Notes on the Revision of SDERGM

April 19, 2021

1 Progress Toward Revision

1.1 Referee 1

Il referee 1 critica l'utilizzo della pseudolikelihood, e suggerisce di usare un metodo "more principled" per la stima. Nel nostro caso la pseudolikelihood serve per la stima, ma anche per definire l'update SD.

Key points della nostra risposta:

- Commentare sulla criticitá nella stima delle incertezze in Van Duijn et al. (2009) per stime pseudo-likelihood, citando Varin et al. (2011).
- Sottolineare similaritá tra i nostri test su filtro di dinamiche misspecified e i tests di Van Duijn et al. (2009), per quanto riguarda la loro analisi del MSE degli stimatori.
- Aggiungere confronto su tra filtro SD-MLE e SD-PMLE, per modello in cui si sa calcolare la funzione di partizione ma likelihood e pseudolikelihood sono diverse.

Possibile struttura della risposta:

- In Handcock (2003) Prof. Mark Handcock pointed out, that *"The value of the maximum pseudolikelihood estimator can then be expediently found by using logistic regression as a computational device. Importantly, the value of the maximum likelihood estimator for the logistic regression will also be the maximum pseudolikelihood estimator. Note, however, that the other characteristics of the maximum likelihood estimator do not necessarily carry over. In particular, the standard errors of the estimates of from the logistic regression will not be appropriate for the maximum pseudolikelihood estimator (MPLE)"*.

  This fact has been studied quantitatively in Van Duijn et al. (2009), paper Co-authored by Prof. Handcock, and published in Social Networks. That work is centered around a comparison of MPLE and MLE for inference in ERGMs.

Their conclusion on bias in the natural parametrization (the one that we use for our dynamical approach) is not very clear. In fact the boxplots show a larger variance of MPLE estimators wrt MLE but a lower bias.

Moreover their analysis on the uncertainty associated with each estimator assumes a distribution for the pseudolikelihood estimators that is known to be wrong. Estimating the variance of a pseudolikelihood estimator with the curvature of the pseudolikelihood, as they do[1], is not justified. Indeed (as discussed in Varin et al., 2011) the variance of PMLEs should be estimated with the Godambe Information Matrix:

$$G(\theta) = H(\theta)\, J(\theta)^{-1}\, H(\theta)$$

where

$$H(\theta) = E\left[-\nabla.\nabla^{(t)\prime}(\theta)\right],$$

the score of the PL is $\nabla^{(t)}(\theta) = \frac{\partial \log PL\left(\mathbf{Y}^{(t)}|\theta\right)}{\partial \theta_s^{(t)\prime}}$, and

$$J(\theta) = Var\left[-\nabla^{(t)}(\theta)\right],$$

In the context of ERGMs, where we tipically have only one observation, estimating matrix $J(\theta)$ is troublesome and tipically computationally expensive. One can either resort to a Jackknife approach (with carefully defined subsamples), apply bootstrap, when multiple observations are available Desmarais and Cranmer (2012), or use parametric bootstrap (as advocated in Schmid and Desmarais, 2017, paper mentioned by referee 2). The latter method amounts to to plain numerical exploration of the estimator's distribution via repeated sampling and estimates.

We believe that the skepticism toward naive applications of PML for inference in ERGMs is well justified. Indeed, obtaining estimates of PMLE's uncertainties, using the curvature of the pseudo-likelihood, is easy but leads to inconsistent confidence intervals. This well known fact should motivate researchers to avoid the bad practice of blindly trusting the results of inference software originally intended for logistic regressions. Nevertheless, it is not a good reason to completely avoid applications of PMLE.

It is important to stress that the SD-ERGM filter based on MPLE uses all observations previous to time t to filter the parameter $\theta^{(t)}$. This is, in our opinion, the main reason behind the clearly better MSE of the SD-ERGM-MPLE wrt to ERGM-MLE. Moreover our simulations ensure that, for the model at hand, bias and variance are moderate, in the dynamical

[1] Citing Van Duijn et al. (2009) at page 56 : *"For each estimator, a standard error estimate is derived from the estimated curvature of the corresponding log-likelihood or log-pseudo-likelihood ((5), (10) and (11)). We refer to these as perceived standard errors because these are the values formally derived as the standard approximations to the true standard errors from asymptotic arguments that have not been justified for these models. These are the values typically provided by standard soft-ware (Handcock et al., 2003; Boer et al., 2003; Wang et al., 2008). "*

context, as much as the results of Van Duijn et al. (2009) justify the MLE approach in the cross sectional case.

