

Study of gender bias in the Spanish film industry

Author: Maria Salceda Servitje, msalcese11@alumnes.ub.edu
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Alessio Cardillo, alessio.cardillo@ub.edu

Abstract: Gender bias affects several (if not all) aspects of our lives, including the film industry. This work examines gender bias in the Spanish film industry before and after the promulgation of the 2007 gender equality law. Using the framework of complexity science, collaboration networks are built from the Sensacine database. Subsequently, their structural properties are measured and compared with those of networks obtained using a suitable null model. The empirical networks exhibit statistically significant gender homophily in both periods, with the latter decreasing significantly after 2007. Nonetheless, homophily still persists after 2007, indicating that gender-neutral collaborations had not been achieved by the end of the studied period. Therefore, we conclude that the 2007 law had a tangible but limited structural impact on gender equity in the film industry.

Keywords: statistical physics, computational physics, complex systems, network science, social networks, gender gap.

SDGs: 5 (Gender equality), 8 (Decent work and economic growth), 10 (Reduced inequalities).

I. INTRODUCTION

Gender inequality manifests globally across social, economic, and professional spheres. The film industry is no exception, and empirical evidence consistently shows that women remain systematically underrepresented [1]. In the Spanish context, this dynamic has been empirically documented for decades. Structural barriers appear to operate not only at the point of entry into the industry, but to accumulate throughout women’s entire professional careers. In response to this issue, the Spanish government introduced a specific legislative intervention. The “*Ley Orgánica 3/2007, 22 de marzo, para la Igualdad efectiva de mujeres y hombres*” established the obligation for all public entities to promote equitable representations across social, cultural, and professional domains. Building on this framework, the “*Ley 55/2007, del cine*” introduced specific funding mechanisms designed to address gender imbalance [2, 3]. However, what remains unclear is whether these laws modified the structure of the industry or whether their impact was simply limited to shifts in participation rates. Assessing these effects requires going *beyond* conventional gender gap measurements. The relevant question is therefore not only how many women participate, but what is the pattern of their *interactions* within the film industry. Statistical physics and complexity science can provide a suitable framework to tackle these questions.

Complex systems consist of a large number of interacting components whose collective behavior cannot be explained from the properties of their individual components alone [4]. The film industry can be thought of as a complex system, where each production connects a diverse set of professionals who, in turn, collaborate across multiple projects, giving rise to an intricate network of interactions. This type of network has been shown to capture the relational structure of professional communities, and network science methods have been successfully

applied to the study of gender bias [5]. Network science thus constitutes the framework adopted in this work to explore how gender bias manifests within the Spanish film industry.

To this end, collaboration networks were constructed from the Sensacine database for two periods, seven years before and after the promulgation of the 2007 law. They were further split into subnetworks based on the gender composition of their links. To characterize the structural properties of these networks, standard metrics (*e.g.*, the number of triangles) are computed alongside gender-specific measures such as the gender assortativity coefficient. These values are then compared against a randomized null model in which gender labels are permuted while preserving the global network topology. This comparison allows us to determine whether the observed patterns are driven by gender, rather than by structural constraints of either the network or the data alone. We find that assortativity decreases significantly after 2007.

II. METHODS

A. Network Measures

One way to represent and study complex systems is to map them as *networks*, where N nodes represent the individual components and L edges (or links) the interactions between them. The structure of a network can be encoded in its *adjacency matrix* $\mathcal{A}$, an $N \times N$ matrix where $a_{ij} = 1$ if nodes i and j are connected and 0 otherwise [6]. A network is **undirected** if edges have no orientation (*i.e.*, $a_{ij} = a_{ji}$), and **unweighted** if all nonzero elements are equivalent (*i.e.*, $a_{ij} \in \{0, 1\}$). In this work, all networks considered are undirected and unweighted. In the following, we provide several indicators used to characterize the networks analyzed.

