Papers in Evolutionary Economic Geography

11.07

Explaining the structure of inter-organizational networks using exponential random graph models: does proximity matter?

Tom Broekel and Matté Hartog

Utrecht University
Urban & Regional research centre Utrecht

Explaining the structure of inter-organizational networks using exponential random graph models: does proximity matter?

Tom Broekel and Matté Hartog

Section of Economic Geography, Faculty of Geosciences, Utrecht University, P.O. Box 80115, 3508 TC, Utrecht, The Netherlands

April 2011

Abstract

A key question raised in recent years is which factors determine the structure of inter-organizational networks. While the focus has primarily been on different forms of proximity between organizations, which are determinants at the dyad level, recently determinants at the node and structural level have been highlighted as well. To identify the relative importance of determinants at these three different levels for the structure of networks that are observable at only one point in time, we propose the use of exponential random graph models.

Their usefulness is exemplified by an analysis of the structure of the knowledge network in the Dutch aviation industry in 2008 for which we find determinants at all different levels to matter. Out of different forms of proximity, we find that once we control for determinants at the node and structural network level, only social proximity remains significant.

Keywords: exponential random graph models, inter-organizational network structure, network analysis, proximity, aviation industry, economic geography

JEL: R11, D85, L14, L62

1. Introduction

The analysis of inter-organizational knowledge networks has become increasingly popular in in recent years (e.g. Giuliani and Bell, 2005; Boschma and Ter Wal, 2007; Giuliani, 2007; Ter Wal and Boschma, 2009; Glückler, 2010; Ter Wal, 2011). As there is a growing awareness of the importance of inter-organizationals networks for innovation, many studies emphasize the necessity of gaining a deeper understanding of the determinants of tie creation to explain the structure of these networks (e.g. Glückler, 2007; Ter Wal and Boschma, 2009; Boschma and Frenken, 2010). In particular the proximity approach by Boschma (2005), which emphasizes the role different proximity types (cognitive, organizational, social, institutional, geographic) between organizations play for tie creation, has recently been the focus of empirical studies (see, e.g., Balland, 2011; Broekel and Boschma, 2011). As proximity is a dyadic concept, this approach naturally focuses on the dyad level. However, determinants at the node level (e.g. the size of an organization) and more recently determinants at the structural network level have been highlighted to impact the structure of networks as well (Glückler, 2007; Boschma and Frenken, 2010; Rivera et al., 2010). For example, the triadic closure hypothesis predicts that partners of partners are more likely to become partners as well in a network, which implies that new tie creation depends on the existing structure of a network (Ter Wal, 2011). A key methodological challenge in this respect is to separate the effects of determinants at the dyad level from effects at the node level and structural network level.

To estimate the relative importance of determinants at all three levels in explaining the structure of a network, they need to be included simultaneously in a model. While this is possible with existing empirical tools (see, e.g., Burk et al., 2007), those tools require longitudinal network data. However, longitudinal data for inter-organizational networks are unavailable in most cases (Baum et al., 2003; Ter Wal and Boschma, 2009). This is especially true for informal knowledge networks as those can usually only be observed by directly interviewing the employees of an organization, who only have a limited memory of the past. Hence, the question is how to determine which determinants explain the structure of an inter-organizational network that is observed at only one point in time.

In this article, we propose that exponential random graph models (ERG-models) may provide an answer to this question. These models are a new set of network analysis techniques that have been developed in the past few years in mathematical sociology (Snijders et al., 2006; Robins et al., 2007; Snijders et al., 2010a) and are increasingly used by scholars across other scientific disciplines to explain the structure of networks (e.g. in biosciences: Saul and

Filkov, 2007; in life sciences: Fowler et al., 2009; in political science: Cranmer and Desmarais, 2011). The reason for their growing popularity is that ERG-models only require cross-sectional network data but allow one to simultaneously estimate the importance of determinants at the node, dyad, and structural network level. For this reason, we believe that they are useful to analyze the determinants of the structure of inter-organizational networks as well.

To illustrate this, we confront ERG-models with one of the most frequently used methods so far to analyze inter-organizational network structure with only cross-sectional network data at hand: a Multiple Regression Quadratic Assignment Procedure, known as the MRQAP-model originally developed by Krackhardt (1987, 1988). We apply both models to explain the structure of the Dutch knowledge aviation network as observed in 2008. We show that the MRQAP-model has limited explanatory value as it can only account for determinants at the dyad level. Instead, with the ERG-model we find that determinants at the node level and structural network level also matter for the structure of the network. Furthermore, including determinants at the node and structural network level renders some determinants at the dyad level insignificant.

The article is structured as follows. The second section gives an overview of determinants at the node, dyad and structural network level that may determine the structure of inter-organizational networks.The models used to analyze the structure of those networks, ERG-models and MRQAP-models, are presented in the third section. In section four, an example of an inter-organizational network is introduced, namely the technological knowledge network of the Dutch aviation industry. In section five, we apply both models to estimate and evaluate the determinants of its structure. Finally, a conclusion is presented in the sixth section.

2. Determinants of the structure of inter-organizational networks

The question what determinants explain the structure of inter-organizational networks has been picked up only recently. The theoretical accounts on this question are framed within a theory of inter-organizational network evolution (Glückler, 2007; Ter Wal and Boschma 2009; Boschma and Frenken, 2010). The most elaborated theory on the formation of inter-organizational networks in space is given by Glückler (2007). He argues that:

"*A theory of network evolution, thus, looks at the changes that every new tie produces in the existing structure and, conversely, at the impact that the structure imposes on the* *formation of the next tie. Note that the unit of analysis is always dyadic tie formation, whereas the object of knowledge is network structure*" (Glückler, 2007, p. 622)

As he points out, central to understanding network formation is the interplay between network structure and tie creation between actors. Accordingly, there are several determinants that may impact tie creation and hence determine the structure of inter-organizational networks. Those are determinants at the (1) dyad (pair) level, (2) the node (organizational) level, and (3) structural network level. We elaborate below on the specific determinants at each of these levels.

First, the dyad level refers to the relation between two network actors. Ter Wal and Boschma (2009) and Boschma and Frenken (2010) argue that organizations with similar attributes are more likely to be tied. In sociology, this effect is known as the homophily effect (McPherson et al., 2001). In particular the proximity approach by Boschma (2005) has received a lot of attention lately. He argues that in the case of inter-organizational networks organizations are more likely to be tied when they are geographically, cognitively, socially, institutionally or organizationally proximate. Geographical proximity may matter because it facilitates frequent face-to-face contacts between organizations' personnel, and hence ties between organizations are more easily established and maintained when organizations are co-located. Accordingly, several studies find that geographical proximity between organizations has a positive impact on the chance of tie establishment (e.g. Maggioni et al., 2007; Ter Wal, 2011). Cognitive proximity between organizations may also matter because organizations only learn from each another when they have some similar knowledge assets (Cohen and Levinthal, 1990; Nooteboom, 2000). Accordingly, organizations prefer to tie to organizations that have a knowledge base that is similar to their own because only then are they able to understand one another.

Furthermore, at the dyad level, social proximity, institutional proximity and organizational proximity may matter for new tie creation. Social proximity refers to socially embedded ties between agents where a certain degree of trust exists between partners (Maskell and Malmberg, 1999; Storper and Venables, 2004). It may matter for new tie creation because ties between organizations are more easily established when managers of organizations trust one another. For example, Agrawal et al. (2006) find that firms are often tied when their employees have worked for a similar organization before and hence already know one another personally. In addition, institutional proximity may be important for tie creation. It refers to the extent to which organizations have related routines and incentive mechanisms (Metcalfe, 1994). If organizations have little institutional proximity they may have a lower chance of being tied. For example, such may be the case for firms and universities because of their different incentives regarding knowledge creation and exchange (keeping new knowledge secret versus publishing new knowledge). Finally, organizational proximity, which refers to the degree of strategic interdependence between organizations, may matter for new tie creation. For example, if organizations are members of the same corporate group (e.g. part of the same parent company), they may be more likely to be tied (Balland, 2010). Recent studies provide empirical evidence that all different proximity types set out above matter for the likelihood of organizationas to link (see, e.g., Balland, 2010; Broekel and Boschma, 2011).

Second, the node level refers to characteristics of network actors (nodes). Particularly, it may be that the size of an organization is relevant in this respect. Large organizations may be more likely to attract new ties because they occupy a more prominent position than small organizations within an industry. Accordingly, large organizations are likely to have more ties in the network. This is supported by Boschma and Ter Wal (2007) who in their study on the knowledge network of footwear producers in Barletta find the size of organizations to matter for their network position, with larger organizations being more central.

Third, the structural network level refers to characteristics of the entire network. Particularly for inter-organizational networks, it may be that new tie creation is affected by two determinants at the structural network level: triadic closure and multi-connectivity. Triadic closure implies that partners of partners are likely to become partners as well. As a result, the network consists of dense cliques of strongly interconnected actors. Such cliques are generally seen as a sign of social capital (Coleman, 1988) and hence enhance trust and willingness among actors to invest in mutual goals such as knowledge sharing. For this reason, it is likely that a tendency towards triadic closure between organizations and hence many triangles should be observed in an inter-organizational network (Ter Wal, 2011). Second, a tendency towards multi-connectivity might explain new tie creation between organizations. For instance, organizations may aim at connecting to other organizations in multiple ways because they want to decrease their dependency on a single link or channel. This implies that multiple independent paths are formed amongst two organizations, something that is known as the multi-connectivity hypothesis (Glückler, 2007). Empirical evidence for this hypothesis is found by Powell et al. (2005).

