Noname manuscript No.
(will be inserted by the editor)

Variational Bayes Model Averaging for Graphon Functions and Motif Frequencies Inference in W-graph Models*

Pierre Latouche · Stéphane Robin

Received: date / Accepted: date

Abstract W-graph refers to a general class of random graph models that can be seen as a random graph limit. It is characterized by both its graphon function and its motif frequencies. In this paper, relying on an existing variational Bayes algorithm for the stochastic block models along with the corresponding weights for model averaging, we derive an estimate of the graphon function as an average of stochastic block models with increasing number of blocks. In the same framework, we derive the variational posterior frequency of any motif. A simulation study and an illustration on a social network complete our work.

Keywords Bayesian model averaging · graphon · network · network motif · stochastic block model · W-graph

Mathematics Subject Classification (2000) 62F15 · 62G05

1 Introduction

W-graph. The W-graph model has been intensively studied in the probability literature. From a theoretical point of view, it defines a limit for dense graphs (Lovász and Szegedy , 2006), but it can also be casted into a general class of inhomogeneous random graph models (Bollobás *et al.* , 2007) involving some hidden latent space. A W-graph is characterized by the so-called 'graphon' function W, where $W(u, v)$ is the probability for two nodes with respective

* The final publication is available at Springer via http://dx.doi.org/[10.1007/s11222-015-9607-0]

P. Latouche
Laboratoire SAMM, EA 4543, Université Paris 1 Panthéon-Sorbonne, France
E-mail: pierre.latouche@univ-paris1.fr

S. Robin
AgroParisTech, UMR 518 MIA, Paris, France
INRA, UMR 518 MIA, Paris, France

latent coordinates u and v (both taken in $[0,1]$) to be connected. The precise definition of a W-graph model is given at the end of this section. Because of very weak assumptions about the graphon function, the W-graph model is very flexible and can result in a large variety of network topologies.

The W-graph model suffers an identifiability issue as, for any measure-preserving transformation σ of $[0,1]$ into $[0,1]$, the graphon function $W_\sigma(u,v) = W(\sigma(u),\sigma(v))$ results in the same W-graph model as with the function W. This issue is often circumvented by further assuming that the mean density $\int W(u,v)\mathrm{d}v$ is an increasing function of u (Bickel and Chen , 2009). However, Diaconis and Janson (2008) showed that subgraphs (called hereafter motifs) frequencies are invariant and constitute intrinsic characteristics of a W-graph.

Interpreting of the graphon. The graphon function provides a two-dimensional representation of the global topology of the network, without any prior assumption as for the form of the degree distribution, or the existence of clusters in the graph. It is the limiting adjacency matrix of the network. For more details, we refer to the work of Lovász and his coauthors (see for instance Lovász and Szegedy (2006)). We emphasize that the connection between the Aldous-Hoover theorem, which is an extension of deFinetti's theorem to exchangeable arrays, and the notion of graph limits, was made by Diaconis and Janson (2008). As shown in Section 5, the graphon function can help in understanding the organization of the network and offers an alternative visualization that is especially useful for large graphs.
Figure 1 provides some examples of graphon functions. A scale free network (Barabási and Albert , 1999) is highly concentrated around a small fraction of nodes with high degree. Such a degree distribution can be retrieved using a graphon similar to this used in the simulation study (see Section 4). The concentration around the central nodes is revealed by the peak in the upper right corner of graphon surface. A community network (Girvan and Newman , 2002), where nodes tend to connect to nodes of the same community, is characterized by a block-diagonal structure. A small-world network (Watts and Strogatz , 1998; Barbour and Reinert , 2006) is defined by a majority of connexions between neighboring nodes revealed by high values of the graphon function along the diagonal. Edges between non-neighbor nodes, which provide the 'small world' property, are made possible by the non-zero value of the graphon function apart from the diagonal. Thus the graphon function summarizes the global topology of the network.
On the other hand, the characterization in terms of motifs provides an information about the local organization of the network. Such a characterization has been used to depict the organization and the functioning of biological networks (Milo *et al.* , 2002). Because the topology of a W-graph only depends on the respective latent location of pairs of nodes, the empirical frequency of motifs of size larger that three can be used to assess the goodness-of-fit of the model.

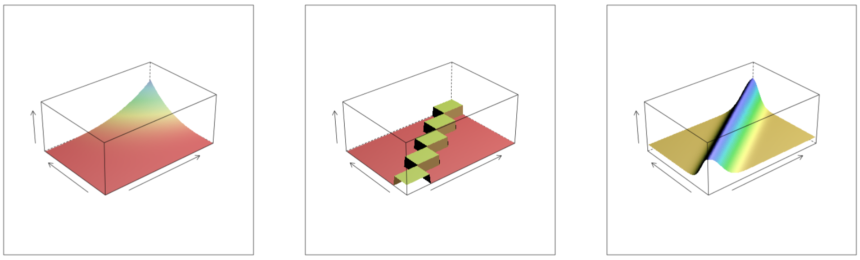


Fig. 1 Graphon function for some typical random graph models. Left: scale-free network. Middle: community network. Right: small world network.

Statistical inference. Until recently, little attention has been paid to the statistical inference of a W-graph model, based on an observed network. The earliest reference on graphon estimation is Kallenberg (1999) who showed the weak convergence of a function, which can be seen as an empirical graphon, to the graphon function. A general framework was considered without further modeling assumptions. Contrary to this work, Palla *et al.* (2010) derived a parametric exchangeable model along with an MCMC-like algorithm to perform inference on real data, resulting in a heavy computational burden. An alternative parametric approach was proposed earlier by Hoff (2008). The graphon estimation as a nonparametric problem was first formulated in Lloyd *et al.* (2012). The authors relied on Gaussian process priors to build the nonparametric scheme and also on an MCMC-like algorithm for the inference. Since then, many approaches have been proposed, in parallel to the present work (Airoldi *et al.* , 2013; Wolfe and Olhede , 2013; Asta and Shalizi , 2014; Chatterjee , 2015; Borgs *et al.* , 2015). We emphasize that a series of methods has considered total variation estimation (Chan and Airoldi , 2014) and the corresponding two or three step inference procedures (Yang *et al.* , 2014).
Wolfe and Olhede (2013) developed a theoretical framework for the nonparametric estimation of the graphon function. This latter approach also relies on the connection between the SBM and W-graph models. However, while they focus on the blockwise constant approximation of the graphon function, we rely in this paper on approximate posterior distributions of the SBM model, which results in a smoother estimate of the graphon function.

Link with SBM. In parallel to W-graph, the stochastic block model (SBM: Nowicki and Snijders (2001)) has been used in a large variety of domains, from sociology to biology, and many efforts have been made in view of its inference. The proposed inference techniques range from MCMC (Nowicki and Snijders , 2001) to a degree based algorithm (Channarond *et al.* , 2012), including variational expectation maximization (VEM) (Daudin *et al.* , 2008) and variational Bayes EM (VBEM) (Latouche *et al.* , 2012). SBM states that each node belongs to a certain class (in finite number) and that the probability for two nodes to be connected depends on the class they belong to. As shown in the next section, SBM corresponds to a W-graph for which the graphon function is blockwise constant. Interestingly, the frequency of motifs in SBM has been studied by Picard *et al.* (2008), who provided explicit formulas.

Contribution. Our purpose in this paper is to rely on some of the statistical works developed for the SBM model in order to carry out the inference of the W-graph model. Considering SBM as a proxy of the W-graph model, we propose a complete inference procedure for both the graphon function and the frequency of any motif. The method is based on the variational Bayes approach proposed by Latouche *et al.* (2012).
The SBM postulates that nodes can belong to a certain number Q of classes, which does not make sense for a W-graph. We show how the posterior distributions conditional to the number of clusters can be integrated out in order to provide an estimation of the posterior distribution of the graphon function. In practice, this integration leads to a smooth version of the blockwise constant approximation. Indeed, this property is desirable when the monotone version of the graphon is expected to be smooth.
In the same spirit, we provide a variational Bayes estimate of the frequency of any network motif (or sub-graph). These frequencies allow us to search for unexpectedly frequent motifs in the network and we suggest to use these results to assess the goodness-of-fit of the model.

