

Beyond Performing Arts: Network Composition and Collaboration Patterns

Yessica Herrera-Guzman
Research Center for Social
Complexity
Universidad del Desarrollo
Santiago, Chile
yherrerag@udd.cl

Eduardo Graells-Garrido
Barcelona Supercomputing Center
Barcelona, Spain
eduardo.graells@bsc.es

Diego Caro
Data Science Institute
Universidad del Desarrollo
Santiago, Chile
dcaro@udd.cl

ABSTRACT

The social network approach has been applied to understand social and cultural phenomena. Previous studies have implemented network metrics to investigate collaborations in different contexts, such as scientific communities. However, collaborations patterns in performing arts have not been fully described. To address this gap, we propose the reconstruction and analysis of the collaboration networks of performing artists registered in Wikidata. Our results suggest that different performing arts share similar collaboration patterns, as well as a mechanism of community formation that is consistent with observed social behaviors. Our study adds value to the use of Wikidata as a source of information to investigate social complex systems.

KEYWORDS

Collaboration networks; Wikidata; Performing arts; Community detection

ACM Reference Format:

Yessica Herrera-Guzman, Eduardo Graells-Garrido, and Diego Caro. 2018. Beyond Performing Arts: Network Composition and Collaboration Patterns. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

Performing arts, according to Wikipedia, are "forms of art in which artists use their voices, bodies or inanimate objects to convey artistic expression". Performing arts mainly include dance, music and theatre, but may also refer to other forms of performances, such as the manipulation of objects. Performing arts are seen in most societies and their prevalence in our cultural heritage has been mostly explained as a luxury and investment good with underlying economic interests of art monopolies [4, 20, 29]. In the particular case of dance, this performing art is considered as a human universal, despite the broad diversity of styles and techniques [22]. More

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

Woodstock '18, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>

specifically, ballet is a dance style that includes bodily expressions from ancient cultures [5], has been consolidated since the sixteenth century and has remained at the core of Western civilisation [17]. The history of ballet shows that collaborations between creators are essential for guiding the final artistic representation [6]. For instance, key creative collaborations might have been essential for the production of leading iconoclastic creations by Russian and European artists in late 19th century and during the further transition from classical to modern styles [21]. It is easy to infer, therefore, that social networks played an important role for the transformation of artistic styles along art history, as well as for the transfer of existing traditions from one generation to the next. Hence, ballet is a reliable performing art to explore the mechanisms of collaborations involved in communities of performing artists, which have not been fully described hitherto.

We propose to use the network analysis approach to explore such mechanisms. We propose to work with data obtained from Wikidata 2019 (www.wikidata.org), a structured source of data derived from Wikipedia, the free multilingual online encyclopedia that captures the information and culture from many countries. From this source, we aim to reproduce the collaboration networks of ballet creators. To confirm that the observed network characteristics described the context of performing arts collaborations, we compared the resulting network with a second network created with data from opera creators. We chose opera because it is a performing art with similar origins from ballet. We also compared the network with that one of the historic repertoire of the Pittsburgh Ballet Theater (PBT) in the United States. We understand, however, that our comparative networks reflect different modes of collaborative work. In the rest of the text, we will refer to "performing arts" to ballet and opera only to avoid repetition, yet we are aware that performing arts include a broader range of artistic expressions.

From our analysis, we aim to test these working hypotheses:

(a) *Social networks of different performing arts have similar collaboration patterns.* The collaborations in performing arts have been historically documented as specific for their context. Thus, we want to use objective measures to reflect the similitudes and differences in these collaborations. We expect to find differences in the amount of nodes and artworks involved in each type of performing art studied, given the cultural influence and origins of each performing art, yet they will present many relevant similitudes such as the distribution of artworks created by artists, the number of artist per artwork, and the existence of a significant giant component.

(b) *In performing arts, communities are formed depending on the strength of dyadic ties.* The historical origins of performing arts

relate to both popular and aristocratic social interactions, yet today performing arts are often considered as an elitist practice. Thus, we want to analyze how communities are formed inside the studied social network. Since the number of collaborations observed per dyad is reflected as the weight or strength of its tie, we expect group formation to satisfy the weak link hypothesis, which implies that strong links tend to find more affiliations inside the community, whereas weak links connect more with other communities [11]. However, we also expect the networks to be assortative, meaning that highly connected artists will also connect with highly connected artists from other communities. That is to say, that we aim to observe strong links within and between communities.

