Universidad del Desarrollo
Facultad de Gobierno
Centro de Investigación en Complejidad Social

Social Complexity of Performing Arts: Quantifying Gender Inequalities and Career Success in Ballet Through Network Science

Author:
Yessica Herrera Guzmán

A thesis presented in fulfillment of the requirements for the degree of Doctor of Philosophy in Complex Social Systems

in the

Research Center for Social Complexity

Principal supervisor:
Cristian Candia Vallejos

Associate supervisor:
Alexander Gates

External supervisor:
Albert-László Barabási

May 23, 2023

Contents

Acknowledgements **5**

Abstract **7**

1 Introduction **9**
References . . . 14

2 Structural effect on gender imbalance in ballet collaboration networks **17**
2.1 Introduction . . . 19
2.2 Methods . . . 20
2.2.1 Data collection and network construction . . . 20
2.2.2 Gender inference . . . 22
2.2.3 Collaboration network analysis . . . 23
2.2.3.1 Gendered differences in centrality . . . 23
2.2.3.2 Null model analysis . . . 25
2.2.4 Perception error on women artists' frequency . . . 26
2.3 Results . . . 26
2.3.1 Collaboration patterns by gender . . . 26
2.3.2 Centrality differences by gender . . . 28
2.3.3 Perception error on the frequency of women artists . . . 31
2.4 Discussion . . . 33
Acknowledgements . . . 36
References . . . 36

3 Quantifying Hierarchy and Prestige in US Ballet Academies as Social Predictors of Career Success **41**
3.1 Introduction . . . 43
3.2 Results . . . 46
3.2.1 Network of ballet schools . . . 46
3.2.2 Career success of ballet dancers . . . 49
3.3 Summary and Discussions . . . 55
3.4 Methods . . . 58
3.4.1 Dataset . . . 58
3.4.2 Measure of social prestige . . . 59
Acknowledgements . . . 59
References . . . 60

A Supplementary Material of Structural effect on gender imbalance in ballet collaboration networks **63**

B Supplementary Material of Quantifying Hierarchy and Prestige in US Ballet Academies as Social Predictors of Career Success **71**
B.1 Youth America Grand Prix . . . 72
B.2 Ballet Schools . . . 72
B.2.1 Distribution of Ballet Schools across the U.S. . . . 73

B.2.2 Schools' achievement . . . 74
B.2.3 Ranking of Ballet Schools . . . 76
B.3 YAGP Alumni . . . 80
B.3.1 Students' achievement . . . 80
B.3.2 Students' success . . . 84
References . . . 86

List of Figures

2.1 **Schematic representation of data processing and network construction.** **a** The collected data contains ballet titles (the title of a ballet creation) and artists' names with their corresponding roles. **b** The data is transformed to a bipartite network, where artists connect to ballet titles that they have been participated. (**c**) From the bipartite network, a projected unipartite network of artists are formed. 22

2.2 **Collaboration patterns by gender in ROH.** **a** The normalized frequency of gender ratio in teams: more than 50% of teams have only men, while teams having only women are nonexistent. **b** The normalized frequency of same gender artists in a team: women collaborate alone in a men-dominated teams, while men collaborate more with other men and form larger teams. **c** The number of ballet creations for each artist. The productivity varies by gender, showing less productivity in women artists' group. Fit line by gender with 95% confidence intervals. 27

2.3 **Distribution of nodes and dyads.** Sample of the the ballet collaboration network of ROH. **a** Nodes are colored yellow or purple for men or women artists, respectively. Node size proportional to degree centrality. Network with dyadic collaborations by gender: **b** man-man; **c** woman-woman. 29

2.4 **Women ratio of $\text{TCA}_{\text{Network}}$ in the empirical data and null models.** Average R_{Women} of $\text{TCA}_{\text{Network}}$ in Edge-shuffled and Gender-shuffled models are shown with standard deviation for ROH sample. The null models show a fairer representation of women artists than the empirical network. 29

2.5 **Comparison of centralities for artists in $\text{TCA}_{\text{Women}}$ and TCA_{Men} with null models.** These figures are for ROH's collaboration network. Z-score of centralities ($Z(C)$) is measured with **a** Edge-shuffled model and **b** Gender-shuffled model. The red line corresponds to the theoretical mean obtained from the the randomized models indicating no difference. **c** The gender gap in the degree centrality (normalized) of TCA is displayed. TCA_{Men} artists have higher degree centrality than their women counterparts. **d** $Z(\Delta C)$ of empirical networks. 31

2.6 **Average perception error $\bar{B}$ for ballet companies.** Average perception error by gender in each company, compared with those from randomized models. **a** ABT, **b** NYCB, **c** NBC, **d** ROH. Line segment in gray guides the difference in perception error by gender group. Most gender groups misconceived the real fraction of women in a network, while the difference is reduced in Edge-shuffled model. The perceived women frequency is more accurate in the gender-shuffled model. . . . 32

3.1 **Network of ballet schools.** Each node is a ballet school, and two schools are connected if they obtained a top student in the same competition venue. Node size and color reflect schools' normalized betweenness centrality, B_k. Node position is determined by the force-directed graph layout with force estimation $\theta = 0.5$ that emphasizes the separation of nodes into clusters. The weak structure (in yellow, network backbone with $\alpha = 0.4$) shows dense connectivity within network clusters and sparser connections between clusters and to the periphery; the strong structure (in red, network backbone with $\alpha = 0.01$), comprises 166 nodes (10%) and 384 edges (0.6%). This network representation illustrates schools' hierarchical structure explaining the role of network position for social prestige. 48

3.2 **Probability of job success in ballet.** Success is defined in a binary fashion, where $P(S=1)$ if the student obtained a job placement in a ballet company, and $P(S=0)$ otherwise. **A** demonstrates a significant positive effect on the predicted probability of a job placement with the increase of school's prestige. **B** shows exponentiated odds ratio with corresponding 95% confidence intervals for the effect on job placement of each additional unit of competition outcomes and institutional prestige. Baseline in 1 indicates no effect. We see that the Grand Prix has the largest effect by type of awards, while long competition trajectories can be detrimental for a job placement, and being a finalist is comparable to be affiliated to ahighly prestigious school. Model coefficients reported in Table 3.1. 53

3.3 **Change in school's prestige for transfer students.** The change in school's prestige (mean with 95% confidence intervals) from students who attend more than one ballet school during their participation in the competition. We observe that students move to more prestigious schools, and students with a successful job placement (labeled as 'success', in yellow) move to even more prestigious schools compared to those without a job placement. 54

S1.1 **Team composition.** Panel **a** shows the probability of a team size in each ballet company. Panel **b** shows the probability of a gender ratio in a team. 64

S1.2 **Collaboration patterns by gender.** Purple/green and dots/triangles represent women/men. Panel **a** shows the fraction of teams by the number of same gender artists in their composition. Panel **b** shows the number of ballet creations and the count of artists by gender (artists' productivity). 65

S1.3 **Proportion of women in ballet collaborations by artistic roles.** Proportion of women written inside purple bar. Group size indicated by bar-width with total count of artists written on top of bar. Figures show women representation by artistic roles for **a** ABT, **b** NYCB, **c** NBC, and **d** ROH. While women face less representation in the Composers group, their representation is consistently low across artistic roles, considering group size. 66

S1.4 **Collaboration networks from ABT, NYCB, and NBC.** 67

S1.5 **Fraction of women of TCA.** Mean of fraction of women, R_{Women} with error bars comparing the empirical networks and the R_{Women} obtained from null models. 68

S1.6 **Gender gap in centrality of TCA.** Reporting normalized values by company. Artists matched by their rank computed in independent gender groups. 69

S1.7 Z**-score of** $C(r)_{real}$**.** Boxplot with error bars. Color code is indicated in each plot. The red line indicates a Z-score equal to zero, corresponding to the mean obtained from each null model. Panel **a** shows the distribution of $Z(C)$, the difference in centrality by rank and gender group $\text{TCA}_{Women,Men}$. Upper boxes display the Edge-shuffled model, lower boxes the Gender-shuffled model. Panel **b** shows the distribution of $Z(\Delta C)$, the Z-score of the gender gap in centrality by rank, separated by randomized model. 70

S2.1 **Network of ballet schools in the US.** Ballet schools are nodes, and two schools are linked by winning in the co-competition of the YAGP. Node size by schools' betweenness centrality, B_k. Schools' location corresponds to their actual geographical location. Nodes are colored by detected communities Γ: Northeast (purple), East (blue), West (orange), South (green), and Midwest (yellow). Edge color/gray for intra/inter community links. 73

S2.2 **Awards and centrality of ballet schools.** **A** shows the fat-tailed distribution of total awards per school, A_k, indicating that from those schools obtaining awards, most of the have only one, and only a few have more than more than 100 awards. **B** shows the log distribution of awards A_k respect to the number of students T_k per school k, i.e. ratio of awards, R_k. Color indicates the corresponding group by ratio of awards and their percentage (plot not displaying schools with 'no medals', $A_k = 0$). Only 56.8% of schools obtained at least one award, and the total number awards increases proportionally to the number of students. However, the ratio of 'high-achievers' is independent from their number of top students. **C** shows the relationship of betweenness centrality, B_k, with each group by ratio of awards. Schools with no medals (in gray) have the lowest centrality value, while only highly central schools are in the 'high-achievers' group. 75

S2.3 **List of top schools.** Selection of top schools by US experts from *Dance Magazine* and *A Ballet Education*. Schools listed in no particular order. 77

S2.4 **AUC for network centrality and ratio of awards.** Panel **A** shows that betweenness centrality is the best indicator of prestige with the largest AUC=0.75 among all centralities in panels **B-D** and schools' achievement, in panel **E**. . . . 78

S2.5 **Ranking of ballet schools.** The list contain the top 5% most prestigious U.S. ballet academies from the network-based ranking. . . . 79

S2.6 **Awards of ballet students.** **A** shows the count of the number number of awards per student A_i, with a power-law distribution. This indicates that most students obtained only one award, while only a few obtained more than 10 awards. **B** shows the percentage of students by group of success ratio, R_i, obtained from dividing the number of awards by the number years they positioned as top student. Taken together, 'no medals' ($R_i = 0$, in gray) and 'under-achievers' ($R_i = (0, 1)$, in cyan) represent the 75.3% of the total number of competitors (6393 students), meaning that most competitors win less than one award per top positioning. About 19.7% of students (1260 students) wins one award per top positioning ($R_i = 1$, in yellow), while only 4.8% of students (309 students) obtains more than one award per time they position as top student. **C** shows the probability of obtaining awards by rank for the finals/semi-finals in orange/blue. We see that a most top students do not obtain an award, and the probability of being awarded decreases as the rank increases, emphasizing scarcity in high rank awards (i.e. Finals). **D** shows the count of students and their number of competitions, T_i by their ratio of awards group, R_i, and demonstrates that all groups display an exponential distribution, suggesting that there is a general pattern of variability in the number of awards obtained per student disregarding of the number of semi-finals competitions. . . . 83

List of Tables

2.1 **Data description and network characteristics of the four ballet companies.** . . . 22

2.2 **Centrality metrics to assess the top-central artists.** Each name denotes the corresponding centrality. . . . 23

2.3 **Network composition by gender.** R_{Women} is the ratio of women in a collaboration network. For the collaborations, the number of woman-woman/man-man/mixed dyadic interactions is counted. . . . 25

3.1 **Model coefficients for the logistic models predicting successful job placement.** Model coefficients labeled by p-value. Standard errors in parentheses. . . . 52

S2.1 Centrality metrics to assess the prestige of ballet schools. Formula in general notation as described in the specified citation. . . . 77

S2.2 **Model Coefficients for the Probability of Success.** Model coefficients labeled by p-value. Standard errors in parentheses. . . . 85

To my beloved husband and my dear mother, Mutti.

In Loving Memory of Pema,

and with heartfelt dedication to my feline companions.

Acknowledgements

I would like to express my deepest gratitude to all those who have supported and contributed to the completion of this thesis. First and foremost, I am immensely grateful to Prof. Albert-László Bárabási, for their guidance, encouragement, and expertise throughout this research. His passion for both arts and science has been a constant source of inspiration for me.

I would also like to extend my gratitude to the members of my thesis committee, Alexander Gates and Cristian Candia, for their valuable feedback, patience, and constructive suggestions, which have greatly enriched the quality of this work. I am grateful for the mentorship from Diego Caro and the time he dedicated to helping me lay the methodological foundation for this work and his encouragement to dive deep into the Data Science world. I appreciate the support from Dorothy Carter during the early stages of this research, her advice and enthusiasm for ballet were the golden seeds of this project. I am specially grateful for the contributions of Eun Lee and Heeate Kim to conduct this research during the difficult times of the COVID-19 pandemic. All their contributions and shared experiences have significantly contributed to the development of my research ideas and methodologies.

I would like to acknowledge the Research Center for Social Complexity at Universidad del Desarrollo for their financial support, which enabled me to carry out this research project. I would also like to express my appreciation to the researchers, staff, and administrators of the Center for Complex Network Research at Northeastern University for their feedback, assistance and logistical support throughout my collaboration with them. I would like to express my sincere gratitude to the Computational Research in Social Sciences Lab for their invaluable support providing resources and dedication for the successful completion of this thesis.

My heartfelt thanks go to my husband for his endless support, and my family and friends for their unconditional love, encouragement, and belief in me. Special gratitude goes to my sister Yerita and to Edgar Onofre, who believed in me from the very beginning of this project. I am deeply grateful for the support and guidance of Prof. Noshir Contractor, Leo Ferres, and Eduardo Graells, who were empathetic companions during the most challenging moments of this research journey. I am immensely grateful to Maria Paz Raveau for sharing meals and fries, and for taking care of Darshan during the pandemic. I am also grateful for the doctors and therapists who provided me with exceptional care, compassion, and guidance for my recovery. Their continuous support and understanding have been my pillars of strength, and I am forever grateful for their presence in my life.

Lastly, I would like to acknowledge all the connected individuals in the field of Complex Social Systems, Network Science, and Science of Science. Their efforts have laid the foundation for my research and have been instrumental in shaping the direction of this thesis. I hold a special gratitude for the people in the artistic community, and Julie Nakagawa and Lauren Huefner, for their motivation and curiosity for this project.

Completing this thesis would not have been possible without the collective support and encouragement from my social network. I am deeply thankful for their contributions and the positive impact they have had on my academic and personal growth.

Abstract

This thesis explores the application of complex systems research to understand the dynamics of the art world, considering art as a complex system and investigating its various components through data-driven methodologies. By studying art as a complex system, we contribute to a systematic understanding of human development and behavior in creative domains.

The social network plays a crucial role in the cultural evolution of art, shaping our cultural identity and collective memory. Analyzing network characteristics provides insights into how individual decisions influence collective dynamics and sustain social phenomena. Previous studies have used network models and data analysis to examine the role of network position and connectedness in artistic collaborations, individual success, and the transmission of artistic knowledge.

In this thesis, we focus on ballet as a unique art form with a rich historical and social structure. Ballet provides an opportunity to investigate the role of the social network in shaping collective dynamics in performing arts. We present two research articles that examine gender inequalities and the role of social connections on the career success of ballet dancers. The first article investigates the social network structure of ballet creations and its potential impact on gender disparities in leading positions. The second article explores the influence of social connections and prestige on the career trajectories of ballet dancers, using network analysis and centrality metrics to uncover hierarchical stratification within ballet academies.

Our research highlights the significance of social dynamics and network effects in understanding complex social phenomena in the art world. It offers insights into gender inequalities and career success in ballet and demonstrates the value of data-centric methodologies in arts research. By generating a unique dataset and applying interdisciplinary approaches, we contribute to the scientific examination of the arts and enhance our understanding of human creativity and cultural heritage.

This thesis contributes to the broader goal of fostering diversity, equity, and inclusion within the arts by shedding light on social structures and suggesting potential avenues for change. Lastly, this work underscores the importance of interdisciplinary research in enriching our understanding of human development and behavior in creative domains.

Chapter 1

Introduction

Complex systems research broadens the landscape of knowledge to disentangle the multiple levels and mechanisms driving the world's social challenges, from human economies to climate change, and the objective exploration of less studied domains, such as the art world. Investigating art as a complex system through data-driven methodologies can deepen the understanding of the different components of the art world: the interaction patterns between artists [1], the response of the audiences [2], or the statistical properties of artistic trends [3], for mentioning a few. Investigating art as a complex system contributes to the systematic understanding of ourselves since art is universally recognized in all societies and unique to humans [4, 5]. Although some scholars consider art as a by-product trait of human evolution, it is still strongly associated as one of the most valuable forms of social expression influencing cultural dynamics [6]. This is because art is a tool driven by technology that is produced from a perceptual brain activity [7, 8], which functions to communicate emotions and aesthetic experiences [9], as well as individual beliefs shaping social interactions [10, 11]. Thus, understanding the social complexity of art can provide insightful findings on the underlying structural mechanisms driving modern human behaviors in creative domains.

Through art, humans convey subjective experiences that create a cultural and cognitive network of minds, ultimately shaping our cultural identity and collective memory [12]. This implies that the social network is essential for the cultural evolution of art. For instance, research on the Florence families from the Renaissance reported the dependency of the social network with different social processes, including the emergence of Renaissance art [13], and such dependency prevails in the modern world [14]. Modern network research highlights the social network as an influential factor from individual decisions, beliefs, and collaborative behaviors, to individual success [15]. Thus, analyzing the network characteristics makes it possible to obtain information on how individual decisions shape collective dynamics that sustain different social phenomena.

Previous studies investigating the arts as a complex social system use network models and data from web repositories to examine the role of network position and connectedness in creative careers. For instance, the social network analysis of artistic collaborations revealed that the structural properties of the network are strong predictors of artists' productivity [16]. Other investigations point out that network centrality plays an essential role in allocating resources and rewards that guarantee the success of individual artists and their teams [1, 17, 18]. In addition, integrating deep learning techniques and network science has facilitated the objective measurement of individual performance and career trajectories, showing that career dynamics and social structures govern the collective impact and success of creative works [19, 20]. Moreover, given that artworks are abstract imitations of ordinary physical objects containing collective and individual symbolism [21], the social network is essential for transmitting, reinforcing, and forgetting artistic knowledge. These effects contribute to our collective memories, which are crucial for social dynamics and, eventually, shape our cultural heritage [22]. These findings highlight network analysis's usefulness in investigating the arts as a complex system and not only as individual behavior, as this can help identify the sociostructural characteristics of cultural endeavors. Therefore, a complex systems approach represents a novel opportunity to explore new methodologies and assess multi-level factors and social drivers shaping the context in which art is produced. In this work, we rely on this approach to use data from web sources, social network analysis, and other computational methodologies to expand the mapping of the art world as a complex social system.

This thesis investigates the network effects on complex social phenomena in the art world using a multidisciplinary perspective. We specifically narrow this research to ballet. Ballet provides a unique opportunity to investigate network effects driving social phenomena because of its characteristic social structure that has prevailed over the years [23]. Many historians have studied the evolution of ballet, and the root of its origins lies within aristocratic festivities, later extended to royal court dances in several European countries. Louis XIV, the father of ballet, formalized the discipline and strongly influenced the style and vocabulary of ballet. Therefore, ballet is an art form that emerged from elitist and hierarchical practices among aristocrats, led by male nobles of the art. Curiously, it is said that ballet evolved thanks to the societal movements seen in human history, from the Renaissance and the Enlightenment periods to the World Wars and the Cold War. These social movements contributed to transforming ballet trends and styles and helped spread the discipline from the core of Western cultures to the rest of the world. In general, the ballet community argues that *every performance we see today, is a long chain tracing back to ballet origins* [24, 25]. Thus, ballet represents a unique opportunity to investigate the role of the social network structure in shaping the collective dynamics of performing arts.

This thesis presents two research articles investigating social dynamics in Ballet. First, in Chapter 2 we study gender inequalities in ballet. Global economic inequality is one of the most complex socioeconomic phenomena broadly investigated [26, 27, 28]. The increase in economic inequality disfavors many social sectors. Still, there is evidence that inequality affects artists and, more importantly, female artists, to enjoy economic growth and access to leading positions in their artistic careers [29, 30, 31]. It is therefore becoming increasingly important to understand the social dynamics of gender inequalities in the arts. The ballet industry has been typically associated with the women figure, often represented as muses [32, 23, 33]. Despite a large majority of women in the workforce, men get higher salaries and leading positions, while women are limited to performing and teaching positions [34, 35, 36]. The existing gender inequalities in ballet suggest a significant segregation of women from leading roles, that have not been improved towards a more equal representation over time. Yet, the problem of gender representation in these positions has not been systematically studied. Thus, we investigate the social network structure of ballet creations and the specific patterns in the collaboration environment. We hypothesize that if the network structure is unbalanced by gender, it will cause the formation of perception errors that reinforce an imbalanced collaboration pattern for the subsequent segregation of females away from leading positions, a social phenomenon known as the glass ceiling effect [37, 38]. This research relies on the stable collaborative structure of ballet concerning other forms of dance to conduct network analysis with scientific validity. Understanding the network effects on gender inequalities would inform policymakers and the management of cultural organizations about potential sociostructural changes for more diversity and equality in the cultural industry and other professional areas leaning on creative work [39, 40].

Second, in Chapter 3, we further investigate the role of social connections on the career success of ballet dancers. The rise to prominence of ballet dancers often hinges on remarkable physical abilities [41, 42, 43]. However, the ballet's historical record indicates that dancers' trajectories were largely contingent upon guild membership —facilitated by court influence— or access to the king, who could bestow privilege by royal decree [23]. In contemporary contexts, ballet academies and institutions may have supplanted the king and court's roles, influencing dancers' career trajectories beyond pure performance ability. We propose that the prestige of ballet academies enhances dancers' professional growth and job placement, while also increasing the academy's social recognition and achievement by the number of professional dancers it produces. This relationship creates an infinite feedback loop, where the academy's reputation enhances dancers' career prospects, and the success of its dancers further increases the academy's prestige. Dancers could leverage this principle by associating with prestigious ballet academies that offer access to professional networks supporting talented dancers. Despite this, there is a lack of research investigating the social drivers of success in ballet.

Our research delves into the ballet academic system and explores its interplay with social prestige in shaping dancers' professional paths. Although awards and high achievement are indisputably pivotal for gaining social recognition [44, 45, 46], we propose that centrality metrics serve as a more precise indicator of social prestige, highlighting the essential role of social connections in bolstering prestige [47, 17]. Through network analysis, we offer an innovative contribution that uncovers the hierarchical social stratification of prestige within the ballet academic milieu. A central contribution of this research is formulating a network-based ranking system for ballet academies, serving as a gauge of social prestige. We employ this ranking to forecast the job placements of ballet students, juxtaposing it with their competition performance to evaluate whether social factors impact selection by company recruiters. This research stresses the significance of examining social prestige's influence on career trajectories in the ballet industry. It highlights the potential of network science methodologies to offer a comprehensive perspective on dancers' professional development, beneficial for educators and researchers intrigued by success patterns in creative careers.

