



Structural gender imbalances in ballet collaboration networks

Yessica Herrera-Guzmán¹, Eun Lee^{2*} and Heetae Kim^{3,4*} 

*Correspondence:

eunlee@pknu.ac.kr;

hkim@kentech.ac.kr

²Department of Scientific Computing, Pukyong National University, Busan, Republic of Korea

³Department of Energy Engineering, Korea Institute of Energy Technology, Naju, Republic of Korea

Full list of author information is available at the end of the article

Abstract

Ballet, a mainstream performing art predominantly associated with women, exhibits significant gender imbalances in leading positions. However, the collaboration's structural composition vis-à-vis gender representation in the field remains unexplored. Our study investigates the gendered labor force composition and collaboration patterns in ballet creations. Our findings reveal gender disparities in ballet creations aligned with gendered collaboration patterns and women's occupation of more peripheral network positions than men. Productivity disparities show women accessing 20–25% of ballet creations compared to men.

Mathematically derived perception errors show the underestimation of women artists' representation within ballet collaboration networks, potentially impacting women's careers in the field. Our study highlights the structural imbalances that women face in ballet creations and emphasizes the need for a more inclusive and equal professional environment in the ballet industry. These insights contribute to a broader understanding of structural gender imbalances in artistic domains and can inform cultural organizations about potential affirmative actions toward a better representation of women leaders in ballet.

Keywords: Social network analysis; Ballet collaboration; Collaboration network; Gender imbalance; Perception error

1 Introduction

Ballet is widely recognized and appreciated around the world and is assumed as a women-dominated profession [1, 2]. However, men still dominate this particular performing art as recent reports show considerable gender imbalances [3]. This issue has been widely discussed in dance communities and questioned gendered representations in the ballet industry [4, 5]. For instance, data from American dance companies reveal the unequal representation of women (less than 40%) in artistic and executive positions [6], while the overall participation of women in the workforce is approximately 70% [7].

These reports have primarily focused on quantifying the percentage or the number of women and men artists involved, while the role of collaboration structures in contributing to gender imbalances in ballet remains poorly understood. The existing literature provides evidence that gendered variations in a social network structure contribute to different professional outcomes for men and women [8], thereby highlighting the importance of inves-

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

tigating the gender representation in collaborative structures. Moreover, many network research reveals, structural properties of collaborations influence the access to information [9, 10], creativity [11], productivity [12], and career success [13].

In particular, homophilic behaviors embedded in an imbalanced social structure can negatively affect the ranking of individuals from minority groups by enhancing segregation effects [14]. In an imbalanced social structure, individuals may inaccurately estimate the frequency of the minority group, resulting in perception errors regarding the representation of attributes in a social network [15, 16], resulting in the representativeness of the minority group can be over or underestimated in respect of what can be expected from the real representation in the network [17]. As perception errors could reinforce assortative patterns in social connections, such as collaborations, understanding the role of the network structure regarding gender imbalances could provide insights into an intervention of equal opportunities in professional positions.

In this work, we investigate the gender imbalance in the perspective of collaboration patterns of ballet creations. We hypothesize that if the network structure is unbalanced by gender, the imbalanced social structure will align with unequal collaborative behaviors and the existence of perception errors, which could explain why women artists are overlooked from leading positions in this industry. We construct collaboration networks from four renowned ballet companies and analyze their gender composition. The collaboration structures studied here mainly comprise a core structure of ballet creators, such as choreographers, composers, and costume and light designers. Then, we compare the real-world collaboration structures with randomized network models. We specifically explore the structural gendered differences and the labor force composition in highly central positions by using the network's centrality measures. We also measure the formation of perception errors on the women's group to examine a possible relationship between gendered collaboration networks and the perceived working environment.

To the best of our knowledge, our study is the first attempt that focuses on the understanding of the structural gender imbalances in major ballet companies. This research will facilitate the comprehension of the underlying social mechanisms driving gender inequalities in a highly collaborative performing art. We hope that this work will shed light on more effective interventions to reduce the segregation of women in creative careers.

2 Methods

2.1 Network of ballet creators

We construct the collaboration networks of ballet creators from four major ballet companies, namely, the American Ballet Theatre (ABT) [18], the New York City Ballet (NYCB) [19], the National Ballet of Canada (NBC) [20], and the Royal Ballet of the Royal Opera House (ROH) [21], based on their worldwide prestige and availability of their historical repertoire on their websites. Company data are collected by a web crawler that automatically visits the web pages of the four companies and parses data from there [22]. The collected data includes original ballet titles, which are listed in each company's repository and refer to ballet works with artistic elements that remain constant across time, performances, and productions (e.g., creators, libretto, music, genre). When appropriate, ballet companies marked revivals (recreated works), and/or company premieres (productions that originally debuted at a different ballet company but that are presented for the first time in the company listing the work), and we included them as a company's original work.

Collaboration networks are formed from the teams of leading artists working together to create a ballet work. Teams of ballet creators are formed from each record of original ballet titles, which includes the credits of leading artists, such as principal creators (choreographer and composer) and specialized roles (librettist, costumes and lighting designer), and excludes the dancers or any other company members. In a few occasions, companies report the producer, designer (unspecified), and media editor of a ballet work, and other team structures vary in size by adding multiple collaborators for the same role (e.g., two or more composers). It is important to note that ballet is strongly recognized for its conventional collaborative structure, comprising a core structure of leading artists, including choreographer, composer, librettist, and costumier and light designer. Hence, in constructing the collaboration network of ballet creators, we consider all listed artists in each ballet title as equal contributors to the ballet creation.

Therefore, a *ballet collaboration* is defined as the creative and collective efforts between leading artists listed by each company for the creation of a ballet work or title. For the conceptualization of network collaborations in performing arts, we follow [11], and one can find a more detailed explanation related to the artistic role in Section S1 of the Additional file 1. Specifically, we considered leading artists, including principal creators (choreographer and composer) and specialized roles (librettist, costumes, producer, media, and lighting designer), listed in Table 1. Regarding classical music composers, we include them in our analysis to determine the cumulative gendered characteristics in the ballet creations as a whole.

The processing of the data unfolds as follows: Fig. 1a illustrates the data showing a list of ballet titles (as an example, “Ballet 1” and “Ballet 2”) with the names of ballet creators (A, B, C, D, and E), and their roles (e.g., Choreography, Music, and Costumes). Then, all artists who collaborate in a ballet creation together are part of the same team. To construct the collaboration network of each company, we first build a bipartite network between ballet

Table 1 Artistic roles included in each company’s dataset

| | ABT | NYCB | NBC | ROH |
|----------------|---|---|---|--|
| Artistic roles | Design, Music, Choreography, Lights, Costumes, Libretto | Design, Choreography, Music, Lights, Costumes | Music, Choreography, Producer, Design, Costumes, Lights | Choreography, Music, Libretto, Producer, Costumes, Design, Lights, Media |

