



HAL
open science

Learning common structures in a collection of networks. An application to food webs

Saint-Clair Chabert-Liddell, Pierre M Barbillon, Sophie Donnet

► **To cite this version:**

Saint-Clair Chabert-Liddell, Pierre M Barbillon, Sophie Donnet. Learning common structures in a collection of networks. An application to food webs. 2022. hal-03685293v1

HAL Id: hal-03685293

<https://hal.science/hal-03685293v1>

Preprint submitted on 2 Jun 2022 (v1), last revised 28 Mar 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning common structures in a collection of networks. An application to food webs

Saint-Clair Chabert-Liddell¹, Pierre Barbillon¹, and Sophie Donnet¹

¹Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA
Paris-Saclay, 75005, Paris, France

Abstract

Let a collection of networks represent interactions within several (social or ecological) systems. Two main issues arise: identifying similarities between the topological structures of the networks or clustering the networks according to the similarities in their structures. We tackle these two questions with a probabilistic model based approach. We propose an extension of the Stochastic Block Model (SBM) adapted to the joint modeling of a collection of networks. The networks in the collection are assumed to be independent realizations of SBMs. The common connectivity structure is imposed through the equality of some parameters.

The model parameters are estimated with a variational Expectation-Maximization (EM) algorithm. We derive an ad-hoc penalized likelihood criterion to select the number of blocks and to assess the adequacy of the consensus found between the structures of the different networks. This same criterion can also be used to cluster networks on the basis of their connectivity structure. It thus provides a partition of the collection into subsets of structurally homogeneous networks.

The relevance of our proposition is assessed on two collections of ecological networks. First, an application to three stream food webs reveals the homogeneity of their structures and the correspondence between groups of species in different ecosystems playing equivalent ecological roles. Moreover, the joint analysis allows a finer analysis of the structure of smaller networks. Second, we cluster 67 food webs according to their connectivity structures and demonstrate that five mesoscale structures are sufficient to describe this collection.

1 Introduction

Context The last few years have seen an increase in the number of interaction networks collected, as networks are popular tools for representing the functioning of a social or ecological system. For a long time the statistical analysis of network data has focused on analyzing a single network at a time. This can be performed either

by looking at local or global topological features, or by setting a probabilistic model inferred from the network (Kolaczyk, 2009). When several networks describing the same kind of interactions are available, a natural question is to assess to what extent they are similar or different. As network data are complex by nature, this comparison of different networks is not an easy task and has mainly focused on comparing statistical topological features on the local, global or mesoscale levels (see Donnat and Holmes, 2018, for a review of graph distances).

In this paper, we focus on the mesoscale structure of the networks by assuming that the nodes can be clustered into groups on the basis of their connectivity pattern (White et al., 1976). We assume that the considered networks have no nodes in common and that the nodes of different networks are not linked as it may be the case in multilayer networks (Kivelä et al., 2014). Furthermore, the networks are assumed to have interactions of the same type (directed or not) and with the same valuation (binary, discrete or continuous). A set of such networks forms what we call in this paper a collection of networks, although some authors may use this terminology in a different meaning. When observing such a collection, we aim to determine if the respective structures of the networks are similar.

A classical tool to infer the mesoscale structure of a single network is the Stochastic Block Model (Holland et al., 1983; Snijders and Nowicki, 1997, SBM). In this model, a latent variable is associated with each node giving its group/block membership. Nodes belonging to the same block share the same connectivity pattern. The SBM has easily interpretable parameters and its framework allows multiple extensions such as modeling the interactions with various distributions (Mariadassou et al., 2010).

Inferring independently the structures for each network and comparing them may be misleading. Indeed, a given network may have several possible clusterings of the nodes (Peel et al., 2017) with similar probabilities (Peixoto, 2014). Furthermore, the observation of a network may be noisy (Guimerà and Sales-Pardo, 2009), especially for ecological networks, the sampling of which is known to be incomplete (Rivera-Hutinel et al., 2012).

Our contribution Thus, we propose to jointly model a collection of networks by extending the SBM. We assume that the networks are independent realizations of SBMs that share common parameters. The natural and interesting consequence is the correspondence between the blocks of the different networks. The proposed model called *colSBM* comes with a few variants, the simplest one assumes that the parameters of the SBMs are identical leading to a collection of i.i.d. networks. As this assumption might be too restrictive for real networks, we introduce two relaxations on this assumption. The first one is to allow the distribution of the block memberships to vary between networks and even to allow some networks to not populate certain blocks. This enables to model a collection of networks where the structure of certain networks is encompassed in the structure of other networks. The second mechanism is to allow networks to have the same structure up to a density parameter. This is particularly useful to model networks with different sampling efforts as it has a direct impact on the density of ecological networks (Blüthgen et al., 2006). The inference of the block memberships, the model

parameters and the model selection are done through an ad-hoc version of classic tools when inferring SBM, namely a Variational EM algorithm for the inference and an adaptation of the integrated classification likelihood (ICL) criterion for the model selection (Daudin et al., 2008).

The interest of the model is then two-folds, the first one is to find a common connectivity pattern which explains the structure of the different networks in the collection and to assess via model selection whether these structures is a reasonable fit for the collection. As a by-product, it allows a fine analysis of the role structure of the different nodes in the networks. In social/ecological networks, individuals/species with the same block membership play the same social/ecological role in its system (White et al., 1976; Allesina and Pascual, 2009). By sharing the blocks between the networks, *colSBM* allows to recover sets of nodes which play the same structural roles in different networks. The second one is to provide a partition of the collection of networks into groups of structurally homogeneous networks. Both aspects have some practical implications in ecology (Ohlsson and Eklöf, 2020; Michalska-Smith and Allesina, 2019).

As a side effect, by modeling these networks together, provided that the networks have common connectivity patterns, we can use the information of certain networks to recover noisy information from other networks by improving the prediction of missing links (Clauset et al., 2008). Hence *colSBM* has a stabilizing effects on the block-clustering of the nodes of the networks and might give a block membership that is closer to the one of the full real network than just a single *SBM* as this will be shown in the numerical studies and application.

Related work As the SBM is a very flexible model, it has been adapted to multilayer networks. To name a few, Matias and Miele (2017) model a collection of networks along a time gradient, the connectivity structure varies from time to time but they integrate a sparsity parameter, which is similar to our density parameter in the binary case. When dealing with networks with no common nodes, Chabert-Liddell et al. (2021) deal with multilevel networks where the networks are linked by a hierarchical relation between the nodes of the different levels. Within the SBM framework the closest work to ours is the strata multilayer SBM (Stanley et al., 2016), in that it looks for both common connectivity patterns and network clustering. However, it does not consider a collection of networks but a multiplex network where all the networks share the same nodes.

Most contributions about collections of networks rely on some node correspondence between the networks. Recently, motivated by the analysis of fMRI data a few works extend the SBM to model population of networks (Paul and Chen, 2018; Pavlović et al., 2020). Le et al. (2018) make the assumption that the networks of the collection are noisy realizations of the true network, while Reyes and Rodriguez (2016) use in a Bayesian framework a hierarchical SBM to model the collection. Signorelli and Wit (2020) propose a mixture of network models which is not restricted to the SBM. The contributions dealing with networks with no node correspondence include a hierarchical mixed membership SBM, using a common bayesian prior on the connectivity parameter of the different networks (Sweet et al., 2014). Finally on network clustering, Mukherjee et al. (2017) propose to fit a graphon when the

networks in the collection have the same number of nodes while they use graph moments when this is not the case.

Outline Section 2 recalls the definition of the Stochastic Block Model. We motivate our new approach by inferring it independently on a collection of food webs. Then in Section 3, we present the different *colSBMs*. The likelihood expression is provided in Section 4, together with some identifiability conditions. We develop the methodology for the parameter estimation and model selection in Section B, while Section 6 deals with network clustering. We finally propose with two applications on food webs in Section 7. First, we compare the structures of 3 networks and show the information transfer between these networks. Second, we seek a partition of a collection of 67 networks. The technical details and numerical studies which demonstrates the efficiency of our inference procedure and the pertinence of our model selection criterion are left in Supplementary Material.

2 Data motivation and the stochastic block model

Consider a collection of M independent networks where each network indexed by m involves its own n_m nodes. The networks are encoded into their adjacency matrices $(X^m)_{m \in \{1, \dots, M\}}$ such that: $\forall m \in \{1, \dots, M\}, \forall (i \neq j) \in \{1, \dots, n_m\}^2$,

$$\begin{cases} X_{ij}^m = 0 & \text{if no interaction is observed between species } i \text{ and } j \text{ of network } m \\ X_{ij}^m \neq 0 & \text{otherwise.} \end{cases}$$

If the networks represent binary interactions then $X_{ij}^m \in \mathcal{K} = \{0, 1\}$, $\forall (m, i, j)$; if the interactions are weighted such as counts, then $X_{ij}^m \in \mathcal{K} = \mathbb{N}$. Moreover, all the networks encompass the same type of interactions (binary, count...) and no self-interaction is considered. Besides, for the sake of simplicity, we assume that all the networks are directed. The extension to undirected networks i.e. such that $X_{ij}^m = X_{ji}^m$ for any $i \neq j$ is straightforward. $\mathbf{X} = (X^1, \dots, X^M)$ denotes the collection of adjacency matrices.

A first ecological example: three stream food webs As a first example, we consider the collection of three stream food webs from [Thompson and Townsend \(2003\)](#). The three networks collected respectively in Martins (Maine USA), Cooper and Herlzier (North-Carolina, USA) involve respectively 105, 58 and 71 species resulting in 343, 126 and 148 binary edges respectively. Classically, the food web edges represent directed trophic links showing the energy flow i.e. $X_{ij}^m = 1$ if species j preys on species i , with no reciprocal interactions. When aiming at unraveling the structure of these networks the Stochastic Block Model (SBM) is an interesting tool which has proven its high flexibility by encompassing a large variety of structures (see [Allesina and Pascual, 2009](#), for the particular case of food webs). When dealing with three networks, the standard strategy that we describe below, is to fit separately one SBM per network.

Separate SBM (sepSBM) The SBM introduces clusters of nodes and assumes that the interaction between two nodes is driven by the clusters the nodes belong to. More precisely, for network m , let the n_m nodes be divided into Q_m clusters. Let $Z^m = (Z_i^m, \dots, Z_{n_m}^m)$ be independent latent random variables such that $Z_i^m = q$ if node i of network m belongs to cluster q with $q \in \{1, \dots, Q_m\}$ and

$$P(Z_i^m = q) = \pi_q^m \quad (1)$$

where $\pi_q^m > 0$ and $\sum_{q=1}^{Q_m} \pi_q^m = 1$. Given the latent variables Z^m , the X_{ij}^m 's are assumed to be independent and distributed as

$$X_{ij}^m | Z_i^m = q, Z_j^m = r \sim \mathcal{F}(\cdot; \alpha_{qr}^m), \quad (2)$$

where \mathcal{F} is referred to as the emission distribution. \mathcal{F} is the Bernoulli distribution for binary interactions, and the Poisson distribution for weighted interactions such as counts. Let f be the density of the emission distribution, then:

$$\log f(X_{ij}^m; \alpha_{qr}^m) = \begin{cases} X_{ij}^m \log(\alpha_{qr}^m) + (1 - X_{ij}^m) \log(1 - \alpha_{qr}^m) & \text{for Bernoulli emission} \\ -\alpha_{qr}^m + X_{ij}^m \log(\alpha_{qr}^m) - \log(X_{ij}^m!) & \text{for Poisson emission} \end{cases} \quad (3)$$

Equations (1), (2) and (3) define the SBM model and we will now use the following short notation:

$$X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q_m, \boldsymbol{\pi}^m, \boldsymbol{\alpha}^m). \quad (\text{sep-SBM})$$

where \mathcal{F} encodes the emission distribution, n_m is the number of nodes, Q_m is the number of blocks and $\boldsymbol{\pi}^m = (\pi_q^m)_{q=1, \dots, Q_m}$ is the vector of their proportions. The $Q \times Q$ matrix $\boldsymbol{\alpha}^m = (\alpha_{qr}^m)_{q,r=1, \dots, Q_m}$ denotes the connection parameters i.e. the parameters of the emission distribution. Moreover, $\alpha_{qr}^m \in \mathcal{A}_{\mathcal{F}}$ where $\mathcal{A}_{\mathcal{F}} = (0, 1)$ (resp. $\mathcal{A}_{\mathcal{F}} = \mathbb{R}^{*+}$) for the Bernoulli (resp. Poisson) emission distribution. In the *sep-SBM* model, each network m is modeled independently with its own parameters $(\boldsymbol{\pi}^m, \boldsymbol{\alpha}^m)$.

Application to the three stream food webs We fit the *sepSBM* on the 3 stream food webs, respectively referred to as Martins, Cooper and Herlzier. To do so, we use the `sbm` R-package (Chiquet et al., 2021; Leger et al., 2020) on each network, which implements a variational version of the EM algorithm to estimate the parameters and selects the number of clusters Q_m using a penalized likelihood criterion ICL. These inference tools will be recalled hereafter.

We obtain respectively $\widehat{Q}_1 = 5$ blocks for Martins, $\widehat{Q}_2 = 3$ blocks for Cooper and $\widehat{Q}_3 = 4$ blocks for Herlzier. The adjacency matrices of the food webs reordered by block membership are plotted in Figure 1. Each food web is composed of 2 blocks of basal species (the 2 bottom blocks). For Cooper, the higher trophic levels are grouped together in the same block, as the lack of statistical power does not allow refinement of the species clustering. For Herlzier the higher trophic level is separated into 2 blocks mainly determined on how much they prey on the less preyed basal block. Martins has a separation into 3 blocks, the third one is a medium trophic level, which preys on basal species and is highly preyed on by species of the first

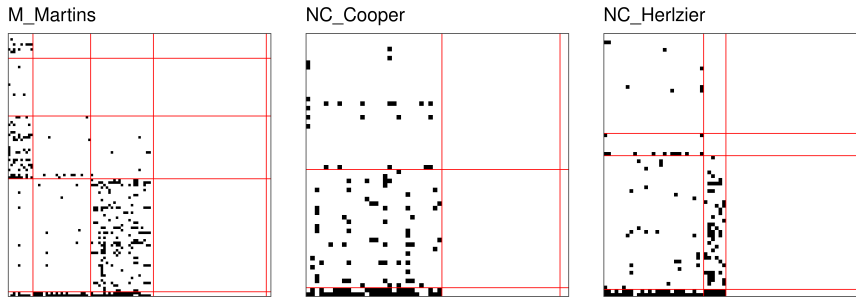


Figure 1: Matricial view of 3 stream food webs. The species are reordered by blocks and blocks are ordered by expected out-degrees to emulate the trophic levels (bottom to top and right to left). The blocks have been obtained by fitting a SBM on each network separately.

block. The first two blocks are made up of higher trophic level species, with the last two blocks being much less connected than the first.