This thesis unveils observations about integrating creative domains into distinct social structures and hierarchical stratifications that govern complex social phenomena. It emphasizes the application of network metrics to elucidate gender inequalities and career success. Broadly, our research paves the way for enhancing the application of data-centric methodologies in the scientific examination of the arts.

Creating our unique dataset is a big step forward in overcoming the lack of data in arts research. We aspire to have demonstrated the value of digitized data for projects at the intersection of computational social sciences, artistic, and humanistic pursuits. Our methodology may inspire platform owners from cultural organizations to maintain data availability and proper organization, fostering a robust data infrastructure for investigating less explored creative domains. We hope our work shines a spotlight on how interdisciplinary research enriches our understanding of human creativity on a societal level, informing us about the underlying mechanics that shape our cultural heritage. In particular, further research on this intersection may help elucidate complex social structures of cultural dynamics, bridging scientific practices with the artistic community, to address questions related to different cultural contexts, artistic disciplines, and cultural production. This knowledge can inform decision-making processes, program development, and policy formulation, ultimately enhancing the organization's ability to engage with audiences, support artists, and contribute to the cultural landscape.

In summary, we offer valuable insights into contemporary social structures of performing arts that may necessitate tailoring to foster increased diversity, equity, and inclusion within creative careers.

References

[1] Brian Uzzi and Jarrett Spiro. “Collaboration and Creativity: The Small World Problem”. en. In: *American Journal of Sociology* 111.2 (Sept. 2005), pp. 447–504. ISSN: 0002-9602, 1537-5390. DOI: 10.1086/432782. URL: http://www.journals.uchicago.edu/doi/10.1086/432782 (visited on 04/02/2019).

[2] Shuo Yan et al. “Exploring audience response in performing arts with a brain-adaptive digital performance system”. In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7.4 (2017), pp. 1–28.

[3] Higor YD Sigaki, Matjaž Perc, and Haroldo V Ribeiro. “History of art paintings through the lens of entropy and complexity”. In: *Proceedings of the National Academy of Sciences* 115.37 (2018), E8585–E8594.

[4] Frans De Waal. *The ape and the sushi master: Cultural reflections of a primatologist.* Basic Books, 2008.

[5] C. Walter. “The First Artists: How creativity made us human.” In: *National Geographic Magazine* 227.1 (2015), pp. 33–57.

[6] Derek Hodgson and Jan Verpooten. “The Evolutionary Significance of the Arts: Exploring the By-product Hypothesis in the Context of Ritual, Precursors, and Cultural Evolution”. In: *Biological Theory* 10.1 (Mar. 1, 2015), pp. 73–85. DOI: 10.1007/s13752-014-0182-y. URL: https://doi.org/10.1007/s13752-014-0182-y.

[7] Pablo PL Tinio. “From artistic creation to aesthetic reception: The mirror model of art.” In: *Psychology of Aesthetics, Creativity, and the Arts* 7.3 (2013), p. 265.

[8] Anjan Chatterjee and Oshin Vartanian. “Neuroaesthetics”. In: *Trends in cognitive sciences* 18.7 (2014), pp. 370–375.

[9] Robin George Collingwood. *The principles of art.* Vol. 11. Oxford University Press, 1958.

[10] Cornelia Geer Leboutillier. “Art as Communication”. In: *The Journal of Aesthetics and Art Criticism* 2.8 (1943), p. 76. ISSN: 00218529. DOI: 10.2307/425948. URL: https://www.jstor.org/stable/10.2307/425948?origin=crossref (visited on 03/08/2019).

[11] Merlin Donald. “Art and Cognitive Evolution”. In: *The Artful Mind: Cognitive Science and the Riddle of Human Creativity.* Oxford University Press, Oct. 2006. ISBN: 9780195306361. DOI: 10.1093/acprof:oso/9780195306361.003.0001. eprint: https://academic.oup.com/book/0/chapter/149231954/chapter-ag-pdf/44987507/book_5920_section_149231954.ag.pdf. URL: https://doi.org/10.1093/acprof:oso/9780195306361.003.0001.

[12] Jan Assmann and John Czaplicka. “Collective memory and cultural identity”. In: *New german critique* 65 (1995), pp. 125–133.

[13] Paul D McLean. *The art of the network: Strategic interaction and patronage in renaissance Florence.* Duke University Press, 2007.

[14] Zizi Papacharissi. *A networked self: Identity, community, and culture on social network sites.* Routledge, 2010.

[15] M.O. Jackson. *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors.* Knopf Doubleday Publishing Group, 2019. ISBN: 9781101871447. URL: https://books.google.com.mx/books?id=X-4sDwAAQBAJ.

[16] Brian Uzzi. “A social network’s changing statistical properties and the quality of human innovation”. In: *Journal of Physics A: Mathematical and Theoretical* 41.22 (2008), p. 224023.

[17] Samuel P. Fraiberger et al. “Quantifying reputation and success in art”. en. In: *Science* 362.6416 (Nov. 2018), pp. 825–829. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aau7224. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aau7224 (visited on 01/28/2019).

[18] Milán Janosov, Federico Battiston, and Roberta Sinatra. “Success and luck in creative careers”. In: *EPJ Data Science* 9.1 (2020), p. 9.

[19] Lu Liu et al. “Hot streaks in artistic, cultural, and scientific careers”. In: *Nature* 559.7714 (2018), p. 396.

[20] Lu Liu et al. “Understanding the onset of hot streaks across artistic, cultural, and scientific careers”. In: *Nature communications* 12.1 (2021), pp. 1–10.

[21] Thomas Adajian. “The Definition of Art”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring, 2022. Metaphysics Research Lab, Stanford University, 2022.

[22] Alin Coman et al. “Mnemonic convergence in social networks: The emergent properties of cognition at a collective level”. In: *Proceedings of the National Academy of Sciences* 113.29 (2016), pp. 8171–8176.

[23] Jennifer Homans. *Apollo’s angels: A history of ballet*. 1st edition. Random House Publishing Group, 2010.

[24] Joan Lawson. *A History of Ballet and its Makers*. New York: Pitman Publishing Corporation, 1964.

[25] Carol Lee. *Ballet in western culture: a history of its origins and evolution*. Psychology Press, 2002.

[26] Gary S Becker. “Human capital, effort, and the sexual division of labor”. In: *Journal of labor economics* 3.1, Part 2 (1985), S33–S58.

[27] Anthony B Atkinson. *Inequality: What can be done?* Harvard University Press, 2015.

[28] Thomas Piketty. “About capital in the twenty-first century”. In: *American Economic Review* 105.5 (2015), pp. 48–53.

[29] Chad M Topaz et al. “Diversity of artists in major US museums”. In: *PloS one* 14.3 (2019), e0212852.

[30] Danielle J Lindemann, Carly A Rush, and Steven J Tepper. “An asymmetrical portrait: Exploring gendered income inequality in the arts”. In: *Social Currents* 3.4 (2016), pp. 332–348.

[31] Philippe Coulangeon, Hyacinthe Ravet, and Ionela Roharik. “Gender differentiated effect of time in performing arts professions: Musicians, actors and dancers in contemporary France”. In: *Poetics* 33.5-6 (2005), pp. 369–387.

[32] Ann Daly. “The Balanchine woman: Of hummingbirds and channel swimmers”. In: *The drama review: TDR* 31.1 (1987), pp. 8–21.

[33] Lisa DeFrank-Cole and Renee K Nicholson. “The slow-changing face of leadership in ballet: an interdisciplinary approach to analysing women’s roles”. In: *Leadership and the Humanities* 4.2 (2016), pp. 73–91.

[34] Elizabeth Yntema. *Dance Data Project First Look Report 2022-2023*. 2022. URL: www.dancedataproject.com.

[35] Elizabeth Yntema. *The ballet world is still male-dominated, research shows*. Aug. 2019. URL: https://www.womensmediacenter.com/news-features/the-ballet-world-is-still-male-dominated-research-shows.

[36] Kim Elsesser. *A Gender Gap In Ballet, Seriously?* Sept. 2019. URL: https://www.forbes.com/sites/kimelsesser/2019/09/12/a-gender-gap-in-ballet-seriously/.

[37] David A Cotter et al. “The glass ceiling effect”. In: *Social forces* 80.2 (2001), pp. 655–681.

[38] Christine L Williams. “The glass escalator: Hidden advantages for men in non traditional occupations”. In: *Social Problems* 39 (1992), pp. 253–267.

[39] Pierre Bourdieu. *The field of cultural production: essays on art and literature*. New York: Columbia University Press, 1993.

[40] Ellen Winner, Thalia R. Goldstein, and Stéphan Vincent-Lancrin. *Art for Art's Sake?: The Impact of Arts Education*. en. Educational Research and Innovation. OECD, June 2013. doi: 10.1787/9789264180789-en. (Visited on 03/06/2019).

[41] Moira Cameron McCormack et al. "The physical attributes most required in professional ballet: a Delphi study". In: *Sports medicine international open* 3.01 (2019), E1–E5.

[42] Nathan Kogan. "Careers in the performing arts: A psychological perspective". In: *Communication Research Journal* 14.1 (2002), pp. 1–16.

[43] Marija Jankovic and Stefan Bogaerts. "Predicting success in the performing arts: Ballet and music". In: *Psychology of Music* 49.4 (2021), pp. 945–957.

[44] Don H Shamblin. "Prestige and the sociology establishment". In: *The American Sociologist* (1970), pp. 154–156.

[45] Bernd Wegener. "Concepts and measurement of prestige". In: *Annual review of sociology* (1992), pp. 253–280.

[46] David Zeitlyn and Daniel W Hook. "Perception, prestige and PageRank". In: *Plos one* 14.5 (2019), e0216783.

[47] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

Chapter 2

Structural effect on gender imbalance in ballet collaboration networks

Ballet is a performing art strongly stereotyped by a women figure and a well known women-dominated profession, yet non-negligible imbalance of gender representations have been reported in leading positions. Moreover, no research considers collaborations' structural influence on this disproportionate gender representation in spite of its collaborative characteristics. In this study, we investigate the gendered labor force composition in ballet creations and collaborations' structural patterns. Our results confirm the numerical gender imbalances in the creations, and show that the existing numerical gender disparity is amplified by collaboration patterns, which ultimately affect women's visibility in the field. Mathematically calculated perception errors on women artists' frequency suggest that gender-assortative structures could be related to the misperception on the size of the women artists, which may negatively impact their professional experiences. This study provides an insight of the role of a gender-diverse collaborative environment in the ballet industry.

Key words: gender inequality, collaborations, social network, perception error, performing arts.

2.1 Introduction

Ballet is an influential performing art that historically has privileged a women figure [1], even though it was originally led and danced only by men [2]. Today, ballet is assumed as a women-dominated industry where employment would favour women at all levels. 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, claiming for equal professional opportunities for women [4, 5]. For instance, the data from American dance companies assessed the unequal representation of women (less than 40%) in artistic and executive positions [6]. On the other hand, the overall women participation in the workforce is about 70% [7]. This difference of gender representation in leadership-level raises a question whether or not the 'glass ceiling' effect exists in ballet industry [8, 9].

Network research explains one reason of gender inequalities with one's benefit optimization of their local network to acquire useful information [10], creativity [11], productivity [12], career success [13], and visibility of individual attributes [14]. In particular, many studies demonstrated gendered dissimilarity in the utilization of one's collaboration network [15, 16]. The collaboration networks with gendered preference can affect one's visibility and rank [17].

Moreover, gendered structure can affect individuals' perceptions on a gender group's frequency and visibility. Studies found that the visibility of an attribute, such as the gender, are correlated to individuals' perception errors on an attribute frequency when the attribute is represented disproportionately in a social network [18, 19, 20]. As a result, a group's importance can be underestimated more of what can be expected from the attribute's real composition [17]. Since the perception errors could reinforce an unequal pattern in social connections (e.g. collaborations), understanding the role of network structure regarding gender imbalance can give an insight to an intervention of gendered imbalance in professional positions.

Despite ballet creation lean on heavy collaborations, most reports have only focused on a numerical inequality in the composition of women and men artists, and little is known about the role of collaboration structures on gender imbalances in ballet. The collaboration structures include diverse core creators, beyond dancers, such as choreographers, composers, and costumes and light designers.

In this study, we use network analysis to investigate gender imbalance in collaboration structure of ballet creations. We construct collaboration networks from four renowned ballet companies and analyze their numerical gender composition. Then, we compare the real-world collaboration structures with randomized network models. We specifically investigate a structural gendered difference in highly centered labor force in collaboration networks. The formation of perception errors on the gender composition has been measured to examine a possible relationship between gendered collaboration networks and perceived working environment.

To the best of our knowledge, our study is the first attempt to understand the structural effects on gender imbalances in major ballet companies. This research will help understand the gender inequalities in a highly collaborative performing arts and will shed light for more effective interventions to reduce the gap of gender representations in general collaborations.

2.2 Methods

2.2.1 Data collection and network construction

We construct collaboration networks of artists from four major ballet companies' records. We select four renowned companies—the American Ballet Theatre (ABT) [21], the New York City Ballet (NYCB) [22], the National Ballet of Canada (NBC) [23], and the Royal Ballet of the Royal Opera House (ROH) [24]—based on their worldwide prestige and the availability of their historical repertoire. The data is collected using a Robotic Process Automation method for web scraping [25]. In this process, we focus only on the *original ballet titles* from each company's online repertoire. The meaning of *original ballet titles* includes a ballet production that is created by a company, recreated works, and company premieres that were originally debuted at a different ballet company but added a new artistic interpretation by a company. Each record of the original ballet production includes participated principal artists' names and their roles.

A typical ballet collaboration in the collected data contains a core structure of artists, such as Choreographer, Composer, and Costume and Light Designer. However, there is no conventional composition of the artist types in the data. For example, a historic creation has simpler configuration than the current form of collaboration, while recent works include more diverse artists with the development of technology (e.g. media editor) or by adding multiple collaborators for the same job in a collaborative manner (e.g. two or more composers).

The composition of the collaboration data is illustrated in Fig. 2.1a, showing the list of ballet productions by title (as an example, 'Ballet 1' and 'Ballet 2') with artists' names (A, B, C, D, and E), and their artistic roles (e.g. Choreography, Music, and Costumes). Here, we refer artists who collaborated in a ballet creation together as a team. Staff positions are not considered for the collaboration networks studied here. To construct the collaboration network of each company, we first build a bipartite network between ballet creations and artists (Fig. 2.1b), where left nodes represent ballet titles and right nodes display artists participated in the ballet title. Then, artists' collaborations are projected to an undirected graph as shown in Fig. 2.1c. Here, each node represents an individual artist, and a link between two artists denotes their collaboration in the same ballet creation. An artist who teamed up in numerous ballet creations connects multiple artists in the same company and becomes a key connector in a collaboration network (Fig. 2.1c).

The resulting empirical networks include around 490–850 artists (nodes) with around 1900–3100 collaborations (links), depending on each company's history. The time of reported ballet creations ranges from 1930's to 2020's. On these networks, we assess basic network properties such as average clustering coefficient [26], average shortest path [27], and small-worldness [28] (see Table 2.1). The 'Giant component' in the Table 2.1 explains the largest component (90% of a collaboration network) of each collaboration network.

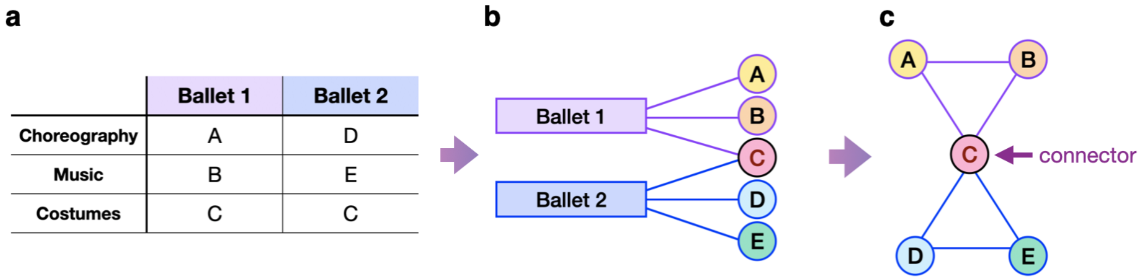


Figure 2.1: **Schematic representation of data processing and network construction.** **a** The collected data contains ballet titles (the title of a ballet creation) and artists' names with their corresponding roles. **b** The data is transformed to a bipartite network, where artists connect to ballet titles that they have been participated. (**c**) From the bipartite network, a projected unipartite network of artists are formed.

Table 2.1: **Data description and 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
Giant Component				
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

2.2.2 Gender inference

Before inferring an artist's gender from one's name, artists' names were processed for misspelling, middle names, and initials to distinguish artists' identity. The names are held constant even though their names are reported in multiple companies. Then, we infer artists' gender by using `gender` package for R. This package contains names from various countries and periods, thus it is adequate for this study since the collected data contains names of artists with diverse nationalities and they were born in the 19th and early 20th centuries.

To estimate an artist's birth year, we assume 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) using the standardized databases of the package (`ssa`, `ipums`, `napp`, and `demo`) [29, 30]. Then, it estimates a probability that a person would have certain gender 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 refer directly the sex of the artist nor the self-assigned gender chosen by each artist. For the names which were not able to assign gender with this method, it is manually assigned after a web search of the artist's identity.

2.2.3 Collaboration network analysis

2.2.3.1 Gendered differences in centrality

From the collaboration networks, we measure four centralities of each artist (degree centrality [31], harmonic centrality [32], betweenness centrality [33], and eigenvector centrality [34]) to explore whether an artist' position in a collaboration network differs by gender (see Table 2.2; here, all equations follow the expression of each reference paper which is independent from this study). All centralities are normalized by the maximum value of the centrality in a gender or a company group to have a linear scaling of $[0, 1]$ range.

Table 2.2: **Centrality metrics to assess the top-central artists.** Each name denotes the corresponding centrality.

Degree	a node's connectivity	$D_i = \frac{k_i}{N-1}$ [31]
Harmonic	a node's closeness to other nodes	$H(i) = \sum_{i \neq j} \frac{1}{d(i,j)}$ [32]
Betweenness	how much a node connects two other nodes	$P_i = \sum_{s,t} \frac{n^i_{s,t}}{n_{s,t}}$ [33]
Eigenvector	a node's influence	$E_i = \frac{1}{\lambda} \sum_k a_{k,i} \, E_k$ [34]

From the centralities, first, we check the ratio of women ranked at Top 20. Specifically, we sort all artists by their normalized centralities in descending order, and selected 20 artists. This Top 20 group is called as *Top-Central Artists* (TCA) in this study. TCA only considers the rank of centralities, $r = \{1, 2, \ldots, 19, 20\}$, so $r = 1$ represents the most central artist having the highest corresponding centrality, and $r = 20$ will have the lowest centrality in a TCA_{group}). We use TCA for two different cases: the first case is for all artists in a company's collaboration network ($\text{TCA}_{\text{Network}}$), and the second case is for a company's artist group by gender (e.g., $\text{TCA}_{\text{Women}}$, TCA_{Men}).

For $\text{TCA}_{\text{Network}}$, we consider tied TCA artists at the same rank, so a variation of the total number of artists is possible if there are artists with equal centrality at each rank. However, for $\text{TCA}_{\text{Women}}$ and TCA_{Men}, the tied centrality is not considered to keep an equivalent number of women and men artists. Then, we quantify the women ratio R_{Women} for a TCA_{group} ($R_{\text{Women}} = \frac{\sum_i^{N_{\text{TCA}}} \theta(i)}{N_{\text{TCA}}}$). Here, i denotes an index for an artist who is in a corresponding TCA, and N_{TCA} represents the total number of artists in a TCA_{group}, and $\theta(i) = 1$ or 0 when an artist is woman or men, respectively.

The difference of centrality ($\Delta C(r)$) between two rank-matched artists from each gender group is measured as $\Delta C(r) = C_{\text{Men},r} - C_{\text{Women},r}$. Here, C denotes a corresponding centrality value of an artist at the same rank r in $\text{TCA}_{\text{Women}}$ and TCA_{Men}. If there is a woman artist ranked 1 in $\text{TCA}_{\text{Women}}$ with a centrality value of 0.4, she is at the most central position in the women artist group, and it can be written as $C_{\text{Women, 1}} = 0.4$. The counterpart of man artist, who is ranked 1 as well in TCA_{Men} will be $C_{\text{Men, 1}} = 0.5$, 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 is located on more central position than the woman counterpart.

A numerical fraction of women artists at the network level of 0.5 is assumed as a gender-balanced collaborations, and we call this situation as 'neutral' composition. Fractions of three types dyadic interactions over the total number collaborations (woman-woman, man-man, or mixed) are assessed to test the distribution of gendered collaboration preferences in collaboration networks (see Table 2.3).

Table 2.3: **Network composition by gender.** R_{Women} is the ratio of women in a collaboration network. For the collaborations, the number of woman-woman/man-man/mixed dyadic interactions is counted.

	ABT	NYCB	NBC	ROH
Artists (nodes)	779	549	579	847
R_{Women}	0.22	0.19	0.18	0.19
Collaborations (links)	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%)

2.2.3.2 Null model analysis

With the help of null models, we can understand how strong the gendered characteristics and patterns in the empirical collaboration networks. The randomized null models remove the collaborator- or gender-preference by shuffling collaborations (links) or artists' attributes in the cumulative collaboration network (see Fig. 2.1c).

The first null model is the edge-shuffled model. It randomizes connections between artists by shuffling edges, while it holds network properties such as the number of collaborations per an artist, and the gender of them. The resulting networks show collaboration structures when there is no gender preference.

The second null model, gender-shuffled model, only rearranges artists' gender. This process preserves all network properties, while the number of participations in ballet creations for a gender group would be changed. Therefore, the resultant networks display an artificial collaboration pattern without a correlation between an artist's gender and productivity as well as a gendered collaboration assortativity.

For a distinction between the original network's centralities and a null model's centralities, we denote the observed centrality by rank in the real network as $C(r)_{\text{real}}$, and that of the null model as $C(r)_{\text{null}}$. Then, TCA$_{group}$'s Z-score of $C(r)_{\text{real}}$ are calculated based on the averaged centrality of 100 null models ($\bar{C}(r)_{\text{null}}$, $Z(C) = \frac{C(r)_{\text{real}} - \bar{C}(r)_{\text{null}}}{\sigma(C(r)_{\text{null}})}$). Z-score of $\Delta C(r)$ in the empirical network is also measured with the values of a randomized network as $Z(\Delta C) = \frac{\Delta C(r)_{\text{real}} - \Delta\bar{C}(r)_{\text{null}}}{\sigma(\Delta C(r)_{\text{null}})}$).