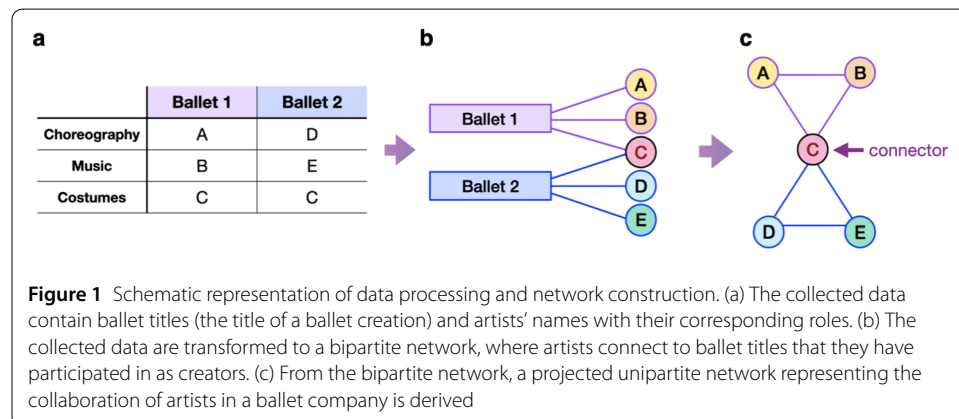


Table 2 Data description and basic network characteristics of the four ballet companies

| | ABT | NYCB | NBC | ROH |
|----------------------------------|--------------|--------------|--------------|--------------|
| Time window | 1940-2020 | 1935-2020 | 1951-2019 | 1931-2012 |
| No. of ballets | 491 | 560 | 349 | 449 |
| Artists (nodes) | 779 | 549 | 579 | 847 |
| Collaborations (links) | 2323 | 2202 | 1913 | 3107 |
| Average degree | 5.9641 | 8.0219 | 6.6079 | 7.3365 |
| Density | 0.0076 | 0.0146 | 0.0114 | 0.0086 |
| Clustering coefficient C_{avg} | 0.7556 | 0.7710 | 0.7776 | 0.8033 |
| Average path length L | 3.52 | 2.63 | 3.60 | 3.49 |
| Small-worldness, S | 97.73 | 67.57 | 75.07 | 102.99 |
| <i>Giant Component</i> | | | | |
| Artists (as percentage) | 715 (91.78) | 643 (98.90) | 540 (93.26) | 816 (96.34) |
| Collaborations (as percentage) | 2251 (96.90) | 2195 (99.68) | 1863 (97.38) | 3071 (98.84) |
| Density | 0.0088 | 0.0149 | 0.0128 | 0.0092 |
| Diameter | 8 | 5 | 10 | 8 |

creations and artists, as seen in Fig. 1b, where the left-hand-side nodes represent ballet titles and the right-hand-side nodes display the artists that created a ballet title. Next, artists' collaborations are projected to an undirected graph, as shown in Fig. 1c, where each node represents one artist, and a link between two artists denotes their collaboration in the same ballet creation. An artist who teams up in more than one ballet creation will connect multiple artists in the same company, hence becoming a connector in the collaboration network.

The resulting empirical networks include approximately 300–560 ballet works, with a range of 490–850 artists (nodes) and 1900–3100 collaborations (links). In addition, the time of reported ballet creations ranges the 1930s to the 2020s, making the networks comparable in terms of size and longevity. The basic network properties, such as size of the giant component, average clustering coefficient [23], average shortest path [24], and small-worldness [25], are presented in Table 2.

2.1.1 Gender inference

Artists' names were processed for misspelling, middle names, and initials to distinguish their identities. The names are held constant if reported across multiple companies. Thereafter, we infer artists' gender by using `gender` package for R [26, 27]. This package not only contains names from various countries and periods but also infers names from standardized databases (`ssa`, `ipums`, `napp`, and `demo`), making it adequate for this study, especially as the collected data comprise names of artists with diverse nationalities and are born in the 19th and 20th centuries.

To estimate an artist's birth year, we assumed that each artist was at least 20 years old when they participated in a ballet creation for the first time. Thus we subtract 20 years from the year of the first ballet production of an artist in our data as a proxy of the minimum age for a productive life in ballet. This method considers a range of 10 years (± 5 years from the estimated birth date). Then, the `gender` package estimates probability that a person would have a certain gender associated with the name. If the probability is larger than or equal to 0.7, the corresponding gender is assigned to each artist. Here, the assigned "gender" is a binary property (woman, man) and does not consider other gender assignments. Note that the inferred gender does not directly refer to the sex of the artist nor the self-assigned gender chosen by each artist but is used as an estimate of the

social construction of gender. The names, which were not able to assign gender with this method, were manually assigned after a web search of the artist's identity.

2.2 Network centralities

To understand the importance or centrality of artists in the collaboration networks, we measure four network metrics over the entire network:

1. Degree centrality is computed following [28] to measure the number of total connections of an individual (i.e., a node). This metric can capture the level of an individual's access to social capital.
2. Harmonic centrality is computed following [29], and it is a variant of closeness centrality created to deal with unconnected graphs to measure the distance one node has with respect to all other nodes in the network. In other words, harmonic centrality captures the position of nodes to reach distant parts of the network efficiently. The larger the value of the harmonic centrality of a node, the closer the node is to others.
3. Betweenness centrality is computed considering all pair of nodes, as described in [30], to measure the number of shortest paths between two pairs of nodes that pass through a node in a network. This metric captures the nodes that are the best intermediaries or bridges between different parts of the network.
4. Eigenvector centrality is computed following [31] and measures the importance of a node based on the centrality of its neighboring nodes. This centrality informs about the nodes that are connected to other influential or central nodes, as these can help gain social prestige in the network.

These metrics are informative on the differential ranking of individuals embedded in the network [32]. For example, an artist with a high degree centrality indicates that the artist has multiple collaboration connections in a network; thus, such an artist is well-positioned to have more access to information, social connections, and professional opportunities. Because of the range difference in centralities, the centralities are normalized and re-scaled in the range of [0, 1]. In a global sense, these centrality metrics help identify structural patterns within a network, providing insights into the underlying relationships between individuals that ultimately shape the network.

2.3 Definitions of top-central artists

From the four aforementioned centrality metrics, we sorted all artists by their centrality in descending order and selected artists ranked at the top 20 highest centrality, referred to as *Top-Central Artists* (TCA) in this study. We test the top 10 to top 40 artists for each centrality and confirm that the top 20 artists capture the largest variation of centrality values in the empirical networks, and between and within gender groups. Hence, we consider the top 20 artists to focus on the highly central individuals per an attributed group and centrality to capture the highly central artists in the ballet collaboration networks and their differences in network positions across gender categories. The top 20 artists for each gender group cover the "core" nodes of the top 3-5% most central individuals in the men's group and the top 10-20% most central women artists [33].

Let us consider the ranking of centralities $C(r)$, where C denotes a corresponding centrality value of an artist at a given rank r , for $r = \{1, 2, \dots, 19, 20 + \alpha\}$, so $r = 1$ represents the most central artist having the highest corresponding centrality (e.g., $C(1) = 0.8$), and

$r = 20$ will have the lowest centrality within the TCA (e.g., $C(20) = 0.01$). Here, α represents that there can be more artists over 20. This is possible when there are artists with the same centrality values. If there are five more artists with the same $C(20)$, then we include that all of them in the TCA group.

