THE RANDOM SUBGRAPH MODEL FOR THE ANALYSIS OF AN ECCLESIASTICAL NETWORK IN MEROVINGIAN GAUL

By Yacine Jernite[*,§], Pierre Latouche[*], Charles Bouveyron[†],
Patrick Rivera[‡], Laurent Jegou[‡] and Stéphane Lamassé[‡]

*Laboratoire SAMM, EA 4543, Université Paris 1 Panthéon–Sorbonne[*],
Laboratoire MAP5, UMR CNRS 8145, Université Paris Descartes and
Sorbonne Paris Cité[†] Laboratoire LAMOP, UMR 8589, Université Paris 1
Panthéon–Sorbonne[‡] and Ecole Polytechnique[§]*

In the last two decades many random graph models have been proposed to extract knowledge from networks. Most of them look for communities or, more generally, clusters of vertices with homogeneous connection profiles. While the first models focused on networks with binary edges only, extensions now allow to deal with valued networks. Recently, new models were also introduced in order to characterize connection patterns in networks through mixed memberships. This work was motivated by the need of analyzing a historical network where a partition of the vertices is given and where edges are typed. A known partition is seen as a decomposition of a network into subgraphs that we propose to model using a stochastic model with unknown latent clusters. Each subgraph has its own mixing vector and sees its vertices associated to the clusters. The vertices then connect with a probability depending on the subgraphs only, while the types of edges are assumed to be sampled from the latent clusters. A variational Bayes expectation-maximization algorithm is proposed for inference as well as a model selection criterion for the estimation of the cluster number. Experiments are carried out on simulated data to assess the approach. The proposed methodology is then applied to an ecclesiastical network in Merovingian Gaul. An R code, called *Rambo*, implementing the inference algorithm is available from the authors upon request.

1. Introduction. Since the original work of Moreno (1934) on sociograms, network data has become ubiquitous in Biology [Albert and Barabási (2002), Milo et al. (2002), Palla et al. (2005)] and computational social sciences [Snijders and Nowicki (1997)]. Applications range from the study of gene regulation processes to that of social interactions. Network analysis was also applied recently to a medieval social network in Villa, Rossi and Truong (2008), where the authors find a clustering of vertices through kernel methods. Both deterministic and probabilistic methods have been used to seek structure in these networks, depending on prior knowledge and assumptions on the form of the data. For example, Hofman and

Wiggins (2008) looked for a partition of the vertices where the clusters exhibit a transitivity property. The model of Handcock, Raftery and Tantrum (2007), on the other hand, assumes the relations to be conditioned on the projection of the vertices in a latent social space. Notable among the community discovery methods, though asymptotically biased [Bickel and Chen (2009)], are those based on the modularity score given by Girvan and Newman (2002).

Many of the other currently used methods derive from the stochastic block model (SBM) [Nowicki and Snijders (2001), Wang and Wong (1987)], which is a probabilistic generalization [Fienberg and Wasserman (1981)] of the method applied by White, Boorman and Breiger (1976) to Sampson's famous monastery data. SBM assumes that each vertex belongs to a hidden cluster and that connection probabilities between a pair of vertices depend exclusively on their clusters, as in Frank and Harary (1982). The parameters and clusters are then inferred to optimize a criterion, usually a lower bound of an integrated log-likelihood. Thus, Latouche, Birmelé and Ambroise (2011) used an approximation of the marginal log-likelihood, while Daudin, Picard and Robin (2008) considered a Laplace approximation of the integrated classification log-likelihood. A nonparametric Bayesian approach was also introduced by Kemp et al. (2006) to estimate the number of clusters while clustering the vertices. SBM was extended by the mixed membership stochastic block model (MMSBM) [Airoldi et al. (2008)], which allows a vertex to belong to different clusters in its relations toward different vertices, and by the overlapping stochastic block model (OSBM) [Latouche, Birmelé and Ambroise (2011)], which allows a vertex to belong to no cluster or to several at the same time. More recent works focused on extending MMSBM to dynamic networks [Xing, Fu and Song (2010)] or dealing with nonbinary networks, such as networks with weighted edges [Soufiani and Airoldi (2012)]. Goldenberg, Zheng and Fienberg (2010) and Salter-Townshend et al. (2012) provide extensive reviews of statistical network models.

In this paper we aim at clustering the vertices of networks with typed edges and for which a partition of the nodes into subgraphs is observed and bears some importance in their behavior. For example, one may be interested in looking for latent clusters in a worldwide social network describing social interactions between individuals where different countries, or at a different scale, different regions of the world, have different connectivity patterns. We might also observe the same kind of phenomenon between different scientific fields in a citation network. This kind of network may be modelled using generalized linear models [Fienberg and Wasserman (1981)] by incorporating the observed partition information as covariates and the clusters serving as random effects or a $p1$ model [Holland and Leinhardt (1981)], where the clusters allow for the estimation of interactions. However, we consider here a different strategy and propose an extension of the SBM model which has the advantage of relying on easy to interpret parameters. Indeed, SBM parameters are not expressed through nonlinear functions like the log or logistic functions and this allows an easy interpretation for nonstatisticians.

This point is of crucial interest in this work because we aim at providing historians with insight into the relationships between ecclesiastics and notable people in the kingdoms that made up Merovingian Gaul, by analyzing a network characterizing their different kinds of relations. Specifically, the data set focuses on the relationships between individuals built during the ecclesiastical councils which took place in Gaul during the 6th century. These councils were convened under the authority of a bishop to discuss specific questions relating to the Church. Though consisting mainly of clergymen, laics would also occasionally take part, as representatives of the secular power or experts in the questions discussed. These assemblies shaped a significant part of that period, and we are interested in discovering how they reflected the relationships between various groups of individuals. For this network, extra information on the vertices, namely, a geographical partition, is available, associating each individual to a specific kingdom. This partition induces a decomposition of the network into subgraphs and we aim at modelling the connection pattern of each subgraph through latent clusters.

Thus, in this paper, we propose a new model, that we call the random subgraph model (RSM), for the analysis of directed networks with typed edges for which a partition of the vertices is available. On the one hand, we consider that the probability of observing an edge between two vertices depends solely on the subgraphs to which the vertices belong. On the other hand, we assume that each vertex belongs to a hidden cluster, with a probability depending on its subgraph. Then, if a relation is present, its type is drawn from a multinomial distribution whose parameters depend on the clusters to which the vertices belong. Let us emphasize that the latter property allows, once the inference is done, to compare the different subgraphs.

The choice of proposing a probabilistic rather than a deterministic model is again motivated by the nature of the historical network we consider. Indeed, as mentioned in Section 4, the data set was built from a collection of data at hand using sources such as council acts or narrative texts. However, the rarity of the sources only allowed an incomplete or approximate characterization of the relations between individuals. Therefore, we rely on the probabilistic framework in order to deal with the uncertainty on the edges. Moreover, we emphasize that probabilistic methods for network analysis are appealing in general because they have been shown to be flexible and capable of retrieving complex heterogeneous structures in networks [see, for instance, Airoldi et al. (2008), Goldenberg, Zheng and Fienberg (2010)].

The article is organized as follows. The random subgraph model is presented along with its inference algorithm in Section 2, then tested on simulated data and compared to other models in Section 3. Our model is then applied to the ecclesiastical network and the results are analyzed from the historical point of view in Section 4. Concluding remarks and possible extensions are finally discussed in Section 5.

2. The random subgraph model. We consider a directed graph $\mathcal{G}$ with N vertices represented by its $N \times N$ adjacency matrix $\mathbf{X}$ along with a known partition $\mathcal{P}$ of the vertices into S classes. Our goal is to cluster the network into K groups with homogeneous connection profiles, that is, estimating a binary matrix $\mathbf{Z}$ such that $Z_{ik} = 1$ if vertex i belongs to cluster k, 0 otherwise.

Let us now detail the notation. Each edge X_{ij}, describing the relation between the vertices i and j, is typed, that is, takes its values in a finite set $\{0, \ldots, C\}$. Note that $X_{ij} = 0$ corresponds to the absence of an edge. We assume that $\mathcal{G}$ does not have any self loop and, therefore, the entries X_{ii} will not be taken into account. In order to simplify the notation when describing the model, we also consider the binary matrix $\mathbf{A}$ with entries A_{ij} such that $A_{i,j} = 1 \iff X_{i,j} \neq 0$.

We also emphasize that the observed partition $\mathcal{P}$ induces a decomposition of the graph into subgraphs where each class of vertices corresponds to a specific subgraph. We introduce the variable s_i which takes its values in $\{1, \ldots, S\}$ and is used to indicate in which of the subgraphs vertex i belongs, for $i \in \{1, \ldots, N\}$.

2.1. *The probabilistic model.* The data is assumed to be generated in three steps. First, the presence of an edge from vertex i to vertex j is supposed to follow a Bernouilli distribution whose parameter depends on the subgraphs s_i and s_j only:

$$A_{i,j} \sim \mathcal{B}(\gamma_{s_i, s_j}).$$

Each vertex i is then associated to a latent cluster with a probability depending on s_i. In practice, if we assume for now that the number K of latent clusters is known, the variable $\mathbf{Z}_i$ is drawn from a multinomial distribution:

$$\mathbf{Z}_i \sim \mathcal{M}(1; \boldsymbol{\alpha}_{s_i}),$$

where

$$\forall s \in 1, \ldots, S \qquad \sum_{k=1}^{K} \alpha_{sk} = 1.$$

A notable point of the model is that we allow each subgraph to have different mixing proportions $\boldsymbol{\alpha}_s$ for the latent clusters. We denote hereafter $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_S)$. Finally, if an edge between i and j is present, that is, $A_{ij} = 1$, its type X_{ij} is sampled from a multinomial distribution with parameters depending on the latent clusters. Thus, if i belongs to cluster k and j to cluster l,

$$X_{i,j} | Z_{ik} Z_{jl} = 1, \qquad A_{ij} = 1 \sim \mathcal{M}(1, \mathbf{\Pi}_{kl}),$$

where the sum over the C types of each vector $\mathbf{\Pi}_{kl} = (\Pi_{kl1}, \ldots, \Pi_{klC})$ is

$$\forall (k, l) \in \{1, \ldots, K\}^2 \qquad \sum_{c=1}^{C} \Pi_{klc} = 1.$$

If there is no edge between the two vertices, the entry X_{ij} is simply set to $X_{ij} = A_{ij} = 0$.