To estimate the relative importance of the determinants at all the three levels above to explain new tie creation and hence the structure of a network, they need to be simultaneously incorporated in a model. With longitudinal data, this can be accomplished with a stochastic actor-based network approach that models the change of a network from one state (point in time) to another as part of an iterative Markov chain process (the technical details are described in Holland and Leinhardt, 1977; Snijders, 1996, 2001; Snijders et al., 2010a). The models applying this approach are derived from mathematical sociology and have been applied very recently to inter-organizational networks as well (Balland 2011; Ter Wal, 2011).

A limitation of using these models, however, is that they require longitudinal network data, which is often not available for inter-organizational networks. The reason for this is that network analysis requires data of high quality that cover all actors of a particular population. If already a small number of links is unobserved (missing data), the structure of the observed network may be very different from the real-world network and hence may not be a valid observation (Ter Wal and Boschma, 2009a). Obtaining high quality data is a major problem when collecting network information at one point in time, but becomes even more of a problem when network data is collected at multiple periods. If data is unavailable at any point in time for some of the actors, the networks cannot be compared over time. This is especially true when the network concerns informal ties between organizations (e.g. social contacts, asking for advice and so on). Those informal inter-organizational networks can realistically only be observed by directly observing (interviewing) organizations' employees, who only have a limited memory of the past.

For this reason, in most instances data on certain inter-organizational networks can only be collected for one particular moment in time. This implies that stochastic actor-based models for longitudinal network data cannot be used. Instead, one has to apply a model that is ment for cross-sectional network data. In the next section we elaborate on two of those models. First, we introduce one of the most frequently used models so far: a Multiple Regression Quadratic Assignment Procedure Model (MRQAP-model). Subsequently, we compare it with and discuss the exponential random graph model (ERG-model).

3. Analyzing the structure of networks: multiple regression quadratic assignment procedure models (MRQAP-models) versus exponential random graph models (ERG-models)

In general, the basis for network analyses are relational variables. Relational variables describe the relationship between two nodes (i.e. organizations), i.e. the extent to which they are distinct, similar, or share certain characteristics. A particular value x_{ij} (i=1…n and j=1…n) indicates the relation between organization i and j with n being the number of observations. In dealing with this type of data to estimate the determinants of the structure of networks, a number of models for cross-sectional network data have been applied so far, particularly binary logit models (e.g. Kaufman et al., 2003; Autant-Bernard et al., 2007), gravity models (e.g. Hoekman et al., 2009) and multiple regression quadratic assignment procedure models (e.g. Cantner and Graf, 2006; Broekel and Boschma, 2011). While these models focus on determinants of network structure, they can only account for determinants at the dyad level.

We go deeper into this issue below by focusing on the multiple regression quadratic assignment procedure as an illustrative example. We compare it to the exponential random graph model, which we believe is a promising alternative since it can also take into account determinants at the node and structural level to explain the structure of a network that is observed at only one point in time.

3.1 Multiple regression quadratic assignment procedure regression model

While the multiple regression quadratic assignment procedure has been developed to deal with network data, it is based on a standard regression, which can be used to analyze the determinants of the structure of a network. However, relational data is characterized by variables that are not vectors but $n*n$ (adjacency) matrices. To apply standard regressions, the matrices need to be vectorized such that the columns are stringed together to form one vector with n^2 elements. Accordingly, the first elements (first row in the adjacency matrix) are the relations of the first organization to all others, next are those of the second organization, and so on.

In the regression underlying the MRQAP-model, the dependent variable is regressed with a standard logit model on the independent variables. For non-valued network data, the logit model is more appropirate than an ordinary least square approach because the dependent variable is a 0/1 variable with 1 indicating the existence of a link between two organizations, and 0 indicating its absence. However, network data are characterized by frequent row/column/block autocorrelation. For this reason, standard tools of inference are invalid (Krackhardt 1987). A solution to this is Krackhardt (1987) and Krackhardt (1988) propse the so-called Quadratic Assignment Procedure. Here, the estimated model statistics are compared to the distribution of such statistics resulting from large numbers of simultaneous row/column permutation of the considered variables (before the vectorization). In most applications the multicolinearity robust "semi-partialling plus" procedure by Dekker is applied (Dekker et. al. 2003). Accordingly the MRQAP-model is a logit model that is able to incorporate network variables and that can deal with the inherent interdependencies of network data. For these reasons they are frequently applied to explain the structure of inter-organizational networks (e.g. Cantner and Graf, 2006; Broekel and Boschma, 2011).

However, the main shortcoming of the MRQAP-model is that it only allows one to identify determinants at the dyad level to explain the structure of a network. This implies that node level determinants cannot be incorporated into a MRQAP-model. Instead, one can try to translate node attributes into dyadic attributes. For instance, to test whether the size of organizations has a positive impact on the chance of being tied to other organizations, with a MRQAP-model, it can only be tested if the probability for two large organizations to be linked is higher than for two small organizations. This is distinct from an argument at the node-level, which could for instance imply that large organizations as such are more likely to have more links. The latter cannot be tested with a MRQAP-model .

In addition, the MRQAP-model also cannot incorporate determinants at the structural network level. For instance, if one wishes to assess whether the existence of a tie between actors Pi and Pj is dependent on whether Pj has ties with actors Pk and Pl, a configuration needs to be included that involves more than two actors (Snijders and Bosker, 1999). Such a configuration, of which triadic closure is an example (partners of partners are more likely to become partners as well) is called a higher order network configuration and cannot be included in a MRQAP-model. Also, translating these determinants to the dyadic level is not possible. It would mean that the newly created dyadic variable is based on the dependent variable, which raises serious concerns regarding the independence assumption underlying the model. Accordingly, the unability to incorporate determinants at the structural network level is a major shortcoming of the MRQAP-model because tendencies towards triadic closure and multi-connectivity are frequently argued to be relevant to explain the structure of inter-organizational networks.

3.2 Exponential random graph models

In response to the shortcomings of the MRQAP-model and other conventional regression approaches used to explain network the structure of networks, recent years have seen the development of exponential random graph models, known as ERG-models (Snijders et al., 2006; Robins et al., 2007; Snijders et al., 2010a). These models allow one to include determinants at the node and structural network level as well. ERG-models are stochastic models that approach new tie creation as a time-continuous process. Specifically, an observed network at one point in time is regarded as one particular realization out of a set of multiple hypothetical networks with similar characteristics. The aim of an ERG-model is to identify the determinants that maximize the probabilty of the emergence of a network with roughly the same characteristics as the structure of the observed network.

The roots of the ERG-models date back to the early 1960s with the development of Bernoulli random graphs (Erdös and Renyi, 1959). These graphs were used to estimate configurations of individual ties between actors. They assume that ties between actors are independent of one another. As this proved to be an inaccurate assumption, in the early 1980s dyadic models, also called p1 models, were developed. These models assumed independence of dyads, which are relations between actors, rather than independence of actors themselves (Holland and Leinhardt, 1981). However, this dependence assumption was soon found to be unrealistic in many circumstances (Newman, 2003). Therefore, Frank and Strauss (1986) introduced the so-called Markov dependence, which assumes that a possible tie between two actors, for example between *a* and *b,* is contingent on other possible ties between other actors involving *a* and *b.* Accordingly, two ties are said to be conditionally dependent in Markov random graph models (Robins et al., 2007). These models serve as the basic building block for ERG-models that been developed until now (Snijders et al., 2010a). Acordingly, in an ERG-model the network is seen as being generated stochastically, whereby relational ties are created in ways that are shaped by the presence or absence other ties and actor attributes. Hence, a network is approached as being a self-organizing system of relational ties, whereby if one tie emerges or disappears in the simulated construction process, other neighboring ties may emerge or disappear as well (Robins et al., 2006).

The general form of ERG-models is defined as follows (Robins et al., 2007):

$$\Pr(Y = y) = \left(\frac{1}{\kappa}\right)\exp\left\{\sum_{A} \eta A \; gA(y)\right\} \qquad \text{(eq. 1)}$$

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). *ή a* is the parameter corresponding to network configuration *A* and *ga (y)* represents the network statistic. Network configurations (*A)* can be determinants at the node level, dyad level and structural network level. Their corresponding network statistic obtain a value of 1 if a configuration is observed in network y and 0 if not.

The aim of solving the equation is to find parameters for the configurations that maximize the probability that the simulated random graphs by the ERG-model are identical to the observed network. This is done by a combination of estimation and simulation methods, which usually involve a Markov Chain Monte Carlo Maximum Likelihood Estimation procedure (Snijders, 2002; Van Duin et al., 2009). This procedure can be performed by modern computer software such as ‘Simulation Investigation for Empirical Network Analysis’, known as SIENA, or pnet and statnet (see Robins et al., 2007; Carter et al., 2008; Snijders et al., 2010b).

