Chapter 4
Determinants of Cross-Regional R&D Collaboration Networks: An Application of Exponential Random Graph Models

Tom Broekel and Matté Hartog

Abstract This study investigates the usefulness of exponential random graph models (ERGM) to analyze the determinants of cross-regional R&D collaboration networks. Using spatial interaction models, most research on R&D collaboration between regions is constrained to focus on determinants at the node level (e.g. R&D activity of a region) and dyad level (e.g. geographical distance between regions). ERGMs represent a new set of network analysis techniques that has been developed in recent years in mathematical sociology. In contrast to spatial interaction models, ERGMs additionally allow considering determinants at the structural network level while still only requiring cross-sectional network data.

The usefulness of ERGMs is illustrated by an empirical study on the structure of the cross-regional R&D collaboration network of the German chemical industry. The empirical results confirm the importance of determinants at all three levels. It is shown that in addition to determinants at the node and dyad level, the structural network level determinant "triadic closure" helps in explaining the structure of the network. That is, regions that are indirectly linked to each other are more likely to be directly linked as well.

T. Broekel (✉)
Institute of Economic and Cultural Geography, Leibniz University of Hanover, Hanover, Germany
e-mail: broekel@wigeo.uni-hannover.de

M. Hartog
Section of Economic Geography, Faculty of Geosciences, Utrecht University, P.O. Box 80115, 3508 TC Utrecht, The Netherlands
e-mail: m.hartog@uu.nl

T. Scherngell (ed.), *The Geography of Networks and R&D Collaborations*, Advances in Spatial Science, DOI 10.1007/978-3-319-02699-2_4,

4.1 Introduction

There is growing scientific interest in the creation of knowledge and its diffusion among organizations. In the new growth theory, new knowledge is regarded as being pivotal to economic growth by generating increasing returns (Romer 1990). In evolutionary economics, the re-combination of existing knowledge from different sources is argued to be crucial for new innovations to occur (Nelson and Winter 1982). These theories and the according empirical evidence also impacted the policy level. For instances, one of the most well known policy instruments to stimulate knowledge diffusion and innovation are the Framework Programmes of the European Union. These programs have been in existence since 1984 and are used to fund thousands of collaborative research projects between organizations in the EU.

Such R&D collaboration networks, which are induced by policy, alter the spatial diffusion of knowledge. This put the investigation of their spatial structures on the agenda of regional economists and economic geographers (Autant-Bernard et al. 2007). The geographical structures of inter-organizational collaboration networks are frequently analyzed from an organizational perspective (cf. Giuliani and Bell 2005) and a regional perspective, the latter focusing on cross-regional R&D collaboration networks (cf. Scherngell and Barber 2009, 2011; Hoekman et al. 2010). In order to investigate factors explaining the structure of cross-regional networks, most commonly used are spatial interaction models, which allow for considering factors at the node and dyad level. An example of a factor at the node level is the size of a region that matters as regions with more organizations are also more likely to have links to regions elsewhere. At the dyad level, most attention has been paid to the effect of geographical distance, which has been found to have a negative impact on the chance of research collaboration (cf. Ponds et al. 2007; Scherngell and Barber 2009; Hoekman et al. 2009, 2010).

In addition to the node and dyad level, factors at the structural network level may also be important, though. That is, the creation of new links might not only depend on attributes of regions or region pairs, but may also be influenced by the existing structure of the cross-regional network. For instance, a key hypothesis in organizational network science is the tendency towards triadic closure (or transitivity), which implies in this context that regions, which are indirectly linked, are more likely to link themselves as well. However, factors at the structural network level cannot be included in spatial interaction models.

This chapter presents exponential random graph models (ERGM) as an alternative empirical tool to investigate this. These models have been developed in mathematical sociology in recent years (Snijders et al. 2006; Robins et al. 2006, 2007; Wang et al. 2012) and are increasingly used across scientific disciplines, for example in bioscience (Saul and Filkov 2007), political science (Desmarais and Cranmer 2012) and organization science (Uddin et al. 2012). The advantage of these models is that they allow for simultaneously estimating the effect of factors at the node, dyad, and structural network level for networks that are observed at one point in time. We illustrate the usefulness of ERGMs by exemplarily investigating the structure and its determinants of the cross-regional R&D collaboration network in the German chemical industry between 2005 and 2010.

The study is structured as follows. The second section gives an overview of the literature on spatial structures of R&D collaboration networks and their determinants. This includes a brief discussion of factors at the node, dyad, and structural network level that may impact network structures. The third section elaborates on the exponential random graph model that we subsequently use to investigate the structure of the cross-regional network. We present the empirical data in the fourth section. It is followed by the discussion of the results in the fifth section and some concluding remarks in the sixth section.

4.2 Determinants of Cross-Regional R&D Collaboration

The structural determinants of cross-regional R&D collaboration networks can be distinguished at three different levels. These are the node level, the dyad level, and the structural network level. In the following, we elaborate on the factors effective at these three different levels.

Node level factors are properties of network entities themselves. With respect to regional R&D collaboration networks, regions' size and research intensity are particularly important. Regions with more organizations can be expected to have more ties because they have more collaboration opportunities. Such a size effect also applies at the firm level, as large organizations are likely to have more ties than small organizations because their position in the industry is more prominent and have more resources at their disposal to create and maintain ties. For instance, Boschma and Ter Wal (2007) find that larger organizations are more central in the knowledge network of footwear producers in Barletta. Secondly, the research intensity of organizations in a region matters. At the firm level, Giuliani and Bell (2005) show that organizations with a more advanced knowledge base are more frequently approached by other organizations to exchange knowledge because they are perceived to be 'technological leaders'. A similar argument can be applied to the regional level: the research intensity of a region is generally characterized by a large number of R&D employees, many organizations being engaged in R&D-intensive activities, and by the presence of universities or other research institutes. These characteristics are likely to increase the number of research collaboration links organizations have with other organizations in the same region (regional collaboration) as well as with organizations located elsewhere (cross-regional collaboration), with the latter representing a region's (degree) centrality in the cross-regional collaboration network. Accordingly, it can be expected that the absolute numbers of regional and cross-regional links are strongly correlated.

Factors at the dyad level are characteristics of relationships between two entities (nodes) in a network. In the context of the paper it refers to the relation between two regions. A key idea in sociology is that entities are more likely to link when they have similar attributes, known as homophily effect (McPherson et al. 2001). For instance, regions with organizations that operate with similar routines and under comparable incentive mechanisms are more likely to be linked in R&D collaboration. Another example are universities, which are subject to different incentive frameworks than firms when it comes to knowledge creation and diffusion as they aim to publish new knowledge, whereas firms have an incentive to keep new knowledge secret. Hence, because of their institutional proximity (Metcalfe 1995), universities are more likely to collaborate with others and especially with other universities (cf. Broekel and Boschma 2012; Broekel and Hartog 2013). This is likely to translate to the regional level as regions rarely house more than one university. Accordingly, university regions have a higher likelihood of being linked to each other.

