Focused model selection for social networks

Pircalabelu E, Claeskens G.

KBI_1607

Focused model selection for social networks

Eugen Pircalabelu[1], Gerda Claeskens

ORSTAT and Leuven Statistics Research Center, KU Leuven
Naamsestraat 69, 3000 Leuven, Belgium
eugen.pircalabelu@kuleuven.be, gerda.claeskens@kuleuven.be

Abstract

We present a focused selection method for social networks. The procedure is driven by a *focus*, the main quantity we want to estimate well. It represents the statistical translation of a research hypothesis into parameters of interest. Given a collection of models, the procedure estimates for each model the mean squared error of the estimator of the focus. The model with the smallest such value is selected. We present focused model selection for (i) exponential random graph models, (ii) network autocorrelation models and (iii) network regression models to investigate existing relations in social networks. Worked-out examples illustrate the methodology.

Key words: Variable selection; Social network; Focused information criterion; Exponential random graphs; Network based models

1 Motivation

Social network analysis aims at understanding and explaining regularities and structures that describe relations linking individuals or any other social units such as organizations, political parties, etc. We present a methodology for model selection in the context of network based parameter estimation that is based on the *focused information criterion* (FIC) introduced and studied under various statistical contexts in the works of Claeskens and Hjort (2003, 2008a), Hjort and Claeskens (2006), Zhang and Liang (2011), Rohan and Ramanathan (2011), Claeskens (2012), Behl et al. (2014) among others. More recently in the works of Pircalabelu et al. (2015a,b) the FIC has been applied to estimate probabilistic graphical models. The goal of the present manuscript is to extend the application of FIC to social network models.

In the focused selection procedure, the focus, which is a function of the model parameters, plays a central role. Throughout the paper we denote the focus by μ. The focused information criterion is constructed to select from a set of models that model where the focus is optimally estimated in terms of *mean squared error* (MSE), which is the sum of the estimator's variance and its squared bias. The FIC selection procedure makes it possible to select explanatory models for focuses especially suited for social network analysis. Often, researchers are not interested in the whole network, but rather in quantities that summarize or describe phenomena such as actor centrality, edge prediction or strength of interpersonal influence between actors. For these quantities of interest (i.e., focuses) that can be formulated as functions of parameters of the underlying model, the FIC may be used to select models that estimate those focuses with small MSE. Since the true MSE is in general unknown, it needs to be estimated. The FIC value is such an estimated MSE, sometimes modulo some

[1]Corresponding author

constants that do not depend on the models. See Section 4 for more details on the relation between FIC and MSE. First, we specify the focus and a list of plausible explanatory models. Next, we estimate the focus parameter and its associated MSE (or FIC) value. Finally, we select as chosen model for this focus that model of the list which has the smallest FIC value.

Unlike the classical information criteria, such as Akaike's information criterion (AIC, Akaike, 1973) and the Schwarz Bayesian information criterion (BIC, Schwarz, 1978) the focused information criterion allows to select a model that is *directed* towards the particular focus. That is, the FIC will select a model that performs well in MSE sense to the estimation of the focus, the quantity of interest. Different focuses, thus different interests, might lead to different models being selected, which is more informative since the selected model is determined based on specific research interests.

To motivate the focused selection approach for model selection for social networks we use three model classes, namely the exponential random graph models (ERGM), network autocorrelation models (NAM) and network regression models (NRM). See Section 2 for details regarding model specifications.

We start from the 'Florentine families' dataset (Breiger and Pattison, 1986; Padgett, 1994) which consists of a social network that records the marriage ties between 16 influential Florentine families (which family is linked with which other family), a social network that records the business ties between the 16 families, the wealth of each family, the number of seats on the civic council for each family, and the total number of ties linking the family to any of the other 116 families from Florence. The networks are represented by two adjacency matrices containing the value 1 on positions (i, j) and (j, i) if members from family i have married (or have business ties with) members from family j. A value of 1 is translated into a tie (edge) between families i and j to denote that the two families have marriage or business ties linking them together. Figure 1 shows both the marriage and business ties that link together the Florentine families.

The data represent the ties formed around 1430, a period when the Strozzi and the Medici were considered to be adversaries. It was a period where the Medici family was powerful (Padgett and Ansell, 1993) and an alliance through marriage between the two families would have been improbable. In 1434 the Strozzi family have been driven into exile by Cosimo

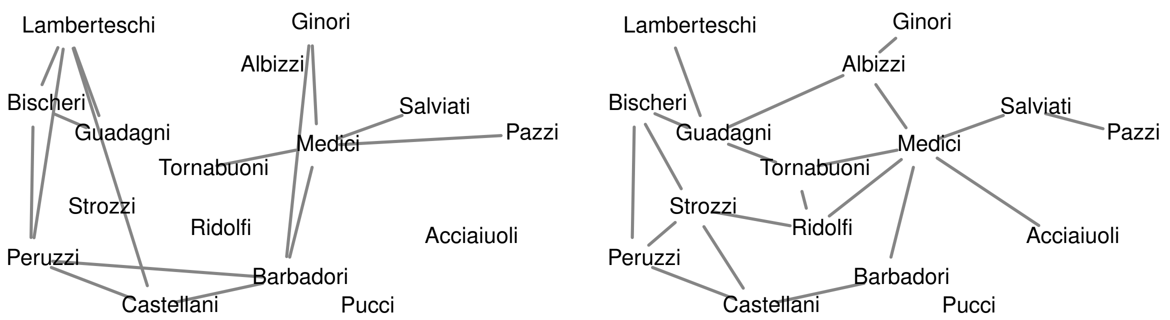

Figure 1: Florentine families data. The left panel displays the business network while the right panel displays the marriage ties between the families.

de' Medici, but due to the radical change in the political context in Florence, in 1508 the two families eventually formed an alliance through marriage (Bullard, 1979). The observed marriage network which is analyzed here does not however, have such a tie present, since it pertains to the period around 1430.

In light of this information, it is of interest to select explanatory models that best express the log odds ratio of such a tie being formed in 1430 between the two families. The log odds ratio is our first *focus* parameter. To estimate this focus, we use an exponential random graph model (ERGM) described in Section 2. A list of $32 = 2^5$ (all possible combinations of predictors) such models was considered, where the narrow, simplest model contained only one parameter, the edges parameter which acted similar to an intercept, and the most complex (full) model contained four extra parameters. All other models, were in between the simplest and the full model. We postpone the proper definition of the predictors to Section 5 as it suffices here to show what the method offers in practice. The aim is to select the suitable collection of parameters that provides the lowest FIC value for the estimated focus, in this case, the log odds ratio. Table 1 shows the two best ranking models using FIC, as well as the full and the narrow model (the latter which coincides with the best scoring AIC and BIC models). The best scoring FIC model to estimate the log odds ratio contains as predictor the wealth of the Florentine families, whereas the second best model suggests adding also the change statistic with respect to the number of triangle configurations and the Gwesp summary statistic (see Section 5). The AIC/BIC selected model (which is insensitive to the focus specification) suggests the usage of the simplest, narrow model without any additional predictors.

Model	$\Delta_{Kstar(2)}$	$\Delta_{Kstar(3)}$	$\Delta_{Triangle}$	$\Delta_{Gwesp(\tau\,=\,.5)}$	*Wealth*	$\widehat{\mu}$	FIC	AIC	BIC
Best FIC	0	0	0	0	1	-1.80	1.64	111.5	117.0
2nd FIC	0	0	1	1	1	-2.24	2.03	115.0	126.1
Full	1	1	1	1	1	-2.22	3.18	119.0	135.7
Best AIC/BIC	0	0	0	0	0	-1.62	7.12	110.1	112.9

Table 1: Florentine family data. ERGM model selection using the focus $\mu = \log(\frac{p_{ij}}{1-p_{ij}})$ where p_{ij} is the probability of a tie occurring between the Medici and Strozzi families. The estimated focus and the FIC, AIC and BIC scores are presented for four different models. The values 0/1 in the columns $\Delta_{Kstar(2)}$– *Wealth* indicate the absence/presence of a predictor in the model. For all models an 'edges' parameter $\Delta_{Kstar(1)}$ is included, by default.

With the FIC we can easily change the focus of the analysis. Using the same ERGM class of models and the same list of potential models, but focusing now on the parameter associated with the triangle predictor (this is the second focus) instead of the log odds ratio, we obtain a different ranking of the models as shown in Table 2. The transitive triangle is an important summary measure for social networks, because it expresses the inclination for actors to form homogenous groups. If actor a has a tie with actor b and b has a tie with c, then under a transitive triangle assumption also a and c will be connected, implying that 'friends of friends are also friends', and as such this summary measure is an important one when describing social networks. When selecting a model that minimizes the MSE expression for the triangle parameter, the best FIC model contains as predictors the change in the transitive triangles statistic, the change in the Gwesp statistic and the wealth of the families showing that different focuses (which embody different research questions) might need different explanatory models.

Model	$\Delta_{Kstar(2)}$	$\Delta_{Kstar(3)}$	$\Delta_{Triangle}$	$\Delta_{Gwesp(\tau=.5)}$	*Wealth*	$\widehat{\mu}$	FIC	AIC	BIC
Best FIC	0	0	1	1	1	-4.62	50.63	115.0	126.1
2nd FIC	0	0	1	1	0	-4.44	51.10	113.6	122.0
Full	1	1	1	1	1	-4.64	53.34	119.0	135.7
Best AIC/BIC	0	0	0	0	0	0.00	1644.08	110.1	112.9

Table 2: Florentine family data. ERGM model selection using the parameter associated with the triangle predictor as focus. The estimated focus and the FIC, AIC and BIC scores are presented for four different models. The values 0/1 in the columns $\Delta_{Kstar(2)}$– *Wealth* indicate the absence/presence of a predictor in the model. For all models an 'edges' parameter $\Delta_{Kstar(1)}$ is included, by default.

