Gender-specific patterns in the artificial intelligence scientific ecosystem

Anahita Hajibabaei [a], Andrea Schiffauerova [a], Ashkan Ebadi [a,b,*]

[a] Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC H3G 2W1 Canada
[b] National Research Council Canada, Montreal, QC H3T 2B2 , Canada

ARTICLE INFO

ABSTRACT

Gender disparity in science is one of the most focused debating points among authorities and the scientific community. Over the last few decades, numerous initiatives have endeavored to accelerate gender equity in academia and research society. However, despite the ongoing efforts, gaps persist across the world, and more measures need to be taken. Using social network analysis, natural language processing, and machine learning, in this study, we comprehensively analyzed gender-specific patterns in the highly interdisciplinary and evolving field of artificial intelligence for the period of 2000–2019. Our findings suggest an overall increasing rate of mixed-gender collaborations. From the observed gender-specific collaborative patterns, the existence of disciplinary homophily at both dyadic and team levels is confirmed. However, a higher preference was observed for female researchers to form homophilous collaborative links. Our core-periphery analysis indicated a significant positive association between having diverse collaboration and scientific performance and experience. We found evidence in support of expecting the rise of new female superstar researchers in the artificial intelligence field.

1. Introduction

During the past decades, gender disparity in science has become one of the most focused debating points among researchers (Shen, 2013). Despite concerted efforts to achieve gender equality (UNESCO, 2018), conclusive evidence has revealed that pervasive gender imbalances still exist in academic and scientific practice (Nelson & Rogers, 2003; Shaw & Stanton, 2012; West, Jacquet, King, Correll & Bergstrom, 2013). Although the number of female and male students pursuing bachelor's and master's degrees is almost equal, the women's under-representation starts at the Ph.D. level, and female attrition can be observed across higher levels of the academic hierarchy (Huyer, 2018)- a phenomenon dubbed the *"leaky pipeline"*[1] (Alper, 1993; Berryman, 1983). Abundant studies have also shown persistent gender bias in many aspects such as salaries (Shen, 2013), hiring and promoting (Moss-Racusin, Dovidio, Brescoll, Graham & Handelsman, 2012; Nelson & Rogers, 2003; Nielsen, 2016), grant funding (Witteman, Hendricks, Straus & Tannenbaum, 2019), scholarly authorship (West et al., 2013), scientific impact (Larivière, Ni, Gingras, Cronin & Sugimoto, 2013), peer reviews (Murray et al., 2019), and collaboration (Uhly, Visser & Zippel, 2015; Zeng et al., 2016).

There are many underlying reasons for the aforementioned imbalances. Family and caring responsibilities have been noted as the main impediment to female academic trajectories (Stack, 2004). While men follow a linear career path, women are more prone to career interruption or relinquishment due to maternity and family issues (Ramos, Cortés & Moreno, 2015), which could also lead to a gender gap in scientific performance (Kyvik & Teigen, 1996). Another viewpoint focusing on personal preferences or innate

---

* Corresponding author at: Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC H3G 2W1 Canada.
E-mail address: ashkan.ebadi@nrc-cnrc.gc.ca (A. Ebadi).
[1] The "leaky pipeline" concept has important drawbacks and methodological limitations (Xie & Shauman, 2003).

individual abilities (Wang & Degol, 2017) may explain the two genders' academic career dichotomy. Research has also reported gender stereotypes (Ceci, Ginther, Kahn & Williams, 2014), inadequate female role models (Nelson & Rogers, 2003), and the degree of specialization (Leahey, 2006) as potential explanations for the gender bias; however, the extent to which factors matter for this issue has remained elusive among researchers (Shaw & Stanton, 2012).

Over the last few decades, numerous initiatives have endeavored to accelerate gender equity in academia; however, women are still underrepresented in most scientific disciplines (Huang, Gates, Sinatra & Barabási, 2020; Thelwall, Bailey, Makita, Sud & Madalli, 2019) and in many countries (Larivière et al., 2013). Several countries enacted effective policies such as introducing flexible and family-oriented programs in the workplace, e.g., in the European Union (EU) countries, India, and Japan (Bonetta & Clayton, 2008; Jurviste et al., 2016), providing financial incentives for appointment and retention of female scientists (Pearson et al., 2015), and promoting gender-balanced organizational structures (Kamraro, 2014). Canada has not been an exception, generous parental leave programs were introduced, including extended parental leave for graduate students and parental sharing benefits, thereby fostering greater equality in family responsibilities and facilitating career recruitment and progression for females (Government of Canada, 2019). Additionally, the Canadian government has recently initiated an internationally recognized program called *"Athena Scientific Women's Academic Network"* (Government of Canada, 2018), aiming to enhance women's career trajectories and adopt an attitude towards gender parity (Ovseiko, Chapple, Edmunds & Ziebland, 2017). Despite these ongoing efforts and policies, gaps persist across the world, and more measures need to be taken on this matter (UNESCO, 2018).

There is a significant body of evidence aiming to assess various aspects of gender differences in academic performance. It has been argued that women are generally less productive than men (Ghiasi, Larivière & Sugimoto, 2015; Huang et al., 2020; Larivière et al., 2013; Xie & Shauman, 1998), in terms of the number of publications, and this gap varies across research disciplines (Duch et al., 2012; Elsevier, 2017). Gender differences in choices of an academic career seemed to be at play in explaining gender imbalances in academia. For example, some scholars indicated that while women usually tend to be active in people-oriented academic fields, men are generally dominant in thing-oriented disciplines (e.g., Lippa, Preston & Penner, 2014; Paswan & Singh, 2020; Su, Rounds & Armstrong, 2009). However, it is suggested that apart from things-people dimensions, other factors such as gender differences in empathizing and systemizing (Luoto, 2020), the number of available career options, and career opportunities to gain higher status or social impact should be considered to explain any gender differences (Thelwall, Bailey, Tobin & Bradshaw, 2019). Myriad studies have been devoted to analyzing and mapping scientific performance and gender patterns through the number of citations. Nonetheless, the research on this area has produced inconclusive and diverse results. For example, while some studies found a citation bias favoring female authors in specific fields (Borrego, Barrios, Villarroya & Ollé Castellà, 2010; Long, 1992; Van Arensbergen, Van der Weijden & Van den Besselaar, 2012), others found no differences for women or men (Bordons, Morillo, Fernández & Gómez, 2003; Cole & Zuckerman, 1984; Mauleón, Bordons & Oppenheim, 2008; Tower, Plummer & Ridgewell, 2007). However, several studies concluded that female authors, on average, have a lower impact compared to male authors (Hunter & Leahey, 2010; Larivière et al., 2013; Larivière, Vignola-Gagné, Villeneuve, Gélinas & Gingras, 2011). Recent research on longitudinal gender disparities in scientific performance has revealed that although women's participation has increased in academia over the past 60 years, gender inequality in publications and citation rates has been growing in favor of men. It is also argued that this gap persists among disciplines and almost all countries and is exacerbated when it comes to highly productive researchers (Huang et al., 2020).

It is widely acknowledged that collaboration has become a hallmark of scientific activities (Cummings, Kiesler, Cummings & Kiesler, 2005; Wood & Gray, 1991). Collaborative research is not only crucial to address complex scientific issues (Bennett & Gadlin, 2012) but also may affect research performance (Ebadi & Schiffauerova, 2016; Uhly et al., 2015), professional recognition (Fox & Mohapatra, 2007), and funding opportunities (Ebadi & Schiffauerova, 2015b; Kwiek, 2020). Owing to the importance of scientific collaboration, existing literature has examined gender-specific patterns in academic collaborations and reported gender homophily effect, i.e., if same-gender authors tend to publish together (Holman & Morandin, 2019; Jadidi, Karimi, Lietz & Wagner, 2018; Lee et al., 2019), and female propensity to domestic and intramural collaboration (Abramo, D'Angelo & Murgia, 2013; Bozeman & Corley, 2004; Larivière et al., 2011; Larivière et al., 2013). Furthermore, whilst women opt to form smaller and short-lived ties, men have more extensive and loyal collaboration networks (Jadidi et al., 2018). Presumably, several causes could be attributed to these differences like family obligations, prejudices against women (Hogan, Zippel, Frehill & Kramer, 2010), and gender inequality in research funding (Larivière et al., 2011; Stack, 2004).

Although prior studies investigated sex differences in research collaboration and productivity, to the best of our knowledge, no study hitherto addresses gender differences in the propensity to collaborate with respect to the similarity or dissimilarity of researchers' academic backgrounds. In this work, we aim to scrutinize the presence of gender-specific patterns in the co-authorship networks in the field of artificial intelligence (AI), an interdisciplinary field with a significant gender gap (World Economic Forum, 2018). To this end, we utilized machine learning, natural language processing, and social network analysis to address the following research questions: (Q1) Do female and male scientists prefer to be involved in different AI subject areas?; (Q2) Do female and male AI scientists differ in collaborating with others with similar or different disciplinary profiles?; (Q3) What is the association between researchers' position in the network and their scientific performance and is it different for female and male researchers? The remainder of the paper is organized as follows: Section 2 describes data and methodology; Section 3 presents empirical results; Section 4 discusses the findings; Section 5 concludes the article; and Section 6 represents the limitations of this study and draws some directions for future research.

2. Data and methods

2.1. Data

Data collection and preparation involved several steps. First, the bibliographic data including but not limited to title, abstract, keywords, date of publication, author list, etc. were retrieved from Elsevier's Scopus, filtering in research articles, conference papers, book chapters, and books published from 2000 to 2019. We only included publications for which both title and abstract were available. We used the ("artificial intelligence" OR "machine learning" OR "deep learning") search query to extract AI-related publications where at least one of the mentioned phrases appeared in the title/abstract of the publication or in the keywords section. This resulted in 45,971 publications. We removed articles with no abstract available ($n = 237$). We also excluded publications for which none of the authors were assigned to a gender ($n = 6,055$ – please refer to Section 2.2.1 for more details). Therefore, the final dataset size is 39,679. It should be noted that we used cross-sectional data and considered AI-related papers published between 2000 and 2019. Since the main objective of our research is not exploring historical gender differences and explaining the reasons behind this issue but investigating gender-specific patterns in AI in the current century wherein the number of AI publications has significantly increased (Tang, Li, Ding, Song & Bu, 2020), and AI research has become more application-oriented (Yuan et al., 2020) due to some technological advances like big data, cloud computing, machine/deep learning, and artificial general intelligence. In the next step, we used a supervised 3-class machine learning classifier to infer the gender of AI scientists. Moreover, we used natural language processing (NLP) techniques for data preprocessing and feature engineering, and applied an unsupervised machine learning topic modeling algorithm to infer the authors' scientific research themes from scientific publications.

We then used social network analysis (SNA) to build the co-authorship network of AI researchers. For this purpose, we first constructed a bipartite co-authorship network, with authors and articles as the two types of nodes, and edges indicating the publication's authorship (De Nooy, Mrvar & Batagelj, 2005). In order to track gender-specific collaborative patterns at the author level, we transformed the bipartite network into a monopartite network wherein nodes represent authors characterized by gender, and two authors were connected if they had a joint publication. The final dataset contains 39,679 publications excluding unknown/unisex only articles ($n = 6,055$ (13%)) as well as those without abstract ($n = 237$ (0.5%)), and 114,371 authors (30,448 (27%) females and 83,923 males (73%)). The gender determination method and authors' disciplinary profiles inference are explained in detail in the following section.