By testing our working hypotheses, we aim to contribute to understand *Wikidata and its structure* as a useful source of historical knowledge and a potential tool to investigate complex social behaviors. In particular, with our network analysis of collaborations in performing arts, we want to test whether Wikidata is a valid and representative open-source to understand specific cultural phenomena of historical relevance.

In summary, the contributions of this paper are: (1) the description of collaboration networks in performing arts and (2) to evidence open-editing communal platforms as potentially advantageous to reflect complex collaborative behaviors that shaped our history.

2 RELATED WORK

Collaborations emerge from interactions, connections or relationships between individuals in a social network [25]. The seminal work of Newman et al. [18, 19] demonstrates the use of statistical physics techniques to compare collaboration patterns in communities from different fields and to understand social communication mechanisms.

Collaborations have also been studied with a dynamic network analysis approach. For example, Barabási et al. [3] found that the metrics applied to characterize the network at different time periods were useful to determine the network growth model governing the evolution of that network, thus being able to systematically study collaborations as evolving complex systems. One application of these methods was also conducted by Moody [16] in the context of knowledge diffusion in the social sciences.

In later work related to creative collaboration networks, Uzzi and Spiro [23] reconstructed the network of artists who created Broadway musicals from 1945 to 1989, and observed a small network effect of a curvilinear shape upon creativity, performance and success. Fraiberger et al. [9] used a network approach to investigate the influential role of reputation upon the career success of visual artists, and demonstrated a reciprocal benefit between the reputation of artists and institutions, as well as a crucial role of reputation for career success and auction price of the artworks.

The use of free web-based datasets, such as Wikipedia, has been validated to investigate collective systems [26]. For instance, Yun et al. [27, 28] proposed a mechanistic model to determine collaboration patterns in the edition of Wikipedia articles. They found fast growth models and inequalities in the formation of collective intelligence, and contributed to the understanding of modern social interactions: online collaborations. Keegan et al. [14] demonstrated

Table 1: Fundamental statistics of collaboration networks in Performing Arts. Numbers in parentheses are standard deviations.

	Ballet	Opera	PBT
Artworks	462	4535	288
Artists (nodes)	349	2563	634
Collaborations (edges)	324	1598	932
Average degree	1.856	1.247	2.93
Density	0.0053	0.0004	0.0046
Mean artworks per artist	2.41(4.41)	2.93(5.71)	4.37(10.16)
Mean artists per artwork	1.82(1.13)	1.65(1.07)	9.6(16.2)
Largest Connected Component			
Artists (as proportion)	143(40.97%)	740(28.2%)	619(97.63%)
Collaborations (as proportion)	228(66.6%)	958(59.9%)	922(98.92%)
Average degree	3.188	2.589	2.979
Density	0.0224	0.0035	0.0048
Diameter	9	31	10
Clustering coefficient	0.1384	0.0025	0.0159
Average clustering	0.320	0.0084	0.0177
Communities	9	20	18

how editors' attributes and article type have a strong influential role on collaboration patterns.

An ultimate advantage of the use of data created in an online collective fashion, is that can be used to investigate cultural interactions and social biases, such as gender inequality [8, 10, 24]. Given the many evidence described here for the study of collaborations, we are interested in inferring collaboration patterns in performing arts using freely shared human knowledge, like Wikidata.

3 DATA SETS

We created collaboration networks by importing binary edgelists from SPARQL searches to retrieve information stored up to January 2019 of all Wikidata. The data included all artists registered in the list of ballets (wd:Q15079786) and operas (wd:Q1344) available in the encyclopaedia. The list of collaborators contains the codes for choreographer (P1809), composer (P86), librettist (P87) and costume designer (P2515).

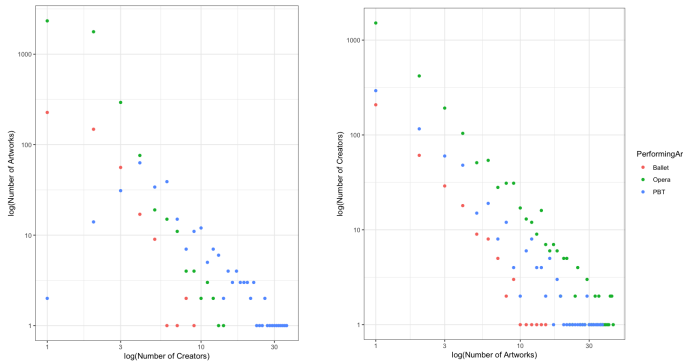
The data for the collaboration network from the historic repertoire of the Pittsburgh Ballet Theater, includes the categories of artists as choreographer, music composer, costume designer, staging, set, and lighting, from the 1969 to 2018 seasons. The networks were constructed using R and Python kernels.