In addition to institutional proximity, other forms of proximity may also be relevant, namely: geographical proximity, technological proximity, and social proximity. Many studies confirm that cross-regional R&D collaboration is more likely when regions are located close to one another in space (e.g. Maggioni et al. 2007; Scherngell and Barber 2009; Hoekman et al. 2009, 2010). This may be due to a variety of reasons, for instance geographical proximity facilitates face-to-face contact, which stimulates the diffusion of information about potential collaboration partners. The likelihood of cross-regional R&D collaboration is shown to increase when regions have similar technological profiles and specializations (Fischer et al. 2006; LeSage et al. 2007; Scherngell and Barber 2009). A potential explanation is that organizations are more prone to collaborate with organizations with related knowledge assets. Similar technological profiles (technological proximity) ensure that two organizations can easily communicate and learn from each other (Cohen and Levinthal 1990; Nooteboom 2000). Social proximity may also increase the likelihood of R&D collaboration (cf. Autant-Bernard et al. 2007). People already knowing each other find it easier to develop trust-based relations, which in turn facilitate knowledge exchange and ease interactions across regional boundaries (Maskell and Malmberg 1999; Sobrero and Roberts 2001; Breschi and Lissoni 2009).

In addition to these factors at the node and dyad level, factors at the structural network level may also matter for the structure of cross-regional R&D collaboration. Such factors relate to properties of the entire network. Three factors are commonly put forward in this context: triadic closure (transitivity), multi-connectivity, and preferential attachment (cf. Ter Wal and Boschma 2009; Glückler 2010). Triadic closure predicts that partners of organizations are likely to become partners themselves as well. As a result, a network will consist of many triangles, i.e. dense cliques of strongly interconnected organizations (Ter Wal 2011). Such cliques can be regarded as a sign of social capital (Coleman 1988) that may enhance trust and willingness among actors to invest in mutual goals, such as research collaboration. In contrast, multi-connectivity suggests that organizations will connect to others in multiple ways to decrease the dependency on a single link. It implies that multiple paths are formed amongst organizations leading to multiple reachability. Evidence for this is found in the creation of inter-firm alliances between US biotech firms (Powell et al. 2005). Preferential attachment means that organizations with many links are more likely to create or attract new links in the future. If a network is shaped by this factor, its degree distribution follows a power law (Barabasi and Albert 1999). Gulati (1999) shows that in the case of multinational firms, the likelihood of creating new alliances increases the better organizations are connected in the network. Hence, the network of alliances among multinational firms is subject to preferential attachment processes.

In contrast to most of the discussed factors at the node and dyad level, these factors are not regional in nature. Concepts like transitivity, preferential attachment or reciprocity do not apply to the regional level. However, in most empirically observed cross-regional networks, links are constructed from regionally aggregated inter-organizational relations. To the extent that these inter-organizational relations involve organizations being located in different regions such effects will naturally be translated to the cross-regional network. Accordingly, they need to be taken into account when analyzing the network structure as multi-connectivity, preferential attachment, and triadic closure also shape the empirically observed cross-regional networks.

To estimate the relative impact of the above factors on the structure of a network, they need to be simultaneously incorporated in the empirical model. This is not possible with the models most frequently used to investigate cross-regional collaboration: spatial interaction models in general and gravity models in particular (cf. Scherngell and Barber 2009). These models can account for factors at the node and dyad level. However, they cannot be used to evaluate factors at the structural network level. In light of the theoretical relevance of factors at the structural network level, we therefore argue that network analysis modeling techniques represent a powerful alternative because they are able to simultaneously incorporate factors at all three levels.

When longitudinal data is available, a stochastic actor-based network approach can be used. It models the change of a network from one point in time to another as part of an iterative Markov chain process (see for technical details: Snijders et al. 2010). When it comes to the analysis of research collaboration networks of regions, however, such an approach is less useful. By aggregating collaboration data to the regional level and creating cross-regional networks, researchers generally are interested in approximating the relational interaction structures of regions and investigate their structures and determinants. Such networks are unlikely to drastically change within short time periods, though, as they are results of long-term social, regional, and industrial evolution processes. Hence, even when longitudinal data on these cross-regional networks structures are available, it is unlikely to cover a sufficiently long time period. It may include multiple time periods (years) and thereby principally allow for employing longitudinal network analysis to study changes in the underlying cross-regional interaction structures.[1] However, the

[1] The relational data derived from the 5th, 6th, and 7th EU-Framework Programmes are (currently) a good example in this respect. While they represent longitudinal data, it covers only a limited

results generated with stochastic actor-based network approaches are unlikely to yield meaningful insights because the empirically observed changes in the network structures are dominated by short-term fluctuations that are of little interest to the researcher. We therefore argue that exponential random graph models are the preferred option when investigating the structure of cross-regional interaction on the basis of data with a cross-sectional nature and factors at the structural network are to be considered. We elaborate on these models in the next section.

4.3 Exponential Random Graph Models

Exponential random graph models are stochastic models that approach link creation as a time-continuous process. They regard a network observed at one point in time as one particular realization out of a set of multiple hypothetical networks with similar properties. This allows applying these models to purely cross-sectional network data.

The aim of exponential random graph models is to identify factors that maximize the probability of the emergence of a network with similar properties as the structure of the observed network. The general form of exponential random graph models is defined as follows (Robins et al. 2007):

$$\Pr(Y = y) = \left(\frac{1}{\kappa}\right)\exp\left\{\sum\nolimits_{\mathrm{A}} \eta_{\mathrm{A}}\, g_{\mathrm{A}}(y)\right\} \tag{4.1}$$

where $\Pr(Y = y)$ is the probability that the network (Y) generated by an exponential random graph is identical to the observed network (y), κ is a normalizing constant to ensure that the equation is a proper probability distribution (summing up to 1), η_{A} is the parameter corresponding to network configuration A, and $g_{\mathrm{A}}(y)$ represents the network statistic. Network configurations can be factors at the node level, dyad level and structural network level.

Estimation of the parameters is done with maximum pseudo likelihood or a Markov Chain Monte Carlo Maximum Likelihood Estimation procedure. The latter has been developed most recently and is regarded as the preferred procedure as it is often most accurate (Snijders 2002; Van Duin et al. 2009). It is based on the generation of a distribution of random graphs by stochastic simulation from a starting set of parameter values, and subsequent refinement of those parameter values by comparing the obtained random graphs against the observed graph. This process is repeated until the parameter estimates stabilize. If they do not, the model might fail to converge and hence becomes unstable (see for technical details, Handcock 2003; Hunter et al. 2008).

time-period (1998–2013). Of course, this may change when data on future programs will become available.