1. Structural features

The set of nodes directly connected to node i is its **neighborhood**, $\mathcal{N}(i)$. The **degree** of a node i , k_i , is the number of its neighbors and can be obtained from the adjacency matrix as:

$$k_i = \sum_{j=1}^N a_{ij}. \quad (1)$$

The **average degree** of the nodes, $\langle k \rangle$, is:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2L}{N}. \quad (2)$$

Another quantity of interest is $k_{\max}$, corresponding to the **maximum degree** in the network. At a global level, the **edge density** is defined as the fraction of existing edges over the maximum possible amount, $\binom{N}{2}$:

$$D = \frac{2L}{N(N-1)}. \quad (3)$$

The values of D span between 0 and 1, where $D = 1$ indicates a *complete* network (*i.e.*, where each node is connected to every other node) and $D = 0$ indicates an *empty* network (*i.e.*, where only nodes exist without any connections between them). Real-world networks are typically *sparse*, with $D \ll 1$ [7]. The **clustering coefficient**, C_i , measures the fraction of pairs of neighbors of node i that are also connected [6]. Equivalently, C_i can be defined as the number of triangles in which node i participates (t_i), normalized by the maximum possible number of these triangles:

$$C_i = \frac{2t_i}{k_i(k_i - 1)}. \quad (4)$$

Moreover, the **average clustering coefficient**, $\langle C \rangle$, provides a global characterization of the network [8]:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i. \quad (5)$$

Finally, a **connected component** is a maximal subset of nodes with at least one *path* between every pair of them (*i.e.*, a sequence of edges that can be followed to go from one origin node to a destination one). The largest connected component of the graph is the so-called **giant component** (GC), and its size compared to the whole network is denoted as $S = \frac{|N_{GC}|}{N}$ [6].

2. Feature-based measures

In addition to the structural metrics described above, two gender-based measures are introduced to characterize the gender mixing patterns of the network. As noted

by Newman and other scholars, individuals tend to associate with others who share similar characteristics, a phenomenon known as *homophily* or *assortative mixing* [9]. To quantify the extent of homophily, we define e_{ij} as the fraction of edges connecting nodes of type i to nodes of type j (with $e_{ij} = e_{ji}$ in an undirected network) and a_i as the fraction of edges attached to nodes of type i . Hence:

$$\sum_{ij} e_{ij} = 1, \quad \text{and} \quad \sum_j e_{ij} = a_i. \quad (6)$$

All the e_{ij} values define the $\mathcal{E}$ matrix. Considering gender as one possible node feature, the **gender assortativity coefficient**, r_G , is then defined as:

$$r_G = \frac{\text{Tr } \mathcal{E} - \|\mathcal{E}^2\|}{1 - \|\mathcal{E}^2\|}, \quad (7)$$

where $\|\cdot\|$ denotes the sum of all elements of a matrix, and $\text{Tr}(\cdot)$ is its trace. The value of r_G ranges from $r_G = -1$, indicating perfect disassortativity, to $r_G = 1$, indicating perfect assortativity, with $r_G = 0$ corresponding to random mixing. Equation (7) coincides with the Pearson correlation coefficient between the gender attributes of adjacent nodes, allowing r_G to be interpreted as a measure of linear correlation between the genders of connected individuals [9]. To complement the macroscopic perspective offered by r_G , we also introduce a node-level measure. Let $\mathcal{V}_x = \{i : \text{gender}(i) = x\}$ denote the set of nodes of gender $x \in \{M, F, U\}$ in the network, with $N_x = |\mathcal{V}_x|$ and $\sum_x N_x = N$. The **neighbor gender fraction** of a node $i \in \mathcal{V}_x$ with respect to gender y , $f_{i,y}$, is:

$$f_{i,y} = \frac{k_{i,y}}{k_i}, \quad (8)$$

where $k_{i,y} = |\{j \in \mathcal{N}(i) : \text{gender}(j) = y\}|$ is the number of neighbors of node i with gender y . The **average neighbor gender fraction**, $\langle f_{x,y} \rangle$, is defined as:

$$\langle f_{x,y} \rangle = \frac{1}{N_x} \sum_{i \in \mathcal{V}_x} f_{i,y}. \quad (9)$$

Together, r_G and $\langle f_{x,y} \rangle$ provide complementary views of gender mixing. On the one hand, r_G captures the overall tendency of nodes to connect within or across gender groups. On the other hand, $\langle f_{x,y} \rangle$ quantifies the average gender composition of each node's neighborhood.

B. Data

The data used in this project have been extracted from the Sensacine website [10]. Two datasets extracted from this source were used: one containing records of the films produced in Spain between 1905 and 2020 and the people involved in their making, and another with meta-information (*e.g.*, date of birth, nationality, etc.) on

TABLE I: Basic features of the dataset for the 2000–2007 and 2008–2014 periods: number of movies, average number of individuals per movie, and number of individuals by gender.

