Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **00**, 1–20
https://doi.org/10.1093/jrsssa/qnad009

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS Journal of the Statistics Society Series A Statistics in Society

Original Article

Modelling the large and dynamically growing bipartite network of German patents and inventors

Cornelius Fritz[1], Giacomo De Nicola[1], Sevag Kevork[1], Dietmar Harhoff[2] and Göran Kauermann[1]

[1]Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, Munich, 80539 Bavaria, Germany
[2]Max Planck Institute for Innovation and Competition, Marstallpl. 1, Munich, 80539 Bavaria, Germany

Address for correspondence: Cornelius Fritz, Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, Munich, 80539 Bavaria, Germany. Email: cornelius.fritz@stat.uni-muenchen.de

Abstract

To explore the driving forces behind innovation, we analyse the dynamic bipartite network of all inventors and patents registered within the field of electrical engineering in Germany in the past two decades. To deal with the sheer size of the data, we decompose the network by exploiting the fact that most inventors tend to only stay active for a relatively short period. We thus propose a Temporal Exponential Random Graph Model with time-varying actor set and sufficient statistics mirroring substantial expectations for our analysis. Our results corroborate that inventor characteristics and team formation are essential to the dynamics of invention.

Keywords: bipartite networks, co-inventorship networks, inventors, knowledge flows, patent collaboration, temporal exponential random graph models

1 Introduction

In the social sciences, bipartite networks are often used to represent and study affiliation of the actors to some groups (such as directors on boards (Friel et al., 2016), or football players in teams (Onody & de Castro, 2004)) and participation of people to events (such as researchers citing papers (Small, 1973), or actors in movies (Ahmed et al., 2007)). Research on bipartite structures initially focused on unimodal projections of the networks (Breiger, 1974), where we consider two nodes of one type to be tied if they share at least one actor of the other kind. This practice forces the researcher to give priority to one type of node over another and thus comes with a loss of possibly relevant information (Koskinen & Edling, 2012). Direct bipartite network analysis has first been considered in Borgatti and Everett (1997), where traditional network analysis techniques are systematically discussed for bipartite networks. Latapy et al. (2008) further adjusted known concepts from unipartite networks, such as clustering and redundancy, to the bipartite case, with a focus on large networks.

For this paper, we consider high-dimensional bipartite networks where actors are related to one another through instantaneous events, which by definition only occur once. In particular, we focus on the network formed by inventors residing in Germany and patents submitted between 1995 and 2015, where a tie between an inventor and a patent is present if the individual is listed among the patent's inventors. The resulting data structure is visualised in Figure 1a, where we can assign each patent (or event, in the jargon of bipartite network analysis) to a time point and a set of co-inventors. For instance, inventors A and B filed the joint patent with ID 1. We may represent the bipartite network structure as an adjacency matrix with entries Y_{ij}, where

Received: January 21, 2022. **Revised:** January 4, 2023. **Accepted:** January 19, 2023

$$Y_{ij} = \begin{cases} 1 & \text{if actor } i \text{ is on patent ID } j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

and $i \in \mathcal{I}$ and $j \in \mathcal{K}$, where we denote the complete set of inventors and patents by $\mathcal{I}$ and $\mathcal{K}$, respectively. In our example, this bipartite network is of massive dimensions, with $|\mathcal{I}| = 78.412$ inventors on a total of $|\mathcal{K}| = 126.388$ filed patents.

The data allow us to gain insight into the dynamics and drivers of innovation, collaboration, and knowledge flows in the private sector. Moreover, inventorship status on a patent is more legally binding than authorship of academic papers, suggesting a greater degree of validity of the results of network analysis in this context. The data, however, present some obstacles to their study. First, the complete network is too massive, making analysis with most traditional network techniques prohibitive. Second, the data carry structural zero entries since not all inventors are active during the entire time period between 1995 and 2015. This phenomenon is partially due to the retirement of inventors, who hence suffer from natural 'actor mortality'. Moreover, inventors may change their career track, e.g., by moving into managerial positions and ending their patenting activities, thus reinforcing the aforementioned actor mortality in our data. Vice versa, new inventors continuously enter the picture by producing their first patent, resulting in what we can call 'actor natality' in the network. These aspects imply that the bipartite network matrix at hand contains structural zeros for inventors which are not active at particular time points. To incorporate this feature into a statistical network model, we consider the network dynamically and discretise the time dimension by looking at yearly data, such that time takes values $t = 1, 2, \dots, T$, as sketched in Figure 1a. In this context, T denotes the number of observed time points. We then allow the actor set to change at each time point. For the adjacency matrix of Figure 1a, this leads to the matrix structure in Figure 1b, where e.g., inventor A retires after time point $t = 1$ and hence does not take part in the patent market at $t = 2$. To encode this information on the changing composition of actors, we define activity sets $\mathcal{I}_t$ to include all actors that are active at time point t. Further, let $\mathcal{K}_t$ denote the event set, containing all patents submitted in a particular time window. We assume that both sets are known for each time point $t = 1, \dots, T$. With this additional information, we decompose the observed massive bipartite network matrix into smaller dimensional bipartite submatrices denoted by

$$\mathbf{Y}_t = (Y_{t,ij} : i \in \mathcal{I}_t,\ j \in \mathcal{K}_t), \tag{2}$$

which are visualised for $t = 1$ and 2 by the grey-shaded areas in Figure 1b and where $Y_{t,ik}$ indicates whether inventor i is a co-owner of patent k at time point t. Instead of modelling the entire bipartite network, we break down our analysis to modelling $\mathbf{Y}_t$ given the previous bipartite networks $\mathbf{Y}_1, \dots \mathbf{Y}_{t-1}$. Incorporating the varying actor set as such in the analysis allows us to structurally account for the observed actor mortality and natality while also making the estimation problem more manageable, thus solving both issues simultaneously.

This change in perspective induces a structure that deviates from conventionally analysed networks. To accommodate for it in a probabilistic modelling framework, we extend the Temporal Exponential Random Graph Model (TERGM, Hanneke et al., 2010) towards dynamic bipartite networks with varying actor set. For TERGMs, we assume that a discrete Markov chain describes the generating process of the networks observed over time. The transition probabilities of jumping from one network to another one are determined by an Exponential Random Graph Model (ERGM, S. Wasserman & Pattison, 1996). ERGMs, on the other hand, were adapted to bipartite data by Faust and Skvoretz (1999), while adjustments to incorporate the model specifications of Snijders et al. (2006) were proposed in Wang, Robins, et al. (2013). These network models were already successfully applied to static (Metz et al., 2019) as well as dynamic networks (Broekel & Bednarz, 2018).

In addition to the dynamically varying actor set, the network at hand presents another particular feature for which we need to account in the modelling. Collaborations generally build up over time, rather than being confined to single time points. To adequately represent these mechanisms, we need to include covariate information from the past and on the pairwise level of one actor set in the model, which has not yet been implemented in the bipartite ERGM framework. We, therefore,

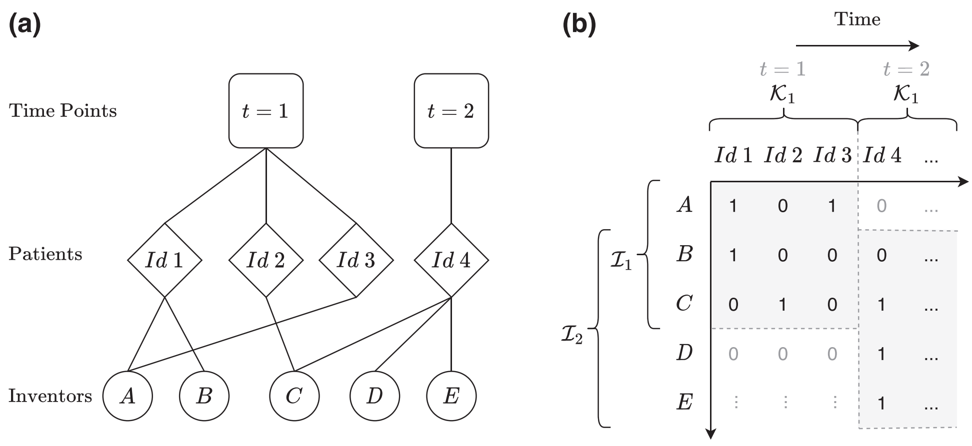

Figure 1. On the left side (a), the tripartite network structure of the patent data is illustrated with an example encompassing four patents (tilted squares) submitted at two different time points (squares with rounded edges) by five inventors (circles). The corresponding adjacency matrix is depicted on the right side (b). The two sub-matrices defined in (2) are shaded in grey are $\mathbf{Y}_1$ in the top left and $\mathbf{Y}_2$ on the bottom right. (a) Network structure and (b) adjacency matrix structure.

define and include sufficient network statistics in our model to account for this particular kind of dynamic interdependence.

Overall, the contributions of this paper are the following: We demonstrate how massive bipartite networks can be broken down in a way that allows their analysis, and propose sufficient statistics for modelling bipartite temporal networks with varying actor set. Using the proposed methodology, we then contribute to the growing literature on innovation by analysing a comprehensive patent and inventor population dataset. In particular, we study the network composed of all electrical engineering patents filed between 1995 and 2015 by inventors located in Germany. The unusually rich dataset allows us to study patterns of team formation in a more refined way than has been feasible to date. In particular, our modelling approach enables us to quantify how factors such as spatial proximity, teamwork, interlocking of collaborations, gender, and seniority affect the output of inventors. By answering these questions, our study contributes empirical findings to current discussions on the role of gender and seniority in innovation and, more generally, in the workplace.

The remainder of the paper is organised as follows: Section 2 gives a literature overview of the research in patent data. In this section, we also describe the data in detail. Section 3 motivates the model and introduces its novelties in more detail from a theoretical perspective. We present the results of our empirical analysis in Section 4, while Section 5 wraps up the paper with some concluding remarks.

