Title	Variational Bayesian inference for the Latent Position Cluster Model for network data
Authors(s)	Salter-Townshend, Michael, Murphy, Thomas Brendan
Publication date	2013-01
Publication information	Salter-Townshend, Michael, and Thomas Brendan Murphy. “Variational Bayesian Inference for the Latent Position Cluster Model for Network Data.” Elsevier, January 2013. https://doi.org/10.1016/j.csda.2012.08.004.
Publisher	Elsevier
Item record/more information	http://hdl.handle.net/10197/12538
Publisher's statement	This is the author’s version of a work that was accepted for publication in Computational Statistics & Data Analysis. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Computational Statistics & Data Analysis (57, 1, (2021)) https://doi.org/10.1016/j.csda.2012.08.004
Publisher's version (DOI)	10.1016/j.csda.2012.08.004

Computational Statistics and Data Analysis 57 (2013) 661–671

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Variational Bayesian inference for the Latent Position Cluster Model for network data

Michael Salter-Townshend *, Thomas Brendan Murphy *

School of Mathematical Sciences & Complex and Adaptive Systems Laboratory, University College Dublin, Belfield, Dublin 4, Ireland

ARTICLE INFO

Article history:
Received 6 January 2012
Received in revised form 13 July 2012
Accepted 3 August 2012
Available online 11 August 2012

Keywords:
Social network analysis
Variational Bayes
Latent Position Cluster Model

ABSTRACT

A number of recent approaches to modeling social networks have focussed on embedding the nodes in a latent "social space". Nodes that are in close proximity are more likely to form links than those who are distant. This naturally accounts for reciprocal and transitive relationships which are commonly found in many network datasets. The Latent Position Cluster Model is one such model that also explicitly incorporates clustering by modeling the locations using a finite Gaussian mixture model. Observed covariates and sociality random effects may also be modeled. However, inference for the model via MCMC is cumbersome and thus scaling to large networks is a challenge. Variational Bayesian methods offer an alternative inference methodology for this problem. Sampling based MCMC is replaced by an optimization that requires many orders of magnitude fewer iterations to converge. A Variational Bayesian algorithm for the Latent Position Cluster Model is therefore developed and demonstrated.

1. Introduction

Hoff et al. (2002) introduced latent space models for networks. In this model, the nodes are embedded in an unobserved "social space"; nodes closer together are more likely to link than nodes far apart and inference is performed on the latent positions of the nodes. Links are modeled as occurring independently given the positions of the nodes (and optionally any observed link or node covariates). One appealing characteristic of such models is that they naturally account for reciprocity and transitivity. In addition, plotting the inferred positions of the nodes gives an intuitive visualization of the network.

Handcock et al. (2007) proposed the Latent Position Cluster Model (LPCM) which extends the latent space models to allow for model based clustering of the nodes. This is to accommodate the clustering of nodes in the network beyond that expected from simple transitivity. Clustering is thus included explicitly in the model rather than found by a post-hoc analysis of the estimated node locations. A spherical Gaussian mixture model structure is assumed for the latent positions.

1.1. Motivation for using the variational method

Currently, inference for the LPCM is via MCMC in a Bayesian setting. The disadvantage of this approach is computational and fitting the model to large or even medium size network datasets is impractical or impossible. Variational Bayesian inference offers one approximate solution to this problem. A closed form posterior is found that is "close" to the intractable posterior. This method has already been exploited successfully for other fully Bayesian social network models. We next motivate the contribution in this paper, namely to develop and assess Variational Bayesian algorithms for the LPCM.

* Corresponding authors.
E-mail address: michael.salter-townshend@ucd.ie (M. Salter-Townshend).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.08.004

In the published discussion of Handcock et al. (2007), the following contributions are amongst those made:

- David Blei and Stephen E. Feinberg: "We have appealed to variational methods for a computationally efficient approximation to the posterior (for a mixed membership blockmodel). These methods can scale to large matrices because of the simplified approximation (but at an unknown cost to accuracy). It would be interesting to understand computational trade-offs for the authors method (LPCM) as the sample size grows and when large numbers of covariates are added".
- Dirk Husmeier and Chris Glasbey: "Although a full reversible jump Markov chain Monte Carlo scheme might be computationally prohibitive, variational methods, which are currently very popular in the machine learning community, would presumably provide a much better approximation to the integration and might therefore provide a promising avenue for future research".
- David S. Leslie: "I congratulate the authors for their interesting paper. However, it seems that the Markov chain Monte Carlo sampling scheme that was used results in extremely slow mixing, requiring 2 million iterations with only every 1000th iteration being used".

This work is further prompted by Airoldi et al. (2008) when the authors state "It would be interesting to develop a variational algorithm for the latent space models".

1.2. Specification of the latent position cluster model

In the LSM and LPCM, a binary interactions sociomatrix $\boldsymbol{Y}$ is modeled using logistic regression in which the probability of a link between two nodes depends on the distance between the nodes in the latent space.

$$\log-\text{odds}(y_{i,j}=1|\underline{z}_i,\underline{z}_j,\beta)=\log\left(\frac{\mathbb{P}\{y_{i,j}=1\}}{\mathbb{P}\{y_{i,j}=0\}}\right)=\beta-|\underline{z}_i-\underline{z}_j|, \tag{1}$$

where $y_{i,j}=1$ if node i links with node j and $y_{i,j}=0$ otherwise, β is an intercept parameter and $|\underline{z}_i-\underline{z}_j|$ is the Euclidean distance between the latent positions $\underline{z}_i$ and $\underline{z}_j$ of nodes i and j. The links are assumed to be independent conditional on the latent positions of the nodes in the latent space.

Hence, the probability of the observed network $\boldsymbol{Y}$ given the latent positions $\boldsymbol{Z}=(\underline{z}_1,\ldots,\underline{z}_N)$ of all of the nodes is

$$\mathbb{P}(\boldsymbol{Y}|\beta,\boldsymbol{Z})=\prod_{i=1}^{N}\prod_{\substack{j\neq i\\ j=1}}^{N}\left[\frac{\exp(\beta-|\underline{z}_i-\underline{z}_j|)}{1+\exp(\beta-|\underline{z}_i-\underline{z}_j|)}\right]^{y_{i,j}}\left[\frac{1}{1+\exp(\beta-|\underline{z}_i-\underline{z}_j|)}\right]^{(1-y_{i,j})}.$$

Note that if the network is undirected then the product term is taken over $i<j$ instead of $i\neq j$.

For the LPCM, in order to represent clustering of nodes in the network, the latent positions $\boldsymbol{Z}$ are modeled as coming from a mixture of G multivariate normal distributions.

$$\underline{z}_i\sim\sum_{g=1}^{G}\lambda_g\,\text{MVN}_d(\underline{\mu}_g,\sigma_g^2\mathbf{I}_d), \tag{2}$$

where λ_g is the probability that a node belongs to the gth group and $\mathbf{I}_d$ is the $d\times d$ identity matrix. We let $\underline{\lambda}=(\lambda_1,\lambda_2,\ldots,\lambda_G)$, $\boldsymbol{\mu}=(\underline{\mu}_1,\underline{\mu}_2,\ldots,\underline{\mu}_G)$ and $\underline{\sigma}^2=(\sigma_1^2,\sigma_2^2,\ldots,\sigma_G^2)$.