The model is therefore defined through the joint distribution

$$p(\mathbf{X}, \mathbf{A}, \mathbf{Z}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}) = p(\mathbf{X}, \mathbf{A}|\mathbf{Z}, \boldsymbol{\gamma}, \boldsymbol{\Pi})p(\mathbf{Z}|\boldsymbol{\alpha})$$
$$= p(\mathbf{X}|\mathbf{A}, \mathbf{Z}, \boldsymbol{\Pi})p(\mathbf{A}|\boldsymbol{\gamma})p(\mathbf{Z}|\boldsymbol{\alpha}),$$

where

$$p(\mathbf{X}|\mathbf{A}, \mathbf{Z}, \boldsymbol{\Pi}) = \prod_{k,l}^{K} \prod_{c=1}^{C} (\Pi_{kl}^c)^{\sum_{i \neq j}^{N} \delta(X_{ij}=c)A_{ij}Z_{ik}Z_{jl}}$$

and

$$p(\mathbf{A}|\boldsymbol{\gamma}) = \prod_{i \neq j}^{N} \gamma_{r_i,r_j}^{A_{ij}} (1 - \gamma_{r_i,r_j})^{1-A_{ij}}.$$

Finally,

$$p(\mathbf{Z}|\boldsymbol{\alpha}) = \prod_{i=1}^{N} \prod_{k=1}^{K} \alpha_{r_i,k}^{Z_{ik}}.$$

We refer to the Appendix for the detailed calculation of the complete data log-likelihood associated to the RSM model and summarize the model parameters in Table 1.

We point out that the choice of separating the role of the known subgraphs and the latent clusters was motivated by historical assumptions on the creation of relationships between individuals in Gaul during the 6th century. These assumptions were at the core of the study of the ecclesiastical network we consider in this paper. An alternative approach would consist in allowing the presence of an edge and

TABLE 1
Summary of the notation used in the paper

Notation	Description
X	Adjacency matrix. $X_{ij} \in \{0, \ldots, C\}$ indicates the edge type
A	Binary matrix. $A_{ij} = 1$ indicates the presence of an edge
Z	Binary matrix. $Z_{ik} = 1$ indicates that i belongs to cluster k
N	Number of vertices in the network
K	Number of latent clusters
S	Number of subgraphs
C	Number of edge types
$\boldsymbol{\alpha}$	α_{sk} is the proportion of cluster k in subgraph s
$\boldsymbol{\Pi}$	Π_{klc} is the probability of having an edge of type c between vertices of clusters k and l
$\boldsymbol{\gamma}$	γ_{rs} probability of having an edge between vertices of subgraphs r and s

its type to depend on both the subgraphs and latent clusters. However, this would dramatically increase the number of model parameters to be estimated. Indeed, for a network with $S = 6$, $K = 6$, and $C = 4$, it would require $K^2 S^2 (C + 1) + SK = 6516$ parameters while RSM only involves $S^2 + K^2 C + SK = 216$ parameters.

2.2. *Bayesian framework.* We consider a Bayesian framework and introduce conjugate prior distributions. Thus, since $\mathbf{Z}_i$ is sampled from a multinomial distribution, we rely on a Dirichlet prior to model the parameters $\boldsymbol{\alpha}_s$:

$$p(\boldsymbol{\alpha}_s) = \mathrm{Dir}(\boldsymbol{\alpha}_s; \chi_{s1}^0, \ldots, \chi_{sK}^0) \qquad \forall s \in \{1, \ldots, S\}.$$

A similar distribution is used as a prior distribution for the parameters $\boldsymbol{\Pi}_{kl}$:

$$p(\boldsymbol{\Pi}_{kl}) = \mathrm{Dir}(\boldsymbol{\Pi}_{kl}; \Xi_{kl1}^0, \ldots, \Xi_{klC}^0) \qquad \forall (k, l) \in \{1, \ldots, K\}^2.$$

If no prior information is available, a common choice in the literature consists in fixing the hyperparameters of the Dirichlet to $1/2$, that is, $\chi_{sk}^0 = 1/2$, $\forall (s, k)$ and $\Xi_{klc} = 1/2$, $\forall (k, l, c)$. Such a distribution corresponds to a noninformative Jeffreys prior distribution which is known to be proper [Jeffreys (1946)]. A uniform distribution can also be obtained by setting the hyperparameters to 1.

Finally, since the presence or absence of an edge between a pair of vertices is drawn from a Bernoulli distribution, we rely on a beta prior for the parameters γ_{rs}:

$$p(\gamma_{rs}) = \mathrm{Beta}(\gamma_{rs}; a_{rs}^0, b_{rs}^0) \qquad \forall (r, s) \in \{1, \ldots, S\}^2.$$

Again, if no prior information is available, both hyperparameters a_{rs}^0 and b_{rs}^0 can be set to $1/2$ or 1 to obtain noninformative prior distributions, respectively, a Jeffreys or a uniform distribution. Figure 1 presents the graphical model associated with the RSM model.

2.3. *Inference with the variational Bayes EM algorithm.* Given the observed matrices $\mathbf{X}$ and $\mathbf{A}$, we aim at estimating the posterior distribution $p(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi} | \mathbf{X}, \mathbf{A})$, which in turn will allow us to compute a maximum a posteriori estimate of the clustering structure $\mathbf{Z}$ as well as the parameters $(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})$.
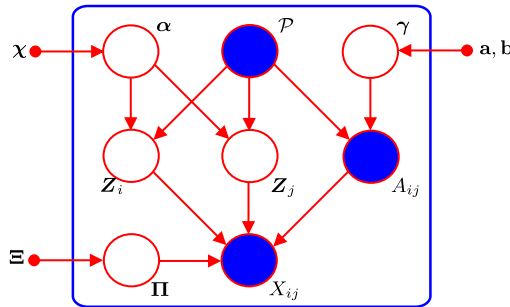

FIG. 1. *Graphical representation of the RSM model.*

Because this distribution is not tractable, approximate inference procedures are required. The Markov chain Monte Carlo (MCMC) sampling scheme is a widely used approach which consists in sampling from tractable conditional distributions. After a burn-in period, samples are assumed to be drawn from the true posterior distribution. One of the main advantages of the MCMC algorithm is that it can characterize the uncertainty in model parameters. Moreover, the convergence of the Markov chain and therefore the quality of the approximation can be tested.

Unfortunately, the MCMC algorithm has a poor scaling with sample sizes. This motivated the work of Daudin, Picard and Robin (2008) who proposed a variational approach for the SBM model which can deal with large networks contrary to the MCMC method of Nowicki and Snijders (2001). In general, the main drawback of variational techniques is that, although they can produce a good estimate of the model parameters or find the mode of the posterior distribution, they usually cannot uncover the uncertainty in the model parameters and tend to underestimate posterior variances. Furthermore, the quality of the variational approximation cannot be tested in most cases since the KL divergence between the true and approximate posterior distribution is not tractable.

However, recent results [Celisse, Daudin and Pierre (2012), Mariadassou and Matias (2014)] gave some new insights on the form of the true posterior distribution in the case of the SBM model and showed that the corresponding variational estimates were consistent. In light of these recent results and because we aim at proposing an inference procedure capable of handling large networks, we rely in the following on a variational Bayes EM (VBEM) algorithm.

Thus, given a distribution $q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})$, the marginal log-likelihood can be computed in two terms,

$$\log p(\mathbf{X}, \mathbf{A}) = \mathcal{L}(q) + \mathrm{KL}\big(q(\cdot)\|p(\cdot|\mathbf{X}, \mathbf{A})\big),$$

where $\mathcal{L}$ is defined as follows:

$$\mathcal{L}(q) = \sum_{\mathbf{Z}} \int_{\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}} q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}) \log\bigg(\frac{p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}{q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}\bigg) d\boldsymbol{\alpha}\, d\boldsymbol{\gamma}\, d\boldsymbol{\Pi}$$

and the KL divergence is given by

$$\mathrm{KL}\big(q(\cdot)\|p(\cdot|\mathbf{X})\big)$$
$$= -\sum_{\mathbf{Z}} \int_{\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}} q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}) \log\bigg(\frac{p(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}|\mathbf{X}, \mathbf{A})}{q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}\bigg) d\boldsymbol{\alpha}\, d\boldsymbol{\gamma}\, d\boldsymbol{\Pi}.$$

Finding the best approximation of the posterior distribution $p(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}|\mathbf{X}, \mathbf{A})$ in the sense of the KL divergence becomes equivalent to finding $q(\cdot)$ that maximizes the lower bound $\mathcal{L}(q)$ of the integrated log-likelihood. To obtain a tractable algorithm, we assume that $q(\cdot)$ can be fully factorized, that is,

$$q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}) = \bigg(\prod_{i=1}^{N} q(\mathbf{Z}_i)\bigg)\bigg(\prod_{s=1}^{S} q(\boldsymbol{\alpha}_s) \prod_{t=1}^{S} q(\gamma_{s,t})\bigg) \prod_{k,l}^{K} q(\boldsymbol{\Pi}_{k,l}).$$

Algorithm 1 VBEM algorithm for the RSM model (see text for details)

Initialize matrix $\boldsymbol{\tau} = \mathbf{Z}$ with k-means
Initialize hyperparameters $\boldsymbol{\theta}^0 = \{\boldsymbol{\chi}^0, (\mathbf{a}^0, \mathbf{b}^0), \boldsymbol{\Xi}^0\}$
Compute $\mathcal{L}(q)$
while $|\boldsymbol{\theta}^{\text{new}} - \boldsymbol{\theta}^{\text{old}}| \geq \varepsilon$ **do**
    E step: update $\boldsymbol{\tau}$
    M step: update $\boldsymbol{\theta}^{\text{new}} = \{\boldsymbol{\chi}, (\mathbf{a}, \mathbf{b}), \boldsymbol{\Xi}\}$
    Compute $\mathcal{L}$
end while

The functional optimization of the lower bound with respect to $q(\cdot)$ is performed using a VBEM algorithm (see Algorithm 1). All the optimization equations are given in the Appendix. We emphasize that the functional form of the prior distributions is preserved through the optimization. In particular, $q(\mathbf{Z})$ is given by

$$q(\mathbf{Z}) = \prod_{i=1}^{N} q(\mathbf{Z}_i) = \prod_{i=1}^{N} \mathcal{M}(\mathbf{Z}_i; 1, \boldsymbol{\tau}_i),$$

where τ_{ik} is a variational parameter denoting the probability of node i to belong to cluster k. The approximate posterior distributions over the other model parameters $(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})$ depend on parameters that we denote $\boldsymbol{\theta} = \{\boldsymbol{\chi}, (\mathbf{a}, \mathbf{b}), \boldsymbol{\Xi}\}$, respectively.

