Stochastic block model based on variational inference and its extensions: An application to examine global migration dynamics

Luca Brusa and Fulvia Pennoni

Abstract Discrete latent variable models are widely used in statistics and related fields because they enable the formulation of flexible and interpretable models for analyzing data with complex dependence structures among variables. We focus specifically on stochastic block (SB) models based on discrete latent variables, which are widely used for modeling network data. We first introduce the SB model for simple graphs, namely undirected binary graphs without self-loops, which allows us to delve into advanced aspects of model formulation and estimation. In this context the SB model assigns nodes to latent blocks with the probability of an edge existing between nodes depending on block membership. We discuss key inferential aspects of this model, including estimation of the parameters, conducted through a variational approximation of the expectation-maximization algorithm, prediction of the latent variable, and model selection. We also present several SB model extensions to realistically represent real-world networks, such as binary and weighted networks, dynamic networks, multiplex networks, bipartite and multipartite networks, and hypergraphs. We illustrate how the dynamic SB model may be applied to identify groups of countries with similar migration flows, using total migrant stock data provided by the United Nations from 1990 to 2015.

Luca Brusa
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milan, Italy e-mail: luca.brusa@unimib.it

Fulvia Pennoni
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milan, Italy e-mail: fulvia.pennoni@unimib.it

1 Introduction

Latent variable models are nowadays widely employed to analyze large and complex data. They postulate the existence of variables that are not directly observable (latent), but are assumed to affect the (observed) manifest variables (Everitt, 1984). In these contexts, latent variables are typically included in statistical models with different purposes, such as to consider: (*i*) the effect of unobservable covariates on the response variables, thus accounting for the unobserved heterogeneity between subjects; (*ii*) the "true" underlying value of the manifest observations, thus accounting for measurement errors; (*iii*) hypothetical constructs for which there exists no operational method for direct measurement (e.g., the quality of life); see Bartolucci, Pandolfi and Pennoni (2022) for a more insightful review on this topic. A possible classification of these models distinguishes between discrete and continuous latent variables. Focusing on the first case, the model may be seen as semi-parametric, since it is not necessary to rely on a parametric assumption of the distribution of the latent variables; in this way, the model is more flexible compared to other proposals. This class of models includes finite mixture (Titterington, Smith, and Mu, 1985), latent class (Goodman, 1974), hidden Markov (Bartolucci, Farcomeni and Pennoni, 2013), and stochastic block (SB, Holland, Laskey, and Leinhardt, 1983) models.

In the following, we focus on SB models, which are widely used in network analysis. Within this context a network is conceived as a set of nodes connected by edges, which represent links between two nodes. Network analysis arises in a variety of fields, from sociology to biology and ecology through computer science, and the nature of interactions between nodes can vary substantially. An important area of research in network analysis focuses on community detection (Newman, 2004; Fortunato, 2010), which aims to identify groups of nodes, or communities, that are more densely connected to each other than to the rest of the network; see Matias and Robin (2014) for more details about the difference between community detection and a more general clustering approach based on connectivity behavior.

SB models offer a more generalized framework for network analysis by assuming an underlying structure defined by a set of (unobservable) latent blocks. In this approach, nodes are grouped into these latent blocks, with nodes within the same block generically exhibiting similar patterns of interaction with other nodes. Starting from the basic formulation of SB models for simple graphs, namely undirected binary graphs without self-loops, several extensions exist, with the aim to handle more complex network structures, including binary, weighted, and dynamic networks, as well as multiplex and multipartite networks, and hypergraphs. Among these, dynamic stochastic block (DSB) models are particularly suited for analyzing networks that evolve over time, capturing changes in latent structures and interaction patterns across different time points. For inference, we consider a maximum likelihood estimation approach, focusing on a variational approximation of the expectation-maximization (EM) algorithm (Dempster, Laird and Rubin, 1977; Daudin, Picard and Robin, 2008).

To illustrate the utility of the DSB model, we present an application analyzing migration patterns among countries. The aim is to identify groups of nations with similar migration behaviors and examine how these patterns evolve over time. The

data, collected by the United Nations every five years, covers the period from 1990 to 2015.

The remainder of the paper is organized as follows: in Section 2, we introduce the notation, the basic SB model for simple graphs, and inference within an approximate maximum likelihood framework. In Section 3, we present several extensions, including binary, weighted, and dynamic SB models. In Section 4, we show the results of an application of the dynamic SB model. In Section 5, we provide concluding remarks and highlight directions for future research.

2 The stochastic block model

Suppose a network is represented by a binary graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} \neq \emptyset$ is the set of nodes, with $n = |\mathcal{N}|$ representing the number of nodes, and $\mathcal{E}$ is the set of edges, defined as pairs (i, j) of connected nodes. A binary graph is commonly represented through an adjacency matrix $A \in \{0, 1\}^{n \times n}$, such that $A_{ij} = 1$ if $(i, j) \in \mathcal{E}$. To keep the discussion as straightforward as possible, highlighting aspects such as model formulation and estimation, we first focus on *simple graphs*. They represent a fundamental class of graphs, characterized by (*i*) undirected interactions (bidirectional connections) between nodes, (*ii*) absence of self-loops (edges connecting a node to itself), and (*iii*) absence of multi-edges (two or more edges incident to the same pair of nodes).

2.1 Stochastic block model for simple graphs

Let Y_{ij} denote the binary response variable such that $Y_{ij} = Y_{ji} = 1$ if there exists an edge connecting nodes i and j, and $Y_{ij} = Y_{ji} = 0$ otherwise, with $Y_{ii} = 0$ for each $i = 1, \ldots, n$. Therefore, Y_{ij} corresponds to the (i, j)-entry of the adjacency matrix of the graph; let $\boldsymbol{Y} = (Y_{ij})_{i,j=1,\ldots,n}$ denote the set of the response variables.

The stochastic block (SB) model (Holland, Laskey, and Leinhardt, 1983; Snijders and Nowicki, 1997) relies on individual specific discrete latent variables denoted by U_i, with k support points identifying unobserved sub-populations of nodes, see also Brusa (2023). Let $\boldsymbol{U} = (U_1, \ldots, U_n)$ denote the vector of node-specific latent variables. The model is characterized by the following parameters: (*i*) the block weights $\pi_u = p(U_i = u)$, $u = 1, \ldots, k$, representing the probability that a node belongs to latent block u, and (*ii*) the connection probabilities $\beta_{uv} = p(Y_{ij}|U_i = u, U_j = v)$, $u, v = 1, \ldots, k$, denoting the probability that a node from latent block u is connected with a node from latent block v. The above probabilities must satisfy the constraint $\sum_{u=1}^{k} \pi_u = 1$ along with that of non-negativity and, since the graph is undirected, $\beta_{uv} = \beta_{vu}$ for each $u, v = 1, \ldots, k$. Likewise other discrete latent variable models, see Bartolucci, Pandolfi and Pennoni (2022), the response variables Y_{ij} are assumed to be conditionally independent given $U_i = u$ and $U_j = v$, so that a form of

local independence is considered. The number of free parameters is equal to

$$\#\text{par} = \underbrace{(k-1)}_{\pi_u} + \underbrace{\frac{k^2+k}{2}}_{\beta_{uv}}.$$

The path diagram of the corresponding SB model is depicted in Figure 1, showing portion of response and latent variables along with the conditional dependencies of the response variables in $\boldsymbol{Y}$ represented by the arrows. A common assumption is that for each pair of distinct nodes $i \neq j$, Y_{ij} is Bernoulli distributed conditionally on the latent variables U_i, U_j:

$$Y_{ij}|\{U_i = u, U_j = v\} \sim \mathcal{B}(\beta_{uv}). \tag{1}$$

Let $\boldsymbol{y} = (y_{ij})_{i,j=1,\dots,n}$ and $\boldsymbol{u} = (u_i)_{i=1,\dots,n}$ represent realizations of $\boldsymbol{Y}$ and $\boldsymbol{U}$, respectively. The SB model is formulated through two sub-models:

(*i*) the measurement sub-model, corresponding to the conditional distribution of the response variables $\boldsymbol{Y}$ given the latent variables $\boldsymbol{U}$, denoted as:

$$p(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{U} = \boldsymbol{u}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p(Y_{ij} = y_{ij} \mid U_i = u_i, U_j = u_j); \tag{2}$$

(*ii*) the latent sub-model, related to the conditional distribution of the latent variables $\boldsymbol{U}$ given by:

$$p(\boldsymbol{U} = \boldsymbol{u}) = \prod_{i=1}^{n} \pi_{u_i}. \tag{3}$$

The observed network distribution, corresponding to the observed data likelihood function, may be expressed as

$$p(\boldsymbol{Y} = \boldsymbol{y}) = \sum_{\boldsymbol{u}} p(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{U} = \boldsymbol{u}) p(\boldsymbol{U} = \boldsymbol{u}), \tag{4}$$

which involves a summation over all possible k^n different latent configurations, and is computationally intractable unless n and k are small. In the next subsection, we address the problem of maximizing this function considering variational inference (Blei, Kucukelbir and McAuliffe, 2017).

