

# Papers in Evolutionary Economic Geography

# 13.25

## **Modeling Knowledge Networks in Economic Geography: A Discussion of Four Empirical Strategies**

Tom Broekel, Pierre-Alexandre Balland, Martijn Burger, Frank van Oort



**Utrecht University**

**Urban & Regional research centre Utrecht**

# Modeling Knowledge Networks in Economic Geography: A Discussion of Four Empirical Strategies

Tom Broekel

*(corresponding author)*

Institute of Economic and Cultural Geography, Leibniz University of Hanover

[broekel@wigeo.uni-hannover.de](mailto:broekel@wigeo.uni-hannover.de)

Pierre-Alexandre Balland

Department of Economic Geography, Utrecht University

[p.balland@uu.nl](mailto:p.balland@uu.nl)

Martijn Burger

Department of Applied Economics, Erasmus University Rotterdam

[mburger@ese.eur.nl](mailto:mburger@ese.eur.nl)

Frank van Oort

Department of Economic Geography, Utrecht University

[f.g.vanoort@uu.nl](mailto:f.g.vanoort@uu.nl)

**Abstract:** The importance of network structures for the transmission of knowledge and the diffusion of technological change has been emphasized in economic geography. Since network structures drive the innovative and economic performance of actors in regional contexts, it is crucial to explain how networks form and evolve over time and how they facilitate inter-organizational learning and knowledge transfer. The analysis of relational dependent variables, however, requires specific statistical procedures. In this paper, we discuss four different models that have been used in economic geography to explain the spatial context of network structures and their dynamics. First, we review gravity models and their recent extensions and modifications to deal with the specific characteristics of networked relations. Second, we discuss the quadratic assignment procedure that has been developed in mathematical sociology for diminishing the bias induced by network dependencies. Third, we present exponential random graph models that not only allow dependence between observations, but also model such network dependencies explicitly. Finally, we deal with dynamic networks, by introducing stochastic actor oriented models. Strengths and weaknesses of the different approaches are discussed together with domains of applicability for the analysis of (knowledge) network structures and their dynamics.

**Key words:** Economic geography, knowledge networks, network models, quadratic assignment procedure, gravity model, exponential random graph model, stochastic actor-oriented model

**JEL codes:** R11, O32, D85

# 1. Introduction

Knowledge networks play a crucial role in the economic development of regions (Van Oort and Lambooy 2013). R&D collaborations among organizations (Hagedoorn 2002), labor mobility (Almeida and Kogut 1999), and personal acquaintances of inventors (Breschi and Lissoni 2009) drive innovation activities, technological change, and economic performance of organizations and regions. Beyond these iconic channels of knowledge transfer, the structure of knowledge networks can more generally be defined as the set of direct and indirect connections that individuals and organizations use to access knowledge (within and outside the region). Given the economic value associated with the structure of knowledge networks and their striking spatial dimension<sup>1</sup>, empirical studies of networks have attracted a growing interest in the geography of innovation over the last twenty years<sup>2</sup> (Grabher 2006, Burger et al. 2009a, Maggioni and Uberti 2011, Ter Wal and Boschma 2009).

The increased interest in the empirics of knowledge networks can be seen as a response to the traditional metaphorical treatment of networks in economic geography and regional science in general and the study of agglomeration economics in particular (Sunley, 2008). Despite over twenty years of research on the benefits of agglomeration, the empirical literature remains inconclusive about the mechanisms and processes that lead to more than proportional regional economic growth. Despite the fact that the micro-foundations (such as knowledge spillovers, labor market pooling, and input sharing) that underlie agglomeration economies have theoretically been specified) agglomeration is often treated as a black box in empirical studies (Burger et al. 2009a, Van Oort and Lambooy, 2013). This is exemplified by the fact that most empirical studies on agglomeration economics merely research the relationship between urban or cluster size and regional economic development (see Melo et al. 2009 and De Groot et al. 2009 for meta-analyses of this literature) and do not examine the different channels through which the concentration of economic activities affect regional economic development.

The analysis of networks, either formal or informal, can help us to identify these channels and get a glimpse of what is in the black box of agglomeration economies (Burger et al 2009a), hereby extending the current discourse on agglomeration externalities in which new conceptual and methodological approaches are needed (Van Oort and Lambooy, 2013). Over the past years, a large literature has developed in economic geography, regional science, management, and sociology that predominantly address the determinants of knowledge and information transfer, focusing on spinoff firms, labor mobility and R&D collaboration (Boschma and Frenken, 2006). One of the main findings of this literature is that firms in agglomerations do not profit automatically from co-location. Instead, knowledge spillovers

---

<sup>1</sup> A burgeoning literature starts to integrate the geographical dimension in sociology and network science: see for instance the special issue 34.1 in *Social Networks* of January 2012 on Capturing Context: Integrating Spatial and Social Network Analysis, edited by Jimi Adams, Katherine Faust and Gina Lovasi.

<sup>2</sup> See the special issue 43.3 in *The Annals of Regional Science* of September 2009 on Embedding Network Analysis in Spatial Studies of Innovation, edited by Edward Bergman.

mainly take place between firms that are networked and strongly locally embedded. A second finding that has come out of this strand of research is that a substantial part of information and knowledge transfer takes place over longer distances as firms have many network relations outside the city or cluster they are located in. From this, it evidently follows that cities and clusters are not spatially isolated entities, but embedded in a system of cities. In the end, an explicit focus on the transfer and network mechanisms of knowledge diffusion cannot only help us to identify the channels through which firms benefit from agglomeration, but also help us to identify (1) which firms profit from knowledge spillovers and (2) the spatial extent of information and knowledge transfer. These are important ingredients of current innovation and network-based (“smart”) growth strategies in the European Union (Thissen et al 2013). In the European Union, knowledge networks, free movement of knowledge workers, information flows and knowledge-based cooperation opportunities in research and development are hypothesized to contribute to local innovation opportunities by academics and policymakers alike (Hoekman et al 2009, Scherngell et al 2013, Basile et al. 2013). Without a network perspective on knowledge, trade and investments, a proper assessment of place-based growth strategies as advocated by the European Union (Barca et al 2012) is impossible (Thissen et al 2013).

In this light, the increased attention for modeling the determinants of network formation is very much needed, especially in order to get a fully-fledged understanding of information and knowledge transfer in and across regions. It enables us to explain why individuals, organizations, and regions differ in their embeddedness in information and knowledge networks, why they vary in their learning and innovation capabilities, and whether this results in variation in their performance. Analyzing the formation and evolution of network structures, however, is more complex than computing structural descriptive statistics like degree, betweenness, clustering coefficient, or average geodesic distance. Explaining the structure of knowledge networks requires an inferential statistics framework, where the dependent variable is related to the overall structure of the network. Even when networks are decomposed into their smallest unit, the dyad, relational data does not fit well into traditional regression frameworks. A fundamental property of network structures lies in the existence of conditional dependencies between observations, especially between dyads that have actors in common (Lindgren et al. 2010). By nature, such network dependencies violate standard statistical inference procedures that assume independence among observations. But more than only correcting for such dependencies, the main challenge is to use the information included in these dependencies to model structural predictors of network formation.

In this paper, we provide a discussion of the main empirical strategies that have been proposed recently in economic geography to explain the formation and structure of networks. Although these strategies are briefly mentioned in a few methodological papers (Ter Wal and Boschma 2009, Broekel and Hartog 2013a, Maggioni and Uberti 2011), a global discussion of their respective range of applicability, strengths, and weaknesses in the context of economic geography is still missing. We believe such a discussion to be useful for economic geographers and regional scientists aiming at modeling network formation, especially because the different models have emerged out of different scientific traditions. Moreover, they are often based on different assumptions, vary in terms of conceptual and empirical issues (like

micro-macro relations, network dynamics, and network-geography interdependencies), and frequently require different types of relational data. This paper provides a discussion and an introduction to four main types of empirical strategies: gravity models (GM), quadratic assignment procedures (QAP), exponential random graph models (ERGM), and stochastic actor-oriented models (SAOM).

Section 2 discusses GM, a class of econometric models generally used in economics to explain the flow between geographical units as a function of supply and demand factors and the distance between the units. These have recently been extended to deal with the specific characteristics of network data. To account for more complex network dependencies, QAP has been developed in mathematical sociology on the principle of bootstrapping procedures. The class of ERGM has been developed on the basis of a Markov chain to include not only dyadic effects but also structural effects at the network level. Lastly, SAOM have been introduced again in mathematical sociology to provide a class of statistical models for network dynamics. This allows for treating of longitudinal rather than cross-sectional data, and therefore the analysis of changing network relationships.