4 METHODOLOGY

The networks are conceptualized as static unipartite graphs, where relationships are from one single level (artists only). In this case, people collaborating with other people (artists with two contributing roles for the same artwork were not considered). For example, there will be a link between two artists that appear in Wikidata as main creators of the same work of ballet or opera. The links have no directionality, thus networks are undirected graphs. The weight of

Table 2: Artists proportion of Performing Arts collaborations

	Ballet	Opera	PBT
Choreographer	0.257	0.002	0.201
Composer	0.546	0.549	0.252
Costume designer	0.035	0.001	0.197
Librettist/other	0.162	0.448	0.35

**Figure 1: Log-log scale distribution of: number of artworks per artists (left), and number of artists per artwork (right).**

the link will be given by the number of shared appearances, namely, collaborations. Since each edge represents one collaboration, multiple edges per node are possible.

We explored each network as static and with no specific time window. The collaboration networks of ballet and opera from Wikidata considered artists creating new works, while the network from the PBT reflects the historical collaborations of artist setting ballet productions on stage. This remarkable difference is a limitation of our work, yet we consider this as a preliminary work to understand how performing arts should be modelled.

To explore the collaboration patterns, we conducted descriptive statistics at the network and largest connected component levels, as proposed by Newman et al. [18, 19]. We used the Louvain method for community detection and the Hierarchical Edge Bundle (HEB) method to generate the visualization of the communities detected and their interactions [13].

5 RESULTS

We observed our networks to differ in node and edges composition, as fundamental statistics show in Table 1. In the case of the networks from Wikidata, the composition may be affected by the existing collective knowledge for opera, which has received more efforts. For example, there are websites devoted to document the creation and production of operas (e.g. theoperadatabase.com, operabase.com), whereas there is not an equivalent work related to ballet.

The average in collaborations (i.e. artworks per artists and artists per artwork) is similar for the networks obtained from Wikidata, but are higher for the PBT. One explanation for this is the belongingness of all PBT artists to the same cultural institution, so that

they collaborate with each other more often. The average of artists per artwork is remarkably higher for the PBT: 9.6 over 1.82 in ballet and 1.85 in opera. This could be an artifact of the different collaborative processes that our networks convey. The distributions of artworks per artists and artists per artworks are shown in Figure 1. Such distributions in a logarithmic scale, allow to infer that: (a) there are few artists with many artworks, and many artists with very few artworks; and (b) few artworks required many artists, and many artworks with very few artists. The observed distribution can be caused by the use of a static network with no specific time window. However, if a dynamic network analysis were conducted, this could be caused by a densification effect of the network over time, based on the Barabási model [2], which implies distributions of a power law shape and a preferential attachment growth of the network. To confirm that, further research is required.

We observed that opera has the largest number of artworks, artists and collaborations, though it has the lowest density and average degree, compared to the ballet networks. The largest average degree and density is seen in the network of the PBT. This difference could be because artists at the PBT work and belong to the same cultural institution, thus making the network denser and higher in collaborations. The differences in density and average degree between networks, might be due to how performing arts are created. For instance, the making of opera usually consists of the collaboration between composer and librettist only, while ballet requires different types of artists, such as choreographers, composers, costume designers and librettists. In setting on stage a ballet performance, additional artistic profiles are also required, such as for lighting, set and staging. Hence, our results confirm the theoretical collaborative proportion of artist types required for the making of performing arts [4]. Ballet has a majority of composers, while opera consists mainly of composers and librettists. As expected, the PBT network has more of other artists: those required to set a live ballet performance. Figure 2 depicts the structure of the network with a color node for each artist type. In ballet, composers tend to be more central in the network, while choreographers tend to be in tree-like positions, leading either composers or librettists. In opera, librettists and composers have an even distribution among the largest connected component and the isolated elements.

Wikidata networks have a larger amount of isolated components, dyads or small groups, resulting in a significant reduction in the proportional size of the Largest Connected Component (LCC) over the size of the complete network (see proportion values in Table 1, visual representation in Figure 2). As a consequence, network metrics such as average degree and density, increase in value with the elimination of disconnected components. Although a network with full connectivity is never seen in reality, the presence of a large amount of disconnected elements is critical for the structure of the relationship among artists. This may reduce the odds of creativity and innovation within each disconnected element and the network as a whole [23]. The diameter of each LCC is relative to network size and number of nodes.