2.2. Methods

2.2.1. Gender determination

Using machine learning techniques and natural language processing, we then applied an automatic gender assignment model trained on a large labeled dataset of names to infer researchers' gender from a diverse set of features such as their first and last names, affiliation, and country of origin. The tool trains a 3-class machine learning classifier on a huge labelled data set of names. It benefits from a tailored feature engineering component that expands the initial feature set (e.g., first and last names and authors' affiliations) to boost the performance of the model. In engineering and creating new features, we considered many factors such as linguistic rules in different countries. For example, female last names in Russian, Czech, and other Slavic languages end with the suffix 'ova'. We implemented such rules in the inference pipeline. Moreover, we used natural language processing techniques to enrich the feature set further by focusing on different parts of the first and last names. For example, we created features representing the location of researchers extracted from their affiliation, the last n characters of the last names, etc. Also, we calculated locations in which a given first name or last name is very frequent and added features representing this information. The machine learning classifier was trained on this enriched data set and output a label from the set of ['female', 'male', 'unisex/unknown']. From the model output, we classified authors' genders into female (F), male (M), and unisex/unknown (U). We labeled authors' genders as unisex/unknown (U) when our method could not detect the gender and excluded these authors ($\sim$27%, $n = 41,936$ authors) from further analysis. We carefully validated the accuracy of the assigned genders by automatic gender identification algorithm through manual verification on a random sample of 1,000 scientists. The algorithm identifies the gender of the researchers with 96% accuracy (94% for females, 98% for males). The purpose of this validation was to ensure that the authors that are used in the analytics pipeline are correctly identified as female/male. It should be noted that this is the accuracy of the automatic gender identification after removing those authors tagged as unisex/unknown by the algorithm. Out of the 39,679 publications included in this research, 24,378 papers have every author's gender identified.

2.2.2. Authors' disciplinary profiles

We adopted a topic modeling approach so as to infer the authors' scientific research themes and domains of interest and identify the degree of interdisciplinarity at the author level. The resulted topic vectors from the topic model would represent disciplinary profiles of the scientists in our target dataset, i.e., the AI-related research publications. Whereas journal disciplinary classification and departmental affiliations may be used to recognize the disciplinary profile (Schummer, 2004; Wang, Thijs & Glänzel, 2015), they cannot accurately represent the authors' academic backgrounds. For instance, computer scientists who apply their knowledge in various fields of studies like mathematics, physics, medicine, etc. cannot be fully and precisely represented by a merely single computer science discipline. On the other hand, academic publications could properly manifest the interdisciplinarity of individual researchers (Porter, Cohen, David Roessner & Perreault, 2007). Taken together, we decided to employ the Latent Dirichlet Allocation (LDA) topic modeling technique, developed by Blei, Ng and Jordan (2003), to derive the researchers' disciplinary profiles from their past publications. We first merged the titles and abstracts of the articles. The rationale for this integration is that titles cover

specific keywords to index papers, while abstracts contain much more succinct information and represent the research's main idea (Ebadi, Tremblay, Goutte & Schiffauerova, 2020). To apply the LDA model, several preprocessing steps were carried out on the corpus including the transformation of the textual data to lowercase, tokenization, removal of non-alphabetic characters, removal of the words with length less than three characters, and eliminating custom stop words. Next, we created a document-term matrix to prepare the data for topic modeling. We built several LDA models with different numbers of topics and then calculated and analyzed several evaluation metrics, such as perplexity and log-likelihood, to assess the quality of the models (Griffiths & Steyvers, 2004). Besides quantitative metrics, three domain experts evaluated the quality of the model by analyzing top topic keywords and document-topic distributions. They concluded that the optimal number of topics for our research objective is 8, including NLP, genomics-drug discovery, Internet of Things (IoT)-energy, decision support systems, computer vision-health informatics, unsupervised learning, machine/deep learning, and cyber security-network. As a result of the LDA model, each publication can be associated with more than one topic with a certain probability. Using the document-topic probability matrix generated by the LDA algorithm, we identified the authors' research fields based on average topic distribution over their past publications. Each author was then represented as a topic distribution vector consisting of 8 disciplines, where each component corresponded to the average topic distribution of the author's past articles under the given field. Moreover, we assumed the topic with the highest probability in the topic distribution vector for each author defines the author's primary discipline. For example, assume that author i has two publications P_1 and P_2 with document-topic probability vectors, including 8 research themes, as follows: P_1(0.5, 0, 0.1, 0, 0, 0.4, 0, 0) and P_2(0.6, 0, 0.2, 0.2, 0, 0, 0, 0), where elements in the vector represent the probability that the publication contains the research theme. We calculated the average topic distribution over author i's past publications (P_1 and P_2) and represented author i's disciplinary profile by a topic distribution vector $\vec{x}_i$(0.55, 0, 0.15, 0.1, 0, 0.2, 0, 0), in which each component shows the probability that author i published under the given discipline.

2.2.3. Descriptive indicators

To examine the impact of network properties on scientific performance and experience, we calculated and utilized several initial descriptive indicators such as the total number of publications published by each author as a proxy for scientific output and the total number of citations received by each article to measure research impact. It is worth noting that publications published in various fields or years accrue citations differently (Waltman, 2016), wherefore, to correct the effects of field and year of publication, we also calculated field normalized citation counts, which is defined by dividing the total citation counts received by each article by the average citation counts for all articles published in the same year and the same field (Paul-Hus et al., 2015). In addition, career length was used as a proxy for academic experience, defined as the difference between the authors' first and last publications' dates in the dataset.

2.2.4. Structural metrics

Having calculated the set of descriptive indicators, we largely adopted the framework proposed by Feng and Kirkley (2020) to calculate the structural metrics, and modified it a bit to satisfy our research objectives. They presented several metrics to capture the degree of disciplinary diversity among researchers at the uni-author, bi-author, and team levels. In this study, we extended their work by taking the gender aspect into account, aiming to assess the existence of disciplinary diversity or homophily among female and male AI researchers in a larger dataset and investigate gender-specific patterns in interdisciplinary collaborations. Moreover, we improved their approach by applying the probabilistic topic modeling technique (LDA) to better capture the research topics and expertise of scientists based on their research output. Following Feng abd Kirkley (2020), we utilized network structure analysis to categorize researchers based on their collaboration patterns and explore how different collaborative behavior could affect research performance and experience. We introduce each metric in detail in the rest of this section.

Uni-author level disciplinary diversity: We applied Shannon's entropy, i.e., one of the common indices to assess diversity (Aydinoglu, Allard & Mitchell, 2016; Feng & Kirkley, 2020; Gray, 2011) to measure male and female scientists' disciplinary diversity. Shannon's entropy index is used to measure the dispersion among various fields of the individual author's research background. The degree of interdisciplinarity at the single author level is then identified by the diversity of fields represented in his/her publications, which is defined by the topic distribution vector $\vec{x}_i$. Hence, the entropy of each author's publication history is computed as follows:

$$H_i = -\frac{1}{log\left(n_d^{(i)}\right)} \sum_{d=1}^{N_d} \vec{x}_{id}\, log\left(\vec{x}_{id}\right) \tag{1}$$

where $n_d^{(i)}$ is the number of unique fields in which researcher i published, and $\vec{x}_{id}$ represents the author i's disciplinary profile based on the average topic distribution of all past publications classified under discipline d (the dth entry of topic distribution vector $\vec{x}_i$). We divided the Shannon entropy metric by the maximum possible diversity index to normalize the Eq. (1) between 0 (no diversity) and 1 (highest diversity) and compensate for the variation in the number of disciplines to which researchers contribute (Shannon, 1948). One may note that authors who contributed to only one field ($n_d = 1$) were excluded from the calculation. Accordingly, an author with a high value of entropy would have more equal publication distribution across different disciplines as well as a higher degree of interdisciplinarity and vice versa. To summarize, the uni-author level diversity metric represents the degree of concentration of researchers in different research themes based on the content of their publications. For uni-author level analysis, we calculated the entropy measure defined in Eq. (1) for all individual authors, with gender identified as female or male, who contributed to 39,679 publications.

Bi-author level disciplinary diversity: To explore whether disciplinary homophily or diversity can be observed as a determining factor in forming a collaborative network in interdisciplinary research, we employed a pairwise similarity measure. The relationship between a pair of researchers can be indicated by edges in the scientific collaboration network, and it could be between authors from similar or diverse academic backgrounds; thus, the more disparate the co-authors' disciplinary profiles, the more heterogeneous their relationships are. To represent dyadic interactions between two genders, we classified edges of the co-authorship network into female-female (FF), female-male (FM), and male-male (MM) collaboration. Since several studies (e.g., Boschma, 2005; Nooteboom, 2000) confirmed that an appropriate level of cognitive proximity, i.e., the degree of similarities in shared "knowledge area and expertise", could provide more efficient communication and better common understanding among team members, we decided to examine the degree of cognitive proximity among AI scientists in following way. As suggested by Feng and Kirkley (2020), using researchers' topic distribution vectors ($\vec{x}$) as indicators of authors' knowledge area and expertise, we applied the cosine similarity to measure the degree of similarity or cognitive proximity between a pair of authors in terms of shared interdisciplinary profile as defined in Eq. (2). Cosine similarity calculates the cosine of two non-zero vectors of n dimensions (Han, Pei & Kamber, 2011). In this case, the two vectors contain the disciplinary profiles of authors.

$$S_{ij} = \frac{\vec{x}_i \cdot \vec{x}_j}{\|\vec{x}_i\| \|\vec{x}_j\|} \tag{2}$$

In the equation, $\|\vec{x}\|$ is the norm of the vector $\vec{x}$, and $\vec{x}_i \cdot \vec{x}_j$ denotes the inner product of $\vec{x}_i$ and $\vec{x}_j$. The cosine similarity measure (S_{ij}) indicates the similarity between the research profile of author i and j, and varies between 0 (authors having collaboration only with those from disparate research backgrounds) and 1 (authors having collaboration only with those with the same research backgrounds). For bi-author level analysis, we excluded papers with only two authors when at least one of the authors had no gender assigned ($n = 7,329$), including 32,350 publications in the calculations. Out of these 32,350 publications, 24,378 articles (~75.4%) have a gender assigned to all the authors.

Team-level disciplinary diversity: Team-level disciplinary diversity can be characterized by either the variety of team members' disciplines, i.e., interpersonal diversity, or the degree of individuals' disciplinary diversity, i.e., intrapersonal diversity (Bunderson & Sutcliffe, 2002). In different terms, a diverse team can be composed of researchers from various disciplines or different interdisciplinary individuals (Wagner et al., 2011), or a combination of both of them. Given such a definition, to investigate gender differences in the tendency to collaborate in interdisciplinary research, we followed a similar approach as in Feng and Kirkley, (2020) and considered both primary disciplines and authors' disciplinary profiles as proxies for interdisciplinarity. Additionally, to compare between two gender groups, we considered the authors who publish the same paper as a team and partitioned teams into three categories, labeled as female-only, male-only, and mixed-gender. Then for each team, within-group entropy is defined based on co-authors' primary disciplines to measure team-level interdisciplinarity. The within-group entropy of publication p is defined as:

$$\tilde{H}_P = -\frac{1}{\log\left(\min\left\{|p|, N_d\right\}\right)} \sum_{d=1}^{N_d} f_{pd} \log\left(f_{pd}\right) \tag{3}$$

where f_{pd} is the fraction of authors with primary discipline d (the discipline with the highest probability in the topic distribution vector $\vec{x}$) in paper p, N_d denotes the number of unique disciplines within each team, and $|p|$ is the total number of authors in a given paper. As a tight upper bound on the entropy is restricted either by N_d or $|p|$, a normalization factor of $(\log(\min\{|p|, N_d\}))^{-1}$ was introduced in Eq. (3).