2.2.4 Perception error on women artists' frequency

An individual's perceptions on the global frequency of an attribute can be inferred from local frequency of the attribute [19, 20, 17]. Even though the social mechanisms of perception are rather complex, here we use a simple mathematical approach to estimate the perception error on women artists' frequency in an entire collaboration network. In practice, the perception error of an individual artist i (B_i) is obtained as the ratio between the local fraction of women among i's collaborators and the actual fraction of women in an entire collaboration network that i is involved in, $B_i = \frac{W_i}{R_{\text{Women}}}$. Here, W_i denotes the local fraction of women among i's collaborators.

Based on the individual artist i's bias, an averaged perception error by gender group is also measured ($\bar{B} = \frac{\sum_i B_i}{N_{group}}$), where N_{group} represents the total number of artists in a group. When $\bar{B} = 1$, a group's perception on a network-level women ratio is accurate on average; when $\bar{B} < 1 (\bar{B} > 1)$, a gender group underestimates (overestimates) the ratio of women artists on average. A gendered homophily is measured by the method in [20] to see a macroscopic gendered preference of the collaboration networks.

2.3 Results

Based on previous reports on gender inequalities of ballet industry in leading positions [7, 35, 6], we analyze the gender composition of the central workforce in ballet creations with a network analysis. We also look into the structural effect of an unequal gender representation on artists' collaboration patterns on the formation of perception errors comparing with null model analysis.

2.3.1 Collaboration patterns by gender

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. There are about 50% of teams having 100%

men artists. While teams having 100% women artists is almost zero (see Figures 2.2a and S1b). In addition, less than 10% of teams have gender neutral ratio of 50% (Fig. S1b). This pattern shows that the majority of teams are composed with less than 50% of women artists, regardless of their sizes. The fraction of teams having at least one additional woman artist is about 60% among women-involved teams (Figures 2.2b and S2a). These results describe that women artists work in men-dominated environments. Conversely, men-alone teams are rather rare ($< 10\%$), as they tend to collaborate with at least other three to five men ($> 20\%$) and form considerably larger teams than women (up to 11 men in one team, at ROH).

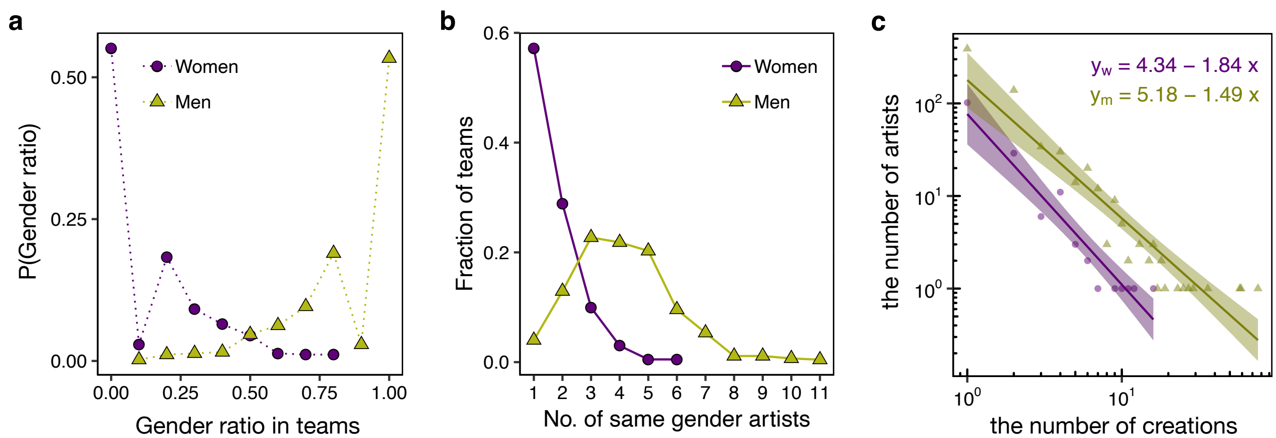


Figure 2.2: **Collaboration patterns by gender in ROH. a** The normalized frequency of gender ratio in teams: more than 50% of teams have only men, while teams having only women are nonexistent. **b** The normalized frequency of same gender artists in a team: women collaborate alone in a men-dominated teams, while men collaborate more with other men and form larger teams. **c** The number of ballet creations for each artist. The productivity varies by gender, showing less productivity in women artists' group. Fit line by gender with 95% confidence intervals.

In terms of the productivity, women artists are less involved in ballet creations than men artists. In NYCB and ROH , the most productive woman participates in about 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 Figures 2.2c, and S2b). 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 to the most productive man artist (ABT's maximum collaborations: Men=35, Women=55). Despite of the exception, most women artists have lower number of ballet creations than men artists in a men-dominated environments.

The average proportion of women by artistic role is also low for Choreographer (13%), Librettist (13%), Composer (4%). Other positions such as Costumes (39%), Lighting (22%), and Design (23%) have relatively large participation of women, but still women artists are less than men artists. Dance communities has specifically reported an overlooking of women in choreographic leads, yet the observed fraction of women indicates that they are actually underrepresented not only for roles in choreography or music, but also in the other artistic domains in ballet creations (see Fig. S3).

The underrepresentation of women artists is clearer in a visual representation of the collaboration networks where men are not only a majority but also more central with high connectivity (see Figures 2.3a, and S4). The node size is proportional to the number of collaborations. The man-man connections are more than 60% across companies (yellow links, Fig. 2.3b) and mixed connections are about 30% on average. On the other hand, woman-woman connections are less than 5% of the total dyadic interactions (purple links, Fig. 2.3c). Men artists are densely co-worked with other artists regardless of gender, locating at the center of the collaboration network (yellow circles). While women artists are sparsely distributed in the periphery of the network (purple circles).

2.3.2 Centrality differences by gender

So far, we have numerically and visually checked a relatively small size of women artists, and their less frequent artistic collaboration than men artists. This observations raise some questions: Does the small number of artists definitely lead a low visibility? To answer this question, four network centralities are compared to capture positional and influential difference of artists by their gender (Table. 2.2). In particular, this comparison is limited to TCA for a company ($TCA_{Network}$), and the TCA by gender (TCA_{Women} and TCA_{Men}) to see the gender gap in top-hierarchical positions. The

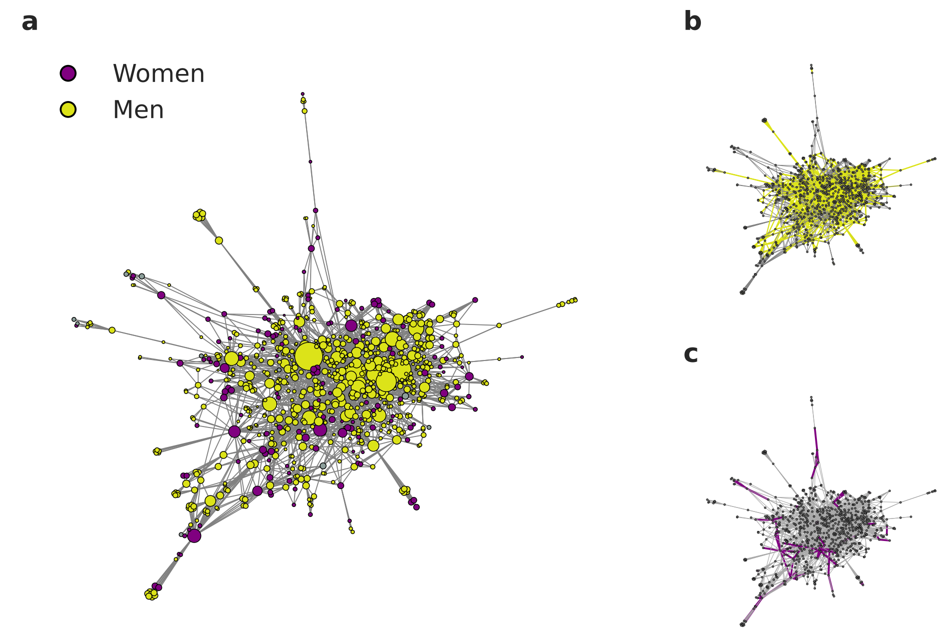


Figure 2.3: **Distribution of nodes and dyads.** Sample of the the ballet collaboration network of ROH. **a** Nodes are colored yellow or purple for men or women artists, respectively. Node size proportional to degree centrality. Network with dyadic collaborations by gender: **b** man-man; **c** woman-woman.

measured fraction of women of TCA sets are analyzed with that of randomized networks to quantify an extent of gender imbalance in the empirical one (see Methods).

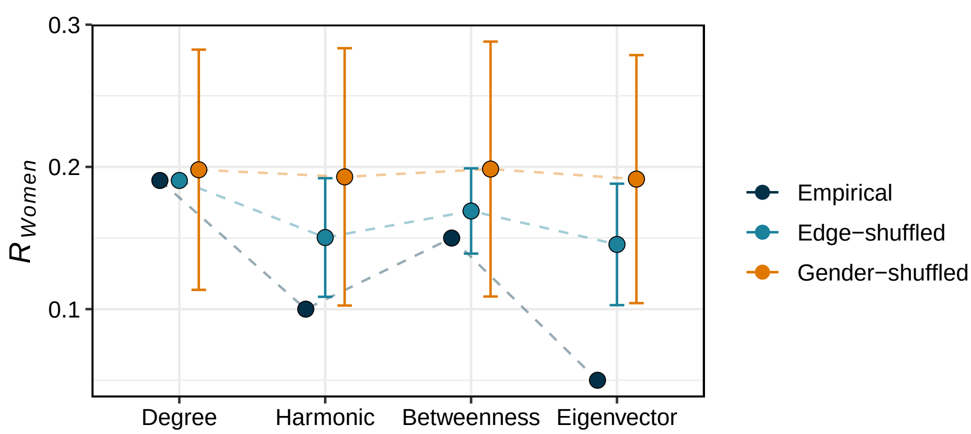


Figure 2.4: **Women ratio of $\text{TCA}_{\text{Network}}$ in the empirical data and null models.** Average R_{Women} of $\text{TCA}_{\text{Network}}$ in Edge-shuffled and Gender-shuffled models are shown with standard deviation for ROH sample. The null models show a fairer representation of women artists than the empirical network.

The fraction of women ranked in $\text{TCA}_{\text{Network}}$ in the empirical network is less than 20% for all centralities. However, R_{Women} increases in the randomized models for all centralities across companies (Fig. S5). For example, Edge-shuffled model improves the representation of women for harmonic centrality from 10% to 15%, and Gender-shuffled model raises it up to 19% in ROH (Fig. 2.4). Note that Edge-shuffled model keeps R_{Women} in TCA regarding degree centrality because the number of collaborations (degree) for an artist and their inferred gender are held constant.

This result suggests that the underrepresentation of women artists could be related to gender assortative collaborations, and the current level of women artists' centrality is not a deterministic outcome of the small size of women artists. That is to say, even when the fraction of women remains small in a network, women artists' representations could be improved if gender-diverse collaborations were encouraged.

For $Z(C)$ in $\text{TCA}_{\text{gender}}$, one can see a clear improvement of women artists' centralities in null models (ROH's sample in Fig. 2.5, all companies in Fig. S7a). In both randomized models (Edge-shuffled: Fig. 2.5a, Gender-shuffled: Fig. 2.5b), men artists' Z-scores are mostly positive for all centralities, representing that men artists are placed at more central positions than the expected locations from null models. In contrast, women artists' Z-scores are mostly negative, showing that their positional importance can be improved in a synthetic collaboration network without gendered collaboration assortativities (Edge-shuffled) and productivity imbalance (Gender-shuffled). This could indicate that the disproportionate gendered productivity and collaboration patterns may limit women artists' visibility, regardless of their small fraction in the field.

The harmonic centrality is the only centrality showing a negative Z-score in Edge-shuffled model (Fig. 2.5a). It is because harmonic centrality denotes an extent of an artist's closeness to other artists on average, so small value represents far distance between nodes. Thus, the negative Z-score suggests that the distance among artists in empirical collaboration networks fall apart farther than the expected distance from the random models.

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 (Fig. 2.5c). This trend of ΔC is

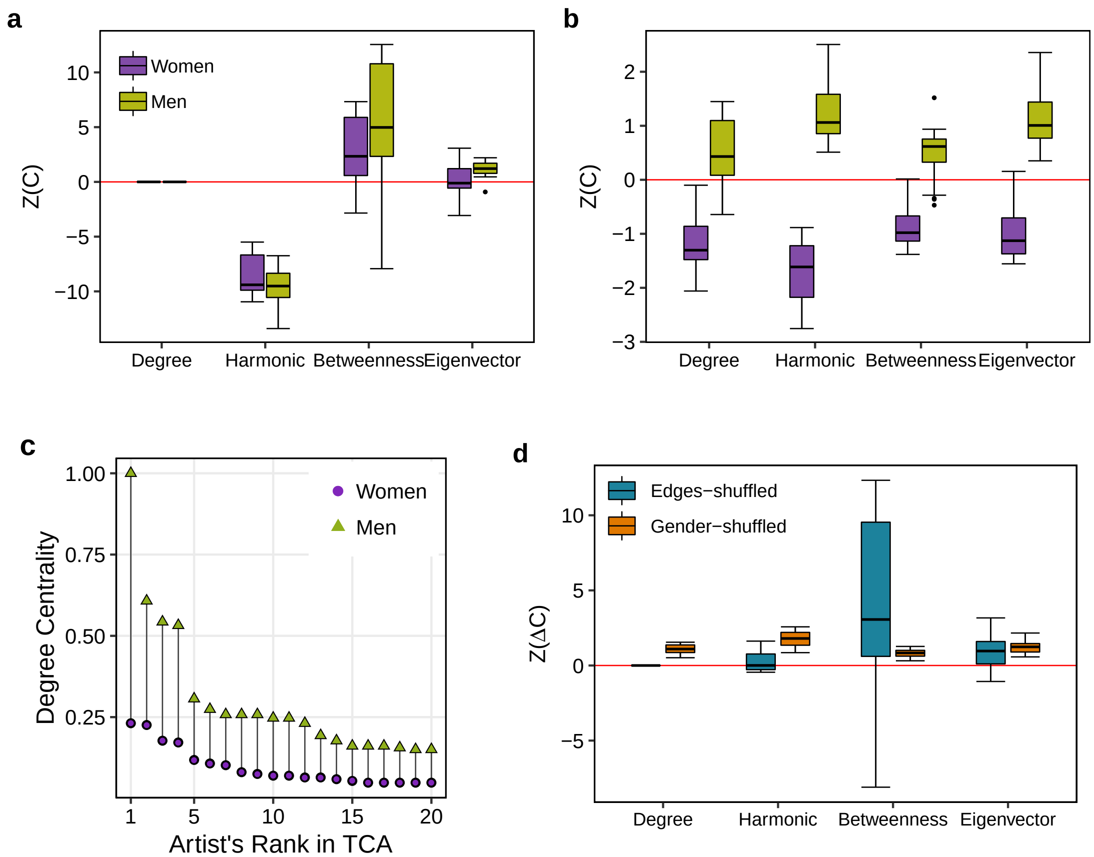


Figure 2.5: **Comparison of centralities for artists in $\text{TCA}_{\text{Women}}$ and TCA_{Men} with null models.** These figures are for ROH's collaboration network. Z-score of centralities ($Z(C)$) is measured with **a** Edge-shuffled model and **b** Gender-shuffled model. The red line corresponds to the theoretical mean obtained from the the randomized models indicating no difference. **c** The gender gap in the degree centrality (normalized) of TCA is displayed. TCA_{Men} artists have higher degree centrality than their women counterparts. **d** $Z(\Delta C)$ of empirical networks.

applied for the rest of centralities with slight variations, confirming men's positional superiority across companies (see Figures 2.5d, and S6). In terms of Z-score, all empirical ΔCs are several standard deviations away compared to the randomized models. It illustrates again that the observed men's high centralities are less likely observed when the gender preference (Edge-shuffled) and gendered productivity correlation (Gender-shuffled) are destroyed.

2.3.3 Perception error on the frequency of women artists

One's view on a frequency of an attribute can be distorted by one's local environment [20, 19]. Considering the previous observations on gendered differences in ballet collaborations, the imbalanced gender collaborations could affect the perceived frequency on women artists. The perception error is defined as the fraction of the observed women frequency in an artist's local collaboration network over the actual fraction of women in the global network (see Methods). That is, the perception error denotes a relative difference of perceived women artist's frequency based on their local collaboration environments and the actual women artists' frequency. From the individual-level perception error B_i, a gender group-level bias $\bar{B}$ compares each gender group's average perception error. 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, and Fig. 2.6).

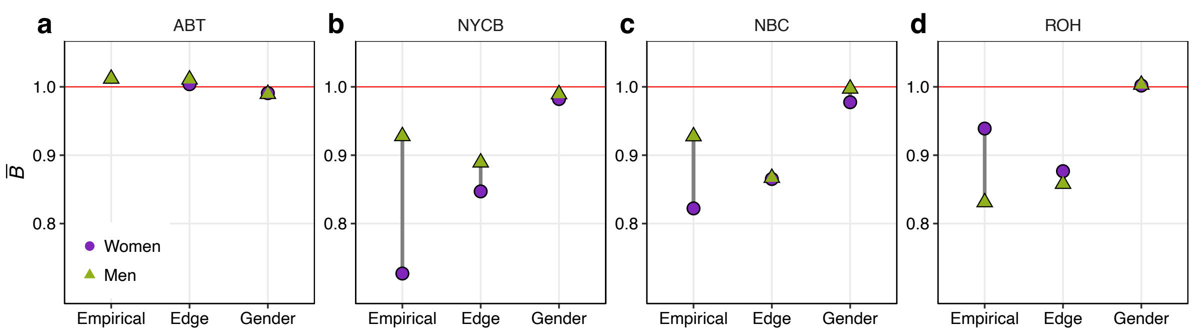


Figure 2.6: **Average perception error $\bar{B}$ for ballet companies.** Average perception error by gender in each company, compared with those from randomized models. **a** ABT, **b** NYCB, **c** NBC, **d** ROH. Line segment in gray guides the difference in perception error by gender group. Most gender groups misconceived the real fraction of women in a network, while the difference is reduced in Edge-shuffled model. The perceived women frequency is more accurate in the gender-shuffled model.

In the empirical collaboration networks of the ABT, NYCB, NBC, and ROH, the women (men) artists' homophily is 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). Based on the measures, one can see that ABT has relatively gender-mixed environment, and as a result both gender groups have relatively accurate perception on the global fraction of women artists (Fig. 2.6a). Whereas, the rest of the companies demonstrate a wide difference in the perception error by gender (Fig. 2.6b–d). For instance, NYCB's men group underestimates women artists about 7%, but their women group underestimates themselves about 27%, showing 20% perception gap of between the groups. It is 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). It leads the perceived underestimation of women artists by themselves. In ROH, women artists have higher estimate of women artists than that of the men artists because 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 since men artists collaborate mostly with men artists (man-man homophily 0.63, man-woman heterophily 0.37), and the assortative collaboration brings the perception gap between gender groups.

Note that Edge-shuffled model displays a reduction in perception gap between gender groups, even though the reduction is limited (Fig. 2.6). This diminution of the perception error suggests that an existence of gender-preferred collaborations can amplify the perception gap between gender groups. Moreover, Gender-shuffled model not only sensibly reduces the perception gap between gender groups, but also achieves a nearly accurate perception on the fraction of women artists. Lowering an extent of imbalanced productivity and gendered preference altogether boosts the representation of women artists, even in the situation of numerically small number of women artists.

2.4 Discussion

Gender imbalance has been investigated for different occupations with the approach of the gender gap in salary and labor force composition [36]. In this context, we find that the numerical composition of women artists in ballet creations is about 18–22%, which is lower than the reported 25% for choreographic leads [6]. These values are far below the gender-neutral ratio (0.5, see Table 2.3) and becomes steeper in highly representative artists (in this study, TCA groups). Not like previously highlighted result, we find that women artists are underrepresented in the overall collaborative structure and artistic roles, not only in choreographic positions.

In addition to the numerical imbalance, we confirm that gendered collaboration structures could potentially aggravate gender imbalances in ballet creations. Crucial roles of individual's social network are associated with the access to important information in scientific and artistic collaborations [37, 10, 12, 38, 11, 15]. Thus, one can infer that gendered collaborations could impact artists' professional experiences, their visibility, and the social perceptions about a gender group. The comparison with randomized network models gives a hint that the observed gender imbalances in terms of central positions and visibility could be explained by systematic inequalities in collaborations rather than random factors.

The social network structures, such as collaborations patterns, play an important role shaping successful careers [39, 40]. Some studies show that men and women utilize different social network structures and behavioral patterns that influence their placement in the job market [16, 41, 42]. Others also show that the formation of one's network and social behaviors over time are related to reinforced perception errors [43, 44]. Taken together, these studies suggest the existence of a permanent feedback for the formation of social relationships, social perception errors, and collaboration patterns, which in return, can influence one's career decisions. For women in ballet, a feedback based on a low representation in a men-dominated environment could negatively impact their decision to undertake a career as ballet creator or engaging in multiple collaborative projects. A future study in this line would find an evidence why women in ballet experience 'glass barrier' in the field.

In addition, the collaboration structure can be crucial for teams [45] and individual performance [46] in terms of creativity and success [47, 11, 48]. A study demonstrates that diversity can improve creative performance [49], and emphasizes the participation of women in collaborative environments because they increase the social sensitivity of the group, making it collectively more intelligent and proficient [50]. In view of this, new policies for more gender-neutral collaborations and more inclusion for women as leading creators should be considered in ballet companies. The more diverse inclusion will boost creative innovation that ultimately benefits the artistic community in general.

At an individual level, perception errors derived from the network structure in the workplace can affect career decisions. In specific, our results show that most companies experience perception errors on the fraction of women artists among group members (Fig. 2.6). The constant underrepresentation of women could reinforce their low visibility, resulting in discourage of women artists in their professional career. This interplay among working environment, perceived possibility of professional career, and their decision could be a pivotal issue to alleviate the gender imbalance in a field.

However, our measure of perception errors is a mathematical approach and can be improved, as there are multiple factors constraining the dependency of these with the local network structure. That is to say, a local network can be described not only by its structure, but also to its embedded social mechanisms, like relationships formed over time, formal and informal norms [51, 52, 53], and individual cognitive processes [54, 55]. Yet our study helps understand the network effects on gender imbalances at the structural level.