## 2. Gravity models

### 2.1. The History of Gravity Models

In economic geography and regional economics, network structures are frequently predicted and elucidated with an analogy to Newton's law of universal gravitation. In its most elementary form, the gravity model predicts that the flow or interaction intensity between two objects (e.g., origin and destination) is assumed to be directly correlated with the masses of the objects and inversely correlated with the physical distance between the objects. More formally,

$$I_{ij} = K \frac{M_i^{\beta_1} M_j^{\beta_2}}{d_{ij}^{\beta_3}} \quad (1)$$

where  $I_{ij}$  is the interaction intensity between object  $i$  and  $j$ ,  $K$  a proportionality constant,  $M_i$  the mass of the object  $i$  (e.g., origin),  $M_j$  the mass of object  $j$  (e.g, destination), and  $d_{ij}$  the physical distance between the two objects.  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  are parameters to be estimated.  $\beta_1$  refers to the potential to generate flows,  $\beta_2$  is related to the potential to attract flows, and  $\beta_3$  is an impedance factor reflecting the rate of increase of the friction of physical distance.

The first appearance of the gravity model in the social sciences dates back to the mid-19th century when it was applied to the study of human interaction patterns (Carey 1858), who used the analogy to Newton's law to answer the question why a city was more likely to attract people than a small town.

The first empirical studies using the gravity model framework appeared at the end of the 19<sup>th</sup> and early 20<sup>th</sup> century, when it was applied to the study of migration (Ravenstein 1885), railway travel (Lill 1891), and retail trade (Reilly 1931). The modern use of the gravity model

was popularized in the school of social physics after the Second World War and formalized by Stewart (1948), Isard (1956), and Tinbergen (1962).<sup>3</sup> Over the course of the years, the model has been applied to a wide variety of spatial interaction patterns, such as international trade, foreign direct investment, tourism, migration, commuting, and shopping. Within the context of the geography of innovation and knowledge transfer, the gravity model framework has been used in studies on inter-alia co-inventorship and co-publishing (Maggioni et al. 2007, Ponds et al. 2007, Hoekman et al. 2009), citation networks (Peri 2005, Fischer et al. 2006), R&D collaboration through European programs (Scherngell and Barber 2009), inventor mobility (Miguelez and Moreno 2012), foreign direct investment in R&D facilities (Castellani et al. 2012), and trade in high-technology products (Liu and Lin 2005). In most empirical research using gravity models, the objects are spatial units, such as cities, regions, or nations. However, disaggregated data at the firm or individual level is increasingly employed to assess the spatial dimension of innovation networks within a gravity model context (see, e.g., Autant-Bernard et al. 2007, Breschi and Lissoni 2009).<sup>4</sup>

## 2.2. The Working Principles of the Gravity Model

Unlike the later introduced QAP, ERGM, and SAOM, the gravity model is a conceptual model and not just a statistical method.<sup>5</sup> Traditionally, the gravity model as presented in equation (1) has been estimated using Ordinary Least Squares (OLS). Taking logarithms of both sides of equation (1) and including a disturbance term, this multiplicative form can be transformed into a linear stochastic form. It results in a testable equation (2), in which  $\varepsilon_{ij}$  is assumed to be identical and independently distributed (i.i.d):

$$\ln I_{ij} = \ln K + \beta_1 M_i + \beta_2 M_j - \beta_3 d_{ij} + \varepsilon_{ij} \quad (2)$$

The model can be extended to a panel data framework, so that it becomes possible to study the development of relational structures over time. In addition, the empirical gravity model can be easily augmented to include other factors that influence network structures. Accordingly, in most of the above-mentioned studies rather than the Newtonian version but a more general form of the gravity model is used, in which the flow between two objects is hypothesized to be dependent on supply factors at the origin that generate flows, demand factors at the destination that attract flows, and by stimulating or restraining factors (e.g., proximity or distance) pertaining to the specific flow between the two objects. For example, it can be argued that the flow of knowledge in networks of R&D collaboration is not only dependent on the physical distance, but also on the cultural, social, and institutional distance between the two regions (Boschma 2005). Likewise, it is not only public investments in R&D that generates knowledge flows, but also the presence of human capital in a region.

---

<sup>3</sup> For an early overview of studies that applied the gravity model in economic geography, see Lukermann and Porter (1960).

<sup>4</sup> However, the term “gravity model” is not often used when studies are conducted at the micro-level. Rather scholars research the effect of geographical proximity on network formation.

<sup>5</sup> In practice, it would be possible to estimate the gravity model with these techniques.

However, there are also some serious problems with the traditional OLS specification of the gravity model. Most importantly, the OLS specification does not control for dependencies present in network data nor is it very well able to model network dependencies. In particular, the traditional equation assumes that flows between two actors are independent from other relationships between actors within the network. Since this strong assumption of structural independency is very unlikely to hold, this can lead to biased estimates of the gravity equation (Anderson and Van Wincoop 2003). Although the fact that the flow between two locations is dependent on the characteristics and the number of alternative locations is well known in the gravity literature (see already the work of Stouffer in 1940), this has until recently not been explicitly addressed in empirical gravity models.

In the recent literature on gravity models, several extensions and modifications of the gravity model have been proposed to deal with this issue (Herrera, 2011). Although most of these originate from the spatial and international economics literature on the gravity model of trade, they can easily be applied to the study of innovation networks.<sup>6</sup> First, researchers have tried to control for network structure by correcting standard errors. More specifically, use has been made of the sandwich style standard errors using multiway clustering on the origin and destination (Lindgren 2010) or a spatial error model (Fischer and Griffith 2008; Scherngell and Lata 2013).<sup>7</sup> These procedures allow for a more careful modeling of the error structure, controlling for correlations that may arise in the error terms. However, as pointed out by Snijders (2011), such empirical strategies mainly treat the network as nuisance and do not modeling network dependencies explicitly.

Second, it is possible to include an indicator for remoteness  $R_i$  to account for third party effects (Head and Mayer, 2000), which proxies the average transaction costs faced by a location:

$$R_i = \sum_j \frac{d_{ij}}{(y_j/y_{world})} \quad (3)$$

Where the numerator represents the bilateral distance between countries  $i$  and  $j$ , and the denominator is for instance the share of country  $j$ 's GDP in the world's GDP. The remoteness variable proxies the full range of potential destinations to a given origin, taking into account the importance of the respective destinations and average distance of a country to all other countries. The advantage of this empirical strategy is that such a remoteness variable is easy to construct. However, as indicated by several authors, this empirical strategy fails to capture other barriers than distance that may hamper interaction (e.g., national borders) (Anderson 2001).

---

<sup>6</sup> Please note that we only discuss problems specifically pertaining to network data. Other problems related to, for example, the fact that the outcome is not always a continuous numeric variable and the many zeros in the network (e.g., Helpman et al 2008; Burger et al 2009b) and causality (e.g., Egger 2004) are discussed elsewhere in the literature. Although these are problems that all empirical researchers are facing, a discussion of these issues is beyond the scope of this paper.

<sup>7</sup> Another (non-spatial) method that controls for the network structure but is not often used in the gravity model literature is the multiple regression quadratic assignment procedure (MRQAP). A more elaborate discussion of this method can be found in the next section.

Third, a fixed-effects specification can be employed to deal with the problem of intervening opportunities. In a cross-sectional setting, this implies including country-specific exporter and importer dummy variables in equation (2). Such specification controls for country-specific fixed effects related to origins and destinations, such as the supply, demand, and origin- and destination-specific transaction costs, which are often difficult to measure, but influence the structure of the network. Following Anderson and Van Wincoop (2003) and Feenstra (2004), such a specification of the gravity equation would be in line with the theoretical concerns regarding the correct specification of the model and yields consistent parameter estimates for the variables of interest. However, when such a strategy is employed, it is impossible to incorporate any origin- or destination-specific (or individual-specific) factors within a cross-sectional setting. In addition, Behrens et al. (2012) point out that such a fixed effects estimation does not fully capture the spatial interdependence among flows, and hence, the assumption of independence of observations might still be violated.

Third, there are a couple of other, more complex strategies to deal with structural dependencies in the gravity model, including estimation of multilateral resistance terms (Anderson and Van Wincoop 2003) and a spatial autoregressive moving average specification for the gravity model (Behrens et al 2012).<sup>8</sup> These strategies have in common that they try to model dependencies present in network data directly and are becoming increasingly popular within the gravity modelling literature, especially within the fields of spatial and international economics. Focusing on trade, Anderson and Van Wincoop (2003) show that bilateral barriers between two countries do not determine the flow of bilateral trade only, but also how easy it is for these countries to trade with the rest of the world. Anderson and Van Wincoop (2003) try to capture these relative barriers by including country-specific price indices, called multilateral resistance terms, which are estimated using a multi-step nonlinear least squares procedure. However, since the method is computationally intensive, it has not been implemented by many researchers, which tend to prefer a fixed effects estimation using OLS or count data models. As an alternative, Behrens and co-authors introduce a spatial econometric estimation of the gravity model<sup>9</sup>, accounting for cross-sectional correlations between flows and controlling for possible cross-sectional correlations in the error terms.

Focusing on trade between US and Canadian regions, the authors find that the exports of any region to a market negatively depend on the exports from the other regions to that market, which themselves depend on the whole distribution of bilateral trade barriers. In addition, the model can incorporate heterogeneous coefficients, allowing relationships to vary across units: for example, the distance decay of trade might differ across regions. Along these lines, the model proposed by Behrens et al. (2012) provides also a subtle link between theory and empirical methods when it comes to trade network research. At the same time, this empirical strategy can be easily extended to other types of flows to capture structural dependencies in general and spatial competition effects in particular. Although these approaches to model structural dependencies are heading into the right direction, there is, however, still no standard

---

<sup>8</sup> Less well known but comparable empirical strategies in this respect are provided by Bikker (2010) and Linders et al. (2010).