Checking whether the parameters predict the observed network well, i.e. evaluating a model's goodness of fit, is done by comparing the structure of the simulated networks to the structure of the observed network. According to Hunter et al. (2008), the comparison consists of the degree distribution, the distribution of edgewise shared partners (the number of links in which two organizations have exactly k partners in common, for each value of k), and the geodesic distribution (the number of pairs for which the shortest path between them is of length k, for each value of k). The more the distributions of the simulated networks are in line with those of the observed network, the more accurate are the parameters of the exponential random graph model. In the next section, we construct an exponential random graph model to investigate the structure of the network of subsidized R&D collaboration in the German chemical industry.

4.4 Determinants of Cross-Regional R&D Collaboration in the German Chemical Industry

4.4.1 Data

We analyze R&D collaboration that has been funded by the German federal government. As in most other advanced countries, the government actively supports public and private R&D activities with subsidies. While the Federal Ministry of Education and Research (BMBF) is the prime source of subsidies, the Federal Ministry of Economics and Technology (BMWi) and the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) contribute as well. The federal ministries publish comprehensive information about subsidized projects in the so-called "Förderkatalog" (subsidies catalog). This catalog contains detailed information on more than 150,000 individual subsidies that have been granted between 1960 and 2012. The catalog also includes information on the cooperative nature of projects. It specifically indicates if projects are joint projects realized by consortia of organizations. Participants in joint projects agree to a number of regulations that guarantee significant knowledge exchange between the partners. Accordingly, two organizations are defined to cooperate if they participate in the same joint project. Hence, the original network is a two-mode network (project-organizations links), which we transform into a one-mode projection of the network (organization-organization links). All organizations can be assigned to labor market regions allowing for regionalizing the network (see for more details on the data Broekel and Graf 2012). The data is comparable to the EU Framework Programmes (EU-FP) data by and large, which is extensively used to model research collaboration networks (cf. Scherngell and Barber 2009). In contrast to the EU-FP data, the data at hand exclusively covers collaboration between German organizations.

To construct the network of subsidized R&D collaboration in the German chemical industry, we first identify all firms in the data that are classified as being involved in the 2-digit NACE code C20 'Manufacture of chemicals and chemical products'. Subsequently, all cooperative projects are extracted in which at least one of these firms participates. On the basis of the joint appearance in a project, we construct the inter-organizational network among all chemical firms participating in these projects. We only consider links among firms: links to universities, research organizations, associations, and to firms belonging to other industries are excluded. We believe that this approach provides the most conservative picture of the (subsidized) R&D collaboration network in the chemical industry. Alternatively one may consider all organizations active in joint projects in which at least one firm of the chemical industry is participating. However, such seems to be a very broad definition of an industry-specific network, which makes the definition of appropriate empirical variables more difficult. We acknowledge that the links to organizations in other industries are also likely to shape the intra-industry network, but as our main focus is on the impact of the factors at the three different levels (node, dyad, structural) rather than on knowledge exchange as such, we leave this for future research.

The corresponding inter-organizational undirected network is subsequently aggregated to the regional level using information on organizations' location in the 270 German labor market regions. The 270 labor market regions are defined by the German Institute for Labor and Employment (e.g. Greif and Schmiedl 2002). We construct the network that existed between 1 January 2005 and 31 December 2010. In this period, 775 projects were subsidized in which at least one firm of the chemical industry was involved. These projects are split into 975 individual funds allocated to 557 German firms belonging to the chemical industry.[2] 133 of the 775 projects are joint projects, which involve on average 2.8 firms. The resulting cross-regional R&D collaboration network is shown in Fig. 4.1.

The network is dichotomized, as we are only interested in whether or not a link exists between regions. The figure shows that the large agglomerations of the Ruhr Area, Frankfurt am Main, and Munich are important nodes in the network. In addition, a number of central regions are located along the Rhine River in the west. The region of Dresden is a central node in East Germany. All these regions are well-known centers of the chemical industry in Germany. Some additional descriptive statistics of the network are presented in Table 4.2 in the Appendix.

[2] This figure is based on the number of executing organizations ("*Ausführende Stelle*") as given in the data. Many of these organizations are part of larger organizations. This has however little relevance for the results as all data are aggregated to the regional level.

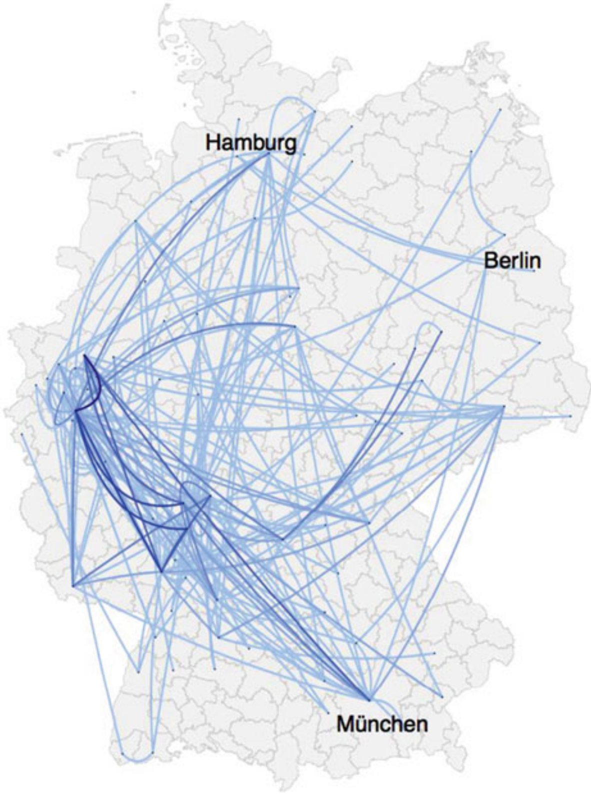

Fig. 4.1 Network of subsidized R&D collaboration among firms in the German chemical industry (2005–2010)

4.4.2 Construction of Empirical Variables

Node Level Variables

The most important node-level factors likely are the intensity of regional R&D and innovation activities in the field of chemistry. Foremost, this is because undertaking R&D activities is necessary to receive R&D subsidies. Regions with large R&D activities are likely to host more organizations that are involved in R&D collaboration. Moreover, such regions may also be the location of the most successful innovators, which are preferred collaboration partners. We therefore consider the number of applied patents in chemistry by regional organizations as proxy for the intensity of regional R&D activities in this field. The regionalized data on patent applications are published in Greif and Schmiedl (2002) and Greif et al. (2006), which include applications to the German as well as to the European Patent Office, with a correction for double counts. The patents are assigned to labor market regions according to the inventor principle. The patent data is organized according to IPC-classes, which is matched to the 2-digit NACE industry using the concordance of Broekel (2007). Lacking the data for the years 2005–2010, we construct the first node-level variable as the summed number of patents of regional firms in the field of chemistry in the years 2001–2005.[3] The variable is denoted as PATS.