In order to fit the model in a Bayesian setting, the following hierarchical priors are assumed for β, $\underline{\lambda}$, $\underline{\sigma}$ and $\boldsymbol{\mu}$:

$$\underline{\lambda}\sim\text{Dirichlet}(\underline{\nu}), \tag{3}$$

$$\beta\sim\text{Normal}(\xi,\psi^2), \tag{4}$$

$$\underline{\mu}_g\sim\text{MVN}_d(0,\omega^2\mathbf{I}_d), \tag{5}$$

and

$$\sigma_g^2\sim\sigma_0^2\text{ Inverse }\chi_\alpha^2 \tag{6}$$

where the values ξ, ψ^2, $\underline{\nu}$, σ_0^2, α and ω^2 are fixed prior hyper-parameters.

Hence, the posterior of the latent positions and the model parameters is given by,

$$p(\boldsymbol{Z},\underline{\lambda},\beta,\boldsymbol{\mu},\underline{\sigma}^2|\boldsymbol{Y})=Cp(\boldsymbol{Y}|\beta,\boldsymbol{Z})p(\boldsymbol{Z}|\underline{\lambda},\boldsymbol{\mu},\underline{\sigma}^2\mathbf{I}_d)p(\underline{\lambda}|\underline{\nu})p(\beta|\xi,\psi^2)p(\boldsymbol{\mu}|0,\omega^2\mathbf{I}_d)p(\underline{\sigma}^2|\sigma_0^2,\alpha) \tag{7}$$

where the proportionality constant C is unknown and therefore the posterior is only known up to proportionality.

1.3. Variational Bayesian inference

We develop a Variational Bayesian inference procedure for approximating the posterior distribution of the latent variables in the LPCM. This approach facilitates the application of the LPCM to larger networks than is currently possible using the existing MCMC methodology for model fitting. The faster algorithm also permits more ambitious exploration of the model space and the development of richer modeling.

We loosely follow the method of Airoldi et al. (2008) who use a variational approximation to fit a mixed-membership stochastic blockmodel for networks. A closed form distribution $q(\boldsymbol{Z}, \boldsymbol{K}, \underline{\lambda}, \beta, \boldsymbol{\mu}, \underline{\sigma}^2|\boldsymbol{Y})$ is formed, with unknown variational parameters. These parameters are then optimized by minimization of the Kullback–Leibler divergence from q to the true posterior p given in Eq. (7), which is known only up to proportionality.

The Kullback–Leibler divergence (KL) is defined by

$$\mathrm{KL} = \mathbb{E}_q[\log(q)] - \mathbb{E}_q[\log(p)]. \tag{8}$$

Note that minimization of KL does not require knowledge of the normalization constant of the true posterior because $\mathbb{E}_q[\log(Cp)] = \mathbb{E}_q[\log(p)] + \log(C)$ for all $C > 0$.

This minimization is achieved via an iterative search algorithm that is similar to the Expectation–Maximization algorithm (Dempster et al., 1977). Although both methods scale as $\mathcal{O}(N^2)$, the computational overhead required to find the (potentially local) optimal variational posterior is far less than for sampling based methods such as MCMC as the conditional optimization requires comparatively few iterations to converge to a minimum. This variational posterior is then a closed-form approximation to the true posterior.

2. Variational Bayesian inference for the LPCM

We apply the Variational Bayesian inference methodology (see Beal, 2003, for example) to the latent position cluster model for network data. We have succeeded in deriving straightforward updates for a subset of the variational parameters in our algorithm. The others involve low dimensional numerical optimizations; see Section 2.3 and Appendix B for details. The non-standard part of the Variational Bayesian method for this model involves an approximation to the log-likelihood for the links and non-links, given the latent positions (see Section 2.2). All other elements of the Kullback–Leibler (KL) divergence calculation are available in the literature, given the form of the variational posterior.

2.1. Specification of the variational model

Using the restricted or quasi Variational Bayesian approach (see for example Šmídl and Quinn (2006, Section 3.4.3)), we propose a variational posterior in the same distributional form as the prior, but with unknown variational parameters. This fully-factorized form is known as a mean-field approximation (see Consonni and Marin, 2007 for details and Armagan and Zaretzki, 2011 for an important note). Fully-factorized forms are a common choice in Variational Bayesian inference as it renders tractable the expectations required when evaluating the Kullback–Leibler divergence in Eq. (8).

We will distinguish the variational parameters with a tilde. Hence, our variational posterior for the latent position cluster model for binary interaction networks is of the form:

$$\begin{aligned} q &= q(\boldsymbol{Z}, \boldsymbol{K}, \underline{\lambda}, \beta, \boldsymbol{\mu}, \underline{\sigma}^2|\boldsymbol{Y}) \\ &= q(\boldsymbol{Z}|\tilde{\boldsymbol{Z}}, \underline{\tilde{\sigma}}^2) q(\boldsymbol{K}|\tilde{\boldsymbol{\lambda}}) q(\underline{\lambda}|\underline{\tilde{\nu}}) q(\beta|\tilde{\xi}, \tilde{\psi}^2) q(\boldsymbol{\mu}|\underline{\tilde{\eta}}, \underline{\tilde{\omega}}^2) q(\underline{\sigma}^2|\underline{\tilde{\alpha}}), \end{aligned} \tag{9}$$

where the six distributions appearing in Eq. (9) are of the same functional form as the priors in Eqs. (2)–(6) respectively.

A variational posterior with fewer terms can be constructed comprising only distributions on the latent positions $\boldsymbol{Z}$, the cluster membership indicators $\boldsymbol{K}$ and the intercept term β. However, we wish to capture as much of the detail in the original MCMC posterior as possible, hence the inclusion of the other terms. Without these, our variational posterior would not inform us on the size and location of the groups/clusters. Eq. (9) represents a fully Bayesian variational posterior. Typically, the factorized variational posterior gives an approximation to the true posterior but has support that is too compact (see Consonni and Marin, 2007; Bishop, 2006, Section 10.1.2).

The expectation of the log-likelihood with respect to this variational posterior is intractable. We next present an approximation to this expectation.