Overall, our results help understand another dimension of gender imbalances in the ballet industry. Yet we are aware of the limitations of this work. Our data depends on the archival of the selected ballet companies, which may not be sufficient to generalize the current results to entire ballet industry. To overcome this, more comprehensive digitized data collections would be needed. Also, with the implementation of computational methods, such as deep learning and network science, it has been possible to objectively measure the impact of individual performance in creative domains [56, 57], and the methods can open the possibility for future research on the relationship between gender, network centrality, and actual ballet creators' impact in the field.

In summary, our research shows that the collaborative structure plays an important role for the development of gender representation disparities in the ballet industry. This investigation can be extended to explore dynamic network factors shaping gender imbalances to propose possible and more adequate interventions for the diversity, equity, and inclusion in cultural organizations.

We hope that this work brings awareness on how social phenomena in creative domains can be systematically studied with network science and data driven methods.

Acknowledgements

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). Y.H.-G. acknowledges the Centro de Investigación en Complejidad at Universidad del Desarrollo, Chile, for the financial support to conduct this research.

References

[1] Ann Daly. "The Balanchine woman: Of hummingbirds and channel swimmers". In: *The drama review: TDR* 31.1 (1987), pp. 8–21.

[2] Jennifer Homans. *Apollo's angels: A history of ballet.* 1st edition. Random House Publishing Group, 2010.

[3] Lisa DeFrank-Cole and Renee K Nicholson. "The slow-changing face of leadership in ballet: an interdisciplinary approach to analysing women's roles". In: *Leadership and the Humanities* 4.2 (2016), pp. 73–91.

[4] Kim Elsesser. *A Gender Gap In Ballet, Seriously?* Sept. 2019. URL: https://www.forbes.com/sites/kimelsesser/2019/09/12/a-gender-gap-in-ballet-seriously/.

[5] Elizabeth Yntema. *The ballet world is still male-dominated, research shows.* Aug. 2019. URL: https://www.womensmediacenter.com/news-features/the-ballet-world-is-still-male-dominated-research-shows.

[6] Elizabeth Yntema. *Dance Data Project First Look Report 2022-2023.* 2022. URL: www.dancedataproject.com.

[7] US Census Bureau. 2021. URL: https://www.census.gov/data.html.

[8] David A Cotter et al. "The glass ceiling effect". In: *Social forces* 80.2 (2001), pp. 655–681.

[9] Christine L Williams. "The glass escalator: Hidden advantages for men in non traditional occupations". In: *Social Problems* 39 (1992), pp. 253–267.

[10] Raj Kumar Pan and Jari Saramäki. "The strength of strong ties in scientific collaboration networks". In: *EPL (Europhysics Letters)* 97.1 (2012), p. 18007.

[11] Brian Uzzi and Jarrett Spiro. "Collaboration and Creativity: The Small World Problem". en. In: *American Journal of Sociology* 111.2 (Sept. 2005), pp. 447–504. ISSN: 0002-9602, 1537-5390. DOI: 10.1086/432782. URL: http://www.journals.uchicago.edu/doi/10.1086/432782 (visited on 04/02/2019).

[12] Alireza Abbasi, Kon Shing Kenneth Chung, and Liaquat Hossain. “Egocentric analysis of co-authorship network structure, position and performance”. In: *Information Processing & Management* 48.4 (2012), pp. 671–679.

[13] Sándor Juhász, Gergő Tóth, and Balázs Lengyel. “Brokering the core and the periphery: Creative success and collaboration networks in the film industry”. In: *Plos one* 15.2 (2020), e0229436.

[14] Marıa Bordons and Isabel Gómez. “Collaboration networks in science”. In: *The web of knowledge: A festschrift in honor of Eugene Garfield* (2000), pp. 197–213.

[15] Austin J Parish, Kevin W Boyack, and John PA Ioannidis. “Dynamics of co-authorship and productivity across different fields of scientific research”. In: *PloS one* 13.1 (2018), e0189742.

[16] Yang Yang, Nitesh V Chawla, and Brian Uzzi. “A network’s gender composition and communication pattern predict women’s leadership success”. In: *Proceedings of the National Academy of Sciences* 116.6 (2019), pp. 2033–2038.

[17] Fariba Karimi et al. “Homophily influences ranking of minorities in social networks”. In: *Scientific reports* 8.1 (2018), pp. 1–12.

[18] Leon Festinger. “A theory of social comparison processes”. In: *Human relations* 7.2 (1954), pp. 117–140.

[19] Kristina Lerman, Xiaoran Yan, and Xin-Zeng Wu. “The" Majority Illusion" in Social Networks”. In: *PLoSO* 11.2 (2016), e0147617.

[20] Eun Lee et al. “Homophily and minority-group size explain perception biases in social networks”. In: *Nature human behaviour* 3.10 (2019), pp. 1078–1087.

[21] Inc. Ballet Theatre Foundation. *ABT Ballet Archive*. URL: https://www.abt.org/explore/learn/repertory-archive/ballets/.

[22] Inc. New York City Ballet. *NYCB The Repertory*. URL: https://www.nycballet.com/discover/ballet-repertory.

[23] The National Ballet of Canada Archives. *NBC Repertoire List*. URL: https://national.ballet.ca/Tickets/Archives/Repertoire-List.

[24] Royal Opera House Collections. *ROH Collections Online*. URL: http://rohcollections.org.uk.

[25] Wil MP Van der Aalst, Martin Bichler, and Armin Heinzl. “Robotic process automation”. In: (2018).

[26] Duncan J Watts and Steven H Strogatz. “Collective dynamics of ‘small-world’networks”. In: *nature* 393.6684 (1998), pp. 440–442.

[27] Fei Chen et al. “The average path length of scale free networks”. In: *Communications in Nonlinear Science and numerical simulation* 13.7 (2008), pp. 1405–1410.

[28] Mark D Humphries and Kevin Gurney. “Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence”. In: *PloS one* 3.4 (2008), e0002051.

[29] Fariba Karimi et al. “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*. 2016, pp. 53–54.

[30] Cameron Blevins and Lincoln Mullen. “Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction.” In: *DHQ: Digital Humanities Quarterly* 9.3 (2015).

[31] Linton C Freeman. “Centrality in Social Networks: I. Conceptual Clarification”. In: *Social networks* 1.3 (1978), pp. 215–239.

[32] Paolo Boldi and Sebastiano Vigna. "Axioms for centrality". In: *Internet Mathematics* 10.3-4 (2014), pp. 222–262.

[33] Linton C Freeman. "A set of measures of centrality based on betweenness". In: *Sociometry* (1977), pp. 35–41.

[34] Phillip Bonacich. "Some unique properties of eigenvector centrality". In: *Social networks* 29.4 (2007), pp. 555–564.

[35] Cesar Deloitte; Datawheel; Hidalgo. 2021. URL: https://datausa.io.

[36] Gary S Becker. "Human capital, effort, and the sexual division of labor". In: *Journal of labor economics* 3.1, Part 2 (1985), S33–S58.

[37] R Duncan Luce. "Connectivity and generalized cliques in sociometric group structure". In: *Psychometrika* 15.2 (1950), pp. 169–190.

[38] Mohsen Jamali and Hassan Abolhassani. "Different aspects of social network analysis". In: *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*. IEEE. 2006, pp. 66–72.

[39] Samuel P. Fraiberger et al. "Quantifying reputation and success in art". en. In: *Science* 362.6416 (Nov. 2018), pp. 825–829. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aau7224. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aau7224 (visited on 01/28/2019).

[40] Milán Janosov, Federico Battiston, and Roberta Sinatra. "Success and luck in creative careers". In: *EPJ Data Science* 9.1 (2020), p. 9.

[41] Orsolya Vasarhelyi and Balazs Vedres. "Gender Typicality of Behavior Predicts Success on Creative Platforms". In: *arXiv preprint arXiv:2103.01093* (2021).

[42] Jasmine Tata and Sameer Prasad. "Social capital, collaborative exchange and microenterprise performance: The role of gender". In: *International Journal of entrepreneurship and small business* 5.3-4 (2008), pp. 373–388.

[43] Matthew O Jackson. "Networks in the understanding of economic behaviors". In: *Journal of Economic Perspectives* 28.4 (2014), pp. 3–22.

[44] Güneş Ertan, Michael D Siciliano, and Deniz Yenigün. "Perception accuracy, biases and path dependency in longitudinal social networks". In: *PloS one* 14.6 (2019), e0218607.

[45] S Gayle Baugh and George B Graen. "Effects of team gender and racial composition on perceptions of team performance in cross-functional teams". In: *Group & Organization Management* 22.3 (1997), pp. 366–383.

[46] Leonard Karakowsky, Kenneth McBey, and You-Ta Chuang. "Perceptions of team performance". In: *Journal of Managerial Psychology* (2004).

[47] Janine Nahapiet and Sumantra Ghoshal. "Social capital, intellectual capital, and the organizational advantage". In: *Academy of management review* 23.2 (1998), pp. 242–266.

[48] Yang Yang et al. "Gender-diverse teams produce more novel and higher-impact scientific ideas". In: *Proceedings of the National Academy of Sciences* 119.36 (2022), e2200841119. DOI: 10.1073/pnas.2200841119. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2200841119. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2200841119.

[49] Barton H Hamilton, Jack A Nickerson, and Hideo Owan. "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation". In: *Journal of political Economy* 111.3 (2003), pp. 465–497.

[50] Anita Williams Woolley et al. "Evidence for a collective intelligence factor in the performance of human groups". In: *science* 330.6004 (2010), pp. 686–688.

[51] Mark S Granovetter. "The strength of weak ties". In: *American journal of sociology* 78.6 (1973), pp. 1360–1380.

[52] Mark Granovetter. "Economic action and social structure: The problem of embeddedness". In: *American journal of sociology* 91.3 (1985), pp. 481–510.

[53] James S Coleman. *Foundations of social theory*. Harvard university press, 1994.

[54] Paul F Lazarsfeld, Robert K Merton, et al. "Friendship as a social process: A substantive and methodological analysis". In: *Freedom and control in modern society* 18.1 (1954), pp. 18–66.

[55] Thorsten Pachur, Ralph Hertwig, and Jörg Rieskamp. "Intuitive judgments of social statistics: How exhaustive does sampling need to be?" In: *Journal of Experimental Social Psychology* 49.6 (2013), pp. 1059–1077.

[56] Lu Liu et al. "Hot streaks in artistic, cultural, and scientific careers". In: *Nature* 559.7714 (2018), p. 396.

[57] Lu Liu et al. "Understanding the onset of hot streaks across artistic, cultural, and scientific careers". In: *Nature communications* 12.1 (2021), pp. 1–10.

Chapter 3

Quantifying Hierarchy and Prestige in US Ballet Academies as Social Predictors of Career Success

In the recent decade, we have seen major progress in quantifying the behaviors and the impact of scientists, resulting in a quantitative toolset capable of monitoring and predicting the career patterns of the profession. It is unclear, however, if this toolset applies to other creative domains beyond the sciences. In particular, while performance in the arts has long been difficult to quantify objectively, research suggests that professional networks and prestige of affiliations play a similar role to those observed in science, hence they can reveal patterns underlying successful careers. To test this hypothesis, here we focus on ballet, as it allows us to investigate in a quantitative fashion the interplay of individual performance, institutional prestige, and network effects. We analyze data on competition outcomes from 6,363 ballet students affiliated with 1,603 schools in the United States, who participated in the Youth America Grand Prix (YAGP) between 2000 and 2021. Through multiple logit models and matching experiments, we provide evidence that schools' strategic network position bridging between communities captures social prestige and predicts the placement of students into jobs in ballet companies. This work reveals the importance of institutional prestige on career success in ballet and showcases the potential of network science approaches to provide quantitative viewpoints for the professional development of careers beyond science.

Key words: social prestige, network centrality, career success, ballet competition, performing arts.

3.1 Introduction

Quantifying the processes and behaviors through which some individuals attain success in creative careers is challenging due to multiple factors, including the subjective valuation of creative performance, the multifaceted ways in which success can become manifest through recognition, and data scarcity [1]. However, the recent proliferation of large digital databases capturing many aspects of scientific careers has fueled advances in data-driven methodological tools to capture career and collaboration patterns, productivity, and impact in science [2]. For instance, the field of *science of science* [2] has unveiled the random impact rule governing the timing of a researchers' most consequential publication [3], how authorship team composition influences productivity patterns [4], and the influence of hierarchy on the faculty job market [5], to name a few. The extension of these methods from science to other creative domains has similarly empowered insights on the timing of creative works [6, 7], the important role of brokerage in collaboration networks [8], and the interplay between rich-get-richer mechanism and professional networks driving the allocation of acting jobs [9]. In particular, quantitative tools from network science have helped to illuminate how social networks influence success in the visual arts, movie, music, and book industries [10, 11]. Yet, due in large part to the subjectivity in the valuation of performance quality, many difficulties remain in quantitatively disentangling the contributions of individual performance and collective patterns of social influence on success in creative professions.

To address the relationship between individual performance, institutional prestige, and career outcomes, here we focus on *ballet*, an influential performing art with a long history dating back to the 17th century, when it was promoted by King Louis XIV as a display of the elegance, power, and perfection of human beings [12]. During this period, the access of dancers to the royal court required artistic talent and physical abilities, yet it was significantly facilitated by their membership in a guild or by access to the king, who could grant the privilege by royal decree. While in the modern era, success in ballet is highly dependent on athletic skills, physical attributes, and specific personality traits [13, 14, 15], success may still be influenced by subjective perceptions of quality [16]. In particular, the social system in ballet, consisting of the network of relationships and hierarchies between dancers and all members of the ballet community, is still driven by access to prestigious institutions and professional connections that could ultimately play an important role in shaping dancers' careers. However, there is a lack of systematic research to quantify the effects of social network connections and prestige on the career success in ballet. Hence, by investigating the social drivers of success in ballet, we do not only inform an important part of our cultural heritage, but also can directly test the tools of *science of science* in another creative domain.

Our research delves deep into the complex world of the ballet academic system and its relationship with social prestige and career success. While awards and high achievement are undoubtedly crucial in attaining social recognition [17, 18, 19], we propose the use of network centrality as a more precise indicator of social prestige, as it underscores the critical role of social connections in enhancing prestige [20]. We hypothesize that the prestige of a school facilitates the professional development and job placement of its students, which ultimately elevates the school's external prestige measured by the number of professional dancers they produce, something that has been observed in other creative fields [21, 22, 10, 23]. Thus, dancers may leverage this principle by affiliating with prestigious ballet academies that provide access to a larger network of dance professionals promoting talented dancers.

As a proxy for dance performance, we use competition outcomes of over 6,000 young dancers competing at the Youth America Grand Prix (YAGP; for more details see SI, Section B.1) from 2000-2021. The YAGP competition system filters the participants to the most promising dancers, hence providing a unique opportunity to capture the desired technical and artistic attributes in the ballet market based on jury assessment. The YAGP awards competition medals (gold, silver, and bronze) based on technical and artistic proficiency; and the *Grand Prix*, an award based on the subjective appreciation of the jury. Although multiple biases in performing arts competitions are possible [24], medals and awards have been long used as an objective metric of performance in different domains [25, 26, 27, 28]. Thus, the YAGP competition outcomes represent an objective instrument derived from an efficient system of expert's opinion evaluating ballet performance.

Using the YAGP data, we build the network of ballet academies from their students participation in the competition and create a ranking of ballet academies by their betweenness centrality, which functions as a validated network-based indicator of prestige. Next, we align students' competition outcomes with the academic ranking of their affiliations to predict the job placement of ballet students. Overall, our analysis unveils the ballet pre-professional landscape by underscoring the critical role of school prestige in the selection of dancers, even at an equal proficiency level of performance.

Ultimately, our research broadens the scope of the *science of science* methodologies to the performing arts, empowering us to identify the impact of institutions on the young dancer's careers. This research also contributes to understanding the multifaceted influences of social prestige on career success. Within ballet, the quantitative understanding of network influences on dancer success may also inform equitable policies for auditions and affirmative action that can support a fairer evaluation of candidates in the ballet industry and other professional areas where creative performance is essential. To the best of our knowledge, our study is the first attempt to systematically investigate the effect of social drivers on success in ballet, and contributes to the general understanding of the social contexts driving human creativity, that also broaden the understanding of the evolution of performing arts and our cultural heritage.

3.2 Results

3.2.1 Network of ballet schools

The Youth America Grand Prix (YAGP) competition plays a pivotal role in supporting young ballet students by fostering connections with a network of dance professionals and academies of international presence. Our hypothesis is that the systematic positioning of schools as top contenders in the competition establishes a hierarchical prestige within the network of ballet schools. This prestige, derived from competition outcomes, subsequently impacts the social system of the ballet industry, leading to a more systematic distribution of awards, job placements, and resources such as scholarships for attracting talented students.

Theories of social stratification vary in their arguments about the formation of social prestige, yet one dominant theme is achievement, conceptualized as a source of social stratification and hierarchical order [18]. Metrics related to achievement (e.g. citations, awards, fellowships, honorary degrees, grants) have been used to understand the role of prestige for career success in academia [2] and faculty hiring [22]. On the other hand, the implied hierarchical differences among individuals in a social context can be captured by network metrics, like network position or connectedness, which are useful as indirect measures of social prestige [20]. For instance, research on the visual arts has demonstrated that network position objectively captures social prestige and is a good predictor of career success [10]. Grounded in this approach, we construct the network of ballet academies from the YAGP data and create a network-based ranking using each school's centrality, a key contribution of this paper.

In the YAGP co-competition network, schools are represented as nodes, and a link is established between two schools if their students were ranked among the top 12 in the same competition venue. This network comprises 1,603 ballet schools and 55,778 links, providing a comprehensive representation of the ballet academy ecosystem (details on the organization of ballet schools in the U.S. can be found in SI, Section B.2). The co-competition network is constructed from both the multiple regional semi-finals and yearly finals competition stages. Thus, the link between two schools captures that both schools were able to produce top dancers under the same competition setting and reflects a degree of similarity in training quality. Connectivity within the co-competition network thus forms an ordered hierarchy in which schools' high achievement contributes to social prestige which is then directly perceived by others [17, 18, 19]. Specifically, highly connected schools in the co-competition are more likely to repeatedly have top dancers in the competition relative to their less connected counterparts. At the same time, these dancers competed against many different schools, thereby increasing their schools' visibility within the community, as opposed to schools who only competed against the same subset of competitors.

Finally, we capture network effects in the perception of prestige by noting that visibility of ballet schools is further influenced by their potential to bridge between communities in the network. We quantify the bridging capacity, and thus the schools' social prestige, by the betweenness centrality in the co-competition network. Betweenness centrality is computed for each node, k, based on the sum of all-pairs shortest paths which pass through that node:

$$B_k = \sum_{\sigma(a,b) \in K} \frac{\sigma(a,b|k)}{\sigma(a,b)} \tag{3.1}$$

where $\sigma(a,b)$ denotes the number of shortest (a,b)-paths, and $\sigma(a,b|k)$ is all the shortest paths passing through node k [29]. To visually capture the role of betweenness centrality in the network structure, we extract the multi-scale network backbone [30]. This method uses a parameter α for the probability of the existence of an edge and reduces the network to the most fundamental structures and hierarchies based on multi-scale interactions and their relative relevance for the network topology. The resulting network is shown in Figure 3.1, where we observe that most schools only attain regional success captured by their low betweenness and weak connections ($\alpha = 0.4$, in yellow edges). On the other hand, the network's strong edges ($\alpha = 0.01$, red edges) connects 166 nodes (10% from the total network), forming a core backbone of ballet schools in strategic positions to gain national attention and prestige.

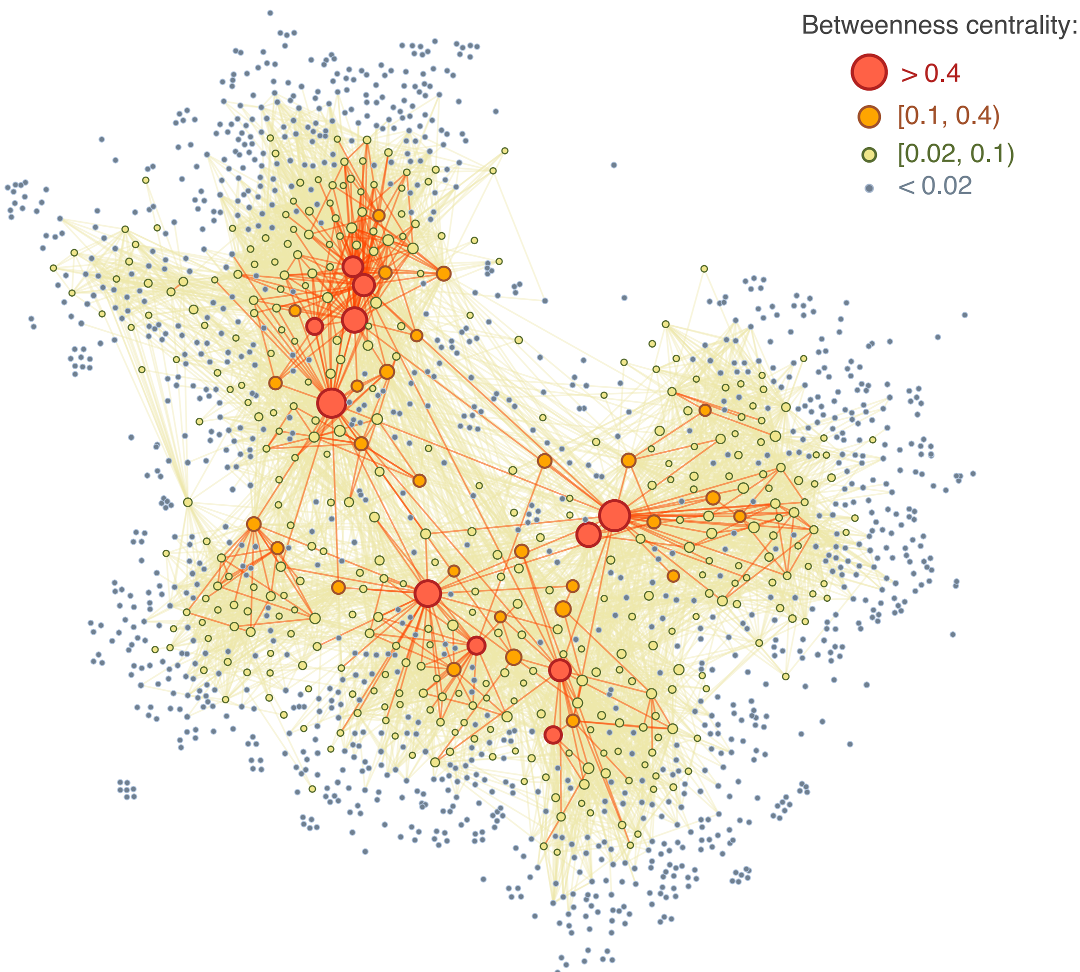


Figure 3.1: **Network of ballet schools.** Each node is a ballet school, and two schools are connected if they obtained a top student in the same competition venue. Node size and color reflect schools' normalized betweenness centrality, B_k. Node position is determined by the force-directed graph layout with force estimation $\theta = 0.5$ that emphasizes the separation of nodes into clusters. The weak structure (in yellow, network backbone with $\alpha = 0.4$) shows dense connectivity within network clusters and sparser connections between clusters and to the periphery; the strong structure (in red, network backbone with $\alpha = 0.01$), comprises 166 nodes (10%) and 384 edges (0.6%). This network representation illustrates schools' hierarchical structure explaining the role of network position for social prestige.