2.4. *Initialization.* The VBEM algorithm, though useful in approximating posterior distributions of graphical models, is only guaranteed to converge to a local optimum [Bilmes (1998)]. Strategies to tackle this issue include simulated annealing and the use of multiple initializations [Biernacki, Celeux and Govaert (2003)]. In this work, we choose the latter option. In order to have a better chance of reaching a global optimum, VBEM is run for several initializations of a k-means like algorithm with the following distance $d(i, j)$ between the vertices i and j:

$$(1) \qquad d(i, j) = \sum_{h=1}^{N} \delta(X_{ih} \neq X_{jh}) A_{ih} A_{jh} + \sum_{h=1}^{N} \delta(X_{hi} \neq X_{hj}) A_{hi} A_{hj}.$$

The first term looks at all possible edges from i and j toward a third vertex h. If both i and j are connected to h, that is, $A_{ih} A_{jh} = 1$, the edge types X_{ih} and X_{jh} are compared. By symmetry, the second term looks at all possible edges from a vertex h to both i as well as j, and compare their types. Thus, the distance computes the number of discordances in the way both i and j connect to other vertices or vertices connect to them. The algorithm starts by sampling the cluster centers among all the vertices of the network. It then iterates a two-step procedure until convergence of the cluster centers. In the first step, the vertices are classified into the cluster with the closest center. Each cluster center is then associated to a vertex minimizing its distance with all the vertices of the corresponding cluster.

2.5. *Choice of K.* So far, the number K of latent clusters has been assumed to be known. Given K, we showed in Section 2.3 how an approximation of the posterior distribution over the latent structure and model parameters could be obtained. We now address the problem of estimating the number of clusters directly from the data. Given a set of values of K, we aim at selecting K^* for which the marginal log-likelihood $\log p(\mathbf{X}|K)$ is maximized. However, because this integrated log-likelihood involves a marginalization over all the model parameters and latent variables, it is not tractable. Therefore, we propose to replace the marginal log-likelihood with its variational approximation, as in Bishop (2006), Latouche, Birmelé and Ambroise (2009), Latouche, Birmelé and Ambroise (2012). Thus, for each value of K considered, the VBEM algorithm is applied. We recall that the maximization of the lower bound induces a minimization of the KL divergence. After convergence of the algorithm, the lower bound is used as an approximation of $\log p(\mathbf{X}|K)$ and K is chosen such that the lower bound is maximized. We prove in the Appendix that, if computed right after the M step of the variational Bayes EM, the lower bound has the following expression:

$$\mathcal{L}(q) = \sum_{r,s}^{S} \log\left(\frac{B(a_{rs}, b_{rs})}{B(a_{rs}^0, b_{rs}^0)}\right) + \sum_{s=1}^{S} \log\left(\frac{C(\boldsymbol{\chi}_s)}{C(\boldsymbol{\chi}_s^0)}\right)$$

$$+ \sum_{k,l}^{K} \log\left(\frac{C(\boldsymbol{\Xi}_{kl})}{C(\boldsymbol{\Xi}_{kl}^0)}\right) - \sum_{i=1}^{N} \sum_{k=1}^{K} \tau_{ik} \log(\tau_{ik}),$$

where $C(x) = \frac{\prod_{d=1}^{D} \Gamma(x_d)}{\Gamma(\sum_{d=1}^{D} x_d)}$ if $x \in \mathbb{R}^D$, $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \forall (a, b) \in \mathbb{R}^2$, and $\Gamma(\cdot)$ is the gamma function. See the Appendix for the definition of a_{rs}, b_{rs}, $\boldsymbol{\chi}_s$, $\boldsymbol{\Xi}_{kl}$ and τ_{ik}.

3. Numerical experiments and comparisons. In this section we first run experiments aimed at proving the validity of our model, focusing on the ability of its inference procedure to find the right clustering. We then compare its performance to that of other stochastic models for graph clustering.

3.1. *Experimental setup.* In order to evaluate the performance of our approach, we applied it on data generated according to the RSM model. To simplify the parameterization and facilitate the reproducibility of the experiments, we constrained the parameters $\boldsymbol{\Pi}$ and $\boldsymbol{\gamma}$ to have the following forms:

$$\boldsymbol{\Pi} = \begin{bmatrix} \mathbf{u} & \mathbf{v} & \cdots & \mathbf{v} \\ \mathbf{v} & \mathbf{u} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{v} \\ \mathbf{v} & \cdots & \mathbf{v} & \mathbf{u} \end{bmatrix}, \qquad \boldsymbol{\gamma} = \begin{bmatrix} \lambda & \varepsilon & \cdots & \varepsilon \\ \varepsilon & \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & \varepsilon \\ \varepsilon & \cdots & \varepsilon & \lambda \end{bmatrix},$$

where $\lambda, \varepsilon \in [0, 1]$ and $\mathbf{u}, \mathbf{v} \in [0, 1]^K$. With such a parameterization, the probability λ of an edge within a subgraph is assumed to be common between subgraphs
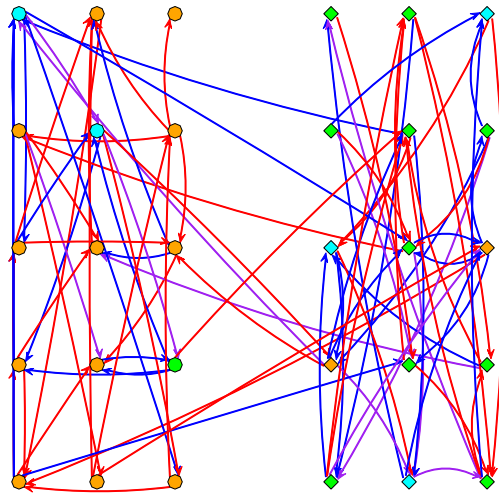
FIG. 2. *Example of a RSM network for S = 2 subgraphs (indicated by the node forms), C = 3 types of edges (indicated by the edge colors) and K = 3 clusters to identify (indicated by the node colors).*

and the probability ε of a connection between different subgraphs is also assumed to be the same for all couples of subgraphs. Similarly, the vector **u** controls the probability of each edge type between nodes of a same cluster, whereas **v** defines the edge type probabilities between nodes of different clusters. We recall that the prior probabilities of each group within each subgraph are given by the parameter $\alpha = (\alpha_1, \ldots, \alpha_S)$.

Figure 2 presents an example of a network generated this way with parameters $S = 2$, $C = 3$, $K = 3$, $\alpha = \left[\begin{smallmatrix} 0.1 & 0.3 & 0.6 \\ 0.6 & 0.3 & 0.1 \end{smallmatrix}\right]$, $\lambda = 0.6$, $\varepsilon = 0.06$, **u** $= (0.8, 0.1, 0.1)$ and **v** $= (0.1, 0.3, 0.6)$. This RSM network is made of 30 nodes with $S = 2$ subgraphs (indicated by the node forms), $C = 3$ types of edges (indicated by the edge colors) and $K = 3$ clusters that have to be identified in practice (indicated by the node colors).

In order to illustrate, on various situations, that RSM is a relevant model and that its corresponding inference procedure provides an accurate estimation of the true clustering structure, we rely in the following paragraphs on three types of graphs, described in Table 2. The three scenarios considered correspond to different situations ranging from an almost classical setup to a more specific one. The first scenario considers networks with no subgraphs ($S = 1$) and with a preponderant proportion of edges of type 1 ($u_1 = 0.8$) and 3 ($u_3 = 0.8$). The second scenario still considers networks with no subgraphs ($S = 1$) but with balanced proportions of edge types. Finally, the third scenario considers networks with several subgraphs ($S = 3$) and balanced proportions for edge types. Therefore, the latter case should be the more complex situation to fit.

TABLE 2
Parameter values for the three types of graphs used in the experiments

Parameters	Scenario 1	Scenario 2	Scenario 3
N	100	100	100
S	1	1	3
C	3	3	3
K	3	3	3
$\boldsymbol{\alpha}$	(0.3, 0.3, 0.4)	(0.3, 0.3, 0.4)	$\begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$
$\mathbf{u}$	(0.8, 0.1, 0.1)	(0.5, 0.45, 0.05)	(0.5, 0.45, 0.05)
$\mathbf{v}$	(0.1, 0.1, 0.8)	(0.1, 0.45, 0.45)	(0.1, 0.45, 0.45)
λ	0.2	0.2	0.2
ε	0.06	0.06	0.1

The VBEM algorithm with multiple initializations, presented in Section 2, is used in the following experiments. For a given value of K, the result with the best value for $\mathcal{L}(q)$ is chosen among the multiple initializations. Then, a clustering partition is deduced from the posterior probabilities τ_{ik} using the maximum *a posteriori* (MAP) rule, that is, a node is assigned to the group with the highest posterior probability.

Since our approach aims to search the unobserved clustering partition of the nodes, we chose here to evaluate the results of our VBEM algorithm by comparing the resulting partition with the actual one (the simulated partition). In the clustering community, the adjusted Rand index (ARI) [Rand (1971)] serves as a widely accepted criterion for the difficult task of clustering evaluation. The ARI looks at all pairs of nodes and checks whether they are classified in the same group or not in both partitions. As a result, an ARI value close to 1 means that the partitions are similar and, in our case, that the VBEM algorithm succeeds to recover the simulated partition.