2.2. Expectation of the log-likelihood

A tractable approximation of the expected log-likelihood is derived by using a series of three first order Taylor expansions to get:

$$\begin{aligned} \mathbb{E}_q[\log(p(\boldsymbol{Y}|\beta, \boldsymbol{Z}))] &= \sum_i^N \sum_j^N y_{i,j} \mathbb{E}_q[\beta - |\underline{z}_i - \underline{z}_j|] - \mathbb{E}_q\left[\log\left(1 + \exp(\beta - |\underline{z}_i - \underline{z}_j|)\right)\right] \\ &\simeq \sum_i^N \sum_j^N y_{i,j} \mathbb{E}_q[\beta - |\underline{z}_i - \underline{z}_j|] - \log\left(1 + \mathbb{E}_q[\exp(\beta - |\underline{z}_i - \underline{z}_j|)]\right) \end{aligned}$$

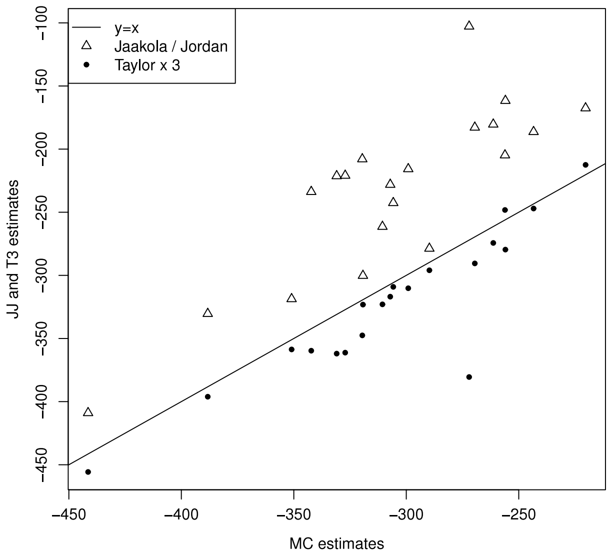

Fig. 1. Comparison of the accuracy of the approximate expectation given by Eq. (10) and the bound derived in Jaakkola and Jordan (2000). 20 draws from the variational posterior given the model fit to Sampson's monks network were drawn. Both approximate expected log-likelihoods were evaluated at these latent variable values along with an unbiased, but expensive Monte-Carlo approximation using 1000 samples. The approximate expectations are plotted against the Monte-Carlo results for various samples from the variational posterior.

$$
\begin{aligned}
&= \sum_i^N \sum_j^N y_{i,j}\left(\tilde{\xi} - \left(|\tilde{\underline{z}}_i - \tilde{\underline{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right) \\
&\quad - \log\left(1 + \mathbb{E}_{q(\beta)}[\exp(\beta)]\mathbb{E}_{q(z)}[\exp(-|\underline{z}_i - \underline{z}_j|)]\right) \\
&\simeq \sum_i^N \sum_j^N y_{i,j}\left(\tilde{\xi} - \left(|\tilde{\underline{z}}_i - \tilde{\underline{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right) \\
&\quad - \log\left(1 + \exp\left(\tilde{\xi} + \frac{\tilde{\psi}^2}{2} + \mathbb{E}_{q(z)}[-|\underline{z}_i - \underline{z}_j|]\right)\right) \\
&\simeq \sum_i^N \sum_j^N y_{i,j}\left(\tilde{\xi} - \left(|\tilde{\underline{z}}_i - \tilde{\underline{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right) \\
&\quad - \log\left(1 + \exp\left(\tilde{\xi} + \frac{\tilde{\psi}^2}{2} - \left(|\tilde{\underline{z}}_i - \tilde{\underline{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right)\right). \qquad (10)
\end{aligned}
$$

This approximation is simpler than the variational bound used in Jaakkola and Jordan (2000) and is quicker to compute. A simple simulation study shows that our approximation is also more accurate (Fig. 1). The three approximations in Eq. (10) are in fact an upper bound, a lower bound and a lower bound respectively. Following a fit of the LPCM to Sampson's monks network (Sampson, 1969), exhaustive Monte-Carlo simulation was used to evaluate the log-likelihood for the network for various q_β and q_z values. These were compared with the approximation in Eq. (10) and the Jaakkola–Jordan bound. For a derivation of the approximate expected log-likelihood using the Jaakkola–Jordan bound see Appendix A. Unlike the Jaakkola–Jordan bound, our approximation may return results that are above or below the true log-likelihood but is closer in expected absolute terms.

2.3. Updating the variational parameters

The variational parameters are optimized to minimize the Kullback–Leibler divergence from q to p. In order to do this, we compute the approximate KL divergence and then differentiate the resulting expression with respect to each of the variational parameters in turn.

Before commencing the minimization procedure, an initial configuration of the variational parameters is chosen (see Section 2.4). The algorithm then iterates through the set of variational parameters performing minimizations with respect to a subset of the parameters, keeping the remaining parameters fixed. Convergence to a steady state (local minimum on the KL surface) is reached after a low number of iterations, depending on the convergence criterion used.

For the variational parameters $\tilde{\boldsymbol{\lambda}}$, $\tilde{\eta}$, $\underline{\tilde{\omega}}^2$, a closed form update equation was found. For all other variational parameters, numerical optimization was required. A fast bracketing-and-bisection algorithm (e.g. Press et al., 1992, Chapter 9.1) was used to find the minimum KL divergence for each scalar variational parameter. The latent position variational means $\underline{\tilde{z}}_i$ are d dimensional variables and are optimized in our algorithm using a conjugate gradient routine (e.g. Press et al., 1992, Chapter 10.6).

The mathematics relating to the variational approximation to the LPCM for networks is provided in Appendix B. This includes calculation of the KL divergence terms, all first order partial derivatives with respect to the variational parameters and the analytical update equations, where available.

The Variational Bayesian inference algorithm proceeds via iterative conditional minimization as outlined below and is an example of Variational Bayes Expectation Maximization. Analytical updates are available only when the partial derivative of the Kullback–Leibler divergence set equal to zero is re-expressible with the updated term appearing only on one side. Otherwise, a more complicated equation with multiple instances of the updated term is found and numerical optimization must be used (see Section 2.2). In cases where the term being updated is a vector or a matrix the entries are cycled through in a random order for each iteration of the variational algorithm.

- `initialize` $\underline{\tilde{z}}_i, \tilde{\sigma}_i^2, \tilde{\lambda}_{i,g}, \tilde{\nu}_g, \tilde{\xi}, \tilde{\psi}^2, \underline{\tilde{\eta}}_g, \tilde{\omega}_g^2, \tilde{\alpha}_g$ `for all` i, g.
- `repeat in a randomized order:`
  - `update` $\tilde{\xi}$ `via bisection.`
  - `update` $\tilde{\psi}^2$ `via bisection.`
  - `update` $\tilde{\boldsymbol{Z}}$ `via conjugate gradient.`
  - `update` $\underline{\tilde{\eta}}$ `via analytical function.`
  - `update` $\underline{\tilde{\sigma}}^2$ `via bisection.`
  - `update` $\tilde{\boldsymbol{\lambda}}$ `via analytical function.`
  - `update` $\underline{\tilde{\omega}}^2$ `via analytical function.`
  - `update` $\underline{\tilde{\alpha}}$ `via bisection.`
  - `update` $\underline{\tilde{\nu}}$ `via bisection.`
- `until convergence.`

2.4. Initialization method

The use of a good initialization of the variational parameters is crucial to the Variational Bayesian method because the Kullback–Leibler divergence may contain multiple local minima. Starting the algorithm at different values for the variational parameters and updating as per Section 2.3 demonstrates that the algorithm is prone to convergence to local minima. Furthermore, the closer the initial values are to the optimum, the fewer iterations will be required for convergence.

