

The Networks: Technical Report for the GEP Analysis Project

Pete Jones, Aresh Dadlani, Deb Verhoeven

<https://doi.org/10.60529/394>

Introduction

This report provides a detailed explanation of the technical and methodological aspects of the network-related work package of the GEP Analysis Project. Our team's aim in this report is to explain the approach taken to preparing and analysing the collaboration network datasets that were generated in the project. In particular, we address key methodological questions and how we resolved them.

Data preparation

We derive our network datasets from the GEP-Analysis key creatives dataset (see Prommer, Radziwill, and Ehrich 2023). This dataset begins in a form such that each row corresponds to a film, and all information about the film, the key creatives that worked on it, and the country/countries that produced it are expressed through variables stored in the columns. However, there are in fact 4 different types of observations in the key creatives dataset that need to be disentangled. First, the dataset contains **film** observations, and variables that describe these films (e.g. title, genre, country of production). Second, the dataset contains **people** observations, and variables that describe these people (e.g. birth year, nationality, gender). Third, the dataset contains **role** observations (which are the occasions on which *people* work on *films*), and a variable describing which role was filled by which person. Finally, the dataset contains **country** observations, with variables describing these countries (e.g. GDP, share of seats in parliament occupied by women). A first step in preparing the dataset for our analysis was to separate each of these datasets out into a separate data table for each type of observation, where rows correspond to observations and columns correspond to those variables that describe the relevant type of observations. We focus primarily on the data tables for films, people and roles, as “*The Networks*” subproject is based on three case study countries and we are not interested in quantitative country-level analysis.

The second data preparation step concerns filtering the film, role and people datasets to contain only the cases where each of our case study countries is listed as a production country (we discuss the criteria for determining each national network in the next section). Thus, we have one set of film, people and role datasets for each of our three case study countries.

The third step entails using the data structures produced by the first two steps to define the networks. For each country, the roles dataset is then used to create an affiliation matrix where rows represent unique people and columns represent unique films, and the value of the cell [p, q] is equal to 1 if person p worked on film q, and 0 otherwise. The row sums of this matrix tell us how many films each person worked on in the given country in the data period; the column sums tell us how many unique people were coded as key creatives for each film. As this is an affiliation matrix, there are two sets of network nodes described here – the people in the rows constitute one type of node, and their node attributes are the variables (e.g. gender, nationality) in the people data table; the films in the columns are a second type of node, and their node attributes are the variables (e.g. genre, title) in the film data table. At this stage, connections only exist between

these two sets of nodes, not within them. The affiliation matrix can thus be used to define a two-mode network where people are connected to the films they worked on. This two-mode network structure is illustrated in Figure 4.1.

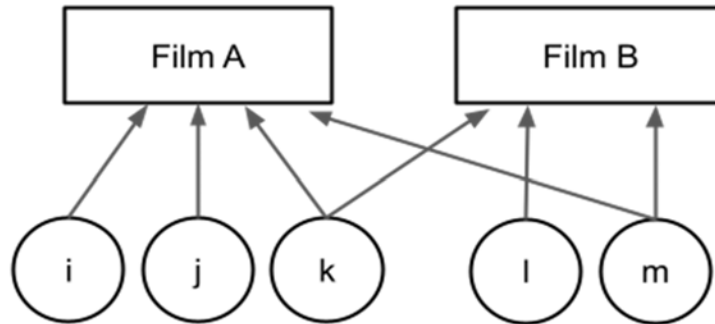


Figure 4.1: Diagram illustrating the two-mode network structure derived from the affiliation matrix. In this hypothetical example, nodes *i*, *j*, *k*, *l* and *m* represent key creatives who worked across two films, Film A and Film B. Only nodes *k* and *m* worked on both films, while *i* and *j* worked only on Film A and *l* only on Film B.

As our interest is in the interpersonal structures that arise from patterns of working together, we want to obtain an adjacency matrix where both the rows and the columns represent the set of unique people. This can be obtained from the affiliation matrix by multiplying it by its transpose. This matrix multiplication gives us an adjacency matrix where both rows and columns now represent the same set of network nodes (people), and the value of the cell $[i, j]$ is equal to the number of films in a given country that person *i* and person *j* both worked on together. This is a one-mode projection of the two-mode network, and its structure is illustrated in Figure 4.2.

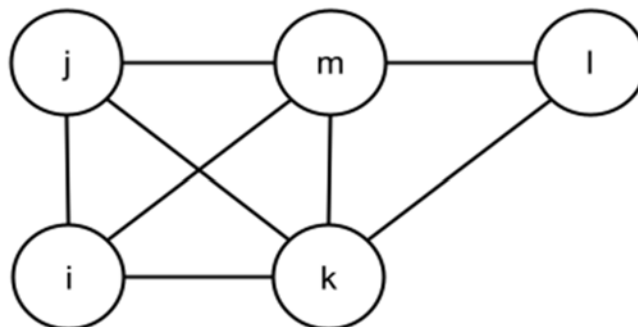


Figure 4.2: Diagram illustrating the one-mode network structure derived from the adjacency matrix obtained via row-projection of the affiliation matrix.

Here, the nodes representing key creatives in Figure 4.1 are now connected by undirected lines which represent the relation “worked on a film with”. Nodes *k* and *m* are connected to all other nodes, as they worked on both films, while *l* is only connected to *m* and *k* and *i* and *j* are only connected to *m*, *k* and each other.

As a result of these data preparation steps, we now have an affiliation matrix and an adjacency matrix for each country (Verhoeven, Jones and Dadlani 2024 a, b, c) as well as data tables describing roles, people and films. Much of the analysis undertaken in this subproject involves using the networks represented in the matrices to derive further variables that can be appended to the data tables. For example, each person’s degree centrality (described in the next section) in the one-mode network is derived from the adjacency matrix and results in a vector of scores which forms a further person-level variable “degree”. Hypothetical “what if” scenarios are enacted by stochastically modifying the values of the affiliation matrix (and in some cases the people-level variables) to reflect the effects of various alternative policy scenarios and then repeating the steps above. Figure 4.3 represents the overall data preparation process diagrammatically.

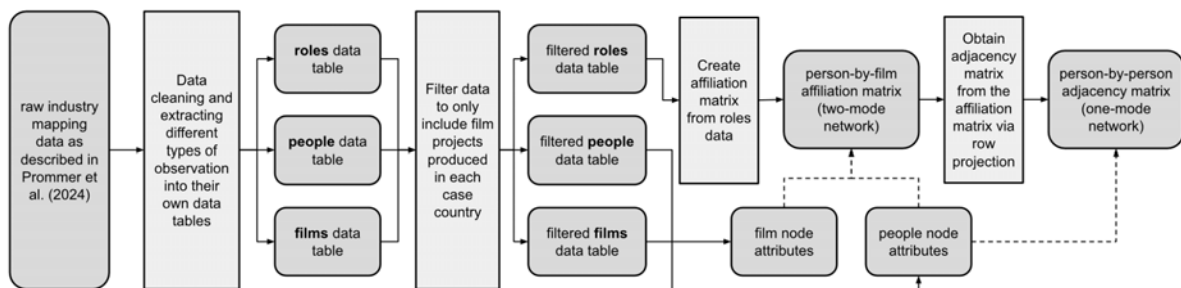


Figure 4.3: Diagram showing the data preparation pipeline to derive the network data from the initial GEP-Analysis industry mapping dataset.

Data objects are depicted in grey rounded boxes, while data processing steps are depicted in lighter grey rectangles.

All our analysis was done using free and open source software: the data preparation and “what is” analysis were undertaken in R (data wrangling utilised Tidyverse packages such as *dplyr* [Wickham et al. 2023], network analysis used many functions from the *igraph* package [Csárdi and Nepusz 2006], and data visualisations were produced using *ggplot2* [Wickham 2016]), while the “what if” analysis was conducted in Python using data visualisation and analysis libraries such as *matplotlib* (Hunter 2007), *networkX* (Hagberg, Schult, and Swart 2008), *seaborn* (Waskom 2021), and *pandas* (McKinney 2010). Lumiere’s data policies restrict the sharing of datasets that are derived from the Lumiere database, so we are not permitted to share the data files created by the process described here. However, all research code used in the project is freely available from the authors of this report upon request.

Analytical approach and key methodological decisions

With the basic data structures used in the analysis outlined, we now turn to detailing our approach to analysing these data. We group our discussion by the scope of the analysis – first, the higher-order questions which affect all subsequent analyses, then the “what is” and “what if” analyses in turn.

The first set of methodological decisions concern questions that affect our analytical approach in both the “What is” and “What if” parts of the analysis. These relate to higher-level conceptualisations of the spatial and temporal dimensions of the networks, as well as to gendered power relations. We summarise these questions and our chosen approach here.