In some sense, our tests of filtering misspecified dynamical DGPs are similar to the approach of Van Duijn et al. (2009). In fact, for each time step, we sample the ERGM and estimate the value of parameter $\theta^{(t)}$ with 2 approaches: SD-ERGM, based on PMLE, and ERGM on a single snapshot, based on MLE. Due to the dynamical nature of our model and the extremely high number of likelihood computations, we had to resort to MPLE. Nevertheless, we stress that the SD-PMLE-ERGM does not require an estimate of the uncertainty of the PMLE at a single snapshot. Indeed in the new analysis motivated by referee's 2 comments we use parametric bootstrap when to account for the parameter uncertainty of the filted ERGMs parameters.

To further motivate the safety of PMLE's applications to SD-ERGM, we consider a simple ERGM for which we are able to estimate both the ML-SD-ERGM and the PML-SD-ERGM, and compare the results. The model considered is an ERG with the total number of links $\sum_{ij} A_{ij}$ and the total number of reciprocal pairs $\sum_{ij} A_{ij}A_{ji}/2$ as statistics.

$$
\begin{aligned}
P(A) = & \quad \frac{e^{\theta \sum_{ij} A_{ij} + \eta \sum_{ij} A_{ij}A_{ji}/2}}{Z(\theta,\eta)} \\
= & \quad \prod_{i>j} \frac{e^{\theta\left(A_{ij}+A_{ji}\right)+\eta A_{ij}A_{ji}}}{Z_{ij}(\theta,\eta)} \\
= & \quad \prod_{i>j} P\left(A_{ij}, A_{ji}\right)
\end{aligned}
\tag{1}
$$

where $Z_{ij}(\theta,\eta) = 1 + 2e^{\theta} + e^{2\theta+\eta}$ is the normalization constant for $P(A_{ij}, A_{ji})$ the MLE can be obtained in closed form but the links are not independent. Let us consider the expected values of $\alpha = \sum_{ij} A_{ij}/(N^2 - N)$ and $\beta = \sum_{ij} A_{ij}A_{ji}/2(N^2 - N)$

$$
\begin{aligned}
E[\alpha] = & \quad E[A_{ij}] = 2\frac{e^{\theta}+e^{2\theta+\eta}}{1+2e^{\theta}+e^{2\theta+\eta}} \\
E[\beta] = & \quad E[A_{ij}A_{ji}/2] = \frac{e^{2\theta+\eta}}{1+2e^{\theta}+e^{2\theta+\eta}},
\end{aligned}
\tag{2}
$$

that, defining $x = e^{\theta}$ and $y = e^{\eta}$, can be rewritten as

$$
\begin{aligned}
E[\alpha] &= 2\frac{x + x^2y}{1 + 2x + x^2y} \\
E[\beta] &= \frac{x^2y}{1 + 2x + x^2y}.
\end{aligned}
\tag{3}
$$

Solving for x and y

$$
\begin{aligned}
x &= \frac{E[\alpha]/2 - E[\beta]}{E[\beta] - E[\alpha] + 1} \\
y &= 2\frac{E[\beta] - E[\alpha] + 1}{\left(E[\alpha]/2 - E[\beta]\right)^2}.
\end{aligned}
\tag{4}
$$

The following constraints

$$0 \leq E\left[A_{ij}\right] \leq 1$$

$$max(0, E\left[A_{ij}\right] - 1/2) \leq E\left[A_{ij}A_{ji}/2\right] \leq E\left[A_{ij}\right]$$

are respected for all values of θ and η, as can be checked rewriting them in terms of x and y and considering that $x, y > 0$. Equation (5) links the ERGM parameters to the expected values of links' density α and reciprocal links' density β. It also relates the MLEs of those parameters with the observed values of α and β.

$$\theta = \log\left(\frac{\alpha/2 - \beta}{\beta - \alpha + 1}\right)$$

$$\eta = \log\left(2\frac{\beta - \alpha + 1}{\left(\alpha/2 - \beta\right)^2}\right). \tag{5}$$