The clustering coefficient describes the existence of a symmetric propensity to have mutual or adjacent connections. Ballet has the highest clustering coefficient from the studied networks, indicating more collaborations of three or more people. Interestingly, the PBT does not show a high clustering coefficient and may indicate that

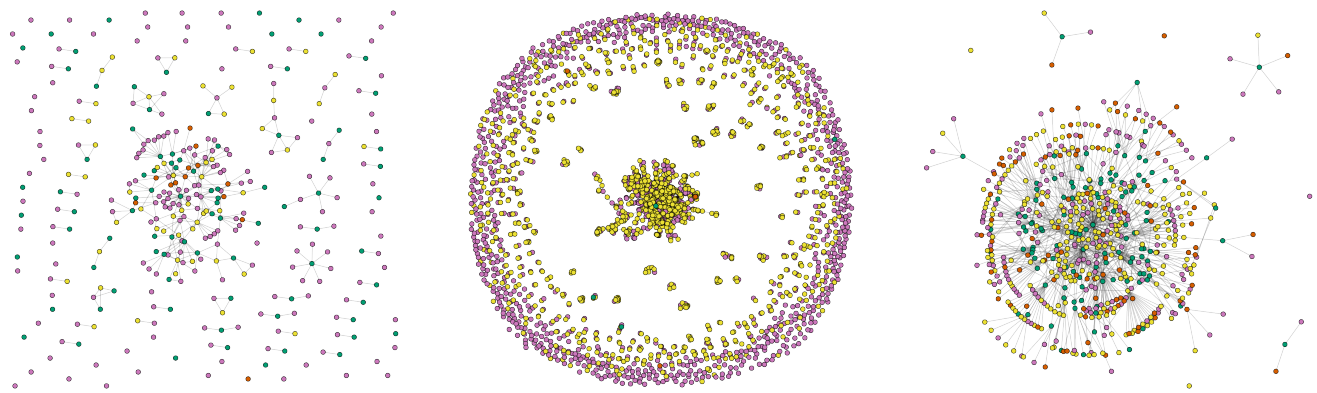


Figure 2: A sample of the networks studied here. Ballet (left) and opera (center) from Wikidata, and the PBT (right). Network composition coded as green for choreographers, purple for composers, orange for costume designers, and yellow for librettists or other staging staff.

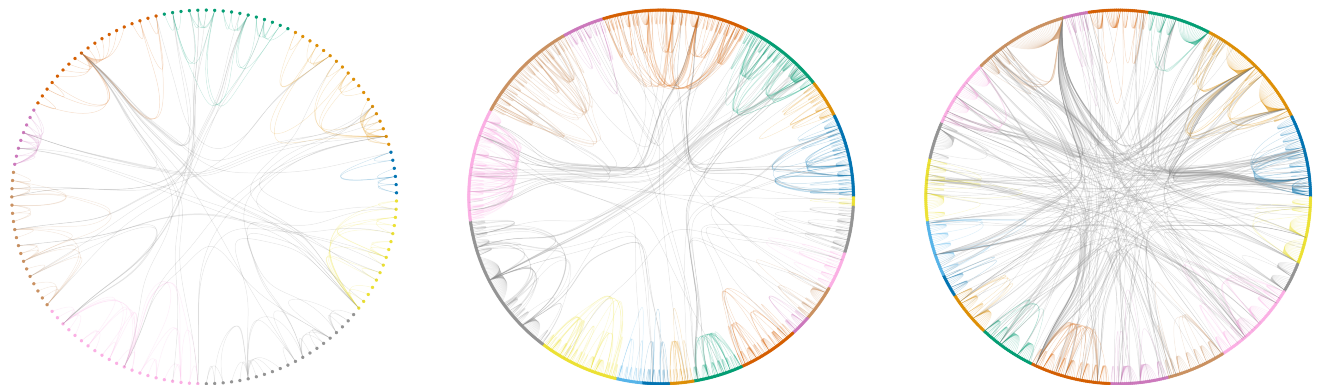


Figure 3: Community formation in performing arts. Ballet (left) and opera (center) from Wikidata, and the PBT (right). Each colour represents a community. Coloured lines are intra community links, grey lines are inter community links.