3.2. *Choice of K and inference results.* In this first simulation study we aim at evaluating the ability of the lower bound $\mathcal{L}(q)$ to serve as a criterion for selecting the appropriate number K of clusters. To this end, the VBEM algorithm for the RSM model was first run on a graph simulated according to scenario 1 for several values of K. The highest criterion values among the different initializations obtained for each value of K are presented in Figure 3. The figure indicates that $K = 3$ seems to be the appropriate number of groups for the studied network, which is the actual number of group.

We then replicated this experiment over 50 networks, still simulated according to scenario 1, for both verifying the consistency of $\mathcal{L}(q)$ and studying the clustering ability of our approach. Figure 4 shows the repartition of the criterion values (left
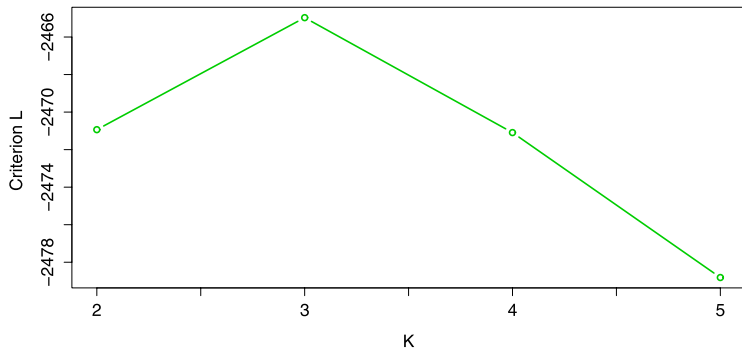
FIG. 3.   *Criterion values $\mathcal{L}(q)$ vs. number K of groups for a graph simulated according to scenario* 1.

panel) as well as the associated ARI values (righ panel). These results confirm that the lower bound $\mathcal{L}(q)$ is a valid criterion for selecting the number of groups. One can also observe that the partition resulting from our VBEM algorithm has, for the selected number of groups, a good adequation with the actual partition of the data.

3.3. *Comparison with the stochastic block model.*   Our second set of experiments compares the performance of RSM to that of other models on data drawn according to its generative process. We were interested in the comparison with the following models:

- *binary SBM* (*presence*): We fit a binary SBM using the R package *mixer* [Ambroise et al. (2010)] on a collapsed version of the data to conform this specific model. The collapsed data were obtained by considering only the presence
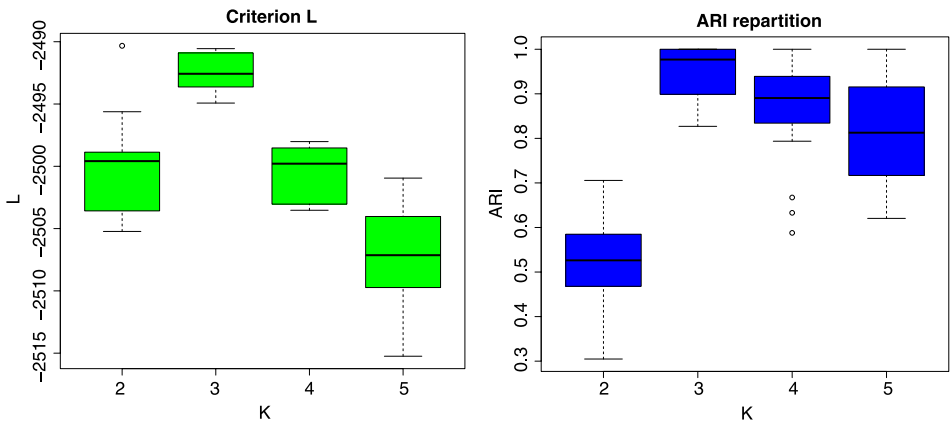

FIG. 4.   *Repartition of the criterion* (*left panel*) *and ARI* (*right panel*) *over* 50 *networks generated with the parameters of the first scenario.*

of the edges and not the type of the edges, that is, $\tilde{X}_{ij} = 0$ if $X_{ij} = 0$ and $\tilde{X}_{ij} = 1$ otherwise.

- *binary SBM* (*type* 1, 2 *or* 3): We fit a binary SBM, still using the *mixer* package, on the networks defined by taking only the edges of one type. For instance, the collapsed network for type $c = 1, 2, 3$ was obtained by considering only the presence of type c edges, that is, $\tilde{X}_{ij} = 1$ if $X_{ij} = c$ and $\tilde{X}_{ij} = 0$ otherwise.
- *typed SBM*: We consider here a SBM with discrete edges. Although SBM was originally proposed in Nowicki and Snijders (2001) with discrete edges, existing softwares only propose to fit a SBM on binary networks. We therefore had to implement a version of the SBM which supports typed edges. Note that, in this case, the types of edges are in $\{0, \ldots, C\}$, where 0 corresponds to the absence of a relation.
- *RSM*: We run the VBEM algorithm, that we proposed in Section 2 for the inference of the RSM model, with the available subgraph partition and with 5 random initializations for each run.

Table 3 presents the average ARI values and standard deviations on 50 simulated graphs for each scenario and with binary SBM, typed SBM and RSM. We point out that the inference is done with the actual number of clusters and this for each method. One can observe that, for the first scenario, the binary SBM based on the link presences and the type 2 SBM always fails, whereas type 1, type 3 and typed SBM work pretty well. Those behaviors can be explained by the nature of scenario 1, which is a rather easy situation with no subgraphs and a predominant presence of type 1 and type 2 links. However, we can remark that it seems easier in this case to fit a binary SBM on type 1 or type 2 edges than to fit a typed SBM. This is due to the high discriminative power of type 1 and type 2 edges in this specific scenario. Let us also remark that RSM works perfectly here even though the network does not contain any subgraphs.

Regarding scenario 2, which considers a situation where there is still no subgraphs but with more balanced proportions of the different edge types, one can

TABLE 3
Average ARI values and standard deviations for binary SBM, typed SBM and RSM according to the three simulation scenarios. The results are averaged on 25 simulated graphs for each scenario

Method	Scenario 1	Scenario 2	Scenario 3
Binary SBM (presence)	0.001 ± 0.012	0.001 ± 0.013	0.239 ± 0.061
Binary SBM (type 1)	0.976 ± 0.071	0.494 ± 0.233	-0.372 ± 0.262
Binary SBM (type 2)	0.001 ± 0.006	-0.003 ± 0.006	0.179 ± 0.097
Binary SBM (type 3)	0.959 ± 0.121	0.519 ± 0.219	0.367 ± 0.244
Typed SBM	0.694 ± 0.232	0.472 ± 0.339	0.360 ± 0.162
RSM	$\mathbf{1.000 \pm 0.000}$	$\mathbf{0.981 \pm 0.056}$	$\mathbf{0.939 \pm 0.097}$

first notice that binary SBM and type 2 SBM fail once again. The type 1 and type 3 SBM have now a behavior closer to the one of typed SBM, whereas RSM gives very accurate results once again. Finally, scenario 3 considers a RSM-type network, that is, with several subgraphs, and all SBM-based algorithms are significantly outperformed by RSM, which succeeds in exploiting both the information carried by different edge types and by the different subgraphs. To summarize, the RSM model and its associated VBEM algorithm turn out to be effective on situations ranging from classical setups without subgraphs to complex scenarios with subgraphs and typed edges.

4. Ecclesiastical network. This section now focuses on applying the RSM model to the ecclesiastical network, that we briefly described in the Introduction and that initially motivated this work, and on analyzing its results from the historical point of view.

Please note that the ecclesiastical network along with a file giving the kingdoms of all vertices in the network and an R code implementing the variational inference approach for the RSM model are provided as supplementary materials in Jernite et al. (2013).

4.1. *Description of the data*. The relational data considered in this section were mainly built from written acts of ecclesiastical councils that took place in Merovingian Gaul during the 6th century. A council is an ecclesiastical meeting, usually called by a bishop, where issues regarding the Church or the faith are addressed. However, since 511, kings could also call for a council to discuss some political, judiciary or legal issues, and that laics (kings, dukes or counts, e.g.) would attend. During the 6th century, 46 councils took place in Gaul. Although there were mostly local or regional councils, attended by individuals from a specific ecclesiastical province, there were some national councils convened under the authority of a king.

The composition of these councils is known thanks to the acts written at the end of the meeting, and which were signed by all attending members. In addition to the council acts, we used narrative texts (among which is the famous *Ten History Books* by Gregory of Tours), hagiographies or letters which also describe these councils. The network, that took over 18 months to build from these historical sources, contains $N = 1331$ individuals who held one or several offices in Gaul between the years 480 and 614, and who we know to have been related or to have met during their lifetime.

The council acts and the other historical sources allowed also to qualify the type of the relationship between the individuals involved in the network. However, the scarcity of the sources only allowed for an approximate characterization of these relationships. As a consequence, $C = 4$ relation types were qualified and the relationships can be either positive, negative, variable or neutral (when the type was unknown). For instance, a positive relationship may describe an agreement

between two clergymen on a question of faith, whereas a negative one may be a disagreement on such a question. Variable relationships usually correspond to relationships which change over time.

Using the different sources, it was also possible to obtain additional information on the individuals. In particular, the geographical positions of the offices hold by the clergymen or the laics allowed us to split the network into $S = 6$ subgraphs. Those 6 subgraphs correspond to the geographical partition of the Gaul at this period (the kingdoms of Neustria, Austrasia and Burgondy, and the provinces of Aquitaine and Provence), completed with an additional subgraph for individuals for whom the information was not available. We also recorded the social positions of the individuals in order to be able to interpret afterward the clusters found by our method. These social positions can be, for instance, ecclesiastical positions (bishops, deacon, archdeacon, abbot, priest, ...) or titles of nobility (king, queen, duke, earl, ...).