<sup>9</sup> For an earlier use of spatial lag models in gravity modeling, see LeSage and Pace (2008).



implementation in software packages.<sup>10</sup> In addition, few empirical studies have applied these methods and, hence, the appropriateness of these methods has to be further evaluated.

### **3. Multiple regression quadratic assignment procedure**

The empirical strategy involving the multiple regression quadratic assignment procedure (MRQAP in the following) starts from a similar viewpoint as the gravity model. In the context of economic geography, the dependent variable of interest is the relational intensity of interaction between individuals, organizations, or spatial units, such as cities or regions. However, in contrast to the gravity model's conceptual basis it can be seen as a purely statistical approach to account for structural dependencies among relational data. In principle, the correction procedure that is proposed can also be put into a gravity model framework, which, to the best of our knowledge, has however not been done so far. More precisely, the multiple regression quadratic assignment procedure-model is a logit or OLS regression model, which incorporates relational variables and considers their inherent interdependencies when assessing their statistical relevance.<sup>11</sup>

MRQAP approaches are applied in a number of studies on inter-organizational networks. However, only recently it found its way into the literature of economic geography. Among the first is the study by Bell (2005) who uses a bivariate quadratic assignment correlation procedures to statistically infer about correlations between friendship, information, and advice networks among executives from within and outside the Toronto industry cluster. Subsequently, the MRQAP procedure has been used to study the intensity of co-inventing among patent inventors located in the region of Jena (Graf and Cantner 2006), to explore the relevance of cognitive, social, institutional, and geographic proximity for the knowledge network connecting Dutch organizations active in the field of aerospace (Broekel and Boschma 2012), and to study the relationship between regional flows of internet hyperlinks, co-patenting relations, EU-funded research collaboration, and the flow of Erasmus exchange students (Maggioni and Uberti 2007). Nevertheless, MRQAP is much less prominent in economic geography than gravity models.

#### **3.1. The History of MRQAP**

Mantel introduced the quadratic assignment procedure in 1967, when he was working at the National Cancer Institute in Maryland and reviewed a number of common empirical approaches used to identify (non-random) time-space clustering of disease (Mantel 1967). The basic statistical problem was the clustering of disease cases in space and in time. While statistical tools were available dealing with spatial or temporal clustering, the simultaneous (two-dimensional) occurrence of the two clustering types remained an empirical challenge.

---

<sup>10</sup> The mathematical appendix and Matlab codes of the approach by Behrens et al. (2012) can be found in the Web Appendix of their article, available at the *Journal of Applied Econometrics* website.

<sup>11</sup> Accordingly, MRQAP is rather a particular permutation method for hypothesis testing and not a model on its own. However, we will refer to it as model in the following to keep a consistent terminology.

Mantel proposed an uncorrected correlation coefficient estimated as the cross-product of the distances in the two dimensions' empirical matrices (spatial distances and temporal distances). To overcome the problem of highly interrelated  $n^2$  values, Mantel constructed repetitively data sets corresponding to the null hypothesis of no correlation between the two matrices by permuting the rows and columns of the two matrices in the same way and such that the values of any row and of a column combination remain together (but change their positions within the matrix). If the null hypothesis is correct than these permutations “*should be equally likely to produce a larger or a smaller coefficient*” (Schneider and Borlund 2007, p. 7). On this basis the Mantel test was developed for estimating the correlation between any two distances matrices (Mantel and Valand 1970). Although Mantel's approach was initially developed for the identification of disease clusters, the procedure can without any difficulties be applied to other contexts (Mantel, 1967).

Hubert and Schultz (1976) introduced the notion of the “quadratic assignment procedure” as an equivalent to the Mantel test. From there, the test statistics were refined and generalized in multiple ways (see for a review Hubert (1987)). In social network analysis this approach became popular through the works of Krackhardt (1987, 1988). He extended the QAP methodology to test the relationship between multiple relational matrices in a regression framework. Since then, the methodology has been subject to multiple refinements including amongst others more advanced approaches to deal with multicollinearity and certain types of autocorrelation (see, e.g., Dekker et al. 2007).

### **3.2. The Working Principles of MRQAP**

At its core, a QAP regression is a combination of the Mantel test, i.e. quadratic assignment procedure and a standard OLS or Logit regression. The dependent variable is hereby the matrix of inter-actor relations. Whether to use an OLS or a Logit model depends on the available network data. For a valued network OLS is appropriate while binary (0/1) network data requires the logit regression. As before, the independent variables are factors whose influence is to be tested on the structure of the network.

To explain the MRQAP, it is best to start from all variables (dependent and independent) being stored in types of adjacency matrices. For the application of standard regressions, these need to be vectorized implying that the matrices are transformed into a column vector of  $n^2$  elements by stacking the columns “on top of each other”.

As pointed out in the previous section, network data are characterized by frequent row/column/block autocorrelation because on dependent observations implying that standard tools of inference are therefore invalid. In the style of the Mantel test, Krackhardt (1987, 1988) therefore suggests comparing the regression statistics to the distribution of such statistics resulting from large numbers of simultaneous row/column permutation of the considered variables (before the vectorization). While still being subject to extensions and further developments, most of today's studies use the multicollinearity robust “semi-partialling plus” procedure by Dekker et al. (2007).

The MRQAP is nonparametric in nature implying that in comparison to standard regression techniques there are few crucial assumptions and issues. However, there are a couple of issues that deserve future research. Among these are spuriousness, multicollinearity, and skewness (see Dekker et al. 2007). The interpretation of the obtained parameter values depends on the type of regression function used (OLS or Logit).

## 4. Exponential random graph models

ERG-models are well known and established in many disciplines. For example in biosciences, Saul and Filkov (2007) use ERGM to explain the structure of cell networks. Fowler et al. (2009) employ ERGM to model genetic variation in human social networks in life science. They are also frequently used in sociology and political science, for instance to analyze the structure of networks of friendship networks (Lubbers and Snijders 2007) or political international conflicts (Cranmer and Desmarais 2011). While there are a number of studies that focus on the role geography plays for the formation of social networks (see, e.g., Daraganova et al. 2012), ERGM have rarely found their way into the analysis of networks in economic geography. Recent contributions to the field of economic geography that use an ERGM-approach include studies on inter-organizational knowledge networks in the Dutch aviation industry (Broekel and Hartog 2013a), networks among biotechnology organizations as created by participating in the EU Framework Programmes (Hazir and Autant-Bernard 2012), and determinants of cross-regional R&D collaboration networks (Broekel and Hartog 2013b).

### 4.1. The History of ERG-Models

In contrast to the previous approaches, the roots of the exponential random graph models are more difficult to identify. Surely the work of Solomonoff and Rapoport (1951) on random graphs was fundamental. Solomonoff a physicist and Rapoport a mathematician conducted the “*the first systematic study on what we would now call a random graph*” (Newman et al. 2006, p.12). These authors already discussed a number of important properties of such graphs (e.g., average component size). However, it took another ten years before Erdős and Rényi (1960) finally popularized the concept of random graphs. By putting forward the Bernoulli random graph distribution, which could be used to estimate configurations of individual links between actors. However, Erdős and Rényi assumed independent links among nodes, which is clearly problematic for many networks. The next major step in the development of ERG-models was the introduction of p1 models by Holland and Leinhardt in 1981. These models are based on the work of Besag (1974, 1975) who showed that a class of probability distributions existed, which are consistent with the (Markovian) condition that the value of one node is dependent only on the values of its adjacent neighbors.<sup>12</sup> Holland and Leinhardt (1981) proposed a family of exponential distribution (p1 distribution) that could be used as

---

<sup>12</sup> Besag (1974, 1975) applied this idea the context of spatial data the idea is, however, also applicable in the context of network data.

null-hypothesis for assessing real-world networks (conditional on the density of the network and the number of links to and from a node). In a direct comment to Holland and Leinhardt’s article, Fienberg and Wasserman (1981) showed that these models can also be estimated using log-linear modeling techniques, which significantly increased the use of  $p1$  models. However, the  $p1$  approach is troubled by the assumption of dyad-independence that is frequently found to be incorrect (Newman 2003). Another major breakthrough represents the work by Ove and Strauss. In their 1986 study they proposed the use of Markov random graphs to overcome the problems related to dyad-independence made in  $p1$  models. In the context of networks, Markov dependence is used to model a link between node A and B being contingently dependent on other possible links of A and B. This marked a significant shift from dyad-independence as two links are assumed to be conditionally (Markov) dependent (Robins et al. 2007). This assumption “*for the first time could be viewed as empirically and theoretically plausible*” (Robins and Morris 2007 p. 170). More than ten years later, Wassermann and Pattison popularized these Markov random graphs in a more generalized form, which are known today as  $p^*$  models (see, e.g., Wasserman and Pattison 1996, Pattison and Wasserman 1999). These models are still the basic building blocks for ERGM (Snijders et al. 2010a). Park and Newman (2004) moreover link ERGM to kinetic mechanics and prove that they “are not merely an ad hoc formulation studied primarily for their mathematical convenience, but a true and correct extension of the statistical mechanics of Boltzmann and Gibbs to the network world” (Park and Newman, 2004, p. 2).