Similar to Eq. (1), a high value of within-group entropy suggests a high degree of disciplinary diversity at the team level. Furthermore, we applied average within-group cosine similarity to measure the level of team interdisciplinarity based on team members' disciplinary profiles, i.e., topic distribution vectors. The average within-group cosine similarity for each publication is expressed as:

$$\tilde{S}_p = \frac{2}{|p|(|p| - 1)} \sum_{(i,j) \in p} S_{ij} \tag{4}$$

where $|p|$ is the total number of authors in a given paper, S_{ij} is the cosine similarity between author i and j vectors, and the summation is taken over all pairs of authors in the paper p. Using $2(|p|(|p| - 1))^{-1}$, we normalized the $\tilde{S}_p$ values between 0 and 1, with 1 indicating the maximum team disciplinary homophily. In summary, Eqs. (3) and (4) help us to explore interdisciplinarity at the team level from two perspectives: (1) based on the diversity of disciplines across team members, utilizing the within-group entropy, and (2) based on the disciplinary diversity of individuals within the team, using the average within-group cosine similarity. For team-level analysis, we excluded papers with only two authors when at least one of the authors had no gender assigned ($n = 7,329$). This resulted in 32,350 publications from which in 24,378 papers (~75.4%) a gender was assigned to all the authors. The remaining articles (32,350 – 24,378 = 7,972) have at least two authors with identified gender as female or male. For example, if there was a paper with 9 authors including 2 unisex authors, 4 females, and 3 males, we excluded 2 unisex authors and considered this paper with 7 authors (4 females and 3 males) and did the calculations.

Core-periphery researcher decomposition: Researchers' position in scientific collaboration networks could affect their performance (Ebadi & Schiffauerova, 2015b, 2016). Therefore, we investigated the relationship between researchers' brokerage role and their academic performance and examined how this relationship could differ between men and women. We used betweenness centrality as a proxy for researchers' influence, importance, and their brokerage role in the network. This measure can identify researchers who act as intermediaries between different groups of researchers, called gatekeepers. Gatekeepers can play a crucial role in the network by connecting different clusters and transferring and/or controlling the flow of information/knowledge between communities
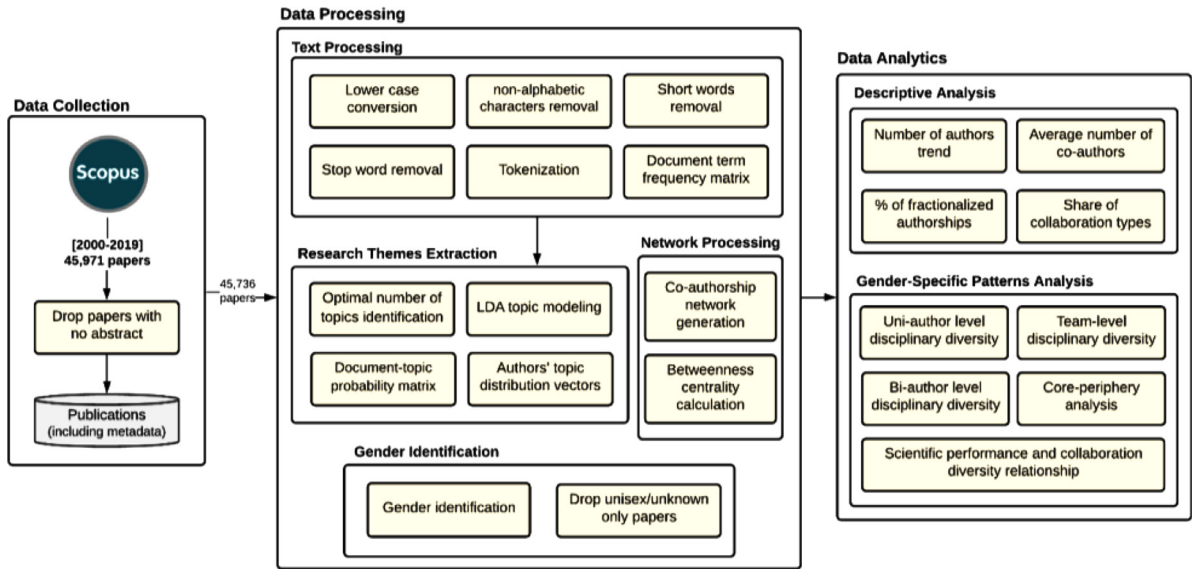
Fig. 1. The conceptual flow. The pipeline contains three main components, i.e., data collection, data processing, and data analytics. In the data collection component, publications, and their metadata such as title, abstract, keywords, date of publication, author list, etc. are collected from 2000 through 2019. In the data processing component, first, the text data (publications' title + abstract) is processed and used to extract authors' research themes represented by topic distribution vectors. Next, the gender of the authors is identified, their co-authorship network is generated, and their betweenness centrality measure is calculated. The target data is finally passed to the data analytics component to analyze gender-specific patterns within the co-authorship network.

(Ebadi & Schiffauerova, 2015a). Betweenness centrality of author i is defined based on the number of times that author i connects two other authors via the shortest path passing through author i (Borgatti, 2005). Hence, the betweenness centrality of the given node (bc_i) is defined as:

$$bc_i = \sum_{i \neq j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{5}$$

In Eq. (5), σ_{jk} denotes the total number of the shortest paths between node j and k, and $\sigma_{jk}(i)$ is the number of those paths containing node i. The betweenness centrality value ranges from 0 to 1, and the higher the value is, the more influence the node has. We performed core-periphery analysis as proposed by Feng and Kirkley (2020) to distinguish researchers based on their influence within the network. We utilized a different network measure from theirs and calculated betweenness centrality to determine the most central and influential female and male scientists. More specifically, we built a co-authorship network for the whole period of 2000–2019, including 114,371 researchers. We then ranked researchers based on their betweenness centrality values and categorized nodes into two roles: (1) The upper 5% of authors in terms of betweenness value are associated with the "*core*" role in the network, and (2) the remaining 95% as the "*periphery*" role. This helped us to identify core researchers who act as a bridge between different communities and funnel the information in the network. Due to their roles, they may benefit from having more opportunities to collaborate more and produce higher impact research (Abbasi, Hossain & Leydesdorff, 2012; Yan & Ding, 2009). Fig. 1 illustrates the conceptual flow of the study.

3. Results

3.1. Descriptive analyses

The constructed co-authorship network contained 83,923 men and 30,448 women who contribute to 39,679 publications throughout 2000–2019. As seen in Fig. 2-a, the number of female and male researchers has been increasing steadily over the years, following an almost similar pattern. Despite the growth, female researchers are still fewer, overall making up only ~27% of the AI research community based on the publications' authorship. We also compared publication outputs of the two genders using fractional authorship counts, where each author is given *1/n* credit for the authorship of a given article where *n* is the number of co-authors in the article (Fig. 2-b). Although female authors' contribution to published articles is slightly increasing over time, they accounted for less than 30% of fractionalized authorships during the whole period. This partially implies that the gender gap persists in the AI research community. However, AI is a male-dominant field, and the gender gap in research productivity might have resulted from women's underrepresentation in this field.

Fig. 3-a shows the average number of distinct co-authors per paper for each gender (left y-axis) as well as the average total team size trend defined by the number of authors per paper (right y-axis). According to Fig. 3-a, females have more distinct co-authors on
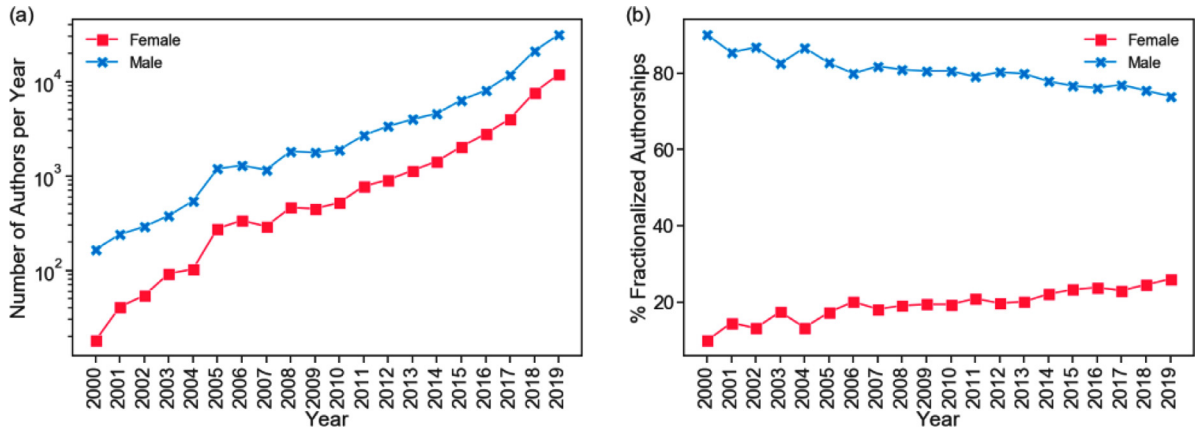
Fig. 2. (a) The total number of female and male authors per year, and b) the percentage of fractionalized authorships for two genders.
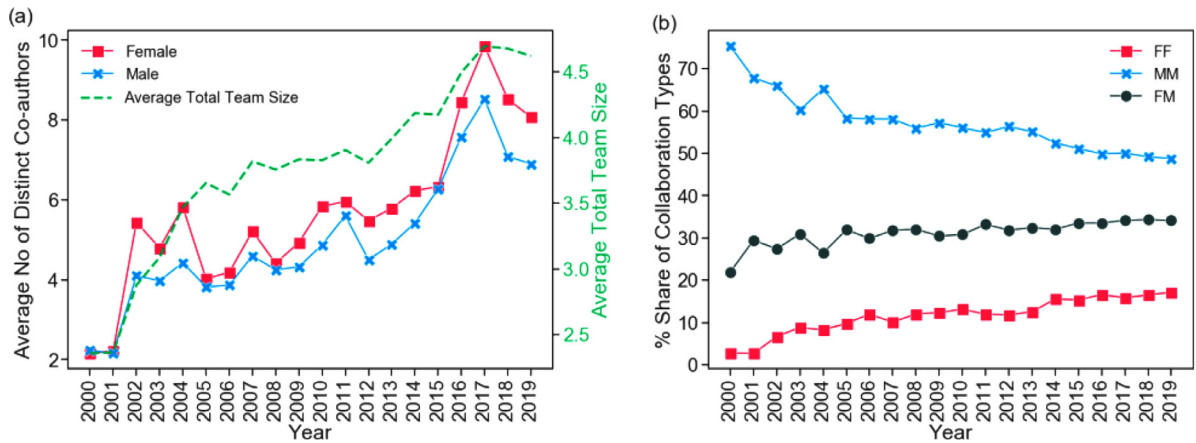

Fig. 3. (a) The average number of distinct co-authors per paper for each gender (left y-axis) and the average number of authors per paper, i.e., average total team size (right y-axis, which is shown by green dashed line), and b) the percentage of share of collaboration types. Note that collaboration types indicate edges in the co-authorship network, classified into female-female (FF), male-male (MM), and female-male (FM) collaborations.

average than their male counterparts, which is consistent with previous studies reporting that women are more collaborative than men (Abramo et al., 2013; Ghiasi, Harsh & Schiffauerova, 2018). Fig. 3-b shows the trends of different collaboration types. From the figure, it can be inferred that the share of male-male (MM) collaboration has followed a downward trend, whereas the shares of female-female (FF) and mixed-gender (FM) collaborations have been increasing over the years. This implies a change in collaboration preferences among the two gender groups and the increasing role and importance of female researchers in the AI community.