Categorising gender

Our ability to analyse gender is dependent on the way gender is captured in the data. In our case we are reliant on the gender variable described in the previous chapter (Prommer et al. 2024), which primarily amounts to name-based guessing and ascription by coders. Throughout our analysis we refer to two main gender categories: “men” and “women and gender minorities (WGM)”. We chose to group the key creatives this way for two main reasons. First, the gender coding scheme described in Prommer et al. (2024) allowed for cases where coders learned that the person did not identify as a man or a woman. Although this resulted in only a few cases being coded as non-binary (0 in Germany, 4 in the UK, and 2 in Canada), we did not want to discard these people from the analysis. To exclude these people from the analysis and report only on men and women renders gender non-conforming people’s experiences of film industry collaboration networks invisible and reinforces a limited binary understanding of gender. However, the small number of non-binary people in the overall data makes granular analysis of gender minorities unethical. Our direct concern was thus to preserve the presence of gender non-conforming people in our analysis and to insist on a categorisation of gender as more than binary.

Second, our primary interest is in the ways in which men’s collaboration behaviours help contribute to their dominance of the way that work is distributed among key creatives in the industry. Rather than focusing on women’s under-representation, we want to focus on men’s role in the marginalisation of women and gender minorities. Grouping everyone who was not categorised by coders as a man into the WGM category allows us to contrast the network behaviours of men with those of everyone else. Power is asymmetrical, and dominance places one group above the others. We wanted our gender variable to capture this.

Defining the boundaries of the German/UK/Canada networks

A key step in our data preparation pipeline is to filter the data to contain only the films that were produced in each of our case countries prior to generating our network datasets. However, the question of what constitutes a German, UK or Canadian film is complex. In the data collected by “*The Numbers*” team, we have two variables with which to assist in identifying films from each country. The first identifies the “majority” country of production, which should usually correspond to the country most responsible for the film’s production. The second simply identifies all countries associated with the production of each film. This gives us two options for how to filter down to the films that are produced in each country: 1) a film produced by Germany/the UK/Canada is any film where Germany/the UK/Canada is listed as the majority production country, or 2) a film produced by Germany/the UK/Canada is any film where Germany/the UK/Canada appears among the list of all countries associated with the film’s production.

The first of these two options results in significantly fewer films as it filters out films where the case countries appear as a “minority” production country. From a network perspective, we find this definition too exclusive – it may be that certain key creatives are very important for bridging the national majority-production industry with the international co-productions, and this importance would be completely lost using the majority definition, where these people would appear to be on the periphery of the industry network. We therefore opted for the more inclusive second option for defining the three case country networks we analyse in the GEP Analysis Project. Throughout the project, we checked our findings against the more restrictive majority-based networks to ensure that our choice of national definition had not completely changed our arguments or interpretations. We found that the findings in both types of network were very similar, with only some small differences here and there. We therefore proceeded to focus only on the larger and more inclusive networks when reporting our results.

Measuring change over time

Evaluating how the network patterns and mechanisms change over time requires a method for temporally aggregating the film productions. There are many options for how we could approach this. All of the temporal information in the key creatives dataset is captured in the film-level variables for “production year” and “release year”. Given this, the most granular form of aggregation would be to construct a network from the films released or produced in a given year, and the least granular approach would be to split the dataset into two big time chunks based on these year variables. Several factors informed our approach to selecting from among these options. Firstly, we want enough information in each time slice that we can begin to see meaningful social structure. Secondly, we want an approach that avoids a clean cut between consecutive time slices, such that our choice of aggregation does not arbitrarily preclude network structure. To illustrate this point, consider an approach where we aggregate each two years of film releases into a network, such that films released in 2005 and 2006 form the first network, films released in 2007 and 2008 form the second network, and so on. Under this approach, two people collaborating on projects released in (using the previous example) early 2007 and late 2008 will be considered part of the same network, while a further collaboration with someone working in early 2009 will be isolated from the network structure created by the earlier collaborations as it will appear separately in the 2009-10 time slice. This would have a distorting effect on our network measures calculated on each time slice. Thirdly, we want to account for the actual pace of film production times, making it more likely that a given time window captures the full production time. Fourthly, we want to capture a notion of prior collaborations persisting structurally through time - collaboration histories do not get wiped from memory after one or two years.

Taking these matters into account, we used two approaches when analysing the networks over time. In the first, we use an overlapping four-year moving window to define time slices for the network. This approach takes all films within a four-year period and the people that worked on them and constructs the network from these projects. As such, it assumes that the most relevant ties are those that occurred in the last few years, with older collaborations no longer being included in the network. Each moving window moves forward two years, so that the final two years of each window is the first two years in the next window. Across the entire 2005-2020 period we have seven overlapping or “rolling” windows. In choosing four years as the temporal basis for defining windows we were guided by several factors. First, we considered the length of time it takes to develop and produce a project. Data on average development times were not available for the jurisdictions under study but we were guided by Stephen Follows’ work on Hollywood development and production which notes that the average time between the first announcement and eventual release date is 871 days – or two years, four months and nineteen days (Follows 2018). Secondly, our choice was informed by the use of rolling windows in other national industry reporting. Screen Australia for example, uses a three-year rolling window to calculate box-office trends and two-year rolling averages to analyse film financing as a way of mitigating the influence of outliers (Screen Australia 2023). Four years was selected as a generous option that would ensure relationships describing key creatives working collaboratively on multiple projects in succession would not be relegated to a separate network for this analysis. To ensure that working relationships formed over a period longer than four years could be also studied we adopted an additional approach.

The second approach to change over time looks at the network cumulatively, such that each year new collaborations are added but all prior observations continue to be part of the network. In the cumulative approach, the period 2005-2009 forms the initial network, and each year the new projects are incorporated as well. This approach allows us to better observe a longitudinal picture of the industry to see where patterns of behaviour are entrenched over time and where relationships are more durable.

Network centrality measures

To evaluate who occupies the strongest positions in the networks, we used several network centrality measures. Network centrality covers a vast range of different conceptualisations and operationalisations of the broad idea of a node's relative importance within network structure. We chose measures which are well-known but that we think capture important aspects of social capital in film industry networks.

The first measure is degree centrality. A node's degree is the number of connections it has to other nodes. It provides a measure of the extent of a node's local influence. In our case, degree is the number of people someone has worked with at the key creative level, and a high degree indicates that a person has many contacts and a relatively high profile among their peers. We use the unweighted degree scores, meaning that we simply count the number of connections, not the weight of those connections; this means that we are measuring the number of different people each person has worked with, regardless of how many times they worked with them.

The second measure we used is eigenvector centrality. Eigenvector is similar to degree but looks at the number of immediate connections as well as how many connections those connections have. Thus, it provides a more diffuse sense of the influence or social capital embedded in one's connections. In collaboration networks such as the film industries we examine, a node with high eigenvector centrality is a person who has worked with particularly well-connected people. As with degree, we ignore edge weights when calculating eigenvector scores.

Third, we calculated betweenness centrality scores. Betweenness measures how often a node lies on the shortest path between pairs of other nodes in the network. It offers a structural sense of a node's importance, where the extent of one's connections is less important than how those connections position one in the overall network structure. A high betweenness node is in a strong position to mediate and control the flow of information through a network as people must go through that node to reach nodes in other parts of the network. We look at both a binary version of betweenness, where the determination of a shortest path considers only whether a tie exists or not, and a weighted version wherein ties with higher weights (i.e. ties between people who have worked together on more projects) are considered "shorter" when determining path lengths.

Fourth, we adopted closeness centrality scores, particularly for the "what if" component of our analysis. Closeness measures the inverse average path length from a node to all other reachable nodes, adjusted according to the number of reachable nodes relative to the network's size. In this context, a person with high closeness centrality is somebody whose collaborations position them such that they could contact another person in the network by going through few intermediary contacts. Hence, the higher the closeness centrality of a person, the greater their ability to access and influence other people in the network. Thus, while betweenness captures the capacity for bridging/gatekeeping in the network, closeness captures overall reachability. Here edge weights are again ignored.

Defining the "network elite"

In terms of "equity" within the network, our focus is not solely on achieving numerical equality, but rather on disrupting men's outsized levels of representation among the most central network positions relative to their overall prevalence in the network when compared with WGM. Network centrality measures are treated as indicative of different kinds of influence in the networks. Specifically, our objective is to address the questions: Are men exerting a disproportionate level of influence compared to their actual representation in the industry across different measures of network influence? How does intervening in the network help align centrality with the statistical presence of gendered individuals in the industry?

