Social Networks 50 (2017) 98–108

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

# Efficient Bayesian inference for exponential random graph models by correcting the pseudo-posterior distribution

Lampros Bouranis*, Nial Friel, Florian Maire

*School of Mathematical Sciences & Insight Centre for Data Analytics, University College Dublin, Ireland*

A R T I C L E I N F O

*Article history:*
Available online 3 May 2017

*Keywords:*
Exponential random graph models
Intractable normalizing constants
Large networks
Logistic regression
Pseudolikelihood
Tractable approximation

A B S T R A C T

Exponential random graph models are an important tool in the statistical analysis of data. However, Bayesian parameter estimation for these models is extremely challenging, since evaluation of the posterior distribution typically involves the calculation of an intractable normalizing constant. This barrier motivates the consideration of tractable approximations to the likelihood function, such as the pseudolikelihood function, which offers an approach to constructing such an approximation. Naive implementation of what we term a pseudo-posterior resulting from replacing the likelihood function in the posterior distribution by the pseudolikelihood is likely to give misleading inferences. We provide practical guidelines to correct a sample from such a pseudo-posterior distribution so that it is approximately distributed from the target posterior distribution and discuss the computational and statistical efficiency that result from this approach. We illustrate our methodology through the analysis of real-world graphs. Comparisons against the approximate exchange algorithm of Caimo and Friel (2011) are provided, followed by concluding remarks.



## 1. Introduction

The study of networks is central to a broad range of applications including epidemiology (dynamics of disease spread), genetics (protein interactions), telecommunications (worldwide web connectivity, phone calls) and social science (Facebook, Twitter, LinkedIn), among others. The high-dimensionality and complexity of such structures poses a real challenge to modern statistical computing methods.

Exponential random graph (ERG) models play an important role in network analysis since they allow for complex correlation patterns between the nodes of the network. However this model presents several difficulties in practice, mainly due to the fact that likelihood function can only be specified up to a parameter dependent normalizing constant. This impacts upon maximum likelihood estimation which is difficult to perform for larger networks, where the full likelihood function is available but it is just too complex to be evaluated. Robins et al. (2007b) presented various approaches for overcoming this problem.

The Bayesian paradigm has been used to infer exponential random graph models. The challenge of carrying out Bayesian estimation for these models has received attention from the statistical community in the recent past (Caimo and Friel, 2011). The main challenge encountered by such a Bayesian setting is the evaluation of the posterior that typically involves the calculation of an intractable normalizing constant. Sampling from distributions with intractable normalization can be done via Markov chain Monte Carlo methods (MCMC) and especially with the celebrated Metropolis–Hastings algorithm (Metropolis et al., 1953; Hastings, 1970).

Nevertheless, the normalizing term in the ERG probability distribution is a function of the model parameters and thus does not cancel as usual in the standard Metropolis–Hastings acceptance probability. This gives rise to a target distribution $\pi(\theta \mid y)$ which is termed *doubly-intractable* (Murray et al., 2006), where at each iteration of the Markov chain Monte Carlo scheme intractability of the normalizing term of the likelihood model within the posterior and intractability of the posterior normalizing term must be handled. More sophisticated Markov chains, such as the Exchange algorithm (Møller et al., 2006; Murray et al., 2006) have been proposed to sample from those doubly-intractable targets. Here again, these methods are not directly applicable in the ERG context as they require independent and identically distributed (iid) draws from the likelihood, which is not feasible for this type of models.

* Corresponding author at: Insight, O'Brien Centre for Science, Science Centre East, University College Dublin, Belfield, Dublin 4, Ireland.
*E-mail addresses:* lampros.bouranis@insight-centre.org (L. Bouranis), nial.friel@ucd.ie (N. Friel), florian.maire@ucd.ie (F. Maire).