Outline. The paper is organized as follows. In Section 2, we present the connexion between SBM and W-graph, we remind the principle of the variational Bayes method and derive the approximate posterior distribution of the graphon function. We follow the same line in Section 3 to derive the approximate posterior mean of motifs frequencies. The performances of the approach are studied via simulation in Section 4 and the proposed method is applied to a subset of the French political blogosphere network, in Section 5.

Notations and definition of the W random graph. All along the paper, we will use the following notations for the W-graph model. We consider n nodes labeled with index $i = 1, \dots n$. A latent variable U_i drawn uniformly over $[0, 1]$ is associated with each node i, the U_i's being mutually independent. The edges $\{X_{ij}\}_{i<j}$ are then drawn independently conditionally on the latent $\{U_i\}$ as $X_{ij}|U_i, U_j \sim \mathcal{B}[W(U_i, U_j)]$, where $W : [0, 1] \times [0, 1] \to [0, 1]$ denotes the graphon function and $\mathcal{B}(\cdot)$ is the Bernoulli distribution.

2 Inference of the graphon function

We propose to estimate the function W via the inference of a stochastic block model. The aim of this section is to recall previous results on the variational Bayes inference of SBM and to show how they can be used to estimate W.

Stochastic block model (SBM). We first recall the definition of the SBM model (Nowicki and Snijders , 2001). The n nodes are supposed to be spread into Q groups with proportions $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$. More precisely, the nodes are associated with independent (unobserved) labels Z_i drawn from a multinomial distribution $\mathcal{M}(1; \boldsymbol{\alpha})$. Connections are ruled by a $Q \times Q$ connectivity matrix $\boldsymbol{\pi} = [\pi_{q\ell}]$, where $\pi_{q\ell}$ is the connection probability between a node from group q and a node from group ℓ ($\boldsymbol{\pi}$ has to be symmetric for undirected graphs). The edges of the graph are then drawn independently from a Bernoulli distribution, conditionally on the labels Z_i, as $X_{ij}|Z_i, Z_j \sim \mathcal{B}(\pi_{Z_i,Z_j})$.
In the sequel, we shall denote $\mathbf{Z} = \{Z_i\}$ the set of unobserved labels, $\mathbf{X} = \{X_{ij}\}$ the set of observed edges and $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\pi})$ the set of model parameters.

Connection between SBM and W-graph. SBM corresponds to the case where W is blockwise constant, with rectangular blocks of size $\alpha_k \times \alpha_\ell$ and height $\pi_{q\ell}$. More precisely, denoting the cumulative proportion

$$\sigma_q = \sum_{j=1}^{q} \alpha_j, \tag{1}$$

if we define the binning function

$$C_{\boldsymbol{\alpha}}(u) = 1 + \sum_{q=1}^{Q} \mathbb{I}\{\sigma_q \leq u\},$$

and if we take

$$W(u, v) = \pi_{C(u),C(v)}, \tag{2}$$

the resulting W-graph model corresponds to the SBM model with parameters $(\boldsymbol{\alpha}, \boldsymbol{\pi})$.

Palla *et al.* (2010) also considered a blockwise constant model; these authors used a self similarity transformation to increase the number of blocks to gain flexibility keeping the number of parameters small. However, the connection with W-graphs is not made explicitly in this article.

Identifiability. As pointed out by Bickel and Chen (2009), the W-graph is not identifiable, as any measure-preserving transformation of the interval $[0, 1]$ would provide the same random graph. Following these authors, we fix the version of W to be estimated using the constraint that the function $D(u) = \int W(u, v)\mathrm{d}v$ is monotonic increasing. For consistency, we require the corresponding condition for the SBM to be fitted, that is: $d_q = \sum_\ell \alpha_\ell \pi_{q\ell}$ increases with q.

2.1 Variational Bayes inference of SBM

The inference of SBM has received many attention in the last decade. Briefly speaking the main pitfalls lies in the determination of the conditional distribution of the labels in $\mathbf{Z}$, given the observation $\mathbf{X}$, which displays an intricate dependency structure. Both Monte-Carlo sampling (Nowicki and Snijders , 2001) and variational approximations (Daudin *et al.* , 2008) have been proposed, but the later scale better. In this paper, we will use the variational Bayes approximation proposed in Latouche *et al.* (2012), which provides a closed-form approximate posterior distribution of the parameters $\boldsymbol{\theta}$ and of the hidden variables in $\mathbf{Z}$ (denoted $\widetilde{p}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ and $\widetilde{p}_{\mathbf{Z}}(\mathbf{Z})$). We recall that this approximation is obtained through the maximization with respect to $\widetilde{p}_{\boldsymbol{\theta}}(\cdot)$ and $\widetilde{p}_{\mathbf{Z}}(\cdot)$ of the functional

$$\mathcal{J} = \log P(\mathbf{X}) - KL(\widetilde{p}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\widetilde{p}_{\mathbf{Z}}(\mathbf{Z})||P(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{X})), \tag{3}$$

where KL stands for the Küllback-Leibler divergence. Our estimate of the function W strongly relies on this approximate posterior distribution, which has been shown to be reliable by Gazal *et al.* (2012). In the sequel, we shall use a tilde to mark approximate posterior variational distributions and probabilities.

The variational Bayes inference of SBM can be achieved using the VBEM algorithm described in Beal and Ghahramani (2003). As SBM can be casted into the exponential family framework, using conjugate priors for the parameters

$$\begin{aligned} \boldsymbol{\alpha} &\sim \mathrm{Dir}(\mathbf{a}^0) \qquad \text{where } \mathbf{a}^0 = (a_1^0, \ldots, a_Q^0), \\ \pi_{q,\ell} &\sim \mathrm{Beta}(\eta_{q,\ell}^0, \zeta_{q,\ell}^0), \end{aligned}$$

(where Dir stands for the Dirichlet distribution), the variational Bayes posterior approximation states that $\boldsymbol{\alpha}$ and the $\pi_{q\ell}$ are all conditionally independent given $\mathbf{X}$ with distributions

$$\begin{aligned} \boldsymbol{\alpha}|\mathbf{X} &\sim \mathrm{Dir}(\mathbf{a}) \qquad \text{where } \mathbf{a} = (a_1, \ldots, a_Q), \\ \pi_{q,\ell}|\mathbf{X} &\sim \mathrm{Beta}(\eta_{q,\ell}, \zeta_{q,\ell}). \end{aligned} \tag{4}$$

The expressions of the a_q, $\eta_{q\ell}$ and $\zeta_{q\ell}$ as functions of a_q^0, $\eta_{q\ell}^0$, $\zeta_{q\ell}^0$ and $\mathbf{X}$ can be found in Latouche *et al.* (2012).

No general guaranty exists about the theoretical properties of variational (Bayes) estimates. Still, SBM appears to be a special case were the consistency of the variational estimates has been shown in a frequentist setting (Celisse *et al.* (2012), Mariadassou *et al.* (2010)). As for the variational Bayes estimates, the simulation study carried out by Gazal *et al.* (2012) shows that the approximate posterior distribution is accurate even for networks with only few tens of nodes.

2.2 Posterior distribution of the function W

We now derive the approximate posterior distribution of the function W, at given coordinate (u, v). To this aim, (2) has to be integrated with respect to the (approximate) posterior distributions of both $\boldsymbol{\pi}$ and $\boldsymbol{\alpha}$.

Proposition 1 *For given $(u, v) \in [0, 1]^2$, $u \leq v$, using a SBM with Q groups, the variational Bayes approximate pdf of $W(u, v)$ is $\widetilde{p}(w(u, v)|\mathbf{X}, Q)$ can be computed exactly with complexity $O(Q^2)$.*

The key point is that the cumulative distribution function (cdf) of a Dirichlet distribution can be calculated via simple recursions given in Gouda and Szántai (2010). The rest of the proof relies on standard algebraic manipulations and is postponed to Appendix A.1.

