UNIVERSITÀ COMMERCIALE “LUIGI BOCCONI”
PHD SCHOOL

PhD program in: Statistics and Computer Science
Cycle: `XXXVII`
Disciplinary Field: `SECS-S/01`

Bayesian hierarchical modeling of array-structured demographic data

Advisor: Daniele Durante
Co-Advisor: Emanuele Aliverti

PhD Thesis by
Giovanni Romanò
ID number: 3172379

Year: 2026

Abstract

Reliable inference for complex demographic phenomena is essential for understanding population health dynamics and informing public policies. The relevance of this endeavor has stimulated increasing interest in rigorous statistical models for high-dimensional demographic data with complex dependence structures. Recent research has provided valuable insights and effective predictive strategies, but it often focuses on specific dimensions at the expense of others. This perspective rules out the possibility to infer more nuanced, yet fundamental, demographic patterns that span multiple dimensions (e.g., calendar years, age classes, causes of death, countries). We contribute to this line of research by developing novel hierarchical Bayesian procedures for joint modeling of cross-sectional and temporal interactions in array-structured demographic data. This thesis addresses, in particular, three main objectives through state-of-the-art methodologies accounting for the core characteristics of demographic processes, while incorporating dynamic partitioning mechanisms.

First, we propose a flexible model for age-period log-mortality rates inducing local clusters of countries. To address the functional nature of the age component, we employ B-spline expansions with dynamic coefficients. Local clustering of the log-mortality rates is achieved through a dependent random partition model on the coefficients, which allows grouping structures to vary flexibly across different combinations of ages and periods. We unveil unexplored relationships between countries, opening new directions for demographic research.

Second, we extend stochastic block models to analyze sequences of directed networks encoding co-occurrences of underlying and contributing causes of death. We handle categorically weighted edges by assuming block-specific Categorical-Dirichlet distributions, implement a double-partition framework to account for asymmetric relationships between underlying and contributing causes, and describe the node clusters through dependent random partitions to ensure the smooth evolution of block structures across age classes. Application to 2019 US data reveals cause partitioning that moves beyond traditional medical classifications toward more nuanced groupings.

Third, we develop a methodology for dynamic clustering of countries based on their age-specific leading cause of death sequences over time. We model each country's sequence as a categorical trajectory and handle their grouping through a mixture model with exponential-distance components based on Hamming distances, which enables the characterization of clusters through a modal sequence and scale parameters describing the heterogeneity of each age-group. The induced partition is allowed to evolve smoothly across years through a temporal random partition model, enabling the identification of clustering structures in leading mortality causes which extend standard epidemiological transition theories.

Contents

Introduction

Reliable estimates of demographic outcomes are essential for understanding population health trends, shaping health and public policy, as well as forming the basis of inference in all areas of social science (Murray and Evans, 2003; Papanicolas et al., 2022). Traditional approaches often struggle to capture the complex dependencies inherent in mortality patterns across regions, age groups, calendar years, or causes of death. The increasing need for more sophisticated models that can accurately represent these intricate relationships motivates the development of new statistical methods and models tailored to the challenges of complex data, such as those arising in a demographic setting.

Recent advances have yielded fundamental insights and successful predictive methods, but often focus on specific dimensions at the expense of others, for instance, examining patterns for isolated countries (Arnold (-Gaille) and Sherris, 2013; Désesquelles et al., 2014b; Woolf and Schoomaker, 2019; Camarda, 2019; Alexopoulos et al., 2019; Pavone et al., 2024), limited to a specific calendar year (Désesquelles et al., 2014b; Egidi et al., 2018) or particular age group populations (Egidi et al., 2018; Woolf and Schoomaker, 2019). As showcased in recent studies (Bergeron-Boucher et al., 2020; Stefanucci and Mazzuco, 2022; Aliverti et al., 2022; Debon et al., 2023; Depaoli et al., 2024; Dharamshi et al., 2025; Varga, 2025), a deeper understanding of demographic phenomena requires comprehensive formulations capable of jointly modeling patterns across multiple dimensions (e.g., countries, age classes, time and causes of death) of demographic data, opening the possibility to explore more complex, yet critical, demographic structures. Our contribution to this line of research lies in developing novel hierarchical Bayesian procedures that simultaneously model cross-sectional and temporal interactions in array-structured demographic data. The thesis tackles three primary objectives using state-of-the-art methodologies that capture the essential features of demographic processes while incorporating dynamic partitioning mechanisms.

First (Chapter 1), we aim to identify and quantify relationships between mortality rates across countries, age classes, sexes, and calendar years, with a focus on the temporal and spatial dimensions of these dependencies. Second (Chapter 2), we investigate the connections between underlying and contributing causes of death in the US population, particularly examining how their mutual influence evolves from the youngest ages to the oldest one. Finally (Chapter 3), we analyze the sequences of the most frequent cause of death in various age classes across different countries worldwide, stratified by sex, aiming to identify clusters of populations with similar trajectories across ages and to track how these patterns have changed over the past two decades.

Although the entities and phenomena under consideration differ, the three contributions share the common goal of identifying communities of homogeneous statistical units with respect to specific characteristics of the problem under investigation, a task known in the statistical literature as *clustering* (see Hennig et al., 2015). Our research is not confined to the analysis of isolated, static, cross-sectional groups (e.g., sets of countries in a given calendar year), rather our interest lies in uncovering the dynamics that drive the evolution of such groups along ordered dimensions, such as time or age. Effectively addressing these objectives requires models that account for both the core structure of the demographic problems and the dynamic clustering patterns among observations, achieved by introducing a suitable latent partitioning mechanism.

Our contribution consists in the design of novel principled Bayesian models for dynamic clustering of mortality data, with applications to log-mortality rates, co-occurrences of causes of death, and life-course trajectories of predominant causes. We now provide a brief overview of the state-of-the-art dynamic clustering methodology employed in our proposals (Page et al., 2022). Finally, we conclude the introduction by summarizing the key contributions of this thesis in the three applications.

Dynamic Clustering

Clustering techniques can be divided into two main categories (Hennig et al., 2015). The first consists of algorithmic distance-based methods, like k-means (MacQueen, 1967) and hierarchical clustering (Ward, 1963), which are procedures that group observations together by measuring how similar they are using a specific distance measure. The second involves model-based schemes that rely on probabilistic frameworks to assign data points to clusters based on distributional assumptions. While these approaches can be implemented from either a frequentist (Pearson, 1894; Day, 1969; Wolfe, 1970; O'Neill, 1978) or Bayesian perspective (Antoniak, 1974; Binder, 1978; Lavine and West, 1992; Diebolt and Robert, 1994), in this thesis we adopt the latter which allows for the incorporation of prior information on the latent partitioning structure as well as the use of posterior distributions for uncertainty quantification.

Routine implementations of model-based clustering rely on mixture representations (see Fruhwirth-Schnatter et al., 2015, for a complete review). Given a random sample $y_1, \ldots, y_n$, a standard mixture model for y_i, $i = 1, \ldots, n$, assumes

$$y_i \overset{\text{ind}}{\sim} \sum_{k=1}^{K} \pi_k p_k(y_i \mid \theta_k),$$

where $p_1(\cdot \mid \theta_1), \ldots, p_K(\cdot \mid \theta_K)$ are the K component distributions indexed by, possibly array-valued, parameters $\theta_1, \dots, \theta_K$, and the set of weights satisfies $\sum_{k=1}^K \pi_k = 1$. Often, the K distributions are assumed to belong to the same family, so $p_k(\cdot \mid \theta_k) \equiv p(\cdot \mid \theta_k)$. Clusters are implicitly induced, depending on which of the K components each observation is assigned to.

An equivalent, two-step description of this formulation that more explicitly models the clustering assignment is available. The first step consists in allocating the observations to one of K possible groups, each with probability π_k, for $k = 1, \dots, K$. Then, conditionally on the group allocations, observations within the generic group k are assumed to be independently and identically distributed according to $p(\cdot \mid \theta_k)$. More formally,

$$y_i \mid c_i = k \overset{\text{ind}}{\sim} p(\cdot \mid \theta_k) \qquad \text{with} \qquad \text{pr}(c_i = k) = \pi_k$$

where $c_1, \dots, c_n$ are latent allocation variables that induce the partition of the n observations $\mathbb{c} = \{\mathcal{C}_1, \dots, \mathcal{C}_K\}$, such that $\mathcal{C}_k = \{i = 1, \dots, n : c_i = k\}$, for $k = 1, \dots, K$.

This alternative formulation naturally leads to another class of widely used models known as random partition models (see, e.g., Grazian, 2025, for a review of recent advances). These models explicitly specify a distribution for the partition $\mathbb{c}$ of data points, thereby replacing the multinomial allocation $\text{pr}(c_i = k) = \pi_k$ with a more flexible framework while preserving the iid assumptions for observations within clusters.

To this end, Gibbs-type priors (see, e.g., De Blasi et al., 2015) provide a rich family of Bayesian nonparametric distributions on exchangeable random partitions that combine flexibility with analytical tractability. A probability distribution for a vector of n unlabeled group indicators $\boldsymbol{c}$ belongs to this family if

$$p(\boldsymbol{c}) = \beta_{n,K} \prod_{k=1}^{K} (1 - \sigma)_{n_k - 1},$$

where n_k denotes the number of individuals in cluster k, $\sigma < 1$ is the discount parameter, $(a)_b = a(a+1) \cdots (a+b-1)$ is the rising factorial, and $\{\beta_{n,K} : 1 \leq K \leq n\}$ represents a set of nonnegative weights satisfying the recursion $\beta_{n,K} = (n - K\sigma)\beta_{n+1,K} + \beta_{n+1,K+1}$, with initial condition $\beta_{1,1} = 1$. This family includes some commonly used examples such as the Gnedin process (obtained by setting $\sigma = -1$ and $\beta_{n,K} = (\eta)_{n-K}[\prod_{k=1}^{K-1}(k^2 - \eta k)] / \prod_{i=1}^{n-1}(i^2 + \eta i)$ for $\eta \in (0, 1)$) and the Dirichlet process (obtained by setting $\sigma = 0$ and $\beta_{n,K} = \eta^K / (\eta)_n$, for $\eta > 0$) (De Blasi et al., 2015; Legramanti et al., 2022). Moreover, this class also comprises the mixtures of Dirichlet processes with respect to η, enabling straightforward inference on the parameter that governs the number and size of induced clusters. Furthermore, under appropriate conditions, both the Gnedin process (Geng et al., 2019) and the mixtures of Dirichlet processes (Ascolani et al., 2023) provide consistent estimators for the number of clusters, a highly desirable property in clustering applications.

It is worth noting that the Gibbs-type priors enjoy an equivalence property with another widespread class of models, known as exchangeable product partition models (PPMs; Hartigan, 1990). These models assign probability to a partition $\mathbb{c}$ as $p(\mathbb{c}) \propto c(\mathcal{C}_1) \cdots c(\mathcal{C}_K)$, where $c(\cdot)$ is a nonnegative *cohesion* function that quantifies within-cluster homogeneity. If we impose the cohesion function $c(\cdot)$ to depend on $\mathcal{C}_1, \dots, \mathcal{C}_k$ only through their cardinality, then these distributions coincide with the Gibbs-type ones (De Blasi et al., 2015). This equivalence is particularly advantageous, as it allows for the extension of the model to include observation-specific information via an auxiliary similarity function (Müller et al., 2011), which can incorporate external knowledge, such as covariates and expert opinion, into the prior on the partition.

While traditional clustering methods focus on identifying static groupings within a single cross-section of the data, our primary interest lies in studying the evolution of partitions across ordered dimensions, such as age groups or calendar years. This dynamic perspective enables us to capture how clusters form, dissolve, or change composition over time or age, providing deeper insights into the underlying demographic processes. Denoting with $\boldsymbol{c}_t = (c_{t,1}, \dots, c_{t,n})$ the latent allocation variables at time t, we consider a Markovian structure for the sequence of allocations, such that the joint distribution of the sequence factorizes as

$$p(\boldsymbol{c}_1, \dots, \boldsymbol{c}_T) = p(\boldsymbol{c}_1) \prod_{t=2}^{T} p(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1}) \tag{1}$$

where $p(\boldsymbol{c}_1)$ is the initial distribution and $p(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1})$ is the transition kernel determining how the partition evolves over $t \in \{2, \dots, T\}$.

To completely characterize such a prior distribution, we need to specify explicit forms for the initial configuration and transition probabilities. To this end, we rely on the time-dependent random partition model proposed by Page et al. (2022)

$$([\mathbf{c}_1, \dots, \mathbf{c}_T] \mid \boldsymbol{\alpha}) \sim \text{tRPM}(\boldsymbol{\alpha}, \eta)$$

which extends the random partition models framework to account for the temporal evolution of clustering patterns through an intuitive yet flexible two-step procedure where only a subset of the observations is allowed to change their cluster allocation at each time point. The vector $\boldsymbol{\alpha} = [\alpha_2, \dots, \alpha_T]$ is a set of time-specific parameters that regulate the temporal persistence of the partitions $\boldsymbol{c}_1, \dots, \boldsymbol{c}_T$ through the transition mechanism underlying $p(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1})$. On the other hand, $\eta > 0$ controls the number and size of the clusters at the first time point $t = 1$ through $p(\boldsymbol{c}_1)$, which also reflects on the subsequent time points, under the Markovian assumption in (1). The parameter η is here assumed to be user-defined, as in Page et al. (2022), and hence does not appear in the conditioning, but we acknowledge the possibility of placing a prior on η to allow for full Bayesian inference on the number and size of the clusters (see Chapter 1 and 2).

An EPPF is assumed for the distribution $p(\boldsymbol{c}_1)$ of the partition at first time point $t = 1$, for instance, one induced by a member of the Gibbs-type prior family. A useful property for $p(\boldsymbol{c}_1)$ is the so-called *Kolmogorov consistency*, or *addition rule*, property, which guarantees the closure of the model family with respect to the marginalization of some units; see, e.g., De Blasi et al. (2015) and Page et al. (2022). In such a setting, this is crucial to provide a characterization of the sequence of distributions, as Page et al. (2022) shows that, under this property, the tRPM construction produces partitions at each subsequent time point that are marginally identically distributed to the initial one. In other words, $\boldsymbol{c}_2, \dots, \boldsymbol{c}_T$ all inherit the same EPPF that was assumed for $\boldsymbol{c}_1$.

Then, subsequent transitions are governed by latent binary indicators γ_{it}, for $i = 1, \dots, n$ and $t = 1, \dots, T$, each determining whether the i–th unit can change its cluster allocation when moving from $t-1$ to t. To allow for a completely unconstrained allocation at the first step $t = 1$, we set $\gamma_{i1} = 0$ for every $i = 1, \dots, n$. Conditionally on the cluster memberships at the previous time step $\boldsymbol{c}_{t-1}$ and the constrained units $\Gamma_t = \{i : \gamma_{it} = 1\}$, the allocations $\boldsymbol{c}_t$ at time t are driven by a mechanism which must satisfy a *compatibility* criterium. Let denote with $\mathbb{C}$ the space of all the partitions of n nodes, $\mathbb{c}_t \in \mathbb{C}$ the partition corresponding to cluster allocations $\boldsymbol{c}_t$ and $\mathbb{c}_t^{\Gamma_t}$ the reduced partition obtained by removing from $\mathbb{c}_t$ the observations not belonging to the index set Γ_t. Then, we define $\mathbb{C}(\boldsymbol{c}_{t-1}, \boldsymbol{\gamma}_t) = \{\mathbb{c}_t \in \mathbb{C} : \mathbb{c}_{t-1}^{\Gamma_t} = \mathbb{c}_t^{\Gamma_t}\}$ as the set of partitions induced by $\boldsymbol{c}_t$ *compatible* with $\boldsymbol{c}_{t-1}$ and $\boldsymbol{\gamma}_t$. Specifically, these are all the partitions of the n units that can be obtained from $\boldsymbol{c}_{t-1}$ by reallocating only those units for which $\gamma_{it} = 0$. Then, for the joint distribution of $(\boldsymbol{\gamma}_t, \boldsymbol{c}_t)$ over $t = 1, \dots, T$, we assume the conditional Markovian factorization

$$p(\boldsymbol{\gamma}_1, \boldsymbol{c}_1, \dots, \boldsymbol{\gamma}_T, \boldsymbol{c}_T) = p(\boldsymbol{c}_1) \prod_{t=2}^{T} p(\boldsymbol{\gamma}_t)\, p(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1}, \boldsymbol{\gamma}_t).$$

For all the auxiliary variables with $t > 1$, we consider independent Bernoulli distributions with time-specific parameter $\alpha_t \in [0, 1]$, such that α_t plays the role of scalar dependence parameter modeling the a priori expected relative frequency of units that can be reallocated in the transition from $t-1$ to t. As such, $\alpha_t = 0$ implies almost sure independence between $\mathbb{c}_t$ and $\mathbb{c}_{t-1}$, conversely, $\alpha_t = 1$ implies that $\mathbb{c}_t = \mathbb{c}_{t-1}$ with probability 1. The formulation is completed with independent and identically distributed Beta priors for α_t, thus obtaining the hierarchical formulation

$$(\gamma_{it} \mid \alpha_t) \overset{\text{iid}}{\sim} \text{Ber}(\alpha_t), \qquad \alpha_t \overset{\text{iid}}{\sim} \text{Beta}(a_\alpha, b_\alpha),$$

for $i \in \{1, \dots, n\}$ and $t \in \{2, \dots, T\}$, where $a_\alpha > 0$ and $b_\alpha > 0$ are non-stochastic prior hyperparameters controlling the expected number of nodes that can be reallocated at t. To conclude the specification, the transition kernel is assumed to have the form

$$\text{pr}(\boldsymbol{c}_t = \boldsymbol{c} \mid \boldsymbol{c}_{t-1}, \boldsymbol{\gamma}_t) \propto \text{pr}(\boldsymbol{c}_t = \boldsymbol{c})\, \mathbb{1}[\mathbb{c} \in \mathbb{C}_t(\boldsymbol{c}_{t-1}, \boldsymbol{\gamma}_t)]$$

where $\mathrm{pr}(\boldsymbol{c}_t = \boldsymbol{c})$ is the EPPF assumed for $\boldsymbol{c}_1$ evaluated at the cluster allocation vector $\boldsymbol{c}$ inducing the partition $\mathbb{c} \in \mathbb{C}$, and $\mathbb{1}[A]$ denotes the indicator function of the event A. This is equivalent to assume that the transition to $\boldsymbol{c}_t$ shares the same distributional behavior as $\boldsymbol{c}_{t-1}$, only restricted to the set of $\boldsymbol{c}$ that can be obtained starting from $\boldsymbol{c}_{t-1}$ and satisfying the compatibility constraints defined by $\boldsymbol{\gamma}_t$.

Summary of the Specific Contributions

In Chapter 1, we propose a flexible model for log-mortality rates that induces local clusters of countries across ages and time points (Romanò et al., 2025). To take into account the functional nature of the data, our approach is based on B-splines expansions, with dynamic coefficients that model the evolution of the temporal trajectories (Pavone et al., 2024). This solution also addresses the high-dimensionality problem affecting age-and-year-specific mortality data (Booth and Tickle, 2008), by analyzing coefficients of the spline expansion that capture mortality behavior for classes of ages, rather than single-age rates. To allow for local partitioning patterns of countries that are specific to combinations of age classes and periods, we cluster the spline coefficients using the dynamic random partition model by Page et al. (2022).

In Chapter 2, we enlarge the class of extended stochastic block models (ESBM) introduced by Legramanti et al. (2022). Our extension (Romanò et al., 2025) infers dynamic groups of nodes with common connectivity patterns in a categorically weighted directed network sequence, motivated by the one encoding the co-occurrences of underlying and contributing causes of death in the 2019 US population across different ages. We introduce several innovations to the ESBM framework. First, we account for networks whose edges are labeled by categorical variables. Second, we include a double partition framework to take into account the directed nature of the problem. This aspect is fundamental in settings with asymmetric relationships, such as in social networks like Instagram, where we want to cluster individuals with two separate partitions depending on their role of *follower* or *followed*. Third, we model the sequence of networks through the dependent random partition proposed by Page et al. (2022) to guarantee the smooth evolution of block structure between adjacent age classes, something unattainable simply by taking a set of independent Gibbs-type priors.

In Chapter 3, we introduce a Bayesian methodology for the dynamic clustering of countries based on their age-specific leading cause-of-death sequences, with clustering patterns that evolve over time. Leveraging the comprehensive WHO mortality database, we analyze annual leading cause-of-death sequences across 183 countries, stratified by sex and age class, from 2000 to 2021. Our approach models each country's sequence across ages as a trajectory of categorical variables and employs the mixture model proposed by Argiento et al. (2024), with exponential-distance components based on the Hamming distance to capture similarities between categorical sequences. We extend this formulation to account for the temporal evolution of countries' similarity patterns, by adopting the temporal random partition model (tRPM) of Page et al. (2022), which allows the clusters to evolve smoothly across years. This framework enables the identification of both persistent and transient clustering structures, providing new insights into global and local trends in cause-specific mortality. The proposed methodology facilitates the detection of converging or diverging epidemiological patterns among countries and offers a principled approach for quantifying uncertainty in both cluster assignments and central cause-of-death trajectories.

Through these novelties, this thesis seeks to contribute to both the formal modeling of complicated dependence structures in complex data that are more and more common, such as functional data, networks, and categorical sequences, and the practical application of advanced statistical methods in demographic analysis.

The research in this thesis is supported by the MUR-PRIN 2022 project "CARONTE" (Prot. 2022KBTEBN), funded by the European Union – Next Generation EU, Mission 4, CUP: J53D23009400001

Chapter 1

Bayesian Local Clustering of Age-Period Mortality Surfaces across Multiple Countries

Joint work with E. Aliverti and D. Durante

1.1 Introduction

The changes in life expectancy, population structure and welfare systems over the past decades have stimulated a growing demand for novel statistical models capable of characterizing heterogeneous mortality patterns across ages, periods and countries, while providing reliable probabilistic forecasts with rigorous uncertainty quantification (e.g., Lutz and Kc, 2010; Raftery et al., 2013; Hunt and Blake, 2021). Advancements along these lines are fundamental in guiding healthcare, social, environmental and retirement policies, thereby motivating active research on mortality modeling within several fields, such as demography (e.g., Lee and Miller, 2001; Li and Lee, 2005; De Jong and Tickle, 2006; Raftery et al., 2013; Li et al., 2013; Hyndman et al., 2013; Mazzuco et al., 2018; Camarda, 2019; Léger and Mazzuco, 2021), statistics (e.g., Lee and Carter, 1992; Hyndman and Ullah, 2007; Alexopoulos et al., 2019; Tang et al., 2022; Aliverti et al., 2022; Lam and Wang, 2023; Pavone et al., 2024; Debon et al., 2023; Dimai, 2025) and actuarial sciences (e.g., Haberman and Renshaw, 2011; Hatzopoulos and Haberman, 2013; Kleinow, 2015; Antonio et al., 2015; Currie, 2016; Enchev et al., 2017; Wong et al., 2018; Dong et al., 2020; Scognamiglio, 2022), among others. As clarified in these contributions, such active research on mortality has witnessed in recent years a progressive shift away from analyzing single countries in isolation and towards joint modeling of age-period mortality surfaces within a multi-country setting.