We take into account the effect of urbanization by including population density (POP) and the gross-domestic product (GDP) of a region in the year 2005. The corresponding data are obtained from the German Federal Institute for Research on Building.

Firms located in regions with strong public research infrastructure may also be more likely to link across regions. For instance, being co-located with public research institutes may induce knowledge spillovers and give better access to highly qualified personnel (e.g. Fritsch and Slavtchev 2007). Accordingly, firms in these regions may be more prone to conduct R&D, engage in R&D collaboration, and be more successful in terms of innovation. In order to approximate this, we measure regions' public R&D infrastructure with three variables. The presence of universities in a region is modeled by counting their numbers of graduates in natural sciences in 2005 (UNI). Similarly, the analysis includes the number of employees working in regional research institutes of the Max Planck Society (MPG) and the Fraunhofer Society (FHG). More precise, only the numbers of employees working in the institutes' technological or natural science institutes in the year 2005 enter the analysis.[4]

Dyad Level Variables

We construct three variables at the dyad level. We measure geographical proximity with the physical distance between two regions' geographic centers. The variable is denoted as (GEO_DIST). The chance of two regions being linked is expected to decrease with geographical distance. Geographical proximity frequently correlates with social proximity (Boschma 2005), which needs to be considered in the interpretation.

We also include the variable SAME_REG that has a value of 1 if both regions are located in the same federal state (i.e. NUTS 1 region), and 0 if not. SAME_REG not only accounts for geographical proximity. It is likely to represent institutional proximity as well, as regions in the same federal state are probably similar in their R&D-related institutional framework. The reason for this is the significant role the federal level is playing in the German R&D landscape. For instance, each federal state ("Bundesland") is responsible for its own resource endowment of universities and has its own R&D policies.

We also take into account that two regions with universities may be more likely to be linked. Firms in such regions are probably structurally more similar than two

[3] The latest version of the "*Patentatlas*" was published in 2006 and includes the patent data up to 2005. We use the aggregated numbers for 2001–2005 to minimize annual fluctuation.

[4] The employment numbers are relatively stable over time. Using data for a single year is therefore considered appropriate.

firms of which one is not located in a university region. It can be expected that firms in university regions are more R&D intensive and technologically more advanced as are more probable to benefit from knowledge spillovers (cf. Jaffe 1989). To take this into account, we include the variable UNI_1, which has a value of one if both regions have a university and zero otherwise.

Notably, we do not construct a measure of technological similarity, which has been shown to make regions more likely to be linked (Scherngell and Barber 2009). This is primarily motivated by data constraints. We analyze a network among firms of the same industry aggregated at the regional level. Hence, for the construction of a meaningful technological similarity measure we need information about the technological profiles of all regional firms in the chemical industry. Unfortunately, we miss such information and have to leave this issue to future research.

Structural Network Level Variables

We include four variables at the structural network level. Triadic closure (or transitivity) is captured by the geometrically weighted edgewise-shared partner statistic (GWESP-statistic: Snijders et al. 2006; Hunter et al. 2008). It measures the number of triangles in the network whilst taking into account the number of links that are involved in multiple triangles (multimodality) (see for details: Hunter et al. 2008). It thereby captures how frequently two nodes are connected by a direct link as well as by an indirect connection of length 2 (i.e. "two-path") through another node (e.g. Hunter 2007). If a positive coefficient is found for this statistic, there is a tendency towards triadic closure in the network.

We consider the geometrically weighted dyad shared partner statistic (GWDSP), which is an advanced version of the alternating k-two-path statistic put forward by Snijders et al. (2006). It measures the extent to which a network shows a tendency of nodes not directly linked to each other being at least indirectly linked. In other words, the statistic approximates whether multiple paths exist between such nodes. Accordingly, it captures multi-connectivity for nodes that are not tied directly.

Another variable at the network level is EDGES. It equals the number of links in the network and should always be included in exponential random graph models. Moreover, EDGES represents a so-called k-star(1) parameter. K-stars are essential configurations in networks. For instance, a k-star(2), or 2-star, corresponds to three nodes of which one is linked to each of the other two. Accordingly, a k-star(3) shows as four nodes with one node being linked to the other three. A triangle, i.e. three mutually connected nodes, logically includes three k-stars(2). This means that these configurations are hierarchically related (cf. Snijders et al. 2006; Hunter 2007). While the EDGES parameter corresponds to a type of intercept parameter in the model, it is especially useful when considering the GWDEGREE statistic as well.

GWDEGREE is the geometrically weighted degree statistic, which helps modeling the observed network's degree distribution. Notably, the statistic can also be seen as an equivalent to the more traditional k-star statistic (Hunter 2007). When being considered alongside the EDGES statistic, GWDEGREE (broadly) allows modeling preferential-attachment processes. More precise, if this statistic obtains a positive coefficient it signals the presence of preferential-attachment and a negative coefficient indicates anti-preferential attachment (Hunter 2007).

For all three statistics, GWESP, GWDSP, and GWD, decay parameters have to be specified. Because few attempts have been made to systematically identify the best fitting parameter combinations (cf. Wright 2010), researchers commonly rely on a manual iterative trail-and-error process of estimating varying model specifications. These specifications differ in terms of included variables as well as decay parameters of the GWDSP, GWESP and GWDEGREE statistics. This process ends when the best fitting model is identified. The best fitting model is a model that is stable and converges (when the Markov Chain Monte Carlo approach is used, the parameter traces should be horizontal) and provides the most appropriate goodness-of-fit statistics (matching degree, edgewise shared partners, and geodesic distributions) given the empirical data (observed network). In other words, the best fitting model most accurately predicts the structure of the observed network.

Once this model is identified the final goodness-of-fit statistics and MCMC trace plots are generated exclude all variables that are not significant at the 0.05 level in the original model. These variables are excluded because they represent noise that may distort the model and thereby bias the according statistics (cf. Wright 2010). This "refined" model is used to generate all goodness-of-fit related statistics. We present the best fitting ERG-model for the cross-regional R&D collaboration network in the next section.

4.5 Results

Table 4.1 presents the results of the final, i.e. best fitting, model and those of its refined variant. Included are factors at the node, dyad, and structural network level. The model is stable and converges. Moreover, it is characterized by appropriate goodness-of-fit statistics (matching degree, edgewise shared partners, and geodesic distributions (Fig. 4.2 in the Appendix) and horizontal parameter traces (Figs. 4.3, 4.4, 4.5, and 4.6 in the Appendix).