Because these centrality measures are very skewed (very few people have high scores, and the vast majority of people have scores that are very low or zero) it is not meaningful to try to capture average centralities. Instead, it is useful to look at the composition of the network elite who do have high scores, which we operationalise as the Top 1% of people ranked by each centrality scores. We are mindful that our analytical focus on the Top 1% of the network can sometimes mean (especially in Canada) very small n numbers of WGM in our analysis, particularly in the rolling window strategy.

To illustrate this skew, we note that the Top 1% of people ranked by degree work on an average of nine to 13 films across countries, while the rest of the network work on an average of 1.3 to 1.7 films across countries; consequently, the average number of connections of the Top 1% is between 26 and 35 across countries, while the RON has between 4.1 and 4.7 connections on average. Similar patterns are seen for the other centrality measures; for each measure, the mean and median number of films as well as centrality scores are substantially higher among the Top 1% than the rest of the network.

“What is?”

The second set of methodological questions concern the “What is?” analysis, which aims to describe the existing gendered patterns of collaboration in the networks. Building on Jones et al. (2024), we identified a number of possible measures and mechanisms of gendered power relations that we wanted to pursue in the observed collaboration networks. We outline those here.

Homophily

One of the key network mechanisms we explored was gender homophily. In social networks, homophily is a preference for forming ties with people similar to oneself. We use Yule’s Q (Yule 1912) as a particularly effective way to measure homophily in cases where the size of the groups is not equal, as it takes into account not only how many similar and dissimilar people a node is connected with, but also the number of similar and dissimilar people not connected to the node (Crossley et al. 2015). The advantage of this is that we can assess the preference of men for working with other men (rather than WGM), taking into account their statistical dominance of the network. So, when the percentage of people that person i works with who have the same gender as i is proportionate to the underlying percentage of available collaborators who have the same gender as i , i ’s score will be 0. Therefore, positive scores represent a *disproportionate* tendency to work with people of the same gender, and negative scores represent a disproportionate tendency to work with people of a different gender. Following Bellotti et al. (2022), for each person, we calculate Yule’s Q for each year that a project came out on which they worked, restricting their available collaboration partners to the people who worked on projects released in that year; each person’s final score is the mean of their yearly scores (see Jones et al. 2024 for a detailed description of the implementation of Yule’s Q in this context). The values can range from -1 (total heterophily) to 1 (total homophily).

Familiarity

The second key mechanism we explored is the tendency for people to work with people they have worked with before. For this part of the analysis we used a team-based measure of recurring cohesion, which was introduced by Lutter (2015), and an individual measure of the propensity to work with familiar people. Lutter’s team familiarity measure first defines a dyadic weight for each pair of people working on each project f . This dyadic weight is defined by looping through all prior

films and for each defining a binary variable which is equal to 1 if both people worked on that film and 0 otherwise. This variable is then divided by the size of the project team on the film minus 1, so that prior collaborations in smaller projects are given more weight than collaborations in larger projects. These per-film values are summed to get the final weight for the dyad on film f . A film's overall team familiarity score is the sum of these dyadic weights for each pair of people on the film team, divided by the overall size of the team to give more weight to collaborations in smaller teams. We adopted Lutter's specification of this measure exactly in our analysis.

For the individual familiarity measure, we had to first identify people who worked more than once as someone who only works once in the data cannot work with the same person twice. Then, for each person, we loop through each project after their first project and separate the people on it into two groups "worked with before" and "hasn't worked with before". We obtain the familiarity score by dividing the cumulative number of people in the first group divided by the number of people in both groups. Thus, for each person who has worked more than once, we have a variable telling us the proportion of people worked with (after their first recorded project) who they have worked with before. This variable ranges between the values of 0 (when the person never works with someone they have worked with before) and 1 (when the person *only* works with people they have worked with before).

Key dyads

Network analysis allows us to go beyond evaluating the positioning and tendencies of individual nodes, widening the analytical focus to dyadic relationships themselves. In this regard, we set out to identify the important dyads in the networks. There are a number of ways that a relationship's importance can be conceptualised in a network, and we chose two that we think represent important ways in which the industries can be constrained or controlled by particular collaborative partnerships.

First, we identified those collaborative partnerships that appeared the most times in the data. In network terms, this amounts to the edge weights in the one-mode network; when projecting from a two-mode affiliation matrix to a one-mode adjacency matrix, the values of the cells become the number of times each pair of nodes were affiliated with the same films in the two-mode network. Thus, each edge between people has a weight representing the number of times the pair worked on the same film together. Our first approach to identifying key dyads ranks each edge by this weight to better understand the types of people that most often work together.

Second, we wanted to gain a more structural sense of which edges were important for holding the overall network structure together. For this we looked at edge betweenness, which is exactly the same principle as node betweenness (described above) but applied to edges: we count the number of times each edge appears on the shortest paths between all other pairs of nodes. Again, we then rank the edges by their betweenness scores to identify the most structurally important collaborative relationships. Edges with a high betweenness are those edges which connect parts of the network together which would otherwise be unconnected. Because of their bridging role, each person from one side of the bridge would have to go through the nodes forming the high-betweenness edge in order to reach any node from the other side of the bridge. This makes the relationship very important for stitching together the "who knows whom" dynamics of collaboration.

"What if?"

The goal of the "What if?" analysis is to simulate the effects of hypothetical gender equity policies on the collaboration networks and evaluate whether the resulting networks are more equitable than the original observed networks. To evaluate this, we compare the composition of the overall

network and the network elite (as defined above) to see whether men continue to disproportionately dominate the strongest network positions following the interventions. This analytical approach is the same for each experiment, so we focus here on describing *how* each hypothetical intervention was implemented, and illustrate the effects each had on the network elite.

Shadowing

In this control mechanism, our approach involves identifying newly entering WGM producers in the industry on a yearly basis. We then randomly connect each of these nodes with highly connected male producers who hold a Top 1% rank in terms of degree centrality and are actively working on projects during that particular year. We limit the number of WGM who shadow to the number of available men in the network elite. So, for example, in Canada there are four men producers in 2007-10 available to mentor one of the 41 WGM newly entering the network each. We randomly assign a WGM candidate to the available man producer (meaning 37 WGM do not receive this opportunity). The WGM who shadow keep their existing projects in the network.

Intuitively, establishing connections between newly entering WGM and influential male producers should have a positive impact on the overall position of WGM. This is because, in addition to their own projects, they instantly become connected to everyone within the teams led by the shadowed male producers. The average number of new WGM selected for shadowing influential male producers annually amounts to 31, 27 and 12 in Germany, the UK, and Canada, respectively. We observe positive outcomes for the gender composition of the Top 1% ranked people across all four centrality measures. Moreover, if this intervention had been implemented in the industry after the initial time window in our data, the network centrality measures for the Top 1% of the film network would have already achieved gender equity in Canada and the UK.