The above perspective is inherently motivated by the fact that countries are not isolated entities, but rather display similar global trends in age-specific mortality patterns (see Figure 1.1) regulated by common exogenous and endogenous factors, such as advancements

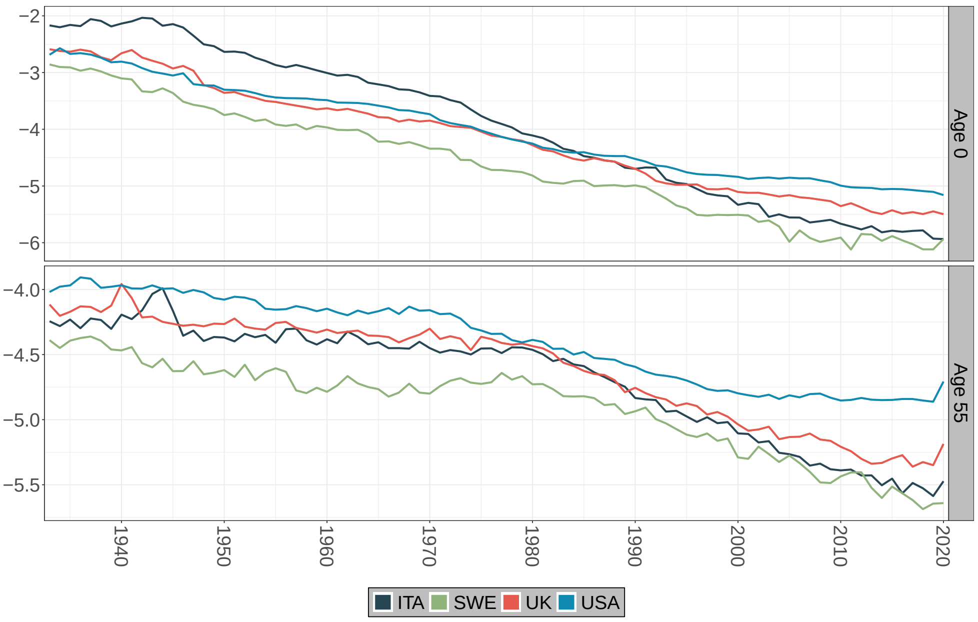


Figure 1.1: Log-mortality rates at age 0 (first panel) and 55 (second panel) in Italy (ITA), Sweden (SWE), the United Kingdom (UK) and the United States of America (USA) between 1933 and 2020.

in healthcare (e.g., Vallin and Meslé, 2004). As such, joint modeling of multiple countries can facilitate effective borrowing of information to improve age-period mortality forecasts, while opening the avenues for inference on local similarities and differences among countries' mortality patterns in specific age classes and periods that could unveil unexplored demographic trends, possibly arising from targeted policies adopted by certain countries. Addressing both objectives would require a unique formulation that accounts not only for the core structures of country-specific age-period mortality surfaces, but also for local clustering patterns between countries arising from overlaps among these country-specific surfaces in particular combinations of ages and periods. Although state-of-the-art multi-country mortality models are still not designed to include these fundamental dynamics, local heterogeneities in country-specific mortality rates across ages and periods are recognized in demographic research (e.g., Vallin and Meslé, 2004; Vaupel et al., 2011) and find strong evidence in mortality data.

For example, the first panel of Figure 1.1 illustrates the infant log-mortality rates in Italy, Sweden, the United Kingdom (UK) and the United States of America (USA) from 1933 to 2020, as retrieved from the Human Mortality Database (HMD: https://www.mortality.org/). During the first forty years, the UK and the USA display overlapping mortality trajectories, which then diverge after 1980. Conversely, Italy follows a distinct path until 1985, after which it aligns with the UK for approximately ten years. Interestingly, as shown within the second panel of Figure 1.1, these grouping structures vary not only across periods, but also with different ages. For instance, the clustering among the UK and the USA at infant ages is no more visible for individuals aged 55, who show, instead, a remarkable overlap between Italy and the UK for a large time window.

Accounting for the above patterns via a principled model-based representation would not only provide a more realistic characterization of age-period mortality surfaces across multiple countries, but could also open the avenues for rigorously answering important scientific questions, e.g., on the differences among low-mortality countries (Oeppen and Vaupel, 2002; Vaupel et al., 2011) and the corresponding socio-economic determinants (Marmot, 2005). As discussed previously, although the literature on multi-country mortality modeling has witnessed extensive advancements over the recent years, state-of-the-art contributions are not designed to address this endeavor. In fact, a common solution in multi-country mortality modeling relies on extending structured bilinear decompositions of age-period mortality surfaces for a single country, such as the one by Lee and Carter (1992), to joint formulations that include country-specific parameters and a common component shared across periods (e.g., Li and Lee, 2005) or ages (e.g., Kleinow, 2015). This general perspective improves forecasts and can be further extended to more sophisticated representations leveraging multi-country generalizations (Hyndman et al., 2013; Lam and Wang, 2023) of functional principal components constructions (Hyndman and Ullah, 2007), or joint decompositions of the country-age-period mortality tensor (Dong et al., 2020); see also Enchev et al. (2017). However, all these solutions are not designed to infer group structures among countries based on similarities in the associated mortality rates. Such an issue is also found in related Bayesian hierarchical formulations relying on conditionally independent models for each country linked by a common prior distribution on shared underlying parameters (e.g., Raftery et al., 2013; Antonio et al., 2015; Aliverti et al., 2022).

An effective direction for overcoming the aforementioned issues is to move towards more recent extensions of the above formulations which explicitly include notions of clustering among countries in terms of the corresponding mortality patterns (Hatzopoulos and Haberman, 2013; Léger and Mazzuco, 2021; Schnürch et al., 2021; Tang et al., 2022; Scognamiglio, 2022; Debon et al., 2023; Dimai, 2025). Albeit providing meaningful representations for model-based clustering of mortality surfaces, the overarching focus of these extensions is on global grouping structures, instead of local ones changing with different combinations of ages and periods. As such, countries are not allowed to display varying clustering behaviors at different ages (Léger and Mazzuco, 2021), time periods (Hatzopoulos and Haberman, 2013), or both dimensions (Tang et al., 2022; Schnürch et al., 2021; Scognamiglio, 2022; Dimai, 2025). Recalling the previous discussion of Figure 1.1, these constraints are not supported by the observed age-period mortality data and hinder the possibility of learning nuanced grouping structures, possibly informing on unexplored localized demographic trends that are visible only for specific combinations of ages and periods.

To our knowledge, the only attempt to remove the constraints imposed by the above global clustering perspectives can be found in the application of latent class clustering models to multi-country age-period mortality data in Debon et al. (2023). Although this contribution has the merit of recognizing the importance of moving beyond global clustering perspectives, a direct application of latent class clustering methods without the additional inclusion of the specific structures of age-period mortality surfaces could undermine the flexibility of the resulting procedure in uncovering local clustering patterns. This potential issue finds evidence in Figure 1 of Debon et al. (2023) where the inferred grouping structures resemble more closely those obtained under a global perspective, than a local one. For example, according to Figure 1 of Debon et al. (2023), at infant ages, all countries share the same cluster in the entire time window analyzed, a behavior which is not in line with the local clustering structures displayed by the observed mortality data in our Figure 1.1. In addition, similarly to all cluster-based extensions of multi-country mortality models, also Debon et al. (2023) requires to specify the unknown number of groups, a challenging task in practice, without a uniquely-accepted solution.

Motivated by the above discussion and by the impact of addressing the aforementioned challenges in multi-country mortality modeling, we propose and develop in Section 1.2 an innovative and principled Bayesian formulation that accounts for the core structures of age-period mortality surfaces, and crucially incorporates clustering patterns among countries which are allowed to vary flexibly across different combinations of ages and periods. This is accomplished by modeling the smooth age patterns of mortality via a flexible linear combination of B-spline bases (e.g., Camarda, 2019; Pavone et al., 2024), whose country-specific dynamic coefficients evolve across calendar years through carefully-designed stochastic processes of time relying on a temporal random partition prior inspired by the general construction in Page et al. (2022). Crucially, these stochastic processes for the joint time trajectories of the country-specific coefficients associated with the different B-spline bases provide a principled characterization of the time changes in the age patterns of mortality across periods, while allowing the grouping structures exhibited by countries to change both in time and across the bases' coefficients associated with the different ages. As such, the resulting representation facilitates the identification of locally converging or diverging trends in multi-country age-period mortality surfaces, while crucially preserving a structured representation that accounts for the core characteristics of these surfaces.

As clarified in Section 1.3, the proposed structured representation, albeit flexible, is amenable to tractable posterior inference via a carefully-designed Gibbs-sampling algorithm that automatically learns the unknown total number of clusters and facilitates both point estimation and uncertainty quantification on mortality patterns and grouping structures. The simulation studies within Section 1.4 and the application to mortality data of 14 countries from 1933 until 2020 in Section 1.5, not only confirm the ability of the proposed model to accurately learn these local grouping structures, but also unveil unique localized similarities among specific countries, which highlight both known demographic phenomena and also yet-unexplored trends acting on selected countries over specific combinations of ages and periods. Concluding remarks can be found in Section 1.6 where we also clarify that, although motivated by multi-country age-period mortality data, the proposed model has broader scope and impact, in that it allows to detect localized overlaps among surfaces associated with different populations. To our knowledge, general methodological contributions exploring this direction are limited.

1.2 Model Formulation

Let us denote with d_{ixt} and E_{ixt}, respectively, the observed death counts and the average number of individuals at risk within country $i = 1, \dots, n$, at age $x \in \mathcal{X}$ and for period $t = 1, \dots, T$ (corresponding, in our case, to calendar years). Consistent with the overarching focus in mortality models for both single and multiple countries (see, for example, Currie, 2016; Enchev et al., 2017; Hunt and Blake, 2021), our interest lies in the analysis of the country-specific age-period log-mortality rates $\log m_{ixt} = \log(d_{ixt}/\mathrm{E}_{ixt})$ for which we assume

$$\log m_{ixt} = f_{it}(x) + \varepsilon_{ixt}, \quad \text{with} \quad \varepsilon_{ixt} \sim \mathrm{N}(0, \sigma_i^2), \tag{1.1}$$

independently for $i = 1, \dots, n$, $x \in \mathcal{X}$ and $t = 1, \dots, T$, where $f_{it}(x) = \mathbb{E}[\log m_{ixt} \mid f_{it}(x)]$ is the expected log-mortality rate surface for the i–th country expressed as a function of age $x \in \mathcal{X}$ that is allowed to vary across periods $t = 1, \dots, T$. The general surface plus Gaussian noise formulation in (1.1) is common to several mortality models for both single and multiple countries (e.g., Currie, 2016; Enchev et al., 2017; Hunt and Blake, 2021). In addition, as recently proved by Pavone et al. (2024, Proposition 2.1) under a single-country perspective, when E_{ixt} is sufficiently large (as in standard demographic settings), the above formulation arises directly from an underlying Poisson log-normal model for the observed death counts d_{ixt} that properly accounts for possible overdispersion; see also Wong et al. (2018). Nonetheless, as discussed in Section 1.1, none of the available multi-country models provides a structured representation for $f_{it}(x)$ that incorporates the core age-period patterns of mortality, while allowing countries to cluster differently as x and t vary.

Addressing the above gap would require the design of a structured representation for the expected log-mortality rate surface which ensures that $f_{it}(x)$ varies *(i)* smoothly as a function of age x, for every $i = 1, \dots, n$ and $t = 1, \dots, T$, and *(ii)* dynamically with periods t, for each $i = 1, \dots, n$ and $x \in \mathcal{X}$, while *(iii)* allowing $f_{it}(x)$ and $f_{i't}(x)$ for any two generic countries i and i' to cluster (i.e., display similar values) only for those combinations of x and t in which there is empirical evidence of overlapping patterns in the log-mortality rates of i and i'. To this end, a natural option for addressing objective *(i)* in a way that facilitates also inclusion of *(ii)* and *(iii)* is to represent $f_{it}(x)$ through the B-spline expansion

$$f_{it}(x) = \textstyle\sum_{j=1}^{p} \beta_{ijt} g_j(x), \qquad i = 1, \dots, n, \; x \in \mathcal{X}, \, t = 1, \dots, T, \tag{1.2}$$

where $[g_1(x), \dots, g_p(x)]$ denotes a set of common B-splines bases (e.g., Eilers and Marx, 1996) with associated coefficients $[\beta_{i1t}, \dots, \beta_{ipt}]$ that are allowed to vary with countries $i = 1, \dots, n$

and periods $t = 1, \dots, T$. Although related representations have been mostly explored in separate analyses of single countries in isolation (e.g., Currie et al., 2004; Camarda, 2019; Pavone et al., 2024), the expansion in (1.2) provides an effective construction in our multi-country setting which ensures that $f_{it}(x)$ is a smooth function of age, for every $i = 1, \dots, n$ and $t = 1, \dots, T$, whose dynamic changes across periods and local clustering patterns among countries can be regulated by a finite set of coefficients $[\beta_{i1t}, \dots, \beta_{ipt}]$, for $i = 1, \dots, n$ and $t = 1, \dots, T$. As such, objective *(i)* is addressed by construction, while *(ii)*–*(iii)* can be accomplished by allowing time changes and ties in these coefficients, respectively. In particular, notice that, when $\beta_{ijt} = \beta_{i'jt}$, then countries i and i' display under (1.2) similar mortality rates in period t for the age interval associated with basis $g_j(x)$.

To formalize this idea, let $c_{ijt} \in \{1, \dots, K_{jt}\}$ be the cluster membership indicator for country i, with respect to the j–th basis in period t, and denote with $\mathbf{c}_{jt} = [c_{1jt}, \dots, c_{njt}] \in \{1, \dots, K_{jt}\}^n$ the vector comprising the memberships for the n countries, for every $j = 1, \dots, p$ and $t = 1, \dots, T$, where $K_{jt} \leq n$ is the total number of clusters at the pair (j, t). Then, ties among the coefficients $\beta_{1jt}, \dots, \beta_{njt}$ can be readily incorporated, for every $j = 1, \dots, p$ and $t = 1, \dots, T$, by letting

$$\beta_{ijt} = \beta^{\star}_{c_{ijt}jt}, \qquad \text{for all } i = 1, \dots, n, \ j = 1, \dots, p, \ t = 1, \dots, T, \tag{1.3}$$

where $\beta^{\star}_{c_{ijt}jt} \in \mathbb{R}$ is the value of the j–th B-spline coefficient in period t associated with the cluster to which country i has been allocated. As such, all countries belonging to the same generic cluster k (for the j–th spline basis within period t) will display the same value $\beta^{\star}_{kjt}$ for the associated coefficient, thereby addressing *(iii)*. Crucially, $\mathbf{c}_{jt}$ varies with both j and t, and therefore, any generic pair of countries i and i' is allowed to cluster locally only for specific ages and periods, consistent with the original motivations underlying the proposed construction.

In order to complete the proposed Bayesian formulation, we require priors for the model parameters in (1.1)–(1.3), namely σ_i^2, for $i = 1, \dots, n$, along with $\mathbf{c}_{jt}$ and $\boldsymbol{\beta}^{\star}_{jt} = [\beta^{\star}_{1jt}, \dots, \beta^{\star}_{K_{jt}jt}]^{\intercal}$, for each $j = 1, \dots, p$, $t = 1, \dots, T$. Regarding σ_i^2, for $i = 1, \dots, n$, we follow common practice and consider conditionally-conjugate Inv-Gamma(a_σ, b_σ) priors for each σ_i^2, independently across $i = 1, \dots, n$. Conversely, for $\mathbf{c}_{jt}$ and $\boldsymbol{\beta}^{\star}_{jt}$ we elicit priors that explicitly account for the temporal structure behind the evolution of both $\mathbf{c}_{jt}$ and $\boldsymbol{\beta}^{\star}_{jt}$ with periods $t = 1, \dots, T$, for each $j = 1, \dots, p$, thereby addressing *(ii)*. Focusing first on $\boldsymbol{\beta}^{\star}_{jt}$, this goal is accomplished by assuming independent Gaussian priors for the entries in such a vector, further centered around a higher-level mean function of time which is assigned a Gaussian process (GP) prior (see, e.g., Williams and Rasmussen, 2006). More specifically, we let

$$\begin{aligned} &(\boldsymbol{\beta}^{\star}_{jt} \mid \psi_{jt}, \delta_j^2) \sim \mathrm{N}_{K_{jt}}(\psi_{jt}\mathbf{1}_{K_{jt}}, \delta_j^2\mathbf{I}_{K_{jt}}), \quad \text{independently for } j = 1, \dots, p, \ t = 1, \dots, T, \\ &(\boldsymbol{\psi}_j = [\psi_{j1}, \dots, \psi_{jT}]^{\intercal} \mid \omega_j^2) \sim \mathrm{N}_T(\boldsymbol{\mu}_j, \omega_j^2\boldsymbol{\Sigma}), \quad \text{independently for } j = 1, \dots, p, \end{aligned} \tag{1.4}$$

where $\mathbf{1}_{K_{jt}}$ and $\mathbf{I}_{K_{jt}}$ denote the $K_{jt} \times 1$ vector of ones and the $K_{jt} \times K_{jt}$ identity matrix, respectively, whereas $\boldsymbol{\mu}_j$ and $\omega_j^2\boldsymbol{\Sigma}$ are the mean vector and the covariance matrix induced by the GP prior on the finite time grid $t = 1, \dots, T$. For $\boldsymbol{\Sigma}$ we consider, in particular, a squared exponential correlation function (see Williams and Rasmussen, 2006, Section 4) which allows the dependence between the generic ψ_{jt} and $\psi_{jt'}$ to progressively decrease as the distance between the time indexes t and t' increases. The mean vector $\boldsymbol{\mu}_j$ is instead elicited under a data-driven perspective (see Section 1.4), whereas δ_j^2 and ω_j^2 are assigned conditionally-conjugate Inv-Gamma(a_δ, b_δ) and Inv-Gamma(a_ω, b_ω) hyperpriors, respectively, independently for $j = 1, \dots, p$.

As discussed above, temporal dependence is expected also in the sequence $\mathbf{c}_{j1}, \dots, \mathbf{c}_{jT}$ of cluster assignment vectors, for each $j = 1, \dots, p$. In particular, it is reasonable to assume that the generic $\mathbf{c}_{j,t-1}$ influences the formation of $\mathbf{c}_{jt}$ through a mechanism in which only a subset of the countries change the corresponding cluster membership when moving from $t-1$ to t, whereas the others preserve it. This Markovian dependence structure implies that $p(\mathbf{c}_{j1}, \dots, \mathbf{c}_{jT} \mid \eta_j) = p(\mathbf{c}_{j1} \mid \eta_j)p(\mathbf{c}_{j2} \mid \mathbf{c}_{j1}, \eta_j) \cdots p(\mathbf{c}_{jT} \mid \mathbf{c}_{j,T-1}, \eta_j)$, where $\eta_j \in \mathbb{R}_+$ regulates the formation of $\mathbf{c}_{j1}$ through $p(\mathbf{c}_{j1} \mid \eta_j)$, which in turn influences those of the subsequent vectors $\mathbf{c}_{j2}, \dots, \mathbf{c}_{jT}$. As such, the mechanism defining $p(\mathbf{c}_{jt} \mid \mathbf{c}_{j,t-1}, \eta_j)$ should be combined with a careful prior on the initial membership vector $\mathbf{c}_{j1}$ at time $t = 1$, which induces a flexible characterization of the joint prior over the entire sequence $\mathbf{c}_{j1}, \dots, \mathbf{c}_{jT}$, for each $j = 1, \dots, p$. A construction of this type can be found in the temporal random partition (tRPM) prior recently proposed by Page et al. (2022) and detailed in Introduction. Adapting this general construction to our specific context, we let

$$([\mathbf{c}_{j1}, \dots, \mathbf{c}_{jT}] \mid \alpha_j, \eta_j) \sim \text{tRPM}(\alpha_j, \eta_j), \qquad \text{independently for } j = 1, \dots, p, \tag{1.5}$$

where $\alpha_j \in [0, 1]$ is a temporal persistence parameter controlling the transition mechanism behind $p(\mathbf{c}_{jt} \mid \mathbf{c}_{j,t-1}, \eta_j)$. Specializing Page et al. (2022) to our setting, such a formation mechanism for $\mathbf{c}_{j1}$ is assumed to be driven by a Chinese restaurant process (CRP) prior with parameter η_j. This prior belongs to the general Gibbs-type class (e.g., De Blasi et al., 2015) and provides a flexible characterization for the formation of the grouping structures in $\mathbf{c}_{j1}$ driven by a tractable urn scheme. In particular, let $\mathbf{c}_{j1}^{(-i)} = [c_{1j1}, \dots, c_{i-1,j1}, c_{i+1,j1}, \dots, c_{nj1}]$ be the cluster membership vector for the j–th spline basis at time $t = 1$, excluding the generic i-th country, and denote with $n_{kj1}^{(-i)}$ and $K_{j1}^{(-i)}$ the cardinality of cluster k and the total number of non-empty clusters in $\mathbf{c}_{j1}^{(-i)}$, respectively. Then, under this scheme, the prior on the cluster memberships for country i, given those of the others, coincides with $\text{pr}(c_{ij1} = k \mid \mathbf{c}_{j1}^{(-i)}, \eta_j) \propto n_{kj1}^{(-i)}$ if k is a cluster already occupied by the other $n-1$ countries, and $\text{pr}(c_{ij1} = k \mid \mathbf{c}_{j1}^{(-i)}, \eta_j) \propto \eta_j$ if k is a new cluster (i.e., $k = K_{j1}^{(-i)} + 1$). Besides illustrating the tractability of this construction, along with the role of the parameter η_j, such an urn scheme also clarifies that the total number of clusters in $\mathbf{c}_{j1}$ can be learned automatically, without the need to pre-specify it. This is a remarkable advantage compared to state-of-the-art cluster-based models for mortality (Hatzopoulos and Haberman, 2013; Léger and Mazzuco, 2021; Schnürch et al., 2021; Tang et al., 2022; Scognamiglio, 2022; Debon et al., 2023; Dimai, 2025) that cannot learn the number of clusters automatically, as part of the estimation process.

To conclude our Bayesian formulation, we select independent hyperpriors $\text{Beta}(a_\alpha, b_\alpha)$ and $\text{Gamma}(a_\eta, b_\eta)$ for the quantities α_j and η_j in (1.5), respectively, for $j = 1, \dots, p$. The $\text{Beta}(a_\alpha, b_\alpha)$ hyperprior is considered also in Page et al. (2022), whereas the $\text{Gamma}(a_\eta, b_\eta)$ for the CRP concentration parameters, although not present in Page et al. (2022), is motivated by recent theoretical results in Ascolani et al. (2023) on the consistency properties of CRP constructions.

Section 1.3 clarifies that the above Bayesian formulation is also amenable to tractable posterior computation and inference leveraging a carefully-designed Gibbs-sampling algorithm.

1.3 Bayesian Computation and Inference

Posterior inference for the Bayesian model in Section 1.2 is conducted via Monte Carlo leveraging the draws of a carefully-designed Gibbs sampler targeting the posterior distribution of the model parameters given the observed mortality rates. Section 1.3.1 derives in detail such a Gibbs sampler, while Section 1.3.2 illustrates how the resulting posterior samples are leveraged to perform Monte Carlo inference on the local clustering structures among countries and the corresponding mortality patterns.

1.3.1 Gibbs Sampler

The proposed Gibbs sampler iterates sequentially between two main steps. First, the temporal cluster allocations are sampled, along with the tRPM hyperparameters in (1.5), from the corresponding full-conditionals by adapting the algorithm of Page et al. (2022, see Section B, Supplementary Materials) to our B-splines construction, and the results in Escobar and West (1995) for the CRP hyperparameters η_j, $j = 1, \dots, p$. Second, conditionally on the group allocations, the cluster-specific B-splines coefficients in (1.4) and the corresponding hyperparameters are updated leveraging Gaussian and inverse-Gamma conjugacy.

In order to update the temporal grouping structures in $\mathbf{c}_{j1}, \dots, \mathbf{c}_{jT}$, for $j = 1, \dots, p$, along with the parameters of the associated tRPM prior, let us first recall that the generic allocation vectors $\mathbf{c}_{j,t-1}$ and $\mathbf{c}_{jt}$ are *compatible*, with respect to $\boldsymbol{\gamma}_{jt}$, if the partition of the n countries encoded within $\mathbf{c}_{jt}$ may be derived from the one associated with $\mathbf{c}_{j,t-1}$ by reallocating only those countries for which $\gamma_{ijt} = 0$. Given $\mathbf{c}_{j,t-1}$ and $\boldsymbol{\gamma}_{jt}$, define with $\mathbb{C}(\mathbf{c}_{j,t-1}, \boldsymbol{\gamma}_{jt})$ the set comprising all partitions $\mathbb{c}$ induced by $\mathbf{c}_{jt}$ that are compatible with $\mathbf{c}_{j,t-1}$, under $\boldsymbol{\gamma}_{jt}$, and let $\Gamma_{jt} = \{i = 1, \dots, n : \gamma_{ijt} = 1\}$ be the set of countries whose group allocation does not change from $t-1$ to t. Furthermore, let us denote with $\mathbb{c}_{jt}^{\Gamma_{jt}}$ the partition induced by $\mathbf{c}_{jt}$, but considering only those countries with indexes in Γ_{jt}. Then, leveraging the results in Page et al. (2022), the full-conditional distribution for the latent indicators γ_{ijt} is a Bernoulli variable with probabilities

$$\operatorname{pr}(\gamma_{ijt}=1 \mid -) = \frac{\alpha_j}{\alpha_j + (1-\alpha_j)p(\mathbb{c}_{jt}^{\Gamma_{jt}^{(+i)}} \mid \eta_j)/p(\mathbb{c}_{jt}^{\Gamma_{jt}^{(-i)}} \mid \eta_j)} \mathbb{1}[\mathbb{c}_{j,t-1}^{\Gamma_{jt}^{(+i)}} = \mathbb{c}_{jt}^{\Gamma_{jt}^{(+i)}}], \tag{1.6}$$

independently for each $i=1,\dots,n$, $j=1,\dots,p$ and $t=1,\dots,T$, where $\mathbb{1}[\cdot]$ denotes the indicator function, while $\Gamma_{jt}^{(-i)} = \Gamma_{jt} \setminus \{i\}$ and $\Gamma_{jt}^{(+i)} = \Gamma_{jt}^{(-i)} \cup \{i\}$.