Before we discuss the variables with significant coefficients, it is worthwhile to take a brief look at the insignificant ones. The insignificance of GDP implies that the economic prosperity of regions does not impact the structure of the cross-regional R&D collaboration network in the German chemical industry. The measure of the absolute physical distance (GEO_DIST) between regions better captures the effect of geographic distance than when considering whether two regions are part of the same federal state (SAME_REG), as the latter's coefficient is insignificant while that of the first is not. The finding moreover questions the role of institutional proximity, which we argued to be reflected by SAME_REG.

The measure of the network's degree distribution (GWDEGREE) does not help in explaining the structure of the network. This means that we do not find evidence for preferential attachment processes, i.e. well-connected regions are not more

Table 4.1 Results of exponential random graph model with dyad level, node level and structural network level variables

	Main model				Refined model		
Variable	Estimate	Std. error	p-Value	Sign.	Estimate	Std. error	Sign.
Node level							
PATS	0.00056	0.00013	< 1e-04	***	0.00028	0.00008	***
UNI	−0.00069	0.00017	< 1e-04	***	−0.00119	0.00015	***
POP_DEN	0.00009	0.00004	0.022735	*	0.00022	0.00001	***
GDP	−0.00113	0.00159	0.478296				
MPG	0.00037	0.00011	0.000882	***	0.00071	0.00009	***
FHG	0.00064	0.00026	0.013101	*	0.00135	0.00016	***
Dyad level							
GEO_DIST	−0.00164	0.00021	< 1e-04	***	−0.00072	0.00018	***
SAME_REG	0.07019	0.10950	0.521505				
Nodematch. UNI_1	0.30200	0.07094	< 1e-04	***	0.14760	0.07873	*
Structural network level							
EDGES	−4.36800	0.17230	< 1e-04	***	−7.24000	0.20440	***
GWESP, 0.69, fix	1.04400	0.06772	< 1e-04	***	2.02	0.00902	***
GWDEGREE	−2.86600	14.81000	0.846554				
GWDSP, 0.15, fix	0.02133	0.02736	0.435589				
Null deviance:	50343.3 on 36,315 degrees of freedom				50343.3 on 36,315 degrees of freedom		
Residual deviance:	1753.3 on 36,302 degrees of freedom				1619.3 on 36,305 degrees of freedom		
Deviance:	48589.o on 13 degrees of freedom				48724.0 on 9 degrees of freedom		
AIC:	1779.3				1639.3		
BIC:	1889.8				1724.3		

*Significant at 95 %; ***Significant at 99 %

prone to gain additional links than sparsely connected regions. The same applies to the GWDSP-statistic suggesting that two regions without a direct link are unlikely to be indirectly connected. Accordingly, we observe insignificant coefficients for variables at all three levels (node, dyad, and structural network level).

Now, we turn towards the significant variables reported in Table 4.1. As expected, regions with R&D intensive firms (PATS) tend to have more links. The same applies to urban regions (POP_DEN) and regions in which institutes of the Max-Planck (MPG) and Fraunhofer (FHG) societies are located. The according coefficients of PATS, POP_DEN, MPG, and FHG are all positive and significant. UNI obtains a negative significant coefficient suggesting that university regions tend to have fewer links. While this contradicts our expectations, it is essential to also consider the positive significant coefficient of the dyad-level variable UNI_1 in the explanation. Accordingly, university regions generally have less links but they are more likely to link to other university regions. The latter is in line with our expectations and signals the presence of a homophily effect.

The dyad-level variable GEO_DIST is characterized by a negative significant coefficient. Hence, geographical distance hampers link creation, which confirms existing empirical studies (cf. Maggioni et al. 2007; Ponds et al. 2007; Scherngell and Barber 2009; Hoekman et al. 2009, 2010; Broekel and Boschma 2012).

We argued above that the main advantage of exponential random graph models is their ability to take into account factors at the structural network level in addition to factors at the node and dyad level. The significant coefficients of two variables at the structural network level empirically confirm this level's relevance. The coefficient of EDGES is negative and significant. By being similar to an intercept variable, EDGES represents the overall density of the network when all other effects are excluded. Its negative coefficient is a common feature of networks established by social processes indicating that such networks tend to be less dense than exponential random networks (cf. Varas 2007).

In addition, we find a positive and significant coefficient of the GWESP-statistic. It means that triangles are a common feature of the network, which corresponding to the visual inspection of the network (see Fig. 4.1). In other words, regions that are directly linked are also more likely to link through indirect connections. Hence, the result suggests that triadic closure is a driving force in the network formation processes. There might however be an alternative explanation. When constructing the empirical network, we transformed a bipartite network into a one-mode type. Such transformation more or less automatically increases the likelihood of triplets in the final one-mode network. Accordingly, the positive GEWSP-statistic might pick up this effect and act as a kind of control parameter for the one-mode projection procedure. However, we pointed out in Sect. 4.1 that on average less than three firms (2.8) are jointly participating in a cooperative project. Hence, it is most likely a combination of both effects that explains the statistic's significance. In any case, this structural network factors significantly helps in modeling the structure of the network.

In sum, we find that the structure of the network is best explained by factors at the node level, dyad level, and structural network level. Moreover, the coefficients (which can be translated into odd-ratios by taking the exponential) underline that in comparison to factors at the dyad, factors at the structural network level have greater explanatory power. It shows the crucial importance of these factors for the structure of the cross-regional R&D collaboration network in the German chemical industry. This result thereby also highlights the usefulness of exponential random graph models as a tool for analyzing the structure of such types of networks.

4.6 Conclusion

The aim of this study was to discuss exponential random graph models (ERGM) as promising tools for the investigation of cross-regional collaboration networks. We pointed out that most existing studies focus on the evaluation of factors at the node and dyad level. However, network science suggests that factors at the structural network level may also be relevant in this respect. Such factors cannot be considered in methods commonly applied in this context. For instance, spatial interaction models allow only for factors at the node and dyad level. We argued that ERG-models represent a powerful alternative as they take into account factors at all three levels and require only cross-sectional network data.

We illustrated the application of ERGMs by analyzing the structure of the cross-regional R&D collaboration network in the German chemical industry between 2005 and 2010. By using an exponential random graph model, we considered factors at all three levels that might influence the network's structure. At the node level, it was shown that urban regions (reflected by population density) and regions with high research intensities are more likely to be linked to other regions. At the dyad level, we found regions to be more likely being linked when they have a university. Moreover, our results confirmed the negative impact of geographical distance on the likelihood of research collaboration. Finally, at the structural network level, evidence was provided for the existence of a triadic closure (transitivity) effect: regions that are indirectly linked to each other are likely to be directly linked as well.