We apply the TCA group to three different attributes: the first group is for all artists in a company’s collaboration network, labeled as $TCA_{Network}$; the other two groups are for a company’s artists grouped by gender, which results in two separate rankings for TCA_{Women} and TCA_{Men} . Each gender subset is computed using the centrality metrics calculated over the whole network. Once the groups are formed, the centralities are normalized by the maximum value of the centrality within the company group (Network) and by company gender groups (Women, Men) to have centralities in the range of $[0, 1]$. This normalization ensures comparability across different centrality metrics and removes the variations by network size and gendered group size by each company.

Please note that the $TCA_{Network}$ selects the top $20 + \alpha$ artists as mentioned before, where tied centrality values receive the same rank. We only found ties in degree centrality for two companies, NBC (22 total artists) and ROH (21 total artists). By keeping the ties in the $TCA_{Network}$, we can examine the representation of women among highly centered TCA in a ballet company.

However, for TCA_{Women} and TCA_{Men} , the tied centrality is not considered. This means that the ranking process sorts each centrality by values, and in case of ties, it assigns the next consecutive rank to the value as found in the original list. With the implementation of this ranking, we maintain an equivalent number of women and men artists (i.e., 20 artists per gender group) and focus on fairly investigating the centrality differences between TCA_{Women} and TCA_{Men} .

From the $TCA_{Network}$, we quantify the women ratio R_{Women} by computing following Eq. (1).

$$R_{Women} = \frac{\sum_i^{N_{TCA}} \theta(i)}{N_{TCA}}. \tag{1}$$

Here, i denotes an index for an artist who is in a corresponding $TCA_{Network}$ N_{TCA} represents the total number of artists in a $TCA_{Network}$, for $\theta(i) = 1$ when an artist is a woman or 0 for men. Then, R_{Women} provides the fraction of women artists who belong to the group of well-positioned individuals in the collaboration network of a ballet company. A numerical fraction of women artists at the network level of 0.5 is assumed as a gender-balanced collaboration, and we call this situation a “neutral” composition. Note that the fraction of women artists in a team with an odd number of artists cannot be 0.5, but we consider the value in the range of $[0.45, 0.55]$ as the neutral composition in this study. The small differences in the N_{TCA} of NBC ($N_{TCA} = 22$) and ROH ($N_{TCA} = 21$) regarding degree centrality makes a difference in R_{Women} of approximately 0.002, which can be marginal in the present analysis.

The difference in centrality $\Delta C(r)$ between two rank-matched artists from each gender group is measured as $\Delta C(r) = C(r)_{Men} - C(r)_{Women}$. Here, each woman artist from TCA_{Women} is matched to her corresponding r -ranked men artist from TCA_{Men} . That is to say, if there is a woman artist ranked 1 in TCA_{Women} with a centrality value of 0.4, she is at the most central position in the women’s group, and it can be written as $C(1)_{Women} = 0.4$.

The counterpart of the man artist, who is ranked 1 in TCA_{Men} , will be $C(1)_{Men} = 0.5$, if he has a centrality of 0.5. Then, $\Delta C(1) = 0.5 - 0.4 = 0.1$. If $\Delta C(r) > 0$, it means that a man artist at the same rank in the gender group is located in a more central position than the woman counterpart.

2.4 Null model analysis

We compute two different null models by simulating 100 synthetic networks derived from the representation of each company's empirical collaboration network. With the help of the null models, we remove the collaborator- or gender-preferences by shuffling collaborations (links) or artists' attributes (gender) in the collaboration network. The overall purpose of the null models is to create a baseline of randomly created networks, which would allow us to determine the absence or existence of randomness in the observed patterns with respect to the empirical network.

- (1) Edge-shuffled model: In this model, edges are randomly rearranged in the network, while preserving artists' degrees. This means that the total number of collaborations per artists is preserved, as well as the total number of artists (nodes) in the network and artists' gender. We use the "random_reference" function of NetworkX [34], which is based on the analysis of [35]. This function generates random graphs by randomly swapping edges between nodes, while preserving the same degree for each node. From this randomization, we remove the gendered correlation from empirical collaboration networks. Therefore, the resulting synthetic networks show collaboration structures where there is no gender preference.
- (2) Gender-shuffled model: This model shuffles the gender of artists, while holding all network properties constant. Here, the empirical network structure is used as a reference, specifically without incorporating nodes' attributes. We randomized artists' genders within an entire collaboration network. This process randomizes the correlations among artists' gender, artistic roles, and productivity while ensuring that the real fraction of women and men in the entire network and the artist types are preserved. In this way, the artists' network position is preserved, but their gender and artist type are randomized in each iteration. Therefore, the resultant networks destroy preferences by artist type and display an artificial collaboration pattern without a correlation between an artist's gender and position, as well as a gendered collaboration assortativity.

To test a null hypothesis distribution, we compute the Z -score for a distinction between the centrality values from the empirical networks and those from the null models. We denote the observed centrality by rank in the real network as $C(r)_{real}$ and that of the null model as $C(r)_{null}$. Then, we determine the Z -score for both TCA_{Women} and TCA_{Men} using the centrality computed over the empirical network, $C(r)_{real}$, and the averaged centrality of 100 null models, $\bar{C}(r)_{null}$; thus the Z -score of a centrality corresponding an artist at rank r can be formulated as

$$Z(C) = \frac{C(r)_{real} - \bar{C}(r)_{null}}{\sigma(C(r)_{null})}. \quad (2)$$

Similarly, the Z -score of the difference in a centrality between artists at the same rank in each gender group ($\Delta C(r)$) can also be measured with the values of the synthetic networks

as

$$Z(\Delta C) = \frac{\Delta C(r)_{\text{real}} - \Delta \bar{C}(r)_{\text{null}}}{\sigma(\Delta C(r)_{\text{null}})}. \quad (3)$$

Applying $Z(\Delta C)$ to the same rank of gender groups can display a possible improvement (or reduction) of an artist's central position, depending on one's gender group.

2.5 Perception error on women artists

To further understand the implications of the gendered differences in the collaborative environment, we use a mathematical approach to measure the existence of perception errors based on [17]. Perception errors refer to the inaccuracy in the estimation of the frequency of an attribute—usually of a minority group—in a social network, as perceived from the frequency of that attribute within the individual local network [14, 16]. In this research, perception errors are the difference in the perceived fraction of women artists from the local network, respect to the fraction in the entire network. For instance, if there are mostly women in the local network, one individual will have a perception error above one that overestimates the size of the women's group, while the opposite happens for the underestimation of the women's group, with a value below 1. Thus, when the perception error is equal to 1, it implies that the perception of the fraction of women in the network is accurate. The perception error B of an individual artist i is thus computed as $B_i = \frac{W_i}{R_{\text{Women}}}$, where W_i denotes the local fraction of women among i 's collaborators, and R_{Women} refers to the real fraction of women in the network, as noted above.

