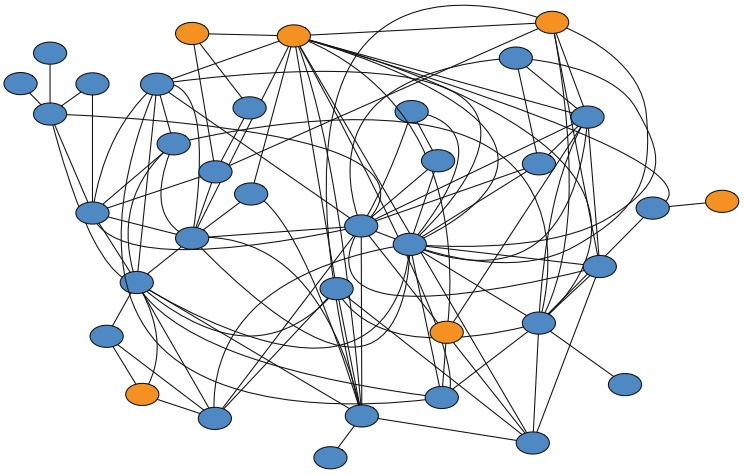

Fig. 1. A subnetwork of the Computer Science network. Vertices represent Computer Science faculty and there is an edge between two faculty members if they are collaborators. All isolated vertices have been removed. Orange vertices represent female faculty members and blue vertices represent male vertices.

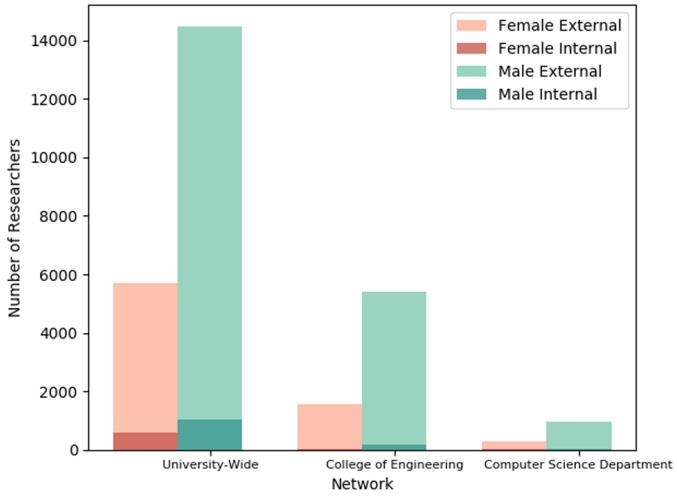

Fig. 2. Counts for female and male internal and external vertices.

5.2 Claim: Women Have Fewer Collaborators Than Men

Ductor, Goyal, and Prummer showed that women have fewer collaborators than men [8].

We tested this hypothesis on each of our three networks by removing multiple edges then comparing the average degree of the female internal vertices with those of the male internal vertices. We found that female researchers have substantially fewer collaborators than the male researchers except for the College of Engineering Network where the difference is not so strong. See Fig. 3.

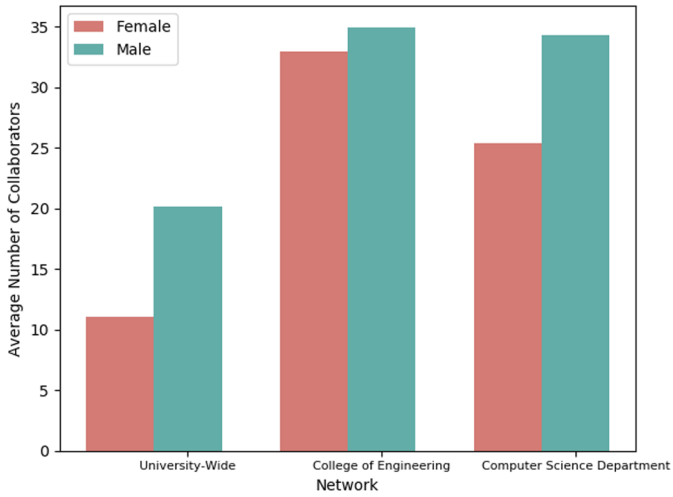

Fig. 3. The average number of female and male collaborators in each of the three networks.

5.3 Claim: Women Collaborate More Often with the Same Co-Authors

Ductor, Goyal, and Prummer showed that women collaborate more often with the same co-authors [8].

