The Annals of Applied Statistics
2010, Vol. 4, No. 2, 715–742
DOI: 10.1214/10-AOAS361

UNCOVERING LATENT STRUCTURE IN VALUED GRAPHS: A VARIATIONAL APPROACH

BY MAHENDRA MARIADASSOU, STÉPHANE ROBIN AND CORINNE VACHER

AgroParisTech and INRA, AgroParisTech and INRA and INRA and University Bordeaux I

As more and more network-structured data sets are available, the statistical analysis of valued graphs has become common place. Looking for a latent structure is one of the many strategies used to better understand the behavior of a network. Several methods already exist for the binary case.

We present a model-based strategy to uncover groups of nodes in valued graphs. This framework can be used for a wide span of parametric random graphs models and allows to include covariates. Variational tools allow us to achieve approximate maximum likelihood estimation of the parameters of these models. We provide a simulation study showing that our estimation method performs well over a broad range of situations. We apply this method to analyze host–parasite interaction networks in forest ecosystems.

1. Introduction. Data sets presenting a network structure are increasingly studied in many different domains such as sociology, energy, communication, ecology or biology [Albert and Barabási (2002)]. Statistical tools are therefore needed to analyze the structure of these networks, in order to understand their properties or behavior. A strong attention has been paid to the study of various topological characteristics such as degree distribution, clustering coefficient and diameter [see, e.g., Barabási and Albert (1999), Newman, Watts and Strogatz (2002)]. These characteristics are useful to describe networks but not sufficient to understand its whole structure.

A natural and intuitive way to capture an underlying structure is to look for groups of edges having similar connection profiles [Getoor and Diehl (2004), Newman, Watts and Strogatz (2002)], which is refereed to as community detection [Girvan and Newman (2002), Newman (2004)]. This usually turns into an unsupervised classification (or clustering) problem which requires efficient estimation algorithms since the data set at hand is ever increasing.

Several attempts at community detection have been proposed in the literature: greedy algorithms for community detection [Girvan and Newman (2002) and Newman (2004)] and clustering based on spectral analysis of the adjacency matrix of a graph [von Luxburg, Belkin and Bousquet (2008)]. Greedy algorithms and spectral clustering both assume that communities are determined by a strong

Received October 2008; revised April 2010.
Key words and phrases. Ecological networks, host–parasite interactions, latent structure, mixture model, random graph, valued graph, variational method.

within connectivity opposed to a low between connectivity. This might be true for so-called communities but need not be true for other groups of nodes. For example, a group of nodes loosely connected to each other but highly connected to a specific group of hubs have the same connection profile and form a homogeneous group but do not form a community. In addition, they do not offer an explicit generative model nor a criterion to select the correct number of communities.

Model-based methods are appealing by contrast: explicit modeling of the heterogeneity between nodes gives different groups an intuitive and easy to understand interpretation. Several probabilistic models exists for random graphs [see Pattison and Robins (2007) for a complete review], ranging from the seminal Erdös–Rényi (ER) model [Erdös and Rényi (1959)] to the sophisticated Stochastic Block Model (SBM) [Nowicki and Snijders (2001)]. The ER model assumes independent and identically distributed edges which entails that all nodes are structurally equivalent and, thus, there is only one community, although a big one. The p_1 model from Holland and Leinhardt (1981) extended the ER model by assuming independent *dyads* instead of *edges*, allowing the breakthrough from undirected to directed graphs. But again, all nodes are structurally equivalent in the p_1 model. Fienberg and Wasserman (1981) and Fienberg, Meyer and Wasserman (1985) lifted these constraints by assuming the nodes are distributed among Q classes with different connectivity profiles. In this model, groups are easily interpreted as nodes belonging to the same class. Unfortunately, Fienberg, Meyer and Wasserman (1985) assumes class assignments are perfectly well known, which rarely happens. The state of the art in terms of graph modeling is the SBM, inspired by Lorrain and White (1971) and introduced by Nowicki and Snijders (2001), which takes advantage of mixture models and unknown latent variables to allow an easy modeling of groups without requiring them to be known in advance.

In the SBM framework, community detection boils down to three crucial steps: assignment of nodes to groups, estimation of the model parameter and selection of the correct number of groups. Several authors offered their method to solve these issues using Bayesian methods. Nowicki and Snijders (2001) work with the original SBM model. Hofman and Wiggins (2008) work in a highly constrained version of SBM in which heterogeneity is strictly limited to intra- and inter-community connection and thus characterized by only two parameters, against Q^2 in the unconstrained SBM. Airoldi et al. (2008) extend the SBM framework by allowing nodes to exhibit multiple communities. By contrast, Daudin, Picard and Robin (2008) use a frequentist approach to estimate the parameters of the SBM. The frequentist approach is less computation intensive than its Bayesian counterpart, whereas the Bayesian approach is supposed to better account for the uncertainty. With the notable exception of Nowicki and Snijders (2001), who use MCMC to estimate the model parameter, both lines of work make heavy use of variational techniques: either Variational EM [Jaakkola (2000)] or Variational Bayes [Attias (2000); Beal and Ghahramani (2003); Xing, Jordan and Russell (2003); Winn, Bishop and Jaakkola (2005)]. MCMC computational cost is prohibitive,

effectively leading to severe size limitations (around 200 nodes). Furthermore, because of the complex likelihood landscape in the SBM, good mixing of the Markov Chain is hard to achieve and monitor. Variational approximations, by contrast, replace the likelihood by a simple surrogate, chosen so that the error is minimal in some sense. Frequentist and Bayesian approach then differ only in the use of this surrogate likelihood: Bayesians combine it to a prior distribution of the parameter (chosen from some suitable distribution), whereas frequentists use it directly. In all these methods, the number of groups is fixed during the estimation procedure and must be selected using some criterion. By contrast, Kemp, Griffiths and Tenenbaum (2004) propose an original approach where the number of groups changes and is selected *during* the estimation process. Both Bayesian and frequentist estimations approaches give the same kind of results: an optimal number of groups and a probabilistic assignment of nodes to groups, depending on their connection profile. However, the Bayesian estimation strategy leads to severe constraints on the choice of prior and hyperprior distributions. The Daudin, Picard and Robin (2008) maximum likelihood approach does not require any prior specification and is more efficient than MCMC estimation [Picard et al. (2007)].

Previous models are all models for binary networks, for which the only information is the presence or absence of an edge. Binary information certainly describes the topology of a network but is a rather poor description. It accounts neither for the intensity of the interaction between two nodes nor for the specific features of an edge. The intensity of an edge may typically indicate the amount of energy transported from one node to another, the number of passengers or the number of common features between two nodes, whereas the specific feature of an edge may be the phylogenetic distance between its two ending nodes. Many networks, such as power, communication, social, ecological or biological networks, are naturally valued and are somehow arbitrarily transformed to a binary graph. This transformation sometimes conceals important results [Tykiakanis, Tscharntke and Lewis (2007)]. Extending binary models and the associated estimation procedures to valued graphs with specific features allows more complexity, and more relevant information with it, to be processed while estimating the structure of the network.

We are motivated by the search of a structure in valued graphs describing the similarity between species within an assemblage according to their biotic interactions. In ecology, an assemblage is defined as a taxonomically related group of species that occurs in the same geographic area [Ricklefs and Miller (2000)]. The species composing an assemblage usually interact with many species belonging to other assemblages and the nature of these interactions is often very diverse (predator–prey interactions, host–parasite interactions, mutualistic interactions, competitive interactions). One of the questions facing ecologists is to understand what determines with whom a species interact. Conventional wisdom is that within an assemblage, two closely related species should share more interactions

than two evolutionary distant species because the range of interactions of a species is constrained by its physiological, morphological and behavioral attributes. In several cases, this conventional wisdom is revealed to be true. Phylogenetically related plant species have been shown to bear similar pathogens and herbivores [Brandle and Brandl (2006); Gilbert and Webb (2007)] and the diet's range of predators has been shown to be phylogenetically constrained [Cattin et al. (2004)]. This tendency for phylogenetically related species to resemble each other is called phylogenetic signal [Blomberg and Garland (2002)]. In other cases, no phylogenetic signal was detected [Rezende et al. (2007); Vacher, Piou and Desprez-Loustau (2008)]. Selection pressures exerted by the environment might account for this absence: species have to adapt to varying environments to survive, diverging from close relatives in their physiology, morphology and behavior, and possibly developing novel interactions [Bersier and Kehrli (2008); Cattin et al. (2004)]. The valued graphs under study have species as nodes and the number of shared interactions as edges. We use a mixture model with phylogenetic distance between species as covariate to measure the strength of the phylogenetic signal. This latter is defined as the decrease in the number of selected groups due to the inclusion of the covariate. Two different assemblages are considered. The first assemblage is composed of 51 tree species occurring in the French forests and the second is composed of 153 parasitic fungal species also occurring in the French forests. The interactions considered are host–parasite interactions. We expect to find a lower phylogenetic signal in the host range of parasitic fungal species [Bersier and Kehrli (2008); Rossberg et al. (2006); Vacher, Piou and Desprez-Loustau (2008)] than in the vulnerability of tree species to parasites [Brandle and Brandl (2006); Gilbert and Webb (2007); Vacher, Piou and Desprez-Loustau (2008)].

In this paper we propose an extension to the stochastic block model, introduced in Fienberg and Wasserman (1981); Fienberg, Meyer and Wasserman (1985); Nowicki and Snijders (2001), and the methods of Airoldi and Carley (2005) and Daudin, Picard and Robin (2008), that deals with valued graphs and accounts for possible covariates. We use a general mixture model describing the connection intensities between nodes spread among a certain number of classes (Section 2). A variational EM approach to get an optimal, in a sense to be defined, approximation of the likelihood is then presented in Section 3. In Section 4 we give a general estimation algorithm and derive some explicit formulas for the most popular distributions. The quality of the estimates is studied on synthetic data in Section 5. Finally, the model is used to elucidate the structure of host–parasite interactions in forest ecosystems and results are discussed in Section 6.

2. Mixture model. We now present the general extension of SBM to valued graphs and discuss the two particular modelings used for the tree species and fungal species interaction networks.