3.2. Uni-Author level disciplinary diversity

As an initial step in investigating gender-specific patterns of research background diversity in collaboration, we used normalized Shannon's entropy to examine the extent whereto disciplinary diversity exists in an individual's research background, which allows us to address the first research question (Q1). We excluded authors who contributed to only one discipline and analyzed the disciplinary profiles for a total of 114,313 authors, consisting of 83,880 (73%) men and 30,433 (27%) women. Fig. 4 shows the probability density distribution of entropies for researchers' publication history, containing eight main research subfields of AI represented by the topic distribution vector. From the smooth density plot it is observed that all authors, irrespective of their gender, tend to have diverse research backgrounds and almost a balanced publication distribution across different disciplines. However, the number of fields (n_d) that each author contributes to could negatively affect their ability to equally contribute to all those fields.

As depicted in Fig. 4-a and -b, both female and male scientists who publish in more fields ($n_d > 5$) have lower entropy values, meaning that they cannot have equal contributions to all the disciplines they are involved in. Within the group of researchers active in more scientific fields ($n_d > 5$), the distribution for both genders is bimodal. While the bimodality for the male group is relatively moderate, the gap between the two peaks for the female group is more pronounced, indicating that the female group is neatly divided into two significant clusters. Notwithstanding the mentioned similarities between the two gender groups, male researchers, on average, have slightly more balanced and diverse research profiles (higher entropy values) than their female counterparts. To
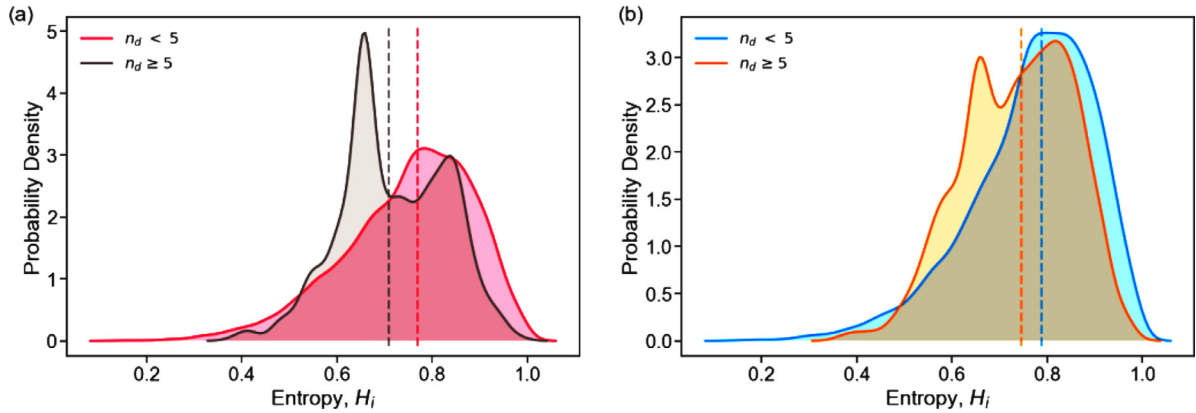
Fig. 4. Probability density of entropies for a) female, and b) male researchers. Dashed lines indicate the median of each distribution, and n_d is the number of unique disciplines in which researchers publish. Note that after examining several thresholds for n_d, we found 5 as the optimal cutoff.
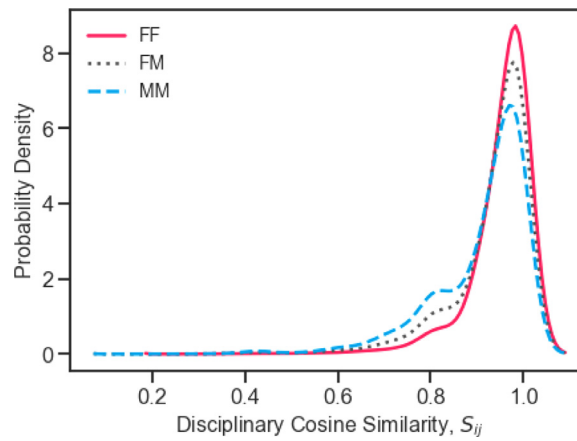

Fig. 5. Probability density of cosine similarity (S_{ij}) for female-female (FF), female-male (FM), and male-male (MM) collaboration types.

statistically support this hypothesis, the one-sided Mann-Whitney U test results revealed that female researchers' median entropy is significantly lower than that of males regardless of the number of fields ($n_d > 5$: Median$_{(female)}$=0.769, Median$_{(male)}$=0.788, $p < 0.001$); ($n_d > 5$: Median$_{(female)}$=0.707, Median$_{(male)}$=0.745, $p < 0.001$). We furthermore performed a permutation test with 10,000 permutations to determine the significance level. This test does not require any distributional and independence assumptions (Edgington, 1980). The results confirmed a significant difference between the medians of the two gender groups.

3.3. Bi-Author level disciplinary diversity

As explained in Section 2.2.2., in order to examine whether authors of each gender prefer to collaborate with researchers with similar research profiles, we calculated cosine similarity between researchers' topic distribution vectors for a total of 1,998,318 edges in the co-authorship network, consisting of 1,137,245 male-male (MM), 145,339 female-female (FF), and 715,734 female-male (FM) collaborations. Assuming the authors' topic distribution vectors as individual disciplinary profiles, we studied the dyadic interdisciplinarity between female and male researchers. Fig. 5 demonstrates the probability densities of the cosine similarity values (S_{ij}) for mixed-gender and same-gender dyads. As seen, cosine similarity values for FF edges are significantly higher than those of MM and FM edges, and mixed-gender dyads tend to have more similar research experience than male dyads. These observations highlight that females have a higher propensity than males to coauthor with other researchers who have similar interdisciplinary research backgrounds. Pairwise comparisons using the Mann-Whitney U test (two-sided) indicated that differences between all groups are statistically significant: (for all dyads including FF and MM, FF and FM, FM and MM: Median=1, $p < 0.001$). However, since the cosine similarity values had a left-skewed distribution and all groups had the same median, we used a one-tailed t-test to determine if the average S_{ij} of FF edges is higher than those of MM and FM edges. Results indicated that differences between all groups are statistically significant, suggesting higher disciplinary homophily levels among FF and FM dyads (FF and MM: Mean$_{(FF)}$= 0.986, Mean$_{(MM)}$= 0.978, $p < 0.001$); (FF and FM: Mean$_{(FF)}$=0.986, Mean$_{(FM)}$=0.981, $p < 0.001$); (FM and MM: Mean$_{(FM)}$=0.981, Mean$_{(MM)}$=0.978, $p < 0.001$). Similar to Section 3.2, a permutation test (10,000 permutations) was performed to identify the significance level. The above analysis
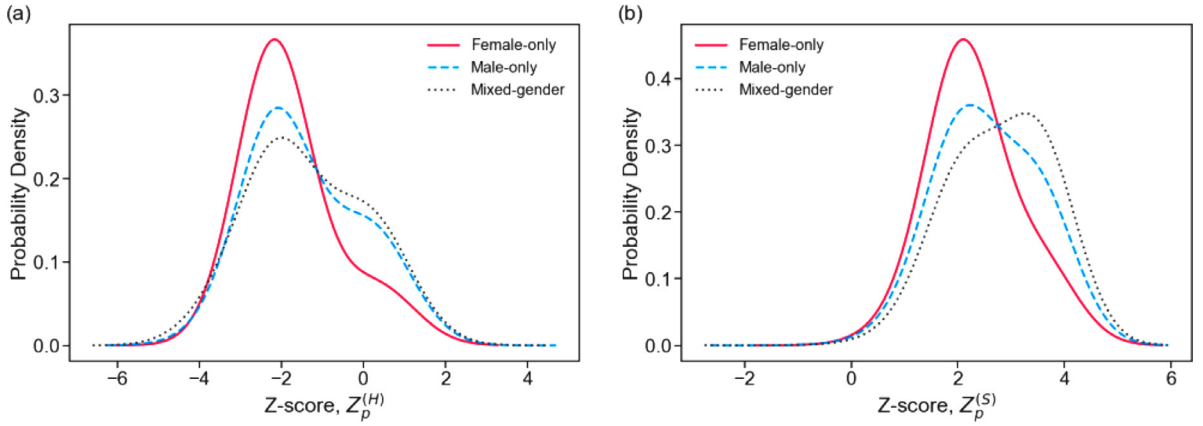
Fig. 6. Team-level disciplinary diversity analysis using 32,350 publications. a) Probability densities of $z_p^{(H)}$ for female-only, male-only, and mixed-gender teams, and b) probability densities of $z_p^{(S)}$ for female-only, male-only, and mixed-gender teams. The z-scores show the deviations from expected values of $\tilde{H}_P$ and $\tilde{S}_p$ in the randomized teams.

and our further investigation of team disciplinary diversity in the next section address our second research question (Q2) concerning the existence of disciplinary homophily in scientific collaborations.

3.4. Team-Level disciplinary diversity

In order to study team disciplinary diversity based on team members' primary disciplines, we applied the following strategy: (1) we considered the total number of unique team size (p) as the population and determined an optimal sample size for each population, assuming 95% confidence interval and 1% margin of error, (2) we generated randomized teams by applying random sampling without replacement from all authors in the real dataset, (3) we calculated within-group entropy ($\tilde{H}_P$) for both real and randomized teams, (4) for each unique team size ($|p|$) in the randomized data, we calculated the average ($\mu_{|p|}^{(H)}$) and standard deviation ($\sigma_{|p|}^{(H)}$) of the entropy values, (5) we compared the real teams with randomized teams by calculating the z-score ($z_p^{(H)}$), which is defined as:

$$z_p^{(H)} = \frac{\tilde{H}_p - \mu_{|p|}^{(H)}}{\sigma_{|p|}^{(H)}}, \tag{6}$$

and, (6) similarly, we replicated our analysis while using $\tilde{S}_p$ instead of $\tilde{H}_P$ and the z-score ($z_p^{(S)}$) to investigate team diversity based on team members' disciplinary profiles. Thereby,

$$z_p^{(S)} = \frac{\tilde{S}_p - \mu_{|p|}^{(S)}}{\sigma_{|p|}^{(S)}} \tag{7}$$

Fig. 6 shows the probability densities of $z_p^{(H)}$ and $z_p^{(S)}$ for female-only, male-only, and mixed-gender teams using 32,350 publications (as explained in Section 2.2.4). In the figure, z-scores indicate how many standard deviations the values of $\tilde{H}_P$ and $\tilde{S}_p$ diverge from what is expected by random. It is observed that for all collaboration teams, the distribution of $z_p^{(H)}$ is concentrated on negative values, suggesting that $\tilde{H}_P$ values for most of the teams are lower than what is expected for randomized groups. On the other hand, the distribution of $z_p^{(S)}$ is concentrated on positive values, meaning that teams with high $\tilde{S}_p$ are more common in the actual dataset than randomized. Additionally, the probability density of both $z_p^{(H)}$ and $z_p^{(S)}$ for female-only teams tend to be higher compared to the other groups. These observations highlight that although co-authorship teams, in general, tend to be composed of researchers with more similar disciplinary profiles, this preference is stronger among females, who have more homophilous collaborations compared to their male counterparts.