2 Patent data

2.1 Research on patents and inventor teams

The analysis of patents and their impact and evolution over time is an important area of current economic research. Hall and Harhoff (2012) provide a general overview of the field and its recent developments. The holder of a patent receives a temporary right (typically for 20 years) to exclude others from using the patented technology. The patent right can be extremely valuable, e.g., when it becomes the foundation of an economic monopoly. Hence, patents can create powerful incentives and induce invention and innovation efforts. In addition, patents require disclosure of the patented invention and thus may invite others to build on the patented technology. These benefits have to be held against the welfare losses due to reduced competition. The study of patents in much of classic economic literature revolves around the trade-offs between these effects. Patent data are also often used in innovation research to explore how new technologies develop and spread, which innovation areas are the most active, how innovation areas and sub-areas are connected with one another, and how productive firms or nations are with regards to their patenting output. Patents contain references to prior patents, so-called patent citations (Alstott et al., 2017). The study of patent citations and the network structures they form have become an important part of innovation economics, since citations can be interpreted as an indicator of knowledge flows. The study of citation networks has been an important area of research at least since the work of Garfield (1955) (see also de Solla Price, 1965; Egghe & Rousseau, 1990). Co-authorship networks have been extensively studied within the area of research publications (see, e.g., Leifeld, 2018; Melin & Persson, 1996; Newman, 2004). The techniques developed for general citation networks can naturally be applied to map patent citation networks as well (see, e.g., Li et al., 2007; Verspagen, 2012; von Wartburg et al., 2005). Moreover, since patent data always indicate the identity of the inventors contributing to the invention, they can be used to study the characteristics of inventor teams and inventor collaboration networks. The focus is then shifted from citations to co-inventorship of patents.

In both cases, i.e., patent citations and inventor teams, modern methods of network analysis can be applied to answer open research questions. In terms of the research questions tackled, our study differs substantially from patent citation studies, since we do not focus on knowledge flows, but rather on the logic of inventor team formation. We share this focus with studies of authorship teams in academia, but we note an important institutional difference: More than 93% of patents are filed by private enterprises (Giuri et al., 2007). Other than in scientific co-authoring, the composition of co-inventor teams does not just reflect the preferences of the authors (inventors), but it involves, in almost all cases, a managerial decision that is guided by profit concerns. Thus, the patterns we uncover in our analysis are not just a reflection of individual preferences, but also of the employer's productivity calculus. This feature of our setting will be particularly important when interpreting results and comparing them to results from other studies (e.g., for gender homophily).

For patent data, it is possible to construct the co-inventorship network in two main ways. One can directly analyse the bipartite network formed by the patents and their inventors (see, e.g., Balconi et al., 2004). Alternatively, one projects the bipartite structure on one of the two modes, which in the context of patent data is usually that of inventors. This entails a network composed only of inventors, in which two nodes (inventors) are connected if they have at least one patent in common (Bauer et al., 2022; Ejermo & Karlsson, 2006). Much of the literature in this area utilises such projections, since models for unimodal networks are developed to a greater extent. Several studies have used unipartite ERGMs to study knowledge diffusion networks in various domains (see e.g., Jiang et al., 2013; Keegan et al., 2012). As ERGMs allow for modelling networks with different types of ties (see Chen, 2021, for an overview), it is also possible to simultaneously model inventor-patent ties in a unipartite, multilayer network context, as done by Jiang et al. (2015). As explained in the introduction, however, projecting everything on one mode inevitably results in a loss of information on the mode that is excluded.

In the case of patents and inventors, the fact that two inventors collaborated on many patents together, together with the size of these patents, brings much information which is not available in the projection, where the inventors are simply linked together. This loss of information is made apparent by the fact that there are many bipartite graphs which lead to the same projection (Latapy et al., 2008). Preserving the original bipartite structure thus enables us to gain more detailed and accurate insight on the mechanisms at play by estimating effects which would not be visible by considering the projection, as will be shown in the application section.

2.2 Data description

We consider patent applications submitted to the European Patent Office or the German Patent and Trademark Office (Deutsches Patent- und Markenamt) between 1995 and 2015. More specifically, we look at patents filed within the main area of electrical engineering, for which at least one of the inventors listed on the patent has a residential address in Germany. For assigning each patent to a single time point, we use the priority date, i.e., the first-time filing date of a patent (which precedes the publication and the grant date). We focus on electrical engineering as it is one of the largest main areas and as it has seen particularly high growth rates since 2010. Moreover, collaborations between inventors are commonplace in this field. For our analyses, we focus on the data starting in 2000 and condition on the information from the first five years considered (i.e., from 1995 to 1999) to derive covariates from them. The dataset can be represented as a massive

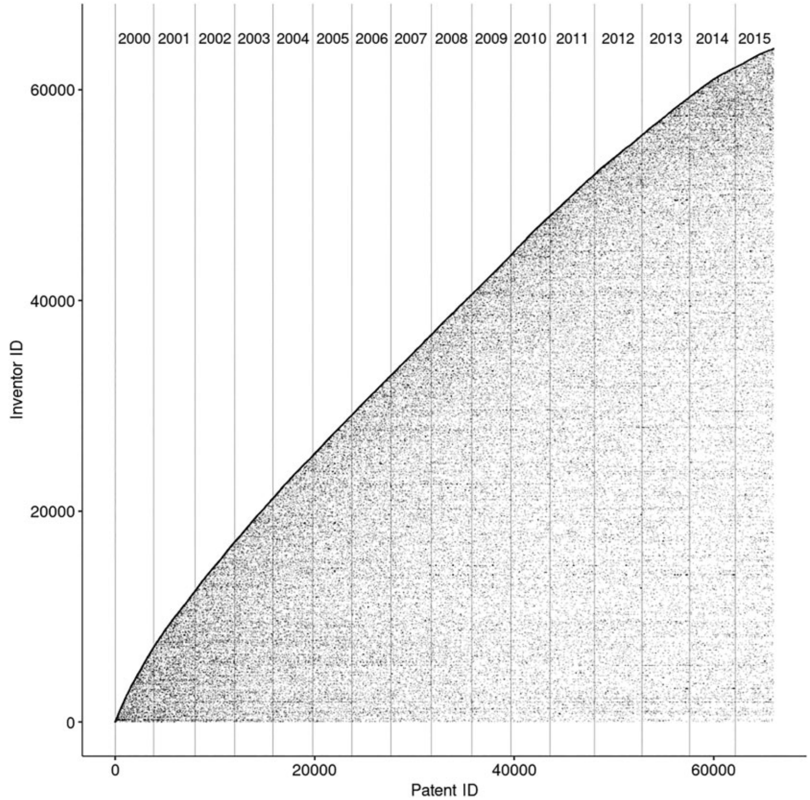

Figure 2. Graphical representation of the adjacency matrix of the patent–inventor network between 2000 and 2015. A dot in position (*i*, *k*) indicates that inventor *i* is a co-owner of patent *k*.

bipartite network, for which the observed adjacency matrix (1) is visualised in Figure 2. From the plot, we can get a clear sense of the previously described actor natality phenomenon, with new inventors becoming active at every time point. Moreover, the figure demonstrates the limits of descriptive analysis when dealing with such large networks, highlighting the need for adequate models to learn something from such data.

As described in Section 1, we instead consider this a dynamic bipartite network, discretising the time steps yearly such that time takes values $t = 1, 2, \ldots, T$. In our notation, $t = 1$ translates to the year 2000. We also allow the actor set to change at each time point so that we end up with T bipartite networks in which the nodes are given by the active inventors at each time point. Resulting from this, we include new inventors that are active for the first time and remove inactive ones from the network at each time point t. The latter point is motivated by the empirical data, which suggests that if previously active inventors do not produce any patents for a long time, it is likely that they will not be active anymore. This phenomenon can stem from a change in career paths (moved up to a management position where writing patents is not among the work tasks) or retirement. To this point, we show the Kaplan–Meier estimate of the time passing between two consecutive patents by the same inventor in Figure 3. As indicated by the dashed grey lines, about 85% of patents by a specific inventor that already had at least one patent are submitted within two years from the previous one. Given this, we define an inventor as active at time t if they had at least one patent in the two years prior to t. Note that by doing so we do not disregard the remaining 15% of the data, but simply label these inventors as inactive for a specific period, i.e., until they appear on another patent.

As we are interested in investigating the drivers of patented innovation and inventor collaboration, we exclude patents developed by a single inventor from the modelled patent set. Moreover, we exclude inventors with no address in Germany from the actor set, as they make up less than 1%

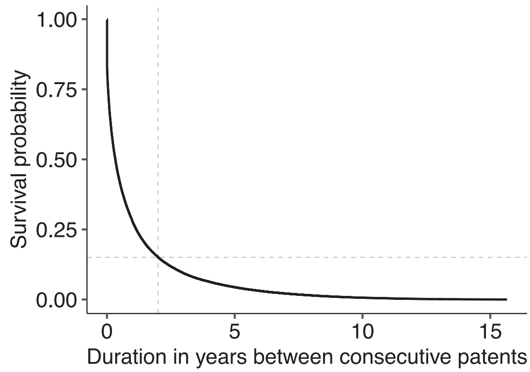

Figure 3. Kaplan–Meier estimate of the duration between consecutive patents submitted between 2000 and 2015.

of the population. In addition to the residence address of each inventor and the date of each patent, we also incorporate information on the gender of each inventor in our model. This set of exogenous covariates aligns with previous work on co-citation networks (Leifeld, 2018).