We now mention some related research. Model (or parameter) selection for social networks is often performed by formal hypothesis testing as in Anderson et al. (1999), Leenders (2002), Robins et al. (2007), by assessing goodness of fit measures as in Goodreau (2007), Hunter et al. (2008a), Wang et al. (2013a), Wang et al. (2013b) or Shore and Lubin (2015), and by using information criteria as in Leenders (2002), Goodreau (2007), Hunter et al. (2008a), Stadtfeld et al. (2011) and Austin et al. (2013) among many other references. Saito et al. (2010) proposed model selection based on average Kullback-Leibler divergence. Bayesian model selection can be found in Koskinen (2004a,b), Zijlstra et al. (2005), Rodríguez (2012) and Caimo and Friel (2013).

2 Social network models

Of all social sciences, sociology and anthropology have been at the forefront of social network analysis, due to the ease with which studying small communities and the interaction between its members, can be reflected to a certain degree by graphical objects. The consequence of such graph oriented representation is that by using basic properties and notions developed for graphs, one can now describe, summarize and also quantify social relations.

A social network consists of a set of units (represented graphically by a set of nodes, one node per unit) and the social connections that exist between the units. Most often the complex relation between the units is reduced to a 'presence or absence' decision, although sometimes one may reduce it to a number that reflects in a way the intensity of the relationship rather than a crude presence/absence representation.

The types of social network models, as summarized nicely in O'Malley and Marsden (2008) vary in complexity and flexibility and reflect different research interests. We use the following three types of models.

(i) Exponential random graph models (ERGM), see Holland and Leinhardt (1981), and Wasserman and Pattison (1996). Local structures in the form of meaningful subgraphs model the global structure of the network. For example, one may use the propensity of forming a triadic configuration (unit i connects with units j and k, and as a transitive result also j and k connect) as a predictor for modeling marriage ties among families. To illustrate the method we have used the 'Florentine marriage' network.

(ii) 'Network autocorelation models' (NAM), see Dow et al. (1982); Leenders (2002). A variable measured for each actor is modeled as a function of other explanatory variables and using the assumption that the underlying errors have a special correlation structure formed as part of an interpersonal influence which is reflected by a social network. For example, suppose one is modeling the wealth of the Florentine families as a function of socio-economic indicators, but since marriage opportunities were intricately linked to wealth, the researcher opts to take this into account by directly using the marriage ties between the families into the model. We use here the 'Florentine marriage' and the 'Florentine business' networks.

(iii) 'Network regression models' (NRM), see Krackhardt (1988). The observed ties in one network are used to predict the ties in another network. For example, a researcher models the exit orders of monks leaving a convent as a function of their influence on one another. To illustrate the method we have used the 'Monastery' data of Sampson (1968).

A description of the available implementations of the above methods is given in Butts (2008a); Hunter et al. (2008b) and Handcock et al. (2008).

As can be seen from the examples given above, each of these models serves a different explanatory purpose and answers related, but different questions. Similar for all three models is the fact that they can all be related to classical, likelihood-based linear or generalized linear models, and as such fast procedures have been developed for estimating coefficients. Most often the assumption of independent errors is violated and a high degree of collinearity is present in social networks models, as for example, all transitive triangles contain also dyads.

To give a statistical formulation of the models that we use, we introduce the following notation. Define $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\mathsf{T}$ to be a column random vector containing measurements of an outcome for n units in the analysis and $\boldsymbol{X}$ a matrix of dimension $n \times p$ for which each row $i = 1, \ldots, n$ represents a vector of p measurements on external predictors. By $\tilde{\boldsymbol{Y}} = [\tilde{Y}_{ij}]$ we denote a random adjacency matrix of size $n \times n$, formed based on binary random variables $\tilde{Y}_{ij}$ which take the value 1 if there is a tie present in the network between unit i and j and 0 otherwise. We denote by $\tilde{\boldsymbol{y}} = [\tilde{y}_{ij}]$ a particular realization of the adjacency matrix $\tilde{\boldsymbol{Y}}$ and by $\tilde{y}_{ij}$ a realization of the random variables $\tilde{Y}_{ij}$. If the strength of the tie between two actors is of interest, rather than the crude presence/absence of ties, we denote by $\tilde{\boldsymbol{Z}} = [\tilde{Z}_{ij}]$ a random matrix of size $n \times n$ that contains real valued entries. The higher the $\tilde{Z}_{ij}$ values, the higher the strength between two actors. We denote by $\tilde{\boldsymbol{z}} = [\tilde{z}_{ij}]$ a particular realization of the adjacency matrix $\tilde{\boldsymbol{Z}}$ and by $\tilde{z}_{ij}$ a realization of the random variables $\tilde{Z}_{ij}$.

By $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_p$ we denote p different adjacency matrices, each of dimension $n \times n$ to be used as possible 'explanatory' networks, in the sense that ties $\tilde{X}_{1,ij}, \ldots, \tilde{X}_{1,ij}$ could be useful in predicting ties $\tilde{Y}_{ij}$. We further construct $\tilde{\boldsymbol{X}}$ as a matrix whose columns are obtained by concatenating the vectorized versions of $\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_p$, i.e., $\tilde{\boldsymbol{X}} = [\mathrm{vec}(\tilde{\boldsymbol{X}}_1), \ldots, \mathrm{vec}(\tilde{\boldsymbol{X}}_p)]$.

Let $\tilde{\boldsymbol{W}}_1$ and $\tilde{\boldsymbol{W}}_2$ be two fixed $n \times n$ adjacency matrices (possibly real valued) that quantify the influence and relations existing between actors in a social network. The elements $\tilde{w}_{ij}$ in each adjacency matrix represent the extent to which a collaboration/relationship or influence is exchanged between actors i and j. By convention the diagonal entries are set to 0.

Mathematically the three models can be represented as follows:

(ERGM): $P(\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{y}}) = \kappa(\boldsymbol{\beta})^{-1} \exp(\boldsymbol{\beta}^\mathsf{T} g(\tilde{\boldsymbol{y}}))$ where $\kappa(\boldsymbol{\beta})$ is a normalizing constant to get a valid probability mass function, $\boldsymbol{\beta}$ is a vector of unknown parameters and $g(\tilde{\boldsymbol{y}})$ is a vector of network summary statistics used as potential explanatory variables;

(NAM): $\boldsymbol{Y} = \alpha \tilde{\boldsymbol{W}}_1 \boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} = \rho \tilde{\boldsymbol{W}}_2 \boldsymbol{\epsilon} + \boldsymbol{\nu}$, where $\boldsymbol{\epsilon}$ is a vector of stochastic errors, $\boldsymbol{\nu} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is a vector of normally distributed stochastic errors, $\boldsymbol{\beta}$ is a vector of regression parameters, α measures the effect of the network $\tilde{\boldsymbol{W}}_1$ on $\boldsymbol{Y}$ and ρ represents the strength of the network autocorrelation or how much the unobserved errors for each actor depend on the errors of its neighbors. Due to this specific formulation, this model is sometimes referred to as an autoregressive and moving average network model;

(NRM): $\mathrm{vec}(\tilde{\boldsymbol{Z}}) = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\tilde{\boldsymbol{Z}}$ real valued, $\boldsymbol{\beta}$ a vector of unknown parameters and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ a vector of stochastic errors.

3 General steps for FIC model selection

Before we present more technical details of the focused information criterion and how it is used in the context of estimating models for social networks, we first provide a more intuitive and less technical description of the grounds on which FIC selection is based.

As in Section 1, one starts by setting one or more research *focuses* for which one desires estimation and consequently models that perform well with respect to the *mean squared error*. In the example in Section 1, one studies family ties as an indirect measure of accumulating wealth and power among Florentine families in the 15th century. At the measurement level the marriage network was used. To understand how families unite in marriages the first focus is the log odds ratio between the probability p_{ij} that family i unites with family j and the probability that family i does not unite with family j. To study the wealth of the families one might be interested in a second focus, the expected value of the wealth, $E(\mathit{Wealth}|\boldsymbol{x})$, as a function of other covariates $\boldsymbol{x}$. A third focus could be the direct effect α, that the business network has on the wealth of the families. These are just a few examples of focuses which translate our research purposes into quantities that can be estimated based on data.

Relevant examples of focuses that are used in this manuscript include:

(i) the probability of observing a tie between two actors as a function of the other covariates (see Section 5),

(ii) the effect of a transitive triad on the probability of two actors forming a tie (see Section 5),

(iii) the magnitude of the autoregressive or autocorrelation effect of a network on the dependent variable (see Section 6),

(iv) the expected wealth of the Medici family as a function of other predictors (see Section 6),

(v) the expected exit orders of monks from a monastery (see Section 7),

(vi) information centrality (Stephenson and Zelen, 1989) of actors (see Section 7).

The above list can easily be enriched with other interesting focuses, more specialized depending on the application area.

Once the focuses are set, we proceed by producing a list of models from which we aim to select the best scoring one. The list should contain at least two models: one which is the smallest, most parsimonious model one is willing to consider, and one which is the largest, most complex model that is appropriate for the data. In between these two models, we list other models, some which contain certain covariates and exclude others. All of the models listed should be deemed appropriate for the data at hand. The list might be based on the set of all possible models in between the smallest and the largest model, as was done in Section 1, but this does not need to be case.