Note that in (1.6), the ratio $p(\mathbb{c}_{jt}^{\Gamma_{jt}^{(+i)}} \mid \eta_j)/p(\mathbb{c}_{jt}^{\Gamma_{jt}^{(-i)}} \mid \eta_j)$ can be computed in closed-form leveraging the results in Page et al. (2022) under the urn scheme of the CRP prior discussed in Section 1.2. This is also a key to update the allocations c_{ijt} of those countries for which $\gamma_{ijt}=0$; if the sampled γ_{ijt} is 1, then c_{ijt} is kept fixed at the allocation drawn for i at the pair $(j, t-1)$. To this end, let $\mathbf{c}_{jt}^{(-i)}$ be the vector of cluster allocations after removing the entry c_{ijt}, and denote with $\mathbf{c}_{jt}^{(c_{ijt}=k)}$ the membership vector $[c_{1jt},\dots,c_{ijt}=k,\dots,c_{njt}]$. Furthermore, let $K_{jt}^{(-i)}$ be the total number of unique clusters in $\mathbf{c}_{jt}^{(-i)}$ and define with $r_{ixt}^{(j)} = \log m_{ixt} - \sum_{j'\neq j} \beta^{\star}_{c_{ij't}j't} g_{j'}(x)$ the partial residuals under model (1.1)–(1.3) without considering the j–th spline. Then, leveraging the Bayes rule and the CRP urn scheme, we have that the full-conditional distribution for those c_{ijt} having $\gamma_{ijt}=0$ is a categorical variable with probabilities, for each $k=1,\dots,K_{jt}^{(-i)}+1$, given by

$$\begin{aligned}&\operatorname{pr}(c_{ijt}=k \mid -) \propto \\ &\quad \operatorname{pr}(c_{ijt}=k \mid \mathbf{c}_{jt}^{(-i)}, \eta_j)\mathbb{1}[\mathbb{c}_{j,t+1} \in \mathbb{C}(\mathbf{c}_{jt}^{(c_{ijt}=k)}, \boldsymbol{\gamma}_{j,t+1})] \textstyle\prod_{x\in\mathcal{X}} \phi(r_{ixt}^{(j)} - \beta^{\star}_{kjt} g_j(x); \sigma_i^2),\end{aligned} \tag{1.7}$$

for every $i=1,\dots,n$, $j=1,\dots,p$ and $t=1,\dots,T$, where $\phi(r_{ixt}^{(j)} - \beta^{\star}_{kjt} g_j(x); \sigma_i^2)$ is the density, evaluated at $r_{ixt}^{(j)} - \beta^{\star}_{kjt} g_j(x)$, of the zero-mean Gaussian distribution with variance σ_i^2, whereas, as discussed in Section 1.2, $\operatorname{pr}(c_{ijt}=k \mid \mathbf{c}_{jt}^{(-i)}, \eta_j)$ can expressed under the CRP urn scheme as

$$\operatorname{pr}(c_{ijt}=k \mid \mathbf{c}_{jt}^{(-i)}, \eta_j) \propto \begin{cases} n_{kjt}^{(-i)} & k=1,\dots,K_{jt}^{(-i)}, \\ \eta_j & k=K_{jt}^{(-i)}+1, \end{cases}$$

where $n_{kjt}^{(-i)}$ denotes the size of the k–th cluster after removing unit i.

Given the samples of $\mathbf{c}_{jt}$ and $\boldsymbol{\gamma}_{jt}$, for all $t=1,\dots T$, the tRPM hyparameters α_j are updated from the full-conditional Beta distributions

$$(\alpha_j \mid -) \sim \operatorname{Beta}\left(a_\alpha + \textstyle\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{ijt}, b_\alpha + nT - \sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{ijt}\right),$$

independently for $j=1,\dots,p$, whereas the CRP concentration parameters η_j, $j=1,\dots,p$, are updated following the data-augmentation scheme described in Escobar and West (1995).

Once the group membership vectors $\mathbf{c}_{j1},\dots,\mathbf{c}_{jT}$ have been updated for each $j=1,\dots,p$, it is possible to sample the cluster-specific coefficients in $\boldsymbol{\beta}^\star_{jt}$, for $j=1,\dots,p$ and $t=1,\dots,T$, along with the corresponding time-varying means in $\boldsymbol{\psi}_j$, for $j=1,\dots,p$. Combining prior (1.4) with the model (1.1)–(1.3), this can be accomplished by leveraging directly Gaussian-Gaussian conjugacy. In particular, the full-conditional distribution for the generic $\beta^\star_{kjt}$ is

$$(\beta^\star_{kjt}\mid -)\sim \mathrm{N}(\omega^{-1}_{\beta^\star_{kjt}}\eta_{\beta^\star_{kjt}},\omega^{-1}_{\beta^\star_{kjt}}),\tag{1.8}$$

independently across every $k=1,\dots,K_{jt}$, $j=1,\dots,p$ and $t=1,\dots,T$, where $\omega_{\beta^\star_{kjt}}=1/\delta^2_j+\sum_{x\in\mathcal{X}}g^2_j(x)\sum_{i:c_{ijt}=k}(1/\sigma^2_i)$, while $\eta_{\beta^\star_{kjt}}=(\psi_{jt}/\delta^2_j)+\sum_{i:c_{ijt}=k}(1/\sigma^2_i)\sum_{x\in\mathcal{X}}r^{(j)}_{ixt}g_j(x)$. Similarly, the full-conditional for each vector $\boldsymbol{\psi}_j$ is

$$(\boldsymbol{\psi}_j\mid -)\sim \mathrm{N}_T(\boldsymbol{\Omega}^{-1}_{\psi_j}\boldsymbol{\eta}_{\psi_j},\boldsymbol{\Omega}^{-1}_{\psi_j}),\tag{1.9}$$

for $j=1,\dots,p$, with $\boldsymbol{\Omega}_{\psi_j}=\omega_j^{-2}\boldsymbol{\Sigma}^{-1}+\delta_j^{-2}\mathrm{diag}(K_{j1},\dots,K_{jT})$ and $\boldsymbol{\eta}_{\psi_j}=\omega_j^{-2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j+\delta_j^{-2}\bar{\boldsymbol{\beta}}_j$, where $\bar{\boldsymbol{\beta}}_j$ is a vector of dimension $T\times 1$ having generic t–th entry $\sum_{k=1}^{K_{jt}}\beta^\star_{kjt}$.

To conclude the Gibbs-sampling routine it remains to update the variance parameters σ^2_i, $i=1,\dots,n$ in (1.1), along with δ^2_j and ω^2_j, $j=1,\dots,p$, entering the priors in (1.4). By conditioning on the quantities sampled in (1.7) and (1.8)–(1.9), the updates for these variance parameters follow directly from inverse-Gamma conjugacy properties, thereby obtaining

$$(\sigma^2_i\mid -)\sim \text{Inv-Gamma}\left(a_\sigma+XT/2, b_\sigma+(1/2)\sum\nolimits_{x\in\mathcal{X}}\sum\nolimits_{t=1}^{T}[\log m_{ixt}-f_{it}(x)]^2\right).$$

for every $i=1,\dots,n$, with $f_{it}(x)$ defined as in (1.2)–(1.3), and

$$(\delta^2_j\mid -)\sim \text{Inv-Gamma}\left(a_\delta+\sum\nolimits_{t=1}^{T}K_{jt}/2, b_\delta+(1/2)\sum\nolimits_{t=1}^{T}\sum\nolimits_{k=1}^{K_{jt}}(\beta^\star_{kjt}-\psi_{jt})^2\right),$$
$$(\omega^2_j\mid -)\sim \text{Inv-Gamma}\left(a_\omega+T/2, b_\omega+(\boldsymbol{\psi}_j-\boldsymbol{\mu}_j)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\psi}_j-\boldsymbol{\mu}_j)/2\right),$$

for every $j=1,\dots,p$.

1.3.2 Monte Carlo Inference

Leveraging the samples produced by the Gibbs routine outlined in Section 1.3.1 posterior inference on the quantities of interest proceeds via Monte Carlo. As discussed in Sections 1.1–1.2, within our context a primary focus is on inferring grouping structures among countries induced by similarities among the corresponding mortality rates, and how these group structures vary locally across ages and periods. This information is contained in the posterior samples of the allocation vectors $\mathbf{c}_{jt}$ for $j=1,\dots,p$ and $t=1,\dots,T$, which we summarize through the $n\times n$ *posterior co-clustering matrices* $\hat{\mathbf{P}}_{jt}$, $j=1,\dots,p$, $t=1,\dots,T$, whose generic element $\hat{\mathbf{P}}_{jt[i,i']}$ is defined as the relative proportion of Gibbs samples in which countries i and i' have the same group allocation, at the pair (j,t). This provides an estimate of the posterior probabilities of co-clustering that is useful for quantifying uncertainty in $\mathbf{c}_{jt}$ for $j=1,\dots,p$, $t=1,\dots,T$, beyond single point estimates. As such, these matrices will be object of study in the mortality data application in Section 1.5.

When a single point estimate $\hat{\mathbf{c}}_{jt}$ of $\mathbf{c}_{jt}$ is of interest for each $j=1,\dots,p$ and $t=1,\dots,T$, the above *posterior co-clustering matrices* can be summarized under the decision-theoretic framework of Wade and Ghahramani (2018) to obtain the estimate

$$\hat{\mathbf{c}}_{jt} = \underset{\mathbf{c}'_{jt}}{\operatorname{argmin}}\, \mathbb{E}_{\mathbf{c}_{jt}}[\textsc{vi}(\mathbf{c}_{jt},\mathbf{c}'_{jt}) \mid \log \boldsymbol{m}], \quad j=1,\dots,p,\ t=1,\dots,T,$$

where $\log \boldsymbol{m}$ is the array of log-mortality rates observed for the n countries across all ages and periods, whereas VI is the variation of information distance (Meilă, 2007), namely a metric measuring the dissimilarity among two generic allocation vectors $\mathbf{c}_{jt}$ and $\mathbf{c}'_{jt}$ based on the associated individual and joint entropies. In practice, the above minimization problem is solved, for each $j=1,\dots,p$ and $t=1,\dots,T$, via `R` package `mcclust.ext` (Wade and Ghahramani, 2018) leveraging, as input, the corresponding *posterior co-clustering matrices* $\hat{\mathbf{P}}_{jt}$, $j=1,\dots,p$, $t=1,\dots,T$.

Besides the grouping structures encoded in $\mathbf{c}_{j1},\dots,\mathbf{c}_{jT}$, for $j=1,\dots,p$, it is also of interest to study the trajectories $\beta_{ij1},\dots,\beta_{ijT}$ of the B-splines coefficients. Recalling (1.2), these trajectories characterize the temporal evolution of the mortality levels for country i at the age interval associated with the j–th spline. Posterior samples for these coefficients can be derived directly from those for $\mathbf{c}_{jt}$ and $\boldsymbol{\beta}^{\star}_{jt}$, after noticing that, under (1.3), $\beta_{ijt} = \beta^{*}_{c_{ijt}jt}$, for every $i=1,\dots,n$, $j=1,\dots,p$ and $t=1,\dots,T$. As such, point estimates for the dynamic country-specific B-spline coefficients can be derived via Monte Carlo by computing the average of the resulting samples for each β_{ijt}, $i=1,\dots,n$, $j=1,\dots,p$ and $t=1,\dots,T$.

1.4 Simulation Study

Before employing the proposed model in the motivating mortality data application, we first assess its performance in recovering the *true* data-generative structures in a simulation study. Recalling Sections 1.1–1.3, the overarching focus is on quantifying to what extent the proposed model can learn accurately realistic grouping structures that vary locally, along with the associated cluster-specific parameters, in different combinations of ages and periods.

Consistent with the above goal, we simulate synthetic log-mortality rates, under the model outlined in (1.1)–(1.3) focusing on $n=5$ countries across $T=10$ periods and for ages $\mathcal{X}=\{0,1,\dots,100\}$. More specifically, $\log m_{ixt}$ are simulated as in (1.1) with $\sigma_i=\sigma=0.05$, for $i=1,\dots,5$, and $f_{it}(x)$ defined through (1.2)–(1.3) considering $p=6$ quadratic B-splines whose cluster-specific coefficients are generated via (1.4), setting $\delta_j=0.05$ for $j\neq 5$,

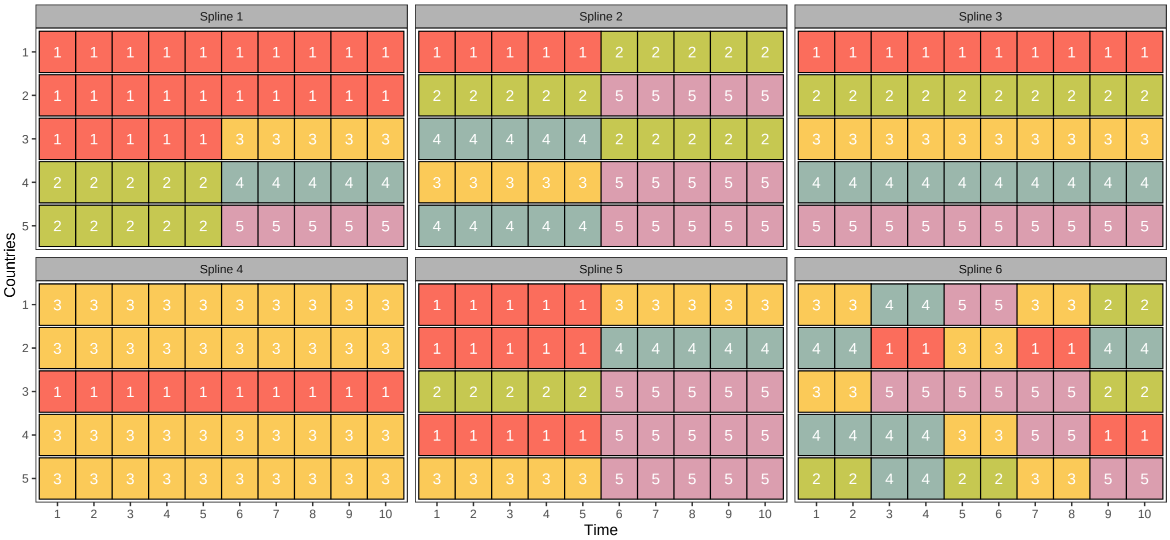


Figure 1.2: Cluster assignments in the simulation study. Colors and numbers represent true cluster memberships.

$\delta_5 = 0.1$, and letting $\boldsymbol{\psi}_j$, $j = 1, \dots, p$ correspond to parallel decreasing lines with slope -0.02. To assess the ability of the proposed model in inferring complex local group structures among countries, we do not produce $\mathbf{c}_{j1}, \dots, \mathbf{c}_{j10}$, for $j = 1, \dots, 6$ from the assumed prior. Rather, we set the group allocations manually as in Figure 1.2, in order to explore a wide spectrum of time-varying local clustering patterns. For example, for the age classes associated with the spline bases 3 (top-right panel) and 4 (bottom-left panel) the synthetic countries are grouped into stable clusters across time, albeit with differing co-clustering patterns. Conversely, the remaining spline bases exhibit more complex dynamic clustering patterns (see, e.g., spline basis 6 where the synthetic countries often change cluster membership).

Leveraging the above simulated data, we perform Bayesian inference under the model proposed in Section 1.2, setting diffuse hyperparameters $a_\sigma = b_\sigma = 10^{-3}$, $a_\delta = b_\delta = a_\omega = b_\omega = 10^{-3}$, $a_\eta = 2 \cdot 10^{-3}, b_\eta = 10^{-3}$ and $a_\alpha = 1$, $b_\alpha = 1$. Consistent with standard practice in Gaussian processes literature (see, e.g., Williams and Rasmussen, 2006), the entries of $\boldsymbol{\Sigma}$ in (1.4) are defined through a squared-exponential kernel with length scale 1.5, i.e. $\boldsymbol{\Sigma}_{[t,t']} = \exp[-0.5(t - t')^2/(1.5)^2]$, for every (t, t'). Finally, to achieve an improved calibration of the proposed model, the mean vectors $\boldsymbol{\mu}_j = [\mu_{j1}, \dots, \mu_{j10}]^\intercal$, $j = 1, \dots, 6$, in (1.4) are defined in a data-driven manner. This is accomplished by first obtaining, separately for every $t = 1, \dots, 10$, an OLS estimate of the splines coefficients under model (1.1) applied to data $\log m_{ixt}$, $i = 1, \dots, 5$, $\mathcal{X} = \{0, 1, \dots, 100\}$, with $f_{ix}(t) = \sum_{j=1}^{6} \beta_{jt} g_j(x)$. For each $j = 1, \dots, 6$, the resulting estimates $\hat{\beta}_{j1}, \dots, \hat{\beta}_{j10}$ are subsequently smoothed via LOESS to obtain the desired data-driven specification for the entries $\mu_{j1}, \dots, \mu_{j10}$ of $\boldsymbol{\mu}_j$. All these settings and hyperparameter choices proved robust also in the mortality data application in Section 1.5, and moderate changes in such quantities did not substantially affect the final conclusions and the performance in the simulation. Under the above settings, posterior

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$
Spline 1 ($j=1$)	0.991	0.998	0.999	0.995	0.976	1.000	1.000	1.000	1.000	1.000
Spline 2 ($j=2$)	0.995	0.999	0.999	1.000	0.998	0.991	0.999	1.000	1.000	0.995
Spline 3 ($j=3$)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Spline 4 ($j=4$)	0.991	1.000	0.999	1.000	0.997	1.000	1.000	0.999	0.987	0.990
Spline 5 ($j=5$)	0.996	1.000	1.000	1.000	0.994	0.997	1.000	0.999	0.999	0.993
Spline 6 ($j=6$)	0.914	0.897	0.968	0.969	0.982	0.979	0.983	0.983	0.989	0.990

Table 1.1: Simulation study. Posterior means of co-clustering accuracies for each combination (j,t).

inference for the proposed model proceeds via Monte Carlo as outlined in Section 1.3.2, leveraging the samples produced by the Gibbs routine derived within Section 1.3.1. Such a routine is run for 20,000 iterations, discarding the first 10,000 as a conservative burn-in. Traceplots and autocorrelation plots indicate satisfactory mixing of the chains. In fact, in this simulation study we observe convergence much before the burn-in employed. Nonetheless, we opted for a more conservative setting that can be considered as a default in general contexts, including in the motivating mortality data application in Section 1.5.

Consistent with our overarching focus, we first assess in Table 1.1 to what extent the proposed model is able to learn the *true* clustering structures displayed in Figure 1.2 among the 5 synthetic countries. To provide a comprehensive assessment of the clustering accuracy achieved by the posterior for $\mathbf{c}_{j1},\ldots,\mathbf{c}_{j10}$, $j=1,\ldots,6$, beyond the one obtained under a single point estimate, we compute, for each Gibbs sample of $\mathbf{c}_{jt}$, $t=1,\ldots,10$, $j=1,\ldots,6$, the percentage of pairs of countries that are correctly co-clustered. The posterior means of these percentages over the 10,000 retained samples is reported in Table 1.1, and confirm the excellent performance of the proposed model in recovering the co-clustering patterns induced

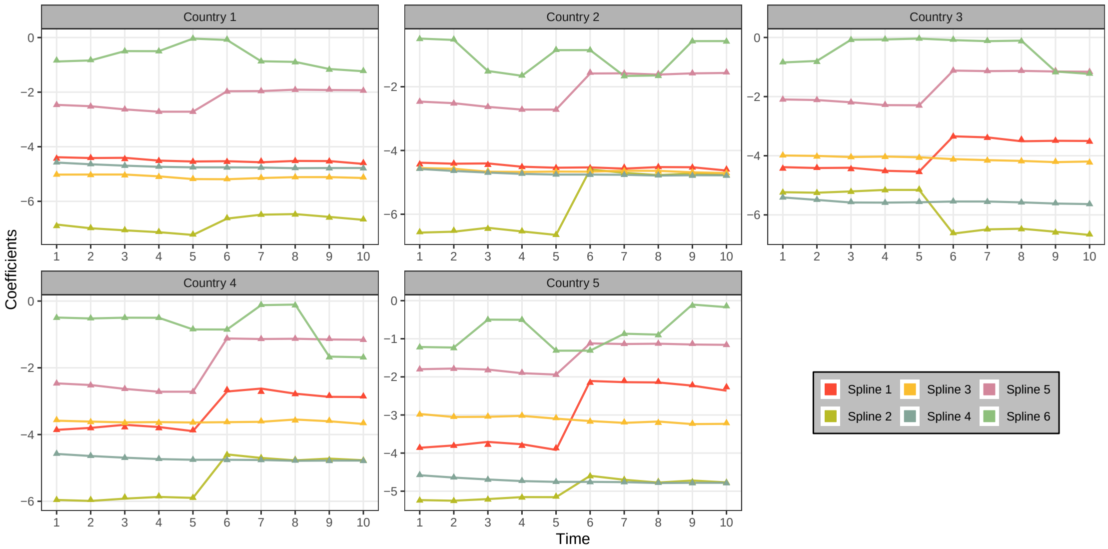


Figure 1.3: Simulation study. Posterior means of $\beta_{ij1},\ldots,\beta_{ij10}$ (lines) and true values (points) for each country $i=1,\ldots,5$ (panels) and basis function $g_j(x)$, $j=1,\ldots,6$ (colors).

by the *true* group structures within Figure 1.2. In particular, the accuracy measures are above 0.975 for all spline bases and time points, except for 6–th one where we observe a slight performance deterioration. This result is expected since such a basis presents the most complex clustering patterns, with frequent changes in group memberships for all countries. Nonetheless, even in this highly challenging regime, we still observe remarkable performance with all accuracy measures across time points above 0.89. As shown within Figure 1.3, the above remarkable performance in learning local clustering structures directly translates into highly accurate point estimates for the country-specific coefficient trajectories $\beta_{ij1} = \beta^\star_{c_{ij1}j1}, \ldots, \beta_{ij10} = \beta^\star_{c_{ij10}j10}$, $i = 1, \ldots, 5$, $j = 1, \ldots, 6$. These point estimates are obtained as detailed in Section 1.3.2, and compared with the associated *true* values within Figure 1.3. Results confirm the ability of the model to characterize the underlying trajectories of each spline coefficient accurately, even for complex underlying temporal dynamics. Under (1.1)–(1.3), this implies effective learning of the data-generative mechanism for the synthetic log-mortality rates.

1.5 Local Clustering of Age-Period Mortality Surfaces for 14 Countries

We conclude by showcasing the performance of the proposed model in learning local clustering structures across ages and periods induced by the log-mortality rates of 14 countries (Australia, Belgium, Canada, Switzerland, Denmark, Spain, Finland, France, Italy, the Netherlands, Norway, Sweden, the United Kingdom and the United States). The original data $\log m_{ixt}$ are retrieved from the Human Mortality Database (Human Mortality Database, 2024) for ages from 0 until 98 years old, over a time horizon of 88 years (1933–2020).