## 4.2. The Working Principles of ERGM

ERGM are stochastic models that perceive link creation being a continuous process, which takes place over time. It implies that an empirically observed network at one particular moment in time can be seen as “*one realization from a set of possible networks with similar important characteristics (at the very least, the same number of actors), that is, as the outcome of some (unknown) stochastic process*” (Robins et al. 2007 p. 175). The basic idea of ERGM is to find a model of a network formation process that maximizes the likelihood of an observed network ( $x$ ) being created at some point in time in this process. As pointed out above, the ERGM builds upon the ideas of exponential graphs, which show in their general specification as (see Robins et al. 2007):

$$\Pr(X = x) = \frac{1}{c} \exp \left( \sum_A \eta_A g_A(x) \right) \quad (4)$$

$\Pr(X=x)$  represents the probability that the network ( $X$ ) created in the exponential random graph process is identical (in terms of a number of specific characteristics) with the empirically observed network ( $x$ ).  $\eta_A$  is the parameter corresponding to network configuration A, and  $g_A(x)$  represents the network statistic. Network configurations can be factors at the node level, dyad level, and structural dependencies. Their corresponding network statistics obtain values of 1 if the corresponding configuration is observed in the

network  $y$  and 0 if not.  $\kappa$  is a normalizing constant ensuring that the equation is a proper probability distribution (summing up to 1). It is defined as

$$\kappa = \sum_{\chi} \exp \left\{ \sum_A \eta_A g_A(x) \right\} \quad (5)$$

with  $\chi(n)$  being the space of all possible networks with  $n$  nodes. Accordingly, the probability  $Pr(X=x)$  depends on the network statistics  $g_A(x)$  in the network  $x$  and on the parameters represented by  $\eta_A$  for all considered configurations  $A$ . The value of  $\eta_A$  indicates the impact of the configuration on the log-odds of the appearance of a tie between two nodes.<sup>13</sup>

In an ERGM estimation equation 4 is solved such that parameter values are identified for each configuration that maximize the probability of the resulting (simulated) network being identical to the one empirically observed. Preferably this is achieved with Maximum Pseudo Likelihood or Markov Chain Monte Carlo Maximum Likelihood Estimation. The latter is nowadays most preferred as it yields the most accurate results (Snijders 2002, Van Duijn et al. 2009). The procedure involves 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. The process is repeated until the parameter estimates stabilize. In case they do not, the model might fail to converge and hence becomes unstable (see for technical details, e.g., Hunter et al. 2008).

An essential part of an analysis using ERGM is the testing of the model's "goodness of fit". This involves checking whether the parameters predict the observed network in a sufficient manner. The structures of the simulated networks are thereby compared to the structure of the observed network. According to Hunter et al. (2008) such involves a comparison of the networks' degree distributions, their distribution of edgewise-shared partners, and their geodesic distributions. The edgewise-shared partner statistic refers to the number of those links in which two organizations have exactly  $k$  partners in common, for each value of  $k$ . The geodesic distribution represents the number of node pairs for which the shortest path in between is of length  $k$ , for each value of  $k$ . The more these statistics are similar for the estimated and empirically observed network the better the former's fit, which implies it being more accurate and hence delivering more reliable parameters for the network statistics of interest.

In addition to these goodness-of-fit tests, the traces of the simulated parameter values over the course of iteration should be relatively stable and vary more or less around the mean value (see for a discussion, Goodreau et al. 2008).

The parameters of the ERGM can be interpreted as non-standardized coefficients obtained from logistic regression analysis, which can be transformed into odd ratios.

---

<sup>13</sup> More details can be found in in Robins et al. (2007).

## 5. Stochastic Actor-Oriented Models

With the growing interest on network dynamics, the availability of longitudinal relational data, and more powerful computers, applications of Stochastic Actor-Oriented Models (SAOM) have started to recently emerge in economic geography. Balland (2012) analyses the influence of different proximity dimensions on the evolution of collaboration networks in the *navigation by satellite industry* in Europe. Balland et al. (2013) and Ter Wal (2013) test the changing influence of network drivers (geographical distance for instance) at different stages of the industry life cycle for the video games and biotech industry, respectively. SAOM are also used by Giuliani (2010) to explain the dynamics of networks within a wine cluster in Chile and by Boschma et al. (2011) to model the structural dynamics of knowledge spillovers (patent citations among firms) in space.

### 5.1. The History of SAOM

In contrast to the previously presented approaches, Stochastic Actor-Oriented Models (SAOM) are a class of statistical models that have been specifically developed for the analysis of network dynamics. The most well-known SAO-models have been proposed by Snijders (2001) in order to provide a statistical model able to analyze empirically the evolution of complex network structures. By combining random utility models, Markov processes and simulation (Van de Bunt and Groenewegen 2007), the SAOM has permitted to study the dynamic of networks and thus to provide recently new results in many fields of social science. A general introduction to SAOM can be found in Snijders et al. (2010), while mathematical foundations of these models are detailed in Snijders (2001). In this discussion, we refer to SAOM implemented in the RSiena statistical package (Ripley et al. 2011).<sup>14</sup>

### 5.2. Working principles of SAOM

The main objective of this class of models is to explain observed changes in the global network structure by modelling choices of actors at a micro-level. More precisely, this statistical model simulates network evolution between observations and estimates parameters for underlying mechanisms of network dynamics by combining discrete choice models, Markov processes, and simulation (Snijders et al. 2010b). Similarly to ERGM, SAOM not only account for statistical dependence of observations, but also explicitly model structural dependencies, like triadic closure. Endogeneity of network structures, i.e. the fact that networks reproduce themselves over time is not perceived as an econometric issue that needs to be corrected, but as a rich source of information used to model the complex evolution of network structures. SAOM are probably the most promising class of models allowing for statistical inference of network dynamics.

---

<sup>14</sup> This class of models is often referred to directly as SIENA models. SIENA stands for "Simulation Investigation for Empirical Network Analysis". The RSiena package is implemented in the R language and can be downloaded from the CRAN website: <http://cran.r-project.org/web/packages/RSiena/>.

The dependent variable in SAOM is not a list of dyads, but the structure resulting from relationships between a set of actors, i.e. the particular way relationships between actors are organized. The dynamic nature of SAOM lies in the fact that the model explains how the observed structure of relations evolves from time  $t$  to time  $t+1$ . Therefore, the dependent variable is a set of consecutive observations of links between actors, which are organized as time series  $x(t), t \in \{t_1, \dots, t_m\}$  for a constant set of organizations  $N = \{1, \dots, n\}$ . These network structures are modeled as a continuous-time Markov chain  $X(t)$ . Thus,  $t_1$  to  $t_m$  are embedded in a continuous set of time points  $T = [t_1; t_m] = \{t \in \mathcal{R} \mid t_1 \leq t \leq t_m\}$ . As specified in Steglich et al (2006, p.3) the basic idea “*is to take the totality of all possible network configurations on a given set of actors as the state space of a stochastic process, and to model observed network dynamics by specifying parametric models for the transition probabilities between these states*”. Each observation is represented by a  $n \times n$  matrix  $x = (x_{ij})$ , where  $x_{ij}$  represents the link from the actor  $i$  to the actor  $j$  ( $i, j = 1, \dots, n$ ). In the simplest specification of the model, the links between the  $n$  actors are represented by directed dichotomous (0/1) linkages implying the analysis of asymmetric adjacency matrices. While the analysis of undirected networks is also possible (Snijders 2010), valued networks are not yet implemented.

*Assumptions of SAOM.* The modelling of the evolution of network structures in SAOM is based on a certain number of underlying assumptions.<sup>15</sup> Most of these assumptions are related to the fact that the evolution of network structures is modelled as a time-continuous Markov chain, driven by probability choices at the actor level. Therefore, this Markov chain is a dynamic process where the network in  $t+1$  is generated stochastically from its architecture in  $t$ , which allows for the existence of statistical dependence between observations. The implication of this modeling strategy is that change probabilities exclusively depend on the current state of the network and not on past configurations. Since history, and memory of past configurations is important, though, it is essential to exogenously include the variables that capture relevant information about joint history or intensity of collaborations to make this assumption more realistic (Steglich et al. 2010). It is also assumed that time runs continuously between observations, which implies that observed change is in fact assumed to be the result of an unobserved sequence of micro steps. Although this assumption is very realistic, it implies that coordination between a set of actors is not modeled. More precisely, at each micro-step, actors can change only one link variable at a time, inducing that a group of actors cannot decide to start relationships simultaneously. If we observed the formation of a closed triangle between  $i, j$  and  $h$  from one period to another, we assume for instance that  $i$  has interacted with  $j$ , then  $j$  with  $h$ , and then  $h$  with  $i$ . Third, and more importantly, it is assumed that network dynamics are based on actors’ choices depending on their preferences and constraints, i.e. the model is “actor-oriented”. This assumption is realistic for most economic

---

<sup>15</sup> For a discussion of these assumptions, see De Federico (2004), while for a summary the reader is referred to Snijders et al. (2010).

networks, and it allows including variables at a structural level, as well as also at a dyadic or individual level. Network structures change because actors develop strategies to create links with others (Jackson and Rogers 2007), which is based on their knowledge of the network configuration. This assumption is not plausible when actors are not able to elaborate their strategic decisions, or in the case where information about relationships of others is impossible to access.