The estimators depend on the predictors one is using and thus they can have a large or small bias and variance. For example, the bias/variance trade-off can be different when one is using the narrow model, the full model or one in between the two models. Having a large number of possible predictors one might be tempted to include all of them for the sake of not missing the important predictors, while another researcher might opt for introducing only few predictors for the sake of having a simple, parsimonious model. Both these strategies might be deficient in case the true model is neither the full model, nor the narrow model. In the first case, one deliberately introduces potential noise variables that might blur the underlying signal, the model becomes less biased, but due of the added 'noise' the variance is increased. On the other hand, working with a simplistic model decreases the variance estimated for that model, but in the same time it increases the bias.

For each model we estimate the focus, the bias and the corresponding variance of the focus estimator. Different models lead to different bias values and to different variances. The purpose is to select that model that provides the best such bias/variance trade-off. Adding the squared bias and the corresponding variance of the estimators, one obtains the value for the MSE of the focus of interest. Next, all models are ranked according to the estimated MSE value and the final choice is for that model with the lowest such quantity, as one generally strives for models with low MSE values.

When one is performing model selection with respect to several focus quantities (as in Section 1 where we focused on the log odds ratio and on the parameter associated with the triangle statistic), it should not be surprising that different final models are selected, because some models might perform well with respect to some focuses and maybe less well with respect to other focuses. A bivariate search is possible too. This should be seen as a strong point of the technique, since it allows for a search of a good model in terms of mean squared error with respect to the quantities one is interested in. This should be contrasted to the traditional AIC/BIC information criteria which result in a single selected model that might not have close connections to the research questions, and is always the same, regardless of the model's use. Indeed, AIC/BIC selection does not require to specify any focus parameter. Those criteria cannot be directed towards a specific focus, as is explicitly included in the FIC. In the end the interpretation step proceeds as in the classical case, with the amendment that the model is targeted to provide low MSE for that particular focus.

To summarize, all main steps are listed below:

1. define focuses of interest related to the research theme (which are mathematical representations of our research interests);
2. conduct research and gather data;
3. construct a sequence of plausible models that allow to estimate the focuses;
4. for each model estimate the focus, the bias, the variance and the MSE of the estimator of the focus of interest;
5. select the model that provides the lowest MSE value as the final model for that focus.

4 Technical description of FIC

4.1 Protected and unprotected parameters

To set the notation, we denote the true value of the parameter vector in the narrow model by $\boldsymbol{\theta}_0$. The full model depends on both $\boldsymbol{\theta}_0$ and an additional parameter vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$. Other models in between the narrow and full model contain $\boldsymbol{\theta}_0$ and only some of the components of $\boldsymbol{\gamma}$. For example, only γ_2 and γ_4 are included. In order to not have to work with different parameter lengths in the different models we denote by $\boldsymbol{\gamma}_0$ the specific, known, value of $\boldsymbol{\gamma}$, often zero, such that when we fill in this value $\boldsymbol{\gamma}_0$ in the full model, we get back the narrow model as a special case. Hence we denote the true parameters of the narrow model by $(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$, where only $\boldsymbol{\theta}_0$ is to be estimated.

Since $\boldsymbol{\theta}_0$ is common to all models, this is called the *protected* parameter vector. The different models contain some or all components of $\boldsymbol{\gamma}$ in addition to $\boldsymbol{\theta}_0$. In other words, variable selection will make a choice of which components of the vector $\boldsymbol{\gamma}$ to include in the model. These are the *unprotected* parameters, subject to selection.

Revisiting the example introduced in Section 1, we have used the following models. The narrow model contained only one protected parameter $\theta_{\Delta_{Kstar(1)}}$, while all other parameters were set to 0. As such, for the narrow model we use the vector $(\boldsymbol{\theta}, \boldsymbol{\gamma}_0) = (\theta_{\Delta_{Kstar(1)}}, 0, 0, 0, 0, 0)$. The next model in the list contained one extra unprotected parameter, namely $\gamma_{\Delta_{Kstar(2)}}$ and as such, the vector of parameters was $(\boldsymbol{\theta}, \boldsymbol{\gamma}) = (\theta_{\Delta_{Kstar(1)}}, \gamma_{\Delta_{Kstar(2)}}, 0, 0, 0, 0)$. The full, most complex, model contained the parameters $(\boldsymbol{\theta}, \boldsymbol{\gamma}) = (\theta_{\Delta_{Kstar(1)}}, \gamma_{\Delta_{Kstar(2)}}, \gamma_{\Delta_{Kstar(3)}}, \gamma_{\Delta_{Triangle}}, \gamma_{\Delta_{Gwesp(\tau=.5)}}, \gamma_{Wealth})$.

4.2 The focus, local misspecification and MSE

For each model, one is interested in estimating a univariate function μ of the corresponding parameters, denoted as $\mu = \mu(\boldsymbol{\theta}, \boldsymbol{\gamma})$. Different choices for which components of $\boldsymbol{\gamma}$ to include, lead to different models, and thus to different estimators of the focus μ. We use MSE to decide on which of those estimators of μ is the best.

Since the true model is unknown, the MSE of the focus estimators needs to be estimated too. For this purpose we make the assumption of a locally misspecified model. This means that the true model is in a 'neighborhood' of the narrow model. With increasing n the neighborhood becomes smaller and in the limit experiment $n \to \infty$, the two models coincide. This is less strict than assuming the true model to be one of the models used.

Under a locally misspecified framework, we assume that the density of a random variable depends on the parameters vectors $\boldsymbol{\theta}_0$, $\boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n}$, with n the sample size. The vector $\boldsymbol{\gamma}_0$ is known and user-specified, e.g., containing all zeros, while $\boldsymbol{\theta}_0$ and $\boldsymbol{\delta}$ are to be estimated. The vector $\boldsymbol{\delta}$ determines the 'proximity' around the narrow model, to indicate the true model. This true model is only used to compute the bias and variance of the focus estimators in the different candidate models.

As such, whenever we model ties between actors we work with the probability mass function $f_n(\tilde{y}_{ij}) = f(\tilde{y}_{ij}, \boldsymbol{\theta}_0, \boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n})$ (in the case of ERGMs and NRMs) and whenever we model actor related continuous variables we work with the density $f_n(y_{ij}) = f(y_{ij}, \boldsymbol{\theta}_0, \boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n})$. By n we denote the sample size which when modeling ties refers to the total number of ties between actors in the network, while in the case of actor related variables, it refers to the number of actors.

As we have argued, these estimators $\widehat{\mu}$ have different biases and different variances. The $\text{MSE}(\widehat{\mu}) = \{\text{bias}(\widehat{\mu})\}^2 + \text{Var}(\widehat{\mu})$ is an objective method of performance as it brings together the bias and the variance of the estimator into one single measure. The smaller the MSE, the better the performance of the estimator.

We now repeat the general derivation of the FIC as in Claeskens and Hjort (2003, 2008b). We start by defining the Fisher information matrix $\boldsymbol{J}$, i.e., the expected value of the matrix of minus second partial derivatives of the log-likelihood with respect to the parameters, for the 'wide' model, as this is one of the key ingredients for calculating the bias and the variance expressions. We further partition $\boldsymbol{J}$ according to the length of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ into $\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_{00} & \boldsymbol{J}_{01} \\ \boldsymbol{J}_{10} & \boldsymbol{J}_{11} \end{pmatrix}$. The submatrix $\boldsymbol{J}_{00}$ relates to the second order partial derivatives of the log-likelihood with respect to the vector $\boldsymbol{\theta}$, the submatrix $\boldsymbol{J}_{11}$ corresponds to using the second order derivatives of the log-likelihood with respect to the vector $\boldsymbol{\gamma}$ and the submatrix $\boldsymbol{J}_{01}$ corresponds to using the second order derivatives of the log-likelihood first with respect to the vector $\boldsymbol{\theta}$ and then with respect to the vector $\boldsymbol{\gamma}$. The partitioning according to the distinction between protected and unprotected parameters is important, as further on in the calculations the different submatrices play different roles. We further define the matrix $\boldsymbol{Q} = (\boldsymbol{J}_{11} - \boldsymbol{J}_{10}\boldsymbol{J}_{00}^{-1}\boldsymbol{J}_{01})^{-1}$ which plays an important role both in the variance (indirectly through the matrix $\boldsymbol{Q}_S^0$ defined below) as well as in the bias (indirectly through the matrix

$\boldsymbol{G}_S$ defined below).

A model is selected from a list of plausible models. Let S be a subset of $\{1, \ldots, p\}$ where p is the number of parameters in the vector $\boldsymbol{\gamma}$. To not have to write down separate notation for each model in the list we identify the model with the index set S. The narrow model corresponds to $S = \emptyset$, the empty set, while the full model corresponds to $S = \{1, \ldots, p\}$. For a certain model S we further introduce the matrices $\boldsymbol{Q}_S = (\boldsymbol{\pi}_S \boldsymbol{Q}^{-1} \boldsymbol{\pi}_S^\mathsf{T})^{-1}$, $\boldsymbol{Q}_S^0 = \boldsymbol{\pi}_S^\mathsf{T} \boldsymbol{Q}_S \boldsymbol{\pi}_S$ (which plays a role in the variance) and $\boldsymbol{G}_S = \boldsymbol{\pi}_S^\mathsf{T} \boldsymbol{Q}_S \boldsymbol{\pi}_S \boldsymbol{Q}^{-1}$ (which plays a role in the bias), where $\boldsymbol{\pi}_S$ is a projection matrix that contains 0s and 1s, such that when multiplied with matrices of interest, it retains those rows and columns that relate to the parameters contained in model S. For example, for a vector $\boldsymbol{v}$, with $\boldsymbol{\pi}_{\{2\}} = (0, 1, 0, \ldots, 0)$, $\boldsymbol{\pi}_{\{2\}} \boldsymbol{v} = v_2$, the second component of $\boldsymbol{v}$ (see Claeskens and Hjort, 2008b, p. 146).