There are two desirable criteria for the initial configuration: the initialization procedure should be fast and the resultant configuration should be as close as possible to the global minimum; this will minimize the chances of the iterative updates becoming stuck in a local minimum. The obvious choice is to use the Maximum Likelihood estimates for the latent positions and the intercept parameter and to apply a clustering algorithm to these. Using *latentnet* we obtain these estimates on a 3 GHz processor machine in 7.2 s for Sampson's monks dataset (18 nodes). We consider this too slow, and therefore present an approximation to the MLEs for the latent positions and the intercept that is an iterative optimization. Our method delivers results similar to the MLEs in a much shorter time (0.38 s for the same dataset on the same machine).

Our initialization procedure is inspired by the Fruchterman and Reingold (1991) network visualization method, which updates a random initial configuration by allowing inter-node forces to act on the nodes. Within the Fruchterman–Reingold method, attractive forces exist between linked nodes which are assumed to be analogous to a spring force. A repulsive force which is analogous to an electrical force exists between all node pairs, whether linked or not.

We adapt the Fruchterman–Reingold method to find initial positions for nodes in the latent space. The method uses the log-likelihood terms for these forces; i.e. there is an attractive force between linked node pairs i and j equal to $\beta - |\underline{z}_i - \underline{z}_j|$ and a repulsive force between all node pairs equal to $\log\left(1 + \exp(\beta - |\underline{z}_i - \underline{z}_j|)\right)$. The positions $\boldsymbol{Z}$ are randomly initialized and then, cycling through all $\underline{z}_i, \underline{z}_j$ pairs, the forces are calculated and the positions updated accordingly. After a number of cycles, the intercept parameter β is updated via maximum-likelihood. These steps may be iterated until convergence. This is followed by the use of the model-based clustering methods implemented in *mclust* (Fraley and Raftery, 2000, 2003) for the initialization of the clustering variables.

The comparison between this initialization method and the direct optimization of the log-likelihood may be extended to a comparison between our overall method and a direct optimization of the total Kullback–Leibler divergence. Again, as speed is our major concern in this work we prefer the iterative algorithm we have presented in Section 2.3 to a simultaneous optimization over all parameters, using off-the-shelf software. Indeed, since the closed-form Variational approximation to the full-posterior is a product of independent parts, our method may be thought of as a conjugate descent algorithm, tailored to the LPCM.

2.5. Case-control approximation to the likelihood

Raftery et al. (in press) demonstrate an approximation to the log-likelihood that enables MCMC based inference for the LPCM on larger networks. This approximation involves stratified case-control sampling of the non-links so that the algorithm scales as $\mathcal{O}(N)$. The computational bottleneck in the LPCM (as in all statistical models of networks) lies in calculating the likelihood as this involves indexing over all possible pairs of nodes which is an $\mathcal{O}(N^2)$ calculation which must be performed every time the likelihood is calculated. The case-control approximation replaces the likelihood with a term that is quicker to compute due to the sparsity of the network. All links are considered but a random sample of non-links are used; this is because most of the information is carried by the links.

A stratification depending on geodesic distance between the non-linked nodes in the network is employed. Within each stratum a random sample of node pairs is considered, rather than the full cohort. In Raftery et al. (in press) they use a set small number of samples per stratum; in our work we decided on a fixed proportion: e.g. for the nodes that are not linked and have 3 degrees of separation we sample 10% of all possible dyads if we set our sample ratio to 0.1. This proportion can be set as a trade off between speed and accuracy with 1 being the full likelihood calculation. Our experimentation with this approximation is in agreement with Raftery et al. (in press) that it works well. Although our **R** package *VBLPCM* makes use of this approximation as an option, we do not employ it in obtaining the results in the next section in order to avoid confounding between it and the approximation inherent in using Variational Bayes.

3. Results

We examine a subset of the Irish blogoshphere; the data come from Wade et al. (2011) and comprise 604 blogs based in Ireland. This is a set of blogs surrounding recipients of 2010 Irish Blog Awards. We examine a network based on the "blogroll". This is the list of links arising from the template for the blog (as such it is nearly static and we examine a single snapshot). They are the other blogs which the blogger chooses to list as recommended. There are a total of 4640 such blogroll links in the network. Such a large network cannot be practically analyzed using the existing MCMC methodology.

3.1. Goodness-of-fit plots

Following Krivitsky and Handcock (2008), we present analysis of our algorithm and the model using goodness-of-fit visualizations based on a comparison of high-order statistics of random networks simulated from the fitted model. The model appears to capture the correct in-degree structure, performing less well on out-degree. The minimum geodesic distances of networks simulated from the fitted model tend to over-sample paths of length 2 and 3 but capture paths of other lengths well. Naturally, these results summarize the fit of the LPCM to the data and do not validate our Variational Bayesian algorithm as such; we present the result as an example of an interesting dataset that is amenable to the LPCM but which requires the use of an algorithm more scalable than MCMC. For example, an out-degree of zero is the most common out-degree in the data, but this is due to not all blogs using the blogroll option to list blogs they are interested in, a fact that is not accounted for by the LPCM. One possible correction for this would be to treat as missing all out-links from blogs with an out-degree of zero. Similarly, the low end of the Edgewise Shared Partner distribution is not captured well (see Fig. 2).

3.2. Groups in the blogosphere

Wade et al. (2011) performed a text-based classification of the content of the blogs using an ensemble non-negative matrix factorization algorithm (Greene et al., 2008) and found 12 groups across 475 of the nodes, with the rest labeled unknown. Comparisons between these textual analysis based groupings and the clusters found by our Variational Bayes algorithm are favorable; in fact we obtain over 69% of total agreement when we model 12 clusters in 3-dimensional latent space with social effects. We calculate this value as the mean percentage of maximally assigned blogs: i.e. we look at each group based on the textual analysis in turn; for that group we then examine the vector of maximally assigned cluster memberships (hard-clusterings) obtained from our algorithm; the final comparison is the mean percentage of blogs correctly assigned by our algorithm. As a control, we randomly generated fifty networks and found a mean agreement of 22.87% with a standard deviation of 2.24%. The text-based classifications do not represent a ground truth validation; the large level of agreement between the two complementary methods merely illustrates the interesting result that the blogroll tends to list blogs of similar textual content.