2.1. *Model and notation.*

Nodes. Consider a graph with n nodes, labeled in $\{1, \ldots, n\}$. In our model the nodes are distributed among Q groups so that each node i is associated to a random vector $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iQ})$, with Z_{iq} being 1 if node i belongs to group q and 0 otherwise. The $\{\mathbf{Z}_i\}$ are supposed to be independent identically distributed observations from a multinomial distribution:

$$\{\mathbf{Z}_i\}_i \text{ i.i.d.} \sim \mathcal{M}(1; \boldsymbol{\alpha}), \tag{2.1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$ and $\sum_q \alpha_q = 1$.

Edges. Each edge from a node i to a node j is associated to a random variable X_{ij}, coding for the strength of the edge. Conditionally to the group of each node, or equivalently knowing the $\{\mathbf{Z}_i\}$, the edges are supposed to be independent. Knowing group q of node i and group ℓ of node j, X_{ij} is distributed as $f(\cdot, \theta_{q\ell}) := f_{q\ell}(\cdot)$, where $f_{\theta_{q\ell}}$ is a probability distribution known up to a finite-dimensional parameter $\theta_{q\ell}$:

$$X_{ij} | i \in q, j \in \ell \sim f(\cdot, \theta_{q\ell}) := f_{q\ell}(\cdot). \tag{2.2}$$

Up to a relabeling of the classes, the model is identifiable and completely specified by both the mixture proportions $\boldsymbol{\alpha}$ and the connectivity matrix $\boldsymbol{\theta} = (\theta_{q\ell})_{q,\ell=1,\ldots,Q}$. We denote $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\theta})$ the parameter of the model.

Directed and undirected graphs. This modeling can be applied to both directed and undirected graphs. In the directed version, the variables X_{ij} and X_{ji} are supposed to be independent conditionally to the groups to which nodes i and j belong. This hypothesis is not always realistic since, for example, the traffic from i to j is likely to be correlated to the traffic from j to i. A way to account for such a dependency is to consider a undirected graph with edges labeled with the bivariate variables $\{(X_{ij}, X_{ji})\}_{1 \le i < j \le n}$. All the results presented in this paper are valid for directed graphs. The results for undirected graphs can easily be derived and are only briefly mentioned.

2.2. *Modeling the number of shared hosts/parasites.* In our tree interaction network, each edge is valued with the number of common fungal species two tree species can host. Our purpose is to understand the structure of this network and it is natural to model the counts X_{ij} as Poisson distributed. The mixture models aims at explaining the heterogeneity of the X_{ij}. However, we would also like to account for some factors that are known to be influential. In our network, we expect two phylogenetically related tree species i and j to share a high number X_{ij} of parasitic species. As such, their average number of shared parasitic species $\mathbb{E}[X_{ij}]$ is expected to decrease with their phylogenetic distance y_{ij}. We consider three alternatives, and compare two of them.

Poisson mixture (PM): In this mixture, we do not account for the covariates and X_{ij} only depends on the classes of i and j:

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{P}(\lambda_{q\ell}).$$

$\lambda_{q\ell}$ is then the mean number of common fungal species (or mean interaction) between a tree species from group q and one from group ℓ and $\theta_{q\ell} = \lambda_{q\ell}$.

Poisson regression mixture with inhomogeneous effects (PRMI): In this mixture, we account for the covariates via a regression model that is *specific* to the classes of i and j:

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{P}(\lambda_{q\ell} e^{\beta_{q\ell}^{\mathsf{T}} \mathbf{y}_{ij}}),$$

where $\mathbf{y}_{ij}$ is a vector of covariates and $\theta_{q\ell} = (\lambda_{q\ell}, \beta_{q\ell})$.

Poisson regression mixture with homogeneous effects (PRMH): In this mixture, the effect of the covariates does not depend on the classes of i and j:

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{P}(\lambda_{q\ell} e^{\beta^{\mathsf{T}} \mathbf{y}_{ij}}),$$

$\theta_{q\ell} = (\lambda_{q\ell}, \beta)$.

We point out that models PRMI and PRMH have different purposes. In PRMI, the link between the covariates and the edges is locally refined *within* each class (q, ℓ), whereas in PRMH, the covariates compete globally with the group structure found by PM. In PRMH, the mixture looks for remaining structure among the residuals of the regression model. If the structure was completely explained by the covariates, the possibly many components found using PM would reduce to a single component when using PRMH. To a lesser extent, we expect the number of components to be smaller with PRMH than with PM if the phylogenetic distance explains part of the structure. As we look for structure beyond the one already explained by the covariates, we consider only models PM and PRMH.

The same models are used for the fungal species interaction network. In our examples, data consist in counts, but other types of data can be handled with similar mixture and/or regression models (see Appendix A.1 for details).

3. Likelihood and variational EM. We now address the estimation of the parameter $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\theta})$. We show that the standard maximum likelihood approach cannot be applied to our model and propose an alternative strategy relying on variational tools, namely, variational EM.

3.1. *Likelihoods.* Let $\mathbf{X}$ denote the set of all edges, $\mathbf{X} = \{X_{ij}\}_{i,j=1,\ldots,n}$, and $\mathbf{Z}$ the set of all indicator variables for nodes, $\mathbf{Z} = \{\mathbf{Z}_i\}_{i=1,\ldots,n}$. In the mixture model literature [McLahan and Peel (2000)] $(\mathbf{X}, \mathbf{Z})$ is referred to as the complete data set, while $\mathbf{X}$ is referred to as the incomplete data set. The conditional independence of the edges knowing $\mathbf{Z}$ entails the decomposition $\log \mathbb{P}(\mathbf{Z}, \mathbf{X}) = \log \mathbb{P}(\mathbf{Z}) +$

$\log \mathbb{P}(\mathbf{X}|\mathbf{Z})$. It then follows from (2.1) and (2.2) that the log-likelihood of the complete data set is

$$\log \mathbb{P}(\mathbf{Z}, \mathbf{X}) = \sum_i \sum_q Z_{iq} \log \alpha_q + \sum_{i \neq j} \sum_{q,\ell} Z_{iq} Z_{j\ell} \log f_{q\ell}(X_{ij}). \tag{3.1}$$

The likelihood of the incomplete data set can be obtained by summing $\mathbb{P}(\mathbf{Z}, \mathbf{X})$ over all possible $\mathbf{Z}$'s: $\mathbb{P}(\mathbf{X}) = \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{Z}, \mathbf{X})$. This summation involves Q^n terms and quickly becomes intractable. The popular E–M algorithm [Dempster, Laird and Rubin (1977)], widely used in mixture problems, allows to maximize $\log \mathbb{P}(\mathbf{X})$ without explicitly calculating it. The E-step relies on the calculation of the conditional distribution of $\mathbf{Z}$ given $\mathbf{X}$: $\mathbb{P}(\mathbf{Z}|\mathbf{X})$. Unfortunately, in the case of network data, the strong dependency between edges makes this calculation untractable.

Undirected graphs. The closed formula (3.1) still holds undirected graphs, replacing the sum over $i \neq j$ by a sum over $i < j$. This is also true for equations (3.6) and (4.3) given below.

3.2. *Variational EM*. We propose to use an approximate maximum likelihood strategy based on a variational approach [see Jordan et al. (1999) or the tutorial by Jaakkola (2000)]. This strategy is also used in Govaert and Nadif (2005) for a biclustering problem. We consider a lower bound of the log-likelihood of the incomplete data set

$$\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma}) = \log \mathbb{P}(\mathbf{X}; \boldsymbol{\gamma}) - KL(R_{\mathbf{X}}(\cdot), \mathbb{P}(\cdot|\mathbf{X}; \boldsymbol{\gamma})), \tag{3.2}$$

where KL denotes the Kullback–Leibler divergence and $R_{\mathbf{X}}$ stands for some distribution on $\mathbf{Z}$. Classical properties of the Kullback–Leibler divergence ensure that $\mathcal{J}$ has a unique maximum $\log \mathbb{P}(\mathbf{X}; \boldsymbol{\gamma})$, which is reached for $R_{\mathbf{X}}(\mathbf{Z}) = \mathbb{P}(\mathbf{Z}|\mathbf{X})$. In other words, if $\mathbb{P}(\mathbf{Z}|\mathbf{X}; \boldsymbol{\gamma})$ was tractable, the maximization of $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ would be equivalent to the maximization of $\log \mathbb{P}(\mathbf{X}; \boldsymbol{\gamma})$. In our case, $\mathbb{P}(\mathbf{Z}|\mathbf{X}; \boldsymbol{\gamma})$ is untractable and we maximize $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ with respect to both $R_{\mathbf{X}}$ and $\boldsymbol{\gamma}$. Jaakkola (2000) shows that $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ can be rewritten as

$$\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma}) = \mathcal{H}(R_{\mathbf{X}}) + \sum_{\mathbf{Z}} R_{\mathbf{X}}(\mathbf{Z}) \log \mathbb{P}(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma}), \tag{3.3}$$

where $\mathcal{H}(\cdot)$ denotes the entropy of a distribution. The last term of (3.3) can be deduced from (3.1):

$$\begin{aligned} &\sum_{\mathbf{Z}} R_{\mathbf{X}}(\mathbf{Z}) \log \mathbb{P}(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma}) \\ &\quad = \sum_i \sum_q \mathbb{E}_{R_{\mathbf{X}}}(Z_{iq}) \log \alpha_q + \sum_{i \neq j} \sum_{q,\ell} \mathbb{E}_{R_{\mathbf{X}}}(Z_{iq} Z_{j\ell}) \log f_{q\ell}(X_{ij}), \end{aligned} \tag{3.4}$$

where $\mathbb{E}_{R_{\mathbf{X}}}$ denotes the expectation with respect to distribution $R_{\mathbf{X}}$. Equation (3.4) requires only the knowledge of $\mathbb{E}_{R_{\mathbf{X}}}(Z_{iq})$ and $\mathbb{E}_{R_{\mathbf{X}}}(Z_{iq} Z_{j\ell})$ for all i, j, q, ℓ. By

contrast, $\mathcal{H}(R_{\mathbf{X}})$ requires all order moments of $R_{\mathbf{X}}$ and is untractable in general. Maximization of $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ in $R_{\mathbf{X}}$ can not be achieved without some restrictions on $R_{\mathbf{X}}$. We therefore limit the search to the class of completely factorized distributions:

$$R_{\mathbf{X}}(\mathbf{Z}) = \prod_i h(\mathbf{Z}_i, \boldsymbol{\tau}_i), \tag{3.5}$$

where h denotes the multinomial distribution and $\boldsymbol{\tau}_i$ stands for a vector of probabilities, $\boldsymbol{\tau}_i = (\tau_{i1}, \ldots, \tau_{iQ})$ (with $\sum_q \tau_{iq} = 1$). In particular, $\mathbb{E}_{R_{\mathbf{X}}}(Z_{iq}) = \tau_{iq}$ and $\mathbb{E}_{R_{\mathbf{X}}}(Z_{iq} Z_{j\ell}) = \tau_{iq}\tau_{j\ell}$. In addition, the entropy is additive over the coordinates for factorized distributions, so that $\mathcal{H}(R_{\mathbf{X}}) = \sum_i \mathcal{H}(h(\cdot, \boldsymbol{\tau}_i)) = -\sum_i \sum_q \tau_{iq} \log \tau_{iq}$. Wrapping everything together,

$$\begin{aligned} \mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma}) = &- \sum_i \sum_q \tau_{iq} \log \tau_{iq} + \sum_i \sum_q \tau_{iq} \log \alpha_q \\ &+ \sum_{i \neq j} \sum_{q,\ell} \tau_{iq}\tau_{j\ell} \log f_{q\ell}(X_{ij}). \end{aligned} \tag{3.6}$$