As can be seen in Figure 1, the connectivity structures of these three networks seem to have a lot of similarities. To explore further this aspect, the following section is dedicated to the presentation of several *colSBM* models assuming common structures among the networks of a given collection.

3 Joint modeling of a collection of networks

We now present a set of probabilistic models designed to introduce structure consensus into a collection of networks of interest. For ease of notation, we develop the models for directed networks; extensions to the undirected cases are straightforward. Note that the networks $(X^m)_{m=1,\dots,M}$ are always assumed to be independent random objects. A summary of the various models is provided in Table 1, from the most to the less constrained model

3.1 A collection of i.i.d. SBM

The first model we propose is the most constrained one and assumes that the networks are independent realizations of the same Q -blocks SBM model with identical parameters. The so-called *iid-colSBM* states that:

$$X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}, \boldsymbol{\alpha}), \quad \forall m = 1, \dots, M, \quad (\textit{iid-colSBM})$$

where $\forall (q, r) \in \{1, \dots, Q\}^2$, $\alpha_{qr} \in \mathcal{A}_{\mathcal{F}}$, $\pi_q \in]0, 1]$ and $\sum_{q=1}^Q \pi_q = 1$. The model involves $(Q-1)+Q^2$ parameters, the first term corresponding to the block proportions and the second term to the connection parameters.

However, assuming that the blocks are represented in the same proportions in each network is a strong assumption that may lead to the model being of little practical use. The following model relaxes this assumption.

3.2 A collection of networks with varying block sizes

π -colSBM assumes that the networks share a common connectivity structure encoded in α , but that the proportions of the blocks are specific to each network. Moreover, by allowing some block proportions to be null, the model encompasses situations where some blocks may not be represented in all the realized networks. More precisely, for $m \in \{1, \dots, M\}$, the X^m are independent and

$$X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}^m, \alpha) \quad \forall m = 1, \dots, M. \quad (\pi\text{-colSBM})$$

where $\alpha_{qr} \in \mathcal{A}_{\mathcal{F}}$. Moreover, we assume that $\sum_{q=1}^Q \pi_q^m = 1, \forall m \in \{1, \dots, M\}$ where $\pi_q^m \in [0, 1]$ and if $\pi_q^m = 0$ then the block q is not represented in network m . In addition, we assume that for any $q \in \{1, \dots, Q\}$, $\exists m \in \{1, \dots, M\}$ such that $\pi_q^m > 0$, meaning that any block q is represented in at least one network. Let S be the $M \times Q$ support matrix such that $\forall(m, q)$

$$S_{mq} = \mathbf{1}_{\pi_q^m > 0}.$$

Then, the set of admissible supports is

$$\mathcal{S}_Q := \left\{ S \in \mathcal{M}_{M,Q}(\{0, 1\}), \sum_{m=1}^M S_{mq} \geq 1 \quad \forall q = 1, \dots, Q \right\}$$

For a given matrix S , the number of parameters of the π -colSBM model is deduced as follows:

$$\text{NP}(\pi\text{-colSBM}) = \sum_{m=1}^M \left(\sum_{q=1}^Q S_{qm} - 1 \right) + \sum_{q,r=1}^Q \mathbf{1}_{(S'S)_{qr} > 0}$$

The first term corresponds to the non-null block proportions in each network. The second quantity accounts for the fact that some blocks may never be represented simultaneously in any network, so the corresponding connection parameters are not useful for defining the model (see the illustration below). The number of parameters is bounded by $M(Q - 1) + Q^2$, this upper bound corresponding to the case where all the blocks are represented in all the networks, but with varying proportions.

Illustration We illustrate the flexibility of this model with three examples, all with $Q = 3$ and $M = 2$.

1. First consider the situation where the 3 blocks are represented in the two networks but with different block proportions:

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \quad \begin{matrix} \pi^1 = [.25, .25, .50] \\ \pi^2 = [.20, .50, .30] \end{matrix}.$$

In that case, $S = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ and the number of parameters is $2(3-1) + 3 \times 3 = 13$.

2. Now imagine two networks with imbricated structures. Blocks 1 and 3 are represented in the two networks while block 2 only exists in network 1.

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \quad \begin{array}{l} \pi^1 = [.25, .25, .50] \\ \pi^2 = [.40, 0, .60] \end{array}.$$

In that case, $S = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ and the number of parameters is $(3 - 1) + (2 - 1) + 3 \times 3 = 12$.

3. Finally, let us consider two networks with partially imbricated structures. The two networks share block 1 (for instance super predators) but the remaining nodes of each network cannot be considered as equivalent in terms of connectivity:

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \cdot \\ \alpha_{31} & \cdot & \alpha_{33} \end{pmatrix} \quad \begin{array}{l} \pi^1 = [.25, .75, 0] \\ \pi^2 = [.40, 0, .60] \end{array}.$$

In that case, $S = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Moreover, blocks 2 and 3 never interact because their elements do not belong to the same network and so α_{23} and α_{32} are not required to define the model. As a consequence, the number of parameters is equal to $(2 - 1) + (2 - 1) + 7 = 9$.

3.3 A collection of networks with varying density (δ -colSBM)

The *iid-colSBM* can be relaxed in another direction, assuming that the M networks exhibit similar intra- and inter- blocks connectivity patterns but with different densities. More precisely, let $\delta_m \in \mathbb{R}$ be a density parameter for network m . The δ -colSBM is defined as follows:

$$X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}, \delta_m \boldsymbol{\alpha}). \quad (\delta\text{-colSBM})$$

with $\pi_q > 0, \forall q = 1, \dots, Q, \sum_{q=1}^Q \pi_q = 1$. Moreover $\forall (m, q, r), \delta_m \alpha_{qr} \in \mathcal{A}_{\mathcal{F}}$ and one of the density parameter equal to one ($\delta_1 = 1$) for identifiability purpose. This model mimics different intensity of connections between networks. δ -colSBM involves $\text{NP}(\delta\text{-colSBM}) = (Q - 1) + Q^2 + (M - 1)$ parameters.

3.4 Collection of networks with varying block sizes and density ($\delta\pi$ -colSBM)

Finally, we propose to mix the models π -colSBM and δ -colSBM to obtain a more complex one which allows each network to have its own block proportions $\boldsymbol{\pi}^m$ as well as a specific scale density parameter δ_m . Then, the $(X^m)_{m \in \{1, \dots, M\}}$ are independent and

$$X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}^m, \delta_m \boldsymbol{\alpha}), \quad (\delta\pi\text{-colSBM})$$

Model name	Block prop.	Connection param.	Nb of param.
<i>iid-colSBM</i>	$\pi_q^m = \pi_q, \pi_q > 0$	$\alpha_{qr}^m = \alpha_{qr}$	$(Q - 1) + Q^2$
π - <i>colSBM</i>	$\pi_q^m, \pi_q^m \geq 0$	$\alpha_{qr}^m = \alpha_{qr}$	$\leq M(Q - 1) + Q^2$
δ - <i>colSBM</i>	$\pi_q^m = \pi_q, \pi_q > 0$	$\alpha_{qr}^m = \delta_m \alpha_{qr}$	$(Q - 1) + Q^2 + (M - 1)$
$\delta\pi$ - <i>colSBM</i>	$\pi_q^m, \pi_q^m \geq 0$	$\alpha_{qr}^m = \delta_m \alpha_{qr}$	$\leq MQ + Q^2 - 1$
sepSBM	$\pi_q^m, \pi_q^m > 0$	α_{qr}^m	$\sum_{m=1}^M (Q_m - 1) + Q_m^2$

Table 1: Summary of the various models defined in Section 3. The last line corresponds to modeling separately each network as presented in Section 2.

where $\forall(m, q, r), \delta_m \alpha_{qr} \in \mathcal{A}_{\mathcal{F}}, \delta_1 = 1, \pi_q^m \geq 0$ and $\sum_{q=1}^Q \pi_q^m = 1$. The number of parameters is given by

$$\text{NP}(\delta\pi\text{-colSBM}) = \text{NP}(\pi\text{-colSBM}) + M - 1,$$

the last term corresponding to the additional proper density of each network. Note that $\text{NP}(\delta\pi\text{-colSBM}) \leq M(Q - 1) + Q^2 + M - 1 = MQ + Q^2 - 1$.

4 Likelihood and identifiability of the models

In this section, we derive the expression of the likelihood of the most complex model $\delta\pi$ -*colSBM* and provide conditions to ensure the identifiability of the parameters for each of the four models.

4.1 Log-likelihood expression

For a given matrix S , let $\boldsymbol{\theta}_S$ be:

$$\boldsymbol{\theta}_S = (\boldsymbol{\pi}^1, \dots, \boldsymbol{\pi}^M, \delta_1, \dots, \delta_M, \boldsymbol{\alpha}) = (\boldsymbol{\pi}, \boldsymbol{\delta}, \boldsymbol{\alpha}),$$

where $\pi_q^m = 0$ for any q such that $S_{mq} = 0$. Let $Z_{iq}^m = \mathbf{1}_{Z_i^m=q}$ be the latent variable such that $Z_{iq}^m = 1$ if node i of network m belongs to block q , $Z_{iq}^m = 0$ otherwise. We define $Z^m = (Z_{iq}^m)_{i=1 \dots n_m, q=1 \dots, Q}$. Then the log likelihood is:

$$\ell(\mathbf{X}; \boldsymbol{\theta}_S) = \sum_{m=1}^M \log \int_{Z^m} \exp \{ \ell(X^m | Z^m; \boldsymbol{\alpha}, \boldsymbol{\delta}) + \ell(Z^m; \boldsymbol{\pi}) \} dZ^m, \quad (4)$$

where

$$\begin{aligned} \ell(X^m | Z^m; \boldsymbol{\alpha}, \boldsymbol{\delta}) &= \sum_{i,j=1}^{n_m} \sum_{(q,r) \in \mathcal{Q}_m} Z_{iq}^m Z_{jr}^m \log f(X_{ij}^m; \delta_m \alpha_{qr}), \\ \ell(Z^m; \boldsymbol{\pi}) &= \sum_{i=1}^{n_m} \sum_{q \in \mathcal{Q}_m} Z_{iq}^m \log \pi_q^m \end{aligned}$$

with $\mathcal{Q}_m = \{q \in \{1, \dots, Q\} | \pi_q^m > 0\}$ and f defined as in Equation 3. The log-likelihood functions of the other models can be deduced from this one, setting $\delta_m = 1$ for *iid-colSBM* and π -*colSBM* and $\boldsymbol{\pi}^m = \boldsymbol{\pi}$ for *iid-colSBM* and δ -*colSBM* with S being a matrix of ones (all blocks are represented in each network).

4.1.1 Identifiability

We aim at giving conditions ensuring the identifiability of the models we propose, in the sense that $\forall \mathbf{X}$, $\ell(\mathbf{X}; \boldsymbol{\theta}) = \ell(\mathbf{X}; \boldsymbol{\theta}')$ implies $\boldsymbol{\theta} = \boldsymbol{\theta}'$. The proof relies on the identifiability for the standard SBM demonstrated by [Celisse et al. \(2012\)](#) and is provided in Section 1 of the Supplementary Material. Note that, like any mixture models, all the models are identifiable upto a label switching of the blocks.

Properties 1.

iid-colSBM The parameters $(\boldsymbol{\pi}, \boldsymbol{\alpha})$ are identifiable upto a label switching of the blocks provided that:

- (1.1) $\exists m^* \in \{1, \dots, M\} : n_{m^*} \geq 2Q$,
- (1.2) $(\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_r \forall (q, r) \in \{1, \dots, Q\}^2, q \neq r$.

δ -colSBM The parameters $(\boldsymbol{\pi}, \boldsymbol{\alpha}, \delta_1, \dots, \delta_M)$ are identifiable upto a label switching of the blocks provided that:

- (2.1) $\exists m^* \in \{1, \dots, M\} : n_{m^*} \geq 2Q$ and $\delta_{m^*} = 1$,
- (2.2) $(\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_r \forall (q, r) \in \{1, \dots, Q\}^2, q \neq r$.

π -colSBM Assume that $\forall m = 1, \dots, M, X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}^m, \boldsymbol{\alpha})$. Let $Q_m = |\mathcal{Q}_m| = |\{q = 1 \dots, Q, \pi_q^m > 0\}|$ be the number of non empty blocks in network m . Then the parameters $(\boldsymbol{\pi}^1, \dots, \boldsymbol{\pi}^M, \boldsymbol{\alpha})$ are identifiable upto a label switching of the blocks under the following conditions:

- (3.1) $\forall m \in \{1, \dots, M\} : n_m \geq 2Q_m$,
- (3.2) $\forall m \in \{1, \dots, M\}, (\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_r \forall (q, r) \in \mathcal{Q}_m^2, q \neq r$,
- (3.3) Each diagonal entry of $\boldsymbol{\alpha}$ is unique.

$\delta\pi$ -colSBM Assume that $\forall m = 1, \dots, M, X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}^m, \delta_m \boldsymbol{\alpha})$. Let $Q_m = |\mathcal{Q}_m| = |\{q = 1 \dots, Q, \pi_q^m > 0\}|$ be the number of non empty blocks in network m . Then the parameters $(\boldsymbol{\pi}^1, \dots, \boldsymbol{\pi}^M, \boldsymbol{\alpha}, \delta_1, \dots, \delta_M)$ are identifiable upto a label switching of the blocks under the following conditions:

- (4.1) $\forall m \in \{1, \dots, M\} : n_m \geq 2|\mathcal{Q}_m|$,
- (4.2) $\delta_1 = 1$,

If $Q \geq 2$:

- (4.3) $(\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_r$ for all $(q \neq r) \in \mathcal{Q}_m^2$,
- (4.4) $\forall m \in \{1, \dots, M\}, Q_m \geq 2$,

(4.5) Each diagonal entry of α is unique,

If $Q \geq 3$:

(4.6) There is no configuration of four indices $(q, r, s, t) \in \{1, \dots, Q\}$ such that $\alpha_{qq}/\alpha_{rr} = \alpha_{ss}/\alpha_{tt}$ with $q \neq s$ or $r \neq t$ and with $q \neq r$ or $s \neq t$,

(4.7) $\forall m \geq 2, |\mathcal{Q}_m \cap \cup_{l:l < m} \mathcal{Q}_l| \geq 2$.

5 Inference of the models

5.1 Variational estimation of the parameters

We now tackle the estimation of the parameters $\theta_S \in \Theta_S$ for a given support matrix S . For ease of reading, the index S is dropped in this section. The likelihood given in Equation (S-2) is not tractable in practice, even for a small collection of networks as it relies on summing over $\sum_{m=1}^M |\mathcal{Q}_m|^{n_m}$ terms. A well-proven approach to handle this problem for the inference of the SBM is to rely on a variational version of the EM (VEM) algorithm. The approach is similar for both Bernoulli and Poisson models.