To validate the use of betweenness centrality within the YAGP co-competition network as a proxy for social prestige of ballet schools, we compared the network ranking to a selection of top schools as identified by leading ballet experts. Ballet experts offer comprehensive understanding of dance and the ballet ecosystem, and are widely recognized for their long-standing existence and influence within the dance community in the United States (see SI, Section B.2.3 for further details). Here, we aggregate a list of Top Ballet Schools selected by *Dance Magazine* and the highly regarded blog *A Ballet Education*, in total capturing the top 60 ballet schools in the United States (List shown in SI, Figure S2.3). We then quantify the extent to which the most prestigious schools as ranked by betweenness centrality recovers the experts' opinions using the AUC, or the probability that our measure ranks a school listed in the Top Ballet Schools higher than a school not on that list. The AUC is a score between 0 and 1, where a value closer to 1 indicates a probability of correct classification, while a score close to 0.5 indicates that the model performs no better than random guessing. We find an AUC of 0.75 indicating a fair alignment between betweenness centrality and experts' assessment of the social prestige of ballet schools. Further, betweenness centrality performs better than simpler measures of school prestige and achievement, including the ratio of winning awards or co-competition degree (see SI, Section B.2.3). While achievement is certainly a crucial factor in attaining social recognition, our findings suggest that betweenness centrality offers a more accurate measure of social prestige, as it captures the critical interplay between social connections and prestige.

As a whole, our results provide evidence that key network patterns, such as bridging between communities, is closely related with school's social prestige in the YAGP co-competition network. These findings highlight the utility of network analysis in understanding the relationship between achievement and social prestige.

3.2.2 Career success of ballet dancers

The hiring process for ballet dancers is limited in opportunities and influenced by a variety of factors including training technique, technical mastery, artistic ability, and even demographics. However, we argue that social prestige plays a significant role beyond performance or ability in predicting the success of dancers' careers.

To understand the influence of social prestige on a successful job placement, we align the aggregated competition outcomes from 6393 students within the professional age range to the highlighted jobs reported in the *Success Stories* by the YAGP. In total, 385 (6%) young YAGP alumni received a job placement in a dance company. Surprisingly, 22% of YAGP alumni with a job placement did not receive any award in the competition while 10% won in both the semi-finals and finals. The majority of dancer who received a job placement (58.5%) won at least one award in the semi-finals but did not advance to the finals. This breakdown suggests that there are different routes and factors other than achievement driving the selection of dancers towards a job placement in a ballet company.

To investigate the intertwined effect of individual achievement and social prestige on job placements, we build a logistic regression model to predict which students are placed into a dance company job. Our dependent variable is success S, measured as a binary outcome, where $S_i = 1$ if student i obtained a job placement in a ballet company and $S_i = 0$ otherwise. The independent variables include the aggregated measures of students' achievement within the YAGP competition, such as total awards by type ($Gold_i$, $Silver_i$, $Bronze_i$, and $GrandPrix_i$) and total number of competitions ($Competitions_i$), as well as the normalized and re-scaled schools' betweenness centrality measure for social prestige ($Prestige_k$). To control for potential confounding factors, we also include a control variable of the student's gender ($Gender_i$). Our primary model is specified as follows:

$$Pr(S_i = 1) = \texttt{Logit}^{-1}(\beta_1 Gender_i + \beta_2 Bronze_i + \beta_3 Silver_i + \beta_4 Gold_i + \beta_5 GrandPrix_i + \beta_6 Competitions_i + \beta_7 Prestige_{ki} + \epsilon_i) \tag{3.2}$$

We observe a strong positive effect of prestige on job placement (Model 1 in Table 3.1). Moreover, our analysis reveals a significant increase in the probability of job placement along with schools' increasing prestige (Figure 3.2A). For example, consider two ballet dancers, Lauren and Juliet, who both won one gold medal after two competition appearances, but who attend schools of differing prestige: Lauren attends a school with prestige 0.87 while Juliet's school has a prestige of 0.09. Our logit model predicts that despite their identical competition performance, Lauren's probability of a job placement is 2.25 times higher than for Juliet.

Next, we test for the potential effect of advancing to the competition finals on job placement by adding a dummy variable for being a finalist ($F_i = 1$) or not ($F_i = 0$). In this second model, we observe a strong effect on the probability of a successful job placement (Model 2 in Table 3.1, AUC = 0.7327), which is comparable to the effect of being affiliated to the most prestigious schools (see Figure 3.2B). This comparison suggests that being a finalist can greatly enhance the career prospects of talented students who attend less prestigious schools, and highlights the significant impact of high performance on a job placement.

Our logistic model can also reveal more detailed effects of medals and competitions on job placement. Intuitively, examining medals by type (model 2 in Table 3.1, Figure 3.2B), the odds ratio increases with medal importance: winning a bronze medal increases the odds of a job placement by 30% compared to a no medal baseline, while one additional gold or silver medal increases those odds by about 50%. The greatest impact of awards on a student's odds of attaining a job placement comes from winning the Grand Prix, a special recognition given on the subjective appreciation of the jury, increasing the odds of a successful job placement by 67%. This suggests that the recognition of a dancer by the jury is highly aligned with the value system adopted by ballet companies, much more than winning multiple competition medals awarded on a technical scoring system.

Our analysis also highlights an unexpected finding: a long competition career may negatively impact job placement. On average, students participate in two semi-final competitions regardless of their job placement outcome. However, our analysis shows that each additional semi-final competition decreases the chances of a job placement by 18% (model 2, Figure 3.2B), which indicates that students who participate in multiple competitions may not improve their chances to be recruited.

Overall, we find that school prestige has the largest effect in determining job placement, with the odds increasing by over 200% for students who attend the most prestigious schools, and this effect is robust across all models (see more model variations in SI, Section B.3.2). Yet, while our results also emphasize the importance of high performance as a key factor for career success, these models are unable to disentangle potential interactions between performance and prestige.

Table 3.1: **Model coefficients for the logistic models predicting successful job placement.** Model coefficients labeled by p-value. Standard errors in parentheses.

	Dependent variable = Success S_i	
	(1)	(2)
Prestige	1.201^{***}	1.136^{***}
	(0.190)	(0.193)
Grand Prix	0.809^{***}	0.515^{***}
	(0.113)	(0.127)
Gold	0.45^{***}	0.344^{***}
	(0.091)	(0.093)
Silver	0.45^{***}	0.413^{**}
	(0.091)	(0.091)
Bronze	0.276^{**}	0.258^{**}
	(0.096)	(0.097)
Finalist F_i		1.0^{***}
		(0.184)
Competitions	-0.238^{***}	-0.208^{***}
	(0.058)	(0.058)
Gender: male	0.754^{***}	0.756^{***}
	(0.121)	(0.121)
Constant	-3.135^{***}	-3.193^{***}
	(0.111)	(0.112)
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$	
Observations	6393	6393
McFadden pseudo R^2	0.085	0.095
AIC	2640.3	2615.3
AUC	0.7327	0.7374

To further elucidate the role of school prestige on individual job placement, we conduct an experiment in which we match students who have identical medal and competition counts, but who differ on their school's prestige. Here, the YAGP medal counts function as a proxy for dancer ability, empowering us to measure the influence of prestige beyond performance.

We consider a binary treatment status denoted as $Y_i = 1$ for the students affiliated to a prestigious school, and $Y_i = 0$ for students who attended a less prestigious school. The subset of prestigious schools comprises the top 5% of the network-based ranking of prestige. Under this

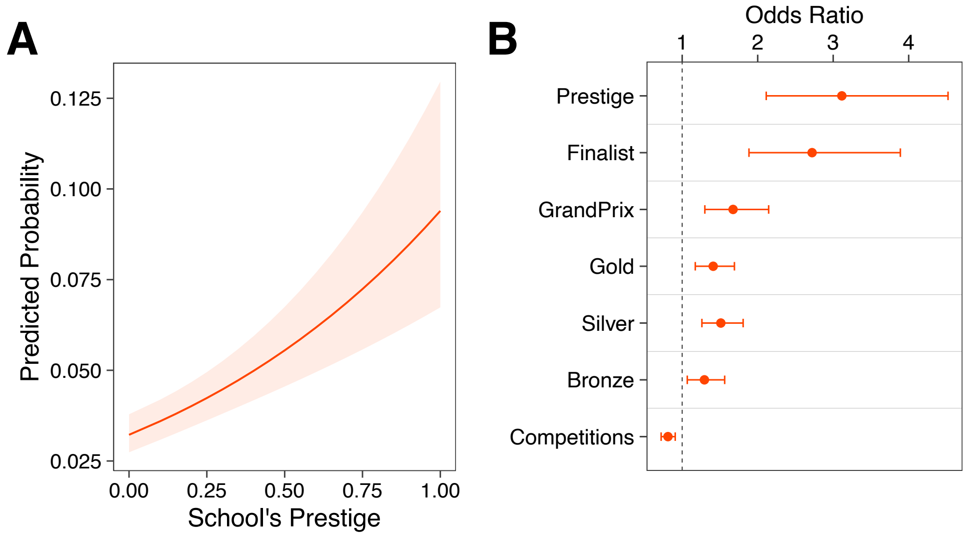


Figure 3.2: **Probability of job success in ballet.** Success is defined in a binary fashion, where $P(S = 1)$ if the student obtained a job placement in a ballet company, and $P(S = 0)$ otherwise. **A** demonstrates a significant positive effect on the predicted probability of a job placement with the increase of school's prestige. **B** shows exponentiated odds ratio with corresponding 95% confidence intervals for the effect on job placement of each additional unit of competition outcomes and institutional prestige. Baseline in 1 indicates no effect. We see that the Grand Prix has the largest effect by type of awards, while long competition trajectories can be detrimental for a job placement, and being a finalist is comparable to be affiliated to ahighly prestigious school. Model coefficients reported in Table 3.1.

criteria, we assign 93 top schools as $Y = 1$, resulting in 2,637 treated students and 4,942 controls. We match the observations with the exact matching method using `MatchIt` [31]. The exact matching is performed over the quantified variables of individual achievement, including: (i) total number of each competition medal (gold, silver, and bronze medals); and (ii) the total number of competitions, both listed only in the semi-finals. Finally, the matching model can be described as:

$$E(S|Y = 1, X) - E(S|Y = 0, X) = \frac{1}{N}\sum_{i=1}^{N}(S_i - S_{j(i)}) \tag{3.3}$$

where S is the job placement outcome, Y is the treatment indicator, X contains the vector of covariates used for exact matching, N is the number of subclasses formed in the matching process, and $j(i)$ is the number of controls used to match the treated observation (i). We compute the matching method specifying the Average Treatment Effect as the estimand and heteroscedasticconsistent standard errors based on subclasses [32].

We observe that, by comparing equally skilled dancers, i.e. students who have exactly the same competition outcomes, there is a significant increase of 65% in the odds of obtaining a job placement for those who attended a prestigious school. These findings indicate that even though dancers can obtain a similar number and type of competition medals, an indicator of similar ability and performance, their affiliations play a crucial role in their careers, which ultimately influences their professional positioning in a ballet company.

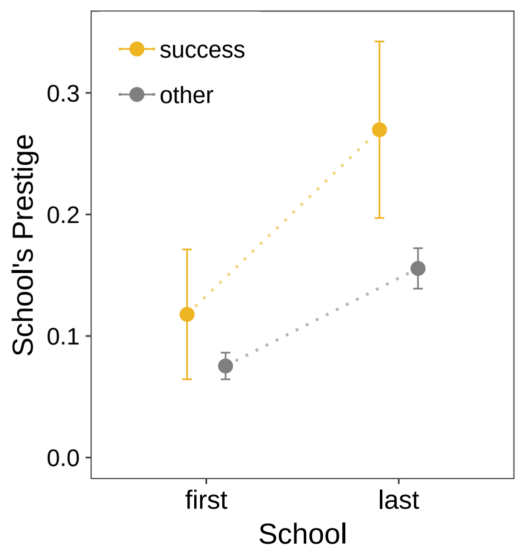


Figure 3.3: **Change in school's prestige for transfer students.** The change in school's prestige (mean with 95% confidence intervals) from students who attend more than one ballet school during their participation in the competition. We observe that students move to more prestigious schools, and students with a successful job placement (labeled as 'success', in yellow) move to even more prestigious schools compared to those without a job placement.

Given the positive impact of school prestige on job placement, we further investigate if students who change schools move to a more prestigious school. While 85% of all participants only reported one school affiliation, the remaining 15% (932 students) attended from two up to five schools. From the students who changed schools, 85 students (10%) obtained a successful job placement. To capture the difference in schools' prestige, we first measure the change in prestige from students' first and last schools. Then, pairing each student by their first and last school, we find a difference of 0.086 in the schools' prestige between the last and first school (last:

mean=0.165, sd=0.258, first: mean=0.079, sd=0.17), indicating that students tend to move to a more prestigious school ($t = 8.49, p < 0.001$). We further examine the school change by comparing the first and last schools by students' job placement outcome. We measure the difference in schools' prestige with a two-way ANOVA test revealing that the change in schools (first/last) shows an increase in school's prestige in general ($F = 73.725, df = 1, 1860, p < 0.001$). Moreover, we observe a larger increase in school's prestige for students who obtained a successful job placement ($F = 19.907, df = 1, 1860, p = p < 0.001$). The interaction effect was also significant ($F = 4.16, df = 1, 1860, p = 0.0414$). The difference in the change of schools' prestige by each group can be seen in Figure 3.3.

Overall, our results emphasize the importance of social prestige for a successful job placement in ballet and show that students may have access to more prestigious institutions over time.

3.3 Summary and Discussions

In summary, our research highlights the usefulness of the *science of science* methods as an efficient tool to quantify career patterns in creative professions that were not possible to elucidate before. The joint use of *science of science* and network science allowed the identification of the leading ballet academies in the US. This contributes to expand the general understanding of the arts academic system in the US and its relationship with reputation and prestige [33]. Moreover, our work also demonstrates that features of artistic careers can be quantified, and emphasizes previous efforts of researchers investigating the different factors driving the evolution of the arts in an objective fashion [27].

Our work unveils the importance of both individual competition performance and schools' prestige as predictors of successful job placements in ballet. By systematically measuring schools' prestige through network analysis, we demonstrate that social prestige is predictive of higher jury's recognition of students, competition advancement, and better career prospects. As a whole, we show that the social network remains essential to shape success in ballet's modern era, and illustrate the potential of data-driven methods to objectively analyze these effects in performing arts.

The pursuit of a successful career in ballet often requires young dancers to give up their childhood, as demanding training regimes are essential to attain the level of athleticism and motor control necessary to execute complex, yet artistic, movements and sequences. Despite the rigorous physical preparation, the history of ballet suggests that the selection and advancement of dancers is influenced by more than just performance ability, and is strongly shaped by the prestige of social and professional connections. In the modern era, dancers can leverage this principle by affiliating with prestigious academies that provide access to the network of experts who play a critical role in identifying and promoting rising stars.

Through our examination of the network of ballet academies in the United States, we provide a network-based ranking of these academies, and reveal the hierarchical social stratification of prestige within the ballet academic environment. This network-based measure of prestige in ballet complements similar measures of prestige in academic careers [21, 22, 23], visual arts [10], and the movie and music industry [11].

The ballet industry is renowned for its limited job opportunities and high competitiveness. Our research shows that ballet companies often exhibit selection biases based on the dancer's affiliations. This is a common issue in competitive settings, where evaluators find it challenging to differentiate between similarly talented candidates [16, 34]. In such cases, evaluators tend to make their selections based on social cues, such as the prestige of affiliation, and personal biases. To counteract these biases, an adequate implementation of blind auditions and affirmative action policies could increase fairness in the selection of candidates [35]. However, the relationship between affiliation prestige and dance ability is complex, as the two variables may influence each other. For instance, a prestigious institution may attract high-quality dancers, which in turn reinforces the dancers' prestige; while a high-quality dancer may also enhance the prestige of their affiliation. It is important to consider the possibility of reverse causality when analyzing the intertwined effect of these two variables for the selection of ballet dancers. In further research, we aim to explore the relationship between performance quality, school choice, and local network dynamics for the career success of dancers.

Several limitations of our research should be taken into consideration. First of all, our data is limited to the YAGP competition outcomes and is unable to capture unobserved variables related to certain standards of beauty, behavior, and technical performance that ballet companies consider in their hiring process. We are also unable to capture the influence of other YAGP rewards including scholarships for summer or yearly training in prestigious academies. Although our findings provide insights into the potential benefits of affiliation with a prestigious school for career success in the ballet industry, it only reflects the hiring process for YAGP participants and may not be representative of the entire population of young ballet dancers. Also, there are several other U.S.-based and international competition substitutes to the YAGP, like the World Ballet Competition or the Prix de Laussane, which could similarly influence student career outcomes.

While our measure of success currently focuses on job placements as company dancers, it is important to recognize that a successful career in ballet can encompass a variety of roles, including teaching, choreography, and administrative duties. Therefore, there is a need for more comprehensive data to investigate the career paths of ballet dancers, from pre-professional to professional levels, allowing our definition of success to include diverse career paths. With richer data, we can disentangle the causality between school choice and career success, to accurately capture dancers' achievements and map the complex relationship between institutional prestige and success.

Lastly, we hope that this work raises the attention of how important is school choice for the dancers' future and how interdisciplinary research contributes to the understanding of human creativity at a social level, which can ultimately inform about the underlying mechanisms driving the evolution of arts and our cultural heritage.

3.4 Methods

3.4.1 Dataset

We use data collected with permission from the YAGP online platform [36] using the `BeautifulSoup` Python library for web scraping [37]. The data contains the competition results of 10,686 students and 2,402 schools participating from 2000 to 2021. We subset the data to only include the 6,363 students listed from competition venues within the United States (namely, Youth America Grand Prix) for a robust representation of the competition system. This selection of students comprises those in the professional age range, which filters out students from 'Pre-Competitive Age Division' after 2014, and 'Junior Age Division' after 2019. All students from the 'Senior Age Division' are considered for this analysis. In total, our student population is 7% Pre-Competitive Age Division, 28% Junior Age Division, and 65% Senior Age Division.

To disambiguate the students and schools, we first checked for misspellings and punctuation. We then performed an exact name matching that leverages middle names and/or initials to distinguish identity, for both students and schools' names. The final data contains 6,475 participants, from which 6,393 students are affiliated to any of the 1,603 ballet schools found in the data. We infer the gender of students using the `gender` package for R, a method of binary gender inference (Woman, Man) that matches names with their gender as found in the package standardized databases (`ssa`, `ipums`, `napp`, and `demo`) [38, 39]. This method's estimation uses the probability of finding a gender assigned to a given name; when the probability is larger than or equal to 0.7, then the gender is assigned to the name tested. Only 0.008% of students' gender was not identified and were removed from the dataset. Its important to emphasize that the inferred gender does not refer to the sex or self-assigned gender of dancers, but serves as an estimate of the social construction of gender. Also, students' reported gender can be confirmed in the YAGP website records if necessary. Overall, women represent 83% of the total population, representing a self-selection gender bias embedded in the competition system.

Our data collection and research methods were approved on January 18th, 2023, by the Institutional Research Ethics Committee of Universidad del Desarrollo, in Chile. In addition, the use of the public data resources was authorized by Larissa Savaliev, director of the YAGP.

3.4.2 Measure of social prestige

The network of ballet schools is represented as $G = (K, V)$, where K is the set of schools (nodes, k) and V the set of connections between schools (edges, v). Hence, there is an edge (v) between two schools $k_1, k_2 \in K$ so that $v(k_1, k_2) \in V$ if their affiliated students are listed as top student in the same competition venue and year.

From this network, we compute the unweighted betweenness centrality, following Eq. 3.1 [29]. Betweenness centrality is then normalized with the Min-Max scaling method to have a linear range between $[0, 1]$, where $B_k = 1$ corresponds to the most central school. We then order the schools by their normalized centrality and create a ranking list using a dense rank function, which generates rank ties for observations with the same centrality values. The rank r of a school k by its centrality ($r = \{1, 2, \ldots, K\}$) is assigned in an ascending fashion, so that $r_k = 1$ is the largest centrality value, and $r_k = 945$ will have the lowest centrality in the set of schools K.

Data availability

Data will be uploaded soon, but its available upon request now.

Acknowledgements

We thank Larissa Savaliev for her support to conduct this research. This work was supported by the ANID FONDECYT INICIACIÓN 11200986. Y.H.-G. acknowledges the Research Center in Social Complexity at Universidad del Desarrollo, Chile, for the financial support to conduct this research. A.-L.B. is also supported by the Templeton Foundation, contract #62452, the European Union's Horizon 2020 research and innovation programme under grant agreement No 810115 –

DYNASNET, by The Eric and Wendy Schmidt Fund for Strategic Innovation, Grant G-22-63228, and NSF grant SES-2219575.

References

[1] Albert-Lászlo Barabási. *The Formula: The Universal Laws of Success*. First Edition. Little, Brown, 2018, p. 380.

[2] Dashun Wang and Albert-László Barabási. *The science of science*. Cambridge University Press, 2021.

[3] Roberta Sinatra et al. "Quantifying the evolution of individual scientific impact". In: *Science* 354.6312 (2016), aaf5239. ISSN: 1095-9203. DOI: 10.1126/science.aaf5239.

[4] Hongwei Zheng, Weihua Li, and Dashun Wang. "Expertise diversity of teams predicts originality and long-term impact in science and technology". In: *arXiv preprint arXiv:2210.04422* (2022).

[5] Aaron Clauset, Samuel Arbesman, and Daniel B. Larremore. "Systematic inequality and hierarchy in faculty hiring networks". In: *Science Advances* 1.1 (2015), e1400005. ISSN: 2375-2548. DOI: 10.1126/sciadv.1400005.

[6] Lu Liu et al. "Hot streaks in artistic, cultural, and scientific careers". In: *Nature* 559.7714 (2018), p. 396.

[7] Lu Liu et al. "Understanding the onset of hot streaks across artistic, cultural, and scientific careers". In: *Nature communications* 12.1 (2021), pp. 1–10.