To summarize, the network is made of $N = 1331$ individuals split into $S = 6$ subgraphs and whose relationships can be of $C = 4$ difference types. Figures 5 and 6 show some parts of the whole adjacency matrix associated to the network where the dot colors indicate the type or relationships. The whole adjacency matrix is provided in a zoomable pdf file as supplementary material [Jernite et al. (2013)]. We expect the statistical analysis with RSM of this network to help us understand how the behavior of an individual can be modeled through their belonging to a group. The use of a probabilistic approach, instead of a deterministic one, makes particular sense here since at least a part of the historical sources are subject to caution due to their nature and age. In History, this kind of approach is more common to modernists or contemporarists than to medievists who rarely
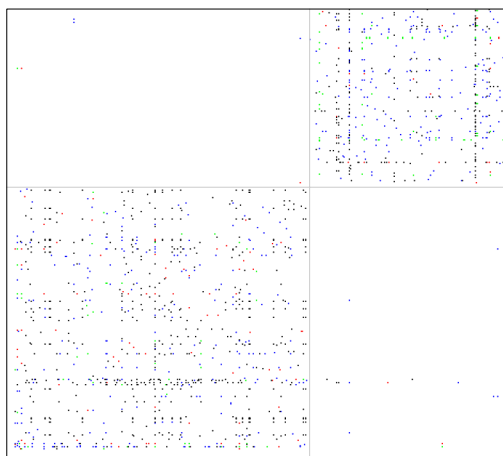

FIG. 5. *Adjacency matrix for the kingdom of Neustria (left block) and the province of Provence (right block). The dot colors indicate the type or relationships*: red = "negative," green = "variable," black = "neutral" and blue = "positive." Zoom on the paper electronic version for details.
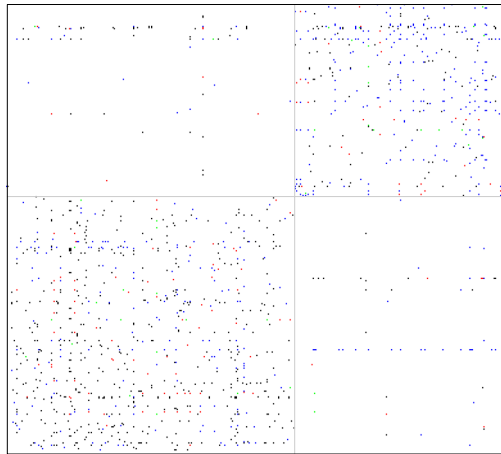
FIG. 6. *Adjacency matrix for the kingdoms of Austrasia* (*left block*) *and Burgundy* (*right block*). *The dot colors indicate the type or relationships*: *red = "negative," green = "variable," black = "neutral" and blue = "positive." Zoom on the paper electronic version for details.*

have access to this kind of data. Let us finally notice that a "source effect" is expected due to the possible overrepresentation in our sources of some places (Neustria by Gregory of Tours or Austrasia by Fredegar) or some individuals (in letters or hagiographies).

4.2. *Results.* The VBEM algorithm that we proposed to infer the RSM model was run on the network defined by these relations, where the subgraphs are the provinces in which the individuals lived (Aquitaine, Austrasia, Burgondy, Neustria, Provence or Unknown). The use of the lower bound $\mathcal{L}(q)$ allowed us to find 6 clusters.

To give some insight into the nature of the found clusters, Figure 7 presents the repartition of the different social positions in the clusters. In view of these results, some historical comments can be done. First, clusters 1 and 3 appear to be made of the people who would attend local assemblies, provincial or diocesan councils. The council of Arles, which took place in 554, would have had the same kind of composition as cluster 3, while that of Auxerre, in 585, could well represent cluster 1. Second, clusters 4 and 5 are more characteristic of aristocratic assemblies, such as the council of Orange in 529. Third, clusters 2 and 6 have the same compositions as councils concerned with more political issues (those usually convened by a king). Such a council took place in Orleans in 511. Let us, however, notice that cluster 2 is composed of very few individuals, which might hurt the relevance of its interpretation. Also, we might be able to further our understanding of the composition of these clusters by taking into account the similarity of certain social positions (such as "duke" and "earl").
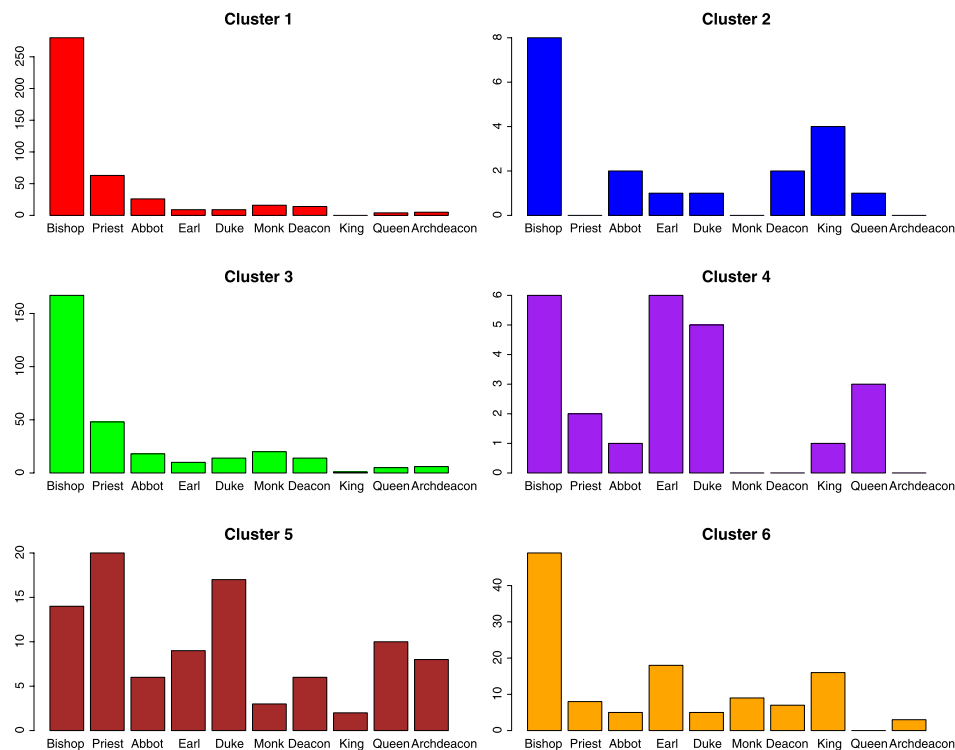
FIG. 7. *Repartition of the different social positions in the found clusters (restricted to the 10 most frequent positions).*

The relations between the different clusters, described by the parameter Π and shown in Figure 8, inform us further. Although the limitations expressed above about the roughness of the relation types still apply, they nevertheless provide us with interesting elements to confirm the coherence of the proposed model. First, it is natural that we should find "neutral" relations at the local level, between clusters 1, 3 and 6. Indeed, local assemblies were the less documented ones in our sources. On the other hand, the links between high level individuals are better known, because councils used to settle conflicts between aristocrats, which explains the presence of "negative" and "variable" relations. Finally, the positive relations between cluster 3, 5 and 6 could represent the personal friendships documented by a collection of letters between bishops.

After having described the political background represented by each of the clusters, we can compare the organization of the different regions. Figure 9 presents the cluster repartition (parameter α) in the different provinces. One can observe that the clergy and noblemen of the different regions were concerned with very different issues: Provence and Burgundy were more concerned with local questions (clusters 1 and 3), and less with political ones (clusters 2 and 6). The clusters
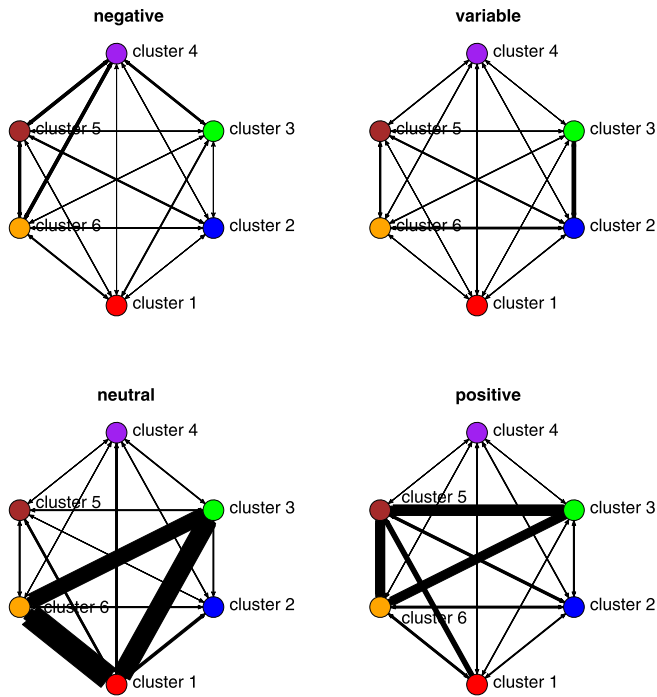
FIG. 8. *Relations between the* 6 *found clusters* (*parameter* Π) *for each relation type*: *negative*, *variable*, *neutral and positive. For visualization purposes*, *the relation weights have been normalized according to relation types.*

concerned both with local (clusters 1 and 3) and high level (cluster 6) questions are represented in Aquitaine. Conversely, all levels of power are represented in Neustria. This could be the result of a "source effect," as mentioned above. Let us also notice that the council structures seem similar in Austrasia and Aquitaine: sovereigns (kings and queens) are involved in the Church and frequently convene councils in order to discuss political questions.