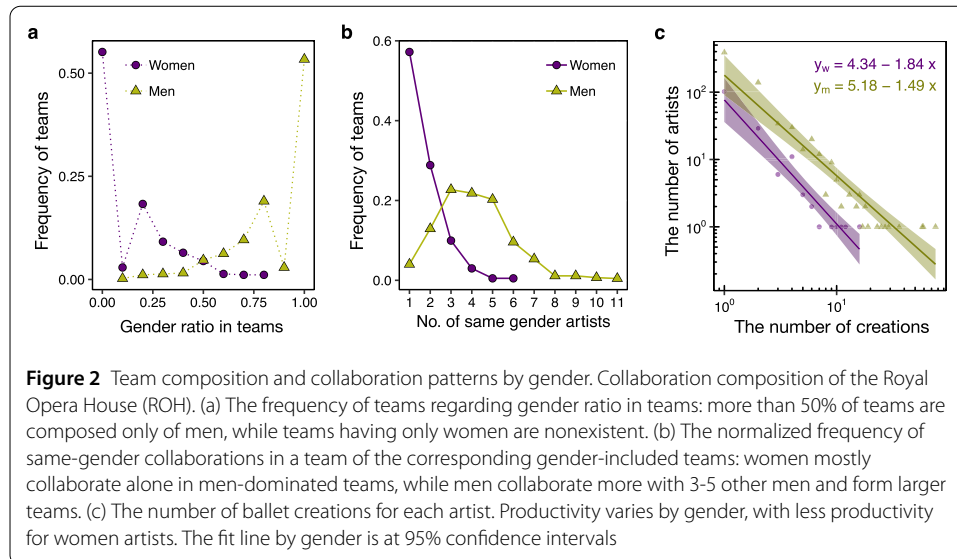
Based on the individual artist's perception error, we measure an averaged perception error by gender group at a network level, so $\bar{B}_{\text{Network}} = \frac{\sum_i B_i}{N_{\text{Network}}}$, where N_{Network} represents the total number of artists in a ballet company. For each gender group' perception error, $\bar{B}_{\text{Women}} = \frac{\sum_i B_i}{N_{\text{Women}}}$ and $\bar{B}_{\text{Men}} = \frac{\sum_i B_i}{N_{\text{Men}}}$ can be defined. Consequently, when $\bar{B} = 1$, it means that the overall perception of women on a company is accurate on average, and when $\bar{B} < 1$ ($\bar{B} > 1$), a group underestimates (overestimates) the ratio of women artists on average. In addition, a gendered homophily is measured following the method in [17] determine the gendered preferences of the collaboration networks.

3 Results

Based on previous reports on the lack of representation of women in leading positions in ballet [6], we explore the general composition of the collaboration networks of ballet creators and the existence of gendered collaboration patterns in the professional environment. We also look into the composition of the most central network positions and the gender gap between men's and women's centralities in the network; in addition, we measure the existence of perception errors of the women's artists group within ballet companies. We compare network position and perception errors from the empirical network structures with two null model analyses.

3.1 Team structure and collaboration patterns

The most common team size for a ballet creation across companies is three to four (20–40%), followed by five members (20%), as shown in Fig. S1a. This indicates that teams of ballet creators are mostly formed by the typical collaborative structure of leading artists. Figure 2a shows a sample of the representation of women in a ballet company (ROH), while



also revealing that there are approximately 50% of teams having 100% men artists, and less than 10% of teams have a gender-neutral ratio of 50%. Conversely, the majority of teams are composed with less than 50% of women artists, regardless of their sizes, and teams having 100% women artists is almost zero.

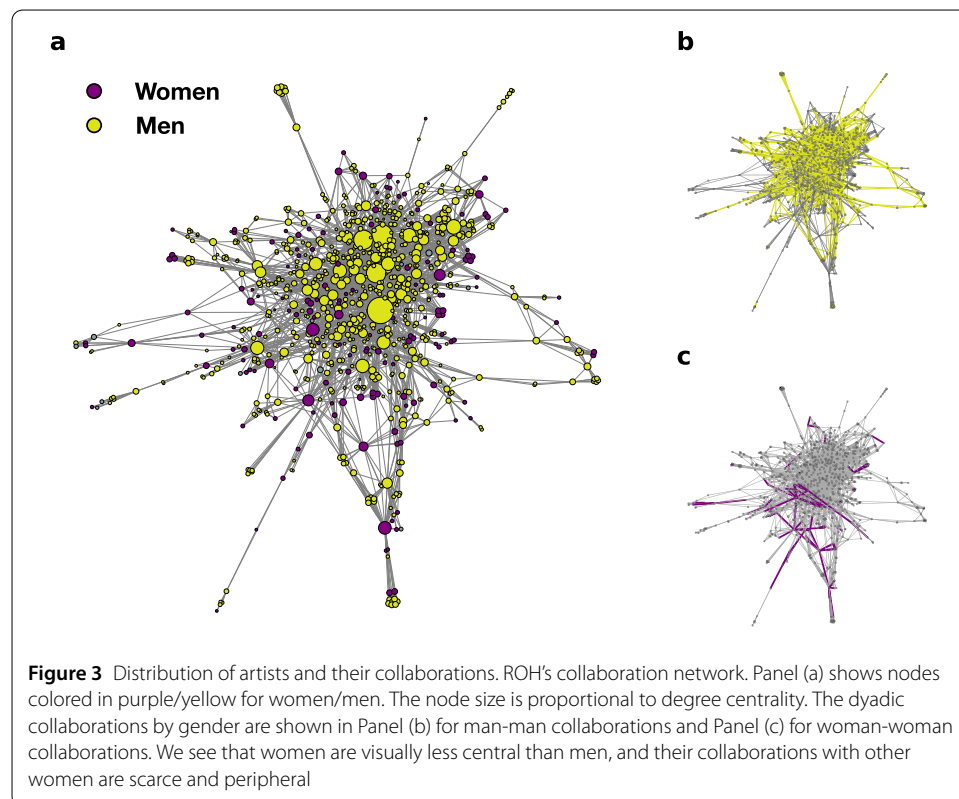
Dance communities have specifically reported an overlooking of women in choreographic leads, and our results suggest that women are less represented than men in general leading roles. Exploring the team composition by artistic role, the proportion of women is considerably low for the choreographer, librettist, and composer groups (Fig. S2). Other positions such as costume, lighting and design have a relatively larger participation of women, but still men are dominant in those roles as well.

Further, in Fig. 2b, we see that when women collaborate in a team, the frequency of working with other women in the same team is actually very low (< 30%). These results describe that women artists mostly work in men-dominated environments. On the other hand, men-alone teams are rather rare (< 10%), as they tend to collaborate with at least three to five other men (> 20%) and participate in considerably larger teams than women (up to 11 men in one team, at ROH).

In terms of productivity, women artists are less involved in ballet creations than men artists. In NYCB and ROH, the most productive woman participates in approximately 20–25% of the creations of the most productive man artist collaborated (ROH's maximum collaborations: Men = 76, Women = 16; NYCB's maximum collaborations: Men = 211, Women = 54, see Figs. 2c, and S3b). For NBC, the highest productivity is a bit similar for both genders. Women artists' highest productivity is just 86% of the most productive man (NBC's maximum collaborations: Men = 38, Women = 33). Only at the ABT, the most productive woman artist exceeded in 20 collaborations relative to the most productive man artist (ABT's maximum collaborations: Men = 35, Women = 55). Notwithstanding the exception, most women artists are less productive than their men counterparts, and the global picture for women is to work in men-dominated creative environments. Team structures, and collaboration and productivity patterns, are similar across all companies studied here (for more details and figures by company, see Section S1).