The focus parameter is estimated in each considered model S. With a slight abuse of notation regarding the order of the components of $\boldsymbol{\gamma}$, this leads to considering estimators $\hat{\mu}_S = \mu(\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\gamma}}_S, \boldsymbol{\gamma}_{0,S^c})$, where $\hat{\boldsymbol{\theta}}_S$ and $\hat{\boldsymbol{\gamma}}_S$ represents the estimated protected and unprotected parameters when using model S, while $\boldsymbol{\gamma}_{0,S^c}$ are the known fixed values of components of $\boldsymbol{\gamma}_0$ that are not considered in the set S. For example, if γ_5 is not in the model, we set $\gamma_{0,5} = 0$.

From Theorem 6.1 in Claeskens and Hjort (2008b) for any model S, the maximum likelihood estimator of the focus parameter obeys

$$\sqrt{n}\,(\hat{\mu}_S - \mu_{true}) \xrightarrow{d} \boldsymbol{\Lambda} \sim N\Big(\boldsymbol{\omega}^\mathsf{T}(\boldsymbol{I} - \boldsymbol{G}_S)\boldsymbol{\delta}, \left(\frac{\partial \mu}{\partial \boldsymbol{\theta}}\right)^\mathsf{T} \boldsymbol{J}_{00}^{-1} \frac{\partial \mu}{\partial \boldsymbol{\theta}} + \boldsymbol{\omega}^\mathsf{T} \boldsymbol{Q}_S^0 \boldsymbol{\omega}\Big).$$

The true focus $\mu_{true} = \mu(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0 + \boldsymbol{\delta}/\sqrt{n})$ is the function μ evaluated at the true parameters. We define $\boldsymbol{\omega} = \boldsymbol{J}_{10} J_{00}^{-1} \frac{\partial \mu}{\partial \boldsymbol{\theta}} - \frac{\partial \mu}{\partial \boldsymbol{\gamma}}$, where $\frac{\partial \mu}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \mu}{\partial \boldsymbol{\gamma}}$ are the vectors of partial derivatives of the focus with respect to the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, while $\boldsymbol{I}$ is the identity matrix of the same dimensions as $\boldsymbol{G}_S$. Adding the squared bias and variance of the estimator using model S leads to the mean squared error expression:

$$\mathrm{MSE}(\hat{\mu}_S) = \left(\frac{\partial \mu}{\partial \boldsymbol{\theta}}\right)^\mathsf{T} \boldsymbol{J}_{00}^{-1} \frac{\partial \mu}{\partial \boldsymbol{\theta}} + \boldsymbol{\omega}^\mathsf{T} \boldsymbol{Q}_S^0 \boldsymbol{\omega} + \boldsymbol{\omega}^\mathsf{T}(\boldsymbol{I} - \boldsymbol{G}_S)\boldsymbol{\delta}\boldsymbol{\delta}^\mathsf{T}(\boldsymbol{I} - \boldsymbol{G}_S)^\mathsf{T}\boldsymbol{\omega}.$$

By plugging-in sample versions of the unknown quantities, the estimated MSE is

$$\widehat{\mathrm{MSE}}(\hat{\mu}_S) = \left(\frac{\partial \mu}{\partial \boldsymbol{\theta}}\right)^\mathsf{T} \hat{\boldsymbol{J}}_{00}^{-1} \frac{\partial \mu}{\partial \boldsymbol{\theta}} + 2\hat{\boldsymbol{\omega}}^\mathsf{T} \hat{\boldsymbol{Q}}_S^0 \hat{\boldsymbol{\omega}} + \hat{\boldsymbol{\omega}}^\mathsf{T}(\boldsymbol{I} - \hat{\boldsymbol{G}}_S)\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^\mathsf{T}(\boldsymbol{I} - \hat{\boldsymbol{G}}_S)^\mathsf{T}\hat{\boldsymbol{\omega}} - \hat{\boldsymbol{\omega}}^\mathsf{T} \hat{\boldsymbol{Q}} \hat{\boldsymbol{\omega}}.$$

Eliminating constants that do not depend on the model S the *focused information criterion* (Claeskens and Hjort, 2003) is obtained as

$$\mathrm{FIC}(\hat{\mu}_S) = \hat{\boldsymbol{\omega}}^\mathsf{T}(\boldsymbol{I} - \hat{\boldsymbol{G}}_S)\hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}^\mathsf{T}(\boldsymbol{I} - \hat{\boldsymbol{G}}_S)^\mathsf{T}\hat{\boldsymbol{\omega}} + 2\hat{\boldsymbol{\omega}}^\mathsf{T} \hat{\boldsymbol{Q}}_S^0 \hat{\boldsymbol{\omega}}. \tag{4.1}$$

In following sections, we show how expression (4.1) can be used to perform model selection within the general classes of ERGMs, NAM and NRM models introduced in Section 2. An implementation of FIC using the R software (R Development Core Team, 2008) is available from the authors' webpage.

5 Empirical example ERGM

To make the previous concepts more concrete, in this section we present examples on how the above quantities can be estimated from data. For this we retake the 'Florentine families' dataset introduced in Section 1. The marriage social network and the wealth of the families are used to exemplify the concepts.

We denote by $\Delta_{\text{network statistic}_{ij}}$ a change in a network statistic when the tie $\tilde{y}_{ij}$ is changed from 0 to 1, or vice versa. As network statistics we consider the number of edges in the network ($\Delta_{Kstar(1)_{ij}}$), the number of two edge configurations ($\Delta_{Kstar(2)_{ij}}$), the number of three edge configurations ($\Delta_{Kstar(3)_{ij}}$), the number of transitive triangles ($\Delta_{Triangle_{ij}}$), the $Gwesp(\tau = 0.5)$ statistic (Hunter and Handcock, 2006) and the absolute difference in wealth (on the log scale) between any two families. For example, the network statistic $\Delta_{Triangle_{ij}}$ records the change in the number of transitive triangles in the network that would occur if the observed value of a tie $\tilde{y}_{ij}$ were changed from 0 to 1, while leaving all other ties intact. The same holds for all other summary statistics.

For this purpose consider the following model:

$$P(\tilde{Y}_{ij} = 1|\tilde{y}^c_{ij}) = p_{ij} =$$
$$= \frac{\exp(\theta_1\Delta_{Kstar(1)_{ij}}+\gamma_1\Delta_{Kstar(2)_{ij}}+\gamma_2\Delta_{Kstar(3)_{ij}}+\gamma_3\Delta_{Triangle_{ij}}+\gamma_4\Delta_{Gwesp(\tau=.5)_{ij}}+\gamma_5 Wealth_{ij})}{1+\exp(\theta_1\Delta_{Kstar(1)_{ij}}+\gamma_1\Delta_{Kstar(2)_{ij}}+\gamma_2\Delta_{Kstar(3)_{ij}}+\gamma_3\Delta_{Triangle_{ij}}+\gamma_4\Delta_{Gwesp(\tau=.5)_{ij}}+\gamma_5 Wealth_{ij})},$$

or more concisely

$$p_{ij} = \frac{\exp(\boldsymbol{u}_{ij}^\top\boldsymbol{\theta} + \boldsymbol{z}_{ij}^\top\boldsymbol{\gamma})}{1 + \exp(\boldsymbol{u}_{ij}^\top\boldsymbol{\theta} + \boldsymbol{z}_{ij}^\top\boldsymbol{\gamma})}, \tag{5.1}$$

where we denote by $\boldsymbol{u}_{ij}$ a column vector of differences of summary statistics whose associated $\boldsymbol{\theta}$ parameters are protected and by $\boldsymbol{z}_{ij}$ a column vector of difference statistics whose associated $\boldsymbol{\gamma}$ parameters are unprotected.

In the above model, $\boldsymbol{\theta} = \theta_1$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_5)$ represent unknown parameters. The parameters $\boldsymbol{\theta}$ represent the parameters that are always present in all considered models and act as protected parameters, whereas the $\boldsymbol{\gamma}$ vector corresponds to all parameters that are unprotected and for which variable selection is desired. We consider here the number of edges in the network as a protected variable, always included in the models through which we search, acting in a sense similar to an intercept in a regression model. The distinction protected/unprotected comes from the scientific intuitions and hypotheses on which the models are built. Extreme cases where all parameters are protected (equivalent to not performing any selection) or unprotected are also possible.

Since (5.1) corresponds to using a logistic regression to model the probability of a tie occurring between two actors we write the log-likelihood as

$$\ell(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_{i,j|i\neq j} y_{ij}\log(p_{ij}) + (1 - y_{ij})\log(1 - p_{ij}),$$

for which the empirical Fisher information matrix is

$$\boldsymbol{J}_n = \frac{1}{n}\sum_{i,j|i\neq j} p_{ij}(1 - p_{ij}) \begin{pmatrix} \boldsymbol{u}_{ij}^\top\boldsymbol{u}_{ij} & \boldsymbol{u}_{ij}^\top\boldsymbol{z}_{ij} \\ \boldsymbol{z}_{ij}^\top\boldsymbol{u}_{ij} & \boldsymbol{z}_{ij}^\top\boldsymbol{z}_{ij} \end{pmatrix}$$

which is the main quantity needed when computing the MSE values.