Some of these observations are confirmed by the estimate of parameter γ, which is given in a log scale by Figure 10. First, it shows a greater frequency of relations between Aquitaine and Neustria, which comes both from a geographical and political proximity (Aquitaine is absorbed into Clovis' kingdom in 507, then divided and absorbed by Neustria in 511). One can also see there another example of "source" effect, as our main source, Gregory of Tours, was bishop in Neustria and raised in Aquitaine (next to his uncle, the bishop of Clermont), which gave him a good knowledge of both provinces. More enlightening is the relative disconnection of Burgondy and Provence, especially in regard to the provinces of Austrasia and Neustria, both heavily connected.
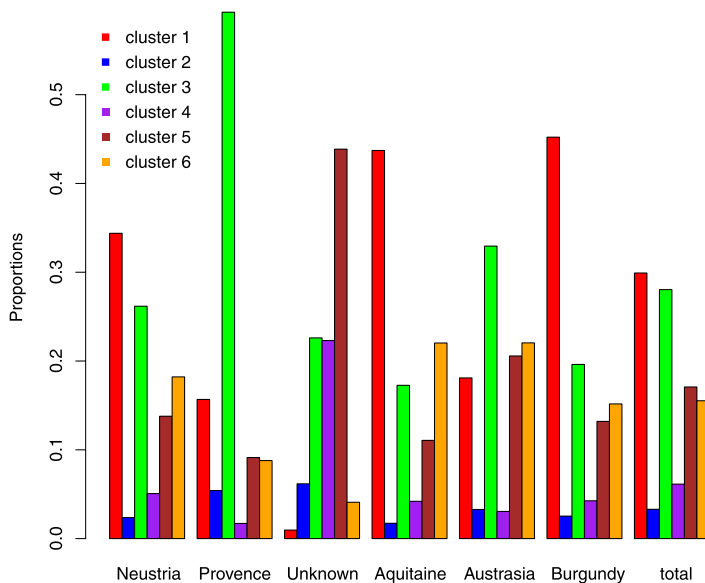
FIG. 9. *Repartition of the 6 found clusters in the different Provinces (parameter α).*

4.3. *Conclusion from the historical point of view.* A first analysis of the results of the RSM model confirms two well-known general facts. Indeed, our re-
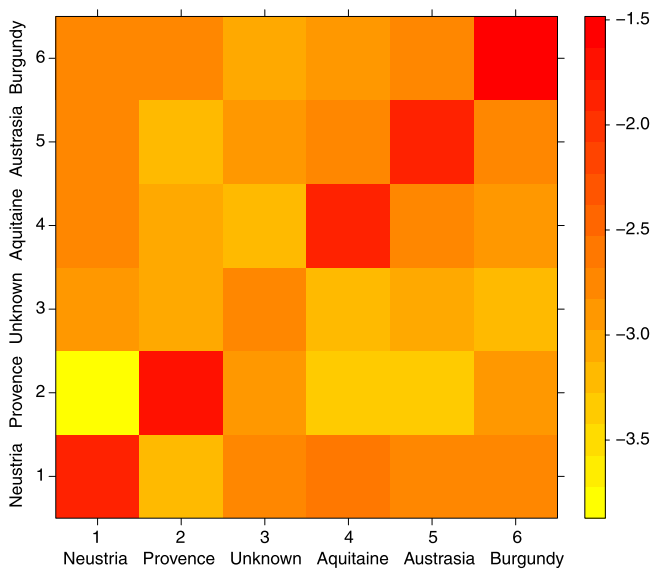

FIG. 10. *Estimated values for the parameter γ (in log scale).*

sults confirm the preponderance of local assemblies in 6th century Gaul and the "source effect." Nevertheless, further analysis of the found clusters and their relations yields a better understanding of the period. In particular, the composition of the found clusters reflect different archetypes of councils and different levels of political concerns. Our results have also highlighted that the types of concerns of each province are closely related to the frequency of their communications with others.

Two limitations to these results remain, however. First, we are limited by the scarcity of the historical documentation. It would be interesting to see whether the use of more precise types of relations (ecclesiastical or secular, through which media, ...) could improve the results. Second, it would also be interesting for the model to take into account temporal evolutions of the relations and clusters. Indeed, one aspect of the data which is currently not addressed by the results of RSM is its temporality. Nevertheless, this lack seems to have a limited impact here since all clusters exhibit the same distribution of individuals over time, reflecting the higher concentration of information in years 550 to 600 (when numerous conflicts were settled by councils: Paris 577, Chalon 579, Berny 580, Lyon 581, ...). The repartition of the different kinds of powers then seems to change little over time on this short period.

5. Conclusion and further work. In this work we proposed a new stochastic graph model, the random subgraph model, to deal with networks where a vertex behavior is influenced by an observed partition variable. We derived a variational Bayes EM algorithm to infer the model parameters from data and applied it to an ecclesiastical network from Merovingian Gaul. The results of the fitted RSM enlightened us on the different levels of power present at this time in Gaul, and on the different power structures of different regions. Let us highlight that the RSM model allows in addition the comparison of subgraphs through the model parameters, in particular, the cluster proportions. We also would like to mention that networks with typed edges and subgraphs can be encountered in many application fields (such as biology, economics, archeology, ...) and the RSM model should be useful in these contexts as well.

One aspect, however, that RSM does not currently address is the temporality of the data. Since this aspect can be found in many of the data sets we wish to apply the RSM model to, we believe that a natural continuation of this work would be a dynamic extension of the RSM model. Moreover, we plan to introduce a Chinese restaurant process on the latent cluster structure in order to automatically estimate the number of clusters while clustering the vertices. Finally, we would like to consider the problem of visualizing such networks with typed edges and known subgraphs.

APPENDIX: VARIATIONAL BAYES

In this final section we detail the computations that lead to the update rules given in Section 2, and provide an explicit expression of the criterion $\mathcal{L}(q)$.

PROPOSITION A.1.   *The complete data log-likelihood the RSM model is given by*

$$\log p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})$$

$$= \sum_{i \neq j}^{N} \sum_{c=1}^{C} \sum_{k,l}^{K} \{\delta(X_{ij} = c) Z_{ik} Z_{jl} \log(\Pi_{klc})\}$$

$$+ \sum_{i=1}^{N} \sum_{k=1}^{K} Z_{ik} \log(\alpha_{r_i,k})$$

$$+ \sum_{i \neq j}^{N} \{A_{ij} \log(\gamma_{r_i,r_j}) + (1 - A_{ij}) \log(1 - \gamma_{r_i,r_j})\}$$

$$+ \sum_{s=1}^{S} \log p(\boldsymbol{\alpha}_s) + \sum_{r,s}^{S} \log p(\gamma_{rs}) + \sum_{k,l}^{K} \log p(\boldsymbol{\Pi}_{kl}).$$

PROOF.

$$\log p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})$$

$$= \log p(\mathbf{X}|\mathbf{A}, \mathbf{Z}, \boldsymbol{\Pi}) + \log p(\mathbf{A}|\boldsymbol{\gamma}) + \log p(\mathbf{Z}|\boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha}) + \log p(\boldsymbol{\gamma})$$

$$+ \log p(\boldsymbol{\Pi})$$

$$= \sum_{i \neq j}^{N} \sum_{c=1}^{C} \sum_{k,l}^{K} \{\delta(X_{ij} = c) Z_{ik} Z_{jl} \log(\Pi_{klc})\}$$

$$+ \sum_{i=1}^{N} \sum_{k=1}^{K} Z_{ik} \log(\alpha_{r_i,k})$$

$$+ \sum_{i \neq j}^{N} \{A_{ij} \log(\gamma_{r_i,r_j}) + (1 - A_{ij}) \log(1 - \gamma_{r_i,r_j})\}$$

$$+ \sum_{s=1}^{S} \log p(\boldsymbol{\alpha}_s) + \sum_{r,s}^{S} \log p(\gamma_{rs}) + \sum_{k,l}^{K} \log p(\boldsymbol{\Pi}_{kl}). \qquad \square$$

PROPOSITION A.2.   *The VBEM update step for the distribution $q(\gamma_{rs})$ is given by*

$$q(\gamma_{rs}) = \mathrm{Beta}(\gamma_{rs}; a_{rs}, b_{rs}) \qquad \forall (r, s) \in \{1, \dots, S\}^2,$$

where

$$a_{rs} = a_{rs}^0 + \sum_{r_i = r, r_j = s} (A_{ij})$$

and

$$b_{rs} = b_{rs}^0 + \sum_{r_i=r,r_j=s} (1 - A_{ij}).$$

PROOF.

$$
\begin{aligned}
&\log q(\gamma_{rs}) \\
&\quad = \mathbb{E}_{\mathbf{Z},\boldsymbol{\alpha},\boldsymbol{\gamma}^{\backslash rs},\boldsymbol{\Pi}} \big[\log p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})\big] + \kappa \\
&\quad = \sum_{r_i=r,r_j=s} \{A_{ij} \log(\gamma_{rs}) + (1 - A_{ij}) \log(1 - \gamma_{rs})\}, \\
&\log p(\gamma_{rs}) + \kappa \\
&\quad = \sum_{r_i=r,r_j=s} \{A_{ij} \log(\gamma_{rs}) + (1 - A_{ij}) \log(1 - \gamma_{rs})\}, \\
&(a_{rs}^0 - 1) \log(\gamma_{rs}) + (b_{rs}^0 - 1) \log(1 - \gamma_{rs}) + \kappa \\
&\quad = \bigg(a_{rs}^0 - 1 + \sum_{r_i=r,r_j=s} A_{ij}\bigg) \log(\gamma_{rs}) \\
&\qquad + \bigg(b_{rs}^0 - 1 + \sum_{r_i=r,r_j=s} (1 - A_{ij})\bigg) \log(1 - \gamma_{rs}) + \kappa,
\end{aligned}
$$

where κ is a constant term. Hence, the functional form of the variational approximation $q(\gamma_{rs})$ corresponds to a Beta distribution with updated hyperparameters:

$$a_{rs} = a_{rs}^0 + \sum_{r_i=r,r_j=s} (A_{ij})$$

and

$$b_{rs} = b_{rs}^0 + \sum_{r_i=r,r_j=s} (1 - A_{ij}). \qquad \square$$

PROPOSITION A.3. *The VBEM update step for the distribution $q(\mathbf{Z}_i)$ is given by*

$$q(\mathbf{Z}_i) = \mathcal{M}(\mathbf{Z}_i; 1, \boldsymbol{\tau}_i) \qquad \forall i,$$

where

$$
\begin{aligned}
\tau_{ik} \propto{} & \exp\bigg(\psi(\chi_{r_i,k}) - \psi\bigg(\sum_{l=1}^{K} \chi_{r_i,l}\bigg)\bigg) \\
& + \exp\bigg\{\sum_{j \neq i}^{N} \sum_{c=1}^{C} \sum_{l=1}^{K} \delta(X_{ij} = c)\tau_{jl}\bigg(\psi(\Xi_{klc}) - \psi\bigg(\sum_{u=1}^{C} \Xi_{klu}\bigg)\bigg)\bigg\} \\
& + \exp\bigg\{\sum_{j \neq i}^{N} \sum_{c=1}^{C} \sum_{l=1}^{K} \delta(X_{ji} = c)\tau_{jl}\bigg(\psi(\Xi_{lkc}) - \psi\bigg(\sum_{u=1}^{C} \Xi_{lku}\bigg)\bigg)\bigg\}.
\end{aligned}
$$