This is done by maximizing a lower (variational) bound of the log-likelihood of the observed data by approximating $p(\mathbf{Z}|\mathbf{X}; \theta)$ with a distribution on \mathbf{Z} named \mathcal{R} issued from a family of factorizable distribution (Daudin et al., 2008):

$$\mathcal{J}(\mathcal{R}, \theta) := \mathbb{E}_{\mathcal{R}}[\ell(\mathbf{X}, \mathbf{Z}; \theta)] + \mathcal{H}(\mathcal{R}(\mathbf{Z})) \leq \ell(\mathbf{X}; \theta),$$

where \mathcal{H} denotes the entropy of a distribution. We further define $\tau_{iq}^m = \mathbb{P}_{\mathcal{R}}(Z_{iq}^m = 1)$ as the parameters of the variational distribution.

The VEM algorithm is a two-step iterative procedure which alternates the variational E step and the M step. The E step consists in optimizing $\mathcal{J}(\mathcal{R}, \theta)$ for a current parameter value θ with respect to \mathcal{R} constrained to be in the family of factorizable distributions and the M step consists in maximizing $\mathcal{J}(\mathcal{R}, \theta)$ with respect to θ for a given variational distribution \mathcal{R} . For the *colSBM* models, networks can be treated independently during the VE-step, while the M-step serves as a link between the structures of the networks in the collection. For δ -*colSBM* and $\delta\pi$ -*colSBM* when $\mathcal{F} = \mathcal{Bernoulli}$, numerical approximations are needed as no explicit expression of $\hat{\delta}$ and $\hat{\alpha}$ can be derived. Further details of the variational procedure and the expression of the parameters estimators are provided in Section 2 of the Supplementary Material.

5.2 Model selection

There are two model selection issues. First, under a fixed *colSBM*, we aim to choose Q and determine the support matrix S for π -*colSBM* and $\delta\pi$ -*colSBM*. This task is tackled in Subsection 5.2.1 by introducing a penalized likelihood criterion. Second, the comparison of the the *colSBM* models –each one introducing various degrees of consensus between the networks– with the *sep-SBM* – which assumes that each network has its own structure– is dealt with in Subsection 5.2.2.

5.2.1 Selecting the number of blocks Q

A classical tool to choose the number of blocks in the SBM context is the Integrated Classified Likelihood (ICL) proposed by [Biernacki et al. \(2000\)](#); [Daudin et al. \(2008\)](#). ICL derives from an asymptotic approximation of the marginal complete likelihood $m(\mathbf{X}, \mathbf{Z}) = \int_{\boldsymbol{\theta}} \exp\{\ell(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})\}p(\boldsymbol{\theta})d\boldsymbol{\theta}$ where the parameters are integrated out against a prior distribution, resulting in a penalized criterion of the form $\max_{\boldsymbol{\theta}} \ell(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta}) - \frac{1}{2}\text{pen}$. In the ICL, the latent variables \mathbf{Z} are integrated out against an approximation of $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})$ obtained via the variational approximation. This leads to the following expression

$$\text{ICL} = \max_{\boldsymbol{\theta}} \mathbb{E}_{\widehat{\mathcal{R}}} [\ell(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})] - \frac{1}{2}\text{pen}$$

Using the fact that $\mathbb{E}_{\widehat{\mathcal{R}}} [\ell(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})] \approx \ell(\mathbf{X}; \boldsymbol{\theta}) - \mathcal{H}(\widehat{\mathcal{R}})$, one understands that, as emphasized in the literature, ICL favors well separated blocks by penalizing for the entropy of the clustering. However, in this work, our goal is not only to cluster the nodes into coherent blocks but also to evaluate the similarity of the connectivity patterns between the different networks. As such we would like to authorize models providing clustering that may be more fuzzy by not penalizing for the entropy. This leads to a BIC-like criterion of the form:

$$\text{BIC-L} = \max_{\boldsymbol{\theta}} \mathbb{E}_{\widehat{\mathcal{R}}} [\ell(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta})] + \mathcal{H}(\widehat{\mathcal{R}}) - \frac{1}{2}\text{pen} = \max_{\boldsymbol{\theta}} \mathcal{J}(\widehat{\mathcal{R}}, \boldsymbol{\theta}) - \frac{1}{2}\text{pen}$$

We now supply the expression of the penalty term for the four models we proposed and discuss possible variations of the criterion.

Selection of Q for *iid-colSBM* and δ -*colSBM*

For *iid-colSBM* and δ -*colSBM*, the derivation of the penalty is a straightforward extension of the classical SBM model, leading to:

$$\text{BIC-L}(\mathbf{X}, Q) = \max_{\boldsymbol{\theta}} \mathcal{J}(\widehat{\mathcal{R}}, \boldsymbol{\theta}) - \frac{1}{2} [\text{pen}_{\pi}(Q) + \text{pen}_{\alpha}(Q) + \text{pen}_{\delta}(Q)],$$

where

$$\begin{aligned} \text{pen}_{\pi}(Q) &= (Q - 1) \log \left(\sum_{m=1}^M n_m \right), \\ \text{pen}_{\alpha}(Q) &= Q^2 \log(N_M), \\ \text{pen}_{\delta}(Q) &= \begin{cases} 0 & \text{for } \textit{iid-colSBM} \\ (M - 1) \log(N_M) & \text{for } \delta\text{-colSBM} \end{cases} \end{aligned}$$

where

$$N_M = \sum_{m=1}^M n_m(n_m - 1) \tag{5}$$

is the number of possible interactions. The first term $\text{pen}_{\pi}(Q)$ corresponds to the clustering part where the $Q - 1$ block proportions have to be estimated from the $\sum_{m=1}^M n_m$ nodes. The terms $\text{pen}_{\alpha}(Q)$ and $\text{pen}_{\delta}(Q)$ are linked to the connection parameters. Finally, Q is chosen as:

$$\widehat{Q} = \operatorname{argmax}_{Q \in \{1, \dots, Q_{\max}\}} \text{BIC-L}(\mathbf{X}, Q).$$

Selection of Q for π -colSBM and $\delta\pi$ -colSBM

Here, in addition to the choice of Q , the collection of support matrices S is considered. In order to penalize the complexity of the model space, we introduce a prior distribution on S defined as follows. Let us introduce $Q_m = \sum_{q=1}^Q S_{mq}$ the number of blocks represented in network m . Assuming independent uniform prior distributions on the (Q_m) 's and a uniform prior distribution on S for fixed numbers of blocks Q_1, \dots, Q_M represented in each network, we obtain the following prior distribution on S :

$$\log p_Q(S) = -M \log(Q) - \sum_{m=1}^M \log \binom{Q}{Q_m}$$

where $\binom{Q}{Q_m}$ is the number of choices of Q_m non-empty blocks among the Q possible blocks in network m . Now, combining the Laplace asymptotic approximation of the marginal complete likelihood (where the parameters have been integrated out) and introducing the prior distribution on S , we obtain the following penalized criterion:

$$\text{BIC-L}(\mathbf{X}, Q) = \max_S \left[\max_{\boldsymbol{\theta}_S \in \Theta_S} \mathcal{J}(\widehat{\mathcal{R}}, \boldsymbol{\theta}_S) - \frac{\text{pen}_\pi(Q, S) + \text{pen}_\alpha(Q, S) + \text{pen}_\delta(Q, S) + \text{pen}_S(Q)}{2} \right],$$

where

$$\begin{aligned} \text{pen}_\pi(Q, S) &= \sum_{m=1}^M (Q_m - 1) \log(n_m), & \text{pen}_\alpha(Q, S) &= \left(\sum_{q,r=1}^Q \mathbf{1}_{(S'_{qr}) > 0} \right) \log(N_M), \\ \text{pen}_\delta(Q, S) &= \begin{cases} 0 & \text{for } \pi\text{-colSBM} \\ (M - 1) \log(N_M) & \text{for } \delta\pi\text{-colSBM} \end{cases}, \\ \text{pen}_S(Q) &= -2 \log p_Q(S). \end{aligned}$$

where N_M has been defined in Equation (5). Finally, Q is chosen such that:

$$\widehat{Q} = \operatorname{argmax}_{Q \in \{1, \dots, Q_{\max}\}} \text{BIC-L}(\mathbf{X}, Q).$$

The details about the derivation of this criterion are provided in Section 3 of the Supplementary Material.

Practical model selection

The practical choice of Q and the estimation of its parameters are computationally intensive tasks. Indeed, we should compare all the possible models through the chosen model selection criterion. Furthermore, for each model, the variational EM algorithm should be initialized at a large number of initialization points (due to its sensitivity to the starting point), resulting in an unreasonable computational cost. Instead, we propose to adopt a stepwise strategy, resulting in a faster exploration of the model space, combined with efficient initializations of the variational EM algorithm. The procedure we suggest is given in Algorithm 1 and is implemented in an R-package colSBM available on GitHub: <https://github.com/Chabert-Liddell/>

colSBM. For initializing a *colSBM* with Q blocks, we start from fitting a *sep-SBM*, then, the Q blocks of the M networks have to be associated. This association can be done in many ways due to label switching within each network which provides us with a lot of possible initializations. Then, the stepwise procedure explores the possible number of blocks by building on the previously fitted models. Note that when fitting the π -*colSBM* or the $\delta\pi$ -*colSBM*, the support S has to be determined which is done through an extra-step that consists in thresholding the parameters π^m related with the block proportions leading to an exploration over the set \mathcal{S}_Q .

Algorithm 1: Model selection algorithm

Data: \mathbf{X} a collection of networks.

begin initialization

- Infer *sep-SBM* on \mathbf{X} , with $Q \in [Q_{min}, Q_{max}]$
- Get $\hat{Z}_{sep-SBM}^m(Q)$
- Fit *colSBMs* with VEM starting from merged $\hat{Z}_{sep-SBM}^m(Q)$ (many initializations as a result of permutations within each $\hat{Z}_{sep-SBM}^m(Q)$)
- Keep the b fitted models with the best BIC-L for each Q

while *BIC-L is increasing* **do**

- Forward loop
- for** $Q = Q_{min} + 1, \dots, Q_{max}$ **do**
 - Fit *colSBM* with Q blocks from initializations obtained by splitting a block in models with $Q - 1$ blocks
 - if** π -*colSBM* or $\delta\pi$ -*colSBM* **then**
 - Fit *colSBM* with $\hat{S}_{qm} = \mathbf{1}_{\hat{\pi}_q^m > t}$ for different value of threshold t
- Backward loop
- for** $Q = Q_{max} - 1, \dots, Q_{min}$ **do**
 - Fit *colSBM* with Q blocks from initializations obtained by merging two blocks in models with $Q + 1$ blocks
 - if** π -*colSBM* or $\delta\pi$ -*colSBM* **then**
 - Fit *colSBM* with $\hat{S}_{qm} = \mathbf{1}_{\hat{\pi}_q^m > t}$ for different value of threshold t
- Among all fitted models, keep the b fitted models with the highest BIC-L for each Q

return $\hat{Q} = \arg \max BIC-L(\mathbf{X}, Q)$, with the corresponding $\hat{\theta}$, $\hat{\mathbf{Z}}$ and \hat{S} for π -*colSBM* and $\delta\pi$ -*colSBM*.

5.2.2 Testing common connectivity structure

We can also use a model selection approach to choose which model from the 4 *colSBMs* and the *sep-SBM* is the most adapted to the collection. The most interesting comparison is to decide whether a collection of networks share the same connectivity structure by comparing the model selection criteria obtained for a given *colSBM* model with the one of *sep-SBM*. We decide that a collection of networks

share the same connectivity structure if:

$$\max_Q \text{BIC-L}_{colSBM}(\mathbf{X}, Q) > \sum_{m=1}^M \max_{Q_m} \text{BIC-L}_{SBM}(X^m, Q_m).$$

5.3 Simulation studies

In Section 4 of the Supplementary Material, we perform a large simulation study in order to test the inference and model selection procedures proposed in this section. More specifically, simulating from a π -*colSBM* model, for various strengths of connectivity structures, we look at our capacity to recover the true connectivity parameter α as well as clustering of the nodes and the true support S . We also assess the quality and the limit of the model selection with the BIC-L criteria. First by testing the ability to select the true number of blocks and to distinguish π -*colSBM* from *iid-colSBM* and *sep-SBM*, and second by varying the block proportions and studying its influence on model comparison and the selection of the number of blocks and their support.

We perform another simulation study to understand how, for particular configuration, using a *colSBM* model on a collection of networks favors the transfer of information between networks and allows to find finer block structures on the networks.

6 Partition of networks according to their mesoscale structure

If the networks in a collection do not have the same connectivity structure, we aim to partition them accordingly. We present hereafter a strategy to perform a clustering of the networks.

Clustering a collection of networks consists in finding a partition $\mathcal{G} = (\mathcal{M}_g)_{g=1, \dots, G}$ of $\{1, \dots, M\}$. Given \mathcal{G} , we set the following model on \mathbf{X} :

$$\forall g \in \{1, \dots, G\}, \quad \forall m \in \mathcal{M}_g, \quad X^m \sim \mathcal{F}\text{-SBM}(Q^g, \boldsymbol{\pi}^m, \delta_m \boldsymbol{\alpha}^g)$$

with $\delta_1 = 1$. Moreover, $\delta_m = 1$ for all m for *iid-colSBM* and π -*colSBM*s and $\boldsymbol{\pi}^m = \boldsymbol{\pi}^g$ for *iid-colSBM* and δ -*colSBM*. In other words, the networks belonging to the subcollection \mathcal{M}_g share the same mesoscale structure given by a particular *colSBM*. To any partition \mathcal{G} we associate the following score:

$$\text{Sc}(\mathcal{G}) = \sum_{g=1}^G \max_{Q^g=1, \dots, Q_{\max}} \text{BIC-L}((X^m)_{m \in \mathcal{M}_g}, Q^g). \quad (6)$$

where $\text{BIC-L}((X^m)_{m \in \mathcal{M}_g}, Q^g)$ is the BIC-L computed on the \mathcal{M}_g subcollection of networks. The best partition \mathcal{G} is chosen as follows:

$$\mathcal{G}^* = \operatorname{argmax}_{\mathcal{G}} \text{Sc}(\mathcal{G}).$$

Computing the BIC-L for all the partitions \mathcal{G} requires to consider the $2^M - 1$ non-empty subcollections of the networks \mathcal{M} , fit the *colSBMs* on these subcollections and then combine the associated BIC-L in order to be able to compute the scores (6). This can be done exhaustively provided that M is not too large but the computational cost becomes prohibitive as M grows.

To circumvent this point, we propose a less computationally intensive forward strategy, starting from $\mathcal{G} = (\{1, \dots, M\})$, and then progressively splitting the collection of networks. In order to explore the space of partitions of $\{1, \dots, M\}$, we define a dissimilarity measure between any pair of networks (m, m') .

Definition of a dissimilarity measure between networks of a collection $(X^m)_{m \in \mathcal{M}}$. This relies on the following steps.