As it is evident from the latter equation, MLEs for both parameters are not defined (they would be infinite) when the observed network is such that α and β satisfy one of the equalities in the constraints defined above. For this reason we define the misspecified dgps in our numerical tests such that the probability of observing networks without a well defined MLE is very low. Note that the SD models that we propose can handle observations for which a single snapshot MLE would not exist. We simply would not expect the filters to attain reasonable performances in reconstructing misspecified dgps resulting in a consistent proportion of such observations. For this reason, in our numerical tests, when comparing the ML-SD-ERGM and the PML-SD-SDERGM as misspecified filters, we consider DGPs for which the probability of observing such pathological realizations is low. Moreover, in our comparison we consider networks of different sizes N, in two density regimes, a dense regime $\alpha \sim const$ and a sparse regime $\alpha \sim 1/N$. We define a dgp for $\theta^{(t)}$ and $\eta^{(t)}$, by first defining a sequence $(\alpha^{(t)}, \beta^{(t)})$ such that, for all $t = 1, \ldots, T$, $\alpha^{(t)} \in [\alpha_m, \alpha_M]$ and $\beta^{(t)} \in [\beta_m, \beta_M]$ and then define $\left(\theta^{(t)}, \eta^{(t)}\right) = \left(\theta^{(t)}, \eta^{(t)}\right)_{\left(\alpha^{(t)}, \beta^{(t)}\right)}$. For networks of different size N, we define the maximum and minimum values for α and β in order to guarantee that the average networks' density $E\left[\alpha\right]$ is compatible with the chosen density regime, and that the sampled networks are unlikely to have values of α and β exactly on the physical constraints, where the MLE is not well defined. In practice, in the dense regime, we set $\alpha_m = 0.2$ and $\alpha_M = 0.3$, while in the sparse regime $\alpha_m = 0.2/N$ and $\alpha_M = 0.3/N$. In both regimes we then define $\beta_m = \alpha_m/5$ and $\beta_m = \alpha_m/2 - \alpha_m/5$. Empirically we found that a sufficient requirement is that the average number of reciprocated pairs is not less than 1, nor greater that $(\sum_{ij} A_{ij}/2) - 1$. This is guaranteed for our choice of $_m$, for $N > 25$, since $(N^2 - N)\alpha_m/5 > 2(N-1)/50$

We point out that, due to the discrete nature of random graphs, for all values of θ and η there is a finite probability of observing each possible graph, included the realizations for which the MLE is not well defined.

A questo proposito, vedi il commento al seguente link https://github.com/domenicodigangi/ScoreDrivenExponentialRandomGraphs/issues/27#issuecomment-757495502

1.2 Referee 2

Suggerimento del referee 2: add confidence intervals to filtered parameters.

Key points della risposta:

- We apply the method proposed in Blasques et. al to add confidence intervals to the filtered time varying parameters. The covariance matrix of the static parameters' estimators is estimated using the robust estimator originally proposed in White, and discussed in Bollerslev and Wooldridge (1992) in application to GARCH models (see also Bollerslev et al., 1994, for a discussion). In short, defining $\psi = (w, \beta, \alpha)$,

$$l_t = \log P\left(\mathbf{Y}^{(t)} | \theta^{(t)} \left(\psi, \left\{\mathbf{Y}^{(t')}\right\}_{t'=1}^{t-1}\right)\right). \tag{6}$$

$$l_T = \sum_{t=1}^{T} l_t \tag{7}$$

the covariance matrix is defined as $B_0^{-1} A_0 B_0^{-1}$, where

$$A_0 = \mathbb{E}\left[\nabla_\psi l_T \nabla'_\psi l_T\right], \qquad B_0 = \mathbb{E}\left[\nabla_\psi \nabla'_\psi l_T\right]$$

are estimated by the following sample means

$$\hat{A}_0 = \sum_t \nabla_\psi l_t \nabla'_\psi l_t, \qquad \hat{B}_0 = \sum_t \nabla_\psi \nabla'_\psi l_t$$

Per l'applicazione delle confidence bands al toy model vedi https://github.com/domenicodigangi/ScoreDrivenExponentialRandomGraphs/issues/32#issuecomment-764656150

1.3 Ulteriori Analisi Necessarie

Il progresso delle analisi rimanenti e tracciato al usando un progetto github associato alla repository: https://github.com/domenicodigangi/ScoreDrivenExponentialRandomGraphs/projects/1

2 Modifiche al Paper

- Non direttamente legato alle richieste dei referees. Cambiare approccio al targeting su SD beta model. Giacomo preferisce proporre la stima della unconditional mean su singola stima statica invece che sulle medie. Testato, ottiene risultati simili al targeting su media da stima single snapshot.