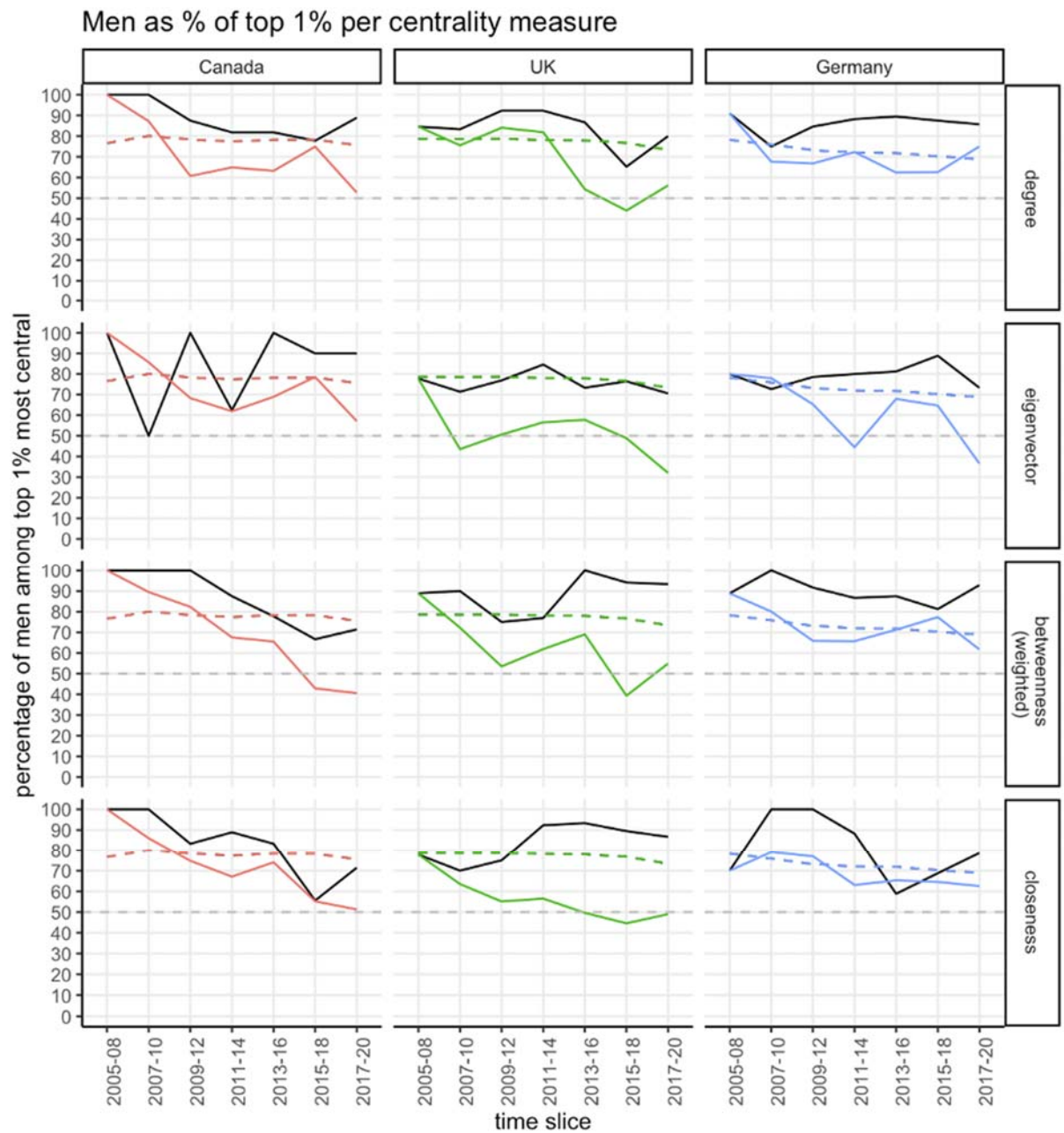


Figure 4.4 Effects of hypothetical shadowing intervention

Graphs indicate the percentage of people in the “network elite” in each cumulative network snapshot that are men, grouped by country and centrality measure used to determine the network elite in the original (black) and controlled (coloured) networks. The dashed reference line indicates the underlying percentage of all people in that cumulative network snapshot that are men.

Gender inclusive quotas

In this approach, we identified all-male projects and randomly toggled the gender category of one team member to WGM, thus simulating a “tokenistic” approach to gender equity policy wherein each film only needs to have one woman or gender minority to meet its quota (commonly referred to as the “Smurfette principle”). Because we are artificially re-coding many men in the data as WGM, the intervention increases the proportion of people that are WGM in Germany, the UK, and Canada, respectively, by 46.63 percent, 56.93 percent, and 61.37 percent. However, in comparison with this compositional change, the intervention’s effect on the access of WGM to the Top 1% of network centrality score ranks is minimal across each country and measure. This can be seen in Figure 4.5, where the reference line and the line representing the network elite in the controlled environment move in a parallel manner with the lines corresponding to the original networks.

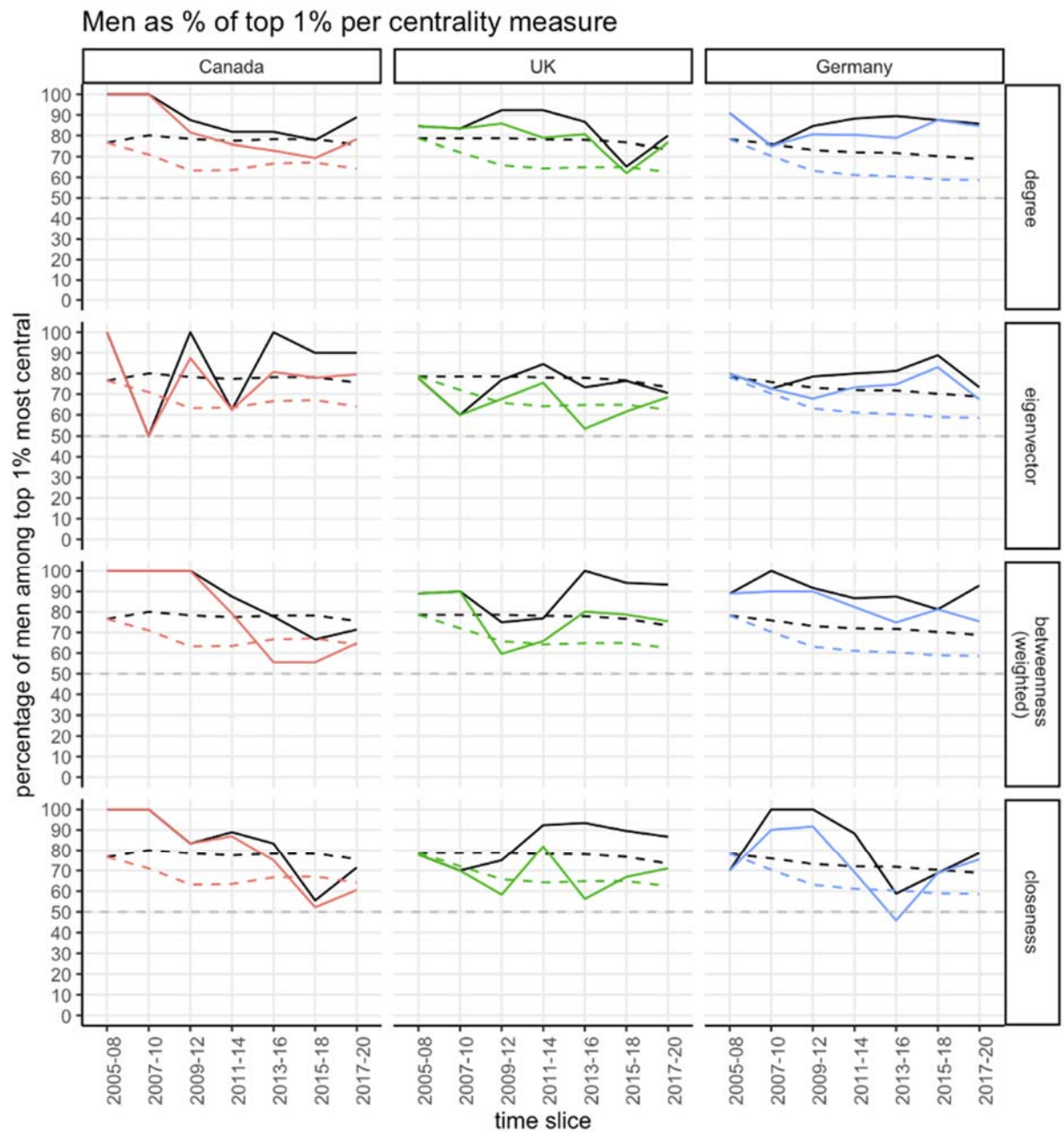


Figure 4.5: Effects of hypothetical gender inclusive quotas intervention

Graphs indicate the percentage of people in the “network elite” in each cumulative network snapshot that are men, grouped by country and centrality measure used to determine the network elite in the original (black) and controlled (coloured) networks. The dashed reference line indicates the underlying percentage of all people in that cumulative network snapshot that are men.

Gender equity quotas

To assess the potential impact of gender equity quotas, we artificially imposed a balanced 50-50 gender distribution in every male-dominated project. To do this, we identified the projects with more men than WGM in their key creative teams and randomly selected men within each team and changed their gender category to WGM until there were as many WGM as men. This gender transformation was carried forward in subsequent time periods to observe the evolutionary dynamics of the network. We found that among all the films produced in Germany, the UK, and Canada, 76.2 percent, 82.9 percent, and 81.8 percent of them respectively consist of teams where men constitute the majority. Moreover, this implementation of gender equity quotas resulted in replacing 25 percent, 29.3 percent, and 29.7 percent of the men in Germany, the UK, and Canada with WGM, respectively. The substantial replacement of men with WGM reduced the overall count of men to nearly 40 percent in all three networks. However, men's over-representation in accessing the network elite remains intact compared to their overall presence in the three jurisdictions. Nevertheless, men continue to over-perform in all four centrality measures in terms of accessing the network elite. Therefore, balancing the gender proportion in men-dominated projects does not disrupt men's disproportionate dominance in the network elite.