Table 3 Network composition by gender. Here, R_{Women} is the fraction of women in the entire collaboration network. For the collaborations, the number of woman-woman/man-man/mixed dyadic interactions is counted

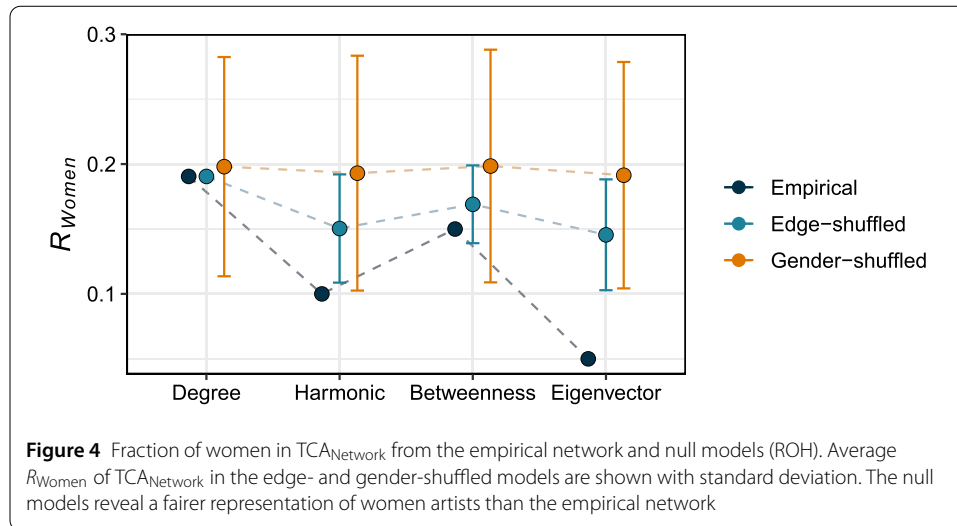
| | ABT | NYCB | NBC | ROH |
|-------------------------------|------------|------------|------------|------------|
| <i>Artists (nodes)</i> | 779 | 549 | 579 | 847 |
| R_{Women} | 0.22 | 0.19 | 0.18 | 0.19 |
| <i>Collaborations (links)</i> | 2589 | 2317 | 1956 | 3385 |
| Woman-woman | 138 (5%) | 64 (3%) | 55 (3%) | 135 (4%) |
| Man-man | 1603 (62%) | 1576 (68%) | 1336 (68%) | 2353 (70%) |
| Mixed | 848 (33%) | 667 (29%) | 656 (29%) | 897 (26%) |



3.2 Centrality differences by gender

Thus far, we have observed a less frequent participation of women with respect to men in ballet collaborations. These observations raise the question: Does the low representation of women relate to their small ratio in the company? To answer this question, we explore the distribution of artists' collaborations in the network. We first compute the fraction of women in the network, R_{Women} , and the proportion of dyadic interactions (see Table 3), showing that most companies only have approximately 20% of women in leading positions.

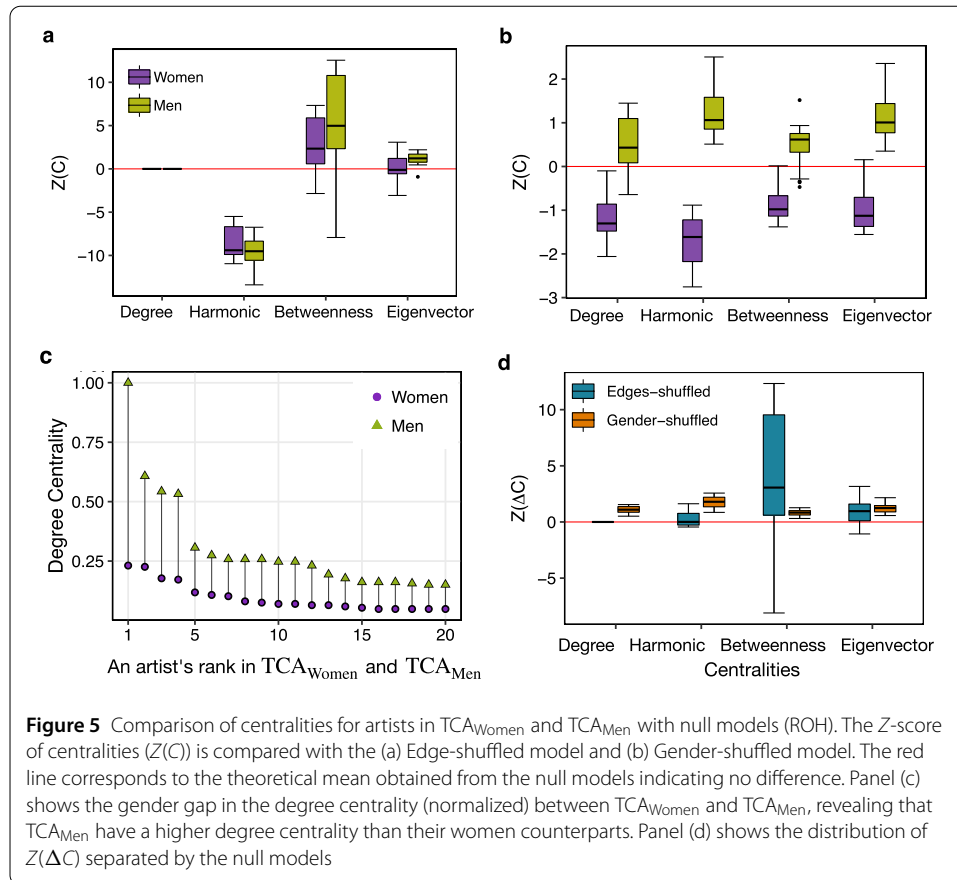
Figure 3a shows a network sample, where men (in yellow) are not only a majority but also with higher connectivity relative to women (in purple) (See all companies' collaboration networks in S4). Moreover, on the one hand, the man-man connections are more than 60% across companies (yellow links, Fig. 3b) and mixed connections are about 30% on average. In contrast, woman-woman connections are less than 5% of the total dyadic interactions (purple links, Fig. 3c). These results inform that, for every four men, there is



only one woman in the network, a collaborative structure in which men artists are densely co-worked with other artists regardless of gender, locating at the center of the collaboration network, while women artists are sparsely distributed in the periphery of the network. The central representation of men artists is also statistically significant for degree, betweenness, harmonic, and eigenvector centrality (Two Sample T-test, see Additional file 1 Table S1).

We then evaluate the proportion of women in the TCA group, $TCA_{Network}$, to focus on the highly central artists in a company by sorting their network centralities, and observe that most companies have a lower central representation of women with respect to R_{Women} in the empirical network. When we compare this ratio with null models, we observe an overall increase of R_{Women} in the randomized models for all centralities (see all companies in Fig. S5). For example, the edge-shuffled model improves R_{Women} in $TCA_{Network}$ for harmonic centrality from 10% to 15%, and gender-shuffled model raises it up to 19% in the ROH (Fig. 4). Note that the edge-shuffled model keeps R_{Women} in $TCA_{Network}$ regarding degree centrality because the number of collaborations (degree) for an artist and their inferred gender are held constant in this model. These results suggest that the low representation of women artists in ballet creations could be related to gender assortative collaborations, and the current level of women artists' centrality is not a deterministic outcome of the small fraction of women artists in the company. Put differently, even when the fraction of women remains small in a network, women artists' representation could be improved if more equal collaborations and diverse artistic roles for women were encouraged.