2.1.1 Maximum likelihood estimation

Maximum likelihood estimation through the EM algorithm (Dempster, Laird and Rubin, 1977) is discussed in Snijders and Nowicki (1997), where they state that this approach is not feasible unless n is relatively small (not greater than 20). As illustrated in Matias and Robin (2014), the conditional distribution of the latent

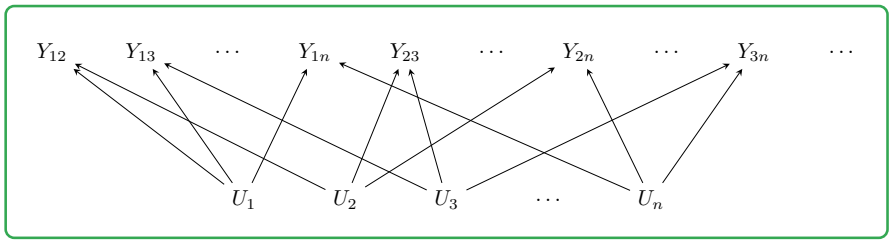

Fig. 1 Path diagram of the stochastic block model for simple graphs, representing a subset of response variables (top row) of latent variables (bottom row); the arrows highlight the dependence structures of the model: response variables Y_{ij} are conditionally independent given the latent variables U_i and U_j, dependencies are represented by directed edges

variables given the responses is intractable. They show that for each pair of latent variables it results as:

$$p(U_i, U_j \mid Y_{ij}) = \frac{p(U_i, U_j, Y_{ij})}{p(Y_{ij})} = \frac{p(Y_{ij} \mid U_i, U_j)p(U_i)p(U_j)}{p(Y_{ij})},$$

which cannot be further factorized since the latent variables U_i and U_j are not independent, conditional on Y_{ij}. Thus, the iterative procedure required for the EM algorithm is computationally infeasible.

A variational approximation of the EM algorithm, first introduced in Jordan et al. (1999), is generally employed. In this approach, rather then maximizing the intractable log-likelihood directly, a suitable lower bound of this function is considered. Let q_τ represent the class of probability distributions defined with respect to the variational parameter $\boldsymbol{\tau}$ as:

$$q_{\boldsymbol{\tau}}(\boldsymbol{U} = \boldsymbol{u}) = \prod_{i=1}^{n} q_{\boldsymbol{\tau}}(U_i = u_i) = \prod_{i=1}^{n} \prod_{u=1}^{k} \tau_{iu}^{I(u_i = u)},$$

where $I(\cdot)$ denotes the indicator function and $\tau_{iu} = q_{\boldsymbol{\tau}}(U_i = u) \in [0, 1]$ is the variational parameter, subject to $\sum_{u=1}^{k} \tau_{iu} = 1$. These distributions are defined in such a way that they are factorized, thus allowing them to approximate the conditional distribution. Let

$$\mathcal{H}(q_{\boldsymbol{\tau}}) = \mathbb{E}_{q_{\boldsymbol{\tau}}}[-\log q_{\boldsymbol{\tau}}(\boldsymbol{U} = \boldsymbol{u})],$$

be an entropy measure (Shannon, 1948), where $\mathbb{E}_{q_{\boldsymbol{\tau}}}$ denote the expected value under $q_{\boldsymbol{\tau}}$. An approximation of the intractable observed data likelihood function $p(\boldsymbol{Y} = \boldsymbol{y})$ may be expressed as:

$$\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \mathbb{E}_{q_{\boldsymbol{\tau}}}[\log p(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{U} = \boldsymbol{u})] + \mathcal{H}(q_{\boldsymbol{\tau}}). \tag{5}$$

Considering the Kullback-Leibler (KL, Kullback and Leibler, 1951)[1] divergence, it results:

$$\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \log p(\boldsymbol{Y} = \boldsymbol{y}) - \mathrm{KL}[q_{\boldsymbol{\tau}}(\boldsymbol{U} = \boldsymbol{u}) \,||\, p(\boldsymbol{U} = \boldsymbol{u} \,|\, \boldsymbol{Y} = \boldsymbol{y})] \leq \log p(\boldsymbol{Y} = \boldsymbol{y}),$$

showing that $\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{\tau})$ is a lower bound of the observed data likelihood function, also named *evidence lower bound* (ELBO). The variational expectation-maximization (VEM) algorithm, once properly initialized, alternates until convergence two steps, which for a generic h-th iteration of the algorithm are defined as follows:

(*i*) a variational expectation (VE) step that maximizes $\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{\tau})$ with respect to $\boldsymbol{\tau}$:

$$\hat{\boldsymbol{\tau}}^{(h)} = \arg\max_{\boldsymbol{\tau}} \mathcal{J}(\hat{\boldsymbol{\theta}}^{(h-1)}, \boldsymbol{\tau}), \quad \text{subject to} \quad \sum_{u=1}^{k} \tau_{iu} = 1 \quad \forall i = 1, \ldots, n,$$

thus minimizing the KL divergence and refining the variational approximation of the log-likelihood function.

(*ii*) a maximization (M) step that maximizes $\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{\tau})$ with respect to $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}^{(h)} = \arg\max_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}, \hat{\boldsymbol{\tau}}^{(h-1)}), \quad \text{subject to} \quad \sum_{u=1}^{k} \pi_u = 1,$$

thus updating the model parameters.

Here $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\theta}}$ denote the estimates for $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$, respectively. We refer to Daudin, Picard and Robin (2008) for explicit solutions of the M-step and additional details on the VE-step, which requires an iterative algorithm like the Newton-Raphson algorithm. To the best of our knowledge, no guarantees about existence nor uniqueness of a solution to this problem exist in the statistical literature.

Another challenge is related to the multi-modality of the log-likelihood function, which can lead the VEM algorithm to converge to a local maximum. Multi-start strategies are often employed, using both deterministic and random rules to initialize model parameters. However, this approach may require substantial computational time. A recent proposal by Brusa, Pennoni and Bartolucci (2024) introduces an updated version of the VEM algorithm based on the evolutionary algorithm (Ashlock, 2004) to provide a refined exploration of the entire parameter space during the iterative estimation process; see also Brusa and Pennoni (2024) for details on this procedure implemented to estimate the SB model.

[1] Note that $\mathrm{KL}[p \,||\, q] \geq 0$ for each pair of distributions p, q, and $\mathrm{KL}[p \,||\, q] = 0$ if and only if $p = q$.

2.2 Further estimation issues

Two additional features of the estimation procedure are worth mentioning. The first concerns the estimated variational parameters ($\hat{\boldsymbol{\tau}}$) through which prediction of the latent variables is achieved. In this way, the SB model is useful for clustering, achieved by assigning nodes to latent blocks such that nodes within each community share similar characteristics. Generally, a maximum a-posteriori rule (MAP, Goodman, 1974) is employed, where each node i is assigned to the latent block with the highest estimate value of $\hat{\tau}_{iu}$, such that:

$$\hat{u}_i = \underset{u=1,\ldots,k}{\arg\max}\ \hat{\tau}_{iu}, \quad i = 1, \ldots, n.$$

A second issue concerns the selection of the number of latent blocks when this number is not known a priori. Information criteria, such as the integrated classification likelihood (ICL) criterion (Biernacki, Celeux and Govaert, 2000), are commonly employed. In the case of SB model for simple graphs, an explicit formulation of the ICL adjusted for the variational approach is based on a penalized complete data log-likelihood function, as proposed by Daudin, Picard and Robin (2008). It can be expressed as:

$$\text{ICL} = \log p(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{U} = \hat{\boldsymbol{u}}) - \frac{1}{2}\frac{k(k+1)}{2}\log\frac{n(n-1)}{2} - \frac{1}{2}(k-1)\log n,$$

where the first penalization term accounts for the number of connection parameters, and the other for the number of block weights. According to this criterion, the optimal SB model is the one with the number of latent blocks that maximizes the ICL value.

3 Extensions and generalizations

We outline some extensions of the SB model proposed to account for the complexity of the observed network data. In particular, we consider both binary and weighted edges, as well as dynamic, multiplex, multipartite networks, and hypergraphs. For detailed illustrations of these and other extensions, see Lee and Wilkinson (2019).