*Modeling change opportunities.* SAOM are built upon the idea that actors can change their relations with other actors by deciding to create, maintain or dissolve links at stochastically determined moments. These opportunities are determined by the so-called rate function (Snijders et al. 2010b), and opportunities to change a link occur according to a Poisson process with rate  $\lambda_i$  for each actor  $i$ . In its simplest specification, all the actors have the same opportunity of change, i.e. equal to a constant parameter  $\lambda_i = p_m$ . In more complex models, heterogeneity in change opportunities can be introduced, in order to consider that actor characteristics (attributes or network positions) may considerably influence opportunities to change relationships. Thus, when individual attribute ( $v_i$ ) and degree ( $\sum_j x_{ij}$ ) are considered for instance, the rate function is given by the following logarithmic link function:

$$\lambda_i(x^0, v) = p_m \exp(\alpha_1 v_i + \alpha_2 \sum_j x_{ij}) \quad (6)$$

The set of permitted new states of the Markov chain, following on a current state  $x^0$ , is  $C(x^0)$  and the product of the two model components  $\lambda_i$  and  $p_i$  ( $p_i$  defines the probability distribution of choices, see equation 6) determines the transition rate matrix (Q-matrix) of which the elements are given by (Snijders 2008):

$$q_{x^0, x} = \lim_{dt \downarrow 0} \frac{P\{X(t+dt) = x | X(t) = x^0\}}{dt} \quad (7)$$

where  $q_{x^0, x} = 0$  whenever  $x_{ij} \neq x_{ij}^0$  for more than one element  $(i, j)$  and  $q_{x^0, x} = \lambda_i(x^0, v, w) p_i(x^0, x, v, w)$  for digraphs  $x$  and  $x^0$ , which differ from each other only in the element with index  $(i, j)$ .

Since the *rate function* sets the frequency of opportunities to change relationships, it models the speed of change of the dependent variable, i.e. network structures with high values implying strong dynamics.

*Modeling choice opportunities.* Given that an actor  $i$  has the opportunity to make a relational change, the choice for this actor is to change one of the link variables  $x_{ij}$ , because actors can only change one link variable at a time. Changing the link variables  $x_{ij}$  will lead to a new state  $x, x \in C(x^0)$ . In order to model choice probabilities, a traditional multinomial logistic regression specified by an objective function  $f_i$  is used (Snijders et al. 2010b):

$$P\{X(t) \text{ changes to } x | i \text{ has a change opportunity at time } t, X(t) = x^0\}$$



$$= p_i(x^0, x, v, w) = \frac{\exp(f_i(x^0, x, v, w))}{\sum_{x' \in C(x^0)} \exp(f_i(x^0, x', v, w))} \quad (8)$$

When actors have the opportunity to change their relations, they choose their partners by trying to maximize their objective function  $f_i$ . This objective function describes preferences and constraints of actors. More formally, collaboration choices are then determined by a linear combination of effects, depending on the current state ( $x^0$ ), the potential new state ( $x$ ), individual attributes ( $v$ ) and attributes at a dyadic level ( $w$ ). Effects related to the current state of the network are endogenous implying a self-reproduction of network structures, like transitive closure. Individual attributes are effects modeling the propensity of certain types of actors to form or to receive more linkages. Dyadic effects express the tendency of actors with similar attributes to form relationships, like actors that are located in the same region. Including these different types of effects, one can then disentangle the effect of geographical proximity from structural, individual other forms of proximity.

$$f_i(x^0, x, v, w) = \sum_k \beta_k s_{ki}(x^0, x, v, w) \quad (9)$$

*Parameter estimates.* The estimation of the different parameters  $\beta_k$  of the objective function is based on simulation procedures. More precisely, as proposed by Snijders (2001), the estimation of the effects  $\beta_k$  is achieved by the mean of an iterative Markov chain Monte Carlo algorithm based on the method of moments. The stochastic approximation algorithm simulates the evolution of the network and estimates the parameters  $\beta_k$  that minimize the deviation between observed and simulated networks. Over the iteration procedure, the provisional parameters of the probability model are progressively adjusted in a way that the simulated networks fit the observed networks. The parameter is then held constant to its final value, in order to evaluate the goodness of fit of the model and the standards errors. Lospinoso and Snijders (2011) provide detailed procedures to assess the goodness of fit. The different parameter estimates of SAOM can be interpreted as non-standardized coefficients obtained from logistic regression analysis (Steglich et al. 2010). Therefore, the parameter estimates are log-odds ratio, and they can be directly interpreted as how the log-odds of link formation change with one unit change in the corresponding independent variable.

## 6. Discussion

Above, we have briefly presented the four different statistical models and we now turn towards comparing their strengths and weaknesses. Moreover, we propose a guideline for the decision to use one model or another in empirical research. The guideline involves seven dimensions: **(I)** the type of relational data dealt with, **(II)** the type of network to be analyzed, **(III)** the size, **(IV)** the dynamic of the network, **(V)** the complex interplay between geography and networks, **(VI)** the main (independent) variables of interest and last but not least **(VII)** practical considerations.

**(I)** *The type of relational data to be analyzed.*

The first point concerns the difference between purely relational and network data. For the first, the independence assumption among links can safely be assumed to hold. For instance, one might argue that when analyzing cross-regional labor flows, the flow between region A and B is independent of the flow between regions B and C or C and D. The content of exchange between actors does not “travel” further than one link in some cases. For instance, while the “diffusion” of bread from the bakery store to the consumer represents a type of relational event, i.e. link, one might reasonably assume that there is no “network” of bread transmission because the consumer will eat the bread and no further transmission takes place. Hence, the object of exchange exclusively “travels” geodesic distances of length one. When such type of relational data is present, one obviously does not need to account for network dependencies and network autocorrelation implying that gravity models are the preferred empirical strategy.

**(II)** *The type of network to be analyzed.*

Two types of knowledge networks are often analyzed in economic geography: networks constructed from links between actors (firms, individuals) and networks constructed from links between geographical units (regions, countries). A fundamental assumption of SAOM (Markov models) is that nodes are actors, and that they control the formation of links, while GM-MRQAP are not built on this assumption. In the case of ERGM this issue is somewhat more complicated, as in general the modeling process seeks to mimic actors’ link formation behavior. However, no actor-based behavioral assumptions are necessarily required in the estimation (Park and Newman, 2004). In case of using SAOM to model knowledge networks between regions with Markov models would hence first require a discussion on the agency of the geographical unit or the reason for aggregating individual and organizational networks to a spatial level. GM and MRQAP do not impose these assumptions and hence, are better choices.<sup>16</sup> As pointed out above, ERGM are somewhat in between.

Related to this issue is whether the observed networks are one-mode or two-mode (bipartite) in nature. The observation of direct interactions between actors (cooperation, trading of goods, etc.) allows for constructing “standard” one-mode networks. In practice, however, so-called two-mode network data are more common. In their case, no direct interactions between actors are observed. Rather it is known that actors participate in the same event. Co-publication is a classic example in this respect. While it is frequently assumed that authors directly interact when writing a paper, all that is actually known is that they participate in the event of “writing a paper”. The actual contributions and interaction intensities remain unobserved and are frequently subject to heavy assumptions. This issue is far from trivial, as it means that on the basis of such assumptions, two-mode network data are commonly projected into one-mode data. However, this can strongly alter the structure of networks, as it

---

<sup>16</sup> See e.g. Liu et al. (2013) for an example of how GM can be used to model regional networks.

tends to increase the cliqueness of the network<sup>17</sup> (see for further discussions Opsahl 2013). If one wishes to avoid all problems associated to the one-mode projects, ERGM and SAOM<sup>18</sup> are preferred because they offer possibilities to directly handle two-mode network data (cf. Wang et al. 2009).

### (III) *Size of networks.*

Another, rather practical, issue is the size of the network of interest. While GM can be used to analyze large networks, MRQAP-ERGM-SAOM are computationally intensive and generally limited to a few thousands nodes.

### (IV) *Static or dynamic?*

From the above presentation, SAOM seem to be the natural choice for studying network dynamics, as they were the only approach introduced for dynamic network data. However, GM is frequently extended to deal with longitudinal relational data within a panel data setting. We refrain from discussing this approach in more detail as all arguments in favor or against its application in the cross-sectional case also apply to the longitudinal case.