[8] Giuseppe Soda, Pier V. Mannucci, and Ronald S. Burt. "Networks, Creativity, and Time: Staying Creative through Brokerage and Network Rejuvenation". In: *Academy of Management Journal* 64.4 (2021), pp. 1164–1190.

[9] Oliver E Williams, Lucas Lacasa, and Vito Latora. "Quantifying and predicting success in show business". In: *Nature communications* 10.1 (2019), pp. 1–8.

[10] Samuel P. Fraiberger et al. "Quantifying reputation and success in art". en. In: *Science* 362.6416 (Nov. 2018), pp. 825–829. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aau7224. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.aau7224 (visited on 01/28/2019).

[11] Milán Janosov, Federico Battiston, and Roberta Sinatra. "Success and luck in creative careers". In: *EPJ Data Science* 9.1 (2020), p. 9.

[12] Jennifer Homans. *Apollo's angels: A history of ballet*. 1st edition. Random House Publishing Group, 2010.

[13] Moira Cameron McCormack et al. "The physical attributes most required in professional ballet: a Delphi study". In: *Sports medicine international open* 3.01 (2019), E1–E5.

[14] Nathan Kogan. "Careers in the performing arts: A psychological perspective". In: *Communication Research Journal* 14.1 (2002), pp. 1–16.

[15] Marija Jankovic and Stefan Bogaerts. "Predicting success in the performing arts: Ballet and music". In: *Psychology of Music* 49.4 (2021), pp. 945–957.

[16] Warren Thorngate, Robyn M Dawes, and Margaret Foddy. *Judging merit*. Psychology Press, 2010.

[17] Don H Shamblin. "Prestige and the sociology establishment". In: *The American Sociologist* (1970), pp. 154–156.

[18] Bernd Wegener. "Concepts and measurement of prestige". In: *Annual review of sociology* (1992), pp. 253–280.

[19] David Zeitlyn and Daniel W Hook. "Perception, prestige and PageRank". In: *Plos one* 14.5 (2019), e0216783.

[20] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[21] Val Burris. "The academic caste system: Prestige hierarchies in PhD exchange networks". In: *American sociological review* 69.2 (2004), pp. 239–264.

[22] Aaron Clauset, Samuel Arbesman, and Daniel B Larremore. "Systematic inequality and hierarchy in faculty hiring networks". In: *Science advances* 1.1 (2015), e1400005.

[23] K Hunter Wapman et al. "Quantifying hierarchy and dynamics in US faculty hiring and retention". In: *Nature* 610.7930 (2022), pp. 120–127.

[24] Herbert Glejser and Bruno Heyndels. "Efficiency and inefficiency in the ranking in competitions: The case of the Queen Elisabeth Music Contest". In: *Journal of cultural Economics* 25 (2001), pp. 109–129.

[25] Filippo Radicchi. "Universality, limits and predictability of gold-medal performances at the Olympic Games". In: *PloS one* 7.7 (2012), e40335.

[26] Marc Santolini et al. "Team success in the iGEM scientific competition". In: *BOOK OF ABSTRACTS*. 2017, p. 27.

[27] David W Galenson. "Quantifying artistic success: Ranking French painters—and paintings—from Impressionism to Cubism". In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 35.1 (2002), pp. 5–19.

[28] Victor Ginsburgh. "Awards, success and aesthetic quality in the arts". In: *Journal of Economic perspectives* 17.2 (2003), pp. 99–111.

[29] Ulrik Brandes. "A faster algorithm for betweenness centrality". In: *The Journal of Mathematical Sociology* 25.2 (2001), pp. 163–177. DOI: 10.1080/0022250X.2001.9990249. eprint: https://doi.org/10.1080/0022250X.2001.9990249. URL: https://doi.org/10.1080/0022250X.2001.9990249.

[30] M Ángeles Serrano, Marián Boguná, and Alessandro Vespignani. "Extracting the multiscale backbone of complex weighted networks". In: *Proceedings of the national academy of sciences* 106.16 (2009), pp. 6483–6488.

[31] Daniel Ho et al. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference". In: *Journal of Statistical Software* 42.8 (2011), pp. 1–28. DOI: 10.18637/jss.v042.i08. URL: https://www.jstatsoft.org/index.php/jss/article/view/v042i08.

[32] Guido W. Imbens. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review". In: *The Review of Economics and Statistics* 86.1 (Feb. 2004), pp. 4–29. ISSN: 0034-6535. DOI: 10.1162/003465304323023651. eprint: https://direct.mit.edu/rest/article-pdf/86/1/4/1613802/003465304323023651.pdf. URL: https://doi.org/10.1162/003465304323023651.

[33] Alexander Michael Petersen et al. "Reputation and impact in academic careers". In: *Proceedings of the National Academy of Sciences* 111.43 (2014), pp. 15316–15321.

[34] Jing Cao and Lynne Stokes. "Evaluation of wine judge performance through three characteristics: Bias, discrimination, and variation". In: *Journal of Wine Economics* 5.1 (2010), pp. 132–142.

[35] Jasmin Droege. "First impression biases in the performing arts: taste-based discrimination and the value of blind auditioning". In: *Journal of Cultural Economics* (2019), pp. 1–47.

[36] *Youth America Grand Prix: The Largest Student Ballet Scholarship Competition*. 2022. URL: https://yagp.org.

[37] Leonard Richardson. "Beautiful soup documentation". In: (Apr. 2007).

[38] Fariba Karimi et al. "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*. 2016, pp. 53–54.

[39] Cameron Blevins and Lincoln Mullen. "Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction." In: *DHQ: Digital Humanities Quarterly* 9.3 (2015).

Appendix A

Supplementary Material of Structural effect on gender imbalance in ballet collaboration networks

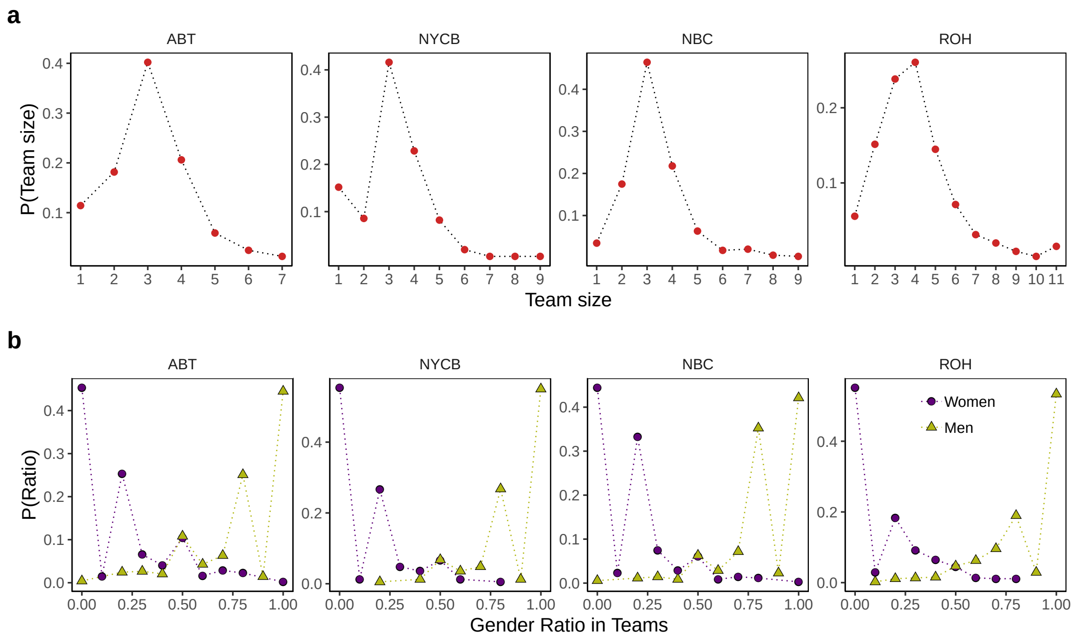


Figure S1.1: **Team composition.** Panel **a** shows the probability of a team size in each ballet company. Panel **b** shows the probability of a gender ratio in a team.

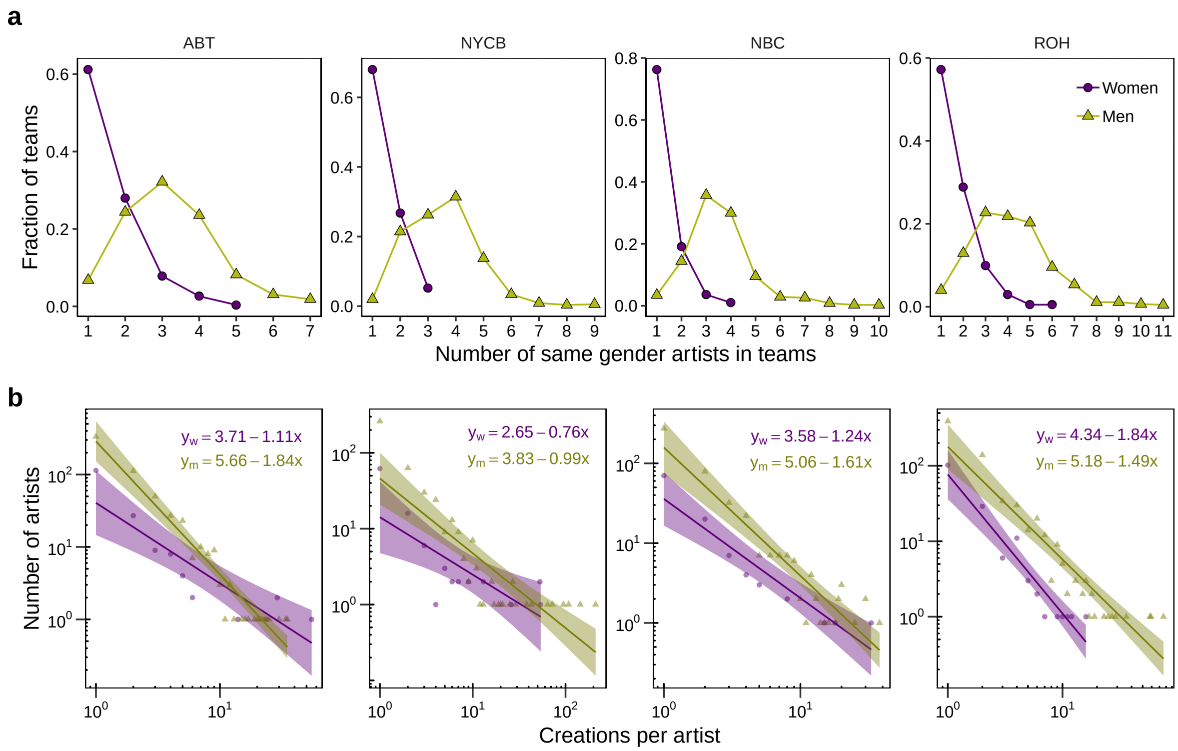


Figure S1.2: **Collaboration patterns by gender.** Purple/green and dots/triangles represent women/men. Panel **a** shows the fraction of teams by the number of same gender artists in their composition. Panel **b** shows the number of ballet creations and the count of artists by gender (artists' productivity).

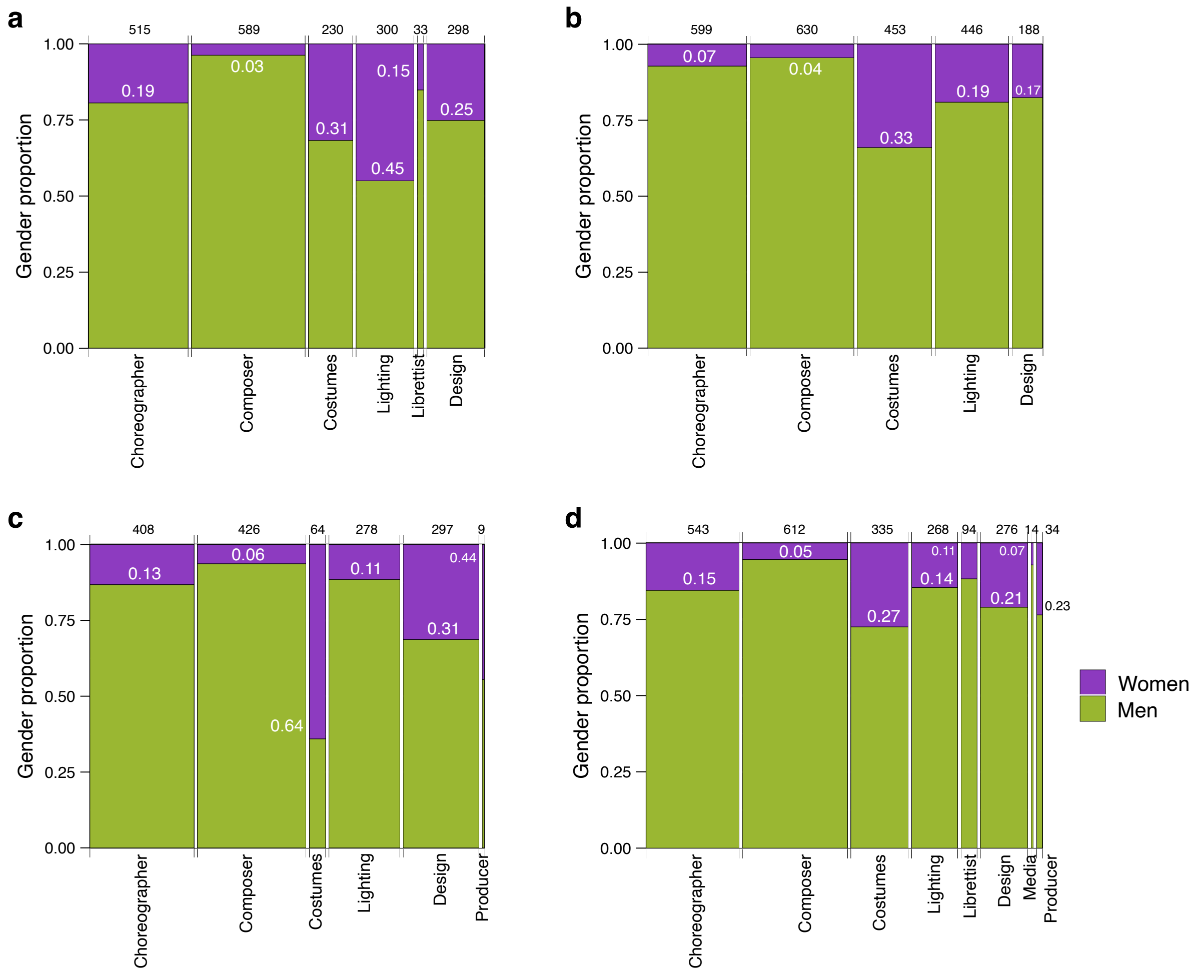
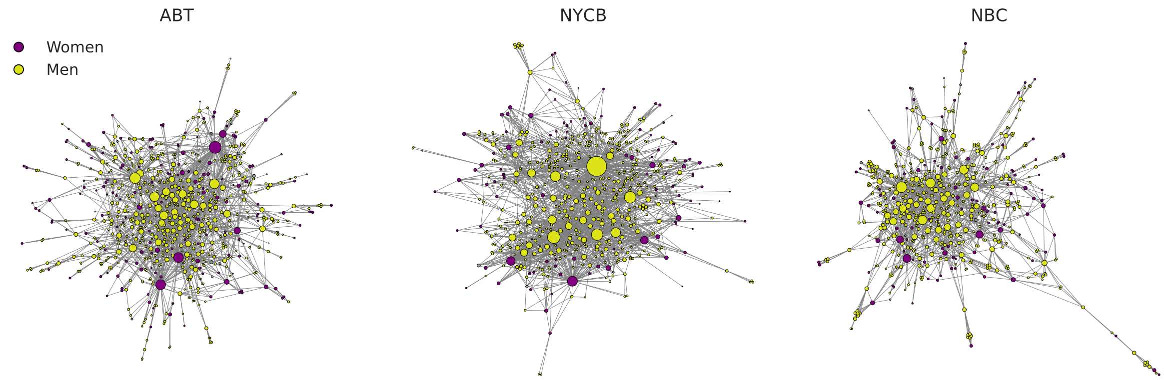
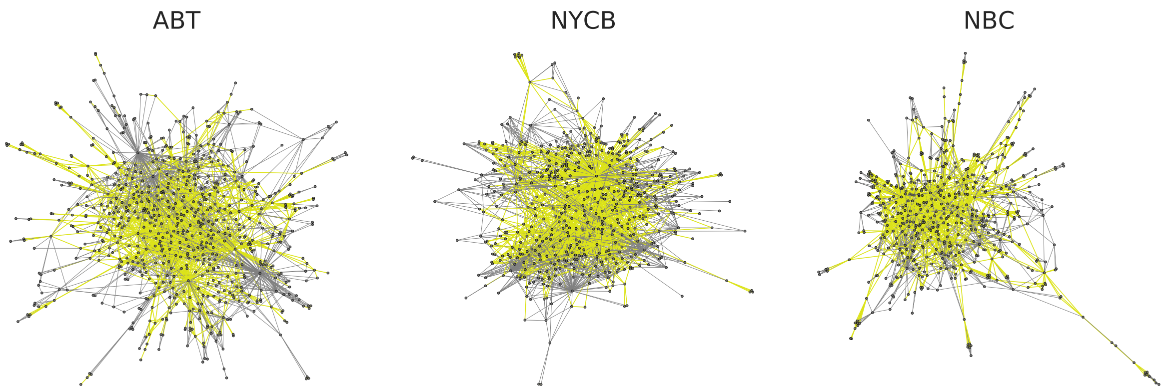


Figure S1.3: **Proportion of women in ballet collaborations by artistic roles.** Proportion of women written inside purple bar. Group size indicated by bar-width with total count of artists written on top of bar. Figures show women representation by artistic roles for **a** ABT, **b** NYCB, **c** NBC, and **d** ROH. While women face less representation in the Composers group, their representation is consistently low across artistic roles, considering group size.

(a) **Degree distribution.**



(b) **Network position of man-man collaborations.**



(c) **Network position of woman-woman collaborations.**

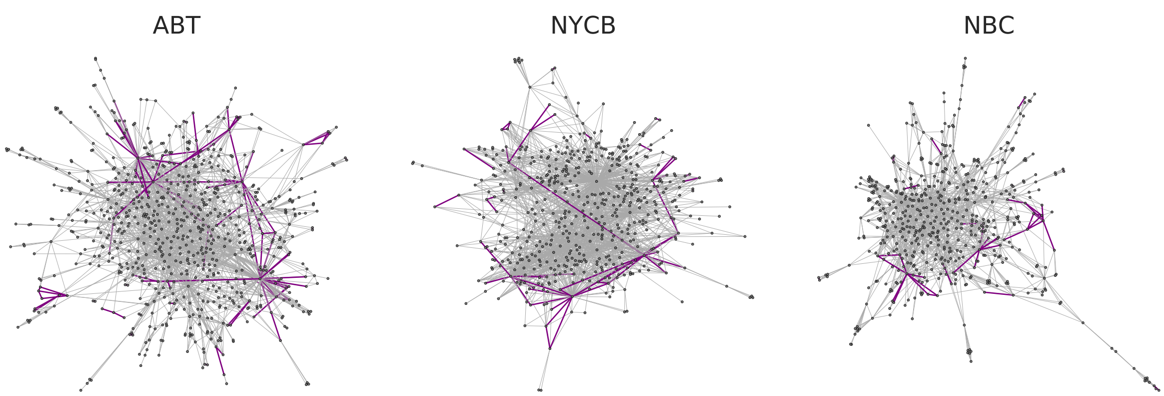


Figure S1.4: **Collaboration networks from ABT, NYCB, and NBC.**

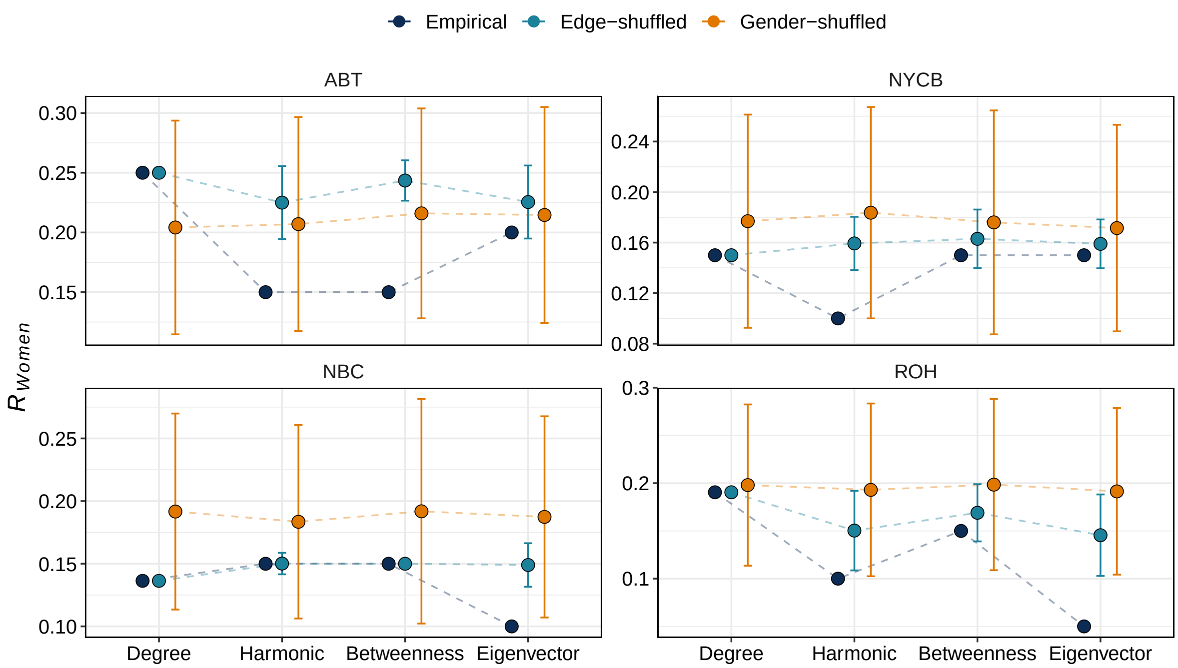


Figure S1.5: **Fraction of women of TCA.** Mean of fraction of women, R_{Women} with error bars comparing the empirical networks and the R_{Women} obtained from null models.