3.3. Exploration of the model space

Model comparison techniques are required to select the number of groups (G) in the Latent Position Cluster Model. Following Handcock et al. (2007), this can be done by selecting the lowest value of the Bayesian Information Criterion (BIC) (Schwarz, 1978) for the number of groups. Using the same calculation of the BIC given in Section 4 of Handcock et al. (2007), we present results for the blogs dataset in a latent space of dimension 2 in Fig. 3. We used multiple random restarts to avoid convergence to a local minimum of the Kullback–Leibler surface. A 25 group model is selected but the BIC has several

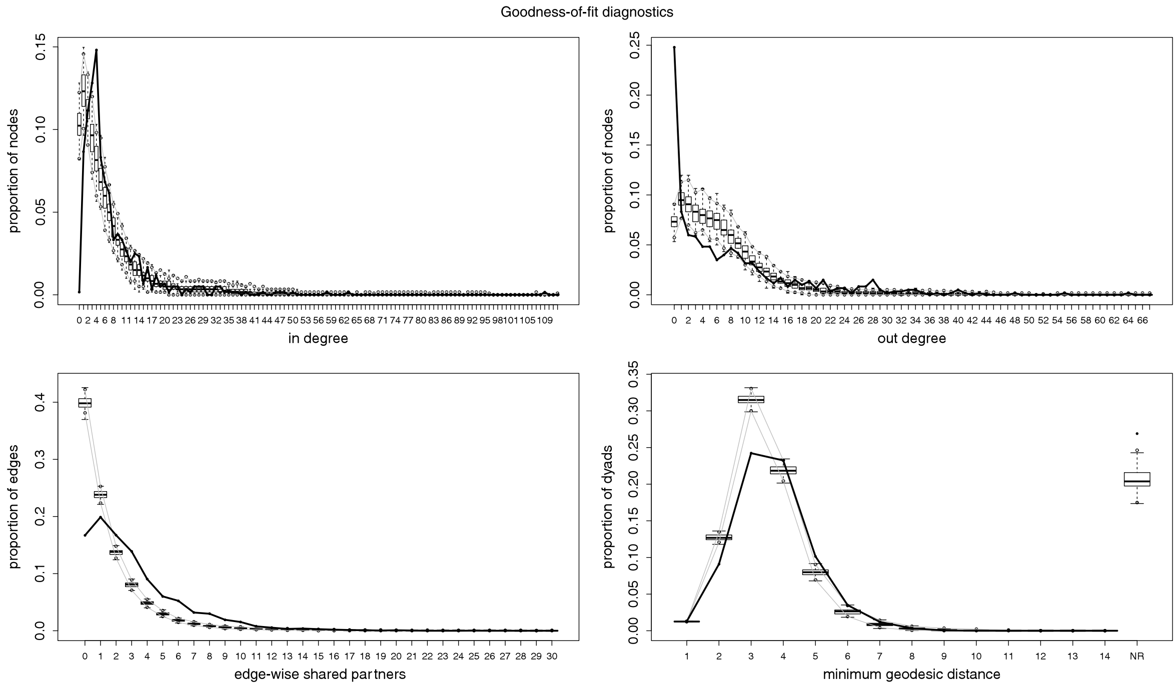

Fig. 2. Goodness of fit diagnostic plots for the blogs dataset with social effects. The heavy line represents summary statistics for the observed network. The boxplots superimposed summarize the same statistics across 100 networks generated from the posterior for all latent variables of the model given the observed network.

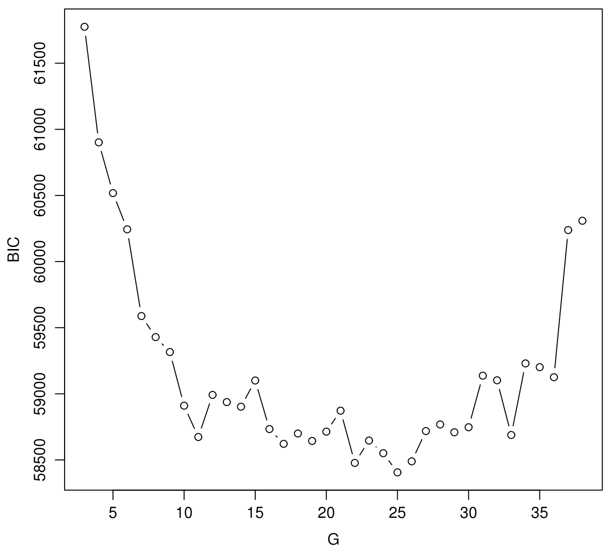

Fig. 3. BIC versus G plot for MCMC and Variational Bayesian inference for the blogs data in 2 latent space dimensions. The 25 group model is selected.

local minima between 10 and 35. This phenomenon is not unexpected; for example, a highly connected set of nodes may be readily modeled using 1 or 4 equi-size spherical groups but less easily so using 2 or 3.

4. Discussion

We have presented a Variational Bayesian algorithm to lessen the computational burden associated with Bayesian analysis of social networks using the Latent Position Cluster Model. Although our method is approximate, it captures the essentials of the current standard MCMC methodology with much less computational burden.

Our algorithm converges in a short number of fast iterations so we can analyse larger graphs than are possible with the MCMC method. Our primary contribution is thus to present a practical inference methodology for the LPCM rather than development of a new model for network data. Although we have only discussed our method in terms of binary link data, our method extends readily to other network data types (as does the MCMC based methodology). Our **R** package for fitting the LPCM using our Variational Bayes algorithm is available on the CRAN *R* repository at http://cran.r-project.org/web/packages/VBLPCM/index.html.

Acknowledgment

This material is based upon works supported by the Science Foundation Ireland under Grant No. 08/SRC/I1407: Clique: Graph & Network Analysis Cluster.

Appendix A. Jaakkola–Jordan Bound

Starting out with the Jaakkola and Jordan (2000) paper (and subbing $-x$ for x):

$$\begin{aligned} -\log(1+\exp(x)) &= \log\left(g(-x)\right) \\ &= -\frac{x}{2} - \log\left(\exp\left(\frac{x}{2}\right) + \exp\left(-\frac{x}{2}\right)\right) \\ &= -\frac{x}{2} + f(x). \end{aligned} \tag{A.1}$$

Note that $f(x)$ is convex in x^2, not x. Use a first order Taylor expansion in x^2 to get a lower bound:

$$\begin{aligned} f(x) &\geq f(a) + \frac{\partial f(a)}{\partial a^2}(x^2 - a^2) \\ &= -\frac{a}{2} + \log(g(-a)) + \frac{1}{4a}\tanh\left(\frac{a}{2}\right)(x^2 - a^2) \end{aligned} \tag{A.2}$$

which is Eq. (6) in Jaakkola and Jordan (2000).

We can now use this bound in finding the expectation w.r.t. $q(\beta, \boldsymbol{Z})$ of $-\log(1+\exp(\beta - |\underline{z}_i - \underline{z}_j|))$ by setting x to be $\beta - |\underline{z}_i - \underline{z}_j|$.