3.1 Binary networks

A simple generalization of the binary SB model presented in Section 2 is proposed in Nowicki and Snijders (2001) for directed graphs. In this case, the conditional probabilities related to the measurement sub-model account for the edge directionality as follows:

$$p(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{U} = \boldsymbol{u}) = \prod_{i=1}^{n} \prod_{j \neq i} p(Y_{ij} = y_{ij} \mid U_i = u_i, U_j = u_j). \tag{6}$$

The symmetry property of the connection probabilities, such that $\beta_{uv} = \beta_{vu}$, no longer holds in the directed case. The possible inclusion of self-loops can be addressed by allowing $i = j$ in Equations (2) and (6), for directed and undirected networks, respectively. These extensions do not substantially affect model formulation or estimation methods. Multi-edges can also be considered, which, while conceptually different from weighted edges[2], are similar from a modeling and estimation perspective, as illustrated in the next subsection.

3.2 Weighted networks

In a wide range of real-world applications, the only information concerning the presence or absence of the connection between two nodes, may be too restrictive for understanding the real structure of the system. For example, in a transportation network for railways or pipelines, a binary graph cannot account for the number of routes or the capacity, nor can it represent sparsity, meaning the absence of connections between most pairs of nodes. To this aim, some extensions are proposed in Kurihara, Kameya and Sato (2006) and Mariadassou, Robin and Vacher (2010), addressing the case of (possibly) sparse, weighted networks. In this context, as defined in Ambroise and Matias (2012), Equation (1) is as follows:

$$Y_{ij} | \{U_i = u, U_j = v\} \sim (1 - \alpha_{uv})\delta_0 + \alpha_{uv} F(\gamma_{uv}), \tag{7}$$

where δ_0 is the Dirac Delta function in 0, $F(\gamma_{uv})$ is a parametric distribution assigning zero probability to the value 0, and $\alpha_{uv} \in [0, 1]$ is the sparsity parameter, with $\alpha_{uv} = 1$ indicating fully connected networks. Notable examples for the distribution $F(\gamma_{uv})$ include the Gaussian distribution (for handling continuous edge weights), the truncated Poisson distribution on $\mathbb{N}\backslash\{0\}$ (for modeling count-based weights), and the multinomial distribution (for edges categorized into discrete types). This general formulation naturally includes also the case of binary networks by setting $F(\gamma_{uv}) = \delta_1$, where δ_1 is the Dirac delta function in 1, representing the presence of an edge. The model is still estimated with the maximum likelihood approach through the VEM algorithm, where both the VE and M steps implemented according to the chosen conditional distribution. For model selection, the ICL criterion is suitably modified with an additional penalty to account for the weighted edges.

More recently, Aicher, Jacobs and Clauset (2015) propose the use of whichever distribution from the exponential family to enhance flexibility in capturing different types of weighted interactions. They also formulate a mixture model (McLachlan and

[2] The former represents multiple distinct occurrences of connections between two nodes, while the latter aggregates these occurrences into a single edge with an associated weight that reflects the number or strength of the connections.

Peel, 2000) for the conditional distribution of the edge weights, allowing for distinct underlying processes that govern the network structure. For example, a mixture SB model could combine two different distributions, such as a Bernoulli distribution for the presence or absence of edges and a Poisson distribution for the edge weights.

It is worth mentioning the approach proposed by Karrer and Newman (2011), which accounts for heterogeneity in node degrees, namely, the number of edges connected to a node, by introducing an additional parameter into the conditional distribution of Y_{ij} given U_i and U_j. This parameter captures relevant differences in nodes degree within the same block. Further extensions in this direction have been proposed, among others, in Peixoto (2014), to accommodate large networks, and in Yan et al. (2014) to jointly select the optimal number of latent blocks and the most suitable SB model choosing between the degree-corrected and standard versions.

3.3 Dynamic networks

An important area of study is the analysis of how a network structure may evolve over time. For example, analyzing temporal patterns of face-to-face contacts can provide valuable insights into the transmission dynamics of infectious diseases. In such a context, when longitudinal network data are available, the dynamic SB (DSB) model provides estimates of time-varying patterns of node connections and dynamic clustering, where nodes may be allocated to different clusters over time; see, among others, Yang et al. (2011), Xu and Hero III (2014), and Ghasemian et al. (2016).

To show some details of this model, we follow the formulation proposed in Matias and Miele (2017). Let $\boldsymbol{Y}^{(t)}$, $t = 1, \ldots, T$ be the set of observed network variables at each time occasion t, where $Y_{ij}^{(t)}$ represents the existence or the weight (in binary and weighted networks, respectively) of the edge between nodes i and j at time t, $i, j = 1, \ldots, n$. We also define $\boldsymbol{Y} = (\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(T)})$. As proposed in Bartolucci and Pandolfi (2020) a possible specification of the DSB model is based on a discrete-time Markovian edge dynamics. In this case, the block membership processes for node i is defined according to discrete latent variables $\boldsymbol{U}_i = (U_i^{(1)}, \ldots, U_i^{(T)})'$ following a first-order Markov-chain with k states. Assuming local independence, each $Y_{ij}^{(t)}$ is conditionally independent given $U_i^{(t)}$ and $U_j^{(t)}$, and Equation (7) can be generalized as:

$$Y_{ij}^{(t)} | \{U_i^{(t)} = u, U_j^{(t)} = v\} \sim (1 - \alpha_{uv}^{(t)})\delta_0 + \alpha_{uv}^{(t)} F(\gamma_{uv}^{(t)}). \tag{8}$$

In addition, the vectors $\boldsymbol{U} = (U_i^{(1)}, \ldots, U_i^{(T)})'_{i=1,\ldots,n}$ are mutually independent and identically distributed, evolving over time according to the initial probabilities π_u, $u = 1, \ldots, k$, and the transition probabilities $\pi_{uv}^{(t)}$, $t = 2, \ldots, T$, $u, v = 1, \ldots, k$. A more parsimonious model is obtained under the assumption of time homogeneity, which means that the model parameters remains constant over time. In this case, the process behaves consistently across different time periods since $\pi_{uv}^{(t)} = \pi_{uv}$,

$\alpha_{uv}^{(t)} = \alpha_{uv}$, and $\gamma_{uv}^{(t)} = \gamma_{uv}$ for each t. The probability mass functions for the measurement and the latent sub-models are thus reformulated as follows:

$$p(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{U} = \boldsymbol{u}) = \prod_{t=1}^{T} \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p(Y_{ij}^{(t)} = y_{ij}^{(t)} \mid U_i^{(t)} = u_i^{(t)}, U_j^{(t)} = u_j^{(t)}),$$

$$p(\boldsymbol{U} = \boldsymbol{u}) = \prod_{i=1}^{n} \pi_{u_i^{(1)}} \prod_{t=2}^{T} \pi_{u_i^{(t-1)} u_i^{(t)}},$$

where $\boldsymbol{y} = (y_{ij}^{(t)})$ and $\boldsymbol{u} = (u_i^{(t)})$ are realizations of $\boldsymbol{Y}$ and $\boldsymbol{U}$, respectively. The observed longitudinal network distribution is specified as in Equation (4).

Maximum likelihood estimation is performed using two different versions of the VEM algorithm: one assumes a posteriori independence across latent variables over both time points and units, following the approach of Yang et al. (2011); the other assumes independence only across units, as in Matias and Miele (2017) and in Bartolucci and Pandolfi (2020). Following the last proposal, the approximate variational inference is obtained by considering:

$$q_{\boldsymbol{\tau}}(\boldsymbol{U} = \boldsymbol{u}) = \prod_{i=1}^{n} \left[q_{\boldsymbol{\tau}}(U_i^{(1)} = u_i^{(1)}) \prod_{t=2}^{T} q_{\boldsymbol{\tau}}(U_i^{(t)} = u_i^{(t)} | U_i^{(t-1)} = u_i^{(t-1)}) \right]$$

$$= \prod_{i=1}^{n} \left[\prod_{u=1}^{k} \tau(i,u)^{I(U_i^{(1)}=u)} \prod_{t=2}^{T} \prod_{u=1}^{k} \prod_{v=1}^{k} \tau(t,i,u,v)^{I(U_i^{(t-1)}=u) I(U_i^{(t)}=v)} \right],$$

where the variational parameters $\tau(i,u)$ and $\tau(t,i,u,v)$ are approximations of $p(U_i^{(1)} = u | \boldsymbol{Y} = \boldsymbol{y})$ and $p(U_i^{(t)} = v | U_i^{(t-1)} = u, \boldsymbol{Y} = \boldsymbol{y})$, respectively. The ELBO function is then defined as in Equation (5), and the VEM algorithm operates as illustrated in Section 2.1.1.