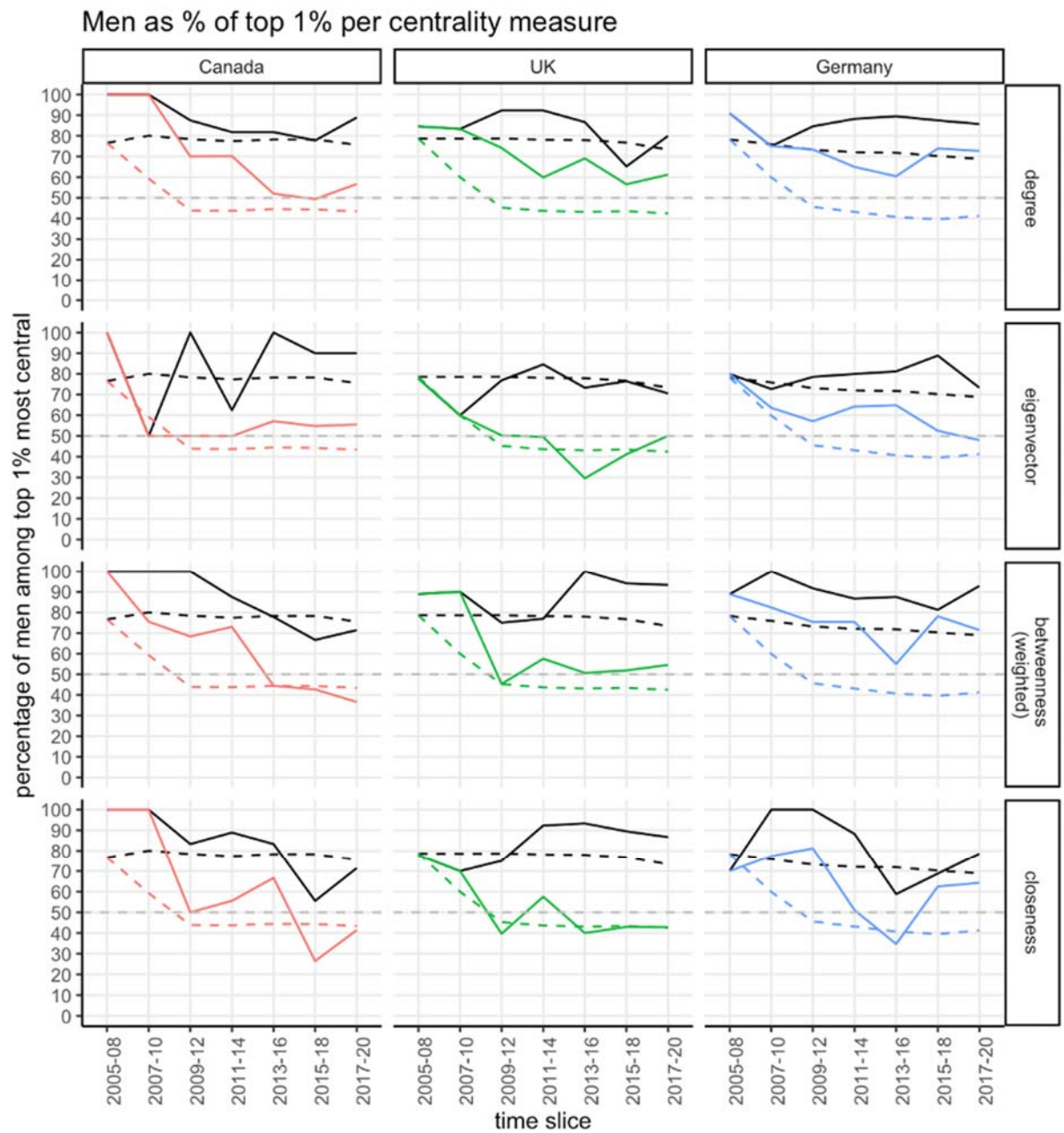


Figure 4.6: Effects of hypothetical gender equity quotas intervention

Graphs indicate the percentage of people in the “network elite” in each cumulative network snapshot that are men, grouped by country and centrality measure used to determine the network elite in the original (black) and controlled (coloured) networks. The dashed reference line indicates the underlying percentage of all people in that cumulative network snapshot that are men.

Removing key players

In this intervention, we remove “key players” from the network – a set of nodes who, when taken as a group, are particularly central in the network structure. The intervention is designed to “open up” the networks by removing the influential nodes who glue the system together. Here we used our own implementation of the fragmentation- and cohesion-based measures offered by Borgatti for identifying sets of key players (Borgatti 2006). The basic idea behind this approach is that the goal is to identify a set of nodes who, when taken as a group, either have the most potential to disrupt and disconnect the network were they to be removed (fragmentation centrality) or who are the most proximate to the other nodes in the network in terms of graph distances (cohesion centrality). The researcher specifies the size k of the key player set they want to find and the algorithm then samples k nodes and continues to swap nodes in and out of the set and evaluate whether the set (treated as if it were one super-node with all its members’ connections combined) is more central after these swaps, until it converges on an optimal solution. Here, we used a key player set size of ten. Each year, we add the new projects to the network, find the key player set, remove the key players, and continue.

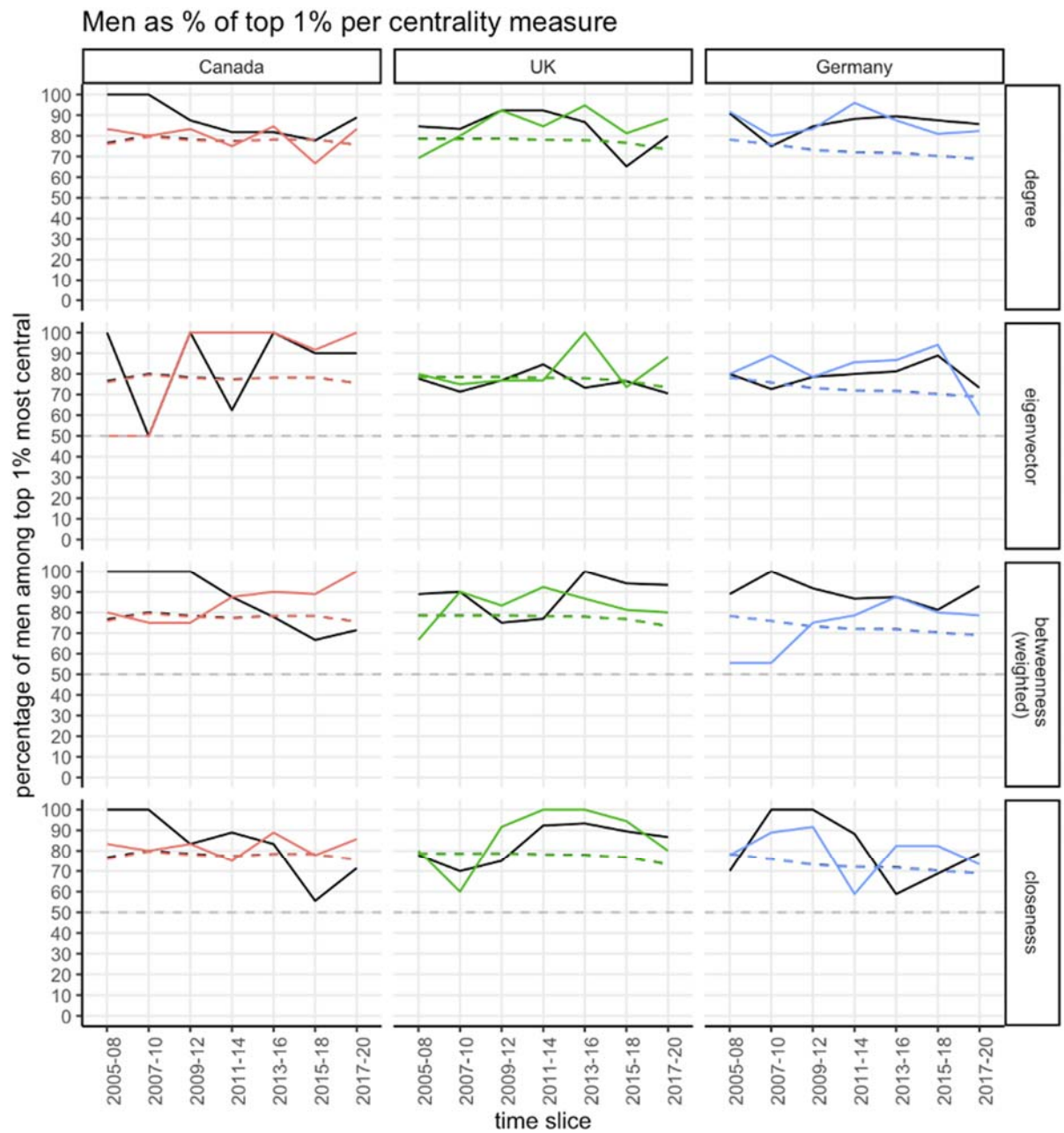


Figure 4.7a: Effects of removing key players (fragmentation-based)

The percentage of people in the “network elite” in each cumulative network snapshot that are men, grouped by country and centrality measure used to determine the network elite in the original (black) and controlled (coloured) networks. The dashed reference line indicates the underlying percentage of all people in that cumulative network snapshot that are men.