A challenge for future research is the projection of networks among individuals and organizations to the regional level. This particularly concerns the question about what factors impact link formation at the level of the individual (e.g. trust, reciprocity), at the level of the organization (e.g. reputation, absorptive capacity), and at the spatial (regional) level (e.g. image, collective identity). In the present paper, and in most of the corresponding literature, these factors are all translated to the same level, i.e. that of the chosen unit of analysis. However, this ignores their relevance at different observational levels. For instance, a general finding is that regions with high research intensity are more likely to be linked to each other, but in theory it could be that the actual linkages between those regions are created by organizations that in contrast to all other organizations in their respective regions, show little or no research intensity (although this is unlikely). The same applies to the factors at the structural network level. For instance, if three organizations in three different regions link with each other a triangle will be observed in the network that might suggest the presence of a triadic closure effects. However, if two of the three organizations are located in the same region, the cross-region network shows a single link instead, which does not supports this interpretation. In this sense, the question of what is the most appropriate unit of analysis (and level of aggregation) becomes evident. This clearly lays the path for future research focusing on changing network structure when moving from one level of node aggregation to another. Researchers will have to adjust the level of node aggregation in correspondence to the objective of their investigation until reliable insights on this matter are available.

Clearly, the study is only a first step towards understanding the role factors at the structural network level play for the formation of cross-regional collaboration networks. It nevertheless underlines the usefulness of exponential random graph models for future research endeavors on this subject.

Appendix

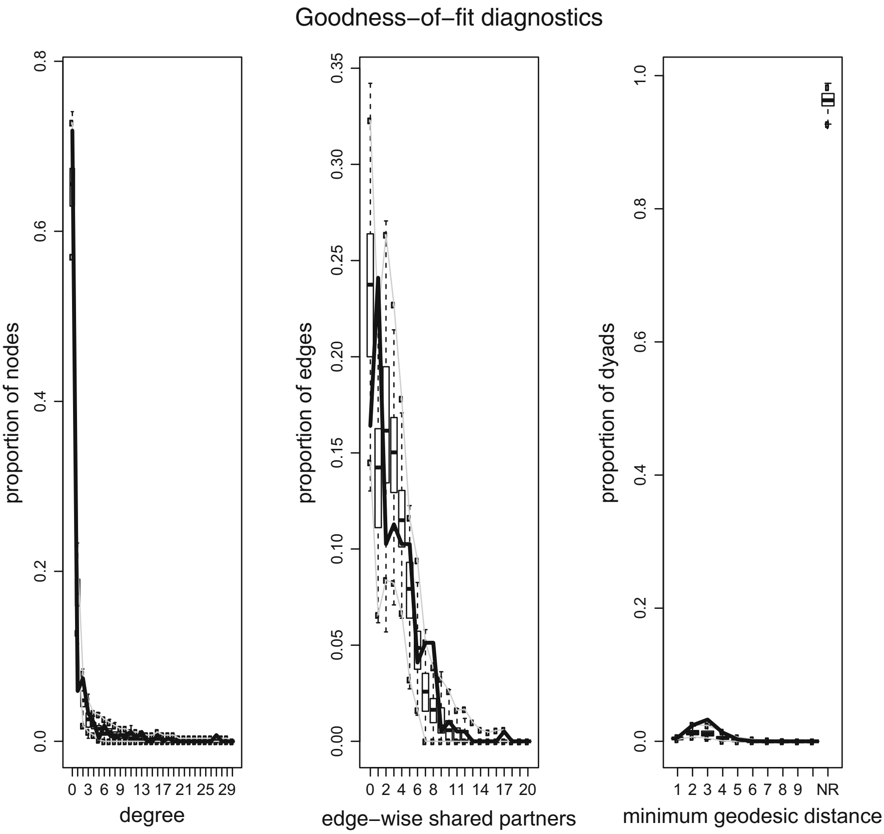

Fig. 4.2 Goodness of fit of exponential random graph model with dyad level, node level + structural network level variables