In applying the proposed model to the above data we follow standard practice in multi-country studies (e.g., Li and Lee, 2005; Aliverti et al., 2022) by considering separate analyses for male and female sub-populations. Posterior inference proceeds under the same hyper-parameters and Gibbs settings as those considered for the simulation studies in Section 1.4, except for the choice of the B-spline bases that are set as in Figure 1.4 to achieve a more

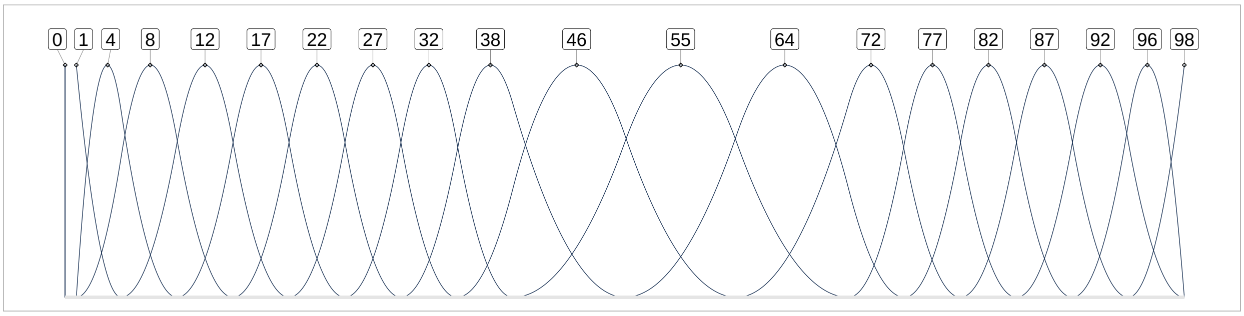


Figure 1.4: Graphical representation of the selected B-splines bases $g_1(x), \ldots, g_{20}(x)$. The number associated with each spline $g_j(x)$ denotes the age at which such a spline takes its maximum value.

realistic characterization of the observed age patterns of mortality. This choice allows increased flexibility in those age ranges where larger local variations are expected (i.e., infant and senescent groups) and has proved effective also in recent single-country analyses (e.g., Pavone et al., 2024), thereby motivating its use also in the newly-developed multi-country model. As in the simulations studies, the traceplots for the quantities of interest and the convergence diagnostics did not provide evidence against convergence. Section 1.5.1 summarizes the results of posterior inference under the proposed model, whereas Section 1.5.2 highlights previously-unexplored local clustering structures with a specific focus on the United States. Finally, Section 1.5.3 explores possible associations between the novel co-clustering patterns inferred by the proposed model and relevant socio-economic variables, including the gross domestic product (GDP) and health expenses.

1.5.1 Results

Figure 1.5 shows the evolution of posterior means for the number of clusters inferred by the proposed model across the age intervals associated with the 20 spline bases within Figure 1.4, over the temporal window analyzed. Such a quantity is generally stable for both infant mortality (age 0) and adult/elder mortality (from age 60 to 90). Instead, considerable variability in the number of clusters is observed for children and adolescents (from age 1 to 25) and late mortality (from age 91), with different patterns. Besides supporting the need of allowing the group structures among countries to vary locally with ages and periods, these trends can be interpreted as a measure of variability in mortality rates across countries, with larger number of groups corresponding to age intervals where the different countries display higher dissimilarities over periods. Consistent with this interpretation,

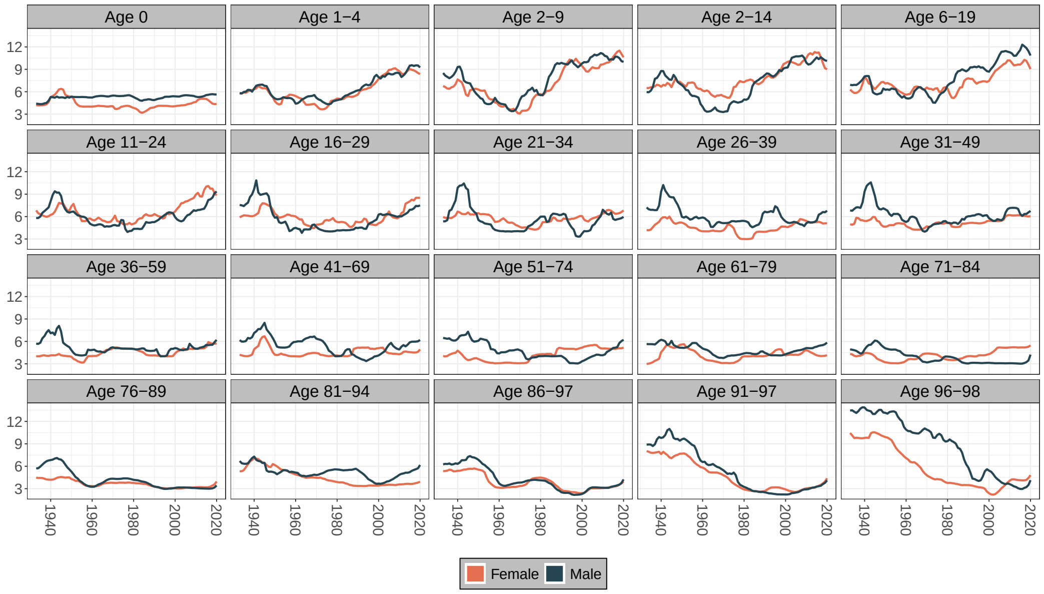


Figure 1.5: Time trajectories for the posterior mean of the number of clusters at the age intervals associated with the 20 selected B-spline bases.

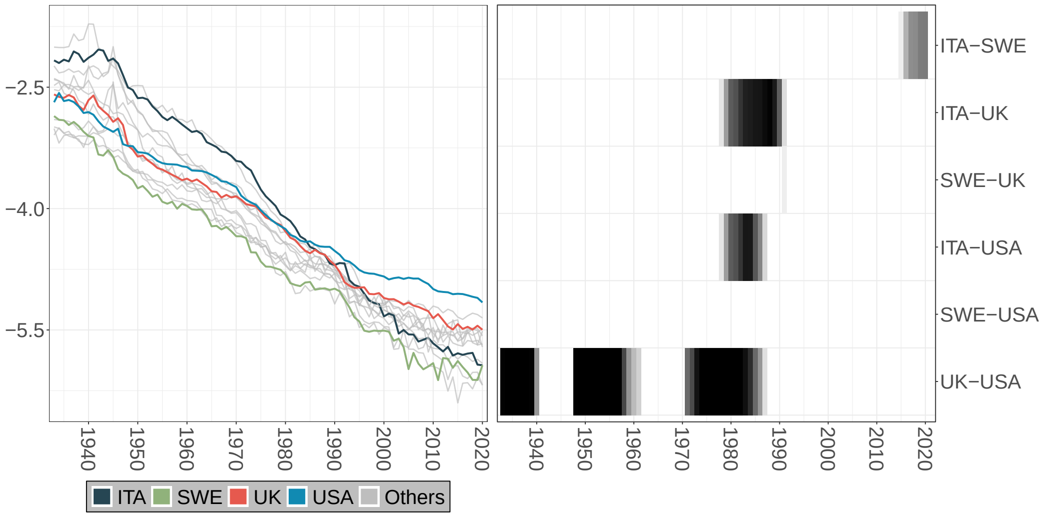
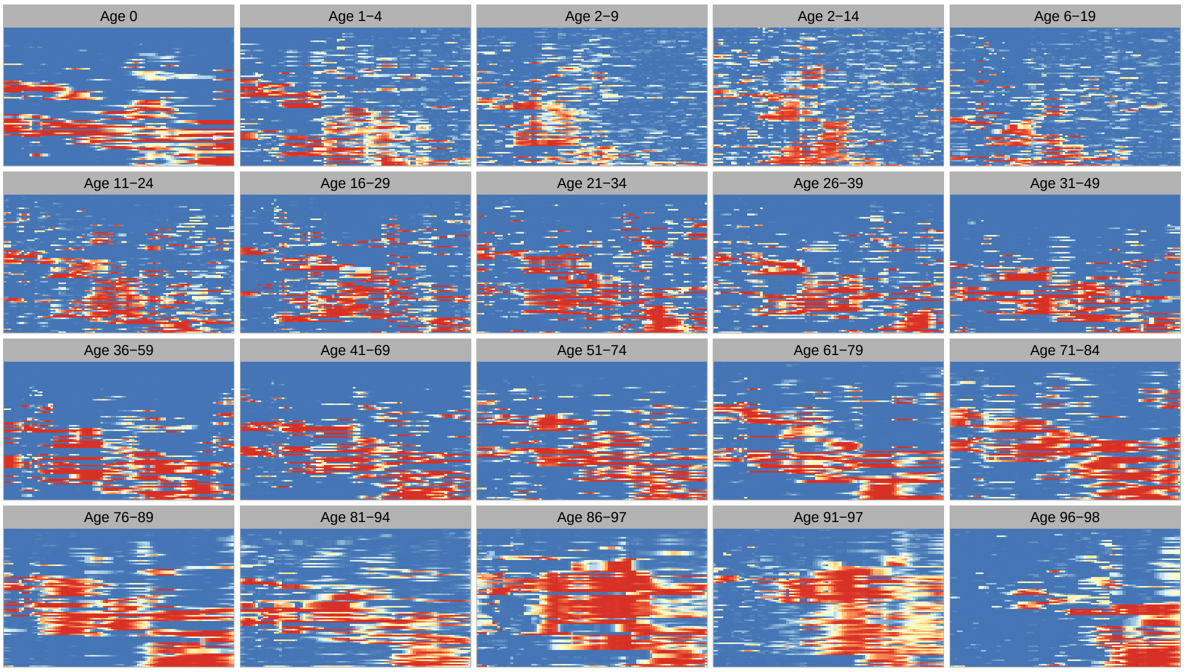


Figure 1.6: Observed male infant log-mortality rates for Italy (ITA), Sweden (SWE), the United Kingdom (UK) and the United States (USA) (left), and estimated dynamic co-clustering probabilities for the corresponding pairs (right).

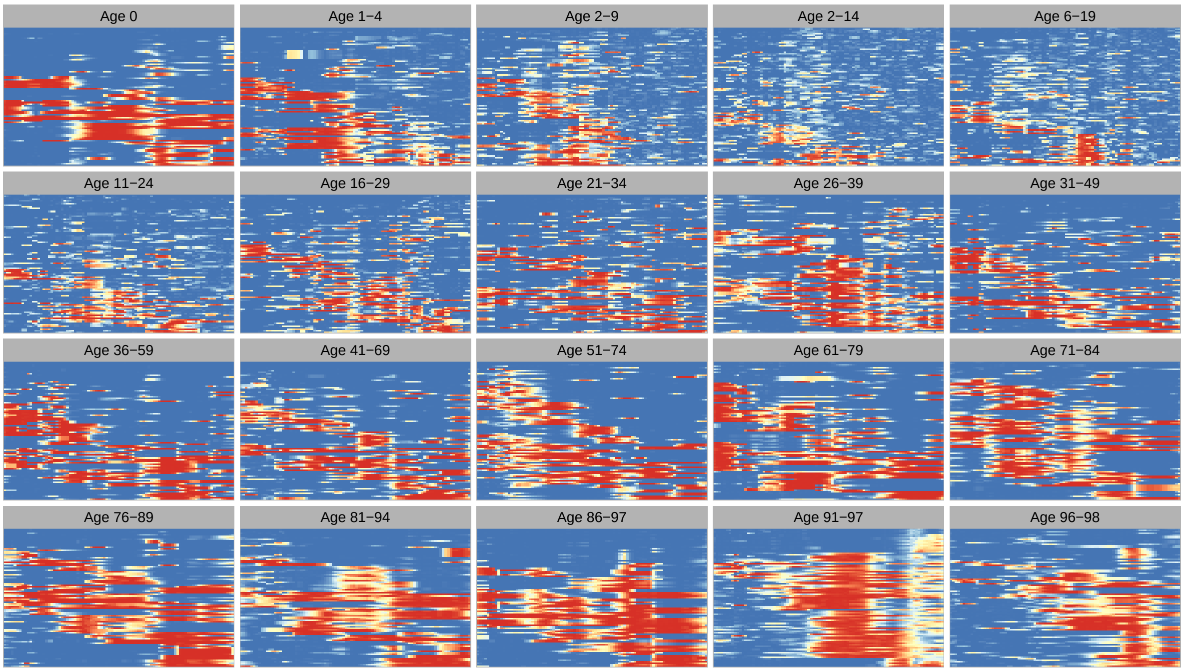
the results in Figure 1.5 suggest that senescent mortality is characterized by an increasing level of similarity among countries over the recent years, whereas child mortality exhibits an opposite trend after the early 1960s, indicating a divergence among countries that might have been driven by social phenomena such as the baby boom. Overall, males and females are characterized by similar patterns, with few notable exceptions. The number of clusters for elders (ages 91–98) converges to few groups for both sub-populations, although males are generally characterized by higher dissimilarities among countries since the first period under investigation. It is also interesting to notice how the mortality of young and adult males (age 16–40) is affected by World War II during the 1940s and reaches a plateau only after the peak associated with the military services.

The ability of the proposed model to effectively capture local clustering patterns is illustrated in Figure 1.6, where the observed infant log-mortality rates in Italy, Sweden, the UK and the USA are compared with the corresponding probabilities of co-clustering estimated from the samples of the Gibbs algorithm presented in Section 1.3.1. This specialized analysis is useful to further illustrate the main advantages of the newly-proposed approach and its implications for the global characterization of mortality rates. In particular, the right panel of Figure 1.6 indicates a fluctuating probability of co-clustering between the UK and the USA (bottom row), with larger values before 1940, in the early 1950s, and more recently from 1970 to 1980, in agreement with the observed trends of mortality rates reported in the left panel. Evidence of co-clustering is observed also for Italy with both the UK and the USA around the 1980s. In contrast, Sweden is characterized by a peculiar and separate trajectory, that overlaps with Italy only in recent years (top row).

To fully explore the evolution of the local co-clustering patterns, beyond the above specialized analysis, Figures 1.7a and 1.7b display the estimated co-clustering probabilities



(a) Male population



(b) Female population

Figure 1.7: Evolution of the estimated co-clustering probabilities for country pairs (rows) and years (columns), over all age intervals (panels). Colors correspond to the estimated probabilities, and range from blue (low) to red (high).

across all pairs of countries, for males and females respectively. Both figures illustrate the co-clustering probabilities for every B-spline basis (corresponding to a different age interval) through a matrix with rows denoting all pairs of countries and columns referring to calendar years. Colors range from blue to red as the estimated co-clustering probabilities vary from 0 to 1. The results in Figures 1.7a and 1.7b indicate, in general, an increasing overlap among countries over periods in terms of the associated age-specific log-mortality rates. Interestingly, evidence of increasing co-clustering is progressively more present both as time advances and as the population ages, demonstrating a dual-directional reinforcement of these similarities. For instance, in the eldest age interval (96–98) of the male sub-population, the proportion of country pairs with a high probability of co-clustering has steadily risen in recent years, forming a distinct block of countries with large probabilities of co-clustering in the bottom-right panel of Figure 1.7a. Such compression is stronger for older ages, with young age classes demonstrating smaller probabilities of co-clustering than elder ones. Furthermore, this phenomenon finds empirical evidence in both the male (Figure 1.7a) and female (Figure 1.7b) sub-populations, thereby supporting previous findings on demographic convergence (e.g., Vaupel et al., 2011; Wilson, 2011).

Focusing on child mortality, the younger age classes (from age 1 to 19) showcase a neat temporal pattern, with most of the co-clustering concentrated between the late 1950s and the early 1980s. This behavior is coherent with the evolution of the number of clusters in Figure 1.5, and suggests that the countries under investigation have been characterized by different levels of child mortality until World War II, followed by the rapid improvement of mortality rates in the early 1960s that created few common clusters with similar mortality patterns. After the 1980s, the co-clustering structures become more irregular and countries deviate to more individual trends. This finding is worth future investigations.

The co-clustering patterns among countries for adult mortality are reported within the third row of Figure 1.7a and Figure 1.7b, and generally show a common diagonal structure with some groups of European countries experiencing important changes of cluster membership around the 1980s. This phenomenon can be more clearly appreciated, for the male sub-population, in the left panel of Figure 1.8, which reports the estimates of the dynamic spline coefficient associated with the ages 61–79. Such estimates show how the Netherlands and Denmark are characterized by low mortality rates until the 1970s, while Belgium, Finland and the UK experienced some of the highest rates during that period. Later on, mortality drops rapidly for most countries, whereas the Netherlands and Denmark experience increments that make these countries co-cluster with higher mortality ones. Structural changes in cross-country relations characterize this period. The potential impact of such changes is visible in the right panel of Figure 1.8, which displays the evolution of the spline coefficient associated with the age interval 91–97, and illustrates that after 1970s mortality levels of elder ages collapse into fewer clusters with stable trends for several years. On the other hand, the initial high volatility of the age-specific log-mortality rates of the older

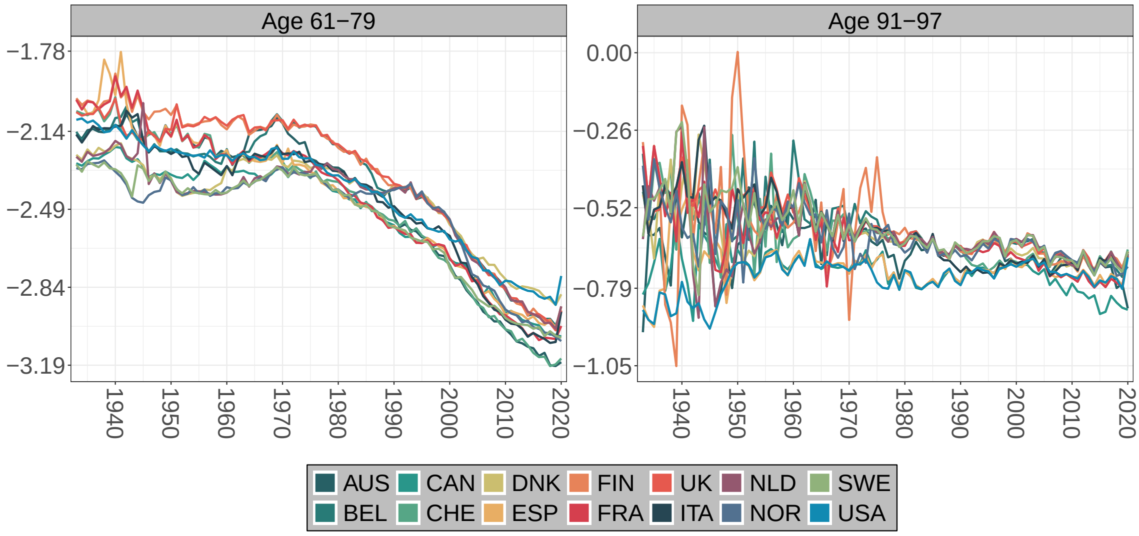


Figure 1.8: Estimates of β_{14} and β_{19} (age classes 61–79 and 91–97) in the male population.

populations is echoed by the estimates of the cluster-specific parameters, whose variability is implied by small cluster sizes, as depicted in Figure 1.5.

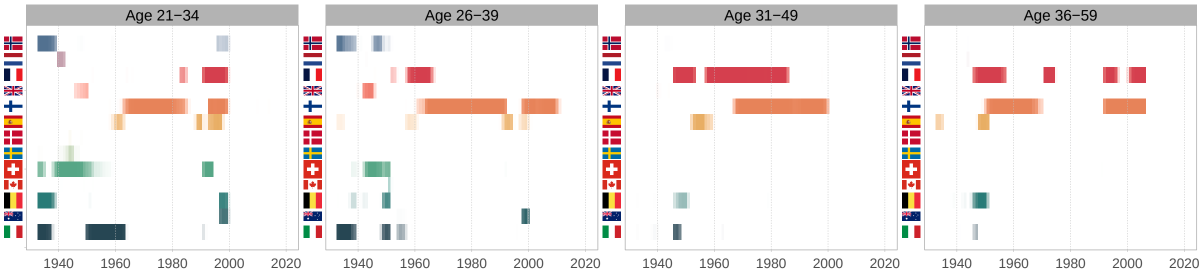
To further study the similarity between the group structures among countries inferred for the male and female log-mortality rates, we have also computed the normalized variation of information distance (NVI) (e.g., Wade and Ghahramani, 2018) among the group membership vectors estimated for these two sub-populations. Overall, the similarity between the partitions is more evident for the two youngest age classes (0 and 1–4), with values of the NVI below 0.44 and 0.58, respectively, throughout the entire time window. For example, female infant mortality in 2020 is divided into four groups. The first comprises the USA, the second Canada, the third Central European countries (Belgium, Switzerland, Denmark, France, the Netherlands and the UK) and Australia, and the fourth encodes both Mediterranean Europe (Italy, Spain) and Scandinavian countries (Finland, Norway, Sweden). The corresponding partition for the male sub-population consists, instead, of five groups, with the same first three as for the female sub-population, and the remaining two obtained by splitting Mediterranean Europe and Scandinavia into two separate clusters. More remarkable differences between male and female sub-populations are observed among middle-aged individuals, with NVI values above 0.5 after the 1960s. Interestingly, for the elderly population, a trend of increasing similarity started in the late 1980s, mirroring the patterns observed for young ages. Overall, these trends reflect a progressive convergence of mortality patterns across countries (Vaupel et al., 2011) for both males and females.

1.5.2 A Specialized Focus on the United States

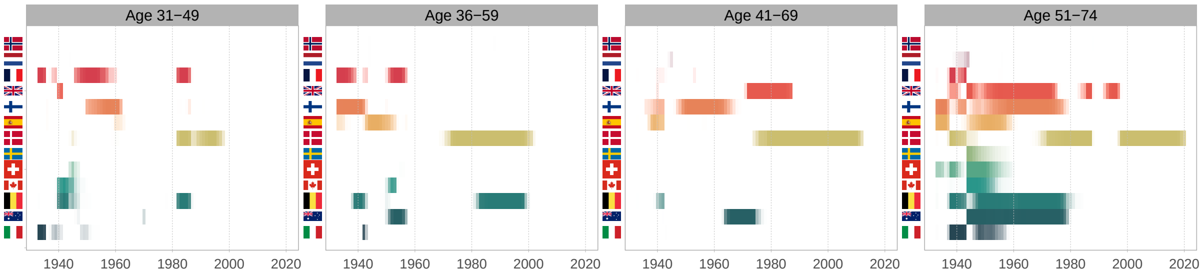
We devote here special attention to the analysis of the co-clustering probabilities between the USA and the other countries, estimated under the proposed model. This specialized focus is motivated by the fact that the USA has exhibited peculiar patterns in recent years, with rising mortality rates among certain demographic groups and a decline of life expectancy (see, e.g., Bergeron-Boucher et al., 2020; Case and Deaton, 2021; Glei, 2022). This phenomenon has been extensively studied in the literature, with a general consensus attributing these unexpected mortality increments to factors such as persistent disparities in healthcare access, increasing suicide rates, and more recently, to the opioid epidemic (see, e.g., Woolf and Schoomaker, 2019). To complement and extend these findings, it is therefore particularly interesting to study how the USA co-clusters with the other countries under investigation in terms of mortality rates.

Consistent with the above goal, Figure 1.9a displays the dynamic probabilities of co-clustering among the USA and the other countries at selected age intervals for the male sub-population, estimated under the proposed model. The results in Figure 1.9a point toward a persistent co-clustering among the USA and Finland. This finding indicates strong similarities in terms of premature male mortality and can be attributed to suicide incidence and cardiovascular diseases. Indeed, although Scandinavian countries provide universal and publicly funded welfare systems for all citizens, significant disparities still persist across the socio-economic spectrum due to the so-called "Nordic Paradox" (see, e.g., Mackenbach, 2017; Højstrup et al., 2023). As such, Figure 1.9a suggests that the "death of despair" phenomenon in the USA (e.g., Case and Deaton, 2021; Glei, 2022) might have interesting similarities with the "Nordic Paradox".

Focusing on the female sub-population, Figure 1.9b provides evidence of a relatively persistent co-clustering among the USA and Denmark females at adult ages, varying across time and with age intervals. Such a peculiar result should be further investigated, since



(a) Male population



(b) Female population

Figure 1.9: Estimated dynamic co-clustering probabilities between the USA and the other countries, for selected age intervals. Colors range from light to dark as the probability varies from low to high.

the interwar generation of Danish females represents a relevant demographic group that comprises cohorts who experienced a distinctive stagnation in life expectancy, resulting in a divergent trajectory compared to other Scandinavian countries (Lindahl-Jacobsen et al., 2016). As such, recent improvements in life expectancy can be linked to a cohort effect associated with this group that is visible also from the co-clustering patterns in Figure 1.9b. Interestingly, Belgian females join these clusters in specific age groups and time periods; a similar trend can be observed for UK at age 41–69. Such patterns might be related to a higher proportion of diseases of the circulatory system for Belgian females (Bergeron-Boucher et al., 2020) and to general declines in life expectancy in the UK (Ho and Hendi, 2018).