To obtain the *repeat collaboration number* for each researcher we summed the number of multiple collaborations that they had with other researchers. For example, if researcher A collaborated with researcher B once, researcher C three times, and researcher D five times, then A's repeat collaboration number is six (two repeats with C and four repeats with D). To test the hypothesis on each of our three networks we averaged the repeat collaboration number for each internal female vertex with the repeat collaboration number for each internal male vertex. We found that the repeat collaboration number is actually higher for males than females in both the University-Wide and Computer Science Department networks. See Fig. 4.

Then we calculate the *repeat collaborator number* for each researcher. This is simply a count of the number of collaborators that the researcher has worked with more than once. For example, suppose researcher A has three collaborators, B, C, and D. Suppose that A has collaborated with B three times, with C one time, and with D seven times. In this case, A's repeat collaborator number is two (one for B and one for D). Again, we found that the repeat collaborator number is higher for males than for females in all three networks. See Fig. 5.

5.4 Claim: A Higher Fraction of Women's Co-Authors Are Co-Authors with Each Other

Ductor, Goyal, and Prummer showed that a higher fraction of women's co-authors are co-authors of each other [8].

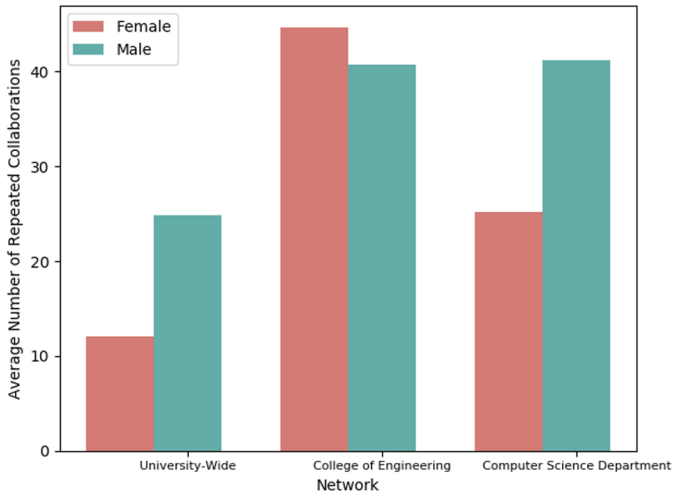

Fig. 4. The average repeat collaboration number for female and male collaborators in each of the three networks.

To test this hypothesis we found the average local clustering coefficient for internal female vertices and compared it with the average local clustering coefficient for internal male vertices. The clustering coefficient for female researchers is higher than that for males, except in the Computer Science Department network where they are very similar. See Fig. 6

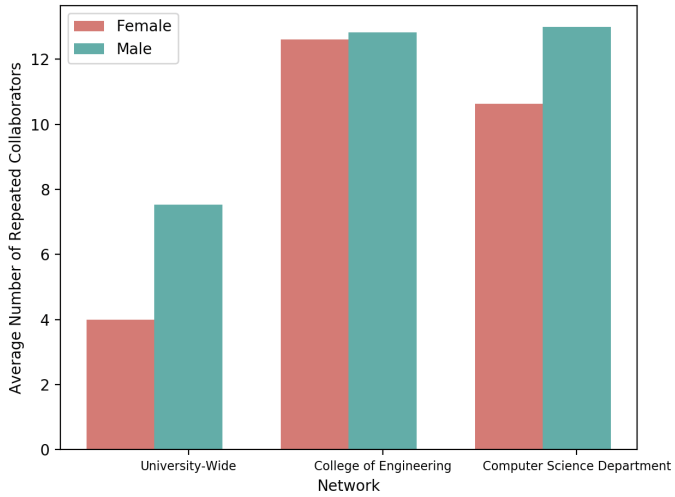

Fig. 5. The average repeat collaborator number for female and male collaborators in each of the three networks.