The approximate posterior mean comes as a direct by-product of Proposition 1: $\widetilde{\mathbb{E}}[W(u, v)|\mathbf{X}] =$

$$\sum_{q \leq \ell} \frac{\eta_{q,\ell}}{\eta_{q,\ell} + \zeta_{q,\ell}} \left[F_{q-1,\ell-1}(u, v; \mathbf{a}) - F_{q,\ell-1}(u, v; \mathbf{a}) - F_{q-1,\ell}(u, v; \mathbf{a}) + F_{q,\ell}(u, v; \mathbf{a})\right],$$

where $F_{q,\ell}(u, v; \mathbf{a})$ denotes the joint cdf of (σ_q, σ_ℓ), as defined in (1), when $\boldsymbol{\alpha}$ has a Dirichlet distribution $\mathrm{Dir}(\mathbf{a})$. The approximate posterior standard deviation can be computed as well.

Variable number of groups. Denoting $\widehat{\mathcal{J}}_Q$ the maximum of the function defined in (3), Latouche *et al.* (2012) derived a close-form expression of $\widehat{\mathcal{J}}_Q$ and showed that it can be used as a model selection criterion, choosing $\widehat{Q} = \arg\max_Q \widehat{\mathcal{J}}_Q$. Thus, in straightforward scenarios where the (hidden) W-graph model could be casted as a unique SBM, *i.e.* graphon function is exactly blockwise constant, this framework provides a way to estimate the number of blocks as $\widehat{Q}^2$. Still, because SBM is mostly used as a proxy for W-graph, it may seem more realistic to assume that no true number of groups Q does actually exist. Therefore we rather consider here a model averaging approach in which the inferred W-graph is an average of a series of SBM with increasing number of groups Q. Volant *et al.* (2012) derived the variational Bayes approximation of $p(Q|\mathbf{X})$ and prove that, if a uniform prior over Q is used, the variational approximation satisfies

$$\widetilde{p}(Q|\mathbf{X}) \propto \exp \widehat{\mathcal{J}}_Q, \qquad \sum_Q \widetilde{p}(Q|\mathbf{X}) = 1.$$

In this case, the variational Bayes approximate posterior distribution of $W(u, v)$, integrated over the number of groups, is simply

$$\widetilde{p}(w|\mathbf{X}) = \sum_Q \widetilde{p}(Q|\mathbf{X})\, \widetilde{p}(w|\mathbf{X}, Q).$$

We remind that $\widehat{\mathcal{J}}_Q$ is the difference between the true marginal likelihood of the data with Q groups, $\log P(X|Q)$ and the KL divergence between the variational approximation of the condition distribution $P(\boldsymbol{\theta}, \mathbf{Z}|\mathbf{X}, Q)$ and this distribution itself. The regular Bayesian model averaging would directly rely on $P(X|Q)$ to weight each considered model (Hoeting *et al.* (1999)). Because of the accuracy of the variational Bayes approach for SBM (Gazal *et al.* (2012), the KL divergence is expected to be small, so the variational weights are close to the theoretical ones.

3 Motif probability

As recalled above, the W-graph model suffers a deep identifiability problem. However, as shown in Diaconis and Janson (2008), the distribution of the number of occurrences of patterns or motifs turns out to be invariant, and therefore characteristic of a given W-graph model. In this section, we show how variational Bayes inference of SBM can be used to estimate a key quantity of such a distribution, namely the occurrence probability of the motif.

The number of occurrences of a given motif in random graphs has been intensively studied in Erdös-Rényi graphs (see for instance Stark , 2001) and some results about W-graphs can be found in Diaconis and Janson (2008) and Bollobás *et al.* (2007). However, the exact distribution for an arbitrary motif can not be determined in general. On the other hand, Picard *et al.* (2008) derived a general approach to derive the moments of the number of occurrences in stationary graphs. These moments only depend on the size of the graph, on the number of automorphisms of the motif and on the occurrence probability of the motif (and of its super-motifs).

The occurrence probability of a motif is therefore a key quantity to characterize the number of occurrences of a motif in a random graph. In this section, we recall the definition of a motif occurrence and of the occurrence probability. Then, we show how variational Bayes inference of SBM can be used to estimate this occurrence probability in a W-graph.

3.1 Motif probability

Definition of a motif. A motif can be defined as a sub-graph with prescribed edges. More precisely, a motif with size k is completely defined by the $k \times k$ 0-1 adjacency matrix $\mathbf{m}$, where $m_{a,b} = 1$ if there is an edge between node a and b, 0 otherwise. Figure 8, given in A.2, displays some typical motifs and their corresponding adjacency matrices $\mathbf{m}$.

As for the occurrence of a motif, we use here the definition used in Picard *et al.* (2008), which defines an occurrence of $\mathbf{m}$ as a set of k nodes in the graph, such that all edges prescribed in $\mathbf{m}$ actually occur. Formally, we consider the occurrence indicator of the motif $\mathbf{m}$ at position $\beta = (i_1, \dots, i_k)$, with $i_1 < \dots < i_k$, as

$$Y_\beta(\mathbf{m}) = \prod_{1 \le a < b \le k} (X_{i_a, i_b})^{m_{ab}}.$$

In a W-graph model, as in all stationary graphs, a given motif $\mathbf{m}$ has the same probability to occur at any position. This probability is called the occurrence probability of the motif $\mathbf{m}$ and we denote it by $\mu(\mathbf{m})$

$$\mu(\mathbf{m}) = \Pr\{Y_\beta(\mathbf{m}) = 1\}.$$

Motif probability in W-graph. Because the edges are independent conditionally to the latent labels, the probability of a motif $\mathbf{m}$ in a W-graph has the following general form

$$\mu(\mathbf{m}) = \int \dots \int \prod_{1 \leq a < b \leq k} [W(u_a, u_b)]^{m_{ab}} \, \mathrm{d}u_1 \dots \mathrm{d}u_k.$$

We provide a close form version of this result in a special case that will be used in the simulation study.

Proposition 2 *If the W function has a symmetric product form $W(u, v) = g(u)g(v)$, then*

$$\mu(\mathbf{m}) = \prod_{1 \leq a \leq k} \xi_{m_{a+}}, \qquad \textit{where} \quad \xi_h = \int g(z)^h \, dz,$$

and $m_{a+} = \sum_{1 \leq b \leq k} m_{ab}$ denotes the degree of vertex a in the motif $\mathbf{m}$.

The proof is given in Appendix A.2.

3.2 Occurrence probability estimate

As shown in Picard *et al.* (2008), for a SBM model, with fixed number Q of groups and with parameters $(\boldsymbol{\alpha}, \boldsymbol{\pi})$, the form of $\mu(\mathbf{m})$ is given by

$$\mu(m|\boldsymbol{\alpha}, \boldsymbol{\pi}) = \sum_{\mathbf{c}} \prod_{1 \leq a \leq k} \alpha_{c_a} \prod_{1 \leq a < b \leq k} \pi_{c_a, c_b}^{m_{ab}}, \tag{5}$$

where $\mathbf{c}$ stands for the labeling of the k nodes: $\mathbf{c} = (c_1, \dots c_k)$, each label c_a being taken in $\{1, \dots Q\}$. Keeping Q fixed, but integrating the uncertainty over $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$, we derive the approximate posterior mean

$$\widetilde{\mathbb{E}}[\mu(\mathbf{m})|\mathbf{X}, Q] = \int \int \mu(m|\boldsymbol{\alpha}, \boldsymbol{\pi}) \widetilde{p}(\boldsymbol{\alpha}, \boldsymbol{\pi}|\mathbf{X}, Q) d\boldsymbol{\alpha} d\boldsymbol{\pi}.$$