http://dx.doi.org/10.1016/j.socnet.2017.03.013
0378-8733/

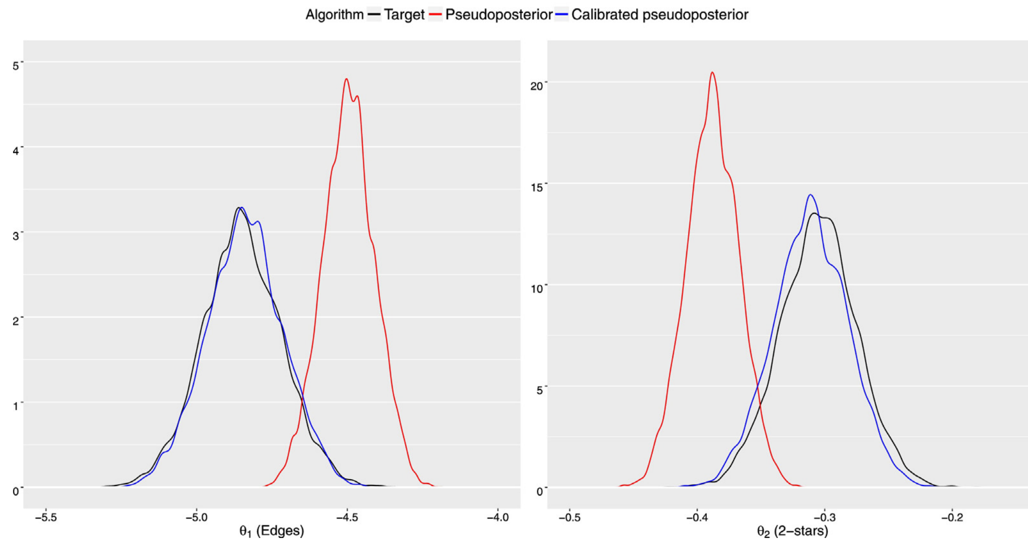


**Fig. 1.** International E-road Network: marginal density estimates of the posterior distribution (based on a long and computationally expensive MCMC run) (black curve); misspecified pseudo-posterior density estimates (where the pseudolikelihood has replaced the likelihood) (red curve); calibrated pseudo-posterior density estimates (blue curve). One can see that the calibration step has resulted in density estimates which are very close to the target posterior. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This motivated the approximate exchange algorithm of Caimo and Friel (2011) which substitutes iid draws from the likelihood with draws from an auxiliary Markov chain admitting the ERG likelihood as limiting distribution. Previous studies have shown that convergence of sampling from the ERG likelihood through Markov chain is likely to be exponentially slow (Bhamidi et al., 2011). This is likely to lead to increased computational burden when analyzing graphs with complex dependencies.

Increasing computational complexity has motivated the development of misspecified but tractable and computationally affordable models. From this perspective, the computational tractability of the pseudolikelihood function and the simplicity in defining the objective function seem to make it a tempting alternative to the full likelihood function when dealing with data with such a complex structure (Robins et al., 2007a; Handcock et al., 2007). The use of such an approximation to the likelihood should be treated with caution, though, as discussed by van Duijn et al. (2009); their work studied the quality of the pseudolikelihood approximation to the true likelihood, concluding that it may underestimate endogenous network formation processes. Despite its use by practitioners in the frequentist setting to fit ERG models, little is known about the pseudolikelihood approximation efficiency when embedded in a Bayesian posterior distribution. Our empirical analysis shows that the Bayesian estimators resulting from using the pseudolikelihood function as a plug-in for the true likelihood function are biased and their variance can be underestimated. However, the calibration procedure which we develop shows how to correct a sample from this so-called pseudo-posterior distribution, so that it is approximately distributed from the true posterior distribution.

We propose a novel methodology that falls into the area of MCMC targeting a misspecified posterior: a Metropolis–Hastings sampler whose stationary distribution we refer to as a *pseudo-posterior* (a posterior distribution where the likelihood is replaced by a pseudolikelihood approximation). Such a method is fast and overcomes double intractability but is also *noisy*, in the sense that it samples from an approximation of the true posterior distribution, $\pi$. A graphical illustration of marginal posterior densities under misspecification can be seen in Fig. 1. This example involves a two-dimensional model which we provide details of in Section 5.2. Here the misspecification of the actual posterior yields a disastrous approximation; compare the black and red curves. It is evident that a sampler which targets an approximate posterior resulting from replacing the intractable likelihood with a pseudolikelihood approximation leads to biased posterior mean estimates and considerably underestimated posterior variances. Nevertheless, we will present a correction method that allows to calibrate the sample to the true density; warping the red to the blue curve which is now a sensible approximation to the black one.

The aim of this paper is to exploit the use of pseudolikelihoods in Bayesian inference of exponential random graph models. Our work explores the computational efficiency of the resulting Markov chains, and the trade-off between computational and statistical efficiency in estimates derived from such pseudo-posteriors in comparison to the approximate exchange algorithm of Caimo and Friel (2011). We present the reader with a viable approach to calibrate the posterior distribution resulting from using a misspecified likelihood function in an efficient manner (Fig. 1), while providing the theoretical framework for this approach.

The outline of the article is as follows. A basic description of exponential random graph models is given in Section 2. The pseudolikelihood function as a surrogate for the true likelihood is introduced in Section 3. In Section 4 we formulate the Bayesian model in the presence of likelihood misspecification, and discuss the theoretical properties and practical aspects of the calibration of the posterior distribution. In Section 5, we illustrate our methods through numerical examples involving real networks. We conclude the paper in Section 6 with final remarks.

## 2. Exponential random graph models

Networks are relational data represented as graphs, consisting of nodes and edges. Many probability models have been proposed in order to understand, summarize and forecast the general structure of graphs by utilizing their local properties. Among those, Exponential random graph models play an important role in network analysis since they can represent transitivity and other structural features in network data that define complicated dependence patterns not easily modeled by more basic probability models (Wasserman and Pattison, 1996, see also Robins et al., 2007b for a review and the references therein for more details).

Let $\mathcal{Y}$ denote the set of all possible graphs on $n$ nodes. The network topology structure is measured by a $n \times n$ random adjacency