PBT artists working for the same institution do not have such personal interactions that enhance further collaborations. In opera, the low value of clustering coefficient is consistent with the network composition.

Our first hypothesis was: *Social networks of different performing arts have similar collaboration patterns*. Our results yield that performing arts networks have similar collaboration patterns, inferred by the similarity of the distribution of artists per artwork and artworks per artist, along with the presence of a largest connected component and an important proportion of disconnected elements.

Tree like shapes are observed in the network structure, suggesting a hierarchical organisation. Thus, through the application of community detection algorithms for weighted networks, we obtained the number of communities formed in our networks (values in Table 1, visual representation in Figure 3). In the opera network, despite its low diversity of artistic profiles, many communities are formed (20). This can be also a result of the high amount of nodes

belonging to its largest connected component (740 artists). Similarly, the PBT network has many communities (18 communities, 619 artists).

As observed from Figure 3, the densities and connectedness differ between networks: the PBT network is denser and its communities are highly connected. This can be due to the belongingness of the artists to the same institution. However, it is noticeable that highly connected nodes hold many intra and inter community links. This supports our second hypothesis to demonstrate that *in performing arts, communities are formed depending on the strength of dyadic ties*. Theorists studying the role of art in society claim that performing arts, such as ballet and opera, are often considered as elitist or only for experts. Several demographic studies reveal a larger preference for performing arts by people belonging to wealthy and highly educated groups, mostly of white racial classification [1, 12] (other demographic references available at culturaldata.org). In the specific case of dance, the review of Christensen et al. [5] discusses several perspectives of its behavioral and social functions, questioning the

related elitist origin and preference for selected groups. Also, recent social claims refer to the clustering of artists and a gender gap in leading positions (e.g. ABT Women's Movement, DDP reports). This suggests that artistic groups form communities that are difficult to access. Our results show strong intra community affiliations and assortative mixing, meaning that highly connected artists connect selectively with highly connected artists from other communities, playing the role of connectors in the social network. However, this characteristic does not satisfy the weak link hypothesis, and the mechanisms underlying such behaviour will require further investigation.

So far, our results suggest that *Wikidata is a useful source of historical knowledge to investigate complex social behaviors*. Data retrieved from Wikidata proved advantageous to recreate the collaboration networks of performing artists. However, the data showed names ambiguities and a notable larger amount of isolated components, affecting network metrics to investigate collaborations. This may be only an artifact of the greater coverage of Wikidata, with more profiles but missing collaborations, and points out the need to extend the efforts in modeling and populating performing arts data in Wikidata.

6 CONCLUSIONS AND FUTURE WORK

The main goal of this research was to describe the collaboration patterns of performing artists, using objective metrics from the network analysis approach. Our results suggest specific characteristics of collaborations in performing arts that show accordance with historical records. Deeper analysis should be conducted considering a more specific network model for performing arts that includes the different types of collaborations (e.g. new works, revivals, stage) and comparing metrics across different sources of data, such as web and performing arts companies. This research helps demonstrate the full potential of Wikidata as a public source of data available to model social behaviors, emphasizing the structural points that should be improved. The completion of a well structured Wikidata would help mitigate the challenges of data scarcity in performing arts. In further research, we want to explore more complex models of collaboration of performing artists, using both Wikidata and Wikipedia. We aim to contribute to the completion of current performing arts data projects, and to identify gender gaps in the wiki coverage.

ACKNOWLEDGMENTS

The authors would particularly like to thank Noshir Contractor, Jorge Fábrega, Mauricio Apablaza, and Leo Ferres for many useful comments and suggestions. The data of the Pittsburgh Ballet Theatre were generously made available by Katie Drozynski (Media Relations & Content Marketing Manager). This work is fully funded by the Universidad del Desarrollo PhD scholarship.