It is immediate from (3.6) that $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ is tractable for distributions $R_{\mathbf{X}}$ of the form (3.5). The $\boldsymbol{\tau}_i$'s must be thought of as variational parameters to be optimized so that $R_{\mathbf{X}}(\mathbf{Z})$ fits $\mathbb{P}(\mathbf{Z}|\mathbf{X}; \boldsymbol{\gamma})$ as well as possible; they depend on the observed data $\mathbf{X}$. Since $R_{\mathbf{X}}$ is restricted to be of the form (3.5), $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ is a lower bound of $\log \mathbb{P}(\mathbf{X})$.

Discussion about tighter bounds. A fully factorized $R_{\mathbf{X}}$ is only one class of distributions we can consider. Broader distribution classes should yield tighter bound of $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$. Unfortunately, for more general distributions, the entropy $\mathcal{H}(R_{\mathbf{X}})$ may not have a simple expression anymore rendering the exact calculation of $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ untractable: better accuracy is achieved at the expense of tractability. A solution to this issue is Bethe free energy [Yedidia, Freeman and Weiss (2005)]. We did not consider it because it relies on an approximation of $\mathcal{H}(R_{\mathbf{X}})$ which disrupts the well-behaved properties of $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$.

Another approach comes from Leisink and Kappen (2001) and Mariadassou (2006). Starting from an exponential inequality, they emphasize the strong connection between fully factorized $R_{\mathbf{X}}$ and first order linear approximation of the exponential function. Using a higher approximation of the exponential and some distribution $S_{\mathbf{X}}$ in addition to $R_{\mathbf{X}}$, it is possible to derive an even tighter bound of $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$. However, the estimation algorithm is then of complexity $\mathcal{O}(n^6 Q^6)$ instead of $\mathcal{O}(n^2 Q^2)$ for a gain which has the same order of magnitude as the computer numerical precision.

4. Parameter estimation. We present here the two-steps algorithm used for the parameter estimation.

4.1. *Estimation algorithm.* As explained in Section 3.2, the maximum likelihood estimator of $\boldsymbol{\gamma}$ is

$$\widehat{\boldsymbol{\gamma}}_{ML} = \arg\max_{\boldsymbol{\gamma}} \log \mathbb{P}(\mathbf{X}; \boldsymbol{\gamma}) = \arg\max_{\boldsymbol{\gamma}} \max_{R_{\mathbf{X}}} \mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma}).$$

In the variational framework, we restrict the last optimization problem to factorized distributions. The estimate we propose is hence

$$\widehat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \max_{R_{\mathbf{X}} \text{ factorized}} \mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma}).$$

The simultaneous optimization with respect to both $R_{\mathbf{X}}$ and $\boldsymbol{\gamma}$ is still too difficult, so we adopt the following iterative strategy. Denoting by $R_{\mathbf{X}}^{(n)}$ and $\boldsymbol{\gamma}^{(n)}$ the estimates after n steps, we compute

$$\begin{cases} R_{\mathbf{X}}^{(n+1)} = \arg\max_{R_{\mathbf{X}} \text{ factorized}} \mathcal{J}\big(R_{\mathbf{X}}, \boldsymbol{\gamma}^{(n)}\big), \\ \boldsymbol{\gamma}^{(n+1)} = \arg\max_{\boldsymbol{\gamma}} \mathcal{J}\big(R_{\mathbf{X}}^{(n+1)}, \boldsymbol{\gamma}\big). \end{cases} \tag{4.1}$$

The next two sections are dedicated to each of these steps.

Initialization step. The optimization procedure (4.1) only ensures the convergence toward a local optimum, so the choice of the starting point for $\boldsymbol{\gamma}$ or $R_{\mathbf{X}}$ is crucial to avoid local optima. This choice is difficult, but, to our experience, hierarchical clustering seems to be a good strategy to get an initial value for $R_{\mathbf{X}}$.

4.2. *Optimal approximate conditional distribution $R_{\mathbf{X}}$.* We consider here the optimization of $\mathcal{J}$ with respect to $R_{\mathbf{X}}$. For a given value of $\boldsymbol{\gamma}$, we denote $\widehat{\boldsymbol{\tau}}$ the variational parameter defining the distribution $\widehat{R}_{\mathbf{X}} = \arg\max_{R_{\mathbf{X}} \text{ factorized}} \mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$. This amounts to maximimizing $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$, given in (3.6), under the condition that, for all i, the τ_{iq}'s must sum to 1. The derivative of $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$ with respect to τ_{iq} is

$$-\log \tau_{iq} - 1 + \log \alpha_q + \sum_{j \neq i} \sum_{\ell} \tau_{j\ell} [\log f_{q\ell}(X_{ij}) + \log f_{\ell q}(X_{ji})] + L_i,$$

where L_i denotes the ith Lagrange multiplier. It results from the previous equation that the optimal variational parameter $\widehat{\boldsymbol{\tau}}$ satisfies the fixed point relation

$$\widehat{\tau}_{iq} \propto \alpha_q \prod_{j \neq i} \prod_{\ell} [f_{q\ell}(X_{ij}) f_{\ell q}(X_{ji})]^{\widehat{\tau}_{j\ell}}. \tag{4.2}$$

The fixed point relation (4.2) can be related to a mean field approximation [see Jaakkola (2000)]. We get $\widehat{\boldsymbol{\tau}}$ simply by iterating this relation until convergence.

Undirected graphs. For a undirected graph, $\widehat{\boldsymbol{\tau}}$ satisfies

$$\widehat{\tau}_{iq} \propto \alpha_q \prod_{j \neq i} \prod_{\ell} [f_{q\ell}(X_{ij})]^{\widehat{\tau}_{j\ell}}.$$

4.3. *Parameter estimates*. We now have to maximize $\mathcal{J}$ with respect to $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\theta})$ for a given distribution $R_{\mathbf{X}}$. Again, this amounts to maximizing $\mathcal{J}(R_{\mathbf{X}}, \boldsymbol{\gamma})$, given in (3.6), under the condition that $\sum_q \alpha_q = 1$. Straightforward calculations show that the optimal $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ are given by

$$\widehat{\alpha}_q = \frac{1}{n}\sum_i \tau_{iq}, \qquad \widehat{\theta}_{q\ell} = \arg\max_\theta \sum_{i\neq j} \tau_{iq}\tau_{j\ell} \log f(X_{ij};\theta). \tag{4.3}$$

Poisson models. Poisson models are of particular interest for our interaction networks. The optimal $\lambda_{q\ell}$ for model PM presented in Section 2.2 is straightforward:

$$\widehat{\lambda}_{q\ell} = \sum_{i\neq j} \tau_{iq}\tau_{j\ell} X_{ij} \Big/ \sum_{i\neq j} \tau_{iq}\tau_{j\ell}.$$

For models PRMH and PRMI presented in the same section, there is no closed formula for $\lambda_{q\ell}$, $\beta_{q\ell}$ or β. However, since the Poisson regression model belongs to the exponential family, $\mathcal{J}$ is only a weighted version of the log-likelihoods of the corresponding generalized linear model. As such, standard optimization procedures can be used.

Exponential family. The optimal $\boldsymbol{\theta}$ is not explicit in the general case, but has a simpler form if the distribution f belongs to the exponential family. Namely, if f belongs to an exponential family with natural parameter θ,

$$f(x;\theta) = \exp\left[\boldsymbol{\Psi}(x)'\theta - A(\theta)\right].$$

According to (4.3), we look for $\widehat{\theta} = \arg\max_\theta \sum_{i\neq j} \tau_{iq}\tau_{j\ell}\boldsymbol{\Psi}(X_{ij})'\theta - A(\theta)$. Maximizing this quantity in θ yields

$$\sum_{i\neq j} \tau_{iq}\tau_{j\ell}\boldsymbol{\Psi}(X_{ij}) - \nabla A(\theta) = \mathbf{0}.$$

If ∇A is invertible, the optimal θ is

$$\widehat{\theta} = (\nabla A)^{-1}\left[\sum_{i\neq j} \tau_{iq}\tau_{j\ell}\boldsymbol{\Psi}(X_{ij})\right]. \tag{4.4}$$

4.4. *Choice of the number of groups*. In practice, the number of groups is unknown and should be estimated. Many criterion have been proposed to select the dimensionality Q of the latent space, ranging from AIC to ICL. AIC, BIC and their variants [Burnham and Anderson (1998)] are based on computing the likelihood of the observed data $\mathbb{P}(\mathbf{X}|m_Q)$ and penalizing it with some function of Q. But the use of variational EM is precisely to avoid computation of $\mathbb{P}(\mathbf{X}|m_Q)$, which is untractable. Given a prior distribution $\mathbb{P}(m_Q)$ over models, and a prior distribution $\mathbb{P}(\boldsymbol{\gamma}|m_Q)$ for each model, variational Bayes [Beal and Ghahramani

(2003)] works by selecting the model with maximum posterior $\mathbb{P}(m_Q|\mathbf{X})$. Estimation of $\mathbb{P}(\mathbf{X}|m_Q)$ is then performed using variational EM and no penalization is required, as complex models are already penalized by diffuse prior $\mathbb{P}(\boldsymbol{\gamma}|m_Q)$. Extension of Deviance Information Criterion (DIC) to finite mixture distributions via variational approximations [McGrory and Titterington (2007)] is even more straightforward: choosing Q^* larger than the expected number of components and running the algorithm, extraneous classes become void as the algorithm converges and the selected number of groups is just the number of nonempty classes. In the context of unknown assignments, Biernacki, Celeux and Govaert (2000) proposed the Integrated Classification Likelihood (ICL), which is an approximation to the complete data likelihood $\mathbb{P}(\mathbf{X}, \mathbf{Z}|m_Q)$. Variational Bayes, BIC and ICL can all be seen as approximations to Bayes factors. Whereas Variational Bayes integrates out the uncertainty about the parameter and the assignment of nodes to groups, ICL replaces them by a point estimate, computed thanks to variational EM. Traditional model selection essentially involves a trade-off between goodness of fit and model complexity, whereas ICL values both goodness of fit and classification sharpness.