- Sia il fitness model che il toy model sono nested nel modello $p1$ di Holland and Leinhardt. Questo potrebbe semplificare l'esposizione nella nuova versione del paper.

- Per difendere la pseudolikelihood discutere confronto del filtro misspecified per modello in cui pseudo-mle $\neq$ mle, ma sappiamo calcolare la funzione di partizione.

- Modificare introduzione SDERGM con pseudolikelihood, riferendosi proprio a Van Duijn et al. (2009), che la critica, sottolineando similaritá tra loro tests e nostri.

- Aggiungere discussione del metodo di Blasques per gli intervalli di confidenza sui filtrati. Ottenuti per il fitness model .

3 Testo della Rejection da Social Networks

- Referee One: This paper tackles a worthwhile extension of ERGMs, but unfortunately it has to resort to pseudo-likelihood estimation to do so. For ERGMs, I simply do not trust that MPLEs are up to the task. That has been the state of knowledge for nearly 20 years, ever since Mark Handcock and Tom Snijders among others showed us good and simple examples where pseudolikelihood estimates were quite misleading. So, for this article, the authors have to either convince me that MPLE is for some reason adequate for their particular model class (and I dont really imagine how this could be); or apply a more principled method of estimation. If they can take either of these steps, then I am happy to work through the detail of a revised paper. But it is not enough simply to state, as they do on page 8, that pseudo-likelihood estimation is computationally superior to MCMC and is the only viable solution for their model.

- Referee Two: In this paper the authors develop a temporal version of the ERGM in which parameters vary over time in a way that is controlled by the score of the likelihood. There is a great need for methodology like this in the literature. First, as with all time series modeling methods, there may be temporal heterogeneity in the parameters. Second, there is a particularly important reason to incorporate this heterogeneity with ERGM, as not doing so can result in degenerate fits to individual time points. The score-driven method represents an important innovation relative to existing methods, which include fitting time interactions with parameters, and trying to find change points in parameter values. This paper offers an important contribution to the methods literature. I do, however, think the authors need to make one major revision before this paper will be ready to offer the contribution it promises. I also suggest two minor revisions.

  First, for the SD-ERGM to offer the a true alternative to, e.g., conventional TERGM or time-by-time ERGM, analysts must have an option for constructing confidence intervals, or at least evaluating whether they can reject the null that the parameter value is equal to 0 in a given time point. Researchers who use ERGM commonly use them for hypothesis testing. I do not see a method proposed in this paper, aside from the LM test for stability, for calculating confidence intervals or p-values. For the SD-ERGM to offer a contribution on-par with its innovativeness, I recommend that the authors develop one or more methods of quantifying uncertainty regarding the parameter values, and test the performance of the uncertainty assessment method(s) in their simulation study. For example, one option would be to use the parametric bootstrap method in the Schmid and Desmarais paper referenced in the article. The development methods for, e.g., calculating confidence intervals, would also increase the contributions offered in the empirical applications presented in this paper, as the authors could compare whether the null hypothesis of a 0 coefficient is rejected with the three different methods.

The second edit that I would recommend is that the authors put the methods that they have developed into an R package, and at least post the package on GitHub, if not submitting to CRAN. There may already be a package in the works, but I think the paper would have a more substantial impact if it referenced a package that users can access to apply the SD-ERGM.

The last edit I suggest regards the forecasting application. Since the SD-ERGM involves estimating extra parameters, when compared to the ERGM, it would be useful to see how the forecasting performance compare as the time series advances. It would be interesting to see when in the time series the SD-ERGM has enough information to out-perform the ERGM.

References

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). Arch models. *Handbook of econometrics 4*, 2959–3038.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews 11*(2), 143–172.

Desmarais, B. A. and S. J. Cranmer (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications 391*(4), 1865–1876.

Handcock, M. S. (2003). Assessing degeneracy in statistical models of social networks. (Working Paper No. 39).

Schmid, C. S. and B. A. Desmarais (2017). Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap. In *Big Data (Big Data), 2017 IEEE International Conference on*, pp. 116–121. IEEE.

Van Duijn, M. A., K. J. Gile, and M. S. Handcock (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.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5–42.