The challenge of detecting changes in the clustering structure of a dynamic network has received considerable attention in the literature. Proposals includes models for repeated connections in continuous time, see, among others, Du Bois, Butts and Smyth et al. (2013) and Matias, Rebafka and Villers (2018), degree-corrected approaches that account for node degree heterogeneity over time (Zhang, Moore and Newman, 2017), and mixed-membership approaches that allow nodes to belong to multiple clusters, see Xing, Fu and L Song (2010).

3.4 More complex network structures

In this section, we briefly review some of the recently proposed extensions of SB models that account for more complex network structures among nodes, including multiplex, bipartite and multipartite networks, and hypergraphs.

3.4.1 Multiplex networks

Multiplex networks frequently arise in applied contexts. For instance, in a social network, layers could represent the type of relationship, such as kinship, friendship, or professional collaboration. Multiplex networks (Kivelä et al., 2014), also referred to as multilevel or multilayer networks, are used to model the simultaneous existence of different types of relationships or interactions among the same block of nodes. Unlike dynamic networks, edges do not change over time, but represent different relational contexts, highlighting the diversity and complexity of connections within multi-faceted networks.

Following the notation proposed in Barbican, Donnet and Bar-Hen (2017), let $\boldsymbol{Y}^1, \ldots, \boldsymbol{Y}^L$ be L networks sharing the same set of nodes $\{1, \ldots, n\}$. In this framework, each network accounts for a different observed type of interaction among the nodes, that is, $Y_{ij}^{\ell} = 1$ if nodes i and j are connect by an edge of type ℓ, $i, j = 1, \ldots, n$. We also let $\boldsymbol{Y} = (\boldsymbol{Y}^L, \ldots, \boldsymbol{Y}^L)$, and consider node-specific latent variables $\boldsymbol{U} = (U_1, \ldots, U_n)$. Under the most general assumption of dependent network layers, it holds:

$$(Y_{ij}^1, \ldots, Y_{ij}^L) | \{U_i = u, U_j = v\} \sim \mathcal{B}_L(\beta_{uv}^{\boldsymbol{\omega}}),$$

where $\mathcal{B}_L$ denotes the L-variate Bernoulli distribution, $\boldsymbol{\omega} \in \{0, 1\}^L$ represents the configuration of connections across the L layers, and $\beta_{uv}^{\boldsymbol{\omega}} = p\big((Y_{ij}^1, \ldots, Y_{ij}^L) = \boldsymbol{\omega} | U_i = u, U_j = v\big)$. Since the model may be over-parameterized, some assumptions are required to reduce the number of parameters and make the model more parsimonious and easier to interpret. In fact, the number of free parameters for a network with undirected edges is $(k-1) + \frac{2^L k(k+1)}{2}$, which reduces to $(k-1) + 2^L k^2$ when the network is directed. To this aim, a straightforward assumption, proposed in Barbican, Donnet and Bar-Hen (2017), is that of conditional independence among layers given the latent variables, such that:

$$Y_{ij}^{\ell} | \{U_i = u, U_j = v\} \sim \mathcal{B}(\beta_{uv}^{\ell}),$$

where the connection probabilities are now defined as $\beta_{uv}^{\ell} = p(Y_{ij}^{\ell} = 1 | U_i = u, U_j = v)$, $\ell = 1, \ldots, L$. The corresponding number of free parameters is equal to $(k - 1) + Lk^2$ when the network is undirected, and $(k - 1) + Lk(k + 1)/2$ for a directed network.

Alternatively, assuming the same conditional distribution, Paul and Chen (2016) propose to model the logit of the connection probabilities β_{uv}^{ℓ} according to the following parameters:

$$\text{logit}(\beta_{uv}^{\ell}) = \phi_{uv} + \eta_{\ell},$$

where ϕ_{uv} captures the clustering effect of the network structure, and η_{ℓ} the effect of the layer. In this case, the number of free parameters further reduces to $(k-1)+(L+k^2)$ and $(k - 1) + (L + k(k + 1)/2)$ for directed and undirected networks, respectively.

The previous proposal is further extended by Peixoto (2015) to account for networks with varying connection intensities and dynamic evolution across layers. Another proposal is that of Stanley et al. (2016), where groups (or strata) of layers are estimated so that all layers within a stratum share the same community structure and model parameters. Finally, Vallès-Català et al. (2016) consider networks where the different layers are hidden, making all types of interactions appear equivalent. Their approach aims to disentangle and uncover the layers, thus revealing the distinct interaction types underlying the observed network.

3.4.2 Bipartite and multipartite networks

In much of the ecological literature, there is interest in understanding species interactions (Dormann and Strauss, 2014); a particular emphasis is placed when they only occur between units of two different species, but not within the same specie, as for example in the interaction between plants and pollinators. Bipartite networks (Newman, Strogatz and Watts, 2001) constitute a generalization of traditional networks, where nodes are divided into two disjoint sets, known as functional groups, and edges are allowed to link only nodes from one set to nodes in the other. Such structures have been extended in multipartite networks considering more than two disjoint sets, with edges permitted only between nodes of different sets.

Bar-Hen, Barbillon and Donnet (2022) propose a multipartite SB model characterized by the following formulation and parameters. Let Q denote the number of functional groups, where $Q = 2$ for bipartite networks and $Q \geq 2$ for multipartite networks, and let n_q be the number of nodes in the q-th functional group, for $q = 1, \ldots, Q$. Denoting by $\mathcal{E}$ the list of pairs (q, s) of functional groups for which we observe interactions, the corresponding network is represented through an adjacency matrix $\boldsymbol{Y}^{qs}$ such that $Y_{ij}^{qs} = 1$ if there exists an edge connecting node i of functional group q to node j of functional group s, and $Y_{ij}^{qs} = 0$, otherwise. Moreover, let $\boldsymbol{Y} = (\boldsymbol{Y}^{qs})_{q,s=1,\ldots,Q}$ denote the whole multipartite network. The SB model assumes that nodes in the q-th functional group belong to k_q hidden blocks. Let U_i^q be the node- and group-specific latent variables; denoting as $\boldsymbol{U} = (\boldsymbol{U}^1, \ldots, \boldsymbol{U}^Q)$, with $\boldsymbol{U}^q = (U_1^q, \ldots, U_{n_q}^q)$, the vector collecting these latent variables, and $\pi_{u_i}^q = p(U_i^q = u)$, $q = 1, \ldots, Q$, $i = 1, \ldots, n_q$, $u = 1, \ldots, k_q$, the probability that node i in functional group q belongs to latent block u, the conditional response distribution may be expressed as

$$Y_{ij}^{qs} | \{U_i^q = u, U_j^s = v\} \sim \mathcal{B}(\beta_{uv}^{qs}),$$

where β_{uv}^{qs} represents the probability that a node in functional group q and latent block u is connected to a node in functional group s and latent block v. Note that $\beta_{uv}^{qs} = 0$ for each $u = 1, \ldots, k_q$ and $v = 1, \ldots, k_s$ if the functional groups q and s can not be connected by edges. This formulation may be easily generalized to a weighted network using Poisson, Gaussian or other conditional distributions; see Bar-Hen, Barbillon and Donnet (2022) for more details.

The measurement and the latent sub-models are expressed as follows:

$$p(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{U} = \boldsymbol{u}) = \prod_{(q,s)\in\mathcal{Q}} \prod_{i=1}^{n_q} \prod_{j=1}^{n_s} p(Y_{ij}^{qs} = y_{ij}^{qs} | U_i^q = u_i^q, U_j^s = u_j^s),$$

$$p(\boldsymbol{U} = \boldsymbol{u}) = \prod_{q=1}^{Q} \prod_{i=1}^{n_q} \pi_{u_i}^q,$$

where $\boldsymbol{y} = (y_{ij}^{qs})$ and $\boldsymbol{u} = (u_i^q)$ are realizations of $\boldsymbol{Y}$ and $\boldsymbol{U}$, respectively. The variational approximation is still employed with the The ELBO function is defined as in Equation (5), and the model parameters are estimated using the VEM algorithm, as illustrated in Section 2.1.1.