3 Modelling patent data as bipartite networks

3.1 Temporal exponential random graph models for bipartite networks

Having laid out the available data, we now formulate a generative network model for the bipartite networks at hand. This framework should allow us to differentiate between random and structural characteristics of the network to support or disregard our substantive expectations, such as, for example, whether or not two inventors that teamed up in the past are likely to produce another patent together in the future. To do so we first need to introduce some additional notation. As a general rule, we write $\mathbf{Y}_t$ to denote the network when viewed as a random variable, and $\mathbf{y}_t = (y_{t,ik} : i \in \mathcal{I}_t,\, k \in \mathcal{K}_t)$ if we relate to the observed counterpart. In this context, $y_{t,ik} = 1$ translates to inventor i being a co-owner of patent k, while $y_{t,ik} = 0$ indicates the contrary. As a result, the observed networks are binary and undirected, i.e., $\mathbf{y}_t \in \{0, 1\}^{|\mathcal{I}_t| \times |\mathcal{K}_t|}$. We denote the space of all networks that could potentially be observed at time point t by $\mathcal{Y}_t$. For our application, as explained in the previous section, the latter is restricted to only allow for patents which have at least two inventors.

We specify the joint probability for the set of networks through

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_1, \ldots, \mathbf{Y}_T) = \prod_{t=1}^{T} \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t \mid \mathcal{H}_t), \tag{3}$$

where $\mathcal{H}_t$ defines the history, composed of the bipartite networks and covariates observed before t. The covariates can encompass dyadic and nodal information, but to make the notation less cumbersome we suppress the explicit inclusion of the covariates in the formulae. Following Hanneke et al. (2010), we simplify (3) by assuming that the temporal dependencies are constrained to a fixed time lag, i.e.,

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t \mid \mathcal{H}_t) = \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t \mid \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \ldots, \mathbf{Y}_{t-s} = \mathbf{y}_{t-s}), \tag{4}$$

for $s \in \mathbb{N}$. The Markov property then allows us to postulate an ERGM for the transition probability (4) in the following form:

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y}_t \mid \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \ldots, \mathbf{Y}_{t-s} = \mathbf{y}_{t-s}) = \frac{\exp\{\boldsymbol{\theta}^{\top} s(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s})\}}{\kappa(\boldsymbol{\theta}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-s})}, \tag{5}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q) \in \mathbb{R}^q$ is a q-dimensional vector of parameters, $s : \mathcal{Y}_t \times \cdots \times \mathcal{Y}_{t-s} \to \mathbb{R}^q$ is the function calculating the vector of sufficient statistics and $\kappa(\boldsymbol{\theta}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-s}) := \sum_{y \in \mathcal{Y}_t} \exp\{\boldsymbol{\theta}^{\top} s(\mathbf{y}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-s})\}$ is a normalising factor (see also

Cranmer et al., 2021, Chapter 6 and Leifeld et al., 2018). We obtain a canonical exponential family model with known characteristics (Barndorff-Nielsen, 1978), which come in handy when quantifying the uncertainty of the estimates of $\boldsymbol{\theta}$. Note that for the application to patent data, the coefficients governing the transition from one time point to another are not necessarily constant over time due to external shocks, such as, for example, the dot-com bubble and the 2008 financial crisis, which may affect the activity of inventors. For this reason, we let $\boldsymbol{\theta}$ in (5) flexibly depend on time and estimate it separately for each time point t, but omit the subscript t from the formulae for notational simplicity. Thurner et al. (2018) and Cranmer et al. (2014) also opted for this parametrisation of dynamic coefficients, while smooth functions over time are employed in Lebacher et al. (2021).

Interpreting the coefficients $\boldsymbol{\theta}$ can be done both at the global network level as well on the single tie level. We illustrate the interpretation for θ_p, defined as the coefficient corresponding to the pth sufficient statistic from (5). For the former, $\theta_p > 0$ implies that networks with higher values of the corresponding sufficient statistic become increasingly more likely, while $\theta_p < 0$ implies the converse. For the latter, we define so-called change statistics, which are the change in the sufficient statistics caused by switching the entry $y_{t,ik}$ from 0 to 1. Formally,

$$\boldsymbol{\Delta}_{t,ik}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) = s(\mathbf{y}^+_{t,ik}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-s}) - s(\mathbf{y}^-_{t,ik}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-s}), \tag{6}$$

where $\mathbf{y}^+_{t,ik}$ is the network $\mathbf{y}_t$ with entry $y_{t,ik}$ fixed at 1, while the entry is set to 0 in $\mathbf{y}^-_{t,ik}$. For each possible inventor-patent connection, we can then state the corresponding probability conditionally on the remaining bipartite network denoted by $\mathbf{y}^C_{t,ik}$, i.e., the complete network $\mathbf{y}_t$ excluding the single entry $y_{t,ik}$. This leads to

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_{t,ik} = 1 \mid \mathbf{Y}^C_{t,ik} = \mathbf{y}^C_{t,ik}) = \frac{\exp\{\boldsymbol{\theta}^\top \boldsymbol{\Delta}_{t,ik}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s})\}}{1 + \exp\{\boldsymbol{\theta}^\top \boldsymbol{\Delta}_{t,ik}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s})\}}. \tag{7}$$

Through this expression we can relate $\boldsymbol{\theta}$, the canonical parameter of (5), to the conditional probability of inventor i to be co-owner of patent k. We can thereby derive an interpretation of the coefficients reminiscent of common logistic regression: if adding the tie $y_{t,ik}$ to the network raises the pth entry of $\boldsymbol{\Delta}_{t,ik}(y_t, \ldots, y_{t-s})$ by one unit, the conditional log-odds of $Y_{t,ik}$ are, ceteris paribus, altered by the additive factor θ_p (Goodreau et al., 2009).

3.2 Sufficient statistics for bipartite patent data

The main ingredient of model (5) is the set of sufficient statistics, which translates to a particular dependence structure assumed for the edges in the observed bipartite network (Wang, Pattison, et al., 2013). A statistic that is typically included is the number of edges at time point t, i.e., $s_{\texttt{edges}}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) = |\mathbf{y}_t|$, which can be comprehended as the equivalent of an intercept term in standard regression models (Goodreau et al., 2009). As we are in a dynamic setting in which additional information on past networks is available, we can define statistics that depend on the past networks, such as the number of patents in the previous s years for each actor active at time point t:

$$s_{\texttt{pastpatent}}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) = \sum_{i \in \mathcal{I}_t} \sum_{k \in \mathcal{K}_t} y_{t,ik} \sum_{u=t-s}^{t-1} \sum_{l \in \mathcal{K}_u} y_{u,il}. \tag{8}$$

As the patent network presents some particular dependence structures, more advanced types of statistics are needed, which we describe in the following.

3.2.1 Pairwise statistics of inventors

One drawback of representing our patent data as a bipartite adjacency matrix instead of the one-mode-projected version is that incorporating information on the pairwise inventor-to-inventor level is not straightforward. We therefore introduce assortative two-star statistics extending the work of Bomiriya (2014, Chapter 2) and Metz et al. (2019) on homophily, which is defined as the mechanism driving ties between similar individuals (McPherson et al., 2001), for bipartite networks. In the context of relational event models for bipartite interactions, Malang et al. (2019) use tie-specific, as opposed to global, variants of these statistics based on exponentially decreasing temporal weights of past events. We take the patent-based two-star statistic as starting point, which for $\mathbf{y}_t$ is defined by

$$s_{\texttt{twostar.patent}}(\mathbf{y}_t) = \frac{1}{2}\sum_{k\in\mathcal{K}_t}\sum_{i\in\mathcal{I}_t} y_{t,ik}\left(\sum_{j\neq i} y_{t,jk}\right). \tag{9}$$

The tendency to interact with one another is often based on the similarity of a factor variable $u_t = (u_{t,i};\, i \in \mathcal{I}_t)$. We therefore define the indicator matrix $\mathbf{x}_t \in \{0, 1\}^{|\mathcal{I}_t|\times|\mathcal{I}_t|}$ with entries $x_{t,ij} = \mathbb{I}(u_{t,i} = u_{t,j})$. In line with Bomiriya (2014, Chapter 2), this allows to augment the two-star statistic (9) in the form

$$s_{\texttt{homophily.x}}(\mathbf{y}_t) = \frac{1}{2}\sum_{k\in\mathcal{K}_t}\sum_{i\in\mathcal{I}_t} y_{t,ik}\left(\sum_{j\neq i} y_{t,jk}x_{t,ij}\right). \tag{10}$$

Next, we follow Metz et al. (2019) and generalise (10) by not restricting ourselves to any particular definition of $\mathbf{x}_t$, but letting the matrix be an arbitrary function of the networks from the past s years and other exogenous information. To further correct for different sizes of patents, i.e., the number of inventors co-owning the patent, we normalise the statistic by the degree of each patent, whereby the resulting statistic is defined through:

$$s_{\texttt{assort.x}}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) = \frac{1}{2}\sum_{k\in\mathcal{K}_t}\sum_{i\in\mathcal{I}_t} y_{t,ik}\left(100\times\frac{\sum_{j\neq i} y_{t,jk}x_{t,ij}}{\sum_{j\neq i} y_{t,jk}}\right). \tag{11}$$

To obtain a less cluttered notation, we keep the dependence of $x_{t,ij}$ on $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-s}$ implicit. The corresponding change statistic for an edge between inventor i and patent k is then

$$\mathbf{\Delta}_{t,ik,\texttt{assort.x}}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) = 100\times\frac{\sum_{j\neq i} y_{t,jk}x_{t,ij}}{\sum_{j\neq i} y_{t,jk}}, \tag{12}$$

which can be interpreted as the percentage of inventors on patent k that match with inventor i in matrix $\mathbf{x}$. We multiply the statistic by 100, which does not affect the model itself but eases interpretation (as a unit increase is now equivalent to a single percentage change). To give an example of a statistic of this type, we can combine (12) with matrix $\mathbf{x}_t^P$, for which entry $x_{t,ij}^P$ is 1 if inventor i and j already had a joint patent in the last s years and 0 otherwise. The resulting statistic measures how previous collaboration among inventors affects the propensity of future collaboration. Section 4 provides more examples of such statistics.