Metric	2000–2007	2008–2014
No. of films	465	489
Avg. individuals per film	4.8	5.8
No. of females	246	370
No. of males	1042	1302
No. of unknown	33	48

those individuals. People participating in acting roles were excluded from our data, as they follow different collaboration dynamics than technical crew members. To focus on the effects of the 2007 law, only films released between 2000 and 2014 have been considered.

Since the second dataset did not include information on individuals’ gender, this was inferred using the `Genderize.io` API [11]; setting the gender attribute to `male` (M) or `female` (F) for confidence thresholds $\lambda \geq 0.7$, and setting it to `unknown` (U) otherwise (see Fig. 3 and further details in Appendix A).

Table I summarizes the descriptive statistics of the dataset divided into two periods. To evaluate the effects of the 2007 law, two temporal snapshots were considered: one covering the period 2000–2007 and another covering the period 2008–2014. The number of films is comparable across both periods, the average number of individuals per film, however, increases from 4.8 to 5.8. This small increase may affect several network metrics, requiring a careful interpretation of the results. In terms of gender composition, females represent 18.6% and 21.5% of the total individuals in each period, respectively. Furthermore, the proportion of individuals with unknown gender remains below 2.8% in both periods and, therefore, is not expected to significantly affect the results.

C. Network construction

After extracting and enriching the data with the gender of individuals, we analyzed them using the toolbox of network science. To this end, we generated two collaboration networks. These networks were constructed by considering individuals associated with films as nodes of the network and adding an edge between two individuals if they have worked together in at least one film. Individuals appearing exclusively in single-member productions were excluded, as they contribute no edges to the network. As each node is labeled according to its gender, the network possesses three types of dyads (*i.e.*, edges): those connecting male nodes, those connecting female nodes, and mixed dyads (corresponding to `male-female`, `male-unknown`, `female-unknown`, and `unknown-unknown` pairs). Thus, besides the complete network, three subnetworks were generated by filtering the data according to the gender composition of the edges. Such

a filter allows us to examine separately the contribution of each gender interaction type. Summing up, six subnetworks were obtained in total: two containing exclusively `female-female` dyads, two with `male-male` dyads, and two comprising mixed interactions. All networks were analyzed using Python’s `NetworkX` library [12].

III. RESULTS

A. Properties of the collaboration networks

A characterization of the global network structure serves as a baseline for the gender analysis that follows. Table II presents the main structural characteristics of the collaboration networks for both periods (see Tab. VIII for the subnetworks). The most recent network has roughly 400 more nodes than the pre-law network, but its L is more than double. As the number of films is approximately constant across periods, this growth can be primarily attributed to the larger average crew size (see Tab. I). Consequently, the values of $\langle k \rangle$, $k_{\max}$, and D may partly reflect this dataset expansion rather than a fundamental structural change. On the other hand, the slight decrease in N_{comp} indicates that the system’s cohesion increases, although the quasi-constant values of S indicate that the bulk of the system belongs to the GC, and that the system’s overall cohesion remained similar. The rise of $\langle C \rangle$ is addressed in Sec. III B via a comparison with values generated with a null model.

TABLE II: Global network characteristics for both periods. We report the number of nodes N , edges L , maximum and average degree $k_{\max}$ and $\langle k \rangle$, density D , average clustering coefficient $\langle C \rangle$, number of components N_{comp} , and giant component size S .

Metric	2000–2007	2008–2014
N	1321	1720
L	5648	11556
$k_{\max}$	91	114
k_{avg}	8.55	13.44
D	0.0065	0.0078
$\langle C \rangle$	0.80	0.82
N_{comp}	72	65
S	0.73	0.72

Table III reports the structural features of the three dyadic subnetworks. All of them grow over time, but this growth is uneven. The F-F subnetwork has the largest relative edge increase (+131%), followed by the mixed subnetwork (+129%). The larger increase in the F-F subnetworks aligns with the increase in female participation. The similar trend in the mixed subnetworks suggests that a higher female presence has also led to more cross-gender collaborations, which is in agreement with the decrease in r_G shown in Tab. IV. Additionally, the edge density and GC size for each subnetwork are

reported in Appendix C (Tabs. VI and VII).

TABLE III: Main features of the **female-female** (F-F), **male-male** (M-M), and mixed filtered networks. For each network, we report the number of nodes, N , and edges, L .