1. Infer *colSBM* on $(X^m)_{m \in \mathcal{M}}$ to get coherent clusterings of the nodes encoded in $\hat{\tau}^m = (\hat{\tau}_{iq}^m)_{i=1, \dots, n_m, q=1, \dots, \hat{Q}}$, for any $m \in \mathcal{M}$, where τ_{iq}^m is defined as in Section B. This step supplies the clusterings of the nodes of each network in terms of mesoscale structure. Note that the inference also supplies the $(\hat{\delta}^m)_{m \in \mathcal{M}}$, these quantities being set to 1 if we work with the π -*colSBM* and *iid-colSBM*.
2. For each network m , compute:

$$\tilde{n}_{qr}^m = \sum_{\substack{i,j=1 \\ i \neq j}}^{n_m} \hat{\tau}_{iq}^m \hat{\tau}_{jr}^m, \quad \tilde{\alpha}_{qr}^m = \frac{\sum_{\substack{i,j=1 \\ i \neq j}}^{n_m} \hat{\tau}_{iq}^m \hat{\tau}_{jr}^m X_{ij}^m}{\tilde{n}_{qr}^m}, \quad \tilde{\pi}_q^m = \frac{\sum_{i=1}^{n_m} \hat{\tau}_{iq}^m}{n_m}$$

with the convention that $\tilde{\pi}_q^m = 0$ if $q \notin \mathcal{Q}_m$ and $\alpha_{qr}^m = 0$ if $\{q, r\} \not\subset \mathcal{Q}_m$. These quantities are the separated estimates of the parameters encoding the mesoscale structure for each network, computed from node clusterings obtained by considering all the networks jointly.

3. Then, for any pair of networks $(m, m') \in \mathcal{M}^2$ compute the dissimilarity:

$$D_{\mathcal{M}}(m, m') = \sum_{(q,r)=1}^Q \max(\tilde{\pi}_q^m, \tilde{\pi}_q^{m'}) \max(\tilde{\pi}_r^m, \tilde{\pi}_r^{m'}) \left(\frac{\tilde{\alpha}_{qr}^m}{\hat{\delta}_m} - \frac{\tilde{\alpha}_{qr}^{m'}}{\hat{\delta}_{m'}} \right)^2. \quad (7)$$

This dissimilarity measure quantifies to what extent the connectivity parameters inferred separately on each network of the pair are different. This is weighted by the size of the blocks and corrected in the case of δ -*colSBM* and $\delta\pi$ -*colSBM* by the density parameter. If this dissimilarity measure is large, it means that enforcing the same connectivity patterns by estimating common connectivity parameters for these two networks is not relevant and the networks cannot be considered to be part of the same group.

An algorithm to cluster the collection of networks Now, we use this dissimilarity to guide the search for the best partition of the collection of networks by using Algorithm 2 which consists in a recursive partitioning of the collection.

Algorithm 2: Clustering a collection of networks into two groups

Call: `Clust2Coll`($\mathbf{X} = (X^m)_{m \in \mathcal{M}}$)

Data: $\mathbf{X} = (X^m)_{m \in \mathcal{M}}$ a collection of networks and $\mathcal{G} = \{\mathcal{M}\}$ the trivial partition in a unique group

begin

 - Fit *colSBM* on \mathbf{X}

 - Compute the score $Sc_0 = Sc(\mathcal{G})$

 - Compute the dissimilarity for all the networks in the collection

$(D_{\mathcal{M}}(m, m'))_{m, m' \in \mathcal{M}}$

 - Apply a 2-medoids algorithm to obtain G_1 and G_2 giving a partition of \mathcal{M} .

 - Compute $Sc^* = Sc(\mathcal{G}^*)$ where $\mathcal{G}^* = \{G_1, G_2\}$.

if $Sc_0 > Sc^*$ **then**

 | **return** \mathcal{G}

else

 | **return** $\{\text{Clust2Coll}((X^m)_{m \in G_1}), \text{Clust2Coll}((X^m)_{m \in G_2})\}$

Simulation study We illustrate our capacity to perform a partition of a collection of networks based on their structure for all 4 *colSBM* models in Section 4 of the Supplementary Material.

7 Application to Food Webs

In this section, we demonstrate the interest of our models on 2 collections of ecological networks. The first one consists of a collection of 3 stream food webs issued from the dataset of [Thompson and Townsend \(2003\)](#) described in Section 2 at page 4. We analyze in detail the different structures given by the different models and we show how using networks with some common structure helps the prediction of missing information in the networks.

The second dataset is a collection of 67 networks issued from the Mangal database ([Vissault et al., 2020](#)). We will use our model to propose a partition of the networks into groups of networks with common mesoscale structures.

7.1 Joint analysis of 3 stream food webs

In Section 2 at page 5, we fitted *sep-SBM* obtained 5 blocks for Martins, 3 blocks for Cooper and 4 blocks for Herlzier. For reminder, a matricial representation of the block reordered food webs was shown in Figure 1. Each food web comprises of 2 blocks of basal species (the 2 bottom blocks).

Finding a common structure between the networks We now fit the four *colSBM* models in order to find a common structure among the 3 networks. First, notice that our model selection criterion greatly favors common network structure

above separated one: BIC-L = -2080 for *sep-SBM* versus respectively -1964 , -1983 , -1970 and -1988 for *iid-colSBM*, π -*colSBM*, δ -*colSBM*, $\delta\pi$ -*colSBM*. The structures of the collection under the different models are represented in Figure 2. In this figure, the square matrix represents the estimated connection matrix $\hat{\alpha}$, while the cumulative bar plot on the right represents the $\tilde{\pi}^m$ (see definition page 16).

For each model, the basal species are separated into 2 blocks (bottom blue blocks in Figure 2), similar to the one obtained with the SBMs. The obtained structure slightly differs as described hereafter.

- *iid-colSBM* highlights 5 blocks in total. Block 3 (light green) is a small block of intermediate trophic level species (ones that prey on basal species and are being preyed on by higher trophic levels) with some within block predation. The higher trophic level is divided into 2 more blocks, block 2 (dark green) only preys on the 2 basal blocks, while block 1 (pink) preys on the intermediate block 3 level but only on the most connected basal species block.
- π -*colSBM* leads also to 5 blocks. There are no empty blocks and the block proportions are roughly corresponding to the ones of *iid-colSBM*. This renders the flexibility of the π -*colSBM* of little use compared to the *iid-colSBM* on this collection.
- With δ -*colSBM*, the species are clustered into 6 blocks and the networks have different estimated density parameters: $\hat{\delta} \approx (1, .9, 1.2)$. Block 1 (red), 2 (pink) and 3 (dark green) correspond to the top trophic levels. Block 4 (light green) is an intermediate trophic level group, well connected with both block 1 and the basal species blocks. Block 2 (pink) is huge and only preys on block 6 while block 3 (dark green) is a small group of species that preys on both basal species blocks.
- Finally, $\delta\pi$ -*colSBM* clusters the species into 5 blocks and the networks have different estimated density parameters: $\hat{\delta} \approx (1, .9, 1.4)$. The connectivity structure is almost similar to the one of π -*colSBM* but the blocks have different proportions. Block 1 (pink) corresponds to block 1 (red) and 2 (pink) of δ -*colSBM* and is the merge of two top trophic levels. Again, the block proportions are still quite homogeneous between networks, and the added flexibility of $\delta\pi$ -*colSBM* compared to δ -*colSBM* is not used.

Remark 1. *On this collection, the entropy of the clustering is much lower on $\delta\pi$ -colSBM than on δ -colSBM. In this model, to ensure homogeneous block proportions between networks, some nodes tend to get a fuzzy clustering and sit between several blocks (the variational parameters do not concentrate on one block). This phenomenon is taken into account by our model selection criterion which tends to favor models with higher entropy than models with well separated clusters.*

Prediction of missing links and dyads Since we have been able to find some common structures between the 3 networks, we now examine if these structures

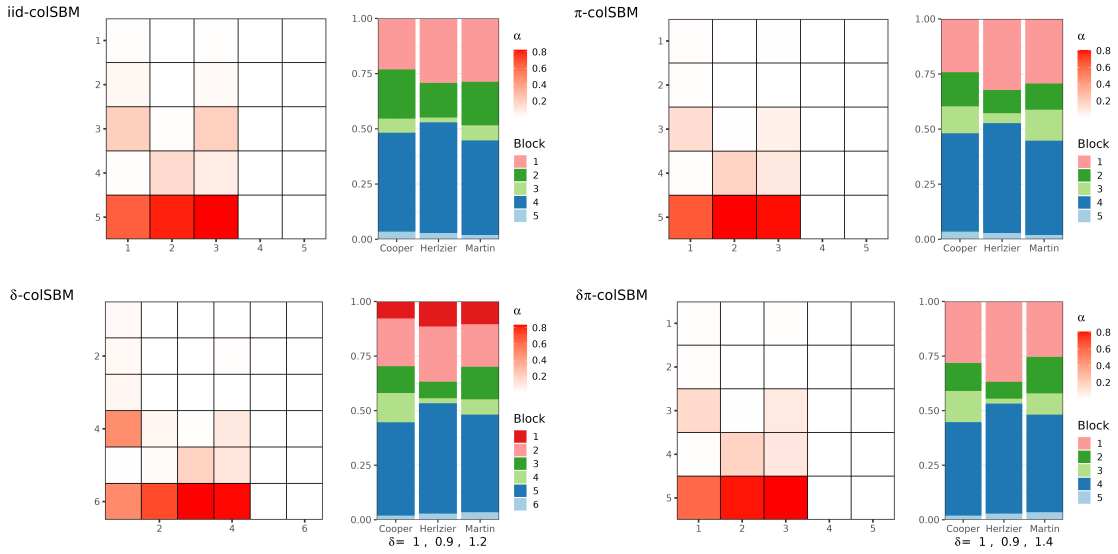


Figure 2: **Estimated structure of the collection of 3 stream food webs with the four $colSBM$ models.** For each network, the matrix on the left is the estimated parameter connectivity $\hat{\alpha}$, the barplot on the right depicts $\hat{\pi}^{(m)}$. The ordering is done by trophic level from bottom to top and right to left. For $(\delta-\delta\pi)colSBMs$ we give $\hat{\delta}$ below the barplot.

could be used to help recover some information on networks with incomplete information. We proceed as follows: we choose a network m and remove a proportion $K \in [.1, .8]$ of

- the existing links uniformly at random *for the missing link experiment*
- or of the existing dyads (both 0 and 1) by encoding them as NA *for the missing dyads experiment*.

Then, for the missing link experiment, we try to recover where the missing links are among all non existing ones. For the missing dyad experiment, we predict the probability of existence of missing dyads (NA entries). Under the $colSBM$, the probability of a link between species i and j for network m is predicted by:

$$\hat{p}_{ij}^m = \sum_{q,r \in \hat{Q}_m} \hat{\tau}_{iq}^m \hat{\tau}_{jr}^m \hat{\delta}^m \hat{\alpha}_{qr},$$

where τ_{iq}^m and τ_{jr}^m are defined as in Section B. We resort to the area under the ROC curve to evaluate the capacity of the different models to recover this information. For each value of K , each experiment is repeated 30 times and the results are shown in Figure 3.

First, let us notice that these stream food webs networks have a structure that is well explained by an SBM. When there is little information missing ($K < .3$) the ROC AUC is over 0.9. Besides, with 70% of missing links or dyads, we still predict better than a random guess (ROC AUC $\approx .75$). As there is a common structure between the 3 networks, there is a lot of information to be taken from the

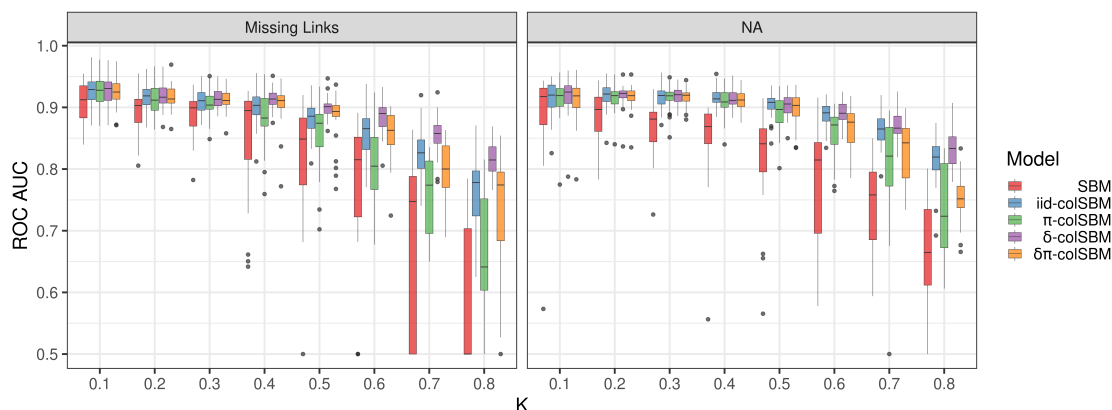


Figure 3: Prediction of missing links and NA entries on stream food webs.

ones with no missing information. Starting from $K \geq .2$, the *colSBMs* outperform the *sep-SBM* on both experiments. Even for $K = .8$, the prediction is still high. About the difference between the *colSBMs*, for the missing links experiment, as we remove links from one of the network, its density decreases and the models with a density parameters, δ -*colSBM* and $\delta\pi$ -*colSBM* have a built-in mechanism that compensates this fact. As a consequence, these models yield to better predictions than *iid-colSBM* or π -*colSBM* for large values of K .

Another noteworthy comment on both the NA and missing links experiments is that as K grows, the amount of information on the modified network gets lower. Hence, π -*colSBM* and $\delta\pi$ -*colSBM* lack the statistical power to separate blocks and will empty blocks on this network. This affects our capability to predict the trophic links. On the other hand, *iid-colSBM* and δ -*colSBM* will force some separation of the species into blocks, and as the information from the other networks is relevant it still has good prediction performance for large K .

7.2 Partition of a collection of 67 predation networks

Now, we consider a collection of 67 predation networks which are all directed networks with more than 30 species, from the Mangal database (Vissault et al., 2020). They are issued from 33 datasets each containing between 1 and 10 networks. The number of species ranges from 31 to 106 (3395 in total) by networks; the networks have density ranging from .01 to .32 (14934 total predation links). This dataset is too heterogeneous to find a common structure that will fit well on all the networks. Therefore we propose to use a π -*colSBM* to look for a partition of the networks into groups sharing common connectivity structures. We focus our investigations on this model since we aim to group together networks which share some blocks with similar features but we do not expect that all the networks in the same group share exactly the same blocks.

We present the obtained network clustering and the connectivity structure of each group in Figure 4 as well as a contingency table of the obtained groups with the different datasets of the Mangal database. Our comments on each group follow.