Nowicki and Snijders (2001) do not propose any criterion to select the number of groups. Hofman and Wiggins (2008) use McGrory's method but in a very specific case of the Stochastic Block Model. They also give no clue as to how to decide that the algorithm has *converged enough*. Airoldi et al. (2008) use either a modification to BIC (for small size networks) or cross-validation (for large size networks) to select the number of groups. Daudin, Picard and Robin (2008) use a modification to ICL criterion. Following along the same line as Daudin, Picard and Robin (2008), we use a modification of ICL adapted to valued graphs to select the number of classes.

ICL criterion: For a model m_Q with Q classes where $\boldsymbol{\theta}$ involves P_Q independent parameters, the ICL criterion is

$$ICL(m_Q) = \max_{\boldsymbol{\gamma}} \log \mathbb{P}(\mathbf{X}, \widetilde{\mathbf{Z}}|\boldsymbol{\gamma}, m_Q) - \frac{1}{2}\{P_Q \log[n(n-1)] - (Q-1)\log(n)\},$$

where the missing data $\mathbf{Z}$ are replaced by their prediction $\widetilde{\mathbf{Z}}$.

Note that the penalty term $-\frac{1}{2}\{P_Q \log[n(n-1)] - (Q-1)\log(n)\}$ is similar to the one of BIC, where the log term refers to number of data. In the case of graphs, the number of data is n (i.e., the number of nodes) for the vector of proportions $\boldsymbol{\alpha}$ ($Q-1$ independent parameters), whereas it is $n(n-1)$ (i.e., the number of edges) for parameter $\boldsymbol{\theta}$ (P_Q independent parameters). For the models PM, PRMI and PRMH (detailed in Section 2.2), P_Q is respectively $Q(Q+1)/2$, $Q(Q+1)$ and $1+Q(Q+1)/2$.

5. Simulation study.

5.1. *Quality of the estimates*.

Simulation parameters. We considered undirected networks of size $n = 100$ and 500 with $Q = 3$ classes. To study balanced and unbalanced proportions, we set $\alpha_q \propto a^q$, with $a = 1, 0.5, 0.2$. $a = 1$ gives uniform proportions, while $a = 0.2$ gives very unbalanced proportions: $\boldsymbol{\alpha} = (80.6\%, 16.1\%, 3.3\%)$. We finally considered symmetric connection intensities λ_{pq}, setting $\lambda_{pp} = \lambda'$ for all p and $\lambda_{pq} = \lambda'\gamma$ for $p \neq q$. Parameter γ controls the difference between within class and between class connection intensities ($\gamma = 0.1, 0.5, 0.9, 1.5$), while λ' is set so that the mean connection intensity λ ($\lambda = 2, 5$) depends neither on γ nor a. γ close to one makes the distinction between the classes difficult. γ larger than one makes the within class connectivities less intense than the between ones. We expect the fitting to be rather easy for the combination $\{n = 500, a = 1, \lambda = 5, \gamma = 0.1\}$ and rather difficult for $\{n = 100, a = 0.2, \lambda = 2, \gamma = 0.9\}$.

Simulations and computations. For each combination of the parameters, we simulated $S = 100$ random graphs according to the corresponding mixture model. We fitted the parameters using the algorithm described in Section 4. To solve the identifiability problem of the classes, we systematically ordered them in descending estimated proportion order: $\widehat{\alpha}_1 \geq \widehat{\alpha}_2 \geq \widehat{\alpha}_3$. For each parameter, we calculated the estimated Root Mean Squared Error (RMSE):

$$RMSE(\widehat{\alpha}_p) = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\widehat{\alpha_p}^{(s)} - \alpha_p\right)^2},$$

$$RMSE(\widehat{\lambda}_{pq}) = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\widehat{\lambda_{pq}}^{(s)} - \lambda_{pq}\right)^2},$$

where the superscript (s) labels the estimates obtained in simulation s. We also calculated the mean posterior entropy

$$H = \frac{1}{S}\sum_{s}\left(-\sum_{i}\sum_{q}\tau_{iq}^{(s)}\ln\tau_{iq}^{(s)}\right),$$

which gives us the degree of uncertainty of the classification.

Results. Figure 1 (resp. 2) gives the RMSE for the proportion α_q (resp. connection intensities λ_{pq}). As expected, the *RMSE* is lower when n is larger. The parameters affecting the *RMSE* are mainly a and γ, whereas λ has nearly no effect. The departures observed for α_1 and α_3 in the balanced case ($a = 1.0$) are due to the systematic reordering of the proportions.

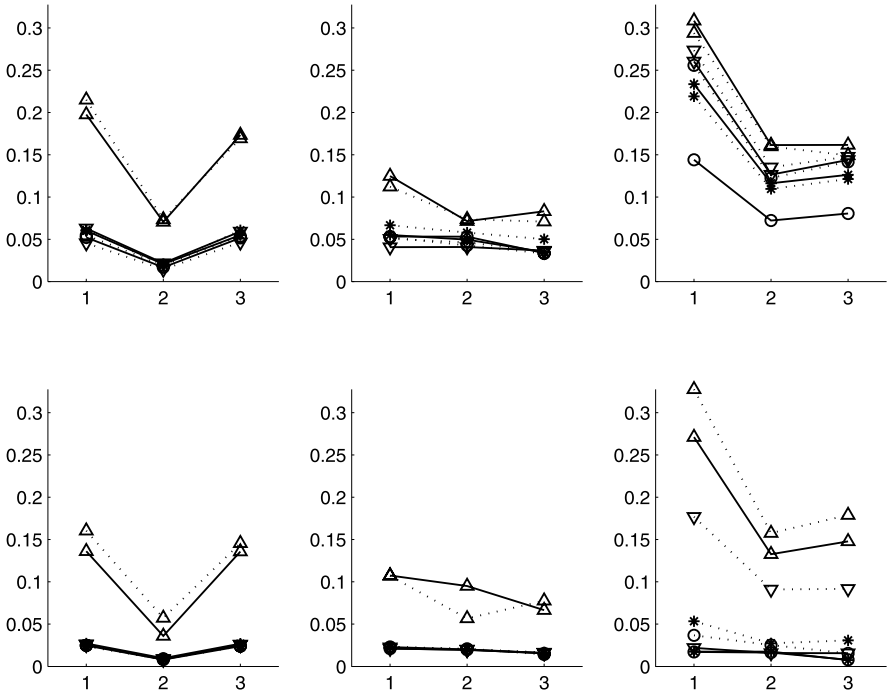


FIG. 1. *RMSE of the estimates* $\widehat{\alpha}_q$. *The* x*-axis refers to* $\alpha_1, \alpha_2, \alpha_3$. *Top*: $n = 100$, *bottom*: $n = 500$, *from left to right*: $a = 1, 0.5, 0.2$. *Solid line*: $\lambda = 5$, *dashed line*: $\lambda = 2$. *Symbols depend on* γ: $\circ = 0.1$, $\triangledown = 0.5$, $\triangle = 0.9$, $* = 1.5$.

Since the graph is undirected, $\lambda_{pq} = \lambda_{qp}$, so only nonredundant parameters are considered in Figure 2. The overall quality of the estimates is satisfying, especially for the diagonal terms λ_{qq}. The within intensity parameter of the smallest class λ_{33} is the most difficult to estimate. The worst case corresponds to a small graph ($n = 100$) with very unbalanced classes ($a = 0.2$) for parameter λ_{12}. In this case, the algorithm is unable to distinguish the two larger classes (1 and 2), so that the estimates extra-diagonal term $\widehat{\lambda}_{12}$ is close to the diagonal ones $\widehat{\lambda}_{11}$ and $\widehat{\lambda}_{22}$, whereas its true value is up to ten times smaller.

Figure 3 gives the mean entropy. Not surprisingly, the most influential parameter is γ: when γ is close to 1, the classes are almost indistinguishable. For small graphs ($n = 100$), the mean intensity λ has almost no effect. Because of the identifiability problem already mentioned, we did not consider the classification error rate.

5.2. *Model selection.* We considered a undirected graph of size $n = 50, 100, 500$ and 1000 with $Q^\star = 3$ classes. We considered the combination $\{a = 0.5, \lambda = 2, \gamma = 0.5\}$ which turned out to be a medium case (see Section 5.1) and com-

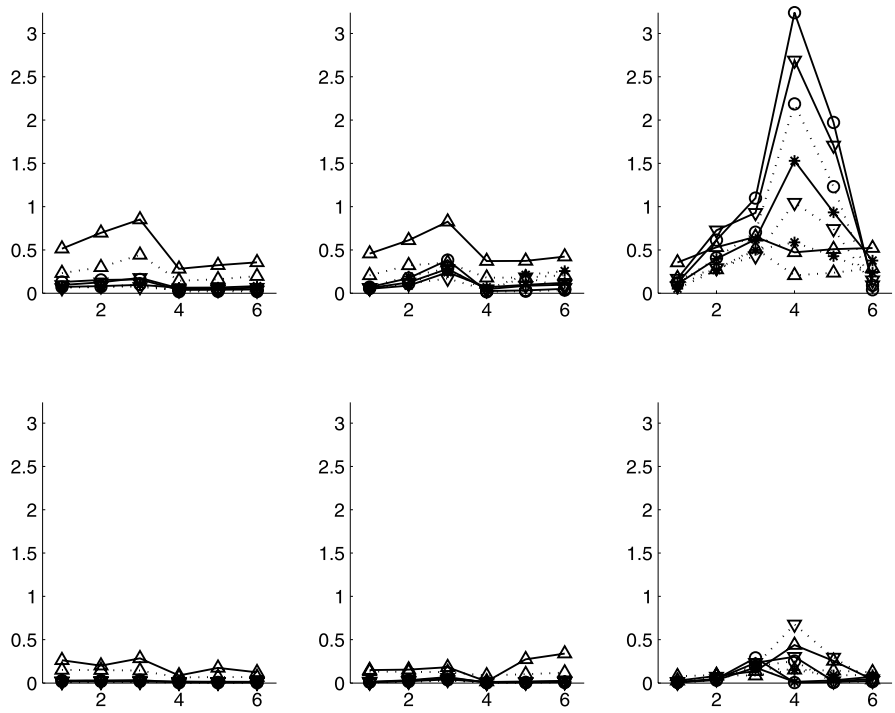


FIG. 2. *RMSE of the estimates* $\widehat{\lambda}_{pq}$. *The x-axis refers to* $\lambda_{11}, \lambda_{22}, \lambda_{33}, \lambda_{12}, \lambda_{13}, \lambda_{23}$. *Same legend as Figure* 1.

puted ICL for Q ranging from 1 to 10 (from 1 to 5 for $n = 1000$) before selecting the Q maximizing ICL. We repeated this for $S = 100$ simulations.