Proposition 3 *Using the same notation as in Proposition 1, the approximate variational Bayes posterior mean of the occurrence probability under SBM with Q groups is*

$$\widetilde{\mathbb{E}}[\mu(\mathbf{m})|\mathbf{X}, Q] = \left\{ \left[\prod_{q\leq\ell}^{Q} \frac{\Gamma(\eta_{q\ell}+\zeta_{q\ell})}{\Gamma(\eta_{q\ell})} \right] \frac{\Gamma(\sum_{q=1}^{Q} n_q)}{\prod_{q=1}^{Q}\Gamma(n_q)} \right\}$$
$$\times \left\{ \sum_{\mathbf{c}} \left[\prod_{q\leq\ell}^{Q} \frac{\Gamma(\eta_{q\ell}+\eta_{q\ell}^{\mathbf{c}})}{\Gamma(\eta_{q\ell}+\eta_{q\ell}^{\mathbf{c}}+\zeta_{q\ell})} \right] \frac{\prod_{q=1}^{Q}\Gamma(n_q+n_q^{\mathbf{c}})}{\Gamma\left[\sum_{q=1}^{Q}(n_q+n_q^{\mathbf{c}})\right]} \right\},$$

where $\mathbf{c} = (c_1, \ldots, c_k)$, $n_q^{\mathbf{c}} = \sum_a \mathbb{I}\{c_a = q\}$, $\eta_{q\ell}^{\mathbf{c}} = \sum_{1\leq a\neq b\leq k} \mathbb{I}\{c_a = q\}\mathbb{I}\{c_b = \ell\}m_{ab}$ for $q \neq \ell$, $\eta_{qq}^{\mathbf{c}} = \sum_{1\leq a<b\leq k} \mathbb{I}\{c_a = q\}\mathbb{I}\{c_b = q\}m_{ab}$ and $\Gamma(\cdot)$ is the gamma function.

The proof is based on the exact calculation of the mean of the occurrence probability (5) using the variational Bayes posterior (4) and is postponed to Appendix A.2.

Therefore, integrating $\widetilde{\mathbb{E}}[\mu(\mathbf{m})|\mathbf{X}, Q]$ over the number Q of groups, as in Section 2.2, leads to the following approximate variational Bayes posterior mean of the occurrence probability of any motif $\mathbf{m}$

$$\widetilde{\mathbb{E}}[\mu(\mathbf{m})|\mathbf{X}] = \sum_{Q} \widetilde{p}(Q|\mathbf{X})\widetilde{\mathbb{E}}[\mu(\mathbf{m})|\mathbf{X}, Q]. \tag{6}$$

3.3 Testing unexpectedly frequent motifs

In the following, we emphasize that (6) can help in characterizing the count of a motif in a network. Let us consider

$$I_k = \{\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\} | i_j \neq i_l, \forall j \neq l\},$$

the set of all potential positions of $\mathbf{m}$ in the graph. Permuting the rows as well as the columns of the adjacency matrix $\mathbf{m}$ can lead to the same motif, at each position $\beta \in I_k$. Therefore, denoting $\mathcal{R}(\mathbf{m})$ the set of non redundant permutations, the count of a motif is defined as

$$N(\mathbf{m}) = \sum_{\beta\in I_k} \sum_{\mathbf{m}'\in\mathcal{R}(\mathbf{m})} Y_\beta(\mathbf{m}').$$

Since the W-graph model is a stationary model, the expectation and variance of $N(\mathbf{m})$ have analytical forms, as shown in Picard *et al.* (2008) for general class of stationary random graph models. While the calculation of $\mathbb{E}[N(\mathbf{m})]$ is straightforward, the derivation of $\mathbb{V}(N(\mathbf{m}))$ is more technical and involves super-motifs which are made of overlaps between occurrences of $\mathbf{m}$. Therefore, for the sake of the discussion, these two quantities are not given here and we refer to Picard *et al.* (2008). A key point is that both $\mathbb{E}[N(\mathbf{m})]$ and

$\mathbb{V}(N(\mathbf{m}))$ involve occurrence probabilities for which we provide estimators (6). Therefore we propose to replace the $\mu(\cdot)$ terms in $\mathbb{E}[N(\mathbf{m})]$ and $\mathbb{V}(N(\mathbf{m}))$ with their corresponding estimators $\widetilde{\mathbb{E}}[\mu(\cdot)|\mathbf{X}]$. Note that an alternative approach consists in approximating the occurrence probabilities themselves using plug-in estimators. We refer to Bickel *et al.* (2011) who studied the asymptotic normality of such plug-in estimates and to Bhattacharyya and Bickel (2015) who considered a resampling-based approach to estimate the variance of the count.

4 Simulation study

We designed a simulation study to assess the quality of the variational Bayes inference we propose. Our study focuses on the estimation of both the graphon and the motifs frequencies. The methodology obviously depends on the choice of prior parameters for the prior distributions. In practice, we set $a_q^0 = 1, \forall q$ and $\eta_{q,\ell}^0 = \zeta_{q,\ell}^0 = 1, \forall (q,\ell)$. Such choices induce uniform prior distributions over all model parameters.

4.1 Simulation design

Simulation model. We considered W-graph models with graphon function $W(u,v) = g(u)g(v)$ where

$$g(u) = \sqrt{\rho}\lambda u^{\lambda-1}. \tag{7}$$

The parameter ρ controls the density of the graph, meaning that ρ is the mean probability for any two nodes to be connected, while λ controls the concentration of the degrees: the higher λ, the more the edges are concentrated around few nodes. Note that $\lambda = 1$ corresponds to the Erdös-Rnyi model with connection probability ρ. Also note that the maximum of W is $\rho\lambda^2$, which has to remain smaller than 1 so $\lambda \leq 1/\sqrt{\rho}$ must hold. Under model (7), the motif probabilities can be computed using Proposition 2 where

$$\xi_h = (\sqrt{\rho}\lambda)^h/(h\lambda - h + 1).$$

We considered graphs of size $n = 100$ to 316 ($\simeq 10^{2.5}$) with log-density $\log_{10}\rho = -2, -1.5, -1$ and concentration $\lambda = 1, 2, 3$ and 5. 100 graphs were sampled for each configuration. For each sampled graph, we fitted SBM models with $Q = 1$ to 10 groups using the VBEM algorithm described above and computed all approximate posterior distributions.

Criteria. We used the variational posterior mean $\widehat{w}(u,v) = \widetilde{\mathbb{E}}(w(u,v)|\widehat{Q},\mathbf{X})$ as an estimate of $W(u,v)$, where $\widehat{Q}$ stands for the maximum *a posteriori* (MAP) estimate of Q

$$\widehat{Q} = \arg\max_Q \widetilde{p}(Q|\mathbf{X}).$$

Marginalizing over Q (i.e. taking $\widehat{w}(u,v) = \widetilde{\mathbb{E}}(w(u,v)|\mathbf{X})$) provided similar results in all configurations (not shown). To assess the quality of this estimation of W, we computed the root mean squared error ($RMSE$) between its true value and its variational posterior mean, that is

$$RMSE = \sqrt{\iint \left[W(u,v) - \widehat{w}(u,v)\right]^2 \mathrm{d}u\mathrm{d}v}.$$

The integral was evaluated on a thin grid over $[0,1]^2 \times [0,1]$.
As for the motif probability, we considered all motifs m with 2, 3 and 4 nodes. For each of them we computed its probability $\mu(\mathbf{m})$ and we used its variational posterior mean as an estimate: $\widehat{\mu}(\mathbf{m}) = \widetilde{\mathbb{E}}(\mu(\mathbf{m})|\widehat{Q}, \mathbf{X})$. To compare the two, we used the Kullback-Leibler divergence between the corresponding Bernoulli distribution, that is

$$KL(\mathbf{m}) = \mu(\mathbf{m}) \log \frac{\mu(\mathbf{m})}{\widehat{\mu}(\mathbf{m})} + (1 - \mu(\mathbf{m})) \log \frac{1 - \mu(\mathbf{m})}{1 - \widehat{\mu}(\mathbf{m})}.$$

4.2 Results

Computational cost. First, in order to give some insight into the computational cost of the proposed methodology, we recorded the running time for the inference of W-graph models in various scenarios. In this section, we set $\lambda = 2$. The results presented in Table 1 were obtained on an Intel Xeon CPU 3.07GHz, a unique core being used. It appears that estimates are obtained in less than 30 seconds, even for dense ($\rho = 10^{-1}$) networks with $n = 316$ nodes. As expected, the running time is lower for sparse networks, *i.e.* as ρ decreases.

size of the network (n)	$\rho = 10^{-1}$	$\rho = 10^{-1.5}$	$\rho = 10^{-2}$
100	5.64 s	5.10 s	5.20 s
147	5.95 s	5.74 s	5.35 s
215	8.71 s	7.85 s	6.49 s
316	22.09 s	19.61 s	14.47 s

Table 1 Averaged running time (in seconds) for the W-graph model inference procedure, for various sizes n of networks and various graph densities ρ.