Because estimation and simulation techniques are used to establish parameter values in an ERG-model, it has to be tested whether the estimated parameters provide a good fit of the observed network. The first step is to assess whether the model is not degenerate. Degeneracy appears when an ERG-model is specified (i.e. the variables and / or the starting parameters of the simulation) that is unlikely to produce the observed network (Handcock 2003a, 2003b; Hunter et al., 2008) .In this case, either the Maximum-Likelihood-Estimates do not exist and the model does not converge, or the estimates exist but they do not provide a good fit to the data. This implies that in the simulated networks a significant number of nodes are either completely linked to each other or totally unconnected (Austad and Friel, 2010). To determine whether the model is degenerate, one has to check the Degeneracy Value, which as a rule of thumb should be below 1 (Goodreau et al., 2009). Moreover,,the traces of the parameter values over the course of iteration should be relative stable and vary more or less around the mean value (see for a discussion, Goodreau et al., 2008).

Another step in checking whether the parameters predict the observed network well is to assess the model’s goodness of fit by comparing the structure of the simulated networks to the structure of the observed network. As Hunter et al. (2008) argue, this can be done by comparing 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 these of the observed network, the more accurate and hence reliable the parameters of the ERG-model are.

Hence, whereas the use of simulation and estimation techniques implies that an ERG-model is not as straightforward as conventional models, it provides one major advantage over them: it allows one to include determinants at the node and structural network level as well as determinations of network structure. With respect to the MRQAP-model, this is a key advantage as it may well be that initially determinants at the dyad level are identified as being relevent whereas it are in fact effects of determinants at the node level or structural network level. For this reason, ERG-models are increasingly used outside mathematical sociology, for example in biosciences to explain the structure of cell networks (Saul and Filkov, 2007), in life sciences to model genetic variation in human social networks (see Fowler et al., 2009) or in political science to analyze the structure of networks of political international conflicts (Cranmer and Desmarais, 2011). However, so far they have not been used to explain the structure of inter-organizational networks. To illustrate their usefulness in this respect as well, in the next section we will apply an ERG-model to an inter-organizational network.

4. Empirical application: determinants of the structure of the knowledge network in the Dutch aviation industry

This section applies an ERG-model to explain the structure of the Dutch aviation knowledge network as observed in 2008 and compares its results to those of a MRQAP-model. As pointed out earlier, in particular the importance of proximity at the dyad level has drawn a lot of attention recently (see, Boschma, 2005, Boschma and Frenken, 2010). We follow this literature and analyze the impact of different types of proximities on the structure of the network. In a common fashion, we conceptualize the proximities as dyad-level determinants (i.e. describing the relation between two organizations) and estimate their impact with a MRQAP-model. Subsequently, we confront these results with a ERG-model where we also include determinants at the node level and structural network level, both of which may provide additional or alternative explanations for the structure of the network.

4.1 Data

The data we use concern network data on technological knowledge sharing between Dutch aviation organizations. These data have been gathered by means of semi-structured interviews held in 2008 and 2009 with members of the Netherlands Aerospace Group (NAG), whose members account for about 95% of total turnover generated by Dutch aviation organizations (NAG, 2008). The interviews focused on profit and non-profit organizations that are activie in aviation related manufacturing and/or engineering since only for those organizations is innovation and the exchange of technological knowledge likely to be of utmost importance. Of the 64 organizations falling into this category, 59 were willing to participate, hence the response rate is 93% and thus we have network data of almost all actors in the population, which is a necessary condition in order to carry out a sound network analysis (Wasserman and Faust, 1994). By means of these relational data, we can construct and analyze the structure of the complete network of technological knowledge sharing of Dutch aviation organizations.

Figure 1 shows the technological knowledge network of the Dutch aviation industry, of which the structure will be analyzed in the remainder of the paper. The organizations were asked to indicate with which other aviation organizations they share technological knowledge, and the links in the network represent the knowledge ties that follow from this. As we assume that knowledge exchange is a mutual process between two organizations, all links are undirected (i.e. going both ways). Some descriptives of the network are presented in Table 1.

<INSERT FIGURE 1 ABOUT HERE>

<INSERT TABLE 1 ABOUT HERE>

4.2 Variables

The variables that we use to explain the structure of the network relate to the node, dyad, and structural network level. As set out before, variables at the node level and structural network level cannot be incorporated in a MRQAP-model and hence are only included in the ERG-model.

4.2.1 Dyad level: social proximity, institutional proximity, geographical proximity, cognitive proximity

At the dyad level we focus on four of the different types of proximities as set out by Boschma (2005) which may matter for the structure of inter-organizational networks: social, institutional, geographical and cognitive proximity. We do not measure organizational proximity because we lack data on this dimension. Below we describe how each of them have been operationalized. The proximity variables are included in both the MRQAP-model and the ERG-model.

Social proximity (SOCPROX): to investigate whether social proximity between organizations matters for tie creation, a dyad-level variable SOCPROX is created. It amounts to a value of 1 if members of the top management of two organizations are former employees of Fokker B.V., and a value of 0 if they are not. The motivation for this is that Fokker B.V. has been the dominant firm in the Dutch aviation industry until 1996 with (at its peak) more than 13,000 employees. Its bancruptcy in 1996 lead to massive job cuts and a reformation of the Dutch aviation industry. However, many of its old employees got employed in new firms or started their own businesses. Broekel and Boschma (2011) show organizations are more likely to be linked if its employees have a shared past in Fokker .They suspect, "old boys" networks might still be in place, which give exclusive knowledge sharing opportunities. Accordingly, it is expected that a shared past in Fokker leads to a higher chance of being tied.

Institutional proximity (INSTPROX): we approximate institutional proximity by differentiating between profit organizations (firms) and non-profit organizations (universities, research institutes, associations, and trade organizations) (see Balland, 2011). With few exceptions, these non-profit organizations are highly connected in the technological knowledge network, see Figure 1. Moreover, they are also frequently named by firms as important technological knowledge sources. A dyadic dichotomous variable (INSTPROX) is constructed that has a value of 1 when both organizations are profit organizations (firms) or if both are non-profit organizations, and a value of 0 when otherwise. It is expected that institutional proximity between organizations increases the chance of being tied.

Geographical proximity (GEOGPROX): to assess the effect of geographical proximity, we first calculate the geographical distance in kilometers between two organizations, which results in a continuous variable. While other studies use travel time (e.g. Ejermo and Karlsson, 2006), the spatial scale of the network of Dutch aviation industry is rather small, which is why the use of travel distances is unlikely to change the results. Then, for computational reasons, the variable is transformed into a dyadic categorical variable. This implies that all distances below fifty kilometers obtain a value of 10, distances between 50 and 100 kilometers obtain a value of 50, and larger distances obtain a value of 100. Accordingly, the variable GEOPROX differentiates between local, regional, and national "distances".[1] It is expected that the more geographical proximity between organizations, the higher the chance of being tied.

Cognitive proximity (COGPROX): we refer to cognitive proximity as to whether two organizations' knowledge bases are technologically similar or not. For its measurement, we rely on the technology classes that are assigned by the Netherlands Aerospace Group (NAG). The NAG defines 15 technologies of which 13 are relevant for the organizations considered in this study. The technological fields and the according number of organizations are listed in Table 2. In case the interviewed organization is not a member of the NAG, the profile was created on the basis of the organization's webpage. The variable COGPROX is defined

[1] We also tested a range of alternative definitions of local, regional, and national distances but the results proved to be fairly robust.

dichotomously, with a value of 1 if both organizations are active in at least one identical technology, and 0 otherwise. It is expected that being cognitively proximate leads to a higher chance of being tied.

The dyadic covariates are included in the ERG-model by:

$$\sum_{i,j} x_{ij} \mathrm{I}\{\upsilon_i = \upsilon_j\} \qquad \text{(eq. 3)}$$

where the indicator function $\mathrm{I}\{\upsilon_i = \upsilon_j\}$ is 1 if the condition $\{\upsilon_i = \upsilon_j\}$ is satisfied, and 0 if this is not the case. Thus, organizations with a similar attribute gain a score of one.

<INSERT TABLE 2 ABOUT HERE>

4.2.2 Node level: Organization size

As set out earlier, at the node level the size of organizations is likely to matter for the structure of a network. Large organizations may be more likely to attract new ties because they occupy a more prominent position in the industry than small organizations, and hence large organizations are likely to have more ties. We define the variable SIZE by the organizations' number of employees.

In the ERG-model, individual covariates are included as follows:

$$\sum_i x_i + \nu_i \qquad \text{(eq. 2)}$$

where i represents the actor in the network ν and x represents the attribute in the network. If a positive parameter is observed for this effect, it means that actor i that scores high on x has a higher chance of being tied to other actor. Accordingly, a positive correlation exists between x (the actor attribute) and the number of ties.

Because the MRQAP-model cannot incorporate variables at the node level, SIZE is only included in the ERG-model. Instead, for the MRQAP-model SIZE is transformed into a dyadic variable called ORGASIZE, which is estimated as the sum of two organizations' number of employees.

4.2.3 Structural network level: Triadic closure and multi-connectivity

We define two variables that relate to the structural network level, namely triadic closure and multi-connectivity, which, as set out earlier, are both likely to matter for the structure of inter-organizational networks. Triadic closure refers to the tendency of organizations' partners becoming partners as well. It may be beneficial because it enhances trust and willingness among actors to invest in mutual goals such as knowledge sharing. Multi-connectivity refers to the tendency of organizations to establish multiple paths amongst each other. It can yield positive effects as multiple paths to other organizations decreases the dependency on a single link or channel. Both of them are higher order network configurations because they involve more than two actors in the network, and hence they can only be included in the ERG-model.