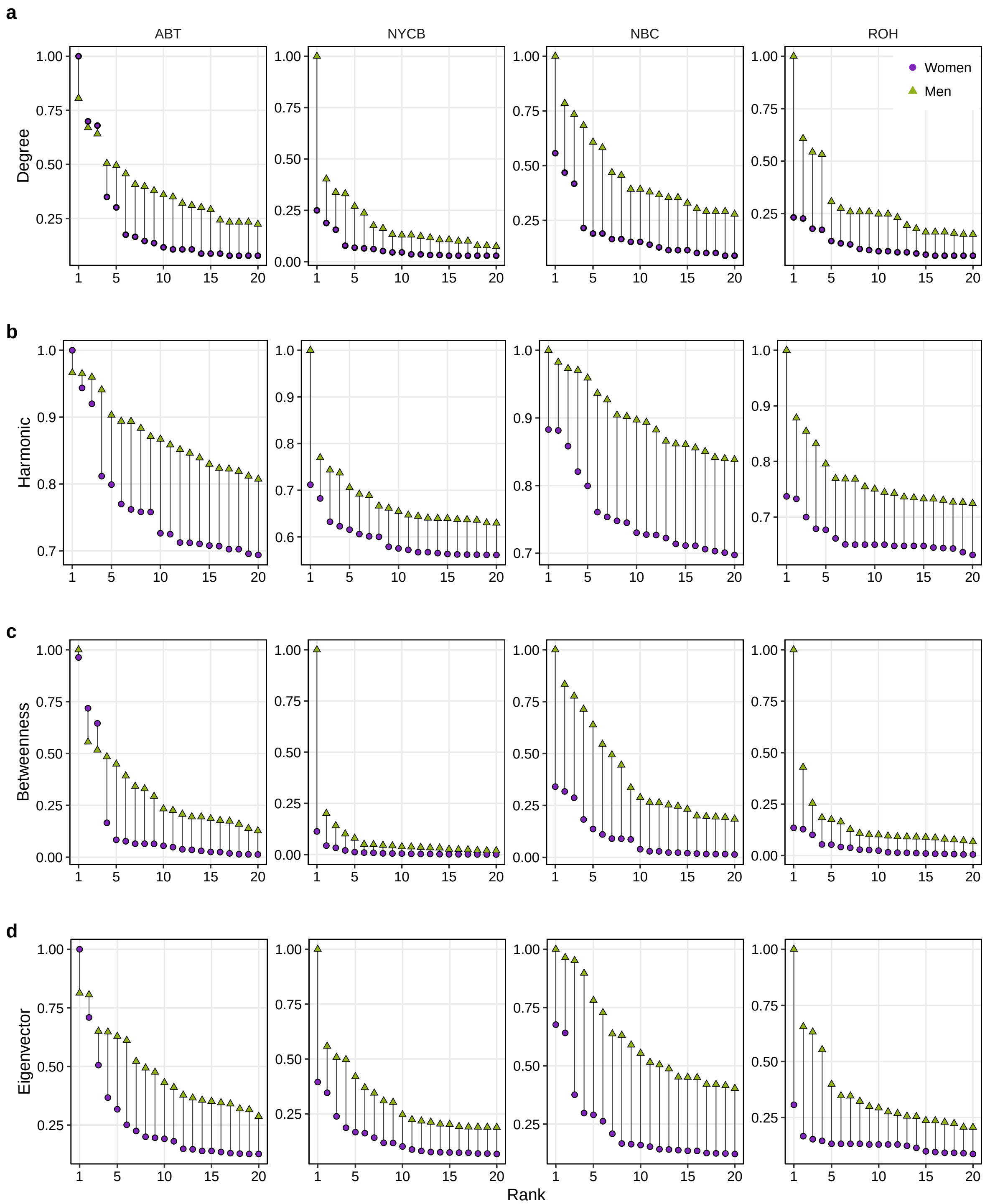


Figure S1.6: **Gender gap in centrality of TCA.** Reporting normalized values by company. Artists matched by their rank computed in independent gender groups.

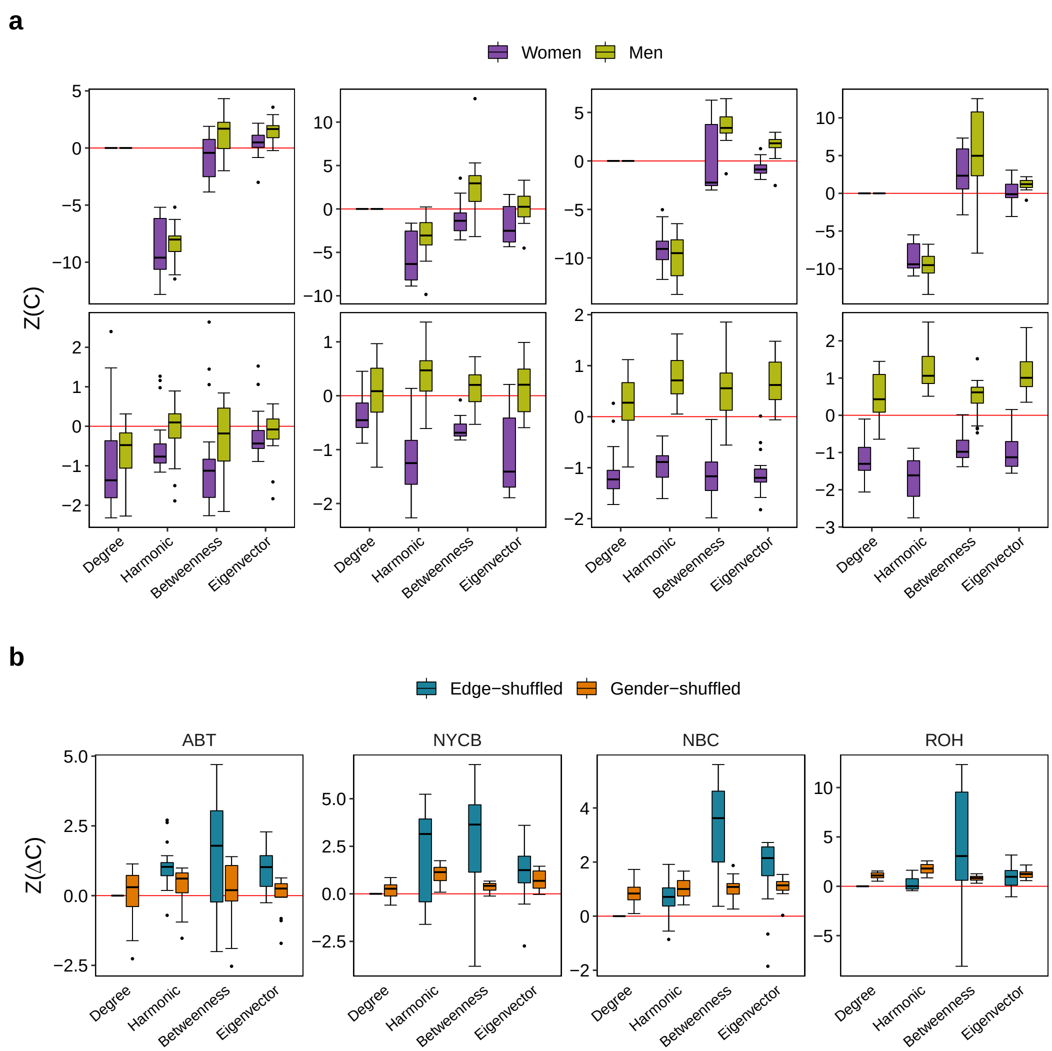


Figure S1.7: Z-**score of** $C(r)_{real}$. Boxplot with error bars. Color code is indicated in each plot. The red line indicates a Z-score equal to zero, corresponding to the mean obtained from each null model. Panel **a** shows the distribution of $Z(C)$, the difference in centrality by rank and gender group TCA$_{Women,Men}$. Upper boxes display the Edge-shuffled model, lower boxes the Gender-shuffled model. Panel **b** shows the distribution of $Z(\Delta C)$, the Z-score of the gender gap in centrality by rank, separated by randomized model.

Appendix B

Supplementary Material of Quantifying Hierarchy and Prestige in US Ballet Academies as Social Predictors of Career Success

B.1 Youth America Grand Prix

The Youth America Grand Prix (YAGP) is a highly influential ballet competition that aims to discover and reward young dancers. This paid event provides education and job opportunities to dance students in pre-professional training, and is seen as an effective way to promote dancers on the path to successful careers. The competition is held in two stages, with multiple regional semi-finals and one final competition each year. Participants are divided into age categories: Pre-Competitive (9-11 years old), Junior (12-14 years old), and Senior (15-19 years old; 20 years old after the COVID-19 pandemic in 2020).

The competition recognizes dancers with two types of awards. The first are the competition medals: gold, silver, and bronze, which are given based on aggregate scores across various judges and dance categories. The second is the Grand Prix, which is a subjective appreciation of exceptional performance, awarded by a committee of judges without explicit criteria. Both awards can be given during the semi-finals or finals; the medals may reflect ties, while the Grand Prix is only awarded to one student per division, or may not be given at all. The YAGP publicly reports the medal and Grand Prix winners, and overall top students (up to top 12) by year and location. As many YAGP competitors are awarded scholarships and professional contracts in ballet companies, the YAGP also reports on the successes of its alumni, including their pre-professional affiliations and current job placements.

B.2 Ballet Schools

We define a 'ballet school' as any organization that offers ballet training. In the US, the structure of ballet schools vary in their organization. There are both university and non-university programs. For instance, the Higher Education Arts Data Services 2020-2021 report lists 76 universities affiliated to the National Association of Schools of Dance. For the context of this research, we must consider that most dancers competing at YAGP would engage in pre-professional ballet training in non-university institutions because of the competitive age divisions, ranging from 10 to 19 years old.

B.2.1 Distribution of Ballet Schools across the U.S.

We explore the geographic location of schools to map the distribution of ballet academies across the US. To do this, we perform a Google search of all schools' addresses to find their georeference. Simultaneously, we confirm the formation of communities in the network of ballet schools using the Louvain algorithm [1], as this helps understand whether schools tend to compete in clusters. We find that the communities detected with the algorithm and their geographical regions strongly overlap. We show the resulting network in Figure S2.1 with colors assigned based on the community detected and node's position based on actual US regions, as: Northeast, East, Midwest, South, and West.

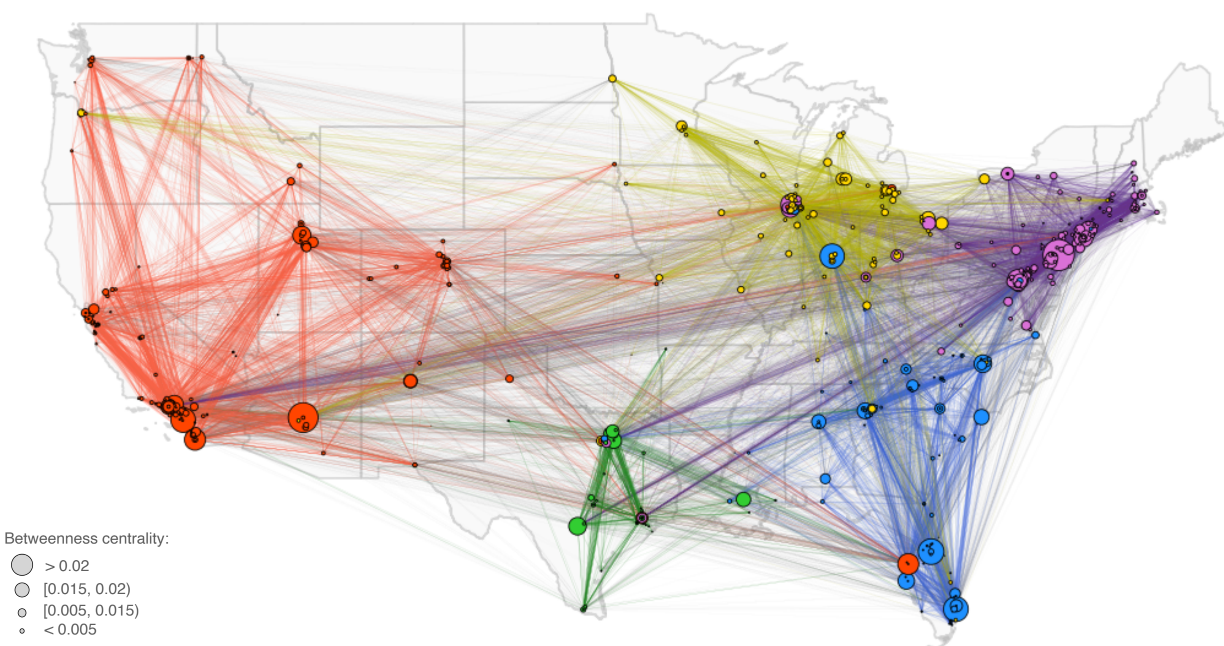


Figure S2.1: **Network of ballet schools in the US.** Ballet schools are nodes, and two schools are linked by winning in the co-competition of the YAGP. Node size by schools' betweenness centrality, B_k. Schools' location corresponds to their actual geographical location. Nodes are colored by detected communities Γ: Northeast (purple), East (blue), West (orange), South (green), and Midwest (yellow). Edge color/gray for intra/inter community links.

B.2.2 Schools' achievement

Achievement, measured by the number of awards received per school, is an important factor influencing prestige because it can capture the level of social recognition that one institution has for its outstanding performance. Here, we explore schools' achievement in the YAGP competition and the existing patterns in the distribution of awards, to test the relationship with network centrality. This analysis provides insights on whether social prestige is rooted in achievement, or it is also related to the richness of professional connections.

From the YAGP's structured data, we quantify schools' achievement as a proxy derived from the number of awards received during their participation in the competition. We compute the following metrics:

i. *School's total number of awards,* A_k: the sum of awards a from the total competitions J (semi-finals and finals), that each student i from the set of students I obtains per competition j when they are affiliated to school k. This can be noted as $A_k = \sum_{i=1}^{I} \sum_{j=1}^{J} A_{ij}\theta_{ijk}$, where A_{ij} is the individual sum of awards per competition, $\theta_{ijk} = 1$ is given if the student i is affiliated to school k for the competition j, and $\theta_{ijk} = 0$ otherwise.

ii. *Ratio of awards per school,* R_k: the ratio of obtained awards derived from the number of competitions where affiliated students ranked as top 12. The ratio of awards for schools is $R_k = \frac{A_k}{T_k}$ where T_k is the total top students affiliated to school k, given by $T_k = \sum_{i=1}^{I} \theta_{ik}$ for $\theta_{ik} = 1$ if a student is affiliated to k, and $\theta_{ik} = 0$ otherwise; then R_k gives the number of awards per student affiliated to each school. Here, $R = 0$ indicates that no awards were obtained, $R = (0, 1)$ indicates less than one award per competition, $R = 1$ gives an even relation between the number of awards and competition, and $R > 1$ for those with more than one award per competition in top 12. We label each group as 'no medals', 'under-achiever', 'break-even', and 'high-achiever', respectively for each ratio group mentioned above.

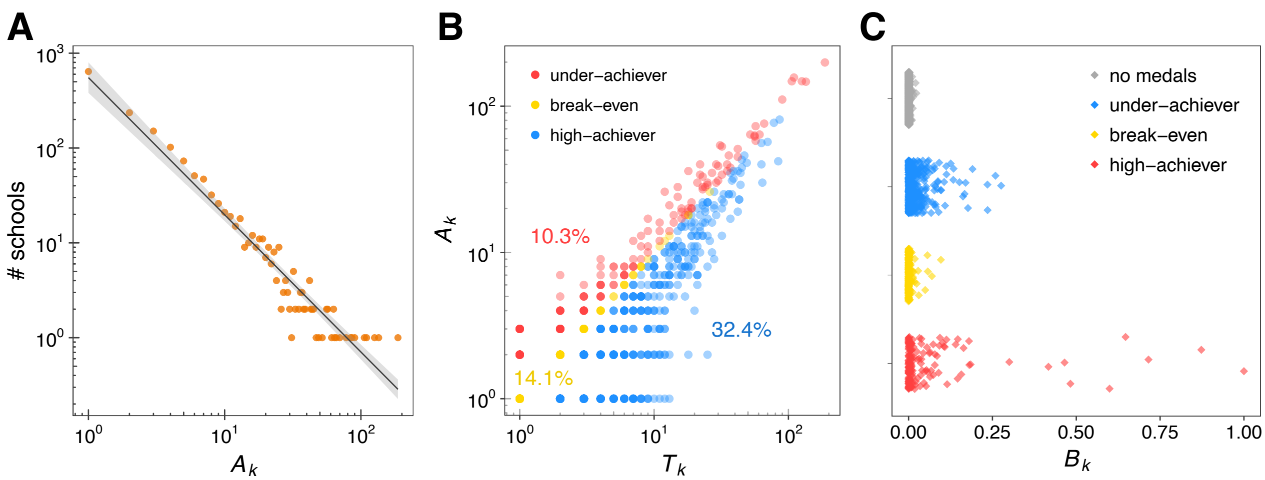


Figure S2.2: **Awards and centrality of ballet schools.** **A** shows the fat-tailed distribution of total awards per school, A_k, indicating that from those schools obtaining awards, most of the have only one, and only a few have more than more than 100 awards. **B** shows the log distribution of awards A_k respect to the number of students T_k per school k, i.e. ratio of awards, R_k. Color indicates the corresponding group by ratio of awards and their percentage (plot not displaying schools with 'no medals', $A_k = 0$). Only 56.8% of schools obtained at least one award, and the total number awards increases proportionally to the number of students. However, the ratio of 'high-achievers' is independent from their number of top students. **C** shows the relationship of betweenness centrality, B_k, with each group by ratio of awards. Schools with no medals (in gray) have the lowest centrality value, while only highly central schools are in the 'high-achievers' group.

Figure S2.2A shows the total number of awards per school, from which we observe a fat-tail distribution, indicating that most schools who have won awards have obtained only one or two awards, while a few schools collected more than 100 awards. To better understand the relationship between the total number of awards (A_k) and the total number of students (T_k) per school k, we compute the ratio of awards as $R_k = A_k/T_k$. From this ratio, four groups of schools are observed (see distribution of in Figure S2.2B). We find that about 75% of schools have a low ratio of awards. For instance, 692 schools (43.2% of all) did not receive any awards ($R_k = 0$), even though they had top students listed by the YAGP; and 519 schools (32.4%) are 'under-achievers', with less than one award per top student ($R_k = (0, 1)$, in blue). Conversely, only 227 schools (14.1%) have a 'break-even' ratio ($R_k = 1$, in yellow), indicating one award per top student; and only 10.3% of schools (165 schools) are 'high-achievers' ($R_k > 1$, in red), meaning that their students obtain more than one award. Noteworthy, the 'high-achiever' group ranges from very low to a higher number of students, indicating that a high ratio of awards is independent of the frequency of schools' participation in the competition.

Next, we examine the relationship between schools' ratio of awards and betweenness centrality. In Figure S2.2C, we observe that most schools are located in the lower quartile of betweeenness centrality, B_k. Schools with no medals (in gray) have the lowest centrality values, closer to zero. On the contrary, only highly central schools are in the 'high-achievers' group, yet this group also includes low centrality schools. Moreover, 'under-achiever' schools have a wider distribution across the lower quartile, indicating that these schools can have higher centrality than the other groups.

Taken together, these findings help contextualize the existing relationship between achievement and its variation by schools' network position.

B.2.3 Ranking of Ballet Schools

To rank schools by their social prestige, we first evaluate different network centralities and schools' achievement to test what measure captures prestige more accurately. The network centralities are listed in Table S2.1 and were selected based on the reported association of network position and social prestige [2] (For more details, see Methods 3.4.2). Achievement is measured by the number of awards received per school (see SI B.3.1), an important factor influencing prestige because it can capture the level of social recognition that one institution has for its outstanding performance [3, 4]. To control the effect of the number of awards (A_k) by the number of students (T_k, a proxy of schools' size), we compute the ratio of awards as $R_k = A_k/T_k$ and include it as a metric of achievement.

Separately, we collect an external selection of prestigious schools by ballet experts from different sources. First, *Dance Magazine*, a leading multimedia platform in the dance world, that also partners with other dance sources in multiple publications (e.g. Pointe, Dance Spirit, and Dance Teacher). We use their selection of top pre-professional academies for the academic year 2022-2023 [9]. Separately, we collect the list of top ballet schools in the US and the 2023 Summer

Table S2.1: Centrality metrics to assess the prestige of ballet schools. Formula in general notation as described in the specified citation.

Network measure	Description	Formula
Betweenness	how much a node connects two other nodes	$B_i = \sum_{s,t} \frac{n^i_{s,t}}{n_{s,t}}$ [5]
Closeness	a node's proximity to other nodes	$Ci = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$ [6]
Degree	a node's connectivity	$D_i = \frac{k_i}{N-1}$ [7]
Eigenvector	a node's influence	$E_i = \frac{1}{\lambda} \sum_k a_{k,i} E_k$ [8]

Intensive Guide from the blog *A Ballet Education*, one of the most reliable online ballet experts for teachers, students, ballet professionals, and ballet lovers [10]. In total, the list of top schools formed from these sources, contains 60 schools' names, shown in Figure S2.3.

1. The Rock School for Dance Education, PA
2.Master Ballet Academy, AZ
3.Indiana Ballet Conservatory Inc, IN
4.Southland Ballet Academy, CA
5.Orlando Ballet School, FL
6.The Art of Classical Ballet, FL
7.Ballet West Academy, UT
8.Ellison Ballet, NY
9.Cary Ballet Conservatory, NC
10.Metropolitan Ballet Academy & Company, PA
11.Anaheim Ballet Academy, CA
12.Colorado Ballet Academy, CO
13.The Sarasota Cuban Ballet School Inc, FL
14.Maryland Youth Ballet, MD
15.Elite Classical Coaching, TX
16.Joffrey Ballet Academy, NY
17.Sarasota Ballet School, FL
18.Joffrey Academy of Dance, IL
19.Cincinnati Ballet Otto M. Budig Academy, OH
20.The Ballet Clinic, AZ
21.Koltun Ballet Boston, MA
22.Harid Conservatory of Music Inc, FL
23.Texas Ballet Theater School, TX
24.Timothy M Draper Center for Dance Education Inc, NY
25.Central Pennsylvania Youth Ballet, PA
26.Kansas City Ballet School, MO
27.Vitacca Vocational School for Dance, TX
28.City Ballet San Francisco, CA
29.BalletCNJ, NJ
30.San Francisco Ballet School, CA
31.Houston Ballet Academy, TX
32.Milwaukee Ballet, WI
33.A&A Ballet, IL
34.Boston Ballet School, MA
35.BalletMet Dance Academy, OH
36.Interlochen Center for the Arts, MI
37.Ballet Conservatory at Skyra, FL
38.Nashville Ballet, TN
39.West Met Classical Training, MN
40.National Ballet Academy Denver, CO
41.Nutmeg Conservatory, CT
42.Pittsburgh Ballet Theatre, PA
43.Pennsylvania Academy of Ballet, PA
44.The School of American Ballet, NY
45.School of Ballet 5:8, IL
46.Dmitri Roudnev Ballet School, IL
47.Oregon Ballet Academy, OR
48.Neglia Ballet, NY
49.Tulsa Ballet Center for Dance Education, OK
50.Sacramento Ballet, CA
51.Alonzo King LINES Ballet School, CA
52.Saint Louis Ballet School, MO
53.Oklahoma City Ballet Yvonne Chouteau School, OK
54.Philadelphia Ballet, PA
55.Ballet Austin Academy, TX
56.Jacqueline Kennedy Onassis School at ABT, NY
57.Miami City Ballet School, FL
58.State Street Ballet, CA
59.The Portland Ballet, OR
60.Princeton Ballet School, NJ

Figure S2.3: **List of top schools.** Selection of top schools by US experts from *Dance Magazine* and *A Ballet Education*. Schools listed in no particular order.