This is somewhat different in the case of ERGM. Very recently, Hanneke and Xing (2007) and Desmarais and Cranmer (2011) put forward the so-called “temporal ERGM” (TERGM), which has been extended by Krivitsky and Handcock (2013) to the “separable temporal ERGM” (STERGM) model, which allows for considering longitudinal network data in the context of ERGM. STERGM is a fascinating new approach that brings the strength of ERGM to longitudinal network analysis. Krivitsky and Handcock (2013) formulate an ERGM for discrete-time network evolution by distinguishing between two processes: the first concerns factors that matter for rate of new link formation. The second process describes the duration of link existence. In essence a STERGM involves formulating two ERGM formulas. Both processes are assumed to be independent of each other within the same time step but might be dependent across time steps. While also two sets of parameters are obtained (one for the formation and one for the dissolution) the two processes are jointly estimated. This method seems to be very promising. However, sound information on how it compares (in particular in practice) to SAOM are still missing. In comparison to SAOM, STERGM particularly circumvents the assumption of actors controlling the formation (and dissolution) of links, which makes it particularly attractive when network nodes are territorial units and alike. However, at the heart of STERGM approach is the separability (of formation and dissolution) assumption, which might become problematic when time steps involve longer time periods or specific types of two-mode network data is used (see for a discussion: Krivitsky and

---

<sup>17</sup> As pointed out by one of the referee, it is possible to avoid complete cliquishness or to go beyond assuming symmetric ties in two mode networks if researchers have detailed data on the level of involvement/learning of actors in a given event.

<sup>18</sup> In SAOM, it is assumed that all agency ruling the dynamics of the network comes from the actors of the first mode of the two-mode network (Snijders et al. 2013). As a result, the second mode is passive and cannot decide to establish a link with the first mode. Besides, no coordination is possible between the first and second mode.

Goodreau, 2012). So far tests on the separability assumption are missing. Currently, STERGM also do require fixed node counts and node attributes. In light of these (in comparison to SOAM) shortcomings and the larger number of existing studies using SOAM, SOAM might still be the better choice when studying network dynamics in the field of Economic Geography. Given the tremendous speed of development in the according research areas, this recommendation needs to be regularly evaluated.

*(V) The main variables of interest.*

When analyzing the geography of knowledge networks, one of the main hypotheses to be tested is the impact of geographical distance on the formation of knowledge links between nodes. All four models can be used for this purpose, and it is also possible to test the influence of other forms of distance, since distance is a dyadic variable (an attribute of a pair of nodes). However, the models primarily differ in their possibilities to consider factors at the node and structural network level. The MRQAP is the most restricted model in this respect as it only allows considering dyad level variables. This means that factors at the node and structural network level can be incorporated only if translated into dyad level factors. For instance, it might be interesting to test the impact of the regions' sizes on the structure of a network. In a MRQAP model, it will be tested if the probability of two large regions being linked is higher than that of two small regions. Such is similar but still distinct from an argument at the node-level, which might rather be that large regions are generally better embedded. In contrast to the MRQAP, such node-level factors can directly be included in GM, ERGM, and SAOM.

However, only ERGM and SAOM are able to simultaneously incorporate node, dyad, and structural network level factors. In order to include factors at the structural network level in MRQAP and Gravity models, these need to be translated to the node or dyad level. An example could be triadic closure. Triadic closure implies that a link between region A and B is more likely if both are also linked to region C. Translating such to the dyadic level is often impossible, even in an approximate fashion. Triadic closure is a good example in this respect because its dyadic representation would have to be based on the dependent variable (the existence of a link between A and C as well as B and C), which raises serious concerns regarding the independence of the independent variables.

Tendencies towards triadic closure and multi-connectivity are frequently argued to be relevant to explain the structure of inter-organizational networks in economic geography (see, e.g., Glückler 2007, Ter Wal 2013). This clearly favors the application of ERGM and SAOM with their abilities to explicitly consider these structural dependencies.

In addition, if the dependent variables concern simultaneously the structure of a network and a node attribute (innovation performance for instance) and longitudinal network data is available, SAOM are to be used because they offer a co-evolution model (to deal with the causality issues between network structure and node attribute).

*(VI) More complex interplay between geography and networks*

Some recent features of these network models can be exploited to better understand the complex interplay between geography and networks. Especially in the case of SAOM, it is possible to separate partner selection from social influence, which is a key question in social science more broadly (Van der Leij 2011). In a geographical context, it means that SAOM offers the opportunity to understand whether actors co-locate (dynamics of geographical proximity) because they already have knowledge ties (or if they start to build relationships because they are already spatially close (network dynamics)). SAOM therefore allows analyzing the co-evolutionary dynamics between geography and networks.

The empirical strategy requires an important level of dynamics both in terms of spatial choices and network ties. We would have to favor a disaggregated level of analysis where actors are spatially mobile (engineers, scientists movements rather than location choices for firms' headquarter). Instead of looking at the (dyadic) physical distance among actors, it is possible to represent choices of actors as a bimodal network (when actor  $i$  move to city  $C$  we draw a spatial ties between  $i$  and  $C$ ), and relation choices as a traditional one-mode network (between  $i$  and  $j$ ). This idea fits with the recent statistical framework proposed by Snijders et al. (2013) and SAOM can be used to analyze the co-evolution of the (spatial) two-mode network and the (knowledge) one-mode networks to disentangle the effects of selection and influence. Similar seems to be possible in light of the new developments in ERGM techniques (TERGM, STERGM). However, these methods still require fixed node counts and node attributes, which do not allow for analyzing the co-evolution of nodes and networks.

The Markov random graphs used in ERGM moreover imply that links between two nodes are assumed to be contingent on their links to other nodes (conditional dependence). Accordingly, ERGM are very appropriate for modeling, for instance, processes involving gatekeeper organizations whose attractiveness as collaboration partner is primarily caused by their (specific) links to other (region external) organizations (cf. Graf, 2010). However, in these models it is also possible to drop all link dependencies and rather assume links being independent of each other. In this case, the model is based on Bernoulli graphs. The empirical strategies presented by Anderson and Van Wincoop (2003) and Behrens et al. (2012) for GM also allow researchers to deal with structural dependencies. Here, especially the approach of Behrens et al. (2012), which uses spatial econometric techniques to account for cross-sectional correlations between flows and cross-sectional correlations in the error terms seems to be highly promising. However, future research is needed to evaluate the appropriateness of this empirical strategy. These aspects represent just some aspects that are possible with advanced models like SAOM, ERGM, and GM using spatial econometric techniques, which will surely be exploited in more detail in future studies in the field.

## **(VII) Practical considerations.**

The greater applicability and power of ERGM and SAOM, and to a lesser extent GM, comes at a price of complexity. In this respect the MRQAP has the advantage of its “*simplicity and accessibility*” (Dekker et al. 2007 p. 564). Moreover, while being advanced in many ways,

ERGM and SAOM are still limited in seemingly simple issues. For instance, SAOM only account for valued links using multiple dependent networks of a binary nature. For ERGM only recently extensions have been put forward that also allow considering valued network data (see, e.g., Krivitsky 2012), which are by now also available in specific software packages.

Another example in this respect is the way researchers can identify the most accurate model. GM and MRQAP offer a wide range of goodness-of-fit statistics. In addition, they can easily be compared across varying parameter specifications and sets of considered independent variables. Despite recent efforts in this direction, similar cannot be said about ERGM and SAOM without restrictions. For instance, ERGMs offer goodness-of-fit, AIC, and BIC statistics, which provide a good assessment of the final model's quality, i.e. the model which converges and represents the observed network well. However, they are of limited value in the process of finding the best parameter combinations and set of explanatory factors to be considered. This is particularly the case when node and dyadic factors as well as structural dependencies correlate, and when (for some reason), the model fail to converge. In this case, researchers have to rely on a manual iterative trial-and-error process of estimating varying model specifications. Given the procedure's substantial computational requirements (in particular in cases of large networks) this frequently turns out to be very cumbersome. The problem is moreover intensified by the necessity of specifying decay parameters for certain structural dependencies such as the geometrically weighted shared partner statistic (see, e.g., Snijders et al. 2006), which can be used to evaluate the importance of triadic closure. The recent development of curved exponential family models may provide some relieve to the latter issue (Hunter, 2007).

Table 1a-b summarizes the main characteristics of the four network modeling strategies, indicating the respective strengths and weaknesses and suggesting the degree of applicability in geography of innovation studies. We also name some software packages that include the according procedures. This list does not however make a claim to be complete.

## 7. Conclusion

We have discussed the scientific roots and the working principles of four main statistical models that are increasingly used to analyze and explicitly model the geography of knowledge networks: gravity models (GM), multiple regression quadratic assignment procedure (MRQAP), exponential random graph models (ERGM), and stochastic actor-oriented models (SAOM). All four research strategies and models turn out to have advantages and disadvantages. They are embedded in their respective epistemic communities of practice – explaining why four varying modeling techniques exist next to each other. GM and MRQAP are more conventional modeling types that require less information and computational efforts – but consequently exploit also less information on the structure of networks than ERGM and SAOM. Also, GM and MRQAP are only implicitly linked to knowledge transmission mechanisms, although gravity modeling (GM) develops into more advanced network analysis applying individual level effects and using spatial econometric techniques. These differences

influence their range of applicability in economic geography. We also derive a guideline helping researchers in this field to decide which model to use in what situations.