Dyad type	Period	N	L
F-F	2000–2007	187	251
	2008–2014	318	579
M-M	2000–2007	1031	3484
	2008–2014	1287	6592
Mixed	2000–2007	981	1913
	2008–2014	1470	4385

B. Comparison with a null model

As mentioned before, the characteristics observed in our empirical networks – and the increase of some indicators like clustering – have little meaning if we do not compare them with a suitable null model. For this reason, we have considered a null case in which we shuffle the nodes’ gender uniformly at random while preserving the global topology of the network. This randomization is equivalent to reassigning gender to each individual while keeping unchanged their role, number of films, crew size per production, and the abundances of **male/female/unknown** individuals. Therefore, this method provides a natural baseline against which gender-driven association patterns can be assessed. The shuffling procedure is repeated 100 times, and the mean and standard deviation are computed over the ensemble of values.

It is worth noting that for each randomized global network one has to extract the gender-oriented subnetworks again, following the same construction procedure introduced in Sec. II C. The reason is that permuting gender labels alters the edges’ assignments to each subnetwork. To assess the statistical significance of the values obtained, z -scores are computed from the values of the randomized ensembles, with a value of $|z| > 2$ denoting statistically significant differences (*i.e.*, p -value < 0.05).

As shown in Tab. IV, the empirical r_G is positive in both periods, whereas the null model gives values consistent with zero. The corresponding z -scores confirm that the values deviate significantly from the random expectation in both cases. These results suggest that gender homophily constitutes an intrinsic non-trivial feature of the system. The assortativity coefficient decreases by 59% between periods, consistent with an expected trend toward more gender-diverse collaboration following the 2007 legislation. However, r_G displays a positive value in the 2008–2014 network, indicating that the network remains far from gender-neutral mixing. It is also worth noting that r_G is normalized by construction and, therefore, is insensitive to the observed growth of the network.

Table V compares the values of $\langle C \rangle$ for each empirical dyadic subnetwork with their randomized counterparts

obtained with our null model. Since all the z -scores are below $|z| = 2$, none of the empirical values deviate significantly from the random expectation. The M-M and mixed subnetworks show no appreciable deviation in either period. The F-F subnetwork displays a 23% increase between periods and the largest positive deviation from the null model. However, two effects prevent a straightforward interpretation: i) the randomization decreases $\langle C \rangle^*$ for minority-group nodes, and ii) both the larger crew sizes and the greater number of females in the second period boost the size of the F-F subnetworks.

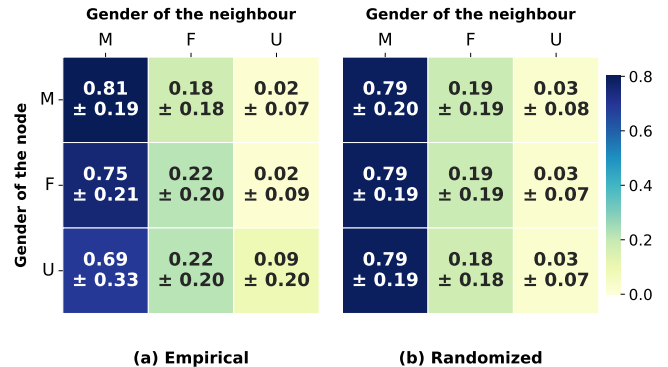


FIG. 1: Average neighbor fraction by gender, $\langle f_{x,y} \rangle$ $x, y \in \{M, F, U\}$ of networks in the period 2000–2007 for empirical (a) and randomized (b) cases.

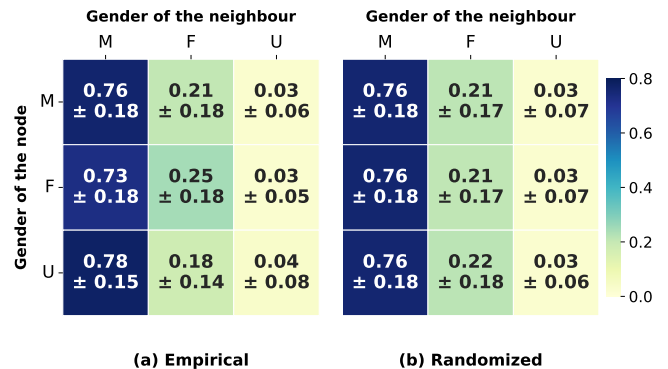


FIG. 2: Average neighbor fraction by gender, $\langle f_{x,y} \rangle$ $x, y \in \{M, F, U\}$ of networks in the period 2008–2014 for empirical (a) and randomized (b) cases.