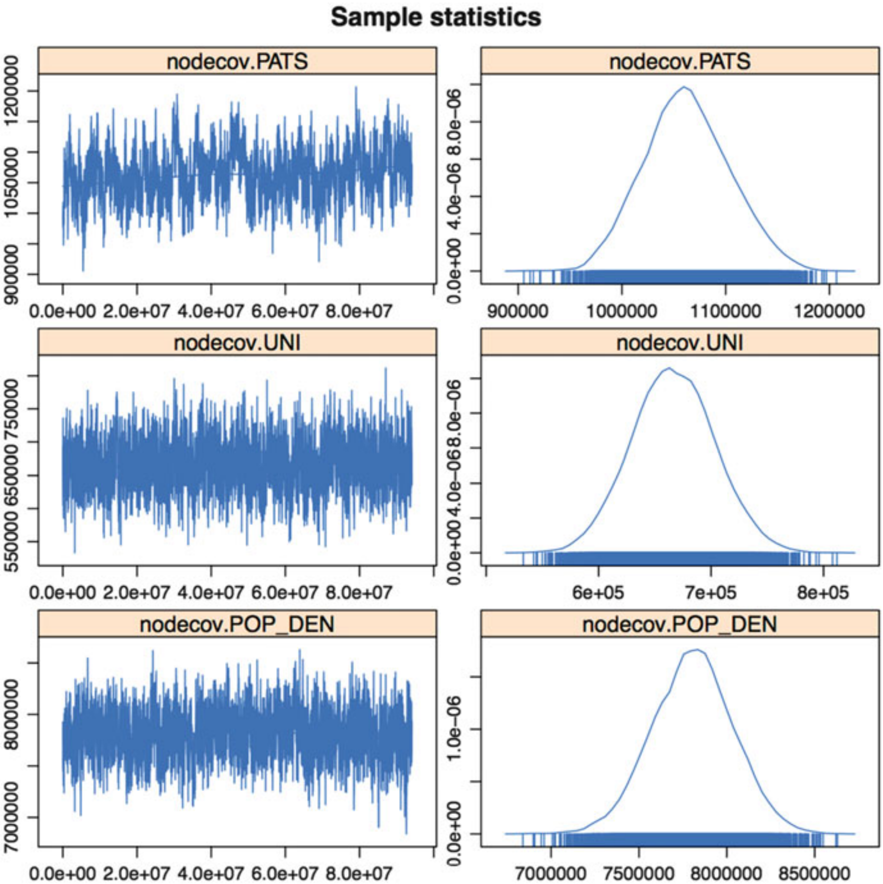

Fig. 4.3 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

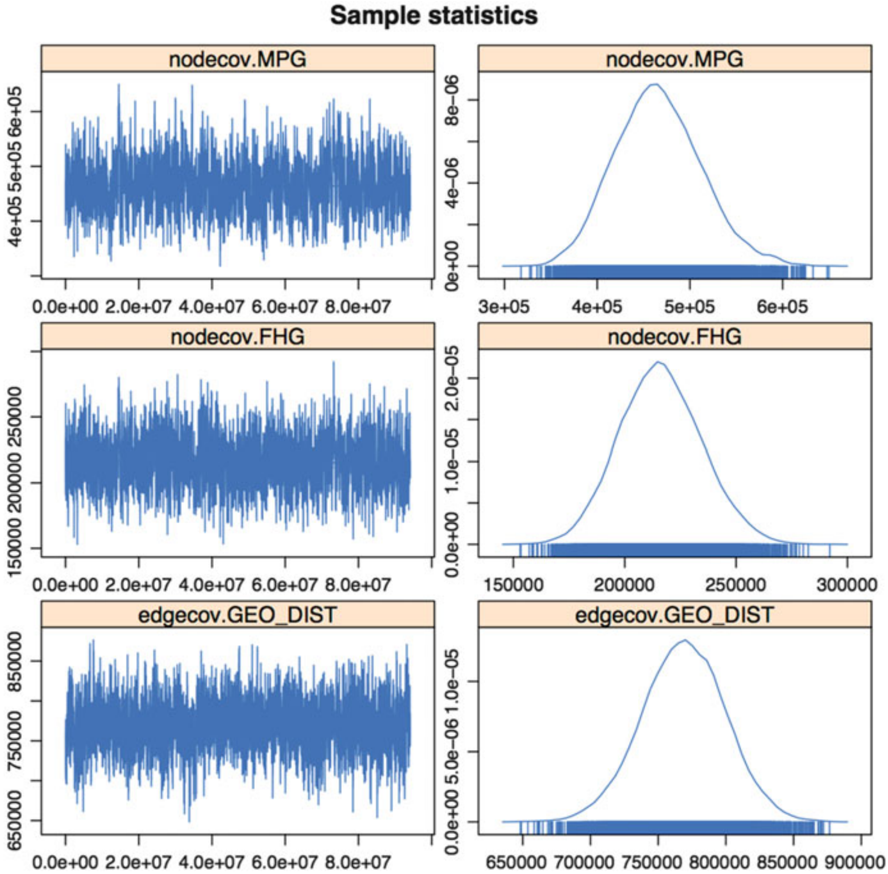

Fig. 4.4 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

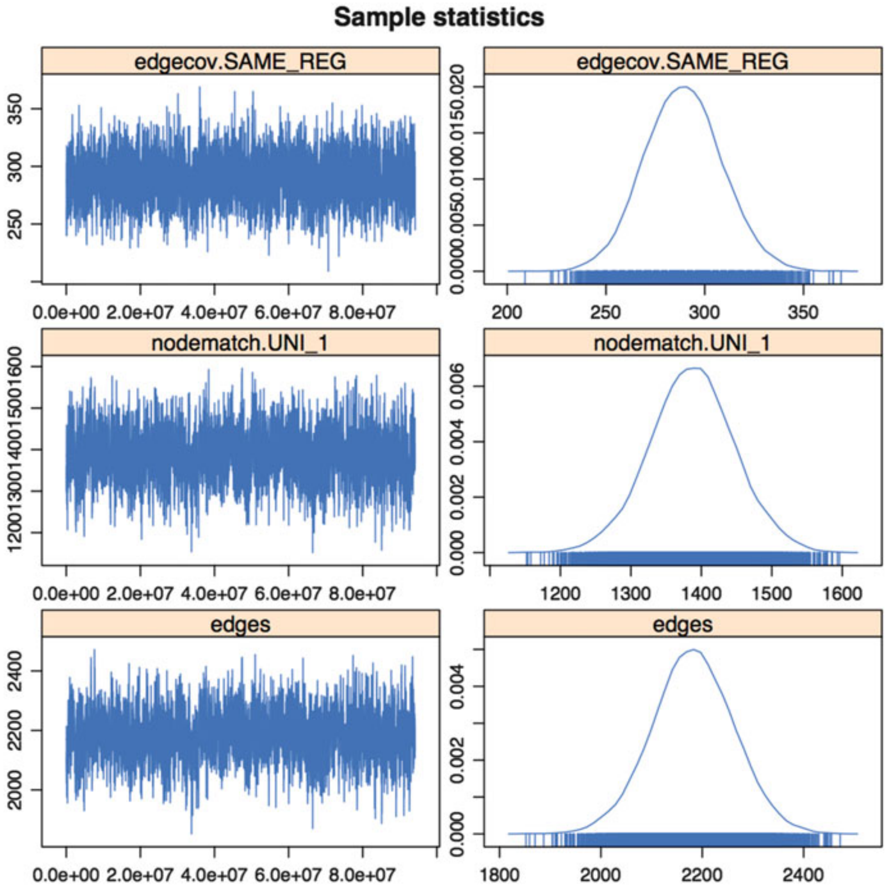

Fig. 4.5 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

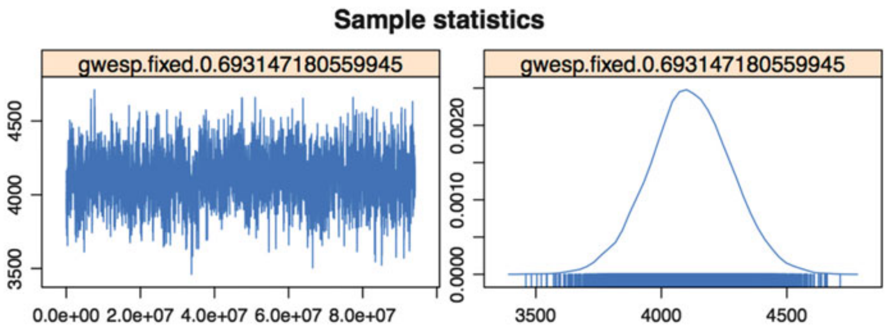

Fig. 4.6 MCMC-Statistics of exponential random graph model with dyad level, node level and structural network level variables

Table 4.2 Descriptives of empirical variables

Variables	n	Mean	St. deviation	Median	Min	Max	Skew	Kurtosis
PATS	270	69.55	199.12	12.48	0	1,691.31	5.34	32.55
POP_DEN	270	825.35	1,265.19	244.5	40	8,523	3.06	11.44
GDP	270	40.46	33.58	26.75	14.1	296.9	3.66	19.83
UNI	270	101.51	244.73	0	0	1,812	3.46	15.55
MPG	270	49.12	248.08	0	0	3,438	10.20	128.50
FHG	270	30.81	123.52	0	0	978	5.22	29.24
GEO_DIST	72,900	379.81	186.03	368.54	0	977.45	0.29	−0.52
SAME_REG	72,900	0.11	0.31	0	0	1	2.49	4.22
UNI_1	72,900	0.62	0.49	1	0	1	−0.47	−1.77

References

Autant-Bernard C, Billand P, Frachisse D, Massard N (2007) Social distance versus spatial distance in R&D co-operation: empirical evidence from European collaboration choices in micro and nanotechnologies. Pap Reg Sci 86:495–519