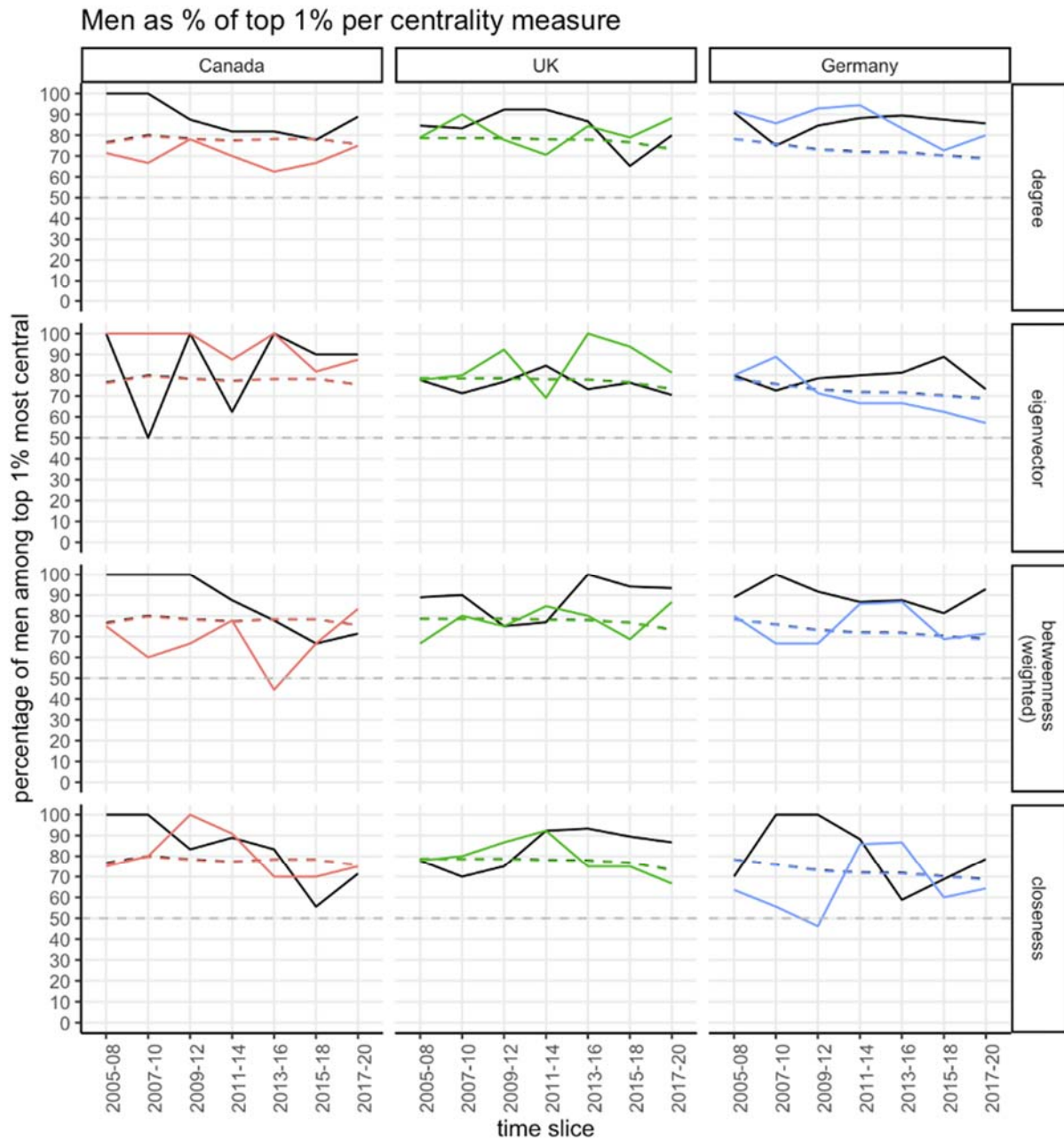


Figure 4.7b: Effects of removing key players (cohesion-based)

The percentage of people in the “network elite” in each cumulative network snapshot that are men, grouped by country and centrality measure used to determine the network elite in the original (black) and controlled (coloured) networks. The dashed reference line indicates the underlying percentage of all people in that cumulative network snapshot that are men.

The fragmentation-based method for identifying key players to remove is almost entirely ineffective at disrupting men’s dominance of the network elite in all cases (fig. 4.7a). The cohesion-based method shows greater impact on the networks, but the results are mixed, with WGM either gaining or losing ground depending on measure and country (fig. 4.7b). Overall, the removal of key players proposed here does not lead to a power shift between genders in all three jurisdictions.

Removing mediocre men

For this intervention we aim to explore the impact of removing projects by “mediocre men” producers, which we defined as male producers who have worked on exactly two projects and have no projects in the first and last time slice of the network (2005-2008 and 2017-2021). In seeking to quantify a definition of “mediocre” we used a simple metric; men who worked on more films than the minimum (i.e. they made more than one movie) but who also did not separate themselves from the majority of their peers during the collection period (i.e. they did not make more than two films). Since we couldn’t assume that filmmakers present in the first time slice did not have a preceding trove of projects and nor could we assume that men isolated to the last time-slice would not proceed to further potential projects, we excluded those men who were only active in these time-slices from the definition of “mediocre”.

If a mediocre man producer is the sole producer of a project, we eliminate the entire project from the network. This however, does not entail removing all the people associated with that particular film from the network if they participate in other projects. If a project involves a team of producers, we selectively remove only the mediocre men producer from the project while retaining the rest of the individuals involved.

The impact of removing mediocre men producers on the overall network sizes is minimal (see fig. 4.8). Germany had the most people eliminated from the network (43 men and seven WGM), followed by the UK (34 men and one WGM), and then Canada (25 men and one WGM). Moreover, it does not reduce the over-representation of men in the network elite. Men who hold structural power in the network are least affected by this strategy and thus continue to dominate the strongest network positions in all jurisdictions for almost all centrality scores. Thus, the removal of mediocre men producers and their projects has minimal effect since mediocre men were unlikely to be in the network elite to begin with. Accordingly, the removal does not lead to any significant improvement in the ability of WGM to access the network elite.