1.5.3 Associations Among Local Clustering Structures and Socio-Economic Indicators

The results in Sections 1.5.1–1.5.2 showcase relevant co-clustering structures among countries. These structures display convergence phenomena over periods for specific age classes (e.g., Oeppen and Vaupel, 2002), along with evidence of growing disparities for other ages. Available studies suggest that such patterns are driven by socio-economic factors such as quality of healthcare, education, and life standards (Marmot, 2005; Vallin and Meslé, 2004; European Commission and Executive Agency for Health and Consumers, 2013).

To provide additional empirical support to the above studies, we conclude our analysis by assessing to what extent the group structures inferred by the proposed model are associated with relevant socio-economic indicators, covering, in particular, the per-capita gross domestic product (GDP) in USD, unemployment rate, health expenditure (as the percentage of a country GDP), nutrition and child vaccination rates (against diphtheria, tetanus, pertussis); refer to OECD (2024) for more details. Data on these indicators are available, for all countries, across varying time periods. The earliest data for the GDP date back to 1970, whereas child vaccination rates and health spending are available starting from 1990. Finally, data on nutrition quality and unemployment rate start from 2010. Following standard practice, the association between these socio-economic indicators and the clustering structures among countries inferred by the proposed model is evaluated using the $\eta^2 \in [0,1]$ coefficient, which quantifies the variability of each indicator between the inferred clusters with respect to the total variability of such an indicator. Values of the η^2 close to 1 imply that the group structures among countries learned on the basis of similarities in the associated mortality rates are also explicative of the variability for the indicator under analysis, thereby suggesting possible associations. Notice that this measure does not inform on the direction of such an association, nor on possible causal interpretations. Nonetheless, it offers a sensible perspective for prioritizing specific indicators within explanatory studies on the socio-economic determinants of multi-country mortality patterns. Moreover, we acknowledge that some of the socio-economic indicators may affect mortality patterns only after some years, but the limited availability makes it difficult to carry out medium- to long-term association analysis.

Figure 1.10 reports, for both the female and male sub-population, the posterior means of the η^2 coefficients in matrix form. Rows correspond to the age intervals $j = 1, \dots, p$ associated with the different spline bases, columns to calendar years $t = 1, \dots, T$, and panels to socio-economic indicators $q = 1, \dots, Q$. To obtain these posterior means we first compute the η^2 between each posterior sample of $\mathbf{c}_{jt} = [c_{1jt}, \dots, c_{njt}]$ and the vector $\mathbf{w}_t^{(q)} = [w_{1t}^{(q)}, \dots, w_{nt}^{(q)}]$, with generic element $w_{it}^{(q)}$ denoting the value of the socio-economic indicator q, for country i in period t. This produces posterior samples of the η^2 for each age interval $j = 1, \dots, p$, period $t = 1, \dots, T$ and socio-economic indicator $q = 1, \dots, Q$. Averaging over these samples yields the posterior means in the matrices in Figure 1.10, which highlight a general concordance in the η^2 coefficients for the female and male sub-populations with non-trivial associations between the inferred clusters and the socio-economic indicators analyzed, particularly at young ages, and with different patterns across calendar years. For example, nutrition, GDP, health spending and unemployment rate appear to be associated with mortality clusters through a lower-diagonal cohort structure spanning a large spectrum of age classes, from childhood until young-adults and adults. This suggests that past incentives have had a lasting impact on specific cohorts, ultimately improving the lifespan of individuals who directly benefited from these initiatives. Not surprisingly, child vaccination rates and nutrition are generally less associated with clusters observed at age 0. In fact, such interventions cannot prevent neonatal and infant mortality, which are caused by specific

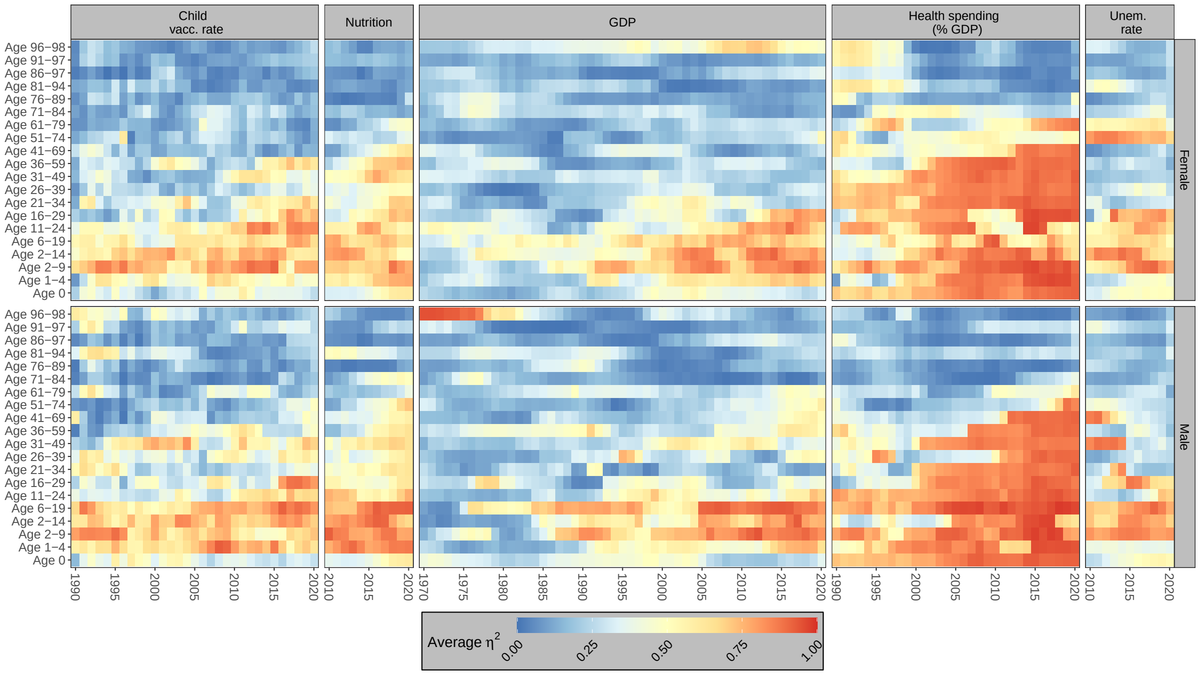


Figure 1.10: Posterior mean of η^2 coefficients between selected socio-economic indicators and the inferred mortality-based clusterings in the female and male sub-populations.

factors (such as birth defects, sudden infant death syndrome or accidents; see MacDorman et al., 2013; Eberstein et al., 1990). Instead, these indicators show a stronger association with the inferred clusters from age 1 until late adolescence, consistent with the protective effects of vaccination and proper nutrition in such ages.

1.6 Conclusions and Future Research Directions

Available statistical models for mortality data are either designed for analyzing single countries in isolation or for inferring global group structures among multiple countries with respect to the entire age-period mortality surface. This perspective prevents from unveiling more nuanced similarity patterns that are observed in practice over specific combinations of ages and calendar years, thereby limiting the possibility to learn and quantify relevant demographic phenomena localized at specific age classes and periods.

In this work, we overcome the above limitations through a novel multi-country model that characterizes the age pattern of mortality via a flexible B-spline expansion, and incorporates both temporal and age-specific clustering structures by allowing the coefficients of the B-spline bases to change in time via separate temporal random partition priors with cluster-specific Gaussian processes regulating the dynamic evolution of such coefficients. The resulting Bayesian formulation is amenable to tractable posterior inference via a Gibbs-sampling algorithm that facilitates interpretable reconstruction of group structures among countries, varying locally with both ages and periods. The unique advantages of the newly-proposed model are illustrated in simulation studies and in an application to mortality data for 14 OECD countries, where our formulation unveils fundamental local clustering structures, including unexplored ones that motivate future research in the area. For example, the in-depth analysis of the USA clustering patterns in Section 1.5.2 reveals distinctive similarities with specific Scandinavian countries for adult populations. In particular, the persistent similarity with Danish female log-mortality rates motivates further investigation. This sub-population is known to have exhibited distinctive mortality behavior due to significant smoking prevalence during World War II. Furthermore, the inferred associations between the local clustering patterns reconstructed by the proposed model and relevant socio-economic indicators (see Section 1.5.3) highlight remarkable cohort effects that are worth future analyses.

Future research includes applying the proposed model to a broader range of countries, while refining the analysis on the association among the group structures induced by mortality patterns and central socio-economic indicators. Both perspectives require overcoming challenges related to the availability of historical data. In fact, the 14 countries considered in this work are the only ones providing mortality rates that date back to our study period; analyzing more countries would imply constraining the analysis to narrower time windows. Improving inference on the associations with socio-economic indicators would require, instead, including such information directly within the proposed model. A possible direction for addressing this goal is to combine the tRPM prior with a reinforcement mechanism favoring the formation of local clusters that are homogeneous also with respect to the associated socio-economic indicators. This could be accomplished by extending the combination among product partition models with covariates (Müller et al., 2011) and tRPM priors proposed in Page et al. (2022) to the case of dynamic covariates. Lastly, it might be worth considering a more flexible specification that incorporates country-age- (and potentially also year-) specific variances σ_{ix} (or σ_{ixt}) of the log-mortality rates, which would allow to account for the dynamic size of the exposed population. However, this would require a larger number of parameters, which could be addressed by exploring a lower-dimensional specification.

Finally, let us emphasize that although our model is motivated by demographic applications, the constructions and results in Sections 1.2–1.3 have broader methodological impact, and can be applied whenever interest lies in the detection of localized overlaps among surfaces associated with different populations. To our knowledge, methodological results in these directions are limited.

Chapter 2

Dependent Stochastic Block Models for Sequences of Directed Networks with Application to Causes of Death Co-Occurrences

JOINT WORK WITH C. CASTIGLIONE AND D. DURANTE

2.1 Introduction

Although demographic research has traditionally focused on investigating mortality phenomena via overall longevity indicators (e.g., Canudas-Romo, 2010; Van Raalte, 2021), recent state-of-the-art studies have suggested that a comprehensive understanding of the core determinants behind modern mortality trends necessarily requires a finer-scale analysis disaggregating such trends across causes of death (e.g., Egidi et al., 2018; Woolf and Schoomaker, 2019; Canudas-Romo et al., 2020; Bergeron-Boucher et al., 2020; Mehta et al., 2020; Grippo et al., 2020; Stefanucci and Mazzuco, 2022; Trias-Llimós and Permanyer, 2023; Calazans and Permanyer, 2023). Besides yielding a deeper understanding of mortality patterns, these analyses are also of paramount importance to devise innovative policies in public health, and evaluate the corresponding effects across multiple, often interrelated, causes of death (e.g., Aburto et al., 2018; Bergeron-Boucher et al., 2020).

The relevance of the above endeavor combined with the availability of increasingly-refined data resources (see, e.g., the "WHO Mortality Database", the "Human Causes of Death Data Series“, the "Global Burden of Disease“, and the "US National Center for Health Statistics“), has produced an unprecedented understanding of modern mortality trends that lacked a clear explanation under classical studies focused on aggregated longevity indicators. These advancements have been achieved through the design of inference and predictive methods for the compositions, determinants, diversification and trends of either the *underlying* cause of death (i.e., “the disease or injury which initiated the chain of events leading directly to death“ (World Health Organization, 2016)) (see, e.g., Foreman et al., 2018; Bergeron-Boucher et al., 2020; Mehta et al., 2020; Stefanucci and Mazzuco, 2022; Depaoli et al., 2024; Ahmad et al., 2024; Huynh and Ludkovski, 2024) or multiple causes of death (MCOD), comprising both the *underlying* one and its *contributing* causes (i.e., “all other significant conditions contributing to death but not resulting in the underlying cause“ (World Health Organization, 2016)) (see, e.g., Désesquelles et al., 2010, 2012, 2014a; Moreno-Betancur et al., 2017; Egidi et al., 2018; Grippo et al., 2020; Trias-Llimós and Permanyer, 2023; Bishop et al., 2023; Grippo et al., 2024). Recalling the comprehensive review by Bishop et al. (2023), among these two perspectives, the second has been object of increasing interest in the recent years since it aligns more closely with the fact that death events are commonly associated with complex systems of multiple interrelated causes, rather than a single one in isolation (e.g., Israel et al., 1986; Redelings et al., 2006; Désesquelles et al., 2014a; Trias-Llimós and Permanyer, 2023). As a consequence, MCOD analyses have potential to provide more refined and realistic understanding of modern mortality patterns, while opening the avenues to study the higher-level system of relational structures among underlying and contributing causes, along with its changes across age classes.

Leveraging the available data from death certificates, the aforementioned relational system can be naturally represented via a sequence of networks whose directed edges measure, for each age class, the strength of the co-occurrence relation “cause i appears as the underlying of the contributing cause j”. However, despite its potential in unveiling yet-unexplored relational structures among causes of death with promising policy impact (e.g., Désesquelles et al., 2010, 2012, 2014a), this network perspective has been mostly overlooked in MCOD studies. As discussed in the following, a key barrier toward advancing along this direction can be found in the lack of suitable statistical models capable of uncovering relevant and interpretable structures that drive the formation and evolution of the complex co-occurrence patterns among underlying and contributing causes of death, across ages. In fact, similarly to epidemiology studies of co-morbidity networks (e.g., Jeong et al., 2017; Fotouhi et al., 2018; Jones et al., 2023), current network-based MCOD analyses (e.g., Egidi et al., 2018; Ukolova and Burcin, 2023) rely on traditional summary measures applied to simplified versions of the original data, which specialize the analysis to a single age class and do not consider the distinction between underlying and contributing causes. Therefore, although these contributions have the merit of showcasing the potential of the network perspective within MCOD studies, the resulting findings are descriptive in nature and do not inform on how relational structures among underlying and contributing causes vary across ages.

As showcased in our application to USA mortality data in 2019 (see Sections 2.1.1 and 2.5), overcoming the above limits necessarily requires a model-based perspective capable of accounting for the full complexity of causes-of-death co-occurrence patterns across several dimensions, while quantifying uncertainty in the inferred structures behind these observed patterns. Advancements along these lines are not only essential to possibly refine current health care investment policies in a phase characterized by a higher diversification, increasing complexity and lower predictability of multiple causes-of-death landscapes (see, e.g., Bergeron-Boucher et al., 2020; Trias-Llimós and Permanyer, 2023), but could also help in resolving recent debates on the determinants behind modern mortality trends. A relevant one, which motivates our focus on USA causes-of-death data, revolves around the recent stagnation in the USA life expectancy (e.g., Woolf and Schoomaker, 2019; Mehta et al., 2020; Case and Deaton, 2021). While such a stagnation has been attributed to a growing mid-life mortality, the determinants underlying this mortality increment have generated a debate around different views that either support causes such as drug overdoses, alcohol abuses and suicides ("death of despair") (e.g., Woolf and Schoomaker, 2019; Case and Deaton, 2021) or identify cardiovascular diseases as main drivers (e.g., Mehta et al., 2020). Albeit different, both views arise from the study of underlying causes. Hence, as illustrated in Section 2.5, the current debate can find a consensus under a refined model-based analysis of causes-of-death networks that is capable of inferring age-specific block interactions among endogenous groups of underlying and contributing causes displaying similar co-occurrence patterns. These inferred clustering structures and the associated block interactions might, in fact, unveil unexplored modules among seemingly unrelated causes that were object of past debates. In addition, as illustrated within Figure 2.5, such a perspective yields an interpretable reconstruction of complex modules in causes-of-death networks that unveil within- and across-group diversification structures in both underlying and contributing causes, along with the associated changes at different age classes. This is a key to shift the focus of current investment policies away from targeting a single underlying cause with a high prevalence in the population and toward jointly prioritizing internally-homogenous groups of underlying and contributing causes characterized by remarkable modular co-occurrences at given age classes.

Motivated by the above endeavor, we generalize stochastic block model (SBM) representations (e.g., Holland et al., 1983; Nowicki and Snijders, 2001) to learn informative structures in age-indexed sequences of categorically-weighted directed networks among underlying and contributing causes of death. As clarified in Sections 2.2–2.3, the proposed formulation (i) learns two separate group structures for underlying and contributing causes, respectively, (ii) allows these structures to change smoothly across age classes via dependent random partition priors (Page et al., 2022) further informed by external macro-classifications of death causes through product partition models (Müller et al., 2011), (iii) automatically estimates the number of groups for each network in the sequence, (iv) accounts for flexible block interactions among these inferred groups and (v) facilitates principled uncertainty quantification and inclusion of expert knowledge under a Bayesian approach to inference.

Although classical SBMs have witnessed effective extensions in several directions over the recent years, a flexible formulation addressing (i)–(v) within a single construction is lacking in the literature. In fact, while state-of-the-art dynamic stochastic block models could be possibly adapted to our motivating application by replacing time with age, available formulations (e.g., Ishiguro et al., 2010; Yang et al., 2011; Xu and Hero, 2014; Xu, 2015; Matias and Miele, 2017; Pensky and Zhang, 2019; Goto et al., 2021) are not designed to infer two separate partitions for the rows and columns of the adjacency matrices characterizing the observed directed networks. Furthermore, these models lack strategies to inform such grouping structures by external node attributes, often focus on binary edges, and generally cannot learn automatically the number of groups for each network in the sequence. While these issues have been addressed separately in the literature (e.g., Tallberg, 2004; Kemp et al., 2006; Mariadassou et al., 2010; Rohe et al., 2016; Zhang et al., 2016, 2022; Geng et al., 2019; Legramanti et al., 2022; Durante et al., 2025), the overarching focus of available contributions has been on static, single-network, settings, rather than on sequences of networks indexed by an ordered covariate (e.g., time or age).

The importance of covering the above methodological gap is illustrated through realistic simulation studies in Section 2.4, where the model we propose in Sections 2.2–2.3 is shown to outperform and extend the inference potential of state-of-the-art dynamic SBMs that could be possibly employed in the applied settings motivating our contribution. These advantages are strengthened in the application to USA mortality data within Section 2.5. In this case, our model unveils yet-unexplored structures in the composition, evolution, diversification and modular interactions among underlying and contributing causes-of-death groups that were hidden to previous demographic studies. Such findings may have important policy implications and open the avenues to achieve a more comprehensive understanding of the determinants underlying the recent stagnation in the USA life expectancy (e.g., Woolf and Schoomaker, 2019; Mehta et al., 2020; Case and Deaton, 2021). Concluding remarks can be found in Section 2.6.

2.1.1 USA Causes of Death Data

As anticipated in Section 2.1, our motivating application arises from the attempt to unveil higher-level and more nuanced determinants behind the alarming stagnation in the USA life expectancy (e.g., Woolf and Schoomaker, 2019; Mehta et al., 2020; Case and Deaton, 2021) through a novel perspective that avoids overly-simplified analyses of the underlying cause of death, but rather studies the complex interaction systems among multiple, interrelated, causes.

To this end, we focus on age-indexed sequences of co-occurrence networks among underlying and contributing causes extracted from the $\approx$ 2,860,000 death certificates recorded in the USA for 2019 (see https://wonder.cdc.gov/mcd-icd10.html). Such certificates are issued as part of the National Vital Statistics System maintained by the National Center for Health Statistics (NCHS), and document, for each death event, key demographic information, such as gender and age, alongside a single underlying cause and its contributing ones. All these causes are identified according to the International Classification of Diseases (ICD) system (now in its 11th revision), which employs different levels of granularity, ranging from a single alphabetic character associated with broad macro-categories, to the finest-scale seven-character code that identifies highly-detailed sub-categorizations for each cause. Such a latter classification yields thousands of mortality factors, with a large portion comprising extremely-specific and highly-rare causes. For this reason, and consistent with our overarching focus, we consider the classification based on the first two characters of the ICD codes, upon discarding never-observed causes and those not related to diseases and pre-existing medical conditions (i.e., injuries and external). This choice yields a total of $n = 139$ causes under analysis, thereby achieving an effective balance between overly-aggregated studies that fail to unveil nuanced patterns, and excessively-fine classifications which do not facilitate interpretable analyses and may be more prone to reporting issues due to increased challenges in disentangling highly-similar causes (e.g., Redelings et al., 2007; Désesquelles et al., 2010, 2012, 2014a; Bishop et al., 2023). Leveraging routine practice in demographic studies (see, e.g., Désesquelles et al., 2010; Lozano et al., 2012; Trias-Llimós and Permanyer, 2023; Depaoli et al., 2024), we further stratify the data by age classes defined as $[0, 1]$, $(1, 10]$, $(10, 20]$, $\dots$, $(90, 100]$, where the first group accounts for the peculiar patterns associated with infant death events.

Leveraging the above stratifications of the death certificates together with the causes-of-death classification considered, it is thus possible to record, for each age class x, the total number of certificates that have cause i as the underlying of the contributing j, for every $i = 1, \dots, n$ and $j = 1, \dots, n$. While these pairwise counts already yield a sequence of causes-of-death co-occurrence networks across ages, we opt for analyzing a discretized version of such directed networks which classifies the original counts into four interpretable levels, defined as `absent` (0), `rare` (1–10), `present` (11–100), and `frequent` (>100). This choice is motivated by two main reasons. First, it facilitates interpretation by providing policy experts with findings based on a simple and intuitive four-level measure of co-occurrence strengths among causes of death. Second, it is beneficial in reducing the noise and partial distortions that may arise from the aggregation of hundreds of thousands of death certificates compiled by different specialists across USA in a year. Note that, although different thresholds may be considered, those we identify are based on the analysis of the empirical distribution of the co-occurrence counts, whose median across ages and causes-of-death pairs is ≈ 15. This motivates our focus on multiples of 10, which proved also robust when considering other connectivity measures among causes of death that normalize the pairwise counts with respect to the degree of appearance of the associated causes within the certificates (e.g., Hidalgo et al., 2009; Chmiel et al., 2014; Fotouhi et al., 2018); see Section 2.6.

Figure 2.1 illustrates the adjacency matrices associated with the resulting networks under analysis, at selected age classes, for both the male and female populations. To provide preliminary quantitative evidence that supports the model developed in Sections 2.2–2.3,

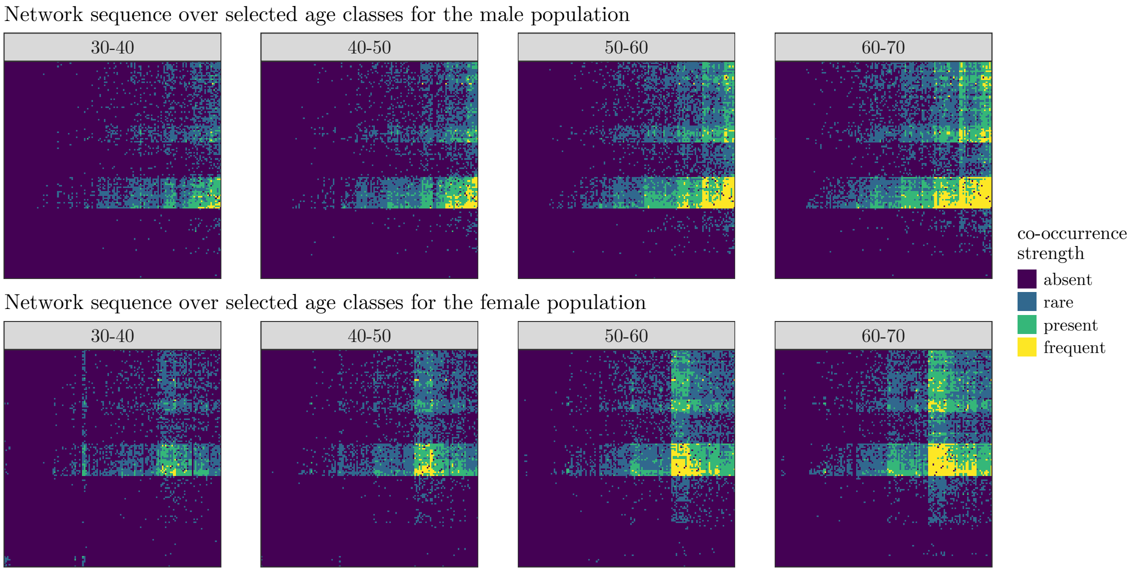


Figure 2.1: For the male and female populations, graphical representations of the adjacency matrices associated with the cause-of-death networks studied, at four selected consecutive age classes. In each matrix, rows (underlying causes) and columns (contributing causes) are re-ordered according to the two group structures identified by a basic hierarchical clustering algorithm. Such an algorithm is applied separately to the two dimensions of each adjacency matrix leveraging the Hamming distance (Hamming, 1950) as a measure of dissimilarity among rows and between columns, respectively.

the entries of each matrix are re-ordered to highlight the row and column groups learned by a basic hierarchical clustering algorithm (see the `R` library `pheatmap`). This algorithm is applied separately to the two dimensions of each adjacency matrix, corresponding, respectively, to underlying (rows) and contributing (columns) causes, and employs the Hamming distance (Hamming, 1950) as a measure of dissimilarity among the rows and between the columns, respectively. As shown in Figure 2.1, there is a clear evidence of block co-occurrence structures in the directed causes-of-death networks under analysis. These blocks are induced by group patterns among rows and between columns that cannot be assumed equal, but rather should be studied as two separate partitions. Furthermore, albeit estimated separately for each age class, such partitions seem to vary smoothly across consecutive ages in the observed networks.