Model complexity. Then, we studied the (approximate) posterior distribution of Q. Figure 2 shows how the SBM model adapts to the graphon shape, using a higher number of classes as the W-graph model becomes more distinct from the Erdös-Renyi model, that is as λ increases. We see that, for a same non-Erdös-Renyi graph ($\lambda > 1$), a more complex SBM can be fitted with a larger graph size n. We also see that for the Erdös-Renyi model ($\lambda = 1$), the posterior distribution Q is more concentrated on the true value $Q = 1$ when n is larger.

The last observation is that all posterior distributions are concentrated around $\widehat{Q}$, resulting in similar results when using the MAP distribution $\widetilde{p}(\cdot|\widehat{Q}, \mathbf{X})$ or the averaged one $\widetilde{p}(\cdot|\mathbf{X}) = \sum_Q \widetilde{p}(Q|\mathbf{X})\widetilde{p}(\cdot|\mathbf{X}, Q)$.

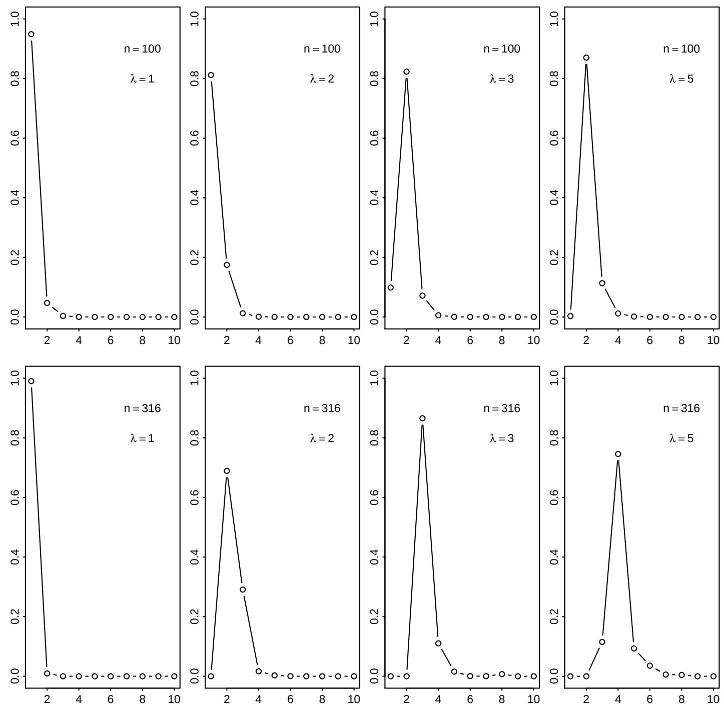


Fig. 2 Approximate posterior distribution $\tilde{p}(Q|\mathbf{X})$ of the number $Q \in \{1, \dots, 10\}$ of classes in the SBM model, for $\rho = 10^{-1.5}$ and various values of n as well as λ.

Estimation of W. Figure 3 shows that the RMSE of the estimate is usually below few percent. As expected, the most difficult configurations are imbalanced ($\lambda \geq 3$) medium size ($n = 100$) graphs. The RMSE also increases with ρ but this only reflects the fact that ρ is the mean value of W, so the error increases with it. However, the relative RMSE ($RMSE/\rho$) actually decreases with ρ (not shown).

Motif probability. We then turned to the motifs probabilities and the results are given in Figure 4. We remind that these quantities are invariant and identifiable in the W-graph, as opposed to the graphon function W. The estimation turns out to be very good, even for very imbalanced shape ($\lambda = 5$) as long as the graph if large ($n = 10^{1.5}$) and dense no too dense ($\rho = 10^{-2}$).

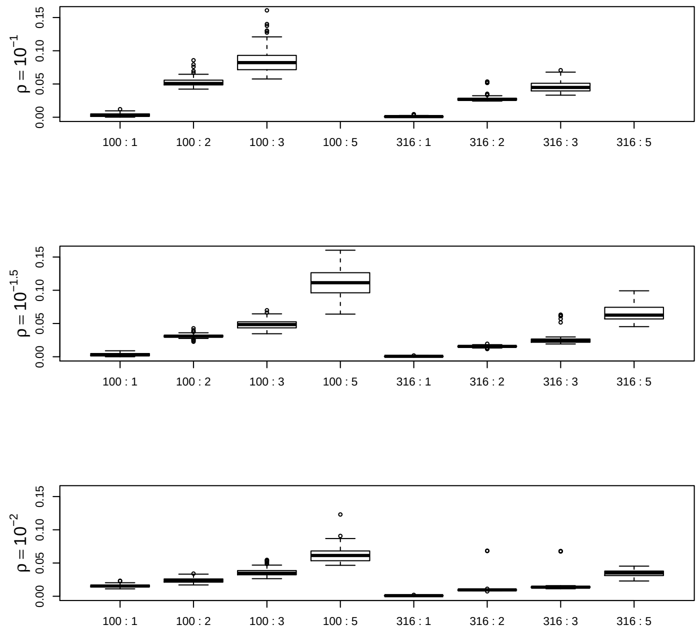


Fig. 3 *RMSE* of the estimate of the graphon function for graph density $\rho = 10^{-2}$, $10^{-1.5}$ and 10^{-1}. x-axis: graph size n and shape λ labeled as $n : \lambda$.

5 French political blogosphere

As in Latouche *et al.* (2011), we consider a subset of the French political blogosphere network. The network is made of 196 vertices connected by 2864 edges. It was built from a single day snapshot of political blogs automatically extracted on 14th october 2006 and manually classified by the 'Observatoire Présidentiel" project (Zanghi *et al.* , 2008). Nodes correspond to hostnames and there is an edge between two nodes if there is a known hyperlink from one hostname to the other. The four main political parties which are present in the data set are the UMP (french "republican"), liberal party (supporters of economic-liberalism), UDF ("moderate" party), and PS (french "democrat"). We run the VBEM algorithm on the data set for $Q \in \{1, \ldots, 20\}$ using the R package *mixer*.

Graphon function. The graphon function estimated using the model averaging approach we proposed in Section 2.2 is given in Figure 5. For this network, we emphasize that the estimated posterior distribution of the number Q of classes is highly concentrated around $Q^* = 12$. As in the preceding section, all prior parameters were set to 1 to induce uniform priors.

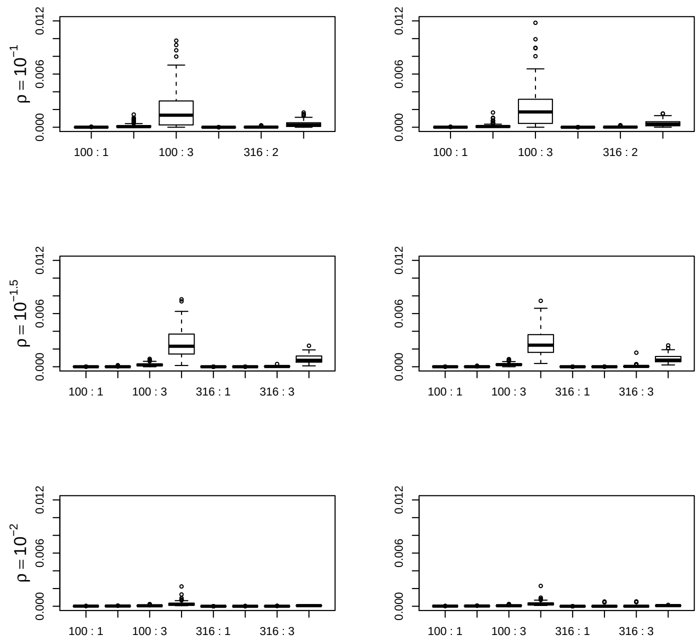


Fig. 4 KL divergence between the true and estimated probabilities for the triangle (left) and square (right) motif. Same legend as Figure 3.