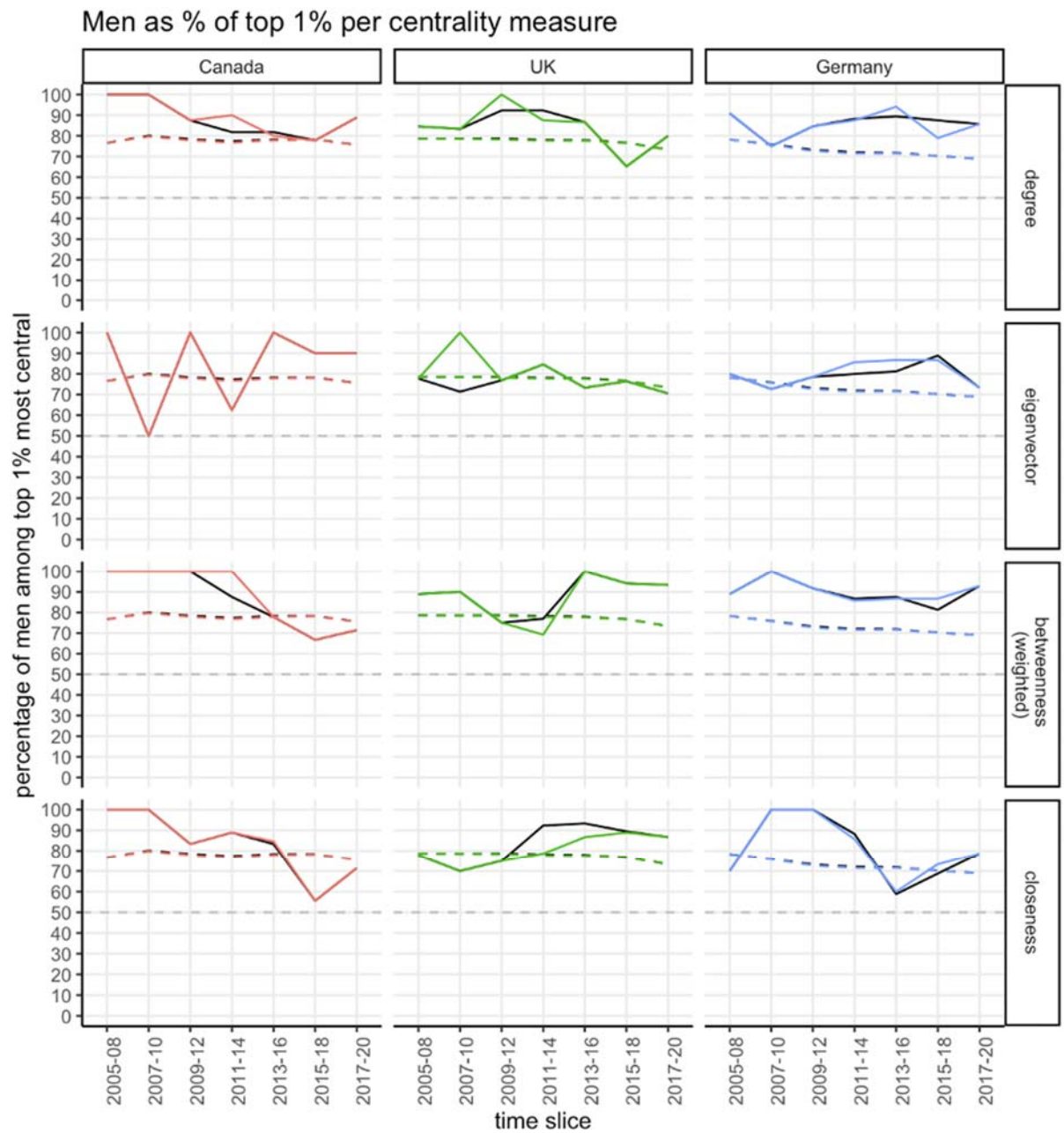


Figure 4.8: The percentage of people in the “network elite” in each cumulative network snapshot that are men, grouped by country and centrality measure used to determine the network elite in the original (black) and controlled (coloured) networks.

The dashed reference line indicates the underlying percentage of all people in that cumulative network snapshot that are men.

Removing projects with all-men key creative teams

A further “what if” scenario we explored related to the removal of films with only men in their key creative teams. This approach retroactively simulates a policy that refuses to allow films with only men occupying all the roles of writer, director and producer on a film. Specifically, we identify films that have only men in the key creative team and remove them from the roles data table prior to creating the affiliation and adjacency matrices; we also filter the films and people data tables to contain only the people and films remaining after deleting the all-men teams. Table 4.1 outlines the statistical impact of this intervention on each network. We found that, under this scenario, some 42 percent of films in the Canadian and UK film industry would not have been made, removing almost half of all the men from the network. In Germany, a lower percentage of films would have been prevented (36.5 %), leading to a lower percentage of men being deleted overall (38.9 %). Statistically this would have a significant impact on the number of men in the networks of each industry. But what would the effect be on their network position?

Table 4.1: The number and percentage of projects and men removed from the original networks as a result of deleting projects with men-only key creative teams.

	Canada	UK	Germany
Number (%) of films with all-men creative teams	366 (42 %)	720 (42.6 %)	710 (36.5 %)
Number (%) of men removed	783 (45.7 %)	1,454 (44.6 %)	1,155 (38.9 %)

When evaluating the network impact of this strategy, we consider the composition of the overall network as well as the network elite in the new networks where all projects have at least one WGM in a key creative role. Even after deleting so many men, the networks still remain male-dominated, with men accounting for 61 percent of the German network, 64 percent of the UK network, and 64.4 percent of people in the Canadian network after the all-male projects are removed. However, as Figure 4.9 shows, men’s capacity to access the network elite at a greater rate than their overall prevalence is significantly reduced when looking at the networks covering the whole data period.

In Canada, men continue to overperform on eigenvector and betweenness centrality, but WGM account for almost 60 percent of the Top 1% of ranked degree scores. In the UK, WGM prove to be excellent “networkers”, overperforming their headcounts by improving to account for 40-50 percent of the top ranked scores across each measure except eigenvector. In Germany, the gender composition of the Top 1% of centrality scores is almost the same as the overall gender composition of the network, deviating by less than 5 percent on each measure. This suggests that, in Germany, removing men-only projects does have the effect of also disrupting men’s disproportionate dominance of the network elite in the overall network covering the full data period.

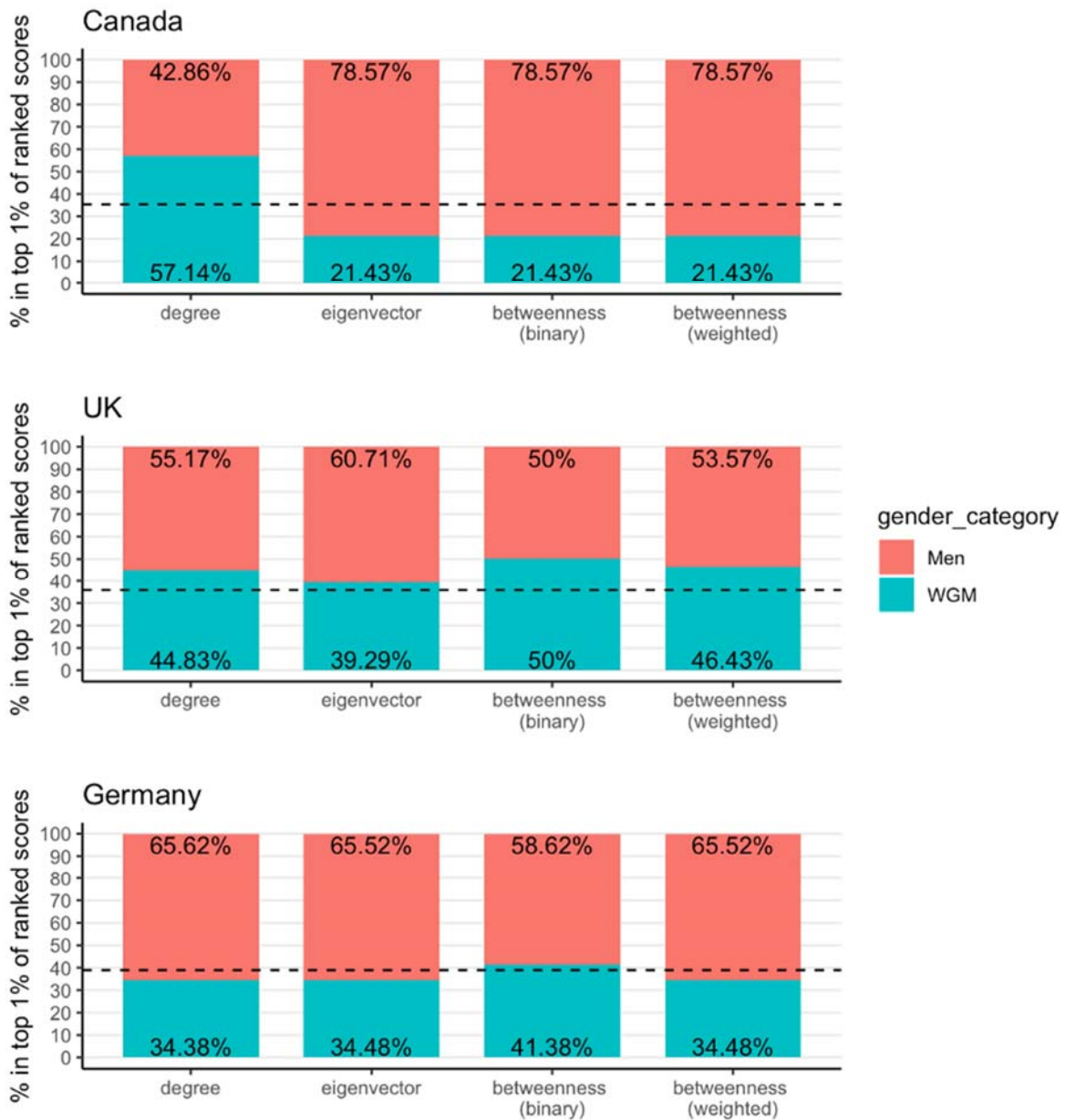


Figure 4.9: Percentage of people in the Top 1% in each country by gender category after removing men-only projects

Dashed reference line indicates the underlying percentage of people overall in the network that are not men after removing men-only projects.

Looking at the network elite over time using the moving windows (fig. 4.10), we can see that the overabundance of men in these positions is largely eliminated in Canada and the UK, where the percentage of the network elite that are men decreases to levels that are consistent with or below the overall percentage of men in the network. In Germany, there is less improvement in terms of the composition of the network elite, as men actually increase their share of the top network positions over time, and are still over-represented for degree and eigenvector centrality.