Triadic closure (TRIADCLOS) is captured by the geometrically weighted edgewise shared partner statistic, known as the GWESP-statistic (Snijders et al., 2006; Hunter et al., 2008). It is formulated as follows:

$$v(y;\tau) = e^{\tau}\sum_{i=1}^{n-2}\left\{1-(1-e^{-\tau})^{i}\right\}\mathrm{EP}_i(y) \qquad \text{(eq. 4)}$$

where, as E denotes the number of edges (existing ties), $\mathrm{EP}_i(y)$ is the edgewise shared partner statistics that indicates the number of edges that share edges with a certain number of other nodes. Accordingly, it represents the number of unordered pairs *(j,k)* where $x_{jk} = 1$ and j and k having exactly *i* neighbours in common in network *y*. In a nutshell, it measures the number of triangles in the network while taking into account the number of edges that are involved in multiple triangles (multimodality) and hence may have the same neighbours in multiple triangles (the technicalities of the statistic are explained in greater detail in Hunter et al., 2008). If for this statistic a positive parameter is found, it means that there is a tendency towards triadic closure in tie creation in the network.

Multi-connectivity (MULTICON) is captured by the alternating independent two-path statistic as introduced by Snijders et al. (2006). It is formulated as follows:

$$c\sum_{i<j}\left\{1-\left(1-\frac{1}{\lambda}\right)^{L_{2ij}}\right\} \qquad \text{(eq. 5)}$$

where for some value *c*, L_{2ij} is the number of two-paths from *i* to *j* (through another node *h*) that is expressed as follows:

$$L_{2ij} = \sum_h x_{ih}x_{hj} \qquad \text{(eq. 6)}$$

Hence, the alternating independent two-path statistic (MULTICON) measures how many partners every pair of nodes shares. Because it does so for pairs of nodes that are not linked themselves, it is a lower order parameter to the TRIADCLOS-statistic. If the value on this statistic is positive, it means that there is a tendency towards creating multiple paths among nodes in the network. The result implies that nodes are better able to reach one another through a greater variety of other nodes, and hence a structural tendency towards multi-connectivity is visible in tie creation in the network.

4.2.4 Control variables

Because the organizations in our sample are heterogeneous, we use a number of control variables that should capture this heterogeneity. First, as private and profit organizations may show distinct cooperation behavior, a node-level variable is created that represents non-profit organizations (NON-PROFIT). Similarly, we define a node-level variable that represent firms with a background in Fokker (FOKKER).

Second, we control for the fact that some organizations are more focused on the aviation industry than others. Because of their shared focus, these organizations may be more likely to be linked in the knowledge network of the industry. This effect is taken into account by the dyadic variable AVIASIM, which indicates whether two organizations are particularly active in the aviation industry or not. For firms, this implies that the share of their turnover attributed to aviation is above the average of all firms in the sample. In case of other organizations, we define them to be "dedicated" to aviation if their focus is mainly on this sector, for which we primarily rely on information derived from the organizations' websites. With this information at hand, the dichotomous variable AVIASIM is created that has a value of 1 if two organizations are dedicated towards aviation and a value of 0 otherwise.

Third, organizations may differ with respect to their openness towards external knowledge. Two organizations that perceive external knowledge as being highly relevant may be more likely to be linked than two organizations that rely more on internal knowledge. Therefore, the variable EXTERNALSIM defined, which has a value of 1 if the relative importance that organizations i and j attribute to external knowledge is above average, and a value of 0 if not. This information is collected by the following question we asked during the interviews: "Please indicate in terms of percentage the relative importance of: a) knowledge acquired inside the company; b) knowledge acquired outside the company (adding up to 100%)".

5. Results

We first show the results of the MRQAP model that only includes the seven variables at the dyadic level. Then, we present the results of the ERG-model. We show that determinants at the node level and structural network level also explain part of the structure of the knowledge network in the Dutch aviation industry.

5.1 Multiple regression quadratic assignment procedure model (MRQAP-model)

The results of the MRQAP-model are shown in Table 3. Given that the MRQAP-model is based on a standard logistic regression, the interpretation of the results is straightforward. Of the four types of proximities, only two are found to increase the probability of tie creation. Institutional proximity as well as geographic proximity have a positive impact on the chance of organizations exchanging knowledge. The large coefficient of INSTPROX reflects the high importance of this variable for the model, implying that it is particularly institutional proximity that determines tie creation. In other words, organizations with the same institutional background (non-profit vs. profit-oriented) are more likely to link to the same kind than connecting to organizations that operate in a different institutional set-up. This particularly regards non-profit organizations that tend to be strongly linked with each other (69 links). In contrast, firms rarely connect with other firms (18 links). Cognitive proximity is insignificant. The same applies to social proximity. In addition, ORGASIZE has a significant positive coefficient, which indicates that larger organizations tend to be linked more frequently with each other. In sum, the results appear to be in line with our expectations. Note however, that we cannot include node or structural network level determinants in the MRQAP-model. In the next part we present the results of the ERG-model to see whether the consideration of these influences the results.

<INSERT TABLE 3 ABOUT HERE>

5.2 Exponential random graph model (ERG-model)

Table 4 shows the results of an ERG-model with the same variables as the MRQAP-model (the dyad-level determinants only). The results are very different from the MRQAP-model in that all the coefficients are negative. However, although the model is not degenerate, the goodness of fit statistics of the ERG-model (Figure 4 of the Appendix) show that these coefficients are unreliable as the fitted models' characteristics (boxplots and dashed lines) depart strongly from the actual distribution in the original network (solid line). The predicted degree distributions, edgewise shared partners distributions and geodesic distributions do not match the distributions in the observed network, which means that the model fits poorly. Moreover, the parameter traces (Figure 2 in the Appendix) reveal that they vary strongly (in particular the parameter of GEOPROX). Accordingly, including only the seven dyadic variables in the ERG-model yields unreliable coefficients. This indicates that other variables might be better in explaining the structure of the Dutch aviation knowledge network. Hence, whereas from the MRQAP-model we can derive statistical associations between dyad level determinants and the chance of being tied, we learn from the ERG-model that when the same variables are also simulated to explain the structure of the network, they turn out to be inaccurate.

<INSERT TABLE 4 ABOUT HERE>

Table 5 shows the results of the an ERG-model that also includes determinants at the node level and structural network level. The reported coefficients represent the model that is (1) not degenerate (2) shows stable and quite narrow parameter traces, and (3) provides the best goodness-of-fit statistics (matching degree, edgewise shared partners and geodesic distributions) of all (theoretically relevant) variable combinations possible with the available data. As such, some determinants that had been included initially are not reported in the final model in Table 5. Particularly, the results of the multi-connectivity variable (MULTICON) are not reported because its inclusion always causes degenerate models. The same applies to three of the proximity variables, namely geographical proximity (GEOGPROX), cognitive proximity (COGPROX) and institutional proximity (INSTPROX) and to two of the control variables, namely the node-level variable FOKKER and the dyadic variable EXTERNALSIM. This means that all of these variables are insignificant drivers of the structure of the Dutch aviation network and their inclusion in the model yields degenerate models in most instances.

<INSERT TABLE 5 ABOUT HERE>

Hence, with the exclusion of the variables above, the model as reported in Table 5 has a degeneracy value that is well below 1 and the model fits the observed network well. To increase the fit of the model, two structural network configurations have been added to it, namely EDGES and ISOLATES. EDGES adds one statistic to the network, which equals the number of links in the network, and ISOLATES accounts for the share of isolates in the original network. As for the whole model's fit, the traces of the parameter values over the course of the iteration process show the required pattern. As shown in Figure 5 and Figure 6, none of them follows a trend away from the mean, and show more or less normal distributions. In other words, the values vary more or less stochastically around the mean, which implies a satisfactory convergence of the model (see for a discussion, Goodreau et al., 2008). Furthermore, the goodness of fit plots (Figure 7) reveal a much better fit than that of the model resembling the MRQAP-model. This is also backed by the lower AIC and BIC values (Akaike and Schwartz criteria), which are two commonly used goodness of fit measures.

The results of the ERG-model in Table 5 are clearly different from the MRQAP-model and can be interpreted as follows. First, non-profit organizations are more likely to be tied to other organizations than profit organizations as the node-level variable NON-PROFIT is positive and significant. This meets our expectations and is in line with the visual inspection of the network in Figure 1 which shows that non-profit organizations generally have more links.

At the dyad level, AVIASIM and SOCPROX turn out to be significant and positive. Hence, the degree of engagement of organizations in this industry matters for the structure of the network (AVIASIM): two organizations that are above average active in the aviation industry are much more likely to share technological knowledge than two organizations that are less focused this industry. Also, social proximity matters for the structure of the Dutch aviation network. As such, if members of the top management of two organizations are former employees of Fokker, these organizations are more likely to be linked to exchange knowledge.

At the structural network level, the TRIADCLOS-variable is positive and significant, which implies that triadic closure is a driver of the structure of the network. This confirms the visual inspection of the network in Figure 1 as it shows the existence of a relatively large number of triangles. Hence, partners of partners are more likely to become partners as well. This structural network determinant of the structure of the network could not be accounted for with the MRQAP-model.

Concluding, our empirical example underlines the two major advantages of the ERG-model. First, we find that some dyadic determinants that are identified as being important in the MRQAP-model turn out to be unimportant when included into the simulation process of the ERG-model. Second, we show that determinants at the node-level (in this case: being a non-profit organization) and determinants at the structural network level (in this case: triadic closure) also matter for the structure of the network, something which cannot be accounted for with the MRQAP-model.