For this example we consider three focuses:

- the log odds ratio $\log(\frac{p_{ij}}{1-p_{ij}})$, where p_{ij} is the probability of a tie occurring between families (or how likely is it to have a tie between two families as opposed to not having it present) at fixed values for the change statistics : $\mu(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \log(\frac{p_{ij}}{1-p_{ij}}) = \boldsymbol{u}_{ij}^{\mathsf{T}}\boldsymbol{\theta} + \boldsymbol{z}_{ij}^{\mathsf{T}}\boldsymbol{\gamma}$. We concentrate here on the Strozzi family and inspect the change statistics when the tie between this family and the Medici or Peruzzi family changes from 0 to 1 in the observed network. The changes with respect to the Medici family are used in the first focus, i.e., $\mu_1(\boldsymbol{\theta}, \boldsymbol{\gamma})$ and the changes with respect to the Peruzzi family are used in the second focus, namely $\mu_2(\boldsymbol{\theta}, \boldsymbol{\gamma})$;

- the effect a transitive triangle has on forming a tie between two families: $\mu_3(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \gamma_3$.

Table 3 presents the results of the analysis. The first two panels present results when focusing on the log odds ratio as a function of the modifications obtained when the Strozzi family would develop a tie with the Medici family (μ_1) or if the link between Strozzi and the Peruzzi family would not be present (μ_2). In the last panel, results are presented for the case when we focus on the effect that a transitive triad (μ_3) has on the probability of observing such a network.

The best five fitting models selected based on FIC when focusing on μ_1 estimate the focus at low values which suggests that based on these models, the probability of observing such a tie is much smaller than the probability of not observing a tie. It suggests thus, that the model does not sustain the presence of such a tie in the network. Alongside the best five scoring FIC models, we have also inspected four other models: the best scoring AIC and BIC models, as well as the full model (containing the predictors $\Delta_{\mathit{Kstar(1)}}$ $\Delta_{\mathit{Kstar(2)}}$, $\Delta_{\mathit{Kstar(3)}}$, $\Delta_{\mathit{Triangle}}$, $\Delta_{\mathit{Gwesp}(\tau\,=\,.5)}$ and *Wealth*) and the narrow model (containing only the predictor $\Delta_{\mathit{Kstar(1)}}$). The AIC and BIC selected the narrow model for these data. As we have mentioned in Section 3, Table 3 illustrates that the full model provided the largest estimated variances 1.65 (μ_1), 2.22 (μ_2) and 26.74 (μ_3), while the narrow model provided the largest estimated biases 2.67 (μ_1), 11.69 (μ_2) and 40.55 (μ_3). Models in between the narrow and the full model provided a better trade-off between squared bias and variance resulting in better FIC performance. Also worth mentioning is that the differences between the scores, especially of the best two fitting models, is small and as such the selection of the final model to be used further might be dictated by other considerations like cost, availability of measurements, time, etc. In general, the question 'how small is a small difference?' that affects all information criteria is still an open question also for the FIC. For an interpretation of the FIC values, it is useful to present the root-estimated MSE, i.e. $\{\widehat{\mathrm{MSE}}(\hat{\mu}_S)\}^{1/2}$, since this bares similarity to the more traditional standard error of an estimator, though now including the bias component too. Following the advice of Burnham and Anderson (2002), we suggest inspecting the top three or top five best scoring models, for which the scores are 'close' together (as a rule of thumb differences between 0 to 2 units are deemed close). If the difference between models is judged to be small, other considerations like parsimony or

Focus	Parameters $(\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5)$	$\widehat{\mu}$	$\widehat{\text{MSE}}^{1/2}$	FIC	$\widehat{\text{bias}}$	$\widehat{\text{Var}}$	AIC	BIC
$\mu_1 : \log(\frac{p_{ij}}{1-p_{ij}})$ Medici								
Best FIC	0 0 0 0 1	-1.80	0.34	1.643	-0.47	0.12	111.5	117.0
2nd FIC	0 0 1 1 1	-2.24	0.71	2.026	-0.75	0.50	115.0	126.1
3rd FIC	1 0 1 0 1	-1.90	1.06	2.660	0.64	1.13	115.4	126.5
4th FIC	1 0 0 1 1	-1.92	1.07	2.663	0.35	1.14	115.3	126.5
5th FIC	1 0 0 0 1	-1.80	1.15	2.841	0.84	1.13	113.5	121.8
Full	1 1 1 1 1	-2.22	1.29	3.181	0.00	1.65	119.0	135.7
Best AIC/BIC	0 0 0 0 0	-1.62	2.68	7.122	2.67	0.06	110.1	112.9
$\mu_2 : \log(\frac{p_{ij}}{1-p_{ij}})$ Peruzzi								
Best FIC	0 0 0 1 0	1.27	1.33	3.856	0.71	1.74	112.0	117.6
2nd FIC	0 0 1 1 1	1.32	1.36	3.939	-0.20	1.84	115.0	126.1
3rd FIC	1 0 0 1 0	1.14	1.47	4.249	0.24	2.15	114.0	122.4
4th FIC	1 0 1 1 1	1.27	1.49	4.311	-0.33	2.16	117.0	130.9
5th FIC	0 1 1 1 1	1.29	1.49	4.322	0.03	2.22	117.0	130.9
Full	1 1 1 1 1	1.28	1.49	4.322	0.00	2.22	119.0	135.7
Best AIC/BIC	0 0 0 0 0	1.62	11.70	136.720	11.69	0.06	110.1	112.9
$\mu_3 = \gamma_3$								
Best FIC	0 0 1 1 1	-4.62	4.90	50.629	0.27	24.02	115.0	126.1
2nd FIC	0 0 1 1 0	-4.44	4.95	51.099	1.79	24.01	113.6	122.0
3rd FIC	1 0 1 1 1	-4.62	5.17	53.302	0.65	26.51	117.0	130.9
4th FIC	0 1 1 1 1	-4.61	5.17	53.325	-0.34	26.67	117.0	130.9
5th FIC	1 1 1 1 1	-4.64	5.17	53.343	0.00	26.74	119.0	135.7
Best AIC/BIC	0 0 0 0 0	0.00	40.55	1644.076	40.55	0.06	110.1	112.9

Table 3: Florentine family data. ERGM based example using three focuses. The narrow model contains only the change in the number of edges statistic $\theta = \Delta_{Kstar(1)}$. The parameters $(\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5)$ correspond to the predictors $(\Delta_{Kstar(2)},\Delta_{Kstar(3)},\Delta_{Triangle},\Delta_{Gwesp(\tau=.5)}, Wealth)$. For μ_1 the vector of change statistics is $(2, 10, 21, 1, 2.4, 0.35)$, corresponding to the changes with respect to the relations Strozzi-Medici, while for μ_2 the vector is (-2, -5, -4, -1, -5.4, -1.1), corresponding to the changes with respect to the relations Strozzi-Peruzzi, which makes the two focuses distinct from one another. The focus μ_3 corresponds to the parameter associated with the transitive triangle network statistic. In the 'Variables' column 0/1 indicate the absence/presence of a predictor in the model.

external expert advice might play a decisive role in selecting the final model.

The analysis in Wasserman and Faust (1994) as well as the historical classification in Molho (1994) have placed the Peruzzi family close to the Strozzi family and as such, we have further focused on the log odds ratio of observing a tie between the two families given the vector of change statistics. The observed adjacency matrix contains this tie, and so we expect a rather high probability of this tie occurring if the identified models are appropriate. In the five best ranked models when focusing on this log odds ratio, the estimated probability of a tie being observed is much larger than the probability of it not being observed.

In the third example, we have focused on the parameter corresponding to the effect a change statistic based on transitive triangles has on forming ties. In this case, all of the five best ranked FIC models suggest including the change statistic based on triangles as it helps reduce the estimated MSE values. All five models estimate a large negative effect of this change statistic on the logit, whereas the best AIC and BIC model would avoid including any such effects.

The main message of the analysis is that different explanatory models performed differently with respect to the bias and the variance of the estimated focus. Using the FIC a researcher is able to guide the model selection process towards that part of the model which is of interest. For comparison, the AIC or BIC criteria are inflexible to such targeted searches, and given data they select a 'best' model which should be then used for all purposes. Using the full model as a safe-guard against missing out important and relevant variables is still a bad idea, because as the theory predicts and Table 3 verifies empirically, this would introduce a large variance, which in turn inflates the estimated MSE. Not including anything is also not a wise option because even though we decrease the variance part, we increase the bias resulting in large estimated MSE values. Models somewhere in between the full and the empty model, for this example, better balance the bias-variance trade-off.

This reflects that the purpose for which a researcher aims to construct a well performing model, is of utmost importance and different models might perform better or worse in offering an answer to the scientific question. In the FIC case, before starting to think about plausible causal relations, one needs to think first for which purpose is the model being developed and to which scientific questions the model should provide an answer.

One of the complications that arise when working with ERGMs is model degeneracy. The subject has been treated in Handcock (2003a,b) and is empirically illustrated in Goodreau et al. (2008), Handcock et al. (2008) and Hunter et al. (2008b). Loosely speaking degeneracy is connected to model misspecification, in the sense that trying to fit a model far from the underlying generating process can lead to non-existence of the maximum likelihood estimator. Being a model selection criterion, FIC is not equipped to alleviate model degeneracy, it makes use of the maximum likelihood estimator, as do the classical AIC and BIC. Inflated standard errors due to non-convergence of the estimation algorithm, are visible in values of the FIC via large values in the Fisher information matrix. To such cases the FIC does not apply.