A This group consists of 7 networks and 12 blocks are required to describe this

cluster. 5 networks are issued from the same dataset (id: 80). These 5 networks populate the 12 blocks, while the other 2 networks only populate parts of them. The average density is about 0.18. From the ecological point of view, the blocks can be divided into 3 heterogeneous sets: block 1 to 3 represent the higher trophic levels, block 4 to 8 the intermediate ones and block 9 to 12 the lower ones.

- B** A group of 26 networks with heterogeneous size and density issued from various datasets. Most networks populate only parts of the 8 blocks. The structure is mainly guided by blocks 2, 3, 5 and 6, from higher to lower trophic levels. Block 4 is represented in only 5 networks where it is either an intermediate or a bottom trophic level. It introduces some symmetry in the connectivity matrix rendering it difficult to order the blocks by trophic order. Species from top trophic levels prey on basal species.
- C** A small group of 6 networks with density ranging from .06 to .11. All networks are represented in 5 or 6 of the 7 blocks, including the first three blocks. The group consists of 3 of the 5 networks of dataset 48, the separation being based on the collecting sites. The top trophic level is divided into 2 blocks, species from those blocks preying only on intermediate trophic level species. One can exhibit two different trophic chains: species from block 2 prey on species from block 4, which prey more on basal species (block 7) than on others intermediate trophic species (block 6), while species from block 1 prey on species from block 3 and 4, block 3 exhibiting the inverse behavior of block 4.
- D** Another heterogeneous group of 23 networks. The 10 networks from dataset 157 (stream food webs from New Zealand) are divided between groups **B** and **D** based on the type of ecosystem. The data from group **B** were collected in creeks, while the one from group **D** were collected on streams. Compared to the other heterogeneous group (**B**), the top trophic species (block 1, 2 and 3) prey mostly on intermediate trophic levels (block 4 and 5). Species from block 4 prey on species from the other intermediate block and the 3 blocks of bottom trophic level species (blocks 6, 7 and 8), while species from block 5 just prey on species from the last two blocks.
- E** The last group consists of 6 networks from various datasets. In the 7 blocks structure, the species of block 1 (represented on 4 of the 6 networks) prey on species from all other blocks with the exception of block 7. The basal species are separated between blocks 6 and 7 depending on whether or not they are preyed by species from the first two blocks.

8 Discussion

In this paper, we proposed a new method to find a common structure and compare different structures of a collection of networks which we do not assume to

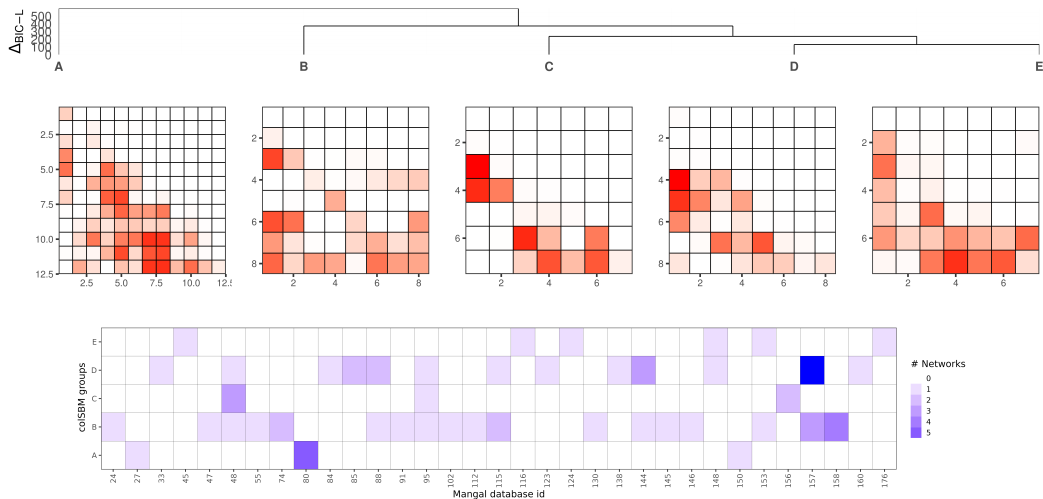


Figure 4: Above: Classification and connectivity structures of a collection of 67 predation networks from the Mangal database into 5 groups. The length of the dendrogram is given by the difference in BIC-L to the best model. Below: Contingency table of the classification found by $\pi\text{-colSBM}$ and the different datasets from the Mangal database.

share common nodes. This method is very general and could be applied to networks sharing common nodes as well, such as temporal or multiplex networks. Starting from our most basic model, – namely the *iid-colSBM* – we refined it by proposing models allowing for different mixture distributions and even empty blocks ($\pi\text{-colSBM}$), models allowing to find common structure for networks of different density ($\delta\text{-colSBM}$) or even models allowing both ($\delta\pi\text{-colSBM}$). The model selection criterion we derived can be used to select the number of blocks but also to choose which *colSBM* fits better the data. We also presented a strategy providing a partition of a collection of networks into groups of networks sharing common connectivity patterns.

The idea behind these models is very general and could be extended to other types of networks. In ecology, bipartite and multipartite networks are common and the model extension (Govaert and Nadif, 2003; Bar-Hen et al., 2020) is straightforward (although some additional modeling choices arise when considering $\pi\text{-colSBM}$, $\delta\text{-colSBM}$ or $\delta\pi\text{-colSBM}$), the main difficulty would then lie in the algorithmic part. Additionally, incorporating the type of ecological interaction as a network covariate (Mariadassou et al., 2010) would help us understand its impact on the structure of the networks and to the robustness of the ecosystems they depict (Chabert-Liddell et al., 2022). The main idea of this article could also be extended to the Degree Corrected *SBM* (Karrer and Newman, 2011) which is quite used in practice.

Finally, we notice during our simulations and applications that the *colSBMs* allow to find a larger number of blocks compared to the *sep-SBM* and so lead to a finer resolution of the mesoscale structure of the networks. This resolution limit problem was one of the motivations of Peixoto (2014) and we believe that this

direction should be explored further for collections of networks.

Acknowledgments

The authors would like to thank Stéphane Robin for his helpful advice. This work was supported by a public grant as part of the Investissement d’avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH. This work was partially supported by the grant ANR-18-CE02-0010-01 of the French National Research Agency ANR (project EcoNet).

References

- Allesina, S. and Pascual, M. (2009). Food web models: a plea for groups. *Ecology letters*, 12(7):652–662.
- Bar-Hen, A., Barbillon, P., and Donnet, S. (2020). Block models for generalized multipartite networks: Applications in ecology and ethnobiology. *Statistical Modelling*, page 1471082X20963254.
- Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence*, 22(7):719–725.
- Blüthgen, N., Menzel, F., and Blüthgen, N. (2006). Measuring specialization in species interaction networks. *BMC ecology*, 6(1):1–12.
- Celisse, A., Daudin, J.-J., and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847–1899.
- Chabert-Liddell, S.-C., Barbillon, P., and Donnet, S. (2022). Impact of the mesoscale structure of a bipartite ecological interaction network on its robustness through a probabilistic modeling. *Environmetrics*, 33(2):e2709.
- Chabert-Liddell, S.-C., Barbillon, P., Donnet, S., and Lazega, E. (2021). A stochastic block model approach for the analysis of multilevel networks: An application to the sociology of organizations. *Computational Statistics & Data Analysis*, 158:107179.
- Chiquet, J., Donnet, S., and Barbillon, P. (2021). *sbm: Stochastic Blockmodels*. R package version 0.4.3.
- Clauset, A., Moore, C., and Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98.
- Daudin, J.-J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Statistics and computing*, 18(2):173–183.

- Donnat, C. and Holmes, S. (2018). Tracking network dynamics: A survey using graph distances. *The Annals of Applied Statistics*, 12(2):971–1012.
- Govaert, G. and Nadif, M. (2003). Clustering with block mixture models. *Pattern Recognition*, 36(2):463–473.
- Guimerà, R. and Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, 106(52):22073–22078.
- Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.
- Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193–218.
- Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107.
- Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3):203–271.
- Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer Publishing Company, Incorporated, 1st edition.
- Le, C. M., Levin, K., and Levina, E. (2018). Estimating a network from multiple noisy realizations. *Electronic Journal of Statistics*, 12(2):4697–4740.
- Leger, J.-B., Barbillon, P., and Chiquet, J. (2020). *blockmodels: Latent and Stochastic Block Model Estimation by a 'V-EM' Algorithm*. R package version 1.1.4.
- Mariadassou, M., Robin, S., and Vacher, C. (2010). Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics*, 4(2):715–742.
- Matias, C. and Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1119–1141.
- Michalska-Smith, M. J. and Allesina, S. (2019). Telling ecological networks apart by their structure: A computational challenge. *PLOS Computational Biology*, 15(6):e1007076.
- Mukherjee, S. S., Sarkar, P., and Lin, L. (2017). On clustering network-valued data. *Advances in neural information processing systems*, 30.
- Ohlsson, M. and Eklöf, A. (2020). Spatial resolution and location impact group structure in a marine food web. *Ecology Letters*, 23(10):1451–1459.
- Paul, S. and Chen, Y. (2018). A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. *Annals of Applied Statistics*, 14(2):993–1029.

- Pavlović, D. M., Guillaume, B. R., Towlson, E. K., Kuek, N. M., Afyouni, S., Vértes, P. E., Yeo, B. T., Bullmore, E. T., and Nichols, T. E. (2020). Multi-subject Stochastic Blockmodels for adaptive analysis of individual differences in human brain network cluster structure. *NeuroImage*, 220:116611.
- Peel, L., Larremore, D. B., and Clauset, A. (2017). The ground truth about meta-data and community detection in networks. *Science advances*, 3(5):e1602548.
- Peixoto, T. P. (2014). Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1):011047.
- Reyes, P. and Rodriguez, A. (2016). Stochastic blockmodels for exchangeable collections of networks. *arXiv preprint arXiv:1606.05277*.
- Rivera-Hutinel, A., Bustamante, R., Marín, V., and Medel, R. (2012). Effects of sampling completeness on the structure of plant–pollinator networks. *Ecology*, 93(7):1593–1603.
- Signorelli, M. and Wit, E. C. (2020). Model-based clustering for populations of networks. *Statistical Modelling*, 20(1):9–29.
- Snijders, T. A. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, 14(1):75–100.
- Stanley, N., Shai, S., Taylor, D., and Mucha, P. J. (2016). Clustering network layers with the strata multilayer stochastic block model. *IEEE transactions on network science and engineering*, 3(2):95–105.
- Sweet, T. M., Thomas, A. C., and Junker, B. W. (2014). *Hierarchical mixed membership stochastic blockmodels for multiple networks and experimental interventions*, chapter chapter22. CRC Press.
- Thompson, R. M. and Townsend, C. R. (2003). Impacts on stream food webs of native and exotic forest: an intercontinental comparison. *Ecology*, 84(1):145–161.
- Vissault, S., Cazelles, K., Bergeron, G., Mercier, B., Violet, C., Gravel, D., and Poisot, T. (2020). *rmangal: An R package to interact with Mangal database*. R package version 2.0.2.
- White, H. C., Boorman, S. A., and Breiger, R. L. (1976). Social structure from multiple networks. i. blockmodels of roles and positions. *American journal of sociology*, 81(4):730–780.

Supplementary Material

A Proof of the identifiability of colSBMs

Properties 1.

iid-colSBM The parameters $(\boldsymbol{\pi}, \boldsymbol{\alpha})$ are identifiable upto a label switching of the blocks provided that:

$$(1.1) \exists m^* \in \{1, \dots, M\} : n_{m^*} \geq 2Q,$$

$$(1.2) (\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_r \quad \forall (q, r) \in \{1, \dots, Q\}^2, q \neq r.$$

δ -colSBM The parameters $(\boldsymbol{\pi}, \boldsymbol{\alpha}, \delta_1, \dots, \delta_M)$ are identifiable upto a label switching of the blocks provided that:

$$(2.1) \exists m^* \in \{1, \dots, M\} : n_{m^*} \geq 2Q \text{ and } \delta_{m^*} = 1,$$

$$(2.2) (\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi})_r \quad \forall (q, r) \in \{1, \dots, Q\}^2, q \neq r.$$

π -colSBM Assume that $\forall m = 1, \dots, M, X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}^m, \boldsymbol{\alpha})$. Let $Q_m = |\mathcal{Q}_m| = |\{q = 1 \dots, Q, \pi_q^m > 0\}|$ be the number of non empty blocks in network m . Then the parameters $(\boldsymbol{\pi}^1, \dots, \boldsymbol{\pi}^M, \boldsymbol{\alpha})$ are identifiable upto a label switching of the blocks under the following conditions:

$$(3.1) \forall m \in \{1, \dots, M\} : n_m \geq 2Q_m,$$

$$(3.2) \forall m \in \{1, \dots, M\}, (\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_r \quad \forall (q, r) \in \mathcal{Q}_m^2, q \neq r,$$

(3.3) Each diagonal entry of $\boldsymbol{\alpha}$ is unique.

$\delta\pi$ -colSBM Assume that $\forall m = 1, \dots, M, X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \boldsymbol{\pi}^m, \delta_m \boldsymbol{\alpha})$. Let $Q_m = |\mathcal{Q}_m| = |\{q = 1 \dots, Q, \pi_q^m > 0\}|$ be the number of non empty blocks in network m . Then the parameters $(\boldsymbol{\pi}^1, \dots, \boldsymbol{\pi}^M, \boldsymbol{\alpha}, \delta_1, \dots, \delta_M)$ are identifiable upto a label switching of the blocks under the following conditions:

$$(4.1) \forall m \in \{1, \dots, M\} : n_m \geq 2|\mathcal{Q}_m|,$$

$$(4.2) \delta_1 = 1,$$

If $Q \geq 2$:

$$(4.3) (\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_q \neq (\boldsymbol{\alpha} \cdot \boldsymbol{\pi}^m)_r \text{ for all } (q \neq r) \in \mathcal{Q}_m^2,$$

$$(4.4) \forall m \in \{1, \dots, M\}, Q_m \geq 2,$$

(4.5) Each diagonal entry of $\boldsymbol{\alpha}$ is unique,

If $Q \geq 3$:

(4.6) There is no configuration of four indices $(q, r, s, t) \in \{1, \dots, Q\}$ such that $\alpha_{qq}/\alpha_{rr} = \alpha_{ss}/\alpha_{tt}$ with $q \neq s$ or $r \neq t$ and with $q \neq r$ or $s \neq t$,

(4.7) $\forall m \geq 2, |\mathcal{Q}_m \cap \cup_{l:l < m} \mathcal{Q}_l| \geq 2$.