6. Conclusion

The aim of this article is to introduce exponential random graph models (ERG-models) as promising tools to explain the structure of inter-organizational networks that are observable at only one point in time. Their main advantage is that they are able to include explanatory determinants of tie creation at the node, dyad, and structural network level to explain the structure of a network. For this reason, ERG-models have grown increasingly popular across scientific disciplines in recent years but they have not been used so far to analyze the relevance of different forms of proximity for the structure of inter-organizational networks. The concept of proximity is dyadic in nature and therefore it puts the dyadic level in the focus. We believe however that ERG-models are particularly useful for evaluating the relative importance of different types of proximity as they allow not only for comparisons with other dyadic factors but also with determinants at the node and structural network level.

To illustrate this, we apply an ERG-model to explain the structure of the Dutch aviation knowledge network.. As many other networks, this is an example of a network of which it is almost unfeasible to collect network data at more than one point in time. Accordingly, to explain its structure, the empirical assessment has to deal with cross-sectional network data. In doing so, we compare the ERG-model to the most conventional model for cross-sectional network data used so far: a multiple regression coupled with the quadratic assignment procedure for statistical inference (MRQAP-model). We show that the MRQAP-model explains only part of the structure of the Dutch aviation knowledge network as it can only handle determinants at the dyadic level. When applying an ERG-model, it is found that determinants at the node level (being a non-profit organization) and structural network level (triadic closure) are relevant as well. Moreover, it is shown that controlling for these renders some determinants at the dyad level insignificant.

This is however not to say that ERG-models are without any drawbacks. A major issue of applying an ERG-model so far is the cumbersom process of finding the model that best fits the observed network if the theoretical framework allows one for some variance in variable selection and specification. While the goodness of fit statistics provide some indications about the choice of variables, a lot of trial and error in recombining variables is involved in practical application. Nevertheless, because of their ability to incorporate determinants at all three levels (node, dyad, structural network) of tie creation to model the structure of a network that is observed at only one point in time, we believe that ERG-models have promising potential for future studies on the structure of inter-organizational networks.

References

Austad, H. and Friel, N. (2010), Deterministic Bayesian inference for the p* model. AISTATS Journal of Machine Learning Research 9, pp 41, 48.

Autant-Bernard C., Billand P., Frachisse D. and N. Massard (2007), Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies. Papers in Regional Science 86, pp. 495-519.

Balland, P.A. (2011), Proximity and the Evolution of Collaboration Networks: Evidence from Research and Development Projects within the Global Navigation Satellite System (GNSS) Industry. Regional Studies, forthcoming. DOI: 10.1080/00343404.2010.529121.

Baum, J.A.C., Shipilov, A.V. and T.J. Rowley (2003), Where do small worlds come from? Industrial and Corporate Change 12, pp. 697-725.

Boschma, R.A. (2005), Proximity and Innovation: A Critical Assesment. Regional Studies 39, pp. 61-74.

Boschma, R.A. and Ter Wal, A.L.J. (2007), Knowledge networks and innovative performance in an industrial district: The case of a footwear district in the south of Italy. Industry &Innovation 14 (2), pp. 177-199.

Boschma, R.A and Frenken, K. (2010) The spatial evolution of innovation networks. A proximity perspective. in R.A. Boschma, R. Martin (eds.), Handbook on Evolutionary Economic Geography. Cheltenham: Edward Elgar, pp. 120–135.

Broekel, T. and Boschma, R. (2011), Knowledge networks in the Dutch aviation industry: the proximity paradox. Journal of Economic Geography, forthcoming.

Breschi, S. and Lissoni, F. (2001), Knowledge Spillovers and Local Innovation Systems: A Critical Survey. Industrial and Corporate Change 10 (4), pp. 975-1005.

Breschi, S. and Lissoni, F. (2003), Mobility and Social Networks: Localised Knowledge Spillovers Revisited. CESPRI Working Paper 142.

Burk, W.J., Steglich, C.E.G. and Snijders, T.A.B. (2007), Beyond Dyadic Interdependence.: actor-oriented models for co-evolving soial networks and individual behaviors.

International Journal of Behavioral Development 31 (4), pp. 397-404.

BVDEP (2010), Business database. Available on the World Wide Web: <intportal.bvdep.com>

Cantner, U. and Graf, H. (2006), The Network of Innovators in Jena: An Application of Social Network Analysis. Research Policy 35(4), pp. 463–480.

Carter, N. Hadlock, C. and D. Haughton (2008), Generating random graphs from a given distribution. Computational Statistics & Data Analysis 52 (8), pp. 3928-3938.

Cohen, W.M. and Levinthal, D.A. (1990), Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly 35, pp. 128-152.

Coleman, J.S. (1988), Social capital in the creation of human capital. American Journal of Sociology 94 (supplement), pp. 95-120.

Cranmer, S.J. and Desmarais, B.A. (2011), Inferential Network Analysis with Exponential Random Graph Models. Political Analysis 19 (1), pp. 66-86.

Dekker, D,. Krackhardt, D. and Snijders, T.A.B. (2003), Mulicollinearity Robust QAP for Multiple Regression. CASOS Working Paper, Carnegie Mellon University.

Ejermo, O. (2003), Patent Diversity as a Predictor of Regional Innovativeness in Sweden. CESPRI Working Paper 140.

Ejermo, O. and Karlsson, C. (2006), Interregional Inventor Networks as Studied by Patent Coinventorships. Research Policy 35, pp. 412–430.

Erdös, P. and Renyi, A. (1959), On random graphs. I. Publicationes Mathematicae Debrecen 6, pp. 290-297.

Fowler, J.H., Dawes, C.T. and Chistakis, N.A. (2009), Model of Genetic Variation in Human Social Networks. Proceedings of the National Academy of Sciences 206 (6), pp. 1720-1724.

Frank, O. and Strauss, D. (1986), Markov graphs. Journal of the American Statistical Association 81, pp. 832-842.

Giuliani, E. and Bell, M. (2005), The micro-determinants of meso-level learning andinnovation: evidence from a Chilean wine cluster. Research Policy 34 (1), pp. 47-68.

Giuliani, E. (2007), The selective nature of knowledge networks in clusters: evidence from the wine industry. Journal of Economic Geography 7, pp. 139-168.

Glückler, J. (2007), Economic geography and the evolution of networks. Journal of Economic Geography 7, pp. 619-634.

Gluckler, J. (2010), The evolution of a strategic alliance network – Exploring the case of German stock photography. In: R. Boschma and R. Martin (eds.), Handbook of Evolutionary Economic Geography. Cheltenham: Edward Elgar, pp. 298-315.

Goodreau, S. M., Kitts, J. A., and 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), pp. 103–125.

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., and Morris, M. (2008), A statnet tutorial. Journal of Statistical Software 24 (9), pp. 1-26.

Grabher, G. (2006), Trading routes, bypasses, and risky intersections: mapping the travels of'networks' between economic sociology and economic geography. Progress in Human Geography 30 (2), pp. 163-189.

Handcock, M.S. (2003a), Assessing Degeneracy in Statistical Models of Social Networks. Working Paper No. 39. University of Washington: Center for Statistics and the Social Sciences.

Handcock. M.S. (2003b), Statistical Models for Social Networks: Degeneracy and Inference. In: R.L. Breiger, K.M. Carley and P. Pattison (eds.), in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. Washington, DC: National Academies Press, pp. 229–40

Hoekman, J., Frenken, K. and Van Oort, F. (2009), The geography of collaborative knowledge production in Europe. The Annals of Regional Science 43 (3), pp. 721-738.

Holland, P. W. and Leinhardt, S. (1977), A dynamic model for social networks. Journal of Mathematical Sociology 5, pp. 5-20.

Holland, P.W. & Leinhardt, S. (1981), An exponential family of probability distributions for directed graphs (with discussion). Journal of the American Statistical Association 76, pp. 33-65.

Hunter, D.R., Goodreau, S.M., Handcock, M.S. (2008), Goodness of Fit for Social Network Models. Journal of the American Statistical Association 103, pp. 248-258.

Kaufman, A., Lehner, P. and Tödtling, F. (2003), Effects of the Internet on the spatial structure of innovation networks. Information Economics and Policy 15, pp. 402-424.

Krackhardt, D. (1987), QAP Partialling as a Test of Spuriousness. Social Networks 9, pp. 171-186.

Krackhardt, D. (1988), Predicting With Networks: Nonparametric Multiple Regression Analyses of Dyadic Data. Social Networks 10, pp. 359-382.

Kriechel, B. (2003), Heterogeneity Among Displaced Workers, Maastricht: ROA.

Ligterink, J. E. (2001), Corporate Financial Risk Management. Amsterdam: Thesis publishers.

Maggioni, M.A., Nosvelli, M. and T.E. Uberti. (2007), Space versus networks in the geography of innovation: a European analysis. Papers in Regional Studies 86 3), pp. 471-493.

Maskell, P. and Malmberg, A. (1999), Localized Learning and Industrial Competitiveness. Cambridge Journal of Economics 23, pp. 167–186.

McPherson, M., Smith-Lovin, L. and J. M. Cook (2001), Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology 27, pp. 415-444.