To further validate the findings, we performed a robustness check by excluding publications with unisex authors and only considering publications for which all the authors have a gender assigned ($n = 24,378$). As seen in Fig. 7, z-score values for within-group entropy are still concentrating on negative values, suggesting that $\tilde{H}_P$ values for most of the teams are lower than what is expected for randomized groups, and the distribution of $z_p^{(S)}$ is yet concentrating on positive values, i.e., that teams with high $\tilde{S}_p$ are more common in the actual dataset than randomized. For mixed-gender teams, in the previous analysis using 32,350 publications, the z-score values were more negative for entropy and more positive for cosine similarity, but the overall pattern is very similar. Comparing Figs. 6 and 7, and considering the same observed patterns, the robustness of the analysis is confirmed.
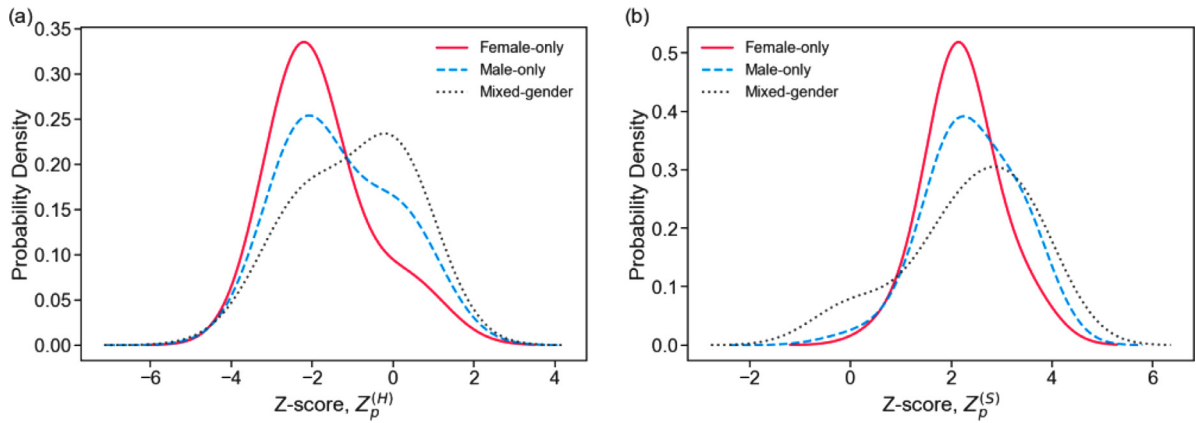
Fig. 7. Team-level disciplinary diversity analysis using 24,378 publications – the robustness check. a) Probability densities of $z_p^{(H)}$ for female-only, male-only, and mixed-gender teams, and b) probability densities of $z_p^{(S)}$ for female-only, male-only, and mixed-gender teams. The z-scores show the deviations from expected values of $\tilde{H}_P$ and $\tilde{S}_p$ in the randomized teams.
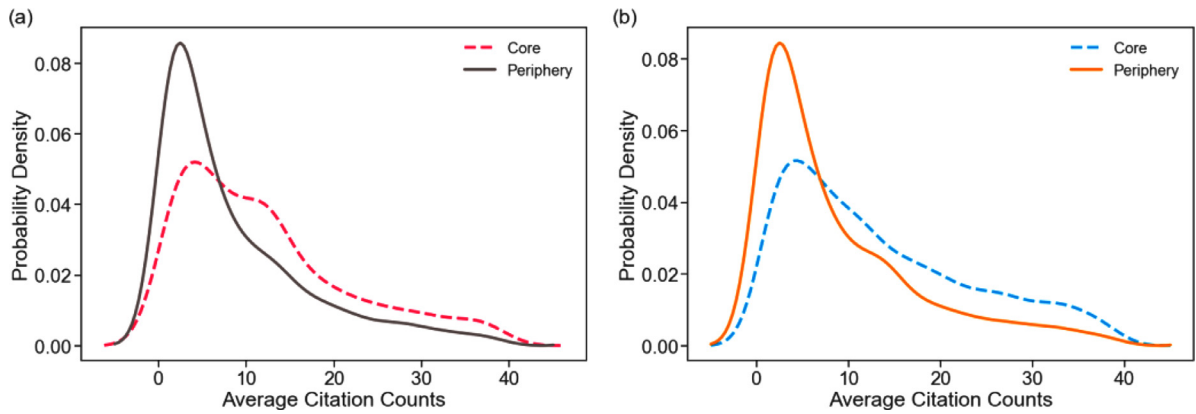

Fig. 8. Average citation counts distribution for (a) female, and (b) male researchers.

3.5. Scientific performance and collaboration diversity relationship

In this section, the last research question (Q3) is addressed by analyzing the relationship between female and male researchers' scientific performance and their collaboration diversity. First, using core-periphery decomposition, all network nodes, i.e., researchers, were divided into core and periphery nodes (see Section 2.2.2. for details). Out of the 114,371 authors, 5,720 researchers (5%) were associated as core researchers, including 4,197 males and 1,523 females. It was observed that male scientists have more brokerage roles in the network as their average betweenness centrality is 1.8 times higher than women's. However, the proportion of core females has increased by over 23% from 2000 to 2019, indicating that women are occupying more influential positions in tandem with men in the AI research community.

Next, we studied the impact of having core roles within the co-authorship network on research performance and experience. Fig. 8 shows the distribution of average citation counts, as a proxy for research impact, for female and male core/periphery researchers. As observed, the core researchers' distribution moves above the periphery researchers' distribution, regardless of gender, suggesting a positive relationship between occupying core positions in the co-authorship network and research impact in terms of citation counts. Notably, from Fig. 8-a and -b, it is seen that the curves for core and periphery researchers are almost the same for female and male authors.

We also investigated the relationship between research output, measured by the number of publications, and researchers' brokerage role in the co-authorship network, measured by betweenness centrality, for female and male researchers (Fig. 9). The same patterns were observed here as well, indicating a positive relationship between having core positions and research output for both female and male researchers. Additionally, the core/periphery density curves were comparable for female and male researchers (Fig. 9-a and -b).

Fig. 10 shows the probability density distributions of researchers' career length as a proxy for academic experience, measured by the difference between authors' first and last publications' dates in the dataset. As seen, the distributions of the core researchers' academic experience for both female and male researchers place above those of the periphery researchers. This suggests a positive relationship between having a core brokerage position in the co-authorship network and longer academic experience, regardless of
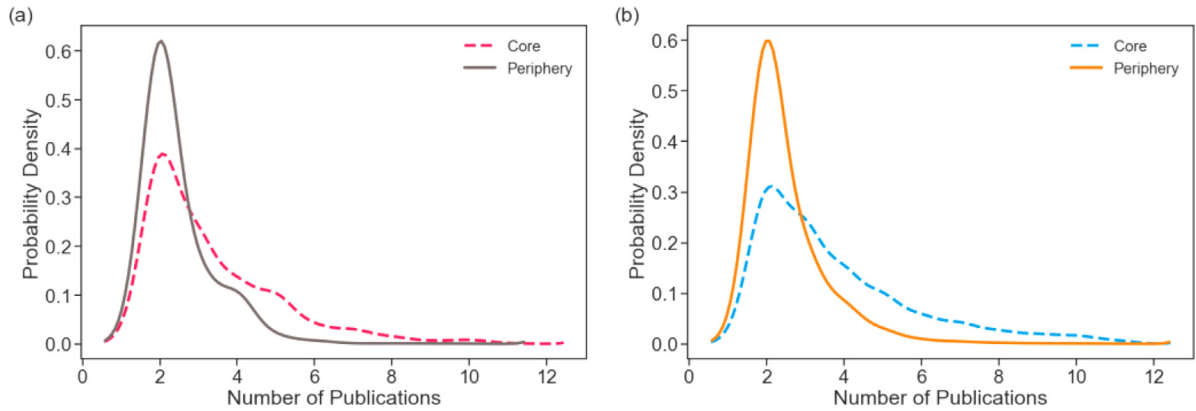
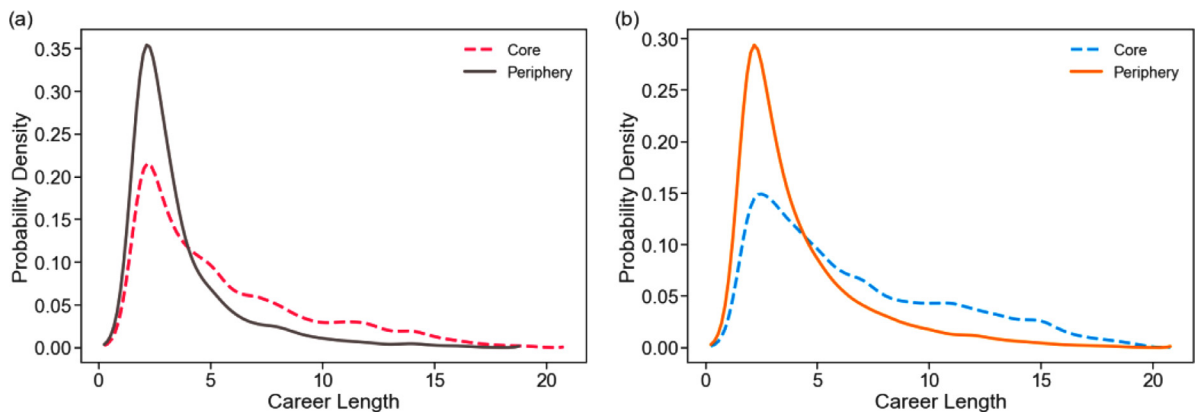**Fig. 9.** Number of publications distribution for (a) female, and (b) male researchers.

Fig. 10. Career length distribution for (a) female, and (b) male researchers.

gender. We validated our findings by performing the one-sided Mann-Whitney U test which resulted in significant differences in all cases ($p < 0.001$).