3.2.2 *Node set statistics*

As a result of the actor natality and mortality described in the Introduction, we can split the set of inventors $\mathcal{I}_t$ at each time step $t = 1, \ldots, T$ into new inventors with their first patent in t, $\mathcal{I}_t^+ = \{i \in \mathcal{I}_t;\, \sum_{u=t-s}^{t-1}\sum_{k\in\mathcal{K}_u} y_{u,ik} = 0\}$, and inventors that were already active prior to t, $\mathcal{I}_t^- = \{i \in \mathcal{I}_t;\, \sum_{u=t-s}^{t-1}\sum_{k\in\mathcal{K}_u} y_{u,ik} > 0\}$. We here use the term 'new inventors' for actors in $\mathcal{I}_t^+$ and 'experienced inventors' for those in $\mathcal{I}_t^-$. Given these sets, we define $\mathbf{y}_t^+ = (y_{t,ik})_{i\in\mathcal{I}_t^+, k\in\mathcal{K}_t}$ and $\mathbf{y}_t^- = (y_{t,ik})_{i\in\mathcal{I}_t^-, k\in\mathcal{K}_t}$ to be the sub-networks of $\mathbf{y}_t$ made up of new and experienced inventors, respectively.

It is apparent that statistics on past behaviour, such as (8), are not meaningful for inventors from $\mathcal{I}_t^+$, since no historical data is available for those inventors at time t. To account for this, we decompose the statistics $\mathbf{s}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s})$ into three types of terms, namely $\mathbf{s}^+(\mathbf{y}_t^+)$, $\mathbf{s}^-(\mathbf{y}_t^-, \ldots, \mathbf{y}_{t-s})$, and $\mathbf{s}^\pm(\mathbf{y}_t)$, which are defined as statistics that only relate to either $\mathbf{y}_t^+$, $\mathbf{y}_t^-$ and past networks or the full set of inventors $\mathbf{y}_t$, respectively. Defining the corresponding coefficients $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \boldsymbol{\theta}^\pm)$ and change statistics $(\boldsymbol{\Delta}^+_{t,ik}, \boldsymbol{\Delta}^-_{t,ik}, \boldsymbol{\Delta}^\pm_{t,ik})$ accordingly yields

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_{t,ik} = 1 \mid \mathbf{Y}^C_{t,ik} = \mathbf{y}^C_{t,ik}) = \begin{cases} \pi^+_{t,ik}(\mathbf{y}_t), & \text{if } i \in \mathcal{I}^+_t \text{ (new inventor)} \\ \pi^-_{t,ik}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}), & \text{if } i \in \mathcal{I}^-_t \text{ (experienced inventor)}, \end{cases} \tag{13}$$

where $\pi^+_{t,ik}(\mathbf{y}_t)$ and $\pi^-_{t,ik}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s})$ are given by

$$\pi^+_{t,ik}(\mathbf{y}_t) = \frac{\exp\{(\boldsymbol{\theta}^+)^\top \boldsymbol{\Delta}^+_{t,ik}(\mathbf{y}^+_t) + (\boldsymbol{\theta}^\pm)^\top \boldsymbol{\Delta}^\pm_{t,ik}(\mathbf{y}_t)\}}{1 + \exp\{(\boldsymbol{\theta}^+)^\top \boldsymbol{\Delta}^+_{t,ik}(\mathbf{y}^+_t) + (\boldsymbol{\theta}^\pm)^\top \boldsymbol{\Delta}^\pm_{t,ik}(\mathbf{y}_t)\}}$$

$$\pi^-_{t,ik}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) = \frac{\exp\{(\boldsymbol{\theta}^-)^\top \boldsymbol{\Delta}^-_{t,ik}(\mathbf{y}^-_t, \ldots, \mathbf{y}_{t-s}) + (\boldsymbol{\theta}^\pm)^\top \boldsymbol{\Delta}^\pm_{t,ik}(\mathbf{y}_t)\}}{1 + \exp\{(\boldsymbol{\theta}^-)^\top \boldsymbol{\Delta}^-_{t,ik}(\mathbf{y}^-_t, \ldots, \mathbf{y}_{t-s}) + (\boldsymbol{\theta}^\pm)^\top \boldsymbol{\Delta}^\pm_{t,ik}(\mathbf{y}_t)\}}.$$

As an example, for the common edge statistic $s_{\texttt{edges}}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s})$, the aforementioned decomposition means we can define $s_{\texttt{New}}(\mathbf{y}^+_t) = |\mathbf{y}^+_t|$ and $s_{\texttt{Experienced}}(\mathbf{y}^-_t, \ldots, \mathbf{y}_{t-s}) = |\mathbf{y}^-_t|$, to allow for new and experienced inventors to generally have a different propensity to be part of a patent. Note that the splitting of the node set as in (13) does not assume any (in)dependence structure between $\mathbf{Y}^+_t$ and $\mathbf{Y}^-_t$, but rather serves as an aid to specify additional terms and interpret the coefficients at a finer level, as just exemplified for the edge statistic.

3.2.3 Adjustment for varying network size

As argued in Krivitsky et al. (2011), the task of comparing estimated coefficients of two models with identical specifications but different network sizes is non-trivial. This behaviour is due to the fact that including the edge count statistic from the previous paragraph in a TERGM assumes density invariance as the network grows. This characteristic seldom holds for real-world networks as it implies a linearly growing mean degree of all involved actors. In the case of our longitudinal patent network, the number and composition of inventors and patents change from year to year, thus correcting for this is of practical importance to be able to compare coefficient estimates at different time points. To solve the issue, we follow the suggestion of Krivitsky et al. (2011) and incorporate the offset term $1/(|\mathcal{I}_t| + |\mathcal{K}_t|)$ to achieve asymptotically constant mean-degree scaling as the composition of inventors and patents change over time.

3.3 Estimation and inference

We now seek to estimate the parameter $\boldsymbol{\theta}$ by maximising the logarithmic likelihood constructed from (5) for the transition between time points $t-1$ and t. Analysing each transition one at a time enables the use of software for static networks, such as `ergm` (Hunter et al., 2013). If some of the coefficients are constant over some periods, one could apply the block-diagonal approach of Leifeld et al. (2018). We follow the Markov Chain Monte Carlo Maximum-Likelihood Estimation procedure introduced by Geyer and Thompson (1992) and adapted to ERGMs by Hunter and Handcock (2006). In our application, we repeat this for each available time step $t = 1, \ldots, T$.

First, note that subtracting any constant from the logarithmic likelihood constructed from (5) does not change its maximum. We can therefore subtract the logarithmic likelihood evaluated at an arbitrary value of the parameter $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}_0$, which yields the equivalent objective function

$$\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top s(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) - \log(\mathbb{E}_{\boldsymbol{\theta}_0}(\exp\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top s(\mathbf{Y}_t, \ldots, \mathbf{y}_{t-s})\})), \tag{14}$$

where $\mathbb{E}_{\boldsymbol{\theta}}(f(\mathbf{X}))$ is the expected value of random variable $\mathbf{X}$ characterised by parameter $\boldsymbol{\theta}$ and transformed through the arbitrary function $f(\cdot)$. As described in Hunter and Handcock (2006), one can evaluate this objective function by approximating the expected value by generating random networks $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \ldots, \mathbf{Y}^{(M)}$ from (5) under $\boldsymbol{\theta}_0$. In particular, we approximate the expected value in (14) through a Monte Carlo quadrature:

$$\mathbb{E}_{\theta_0}(\exp\{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top s(\mathbf{Y}_t,\ \ldots,\ \mathbf{y}_{t-s})\}) \approx \frac{1}{M}\sum_{m=1}^{M} \exp\{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top s(\mathbf{y}^{(m)},\ \mathbf{y}_{t-1},\ \ldots,\ \mathbf{y}_{t-s})\}. \tag{15}$$

For sufficiently large M, the convergence of this expectation is guaranteed, and we can plug (15) into (14) and apply Newton–Raphson-type methods to maximise it with respect to $\boldsymbol{\theta}$. Sampling from a probability distribution with intractable normalisation constant, such as (5), is achieved by a Metropolis–Hastings algorithm. In particular, we first sample an edge, defined as the tuple (i, k), at random, and consecutively toggle the corresponding entry of $\mathbf{Y}_t$ from 0 to 1 with probability equal to (7) (for more details see Hunter et al., 2013). Due to the large size of the patent networks, we start with the observed network, propose 15.000 of such changes and then stop the Markov chain. This procedure is hence equivalent to contrastive divergence as introduced by Hinton (2002) and adapted to ERGMs by Krivitsky (2017).

Inference on the estimates is drawn based on the Fisher matrix $\mathbf{I}(\boldsymbol{\theta})$, which equals the variance of the sufficient statistics for exponential family distributions (L. Wasserman, 2004). Thus, we can approximate the Fisher matrix through

$$\begin{aligned}\widehat{\mathbf{I}}(\boldsymbol{\theta}) = \mathrm{Var}_{\boldsymbol{\theta}}(s(\mathbf{Y}_t,\ \ldots,\ \mathbf{y}_{t-s})) \approx \frac{1}{M}\sum_{m=1}^{M} &(s(\mathbf{y}^{(m)},\ \mathbf{y}_{t-1},\ \ldots,\ \mathbf{y}_{t-s}) - \bar{s}(\mathbf{y}^{(1)},\ \ldots,\ \mathbf{y}^{(M)}))\\ &\times (s(\mathbf{y}^{(m)},\ \mathbf{y}_{t-1},\ \ldots,\ \mathbf{y}_{t-s}) - \bar{s}(\mathbf{y}^{(1)},\ \ldots,\ \mathbf{y}^{(M)}))^\top,\end{aligned}$$

where $\bar{s}(\mathbf{y}^{(1)},\ \ldots,\ \mathbf{y}^{(M)}) = (1/M)\sum_{m=1}^{M} s(\mathbf{y}^{(m)},\ \mathbf{y}_{t-1},\ \ldots,\ \mathbf{y}_{t-s})$ is the vector containing the averages of the sufficient statistics from the simulated networks $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}$, which are, in turn, drawn from the fitted model, with the parameter $\boldsymbol{\theta}$ set to its maximum-likelihood estimate.