Metcalfe, S. (1994), The Economic Foundations of Technology Policy: Equilibrium and Evolutionary Perspectives. In: Dodgson and M. Rothwell R (Eds.), The Handbook of Industrial Innovation. Cheltenham: Edward Elgar, pp. 409-512.

NAG (2008), Your Guide to the Netherlands Aviation Industry, Netherlands Aviation Group.

Newman, M.E.J. (2003), The Structure and Function of Complex Networks. SIAM Review 45 (2), pp. 167-256.

Nooteboom, B. (2000), Learning and innovation in organizations and economies. Oxford: Oxford University Press.

Powell, W.W., Koput, K.W., and L. Smith-Doerr (1996), Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. Administrative Science Quarterly 41, pp. 116.145.

Powell, W. W., White, D. R., Koput, K. W., and J. Owen-Smith (2005), Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. American Journal of Sociology, 110(4), pp. 1132-1206.

Reuters (1996), International Business; Core Units Are Closed by Fokker. New York: The New York Times.

Rivera, M., Soderstrom, S., and B. Uzzi (2010), Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. Annual Review of Sociology 36, pp. 91–115.

Robins, R., Snijders, T., Wang, P., Handcock, and P. Pattison (2006), Recent developments in exponential random graph (p*) models for social networks. Social networks 29 (2), pp. 192-215.

Robins, G., Pattison P., Kalish Y. and Lusher D. (2007), An introduction to exponential random graph (p*) models for social networks. Social Networks 29 (2), pp 173-191.

Saul, Z.M. and Filkov, V (2007), Exploring biological network structure using exponential random graph models. Bioinformatics 23 (19), pp. 2604- 2611.

Snijders, T.A.B. (1996), Stochastic Actor-oriented Models for Network Change. Journal of Mathematical Sociology 21, pp. 149-172.

Snijders, T.A.B. and Bosker, R.J. (1999), Multilevel Analysis. An introduction to basic and advanced multilevel modeling. London: SAGE Publications.

Snijders, T.A.B. (2001), The statistical evolution of social network dynamics. Sociological Methodology 31 (1), pp. 361-395.

Snijders, T.A.B. (2002), Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. Journal of Social Structure 3 (2). Web journal available from: <http://www2.heinz.cmu.edu/project/INSNA/joss/index1.html.>

Snijders, T.A.B, Pattison P.E., Robins G. and Handcock, M.S. (2006), New specifications for exponential random graph models. Sociological Methodology 36 (1), pp. 99-153.

Snijders, T.A.B., van de Bunt, G.G., Steglich, C.E.G. (2010a), Introduction to stochastic actor-based models for network dynamics. Social Networks 32 (1), pp. 44-60.

Snijders, T.A.B., Steglich, C., Schweinberger, M. and Huisman, M. (2010b), Manual for SIENA version 3.2. Provisional version. Oxford: University of Oxford, Department of Statistics.

Storper, M. and Venables, A.J. (2004), Buzz: face-to-face contact and the urban economy. Journal of Economic Geography 4 (4), pp. 351-370.

Ter Wal, A.L.J. & R.A. Boschma (2009a), Applying social network analysis in economic geography: framing some key analytic issues. Annals of Regional Science 43 (3), pp. 739-756.

Ter Wal, A.L. J. (2011), The Dynamics of Inventor Networks in German Biotechnology: Geographical Proximity versus Triadic Closure. Papers in Evolutionary Economic Geography 11.02.

Van Burg, E., van Raaij, E. and Berends, H. (2008), The Fibres that Hold an Innovation Network: An Analysis of Knowledge-Sharing in the Glare Network. In: de Man, A.-P. (Ed.), Knowledge Management and Innovation in Networks. Cheltenham: Edward Elgar.

Van Duijn, M.A.J., Gile, K.J. and Handcock, M.S. (2009), A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Social Networks 31 (1), pp. 52-62.

Wasserman, S. and Faust, K. (1994), Social Network Analysis. Cambridge: Cambridge University Press.

Tables and Figures

Table 1: Network descriptives, technological knowledge network of Dutch aviation organizations, 2008	
Attributes	Value
Nodes	59
Links	146
Density	0.085
Max. Component	47
Isolates	12
Degree Centralization	0.411
Between. Centralization	0.213
Mean degree	9.89
Average distance in main component	2.122

Table 2: NAG technological fields	
Technological field according to NAG	Number of firms
Airframe subsystems & components	17
Interiors	10
Propulsion & engine components	15
Auxiliary systems	5
Avionics, simulation & control	12
Education & training	13
General services	3
Engineering & R&D	31
Space subsystems & components	15
Maintenance & overhaul	11
Spare parts	10
Special materials.	10
Consultancy	5

Table 3: QAP-model results

	Estimate	Exp(b)	Pr(<=b)	Pr(>=b)	Pr(>=|b|)
(intercept)	-6.09***	0.00	0.00	1.00	0.00
ORGASIZE	0.08***	1.08	1.00	0.00	0.00
SOCPROX	-0.36	0.70	0.33	0.67	0.61
OPEN	0.52	1.69	0.86	0.15	0.30
AVIASIM	0.34	1.41	0.76	0.24	0.52
INSTPROX	1.19***	3.27	1.00	0.00	0.00
COGPROX	0.43	1.53	0.85	0.15	0.28
GEOPROX	-0.01**	0.99	0.05	0.95	0.10
Chi-Square	1501.46		8 degrees of freedom		
AIC	886.48				
BIC	930.04				
Pseudo-R^2 Measures					
(Dn-Dr)/(Dn-Dr+dfn)	0.47				
(Dn-Dr)/Dn	0.633				

* Significant at 90%; **Significant at 95%; *** Significant at 99%

Table 4: Results ERG-model resembling the QAP-model

Variable	Estimate	Std. Error	MCMC s.e.	p-value	
ORGASIZE	-0.008	185.499	368000.000	1.000	
EXTERNALSIM	-0.373	0.211	0.219	0.077	.
AVIASIM	-0.827	0.211	2.639	<1e-04	***
SOCPROX	-0.329	0.420	1.498	0.434	
INSTPROX	-0.705	0.151	2.284	<1e-04	***
COGPROX	-0.004	185.498	368000.000	1.000	
GEOPROX	-0.025	0.002	0.186	<1e-04	***
AIC	1137.7				
BIC	1175.8				
Degeneracy value	0.167				
Log likelihood	-561.853				

* Significant at 90%; **Significant at 95%; *** Significant at 99%

Table 5: ERG-model with node level and structural network level variables					
Variables	Estimate	Str. Error	MCMC s.e.	p-value	
NON-PROFIT	1.678	0.195	0.023	< 1e-04	***
SOCPROX	1.508	0.415	0.068	0.003	***
AVIASIM	0.775	0.232	0.022	0.001	***
EDGES	-5.021	0.332	0.045	< 1e-04	***
ISOLATES	1.084	0.518	0.139	0.037	*
TRIADCLOS	0.805	0.198	0.044	< 1e-04	***
AIC BIC	754.86 782.08				
Degeneracy value	0.524				
Log likelihood	-372.492				

* Significant at 90%; **Significant at 95%; *** Significant at 99%

Figure 1: Technological knowledge network of Dutch aviation organizations, 2008

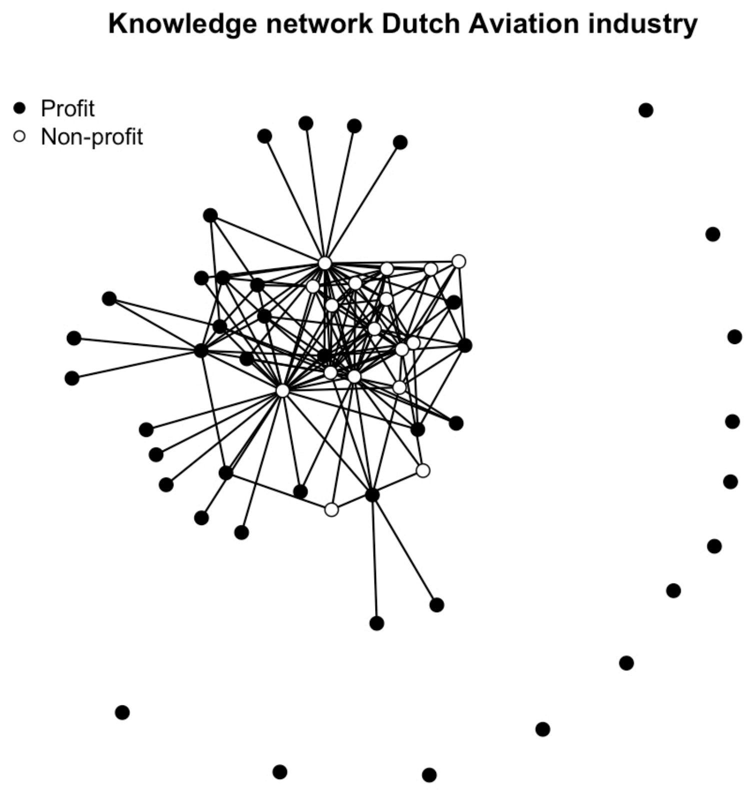

Figure 2: Goodness of fit ERG-model resembling the QAP model (a)