Proof. [Celisse et al. \(2012\)](#) proved that the parameters $(\boldsymbol{\pi}^m, \boldsymbol{\alpha}^m)$ of the \mathcal{F} -SBM $_{n_m}(Q_m, \boldsymbol{\pi}^m, \boldsymbol{\alpha}^m)$ are identifiable up to a label switching of the blocks from the observation of a single network X^m when \mathcal{F} is the Bernoulli distribution and provided that

1. $n_m > 2Q_m$,
2. $(\boldsymbol{\alpha}^m \cdot \boldsymbol{\pi}^m)_q \neq (\boldsymbol{\alpha}^m \cdot \boldsymbol{\pi}^m)_r$ for all $(q \neq r) \in \{1, \dots, Q_m\}^2$.

Although they only consider the case where the emission distribution is the Bernoulli distribution, the extension to the Poisson distribution is straightforward. We prove the identifiability of our *colSBM* models by using their result, \mathcal{F} being either the Bernoulli or the Poisson distribution. The proofs for *iid-colSBM* and δ -*colSBM* are straightforward while the proofs for π -*colSBM* and $\delta\pi$ -*colSBM* are more complicated due to the possible existence of empty clusters in some networks X^m .

iid-colSBM Under this model, for all $m = 1, \dots, M$, $X^m \sim \mathcal{F}$ -SBM $_{n_m}(Q, \boldsymbol{\pi}, \boldsymbol{\alpha})$. As a consequence, following [Celisse et al. \(2012\)](#), the identifiability of $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$ is derived from the distribution of X^{m^*} under assumptions (1.1) and (1.2).

δ -*colSBM* Under this model, for all $m = 1, \dots, M$, $X^m \sim \mathcal{F}$ -SBM $_{n_m}(Q, \boldsymbol{\pi}, \delta_m \boldsymbol{\alpha})$. Under assumptions (2.1) and (2.2), we obtain the identifiability of $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$ from network X^{m^*} ([Celisse et al., 2012](#)). Now, for any $m \neq m^*$, by definition of \mathcal{F} -SBM $_{n_m}(Q, \boldsymbol{\pi}, \delta_m \boldsymbol{\alpha})$, we have:

$$\mathbb{E}[X_{ij}^m] = \delta_m \boldsymbol{\pi}' \boldsymbol{\alpha} \boldsymbol{\pi}.$$

This proves that δ_m is identifiable.

π -*colSBM* Note that under π -*colSBM*, we have, for all m ,

$$X^m \sim \mathcal{F}$$
-SBM $_{n_m}(Q_m, \tilde{\boldsymbol{\pi}}^m, \tilde{\boldsymbol{\alpha}}_m)$

where $\tilde{\boldsymbol{\pi}}^m$ is a vector of non-zero proportions of length Q_m and $\tilde{\boldsymbol{\alpha}}^m$ is the restriction of $\boldsymbol{\alpha}$ to \mathcal{Q}_m . Under assumptions (3.1) and (3.2) and applying [Celisse et al. \(2012\)](#) we obtain the identifiability of the parameters $\tilde{\boldsymbol{\pi}}^m$ and $\tilde{\boldsymbol{\alpha}}^m$ from each network X^m separately. However, the identifiability of each $\tilde{\boldsymbol{\pi}}^m$ and $\tilde{\boldsymbol{\alpha}}^m$ is established up to a label switching of the blocks in each network. We now have to find a coherent reordering between the networks which takes into account that some blocks are not represented in all the networks.

Let us build the complete matrix $\boldsymbol{\alpha}$ using the $\tilde{\boldsymbol{\alpha}}^m$. We fill the diagonal of $\boldsymbol{\alpha}$ which is composed of $(\text{diag}(\tilde{\boldsymbol{\alpha}}^m))_{m=1, \dots, M}$, taking the unique values and sorting them

by increasing order, i.e. $\alpha_{11} < \alpha_{22} < \dots < \alpha_{QQ}$. This task is possible because of assumption (3.3).

Now, we get back to $\tilde{\pi}^m$ and reorganize them to match with α . For any m , we define $\phi_m : \{1, \dots, Q_m\} \rightarrow \{1, \dots, Q\}$ such that $\alpha_{\phi_m(q), \phi_m(q)} = \tilde{\alpha}_{qq}^m$. With the (ϕ_m) we are able to fill the rest of α as: $\alpha_{\phi_m(q)\phi_m(r)} = \tilde{\alpha}_{qr}^m$ for all $(q, r) \in \{1, \dots, Q_m\}^2$. Finally, we define π^m a vector of size Q such that:

$$\pi_q^m = \begin{cases} 0 & \forall q \in \{1, \dots, Q\} \setminus \phi^m(\{1, \dots, Q_m\}) \\ \tilde{\pi}_{\phi_m^{-1}(q)}^m & \forall q \in \phi^m(\{1, \dots, Q_m\}) \end{cases}.$$

$\delta\pi$ -colSBM We now consider the model where

$$X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q, \pi^m, \delta_m \alpha). \quad (\text{S-1})$$

Like in model π -colSBM, Equation (S-1) implies that marginally,

$$X^m \sim \mathcal{F}\text{-SBM}_{n_m}(Q_m, \tilde{\pi}^m, \tilde{\alpha}^m) \quad \text{where} \quad \tilde{\alpha}^m = \delta_m(\alpha_{qr})_{q,r \in Q_m}.$$

Applying [Celisse et al. \(2012\)](#) on the distribution of X^m , we obtain the identifiability of the parameters $\tilde{\pi}^m$ and $\tilde{\alpha}^m$ (assumptions (4.1) and (4.3)). We now have to match the structures of the networks and to take into account the empty blocks. We separate the cases where $Q = 2$ from the ones where $Q > 3$.

The proof relies on a sequential identification the parameters (with respect to m).

- For $Q = 2$, we do not allow empty clusters (assumption (4.4), $Q_m \geq 2$) so $\tilde{\alpha}^m = \delta_m \alpha$ and $\tilde{\pi}^m = \pi^m$. Using the fact that $\delta_1 = 1$ (assumption (4.2)), we identify π^1 and α . Since we know that the diagonal entries of α are unique (assumption (4.5)), α can be chosen such that $\alpha_{11} > \alpha_{22}$. This provides the ordering of the blocks in each network. Then, we identify the π^m in a unique manner and not up to label switching.

- For $Q \geq 3$, for each $m \in \{1, \dots, M\}$, by assumptions (4.1) and (4.3) and using the marginal distributions, we are able to identify $\tilde{\alpha}^m$ and $\tilde{\pi}^m$. Using the fact that $\delta_1 = 1$ (assumption (4.2)) and the fact that the entries of the diagonal of α are unique, we can do as in π -colSBM and identify π^1 and $(\alpha_{qr})_{(q,r) \in Q_1}$.

Then for $m = 2$, up to a relabelling of the blocks in $\tilde{\alpha}^2$, we can define $\delta_2 = \tilde{\alpha}_{11}^2/\alpha_{11} = \tilde{\alpha}_{22}^2/\alpha_{22}$ since there are at least two blocks in network $m = 2$ that correspond to two blocks already identified in network $m = 1$ by assumption (4.7). We then need to prove that this parameter δ_2 is uniquely defined. Let us assume that there exists a parameter δ'_2 such that $\delta'_2 \neq \delta_2$ and $\delta'_2 = \tilde{\alpha}_{ii}^2/\alpha_{kk} = \tilde{\alpha}_{jj}^2/\alpha_{ll}$ with $i \neq j$. By definition of the $\delta\pi$ -colSBM $\alpha_{uu} := \tilde{\alpha}_{ii}^2/\delta_2$ which is either an already identified parameter from network $m = 1$ or corresponds to a new block represented in network $m = 2$ but not in network $m = 1$. Note that, since $\delta'_2 \neq \delta_2$, then $u \neq k$. For the same reason, $\exists v \neq l$ such that $\alpha_{vv} = \tilde{\alpha}_{jj}^2/\delta_2$. Since $i \neq j$, the parameters $\tilde{\alpha}_{ii}^2$ and $\tilde{\alpha}_{jj}^2$ are not equal which implies that $\alpha_{kk} \neq \alpha_{ll}$ and $\alpha_{uu} \neq \alpha_{vv}$ and finally that $k \neq l$ and $u \neq v$. By computing:

$$\frac{\alpha_{uu}}{\alpha_{vv}} = \frac{\tilde{\alpha}_{ii}^2/\delta_2}{\tilde{\alpha}_{jj}^2/\delta_2} \cdot \frac{\delta'_2}{\delta'_2} = \frac{\alpha_{kk}}{\alpha_{ll}}$$

we obtain a contradiction with assumption (4.6). Therefore, $\delta_2 = \delta'_2$.

We can then identify the blocks in network $m = 2$ by matching $\tilde{\alpha}_{qq}^2/\delta_2$ with the α_{qq} already identified. The $\tilde{\alpha}_{qq}^2/\delta_2$ that do not match with the previously identified parameters complete the matrix $\boldsymbol{\alpha}$. The process is iterated with networks $m = 3, \dots, M$. Once the matrix $\boldsymbol{\alpha}$ and the parameters in $\boldsymbol{\delta}$ are identified, injections from $\{1, \dots, Q_m\} \rightarrow \{1, \dots, Q\}$, corresponding to the matching of the blocks, provide the $\boldsymbol{\pi}^m$. □

B Variational estimation of the parameters

We provide here some details for the estimation of the parameters $\boldsymbol{\theta}_S \in \Theta_S$ for a given support matrix S . For ease of reading, the index S is dropped in this section. The likelihood

$$\ell(\mathbf{X}; \boldsymbol{\theta}) = \sum_{m=1}^M \log \int_{Z^m} \exp \{ \ell(X^m | Z^m; \boldsymbol{\alpha}, \boldsymbol{\delta}) + \ell(Z^m; \boldsymbol{\pi}) \} dZ^m \quad (\text{S-2})$$

is not tractable in practice, even for a small collection of networks as it relies on summing over $\sum_{m=1}^M |\mathcal{Q}_m|^{n_m}$ terms. A well-proven approach to handle this problem for the inference of the SBM is to rely on a variational version of the EM algorithm. This is done by maximizing a lower (variational) bound of the log-likelihood of the observed data (Daudin et al., 2008). More precisely,

$$\begin{aligned} \ell(\mathbf{X}; \boldsymbol{\theta}) &= \sum_{m=1}^M \ell(X^m; \boldsymbol{\theta}) \\ &\geq \sum_{m=1}^M \left(\ell(X^m; \boldsymbol{\theta}) - D_{\text{KL}}(\mathcal{R}_m(Z^m) \| p(Z^m | X^m; \boldsymbol{\theta})) \right) \end{aligned}$$

where D_{KL} is the Kullback-Leibler divergence and \mathcal{R}_m stands for any distribution on Z^m . The last equation can be reformulated as:

$$\ell(\mathbf{X}; \boldsymbol{\theta}) \geq \sum_{m=1}^M \left(\mathbb{E}_{\mathcal{R}_m} [\ell(X^m, Z^m; \boldsymbol{\theta})] + \mathcal{H}(\mathcal{R}_m) \right) =: \mathcal{J}(\mathcal{R}, \boldsymbol{\theta}). \quad (\text{S-3})$$

where $\mathcal{R} = \otimes_{m=1}^M \mathcal{R}_m$ and \mathcal{H} denotes the entropy of a distribution. Now, if, for all $m \in \{1, \dots, M\}$, \mathcal{R}_m is chosen in the set of fully factorizable distributions and if one sets $\tau_{iq}^m = \mathbb{P}_{\mathcal{R}_m}(Z_{iq}^m = 1)$ then $\mathcal{H}(\mathcal{R}_m)$ is equal to:

$$\mathcal{H}(\mathcal{R}_m) = - \sum_{i=1}^{n_m} \sum_{q \in \mathcal{Q}_m} \tau_{iq}^m \log \tau_{iq}^m. \quad (\text{S-4})$$

Besides, the complete likelihood of network m for the $\delta\pi$ -colSBM marginalized over \mathcal{R}_m is given by:

$$\mathbb{E}_{\mathcal{R}_m}[\ell(X^m, Z^m; \boldsymbol{\theta})] = \sum_{\substack{i,j=1 \\ i \neq j}}^{n_m} \sum_{(q,r) \in \mathcal{Q}_m} \tau_{iq}^m \tau_{jr}^m \log f(X_{ij}^m; \delta_m \alpha_{qr}) + \sum_{i=1}^{n_m} \sum_{q \in \mathcal{Q}_m} \tau_{iq}^m \log \pi_q^m. \quad (\text{S-5})$$

Finally, the variational lower bound $\mathcal{J}(\mathcal{R}, \boldsymbol{\theta}) := \mathcal{J}(\boldsymbol{\tau}, \boldsymbol{\theta})$ is obtained by plugging Equations (S-4) and (S-5) into the right member of Equation (S-3). Note that the lower bound $\mathcal{J}(\boldsymbol{\tau}, \boldsymbol{\theta})$ is equal to the log-likelihood if $\mathcal{R}_m(Z^m) = p(Z^m|X^m; \boldsymbol{\theta})$ for all $m \in \{1, \dots, M\}$.

The variational EM (VEM) algorithm consists in optimizing the lower bound $\mathcal{J}(\boldsymbol{\tau}, \boldsymbol{\theta})$ with respect to $(\boldsymbol{\tau}, \boldsymbol{\theta})$, by iterating two optimization steps with respect to $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ respectively, also referred to as VE-step and M-step. The details of each step are specific to the model at stake and are detailed hereafter.