6 Empirical example NAM

The example in Section 5 presented models that describe the process of presence or absence of ties as a function of other network summary statistics. In this section we concentrate on estimating models that describe the relations determining the wealth of a family, using the same 'Florentine family' dataset, as a function of several other family characteristics.

The added complications arise from the fact that one is also interested in quantifying the effect a social network has on determining the family wealth. It is quite plausible that since marriages were seen more as alliances between families in order to increase their social and financial power, this might also affect their wealth. It is sensible to think that in the 15th century, wealthier families would have an inclination to marry other wealthy families to maintain their social status and financial power and also to gain access to other resources. Alongside the marriage network used in Section 5 we use here also the 'Florentine business' network which records which of the 16 families have conducted business transactions together. It is quite plausible that the interactions between actors involved in such a network would have a direct consequence on the wealth of a family, maybe more importantly than the effect of the marriage network.

The model we put forward is

$$\boldsymbol{Y} = \alpha\tilde{\boldsymbol{W}}_1\boldsymbol{Y} + \beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \text{with } \boldsymbol{\epsilon} = \rho\tilde{\boldsymbol{W}}_2\boldsymbol{\epsilon} + \boldsymbol{\nu},$$

where $\boldsymbol{Y}$ represents a vector corresponding to the wealth of the families, $\boldsymbol{X}$ is a matrix of explanatory predictors and $\tilde{\boldsymbol{W}}_1$ and $\tilde{\boldsymbol{W}}_2$ are the business and marriage networks. Here the $\boldsymbol{X}$ matrix contains as columns the measurements for 'Priorates' (the number of seats on the civic council), 'Totalties' (the number of all ties linking the family to any of the other 116 families from Florence), the degree centrality of the families involved in marriage network, the average betweenness score of each family and the average flow betweenness scores of each families. The averages are obtained by taking the respective measures estimated from the marriage and business networks and averaging them for each family. The vector $(\alpha, \beta_0, \boldsymbol{\beta}, \rho)$ collects unknown coefficients to be estimated from the data, $\boldsymbol{\beta} = (\beta_{Priorates}, \beta_{Totalties}, \beta_{Degree}, \beta_{Betweenness}, \beta_{Flow})$ is the vector of unknown parameters associated to predictors stored in $\boldsymbol{X}$, $\boldsymbol{\epsilon}$ represents a vector of unknown errors and $\boldsymbol{\nu}$ is a vector of normal random variables, $\boldsymbol{\nu} \sim N(\boldsymbol{0}, \sigma_\nu^2\boldsymbol{I})$.

In this particular application $\boldsymbol{\theta} = (\sigma_\nu^2, \beta_0)$ which corresponds to having only an intercept in the protected group of predictors and to saying that all models should contain the variance parameter. The parameters related to the possible predictors are $\boldsymbol{\gamma} = (\alpha, \rho, \boldsymbol{\beta})$. We want to perform model selection and thus the different models set some of the $\boldsymbol{\gamma}$ entries to 0. For example, the narrow model uses the parameter vector $(\boldsymbol{\theta}, \boldsymbol{\gamma}_0) = (\sigma_\nu^2, \beta_0, 0, 0, 0, 0, 0, 0, 0)$, whereas the full model uses the parameter vector $(\boldsymbol{\theta}, \boldsymbol{\gamma}) = (\sigma_\nu^2, \beta_0, \alpha, \rho, \beta_{Priorates}, \beta_{Totalties}, \beta_{Degree}, \beta_{Betweenness}, \beta_{Flow})$.

Depending on the research interest one may also consider one or both components of the vector (α, ρ) that is associated with the network effects, as being protected if one desires models that should contain such dependencies, but for didactic reasons we have left these two parameters unprotected, reflecting interest in allowing for some focuses a possible exclusion of such dependencies, if simpler models would perform better in terms of FIC scores.

We focus on two parameters:

1. the expected wealth of the Medici family as a function of all other predictors and social networks i.e. $\mu_1(\boldsymbol{\theta}, \boldsymbol{\gamma}) = E(Y_{Medici}|\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \boldsymbol{x})$;

2. α the autoregressive effect mediating the effect of the business network on the wealth, i.e. $\mu_1(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \alpha$.

These two focuses represent the statistical translation of our research questions. Estimating the expected wealth with small MSE seems a natural choice for a focus. There might be a connection between the wealth of the family and the business network a family is involved in, since this is directly linked to revenue. There might also be an indirect connection between the wealth of the family and the marriage network a family belongs to, since there might be a tendency for wealthier families to be united by marriage ties with other wealthy families.

Table 4 presents the obtained results. The first two parts present the selected models when it is desired to have good performing models for the two focuses separately, while the third part contains models which are ranked according to the average of the estimated MSE for both focuses, $0.5\{\widehat{\text{MSE}}(\hat{\mu}_1) + \widehat{\text{MSE}}(\hat{\mu}_2)\}$. As in Section 5 the full and empty model as well as the best scoring AIC and BIC models are included for comparison.

As before, the main message is that different focuses are better estimated by different models. With respect to the best scoring FIC models, they all estimate the wealth of the Medici family closer to the actual observed wealth (see the 'Abs.Diff.' column in Table 4, which presents the absolute value of the difference between the predicted wealth of the family based on the model and the actual observed wealth) either when estimating models targeted solely for this focus, or for the joint combination of focuses (last panel from Table 4). As in the previous example, the AIC and BIC models perform poorly and result in estimating high values for the MSE.

All five best FIC models, when compared to the AIC/BIC model, estimate a higher effect (α) of the business network on the estimated wealth suggesting that business ties seem to have some importance in determining the wealth of a family. As in the previous case, the empty model seems to be the worst performing model.

7 Empirical example NRM

We perform a small simulation study, starting from a real data example, where we evaluate the performance of the models selected by FIC and AIC in a controlled experiment.

The general considered model is

$$\text{vec}(\tilde{\boldsymbol{Z}}) \sim N(\beta_0 + \tilde{\boldsymbol{X}}\boldsymbol{\beta}, \sigma^2\{(\boldsymbol{I} - \rho\tilde{\boldsymbol{W}}_2^{\mathsf{T}})(\boldsymbol{I} - \rho\tilde{\boldsymbol{W}}_2)\}^{-1}) \equiv N(\beta_0 + \tilde{\boldsymbol{X}}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Sigma}_\rho),$$

where $\beta_0, \boldsymbol{\beta}, \rho$ and σ^2 are unknown parameters and $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{W}}_2$ are known matrices. The negative log-likelihood function to be minimized is proportional to $\log(\boldsymbol{e}_\rho^{\mathsf{T}}\sigma^{-2}\boldsymbol{\Sigma}_\rho\boldsymbol{e}_\rho) + \log\det(\sigma^2\boldsymbol{\Sigma}_\rho)$,

Focus	Parameters $(\alpha, \rho, P, T, D, B, F)$	$\widehat{\mu}$	Abs.Diff.	$\widehat{\text{MSE}}^{1/2}$	FIC	$\widehat{\text{bias}}$	$\widehat{\text{Var}}$	AIC	BIC
μ_1									
Best FIC	0 1 1 1 1 1 0	5.48	0.85	0.58	0.644	0.00	0.33	44.2	49.6
2nd FIC	0 1 0 1 1 1 0	5.42	0.79	0.62	0.674	0.07	0.31	42.5	47.1
3rd FIC	0 1 0 1 1 1 1	5.46	0.83	0.61	0.705	0.01	0.35	44.4	49.8
4th FIC	0 1 0 1 1 0 1	5.59	0.96	0.68	0.711	0.19	0.27	42.7	47.3
5th FIC	1 1 0 1 1 1 1	5.38	0.75	0.73	0.871	0.17	0.36	45.7	51.9
Full	1 1 1 1 1 1 1	5.48	0.85	0.74	1.075	0.00	0.55	47.6	54.6
Narrow	0 0 0 0 0 0 0	3.37	1.26	7.32	53.508	53.51	0.01	48.4	49.9
Best AIC/BIC	0 1 0 1 0 0 0	5.73	1.45	1.10	2.335	1.85	0.25	41.8	44.8
μ_2									
Best FIC	1 0 0 0 1 0 0	0.599	/	0.41	0.322	0.02	0.15	46.4	49.5
2nd FIC	1 1 0 0 1 0 0	0.603	/	0.48	0.378	0.08	0.15	45.1	49.0
3rd FIC	1 0 0 0 0 1 0	0.598	/	0.50	0.497	0.00	0.25	49.2	52.3
4th FIC	1 0 1 0 1 0 0	0.767	/	0.58	0.510	0.16	0.17	48.2	52.0
5th FIC	1 1 0 0 0 1 0	0.663	/	0.55	0.555	0.06	0.25	50.5	54.4
Full	1 1 1 1 1 1 1	0.531	/	0.92	1.680	0.00	0.84	47.6	54.6
Narrow	0 0 0 0 0 0 0	0.000	/	1.77	3.137	3.14	0.00	48.4	49.9
Best AIC/BIC	0 1 0 1 0 0 0	0.000	/	2.10	4.409	4.41	0.00	41.8	44.8
(μ_1, μ_2)									
Best FIC	1 1 0 1 1 1 0	(5.31 ,0.24)	(0.68,/)	0.31	0.132	/	/	44.2	49.7
2nd FIC	1 1 0 1 1 1 1	(5.38 ,0.65)	(0.74,/)	0.33	0.142	/	/	45.7	51.9
3rd FIC	1 1 1 1 1 1 0	(5.40 ,0.13)	(0.77,/)	0.38	0.175	/	/	46.2	52.4
4th FIC	1 1 1 1 1 1 1	(5.48 ,0.53)	(0.84,/)	0.33	0.182	/	/	47.6	54.6
5th FIC	1 1 0 1 1 0 1	(5.35 ,0.59)	(0.71,/)	0.31	0.190	/	/	43.7	49.1
Full	1 1 1 1 1 1 1	(5.48 ,0.53)	(0.84,/)	0.32	0.199	/	/	47.6	54.6
Narrow	0 0 0 0 0 0 0	(3.37 ,0.00)	(1.27,/)	3.51	12.352	/	/	48.4	49.9
Best AIC/BIC	0 1 0 1 0 0 0	(5.73 ,0.00)	(1.09,/)	0.88	0.797	/	/	41.8	44.8