The above preliminary quantitative findings motivate the dependent SBM for age-indexed sequences of directed causes-of-death networks developed in Sections 2.2–2.3. In fact, while the results in Figure 2.1 already provide relevant insights, basic hierarchical clustering is not designed to borrow information between row and column partitions across ages, and does not provide a principled model-based representation of the complex joint system of interactions among underlying and contributing causes across age classes. These issues do not facilitate rigorous inference, uncertainty quantification, supervision by meaningful external information and, possibly, assessment of the effects of different policy scenarios via simulations from the generative model. Moreover, the results in Figure 2.1 are descriptive in nature, and hence, are also more prone to suffer from the possible distortions that arise from reporting practices of death certificates, an issue which is not specific to our analysis, but rather common to all causes-of-death studies (see, e.g., Redelings et al., 2007; Désesquelles et al., 2010, 2012, 2014a; Bishop et al., 2023). Although recent advancements and automated reporting procedures have substantially improved the quality of death certificates, mitigating possible biases through careful data-preprocessing and principled statistical models that account for uncertainty is still important. Our contribution moves along these lines, while further addressing potential reporting issues through the supervision by biologically-meaningful external classifications of the causes of death. In this respect, it shall be emphasized that certain groups inferred by the novel SBM we propose in Sections 2.2 and 2.3 may be also useful in unveiling systematic reporting practices associated with specific underlying and contributing causes at given age classes. As a result, our contribution can also help in improving the understanding of such practices, thereby motivating the design of additional guidelines in causes-of-death reporting.

2.2 Dependent Stochastic Block Models for Directed Networks

Define the four levels `absent`, `rare`, `present` and `frequent` introduced in Section 2.1.1 through the numerical labels $w = 1, \dots, 4$. Moreover, recode the age classes $[0, 1]$, $(1, 10]$, $(10, 20]$, $\dots$, $(90, 100]$ into the ordered indexes $x = 1, 2, \dots, X$ (with $X = 11$ in our application). Then, the networks analyzed are available in the form of a sequence $\mathbf{Y}_1, \dots, \mathbf{Y}_X$ of $n \times n$ categorically-weighted asymmetric adjacency matrices, where the generic $\mathbf{Y}_x$ has entries $y_{ijx} = w$ if, at age class x, the strength of the directed co-occurrence relation "cause i appears as the underlying of the contributing cause j" is equal to w.

In Section 2.2.1, we generalize state-of-the-art SBMs to define a probabilistic generative mechanism for $\mathbf{Y}_1, \dots, \mathbf{Y}_X$, which addresses points (i)–(v) discussed in Section 2.1 via a single formulation. Such a joint model is designed to infer age-specific block interactions among endogenous groups of underlying and contributing causes displaying similar co-occurrence patterns. To this end, we employ, and learn, two separate grouping structures that are allowed to change smoothly across age classes via dependent random partition priors further informed by macro-classifications of death causes. Such priors are inspired by contributions of Page et al. (2022) and Müller et al. (2011), and are presented in Section 2.2.2.

2.2.1 Model Formulation

Guided by the empirical evidence in Figure 2.1, we extend available SBMs to infer separate group structures for the underlying and contributing causes, respectively, from the block patterns displayed by the observed adjacency matrices $\mathbf{Y}_1, \dots, \mathbf{Y}_X$. Recalling Section 2.1, current SBMs (e.g., Holland et al., 1983; Nowicki and Snijders, 2001; Schmidt and Morup, 2013; Lee and Wilkinson, 2019; Geng et al., 2019; Legramanti et al., 2022) mostly focus on binary undirected networks, thus requiring only a single partition to flexibly characterize the modular architectures in the observed network. Such a single partition is often employed also in directed settings under the assumption that the rows and columns of the asymmetric adjacency matrix display a shared group structure (see, e.g., Wang and Wong, 1987; Newman and Leicht, 2007; McDaid et al., 2013; Peixoto, 2022). Although this perspective simplifies the model, in practice, it may provide an unrealistic characterization of the data generative mechanisms in directed settings, thus raising misspecification issues that yield to biased inferences. In fact, as illustrated in Figure 2.1, it is reasonable to expect that two generic causes of death having similar co-occurrence patterns when treated as underlying, may not display the same similarities when considered as contributing. In addition, these co-clustering relations might also vary across different age classes.

Consistent with the above discussion we employ, for each age-class index $x = 1, \dots, X$, two separate partitions (one for the underlying causes and the other for the contributing ones), whose Cartesian product yields a block structure on the adjacency matrix $\mathbf{Y}_x$ that clusters together, within each block, pairs of underlying and contributing causes with similar co-occurrence strengths. More specifically, let $\boldsymbol{c}_x = (c_{1x}, \dots, c_{nx})$ and $\boldsymbol{d}_x = (d_{1x}, \dots, d_{nx})$ be the group membership vectors associated to the generic row and column partitions $\mathbb{c}_x = \{\mathcal{C}_{1x}, \dots, \mathcal{C}_{H_x x}\}$ and $\mathbb{d}_x = \{\mathcal{D}_{1x}, \dots, \mathcal{D}_{K_x x}\}$, such that $c_{ix} = h$, for $h = 1, \dots, H_x$, if and only if $i \in \mathcal{C}_{hx}$ and, analogously, $d_{jx} = k$, for $k = 1, \dots, K_x$, if and only if $j \in \mathcal{D}_{kx}$. Then, extending the classical Bernoulli likelihoods for binary edges to the categorically-weighted interactions characterizing the networks under analysis, we assume that $(y_{ijx} \mid c_{ix} = h, d_{jx} = k, \boldsymbol{\theta}_{hkx}) \sim \text{Cat}_{1:4}(\boldsymbol{\theta}_{hkx})$, independently for every $i = 1, \dots, n$, $j = 1, \dots, n$ and $x = 1, \dots, X$, where $\text{Cat}_{1:4}(\boldsymbol{\theta}_{hkx})$ is a categorical variable indexed by the probability vector $\boldsymbol{\theta}_{hkx}$, with entries $\theta_{hkxw} = \text{pr}(y_{ijx} = w \mid c_{ix} = h, d_{jx} = k) \in (0, 1)$, for $w = 1, \dots, 4$. Hence, consistent with general SBMs, within-block homogeneity is expressed by assuming that the distribution for each entry y_{ijx} of $\mathbf{Y}_x$ only depends on the corresponding row and column groups along with the probability vector characterizing the block associated to such a pair of groups. This means that the co-occurrence strengths among all underlying and contributing causes in groups h and k, respectively, are generated from the same categorical distribution, whose parameters can change across blocks (i.e., pairs of groups), but not within block, thereby accounting for flexible modular structures in the observed networks.

To complete our Bayesian formulation, we require priors on the group membership vectors $\boldsymbol{c}_x$ and $\boldsymbol{d}_x$, $x = 1, \dots, X$, together with the block-specific parameters $\boldsymbol{\theta}_{hkx}$, for $h = 1, \dots, H_x$, $k = 1, \dots, K_x$ and $x = 1, \dots, X$. Extending routine Bayesian SBMs (e.g., Nowicki and Snijders, 2001; Mariadassou et al., 2010; Legramanti et al., 2022) from binary to categorical relationships, it is natural to consider independent Dirichlet($\mathbf{a}^{\theta} = (a_1^{\theta}, \dots, a_4^{\theta})$) priors for the block-specific parameters $\boldsymbol{\theta}_{hkx}$, $h = 1, \dots, H_x$, $k = 1, \dots, K_x$ and $x = 1, \dots, X$. Besides corresponding to the multivariate generalization of the Beta priors employed for the block probabilities in binary networks, such a choice achieves conjugacy with the $\text{Cat}_{1:4}(\boldsymbol{\theta}_{hkx})$ distribution assumed for the entries of $\mathbf{Y}_x$. As such, the block-specific parameters can be integrated out analytically to obtain a Dirichlet-categorical joint likelihood for $\mathbf{Y}_1, \dots, \mathbf{Y}_X$ conditioned on the group membership vectors $\boldsymbol{c}_x$ and $\boldsymbol{d}_x$, $x = 1, \dots, X$. In particular

$$\begin{aligned} p(\mathbf{Y}_1, \dots, \mathbf{Y}_X \mid (\boldsymbol{c}_x, \boldsymbol{d}_x), x = 1, \dots, X) &= \\ &= \prod_{x=1}^{X} \left(\prod_{h=1}^{H_x} \prod_{k=1}^{K_x} \left[\frac{\Gamma(a^{\theta}_{\bullet})}{\Gamma(a^{\theta}_{\bullet} + n_{hkx\bullet})} \prod_{w=1}^{4} \frac{\Gamma(a^{\theta}_{w} + n_{hkxw})}{\Gamma(a^{\theta}_{w})} \right] \right), \end{aligned} \tag{2.1}$$

where n_{hkxw} represents the number of pairs (i, j) such that $c_{ix} = h$, $d_{jx} = k$ and $y_{ijx} = w$, while $n_{hkx\bullet} = \sum_{w=1}^{4} n_{hkxw}$ and $a^{\theta}_{\bullet} = \sum_{w=1}^{4} a^{\theta}_{w}$. The likelihood in (2.1) formally treats the quantities $\boldsymbol{\theta}_{hkx}$, $h = 1, \dots, H_x$, $k = 1, \dots, K_x$, $x = 1, \dots, X$ as nuisance parameters and focuses inference on the group structures encoded in $\boldsymbol{c}_x$ and $\boldsymbol{d}_x$, for $x = 1, \dots, X$. While learning the block-specific probability vectors $\boldsymbol{\theta}_{hkx}$ is also of interest, this perspective is common in SBM formulations (e.g., Wyse and Friel, 2012; McDaid et al., 2013; Schmidt and Morup, 2013; Legramanti et al., 2022; Durante et al., 2025), whose primary interest lies in uncovering groups of nodes that display similar connectivity patterns within the observed network. In our context, such an information is encoded in $\boldsymbol{c}_x$ and $\boldsymbol{d}_x$, for $x = 1, \dots, m$, thereby motivating our main focus on these two vectors under the Dirichlet-categorical joint likelihood in (2.1). As discussed in Section 2.3, this choice further facilitates posterior computation, and does not prevent from obtaining ex-post sensible estimates of $\boldsymbol{\theta}_{hkx}$, $h = 1, \dots, H_x$, $k = 1, \dots, K_x$, $x = 1, \dots, X$.

Notice that the likelihood in (2.1) factorizes across the age-class indexes $x = 1, \dots, X$. Although this formulation is useful to avoid overly-sophisticated and intractable representations, as shown within Figure 2.1, causes-of-death networks exhibit a form of dependence across ages, which is visible in terms of smooth transitions for the corresponding group structures over contiguous age classes. In Section 2.2.2, we incorporate this dependence through carefully designed priors for the two sequences $\boldsymbol{c}_1, \dots, \boldsymbol{c}_X$ and $\boldsymbol{d}_1, \dots, \boldsymbol{d}_X$ that further include external information on cause-of-death macro-classifications.

2.2.2 Prior for Dependent Sequences of Random Partitions

As anticipated in Section 2.2.1, we elicit priors for the two sequences $\boldsymbol{c}_1, \dots, \boldsymbol{c}_X$ and $\boldsymbol{d}_1, \dots, \boldsymbol{d}_X$ which enforce smooth, yet flexible, transitions for the group membership vectors across contiguous age classes, while informing these vectors through external macro-classifications of causes of death. Such a latter exogenous information is encoded in the vector $\mathbf{z} = (z_1, \dots, z_n)$, whose generic categorical entry z_i identifies the macro-category $l = 1, \dots, L$ (e.g., neoplasms, malformations, circulatory diseases, mental health problems, etc...) of cause $i = 1, \dots, n$. This macro-classification can be retrieved from the ICD system presented in Section 2.1.1, yielding a total of $L = 19$ macro-categories.

Consistent with the above discussion, we leverage the dependent random partition priors of Page et al. (2022) in combination with product partition model constructions (Müller et al., 2011) to obtain

$$\begin{aligned}(\boldsymbol{c}_1, \dots, \boldsymbol{c}_X \mid \mathbf{z}, \boldsymbol{\alpha}^c, \eta_c) &\sim \textsc{drpm-z}(\boldsymbol{\alpha}^c, \eta_c),\\ (\boldsymbol{d}_1, \dots, \boldsymbol{d}_X \mid \mathbf{z}, \boldsymbol{\alpha}^d, \eta_d) &\sim \textsc{drpm-z}(\boldsymbol{\alpha}^d, \eta_d),\end{aligned} \tag{2.2}$$

where $\boldsymbol{\alpha}^c, \boldsymbol{\alpha}^d, \eta_c$ and η_d are prior hyper-parameters whose meaning will be clarified in the following, whereas DRPM-Z denotes the assumed dependent random partition prior informed by $\mathbf{z}$. Such a prior is presented in detail in Section 2.2.2 for the sequence $\boldsymbol{c}_1, \dots, \boldsymbol{c}_X$, where the subscripts and superscripts c are removed for the sake of generality.

Constructive Representation

The DRPM-Z$(\boldsymbol{\alpha}, \eta)$ prior for a generic sequence of group allocations $\boldsymbol{c}_1, \dots, \boldsymbol{c}_X$ is defined hierarchically through a Markovian mechanism, which combines two main constructions. The first one is a flexible, yet smooth, transition mechanism from $\boldsymbol{c}_{x-1}$ to $\boldsymbol{c}_x$ regulated by n independent Bernoulli(α_x) variables $\gamma_{1x}, \dots, \gamma_{nx}$, with the generic γ_{ix} controlling whether the i–th cause of death is allowed to possibly change group allocation from age class $x-1$ to x (i.e., $\gamma_{ix} = 0$), or not (i.e., $\gamma_{ix} = 1$) (Page et al., 2022). The second is a carefully-designed supervised Chinese restaurant process prior CRP-Z(η) (Müller et al., 2011; Page et al., 2022; Legramanti et al., 2022) regulating both the formation of the group structures among causes at the first age class, i.e., $\boldsymbol{c}_1$, and the mechanism through which the subset of causes with $\gamma_{ix} = 0$, for $i = 1, \dots, n$, are re-allocated to clusters from $\boldsymbol{c}_{x-1}$ to $\boldsymbol{c}_x$, for each age $x = 2, \dots, X$.

Recalling classical results from Bayesian nonparametric statistical literature (e.g., see Gershman and Blei (2012) for an introductory overview) in connection with covariate-dependent product partition models (Müller et al., 2011), the assumed CRP-Z(η) prior has probability mass function for the generic group membership vector $\boldsymbol{c}$ of length n defined as $p(\boldsymbol{c} \mid \eta, \mathbf{z}) \propto [\eta^H \Gamma(\eta)/\Gamma(\eta + n)] \prod_{h=1}^{H} \rho_{\text{DC}}(\mathbf{z}_h, \mathbf{a}^{\mathbf{z}})(n_h - 1)!$. In this expression, $\Gamma(\cdot)$ corresponds to the Gamma function, n_h denotes the number of causes in group h, and $\mathbf{z}_h$ is the vector encoding the memberships to the L external macro-categories of the n_h causes in group h. This exogenous information enters $p(\boldsymbol{c} \mid \eta, \mathbf{z})$ via the Dirichlet-categorical cohesion function $\rho_{\text{DC}}(\cdot, \mathbf{a}^{\mathbf{z}})$ with parameters $\mathbf{a}^{\mathbf{z}} = (a_1^{\mathbf{z}}, \dots, a_L^{\mathbf{z}})$ (Müller et al., 2011, Ch. 4), which increases the prior probabilities of those groups that are homogenous with respect to the external macro-classification of the causes. As such, by including $\rho_{\text{DC}}(\mathbf{z}_h, \mathbf{a}^{\mathbf{z}})$, the CRP-Z$(\eta)$ supervises the classical CRP(η) prior by the external covariate $\mathbf{z}$ in a way that favors internally-homogenous groups of causes with respect to the associated macro-classification. Such a mechanism is evident when studying the conditional distribution for each c_i given $\boldsymbol{c}^{(-i)} = (c_1, \dots, c_{i-1}, c_{i+1}, c_n)$. Under the CRP-Z$(\eta)$ prior, this conditional distribution is

$$\mathrm{pr}(c_i = h \mid \boldsymbol{c}^{(-i)}, \eta, \mathbf{z}) \propto \begin{cases} n_h^{(-i)} \dfrac{\bar{n}_{hz_i}^{(-i)} + a^{\mathbf{z}}_{z_i}}{\bar{n}_{h\bullet}^{(-i)} + a^{\mathbf{z}}_{\bullet}} & \text{for } h = 1, \dots, H^{(-i)}, \\ \eta \dfrac{a^{\mathbf{z}}_{z_i}}{a^{\mathbf{z}}_{\bullet}} & \text{for } h = H^{(-i)} + 1, \end{cases} \tag{2.3}$$

for every $i = 1, \dots, n$, where $H^{(-i)}$ and $n_h^{(-i)}$ are the total number of non-empty groups and the cardinality of the h–th cluster, respectively, after removing cause i, while $\bar{n}_{hz_i}^{(-i)}$ corresponds to the number of causes in group h with the same macro-category of the i–th one. Finally, $\bar{n}_{h\bullet}^{(-i)} = \sum_{l=1}^{L} \bar{n}_{hl}^{(-i)} = |\mathbf{z}_h^{(-i)}|$, $a^{\mathbf{z}}_{\bullet} = \sum_{l=1}^{L} a^{\mathbf{z}}_l$ and $a^{\mathbf{z}}_{z_i}$ is the entry of $\mathbf{a}^{\mathbf{z}}$ corresponding to the macro-category of cause i. According to (2.3), conditioned on $\boldsymbol{c}^{(-i)}$, the i–th cause can either occupy a group already observed for the other causes, or generate a new one. The former event has probability which depends on the cardinality of the group, as for the classical CRP prior, further reinforced by a factor that favors the attribution of the i–th cause to those existing groups that have a higher fraction of causes with its same macro-category. The creation of a new cluster is instead regulated by the CRP concentration parameter $\eta > 0$ and by those of the Dirichlet-categorical cohesion function (i.e., $\mathbf{a}^{\mathbf{z}} = (a^{\mathbf{z}}_1, \dots, a^{\mathbf{z}}_L)$, with $a_l > 0$, $l = 1, \dots, L$). Consistent with (2.3), larger values for η yield a higher expected number of clusters in $\boldsymbol{c}$.

Besides illustrating the tractability of the CRP-Z(η) prior, along with the meaning of its parameters, the scheme in (2.3) plays also a key role in the Gibbs sampling algorithm designed in Section 2.3, and clarifies that the total number of groups in the generic $\mathbf{z}$ does not need to be pre-specified, but rather can be learned automatically. This is a substantial gain relative to popular SBMs for dynamic networks that could be possibly applied to our motivating application (see, e.g., Yang et al., 2011; Xu and Hero, 2014; Xu, 2015; Matias and Miele, 2017). In fact, such models assume knowledge of the total number of groups, or leverage information criteria which require estimation under different settings for H. This is a computationally challenging task in contexts where the number of groups varies with x.

Combining the above CRP-Z(η) formulation with the previously-introduced Bernoulli transition mechanism yields the following hierarchical construction for the DRPM-Z($\boldsymbol{\alpha}, \eta$) prior on the generic sequence $\boldsymbol{c}_1, \dots, \boldsymbol{c}_X$ of group membership vectors

$$\begin{aligned} &(\gamma_{ix} \mid \boldsymbol{\alpha} = (\alpha_2, \dots, \alpha_X)) \overset{\text{indep.}}{\sim} \text{Bernoulli}(\alpha_x), \qquad i = 1, \dots, n, \quad x = 2, \dots, X, \\ &(\boldsymbol{c}_1 \mid \eta, \mathbf{z}) \sim \text{CRP-Z}(\eta), \qquad (\boldsymbol{c}_x \mid \boldsymbol{c}_{x-1}, \boldsymbol{\gamma}_x, \eta, \mathbf{w}) \sim \text{CRP-Z}(\eta)|_{\mathbb{C}(\boldsymbol{c}_{x-1}, \boldsymbol{\gamma}_x)}, \quad x = 2, \dots, X, \end{aligned} \tag{2.4}$$

where $\boldsymbol{\gamma}_x = (\gamma_{1x}, \dots, \gamma_{nx})$, while CRP-Z$(\eta)|_{\mathbb{C}(\boldsymbol{c}_{x-1}, \boldsymbol{\gamma}_x)}$ is the supervised Chinese restaurant process prior constrained to the set of partitions $\mathbb{C}(\boldsymbol{c}_{x-1}, \boldsymbol{\gamma}_x)$ *compatible* with $\boldsymbol{c}_{x-1}$ under $\boldsymbol{\gamma}_x$, namely all the partitions among the n causes of death that can be derived from the one associated with the group membership vector $\boldsymbol{c}_{x-1}$ by reallocating only those causes for which $\gamma_{ix} = 0$. Recalling Page et al. (2022), under this constraint, $\mathrm{pr}(\boldsymbol{c}_x = \boldsymbol{c} \mid \boldsymbol{c}_{x-1}, \boldsymbol{\gamma}_x, \eta, \mathbf{z}) \propto p(\boldsymbol{c} \mid \eta, \mathbf{z}) \mathbb{1}(\mathbb{c} \in \mathbb{C}(\boldsymbol{c}_{x-1}, \boldsymbol{\gamma}_x))$, where $p(\boldsymbol{c} \mid \eta, \mathbf{z})$ is the previously-defined probability mass function of the CRP-Z(η) prior, whereas $\mathbb{c}$ is the partition of the n causes associated with the generic group membership vector $\boldsymbol{c}$. As such, the prior on $\boldsymbol{c}_x$ is anchored to $\boldsymbol{c}_{x-1}$ by allowing only a subset of causes to change group allocation, with the size of this subset regulated by the parameter $\alpha_x \in [0, 1]$. When $\alpha_x = 1$ the clustering structures among causes of death at age classes $x-1$ and x perfectly overlap, whereas a value of $\alpha_x = 0$ forces all causes to be re-allocated according to an unrestricted CRP-Z(η) prior, thereby removing the dependence on $\boldsymbol{c}_{x-1}$. Therefore, as α_x ranges from 0 to 1, the vectors $\boldsymbol{c}_{x-1}$ and $\boldsymbol{c}_x$ are favored to be increasingly similar.

Note that the above notion of smoothness is allowed to change flexibly across age classes via transition-specific parameters $\alpha_2, \ldots, \alpha_X$, which are assigned conjugate Beta hyperpriors, i.e., $\alpha_x \sim \text{Beta}(a_\alpha, b_\alpha)$, for $x = 2, \ldots, X$. This allows smoothness to be inferred adaptively from the observed matrices $\mathbf{Y}_1, \ldots, \mathbf{Y}_X$. Such a data-oriented perspective is considered also for the CRP concentration parameter η on which we assume a conjugate Gamma hyperprior (Escobar and West, 1995), namely $\eta \sim \text{Gamma}(a_\eta, b_\eta)$.

As illustrated in Section 2.3, this prior elicitation facilitates the design of a tractable tempered Gibbs-sampler for posterior inference on the sequences of partitions $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_X$ and $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_X$.

2.3 Bayesian Computation and Inference

Posterior inference for the sequences of partitions $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_X$ and $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_X$ that parameterize the Bayesian model presented in Section 2.2 is performed via Monte Carlo leveraging samples produced by a carefully-designed collapsed Gibbs sampler in combination with adaptive parallel tempering. This algorithm is derived in detail in Sections 2.3.1–2.3.2 by exploiting the tractability of the likelihood in (2.1) along with the hierarchical representation (2.4) of the priors in (2.2). Section 2.3.3, concludes by presenting useful inferential strategies to perform point estimation and uncertainty quantification on the group structures encoded in $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_X$ and $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_X$ leveraging the Gibbs samples.