The $Z(C)$ reveals a general change in artists' centrality with the null models (see ROH's sample in Fig. 5, all companies in Fig. S6a). For the edge-shuffled model, only the harmonic centrality displays a negative Z-score for both women and men. Harmonic centrality denotes an extent of an artist's closeness to other artists on average, so a small value represents a far distance between artists. The negative Z-score suggests that the distance among artists in empirical collaboration networks falls apart farther than the expected distance from the null models. In other words, TCA in the empirical networks is more concentrated themselves, separating other artists than the expected distances in the null models.



For the gender-shuffled model, the negative women artists' Z-scores for all four centralities indicate that their positional importance can be improved in a synthetic network with collaboration imbalances (see Fig. 5b). Altogether, our results suggest that differences in centrality among TCA may not be derived by random factors, but there may be underlying systematic social behaviors limiting women artists' collaborations and network position, regardless of their small fraction in the network.

The difference in degree centrality (ΔC) highlights that a man artist locates at more central position than the same-ranked women artist in her TCA group. Figure 5c shows a sample for degree centrality and reveals that the most central man is considerably more central than the most central woman. This ΔC trend is observed across centralities with slight variations, thereby confirming that men are considerably better positioned with respect to women across companies (see gender gap in centrality for all companies in Fig. S7). Note that all empirical Z-scores for ΔC are several standard deviations away compared to the null models (see all companies in Fig. S6b). Figure 5c illustrates the variations by null model, and showcases that a large gender gap is less likely observed when the gender preference (edge-shuffled) and gendered productivity and artistic roles correlations (gender-shuffled) are destroyed.

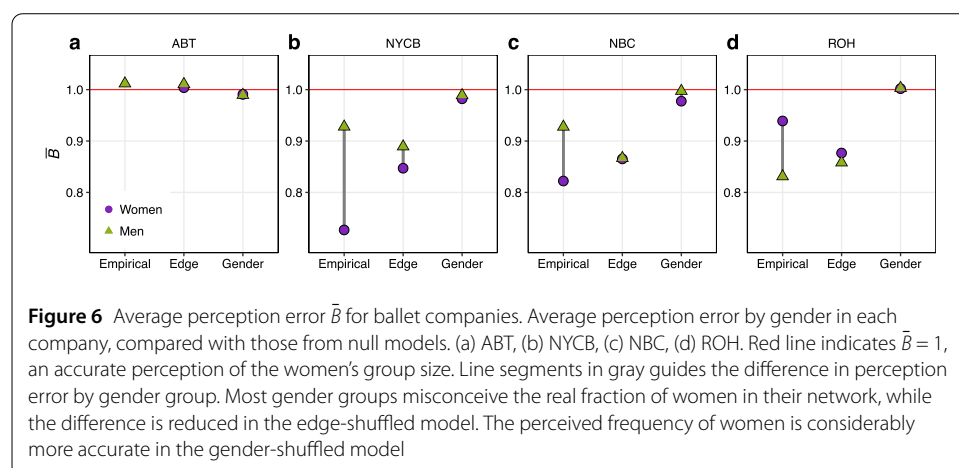
3.3 Perception error on women artists

Given the observed structural imbalances in ballet collaboration networks, the low participation of women in professional collaborations could affect the perceived frequency

on women artists in the entire network. Perception errors are the distorted frequency estimations of an attribute in a social network by the individual local environment [14, 16]. Here, perception error is defined as the fraction of the observed frequency of women in an artist's local collaboration network over the real fraction of women in the global network (see Methods). This implies that perception error denotes a relative difference of women artists in the local collaboration environment of each artist and the actual women artists' frequency in each ballet company. From the individual-level perception error B_i , a gender group-level error \bar{B} compares the average perception error for women and men. If $\bar{B} > 1.0$ ($\bar{B} < 1.0$), it means a gender group overestimates (underestimates) the global frequency of women artist. When $\bar{B} = 1.0$, it denotes an accurate perception on the women frequency (see Methods). We complement perception error with a measure of homophily.

Our results show that for the empirical collaboration networks of the ABT, NYCB, NBC, and ROH, the women (men) artists' homophily values are 0.56 (0.53), 0.47 (0.55), 0.45 (0.57), and 0.56 (0.63), respectively (1 is a perfect homophily, and 0 is a perfect heterophily situation). The ABT has a relatively gender-mixed environment, resulting in both gender groups having a relatively accurate perception on the global fraction of women artists, as shown in Fig. 6a. Conversely, the rest of the companies demonstrate a wide difference in perception error by gender, as shown in Fig. 6b–d. For instance, the NYCB's men group underestimates women artists by approximately 7%, but their women group underestimates themselves by approximately 27%, showing a 20% difference in the perception of women between the two groups. Such a difference may be related with men artists' strong homophily in NYCB collaborations and women artists' gender-heterophilic collaborations (woman-man heterophily 0.53 > woman-woman homophily 0.47), indicating a perceived underestimation of women artists by themselves. In ROH, women artists have a more accurate estimation of women artists with respect to men artists, which aligns with their collaborative behaviors, where women artists collaborate more with other women artists than other men artists (woman-man heterophily 0.44 < woman-woman homophily 0.56). Yet, the difference in perception still exists, especially as men artists collaborate mostly with men artists (man-man homophily 0.63, man-woman heterophily 0.37), and the assortative collaboration widens the difference in perception between gender groups.

To investigate the significance of perception errors, we conducted a series of mean comparisons. With the aggregated data, we conducted a two-way ANOVA with gender and the



null models as factors. We observed a statistical effect of the network model on the perception error ($F(2) = 86.212, p < 0.001^{***}$), but no effect by gender ($F(1) = 0.038, p = 0.843$) or the interaction between gender and network model ($F(2) = 0.36, p = 0.697$). In more detail, the edge-shuffled model displays a reduction in perception error difference between women and men, even though the reduction is limited. The reduction suggests the correlation of gender assortative collaboration structures in ballet creations. Moreover, the gender-shuffled model not only sensibly reduces the difference in the average perception error for women and men but also achieves a nearly accurate perception on the fraction of women artists. This strongly suggests that lowering an extent of imbalanced productivity and gendered preferences altogether boosts the representation of women artists, even considering the small representation of women artists in the company.

4 Discussion

Gendered inequalities have been investigated for different occupations with numerical differences of labor force compositions and salary [36]. To expand the investigation from the numerical imbalance, the current study investigates gendered collaboration structures and their correlations with gender imbalances in ballet creations. The results demonstrate that fewer number of women artists are positioned at the top central artists, and gendered collaboration patterns (edge-shuffled model) and gendered productivity correlations (gender-shuffled model) can aggravate their visibility in terms of centralities and perceived frequency.