PROOF.

$$\log q(\mathbf{Z}_i) = \mathbb{E}_{\mathbf{Z}^{\backslash i}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi}}\big[\log p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})\big] + \kappa$$

$$= \mathbb{E}_{\mathbf{Z}^{\backslash i}, \boldsymbol{\Pi}}\left[\sum_{j=1}^{N}\sum_{c=1}^{C}\left\{\delta(X_{ij} = c)\sum_{k,l}^{K} Z_{ik}Z_{jl}\log(\Pi_{klc})\right\}\right]$$

$$+ \mathbb{E}_{\mathbf{Z}^{\backslash i}, \boldsymbol{\Pi}}\left[\sum_{j=1}^{N}\sum_{c=1}^{C}\left\{\delta(X_{ji} = c)\sum_{k,l}^{K} Z_{jk}Z_{il}\log(\Pi_{klc})\right\}\right]$$

$$+ \mathbb{E}_{\mathbf{Z}^{\backslash i}, \boldsymbol{\alpha}}\left[\sum_{k=1}^{K} Z_{ik}\log(\alpha_{r_i, k})\right] + \kappa$$

$$= \sum_{k=1}^{K} Z_{ik}\mathbb{E}_{\boldsymbol{\alpha}}\big[\log(\alpha_{r_i, k})\big]$$

$$+ \sum_{j=1}^{N}\sum_{c=1}^{C}\sum_{l,k}^{K} Z_{ik}\delta(X_{ij} = c)\mathbb{E}_{\mathbf{Z}^{\backslash i}, \boldsymbol{\Pi}}\big[Z_{jl}\log(\Pi_{klc})\big]$$

$$+ \sum_{j=1}^{N}\sum_{c=1}^{C}\sum_{l,k}^{K} Z_{ik}\delta(X_{ji} = c)\mathbb{E}_{\mathbf{Z}^{\backslash i}, \boldsymbol{\Pi}}\big[Z_{jl}\log(\Pi_{lkc})\big] + \kappa$$

$$= \sum_{k=1}^{K} Z_{ik}\left(\psi(\chi_{r_i, k}) - \psi\left(\sum_{l=1}^{K}\chi_{r_i, l}\right)\right)$$

$$+ \sum_{k=1}^{K} Z_{ik}\left\{\sum_{j\neq i}^{N}\sum_{c=1}^{C}\sum_{l=1}^{K}\delta(X_{ij} = c)\tau_{jl}\left(\psi(\Xi_{klc}) - \psi\left(\sum_{u=1}^{C}\Xi_{klu}\right)\right)\right\}$$

$$+ \sum_{k=1}^{K} Z_{ik}\left\{\sum_{j\neq i}^{N}\sum_{c=1}^{C}\sum_{l=1}^{K}\delta(X_{ji} = c)\tau_{jl}\left(\psi(\Xi_{lkc}) - \psi\left(\sum_{u=1}^{C}\Xi_{lku}\right)\right)\right\}$$

$$+ \kappa,$$

where κ is a constant term. Hence, the functional form of the variational approximation $q(\mathbf{Z}_i)$ corresponds to a multinomial distribution, with updated parameters:

$$\tau_{ik} \propto \exp\left(\psi(\chi_{r_i, k}) - \psi\left(\sum_{l=1}^{K}\chi_{r_i, l}\right)\right)$$

$$+ \exp\left\{\sum_{j\neq i}^{N}\sum_{c=1}^{C}\sum_{l=1}^{K}\delta(X_{ij} = c)\tau_{jl}\left(\psi(\Xi_{klc}) - \psi\left(\sum_{u=1}^{C}\Xi_{klu}\right)\right)\right\}$$

$$+ \exp\left\{\sum_{j\neq i}^{N}\sum_{c=1}^{C}\sum_{l=1}^{K}\delta(X_{ji} = c)\tau_{jl}\left(\psi(\Xi_{lkc}) - \psi\left(\sum_{u=1}^{C}\Xi_{lku}\right)\right)\right\}. \qquad \square$$

PROPOSITION A.4.   *The VBEM update step for the distribution $q(\boldsymbol{\alpha}_s)$ is given by*

$$q(\boldsymbol{\alpha}_s) = \mathrm{Dir}(\boldsymbol{\alpha}_s; \chi_s) \qquad \forall s \in \{1, \ldots, S\},$$

where

$$\chi_{sk} = \chi_{sk}^0 + \sum_{i=1}^N \delta(r_i = s)\tau_{ik} \qquad \forall k \in \{1, \ldots, K\}.$$

PROOF.

$$\log q(\boldsymbol{\alpha}_s) = \mathbb{E}_{\mathbf{Z},\boldsymbol{\alpha}\setminus s,\boldsymbol{\gamma},\boldsymbol{\Pi}}\big[\log p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})\big] + \kappa$$

$$= \mathbb{E}_{\mathbf{Z},\boldsymbol{\alpha}\setminus s}\left[\sum_{i=1}^N \sum_{k=1}^K Z_{ik}\log(\alpha_{r_i,k})\right] + \log p(\boldsymbol{\alpha}_s) + \kappa$$

$$= \sum_{k=1}^K \sum_{i=1}^N \delta(r_i = s)\log(\alpha_{sk})\mathbb{E}_{\mathbf{Z}}[Z_{ik}] + \sum_{k=1}^K \log(\alpha_{sk})(\chi_{sk}^0 - 1) + \kappa$$

$$= \sum_{k=1}^K \log(\alpha_{sk})\left\{\chi_{sk}^0 - 1 + \sum_{i=1}^n \delta(r_i = s)\tau_{ik}\right\} + \kappa,$$

where κ is a constant term. Hence, the functional form of the variational approximation $q(\boldsymbol{\alpha}_s)$ corresponds to a Dirichlet distribution with updated hyperparameters:

$$\chi_{sk} = \chi_{sk}^0 + \sum_{i=1}^N \delta(r_i = s)\tau_{ik} \qquad \forall k \in \{1, \ldots, K\}. \qquad \square$$

PROPOSITION A.5.   *The VBEM update step for the distribution $q(\boldsymbol{\Pi}_{kl})$ is given by*

$$q(\boldsymbol{\Pi}_{kl}) = \mathrm{Dir}(\boldsymbol{\Pi}_{kl}; \Xi_{kl}) \qquad \forall (k, l) \in \{1, \ldots, K\}^2,$$

where

$$\Xi_{klc} = \Xi_{klc}^0 + \sum_{i \neq j}^N \delta(X_{ij} = c)\tau_{ik}\tau_{jl} \qquad \forall c \in \{1, \ldots, C\}.$$

PROOF.

$$\log q(\boldsymbol{\Pi}_{k,l}) = \mathbb{E}_{\mathbf{Z},\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\Pi}\setminus kl}\big[\log p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})\big] + \kappa$$

$$= \mathbb{E}_{\mathbf{Z},\boldsymbol{\Pi}\setminus kl}\left[\sum_{i \neq j}^N \sum_{c=1}^C \delta(X_{ij} = c)Z_{ik}Z_{jl}\log(\Pi_{klc})\right] + \log p(\boldsymbol{\Pi}_{kl}) + \kappa$$

$$= \sum_{c=1}^{C} \log(\Pi_{klc}) \left\{ \sum_{i \neq j}^{N} \delta(X_{ij} = c)\tau_{ik}\tau_{jl} \right\}$$

$$+ \sum_{c=1}^{C} \log(\mathbf{\Pi}_{klc})(\Xi_{klc}^{0} - 1) + \kappa$$

$$= \sum_{c=1}^{C} \log(\Pi_{klc}) \left\{ \Xi_{klc}^{0} - 1 + \sum_{i \neq j}^{N} \delta(X_{ij} = c)\tau_{ik}\tau_{jl} \right\} + \kappa,$$

where κ is a constant term. Hence, the functional form of the variational approximation $q(\mathbf{\Pi}_{kl})$ corresponds to a Dirichlet distribution with updated hyperparameters:

$$\Xi_{klc} = \Xi_{klc}^{0} + \sum_{i \neq j}^{N} \delta(X_{ij} = c)\tau_{ik}\tau_{jl} \qquad \forall c \in \{1, \dots, C\}.$$

$\square$

PROPOSITION A.6. *When computed right after the M step, the lower bound of the marginal log-likelihood is given by*

$$\mathcal{L}(q) = \sum_{r,s}^{S} \log\left(\frac{B(a_{rs}, b_{rs})}{B(a_{rs}^{0}, b_{rs}^{0})}\right) + \sum_{s=1}^{S} \log\left(\frac{C(\chi_s)}{C(\chi_s^{0})}\right)$$

$$+ \sum_{k,l}^{K} \log\left(\frac{C(\Xi_{kl})}{C(\Xi_{kl}^{0})}\right) - \sum_{i=1}^{N} \sum_{k=1}^{K} \tau_{ik} \log(\tau_{ik}),$$

where $C(x) = \frac{\prod_{d=1}^{D} \Gamma(x_d)}{\Gamma(\sum_{d=1}^{D} x_d)}$ if $x \in \mathbb{R}^D$ and $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \forall(a, b) \in \mathbb{R}^2$.