VE-step At iteration (t) of the VEM algorithm, the VE-step consists in maximizing the lower bound with respect to $\boldsymbol{\tau}$:

$$\widehat{\boldsymbol{\tau}}^{(t+1)} = \arg \max_{\boldsymbol{\tau}} \mathcal{J}(\boldsymbol{\tau}, \widehat{\boldsymbol{\theta}}^{(t)}).$$

Note that by doing so, one minimizes the Kullback-Leibler divergences between $\mathcal{R}_m(Z^m)$ and $p(Z^m|X^m)$, and so approximates the true conditional distribution $p(Z^m|X^m)$ in the space of fully factorizable probability distributions. The $\boldsymbol{\tau}^m$'s can be optimized separately by iterating the following fixed point systems for all $m \in \{1, \dots, M\}$:

$$\widehat{\tau}_{iq}^{m(t+1)} \propto \widehat{\pi}_q^{m(t)} \prod_{\substack{j=1 \\ j \neq i}}^{n_m} \prod_{r \in \mathcal{Q}_m} f(X_{ij}^m; \widehat{\delta}_m^{(t)} \widehat{\alpha}_{qr}^{(t)}) \widehat{\tau}_{jr}^{m(t+1)} \quad \forall i = 1, \dots, n_m, q \in \mathcal{Q}_m. \quad (\text{S-6})$$

M-Step At iteration (t) of the VEM algorithm, the M-step maximizes the variational bound with respect to the model parameters $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} \mathcal{J}(\widehat{\boldsymbol{\tau}}^{(t+1)}, \boldsymbol{\theta}).$$

The update depends on the chosen model and the estimations are derived by canceling the gradient of the lower bound. For the sake of simplicity, the iteration index (t) is dropped in the following formulae. The obtained formulae involve the following quantities:

$$e_{qr}^m = \sum_{\substack{i,j=1 \\ i \neq j}}^{n_m} \tau_{iq}^m \tau_{jr}^m X_{ij}^m, \quad n_{qr}^m = \sum_{\substack{i,j=1 \\ i \neq j}}^{n_m} \tau_{iq}^m \tau_{jr}^m, \quad n_q^m = \sum_{i=1}^{n_m} \tau_{iq}^m.$$

On the one hand, the $(\pi_q^{(m)})_{q \in \mathcal{Q}_m}$ are estimated as

$$\widehat{\pi}_q^m = \frac{n_q^m}{n_m} \quad \text{for } \pi\text{-colSBM and } \delta\pi\text{-colSBM},$$

which is the expected proportion of the nodes in each allowed block for network m . On the other hand,

$$\hat{\pi}_q = \frac{\sum_{m=1}^M n_q^m}{\sum_{m=1}^M n_m} \quad \text{for } iid\text{-colSBM} \text{ and } \delta\text{-colSBM},$$

taking into account all the networks at the same time. The connection parameters α_{qr} of $iid\text{-colSBM}$ and $\pi\text{-colSBM}$ are estimated as the ratio of the number of interactions between blocks q and r among all networks over the number of possible interactions:

$$\hat{\alpha}_{qr} = \frac{\sum_{m=1}^M e_{qr}^m}{\sum_{m=1}^M n_{qr}^m} \quad \text{for } iid\text{-colSBM} \text{ and } \pi\text{-colSBM}.$$

For the $\delta\text{-colSBM}$ and $\delta\pi\text{-colSBM}$, there is no closed form for $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\alpha}}$ for a given value of $\boldsymbol{\tau}$. If $\mathcal{F} = \mathcal{Poisson}$, then $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\alpha}}$ can be iteratively updated using the following formulae:

$$\hat{\alpha}_{qr} = \frac{\sum_{m=1}^M e_{qr}^m}{\sum_{m=1}^M n_{qr}^m \hat{\delta}^m} \quad \text{and} \quad \hat{\delta}^m = \frac{\sum_{q,r \in \mathcal{Q}_m} e_{qr}^m}{\sum_{q,r \in \mathcal{Q}_m} n_{qr}^m \hat{\alpha}_{qr}}$$

If $\mathcal{F} = \mathcal{Bernoulli}$, no explicit expression can be derived and one has to rely on a gradient ascent algorithm to update the parameters at each M-Step.

Remark S-1. *In the VE-Step, each network can be treated independently, so the computation can be parallelized with ease. Also, it can be more efficient to update only a subset of networks at each step to avoid being stuck in local maxima. So we use a slightly modified VEM algorithm where we just compute the VE-step on one network at a time (the order of which is taken uniformly at random) before updating the corresponding parameters in the M-Step.*

C Details of the model selection procedure when allowing for empty blocks

For $\pi\text{-colSBM}$ and $\delta\pi\text{-colSBM}$, the model is described by its support S . We can compute the likelihood for a given support. We recall that $\boldsymbol{\theta}_S = \{\boldsymbol{\alpha}_S, \boldsymbol{\delta}, \boldsymbol{\pi}_S\}$ are the parameters restricted to their support. Then for the model represented by S , the complete likelihood is given by:

$$\begin{aligned} p(\mathbf{X}, \mathbf{Z}|S) &= \int_{\boldsymbol{\theta}_S} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}_S, S) p(\boldsymbol{\theta}_S) d(\boldsymbol{\theta}_S) \\ &= \int_{(\boldsymbol{\alpha}_S, \boldsymbol{\delta})} \int_{\boldsymbol{\pi}_S} p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\alpha}_S, \boldsymbol{\delta}, S) p(\mathbf{Z}|\boldsymbol{\pi}_S, S) p(\boldsymbol{\alpha}_S, \boldsymbol{\delta}) p(\boldsymbol{\pi}_S) d(\boldsymbol{\alpha}_S, \boldsymbol{\delta}) d(\boldsymbol{\pi}_S) \\ &= \underbrace{\int_{(\boldsymbol{\alpha}_S, \boldsymbol{\delta})} p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\alpha}_S, \boldsymbol{\delta}, S) p(\boldsymbol{\alpha}_S, \boldsymbol{\delta}) d(\boldsymbol{\alpha}_S, \boldsymbol{\delta})}_{B1} \underbrace{\int_{\boldsymbol{\pi}_S} p(\mathbf{Z}|\boldsymbol{\pi}_S, S) p(\boldsymbol{\pi}_S) d(\boldsymbol{\pi}_S)}_{B2}, \end{aligned} \tag{S-7}$$

where the prior on the emission parameters and on the mixture parameters are assumed to be independent.

The restriction of the parameter space to the one associated with the support S is needed. Otherwise, some parameters would not be defined or would lie on the boundary of the parameters space, and the following asymptotic derivation would not be properly defined. We use a BIC approximation on $B1$ where we rewrite:

$$\begin{aligned} p(\mathbf{X}|\mathbf{Z}, S) &= \int_{(\boldsymbol{\alpha}_S, \boldsymbol{\delta})} \left(\prod_{m=1}^M p(X^m|Z^m, \boldsymbol{\alpha}_S, \boldsymbol{\delta}, S) \right) p(\boldsymbol{\alpha}_S, \boldsymbol{\delta}) d(\boldsymbol{\alpha}_S, \boldsymbol{\delta}) \\ &= \max_{(\boldsymbol{\alpha}_S, \boldsymbol{\delta})} \exp \left(\sum_{m=1}^M \ell(X^m|Z^m; \boldsymbol{\alpha}_S, \boldsymbol{\delta}, S) - \frac{1}{2} \nu(\boldsymbol{\alpha}_S, \boldsymbol{\delta}) \log \left(\sum_m n_m(n_m - 1) \right) + \mathcal{O}(1) \right), \end{aligned}$$

where

$$\nu(\boldsymbol{\alpha}_S, \boldsymbol{\delta}) = \begin{cases} \nu(\boldsymbol{\alpha}_S) = \sum_{q,r=1}^Q \mathbf{1}_{(S^r S)_q > 0}, & \text{for } \pi\text{-colSBM} \\ \nu(\boldsymbol{\alpha}_S) + M - 1 & \text{for } \delta\pi\text{-colSBM}. \end{cases}$$

For $B2$, we use a Q_m -dimensional Dirichlet prior for each π_S^m :

$$\begin{aligned} p(\mathbf{Z}|S) &= \prod_{m \in \mathcal{M}} \int_{\pi_S^m} p(Z^m|\pi_S^m) p(\pi_S^m) d(\pi_S^m) \\ &= \max_{\pi_S} \exp \left(\sum_{m=1}^M \ell(Z^m; \pi_S^m) - \frac{Q_m - 1}{2} \log(n_m) + \mathcal{O}(1) \right). \end{aligned}$$

Then, we input $B1$ and $B2$ into Equation (S-7) to obtain:

$$\log p(\mathbf{X}, \mathbf{Z}|S) \approx \max_{\boldsymbol{\theta}_S} \ell(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta}_S) - \frac{1}{2} (\text{pen}_\pi(Q, S) + \text{pen}_\alpha(Q, S) + \text{pen}_\delta(Q, S)), \quad (\text{S-8})$$

with the penalty terms as given in the main text.

As \mathbf{Z} is unknown we replace each Z_{iq}^m by the variational parameters $\hat{\tau}_{iq}^m$ which maximizes the variational bound for a given support S . Then, we add the entropy of the variational distribution $\mathcal{H}(\hat{\mathcal{R}})$ to Equation (S-8). This leads to the variational bound of Equation (S-3), as

$$\max_{\boldsymbol{\theta}_S} \mathcal{J}(\hat{\boldsymbol{\tau}}, \boldsymbol{\theta}_S) = \max_{\boldsymbol{\theta}_S} \ell(\mathbf{X}, \mathbb{E}_{\hat{\mathcal{R}}}[\mathbf{Z}]; \boldsymbol{\theta}_S) + \mathcal{H}(\hat{\mathcal{R}}),$$

which we recall is a surrogate of the log-likelihood of the observed data. We define

$$\text{BIC-L}(\mathbf{X}, Q, S) = \max_{\boldsymbol{\theta}_S} \mathcal{J}(\hat{\boldsymbol{\tau}}, \boldsymbol{\theta}_S) - \frac{1}{2} (\text{pen}_\pi(Q, S) + \text{pen}_\alpha(Q, S) + \text{pen}_\delta(Q, S))$$

which is a penalized likelihood criterion when the support S is known.

Finally to obtain the criterion $\text{BIC-L}(\mathbf{X}, Q)$ for $\pi\text{-colSBM}$ and $\delta\pi\text{-colSBM}$, we need to penalize for the size of the space of possible models that depends on the support S . For a given Q corresponding to the number of different blocks in the collection of network \mathbf{X} , we set the prior on S decomposed as the product of uniform priors on the numbers of blocks (between 1 and Q) actually represented in each network

and uniform priors for the choice of these Q_m blocks among the Q possible blocks (Q_m is the number of blocks that are represented in network m):

$$p_Q(S) = p_Q(Q_1, \dots, Q_M) \cdot p_Q(S|Q_1, \dots, Q_M) = \frac{1}{Q^M} \cdot \prod_{m=1}^M 1 / \binom{Q}{Q_m}.$$

The prior is given on the space \mathcal{S}_Q of admissible support.

Using a BIC approximation and under a concentration assumption on the correct support, we derive

$$\begin{aligned} \log p(\mathbf{X}, \mathbf{Z}|\mathbf{Q}) &= \log \int_S p(\mathbf{X}, \mathbf{Z}|S, \mathbf{Q}) p_Q(S) dS \\ &\approx \log \int_S \exp(\text{BIC-L}(\mathbf{X}, Q, S)) p_Q(S) dS \\ &\approx \max_{S \in \mathcal{S}_Q} (\text{BIC-L}(\mathbf{X}, Q, S) - \log p_Q(S)). \end{aligned}$$

Thus, by denoting $\text{pen}_S(Q) = -2 \log p_Q(S)$ in the equation above, we obtain:

$$\text{BIC-L}(\mathbf{X}, Q) = \max_{S \in \mathcal{S}_Q} \left[\max_{\hat{\boldsymbol{\tau}}, \boldsymbol{\theta}_S} \mathcal{J}(\hat{\boldsymbol{\tau}}, \boldsymbol{\theta}_S) - \frac{1}{2} [\text{pen}_\pi(Q, S) + \text{pen}_\alpha(Q, S) + \text{pen}_\delta(Q, S) + \text{pen}_S(Q)] \right],$$

where

$$\begin{aligned} \text{pen}_\pi(Q, S) &= \sum_{m=1}^M (Q_m - 1) \log(n_m), \\ \text{pen}_\alpha(Q, S) &= \left(\sum_{q,r=1}^Q \mathbf{1}_{(S' S)_{qr} > 0} \right) \log(N_M), \\ \text{pen}_\delta(Q, S) &= \begin{cases} 0 & \text{for } \pi\text{-colSBM} \\ (M-1) \log(N_M) & \text{for } \delta\pi\text{-colSBM} \end{cases}, \\ \text{pen}_S(Q) &= -2 \log p_Q(S). \end{aligned}$$

D Simulation studies

In this section, we perform a large simulation study. The first study aims at testing the ability of the inference method to recover the number of blocks and the parameters for the $\pi\text{-colSBM}$ model. The second study highlights the performances in terms of clustering of networks based on their mesoscale structure.

D.1 Efficiency of the inference procedure

Simulation paradigm Let us simulate data under the π -colSBM model with $M = 2$, $n_m = 120$ and $Q = 4$. α and π are chosen as:

$$\alpha = .25 + \begin{pmatrix} 3\epsilon_\alpha & 2\epsilon_\alpha & \epsilon_\alpha & -\epsilon_\alpha \\ 2\epsilon_\alpha & 2\epsilon_\alpha & -\epsilon_\alpha & \epsilon_\alpha \\ \epsilon_\alpha & -\epsilon_\alpha & \epsilon_\alpha & 2\epsilon_\alpha \\ -\epsilon_\alpha & \epsilon_\alpha & 2\epsilon_\alpha & 0 \end{pmatrix}, \quad \pi^1 = \sigma_1(.2, .4, .4, 0), \quad \pi^2 = \sigma_2(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}). \quad (\text{S-9})$$

with ϵ_α taking eight equally spaced values ranging from 0 to 0.24. For each value of ϵ_α , 30 datasets (X^1, X^2) are simulated, resulting in $8 \times 30 = 240$ datasets. More precisely, for each dataset, we pick uniformly at random two permutations of $\{1, \dots, 4\}$ (σ_1, σ_2) with the constraint that $\sigma_1(4) \neq \sigma_2(1)$. This ensures that each of the two networks have a non-empty block that is empty in the other one. Then the networks are simulated with $\mathcal{B}\text{ern-SBM}_{120}(4, \alpha, \pi^m)$ with the previous parameters.

Each network has 2 blocks in common and their connectivity structures encompass a mix of core-periphery, assortative community and disassortative community structures, depending on which 3 of the 4 blocks are selected for each network. ϵ_α represents the strength of these structures, the larger, the easier it is to tell apart one block from another.

Inference On each simulated dataset, we fit the *iid-colSBM*, π -colSBM and *sep-SBM* models. The inference is performed with the VEM algorithm and the BIC-L criterions presented in the main manuscript.

Quality indicators To assess the quality of the inference, we compute the following set of indicators for each simulated dataset.

- First, for each dataset, we put in competition π -colSBM with *sep-SBM* and *iid-colSBM* respectively. To do so, for each dataset, we compute the BIC-L of each model π -colSBM is preferred to *sep-SBM* (resp. *iid-colSBM*) if its BIC-L is greater.
- Secondly, when considering the π -colSBM, we compare \hat{Q} to its true value ($Q = 4$).
- For π -colSBM and Q fixed to its true value ($Q = 4$), we evaluate the quality of recovery of the support matrix S by calculating:

$$\text{Rec}(\hat{S}, S) = \max_{\sigma \in \mathfrak{S}_4} \mathbf{1}_{\{\forall q, m S_{mq} = \hat{S}_{m\sigma(q)}\}} \quad (\text{S-10})$$

the greater the better.

- In order to evaluate the ability to recover the true connectivity parameter in the π -colSBM model, we compare $\hat{\alpha}$ to its true value for the true number of blocks $Q = 4$ through:

$$\text{RMSE}(\hat{\alpha}, \alpha) = \min_{\sigma \in \mathfrak{S}_4} \sqrt{\frac{1}{16} \sum_{1 \leq q, r \leq 4} (\hat{\alpha}_{\sigma(q)\sigma(r)} - \alpha_{qr})^2}, \quad (\text{S-11})$$

the σ being there to correct the possible label switching of the blocks.