Motivated by these findings, we did further statistical tests to examine whether there are differences in research performance and experience between female and male researchers who are holding core positions in the network. The two-sided Mann-Whitney U test revealed that the difference between men and women in core positions was statistically significant in all cases ($p < 0.001$). Furthermore, the one-tailed *t*-test was performed to compare research performance between the two groups. Pairwise comparisons between the two groups showed that means of the career length and average citation counts of core females were significantly lower than those of core males, and that women in core positions, on average, publish less than their male counterparts (average raw citation counts: $\text{Mean}_{(core\ female)}=20.584$, $\text{Mean}_{(core\ male)}=28.236$, $p < 0.001$); (career length: $\text{Mean}_{(core\ female)}=4.912$, $\text{Mean}_{(core\ male)}=6.093$, $p < 0.001$); (number of publications: $\text{Mean}_{(core\ female)}=3.462$, $\text{Mean}_{(core\ male)}=4.126$, $p < 0.001$). We moreover calculated field normalized citation counts and observed similar relationships as the ones seen for the average citation count. Our findings show that authors who collaborate with diverse groups are more likely to have higher seniority levels and scientific performance in terms of both quantity and impact. Weaker research performance was observed for core females compared to the core male researchers; however, it is worthy to mention that having shorter career length as well as having a fewer number of female researchers in the AI scientific ecosystem, as suggested by our data, may have a negative impact on the academic performance of female researchers.

Finally, we closely investigated the collaborative relationships among top core AI researchers, defined as the top-0.1% researchers with the highest betweenness centrality ($n = 114$, 15 females and 99 males). Fig. 11-a shows their co-authorship network. The red nodes represent female researchers and the blue nodes represent male researchers. The higher the betweenness centrality value, the bigger the node is in the network. Edges show co-authorship relationships, i.e., if two authors have a joint paper they are directly connected in the graph, and the thickness of edges indicates the collaboration frequency. As seen, not only is the network male dominant, but the top core researchers are also male, i.e., the biggest nodes in the graph. In addition, the gender homophily effect can be observed among both genders meaning that they seem to be collaborating more with researchers of the same gender as theirs. Nevertheless, some female core researchers are creating mixed-gender clusters by joining male-dominant clusters. Fig. 11-b compares the performance of top core AI researchers. As one may observe, top male researchers have more direct connections than their female counterparts, measured by the average degree, however, the gap is negligible. Nonetheless, in terms of the average number
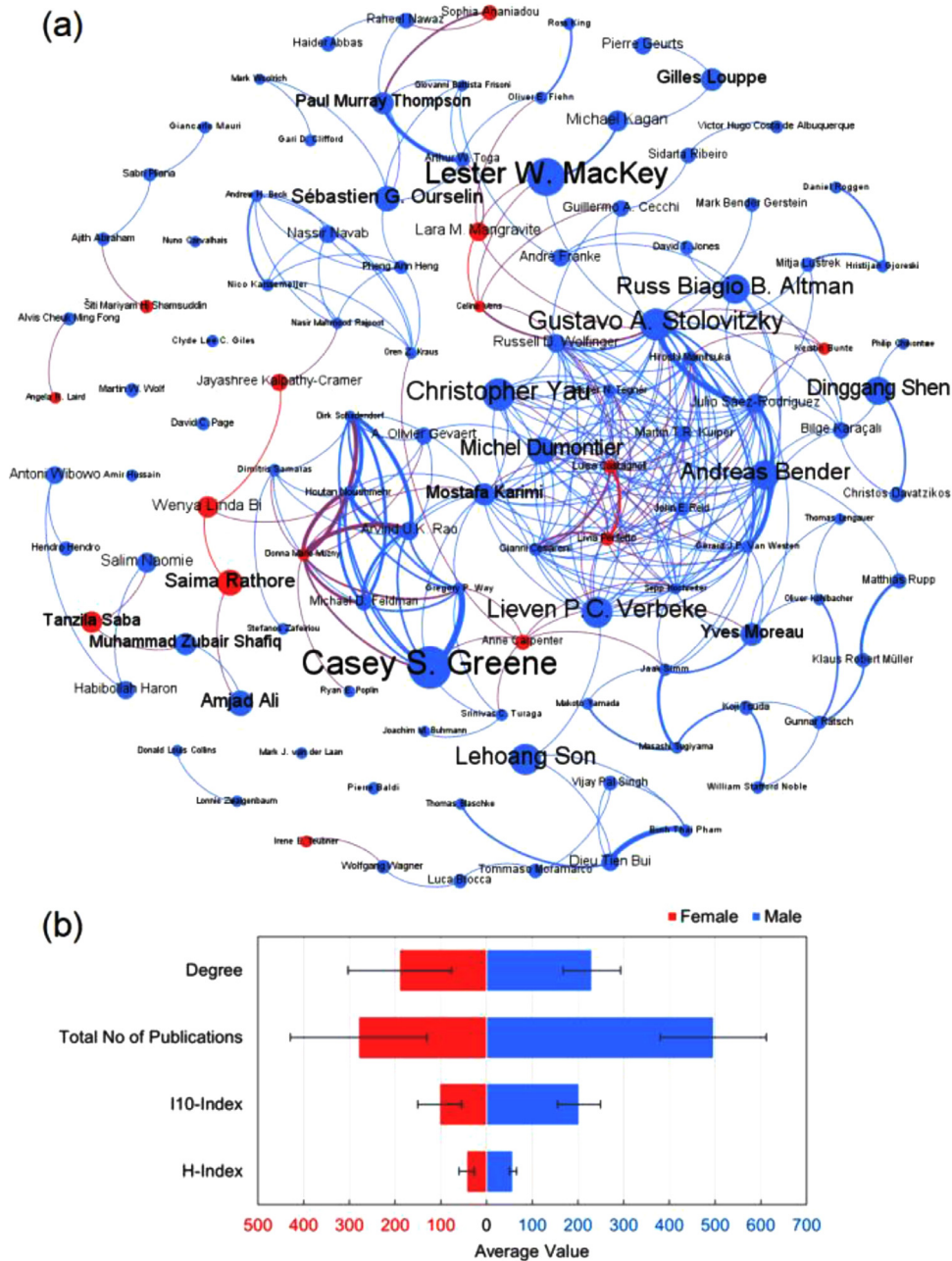
Fig. 11. (a) The co-authorship network of most central AI researchers, i.e., the top-0.1% researchers with the highest betweenness centrality (*n* = 114, 15 females and 99 males). Female nodes are colored red and male nodes blue, and b) comparison of male and female AI researchers based on their degree, the total number of publications, h-index, and i10-index. The error bar represents a 95% confidence interval for the mean value per gender.

of publications, the male researchers produced almost twice the female researchers. Moreover, papers published by males have been more cited, according to their average h-index and i-10 index. These observations are akin to our findings from the core-periphery analysis, indicating the weaker academic performance for women in brokerage roles in the AI community. Regardless of gender, these top core researchers are mostly active in health and biology, partially explaining their high number of direct connections/co-authors. We also compared the ratio of their AI publications in the Scopus database over their total number of publications. The average ratio is slightly higher for men, 4.1% vs 3.7% with 95% confidence intervals of [2.9%, 5.3%] and [2.5%, 5.1%] for males and females, respectively.

4. Discussion

Despite notable progress, gender disparity in science and scientific activities remains, calling for a more systematic and comprehensive approach to tackle and investigate the problem. Many initiatives have endeavored over the last decades to fill the gender gap in the academic and research community. As a highly interdisciplinary and evolving domain, artificial intelligence is constantly affecting many angles of human life, not only attracting the attention of researchers but also the governments and decision-/policymakers. This study complements existing literature on gender differences in scientific collaboration by analyzing disciplinary diversity among female and male researchers at the uni-author, bi-author, and team levels, as well as assessing the influence of disciplinary homophily or diversity on scientific research team formation. We applied multiple techniques, i.e., natural language processing, machine learning, social network analysis, and statistical analysis, to comprehensively investigate gender-specific patterns in the AI scientific ecosystem between 2000 and 2019. To the best of the authors' knowledge, this is the first study that addresses gender differences in the propensity to collaborate concerning the similarity or dissimilarity of researchers' academic backgrounds in a highly-interdisciplinary and evolving field such as AI.

In line with general conclusions of a recent report on the existence of a global gender gap in AI published by the World Economic Forum (World Economic Forum, 2018), we also found that the number of male researchers publishing in the AI domain has been constantly higher than their female counterparts in the entire examined period, with an increasing trend observed for both genders. Notwithstanding, our findings suggest that male and female AI researchers are collaborating together more, indicated by the increasing trend for mixed-gender collaborations. This property was recently observed in some other interdisciplinary fields such as nanotechnology (Ghiasi et al., 2018). This property, along with our observed higher preference of female researchers toward collaboration, may facilitate the knowledge transfer between genders, resulting in generating more female AI superstar researchers that itself could inspire and engage more women at the forefront of the AI domain. Thus, our finding provides new insights into the changing role of female AI scientists and their active involvement in the field, which is in line with the high-level objectives of the frontline countries investing in AI and partially confirms the effectiveness of the policies taken to generate and support women AI leaders. As an example, in Canada, a new initiative has recently been put in place to tackle diversity in AI and foster the next generation of female leaders (CIFAR, 2019), as only 14% of Canadian AI researchers are women (Shelementaicom, 2019). In our dataset, we observed a 22% share for female Canadian AI researchers. This could indicate that women are still underrepresented in the Canadian AI research community despite implementing new initiatives. The possible explanation could be that such initiatives require more time to bear fruit and narrow the gender gap. Of note, the difference between the share of female Canadian AI researchers (14% versus 22%) might be due to using different datasets.

Analyzing male and female scientists' disciplinary diversity revealed that although female scientists have slightly less balanced and less diverse research profiles than males, all authors, regardless of gender, tend to have a wide range of experience in different AI subject areas. In contrast, the tendency to have equal contribution to all fields is decreased as they engage in more research areas. This is in accordance with Feng and Kirkley (2020), demonstrating that individual-level disciplinary diversity is a common indicator of interdisciplinary research collaborations. Generally, several studies reported that researchers prefer working with others who have a certain degree of similarity between different characteristics (AlShebli, Rahwan & Woon, 2018; Holman & Morandin, 2019). Our findings denoted that while individual AI researchers are generally involved in diverse AI subfields, they are more likely to collaborate with authors engaged in a similar set of AI subfields as theirs. This could imply that the interdisciplinary nature of AI research mainly stems from individual researchers' disciplinary diversity, i.e., intrapersonal diversity, rather than diversity among researchers within the team, i.e., interpersonal diversity. Thus, uni-author level disciplinary diversity plays a significant role in the formation of interdisciplinary collaborations within this field. In other words, a given AI research team is more likely to consist of researchers who have diverse research expertise individually but have a similar set of expertise with other team members. To borrow from Bunderson and Sutcliffe (2002), teams composed of interdisciplinary individuals might be better capable of sharing their knowledge and overcoming common collaboration barriers, e.g., different mindsets, lack of mutual understanding, and communication problems, which could lead to higher team effectiveness. These findings provide insights for both researchers and organizations about team composition characteristics in the AI scientific ecosystem from a gender perspective.