We expand around the expectation of $\beta - |\underline{z}_i - \underline{z}_j|$ squared: i.e. set

$$a = \tilde{\xi} - \left(|\underline{\tilde{z}}_i - \underline{\tilde{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}. \tag{A.3}$$

Therefore:

$$\begin{aligned} -\log(1+\exp(\beta - |\underline{z}_i - \underline{z}_j|)) &= \log\left(g(-\beta + |\underline{z}_i - \underline{z}_j|)\right) \\ &= -\frac{(\beta - |\underline{z}_i - \underline{z}_j|)}{2} + f(\beta - |\underline{z}_i - \underline{z}_j|) \\ &\geq \frac{(\beta - |\underline{z}_i - \underline{z}_j|)}{2} + f(a) + \frac{\partial f(a)}{\partial a^2}(x^2 - a^2) \\ &= -\frac{(\beta - |\underline{z}_i - \underline{z}_j|)}{2} - \frac{a}{2} + \log(g(-a)) + \frac{1}{4a}\tanh\left(\frac{a}{2}\right)(x^2 - a^2) \\ &= -\frac{(\beta - |\underline{z}_i - \underline{z}_j|)}{2} - \frac{a}{2} + \frac{a}{2} - \log\left(\exp\left(\frac{a}{2}\right) + \exp\left(-\frac{a}{2}\right)\right) \\ &\quad + \frac{1}{4a}\tanh\left(\frac{a}{2}\right)(x^2 - a^2) \\ &= -\frac{(\beta - |\underline{z}_i - \underline{z}_j|)}{2} - \log\left(\exp\left(\frac{a}{2}\right) + \exp\left(-\frac{a}{2}\right)\right) \\ &\quad + \frac{1}{4a}\tanh\left(\frac{a}{2}\right)(x^2 - a^2). \end{aligned} \tag{A.4}$$

We require the expectation of the above, w.r.t. $q(\beta)$ and $q(\boldsymbol{Z})$. There are still occurrences of $x = \beta - |\underline{z}_i - \underline{z}_j|$ in this equation. The expectation of x is a. The expectation of $x^2 - a^2$ is:

$$\begin{aligned} \mathbf{E}[\beta^2 + |\underline{z}_i - \underline{z}_j|^2 - 2\beta|\underline{z}_i - \underline{z}_j| - a^2] &= \tilde{\xi}^2 + \tilde{\Psi}^2 + |\underline{\tilde{z}}_i - \underline{\tilde{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2) - 2\tilde{\xi}\sqrt{|\underline{\tilde{z}}_i - \underline{\tilde{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)} - a^2 \\ &= \tilde{\psi}^2. \end{aligned} \tag{A.5}$$

Eq. (A.4) then becomes:

$$-\frac{a}{2} - \log\left(\exp\left(\frac{a}{2}\right) + \exp\left(-\frac{a}{2}\right)\right) + \frac{1}{4a}\tanh\left(\frac{a}{2}\right)(\Psi^2). \tag{A.6}$$

This is the approximation used in the simulation experiment depicted in Fig. 1.

Appendix B. Variational parameter updates

B.1. Kullback–Leibler divergence

$$KL_{q\|p} = \mathbb{E}_q[\log(p)] - \mathbb{E}_q[\log(q)]. \tag{B.1}$$

Setting the notation that:

$$\hat{p}_i = \mathbb{E}_q[\log(p_i)]. \tag{B.2}$$

$$\hat{q}_i = \mathbb{E}_q[\log(q_i)]. \tag{B.3}$$

Note that the log of the factorizable p and q posteriors is a sum of logs.

It now remains to find the sum of $\hat{p}_1, \hat{p}_2, \hat{q}_2, \hat{p}_3, \hat{q}_3$ with the differences $\hat{p}_4 - \hat{q}_4, \hat{p}_5 - \hat{q}_5, \hat{p}_6 - \hat{q}_6, \hat{p}_7 - \hat{q}_7$ simply available from the literature on Kullback–Leibler divergences between these forms of distributions (the other terms involving more than one expectation over the model parameters). Note that $\hat{p}_i$ is the expected log-likelihood.

$$KL_{q\|p} = \hat{p}_1 + \hat{p}_2 - \hat{q}_2 + \hat{p}_3 - \hat{q}_3 + (\hat{p}_4 - \hat{q}_4) + (\hat{p}_5 - \hat{q}_5) + (\hat{p}_6 - \hat{q}_6) + (\hat{p}_7 - \hat{q}_7). \tag{B.4}$$

$$\begin{aligned}
\hat{p}_1 &= \sum_i^N \sum_j^N y_{i,j} \mathbf{E}_{q_{2,5}}[\beta - |\underline{z}_i - \underline{z}_j|] - \mathbf{E}_{q_{2,5}}\left[\log\left(1 + \exp(\beta - |\underline{z}_i - \underline{z}_j|)\right)\right] \\
&\simeq \sum_i^N \sum_j^N y_{i,j} \mathbf{E}_{q_{2,5}}[\beta - |\underline{z}_i - \underline{z}_j|] - \log\left(1 + \mathbf{E}_{q_{2,5}}[\exp(\beta - |\underline{z}_i - \underline{z}_j|)]\right) \\
&= \sum_i^N \sum_j^N y_{i,j}\left(\tilde{\xi} - \left(|\underline{\tilde{z}}_i - \underline{\tilde{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right) - \log\left(1 + \mathbf{E}_{q_5}[\exp(\beta)]\mathbf{E}_{q_2}[\exp(-|\underline{z}_i - \underline{z}_j|)]\right) \\
&\simeq \sum_i^N \sum_j^N y_{i,j}\left(\tilde{\xi} - \left(|\underline{\tilde{z}}_i - \underline{\tilde{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right) - \log\left(1 + \exp\left(\tilde{\xi} + \frac{\tilde{\psi}^2}{2} + \mathbf{E}_{q_2}[-|\underline{z}_i - \underline{z}_j|]\right)\right) \\
&\simeq \sum_i^N \sum_j^N y_{i,j}\left(\tilde{\xi} - \left(|\underline{\tilde{z}}_i - \underline{\tilde{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right) \\
&\quad - \log\left(1 + \exp\left(\tilde{\xi} + \frac{\tilde{\psi}^2}{2} - \left(|\underline{\tilde{z}}_i - \underline{\tilde{z}}_j|^2 + d(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2)\right)^{\frac{1}{2}}\right)\right).
\end{aligned} \tag{B.5}$$

$$\begin{aligned}
\hat{p}_2 &= \mathbf{E}_{q_{3,7,6,2}}[\log(p_2)] \\
&= \mathbf{E}_{q_{3,7,6,2}}\left[-\sum_i^N \frac{(\underline{z}_i - \underline{k}_i\underline{\mu})^T(\underline{z}_i - \underline{k}_i\underline{\mu})}{2\underline{k}_i\sigma^2} - d\log \underline{k}_i\sigma^2\right] + \text{constant} \\
&= \mathbf{E}_{q_{3,7,6}}\left[-\sum_i^N \tilde{\sigma}_i^2 + \frac{(\underline{\tilde{z}}_i - \underline{k}_i\underline{\mu})^T(\underline{\tilde{z}}_i - \underline{k}_i\underline{\mu})}{2K_i\sigma^2} - d\log \underline{k}_i\sigma^2\right] + \text{constant} \\
&= \mathbf{E}_{q_{3,7}}\left[-\sum_i^N \tilde{\sigma}_i^2 + \underline{k}_i\underline{\tilde{\omega}}^2 + \frac{(\underline{\tilde{z}}_i - \underline{k}_i\underline{\tilde{\eta}})^T(\underline{\tilde{z}}_i - \underline{k}_i\underline{\tilde{\eta}})}{2K_i\sigma^2} - d\log \underline{k}_i\sigma^2\right] + \text{constant} \\
&= \mathbf{E}_{q_3}\left[-\sum_i^N \frac{\underline{k}_i\underline{\tilde{\alpha}}}{2\sigma_0^2}\left(\tilde{\sigma}_i^2 + \underline{k}_i\underline{\tilde{\omega}}^2 + (\underline{\tilde{z}}_i - \underline{k}_i\underline{\tilde{\eta}})^T(\underline{\tilde{z}}_i - \underline{k}_i\underline{\tilde{\eta}})\right) + d\Psi\left(\frac{\underline{k}_i\underline{\tilde{\alpha}}}{\sigma_0^2}\right)\right] + \text{constant} \\
&= \sum_g^G \sum_i^N \tilde{\lambda}_{i,g}\left(d\Psi\left(\frac{\tilde{\alpha}_g}{\sigma_0^2}\right) - \frac{\tilde{\alpha}_g}{2\sigma_0^2}\left(\tilde{\sigma}_i^2 + \tilde{\omega}_g^2 + (\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)^T(\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)\right)\right) + \text{constant}.
\end{aligned} \tag{B.6}$$