Many studies have reported the crucial roles of an individual's social network are associated with access to information and professional success in creative collaborations [10–12, 37–41]. Men and women utilize different social network structures and behavioral patterns that influence their placement in the job market [8, 42, 43], and the formation of a personal network and social behaviors over time are related to reinforced perception errors [44, 45]. The existence of feedback among social relationships, perception errors, and collaboration patterns can consequently influence individual career decisions. For women in ballet, a feedback based on an actual and perceived low representation within men-dominated collaborations can negatively impact their decision to undertake a career as ballet creators or engage in multiple collaborative projects. A future study of the interplay of those elements could provide more insights into the career decisions of women in ballet in the long term.

Practically, the diverse collaboration structure can be crucial for teams [46] and individual performance [47] in terms of creativity and success [11, 48]. A study demonstrates that diversity can improve creative performance [49] and emphasizes the importance of women's participation in collaborative environments because they increase the social sensitivity of the group, making the team collectively more intelligent and proficient [50]. Accordingly, new policies for more equal collaborations and a more inclusive environment for women as leading creators can be considered in the creative industries.

Regardless of the many insights the current results can provide, the current measure of perception errors is a mathematical approach and can be improved, as multiple factors influence the perception of a local network structure. That is to say, a local network can be described not only by its structure, but also by its embedded social mechanisms, like the strength of relationships formed over time, access to information, formal and informal norms [9, 51, 52], and individual cognitive processes and preferences [53, 54]. In addition,

ballet is strongly influenced by biological constraints, such as the physical demands of the art form, including strength, flexibility, and technical requirements. These constraints, combined with the distribution of labor in family responsibilities, may be stronger for women and may contribute to fewer women overcoming social barriers in the workplace and hinder the professional development of women artists in ballet [36]. Analysis of the gendered imbalance with artists' life cycles would open another perspective to a better understanding of the formation of creative collaborations.

Also, the data analyzed in this study depends on the archival of the selected ballet companies, which may not be sufficient to generalize the current results to the entire ballet industry. Moreover, artists may hold different types and duration of contracts within a company, which can result in variations in observed professional collaborations. To overcome this, more comprehensive digitized data collection would be needed. The improved collection of data can be implemented with deep learning and network science. Since it has been possible to objectively measure career success [55] and the impact of individual performance in creative domains [56, 57], applications of those methods can open the possibility for future research on the relationship between gender, network centrality, and actual ballet creators' impact in the field.

To sum up, this study highlights the low representation of women as ballet creators and sheds light on their peripheral network position and gendered collaboration preferences within the ballet industry. This investigation can be extended to explore the dynamic network factors shaping gender imbalances to propose possible and more adequate interventions for diversity, equity, and inclusion in cultural organizations. We hope that this work brings awareness to how social phenomena and inequalities in creative domains can be systematically studied with network science and data-driven methods.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-023-00428-z>.

Additional file 1. (PDF 1.3 MB)

Acknowledgements

We thank Felipe Salgado for assisting with data collection. Y.H.-G. acknowledges the Centro de Investigación en Complejidad at Universidad del Desarrollo, Chile, for the financial support to conduct this research. Y.H.-G. also acknowledges Elizabeth Yntema and Rebecca Ferrell from the Dance Data Project for the fruitful discussions about this work.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2022R1C1C1005856), the National Agency of Investigation and Development, ANID, through the grant FONDECYT No. 11190096, the KENTECH Research Grant (KRG 2021-01-003), and the Pukyong National University Research Fund in 2022(202203530001).

Abbreviations

ABT, American Ballet Theatre; NBC, National Ballet of Canada; NYCB, New York City Ballet; ROH, Royal Ballet of the Royal Opera House; TCA, Top-Central Artists.

Availability of data and materials

The datasets generated and analysed during the current study are available in the web repository, <https://doi.org/10.5281/zenodo.8374171>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author contributions

All authors contributed to the research design and writing of the paper. YH-G contributed with art-specific knowledge, constructed the data and networks, developed and performed the models, analyzed the data, and performed data visualizations; EL was mainly responsible for the measurement of perception errors and homophily; and HK contributed to data construction and network analysis. EL and HK supervised the research. All authors discussed the results and contributed to writing the manuscript. All authors read and approved the final manuscript.

Author details

¹Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

²Department of Scientific Computing, Pukyong National University, Busan, Republic of Korea. ³Department of Energy Engineering, Korea Institute of Energy Technology, Naju, Republic of Korea. ⁴Instituto de Data Science, Facultad de Ingeniería, Universidad del Desarrollo, Santiago, Chile.

Received: 30 June 2022 Accepted: 13 November 2023 Published online: 23 November 2023

References

1. Daly A (1987) The balanchine woman: of hummingbirds and channel swimmers. *Drama Rev* 31(1):8–21. <https://doi.org/10.2307/1145763>
2. Homans J (2010) *Apollo's angels: a history of ballet*, 1st edn. Random House, New York
3. DeFrank-Cole L, Nicholson RK (2016) The slow-changing face of leadership in ballet: an interdisciplinary approach to analysing women's roles. *Leadersh Humanit* 4(2):73–91. <https://doi.org/10.4337/lath.2016.02.01>
4. Elssesser K (2019) A gender gap in ballet, seriously? <https://www.forbes.com/sites/kimelssesser/2019/09/12/a-gender-gap-in-ballet-seriously/>
5. Yntema E (2019) The ballet world is still male-dominated, research shows. <https://www.womensmediacenter.com/news-features/the-ballet-world-is-still-male-dominated-research-shows>
6. Dance Data Project (2015). www.dancedataport.com
7. United States Census Bureau (2021). <https://www.census.gov/data.html>
8. Yang Y, Chawla NV, Uzzi B (2019) A network's gender composition and communication pattern predict women's leadership success. *Proc Natl Acad Sci* 116(6):2033–2038. <https://doi.org/10.1073/pnas.1721438116>
9. Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
10. Pan RK, Saramäki J (2012) The strength of strong ties in scientific collaboration networks. *Europhys Lett* 97(1):18007. <https://doi.org/10.1209/0295-5075/97/18007>
11. Uzzi B, Spiro J (2005) Collaboration and creativity: the small world problem. *Am J Sociol* 111(2):447–504. <https://doi.org/10.1086/432782>. Accessed 2019-04-02
12. Abbasi A, Chung KSK, Hossain L (2012) Egocentric analysis of co-authorship network structure, position and performance. *Inf Process Manag* 48(4):671–679. <https://doi.org/10.1016/j.ipm.2011.09.001>
13. Juhász S, Tóth G, Lengyel B (2020) Brokering the core and the periphery: creative success and collaboration networks in the film industry. *PLoS ONE* 15(2):0229436. <https://doi.org/10.1371/journal.pone.0229436>
14. Karimi F, Génois M, Wagner C, Singer P, Strohmaier M (2018) Homophily influences ranking of minorities in social networks. *Sci Rep* 8(1):1–12. <https://doi.org/10.1038/s41598-018-29405-7>
15. Festinger L (1954) A theory of social comparison processes. *Hum Relat* 7(2):117–140. <https://doi.org/10.1177/001872675400700202>
16. Lerman K, Yan X, Wu X-Z (2016) The 'majority illusion' in social networks. *PLoS ONE* 11(2):0147617. <https://doi.org/10.1371/journal.pone.0147617>
17. 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. *Nat Hum Behav* 3(10):1078–1087. <https://doi.org/10.1038/s41562-019-0677-4>
18. American Ballet Theatre. Ballet Archive. <https://www.abt.org/explore/learn/repertory-archive/ballets/>
19. New York City Ballet. The Repertory. <https://www.nycballet.com/discover/ballet-repertory>
20. The National Ballet of Canada Archives. Repertoire List. <https://national.ballet.ca/Tickets/Archives/Repertoire-List>
21. Royal Opera House Collections. <http://rohcollections.org.uk>
22. Van der Aalst WM, Bichler M, Heinzl A (2018) *Robotic process automation*. Springer, Berlin. <https://doi.org/10.1007/s12599-018-0542-4>
23. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442. <https://doi.org/10.1038/30918>
24. Chen F, Chen Z, Wang X, Yuan Z (2008) The average path length of scale free networks. *Commun Nonlinear Sci Numer Simul* 13(7):1405–1410. <https://doi.org/10.1016/j.cnsns.2006.12.003>
25. Humphries MD, Gurney K (2008) Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PLoS ONE* 3(4):0002051. <https://doi.org/10.1371/journal.pone.0002051>
26. Karimi F, Wagner C, Lemmerich F, Jadidi M, Strohmaier M (2016) Inferring gender from names on the web: a comparative evaluation of gender detection methods. In: *Proceedings of the 25th international conference companion on World Wide Web*, pp 53–54. <https://doi.org/10.1145/2872518.2889385>
27. Blevins C, Mullen L (2015) Jane, john... leslie? A historical method for algorithmic gender prediction. *DHQ: Digital Humanities Quarterly* 9(3). R package version 0.6.0
28. Freeman LC (1978) Centrality in social networks: I. conceptual clarification. *Soc Netw* 1(3):215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
29. Boldi P, Vigna S (2014) Axioms for centrality. *Internet Math* 10(3–4):222–262. <https://doi.org/10.1080/15427951.2013.865686>
30. Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41. <https://doi.org/10.2307/3033543>
31. Bonacich P (2007) Some unique properties of eigenvector centrality. *Soc Netw* 29(4):555–564. <https://doi.org/10.1016/j.socnet.2007.04.002>
32. Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Structural analysis in the social sciences. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511815478>