4 Application to inventor team formation

We now present the results of our application, in which we model inventor team formation using the patent data introduced in Section 2. For each statistic included in the model, we explain its meaning, interpret the corresponding estimated coefficient, and then discuss the relationship of our results to prior literature. Further details on the specification of each sufficient statistic can be found in Appendix A. We further provide MCMC diagnostics and goodness-of-fit assessments as proposed by Hunter et al. (2008) in the online Supplementary Material. Due to the slow inertia of patent submissions visible in Figure 3, we set $s = 5$, i.e., consider data from the last five years to be relevant for modelling the current network. This allows us to have enough information for capturing long-range dependence in the networks involving repeated patent submissions of single actors as well as groups of actors.

4.1 Network effects

4.1.1 Propensity to invent

To account for the changing activity levels over time, we incorporate a statistic that counts how many edges are in the network. Following Section 3.2, we split this term into separate statistics for experienced and new inventors. Heuristically, one can interpret the corresponding coefficients as the general propensity to form ties, i.e., participate in a patent, for the two inventor sets, respectively. Note that it would not be possible to estimate this effect by modelling a unipartite projection on inventors: in that case, the intercept term would only measure the propensity for inventors to collaborate, regardless of the number of patents produced. The plot of the estimates for the propensity to invent over time is shown in the upper left panel of Figure 4. It exhibits a different level of activity for new and experienced inventors. We expect this by design, as new inventors enter the network precisely because they are active at time t, while experienced ones might only have been active in the past. Overall, we observe a steady increase in activity in the network from 2008 onward for both sets of inventors.

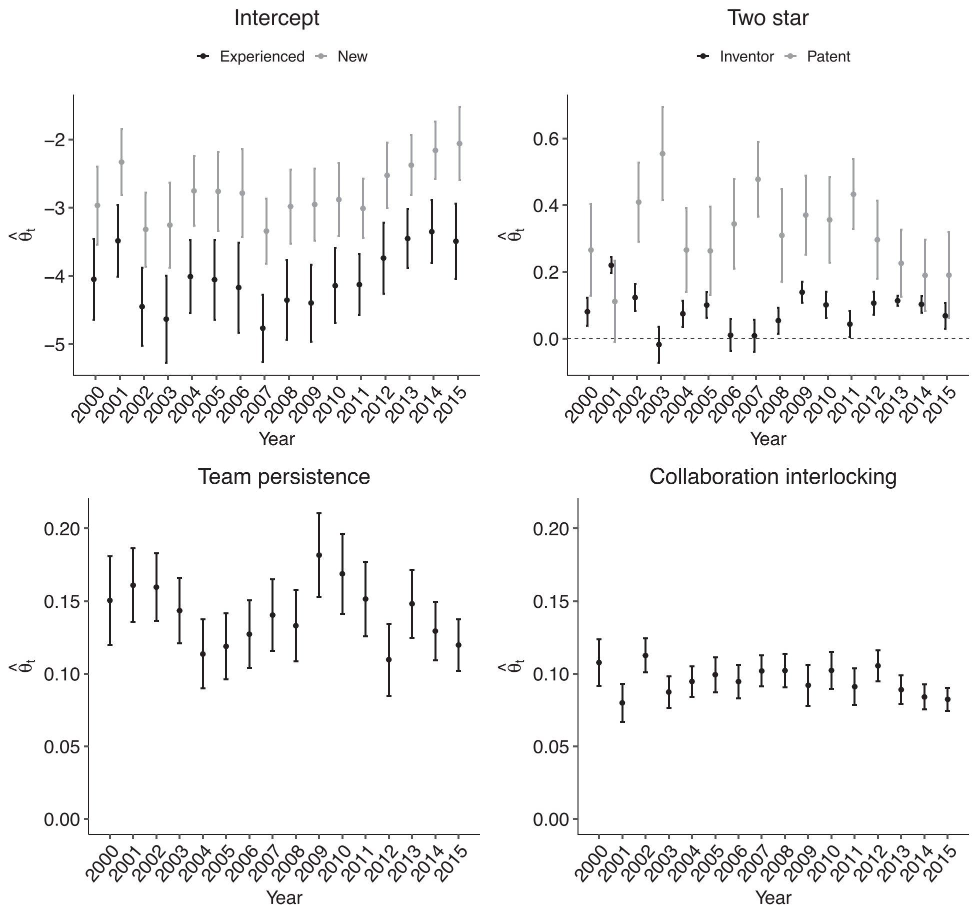

Figure 4. Estimated time-varying coefficients regarding the propensity to invent, two-star statistics, team persistence and collaboration interlocking.

4.1.2 Two-star statistics

Two-star statistics relate to the concept of centrality (S. Wasserman & Faust, 1994). For bipartite networks, they can be defined with respect to each of the two modes (inventors and patents, respectively). For inventors the statistic is given in Appendix A and expresses whether inventor i is more or less likely to invent an additional patent in year t, given that he/she is (co-)owner of at least another patent in that year. For patents, the statistic relates to the number of inventors per patent and is given in (9). These effects could not be estimated for a unipartite projection on inventors: in that case, the two-star statistic would simply relate to the propensity for inventors to have additional collaborators, with information on the number of patents and their size being lost. The top right panel of Figure 5 depicts both estimates for the two-star statistics. For inventors, the estimates take small positive values for most time points, without much temporal variation. This indicates a slight tendency towards centralisation for inventors, i.e., inventors aiming to submit multiple patents per year. For patents the corresponding two-star estimates are larger, i.e., patents tend to be owned by multiple inventors. The two-star effect slowly decreases since 2011, meaning that the number of owners per patent is getting smaller. The variance for the estimated two-star patent effect is generally larger than the estimate of the corresponding two-star inventor effect, which stems from the fact that there are fewer patents than inventors in a single year.

4.1.3 Team persistence

Most patented inventions are the result of team work (Giuri et al., 2007), which leads to the build-up of valuable team-specific capital (Jaravel et al., 2018). We therefore expect past collaboration

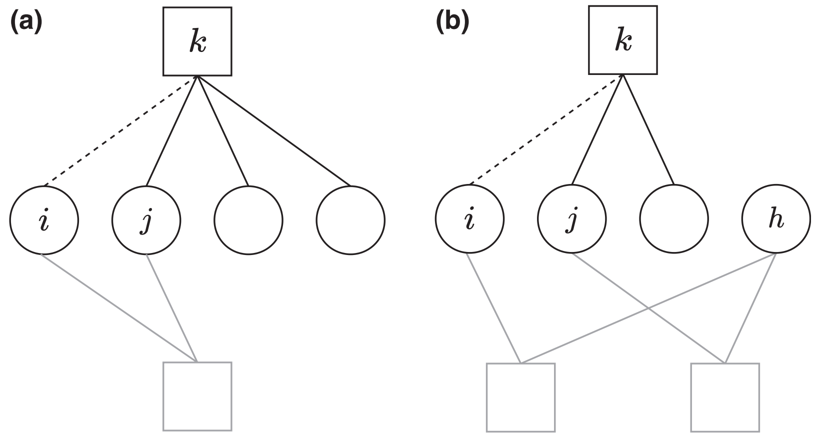

Figure 5. Illustration of the change statistics related to assortative network statistics for team persistence (a) and collaboration interlocking (b). Circles represent inventors, and squares are patents. The dashed line indicates a possible edge at time point t, while black lines represent edges given at time point t. Grey lines, on the other hand, display past connections, and grey squares stand for past patents. (a) Team persistence and (b) collaboration interlocking.

to positively affect the propensity for two inventors to collaborate again. To account for this effect, we include a team persistence statistic based on the pairwise statistics of inventors proposed in 3.2 in the model. The statistic, which could also be termed 'repetition' (or 'reciprocity', as defined in Leifeld & Brandenberger, 2019), is visually represented in Figure 5a, and rests on the definition of matrix $\mathbf{x}_t^{\mathrm{P}}$, whose (i, j)th entry is 1 if inventors i and j have already co-invented a patent in the previous five years, and 0 otherwise. The bottom left panel of Figure 4 depicts the corresponding coefficient estimate, which is positive and significantly different from zero over time. This finding corroborates our anticipations that, controlling for the other factors, two inventors are more likely to jointly produce a patent if they already worked on an invention together in the past. Hence, teams of inventors play an important role in patent creation.

4.1.4 Collaboration interlocking

In addition to investigating the persistence of collaborations, it is of interest to understand how having had a common partner in the past influences the tendency to develop a joint patent in the present. We account for this by including the collaboration interlocking statistic in our model. By common partners we are referring to actors such as inventor h for inventors i and j in Figure 5b. We define the statistic again by pairwise statistics of inventors through the matrix $\mathbf{x}_t^{\mathrm{CI}}$, where the binary information of whether or not inventors i and j have at least one common partner is encoded in the (i, j)th entry. The related coefficient estimates are shown in the bottom right panel of Figure 4, where we notice that the estimate attains significantly positive values throughout the observational period. This result suggests that if two inventors i and j both had a patent with the same inventor h, they are generally more likely to co-invent in the future. Our finding holds controlling for all other features in our model (including the previously described team persistence statistic). This effect can be considered similar to triadic closure in unimodal networks, i.e., 'a collaborator of my collaborator is more likely to become my collaborator'. The result thus supports the idea that the creation of inventor teams is often promoted via common colleagues and that informal knowledge flows are key to the invention process (see Giuri & Mariani, 2013 and references cited therein).