3.4.3 Hypergraphs

In contexts such as social networks, where interactions commonly occur among three or more individuals, or in studies examining simultaneous interactions among groups of more than two neurons or brain regions, more complex structures and models are required to capture these multifaceted interactions; see, among others, Yang (2017). Hypergraphs (Battiston et al., 2020; Bick et al., 2023) represent an important generalization of networks, since they can account for high-order interactions, involving groups of three or more nodes. Similarly to a graph, a simple hypergraph is denoted as $\mathcal{HG} = (\mathcal{N}, \mathcal{HE})$, where $\mathcal{N} \neq \emptyset$ is the set of n nodes and $\mathcal{HE}$ is the set of hyperedges. Each hyperedge consists of m distinct nodes involved in a common interaction. Let M denote the largest possible size of hyperedges in $\mathcal{HE}$. Note that, in simple hypergraphs, self-loops and weighted hyperedges are not allowed. Considering the formulation of the hypergraph SB (HSB) model provided in Brusa and Matias (2024), for each node subset $\{i_1, \ldots, i_m\}$ of size m we define the following indicator variable:

$$Y_{i_1,\ldots,i_m} = I\Big(\{i_1, \ldots, i_m\} \in \mathcal{HE}\Big) = \begin{cases} 1 & \text{if } \{i_1, \ldots, i_m\} \in \mathcal{HE} \\ 0 & \text{if } \{i_1, \ldots, i_m\} \notin \mathcal{HE} \end{cases},$$

with $i_1 \neq \ldots \neq i_m$, and $m = 2, \ldots, M$. Let $\boldsymbol{Y} = (Y_{i_1,\ldots,i_m})$ denote the vector collecting all the variables; also in this case block membership depends on node specific discrete latent variables $\boldsymbol{U} = (U_1, \ldots, U_n)$ with $\pi_u = p(U_i = u)$ as block weights, subject to $\sum_{u=1}^{k} \pi_u = 1$. Local independence is still assumed since each indicator variable $Y_{i_1,\ldots,i_m}$ is conditionally independent given $\boldsymbol{U}$ and follows a Bernoulli distribution:

$$Y_{i_1,\ldots,i_m} | \{U_{i_1} = u_1, \ldots, U_{i_m} = u_m\} \sim \mathcal{B}(\beta_{u_1,\ldots,u_m}),$$

where $\beta_{u_1,\dots,u_m}$ is the probability that m unordered nodes, with latent configuration $\{u_1,\dots,u_m\}$, are connected into a certain hyperedge[3]. Note that $\beta_{u_1,\dots,u_m} = \beta_{u_{\sigma(1)},\dots,u_{\sigma(m)}}$, for each $u_1,\dots,u_m$ and each permutation σ of $1,\dots,m$. The total number of free parameters in the HSB model is equal to $(k-1) + \sum_{m=2}^{M} \binom{k+m-1}{m}$. The probability mass function for the measurement sub-models is

$$p(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{U} = \boldsymbol{u}) = \prod_{m=2}^{M} \prod_{\{i_1,\dots,i_m\}} p(Y_{i_1,\dots,i_m} = y_{i_1,\dots,i_m} \mid U_{i_1} = u_{i_1},\dots,U_{i_m} = u_{i_m});$$

the latent sub-model distribution and the observed network distribution remain as those defined in Equations (3) and (4), respectively. Estimation of model parameters is conducted through the VEM algorithm, relying on the variational procedure defined in Section 2.1.1; see Brusa and Matias (2024) for more details.

This model can be easily extended to handle self-loops, by allowing for $m = 1$, and to weighted hyperedges, by replacing the conditional Bernoulli distribution of the hyperedges with any suitable parametric distribution. For instance, Contisciani, Battiston and Bacco (2022) consider a Poisson HSB model to address the specific case of multi-hyperedges (integer-valued weights). They also propose to reduce the complexity of the inferential procedure by constraining the connectivity parameters to be nonzero only between nodes within the same cluster. Finally, the model has been extended in Chodrow, Veldt and Benson (2021) to the case of multiset hypergraphs, where hyperedges may include repeated nodes with specific multiplicities.

4 Application: Migratory flows

In the following, we present an application of the DSB model related to migration patterns among countries aiming to identify groups of nations with similar migration trends and analyze their changes over time. As illustrated in Section 3.3, in this model nodes are assumed conditionally independent given the latent blocks, and changes are allowed between blocks via a discrete-time Markov chain of first-order. The data are collected by the United Nations, and are publicly available[4], see United Nations, Department of Economic and Social Affairs (2015) for more details. Migration trends are provided every five years, and a time window from 1990 to 2015 is considered in the following application.

The adjacency array, denoted as $\boldsymbol{Y}$, describes the migration patterns over time, with $Y_{ij}^{(t)} \geq 0$ denoting the overall number of individuals born in country j and living in country i at time t, with $i, j = 1,\dots,232$ and $t = 1,\dots,6$. A way to choose a suitable parameterization for the measurement sub-model is to focus on the empirical

[3] For each value of $m = 2,\dots,M$, the set of all probabilities $\beta_{u_1,\dots,u_m}$ is a fully symmetric tensor of rank m.

[4] Data are available at the following website (accessed on September 2024): https://www.un.org/development/desa/pd/content/international-migrant-stock

distribution of the edge weights. Approximately 80.1% of the $Y_{ij}^{(t)}$ entries for all i, j and t are equal to 0, indicating absence of individuals born in country j and living in country i (i.e., the lack of an edge from node i to node j) at time t. For pairs of countries with non-zero migrant stock, we observe the following percentages within these intervals: 7.0% in $(0, 10^2]$, 6.0% in $(10^2, 10^3]$, and 4.2% in $(10^3, 10^4]$. Notably, only 2.7% of entries show values higher than 10^4, with just three exceeding 10^7, corresponding to the Mexican-born individuals residing in the USA in years 2005, 2010, and 2015. Therefore, data present a zero-inflated and left-skewed empirical distribution. As shown in Section 3.3, using the parameterization defined in Equation (8), we account for excess of zeros, and a Box-Cox logarithmic transformation (Box and Cox, 1964) is applied to the non-zero elements of the adjacency array to correct for skewness. The conditional distribution of $Y_{ij}^{(t)}$ given U_i and U_j is then as follows:

$$Y_{ij}^{(t)} | \{U_i^{(t)} = u, U_j^{(t)} = v\} \sim (1 - \beta_{uv})\delta_0 + \beta_{uv} N(\mu_{uv}, \sigma^2),$$

where β_{uv} is the sparsity parameter, and the conditional Gaussian distribution is assumed with means μ_{uv} specific to each pair of latent blocks $u, v = 1, \ldots, k$, and variance σ^2 which is constant across latent blocks. The model is time-homogeneous, as transition probabilities, sparsity parameters, and means are not time-depending.

Once the model is estimated, we consider the following measures to better characterize the blocks that have been identified. Let $i \in \hat{u}^{(t)}$ denote node i belonging to the estimated latent block $\hat{u}$ at time t, and let $n_{\hat{u}^{(t)}}$ be the number of nodes in estimated block $\hat{u}$ at time t. For each block $\hat{u}$ we define the following measures:

(*i*) the average number of edges starting from nodes in a certain latent block $\hat{u}$:

$$\bar{\nu}_{\hat{u}} = \frac{\sum_{t=1}^{T} \sum_{i \in \hat{u}^{(t)}} I(Y_{ij}^{(t)} > 0)}{\sum_{t=1}^{T} n_{\hat{u}^{(t)}}}, \qquad \hat{u} = 1, \ldots, k.$$

With respect to the applicative context, this provides a measure of the average number of distinct countries from which people emigrated to the countries within $\hat{u}$. In this way, it is possible to estimate how many different native country contribute to the migrant population residing in the countries of a certain block;

(*ii*) the average weight of edges starting from nodes in a certain latent block $\hat{u}$:

$$\bar{\mu}_{\hat{u}} = \frac{\sum_{t=1}^{T} \sum_{i \in \hat{u}^{(t)}} Y_{ij}^{(t)}}{\sum_{t=1}^{T} n_{\hat{u}^{(t)}}}, \qquad \hat{u} = 1, \ldots, k.$$

With respect to the applicative context, this provides a measure of the average number of individuals who emigrated to the countries within $\hat{u}$. In this way, it is possible to estimate the overall volume of the migrant population residing in the countries of a certain block;

(*iii*) the destination of edges starting from nodes in certain latent block $\hat{u}$. In this way, it is possible to enumerate the countries of origin of the inhabitants of the countries identified in each block.

4.1 Results

The proposed DSB model is estimated with a number of components k ranging from 1 to 20, with model selection based on the ICL criterion described in Section 2.2. Estimation of the model is carried out in R (R Core Team, 2024) using the `dynsbm` package (Matias and Miele, 2020). Convergence of the VEM algorithm is assumed when the relative difference between the log-likelihood function at two consecutive iterations is smaller than 10^{-8}. The maximum number of iterations for the VEM algorithm is fixed to 25. To mitigate the problem of convergence to local maxima, each model is estimated 50 times, each time using a different initialization as explained in Section 2.1.1. The code to estimate the proposed DSB model is available at link github to be added.