Table 4: Florentine family data. NAM example using two focuses. The focus μ_1 represents the expected wealth of the Medici family given its observed network ties, i.e. $E(Y_{Medici}|\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \boldsymbol{x})$ and covariate values. The focus μ_2 corresponds to the parameter associated with the autoregressive effect mediating the effect of the business network on wealth, i.e. α. The narrow model contains only the intercept. The vector of parameters $(\alpha, \rho, P, T, D, B, F)$ stands for $(\alpha, \rho, \beta_{Priorates}, \beta_{Totalties}, \beta_{Degree}, \beta_{Betweenness}, \beta_{Flow})$.

where $\boldsymbol{e}_\rho = \tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}^\mathsf{T}\sigma^{-2}\boldsymbol{\Sigma}_\rho^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^\mathsf{T}(\sigma^{-2}\boldsymbol{\Sigma}_\rho)\tilde{\boldsymbol{Y}}$. The likelihood function depends only on ρ after inserting the maximum likelihood estimators for $\boldsymbol{\beta}$ and σ^2. The estimator of ρ is obtained by numerical optimization because a closed form expression is not available.

The Fisher information matrix (see Waller and Gotway, 2004), needed for calculating the MSE values for the considered models, takes the form:

$$\boldsymbol{J}(\rho, \sigma^2, \beta_0, \boldsymbol{\beta}) = \frac{1}{\sigma^2} \begin{pmatrix} \sum_i \left(\frac{\vartheta_i^2}{(1-\rho\vartheta_i)^2}\right) + \text{trace}(\boldsymbol{G}_\rho^\top \boldsymbol{G}_\rho) & \text{trace}(\boldsymbol{G}_\rho) & \boldsymbol{0} \\ \text{trace}(\boldsymbol{G}_\rho) & \frac{N\sigma^2}{2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & (\boldsymbol{1}, \tilde{\boldsymbol{X}})^\top \boldsymbol{A}_\rho (\boldsymbol{1}, \tilde{\boldsymbol{X}}) \end{pmatrix},$$

where $\boldsymbol{A}_\rho = (\boldsymbol{I} - \rho\tilde{\boldsymbol{W}}_2)^{-1}(\boldsymbol{I} - \rho\tilde{\boldsymbol{W}}_2^\top)^{-1}$, $\boldsymbol{G}_\rho = \tilde{\boldsymbol{W}}_2(\boldsymbol{I} - \rho\tilde{\boldsymbol{W}}_2)^{-1}$, ϑ_i are the eigenvalues of $\tilde{\boldsymbol{W}}_2$ and trace$(\cdot)$ denotes the trace operator, the sum of the diagonal elements of a square matrix. Under this model we optimize the log-likelihood with respect to all unknown parameters, considering $\boldsymbol{\theta} = (\beta_0, \rho, \sigma^2)$ as protected parameters included in all models and $\boldsymbol{\gamma} = \boldsymbol{\beta}$ as the unprotected parameters.

In this section we start from the 'Monastery' data of Sampson (1968) which contains information regarding 18 monks entering a monastery. Data were collected regarding the order in which the monks have left the monastery (due to ideological and political tensions some of the monks have been expelled from the monastery or have quit voluntarily) and as well eight different social relations between monks have been recorded. Each monk was asked to list three other monks which he likes, dislikes, praises or blames, esteems the most, esteems the least, has had the most influence on him or has had the least influence on him. The three nominations were afterward recoded as scores and eight social networks which embody monk-to-monk relations have been constructed. All eight networks are regarded as potential explanatory social networks.

In our analysis, the dependent network $\tilde{\boldsymbol{Z}}$ is given by $\tilde{Z}_{ij} =\mid z_i - z_j \mid$, where z_i and z_j represent the exit orders of monk i and monk j respectively. Note that the elements $\tilde{Z}_{ij}$ reflect the strength of the relation between the actors, rather than presence/absence of ties. The purpose is to model vec$(\tilde{\boldsymbol{Z}})$ as a linear function of the other eight explanatory social networks. The assumption of independent ties is questionable, because leaving the convent has been registered to occur in groups: first a group of four monks has been expelled and five other monks have soon afterward left on a voluntary base. The monks continued to leave the convent and as such, of the original 18 monks only four remained until the end of the seminar.

It is expected that not all the predictor social matrices have an effect on the order of exit, but it is unknown to the researcher which social networks to keep or exclude in order to obtain a good model.

In a second step, for each real value in the social network $\tilde{\boldsymbol{Z}}$ we have generated random errors that have been elementwise added to each tie. For the generation of errors, we first set $k_{ij} (0 \leq k_{ij} \leq 1)$ and σ_{sim} to particular values. We consider $k_c = 1 - k_{ij}$ as measuring the autocorrelation within columns and generate two random variables u_j and u_{ij} from the normal distribution $N(0, \sigma_{sim}^2)$. In the final steps, errors for each tie (i, j) are generated as $\epsilon_{ij} = k_c u_j + k_{ij} u_{ij}$ and added to $\tilde{Z}_{ij}$. The larger k_c, the stronger the correlation between monks; the larger σ_{sim}, the more noise is added to measurements for a particular monk j and the more noise is added to ties between actors i and j. In this example, k_{ij} took a value in the set $\{.1, .5, .7, .9\}$ and σ_{sim}^2 took a value in the set $\{1, 2, 3, 5\}$. This resulted in 16 different scenarios for generating errors and the noisy version of $\tilde{\boldsymbol{Z}}$ has been used throughout the calculations as the 'response' network.

In the above model, the likelihood expression requires the specification of the $\tilde{\boldsymbol{W}}_2$ matrix which is assumed to be a known matrix that represents the autocorrelation structure existing between ties, however its precise form is almost always unknown to the researcher, and it is not excluded that the ties are in some situations uncorrelated to each other. For this reason, we choose to add generated noise with correlated ties following the above procedure, but since the actual $\tilde{\boldsymbol{W}}_2$ matrix is unknown, we chose to fit models where such dependencies are ignored. More specifically, we wanted to see how detrimental modeling is under the assumption of independent ties, when this assumption is violated.

For this example we focus on two parameters:

1. the expected difference between exit orders of the monks i and j as a function of all other social networks ties i.e. $\mu_1(\boldsymbol{\theta}, \boldsymbol{\gamma}) = E(\tilde{Z}_{ij}|\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \tilde{\boldsymbol{x}})$; we repeat the procedure for each tie and select the model that provides on average the lowest FIC score;

2. the *information centrality* scores (Stephenson and Zelen, 1989) of each monk.

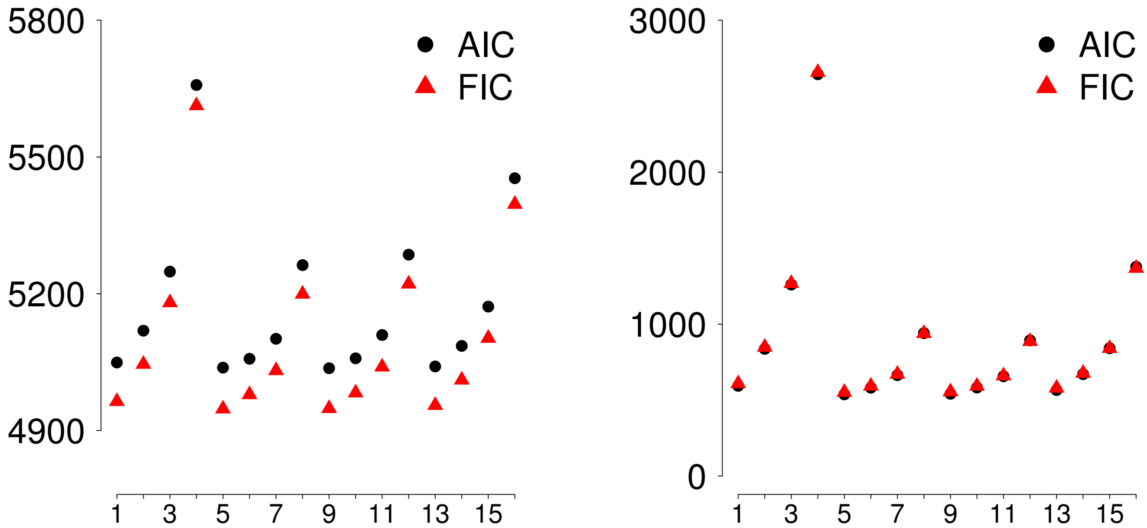

Figure 2: Sampson data. SSE_{Exit} (left panel) and $SSE_{InfoCent}$ (right panel), on the y-axis, for FIC and AIC selected models across 16 different simulation settings, on the x-axis.