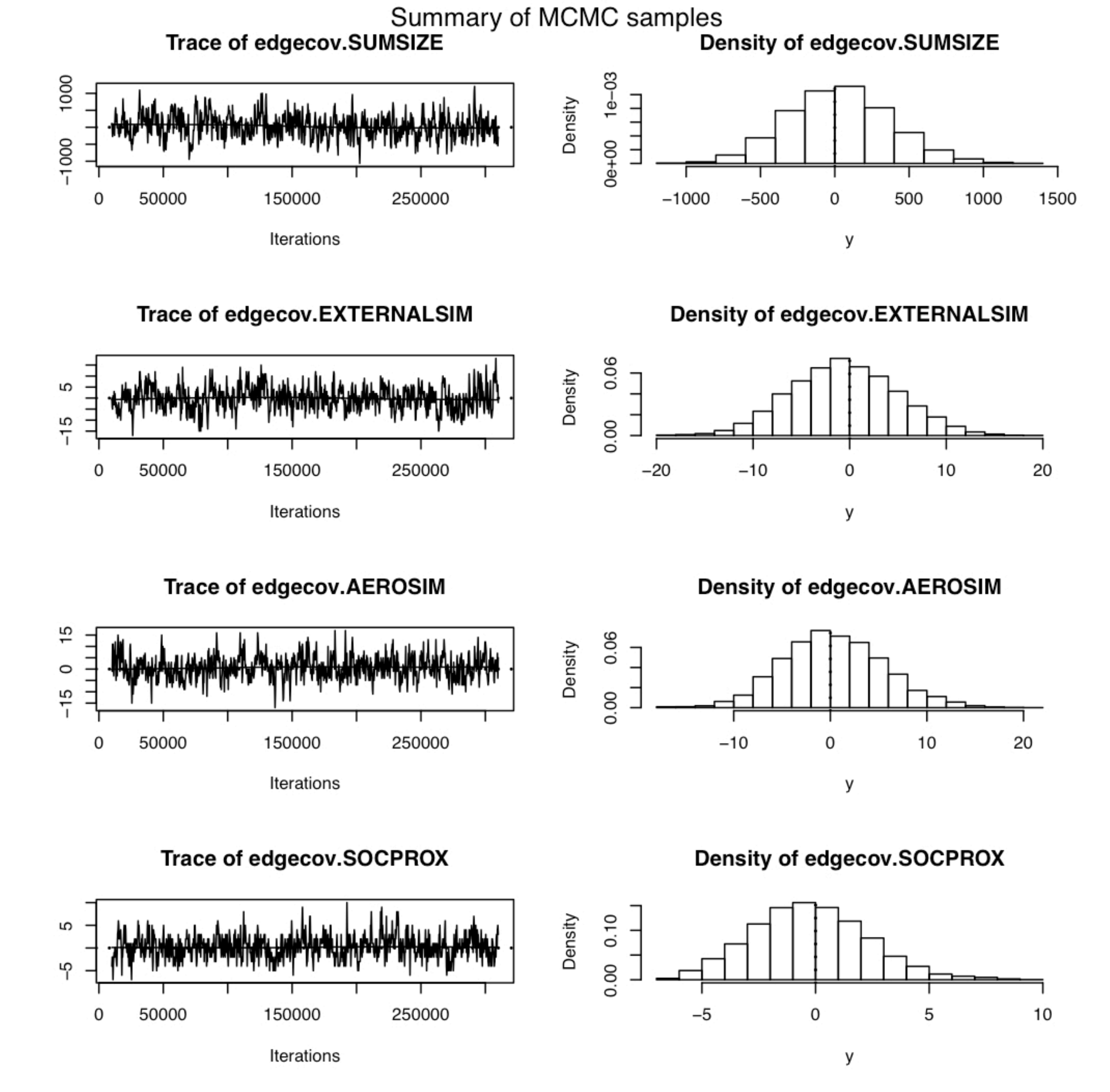

Figure 2: Goodness of fit ERG-model resembling the QAP model (b)