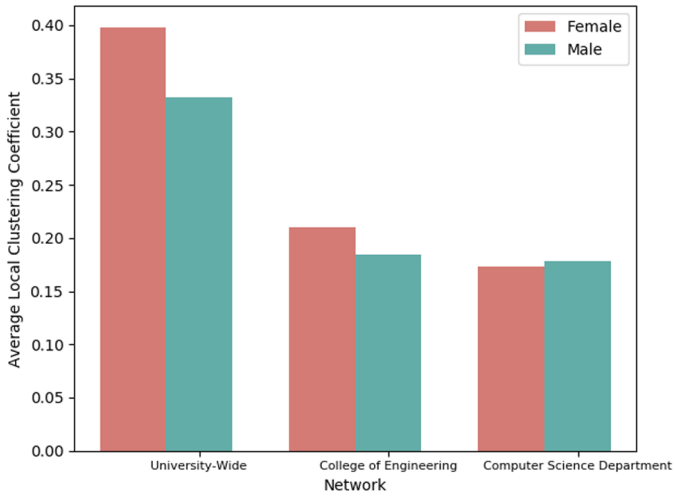

Fig. 6. The average local clustering coefficient for female and male collaborators in each of the three networks.

5.5 Claim: Researchers Preferentially Co-Publish with Authors of the Same Gender

Holman and Morandin found that researchers preferentially co-publish with colleagues of the same gender [12]. However Araújo, Araújo, Moreira, Herrmann, and Andrade show that while men are more likely to collaborate with other men, women are more egalitarian regardless of how many collaborators each scientist has [3].

To test homophily, we took the average over all internal female researchers the ratio of the number of their female collaborators to the number of their

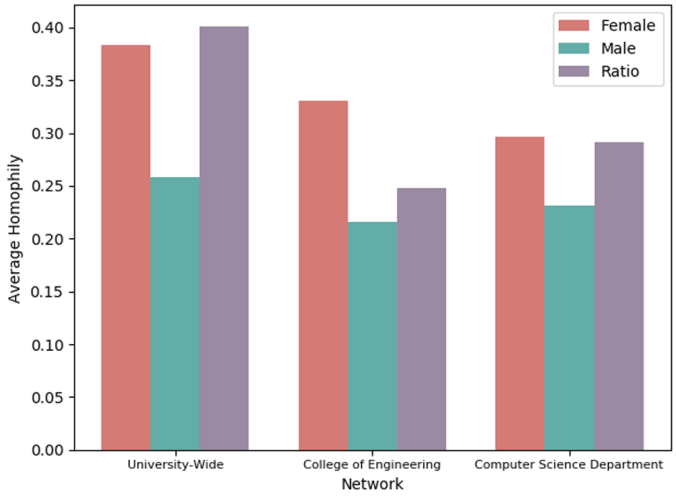

Fig. 7. The average homophily for female and male collaborators in each of the three networks.

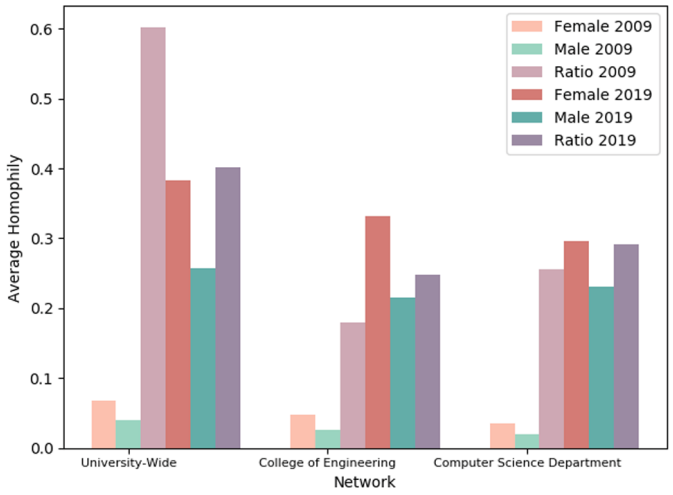

Fig. 8. The average homophily for female and male collaborators with publications until 2019 compared with the average homophily for female and male collaborators with publications until 2009 in each of the three networks

male collaborators. We took the same average for internal male researchers. For reference we include the simple ratio of total female researchers to total male researchers. See Fig. 7.

5.6 Claim: 'Gender Homophily' Increases over Time

Holman and Morandin found that 'Gender homophily' is slightly stronger today than it was 10 years ago [12].

To test this, we reduced our three networks by including only edges from collaborations that happened before 2009, we removed any isolated vertices then tested homophily in the same way as the previous section. See Fig. 8.