REFERENCES

- [1] Alan R. Andreasen and Russell W. Belk. 1980. Predictors of Attendance at the Performing Arts. *Journal of Consumer Research* 7, 2 (Sept. 1980), 112. <https://doi.org/10.1086/208800>
- [2] Albert-László Barabási. 2016. *Network science*. Cambridge University Press.
- [3] Albert-László Barabási, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. 2002. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications* 311, 3-4 (2002), 590–614.
- [4] Pierre Bourdieu. 1993. *The field of cultural production: essays on art and literature*. Columbia University Press, New York.
- [5] Julia F Christensen, Camilo José Cela-Conde, and Antoni Gomila. 2017. Not all about sex: neural and biobehavioral functions of human dance. *Annals of the New York Academy of Sciences* 1400, 1 (2017), 8–32.
- [6] Noyale Colin and Stefanie Sachsenmaier. 2016. Collaborative performance-making in context. In *Collaboration in Performance Practice*. Springer, 1–24.
- [7] Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2011. Generalized louvain method for community detection in large networks. In *2011 11th International Conference on Intelligent Systems Design and Applications*. IEEE, 88–93.
- [8] Young-Ho Eom, Pablo Aragón, David Laniado, Andreas Kaltenbrunner, Sebastian Vigna, and Dima L Shepelevsky. 2015. Interactions of cultures and top people of Wikipedia from ranking of 24 language editions. *PLoS one* 10, 3 (2015), e0114825.
- [9] Samuel P. Fraiberger, Roberta Sinatra, Magnus Resch, Christoph Riedl, and Albert-László Barabási. 2018. Quantifying reputation and success in art. *Science* 362, 6416 (Nov. 2018), 825–829. <https://doi.org/10.1126/science.aau7224>
- [10] Eduardo Graells-Garrido, Mounia Lalmas, and Filippo Menczer. 2015. First women, second sex: Gender bias in Wikipedia. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. ACM, 165–174.
- [11] Mark Granovetter. 1973. The strength of weak ties. *Amer. J. Sociology* 78, 6 (1973), 1360–1380.
- [12] Mark A Hager and Mary Kopczynski Winkler. 2012. Motivational and demographic factors for performing arts attendance across place and form. *Nonprofit and Voluntary Sector Quarterly* 41, 3 (2012), 474–496.
- [13] Yuntao Jia, Michael Garland, and John C Hart. 2011. Social network clustering and visualization using hierarchical edge bundles. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 2314–2327.
- [14] Brian Keegan, Darren Gergle, and Noshir Contractor. 2012. Do editors or articles drive collaboration?: multilevel statistical network analysis of wikipedia coauthorship. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. ACM, 427–436.
- [15] Stefan Krätke. 2002. Network analysis of production clusters: the Potsdam/Babelsberg film industry as an example. *European Planning Studies* 10, 1 (2002), 27–54.
- [16] James Moody. 2004. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American sociological review* 69, 2 (2004), 213–238.
- [17] Maureen Needham. 1997. Louis xiv and the Académie Royale de Danse, 1661–A commentary and translation. *Dance Chronicle* 20, 2 (1997), 173–190.
- [18] Mark EJ Newman. 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E* 64, 1 (2001), 016132.
- [19] Mark EJ Newman. 2001. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences* 98, 2 (2001), 404–409.
- [20] Jörg Schimmelpfennig. 2003. *Ballet*. Technical Report.
- [21] Tim Scholl. 2003. *From Petipa to Balanchine: Classical Revival and the Modernisation of Ballet*. Routledge. Google-Books-ID: N7eHAgAAQBAJ.
- [22] Paul Spencer et al. 1985. *Society and the dance: The social anthropology of process and performance*. Cambridge University Press.
- [23] Brian Uzzi and Jarrett Spiro. 2005. Collaboration and Creativity: The Small World Problem. *Amer. J. Sociology* 111, 2 (Sept. 2005), 447–504. <https://doi.org/10.1086/432782>
- [24] Claudia Wagner, David Garcia, Mohsen Jadidi, and Markus Strohmaier. 2015. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. In *Ninth international AAAI conference on web and social media*.
- [25] Stanley Wasserman, Katherine Faust, et al. 1994. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.
- [26] Dennis M Wilkinson and Bernardo A Huberman. 2007. Assessing the value of cooperation in wikipedia. *arXiv preprint cs/0702140* (2007).
- [27] Jinhyuk Yun, Sang Hoon Lee, and Hawoong Jeong. 2016. Early onset of structural inequality in the formation of collaborative knowledge, Wikipedia. *arXiv preprint arXiv:1610.06006* (2016).
- [28] Jinhyuk Yun, Sang Hoon Lee, Hawoong Jeong, et al. 2016. Intellectual interchanges in the history of the massive online open-editing encyclopedia, Wikipedia. *Physical Review E* 93, 1 (2016), 012307.
- [29] Alessia Zorloni. 2013. *The Economics of Contemporary Art*. Springer Berlin Heidelberg, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-32405-5>