33. Hajibabaei A, Schiffauerova A, Ebadi A (2023) Women and key positions in scientific collaboration networks: analyzing central scientists' profiles in the artificial intelligence ecosystem through a gender lens. *Scientometrics* 128(2):1219–1240. <https://doi.org/10.1007/s11192-022-04601-5>
34. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296(5569):910–913. <https://doi.org/10.1126/science.1065103>
35. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296:910–913. <https://doi.org/10.1126/science.1065103>
36. Becker GS (1985) Human capital, effort, and the sexual division of labor. *J Labor Econ* 3(1):33–58
37. Luce RD (1950) Connectivity and generalized cliques in sociometric group structure. *Psychometrika* 15(2):169–190. <https://doi.org/10.1007/BF02289199>
38. Jamali M, Abolhassani H (2006) Different aspects of social network analysis. In: 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings) (WI'06). IEEE, Los Alamitos, pp 66–72. <https://doi.org/10.1109/WI.2006.61>
39. Parish AJ, Boyack KW, Ioannidis JP (2018) Dynamics of co-authorship and productivity across different fields of scientific research. *PLoS ONE* 13(1):0189742. <https://doi.org/10.1371/journal.pone.0189742>
40. Fraiberger SP, Sinatra R, Resch M, Riedl C, Barabási A-L (2018) Quantifying reputation and success in art. *Science* 362(6416):825–829. <https://doi.org/10.1126/science.aau7224>. Accessed 2019-01-28
41. Janosov M, Battiston F, Sinatra R (2020) Success and luck in creative careers. *EPJ Data Sci* 9(1):9. <https://doi.org/10.1140/epjds/s13688-020-00227-w>
42. Vasarhelyi O, Vedres B (2021) Gender typicality of behavior predicts success on creative platforms. arXiv preprint [arXiv:2103.01093](https://arxiv.org/abs/2103.01093). <https://doi.org/10.48550/arXiv.2103.01093>
43. Tata J, Prasad S (2008) Social capital, collaborative exchange and microenterprise performance: the role of gender. *Int J Entrepreneurship Small Bus* 5(3–4):373–388. <https://doi.org/10.1504/IJESB.2008.01731>
44. Jackson MO (2014) Networks in the understanding of economic behaviors. *J Econ Perspect* 28(4):3–22. <https://doi.org/10.1257/jep.28.4.3>
45. Ertan G, Siciliano MD, Yenigün D (2019) Perception accuracy, biases and path dependency in longitudinal social networks. *PLoS ONE* 14(6):0218607. <https://doi.org/10.1371/journal.pone.0218607>
46. Baugh SG, Graen GB (1997) Effects of team gender and racial composition on perceptions of team performance in cross-functional teams. *Group Organ Manage* 22(3):366–383. <https://doi.org/10.1177/1059601197223004>
47. Karakowsky L, McBey K, Chuang Y-T (2004) Perceptions of team performance: the impact of group composition and task-based cues. *J Manag Psychol*. <https://doi.org/10.1108/02683940410543597>
48. Yang Y, Tian TY, Woodruff TK, Jones BF, Uzzi B (2022) Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proc Natl Acad Sci* 119(36):2200841119. <https://doi.org/10.1073/pnas.2200841119>
49. Hamilton BH, Nickerson JA, Owan H (2003) Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. *J Polit Econ* 111(3):465–497
50. Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *Science* 330(6004):686–688. <https://doi.org/10.1126/science.1193147>
51. Granovetter M (1985) Economic action and social structure: the problem of embeddedness. *Am J Sociol* 91(3):481–510
52. Coleman JS (1994) *Foundations of social theory*. Harvard University Press, Cambridge. <https://doi.org/10.2307/2579680>. <https://www.jstor.org/stable/2579680>
53. Lazarsfeld PF, Merton RK (1954) Friendship as a social process: a substantive and methodological analysis. *Freedom Control Mod Soc* 18(1):18–66
54. Pachur T, Hertwig R, Rieskamp J (2013) Intuitive judgments of social statistics: how exhaustive does sampling need to be? *J Exp Soc Psychol* 49(6):1059–1077. <https://doi.org/10.1016/j.jesp.2013.07.004>
55. Herrera-Guzmán Y, Gates A, Candia C, Barabási A-L (2023) Quantifying hierarchy and prestige in us ballet academies as social predictors of career success. SocArXiv preprint RePEc:osf:socar:x9zwn. <https://doi.org/10.31219/osf.io/x9zwn>
56. Liu L, Wang Y, Sinatra R, Giles CL, Song C, Wang D (2018) Hot streaks in artistic, cultural, and scientific careers. *Nature* 559(7714):396–399. <https://doi.org/10.1038/s41586-018-0315-8>
57. Liu L, Dehmamy N, Chown J, Giles CL, Wang D (2021) Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nat Commun* 12(1):1–10. <https://doi.org/10.1038/s41467-021-25477-8>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.