2.3.1 Collapsed Gibbs Sampler

The proposed Gibbs algorithm samples iteratively from the full-conditional distributions of the group membership vectors $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_X$ and $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_X$, the persistence variables $\boldsymbol{\gamma}_2^c, \ldots, \boldsymbol{\gamma}_X^c$ and $\boldsymbol{\gamma}_2^d, \ldots, \boldsymbol{\gamma}_X^d$ appearing in the hierarchical representation (2.4) of the DRPM-Z priors in (2.2), and finally, the hyperparameters $\boldsymbol{\alpha}^c$, η_c, $\boldsymbol{\alpha}^d$, and η_d. As clarified in the following, by combining the priors presented in Section 2.2.2 with the collapsed likelihood in (2.1) yields closed-form full-conditionals for all these quantities, thereby facilitating the design of a tractable sampling scheme.

Focusing on the row-specific quantities $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_X$, $\gamma_2^c, \ldots, \gamma_X^c$, $\boldsymbol{\alpha}^c$ and η_c, let us first introduce some useful notation. In particular, let $\Gamma_x^c = \{i = 1, \ldots, n : \gamma_{ix}^c = 1\}$ be the set of underlying causes whose group allocation does not change from $x-1$ to x. Moreover, denote with $\mathbb{c}_x^{\Gamma_x^c}$ the partition induced by $\boldsymbol{c}_x$ considering only those underlying causes with indexes in Γ_x^c. Then, adapting the derivations in Page et al. (2022) to our construction, the full-conditional distributions for the binary persistence variables γ_{ix}^c, $i = 1, \ldots, n$, $x = 2, \ldots, X$, are Bernoulli with probabilities

$$\mathrm{pr}(\gamma_{ix}^c = 1 \mid -) = \frac{\alpha_x^c}{\alpha_x^c + (1-\alpha_x^c)\dfrac{p(\mathbb{c}_x^{\Gamma_x^{c(+i)}} \mid \eta_c, \mathbf{z})}{p(\mathbb{c}_x^{\Gamma_x^{c(-i)}} \mid \eta_c, \mathbf{z})}} \mathbb{1}\left[\Gamma_x^{c(+i)} = \Gamma_{x-1}^{c(+i)}\right], \tag{2.5}$$

for each node $i = 1, \ldots, n$ and age class $x = 2, \ldots, m$, where $\mathbb{1}[\cdot]$ corresponds to the indicator function, whereas $\Gamma_x^{c(-i)} = \Gamma_x^c \setminus \{i\}$ and $\Gamma_x^{c(+i)} = \Gamma_x^{c(-i)} \cup \{i\}$. Notice that within the above expression the ratio $p(\mathbb{c}_x^{\Gamma_x^{c(+i)}} \mid \eta_c, \mathbf{z})/p(\mathbb{c}_x^{\Gamma_x^{c(-i)}} \mid \eta_c, \mathbf{z})$ relates directly to the conditional distribution for c_{ix} given $\boldsymbol{c}_x^{(-i)}$, η_c and $\mathbf{z}$. As such, recalling also Page et al. (2022), it can be computed from (2.3) after replacing the generic quantities $n_h^{(-i)}$, $\bar{n}_{hw_i}^{(-i)}$, $\bar{n}_{h\bullet}^{(-i)}$ and η with those specific to the reduced partition among the underlying causes at age class x.

The results in (2.3) are also useful to sample the allocations c_{ix} at each age class $x = 1, \ldots, X$ for those underlying causes whose sampled γ_{ix}^c is equal to 0; if $\gamma_{ix}^c = 1$, then c_{ix} is kept fixed to the group allocation drawn for cause i at $x-1$. More specifically, denote with $\mathbb{c}_x^{c_{ix}=h}$ the partition induced by the group membership vector $(c_{1x}, \ldots, c_{ix} = h, \ldots c_{nx})$ with the i–th underlying cause allocated to cluster h. Then, direct application of the Bayes rule in combination with results from Page et al. (2022) yields a full-conditional categorical distribution for each c_{ix} with probabilities

$$\begin{aligned} &\mathrm{pr}(c_{ix} = h \mid -) \propto \\ &\qquad \mathrm{pr}(c_{ix} = h \mid \boldsymbol{c}_x^{(-i)}, \eta_c, \mathbf{z}) \mathbb{1}[\mathbb{c}_{x+1} \in \mathbb{C}(\boldsymbol{c}_x^{(c_{ix}=h)}, \boldsymbol{\gamma}_{x+1}^c)] \frac{p(\mathbf{Y}_x \mid c_{ix} = h, \boldsymbol{c}_x^{(-i)}, \boldsymbol{d}_x)}{p(\mathbf{Y}_x^{(-i\cdot)} \mid \boldsymbol{c}_x^{(-i)}, \boldsymbol{d}_x)}, \end{aligned} \tag{2.6}$$

for every $h = 1, \ldots, H_x^{(-i)} + 1$, $i = 1, \ldots, n$ and $x = 1, \ldots, X$, where $\mathbf{Y}_x^{(-i\cdot)}$ corresponds to the adjacency matrix at age class x without the i–th row. In the above expression, the prior probabilities $\mathrm{pr}(c_{ix} = h \mid \boldsymbol{c}_x^{(-i)}, \eta_c, \mathbf{z})$ are available directly from (2.3) after replacing the involved generic quantities with those specific to the partition of the underlying causes at age x, whereas the constraint included through the indicator function guarantees the compatibility discussed in Section 2.2.2 for contiguous partitions (notice that compatibility with the partition at age class $x-1$ holds by prior construction, and hence, it does not need to be checked). Finally, generalizing results on Beta-Binomial distributions (e.g., Schmidt and Morup, 2013; Legramanti et al., 2022) to the Dirichlet-categorical one in (2.1), also the likelihood factor within (2.6) admits the closed-form expression

$$\frac{p(\mathbf{Y}_x \mid c_{ix} = h, \boldsymbol{c}_x^{(-i)}, \boldsymbol{d}_x)}{p(\mathbf{Y}_x^{(-i\cdot)} \mid \boldsymbol{c}_x^{(-i)}, \boldsymbol{d}_x)} = \prod_{k=1}^{K_x} \frac{\Gamma(a_\bullet^\theta + n_{hkx\bullet}^{(-i)})}{\Gamma(a_\bullet^\theta + n_{hkx\bullet})} \prod_{w=1}^{4} \frac{\Gamma(a_w^\theta + n_{hkxw})}{\Gamma(a_w^\theta + n_{hkxw}^{(-i)})}, \tag{2.7}$$

for every $h = 1, \ldots, H_x^{(-i)} + 1$, $i = 1, \ldots, n$ and $x = 1, \ldots, X$.

Given the persistence variables γ_{ix}^c, $i = 1, \ldots, n$, $x = 2, \ldots, X$, also the prior hyperparameters α_x^c, $x = 2, \ldots, X$ admit closed-form full-conditional distributions. More specifically, combining the Bernoulli likelihood for each γ_{ix}^c with the Beta priors assumed in Section 2.2.2 for every α_x^c, yields

$$(\alpha_x^c \mid -) \sim \text{Beta}\left(a_\alpha + \textstyle\sum_{i=1}^n \gamma_{ix}^c, b_\alpha + n - \sum_{i=1}^n \gamma_{ix}^c\right), \tag{2.8}$$

independently for $x = 2, \ldots, X$. Similarly-tractable full-conditionals are also available for the CRP concentration parameter η_c via a direct application of the results in Escobar and West (1995).

To conclude, it remains to update the column-specific quantities $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_X$, $\boldsymbol{\gamma}_2^d, \ldots, \boldsymbol{\gamma}_X^d$, $\boldsymbol{\alpha}^d$ and η_c. Due to the model symmetry, these updates share the same structure of those derived for the row-specific counterparts. As a consequence, it suffices to apply again (2.5)–(2.8) by replacing the quantities and indexes related to the underlying causes with those of the contributing ones, and vice-versa.

2.3.2 Adaptive Parallel Tempering

Albeit tractable, the Gibbs routine presented in Section 2.3.1 requires exploration of a high-dimensional discrete space involving two sequences of dependent random partitions. To mitigate the possible mixing issues that may arise in this challenging computational setting, we combine the previously-derived Gibbs sampler with an adaptive parallel tempering implementation.

Parallel tempering (PT) is a general methodology to improve the mixing of MCMC algorithms, especially in settings characterized by complex high-dimensional posteriors that may exhibit multiple local modes (e.g., Earl and Deem, 2005; Syed et al., 2022). In its general form, PT runs a sequence of independent MCMC samplers at different *temperatures*, where higher-temperature chains allow for an improved exploration of the state space by flattening the posterior distribution, and hence, also its local modes. In contrast, lower-temperature chains sample from distributions progressively more similar to the actual posterior, thereby targeting the original distribution of interest. At each iteration, all chains are independently updated performing a local move, which, in our case, uses the Gibbs sampling steps outlined in Section 2.3.1. Then, at periodic intervals, the algorithm attempts to swap states between chains via a Metropolis-Hastings acceptance criterion that ensures detailed-balance is maintained. This swap allows lower-temperature chains to benefit from the broader exploration of higher-temperature ones.

Within our implementation we consider, in particular, non-reversible deterministic swap proposals (Syed et al., 2021) to guarantee a more rapid information exchange among high- and low-temperature chains. Moreover, we implement the online stochastic optimization method proposed by Miasojedow et al. (2013) to dynamically adapt the temperature grid during the PT runs. This strategy builds a decreasing sequence of inverse temperatures and, at each swap, it uses the associated acceptance probabilities to tune the inverse-temperature grid via a Robbins-Monroe update targeting the desired acceptance level. Such a scheme robustifies PT against suboptimal specifications of the initial temperature schedule, thus allowing for a more efficient exchange of information from high- to low-temperature chains, which is fundamental to obtain an effective exploration of the posterior distribution.

2.3.3 Monte Carlo Inference

Posterior inference on the quantities of interest is performed via Monte Carlo, relying on the samples produced by the routine outlined in Sections 2.3.1–2.3.2. As discussed in Sections 2.1–2.2, our primary focus lies on the group membership structures encoded in the sequences $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_X$ and $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_X$, which unveil the block patterns displayed by the observed causes-of-death networks across age classes. Extending general guidelines in Bayesian model-based clustering (e.g., Wade and Ghahramani, 2018) to our specific context, a natural and interpretable option for summarizing the posterior distribution over these group structures is to compute the so-called posterior *similarity* (or *co-clustering*) matrices $\hat{\mathbf{P}}^c_1, \ldots, \hat{\mathbf{P}}^c_X$ and $\hat{\mathbf{P}}^d_1, \ldots, \hat{\mathbf{P}}^d_X$, whose generic entries $\hat{\mathbf{P}}^c_{x[i,i']}$ and $\hat{\mathbf{P}}^d_{x[j,j']}$ estimate $\text{pr}(c_{ix} = c_{i'x} \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_X)$ and $\text{pr}(d_{jx} = d_{j'x} \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_X)$ via the relative frequency of Gibbs samples in which $c_{ix} = c_{i'x}$ and $d_{jx} = d_{j'x}$, respectively. This yields an interpretable probabilistic summary of the co-clustering relationships supported by the posterior across age classes, along with a quantification of uncertainty in the inferred group structures.

Leveraging the above *similarity* matrices in combination with the decision-theoretic framework of Wade and Ghahramani (2018) it is also possible to obtain the posterior point estimates $\hat{\mathbf{z}}_{(1)x}$ and $\hat{\mathbf{z}}_{(2)x}$ for $\mathbf{z}_{(1)x}$ and $\mathbf{z}_{(2)x}$, respectively, at each age class $x = 1, \ldots, m$, via

$$\hat{\boldsymbol{c}}_x = \underset{\boldsymbol{c}'_x}{\text{argmin}}\, \mathbb{E}_{\boldsymbol{c}_x}[\,\text{VI}(\boldsymbol{c}_x, \boldsymbol{c}'_x) \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_X],$$

$$\hat{\boldsymbol{d}}_x = \underset{\boldsymbol{d}'_x}{\text{argmin}}\, \mathbb{E}_{\boldsymbol{d}_x}[\,\text{VI}(\boldsymbol{d}_x, \boldsymbol{d}'_x) \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_X],$$

where $\text{VI}(\cdot,\cdot)$ is the variation of information distance introduced by Meilă (2007) as a metric to measure differences between two generic partitions via a comparison among individual and joint entropies. The solutions to the above minimization problems can be obtained under the `R` library `mcclust.ext` (Wade and Ghahramani, 2018), which requires, as inputs, the matrices $\hat{\mathbf{P}}^c_x$ and $\hat{\mathbf{P}}^d_x$, respectively.

Besides studying the age-specific group structures among underlying and contributing causes encoded within $\hat{\boldsymbol{c}}_x$ and $\hat{\boldsymbol{d}}_x$, respectively, in our specific context it is also of interest to assess the stability of these structures across age classes. To this end, in Section 2.5 we will further study the *meet* of the estimated partitions across selected contiguous ages. Recalling, e.g., Wade and Ghahramani (2018), such a meet corresponds to a finer partition where two causes belong to the same *meet cluster* if these causes co-cluster in all the considered partitions, thereby highlighting sets of underlying and contributing causes that display stable group behaviors across the selected age classes.

To conclude, notice that while the block-specific vectors $\boldsymbol{\theta}_{hkx}$ are integrated out to obtain the likelihood in (2.1), when such quantities are also of interest a plug-in estimate can be readily derived. In particular, denote with $\hat{n}_{hkxw}$ the total number of pairs (i, j) such that $\hat{c}_{ix} = h$, $\hat{d}_{jx} = k$ and $y_{ijx} = w$. Then, leveraging Dirichlet-categorical conjugacy, a sensible point estimate for each θ_{hkxw} is

$$\hat{\theta}_{hkxw} = \mathbb{E}(\theta_{hkxw} \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_X, \hat{\boldsymbol{c}}_x, \hat{\boldsymbol{d}}_x) = \frac{(a^{\theta}_{w} + \hat{n}_{hkxw})}{(a^{\theta}_{\bullet} + \hat{n}_{hkx\bullet})}, \tag{2.9}$$

for every $h = 1, \ldots, \hat{H}_x$, $k = 1, \ldots, \hat{K}_x$, $x = 1, \ldots, X$ and $w = 1, \ldots, 4$.

2.4 Simulation Study

In this section, we present extensive simulation studies that illustrate the performance of the model proposed in Section 2.2 and quantify its gains with the respect to state-of-the-art alternatives that could be possibly employed within our motivating application. As a benchmark competitor, we consider, in particular, the dynamic SBM proposed by Matias and Miele (2017) and implemented in the `R` library `dynsbm`. This formulation has emerged in recent years as one of the most widely adopted implementations for inference on group structures among nodes varying across a temporal index (Peixoto, 2018; Kim et al., 2018; Lee and Wilkinson, 2019), making it a natural candidate for benchmarking in our simulation study. Note that, unlike for our proposed model, the one designed by Matias and Miele (2017) does not allow for two separate partitions on the rows and columns, respectively, and further assumes that the total number of non-empty groups is constant across the sequence of networks. As such, the comparison against this competitor is also useful to assess whether the increased flexibility provided by the model we propose yields practical gains in settings aligned with our motivating application.

To provide a comprehensive assessment, the above models are tested in two different simulation scenarios. The first resembles the structures discussed in Section 2.1 for the motivating causes-of-death networks, where the data are generated from a sequence of categorically-weighted directed networks regulated by two distinct partitions on the rows and columns of the adjacency matrices. As such, we simulate edges $(y_{ijx} \mid c^{(0)}_{ix} =$

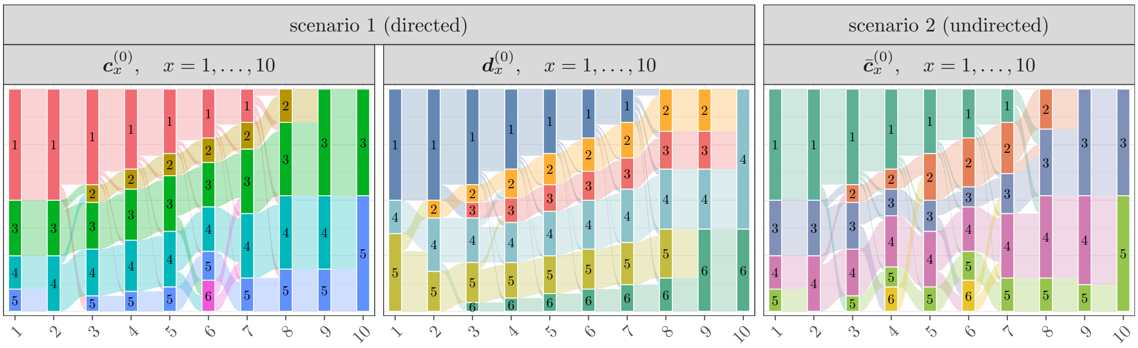


Figure 2.2: Riverplots representing the evolution of $\boldsymbol{c}_x^{(0)}$ and $\boldsymbol{d}_x^{(0)}$, $x = 1, \ldots, 10$ in the first scenario, and $\bar{\boldsymbol{c}}_x^{(0)}$, $x = 1, \ldots, 10$ in the second scenario.

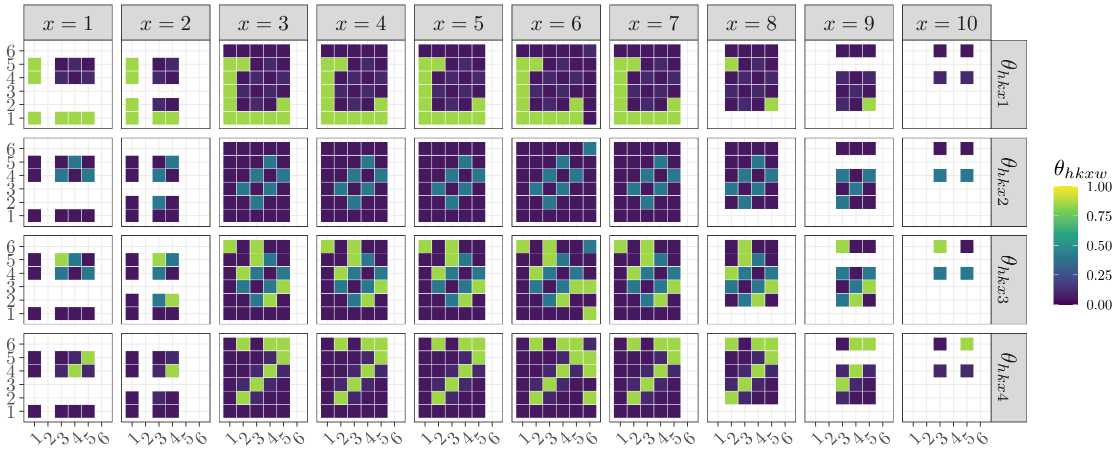


Figure 2.3: For the first simulation scenario (directed), graphical representation of the block-specific parameters, $\theta_{hkxc}^{(0)}$, evolving over age classes $x = 1, \ldots, 10$ and stratified by category $w = 1, \ldots, 4$. The numbered rows and columns correspond to row- and column-specific groups. White cells correspond to pairs of empty groups at the specific ages.

$h, d_{jx}^{(0)} = k, \boldsymbol{\theta}_{hkx}^{(0)}) \sim \text{Cat}_{1:4}(\boldsymbol{\theta}_{hkx}^{(0)})$, independently for every $i = 1, \ldots, 100$, $j = 1, \ldots, 100$ and $x = 1, \ldots, 10$, where each $\boldsymbol{\theta}_{hkx}^{(0)}$ is selected from four possible configurations — namely, $[0.85, 0.05, 0.05, 0.05]$, $[0.05, 0.05, 0.85, 0.05]$, $[0.05, 0.05, 0.05, 0.85]$ and $[0.1, 0.4, 0.4, 0.1]$ — (see Figure 2.3), whereas $\boldsymbol{c}_x^{(0)}$ and $\boldsymbol{d}_x^{(0)}$ are obtained by manually fixing an evolution across ages for the total of non-empty clusters and the number of causes allowed to change group among consecutive ages. Conversely, the selection of which specific causes are considered for these transitions and the updated group allocations for such causes are performed randomly (see Figure 2.2 for a graphical representation of $\boldsymbol{c}_x^{(0)}$ and $\boldsymbol{d}_x^{(0)}$, $x = 1, \ldots, 10$). In this way, the partitions to be inferred are not simulated from the assumed prior, thereby providing a more realistic assessment of the model proposed. For this same reason, we further test our construction in a second simulation scenario consisting of sequences of undirected networks with block structures regulated by a single partition $\boldsymbol{c}_x^{(0)}$, for $x = 1, \ldots, 10$, that is common to both rows and columns. Recalling our previous discussion, this setting is favorable to the model of Matias and Miele (2017) and can be simulated similarly to the directed case by forcing $\boldsymbol{c}_x^{(0)} = \boldsymbol{d}_x^{(0)} = \bar{\boldsymbol{c}}_x^{(0)}$, $\boldsymbol{\theta}_{hkx}^{(0)} = \boldsymbol{\theta}_{khx}^{(0)}$, and $y_{ijx} = y_{jix}$ for each $x = 1, \dots, 10$ (see Figure 2.2 for a graphical illustration of $\bar{\boldsymbol{c}}_x^{(0)}$, $x = 1, \dots, 10$). Notice that, in both scenarios, we do not consider supervision by external covariates $\mathbf{z}$ to avoid an overly-penalized treatment of the model by Matias and Miele (2017), which is not designed to include this additional source of information. When $\mathbf{z}$ is not available, it suffices to remove $(\bar{n}_{hz_i}^{(-i)} + a_{z_i}^{\mathbf{z}})/(\bar{n}_{h\bullet}^{(-i)} + a_{\bullet}^{\mathbf{z}})$ and $a_{z_i}^{\mathbf{z}}/a_{\bullet}^{\mathbf{z}}$ in (2.3) for adapting our formulation to the unsupervised case.

Given the above simulated data, we perform posterior inference under the Bayesian model proposed within Section 2.2 leveraging diffuse Dirichlet$(1,1,1,1)$, Beta$(1,1)$ and Gamma$(0.002, 0.001)$ priors on the block-specific parameters vectors, the transition probabilities and the CRP concentration parameters, respectively. Under such settings, posterior samples for $\boldsymbol{c}_1, \dots, \boldsymbol{c}_{10}$ and $\boldsymbol{d}_1, \dots, \boldsymbol{c}_{10}$ in the first simulation scenario can be obtained via the tempered Gibbs-sampler developed in Sections 2.3.1–2.3.2. Such a routine can be also adapted to the second scenario by sampling from the posterior of a single sequence $\bar{\boldsymbol{c}}_1, \dots, \bar{\boldsymbol{c}}_{10}$ (having the same prior as $\boldsymbol{c}_1, \dots, \boldsymbol{d}_{10}$ and $\boldsymbol{d}_1, \dots, \boldsymbol{d}_{10}$), while replacing the likelihood induced by directed networks with the one arising in undirected settings, where $y_{ijx} = y_{jix}$, and hence, $\boldsymbol{\theta}_{hkx} = \boldsymbol{\theta}_{khx}$. As a consequence, in the directed case, $p(\mathbf{Y}_1, \dots, \mathbf{Y}_X \mid \boldsymbol{c}_x,\ x = 1, \dots, X) = \prod_{x=1}^{X}(\prod_{h=1}^{H_x}\prod_{k=1}^{h}[\{\Gamma(a_{\bullet}^{\theta})/\Gamma(a_{\bullet}^{\theta} + n_{hkx\bullet})\}\prod_{w=1}^{4}\{\Gamma(a_w^{\theta} + n_{hkxw})/\Gamma(a_w^{\theta})\}])$.

Under the above routines, we consider 4,000 posterior samples for the sequences of partitions analyzed. Such samples are obtained after a conservative burn-in of 10,000 and thinning by 10 the subsequent 40,000 draws. The study of the traceplots for the logarithm of the likelihoods associated with the two simulation scenarios does not provide evidence against convergence, and showcase adequate mixing. As such, the produced samples are leveraged to obtain a posterior point estimate for $\boldsymbol{c}_1, \dots, \boldsymbol{c}_{10}$ and $\boldsymbol{d}_1, \dots, \boldsymbol{d}_{10}$ in the first scenario, and $\bar{\boldsymbol{c}}_1, \dots, \bar{\boldsymbol{c}}_{10}$ in the second, via the strategies in Section 2.3.3.