Results are reported in Tables 1, 2, and 3, and in Figures 2 and 3. As shown in Table 1, the minimum value of the ICL criterion is reached for the model with 17 latent blocks. This model shows a log-likelihood value at convergence equal to $-218,992.8$ with 867 parameters. Table 2 reports the main estimated quantities. The average weight $\bar{\mu}$ of edges originating from each of the 17 blocks of the selected model are ordered from highest to lowest and they characterize each identified cluster. Countries are allocated to latent blocks through the MAP approach presented in Section 2.2. Figure 3 shows the world map for years 1990 and 2015, where each country is colored according to the resulting posterior allocation obtained for the initial observational year and the final year.

The 1st group shows the highest values of both the average number and the average weight of edges ($\bar{\nu}_u$ and $\bar{\mu}_u$, respectively). As shown in Table 2, in countries within this block live, on average, around $10,469,159$ people, from a mean of 195 foreign countries. Looking at the first row from the bottom in Figure 2, we notice that countries in this group, which are Australia, Canada, France, the United Kingdom, and the United States since 1990, as well as Italy from 2000, have strong connections with other countries. They are key destinations for global migration flows that extend far beyond their geographical region. Looking at the darkest colors in this row in Figure 2, we notice that the flows from countries in the 2nd block (Germany and Russia), in the 4th block (among which, China, India, and Japan), and in the 8th block (main Latin America countries, such as Mexico) have the highest values.

The 2nd latent block, including Russia and Germany, shows a very high average number of foreign-born residents, around 10,413,101 people. However, unlike the 1st block, this cluster has a lower value of $\bar{\nu}_2$ (equal to 142), indicating fewer and more geographically localized migration flows. The corresponding row in Figure 2 highlights the absence of significant migration from Latin American countries (8th, 14th, and 17th blocks). Countries in the 5th block, such as countries of the ex Union of Soviet Socialist Republics (USSR), excluding the Baltic Republics, present strong connections having extremely high values for both $\hat{\mu}_{2-5}$ and $\hat{\mu}_{5-2}$. This highlights a significant presence in Russia and Germany of individuals born in the countries of the 5th latent block, and also a notable group of people born in Russia and Germany

Table 1 Log-likelihood value at convergence ($\hat{\ell}$), number of parameters and ICL value resulting from fitting the DSB model for different values of k; values in bold identify the selected model

k	$\hat{\ell}$	#par	ICL	k	$\hat{\ell}$	#par	ICL
1	-316,812.5	3	-316,824.5	11	-228,456.8	363	-232,341.4
2	-271,129.6	12	-271,221.8	12	-226,566.2	432	-231,213.4
3	-260,120.0	27	-260,360.7	13	-225,056.2	507	-230,534.2
4	-250,593.8	48	-251,051.2	14	-222,997.9	588	-229,374.9
5	-247,333.8	75	-248,076.0	15	-221,422.0	675	-228,766.2
6	-242,804.3	108	-243,899.8	16	-220,313.0	768	-228,692.6
7	-239,729.9	147	-241,246.7	**17**	**-218,992.8**	**867**	**-228,476.2**
8	-237,650.2	192	-239,656.6	18	-218,236.1	972	-228,891.4
9	-236,338.5	243	-238,902.7	19	-217,793.0	1083	-229,688.5
10	-231,301.6	300	-234,492.0	20	-216,609.9	1200	-229,813.8

who, instead, reside in the ex USSR countries[5]. Countries in the 5th group host an average of around $1,136,739$ immigrants, with migration connections from a more limited number of origin countries ($\bar{\nu}_5 = 24$), primarily from the 2nd and 5th blocks.

The 3rd latent block includes mainly countries of the Arabian Peninsula [6]. These countries show a high number of foreign-born residents ($\bar{\mu}_3 = 2,313,922$) from a relatively low number of countries ($\bar{\nu}_3 = 27$), primarily from the 6th block, including, among others, India, Bangladesh, and Pakistan.

A similar pattern is observed in the 4th group, including South, East, and Southeast Asia, as well as New Zealand. These countries host an average of $1,428,489$ foreign-born residents from a mean of around 35 countries, mostly within the same latent block. There is also a significant presence of individuals from countries in the 11th group, which includes smaller nations in the same geographical region, as well as islands in the Indian and Pacific Oceans.

The 6th latent block stands out for the large number of countries its residents originate from, with an average of 179 source countries (the second highest after the 1st block) and an average foreign population of around $1,005,211$. Countries in this group are mainly from Western and Northern Europe, like the Scandinavian countries, the Netherlands, Belgium, and Spain, alongside Chile and South Africa, which, despite their geographical distance, share similar migration patterns. For instance, in 1990, South Africa had foreign-born residents from 148 countries, while neighboring countries in the 12th group, had significantly fewer, with an average of 16 source countries and a maximum of 33.

[5] For example, in year 1990 Kazakhstan was the second emigration target for Germany, after the United States; see the report of the United Nations, Department of Economic and Social Affairs (2015).

[6] Yemen is the only Arabian country absent from this group, reflecting its distinct migration patterns compared to other nations in the region. Indeed, unlike its Gulf neighbors, Yemen serves primarily as a transit point for migrants rather than a destination (United Nations High Commissioner for Refugees, 2015).

Table 2 Average number ($\bar{\nu}_u$) and weight ($\bar{\mu}_u$) of edges originating from nodes in each latent block u for the DSB model with 17 latent blocks, estimated on the migrant stock data, ranging from 1990 to 2015. For each block, the percentage of countries in the different continents is also reported

u	$\bar{\mu}_u$	$\bar{\nu}_u$	Continent (%)
1	10,469,159	195	Europe: 47%; Americas: 35%; Oceania: 18%
2	10,413,101	142	Europe: 100%
3	2,313,922	28	Asia 100%
4	1,428,489	35	Asia: 92%; Oceania: 8%
5	1,136,739	24	Asia: 69%; Europe: 31%
6	1,005,211	179	Europe: 84%; Americas: 9%; Africa: 7%
7	639,544	126	Europe: 49%; Asia: 31%; Africa: 20%
8	491,058	105	Americas: 100%
9	444,541	13	Asia: 44%; Africa: 38%; Europe: 19%
10	326,613	71	Europe: 84%; Africa: 8%; Asia: 8%
11	302,750	18	Asia: 69%; Africa: 19%; Oceania: 12%
12	291,123	16	Africa: 100%
13	290,413	21	Africa: 100%
14	97,581	36	Americas: 100%
15	28,006	20	Americas 94%; Europe: 6%
16	26,191	8	Oceania: 47%; Africa: 22%; Europe: 19%; Americas: 7%; Asia: 5%
17	17,283	31	Americas: 100%

The remaining European countries are mostly allocated in 7th and 10th latent blocks, having a low quota of foreign-born residents, with an average of $639,543$ in the 7th block and of $326,613$ in the 10th block. The average number of source countries reduces from 126 and 71 for the two blocks, respectively. The 7th group is more heterogeneous and includes European countries such as Poland, Hungary, and the Czech Republic, along with Asian (Turkey, Jordan, and Cyprus) and African countries (Egypt and Libya). Despite being located in different geographical areas, these countries serve as regional migration hubs, hosting moderate numbers of foreign residents from a broad range of origin countries. The 10th block stands out for the presence, alongside European countries such as Croatia, Slovenia, and Romania, of Namibia. Unlike other southern African nations, indeed, Namibia is a destination for migrants from 83 countries, well above the regional average of 36, though still below South African levels. Its historical ties to Germany (as a former colony) determine its grouping in this community.

Latin American countries are primarily in the 8th, 14th, 15th, and 17th latent blocks. Among them, only the 8th group, including main South American countries, displays a notable presence of foreign-born residents, with an average of around 491,058 individuals coming from approximately 105 source countries. Significant migration flows are particularly evident between countries within this block, as well as from nations in the 14th block including Cuba and other Caribbean countries.