Our analysis of gender-specific collaborative patterns confirmed the existence of homophily in terms of disciplinary profiles at both bi-author and team levels, consistent with previous findings by Feng and Kirkley (2020). Broadly, while AI researchers have, on average, a high degree of interdisciplinarity at the individual level, they are inclined to collaborate with authors who are like themselves in terms of shared interdisciplinary profiles; therefore, our results complement previous studies reporting the presence of homophily in academic collaborations in terms of gender (Holman & Morandin, 2019; Jadidi et al., 2018), affiliation, discipline, and ethnicity (AlShebli et al., 2018). Further, we found that women exhibit a stronger tendency toward disciplinary homophily compared to men in the AI scientific ecosystem. Female scientists tend to form more homophilous collaborative ties, i.e., interacting more with scientists from similar research backgrounds, and this tendency becomes stronger when it comes to female-female collaborations. Several studies (e.g., Brewer, 1999; The National Academies., 2005) indicated that although interdisciplinary collaboration, composed of individuals from diverse disciplines, could positively impact knowledge and innovation production in science, it could also propose some daunting challenges such as communication barriers, lack of common ground, etc. Given mentioned challenges, we speculate that uni-author level disciplinary diversity, along with homogeneity in AI research teams, may facilitate information/knowledge sharing among team members while maintaining the interdisciplinary nature of teams. This could be in accordance with some literature (e.g., Boschma, 2005; Nooteboom, 2000) showing that if members have an appropriate level of similarities in their knowledge area and expertise, known as cognitive proximity, they could communicate more effectively to exchange their knowledge. However, we also assume that researchers who are primarily collaborating with those of the same disciplinary profile, in a highly evolving and

interdisciplinary ecosystem such as AI, might be less exposed to the cutting-edge advancements in the field that could affect the gender gap in time. As suggested by Boschma (2005) and Nooteboom (2000), while having a similar knowledge base could positively affect collaborative activities, too many similarities could also impede novelty. Therefore, it is suggested that researchers of both genders tend to maintain a balance between collaborating with others with similar and diverse disciplinary profiles to benefit from both efficient communication and accessing new knowledge leading to novel research.

Lastly, the core-periphery analysis indicated a significant positive association between having diverse collaboration and scientific performance and experience. In line with the existing literature (Abbasi et al., 2012; Feng & Kirkley, 2020; Yan & Ding, 2009), our findings demonstrated that having core positions within the scientific network could positively enhance scholars' academic performance and, as a result, their success. Core researchers can play an exceedingly important role in developing new collaboration ties with researchers from diverse communities, enabling them to reap the benefits of a wider range of knowledge and experience as well as higher scientific performance. In addition, our findings suggest tangible differences between the two genders occupying core network positions. On average, core female scientists have lower seniority levels and scientific performance in terms of both productivity and impact. We speculate that this lower performance might be due to the fact that women are underrepresented in AI disciplines and they are more likely to leave research careers or pause their scientific activities for a plethora of reasons such as family obligations (Ramos et al., 2015), inadequate female role models (Nelson & Rogers, 2003), masculinist working atmosphere, and professional preferences (Wang & Degol, 2017). If true, this would aid in explaining fewer influential women in the AI community, impeding females' academic success.

5. Conclusion

Undoubtedly being the main driver of many emerging technologies, artificial intelligence is changing the world and many aspects of the way we live, impacting tremendously our future. As an evolving and complex ecosystem that is attracting significant investments annually, training leading AI researchers is also complex and costly. In this study, we comprehensively analyzed gender-specific patterns in the AI scientific ecosystem from 2000 to 2019 and shed light on gender differences in collaborative behavior and how these differences could impact academic success. We also highlighted that disciplinary homophily is one of the main team composition characteristics in the AI scientific ecosystem. In spite of the gap observed between the female and male researchers active in the field, we found promising indications of growth in the number of female scientists in the AI ecosystem. The differences observed between genders in performing scientific activities and collaboration are of crucial importance. We believe our analysis provides a comprehensive picture of the gender-specific patterns in AI that could help the policymakers to adjust existing or set new strategies to support female researchers more sustainably and efficiently by considering all the aspects of the differences between females and males not only in initiating, leading, and maintaining collaboration and scientific activities but other influencing factors such as social and family responsibilities that could impact their professional career.

6. Limitations and future work

We focused on gender-specific patterns in the AI field as a highly interdisciplinary male-dominated field (World Economic Forum, 2018), and this characteristic might partially affect our results. Thus, future research could analyze other emerging technologies to compare the results. We used Scopus to extract the target scientific publications. Scopus and other similar data sources are English biased and non-English articles could be underrepresented. Moreover, while a recent study (Huang et al., 2020) utilized longitudinal data to investigate gender inequality in academia and reported that career-wise differences could mainly explain gender disparity in research performance, we used cross-sectional data in this work and focused on only researchers' AI-related activities. Hence, future research may consider a longitudinal perspective and examine the entire research activities of scientists to explore gender-specific patterns and identify any possible explanations. We applied an automatic gender assignment algorithm and excluded authors for which we were unable to assign gender. However, we ensured that the assigned genders are validated by verifying a random sample of 1,000 authors. Manual verification is a time-consuming and costly process. Future research could explore other gender assignment strategies such as manual and/or semi-automatic. Another limitation related to the automatic gender identification is we observed that the majority of the authors for whom the algorithm could not identify their gender are Asian researchers (mainly Chinese and Korean), hence, AI-related research performed by Asian people could be underrepresented in our analysis. We partially addressed this limitation for Asian researchers who reside outside of Asia since some of them used western nicknames that were easily identified by our gender identification algorithm. But for Asian researchers with no western nicknames the automatic gender identification tool could not highly perform. Of course, manual verification and search per author over the internet could address this limitation (although manual identification is also prone to errors), however, it was very time-consuming and costly considering the size of our examined dataset. In this study, we analyzed collaboration patterns among researchers through a co-authorship network. Although the co-authorship network is regarded as one of the most common and tangible indicators to measure scientific collaboration, it cannot capture all collaboration types inasmuch as collaborative efforts do not always result in a joint publication (Katz & Martin, 1997). Another future direction could be addressing this issue by taking other approaches to measure research collaboration. We used betweenness centrality to measure researchers' brokerage role in the co-authorship network. Future studies may consider other centrality measures and analyze their relationships with researchers' scientific performance. Additionally, we considered the time difference between the authors' first and last publications as a proxy for their career age as we did not have data on researchers' age and/or seniority level. One may use other proxies for authors' seniority level in future research and compare the results. Lastly, we

used articles' abstracts and titles to infer researchers' disciplinary profiles. Future research could further explore the authors' research disciplines by taking the entire publications' text into the account.

Data availability

Data is available upon request. Please contact the corresponding author.

Declaration of Competing Interest

The authors have no competing interests.

CRediT authorship contribution statement

Anahita Hajibabaei: Data curation, Methodology, Visualization, Formal analysis, Writing – original draft, Writing – review & editing. **Andrea Schiffauerova:** Formal analysis, Resources, Supervision, Writing – original draft, Writing – review & editing. **Ashkan Ebadi:** Conceptualization, Methodology, Visualization, Formal analysis, Supervision, Writing – original draft, Writing – review & editing.

References

Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics, 6*(3), 403–412. 10.1016/j.joi.2012.01.002.

Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). Gender differences in research collaboration. *Journal of Informetrics, 7*(4), 811–822. 10.1016/j.joi.2013.07.002.

Alper, J. (1993). The pipeline is leaking women all the way along. *Science, 260*(5106), 409–411 (New York, N.Y.). 10.1126/science.260.5106.409.

AlShebli, B. K., Rahwan, T., & Woon, W. L. (2018). The preeminence of ethnic diversity in scientific collaboration. *Nature Communications, 9*(1), 5163. 10.1038/s41467-018-07634-8.

Aydinoglu, A. U., Allard, S., & Mitchell, C. (2016). Measuring diversity in disciplinary collaboration in research teams: An ecological perspective. *Research Evaluation, 25*(1), 18–36. 10.1093/reseval/rvv028.

Bennett, L. M., & Gadlin, H. (2012). Collaboration and team science: From theory to practice. *Journal of Investigative Medicine, 60*(5), 768–775.

Berryman, S.E. (1983) Who will do science? Trends, and their causes in minority and female representation among holders of advanced degrees in science and mathematics. A special report.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022. https://www.jmlr.org/papers/v3/blei03a.

Bonetta & Clayton. (2008) Remarkable women in science.

Bordons, M., Morillo, F., Fernández, M. T., & Gómez, I. (2003). One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. *Scientometrics, 57*(2), 159–173. 10.1023/A:1024181400646.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks, 27*(1), 55–71. 10.1016/j.socnet.2004.11.008.

Borrego, Á., Barrios, M., Villarroya, A., & Ollé Castellà, C. (2010). Scientific output and impact of postdoctoral scientists. *Scientometrics, 83*, 93–101. 10.1007/s11192-009-0025-y.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies, 39*, 61–74. 10.1080/0034340052000320887.

Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: Implications for scientific and technical human capital. *Research Policy, 33*, 599–616. 10.1016/j.respol.2004.01.008.

Brewer, G. D. (1999). The challenges of interdisciplinarity. *Policy Sciences, 32*(4), 327–337.

Bunderson, J. S., & Sutcliffe, K. M. (2002). Comparing alternative conceptualizations of functional diversity in management teams: process and performance effects. *The Academy of Management Journal, 45*(5), 875–893 JSTOR. 10.2307/3069319.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest, 15*(3), 75–141. 10.1177/1529100614541236.

CIFAR. (2019) How Canada is supporting the next generation of women AI leaders. https://www.cifar.ca/cifarnews/2019/05/31/how-canada-is-supporting-the-next-generation-of-women-ai-leaders.

Cole, J., & Zuckerman, H. (1984). The productivity puzzle. *Advances in Motivation and Achievement*, 217–258.

Cummings, J. N., Kiesler, S., Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science*, 703–722.

De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with pajek: Revised and expanded edition for updated software* p. 509.

Duch, J., Zeng, X. H. T., Sales-Pardo, M., Radicchi, F., Otis, S., Woodruff, T. K., et al. (2012). The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS One, 7*(12), e51332. 10.1371/journal.pone.0051332.

Ebadi, A., & Schiffauerova, A. (2015a). How to become an important player in scientific collaboration networks? *Journal of Informetrics, 9*(4), 809–825. https://ideas.repec.org/a/eee/infome/v9y2015i4p809-825.html.

Ebadi, A., & Schiffauerova, A. (2015b). How to receive more funding for your research? Get connected to the right people!. *PLoS One, 10*(7), Article e0133061. 10.1371/journal.pone.0133061.

Ebadi, A., & Schiffauerova, A. (2016). How to boost scientific production? A statistical analysis of research funding and other influencing factors. *Scientometrics, 106*(3), 1093–1116. 10.1007/s11192-015-1825-x.

Ebadi, A., Tremblay, S., Goutte, C., & Schiffauerova, A. (2020). Application of machine learning techniques to assess the trends and alignment of the funded research output. *Journal of Informetrics, 14*(2), Article 101018. 10.1016/j.joi.2020.101018.

Edgington, E. S. (1980). *Randomization tests*: 31. Boca Raton, FL: Chapman & Hall/CRC https://zbmath.org/?q=an%3A0439.62030.

Elsevier. (2017) Gender in the global research landscape. https://www.elsevier.com/research-intelligence/campaigns/gender-17

Feng, S., & Kirkley, A. (2020). Mixing patterns in interdisciplinary Co-authorship networks at multiple scales. *Scientific Reports, 10*(1), 7731. 10.1038/s41598-020-64351-3.

Fox, M. F., & Mohapatra, S. (2007). Social-organizational characteristics of work and publication productivity among academic scientists in doctoral-granting departments. *The Journal of Higher Education, 78*(5), 542–571. 10.1353/jhe.2007.0032.

Ghiasi, G., Harsh, M., & Schiffauerova, A. (2018). Inequality and collaboration patterns in Canadian nanotechnology: Implications for pro-poor and gender-inclusive policy. *Scientometrics, 115*. 10.1007/s11192-018-2701-2.