First, we notice that high connectivity regions appear in two series of hills, one along the diagonal and one parallel to the y-axis, close to $x = 1$ (we recall that this function is symmetric). The series of hills on the diagonal each corresponds to a specific political party. In terms of connection patterns, the diagonal structure reveals that blogs of the given community more likely connect to blogs of the same community. Moreover, we emphasize that the plateau as the very bottom left hand side of the graphon function represents blogs from various political parties, from the left wing to the right wing, having very weak connection profiles. Conversely, the series of hills parallel to the axes correspond to blogs, and in particular blogs of political analysts, having strong connections with the different political parties. Because of the identifiability rule which makes the degree $D(x) = \int W(x, y)\mathrm{d}y$ increasing, this region also corresponds to nodes with highest degree. From a global point of view, this region of the graphon plays a critical role as it ensures the connectivity of the whole network. Yet, a closer look at the contour plot given in Figure 6 shows thats the modes of these hills all have a x coordinate close to 1 but also have very different y coordinates, which reveals that some of these blogs have themselves preferential connections with specific political parties.

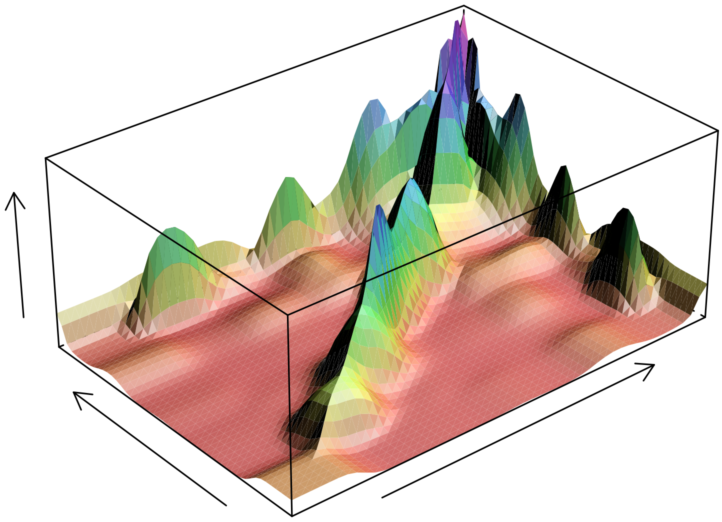


Fig. 5 Graphon function of the French political blogosphere network estimated using the estimated posterior mean derived in Section 2.2.

Motif frequency. To complete the analysis of the French political blogosphere network, we also computed the estimated mean and standard deviation of the motif count $N(\mathbf{m})$, for various motifs. As pointed out in Section 3.3, both quantities involve motif occurrence probabilities $\mu(\cdot)$ for which we provide estimators in Section 3.2, using the variational inference procedure. Our results are summarized in Table 7.
First, its appears that the three motifs which are mainly present in the network are motif 3 (4-edges path), motif 4 (3-branch star), and motif 6 (triangle plus an edge). However, none of the motifs are seen as unexpectedly frequent motifs. Indeed, we found that their observed counts are less than 1.5 standard deviation away for their means under the W-graph model. This means that the W-graph model, estimated using the variational approach, explains reasonably well the presence of the motifs in the network, such that no counts N_{obs} are seen as unexpected. This tends to illustrate the goodness of fit of the estimated W-graph model.

In the same vein, we would like to stress that random graph models in social sciences often consider specific parameters to explain the presence of triangles in networks (see for instance Robins *et al.* (2007)). The additional parameter dedicated to triangles aims at accounting for the 'friends of my friends are my friends' effect. Conversely, a W-graph model focuses on modeling edges between pairs of nodes. Triangles are not specifically modeled and only result from the construction of edges between triads. Interestingly, for the social network of blogs we considered, we found that the observed count is less than one standard deviation away from its mean under the W-graph model. Again,

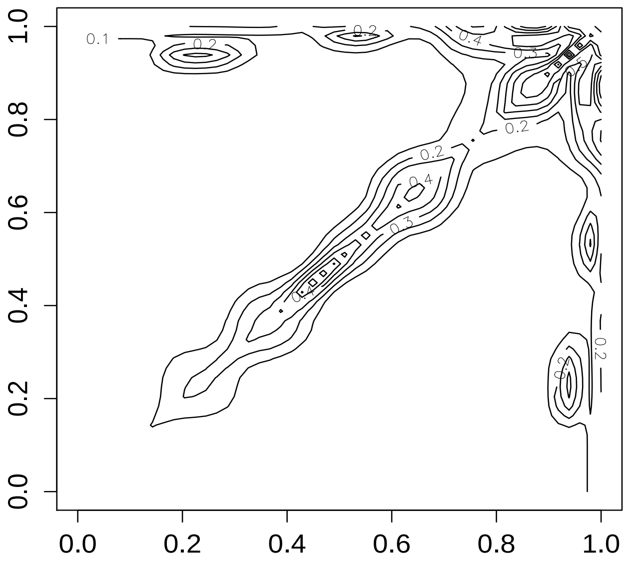


Fig. 6 Contour plot of the French political blogosphere network graphon function estimated using the estimated posterior mean derived in Section 2.2.

this tends to show that the presence of triangles in the network is sufficiently explained by the estimated W-graph model and that the 'friends of my friends' effects is accounted for by the latent position of the actors.

6 Conclusion

We considered the W-graph model which generalizes most of the random graph models commonly used in the literature to extract knowledge from network topologies. The model is defined through a graphon function W which has to be inferred in practice while working on real data.
To this aim, we relied on a variational approximation procedure originally developed for the SBM model that can be seen as a W-graph model with blockwise constant graphon function. Then, we showed how the approximate posterior distribution over the SBM model parameters (including the number of blocks) could be integrated out analytically to obtain an estimate of the posterior distribution of the graphon function.
Using the same approach, we derived the approximate posterior mean of motifs frequencies. We propose to use this expected frequencies under the W-graph as a goodness-of-fit criterion for this model. In the blogosphere application,

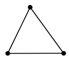
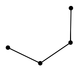
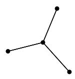
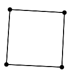
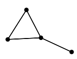
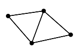
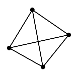
	Motifs	N_{obs}	$\mathbb{E}[N(\mathbf{m})]$	$\sqrt{\mathbb{V}(N(\mathbf{m}))}$	Sd. diff.
1		29715	39722.11	8259.28	-1.21
2		3821	4512.51	1276.13	-0.54
3		608708	968364.08	336800.24	-1.07
4		279771	428867.52	153962.02	-0.97
5		47415	74533.94	35075.09	-0.77
6		270497	397053.82	177049.01	-0.71
7		62071	87849.83	47407.22	-0.54
8		6523	8818.95	5385.87	-0.43

Fig. 7 N_{obs} true counts of motifs in the French political blogosphere network; $\mathbb{E}[N(\mathbf{m})]$ estimated means; $\sqrt{\mathbb{V}(N(\mathbf{m}))}$ estimated standard errors; Sd. diff. (standardized difference $(N_{obs} - \mathbb{E}[N(\mathbf{m})])/\sqrt{\mathbb{V}(N(\mathbf{m}))}$.

we conclude that the most widely studied motifs display a frequency that is consistent with the W-graph model.