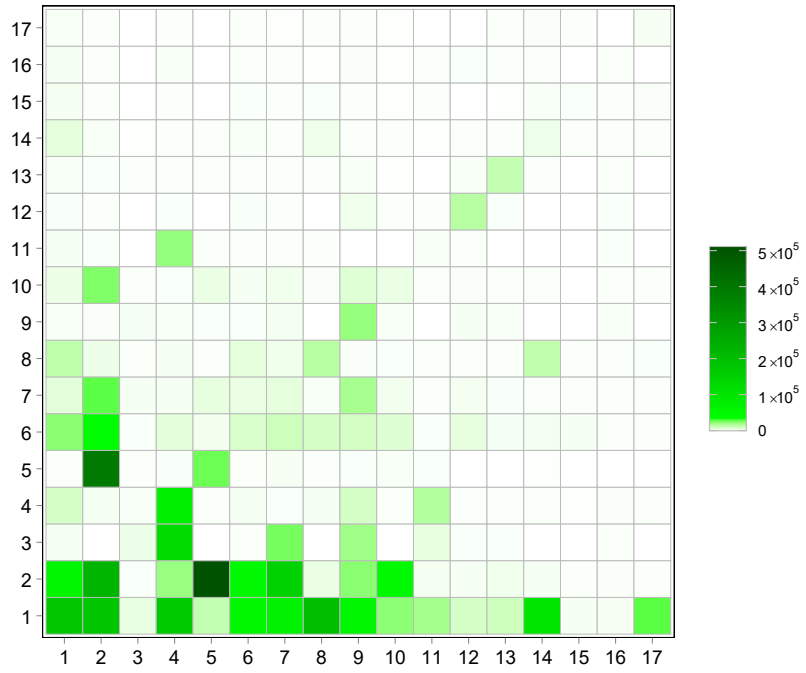

Fig. 2 Estimated average weight ($\hat{\mu}_{u-v}$) of edges from block u to block v for the DSB model with 17 latent blocks estimated on the migrant stock data ranging from 1990 to 2015

Most of the remaining African countries are grouped into the 12th and 13th latent blocks. In both cases, the only significant connections observed are between countries within the same block. Both blocks have an average foreign-born population of approximately 290,000, with residents originating from an average of 16 and 21 countries, respectively. The 9th latent block, despite comprising countries from various regions, is characterized by Muslim-majority populations. This block spans North Africa and Asia and includes Bosnia and Herzegovina and Albania, the only Muslim-majority countries in Europe. While the average number of origin countries is relatively low (13), this cluster has a notably high foreign-born population, averaging approximately 444, 541. Finally, the 16th group has the lowest value of $\bar{v}_u$, indicating the smallest number of origin countries. This heterogeneous group spans multiple continents and regions and is characterized by exceptionally low migration volumes.

Considering the estimated transition probability matrix, which is not reported due to space constraints, we notice high persistence in each latent block; Table 3 reports the countries estimated to transit across blocks over the observed period. The 7th cluster, shows the highest transition probability, equal to 0.08. Italy transitions from the 6th to the 1st group, reflecting a rise in foreign-born residents, while countries

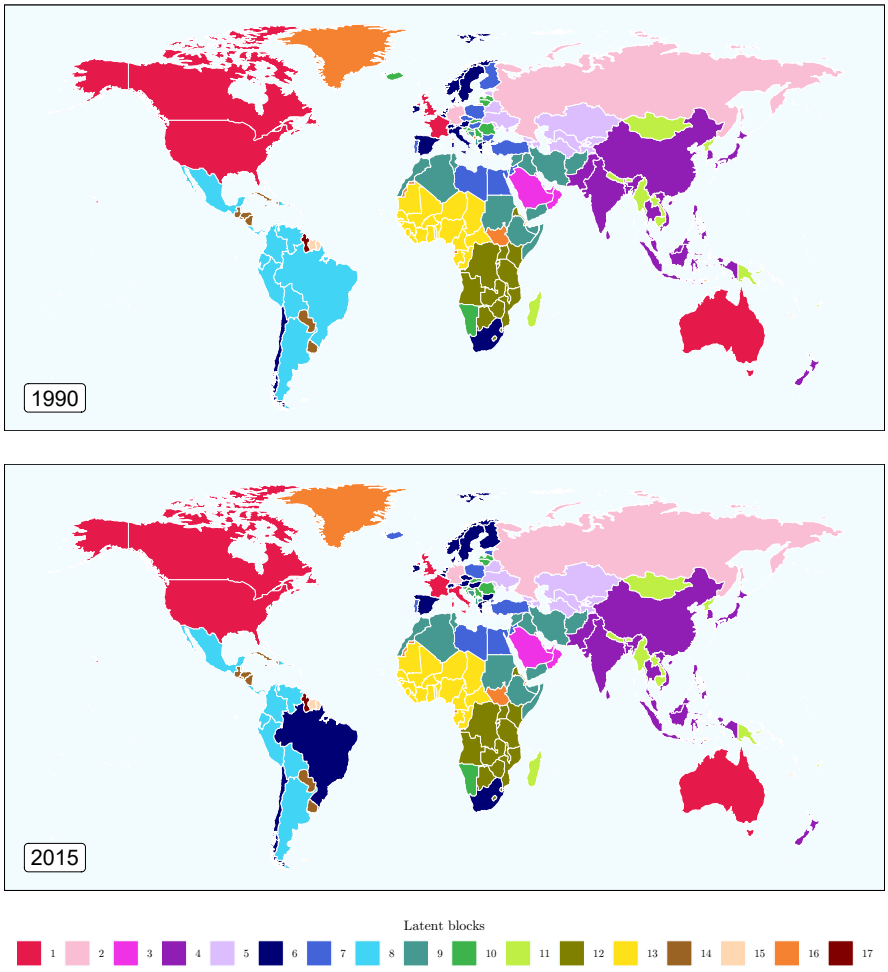

Fig. 3 World map colored according to the classification of each country within the latent blocks obtained using the MAP approach under the selected DSB model. The maps depict the classifications for the years 1990 (top) and 2015 (bottom)

like Brazil, Bulgaria, and Hungary move to the 6th group. Montenegro also shifts from a very low-migration latent block to the 9th, marking a significant change in its migration patterns.

Table 3 Countries transitioning between estimated clusters over the observed period from 1990 to 2015 according to the MAP approach under the estimated DSB model

Country	1990	1995	2000	2005	2010	2015
Brazil	8	8	8	8	6	6
Bulgaria	7	7	7	6	6	6
Czech Republic	7	7	7	6	6	6
Estonia	5	5	5	5	7	7
Finland	7	7	6	6	6	6
Hungary	7	7	6	6	6	6
Iceland	10	7	7	7	7	7
Italy	6	6	1	1	1	1
Montenegro	16	16	16	16	10	10

5 Concluding remarks

Stochastic block (SB) models are advanced statistical methods for analyzing complex network data. In this paper, we focus on their specific formulation in terms of discrete latent variables, which offers a flexible distribution and enhances the interpretability of the resulting inferential procedures. Starting from the basic formulation of the SB model for simple graphs, we have presented a brief review of various extensions to encompass binary, weighted, and dynamic networks, along with more complex structures such as multiplex and multipartite networks, and hypergraphs. In particular, we have focused on the maximum likelihood estimation approach using a variational approximation of the expectation-maximization (VEM) algorithm, the integrated classification likelihood (ICL) criterion for the selection of the number of latent blocks, and the maximum a-posteriori rule for the prediction of latent states for each node to uncover the estimated hidden structure of the network. The usefulness of the dynamic stochastic block (DSB) model is shown through the analysis of global migration data. Using a weighted DSB model, we identified 17 latent blocks of countries with similar migration patterns and tracked their evolution from 1990 to 2015. The obtained clusters highlighted key migration destinations, such as the United States, France, and Australia, which attract migrants from a wide range of countries far beyond their own geographical regions. In contrast, they also revealed groups of countries with smaller, more localized migration flows, mostly confined within the same geographical area, such as Latin American nations and Arabian Peninsula countries.

Considering a first future research direction that can be explored, convergence to local maxima is a significant challenge in the estimation of all the aforementioned SB models. As discussed in Section 2.1.1, the evolutionary version of the VEM algorithm has proven effective in addressing this issue for simple and dynamic networks (Brusa and Pennoni, 2024), and its application should be further investigated for more complex network structures. The evolutionary approach could also be extended to

select the number of latent blocks as an alternative to the ICL criterion. Another aspect of research that needs to be addressed is the scalability of the estimation algorithms, in order to handle the analysis of large-scale networks. Finally, we mention that, even though in this work we focused solely on maximum likelihood estimation, these models can also be estimated using a Bayesian approach, typically via Markov chain Monte Carlo; we refer, among others, to Nowicki and Snijders (2001), McDaid, et al. (2013), Peixoto (2017), and Ludkin, Eckley and Neal (2018).

Acknowledgements The authors acknowledge the financial support from the grant "Hidden Markov Models for Early Warning Systems" of Ministero dell'Università e della Ricerca (PRIN 2022TZEXKF) funded by European Union - Next Generation EU, Component 2, Mission 4.

References

Aicher, C., A. Z. Jacobs, and A. Clauset. 2015. Learning latent block structure in weighted networks. *IEEE Journal of Complex Networks 3*, 221–248.

Ambroise, C. and C. Matias. 2012. New consistent and asymptotically normal parameter estimates for random-graph mixture models. *Journal of the Royal Statistical Society, Series B 74*, 3–35.