Figure 4 gives ICL as a function of Q, while Table 1 returns the frequency with which each Q is selected. As soon as n is larger than 100, ICL almost always selects the correct number of classes; for smaller graphs ($n = 50$), it tends to underestimate it. The proposed criterion is thus highly efficient.

6. Uncovering the structure of host–parasite interactions in forest ecosystems. Here we use mixture models to highlight the factors governing with whom a species interact in an ecosystem. The factors which may account for species interactions are introduced as covariates in the mixture models. The explanatory power of each factor is measured as the decrease in the number of groups selected. Our study focuses on host–parasite interactions in forest ecosystems. We address the two following questions: (1) Is similarity in the parasite assemblages of two tree species explained by their phylogenetic relatedness rather than by the degree of overlap of their distributional range? (2) Is similarity in the host range of two parasitic fungal species explained by their phylogenetic relatedness rather than their

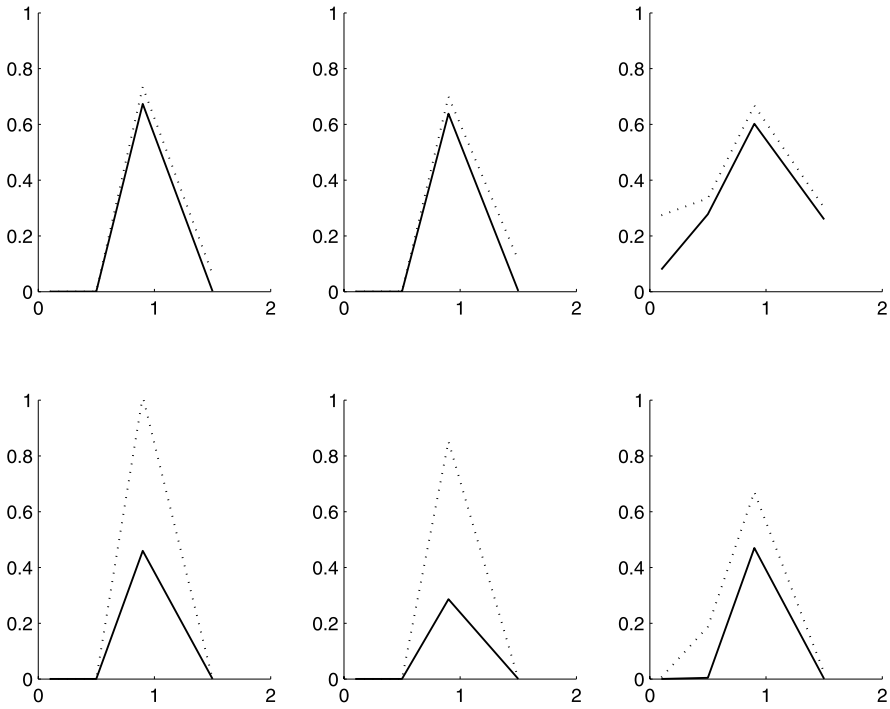


FIG. 3. *Mean (normalized) entropy H/n as a function of γ. Top*: $n = 100$, *bottom*: $n = 500$, *from left to right*: $a = 1, 0.5, 0.2$. *Solid line*: $\lambda = 5$, *dashed line*: $\lambda = 2$.

common nutritional strategy? The explanatory power of phylogenetic relatedness is subsequently called phylogenetic signal, as in the ecological literature [Rezende et al. (2007); Vacher, Piou and Desprez-Loustau (2008)].

6.1. *Data*.

Host–parasite interaction records. We considered two undirected, valued networks having parasitic fungal species ($n = 154$) and tree species ($n = 51$) as nodes, respectively. Edges strength was defined as the number of shared host species and the number of shared parasitic species, respectively [Mariadassou, Robin and Vacher (2010)].

The methods used for collecting data on tree–fungus interactions are fully described in Vacher, Piou and Desprez-Loustau (2008). Fungal species names were checked since then in the Index Fungorum database (www.indexfungorum.org): 17 names were updated, yielding to 3 new species synonymies. The fusion of synonym species accounts for the lower number of fungal species in the present study than in the original publication.

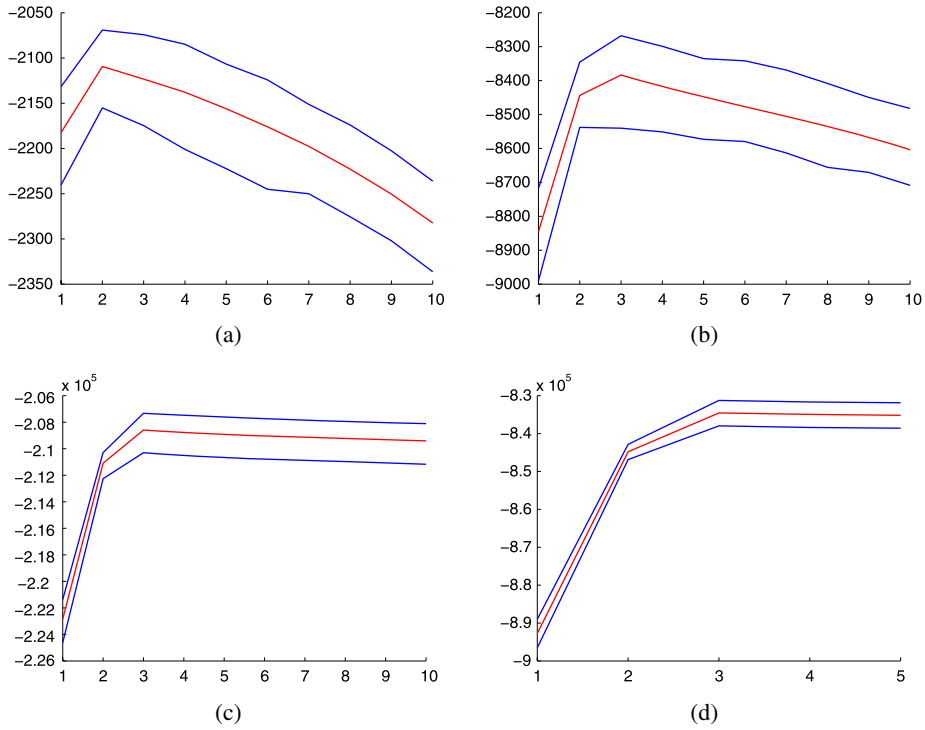


FIG. 4. *Mean ICL and 90% confidence interval as a function of Q.* (a) $n = 50$, (b) $n = 100$, (c) $n = 500$, (d) $n = 1000$.

Phylogenetic relatedness between species. In order to verify the existence of a phylogenetic signal in the parasite assemblages of tree species, we estimated genetic distances between all pairs of tree species. The maximally resolved seed plant tree of the software Phylomatic2 [Webb and Donoghue (2005)] was used to produce a phylogenetic tree for the 51 tree species included in our study. Then, pairwise genetic distances (in million years) were extracted by using the cophenetic.phylo function of the R ape package [Paradis, Claude and Strimmer

TABLE 1
Frequency (in %) at which Q is selected for various sizes n

	n			
Q	**50**	**100**	**500**	**1000**
2	82	7	0	0
3	17	90	100	100
4	1	3	0	0

(2004)]. Because the phylogenetic tree was loosely resolved for gymnosperms, we also used taxonomic distances to estimate phylogenetic relatedness between tree species. Since all tree species included in the study belong to the phylum Streptophyta, we used the finer taxonomic ranks of class, order, family and genus to calculate pairwise taxonomic distances. Based on the NCBI Taxonomy Browser (www.ncbi.nlm.nih.gov/Taxonomy/), we found that the species are evenly distributed into two taxonomic classes (Magnoliophyta and Conipherophyta) and further subdivided in 8 orders, 13 families and 26 genera. Following Poulin (2005), we considered that the taxonomic distance is equal to 0 if species are the same, 1 if they belong to the same genus, 2 to the same family, 3 to the same order, 4 to the same taxonomic class and 5 if their only common point lies in belonging to the phylum Streptophyta.

In order to investigate the existence of a phylogenetic signal in the host range of parasitic fungal species, we estimated taxonomic distances between all pairs of fungal species. Pairwise genetic distances could not be calculated because genetic data were not available for all the species. Since the 153 fungal species at hand span a wider portion of the tree of life than the tree species, we had to use the higher order rank of kingdom. The taxonomic distance for fungal species thus ranges from 0 to 6 (kingdom level) when compared to 0 to 5 for trees. The taxonomy was retrieved from Index Fungorum (www.indexfungorum.org). All fungal species included in the study belong to the Fungi kingdom, are divided in two phyla (Ascomycota and Basidiomycota) and further subdivided in 9 taxonomic classes, 21 orders, 48 families and 107 genera. When pairs included a species whose taxonomic is uncertain for a given taxonomic rank, this rank was skipped and upper ranks were used to estimate distance.

Other explanatory factors. Other factors than phylogenetic relatedness may account for pairwise similarities in parasite assemblages between tree species. In particular, two tree species having overlapping distributional range are exposed to similar pools of parasitic species and may therefore share more parasitic species than two tree species with nonoverlapping distributions [Brandle and Brandl (2006)]. We tested this hypothesis by calculating the geographical distance between all pairs of tree species. The geographical distance is the Jaccard distance [Jaccard (1901)] computed on the profiles of presence/absence in 309 geographical units covering the entire French territory.