All four modeling types bring geography and network structures together in explaining the web of knowledge links between nodes (actors or regions). The reverse relationship – network positions partly determining local and regional development opportunities – constitutes an interesting research direction in economic geography. In such a conceptualization, local development is determined by hub positions in key networks next to hotspot characteristics in regions. Although the models discussed mainly focus in bringing (exogenous) geography into (endogenous) networks, analyzing endogenous regional development from (also endogenous) networks seems to be a major challenge for future research. For this relationship to be studied, additional (spatial) econometric modeling techniques are required, in which networked interactions between regions and actors function as carriers of knowledge and innovation diffusion.

<b>Characteristics</b>	<b>Gravity Models (GM)</b>	<b>Quadratic Assignment Procedure (QAP)</b>	<b>Exponential Random Graph Models (ERGM)</b>	<b>Stochastic Actor-Oriented Models (SAOM)</b>
<b>Error type 1 (underestimation of standard errors)</b>	Can be corrected	Corrected	Corrected	Corrected
<b>Structural dependencies (i.e. triadic closure)</b>	Can be controlled for to some extent, but additional econometric modeling needed	Not modeled	Modeled	Modeled
<b>Proximity (geographical or other dyadic variables)</b>	Available	Available	Available	Available
<b>Node level variables (i.e. R&amp;D expenditures)</b>	Available	Not available	Available	Available
<b>Computational time</b>	Low	High	High	High
<b>Flexibility of assumptions</b>	High	High	High	Low
<b>Type of nodes</b>	All	All	All	Actors (firms, individuals...)
<b>Distribution assumption</b>	Parametric	Non Parametric	Parametric	Parametric
<b>Type of <math>\beta</math> (interpretation of estimated coefficients)</b>	Flexible, large range of interpretation possible	Flexible, large range of interpretation possible	Non standardized Log-odds ratio	Non standardized Log-odds ratio
<b>Type of data</b>	Cross sectional and longitudinal	Cross sectional	Cross sectional and longitudinal	Longitudinal

**Table 1a: Statistical models for the formation of knowledge networks in economic geography**



<b>Characteristics</b>	<b>Gravity Models (GM)</b>	<b>Quadratic Assignment Procedure (QAP)</b>	<b>Exponential Random Graph Models (ERGM)</b>	<b>Stochastic Actor-Oriented Models (SAOM)</b>
<b>Type of relations</b>	One-mode	One-mode	One and two-mode	One and two-mode
<b>Goodness of fit</b>	Available	Available	Available	Available
<b>Multiple dependent variables (node attribute, other network)</b>	Not available	Not available	Not available, however two networks as dependent variables possible	Available
<b>Statistical Programs</b>	Stata, SPSS, SAS, R	Stata, Ucinet, R	R, PNET (and variants of PNET)	R, Stocnet
<b>Valued Networks</b>	Available	Available	Available	Available
<b>Directed Networks</b>	Available	Available	Available	Available
<b>Conceptualizing knowledge transfer</b>	Implicit	Implicit, due to lack of individual effects modeling	Explicit, modeling of access to external knowledge	Explicit, modeling of access to external knowledge
<b>Geography &gt; Network interaction</b>	Suitable, static and increasingly dynamic	Suitable, static	Suitable, static, (STERGM for dynamics)	Suitable, dynamic
<b>Network &gt; regional development interaction</b>	Limited possibility, additional econometric modeling needed	Limited possibility, additional econometric modeling needed	Possible, additional econometric modeling needed	Can be implemented at the actor level, in a co-evolution model
Note that some options may constrain the availability of others. For instance, currently ERGM can only model two-mode networks for cross-sectional data. Similar holds for other modeling options and other models as well.				

**Table 1b: Statistical models for the formation of knowledge networks in economic geography**

## References

- Almeida P, Kogut B (1999) Localization of knowledge and the mobility of engineers in regional networks. *Management science* 45: 905-917
- Autant-Bernard C, Billand P, Frachisse D, Massard N (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: 495-519
- Balland P A (2012) Proximity and the Evolution of Collaboration Networks: Evidence from Research and Development Projects within the Global Navigation Satellite System (GNSS) Industry. *Regional Studies* 46: 741-756
- Balland P A, de Vaan M, Boschma R (2013) The Dynamics of Interfirm Networks along the Industry Life Cycle: The Case of the Global Video Games Industry 1987-2007. *Journal of Economic Geography*, forthcoming
- Barca F, McCann P and Rodriguez-Pose A. (2012) The case for regional development intervention: place-based versus place-neutral approaches, *Journal of Regional Science* 52: 134-152.
- Basile R, Capello R and Caragliu A. (2012) Technological interdependence and regional growth in Europe: Proximity and synergy in knowledge spillovers, *Papers in Regional Science* 91: 697-722.
- Behrens K, Ertur C, Koch W (2012) Dual' gravity: using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics*, 27(5): 773-794.
- Bell, G G (2005) Clusters, networks, and firm innovativeness. *Strategic Management Journal* 24:287-295
- Besag J (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 36(2):192-236
- Besag J (1975) Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician)* 24:179-95.
- Bikker, J.A. (2010) An extended gravity model with substitution applied to international trade flows. In: P.A.G. van Bergeijk and S. Brakman (eds.) *The Gravity Model in International Trade: Advances and Applications*. Cambridge University Press, pp. 135-162.
- Boschma R (2005) Proximity and innovation: a critical assessment. *Regional Studies* 39: 61-74
- Boschma R, Balland P A, Kogler D (2011) A relational approach to knowledge spillovers in biotech. Network structures as drivers of inter-organizational citation patterns, *Papers in Evolutionary Economic Geography* 11.20, Utrecht University, Utrecht
- Boschma R, Frenken K (2006) Why is economic geography not an evolutionary science? Towards an evolutionary economic geography, *Journal of Economic Geography*, 6: 273-302.
- Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9:439-468
- Bröcker J (1989) Partial equilibrium theory of interregional trade and the gravity model. *Papers of the Regional Science Association* 66: 7-18

- Broekel T, Hartog M (2013a) Explaining the structure of inter-organizational networks using exponential random graph models: does proximity matter? *Industry and Innovation*, 20(3):277-295
- Broekel T, Hartog M (2013b) Determinants of cross-regional R and D collaboration networks: an application of exponential random graph models. *Working Papers on Innovation and Space*, #13.4
- Broekel T, Boschma R A (2012) Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography* 12: 409–433
- Burger M J, Van Oort F G, Frenken K, Van der Knaap B. (2009a) Networks and economic agglomerations: introduction to the special issue. *Tijdschrift voor Economische en Sociale Geografie* 100: 139-144
- Burger M J, Van Oort F G, Linders G J M (2009b) On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis* 4: 167-190.
- Carey H C (1858) *Principles of Social Science*. Philadelphia: Lippincott.
- Castellani D, Palmero A J, Zanfei A (2012) The gravity of R&D FDI. Working paper, University of Urbino
- Cranmer S J, Desmarais B A (2011) Inferential Network Analysis with Exponential Random Graph Models. *Political Analysis* 19:66-86
- Daraganova G, Pattison P, Koskinen J, Mitchell B, Bill A, Watts M, Baum S (2012) Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. *Social Networks* 34, 6-17.
- De Federico A (2004) L'analyse longitudinal de réseaux sociaux totaux avec SIENA: méthode, discussion et application, *Bulletin de Méthodologie Sociologique* 84, 5–39
- De Groot, H L F, Poot J, Smit M J (2009) Agglomeration externalities, innovation and regional growth: theoretical perspectives and meta-analysis. In : Capello, R and Nijkamp, P (eds) *Handbook of Regional Growth and Development Theories*. Edward Elgar, Cheltenham, UK.
- Dekker D, Krackhardt D, Snijders T A B (2007) Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika* 72: 564-581
- Egger P (2004) On the Problem of Endogenous Unobserved Effects in the Estimation of Gravity Models. *Journal of Economic Integration* 19: 182-191
- Erdős P, Rényi A (1960) On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5:17–61
- Feenstra R C (2004) *Advanced International Trade: Theory and Evidence*, Princeton University Press, Princeton.
- Fienberg S, Wasserman S (1981) An exponential family of probability distributions for directed graphs: comment. *Journal of the American Statistical Association* 76:54-57.
- Fischer M M, Griffith D A (2008) Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union. *Journal of Regional Science* 48: 969-989