Figures 1 and 2 display the $\langle f_{x,y} \rangle$ matrices for both

TABLE IV: Gender assortativity for the complete networks considered: we report the values measured in the empirical network, r_G , alongside the average of the ensemble of randomized networks, r_G^* , and the corresponding z -score.

Period	$r_G (\times 10^{-2})$	$r_G^* (\times 10^{-2})$	z -score
2000–2007	3.9	-0.2 ± 1.3	3.15
2008–2014	1.6	-0.2 ± 0.8	2.25

TABLE V: Average clustering coefficient, $\langle C \rangle$, by period and dyadic type. We display the empirical value, $\langle C \rangle$, along with its randomized counterpart, $\langle C \rangle^*$, and the corresponding z -scores.

Dyad type	Period	$\langle C \rangle$	$\langle C \rangle^*$	z -score
F-F	2000–2007	0.52	0.45 ± 0.05	+1.40
	2008–2014	0.64	0.60 ± 0.03	+1.33
M-M	2000–2007	0.77	0.77 ± 0.01	0.00
	2008–2014	0.79	0.80 ± 0.01	−1.43
Mixed	2000–2007	0.10	0.10 ± 0.02	0.00
	2008–2014	0.14	0.12 ± 0.02	+1.25

periods compared to the null model. The analysis of the z -scores (see Appendix B, Fig. 4) reveals that none of the observed deviations are statistically significant in either period. However, a slight tendency can be noted: i) in both periods, **female** nodes show a slightly higher fraction of **female** neighbors and a slightly lower fraction of **male** neighbors than expected under random mixing. ii) **male** nodes connect with other **male** nodes slightly more than expected under random mixing before 2007, with this deviation disappearing after 2007.

IV. CONCLUSIONS

The Spanish film industry collaboration network exhibits statistically significant gender homophily in both periods. This finding indicates that homophily is an intrinsic property of the system rather than an artifact of the abundance of individuals. After the implementation of the 2007 legislation, r_G dropped by 59%. Despite this, r_G remains positive in the post-law period, suggesting that the industry is still far from attaining gender-neutral collaborations. The increase in female participation from 18.6% to 21.5% between periods is reflected in the growth of the F-F and mixed subnetworks. No sta-

tistically significant deviations from the null hypothesis were found for $\langle C \rangle$ or $\langle f_{x,y} \rangle$ in either period. However, the observed trends are consistent with the homophily pattern captured by r_G . Despite some limitations, our findings may suggest that the increase in female participation could drive structural changes favoring greater cross-gender collaborations.

The above results should be interpreted with caution. First, the increase in average crew size may inflate some structural indicators independently of any legislative effect. Second, the analysis relies on a single data source that may present uneven coverage across years and production types, potentially biasing the network. In addition, the gender classification used by **Genderize.io** does not capture gender diversity beyond the binary **male-female** paradigm.

In general, the findings suggest that the 2007 legislation had a *tangible* but *limited* impact on the Spanish film industry. Reducing gender inequality requires going beyond increasing women’s numerical representation. Therefore, effective policies should also ensure the structural inclusion of women, so that participation and collaboration opportunities are equally distributed. Since this work is limited by the characteristics of the data described above, we leave as future work the implementation of a null model that takes into account the possible biases existing within the data. Additionally, a role-based analysis would allow one to examine whether gender segregation differs across professional roles.

Acknowledgments

I would like to sincerely thank my advisor, Alessio Cardillo, for his thoughtful guidance and for always finding time to help throughout this work. I also thank my family and friends for their support and encouragement throughout this journey.