In the case of fungal species, other factors may also account for similarity in host range. Here we investigated whether fungal species having similar nutritional strategies also have similar host ranges. Fungal species were classified into ten nutritional strategies based on their parasitic lifestyle (biotroph or necrotroph) and on the plant organs and tissues attacked. Five strategies (strict foliar necrotroph parasites, canker agents, stem decay fungi, obligate biotroph parasites and root decay fungi) accounted for 87% of the fungal species. We considered that nutritional distance between two species equals one if the strategies are the same and 0 otherwise.

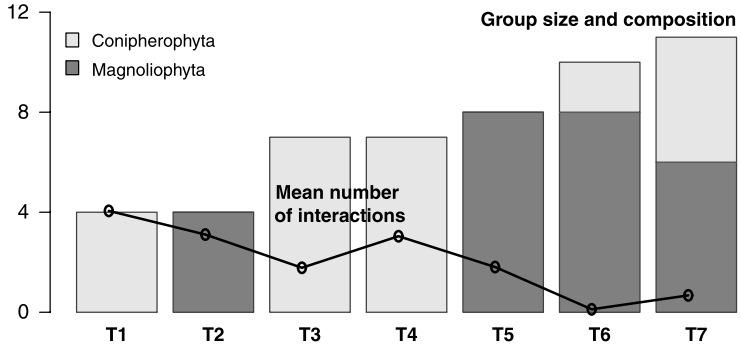
6.2. *Identification of groups of species sharing similar interactions.*

Model. For both networks, we used the mixture model to define groups of tree species and fungal species having similar interactions We assumed that, in each network, the edge intensities were Poisson distributed. For both networks, we considered the PM and PRMH models (see Section 2.2) using pairwise distance between species (genetic, taxonomic, geographic or nutritional) as a covariate.

PM model: *No covariate.* In the absence of covariates, the ICL criterion selected 7 groups of tree species. Two groups of tree species (T2 and T5) were exclusively composed of species belonging to the Magnoliophyta, whereas three other groups (T1, T3 and T4) were exclusively composed of species belonging to the Conipherophyta. The two last groups (T6 and T7) were mixed (Table 2). According to the mean number of interactions per species and the parameters estimates of the model (Table 2), they were composed of tree species having few parasitic species and sharing few of them with other tree species.

TABLE 2

Top: *Size, mean number of interactions and Magnoliophyta content for each group found with PM. Bottom*: *Parameter estimates for the tree network*: $\lambda_{q\ell}$ = *mean number of shared parasitic species*, α_q = *group proportion* (%) *with PM* (*no covariate*)



$\widehat{\lambda}_{q\ell}$	T1	T2	T3	T4	T5	T6	T7
T1	14.46	4.19	5.99	7.67	2.44	0.13	1.43
T2	4.19	14.13	0.68	2.79	4.84	0.53	1.54
T3	5.99	0.68	3.19	4.10	0.66	0.02	0.69
T4	7.67	2.79	4.10	7.42	2.57	0.04	1.05
T5	2.44	4.84	0.66	2.57	3.64	0.23	0.83
T6	0.13	0.53	0.02	0.04	0.23	0.04	0.06
T7	1.43	1.54	0.69	1.05	0.83	0.06	0.27
$\widehat{\alpha}_q$	7.8	7.8	13.7	13.7	15.7	19.6	21.6

It is noteworthy that group T2 was composed of four species belonging to the same order (Fagales) and also to the same family (Fagaceae). Groups T1, T3 and T4 were also composed of species belonging to the same family (Pinaceae) since the only three coniferous species belonging to another family were classified in groups T6 and T7. These results confirm that two plant species with a similar evolutionary history are likely to share the same set of parasitic species [Brandle and Brandl (2006), Gilbert and Webb (2007), Vacher, Piou and Desprez-Loustau (2008)].

PRMH model: *Accounting for phylogenetic relatedness*. When accounting for taxonomy, ICL selected only 4 groups of tree species. The estimated regression coefficient was $\widehat{\beta} = -0.317$, which means that, for the mean taxonomic distance $\overline{y} = 3.82$, the mean connexion intensity is reduced of 70% ($e^{\widehat{\beta}\overline{y}} = 0.298$). The cross classification table (Table 3) shows that the taxonomic distance reduces the number of class by merging groups T1 and T2 with most of the trees of T4 and T5. T'3 essentially consists of T6, T'1 of T7 and T'2 is made of trees from T3 completed with leftovers from other classes. Interestingly and unlike the groups obtained with no covariates, no group has species belonging exclusively to one or the other of the taxonomic classes (Magnoliophyta or Conipherophyta): the association between group of trees and taxonomy was cropped out by the covariate (Table 4). The same results hold when using the genetic distance as a covariate instead of the taxonomic distance (results not shown).

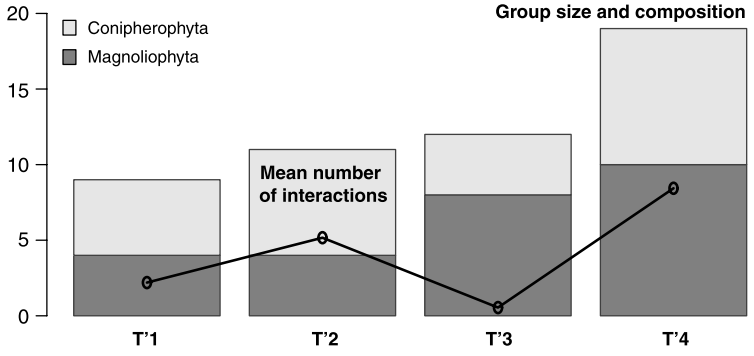
Therefore, the inclusion of taxonomic (or genetic) distance as a covariate shows that the phylogenetic relatedness between tree species accounts for a large part of the structure of tree–parasitic fungus interactions in forest ecosystems, but not for all the structure. Indeed, even after controlling for the evolutionary history through the taxonomic (or genetic) distance, ICL still finds 4 groups of trees, whereas we would expect only one group if the phylogeny was the sole source of structure. Below we investigate whether the distributional overlap between tree species is another source of structure.

TABLE 3
Cross classification of the groups of tree selected found by PM and PRMH (*with taxonomic variate as a covariate*)

	T'1	**T'2**	**T'3**	**T'4**
T1	0	0	0	4
T2	0	0	0	4
T3	2	5	0	0
T4	0	2	0	5
T5	0	2	0	6
T6	0	0	10	0
T7	7	2	2	0

TABLE 4

Top: *Size, mean number of interactions (scaled by* $\times 3$*) and Magnoliophyta content for each group found with PRMH. Bottom*: *Parameter estimates for the tree network*: $\lambda_{q\ell}$ = *mean number of shared parasitic species*, α_q = *group proportion* (%) *with PRMH* (*with covariate*), $\widehat{\beta}$ = *covariate regression coefficient*



$\widehat{\lambda}_{q\ell}$	T'1	T'2	T'3	T'4
T'1	0.75	2.46	0.40	3.77
T'2	2.46	4.30	0.52	8.77
T'3	0.40	0.52	0.080	1.05
T'4	3.77	8.77	1.05	14.22
$\widehat{\alpha}_q$	17.7	21.5	23.5	37.3
$\widehat{\beta}$		−0.317		

PRMH model: *Accounting for distributional overlap*. In contrast with the taxonomic and genetic distance, the geographical distance between species does not reduce the number of groups (not shown). This result suggests that the current distributional overlap between tree species does not account for the similarity in their parasite assemblages. This result is opposite to the conventional wisdom in the field of community ecology, which favors ecological processes, taking place over short time scale, over evolutionary processes, taking place over longer time scales, as the main source of biotic interaction diversity. Our findings point out that the relative importance of these processes might be the other way round.

6.3. *Factors accounting for the host ranges of parasitic fungal species*.

PM model: *No covariate*. The ICL criterion selected 9 groups of parasitic fungal species. The estimates intensities $\widehat{\lambda}_{q\ell}$ range from almost zero (1.4×10^{-3}) to 12.1, while the group proportions $\widehat{\alpha}_q$ range from 1.3% to 40.2% (Table 7).

PRMH model: *Accounting for phylogenetic relatedness*. Accounting for taxonomic distance does not reduce the number of groups (not shown), indicating a lack of phylogenetic signal in the host range of fungal species. These results parallel those obtained with another clustering approach [Newman (2004)] for the same tree–fungus network [Vacher, Piou and Desprez-Loustau (2008)]. They are congruent with the results obtained for other bipartite networks since asymmetries in the phylogenetic signal have been found in numerous plant–animal mutualistic networks [Rezende et al. (2007)] and in a host–parasite network between leaf-miner moths and parasitoid insects [Ives and Godfray (2006)]. In the latter case, the authors also observed a lack of signal through the parasite phylogeny. In the case of the tree–fungus network, we proposed that the very early divergence of the major fungal phyla may account for the asymmetric influence of past evolutionary history [Vacher, Piou and Desprez-Loustau (2008)]: the lack of signal through the fungal phylogeny may be the result of parasitic fungal species splitting into two groups when the Conipherophyta and the Magnoliophyta diverged (both groups containing Ascomycota and Basidiomycota species) and the subsequent coevolution of each set of fungal species with its plant phylum. Stronger selection pressures on parasitic species than on host species might also account for the asymmetry of the signal [Bersier and Kehrli (2008); Rossberg et al. (2006)].

PRMH model: *Accounting for nutritional strategies*. Fungal Correlation analysis showed an association between the 9 groups selected with the PM model and the nutritional type. In particular, two groups of fungal species (F2 and F3, see Appendix A.3) contained a high proportion of root decay fungi (100% and 75%, respectively). However, taking the nutritional strategy as a covariate does not reduce the number of groups, indicating the lack of 'nutritional signal' in the host range of parasitic fungal species.

6.4. *Goodness of fit*. Since no covariate decreases the number of mixture components in the fungus interaction network, we assessed goodness of fit only for the tree interaction network. The goodness is assessed in two ways: in terms of likelihood with the ICL criterion and in terms of predictive power for the strength of an interaction. The ICL criterion is -2876.6 for the base model with no class. It jumps to -1565.6 ($\Delta ICL = 1212.8$) when allowing a mixture structure (with 7 classes). It jumps again to -1449.6 ($\Delta ICL = 116$) when adding the taxonomic distance as a covariate in the model (with 4 classes). Interestingly, adding a covariate to the 4 class mixture model provides a gain in goodness of fit twice as big as the gain of adding three additional classes ($\Delta ICL = 214.2$ against 98.2). But adding a covariate only requires one additional parameter (β), against 21 for the three additional classes.