We implement a classification model using a binary system for the list of Top Ballet Schools (1:top school, 0:no top school) to test the schools' centrality and achievement measures' accuracy to capture social prestige. Considering each metric in separate models, we measure the area under the cure (AUC) of the classification model (Receiver Operating Characteristic Curve, ROC) using the pROC package for R [11]. The AUC is a score between 0 and 1, where a value closer to 1 indicates a probability of correct classification, while a score lower than 0.5 indicates that the model performs no better than random guessing. AUC and ROC for each centrality are shown in Figure S2.4. The AUC=0.75 suggests that betweenness centrality (Fig. S2.4A) is the most accurate centrality measure capturing social prestige, respect to the other centrality metrics (Fig. S2.4B-D). On the other hand, when we explore the ratio of awards (number of total awards controlled by school size) we observe the lowest accuracy in predicting schools' prestige (Fig. S2.4F), with an AUC=0.71.

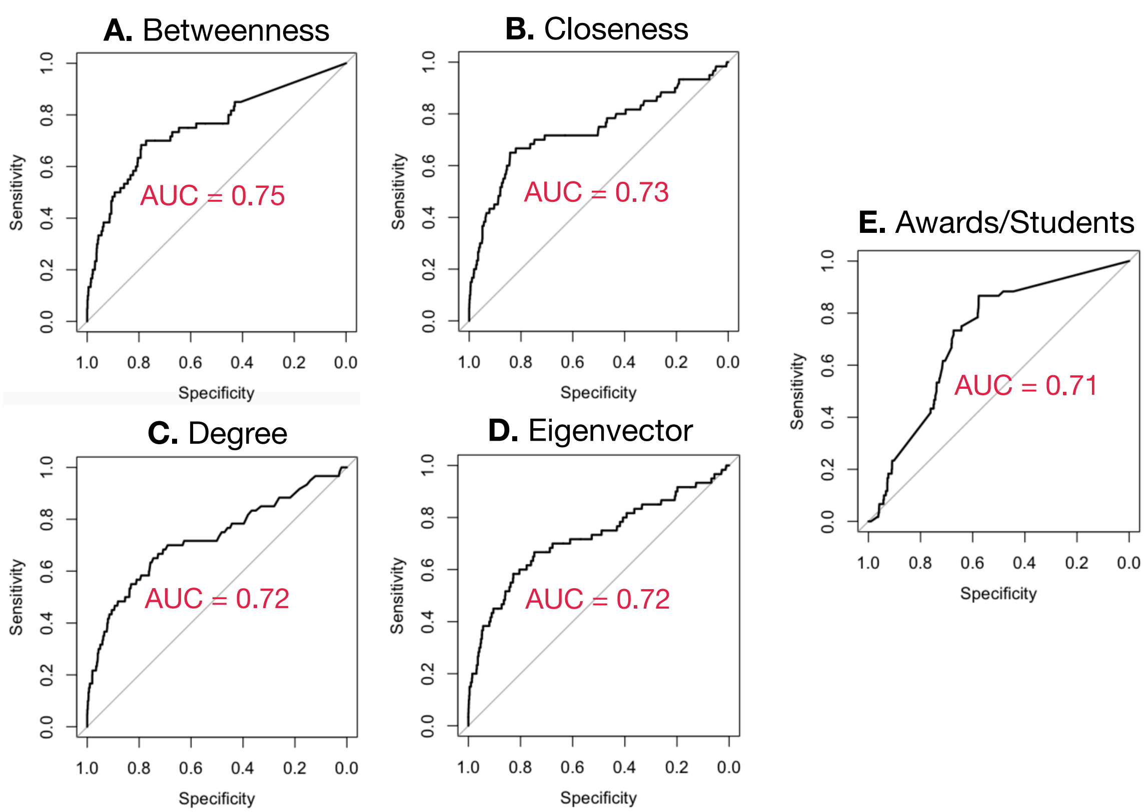


Figure S2.4: **AUC for network centrality and ratio of awards.** Panel **A** shows that betweenness centrality is the best indicator of prestige with the largest AUC=0.75 among all centralities in panels **B-D** and schools' achievement, in panel **E**.

Further, we order the schools by their betweenness centrality and create a ranking list. In this rank, a school k with $r_k = 1$ has the largest centrality value, i.e. is more central or prestigious, and the largest rank value (e.g. $r_k = 945$) has the lowest centrality in the set of schools K. Based on the number of observations per school, we select the upper 5% of schools from the network-based ranking. The list of schools is shown in Figure S2.5. We use this selection for the further treatment effect of being in a top school on job placement, notated as $Y_i = 1$ for students affiliated to a prestigious school, and $Y_i = 0$ for students who attended a less prestigious school.

1.The Rock School for Dance Education, PA
2.Master Ballet Academy, AZ,
3.Indiana Ballet Conservatory Inc, IN
4.Dmitri Kulev Classical Ballet Academy, CA
5.The Kirov Academy of Washington, DC
6.Southland Ballet Academy, CA
7.Orlando Ballet School, FL
8.Marat Daukayev Ballet Theatre Inc, CA
9.Faubourg Theatre Inc, IL
10.The Art of Classical Ballet, FL
11.Danceology Academy of Classical Ballet, CA
12.Next Generation Ballet, FL
13.University of North Carolina School of the Arts, NC
14.Ballet West Academy, UT
15.Pacific Coast Academy of Dance, CA
16.Ellison Ballet, NY
17.The Dallas Conservatory, TX
18.The Academy of Ballet Arts Inc
19.The Washington Ballet School, DC
20.The Academy of Dance Arts, IL
21.International Ballet School, CO
22.CityDance School & Conservatory, MD
23.Dance Center of San Antonio, TX
24.V&T Classical Ballet & Dance Academy, CA
25.The Rock Center for Dance, NV
26.International City School of Ballet, GA
27.Alwin School of the Dance, NM
28.Ballet Conservatory, TX
29.Cary Ballet Conservatory, NC
30.Westlake School for the Performing Arts, CA
31.Metropolitan Ballet Academy & Company, PA
32.Akhmedova Ballet Academy, MD
33.Greenwich Ballet Academy Inc, CT
34.Jacqueline's School of Ballet, UT
35.Classical Ballet Academy, CA
36.Odasz Dance Center, NY
37.Stars Dance Studio, FL
38.Los Angeles Ballet Academy, CA
39.Fishback Studio of the Dance, NM
40.Anaheim Ballet Academy, CA
41.Mid-Atlantic Center for the Performing Arts, MD
42.MorningStar Dance Academy of Atlanta, GA
43.Americas Ballet School, FL
44.Colorado Ballet Academy, CO
45.Yuri Grigoriev School of Ballet, CA
46.Marina Almayeva School of Classical Ballet, TX
47.Steps On Broadway, NY
48.Festival Ballet Providence, RI
49.Summit Dance Shoppe, MN
50.Julianas Academy of Dance, MI
51.Coastal Dance Centre, SC
52.South Carolina Governors School for the Arts and Humanities, SC
53.Valentina Kozlova Dance Conservatory of New York, NY
54.Zamuel Ballet School, CO
55.Long Beach Ballet, CA
56.International Ballet Academy, NC
57.The Sarasota Cuban Ballet School Inc, FL
58.Academy of Russian Ballet, VA
59.Maryland Youth Ballet, MD
60.Peninsula School for Performing Arts, CA
61.Aika Ballet School, NY
62.Litchfield Dance Arts Academy, SC
63.Academy of Russian Classical Ballet, MI
64.All American Classical Ballet School, FL
65.First State Ballet Theatre Inc, DE
66.Westchester Dance Academy, NY
67.Olga Kresin Ballet School and Studio, PA
68.Elite Classical Coaching, TX
69.Joffrey Ballet Academy, NY
70.Russian Ballet Academy, AZ
71.Sarasota Ballet School, FL
72.Littleton Ballet Academy, CO
73.Joffrey Academy of Dance, IL
74.Columbia Conservatory of Dance, SC
75.Dance Industry Performing Arts Center, TX
76.Art Ballet Academy, TX
77.Florida School for Dance Education, FL
78.Pittsburgh Ballet House, PA
79.Cincinnati Ballet Otto M. Budig Academy, OH
80.Ballet Arte Academy of Classical Ballet, CA
81.Columbus Youth Ballet, OH
82.Ballet Tech Ohio, OH
83.The Ballet Clinic, AZ
84.Russian Ballet Academy of Indiana, IN
85.Dancing Arts Center, MA
86.Burbank Dance Academy, CA
87.Hudson Conservatory of Ballet, OH
88.Menlo Park Academy of Dance, CA
89.School of Classical Ballet & Dance, IA
90.Emerald Ballet Theatre, WA
91.Artistry in Motion Performing Arts Center, TX
92.North County Academy of Dance, CA, USA
93.Ohio Conservatory of Ballet, OH, USA

Figure S2.5: **Ranking of ballet schools.** The list contain the top 5% most prestigious U.S. ballet academies from the network-based ranking.

B.3 YAGP Alumni

B.3.1 Students' achievement

We quantify students' performance in the competition by means of the number of awards and the times they are listed as top students by the YAGP. We use these measures as a proxy of performance in ballet. We compute the following metrics:

i. *Student's total competitions,* T_i: the number of semi-final competitions as a top student. Multiple participation in the competition per year are possible, thus T_i denotes only competitions in the semi-finals.

ii. *Student's total number of awards,* A_i: the sum of awards a each student i obtains per competition j, considering the medals earned in each category (c = classical, m = contemporary), and the *Grand Prix* (g). Then, $A_i = A_{ic} + A_{im} + A_{ig}$, where for $A_{ic} = \sum_{j=1}^{T_i} a_{ij}$, $a_{ij} = 1$ when a dancer was awarded a medal in the classical category for the competitions j up to T_i for all semi-final competitions in C, and $a_{ij} = 0$ meaning that the student did not obtain a medal. The notation is the same for all possible awards in c, m, and g. One student cannot win more than one medal per category in each competition, then the possible number of awards in a given competition would be $A_{ij} = \{0, 1, 2, 3\}$. In a similar fashion, competition medals per student are counted by type (gold, silver, bronze).

iii. *Ratio of awards per student,* R_i: the ratio of obtained awards derived from the number of competitors ranked as top 12. The ratio of awards for students is given by $R_i = \frac{A_i}{T_i}$ and averages the awards obtained per student in each competition. Similarly to schools' ratio of awards, $R = 0$ indicates that no awards were obtained, $R = (0, 1)$ indicates less than one award per student, $R = 1$ gives an even relation between the number of awards and student (e.g. one student won one award every time he was listed as top 12), and $R > 1$ for those with more than one award per student in top 12. We label each group as 'no medals', 'under-achiever', 'break-even', and 'high-achiever', respectively for each ratio group mentioned above.

In Figure S2.6A, we observe the power-law distribution of students' number of awards, A_i. This indicates that only a few students win multiple awards, while the majority can achieve at least one medal. This distribution of awards is consistent with the one observed in the number of awards per school (Figure S2.2A). Most students (75%, 4824 students) have a success ratio lower than one, indicating that they win less than one award per competition. From this fraction, 3642 students (57% of all) obtained 'no medals', $R_i = 0$, even though they were listed in the top 12. Figure S2.6B shows that, from those who obtained at least one medal, a 18.5% (1182 students, in blue) are 'under-achievers', $R_i = (0, 1)$, meaning that they obtained less awards than the number of times they were listed in the top 12 ($A_i < T_i$). Conversely, a very small fraction of students (4.8%, 309 students, in red) are 'high-achievers', characterized by a high success ratio, $R_i > 1$, with more awards than the number of times they were listed in the top 12 ($A_i > T_i$). Only 19.7% of students (1260 students, in yellow) obtained a 'break-even' success ratio ($R_i = 1$). One would expect that the number of awards per student (A_i) strongly correlates with the total times in the top 12 (T_i), indicating that awards accumulate as students participate in the competition. Such relationship is moderate (Pearson's corr. coefficient=0.68), suggesting that being listed in the top 12 multiple times is not a strong but a mild predictor of the total number of awards. The majority of students (3511 students, 55%) have only one reported competition in the top 12 ($T_i = 1$), but we see in Figure S2.6D that all groups by ratio of awards have an exponential distribution respect to T_i. This indicates that there is a general pattern of variability in T_i that is independent of students' success ratio, S_i and also illustrates that the awards are actually scarce, even among top contenders.

To explore the distribution of awards, we compute the observed probability of obtaining an award, $P(\Lambda)$, where Λ represents each award by their implied rank. Such implied rank Λ is for semi-finals (SF): no medals, Bronze, Silver, Gold, Grand Prix; and for the finals (F): no medals, Bronze, Silver, Gold, Grand Prix (this rank does not assume the award value given by individuals). We divide the total number of awards per rank by the total number of top 12 students in the competition pool. Figure S2.6C shows that YAGP competitors have a 0.62 probability of being positioned in the top 12 without winning an award, while winning a competition medal in the semi-finals has about 0.1 of probability for gold, silver, and bronze. Interestingly, the probability of winning the *Grand Prix* in the semi-finals is rather low (0.03) but higher than being in the top 12 of the finals without obtaining an award (0.017 probability). Winning any competition medal in the YAGP finals has a very small probability ($Pr(\Lambda = \{\text{F: bronze, silver, gold}\}) < 0.005$), and the probability of winning the Grand Prix at this competition level is even smaller (0.002). Taking all together, our findings draw a picture of the level of competitiveness that YAGP competitors face and shows how repeated competitions do not imply a higher success ratio, even among top contenders, and raise the question about what factors different from persistence –and accumulated practice/experience– can determine dancers to be recognized by the jury in the YAGP competition venues.

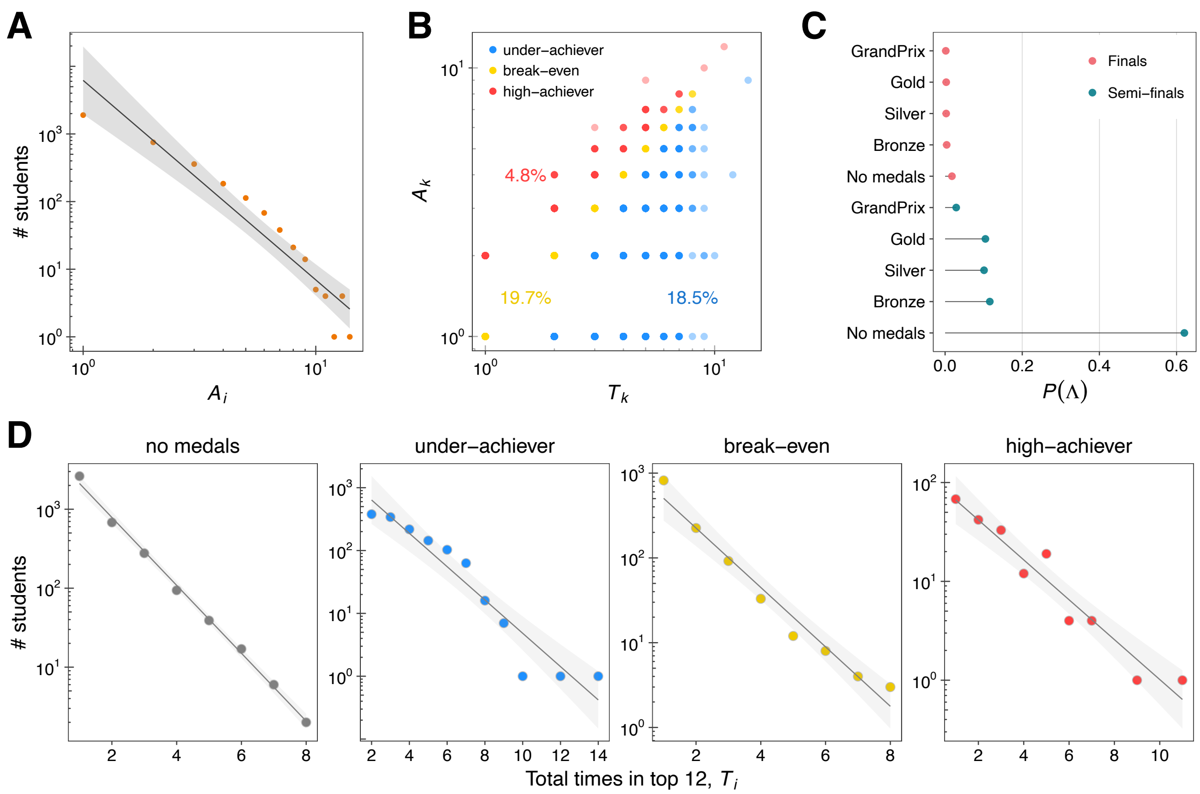


Figure S2.6: **Awards of ballet students. A** shows the count of the number number of awards per student A_i, with a power-law distribution. This indicates that most students obtained only one award, while only a few obtained more than 10 awards. **B** shows the percentage of students by group of success ratio, R_i, obtained from dividing the number of awards by the number years they positioned as top student. Taken together, 'no medals' ($R_i = 0$, in gray) and 'under-achievers' ($R_i = (0, 1)$, in cyan) represent the 75.3% of the total number of competitors (6393 students), meaning that most competitors win less than one award per top positioning. About 19.7% of students (1260 students) wins one award per top positioning ($R_i = 1$, in yellow), while only 4.8% of students (309 students) obtains more than one award per time they position as top student. **C** shows the probability of obtaining awards by rank for the finals/semi-finals in orange/blue. We see that a most top students do not obtain an award, and the probability of being awarded decreases as the rank increases, emphasizing scarcity in high rank awards (i.e. Finals). **D** shows the count of students and their number of competitions, T_i by their ratio of awards group, R_i, and demonstrates that all groups display an exponential distribution, suggesting that there is a general pattern of variability in the number of awards obtained per student disregarding of the number of semi-finals competitions.

B.3.2 Students' success

Table S2.2 shows four variations of our base model. Model 1 corresponds to model described in Eq. 3.2. In model 2, we test for the potential effect of advancing to the competition finals by adding a dummy variable for being a finalist ($F_i = 1$) or not ($F_i = 0$). Models 1 and 2 are discussed in the results section.

In model 3, we examine the role of an affiliation for a job placement. Here, we replace the measure of school's prestige with a dummy variable for being affiliated ($D_i = 1$) or being an independent competitor ($D_i = 0$). This model shows no statistical effect for attending a school versus being an independent competitor, suggesting that there is no statistical difference of being affiliated or not regarding the chances of obtaining a job placement. Separately, in model 4 we explore the effect of total competition medals and observe that winning any medal has a significant but small positive effect on the chances of obtaining a job placement, increasing only by 49%.

In addition, we check for robustness of our treatment effect analysis on being affiliated to a top school, and observe that when we use the top 10% of schools (148 schools, 4217 controls, 3501 treated), the chances of job placement increase by 43% ($p = 0.0015$).

In sum, our analyses show that school prestige has a robust and significant positive effect on dancers' job placement, a comparable effect respect to being a finalist, but more importantly, that even when performance is similar, there are social factors such as prestige driving the selection of dancers towards successful company positions.

Table S2.2: **Model Coefficients for the Probability of Success.** Model coefficients labeled by p-value. Standard errors in parentheses.

	Dependent variable = Success S_i (job placement)			
	(1)	(2)	(3)	(4)
Affiliation D_i			0.319	
			(0.391)	
Prestige	1.201^{***}	1.136^{***}		1.203^{***}
	(0.190)	(0.193)		(0.190)
Grand Prix	0.809^{***}	0.515^{***}	0.862^{***}	0.831^{***}
	(0.113)	(0.127)	(0.110)	(0.112)
Gold	0.45^{***}	0.344^{***}	0.448^{***}	
	(0.091)	(0.093)	(0.088)	
Silver	0.45^{***}	0.413^{**}	0.472^{***}	
	(0.091)	(0.091)	(0.089)	
Bronze	0.276^{**}	0.258^{**}	0.295^{**}	
	(0.096)	(0.097)	(0.095)	
Total				0.403^{***}
				(0.053)
Finalist F_i		1.0^{***}		
		(0.184)		
Competitions	-0.238^{***}	-0.208^{***}	-0.234^{**}	-0.244^{***}
	(0.058)	(0.058)	(0.059)	(0.058)
Gender: male	0.754^{***}	0.756^{***}	0.915^{***}	0.751^{***}
	(0.121)	(0.121)	(0.116)	(0.121)
Constant	-3.135^{***}	-3.193^{***}	-3.608^{***}	-3.138^{***}
	(0.111)	(0.112)	(0.393)	(0.111)
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$			
Observations	6393	6393	6475	6393
McFadden pseudo R^2	0.085	0.095	0.075	0.08
AIC	2640.3	2615.3	2911.5	2638.9
AUC	0.7327	0.7374	0.7286	0.7312

References

[1] Vincent D Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10 (2008), P10008.

[2] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[3] Don H Shamblin. "Prestige and the sociology establishment". In: *The American Sociologist* (1970), pp. 154–156.

[4] Bernd Wegener. "Concepts and measurement of prestige". In: *Annual review of sociology* (1992), pp. 253–280.

[5] Linton C Freeman. "A set of measures of centrality based on betweenness". In: *Sociometry* (1977), pp. 35–41.

[6] Gert Sabidussi. "The centrality index of a graph". In: *Psychometrika* 31.4 (Dec. 1, 1966), pp. 581–603. DOI: 10.1007/BF02289527. URL: https://doi.org/10.1007/BF02289527.

[7] Linton C Freeman. "Centrality in Social Networks: I. Conceptual Clarification". In: *Social networks* 1.3 (1978), pp. 215–239.

[8] Phillip Bonacich. "Some unique properties of eigenvector centrality". In: *Social networks* 29.4 (2007), pp. 555–564.

[9] *Dance Magazine: Pre-professional Program Guide 2022-23*. 2022. URL: https://www.dancemagazine.com/guide/dance-magazines-2022-23-pre-professional-program-guide/.

[10] David King. *A Ballet Education: The Blog*. URL: https://aballeteducation.com.

[11] Xavier Robin et al. "pROC: an open-source package for R and S+ to analyze and compare ROC curves". In: *BMC Bioinformatics* 12 (2011), p. 77.