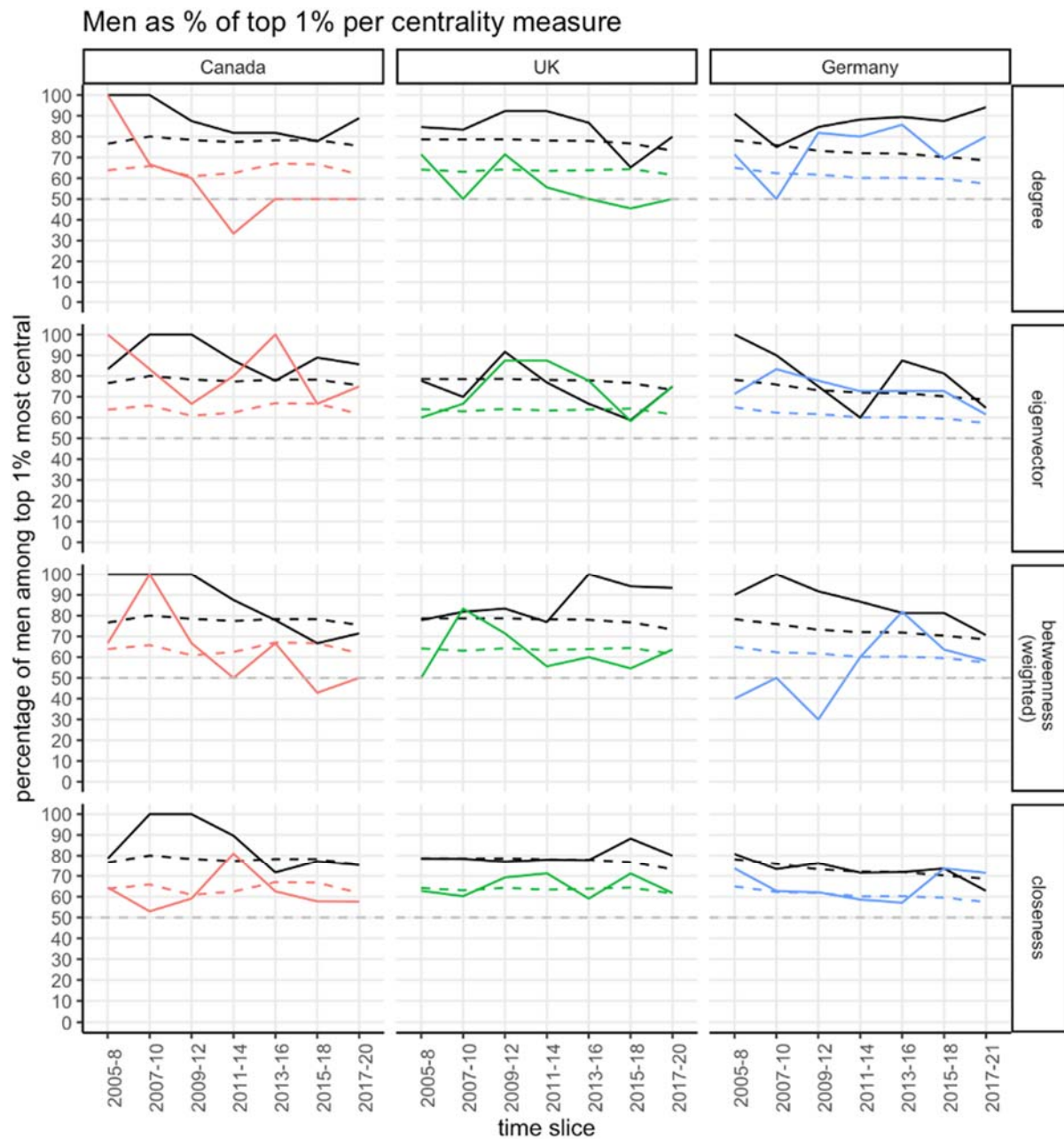


Figure 4.10: The percentage of people in the “network elite” in each time window that are men, grouped by country and centrality measure used to determine the network elite.

Black lines correspond to the original, unaltered networks; coloured lines correspond to the “what if” networks after removing men-only projects. The dashed reference line indicates the underlying percentage of all people in that time window that are men.

References

- Bellotti, Elisa, Dominika Czerniawska, Martin G. Everett, and Luigi Guadalupi. 2022. "Gender Inequalities in Research Funding: Unequal Network Configurations, or Unequal Network Returns?" *Social Networks* 70 (July): 138–51. <https://doi.org/10.1016/j.socnet.2021.12.007>.
- Borgatti, Stephen P. 2006. "Identifying Sets of Key Players in a Social Network." *Computational and Mathematical Organization Theory* 12 (April): 21–34. <https://doi.org/10.1007/s10588-006-7084-x>.
- Crossley, Nick, Elisa Bellotti, Gemma Edwards, Martin G Everett, Johan Koskinen, and Mark Tranmer. 2015. *Social Network Analysis for Ego-Nets*. London: SAGE Publications Ltd. <https://doi.org/10.4135/9781473911871>.
- Csárdi, Gábor, and Tamás Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems*. <https://igraph.org>.
- Follows, Stephen. 2018. "How Long Does the Average Hollywood Movie Take to Make?" Stephen Follows. 7 May 2018. <https://stephenfollows.com/how-long-the-average-hollywood-movie-take-to-make/>.
- Guyan, Kevin, Doris Ruth Eikhof, and Amanda Coles. 2024. "The Policies: Technical Report for the GEP Analysis Project." In *Technical Appendix of the Gender Equity Policy (GEP) Analysis Project Report "Re-Framing the Picture"*, edited by Skadi Loist and Martha E. Ehrich, 6–33. Potsdam: Film University Babelsberg KONRAD WOLF.
- Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart. 2008. "Exploring Network Structure, Dynamics, and Function Using NetworkX." In *Proceedings of the 7th Python in Science Conference*, edited by Gaël Varoquaux, Travis Vaught, and Jarrod Millman, 11–15. Pasadena, CA USA.
- Hunter, John D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Jones, Pete, Deb Verhoeven, Aresh Dadlani, and Vejune Zemaityte. 2024. "She Must Be Seeing Things! Gender Disparity in Camera Department Networks." *Social Networks* 76 (January): 120–34. <https://doi.org/10.1016/j.socnet.2023.09.004>.
- Lutter, Mark. 2015. "Do Women Suffer from Network Closure? The Moderating Effect of Social Capital on Gender Inequality in a Project-Based Labor Market, 1929 to 2010." *American Sociological Review* 80 (2): 329–58. <https://doi.org/10.1177/0003122414568788>.
- McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman, 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.
- Prommer, Elizabeth, Sophie Radziwill, and Martha E. Ehrich. 2024. "The Numbers: Technical Report for the GEP Analysis Project." In *Technical Appendix of the Gender Equity Policy (GEP) Analysis Project Report "Re-Framing the Picture"*, edited by Skadi Loist and Martha E. Ehrich, 34–42. Potsdam: Film University Babelsberg KONRAD WOLF.
- Screen Australia. 2023. "Cinema Industry Trends – Box Office in Australia, 1977–2022." Screen Australia. 2023. <https://www.screenaustralia.gov.au/fact-finders/cinema/australian-films/feature-film-releases/box-office-share>.
- Verhoeven, Deb, Pete Jones, and Aresh Dadlani. 2024 a. "GEP-Analysis Project: Film Industry Affiliation and Adjacency Matrices (Canada)." *Figshare*. <https://doi.org/10.6084/m9.figshare.25194251.v1>.

- Verhoeven, Deb, Pete Jones, and Aresh Dadlani. 2024 b. "GEP-Analysis Project: Film Industry Affiliation and Adjacency Matrices (Germany)." *Figshare*. <https://doi.org/10.6084/m9.figshare.25194254.v1>.
- Verhoeven, Deb, Pete Jones, and Aresh Dadlani. 2024 c. "GEP-Analysis Project: Film Industry Affiliation and Adjacency Matrices (UK)." *Figshare*. <https://doi.org/10.6084/m9.figshare.25201709.v1>.
- Waskom, Michael. 2021. "Seaborn: Statistical Data Visualization." *Journal of Open Source Software* 6 (60): 3021. <https://doi.org/10.21105/joss.03021>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. <https://dplyr.tidyverse.org>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016. Use R! Cham: Springer. <https://doi.org/10.1007/978-3-319-24277-4>.
- Yule, G. Udny. 1912. "On the Methods of Measuring Association Between Two Attributes." *Journal of the Royal Statistical Society* 75, no. 6: 579–652. <https://doi.org/10.2307/2340126>.