We also assessed goodness of fit in terms of predictive power. For the PRMH model with 4 classes and taxonomic distance as a covariate, we can predict both the weighted degree $K_i = \sum_{j \neq i} X_{ij}$ of node i as $\widehat{K}_i = \sum_j \sum_{q,\ell} \tau_{iq} \tau_{jl} \lambda_{ql} e^{\beta^\intercal \mathbf{y}_{ij}}$ and

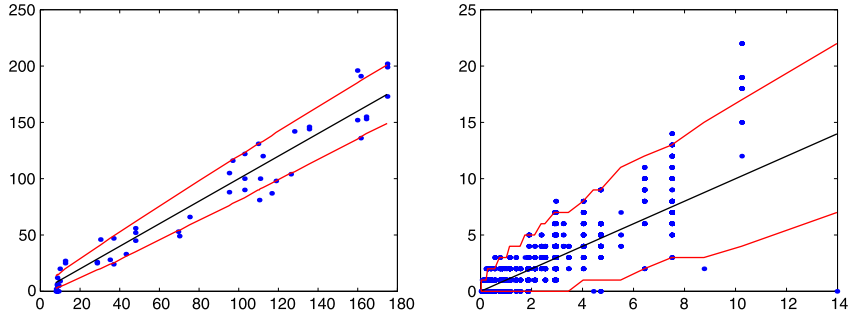


FIG. 5. *Left*: *Observed versus predicted graph of the weighted degree* K_i *of node* i ($R^2 = 0.94$). *Right*: *Observed versus predicted graph of single edge values* X_{ij} ($R^2 = 0.56$). *Black*: *regression line*; *red*: *Poisson* 95% *confidence interval*.

the value X_{ij} of a single edge fungal as $\widehat{X}_{ij} = \sum_{q,\ell} \tau_{iq}\tau_{j\ell}\lambda_{ql}e^{\beta^\intercal \mathbf{y}_{ij}}$. The prediction of K_i using $\widehat{K}_i$ is pretty accurate (Figure 5 left, $R^2 = 0.94$). The prediction of X_{ij} using $\widehat{X}_{ij}$ is less accurate, but the confidence region is still pretty good (Figure 5 right, $R^2 = 0.56$).

6.5. *Conclusion*. The structure of host–parasite interactions in forest ecosystems is a complex one. Some tree species share more parasites than others and this variability is well captured by a mixture model. However and as shown in Table 5, the naive mixture model deceptively captures part of the variability readily explained by other factors, such as the phylogenetic relatedness (measured either by taxonomic or genetic distance) and artificially increases the number of groups in the mixture. Accounting for relevant factors decreases the number of groups selected. Using group reduction as a yardstick (Table 5), we conclude that similarity in the parasite assemblages of tree species is explained by their phylogenetic relatedness rather than their distributional overlap, indicating the importance of evolutionary processes for explaining the current patterns of inter-specific interactions.

TABLE 5

Tree interaction network. Effect of different factors on the similarity in parasite assemblages between tree species. ΔICL *is the gain (in log-likelihood units) obtained when switching from the best PM model to the best PRMH model for a given covariate*

Factor	**Covariate**	**Nb. groups (PM)**	**Nb. groups (PRMH)**	ΔICL
Phylogenetic relatedness	Taxon. dist.	7	4	116.0
	Genetic dist.	7	4	94.8
Distributional overlap	Jaccard dist.	7	7	−8.6

Our study is however inconclusive on the relative contribution of phylogenetic relatedness and nutritional strategy to the similarity in the host ranges of parasitic fungal species parasites of two parasitic fungus (Table 8 in Appendix A.3). In either case, since the PRMH model still finds 4 (resp. 9) classes for the tree species (resp. fungal species) interaction network, a significant fraction of the variability remains unexplained by our predictors.

APPENDIX

A.1. Other mixture models. We examine here some other classical distributions which can be used in our framework.

Bernoulli. In some situations such as co-authorship or social networks, the only available information is the presence or absence of the edge. X_{ij} is then supposed to be Bernoulli distributed:

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{B}(\pi_{q\ell}).$$

It is equivalent to the stochastic block model of Nowicki and Snijders (2001) or Daudin, Picard and Robin (2008).

Multinomial. In a social network, X_{ij} may specify the nature of the relationship: colleague, family, friend, etc. The X_{ij}'s can then be modeled by multinomial variables:

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{M}(1; \mathbf{p}_{q\ell}).$$

The parameter $\theta_{q\ell}$ to estimate is the vector of probability $\mathbf{p}_{q\ell} = (p^1_{q\ell}, \ldots, p^m_{q\ell})$, m being the number of possible labels.

In directed random graphs, this setting allows to account for some dependency between symmetric edges X_{ij} and X_{ji}. We only need to consider the equivalent undirected graphs where edge (i, j) is labeled with the couple (X_{ij}, X_{ji}). $m = 4$ different labels can the be observed: $(0, 0)$ if no edge exists, $(1, 0)$ for $i \to j$, $(1, 1)$ for $i \leftarrow j$ and $(1, 1)$ for $i \leftrightarrow j$.

Gaussian. Traffic networks describe the intensity of the traffic between nodes. The airport network is a typical example where the edges are valued according to the number of passengers traveling from airport i to airport j. The intensity X_{ij} of the traffic can be assumed to be Gaussian:

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{N}(\mu_{q\ell}, \sigma^2_{q\ell}), \qquad \theta_{q\ell} = (\mu_{q\ell}, \sigma^2_{q\ell}).$$

Bivariate Gaussian. The correlation between symmetric edges X_{ij} and X_{ji} can be accounted for, considering the undirected valued graph where edge (i, j) is valued by (X_{ij}, X_{ji}), which is assumed to be Gaussian. Denoting $\mathbf{X}_{ij} =$

$[X_{ij}X_{ji}]'$,

$$\mathbf{X}_{ij}|i \in q, j \in \ell \sim \mathcal{N}(\boldsymbol{\mu}_{q\ell}, \boldsymbol{\Sigma}_{q\ell}), \qquad \theta_{q\ell} = (\boldsymbol{\mu}_{q\ell}, \boldsymbol{\Sigma}_{q\ell}).$$

Linear regression. When covariates are available, the linear model, either Gaussian for real valued edges or generalized for integer valued (e.g., Poisson or Bernoulli) allows to include them. For example, for Gaussian valued edges, denoting $\mathbf{y}_{ij}$ the $p \times 1$ vector of covariates describing edge (i, j), we set

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{N}(\beta_{q\ell}^{\mathsf{T}}.\mathbf{y}_{ij}, \sigma_{q\ell}^2).$$

Simple linear regression. A case of specific interest for plant ecology is the simple linear homoskedastic regression with group specific intercept $a_{q\ell}$ but constant regression coefficient b. It is particularly useful when controlling for the effect of geography, which is assumed to be the same for all groups of plants. We then set

$$X_{ij}|i \in q, j \in \ell \sim \mathcal{N}(a_{q\ell} + by_{ij}, \sigma^2).$$

The model can again be extended to Poisson or Bernoulli valued edges using adequate link function.

A.2. Parameter estimates for other distributions. Table 6 gives the parameter estimates for the model listed in Section A.1. The estimates of the mean parameter for Gaussian ($\mu_{q\ell}$) distributions are the same as the estimate of the probability $\pi_{q\ell}$ in the Bernoulli case. The results displayed in this table are all straightforward. Note that all estimates are weighted versions of the intuitive ones.

TABLE 6

Estimates of $\theta_{q\ell}$ for some classical distributions. Notation is defined in Section A.1. $\kappa_{q\ell}$ stands for $1/\sum_{i\neq j}\tau_{iq}\tau_{j\ell}$. $\mathbf{W}_{q\ell}$ is the diagonal matrix with diagonal term $\tau_{iq}\tau_{j\ell}$. # param. is the number of independent parameters in the case on directed graph, except for the bivariate Gaussian only defined for a nonoriented graph

Distribution	**Estimate**	**# param.**
Bernoulli	$\widehat{\pi}_{q\ell} = \kappa_{q\ell}\sum_{i\neq j}\tau_{iq}\tau_{j\ell}X_{ij}$	Q^2
Multinomial	$\widehat{p}^k_{q\ell} = \kappa_{q\ell}\sum_{i\neq j}\tau_{iq}\tau_{j\ell}\mathbb{I}(X_{ij}=k)$	$(m-1)Q^2$
Gaussian	$\widehat{\sigma}^2_{q\ell} = \kappa_{q\ell}\sum_{i\neq j}\tau_{iq}\tau_{j\ell}(X_{ij}-\widehat{\mu}_{q\ell})^2$	Q^2
Bivariate Gaussian	$\widehat{\boldsymbol{\mu}}_{q\ell} = \kappa_{q\ell}\sum_{i\neq j}\tau_{iq}\tau_{j\ell}\mathbf{X}_{ij}$	$Q(Q+1)$
	$\widehat{\boldsymbol{\Sigma}}_{q\ell} = \kappa_{q\ell}\sum_{i\neq j}\tau_{iq}\tau_{j\ell}(\mathbf{X}_{ij}-\widehat{\boldsymbol{\mu}}_{q\ell})(\mathbf{X}_{ij}-\widehat{\boldsymbol{\mu}}_{q\ell})'$	$\frac{3}{2}Q(Q+1)$
Linear regression	$\widehat{\boldsymbol{\beta}}_{q\ell} = (\mathbf{Y}'\mathbf{W}_{q\ell}^{-1}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{W}_{q\ell}^{-1}\mathbf{X}$	pQ^2
	$\widehat{\sigma}^2_{q\ell} = \kappa_{q\ell}\sum_{i\neq j}\tau_{iq}\tau_{j\ell}(X_{ij}-\mathbf{y}'_{ij}\widehat{\beta}_{q\ell})^2$	Q^2
Simple regression	$\widehat{b} = \frac{\sum_{i\neq j}\sum_{q,\ell}\tau_{iq}\tau_{j\ell}(X_{ij}-\overline{X}_{ql})(y_{ij}-\bar{y}_{ql})}{\sum_{i\neq j}\sum_{q,\ell}\tau_{iq}\tau_{j\ell}(y_{ij}-\overline{y}_{ql})^2}$	1
	$\widehat{\alpha}_{ql} = \overline{X}_{q\ell} - \widehat{b}\overline{y}_{ql}$	Q^2
	$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i\neq j}\sum_{q,\ell}\tau_{iq}\tau_{j\ell}(X_{ij}-\widehat{\alpha}_{q\ell}y_{ij})^2$	1

A.3. Parameter estimates for the fungus interaction network.