4.2 Effects of inventor-specific covariates

4.2.1 Spatial proximity

Many patents are created in a workplace environment (Giuri et al., 2007). For this reason, we would expect inventors that live close to each other to be more likely to invent together. Moreover, there is empirical evidence that collaboration is more likely between inventors that

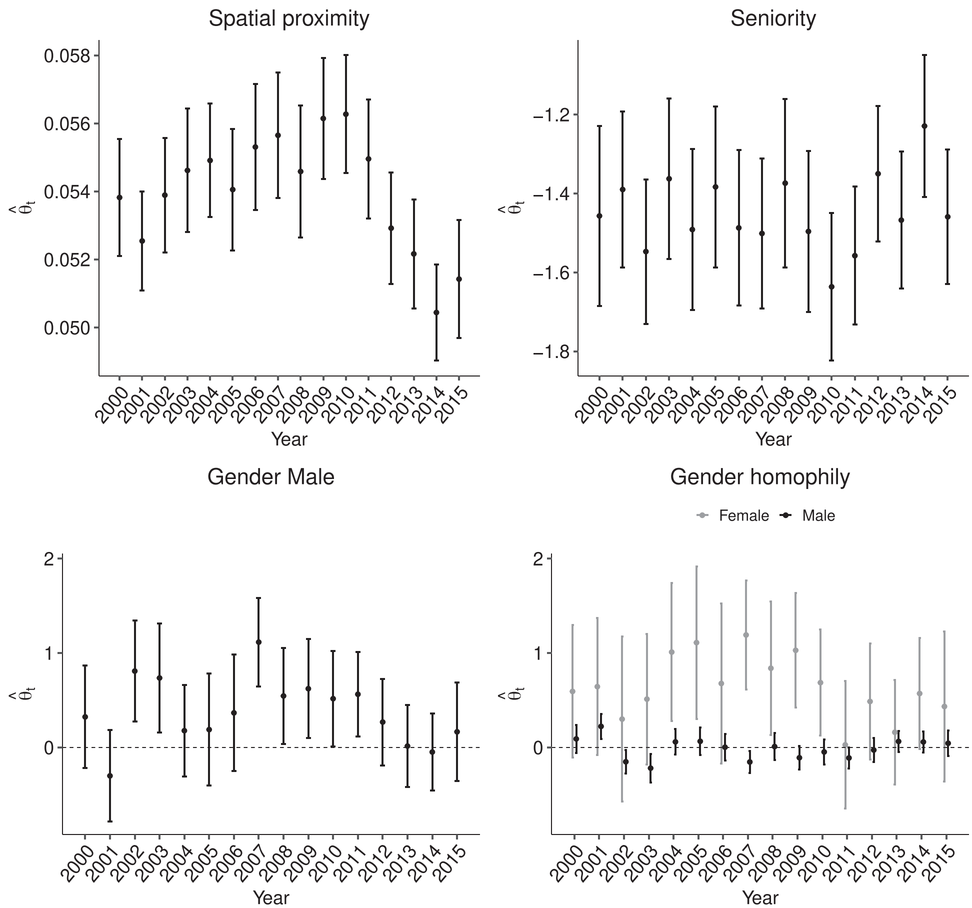

Figure 6. Estimated time-varying coefficients regarding the spatial proximity, seniority, and gender of inventors.

live close to one another even if they do not share the same employer (e.g., Crescenzi et al., 2016). For these reasons, we include a spatial proximity statistic in our model, where we define spatial proximity as living within a range of 50 km. We encode this proximity information in a binary matrix $\mathbf{x}^{SP}$ and incorporate it in the model as a pairwise statistic of inventors. The top left panel of Figure 6 depicts the estimated coefficients for the statistic. The positive values attained over time confirm that inventors living near each other have a higher chance to collaborate. We can also see that the effect goes down over time from 2010 onward; this makes sense in an increasingly interconnected society, where more and more connections are formed through the web in addition to physical ones.

4.2.2 *Seniority*

The top right panel in Figure 6 depicts the effect of the number of previously owned patents by each inventor in the past five years. The corresponding statistic can be viewed as a measure of inventor seniority, where inventors with more patents in the past are considered to be senior. As this statistic would trivially be a structural zero for new inventors, it is only computed for the set of inventors which were previously active in the network (experienced inventors). This is another effect we would not be able to estimate if we only considered the unipartite projected network of inventors. The negative coefficient estimate here suggests that, conditional on all other statistics included in the model, senior inventors have a lower propensity to create new patents. Prior research has shown that career dynamics of inventors are complex as economic opportunities, productivity and personal preferences interact (see, e.g., Allen & Katz, 1992; Bell et al., 2019). But our results are consistent with earlier results indicating that with greater seniority, inventors take over managerial responsibilities within the same firm, or that high visibility of their invention output also leads them to move to new employers and tasks, thus lowering (or halting) their invention output.

4.2.3 Gender and gender homophily

Another variable of interest in the realm of innovation research is gender. Many researchers have expressed concerns about the sparse representation of women among inventors (typically far less than 10%) and possible wage discrimination (see, e.g., Hoisl & Mariani, 2017; Jensen et al., 2018). These studies established gender as an essential topic in innovation economics. We incorporate gender in our model in two ways, i.e., as a main effect and as a homophily effect (as introduced in (10)). The two plots at the bottom of Figure 6 show the effects of gender on the propensity to create patents (left) and on homophily, i.e., the tendency of inventing together with people of the same gender (right). Note that both effects need to be interpreted keeping in mind that the vast majority of the actors in the network are male (96%). From the plot on the bottom left, we can see how, while male inventors seem to be slightly more active, all in all male and female inventors did not show significant differences in their propensity to invent. Note that this holds given the inclusion of those inventors in the network, i.e., given that they were already inventors. The gender homophily plot shows different results; here we see that, while male inventors seem to have the same likelihood to form patents with both genders, female inventors tend to have more collaborations with other females than with males. While the effect is quite sizeable in absolute value, the uncertainty here is considerable given the small number of female actors in the network. Still, we can see this as weak evidence for a gender homophily effect for female inventors. These results are consistent with earlier findings by Whittington (2018), who studies the role of gender in life science inventor teams.

5 Discussion

This paper analyses a massive bipartite network, consisting of all inventors and collaborative patents filed between 1995 and 2015 in electrical engineering. To account for the sheer size of the complete network and the structural zeros in the related bipartite adjacency matrix, we suggested a temporal decomposition of the data into multiple smaller networks. Guided by substantive questions posed by innovation research, we then proposed a set of bipartite network statistics focused on gender issues, team persistence, collaboration interlocking, and spatial proximity.

Time-varying actor sets due to actor mortality and natality are often observed in networks beyond the realm of patent data. For instance, scientific collaboration behaves similarly, as many PhD students do not pursue an academic career and hence have a short lifespan in the scientific collaboration network. At the same time, new PhD students continuously enter the scientific world. Therefore, the proposed temporal decomposition and the employed network terms exploiting pairwise information on either mode of actors can also be used in other settings.

In addition to the methodological contributions, our study offers several novel results concerning the substantive analysis. We utilise a population dataset spanning 20 years (1995–2015). The time span and the availability of population data are crucial to assess the team formation process reliably. Using a population dataset of this size is unique in the literature on inventor team formation. Moreover, while much of the literature has focused on the relationship between team characteristics and performance, there are very few studies on the actual process of inventor team formation. While some of the variables we are using have been discussed and utilised in other domains, we are unaware of inventor team studies employing data with a similar breadth of team and inventor descriptors. This breadth adds to the novelty of our study. We also note that our variable set reflects a number of meaningful concerns such as inclusiveness, gender equality, and seniority. The results should therefore be of considerable interest to policymakers.

Still, we want to address some limitations in our analysis, which would benefit from further research. First, our definition of the actor sets is based on a simple heuristic we determined in a data-driven manner. However, this practice might bias our findings concerning degree-related statistics since the exact number of isolated inventors is not known but assumed. More complex methods to identify active inventors based on further exogenous data, such as job histories, might be a fruitful future endeavour. Second, we assumed the parameters to be different each year. Extending the approach of Cranmer et al. (2014), one could incorporate a change-point detection directly into the TERGM framework to identify periods over which the coefficients are constant from the observed data. Note that, to facilitate building on our research, we make our implementation available through the R software package `patent.ergm`. Moreover, to guarantee the replicability of our results, we make the full data and code available online on a GitHub repository. This repository also includes the R package `patent.ergm`.

All in all, we show how spatial proximity, teamwork and interlocking of collaborations positively impact the output of inventors. Further, we demonstrate how inventors' characteristics, such as gender and seniority, play a significant role in the process, and identify gender homophily as a critical determinant of inventor team formation. Our application to inventor teams presents an alternative to classical forms of analysis of patenting and inventorship networks. While prior studies are almost exclusively focused on analysing the underlying mechanisms one at a time, we model them simultaneously in the framework of bipartite networks. Our study thus provides an effective alternative to classical forms of regression-based analysis of innovation and the mechanisms driving it.

Conflict of interest: No potential conflict of interest was reported by the author(s).

Funding

The work has been partially supported by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibility for its content.

Data availability

We provide full replication code and materials, including data, goodness-of-fit analysis and MCMC diagnostics, in our GitHub repository, available at https://github.com/corneliusfritz/Modelling-German-patents-and-inventors. Moreover, this repository includes the package `ergm.patent` that implements the pairwise statistics introduced in Section 3.2.1.

Appendix A. Sufficient statistics

In the following, we detail the mathematical definitions of all sufficient statistics incorporated in our model.

Propensity to invent: As already stated in Section 3.1, the standard term to incorporate in any ERGM specification is an edge statistic that counts how many edges are realised in the network. In accordance with Section 3.2, we split this term into the statistics $s_{\texttt{New}}(\mathbf{y}_t^+) = |\mathbf{y}_t^+| = \sum_{i \in \mathcal{I}_t^+} \sum_{k \in \mathcal{K}_t} y_{t,ik}$ and $s_{\texttt{Experienced}}(\mathbf{y}_t^-, \ldots, \mathbf{y}_{t-s}) = |\mathbf{y}_t^-| = \sum_{i \in \mathcal{I}_t^-} \sum_{k \in \mathcal{K}_t} y_{t,ik}$. Figures A1a and A1b visualise the corresponding two network configurations.