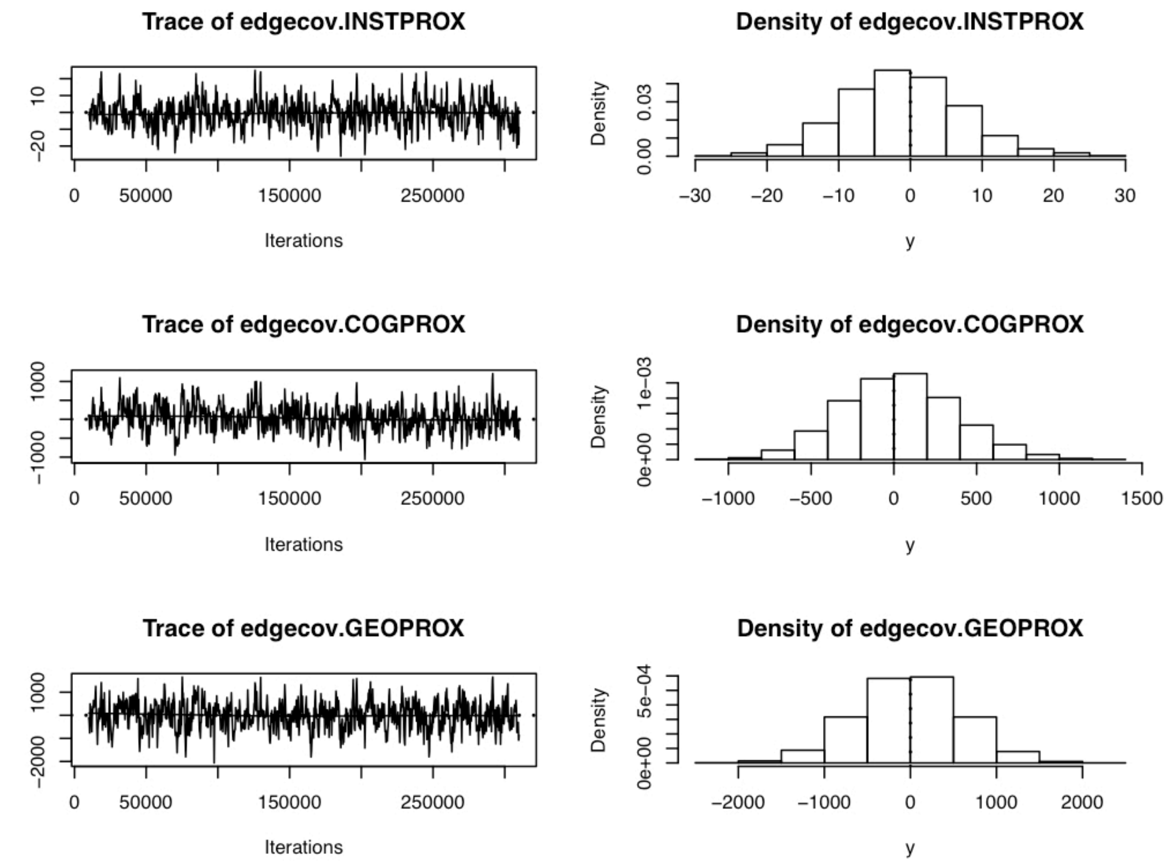

Figure 3: Goodness of fit ERG-model resembling the QAP model

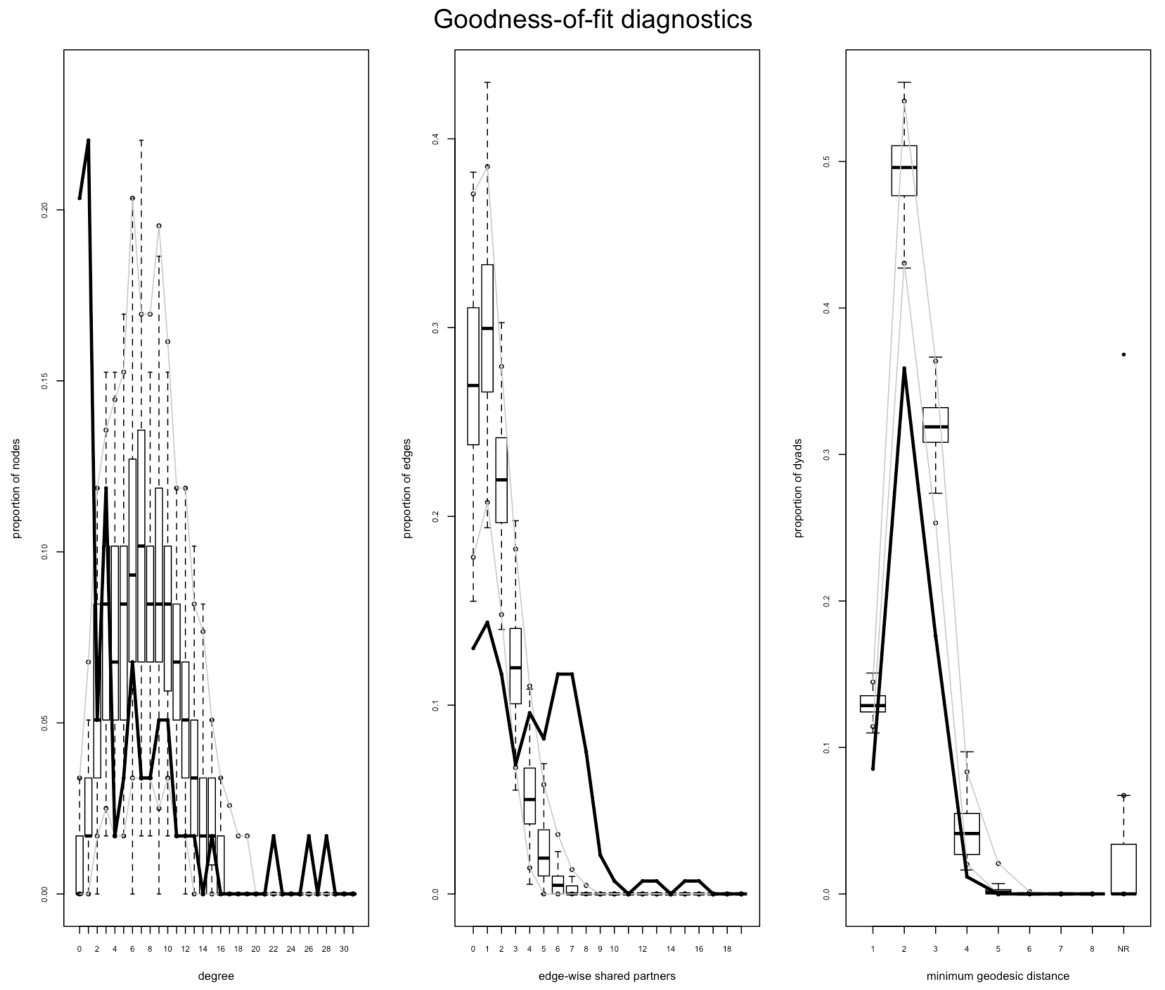

Figure 4: Goodness of fit ERG-model with node level and structural network level variables (a)

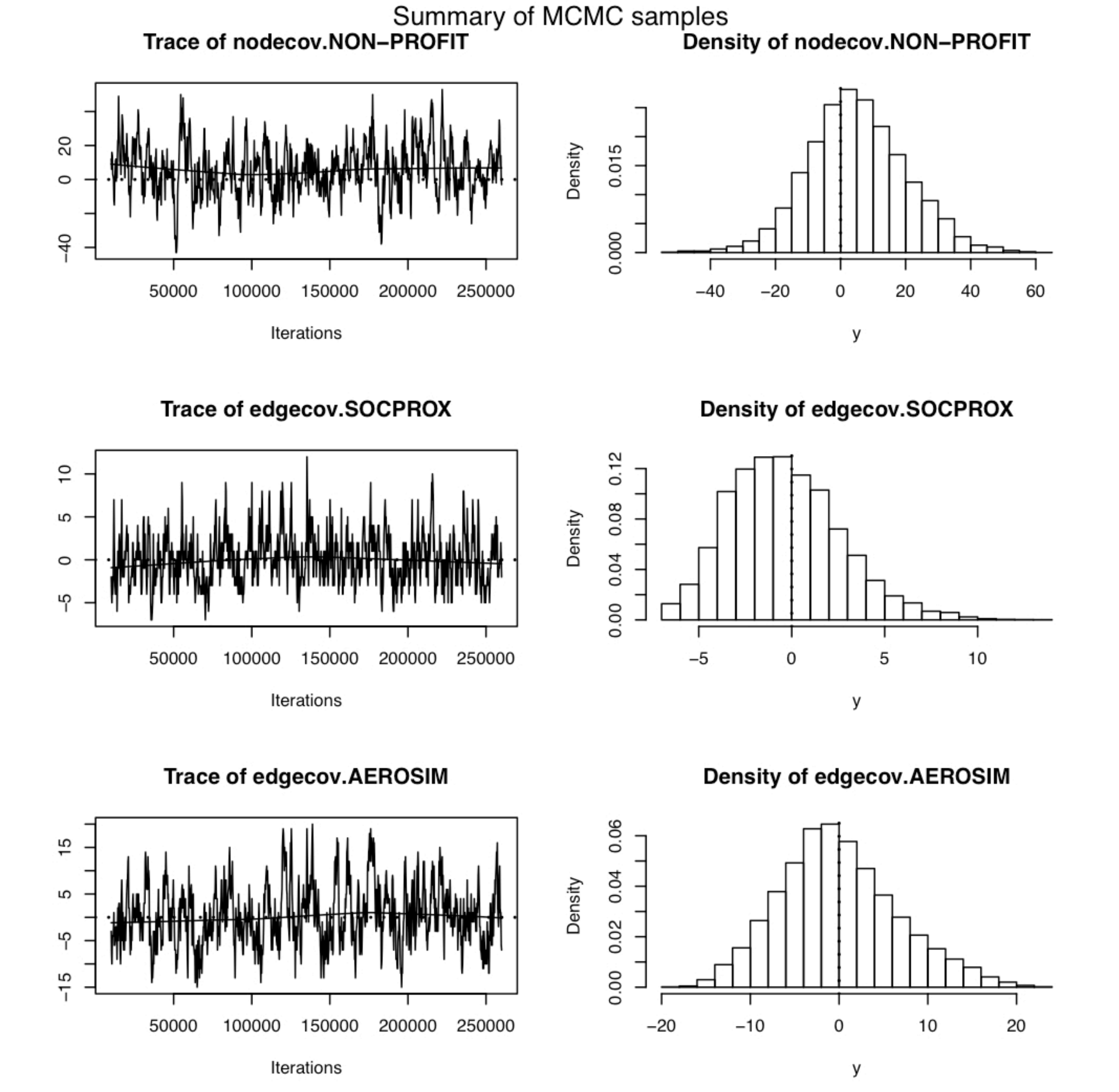

Figure 4: Goodness of fit ERG-model with node level and structural network level variables (b)

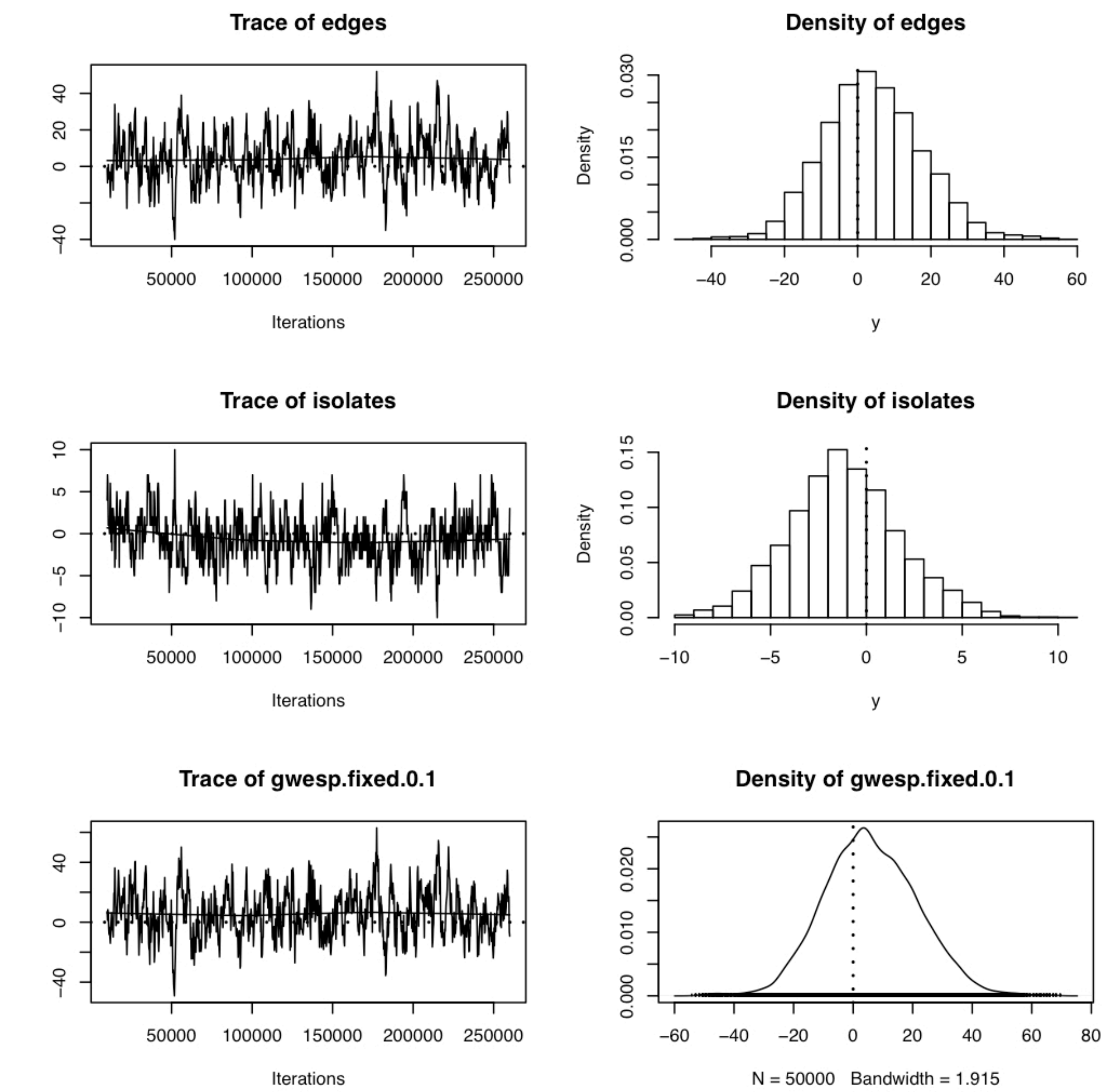

Figure 5: Goodness of fit ERG-model with node level and structural network level variables