- Finally, we judge the quality of our clustering with the Adjusted Rand Index (Hubert and Arabie, 1985, $\text{ARI} = 0$ for a random clustering and 1 for a perfect recovery). For each network, for the $\pi\text{-colSBM}$, using \hat{Q} , we compare the block memberships to the real ones by taking the average over the two networks

$$\overline{\text{ARI}} = \frac{1}{2} \left(\text{ARI}(\hat{\mathbf{Z}}^1, \mathbf{Z}^1) + \text{ARI}(\hat{\mathbf{Z}}^2, \mathbf{Z}^2) \right)$$

and by computing it on the whole collection of nodes

$$\text{ARI}_{1,2} = \text{ARI} \left((\hat{\mathbf{Z}}^1, \hat{\mathbf{Z}}^2), (\mathbf{Z}^1, \mathbf{Z}^2) \right).$$

All these quality indicators are averaged among the 30 simulated datasets. The results are provided in Table S-1. Each line corresponds to the 30 datasets simulated with a given value of ϵ_α . The first columns concatenate the results of the model comparison task. The following set of columns is about the selection of Q and the estimation of S . The last columns supply the RMSE on α and the ARI.

Results For the model comparison, when ϵ_α is small ($\epsilon_\alpha \in [0, .04]$), the simulation model is close to the Erdős-Rényi network and it is very hard to find any structure beyond the one of a single block. As such, the *iid-colSBM* and $\pi\text{-colSBM}$ models are equivalent and *iid-colSBM* is preferred to *sep-SBM*.

We observe a transition when $\epsilon_\alpha = .08$ where we become able to recover the true number of blocks $\hat{Q} = 4$ and the support of the blocks given the true number of blocks. During this transition, the model selection criterion is about half of the time in favor of *sep-SBM* i.e. the model with no common connectivity structure between the networks.

From $\epsilon_\alpha = .16$, we recover the true number of blocks and their support most of the time and the common structure obtained by the $\pi\text{-colSBM}$ is found to be relevant. Note that when we are able to recover the true number of blocks, we are also able to recover their support almost every time.

For both the estimation of the parameters and the ARIs, the results mainly follow our ability to recover the true number of blocks, with the error of estimation of the parameters slowly decreasing from $\epsilon_\alpha = 0.12$. $\overline{\text{ARI}}$ goes to 1 a bit faster than $\text{ARI}_{1,2}$, denoting our ability to recover faster the real clustering of each network than to match the blocks between the networks. This is directly linked with the detection of the true number of blocks and their support. Indeed, to get $\text{ARI}_{1,2} = 1$, we need $\text{Rec}(\hat{S}, S) = 1$ while the effective block number for each network is of only $Q = 3$, meaning that even with the wrong selected model we can still reach $\overline{\text{ARI}} = 1$.

D.2 Capacity to distinguish $\pi\text{-colSBM}$ from *iid-colSBM*

We aim to understand how well we are able to differentiate *iid-colSBM* from $\pi\text{-colSBM}$ depending on the block proportions. To do so, we fix $n_m = 90$ and

ϵ_α	Model comparison (π -colSBM vs \cdot)		Estimation of Q and S under π -colSBM				Parameter & Clustering accuracy under π -colSBM (mean \pm sd)		
	<i>sep</i> -SBM	<i>iid</i> -colSBM	$\mathbf{1}_{\hat{Q}<4}$	$\mathbf{1}_{\hat{Q}=4}$	$\mathbf{1}_{\hat{Q}>4}$	$\text{Rec}(\hat{S}, S)$	$\text{RMSE}(\hat{\alpha}, \alpha)$	$\overline{\text{ARI}}$	$\text{ARI}_{1,2}$
0	1	0	1	0	0	0	.1 \pm .002	0	0
.04	.83	0	1	0	0	0	.13 \pm .003	0	0
.08	.27	.43	.97	.03	0	.03	.13 \pm .044	.42 \pm .3	.25 \pm .21
.12	.4	.73	.3	.67	.03	.67	.1 \pm .076	.95 \pm .04	.64 \pm .29
.16	.9	1	0	.93	.07	.9	.03 \pm .04	.99 \pm .01	.97 \pm .1
.2	.93	1	0	.97	.03	.97	.02 \pm .04	1	.99 \pm .06
.24	.97	1	0	1	0	1	.01 \pm .003	1	1

Table S-1: **Accuracy of the inference for varying α .** All the quality indicators are averaged over the 30 simulated datasets.

$Q = 3$ and set α and π as follows:

$$\alpha = .25 + \begin{pmatrix} 3\epsilon_\alpha & 2\epsilon_\alpha & \epsilon_\alpha \\ 2\epsilon_\alpha & 2\epsilon_\alpha & -\epsilon_\alpha \\ \epsilon_\alpha & -\epsilon_\alpha & \epsilon_\alpha \end{pmatrix}, \pi^1 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \quad \text{and} \quad \pi^2 = \sigma \left(\frac{1}{3} - \epsilon_\pi, \frac{1}{3}, \frac{1}{3} + \epsilon_\pi\right),$$

with $\epsilon_\alpha = 0.16$ and ϵ_π taking 8 values equally spaced in $[0, .28]$. σ is a random permutation of the blocks. We simulate 30 different collections for each value of ϵ_π .

Here again, we put in competition π -colSBM with *iid*-colSBM and *sep*-SBM and select a model if its BIC-L is the greater than the two other ones. Then, for π -colSBM we compare \hat{Q} to 3 and evaluate our ability to recover S . The results are provided in Table S-2.

First notice that, since we chose $\epsilon_\pi \ll \frac{1}{3}$, we do not simulate any empty block. As a consequence, the inference of the model is quite easy and we are able to recover the true number of blocks and the right support for the π -colSBM model almost always. When $\epsilon_\pi = 0$, $\pi^1 = \pi^2$ and the model reduces to *iid*-colSBM. This remark explains why *iid*-colSBM is preferred to π -colSBM when $\epsilon_\pi < .2$. As ϵ_π increases, π -colSBM gets more and more selected, highlighting our capacity to recover the simulated structure.

D.3 Partition of networks

The third simulation experiment aims to illustrate our capacity to perform a partition of a collection of networks based on their structure, as presented in Section 6.

Simulation scenario For *iid*-colSBM, π -colSBM and δ -colSBM and $\delta\pi$ -colSBM, we simulate $M = 9$ undirected networks with 75 nodes and $Q = 3$ blocks. The block proportions are chosen as follows:

$$\pi^1 = (.2, .3, .5)$$

and for all $m = 2, \dots, 9$

$$\pi^m = \begin{cases} \pi^1 & \text{for } iid\text{-colSBM} \quad \text{and} \quad \delta\text{-colSBM} \\ \sigma_m(\pi^1) & \text{for } \pi\text{-colSBM} \quad \text{and} \quad \delta\pi\text{-colSBM} \end{cases}$$

ϵ_π	Model comparison			Estimation of Q and S under π -colSBM	
	<i>iid-colSBM</i>	π -colSBM	<i>sep-SBM</i>	$\mathbf{1}_{\hat{Q}=3}$	$\text{Rec}(\hat{S}, S)$
0	.97	.03	0	1	1
.04	.97	.03	0	1	1
.08	1	0	0	1	1
.12	.77	.23	0	1	1
.16	.77	.23	0	1	1
.2	.5	.5	0	1	1
.24	.2	.8	0	1	1
.28	.03	.97	0	1	1

Table S-2: **Model selection for varying mixture parameters.** The number of blocks \hat{Q} is given for the π -colSBM. The similarity of the block support to the true one $\text{Rec}(\hat{S}, S)$ is given for π -colSBM with $Q = 4$.

where σ_m is a permutation of $\{1, 2, 3\}$ proper to network m and $\sigma(\boldsymbol{\pi}) = (\pi_{\sigma(i)})_{i=1,\dots,3}$. The networks are divided into 3 groups of 3 networks with connectivity parameters as follows:

$$\boldsymbol{\alpha}^{\text{as}} = .3 + \begin{pmatrix} \epsilon & -\frac{\epsilon}{2} & -\frac{\epsilon}{2} \\ -\frac{\epsilon}{2} & \epsilon & -\frac{\epsilon}{2} \\ -\frac{\epsilon}{2} & -\frac{\epsilon}{2} & \epsilon \end{pmatrix}, \quad \boldsymbol{\alpha}^{\text{cp}} = .3 + \begin{pmatrix} \frac{3\epsilon}{2} & \epsilon & \frac{\epsilon}{2} \\ \epsilon & \frac{\epsilon}{2} & 0 \\ \frac{\epsilon}{2} & 0 & -\frac{\epsilon}{2} \end{pmatrix}, \quad \boldsymbol{\alpha}^{\text{dis}} = .3 + \begin{pmatrix} -\frac{\epsilon}{2} & \epsilon & \epsilon \\ \epsilon & -\frac{\epsilon}{2} & \epsilon \\ \epsilon & \epsilon & -\frac{\epsilon}{2} \end{pmatrix} \quad (\text{S-12})$$

with $\epsilon \in [.1, .4]$. $\boldsymbol{\alpha}^{\text{as}}$ represents a classical assortative community structure, while $\boldsymbol{\alpha}^{\text{cp}}$ is a layered core-periphery structure with block 2 acting as a semi-core. Finally, $\boldsymbol{\alpha}^{\text{dis}}$ is a disassortative community structure with stronger connections between blocks than within blocks. If $\epsilon = 0$, the three matrices are equal and the 9 networks have the same connection structure. Increasing ϵ differentiates the 3 clusters of networks. For δ -colSBM an $\delta\pi$ -colSBM, we add density parameters $\delta^1 = \delta^4 = \delta^7 = 1$, $\delta^2 = \delta^5 = \delta^8 = 0.75$ and $\delta^3 = \delta^6 = \delta^9 = 0.5$.

Each of these configurations is simulated 30 times. We apply the strategy exposed in Section 6 of the main manuscript and evaluate the recovery of the simulated network partition.

Results We assess the quality of our procedure by comparing the obtained classifications of the network collection with the simulated one through the ARI index. As ϵ grows we are able to better differentiate the networks and do so almost perfectly on all colSBM setup. Note that Adding complexity slightly deteriorates the results as we recover the partition better for *iid-colSBM* and π -colSBMs than for δ -colSBM and $\delta\pi$ -colSBM .

D.4 Finding finer block structures

Finally, we perform a last simulation study in order to illustrate the fact that, for particular configurations, using a colSBM model on a collection of networks

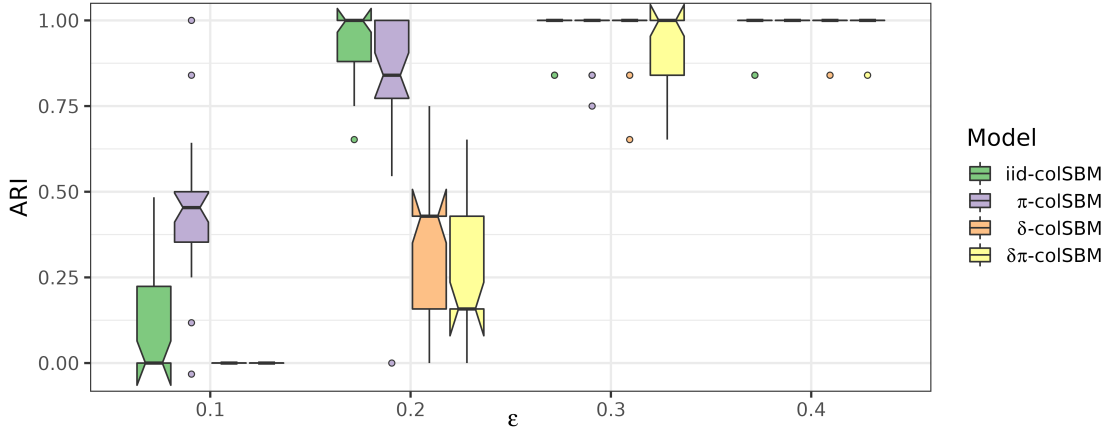


Figure S-1: **Partition of networks.** ARI of the recovered partition of networks.

favours the transfer of information between networks and allows to find finer block structures on the networks. We consider the core-periphery structure configuration described in Equation (S-12) with $\epsilon = .4$. In that case $Q = 3$.

We simulate a collection of 5 networks. 4 networks are of respective size $(90, 90, 120, 120)$. The last network is smaller with only 60 nodes and has a less marked structure ($\delta = .5$) for the $\delta\text{-colSBM}$ and $\delta\pi\text{-colSBM}$ models.

Our goal is to recover the true connection structure of this last network X^5 . To do so, we compare the results obtained using either a standard single SBM on X^5 , or using the corresponding $colSBM$ inferred with $M = 2, 3$ and 5 networks. We study Q in the various scenarii. In the simulation experiment, we obtained only $\hat{Q} = 2$ or 3. The experiment is repeated 30 times. The results are depicted in Figure S-2.

For the 4 models of simulation, the simple SBM recovers 2 blocks most of the time. For $iidcolSBM$ and $\pi\text{-colSBM}$, we always recover the 3 blocks while for the other case, we improve the ability to recover the true number of blocks when the quantity of information available from the other networks grows, either by augmenting the number of networks or by augmenting the number of nodes.

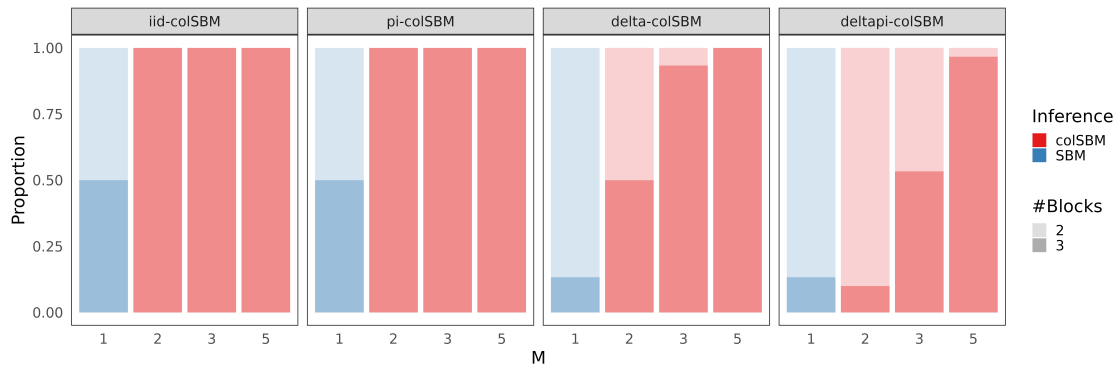


Figure S-2: **Finding finer block structures.** Cumulative barplot of \hat{Q} by the *SBMs* (blue) and the adequate *colSBM* (red) under the different simulation scenario. The number of blocks to be recovered is 3 and the darkest shade corresponds to $\hat{Q} = 3$.