-
- | | |
|--|--|
| <p>[1] M. Quick, BBC News (2022), URL https://www.bbc.com/culture/article/20180508-the-data-that-reveals-the-film-industrys-woman-problem.</p> <p>[2] Y. González Osuna and E. Barrera Martínez, <i>Revista de Comunicación de la SEECI</i> 58, 1 (2025).</p> <p>[3] C. Coronado Ruiz, Área Abierta. <i>Revista de comunicación audiovisual y publicitaria</i> 22, 155 (2022).</p> <p>[4] G. Bianconi <i>et al.</i>, <i>Journal of Physics: Complexity</i> 4, 010201 (2023).</p> <p>[5] D. Verhoeven <i>et al.</i>, <i>PLOS ONE</i> 15, 1 (2020).</p> <p>[6] M. E. J. Newman, <i>Networks: An Introduction</i> (Oxford University Press, 2010).</p> <p>[7] G. Casiraghi and G. Andres, <i>PNAS Nexus</i> 4, pgaf001 (2025).</p> <p>[8] M. Kaiser, <i>New Journal of Physics</i> 10, 083042 (2008).</p> <p>[9] M. E. J. Newman, <i>Phys. Rev. E</i> 67, 026126 (2003).</p> | <p>[10] <i>SensaCine</i>, Available at: https://www.sensacine.com (2026).</p> <p>[11] <i>Genderize</i>, Available at: https://genderize.io/ (2026).</p> <p>[12] A. A. Hagberg <i>et al.</i>, in <i>Proceedings of the 7th Python in Science Conference</i>, edited by G. Varoquaux, T. Vaught, and J. Millman (Pasadena, CA USA, 2008), pp. 11 – 15.</p> |
|--|--|

Estudi del biaix de gènere en la indústria del cinema espanyol

Author: Maria Salceda Servitje, msalcese11@alumnes.ub.edu

Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Alessio Cardillo, alessio.cardillo@ub.edu

Resum: El biaix de gènere afecta diversos (si no tots) aspectes de les nostres vides, inclosa la indústria cinematogràfica. Aquest treball estudia el biaix de gènere en la indústria del cinema espanyol abans i després de la promulgació de la llei d'igualtat de gènere del 2007. Mitjançant el marc que ofereixen els sistemes complexos, es construeixen xarxes de col·laboració a partir de la base de dades de Sensacine. Posteriorment, se'n mesuren les propietats estructurals i es comparen les xarxes obtingudes amb un model nul adequat. Les xarxes empíriques presenten una homofília de gènere estadísticament significativa en ambdós períodes, i aquesta disminueix significativament després del 2007. No obstant això, l'homofília persisteix després del 2007, la qual cosa indica que al període posterior a la llei no es van aconseguir col·laboracions neutres pel que fa al gènere. Per tant, concloem que la llei del 2007 va tenir un impacte estructural tangible però limitat en l'equitat de gènere a la indústria cinematogràfica.

Paraules clau: física estadística, física computacional, sistemes complexos, ciència de xarxes, xarxes socials, escletxa de gènere

ODSs: Aquest TFG està relacionat amb els Objectius de Desenvolupament Sostenible (SDGs)

Objectius de Desenvolupament Sostenible (ODSs o SDGs)

1. Fi de les desigualtats		10. Reducció de les desigualtats	X
2. Fam zero		11. Ciutats i comunitats sostenibles	
3. Salut i benestar		12. Consum i producció responsables	
4. Educació de qualitat		13. Acció climàtica	
5. Igualtat de gènere	X	14. Vida submarina	
6. Aigua neta i sanejament		15. Vida terrestre	
7. Energia neta i sostenible		16. Pau, justícia i institucions sòlides	
8. Treball digne i creixement econòmic	X	17. Aliança pels objectius	
9. Indústria, innovació, infraestructures			

Aquest treball de fi de grau s'emmarca en els ODSs 5, 8 i 10. Atès que l'eix principal del treball és l'estudi de la desigualtat entre homes i dones d'un sector professional concret, podem dir que l'ODS 5 (igualtat de gènere) està fortament vinculat. A més, l'estudi no es limita a mesurar la participació femenina, sinó que examina com es manifesten aquestes desigualtats de forma estructural en els patrons de col·laboració. Per aquesta raó, aquest TFG també es pot relacionar amb l'ODS 10, ja que mitjançant l'anàlisi d'aquests patrons es pot fomentar la reducció de les desigualtats de gènere. Finalment, aquest TFG es pot vincular també amb l'ODS 8 perquè estudia les desigualtats de gènere dins del marc laboral, més concretament, el biaix de gènere present en la indústria del cinema.

Appendix A: Gender assignment

Given the first name and the nationality of an individual, the `Genderize.io` API [11] assigns a gender with a given confidence value, λ . Fig. 3 displays the probability distribution of λ for the names appearing in our data. Most values accumulate around $\lambda = 1$, reflecting high confidence in the gender assignments.