We evaluate the performance of the models in terms of sums of squared errors (smaller is better) of the form

$$SSE_{Exit} = \sum_{i=1}^{18}\sum_{j=1}^{18}(\tilde{Z}_{ij} - \tilde{Z}_{ij}^{pred})^2;$$

$$SSE_{InfoCent} = \sum_{i=1}^{18}(InfoCent_i - InfoCent_i^{pred})^2.$$

where $\tilde{Z}_{ij}$ and $InfoCent_i$ represent the true order of exit differences and information centrality score and $\tilde{Z}_{ij}^{pred}$, $InfoCent_i^{pred}$ represent their predicted counterparts based on the selected models.

The final selected models correspond to the models that have the smallest value of the FIC and the AIC from the list of 2^8 possible different models. Figure 2 presents the SSE_{Exit} (left panel) and the $SSE_{InfoCent}$ (right panel) performance of both the FIC and AIC and under all types of autocorrelation structure the FIC model produces smaller SSE_{Exit} values than AIC. The values of $SSE_{InfoCent}$ are nearly the same for FIC and AIC, suggesting that the selected models do not influence to a high degree the centrality scores of the monks. The close values of $SSE_{InfoCent}$ show that in some cases, depending on the focus under study the FIC models behave similar to AIC selected models and it is not necessarily the case that FIC always improves on the AIC models, it all depends on the focus under study.

8 Discussion

We provide examples of how the FIC can be applied for model selection for social networks for three classes of models: ERGM, NAM and NRM. The procedure allows for selecting targeted models for certain quantities of interest which are linked to specific research questions. The advantage of using FIC is that the selected model represents best the question of interest. Different research questions can be offered possibly different selected models, because the final models are selected to perform optimally with respect to the focused parameters that statistically summarize one's research questions. Through the FIC one has a direct access to models that are more closely linked to the subject of the study and the researcher's interests as well as flexibility in using several models, each of them fine-tuned for a particular focus. A possible interesting extension of the applications of FIC is that of community detection (see Amini et al., 2013) where one could focus on probabilities for node labels to differentiate between communities of nodes in the graph. A further interesting development of the FIC can be towards dynamic networks where one might focus on transition probabilities (see Fan and Shelton, 2009) that link the existence of edges at time t with the existence of edges at future time points. Roughly in the same spirit, the FIC can also be extended towards relational event models (Butts, 2008b), where the focus can be related to parameters modeling survival functions. These possible extensions and similar others illustrate that the FIC is quite versatile with respect to model selection aspects and at the same time might open the way for further research.

Acknowledgements

The authors thank the reviewers for providing helpful comments to improve the manuscript. E. Pircalabelu is postdoctoral researcher of the Research Foundation Flanders (FWO) of which the support is acknowledged, together with support from KU Leuven grant GOA/12/14, and of the IAP Research Network P7/06 of the Belgian Science Policy. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Hercules Foundation and the Flemish Government - department EWI.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. and Csáki, F., editors, *Second International Symposium on Information Theory*, pages 267–281. Akadémiai Kiadó, Budapest.

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122.

Anderson, C. J., Wasserman, S., and Crouch, B. (1999). A p* primer: logit models for social networks. *Social Networks*, 21(1):37–66.

Austin, A., Linkletter, C., and Wu, Z. (2013). Covariate–defined latent space random effects model. *Social Networks*, 35(3):338–346.

Behl, P., Claeskens, G., and Dette, H. (2014). Focused model selection in quantile regression. *Statistica Sinica*, 24(2):601–624.

Breiger, R. L. and Pattison, P. E. (1986). Cumulated social roles: The duality of persons and their algebras. *Social Networks*, 8(3):215–256.

Bullard, M. (1979). Marriage politics and the family in Florence: The Strozzi-Medici alliance of 1508. *The American Historical Review*, 84(3):668–687.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference : A practical information-theoretic approach*. Springer.

Butts, C. (2008a). Social network analysis with sna. *Journal of Statistical Software*, 24(6):1–51.

Butts, C. T. (2008b). A relational event framework for social action. *Sociological Methodology*, 38(1):155–200.

Caimo, A. and Friel, N. (2013). Bayesian model selection for exponential random graph models. *Social Networks*, 35(1):11–24.

Claeskens, G. (2012). Focused estimation and model averaging with penalization methods, an overview. *Statistica Neerlandica*, 66(3):272–287.

Claeskens, G. and Hjort, N. (2003). The focused information criterion. *Journal of the American Statistical Association*, 98:900–916.

Claeskens, G. and Hjort, N. (2008a). Minimising average risk in regression models. *Econometric Theory*, 24:493–527.

Claeskens, G. and Hjort, N. (2008b). *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.

Dow, M., Burton, M., and White, D. (1982). Network autocorrelation: A simulation study of a foundational problem in regression and survey research. *Social Networks*, 4(2):169–200.

Fan, Y. and Shelton, C. R. (2009). Learning continuous–time social network dynamics. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 161–168.

Goodreau, S., Handcock, M., Hunter, D., Butts, C., and Morris, M. (2008). A statnet tutorial. *Journal of Statistical Software*, 24(1):1–26.

Goodreau, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Social Networks*, 29(2):231–248.

Handcock, M. (2003a). Assessing degeneracy in statistical models of social networks. Technical report, Working Paper 39, Center for Statistics and the Social Sciences, University of Washington.

Handcock, M. (2003b). Statistical models for social networks: Inference and degeneracy. In Breiger, R., Carley, K., and Pattison, P., editors, *Dynamic Social Network Modeling and Analysis*, volume 126, pages 229–252. Committee on Human Factors, Board on Behavioral, Cognitive, and Sensory Sciences, National Academy Press, Washington, DC.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1):1–11.

Hjort, N. and Claeskens, G. (2006). Focussed information criteria and model averaging for Cox's hazard regression model. *Journal of the American Statistical Association*, 101:1449–1464.

Holland, P. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50.

Hunter, D., Goodreau, S., and Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103:248–258.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15:565–583.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential–family models for networks. *Journal of Statistical Software*, 24(3):1–29.

Koskinen, J. H. (2004a). Bayesian analysis of exponential random graphs: Estimation of parameters and model selection. Technical report.

Koskinen, J. H. (2004b). Model selection for longitudinal social networks. Technical report.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social networks*, 10(4):359–381.

Leenders, R. T. A. J. (2002). Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, 24(1):21–47.

Molho, A. (1994). *Marriage Alliance in Late Medieval Florence*. Harvard University Press.

O'Malley, A. and Marsden, P. (2008). The analysis of social networks. *Health Services and Outcomes Research Methodology*, 8(4):222–269.

Padgett, J. and Ansell, C. (1993). Robust action and the rise of the Medici, 1400–1434. *The American Journal of Sociology*, 98(6):1259–1319.

Padgett, J. F. (1994). Marriage and elite structure in Renaissance Florence, 1282–1500. Technical report.

Pircalabelu, E., Claeskens, G., Jahfari, S., and Waldorp, L. (2015a). A focused information criterion for graphical models in fMRI connectivity with high–dimensional data. *The Annals of Applied Statistics*, 9(4):2179–2214.

Pircalabelu, E., Claeskens, G., and Waldorp, L. (2015b). A focused information criterion for graphical models. *Statistics and Computing*, 25(6):1071–1092.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robins, G., Snijders, T., Wang, P., and Handcock, M. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29:192–215.

Rodríguez, A. (2012). Modeling the dynamics of social networks using Bayesian hierarchical blockmodels. *Statistical Analysis and Data Mining*, 5(3):218–234.

Rohan, N. and Ramanathan, T. (2011). Order selection in ARMA models using the focused information criterion. *Australian & New Zealand Journal of Statistics*, 53(2):217–231.

Saito, K., Kimura, M., Ohara, K., and Motoda, H. (2010). Selecting information diffusion models over social networks for behavioral analysis. In Balcázar, J. L., Bonchi, F., Gionis, A., and Sebag, M., editors, *Machine Learning and Knowledge Discovery in Databases*, volume 6323, pages 180–195. Springer.

Sampson, F. S. (1968). *A novitiate in a period of change: An experimental and case study of social relationships*. PhD thesis, Cornell University.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shore, J. and Lubin, B. (2015). Spectral goodness of fit for network models. *Social Networks*, 43:16–27.

Stadtfeld, C., Geyer-Schulz, A., and Allmendinger, O. (2011). The influence of distance, time and communication network structures on the choice of communication partners. In *Proceedings of the 2011 IEEE International Conference on Social Computing*, pages 402–409.

Stephenson, K. and Zelen, M. (1989). Rethinking centrality: Methods and applications. *Social Networks*, 11:1–37.

Waller, L. and Gotway, C. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley.

Wang, P., Pattison, P., and Robins, G. (2013a). Exponential random graph model specifications for bipartite networks: A dependence hierarchy. *Social Networks*, 35(2):211–222.

Wang, P., Robins, G., Pattison, P., and Lazega, E. (2013b). Exponential random graph models for multilevel networks. *Social Networks*, 35(1):96–115.

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

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, 61:401–425.

Zhang, X. and Liang, H. (2011). Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics*, 39(1):174–200.

Zijlstra, B. J. H., Duijn, M. A. J., and Snijders, T. A. B. (2005). Model selection in random effects models for directed graphs using approximated Bayes factors. *Statistica Neerlandica*, 59(1):107–118.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be