A Appendix

A.1 Inference of the function W

Proof of Proposition 1. The first part is straightforward, based on a conditioning of the binnings of u and v

$$\begin{aligned}
\widetilde{p}(w(u,v)|\mathbf{X},Q) &= \widetilde{p}(\pi_{C(u),C(v)}|\mathbf{X},Q) \\
&= \sum_{q\le\ell} \widetilde{p}(\pi_{q,\ell}|\mathbf{X},Q,C(u)=q,C(v)=\ell)\widetilde{\Pr}\{C(u)=q,C(v)=\ell|\mathbf{X},Q\} \\
&= \sum_{q\le\ell} b(w;\eta_{q,\ell},\zeta_{q,\ell})\widetilde{\Pr}\{C(u)=q,C(v)=\ell|\mathbf{X},Q\}.
\end{aligned}$$

We are now left with the calculation of

$$\begin{aligned}
\widetilde{\Pr}\{C(u)=q,C(v)=\ell|\mathbf{X},Q\} &= \widetilde{\Pr}\{\sigma_{q-1}<u<\sigma_q,\sigma_{\ell-1}<v<\sigma_\ell|\mathbf{X},Q\} \\
&= F_{q-1,\ell-1}(u,v;\mathbf{a}) - F_{q,\ell-1}(u,v;\mathbf{a}) - F_{q-1,\ell}(u,v;\mathbf{a}) \\
&\quad + F_{q,\ell}(u,v;\mathbf{a})
\end{aligned}$$

where

- $\mathbf{a},\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ are the parameters of the variational Bayes posterior distributions;
- $b(\cdot;\boldsymbol{\eta},\boldsymbol{\zeta})$ stands for the pdf of the Beta distribution Beta$(\boldsymbol{\eta},\boldsymbol{\zeta})$;
- $F_{q,\ell}(u,v;\mathbf{a})$ denotes the joint cdf of (σ_q,σ_ℓ), as defined in (1), when $\boldsymbol{\alpha}$ has a Dirichlet distribution Dir$(\mathbf{a})$.

The last argument comes from Gouda and Szántai (2010) who give explicit recursions to compute the uni- and bi-variate cdf for the Dirichlet Dir$(\mathbf{a})$, denoted $G_q(u;\mathbf{a})$ and $G_{q,\ell}(u,v;\mathbf{a})$ respectively.
Reminding that the approximate variational posterior of $\boldsymbol{\alpha}$ is Dir$(\mathbf{a})$ and using a simple property of the Dirichlet distribution

$$(\boldsymbol{\alpha}) \sim \mathrm{Dir}(\mathbf{a}) \quad \Rightarrow \quad \left(\sum_{j=1}^{q}\alpha_j,\ \sum_{j=q+1}^{\ell}\alpha_j,\ \sum_{j=\ell+1}^{Q}\alpha_j\right) \sim \mathrm{Dir}\left(\sum_{j=1}^{q}a_j,\ \sum_{j=q+1}^{\ell}a_j,\ \sum_{j=\ell+1}^{Q}a_j\right),$$

the calculation of $F_{q,\ell}(u,v)$ follows as

$$\begin{aligned}
F_{q,\ell}(u,v) &= \widetilde{\Pr}\{\sigma_q<u,\sigma_\ell<v|\mathbf{X},Q\} \\
&= \widetilde{\Pr}\{\sigma_q<u,1-\sigma_\ell>1-v|\mathbf{X},Q\} \\
&= \widetilde{\Pr}\{\sigma_q<u|\mathbf{X},Q\} - \Pr\{\sigma_q<u,\sigma_\ell<1-v|\mathbf{X},Q\} \\
&= G_1(u;[s_q,s_\ell-s_q,s_Q-s_\ell]) - G_{1,3}(u,1-v;[s_q,s_\ell-s_q,s_Q-s_\ell]),
\end{aligned}$$

where the (s_q) are the cumulated parameters: $s_q=\sum_{j=1}^{q}a_j$. ■

A.2 Motif probability

Proof of Proposition 3. We directly write the approximate variational expectation

$$\begin{aligned}
\widetilde{\mathbb{E}}[\mu(\mathbf{m})|\mathbf{X},Q] &= \int\int \mathbb{E}[\mu(\mathbf{m})|\boldsymbol{\alpha},\boldsymbol{\pi}]\widetilde{p}(\boldsymbol{\alpha},\boldsymbol{\pi}|\mathbf{X},Q)\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\pi} \\
&= \int\int\left\{\sum_{\mathbf{c}}\mathbb{E}[\mu(\mathbf{m})|\mathbf{c},\boldsymbol{\pi}]p(\mathbf{c}|\boldsymbol{\alpha})\right\}\widetilde{p}(\boldsymbol{\alpha},\boldsymbol{\pi}|\mathbf{X},Q)\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\pi},
\end{aligned}$$

motif m

adjacency matrix $\mathbf{m}$

$$\begin{bmatrix} 0&1&1\\ 1&0&1\\ 1&1&0 \end{bmatrix} \qquad \begin{bmatrix} 0&1&0\\ 1&0&1\\ 0&1&0 \end{bmatrix} \qquad \begin{bmatrix} 0&1&1&1\\ 1&0&0&0\\ 1&0&0&0\\ 1&0&0&0 \end{bmatrix} \qquad \begin{bmatrix} 0&1&0&1\\ 1&0&1&0\\ 0&1&0&1\\ 1&0&1&0 \end{bmatrix}$$

Fig. 8 Adjacency matrix $\mathbf{m}$ for four typical motifs.

where

$$p(\mathbf{c}|\boldsymbol{\alpha}) = \prod_{1\leq a\leq k} p(c_a|\boldsymbol{\alpha}) = \prod_{1\leq a\leq k}\prod_{1\leq q\leq Q} \alpha_q^{\mathbb{I}\{c_a=q\}} = \prod_{1\leq q\leq Q} \alpha_q^{n_q^{\mathbf{c}}}.$$

Furthermore, we have

$$\begin{aligned}
\mathbb{E}[\mu(\mathbf{m})|\mathbf{c},\boldsymbol{\pi}] &= \Pr\left\{\prod_{1\leq a<b\leq k} X_{ab}^{m_{ab}} = 1|\mathbf{c},\boldsymbol{\pi}\right\} = \prod_{1\leq a<b\leq k} \Pr\left\{X_{ab}=1|c_a,c_b,\boldsymbol{\pi}\right\}^{m_{ab}}\\
&= \prod_{1\leq a<b\leq k}\prod_{1\leq q,\ell\leq Q} \pi_{q\ell}^{\mathbb{I}\{c_a=q\}\mathbb{I}\{c_b=\ell\}m_{ab}}\\
&= \prod_{1\leq q<\ell\leq Q}\prod_{a\neq b} \pi_{q\ell}^{\mathbb{I}\{c_a=q\}\mathbb{I}\{c_b=\ell\}m_{ab}} \prod_{1\leq q\leq Q}\prod_{1\leq a<b\leq k} \pi_{qq}^{\mathbb{I}\{c_a=q\}\mathbb{I}\{c_b=q\}m_{ab}} = \prod_{1\leq q\leq \ell\leq Q} \pi_{q\ell}^{\eta_{q\ell}^{\mathbf{c}}},
\end{aligned}$$