Two-star statistics: Two-star statistics can be stated with regards to either set of actors in the case of bipartite networks. The definition of the two-star statistic for the patents is shown in Figure A1c and given by

$$s_{\texttt{twostar.patent}}(\mathbf{y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} \right),$$

while the version for the inventors is visualised in Figure A1d and defined as:

$$s_{\texttt{twostar.inventor}}(\mathbf{y}_t) = \frac{1}{2} \sum_{i \in \mathcal{I}_t} \sum_{k \in \mathcal{K}_t} y_{t,ik} \left(\sum_{l \neq k} y_{t,il} \right).$$

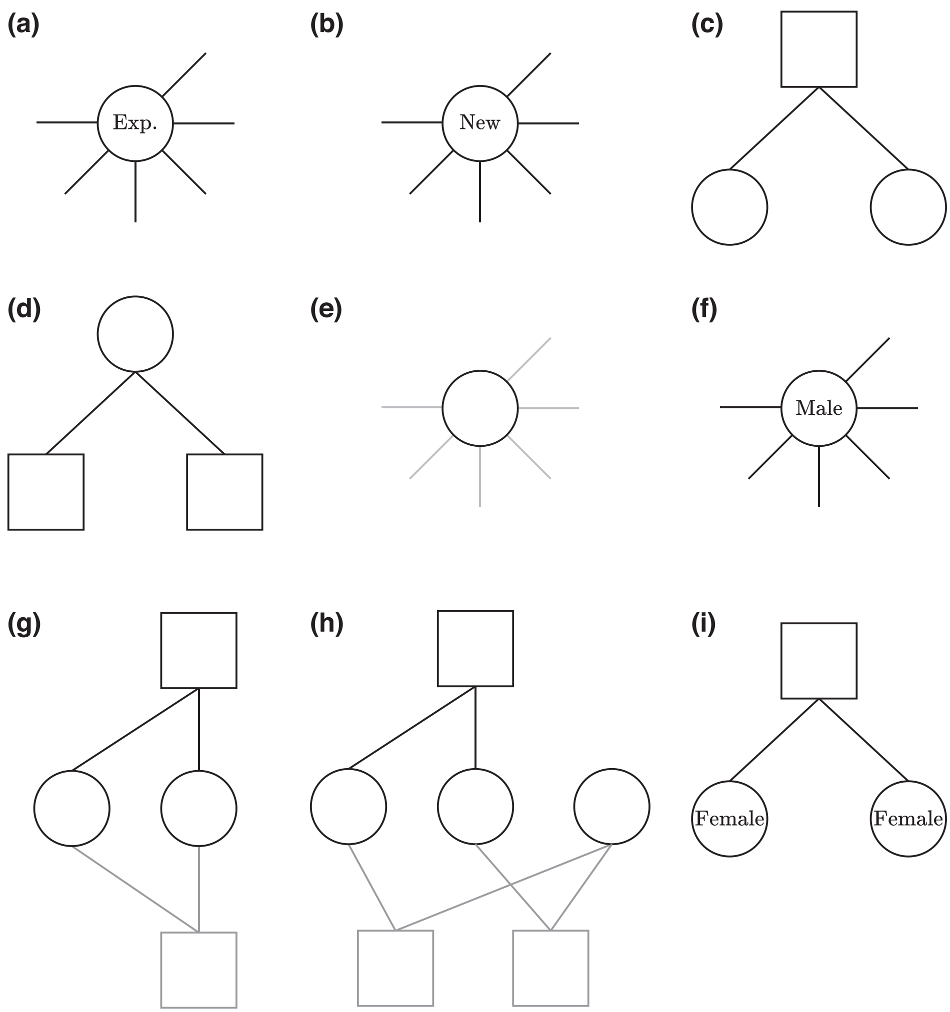

Figure A1. Network configurations for the general edge and two-star terms. Circles are inventors and squares patents and black lines are observed edges in the network at time point t, while grey lines are edges in the past. (a) Experienced inventors, (b) new inventors, (c) patent two-stars, (d) inventor two-stars, (e) seniority, (f) male inventors, (g) team persistence, (h) collaborative interlocking and (i) homophily of females.

Pairwise statistics of inventors: We include three versions of pairwise statistics of inventors introduced in Section 3.2. The statistics are given by

$$s_{\texttt{assort.x}}(\mathbf{y}_t, \ \ldots, \mathbf{y}_{t-s}) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(100 \times \frac{\sum_{j \neq i} y_{t,jk} x_{t,ij}}{\sum_{j \neq i} y_{t,jk}} \right).$$

Note that, in general, the matrix $\mathbf{x}$ can be an arbitrary function of the past networks and nodal or dyadic exogenous information. Its definition differs between the three statistics of pairwise statistics of inventors:

1. Team persistence: For $i, j \in \mathcal{I}_t$ and $i \neq j$ the entries of $\mathbf{x}_t^{\mathrm{P}}$ are given by

$$x_{t,ij}^{\mathrm{P}} = \begin{cases} 1, & \text{if } \sum_{u=t-s}^{t-1} \sum_{k \in \mathcal{K}_u} y_{u,ik} y_{u,jk} > 0 \\ 0, & \text{else} \end{cases}$$

and a graphical illustration of the statistic is provided in Figure A1g. Note that Leifeld and Brandenberger (2019) and Metz et al. (2019) describe a closely related mechanism as *reciprocity* and *collaboration*, respectively. One can comprehend this statistic as a particular type of the four-cycle statistic (Wang, Pattison, et al., 2013) where one half already occurred in the past, and the other half might occur in the present.

2. Collaboration interlocking: For $i, j \in \mathcal{I}_t$ and $i \neq j$, the entries of $\mathbf{x}_t^{\mathtt{CI}}$ are defined by

$$x_{t,ij}^{\mathtt{CI}} = \begin{cases} 1, & \text{if } \sum_{u=t-s}^{t-1} \sum_{h \in \mathcal{I}_t} \sum_{k,l \in \mathcal{K}_u} y_{u,ik} y_{u,hk} y_{u,jl} y_{u,hl} > 0 \\ 0, & \text{else} \end{cases}$$

and a graphical illustration of the statistic is provided in Figure A1h. Coming back to the representation as cycle-statistics, this term is a six-cycle statistic in which four of the six edges happened in the time frame from $t-5$ to $t-1$ and two in year t.

3. Spatial proximity: For $i, j \in \mathcal{I}_t$ and $i \neq j$ the entries of $\mathbf{x}_t^{\mathtt{SP}}$ are defined as

$$x_{t,ij}^{\mathtt{SP}} = \begin{cases} 1, & \text{if } \mathrm{dist}(x_{\mathtt{coord},i}, x_{\mathtt{coord},j}) > 50\,\mathrm{km} \\ 0, & \text{else} \end{cases}$$

where $x_{\mathtt{coord},i}$ and $x_{\mathtt{coord},j}$ define the longitude and latitude of inventors i and j, respectively, and the function $\mathrm{dist}(x_{\mathtt{coord},i}, x_{\mathtt{coord},j})$ computes the distance in kilometres between them via the haversine formula. A continuous form of this statistic based on the Euclidean distance itself was employed in Metz et al. (2019).

Seniority: The respective binary indicator is based on the `pastpatent` statistic given in (8), but in this case we define it on the inventor level:

$$s_{\mathtt{seniority},i}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s}) = \sum_{u=t-s}^{t-1} \sum_{k \in \mathcal{K}_u} y_{u,ik}$$

We binarise this inventor-specific covariate by first computing the median of $s_{\mathtt{seniority},i}(\mathbf{y}_t, \ldots, \mathbf{y}_{t-s})$ over all inventors and then using this value to split the inventors into two groups (i.e., seniors and juniors). The resulting categorical covariate relates to the number of patents in the past and is represented in Figure A1e.

Gender and gender homophily: The main effect of gender is depicted in Figure A1f and defined by:

$$s_{\mathtt{gender}}(\mathbf{y}_t) = \sum_{i \in \mathcal{I}_t} \sum_{k \in \mathcal{K}_t} y_{t,ik} \mathbb{I}(x_{\mathtt{gender},i} = \text{'male'}),$$

where $x_{\mathtt{gender},i} \in \{\text{'male'}, \text{'female'}\}$ indicates the gender of inventor i. The homophily effect, on the other hand, is for males defined by:

$$s_{\mathtt{homophily.male}}(\mathbf{y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} \mathbb{I}(x_{\mathtt{gender},i} = \text{'male'}) \mathbb{I}(x_{\mathtt{gender},j} = \text{'male'}) \right).$$

and for females the formula reads:

$$s_{\mathtt{homophily.female}}(\mathbf{y}_t) = \frac{1}{2} \sum_{k \in \mathcal{K}_t} \sum_{i \in \mathcal{I}_t} y_{t,ik} \left(\sum_{j \neq i} y_{t,jk} \mathbb{I}(x_{\mathtt{gender},i} = \text{'female'}) \mathbb{I}(x_{\mathtt{gender},j} = \text{'female'}) \right).$$

Figure A1i visualises the homophily statistic for females. The equivalent statistic for males can be defined in the same manner.

References

Ahmed A., Batagelj V., Fu X., Hong S.-H., Merrick D., & Mrvar A. (2007). Visualisation and analysis of the internet movie database. In *6th International Asia-Pacific Symposium on Visualization* (pp. 17–24). IEEE.

Allen T. J., & Katz R. (1992). Age, education and the technical ladder. *IEEE Transactions on Engineering Management*, *39*(3), 237–245. https://doi.org/10.1109/17.156557

Alstott J., Triulzi G., Yan B., & Luo J. (2017). Mapping technology space by normalizing patent networks. *Scientometrics*, *110*(1), 443–479. https://doi.org/10.1007/s11192-016-2107-y

Balconi M., Breschi S., & Lissoni F. (2004). Networks of inventors and the role of academia: An exploration of Italian patent data. *Research Policy*, *33*(1), 127–145. https://doi.org/10.1016/S0048-7333(03)00108-2

Barndorff-Nielsen O. (1978). *Information and exponential families in statistical theory*. Wiley.