$$\hat{p}_3 = \sum_g^G \sum_i^N \tilde{\lambda}_{i,g}\left(\Psi\left(\tilde{\nu}_g\right) - \Psi\left(\sum_j^G \tilde{\nu}_j\right)\right) \quad \text{(Schervish, 1995)}. \tag{B.7}$$

$$-\hat{q}_2 = \frac{N}{2}\left(1 + \log(2\pi) + \frac{d}{2}\sum_i^N \log(\tilde{\sigma}_i^2)\right). \tag{B.8}$$

$$-\hat{q}_3 = -\sum_g^G \sum_i^N \tilde{\lambda}_{i,g}\log(\tilde{\lambda}_{i,g}). \tag{B.9}$$

$$\hat{p}_4 - \hat{q}_4 = -\log\left(\Gamma\left(\sum_g \tilde{\nu}_g\right)\right) + \log\left(\Gamma\left(\sum_g \nu_g\right)\right) + \sum_g (\log(\Gamma(\tilde{\nu}_g)) - \log(\Gamma(\nu_g)) - (\tilde{\nu}_g - \nu_g)(\Psi(\tilde{\nu}_g) - \Psi(\nu_g))). \tag{B.10}$$

$$\hat{p}_5 - \hat{q}_5 = \frac{1}{2}\left(d\log\left(\frac{\tilde{\psi}^2}{\psi^2}\right) - d\left(\frac{\tilde{\psi}^2}{\psi^2}\right) - \frac{(\tilde{\xi}-\xi)^T(\tilde{\xi}-\xi)}{\psi^2} + d\right). \tag{B.11}$$

$$\hat{p}_6 - \hat{q}_6 = \sum_g^G \frac{1}{2}\left(d\log\left(\frac{\tilde{\omega}_g^2}{\omega^2}\right) - d\left(\frac{\tilde{\omega}_g^2}{\omega^2}\right) - \frac{(\underline{\tilde{\eta}}_g - 0)^T(\underline{\tilde{\eta}}_g - 0)}{\omega^2} + d\right). \tag{B.12}$$

$$\hat{p}_7 - \hat{q}_7 = \sum_g^G \log\left(\Gamma\left(\frac{\tilde{\alpha}_g}{2}\right)\right) - \log\left(\Gamma\left(\frac{\alpha_g}{2}\right)\right) + \frac{(\alpha_g - \tilde{\alpha}_g)}{2}\left(\Psi\left(\frac{\tilde{\alpha}_g}{2}\right)\right). \tag{B.13}$$

B.2. First order partial derivatives

with $a = \tilde{\xi} - \left((\underline{\tilde{z}}_i - \underline{\tilde{z}}_j)^T(\underline{\tilde{z}}_i - \underline{\tilde{z}}_j) + d(\tilde{\sigma}_i^2 + S_j^2)\right)^{\frac{1}{2}}$:

$$\frac{\partial \mathrm{KL}}{\partial \underline{\tilde{z}}_i} = \sum_j^N \frac{\underline{\tilde{z}}_i - \underline{\tilde{z}}_j}{a}\left(y_{i,j} - \frac{1}{1+\exp\left\{-a - \frac{\tilde{\psi}^2}{2}\right\}}\right) + \sum_g^G (\underline{\tilde{\eta}}_g - \underline{\tilde{z}}_i)\tilde{\lambda}_{i,g}\frac{\tilde{\alpha}_g}{2\sigma_0^2}. \tag{B.14}$$

$$\frac{\partial \mathrm{KL}}{\partial \tilde{\sigma}_i^2} = \sum_j \frac{d}{\tilde{\sigma}_i^2\left(1+\exp\left\{-a-\frac{\tilde{\psi}^2}{2}\right\}\right)} - y_{i,j}d - \sum_g^G \tilde{\lambda}_{i,g}\frac{\tilde{\alpha}_g}{2\sigma_0^2} + \frac{d}{\tilde{\sigma}_i^2}. \tag{B.15}$$

$$\frac{\partial \mathrm{KL}}{\partial \tilde{\lambda}_{i,g}} = d\Psi\left(\frac{\tilde{\alpha}_g}{2\sigma_0^2}\right) - \frac{\tilde{\alpha}_g}{2\sigma_0^2}(\tilde{\sigma}_i^2 + \tilde{\omega}_g^2 + (\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)^T(\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)) + \Psi(\tilde{\nu}_g) - \Psi\left(\sum_j^G \tilde{\nu}_j\right) - 1 - \log(\tilde{\lambda}_{i,g}). \tag{B.16}$$

$$\frac{\partial \mathrm{KL}}{\partial \underline{\tilde{\eta}}_g} = \sum_i^N \tilde{\lambda}_{i,g}\frac{\tilde{\alpha}_g}{2\sigma_0^2}(\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)^T - \frac{1}{2\omega^2}(\underline{\tilde{\eta}}_g - 0)^T. \tag{B.17}$$

$$\frac{\partial \mathrm{KL}}{\partial \tilde{\omega}_g^2} = \sum_i^N -\frac{\tilde{\alpha}_g}{2\sigma_0^2}\tilde{\lambda}_{i,g} - \frac{d}{2\omega^2} + \frac{d}{2\tilde{\omega}_g^2}. \tag{B.18}$$

$$\frac{\partial \mathrm{KL}}{\partial \tilde{\alpha}_g} = \sum_i^N \tilde{\lambda}_{i,g}\left(\frac{d}{\sigma_0^2}\Psi'\left(\frac{\tilde{\alpha}_g}{2\sigma_0^2}\right) - \frac{1}{2\sigma_0^2}(\tilde{\sigma}_i^2 + \tilde{\omega}_g^2 + (\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)^T(\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g))\right) + \frac{(\tilde{\alpha}_g - \alpha_g)}{2}\Psi'\left(\frac{\tilde{\alpha}_g}{2}\right) - \Psi\left(\frac{\tilde{\alpha}_g}{2}\right) + \frac{\Psi\left(\frac{\tilde{\alpha}_g}{2}\right)}{\log\left(\Gamma\left(\frac{\tilde{\alpha}_g}{2}\right)\right)}. \tag{B.19}$$