so we end up with

$$\begin{aligned}
\widetilde{\mathbb{E}}[\mu(\mathbf{m})|\mathbf{X},Q] &= \int\int\sum_{\mathbf{c}}\prod_{1\leq q\leq\ell\leq Q}\pi_{q\ell}^{\eta_{q\ell}^{\mathbf{c}}}\prod_{1\leq q\leq Q}\alpha_q^{n_q^{\mathbf{c}}}\widetilde{p}(\boldsymbol{\alpha},\boldsymbol{\pi}|Q)\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\pi}\\
&= \int\int\sum_{\mathbf{c}}\prod_{1\leq q\leq\ell\leq Q}\pi_{q\ell}^{\eta_{q\ell}^{\mathbf{c}}}\prod_{1\leq q\leq Q}\alpha_q^{n_q^{\mathbf{c}}}\prod_{1\leq q\leq\ell\leq Q}\frac{\Gamma(\eta_{q\ell}+\zeta_{q\ell})}{\Gamma(\eta_{q\ell})\Gamma(\zeta_{q\ell})}\pi_{q\ell}^{\eta_{q\ell}-1}(1-\pi_{q\ell})^{\zeta_{q\ell}-1}\\
&\qquad \frac{\Gamma(\sum_{1\leq q\leq Q}n_q)}{\prod_{1\leq q\leq Q}\Gamma(n_q)}\prod_{1\leq q\leq Q}\alpha_q^{n_q-1}\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{\pi}\\
&= \sum_{\mathbf{c}}\prod_{1\leq q\leq\ell\leq Q}\frac{\Gamma(\eta_{q\ell}+\zeta_{q\ell})}{\Gamma(\eta_{q\ell})\Gamma(\zeta_{q\ell})}\int\pi_{q\ell}^{\eta_{q\ell}+n_{q\ell}^{\mathbf{c}}-1}(1-\pi_{q\ell})^{\zeta_{q\ell}-1}\,\mathrm{d}\pi_{q\ell}\\
&\qquad \frac{\Gamma(\sum_{1\leq q\leq Q}n_q)}{\prod_{1\leq q\leq Q}\Gamma(n_q)}\prod_{1\leq q\leq Q}\int\alpha_q^{n_q+n_q^{\mathbf{c}}-1}\,\mathrm{d}\boldsymbol{\alpha}_q\\
&= \sum_{\mathbf{c}}\prod_{1\leq q\leq\ell\leq Q}\frac{\Gamma(\eta_{q\ell}+\zeta_{q\ell})}{\Gamma(\eta_{q\ell})\Gamma(\zeta_{q\ell})}\frac{\Gamma(\eta_{q\ell}+\eta_{q\ell}^{\mathbf{c}})\Gamma(\zeta_{q\ell})}{\Gamma(\eta_{q\ell}+\eta_{q\ell}^{\mathbf{c}}+\zeta_{q\ell})}\frac{\Gamma(\sum_{1\leq q\leq Q}n_q)}{\prod_{1\leq q\leq Q}\Gamma(n_q)}\frac{\prod_{1\leq q\leq Q}\Gamma(n_q+n_q^{\mathbf{c}})}{\Gamma\sum_{1\leq q\leq Q}(n_q+n_q^{\mathbf{c}})},
\end{aligned}$$

and the proof is completed. ■

Proof of Proposition 2. Because the Z_i's are uniformly distributed over $[0; 1]$, we have

$$\begin{aligned}
\mu(\mathbf{m}) &= \Pr\{Y(i_1, \dots i_k; m) = 1\} \\
&= \int \cdots \int \Pr\left\{ \prod_{1 \le a < b \le k} X_{i_a i_b}^{m_{ab}} = 1 | Z_{i_1} = z_1, \dots Z_{i_k} = z_k \right\} \mathrm{d}z_1 \dots \mathrm{d}z_k \\
&= \int \cdots \int \prod_{1 \le a < b \le k} [w(z_a) w(z_b)]^{m_{ab}} \mathrm{d}z_1 \dots \mathrm{d}z_k \\
&= \int \cdots \int \prod_{1 \le a \le k} w(z_a)^{m_{a+}} \mathrm{d}z_1 \dots \mathrm{d}z_k \quad = \prod_{1 \le a \le k} \int w(z)^{m_{a+}} \mathrm{d}z.
\end{aligned}$$

■

Acknowledgements The authors thanks Stevenn Volant for helpful comments and discussions. The authors also thank the anonymous reviewer for his helpful remarks on our work.

References

Airoldi, E. M., Costa, T. B. and Chan, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems*, 692–700.

Asta, D. and Shalizi, C. R. (2014), Geometric network comparison. Technical report, arXiv:1411.1350v1.

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

Barbour, A. and Reinert, G. (2006). Discrete small world networks. *Electron. J. Probab.* **11** no. 47, 1234–1283.

Beal, J., M. and Ghahramani, Z. (2003). The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayes. Statist.* **7** 543–52.

Bhattacharyya, S. and Bickel, P. J. (12, 2015). Subsampling bootstrap of count features of networks. *Ann. Statist.* **43 (6)** 2384–2411.

Bickel, P. and Chen, A. (2009). A non parametric view of network models and Newman-Girvan and other modularities. In *Proceedings of the National Academy of Sciences*, volume 106, 21068–21073.

Bickel, P., Chen, A. and Levina, E. (2011). The method of moments and degree distributions for network models. *The Annals of Statistics.* **39 (5)** 2280–2301.

Bollobás, B., Janson, S. and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Rand. Struct. Algo.* **31 (1)** 3–122.

Borgs, C., Chayes, J., Cohn, H. and Ganguly, S. (2015), Consistent nonparametric estimation for heavy-tailed sparse graphs. Technical report, arXiv:1508.06675.

Celisse, A., Daudin, J.-J. and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Statis.* **6** 1847–99.

Chan, S. and Airoldi, E. (2014). A consistent histogram estimator for exchangeable graph models. In *Journal of Machine Learning Research and Conference Proceedings*, 208–216.

Channarond, A., Daudin, J.-J. and Robin, S. (2012). Classification and estimation in the stochastic block model based on the empirical degrees. *Electron. J. Stat.* **6** 2574–601.

Chatterjee, S. (2015). Matrix estimation by Universal Singular Value Thresholding. *The Annals of Statistics.* **43 (1)** 177–214.

Daudin, J.-J., Picard, F. and Robin, S. (Jun, 2008). A mixture model for random graphs. *Stat. Comput.* **18 (2)** 173–83.

Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* **7 (28)** 33–61.

Gazal, S., Daudin, J.-J. and Robin, S. (2012). Accuracy of variational estimates for random graph mixture models. *Journal of Statistical Computation and Simulation.* **82 (6)** 849–862.

Girvan, M. and Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences.* **99 (12)** 7821.

Gouda, A. and Szántai, T. (2010). On numerical calculation of probabilities according to Dirichlet distribution. *Ann. Oper. Res.* **177** 185–200. DOI: 10.1007/s10479-009-0601-9.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. **14 (4)** 382–417.

Hoff, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems*, 657–664.

Kallenberg, O. (1999). Multivariate sampling and the estimation problem for exchangeable arrays. *Journal of Theoretical Probability.* **12 (3)** 859–883.

Latouche, P., Birmelé, E. and Ambroise, C. (2011). Overlapping stochastic block models with application to the french political blogosphere. *Annals of Applied Statistics.* **5 (1)** 309–336.

Latouche, P., Birmelé, E. and Ambroise, C. (2012). Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling.* **12 (1)** 93–115.

Lloyd, J., Orbanz, P., Ghahramani, Z. and Roy, D. (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems*, 998–1006.

Lovász, L. and Szegedy, B. (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B.* **96 (6)** 933 – 957.

Mariadassou, M., Robin, S. and Vacher, C. (06, 2010). Uncovering latent structure in valued graphs: A variational approach. *Ann. Appl. Stat.* **4 (2)** 715–742.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Networks motifs: simple building blocks of complex networks. *Science.* **298** 824–827.

Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–87.

Palla, G., Lovasz, L. and Vicsek, T. (Apr, 2010). Multifractal network generator. *Proc. Natl. Acad. Sci. U.S.A.* **107 (17)** 7640–7645.

Picard, F., Daudin, J.-J., Koskas, M., Schbath, S. and Robin, S. (2008). Assessing the exceptionality of network motifs,. *J. Comput. Biol.* **15 (1)** 1–20.

Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007). An introduction to exponential random graph models for social networks. *Social Networks.* **29** 173–191.

Stark, D. (2001). Compound poisson approximations of subgraph counts in random graphs. *Random Struct. and Algo.* **18 (1)** 39–60.

Volant, S., Magniette, M.-L. M. and Robin, S. (2012). Variational bayes approach for model aggregation in unsupervised classification with markovian dependency. *Comput. Statis. & Data Analysis.* **56 (8)** 2375 – 2387.

Watts, D. J. and Strogatz, S. H. (Jun, 1998). Collective dynamics of 'small-world' networks. *Nature.* **393 (6684)** 440–442.

Wolfe, P. J. and Olhede, S. C. (2013), Nonparametric graphon estimation. Technical report, arXiv:1309.5936.

Yang, J., Han, Q. and Airoldi, E. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Journal of Machine Learning Research and Conference Proceedings*, 1060–1067.

Zanghi, H., Ambroise, C. and Miele, V. (2008). Fast online graph clustering via erdös renyi mixture. *Pattern Recognition.* **41 (12)** 3592–3599.