- Fischer M M, Scherngell T, Jansberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe: Evidence from a spatial interaction modeling perspective. *Geographical Analysis* 38: 288-309
- Fowler J H, Dawes, C T, Chistakis N A (2009) Model of Genetic Variation in Human Social Networks. *Proceedings of the National Academy of Sciences* 206: 1720-1724
- Giuliani E (2010) Network Dynamics in Regional Clusters: The Perspective of an Emerging Economy, *Papers in Evolutionary Economic Geography* 10-14, Section of Economic Geography, Utrecht University
- Glückler J (2007) Economic Geography and the Evolution of Networks. *Journal of Economic Geography* 7: 619 - 634
- 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: 163-189
- Graf, H., (2010) Gatekeepers in Regional Networks of Innovation. *Cambridge Journal of Economics*, online account (doi:10.1093/cje/beq001), pp.173–198.
- Hagedoorn J (2002) Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research Policy* 31: 477-492
- Hanneke S, Xing E P (2007) Discrete temporal models of social networks. In *Statistical network analysis: Models, issues, and new directions*. Vol. 4503 of *Lecture Notes in Computer Science*, 115–125. Berlin, Germany: Springer
- Hazir C S, Autant-Bernard C (2012) Using Affiliation Networks to Study the Determinants of Multilateral Research Cooperation: Some Empirical Evidence from EU Framework Programs in Biotechnology. GATE Working Paper No. 1212
- Head K, Mayer T (2000) Non-Europe: the magnitude and causes of market fragmentation in the EU. *Review of World Economics* 136: 285-314.
- Helpman E, Melitz M, Rubinstein Y (2008) Estimating trade flows: trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2): 441-487.
- Hoekman J, Frenken K, Van Oort F G (2009) The geography of collaborative knowledge production in Europe. *Annals of Regional Science* 43: 721-738
- Holland P W, Leinhardt S (1981) An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association* 76: 33-65
- Hubert L J (1987) *Assignment Methods in Combinatorial Data Analysis*. New York: Marcel Dekker
- Hubert L J, Golledge R G (1981) A heuristic method for the comparison of related structures. *Journal of Mathematical Psychology* 23:214-226
- Hunter D R (2007) Curved exponential family models for social networks. *Social Networks* 29:216–230
- Hunter D R, Goodreau S M, Handcock M S (2008) Goodness of Fit for Social Network Models. *Journal of the American Statistical Association* 103: 248-258
- Isard W (1956) *Location and Space-Economy*. Cambridge, MA: MIT Press
- Jackson M O, Rogers B W (2007) Meeting strangers and friends of friends: how random are social networks?, *American Economic Review* 97, 890–915

- Krackhardt D (1987) QAP partialling as a test of spuriousness. *Social Networks*, 9:171-186
- Krackhardt D (1988) Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10:359- 381
- Krivitsky P N (2012) Exponential-family random graph models for valued networks, *Electronic Journal of Statistics*, 6: 1100-1128
- Krivitsky P N, Goodrea S M (2012): STERGM – Separable Temporal ERGMs for modeling discrete relational dynamics with statnet, STERGM Tutorial: <https://statnet.csde.washington.edu/trac/raw-attachment/wiki/Resources/STERGMtutorial.pdf> (21.11.2013)
- Krivitsky P N, Handcock M S (2013) A separable model for dynamic networks. *Journal of the Royal Statistical Society B* (2013): 75(4)
- LeSage J P, Pace R K (2008) Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48: 941-967
- Lill E (1891) *Das Reisegesetz und seine Anwendung auf den Eisenbahnverkehr*, Vienna
- Linders, G-J, Patuelli R, Griffith, D.A. (2010) The space of gravity: spatial filtering estimation of a gravity model for bilateral trade. Mimeo
- Lindgren K O (2010) Dyadic regression in the presence of heteroscedasticity – an assessment of alternative approaches, *Social Networks*, 32, 279-289
- Liu W H, Lin Y C (2005) Foreign patent rights and high-tech exports: evidence from Taiwan. *Applied Economics*, 37:1534-1555
- Liu X, Derudder B, Liu Y (2013) Regional geographies of intercity corporate networks: The use of exponential random graph models to assess regional network-formation, *Papers in Regional Science* (forthcoming).
- Lospinoso J A, Snijders T A B (2011) Goodness of fit for stochastic actor oriented models. Working paper
- Lubbers M J, Snijders T A B (2007) A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, 29, 489-507
- Lukermann F, Porter P W (1960) Gravity and potential models in economic geography. *Annals of the Association of American Geographers*, 50: 493-504
- Maggioni M A, Uberti T E (2007) Inter-regional knowledge flows in Europe: An econometric analysis. In: Frenken K (ed.) *Applied evolutionary economics and economic geography*, Edward Elgar, Cheltenham, UK
- Maggioni M, Uberti E (2011) Networks and geography in the economics of knowledge flows. *Quality and Quantity* 45, forthcoming
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2 Part 1):209–220
- Mantel N, Valand R S (1970) A technique of nonparametric multivariate analysis. *Biometrics*, 26:547–558
- Melo, P C, Graham D J, Noland, R B (2009) A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39: 332–342.

- Newman M, Barabási A L, Watts D J (2006) *The structure and dynamics of networks*. Princeton University Press, Princeton, USA
- Opsahl T (2013) Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks* 35, doi: 10.1016/j.socnet.2011.07.001
- Park J, Newman M E J (2004) *The Statistical Mechanics of Networks*. *Physical Review E*, 70(6)
- Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks. II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology* 52:169-194
- Peri G (2005) Determinants of knowledge flows and their effect on innovation. *The Review of Economics and Statistics*, 87: 308-322
- Ponds R, Van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration, *Papers in Regional Science*, 86: 423-444.
- Ravenstein E G (1885) The laws of migration. *Journal of the Statistical Society* 48:167-235
- Ripley R M, Snijders T A B, Preciado Lopez P (2011) *Manual for RSiena*, University of Oxford, Department of Statistics; Nuffield College, October 7
- Robins G, Morris M (2007) Advances in exponential random graph ( $p^*$ ) models. *Social Networks* 29:169-172
- Saul Z M, Filkov V (2007) Exploring biological network structure using exponential random graph models. *Bioinformatics* 23 (19), pp. 2604- 2611
- Scherngell T, Barber M J (2009) Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU Framework Programme. *Papers in Regional Science*, 88: 531-546
- Scherngell T, Lata, R (2013) Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. *Papers in Regional Science*, 92: 555-577.
- Schneider J W, Borlund P (2007) Matrix comparison, part 2: Measuring the resemblance between proximity measures or ordination results by use of the mantel and procrustes statistics. *Journal of the American Society for Information Science and Technology*, 58(11):1596–1609
- Snijders T A B (2001) The statistical evaluation of social network dynamics, In: Sobel, M., Becker M (Eds.), *Sociological Methodology*, Basil Blackwell, Boston and London, 361–395.
- Snijders T A B (2002) Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure* 3 (2)
- Snijders T A B (2008) Longitudinal methods of network analysis, in MEYERS B. (Ed.) *Encyclopedia of Complexity and System Science*, Springer, Berlin
- Snijders T A B (2010) *Actor-Based Models for Network Dynamics*, Working Paper.
- Snijders T A B (2011) Statistical models for social networks. *Annual Review of Sociology*, 37: 131-153.
- Snijders T A B, van de Bunt G, Steglich C (2010a) Introduction to stochastic actor-based models for network dynamics. *Social Networks* 32 (1):44-60
- Snijders T A B, Steglich C, Schweinberger M, Huisman M (2010b) *Manual for SIENA version 3.2*. Provisional version. Oxford: University of Oxford, Department of Statistics

- Snijders T A B, Pattison P E, Robins G, Handcock M S (2006) New specifications for exponential random graph models, *Sociological Methodology*, 36, pp. 99-153
- Snijders, T A B, Lomi A, Torló V J (2013) A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social networks* 35(2): 265-276.
- Solomonoff R, Rapoport A (1951) Connectivity of random nets. *Bulletin of Mathematical Biophysics*, 13:107-117
- Steglich C E G, Snijders, T A B, West P (2006) Applying SIENA: An illustrative analysis of the co-evolution of adolescents' friendship networks, taste in music, and alcohol consumption. *Methodology*, 2:48-56
- Steglich C, Snijders T A B, Pearson M (2010) Dynamic networks and behavior: separating selection from influence, *Sociological Methodology* 40:329-393
- Stewart J Q (1948) Demographic gravitation: evidence and application. *Sociometry* 1: 31-58.
- Stouffer S A (1940) Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5, 845-867.
- Sunley P (2008) Relational economic geography: a partial understanding or a new paradigm? *Economic Geography*, 84: 1-26
- Ter Wal A L J (2013) The spatial dynamics of the inventor network in German biotechnology: Geographical proximity versus triadic closure, *Journal of Economic Geography*, forthcoming.
- Ter Wal A., Boschma R (2009) Applying social network analysis in economic geography: theoretical and methodological issues. *The Annals of Regional Science*, 43:739-756.
- Thissen M, Van Oort F, Diodato D and Ruijs A (2013) Regional competitiveness and smart specialization in Europe. *Place-based development in international economic networks*. Cheltenham: Edward Elgar.
- Tinbergen J (1962) *Shaping the World Economy*. New York: Twentieth Century Fund.
- Van de Bunt G, Groenewegen P (2007) An actor-oriented dynamic network approach: The case of interorganizational network evolution, *Organizational Research Methods* 10(3):463
- Van der Leij M J (2011) Experimenting with buddies. *Science* 334: 1220-1221.
- Van Duijn M A J, Gile K J, 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):52-62.
- Van Oort F, Lambooy J (2013) Cities, knowledge and innovation. Chapter in: M. Fischer and P. Nijkamp (eds.), *Handbook of Regional Science*. Berlin: Springer, p. 475-488
- Wang, P., Sharpe, K., Robins, G., and Pattison, P. (2009) Exponential random graph ( $p^*$ ) models for affiliation networks. *Social Networks*, 31(1):12-25.
- Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and  $p^*$ . *Psychometrika* 61:401-425