TABLE 7

Top: *Size*, *mean number of interactions* ($\bar{\lambda}$) *for each group found with PM*. *Bottom*: *Parameter estimates for the fungus network*: $\lambda_{q\ell}$ = *mean number of shared host species*, α_q = *group proportion* (%) *with PM* (*no covariate*). * *stand for* $\lambda_{q\ell}$ = *lower than* 5*e*–3

	F1	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**
Size	2	3	5	6	7	19	24	26	62
$\bar{\lambda}$	1.68	1.95	1.65	0.59	0.85	0.57	0.50	0.20	0.12
$\widehat{\lambda}_{q\ell}$									
F1	5.87	7.43	7.64	2.24	3.26	2.88	1.70	0.96	0.47
F2	7.43	9.88	7.29	3.59	4.45	1.54	2.77	1.03	0.71
F3	7.64	7.29	12.1	4.18	1.54	3.59	0.31	1.47	0.09
F4	2.24	3.59	4.18	2.92	0.50	0.47	0.05	0.81	0.03
F5	3.26	4.45	1.54	0.50	2.66	0.41	1.91	0.17	0.38
F6	2.88	1.54	3.59	0.47	0.41	2.35	*	0.32	*
F7	1.70	2.77	0.31	0.05	1.91	*	1.61	0.01	0.18
F8	0.96	1.03	1.47	0.81	0.17	0.38	0.01	0.25	*
F9	0.47	0.71	0.09	0.03	0.38	*	0.18	*	0.13
$\widehat{\alpha}_q$	1.3	2.0	3.3	3.9	4.6	12	16	17	40

TABLE 8

Fungus interaction network. *Effect of different factors on the similarity of host ranges between fungal species*. ΔICL *is the gain* (*in log-likelihood units*) *obtained when switching from the best PM model to the best PRMH model for a given covariate*

Factor	**Covariate**	**Nb. of groups (PM)**	**Nb. of groups (PRMH)**	**Δ*ICL***
Phylogenetic relatedness	Taxonomic distance	9	9	NA
Nutritional strategy	Trivial distance	9	9	NA

Acknowledgments. We thank the Département Santé des Forêts (DSF) of the French Ministère de l'Agriculture et de la Pêche for allowing us to use their database. We thank Dominique Piou and Marie-Laure Desprez-Loustau for checking the data and for helpful comments on the results.

SUPPLEMENTARY MATERIAL

Interaction network between tree and fungal species (DOI: 10.1214/07-AOAS361SUPP; .csv). This file contains:

- The adjacency matrix of interactions between tree and fungal species.
- The list of the tree species.
- The list of the fungal species.
- The matrix of genetic distances between tree species.
- The matrix of geographical distances between tree species.
- The matrix of taxonomic distances between fungal species.
- The matrix of nutritional type of the fungal species.

REFERENCES

AIROLDI, E. M. and CARLEY, K. M. (2005). Sampling algorithms for pure network topologies. *ACM KDD Explorations* **7** 13–22.

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

ATTIAS, H. (2000). A variational Bayesian framework for graphical models. In *Advances in Neural Information Processing Systems* **12** 209–215. MIT Press, Cambridge.

BARABÁSI, A. L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

BEAL, M. J. and GHAHRAMANI, Z. (2003). The variational Bayesian EM algorithm for incomplete data: With application to scoring graphical model structures. In *Bayesian Statistics 7* (J. M. Bernardo et al., eds.) 543–552. Oxford Univ. Press, Oxford. MR2003189

BERSIER, L. F. and KEHRLI, P. (2008). The signature of phylogenetic constraints on food-web structure. *Ecol. Complex.* **5** 132–139.

BIERNACKI, C., CELEUX, G. and GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Machine Intel.* **22** 719–725.

BLOMBERG, S. P. and GARLAND, T. J. (2002). Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *J. Evol. Biol.* **15** 899–910.

BRANDLE, M. and BRANDL, R. (2006). Is the composition of phytophagous insects and parasitic fungi among trees predictable? *Oikos* **113** 296–304.

BURNHAM, K. P. and ANDERSON, R. A. (1998). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Wiley, New York.

CATTIN, M. F., BERSIER, L. F., BANASEK-RICHTER, C. C., BALTENSPERGER, R. and GABRIEL, J. P. (2004). Phylogenetic constraints and adaptation explain food-web structure. *Nature* **427** 835–839.

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

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B* **39** 1–38. MR0501537

ERDÖS, P. and RÉNYI, A. (1959). On random graphs, i. *Publ. Math.* **6** 290–297. MR0120167

FIENBERG, S. E. and WASSERMAN, S. (1981). Categorical data analysis of single sociometric relations. In *Sociological Methodology 1981* 156–192. Jossey-Bass, San Francisco.

FIENBERG, S. E., MEYER, M. M. and WASSERMAN, S. S. (1985). Statistical analysis of multiple sociometric relations. *J. Amer. Statist. Assoc.* **80** 51–67.

GETOOR, L. and DIEHL, C. P. (2004). Link mining: A survey. *SIGKDD Explor.* **7** 3–12.

GILBERT, G. S. and WEBB, C. O. (2007). Phylogenetic signal in plant pathogen-host range. *Proc. Natl. Acad. Sci. USA* **104** 4979–4983.

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

GOVAERT, G. and NADIF, M. (2005). An EM algorithm for the block mixture model. *IEEE Trans. Pattern Anal. Machine Intel.* **27** 643–647.

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

HOLLAND, P. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–50. MR0608176

IVES, A. R. and GODFRAY, H. C. J. (2006). Phylogenetic analysis of trophic associations. *Am. Nat.* **16** E1–E14.

JAAKKOLA, T. (2000). Tutorial on variational approximation methods. In *Advanced Mean Field Methods: Theory and Practice*. MIT Press, Cambridge.

JACCARD, P. (1901). Tude comparative de la distribution florale dans une portion des alpes et des jura. *Bullet. Soc. Vaud. Sci. Natur.* **37** 547–579.

JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

KEMP, C., GRIFFITHS, T. H. and TENENBAUM, J. B. (2004). Discovering latent classes in relational data. Technical report, MIT Computer Science and Artificial Intelligence Laboratory.

LEISINK, M. A. R. and KAPPEN, H. J. (2001). A tighter bound for graphical models. *Neural Comput.* **13** 2149–2171.

LORRAIN, F. and WHITE, H. C. (1971). Structural equivalence of individuals in social networks. *J. Math. Soc.* **1** 49–80.

MARIADASSOU, M. (2006). Estimation paramétrique dans le modèle ERMG. Master's thesis, Univ. Paris XI/Ecole Nationale Supèrieure.

MARIADASSOU, M., ROBIN, S. and VACHER, C. (2010). Supplement to "Uncovering latent structure in valued graphs: A variational approach." DOI: 10.1214/07-AOAS361SUPP.

MCGRORY, C. A. and TITTERINGTON, D. M. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Comput. Statist. Data Anal.* **51** 5352–5367. MR2370876

MCLAHAN, G. and PEEL, D. (2000). *Finite Mixture Models*. Wiley, New York. MR1789474

NEWMAN, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **69** 066133.

NEWMAN, M. E. J., WATTS, D. J. and STROGATZ, S. H. (2002). Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* **99** 2566–2572.

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

PARADIS, E., CLAUDE, J. and STRIMMER, K. (2004). Ape: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20** 289–290.

PATTISON, P. E. and ROBINS, G. L. (2007). Probabilistic network theory. In *Handbook of Probability Theory with Applications*. Sage, Thousand Oaks, CA.

PICARD, F., DAUDIN, J.-J., MIELE, V., MARIADASSOU, M. and ROBIN, S. (2007). A novel framework for random graph models with heterogeneous connectivity structure. Submitted.

POULIN, R. (2005). Relative infection levels and taxonomic distances among the host species used by a parasite: Insights into parasite specialization. *Parasitology* **130** 109–115.

REZENDE, E. L., LAVABRE, J. E., GUIMARAES, P. R., JR., JORDANO, P. and BASCOMPTE, J. (2007). Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448** 925–928.

RICKLEFS, R. E. and MILLER, G. L. (2000). Community ecology. In *Ecology*, 4th ed. Freeman, San Francisco, CA.

ROSSBERG, A. G., ISHII, R., AMEMIYA, T. and ITOH, K. (2006). Food webs: Experts consuming families of experts. *J. Theoret. Biol.* **241** 552–563. MR2254907

TYKIAKANIS, J. M., TSCHARNTKE, T. and LEWIS, O. T. (2007). Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature* **51** 202–205.

VACHER, C., PIOU, D. and DESPREZ-LOUSTAU, M.-L. (2008). Architecture of an antagonistic tree/fungus network: The asymmetric influence of past evolutionary history. *PLoS ONE* **3** 1740.

VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. MR2396807

WEBB, C. O. and DONOGHUE, M. J. (2005). Phylomatic: Tree assembly for applied phylogenetics. *Mol. Ecol. Notes* **5** 181–183.

WINN, J., BISHOP, C. M. and JAAKKOLA, T. (2005). Variational message passing. *J. Mach. Learn. Res.* **6** 661–694. MR2249835

XING, E., JORDAN, M. and RUSSELL, S. (2003). A generalized mean field algorithm for variational inference in exponential families. In *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)* 583–591. Morgan Kaufmann, San Francisco, CA.

YEDIDIA, J. S., FREEMAN, W. T. and WEISS, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Inform. Theory* **15** 2282–2312. MR2246363

M. MARIADASSOU
S. ROBIN
UMR 518 AGROPARISTECH/INRA MIA
16, RUE C. BERNARD, F-75005 PARIS
FRANCE
E-MAIL: mariadas@agroparistech.fr
robin@agroparistech.fr

C. VACHER
UMR 1202 UNIV. BORDEAUX I/INRA BIOGECO
69, ROUTE D'ARCACHON, F-33612 CESTAS
FRANCE
E-MAIL: corinne.vacher@pierroton.inra.fr