Ghiasi, G., Larivière, V., & Sugimoto, C. R. (2015). On the compliance of women engineers with a gendered scientific system. *PLoS One, 10*(12), Article e0145931. 10.1371/journal.pone.0145931.

Government of Canada. (2018,. July 25) NSERC - made in Canada athena SWAN consultation. https://www.nserc-crsng.gc.ca/NSERC-CRSNG/EDI-EDI/Athena-SWAN_eng.asp

Government of Canada. (2019,. March 19) Department of finance-budget 2019. https://www.budget.gc.ca/2019/docs/plan/chap-01-en.html#improving-gender-and-diversity-outcomes-in-skills-programs

Gray, R. M. (2011). *Entropy and information theory* (2nd ed.). Springer US. 10.1007/978-1-4419-7970-4.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(suppl 1), 5228–5235.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Elsevier Science & Technology http://ebookcentral.proquest.com/lib/concordia-ebooks/detail.action?docID=729031.

Hogan, A., Zippel, K., Frehill, L., & Kramer, L. (2010). *Report of the international workshop on international research collaboration*. Arlington, VA: National Science Foundation.

Holman, L., & Morandin, C. (2019). Researchers collaborate with same-gendered colleagues more often than expected across the life sciences. *PloS One, 14*(4), Article e0216128. 10.1371/journal.pone.0216128.

Huang, J., Gates, A. J., Sinatra, R., & Barabási, A. L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences, 117*(9), 4609–4616. 10.1073/pnas.1914221117.

Hunter, L., & Leahey, E. (2010). Parenting and research productivity: New evidence and methods. *Social Studies of Science, 40*, 433–451 SOC STUD SCI. 10.1177/0306312709358472.

Huyer, S. (2018) Gender equality will encourage new solutions and expand the scope of research; it should be considered a priority by all if the global community is serious about reaching the next set of development goals. 21.

Jadidi, M., Karimi, F., Lietz, H., & Wagner, C. (2018). Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *Advances in Complex Systems, 21*(03n04), Article 1750011. 10.1142/S0219525917500114.

Jurviste et al.M(2016).aternity and paternity leave in the EU. https://scholar.google.com/scholar_lookup?title=Maternity%20and%20Paternity%20Leave%20in%20the%20EU&publication_year=2016&author=U.%20Jurviste&author=M.%20Prpic&author=G.%20Sabbati

Kamraro. (2014,. March 26) Promoting gender equality in research and innovation [Text]. Horizon 2020 - European commission. https://ec.europa.eu/programmes/horizon2020/en/h2020-section/promoting-gender-equality-research-and-innovation

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18.

Kwiek, M. (2020). Internationalists and locals: International research collaboration in a resource-poor system. *Scientometrics, 124*(1), 57–105. 10.1007/s11192-020-03460-2.

Kyvik, S., & Teigen, M. (1996). Child care, research collaboration, and gender differences in scientific productivity. *Science, Technology & Human Values, 21*(1), 54–71. 10.1177/016224399602100103.

Larivière, V., Diepeveen, S., Ni Chonaill, S., Macaluso, B., Pollitt, A., & Grant, J. (2013a). International comparative performance of mental health research, 1980–2011. *European Neuropsychopharmacology, 23*(11), 1340–1347. 10.1016/j.euroneuro.2013.01.006.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013b). Bibliometrics: Global gender disparities in science. *Nature News, 504*(7479), 211. 10.1038/504211a.

Larivière, V., Vignola-Gagné, E., Villeneuve, C., Gélinas, P., & Gingras, Y. (2011). Sex differences in research funding, productivity and impact: An analysis of Québec University professors. *Scientometrics, 87*(3), 483–498. 10.1007/s11192-011-0369-y.

Leahey, E. (2006). Gender differences in productivity: Research specialization as a missing link. *Gender & Society, 20*(6), 754–780. 10.1177/0891243206293030.

Lee, E., Karimi, F., Wagner, C., Jo, H. H., Strohmaier, M., & Galesic, M. (2019). Homophily and minority-group size explain perception biases in social networks. *Nature Human Behaviour, 3*(10), 1078–1087. 10.1038/s41562-019-0677-4.

Lippa, R. A., Preston, K., & Penner, J. (2014). Women's Representation in 60 occupations from 1972 to 2010: More women in high-status jobs, few women in things-oriented jobs. *PloS One, 9*(5), e95960. 10.1371/journal.pone.0095960.

Long, J. S. (1992). Measures of sex differences in scientific productivity. *Social Forces, 71*(1), 159–178. 10.1093/sf/71.1.159.

Luoto, S. (2020). Sex differences in people and things orientation are reflected in sex differences in academic publishing. *Journal of Informetrics, 14*(2), Article 101021.

Mauleón, E., Bordons, M., & Oppenheim, C. (2008). The effect of gender on research staff success in life sciences in the Spanish National Research Council. *Research Evaluation, 17*, 213–225 Res Evaluat. 10.3152/095820208X331676.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474–16479. 10.1073/pnas.1211286109.

Murray, D., Siler, K., Larivière, V., Chan, W. M., Collings, A. M., Raymond, J., et al. (2019). Author-reviewer homophily in peer review. *BioRxiv*, Article 400515. 10.1101/400515.

Nelson, D. J., & Rogers, D. C. (2003). *A national analysis of diversity in science and engineering faculties at research universities*. National Organization for Women Washington DC.

Nielsen, M. W. (2016). Limits to meritocracy? Gender in academic recruitment and promotion processes. *Science and Public Policy, 43*(3), 386–399. 10.1093/scipol/scv052.

Nooteboom, B. (2000). *Learning and innovation in organizations and economies*. Oxford University Press. 10.1093/acprof:OSo/9780199241002.001.0001.

Ovseiko, P. V., Chapple, A., Edmunds, L. D., & Ziebland, S. (2017). Advancing gender equality through the Athena SWAN Charter for Women in Science: An exploratory study of women's and men's perceptions. *Health Research Policy and Systems, 15*. 10.1186/s12961-017-0177-9.

Paswan, J., & Singh, V. K. (2020). Gender and research publishing analyzed through the lenses of discipline, institution types, impact and international collaboration: A case study from India. *Scientometrics, 123*(1), 497–515. 10.1007/s11192-020-03398-5.

Paul-Hus, A., Bouvier, R. L., Ni, C., Sugimoto, C. R., Pislyakov, V., & Larivière, V. (2015). Forty years of gender disparities in Russian science: A historical bibliometric analysis. *Scientometrics, 102*(2), 1541–1553. 10.1007/s11192-014-1386-4.

Pearson, W., Frehill, L. M., & McNeely, C. L. (2015). *Advancing women in science: An international perspective*. Springer International Publishing. 10.1007/978-3-319-08629-3.

Porter, A. L., Cohen, A. S., David Roessner, J., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics, 72*(1), 117–147. 10.1007/s11192-007-1700-5.

Ramos, A. M. G., Cortés, J. N., & Moreno, E. C. (2015). Dancers in the dark: Scientific careers according to a gender-blind model of promotion. *Interdisciplinary Science Reviews, 40*(2), 182–203 Scopus. 10.1179/0308018815Z.000000000112.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics, 59*(3), 425–465. 10.1023/B:SCIE.0000018542.71314.38.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379–423.

Shaw, A. K., & Stanton, D. E. (2012). Leaks in the pipeline: Separating demographic inertia from ongoing gender differences in academia. *Proceedings of the Royal Society B: Biological Sciences, 279*(1743), 3736–3741. 10.1098/rspb.2012.0822.

Shelementaicom. (2019) Global AI talent report 2019. Jfgagne. https://jfgagne.ai/talent-2019/

Shen, H. (2013). Inequality quantified: Mind the gender gap. *Nature News, 495*(7439), 22. 10.1038/495022a.

Stack, S. (2004) Gender, children and research productivity. Research in Higher Education, 45(8), 891–920. 10.1007/s11162-004-5953-z

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*(6), 859–884. 10.1037/a0017364.

Tang, X., Li, X., Ding, Y., Song, M., & Bu, Y. (2020). The pace of artificial intelligence innovations: Speed, talent, and trial-and-error. *Journal of Informetrics, 14*(4), Article 101094. 10.1016/j.joi.2020.101094.

The National Academies. (2005). *Facilitating interdisciplinary research*. National Academies Press.

Thelwall, M., Bailey, C., Makita, M., Sud, P., & Madalli, D. P. (2019a). Gender and research publishing in India: Uniformly high inequality? *Journal of Informetrics, 13*(1), 118–131. https://ideas.repec.org/a/eee/infome/v13y2019i1p118-131.html.

Thelwall, M., Bailey, C., Tobin, C., & Bradshaw, N. A. (2019b). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics, 13*(1), 149–169. https://ideas.repec.org/a/eee/infome/v13y2019i1p149-169.html.

Tower, G., Plummer, J., & Ridgewell, B. (2007). A multidisciplinary study of gender-based research productivity in the worlds best journals. *Journal of Diversity Management (JDM), 2*(4), 23–32. 10.19030/jdm.v2i4.5020.

Uhly, K. M., Visser, L. M., & Zippel, K. S. (2015). Gendered patterns in international research collaborations in academia. *Studies in Higher Education*, 1–23. 10.1080/03075079.2015.1072151.

UNESCO. (2018) Improving measurement and policies for gender equality in science, technology and innovation-UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000266102

Van Arensbergen, P., Van der Weijden, I., & Van den Besselaar, P. (2012). Gender differences in scientific productivity: A persisting phenomenon? *Scientometrics, 93*(3), 857–868. 10.1007/s11192-012-0712-y.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics, 5*(1), 14–26.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics, 10*(2), 365–391. 10.1016/j.joi.2016.02.007.

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: distinct effects of variety, balance, and disparity. *PloS one, 10*(5). 10.1371/journal.pone.0127298.

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review,, 29*(1), 119–140. 10.1007/s10648-015-9355-x.

West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PloS One, 8*(7), e66212. 10.1371/journal.pone.0066212.

Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2019). Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet, 393*(10171), 531–540. 10.1016/S0140-6736(18)32611-4.

Wood, D. J., & Gray, B. (1991). Toward a comprehensive theory of collaboration. *The Journal of Applied Behavioral Science, 27*(2), 139–162. 10.1177/0021886391272001.

World Economic Forum. (2018) The global gender gap report 2018. http://www3.weforum.org/docs/WEF_GGGR_2018.pdf

Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review, 63*(6), 847–870. 10.2307/2657505.

Xie, Y., & Shauman, K. A. (2003). Women in science: Career processes and outcomes. *Harvard University Press, 82*(4), 1669–1671.

Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology, 60*(10), 2107–2118. 10.1002/asi.21128.

Yuan, S., Shao, Z., Wei, X., Tang, J., Hall, W., Wang, Y., et al. (2020). Science behind AI: The evolution of trend, mobility, and collaboration. *Scientometrics, 124*(2), 993–1013. 10.1007/s11192-020-03423-7.

Zeng, X. H. T., Duch, J., Sales-Pardo, M., Moreira, J. A. G., Radicchi, F., Ribeiro, H. V., et al. (2016). Differences in collaboration patterns across discipline, career stage, and gender. *PLOS Biology,, 14*(11), Article e1002573. 10.1371/journal.pbio.1002573.