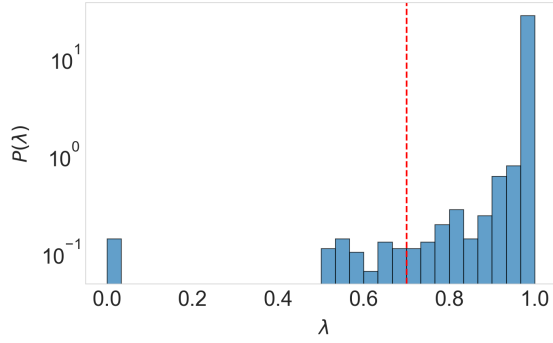


FIG. 3: Probability distribution, $P(\lambda)$, of predicted gender confidence values, λ , for all the individuals appearing in the dataset. The vertical axis is shown in logarithmic scale. In red, the threshold value $\lambda = 0.7$.

Appendix B: z -scores of the average neighbor fraction by gender

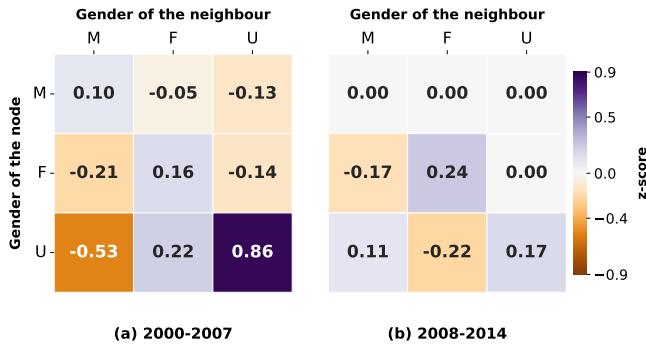


FIG. 4: z -scores of the empirical average neighbor fraction by gender, $\langle f_{x,y} \rangle$ $x, y \in \{M, F, U\}$, relative to the randomized null model for the 2000–2007 period (a) and for the 2008–2014 period (b). Purple hue indicates pairs connecting more often than expected by chance, whereas orange hue indicates the opposite. Average values are computed over an ensemble of 100 realizations.

Appendix C: Other structural features of the networks

Here we report the tables with the values of density, D (Tab. VI), and of the relative size of the giant connected component, S (Tab. VII).

TABLE VI: Edge density, D , by period and dyadic type. We display the empirical value, D , and its randomized counterpart, D^* .

Dyadic type	Period	D	D^*
F-F	2000–2007	0.0144	0.0132 ± 0.0014
	2008–2014	0.0115	0.0119 ± 0.0010
M-M	2000–2007	0.0066	0.0067 ± 0.0002
	2008–2014	0.0080	0.0081 ± 0.0002
Mixed	2000–2007	0.0040	0.0036 ± 0.0002
	2008–2014	0.0041	0.0039 ± 0.0002

TABLE VII: GC proportion, S , by period and dyadic type. We display the empirical value, S with its randomized counterpart, S^* .

Dyadic type	Period	S	S^*
F-F	2000–2007	0.29	0.27 ± 0.11
	2008–2014	0.40	0.37 ± 0.07
M-M	2000–2007	0.67	0.68 ± 0.03
	2008–2014	0.70	0.68 ± 0.02
Mixed	2000–2007	0.75	0.73 ± 0.03
	2008–2014	0.70	0.70 ± 0.02

Appendix D: Properties of the collaboration subnetworks

TABLE VIII: Subnetwork characteristics for both periods and types of dyadic. We report number of nodes N , edges L , maximum and average degree $k_{\max}$ and $\langle k \rangle$, density D , average clustering coefficient $\langle C \rangle$, number of components N_{comp} , and giant component size S .

Metric	F-F		M-M		Mixed	
	2000–2007	2008–2014	2000–2007	2008–2014	2000–2007	2008–2014
N	187	318	1031	1287	981	1470
L	251	579	3484	6592	1913	4385
$k_{\max}$	18	20	73	90	50	91
k_{avg}	2.7	3.6	6.8	10.2	3.9	6.0
D	0.014	0.0115	0.007	0.008	0.004	0.004
$\langle C \rangle$	0.52	0.64	0.76	0.79	0.11	0.14
N_{comp}	41	56	72	72	44	48
S	0.29	0.40	0.67	0.70	0.75	0.70