PROOF. The lower bound is given by

$$\mathcal{L}(q) = \mathbb{E}_{\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{\Pi}} \left[\log\left(\frac{p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{\Pi})}{q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{\Pi})}\right) \right],$$

where

$$\log\left(\frac{p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{\Pi})}{q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{\Pi})}\right) = \sum_{i \neq j}^{N} \{A_{ij} \log(\gamma_{r_i, r_j}) + (1 - A_{ij}) \log(1 - \gamma_{r_i, r_j})\}$$

$$+ \sum_{i \neq j}^{N} \sum_{c=1}^{C} \sum_{k,l}^{K} \{\delta(X_{ij} = c)Z_{ik}Z_{jl} \log(\Pi_{klc})\}$$

$$+ \log\left(\frac{p(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{\Pi})}{q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{\Pi})}\right)$$

and

$$\log\left(\frac{p(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}{q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}\right)$$

$$= \log\left(\frac{p(\mathbf{Z}|\boldsymbol{\alpha})}{q(\mathbf{Z})}\right) + \log\left(\frac{p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}{q(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}\right)$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} Z_{ik} \log\left(\frac{\alpha_{r_i,k}}{\tau_{ik}}\right) + \sum_{r,s} \log\left(\frac{\text{Beta}(\gamma_{rs}; a_{rs}^0, b_{rs}^0)}{\text{Beta}(\gamma_{rs}; a_{rs}, b_{rs})}\right)$$

$$+ \sum_{s=1}^{S} \log\left(\frac{\text{Dir}(\boldsymbol{\alpha}_s; \boldsymbol{\chi}_s^0)}{\text{Dir}(\boldsymbol{\alpha}_s; \boldsymbol{\chi}_s)}\right) + \sum_{k,l}^{K} \log\left(\frac{\text{Dir}(\boldsymbol{\Pi}_{kl}; \boldsymbol{\Xi}_{kl}^0)}{\text{Dir}(\boldsymbol{\Pi}_{kl}; \boldsymbol{\Xi}_{kl})}\right)$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} Z_{ik} \log\left(\frac{\alpha_{r_i,k}}{\tau_{ik}}\right) + \sum_{r,s} \log\left(\frac{B(a_{rs}, b_{rs})}{B(a_{rs}^0, b_{rs}^0)\gamma_{rs}^{a_{rs}-a_{rs}^0}(1-\gamma_{rs})^{b_{rs}-b_{rs}^0}}\right)$$

$$+ \sum_{s=1}^{S} \log\left(\frac{C(\boldsymbol{\chi}_s)}{C(\boldsymbol{\chi}_s^0)\prod_{k=1}^{K}\alpha_k^{\chi_{sk}-\chi_{sk}^0}}\right) + \sum_{k,l}^{K} \log\left(\frac{C(\boldsymbol{\Xi}_{kl})}{C(\boldsymbol{\Xi}_{kl}^0)\prod_{c=1}^{C}\Pi_{klc}^{\Xi_{klc}-\Xi_{klc}^0}}\right).$$

If $x \in \mathbb{R}^D$, then $C(x) = \frac{\prod_{d=1}^{D}\Gamma(x_d)}{\Gamma(\sum_{d=1}^{D}x_d)}$, where $\Gamma(\cdot)$ is the gamma function. Moreover, if $(a, b) \in \mathbb{R}^2$, then $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. Finally,

$$\log\left(\frac{p(\mathbf{X}, \mathbf{A}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}{q(\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Pi})}\right)$$

$$= \sum_{r,s}^{S} \left\{\left(a_{rs}^0 - a_{rs} + \sum_{r_i=r, r_j=s} A_{ij}\right)\log(\gamma_{rs})\right\}$$

$$+ \sum_{r,s}^{S} \left\{\left(b_{rs}^0 - b_{rs} + \sum_{r_i=r, r_j=s}(1 - A_{ij})\right)\log(1 - \gamma_{rs})\right\}$$

$$+ \sum_{s=1}^{S}\sum_{k=1}^{K} \left\{\left(\chi_{sk}^0 - \chi_{sk} + \sum_{i=1}^{N}\delta(r_i=s)Z_{ik}\right)\log(\alpha_{sk})\right\}$$

$$+ \sum_{k,l}^{K}\sum_{c=1}^{C} \left\{\left(\Xi_{klc}^0 - \Xi_{klc} + \sum_{i\neq j}^{N}\delta(X_{ij}=c)Z_{ik}Z_{jl}\right)\log(\Pi_{klc})\right\}$$

$$- \sum_{i=1}^{N}\sum_{k=1}^{K} Z_{ik}\log(\tau_{ik}) + \sum_{r,s}^{S}\log\left(\frac{B(a_{rs}, b_{rs})}{B(a_{rs}^0, b_{rs}^0)}\right)$$

$$+ \sum_{s=1}^{S}\log\left(\frac{C(\boldsymbol{\chi}_s)}{C(\boldsymbol{\chi}_s^0)}\right) + \sum_{k,l}^{K}\log\left(\frac{C(\boldsymbol{\Xi}_{kl})}{C(\boldsymbol{\Xi}_{kl}^0)}\right).$$

Therefore, if $\mathcal{L}(q)$ is computed right after the M step,

$$\mathcal{L}(q) = \sum_{r,s}^{S} \log\left(\frac{B(a_{rs}, b_{rs})}{B(a_{rs}^0, b_{rs}^0)}\right) + \sum_{s=1}^{S} \log\left(\frac{C(\boldsymbol{\chi}_s)}{C(\boldsymbol{\chi}_s^0)}\right)$$
$$+ \sum_{k,l}^{K} \log\left(\frac{C(\boldsymbol{\Xi}_{kl})}{C(\boldsymbol{\Xi}_{kl}^0)}\right) - \sum_{i=1}^{N}\sum_{k=1}^{K} \tau_{ik} \log(\tau_{ik}). \qquad \square$$

SUPPLEMENTARY MATERIAL

Data and code (DOI: 10.1214/13-AOAS691SUPP; .zip). We provide the original ecclesiastical network along with a file giving the kingdoms of all vertices in the network and an R code implementing the variational inference approach for the RSM model.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** 47–97. MR1895096

AMBROISE, C., GRASSEAU, G., HOEBEKE, M., LATOUCHE, P., MIELE, V. and PICARD, F. (2010). The mixer R package. Available at http://cran.r-project.org/web/packages/mixer/.

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

BIERNACKI, C., CELEUX, G. and GOVAERT, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Comput. Statist. Data Anal.* **41** 561–575. MR1968069

BILMES, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute* **4** 126.

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York. MR2247587

CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. MR2988467

DAUDIN, J. J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. MR2390817

FIENBERG, S. E. and WASSERMAN, S. S. (1981). Categorical data analysis of single sociometric relations. *Sociol. Method.* **12** 156–192.

FRANK, O. and HARARY, F. (1982). Cluster inference by using transitivity indices in empirical graphs. *J. Amer. Statist. Assoc.* **77** 835–840. MR0686407

GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826 (electronic). MR1908073

GOLDENBERG, A., ZHENG, A. X. and FIENBERG, S. E. (2010). *A Survey of Statistical Network Models*. Now Publishers, Hanover, MA.

HANDCOCK, M. S., RAFTERY, A. E. and TANTRUM, J. M. (2007). Model-based clustering for social networks. *J. Roy. Statist. Soc. Ser. A* **170** 301–354. MR2364300

HOFMAN, J. M. and WIGGINS, C. H. (2008). Bayesian approach to network modularity. *Phys. Rev. Lett.* **100** 258701.

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London. Ser. A.* **186** 453–461. MR0017504

JERNITE, Y., LATOUCHE, P., BOUVEYRON, C., RIVERA, P., JEGOU, L. and LAMASSÉ, S. (2013). Supplement to "The random subgraph model for the analysis of an ecclesiastical network in Merovingian Gaul." DOI:10.1214/13-AOAS691SUPP.

KEMP, C., TENENBAUM, J. B., GRIFFITHS, T. L., YAMADA, T. and UEDA, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the National Conference on Artificial Intelligence* **21** 381. AAAI Press, Boston, MA.

LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2009). Advances in Data Analysis Data Handling and Business Intelligence Bayesian methods for graph clustering, 229–239. Springer, Berlin.

LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2011). Overlapping stochastic block models with application to the French political blogosphere. *Ann. Appl. Stat.* **5** 309–336. MR2810399

LATOUCHE, P., BIRMELÉ, E. and AMBROISE, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Stat. Model.* **12** 93–115. MR2953099

MARIADASSOU, M. and MATIAS, C. (2014). Convergence of the groups posterior distribution in latent or stochastic block models. *Bernoulli*. To appear.

MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, D., CHKLOVSKII, D. and ALON, U. (2002). Network motifs: Simple building blocks of complex networks. *Science* **298** 824–827.

MORENO, J. L. (1934). *Who Shall Survive?*: *A New Approach to the Problem of Human Interrelations*. Nervous and Mental Disease Publishing Co, Washington, DC.

NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255

PALLA, G., DERÉNYI, I., FARKAS, I. and VICSEK, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435** 814–818.

RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* 846–850.

SALTER-TOWNSHEND, M., WHITE, A., GOLLINI, I. and MURPHY, T. B. (2012). Review of statistical network analysis: Models, algorithms, and software. *Stat. Anal. Data Min.* **5** 260–264. MR2958152

SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *J. Classification* **14** 75–100. MR1449742

SOUFIANI, H. A. and AIROLDI, E. M. (2012). Graphlet decomposition of a weighted network. *J. Mach. Learn. Res.* **22** 54–63.

VILLA, N., ROSSI, F. and TRUONG, Q. D. (2008). Mining a medieval social network by kernel SOM and related methods. In *Proceedings of MASHS 2008* (*Modèles et Apprentissage en Sciences Humaines et Sociales*), Créteil, France, June 2008.
KEMP, C., TENENBAUM, J. B., GRIFFITHS, T. L., YAMADA, T. and UEDA, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the National Conference on Artificial Intelligence* **21** 381. AAAI Press, Boston, MA.

WANG, Y. J. and WONG, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.* **82** 8–19. MR0883333

WHITE, H. C., BOORMAN, S. A. and BREIGER, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *Am. J. Sociol.* 730–780.

XING, E. P., FU, W. and SONG, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.* **4** 535–566. MR2758639

Y. JERNITE
LABORATOIRE SAMM, EA 4543
UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE
90 RUE DE TOLBIAC
75634 PARIS CEDEX 13
AND
ECOLE POLYTECHNIQUE
91128 PALAISEAU CEDEX
FRANCE

C. BOUVEYRON
LABORATOIRE MAP5, UMR CNRS 8145
UNIVERSITÉ PARIS DESCARTES
   AND SORBONNE PARIS CITÉ
45 RUE DES SAINTS PÈRES
75006 PARIS
FRANCE

P. LATOUCHE
LABORATOIRE SAMM, EA 4543
UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE
90 RUE DE TOLBIAC
75634 PARIS CEDEX 13
FRANCE
E-MAIL: pierre.latouche@univ-paris1.fr

P. RIVERA
L. JEGOU
S. LAMASSÉ
LABORATOIRE LAMOP, UMR 8589
UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE
17 RUE DE LA SORBONNE
75005 PARIS
FRANCE