Ashlock, D. 2004. *Evolutionary Computation for Modeling and Optimization*. New York: Springer.

Bar-Hen, A., P. Barbillon and S. Donnet. 2022. Block models for generalized multipartite networks: Applications in ecology and ethnobiology. *Statistical Modelling 22*, 273–296.

Barbillon, P., S. Donnet, E. Lazega and A. Bar-Hen. 2017. Stochastic block models for multiplex networks: an application to a multilevel network of researchers. *Journal of the Royal Statistical Society, Series A 180*, 295–314.

Bartolucci, F. and S. Pandolfi. 2020. An exact algorithm for time-dependent variational inference for the dynamic stochastic block model. *Pattern Recognition Letter 138*, 362–369.

Bartolucci, F., A. Farcomeni and F. Pennoni. 2013. *Latent Markov Models for Longitudinal Data*. Boca Raton: Chapman and Hall/CRC.

Bartolucci, F., S. Pandolfi and F. Pennoni. 2022. Discrete latent variable models. *Annual Review of Statistics and its Application 9*, 425–452.

Battiston, F., G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young and G. Petri. 2020. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports 874*, 1–92.

Bick, C., E. Gross, H. A. Harrington and M. T. Schaub. 2023. What are higher-order networks?. *SIAM Review 65*, 686–731.

Biernacki, C., G. Celeux, and G. Govaert. 2000. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 719–725.

Blei, D. M., A. Kucukelbir and J. D. McAuliffe. 2017. Variational inference: A review for statisticians. *Journal of the American Statistical Association 112*, 859–877.

Box, G. E. and D. R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B 26*, 211–243.

Brusa, L. 2023. Developments in discrete latent variable models: dealing with likelihood multimodality and clustering of simple hypergraphs. Ph.D. Thesis, University of Milano-Bicocca.

Brusa, L. and C. Matias. 2024. Maximum likelihood estimation for discrete latent variable models via evolutionary algorithms. *Scandinavian Journal of Statistics 51*, 1661–1684.

Brusa, L. and F. Pennoni. 2024. Variational inference for estimating dynamic stochastic block models through an evolutionary algorithm. *Submitted*.

Brusa, L., F. Pennoni, and F. Bartolucci. 2024. Maximum likelihood estimation for discrete latent variable models via evolutionary algorithms. *Statistics and Computing 34*, 62.

Chodrow, P. S., N. Veldt, and A. R. Benson. 2021. Generative hypergraph clustering: From blockmodels to modularity. *Science Advances 7*, eabh1303.

Contisciani, M., F. Battiston, and C. De Bacco. 2022. Inference of hyperedges and overlapping communities in hypergraphs. *Nature Communications 13*, 7229.

Daudin, J. J., F. Picard, and S. Robin. 2008. A mixture model for random graphs. *Statistics and Computing 18*, 173–183.

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

Dormann, C. F., and R. Strauss. 2014. A method for detecting modules in quantitative bipartite networks. *Methods in Ecology and Evolution 5*, 90–98.

Du Bois, C., C. T. Butts and P. Smyth. 2013. Stochastic blockmodeling of relational event dynamics. *Journal of Machine Learning Research 31*, 238–246.

Everitt B. 1984. An Introduction to Latent Variable Models. Boca Raton: Chapman and Hall/CRC.

Fortunato S. 2010. Community detection in graphs. *Physics Reports 486*, 75–174.

Ghasemian, A., P. Zhang, A. Clauset, C. Moore and L. Peel. 2016. Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Physical Review X 6*, 031005.

Goodman, L. A. 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika 61*, 215–231.

Holland, P. W., K. B. Laskey, and S. Leinhardt. 1983. Stochastic blockmodels: First steps. *Social Networks 5*, 109–137.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola and L. K. Saul. 1999. An introduction to variational methods for graphical models. *Machine Learning 37*, 183–233.

Karrer, B. and M. E. J. Newman. 2011. Stochastic blockmodels and community structure in networks. *Physical Review E 83*, 016107.

Kivelä M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno and M. A. Porter. 2014. Multilayer networks. *Complex Networks 2*, 203–271.

Kullback, S. and R. A. Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics 22*, 79–86.

Kurihara, K., Y. Kameya and T. Sato. 2006. A frequency-based stochastic blockmodel. In:*Workshop on information-based induction sciences*.

Lee, C. and D. J. Wilkinson. 2019. A review of stochastic block models and extensions for graph clustering. *Applied Network Science 4*, 122.

Ludkin, M., I. Eckley and P. Neal. 2018. Dynamic stochastic block models: Parameter estimation and detection of changes in community structure. *Statistics and Computing 28*, 1201–1213.

Mariadassou, M., S. Robin and C. Vacher. 2010. Uncovering latent structure in valued graphs: A variational approach. *The Annals of Statistics 4*, 715–742.

Matias, C. and V. Miele. 2017. Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society, Series B 79*, 1119–1141.

Matias, C. and V. Miele. 2020. `dynsbm`: Dynamic Stochastic Block Models. *R package version 0.7*, https://CRAN.R-project.org/package=dynsbm.

Matias, C. and S. Robin. 2014. Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM: Proceedings and Surveys 47*, 55–74.

Matias, C., T. Rebafka and F. Villers. 2018. A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika 105*, 665–680.

McDaid, A. F., T. B. Murphy, N. Friel and N. J. Hurley. 2013. Improved Bayesian inference for the stochastic block model with application to large networks. *Computational Statistics & Data Analysis 60*, 12–31.

McLachlan, G. and D. Peel. 2000. *Finite Mixture Models*. New York: John Wiley.

Newman, M. E. J. 2004. Fast algorithm for detecting community structure in networks. *Physical Review E 69*, 066133.

Newman, M. E. J., S. H. Strogatz and D. J. Watts. 2001. Random graphs with arbitrary degree distributions and their applications. *Physical Review E 64*, 026118.

Nowicki, K. and T. A. B. Snijders. 2001. Estimation and prediction for stochastic blockstructure. *Journal of Classification 14*, 75–100.

Paul, S. and Y. Chen. 2016. Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electronic Journal of Statistics 10*, 3807–3870.

Peixoto, T. P. 2014. Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X 4*, 011047.

Peixoto, T. P. 2015. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Physical Review E 92*, 042807.

Peixoto, T. P. 2017. Nonparametric Bayesian inference of the microcanonical stochastic block model. *Physical Review E 95*, 012317.

Shannon, C. E. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal 27*, 379–423.

Snijders, T. A. B. and K. Nowicki. 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification 14*, 75–100.

Stanley, N., S. Shai, D. Taylor and P. J. Mucha. 2016. Clustering network layers with the strata multilayer stochastic block model. *IEEE Transactions on Network Science and Engineering 3*, 95–105.

R Core Team. 2024. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Titterington, D. M., A. F. M. Smith and E. Mu. 1985. *Statistical analysis of finite mixture distributions*. New York: John Wiley.

United Nations, Department of Economic and Social Affairs. 2015. Trends in International Migrant Stock: Migrants by Destination and Origin. *United Nations database*, POP/DB/MIG/Stock/Rev.2015.

United Nations High Commissioner for Refugees. 2015. Mixed migration update: Yemen. https://www.unhcr.org/media/yemen-mixed-migration-update.

Vallès-Català, T., F. A. Massucci, R. Guimerà and M. Sales-Pardo. 2016. Multilayer stochastic block models reveal the multilayer structure of complex networks. *Physical Review X 3*, 011036.

Xing, E. P., W. Fu and L. Song. 2010. A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics 4*, 535–566.

Xu, K. S. and A. O. Hero III. 2014. Dynamic stochastic blockmodels for time-evolving social networks. *IEE Journal on Selected Topics in Signal Processing 8*, 552–562.

Yan, X., C., Shalizi, J. E. Jensen, F. Krzakala, C. Moore, L. Zdeborová, P. Zhang and Y. Zhu. 2014. Model selection for degree-corrected block models. *Journal of statistical mechanics 5*, P05007.

Yang, X. 2017. Hypergraph partitioning for social networks based on information entropy modularity. *Journal of Network and Computer Applications 86*, 59–71.

Yang, T., Y., Chi, S. Zhu, Y. Gong and R. Jin. 2011. Detecting communities and their evolutions in dynamic social networks - a Bayesian approach. *Machine Learning 82*, 157–189.

Zhang, X., C., Moore and M. E. J. Newman. 2017. Random graph models for dynamic networks. *The European Physical Jouornal B 90*, 200.

Zhao, Y., E., Levina and J. Zhu. 2012. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics 40*, 2266–2292.