Bauer V., Harhoff D., & Kauermann G. (2022). A smooth dynamic network model for patent collaboration data. *AStA Advances in Statistical Analysis*, *106*(1), 97–116. https://doi.org/10.1007/s10182-021-00393-w

Bell A., Chetty R., Jaravel X., Petkova N., & Van Reenen J. (2019). Do tax cuts produce more einsteins? The impacts of financial incentives versus exposure to innovation on the supply of inventors. *Journal of the European Economic Association*, *17*(3), 651–677. https://doi.org/10.1093/jeea/jvz013

Bomiriya R. P. (2014). *Topics in exponential random graph modeling* [Phd thesis]. Pennsylvania State University.

Borgatti S. P., & Everett M. G. (1997). Network analysis of 2-mode data. *Social Networks*, *19*(3), 243–269. https://doi.org/10.1016/S0378-8733(96)00301-2

Breiger R. L. (1974). The duality of persons and groups. *Social Forces*, *53*(2), 181–190. https://doi.org/10.2307/2576011

Broekel T., & Bednarz M. (2018). Disentangling link formation and dissolution in spatial networks: An application of a two-mode STERGM to a project-based R&D network in the German biotechnology industry. *Networks and Spatial Economics*, *18*(3), 677–704. https://doi.org/10.1007/s11067-018-9430-1

Chen T. H. Y. (2021). Statistical inference for multilayer networks in political science. *Political Science Research and Methods*, *9*(2), 380–397. https://doi.org/10.1017/psrm.2019.49

Cranmer S. J., Desmarais B. A., & Morgan J. W. (2021). *Inferential network analysis*. Cambridge University Press.

Cranmer S. J., Heinrich T., & Desmarais B. A. (2014). Reciprocity and the structural determinants of the international sanctions network. *Social Networks*, *36*(1), 5–22. https://doi.org/10.1016/j.socnet.2013.01.001

Crescenzi R., Nathan M., & Rodríguez-Pose A. (2016). Do inventors talk to strangers? on proximity and collaborative knowledge creation. *Research Policy*, *45*(1), 177–194. https://doi.org/10.1016/j.respol.2015.07.003

de Solla Price D. J. (1965). The science of science. *Bulletin of the Atomic Scientists*, *21*(8), 2–8. https://doi.org/10.1080/00963402.1965.11454842

Egghe L., & Rousseau R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Elsevier.

Ejermo O., & Karlsson C. (2006). Interregional inventor networks as studied by patent coinventorships. *Research Policy*, *35*(3), 412–430. https://doi.org/10.1016/j.respol.2006.01.001

Faust K., & Skvoretz J. (1999). Logit models for affiliation networks. *Sociological Methodology*, *31*(1), 253–280.

Friel N., Rastelli R., Wyse J., & Raftery A. E. (2016). Interlocking directorates in Irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences*, *113*(24), 6629–6634. https://doi.org/10.1073/pnas.1606295113

Garfield E. (1955). Citation indexes for science. *Science*, *122*(3159), 108–111. https://doi.org/10.1126/science.122.3159.108

Geyer C. J., & Thompson E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *54*(3), 657–699. https://doi.org/10.1111/j.2517-6161.1992.tb01443.x

Giuri P., *et al.* (2007). Inventors and invention processes in Europe: Results from the PatVal-EU survey. *Research Policy*, *36*(8), 1107–1127. https://doi.org/10.1016/j.respol.2007.07.008

Giuri P., & Mariani M. (2013). When distance disappears: Inventors, education, and the locus of knowledge spillovers. *Review of Economics and Statistics*, *95*(2), 449–463. https://doi.org/10.1162/REST_a_00259

Goodreau S., Kitts J. A., & Morris M. (2009). Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. *Demography*, *46*(1), 103–125. https://doi.org/10.1353/dem.0.0045

Hall B. H., & Harhoff D. (2012). Recent research on the economics of patents. *Annual Review of Economics*, *4*(1), 541–565. https://doi.org/10.1146/annurev-economics-080511-111008

Hanneke S., Fu W., & Xing E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, *4*, 585–605. https://doi.org/10.1214/09-EJS548

Hinton G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation*, *14*(8), 1771–1800. https://doi.org/10.1162/089976602760128018

Hoisl K., & Mariani M. (2017). It's a man's job: Income and the gender gap in industrial research. *Management Science*, *63*(3), 766–790. https://doi.org/10.1287/mnsc.2015.2357

Hunter D. R., Goodreau S. M., & Handcock M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248–258. https://doi.org/10.1198/016214507000000446

Hunter D. R., Goodreau S. M., & Handcock M. S. (2013). Ergm.userterms: A template package for extending statnet. *Journal of Statistical Software*, *52*(2), 1–25. https://doi.org/10.18637/jss.v052.i02

Hunter D. R., & Handcock M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, *15*(3), 565–583. https://doi.org/10.1198/106186006X133069

Jaravel X., Petkova N., & Bell A. (2018). Team-specific capital and innovation. *American Economic Review*, *108*(4–5), 1034–1073. https://doi.org/10.1257/aer.20151184

Jensen K., Kovács B., & Sorenson O. (2018). Gender differences in obtaining and maintaining patent rights. *Nature Biotechnology*, *36*(4), 307–309. https://doi.org/10.1038/nbt.4120

Jiang S., Gao Q., & Chen H. (2013). Statistical modeling of nanotechnology knowledge diffusion networks. *ICIS 2013 Proceedings*. Atlanta, GA, United States: Association for Information Systems.

Jiang S., Gao Q., Chen H., & Roco M. C. (2015). The roles of sharing, transfer, and public funding in nanotechnology knowledge-diffusion networks. *Journal of the Association for Information Science and Technology*, *66*(5), 1017–1029. https://doi.org/10.1002/asi.23223

Keegan B., Gergle D., & Contractor N. (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* (pp. 427–436). New York, NY, United States: Association for Computing Machinery.

Koskinen J., & Edling C. (2012). Modelling the evolution of a bipartite network – peer referral in interlocking directorates. *Social Networks*, *34*(3), 309–322. https://doi.org/10.1016/j.socnet.2010.03.001

Krivitsky P. N. (2017). Using contrastive divergence to seed Monte Carlo MLE for exponential-family random graph models. *Computational Statistics and Data Analysis*, *107*, 149–161. https://doi.org/10.1016/j.csda.2016.10.015

Krivitsky P. N., Handcock M. S., & Morris M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology*, *8*(4), 319–339. https://doi.org/10.1016/j.stamet.2011.01.005

Latapy M., Magnien C., & Del Vecchio N. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks*, *30*(1), 31–48. https://doi.org/10.1016/j.socnet.2007.04.006

Lebacher M., Thurner P. W., & Kauermann G. (2021). A dynamic separable network model with actor heterogeneity: An application to global weapons transfers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *184*(1), 201–226. https://doi.org/10.1111/rssa.12620

Leifeld P. (2018). Polarization in the social sciences: Assortative mixing in social science collaboration networks is resilient to interventions. *Physica A: Statistical Mechanics and its Applications*, *507*, 510–523. https://doi.org/10.1016/j.physa.2018.05.109

Leifeld P., & Brandenberger L. (2019). Endogenous coalition formation in policy debates. Preprint, arXiv:1904.05327.

Leifeld P., Cranmer S. J., & Desmarais B. A. (2018). Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals. *Journal of Statistical Software*, *83*(6), 1–36. https://doi.org/10.18637/jss.v083.i06

Li X., Chen H., Huang Z., & Roco M. C. (2007). Patent citation network in nanotechnology (1976–2004). *Journal of Nanoparticle Research*, *9*(3), 337–352. https://doi.org/10.1007/s11051-006-9194-2

Malang T., Brandenberger L., & Leifeld P. (2019). Networks and social influence in european legislative politics. *British Journal of Political Science*, *49*(4), 1475–1498. https://doi.org/10.1017/S0007123417000217

McPherson M., Smith-Lovin L., & Cook J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Melin G., & Persson O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, *36*(3), 363–377. https://doi.org/10.1007/BF02129600

Metz F., Leifeld P., & Ingold K. (2019). Interdependent policy instrument preferences: A two-mode network approach. *Journal of Public Policy*, *39*(4), 609–636. https://doi.org/10.1017/S0143814X18000181

Newman M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5200–5205. https://doi.org/10.1073/pnas.0307545100

Onody R. N., & de Castro P. A. (2004). Complex network study of Brazilian soccer players. *Physical Review E*, *70*(3), 037103. https://doi.org/10.1103/PhysRevE.70.037103

Small H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. https://doi.org/10.1002/asi.4630240406

Snijders T. A., Pattison P. E., Robins G. L., & Handcock M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, *36*(1), 99–153. https://doi.org/10.1111/j.1467-9531.2006.00176.x

Thurner P. W., Schmid C. S., Cranmer S. J., & Kauermann G. (2018). Network interdependencies and the evolution of the international arms trade. *Journal of Conflict Resolution*, *63*(7), 1736–1764. https://doi.org/10.1177/0022002718801965

Verspagen B. (2012). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, *10*(1), 93–115. https://doi.org/10.1142/S0219525907000945

von Wartburg I., Teichert T., & Rost K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, *34*(10), 1591–1607. https://doi.org/10.1016/j.respol.2005.08.001

Wang P., Pattison P., & Robins G. (2013). Exponential random graph model specifications for bipartite networks: A dependence hierarchy. *Social Networks*, *35*(2), 211–222. https://doi.org/10.1016/j.socnet.2011.12.004

Wang P., Robins G., Pattison P., & Lazega E. (2013). Exponential random graph models for multilevel networks. *Social Networks*, *35*(1), 96–115. https://doi.org/10.1016/j.socnet.2013.01.004

Wasserman L. (2004). *All of statistics*. Springer.

Wasserman S., & Faust K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Wasserman S., & Pattison P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*. *Psychometrika*, *61*(3), 401–425. https://doi.org/10.1007/BF02294547

Whittington K. B. (2018). A tie is a tie? Gender and network positioning in life science inventor collaboration. *Research Policy*, *47*(2), 511–526. https://doi.org/10.1016/j.respol.2017.12.006