6 Future Work

We intend to manually verify the curriculum vitae of every faculty member at the university in an attempt to form a more complete network. At the time of writing, only the Computer Science faculty members' CVs have been verified by humans. We further wish to request demographic and job satisfaction information for each university faculty member in the network in the form of an online survey. This survey will include gender and ethnicity. We will also request that each university faculty member states their satisfaction with their position on a Likert scale in an attempt to answer the question: "Does a higher clustering coefficient indicate greater satisfaction?". In this same survey we also wish for authors to self-identify additional collaborative experiences that may not be documented by our existing collaborative experiences in order to create a more complete network. For example, two faculty members who wrote a paper together that

did not get accepted to any conference or journal. We also intend to study this network as it evolves over the years. NSF found that between 2000 and 2013, the percentage of publications with authors from multiple countries rose from 13.2% to 19.2% [20]. Noting that research is becoming increasingly international, we further wish to analyze the collaborations in our network with respect to location.

Acknowledgements. The authors would like to express their sincerest thanks to Professor Zoë Wood for inspiring this work based on her work with The Center for Research, Excellence and Diversity in Team Science (CREDITS), an integrated research and training program to increase and enhance Team Science and collective intelligence capacity, effectiveness, and excellence in California [5]. We would also like to express our thanks to Professor Foaad Khosmood for teaching our group about data sustainability and data scraping.

References

1. Abramo, G., DâAngelo, C.A., Murgia, G.: Gender differences in research collaboration. J. Inf. **7**(4), 811–822 (2013)
2. Allmendinger, J.: Quests for interdisciplinarity: a challenge for the era and horizon 2020. Science Europe Position Statement (2012)
3. Araújo, E.B., Araújo, N.A.M., Moreira, A.A., Herrmann, H.J., Andrade, J.S.: Gender differences in scientific collaborations: women are more egalitarian than men. PLoS ONE **12**(5), 1–10 (2017)
4. Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. Physica A: Stat. Mech. Appl. **311**(3), 590–614 (2002)
5. Campbell, L.G., Mehtani, S., Dozier, M.E., Rinehart, J.: Gender-heterogeneous working groups produce higher quality science. PLoS ONE **8**(10), 1–6 (2013)
6. De Castro, R., Grossman, J.: Famous trails to Paul Erdös. In: Vitanyi, P.M.B. (ed.) The Mathematical Intelligencer, vol. 21, January 1999
7. Ding, Y.: Scientific collaboration and endorsement: network analysis of coauthorship and citation networks. J. Inf. **5**(1), 187–203 (2011)
8. Ductor, L., Goyal, S., Prummer, A.: Gender & collaboration. Cambridge working papers in economics, Faculty of Economics, University of Cambridge (2018)
9. Friedkin, N.E.: University social structure and social networks among scientists. Am. J. Sociol. **83**(6), 1444–1465 (1978)
10. Grossman, J.W., Ion, P.D.F.: On a portion of the well-known collaboration graph. Congressus Numerantium **108**, 129–131 (1995)
11. Harzing, A.-W., Alakangas, S.: Google scholar, scopus and the web of science: a longitudinal and cross-disciplinary comparison. Scientometrics **106**(2), 787–804 (2016)
12. Holman, L., Morandin, C.: Researchers collaborate with same-gendered colleagues more often than expected across the life sciences. PLoS ONE **14**(4), 1–19 (2019)
13. Lyall, C., Meagher, L.R.: A masterclass in interdisciplinarity: research into practice in training the next generation of interdisciplinary researchers. Futures **44**(6), 608–617 (2012). Special Issue: Politics, Democracy and Degrowth
14. Memon, A.R.: Researchgate is no longer reliable: leniency towards ghost journals may decrease its impact on the scientific community. J. Pak. Med. Assoc. **66**(12), 1643–1647 (2016)

15. Newman, M.E.J.: Coauthorship networks and patterns of scientific collaboration. Proc. Nat. Acad. Sci. **101**(suppl 1), 5200–5205 (2004)
16. Newman, M.E.J.: Scientific collaboration networks. i. network construction and fundamental results. Phys. Rev. E, Stat. Nonlinear Soft Matter Phys. **64**, 0416131 (2001)
17. Santamaría, L., Mihaljevic, H.: Comparison and benchmark of name-to-gender inference services. Peer J. Comput. Sci. **4**, e156 (2018)
18. Watts, D.J.: Small Worlds: The Dynamics of Networks Between Order and Randomness. Princeton studies in complexity. Princeton University Press, Princeton (2004)
19. West, J.D., Jacquet, J., King, M.M., Correll, S.J., Bergstrom, C.T.: The role of gender in scholarly authorship. PLoS ONE **8**(7), 1–6 (2013)
20. Witze, A.: Research gets increasingly international. Nature, January 2016