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 15:509–512

Boschma R (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Boschma RA, Ter Wal ALJ (2007) Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the south of Italy. Ind Innov 14:177–199

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. J Econ Geogr 9:439–468

Broekel T (2007) A concordance between industries and technologies – matching the technological fields of the Patentatlas to the German industry classification. Jenaer economic research papers, 2007-013

Broekel T, Boschma R (2012) Knowledge networks in the Dutch aviation industry: the proximity paradox. J Econ Geogr 12:409–433

Broekel T, Graf H (2012) Public research intensity and the structure of German R&D networks: a comparison of 10 technologies. Econ Innov New Technol 21(4):345–372

Broekel T, Hartog M (2013) Explaining the structure of inter-organizational networks using exponential random graph models: does proximity matter? Industry Innov 20(3):277–295

Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. Adm Sci Q 35:128–152

Coleman JS (1988) Social capital in the creation of human capital. Am J Sociol 94(Suppl): S95–S120

Desmarais BA, Cranmer SJ (2012) Micro-level interpretation of exponential random graph models with an application to estuary networks. Policy Stud J 40. doi: 10.1111/j.1541-0072.2012.00459.x

Fischer MM, Scherngell T, Jansenberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe. Evidence from a spatial interaction modeling perspective. Geogr Anal 38:288–309

Fritsch M, Slavtchev V (2007) Universities and innovation in space. Ind Innov 14:201–218

Giuliani E, Bell M (2005) The micro-determinants of meso-learning and innovation: evidence from a Chilean wine cluster. Res Policy 34:47–68

Glückler J (2010) The evolution of a strategic alliance network: exploring the case of stock photography. In: Boschma R, Martin R (eds) Handbook of evolutionary economic geography. Edward Elgar, Cheltenham, pp 298–315

Greif S, Schmiedl D (2002) Patentatlas 2002 Dynamik und Strukturen der Erfindungstätigkeit. Deutsches Patent- und Markenamt, München
Greif S, Schmiedl D, Niedermeyer G (2006) Patentatlas 2006. Regionaldaten der Erfindungstätigkeit. Deutsches Patent- und Markenamt, München
Gulati R (1999) Network location and learning: the influence of network resources and firm capabilities on alliance formation. Strateg Manag J 20:397–420
Handcock MS (2003) Statistical models for social networks: degeneracy and inference. In: Breiger RL, Carley KM, Pattison P (eds) Dynamic social network modeling and analysis: workshop summary and papers. National Academies Press, Washington, DC, pp 229–240
Hoekman J, Frenken K, Van Oort F (2009) The geography of collaborative knowledge production in Europe. Ann Reg Sci 43:721–738
Hoekman J, Frenken K, Tijssen RJW (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe. Res Policy 39:662–673
Hunter DR (2007) Curved exponential family models for social networks. Soc Netw 29 (2):216–230
Hunter DR, Goodreau SM, Handcock MS (2008) Goodness of fit for social network models. J Am Stat Assoc 103:248–258
Jaffe A (1989) Real effects of academic research. Am Econ Rev 79(5):957–970
LeSage J, Fischer MM, Scherngell T (2007) Knowledge spillovers across Europe. Evidence from a Poisson spatial interaction model with spatial effects. Pap Reg Sci 86:393–421
Maggioni MA, Nosvelli M, Uberti TE (2007) Space versus networks in the geography of innovation: a European analysis. Pap Reg Sci 86:471–493
Maskell P, Malmberg A (1999) Localized learning and industrial competitiveness. Camb J Econ 23:167–186
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Sociol 27:415–444
Metcalfe S (1995) The economic foundations of technology policy: equilibrium and evolutionary perspectives. In: Stoneman P (ed) Handbook of the economics of innovation and technological change. Basil Blackwell, Oxford, pp 409–512
Nelson R, Winter S (1982) An evolutionary theory of economic change. Belknap, Cambridge, MA
Nooteboom B (2000) Learning and innovation in organizations and economies. Oxford University Press, Oxford
Ponds R, Van Oort F, And Frenken K (2007) The geographical and institutional proximity of research collaboration. Pap Reg Sci 86:423–443
Powell WW, White DR, Koput KW, Owen-Smith J (2005) Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. Am J Sociol 110:1132–1206
Robins R, Snijders T, Wang P, Handcock M, Pattison P (2006) Recent developments in exponential random graph (p*) models for social networks. Soc Netw 29:192–215
Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. Soc Netw 29:173–191
Romer PM (1990) Endogenous technological change. J Polit Econ 98:71–102
Saul ZM, Filkov V (2007) Exploring biological network structure using exponential random graph models. Bioinformatics 23:2604–2611
Scherngell T, Barber MJ (2009) Spatial interaction modeling of cross-region R&D collaboration: empirical evidence from the 5th EU framework programme. Pap Reg Sci 88:531–546
Scherngell T, Barber MJ (2011) Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU framework programme. Ann Reg Sci 46:247–266
Snijders TAB (2002) Markov Chain Monte Carlo estimation of exponential random graph models. J Soc Struct 3. Web journal available from: http://www2.heinz.cmu.edu/project/INSNA/joss/index1.html
Snijders TAB, Pattison PE, Robins G, Handcock MS (2006) New specifications for exponential random graph models. Sociol Methodol 36:99–153

Snijders TAB, van de Bunt GG, Steglich CEG (2010) Introduction to stochastic actor-based models for network dynamics. Soc Netw 32:44–60

Sobrero M, Roberts EB (2001) The trade-off between efficiency and learning in interorganizational relationships for product development. Manag Sci 47:493–511

Ter Wal ALJ (2011) The dynamics of inventor networks in German biotechnology: geographical proximity versus triadic closure. Pap Evol Econ Geogr 11:02

Ter Wal ALJ, Boschma RA (2009) Applying social network analysis in economic geography: framing some key analytic issues. Ann Reg Sci 43:739–756

Uddin S, Hamra J, Hossain L (2012) Exploring communication networks to understand organizational crisis using exponential random graph models. Comput Math Organ Theor. doi:10.1007/s10588-011-9104-8

Van Duin MAJ, Gille KJ, Handcock MS (2009) A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Soc Netw 31:52–62

Varas MLL (2007) Essays in social space: applications to Chilean communities on inter-sector social linkages, social capital, and social justice. Dissertation at University of Illinois, Urbana-Champaign

Wang P, Pattison P, Robins G (2012) Exponential random graph model specifications for bipartite networks – a dependence hierarchy. Soc Netw. http://dx.doi.org/10.1016/j.socnet.2011.12.004

Wright D (2010) Repression and network science: tools in fight against terrorism. Dissertation at University of Michigan: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/77898/dewright_1.pdf?sequence=1