$$\frac{\partial \mathrm{KL}}{\partial \tilde{\nu}_g} = \sum_i^N \tilde{\lambda}_{i,g}\left(\Psi'(\tilde{\nu}_g) - \Psi'\left(\sum_j^G \tilde{\nu}_j\right)\right) - \Psi\left(\sum_j^G \tilde{\nu}_j\right) - \Psi(\nu_g) - \tilde{\nu}_g\Psi'(\tilde{\nu}_g) + \nu_g\Psi'(\tilde{\nu}_g). \tag{B.20}$$

$$\frac{\partial \mathrm{KL}}{\partial \tilde{\xi}} = \sum_i^N \sum_j^N y_{i,j} - \frac{1}{1+\exp\left\{-a-\frac{\tilde{\psi}^2}{2}\right\}} - \frac{\tilde{\xi}-\xi}{2\psi^2}. \tag{B.21}$$

$$\frac{\partial \mathrm{KL}}{\partial \tilde{\psi}^2} = -\frac{1}{2\left(1+\exp\left(-a-\frac{\tilde{\psi}^2}{2}\right)\right)} + \frac{D}{2}\left(\frac{1}{\tilde{\psi}^2} - \frac{1}{\psi^2}\right). \tag{B.22}$$

Eqs. (B.16)–(B.18) may be updated directly. The $\tilde{\boldsymbol{Z}}$ are optimized using conjugate gradient. Updates based on minimization are denoted with an $\leftarrow$. Others are with an = sign. Thus:

$$\underline{\tilde{z}}_i \leftarrow \text{minimize KL}(\underline{\tilde{z}}_i) \text{ w.r.t. } \underline{\tilde{z}}_i \text{ using Eq. (B.14).} \tag{B.23}$$

$$\tilde{\sigma}_i^2 \leftarrow \text{minimize KL}(\tilde{\sigma}_i^2) \text{ w.r.t. } \tilde{\sigma}_i^2 \text{ using Eq. (B.15)}. \tag{B.24}$$

$$\tilde{\lambda}_{i,g} \propto \exp\left\{-1 + \Psi\left(\frac{\tilde{\alpha}_g}{2\sigma_0^2}\right) - \frac{\tilde{\alpha}_g}{2\sigma_0^2}(\tilde{\sigma}_i^2 + \tilde{\omega}_g^2 + (\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)^T(\underline{\tilde{z}}_i - \underline{\tilde{\eta}}_g)) + \Psi(\tilde{\nu}_g) - \Psi\left(\sum_j^G(\tilde{\nu}_j)\right)\right\}. \tag{B.25}$$

$$\underline{\tilde{\eta}}_g = \frac{\sum_i^N \tilde{\lambda}_{i,g}\frac{\tilde{\alpha}_g}{2\sigma_0^2}\underline{\tilde{z}}_i + \frac{0}{2\omega^2}}{\sum_i^N \tilde{\lambda}_{i,g}\frac{\tilde{\alpha}_g}{2\sigma_0^2} + \frac{1}{2\omega^2}}. \tag{B.26}$$

$$\tilde{\omega}_g^2 = \left(\frac{\sum_i^N(\tilde{\lambda}_{i,g}\tilde{\alpha}_g)}{d\sigma_0^2} + \frac{1}{\omega^2}\right)^{-1}. \tag{B.27}$$

$$\tilde{\alpha}_g \leftarrow \text{minimize KL}(\tilde{\alpha}_g) \text{ w.r.t. } \tilde{\alpha}_g \text{ using Eq. (B.19)}. \tag{B.28}$$

$$\tilde{\nu}_g \leftarrow \text{minimize KL}(\tilde{\nu}_g) \text{ w.r.t. } \tilde{\nu}_g \text{ using Eq. (B.20)}. \tag{B.29}$$

$$\tilde{\xi} \leftarrow \text{minimize KL}(\tilde{\xi}) \text{ w.r.t. } \tilde{\xi} \text{ using Eq. (B.21)}. \tag{B.30}$$

$$\tilde{\psi}^2 \leftarrow \text{minimize KL}(\tilde{\psi}^2) \text{ w.r.t. } \tilde{\psi}^2 \text{ using Eq. (B.22)}. \tag{B.31}$$

References

Airoldi, E., Blei, D., Fienberg, S., Xing, E., 2008. Mixed-membership stochastic blockmodels. Journal of Machine Learning Research 9, 1981–2014.

Armagan, A., Zaretzki, R.L., 2011. A note on mean-field variational approximations in Bayesian probit models. Computational Statistics & Data Analysis 55 (1), 641–643.

Beal, M., 2003. Variational algorithms for approximate Bayesian inference. Ph.D. Thesis. University College London.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer.

Consonni, G., Marin, J.-M., 2007. Mean-field variational approximate Bayesian inference for latent variable models. Computational Statistics & Data Analysis 52 (2), 790–798.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39 (1), 1–38.

Fraley, C., Raftery, A.E., 2000. Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association 97, 611–631.

Fraley, C., Raftery, A.E., 2003. Enhanced model-based clustering, density estimation and discriminant analysis software: mclust. Journal of Classification 20, 263–286.

Fruchterman, T., Reingold, E., 1991. Graph drawing by force-directed placement. Software—Practice and Experience 21 (11), 1129–1164.

Greene, D., Cagney, G., Krogan, N., Cunningham, P., 2008. Ensemble non-negative matrix factorization methods for clustering protein–protein interactions. Bioinformatics 24, 1722–1728.

Handcock, M.S., Raftery, A., Tantrum, J., 2007. Model-based clustering for social networks. Journal of the Royal Statistical Society: Series A 170 (2), 1–22.

Hoff, P., Raftery, A., Handcock, M.S., 2002. Latent space approaches to social network analysis. Journal of the American Statistical Association 97, 1090–1098.

Jaakkola, T.S., Jordan, M.I., 2000. Bayesian parameter estimation via variational methods. Statistics and Computing 10, 25–37.

Krivitsky, P.N., Handcock, M.S., 2008. Fitting position latent cluster models for social networks with latentnet. Journal of Statistical Software 24 (5), 1–23.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Computing, second ed. Cambridge University Press, Cambridge, UK.

Raftery, A.E., Niu, X., Hoff, P., Yeung, K.Y., 2012. Fast inference for the latent space network model using a case-control approximate likelihood. Journal of Computational and Graphical Statistics, in press (http://dx.doi.org/10.1080/10618600.2012.679240).

Sampson, S., 1969. Crisis in a cloister. Ph.D. Thesis. Cornell University.

Schervish, M.J., 1995. Theory of Statistics. Springer-Verlag.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Šmídl, V., Quinn, A., 2006. The Variational Bayes Method for Signal Processing. Springer Verlag, Berlin.

Wade, K., Greene, D., Lee, C., Archambault, D., Cunningham, P., 2011. Identifying representative textual sources in blog networks. In: Proc. 5th International AAAI Conference on Weblogs and Social Media, ICWSM'11.