Figure 2.4 illustrates the quality of these estimated partitions in recovering the true ones behind the generative mechanism of the simulated data in scenario one (i.e., $\boldsymbol{c}_1^{(0)}, \dots, \boldsymbol{c}_{10}^{(0)}$ and $\boldsymbol{d}_1^{(0)}, \dots, \boldsymbol{d}_{10}^{(0)}$) and two (i.e., $\bar{\boldsymbol{c}}_1^{(0)}, \dots, \bar{\boldsymbol{c}}_{10}^{(0)}$). Such a quality is quantified in 50 replicated experiments under both scenarios via the most widely-implemented measures of clustering accuracy, namely, the rand index (RI) (Rand, 1971), the adjusted rand index (ARI) (Hubert and Arabie, 1985), and the normalized mutual information (NMI) (Strehl and Ghosh, 2002). These three measures take values below 1 (which corresponds to perfect overlap among the two partitions compared), and are computed also for the estimates produced by the model of Matias and Miele (2017) under its implementation in the `R` library `dynsbm`. As discussed previously, this competitor does not automatically infer the number of non-empty clusters. As such, we explore three specifications with this quantity set equal to 3, 6 and 9, where 3 corresponds to an underparameterized model, 6 is the true maximum number of clusters in the data, and 9 represents a flexible overparameterized formulation. Moreover, as suggested by the authors, we obtain the maximum likelihood estimate of the partitions via a multi-start routine from 25 different initial points.

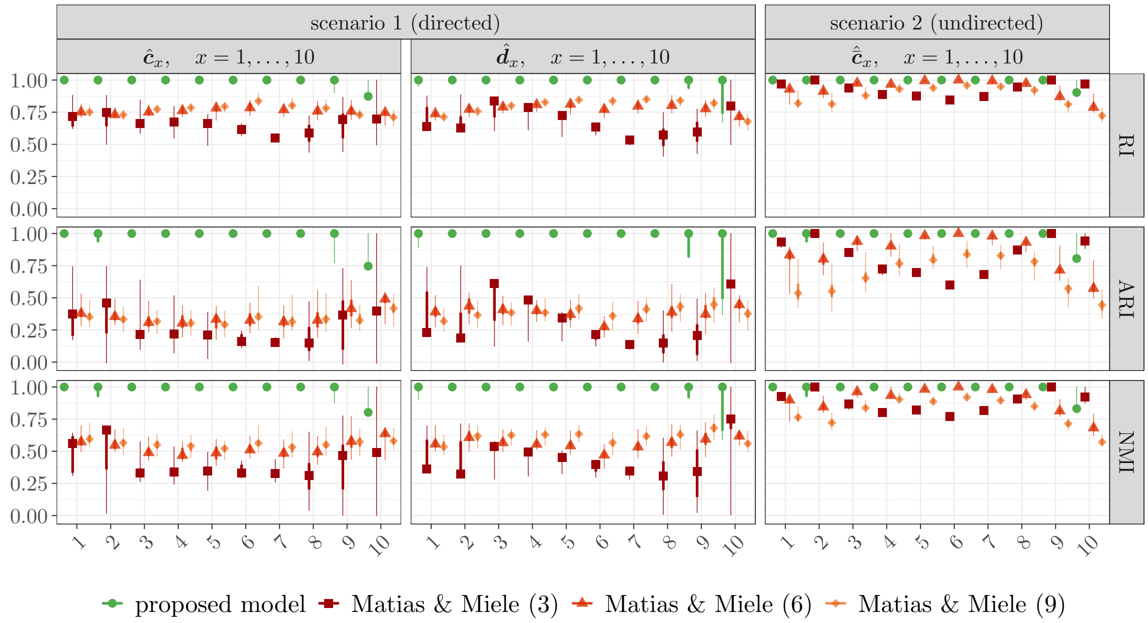


Figure 2.4: Accuracy of the proposed model and of the one designed by Matias and Miele (2017) (under three specifications for the total number of clusters) in estimating the true group membership vectors $\boldsymbol{c}_x^{(0)}$ and $\boldsymbol{d}_x^{(0)}$, $x = 1, \ldots, 10$ in the first scenario, and $\bar{\boldsymbol{c}}_x^{(0)}$, $x = 1, \ldots, 10$ in the second scenario. The performance is measured under the rand index (RI), adjusted rand index (ARI), and normalized mutual information (NMI). The points, thick lines, and thin lines represent, respectively, the medians, the inter-quartile ranges, and the ranges of the considered measures over 50 replicated experiments of the two simulation scenarios under analysis.

As shown in Figure 2.4, our proposed model almost perfectly estimates all the partitions across ages, in both scenarios, and systematically outperforms the clustering accuracy achieved by the state-of-the-art alternative in Matias and Miele (2017). The remarkable gains obtained in the first scenario clarify the importance of allowing for two separate partitions on the rows and columns in directed networks displaying asymmetric block patterns within the adjacency matrix. As discussed in Section 2.1.1 (see also Figure 2.1), this is the case of our causes-of-death networks, thereby highlighting the need for a model as the one we propose in Section 2.2 to avoid the substantial bias that would arise when adapting available solutions to our motivating application. Interestingly, our model outperforms Matias and Miele (2017) also in the second scenario, where the focus is on those undirected settings favorable to the competitor under analysis. This clarifies that, besides the increased flexibility provided by the inclusion of two separate partitions in directed settings, the evolution process induced by the prior in Section 2.2.2 on the sequence of partitions appears to be a superior construction in more general settings, including in undirected ones. Unlike for Matias and Miele (2017), the process we employ allows the number of non-empty groups to change with age and to be automatically inferred. These advantages motivate the use of the proposed model in the analysis of the causes-of-death networks introduced in Section 2.1. A detailed presentation of the important findings obtained from this analysis is provided in Section 2.5.

2.5 Application to USA Cause of Death Networks

We conclude by illustrating the potential of the proposed model in unveiling group structures and modular interactions among causes of death that were hidden to previous studies of the USA data presented in Section 2.1.1. In line with standard practice in demography, we apply the model separately to male and female populations. Posterior inference relies on the same hyperparameters and MCMC settings as in the simulation studies to assess the robustness of these default choices in general contexts.

2.5.1 Empirical Results and Findings

As a first important assessment, Figure 2.5 displays the age-specific adjacency matrices of the observed co-occurrence networks among underlying and contributing causes, with rows and columns re-ordered according to the groups estimated by our model. This representation clarifies that the inclusion of two separate partitions varying with age classes is essential to obtain an accurate characterization of the block structures in the observed networks for both males and females. In addition, it confirms that our model provides an highly-effective construction in achieving such an objective, thus motivating an in-depth analysis of the composition and evolution across ages of the estimated partitions.

Consistent with the above discussion, Figure 2.6 presents the minimum-VI point estimates of the partitions for the underlying and contributing causes across age classes in both the female and male populations. The slowly-varying number of groups across different life stages — ranging from 3 to 8 for females, and from 3 to 10 for males — further supports the need for a model that allows for smoothly-evolving partition structures across ages. Notice how the total number of non-empty groups increases progressively with age classes, particularly in the middle and late adulthood, highlighting the growing complexity of the causes-of-death landscape, which reflects the broadening range of health conditions contributing to mortality. Correspondingly, the average group size declines with age classes, thereby unveiling the increasing heterogeneity in the interactions between underlying and contributing causes (see also Figure 2.5). As is evident from Figure 2.6, such an heterogeneity results from smooth fragmentation of medium-to-large underlying and contributing groups driven by diverging interaction patterns between specific subsets of causes. These peculiar transition patterns are studied in detail in the following through the analysis of specific meet clusters, and crucially expand available findings on underlying (Bergeron-Boucher et al., 2020; Calazans and Permanyer, 2023) and multiple (Trias-Llimós and Permanyer, 2023) causes-of-death diversity by clarifying that systematic heterogeneity increments are found also in higher-level block-interactions among underlying and contributing causes over ages.

Among the aforementioned transition and fragmentation patterns, a relevant one, which has received limited attention in the literature, refers to evolution of group 1 displayed

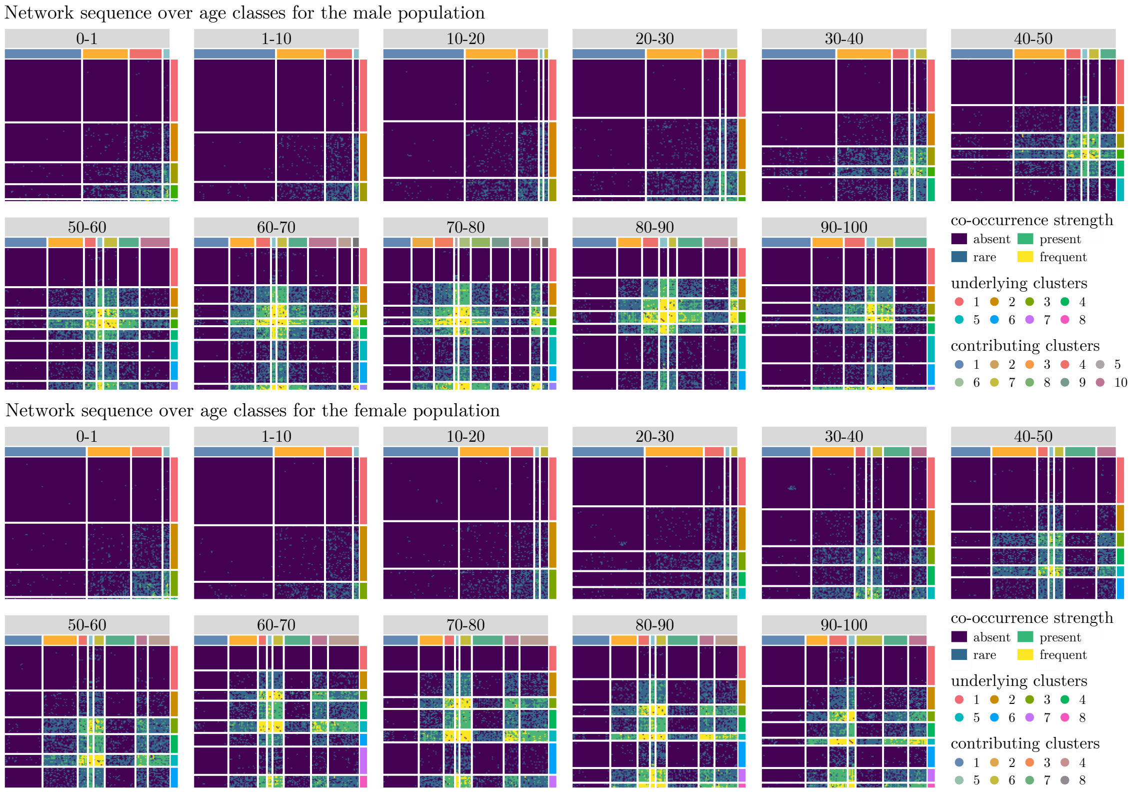


Figure 2.5: Observed co-occurrence networks among underlying and contributing causes of death at different age classes in the USA male (top) and female (bottom) populations. Rows and columns of the adjacency matrices are re-ordered according to the group structures estimated under the dependent stochastic block model proposed in Section 2.2.

in Figures 2.5–2.6 for both the underlying and contributing causes. As illustrated within Figure 2.5, such a group comprises non-interacting causes which are often removed from the analysis when the focus is on a single age class or on an aggregated network for a subset of ages. Albeit common, this practice neglects the fact that causes appearing as non-interacting at given age classes may display meaningful co-occurrences at other ages, thereby hindering the possibility to infer such peculiar transition mechanism from non-interacting to interacting blocks across age classes. In fact, addressing this objective is a key to understand at which ages specific causes join or exit inactive groups, a crucial information in the design of targeted policies. Consistent with this aim, we retain all the causes of death in our analysis, and let the proposed model infer non-interacting groups along with the associated evolution across ages.

As illustrated in Figures 2.5–2.6, the above information is present in the estimated group 1 across age classes, for both the underlying and contributing causes. At age 0–1, this group includes a large number of causes not typically associated with perinatal mortality, and then progressively fades at subsequent ages, further confirming the increasing heterogeneity of the causes-of-death landscape (which is apparent also in terms of additional underlying and contributing causes joining the active groups). Interestingly, Figure 2.6 highlights a slight increase in the size of group 1 at age classes 80–90 and 90–100. While this pattern deserves

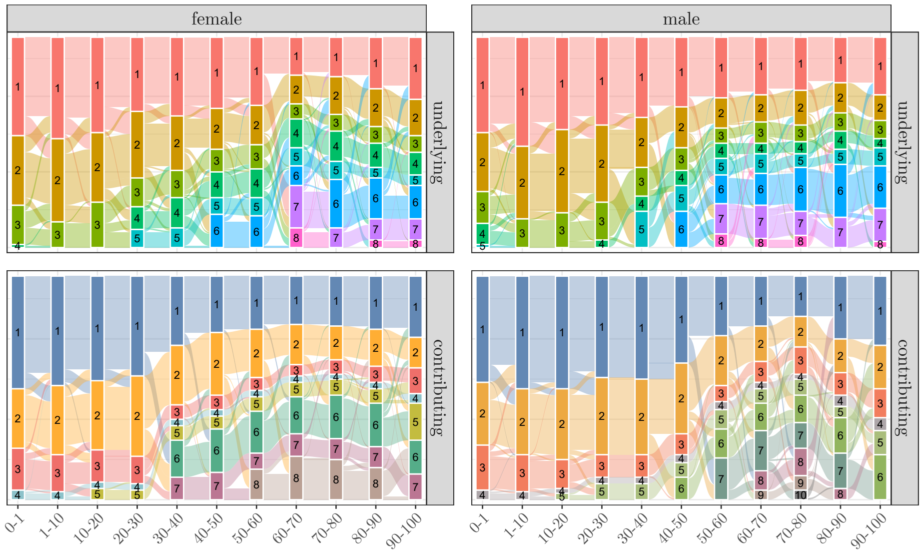


Figure 2.6: Riverplots representing the evolution of the estimated partitions $\hat{\mathbf{z}}_{(1)x}$ and $\hat{\mathbf{z}}_{(2)x}$ over age classes $x = 1, \ldots, 11$ in the female and male populations. The size of each bar is proportional to the cardinality of the inferred group.

further exploration, it may evidence selection effects in late life combined with susceptibility to fewer interactions among prevalent underlying and contributing causes.

Note that the change in size of group 1 is not only due to the emergence of causes exiting the inactive set at age 0–1. Rather, causes both appear and disappear with ages. For example, conditions associated with perinatal issues and malformations enter this group beyond the initial age class. Notably, while the most perinatal causes regarded as contributing factors are rapidly absorbed in group 1 by the second age class, many of these conditions (such as respiratory and cardiovascular disorders, and infections specific to the perinatal period) persist longer when regarded as underlying causes, joining group 1 only at later stages. This reflects the possibility that conditions emerging in the perinatal period may not turn out to be fatal at the same stage, but may persist into later life to become the cause that initiated the course of morbid events that led to death. In contrast, the same causes are unlikely to assume a contributing role in adulthood. Conversely, the causes exiting group 1 to join active clusters allow us to detect when a certain disease starts entering the mortality process and which already-active causes showcase its similar co-occurrence patterns. For example, in the female population, neoplasms of the respiratory organs, bones, and urinary tract exit group 1 in the transition between 0–1 years and 1–10 years, both for the underlying and contributing causes. Other forms of neoplasms, such as those of the breast and female genital organs, enter the mortality process when transitioning from 1–10 to 10–20 years for the underlying causes, while this transition only manifests from 10–20 to 20–30 years in the contributing causes. This suggests an asymmetry in the process, which may be due to the fact that early breast cancer is unlikely to be detected and thus is more likely to cause death as underlying factor rather than contributing one. Other gender-specific phenomena also emerge, particularly in older age groups for contributing causes. These differences, motivate further research on how contributing causes interact differently with underlying ones in males and females.

To further expand the results in Figures 2.5–2.6, we quantify in Figure 2.7 the evolution of the similarity among the estimated partitions not only across consecutive age classes, but also with respect to contemporary group structures of the underlying and contributing causes at the same age. These patterns are measured through $\text{NMI}(\hat{\boldsymbol{c}}_x, \hat{\boldsymbol{c}}_{x-1})$, $\text{NMI}(\hat{\boldsymbol{d}}_x, \hat{\boldsymbol{d}}_{x-1})$ and $\text{NMI}(\hat{\boldsymbol{c}}_x, \hat{\boldsymbol{d}}_x)$, respectively, for each x, and provide additional empirical support in favor of our model. In particular, the evolution of $\text{NMI}(\hat{\boldsymbol{c}}_x, \hat{\boldsymbol{d}}_x)$ evidences marked and stable differences among the partition structures of underlying and contributing causes. These asymmetries confirm the need for two separate partitions and highlight how causes display different group behaviors when acting as underlying or contributing. The importance of incorporating adaptive smoothness through age-specific persistence parameters $\alpha_{(1)x}$ and $\alpha_{(2)x}$ is instead confirmed by the evolution of $\text{NMI}(\hat{\boldsymbol{c}}_x, \hat{\boldsymbol{d}}_{x-1})$ and $\text{NMI}(\hat{\boldsymbol{d}}_x, \hat{\boldsymbol{d}}_{x-1})$ which showcase higher similarities among consecutive partitions of underlying and contributing causes, respectively, with an increasing trend that tends to stabilize at elder ages.

While Figures 2.5, 2.6 and 2.7 inform on the overall structure and transitions of the inferred groups, Figure 2.8 summarizes the relevance of such groups in terms of co-occurrence strengths and the associated composition with respect to three macro-categories of interest, namely, circulatory system conditions, mental health, and pregnancy-related causes. These macro-categories appear in the different panels of Figure 2.8, where each point corresponds to a group and its size is proportional to the ratio between the number of causes from the macro-category analyzed that have been allocated to that group and the total number of causes in such a macro-category. Hence, points of large size correspond to groups absorbing most of the causes of death from the macro-category analyzed. The vertical axis measures instead the overall intensity displayed by the underlying (contributing) causes of a given group in co-occurring with those of all the contributing (underlying) clusters. These marginal connectivity scores (MCS) are defined as the averaged probability of observing a `present` or `frequent`

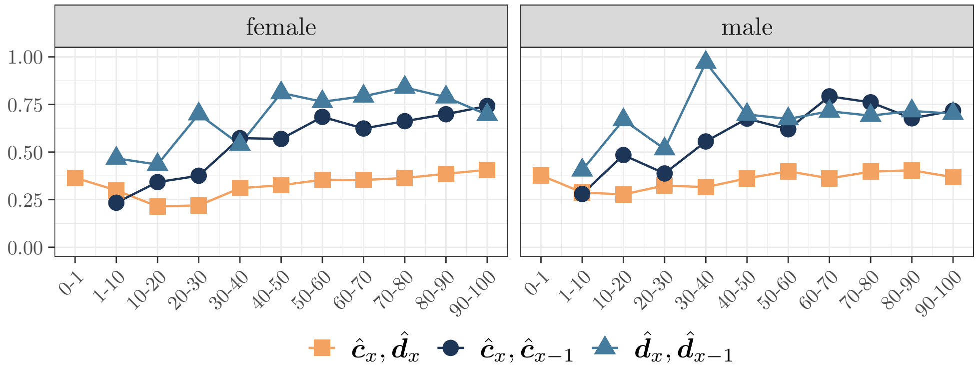


Figure 2.7: Normalized mutual information across ages between selected pairs (see the legend) of estimated partitions.

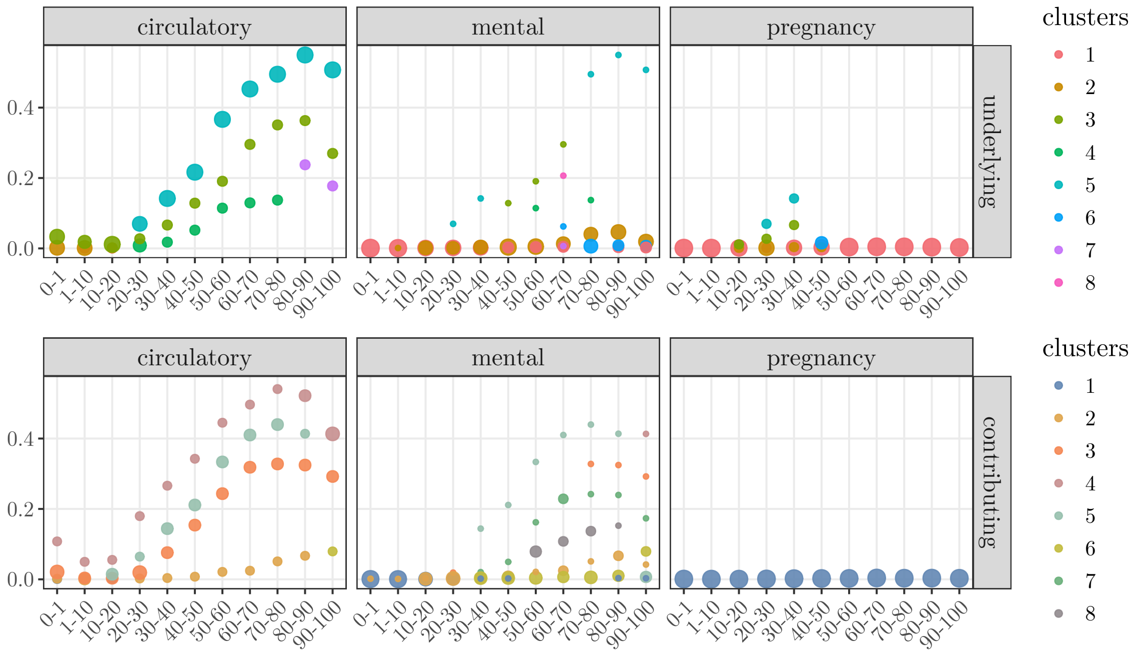


Figure 2.8: Group-specific marginal connectivity scores (MCS) over age classes for underlying (top) and contributing (bottom) causes in the female population, stratified by three macro-categories (circulatory, mental, and pregnancy). Point size is proportional to the number of causes in each macro-category belonging to a given group.

interaction between the causes of a given underlying (contributing) group h (k) and those belonging to a generic contributing (underlying) cluster $k = 1, \ldots, K_x$ $(h = 1, \ldots, H_x)$. More specifically, at each age class we compute $\textsc{mcs}_{(1)hx} = (1/\hat{K}_x)\sum_{k=1}^{\hat{K}_x}(\hat{\theta}_{hkx3} + \hat{\theta}_{hkx4})$, for $h = 1, \ldots, \hat{H}_x$, and $\textsc{mcs}_{(2)kx} = (1/\hat{H}_x)\sum_{h=1}^{\hat{H}_x}(\hat{\theta}_{hkx3} + \hat{\theta}_{hkx4})$, for $k = 1, \ldots, \hat{K}_x$, where $\hat{\theta}_{hkx3}$ and $\hat{\theta}_{hkx4}$ are obtained as in (2.9). The analysis of such measures in Figure 2.8, interestingly shows that pregnancy-related causes, while rare overall, appear almost exclusively as underlying causes among women of childbearing age, with minimal presence as contributing factors. Mental health-related causes also display distinct patterns across panels, reflecting an evolving role in the progression of disease. As age grows, these conditions become increasingly common as contributing causes. In contrast, among women over 70, high prevalence as underlying causes is limited to organic mental disorders, including symptomatic conditions such as Alzheimer's and dementia, concentrated in cluster 5. Circulatory system conditions are infrequent at younger ages but become more prominent in later stages for both roles. Still, underlying and contributing patterns diverge. For instance, five of the six circulatory causes (hypertensive, ischemic, pulmonary, cerebrovascular, and other heart diseases) co-cluster early (by ages 30–40) as underlying causes with high $\textsc{mcs}_{(1)hx}$. In contrast, contributing roles are more fragmented, with these conditions spreading across groups of varying $\textsc{mcs}_{(2)hx}$.

We conclude by exploring in more detail persistent groups and peculiar fragmentation patterns that arise in the female and male population at elder (60–70, 70–80, 80–90) and

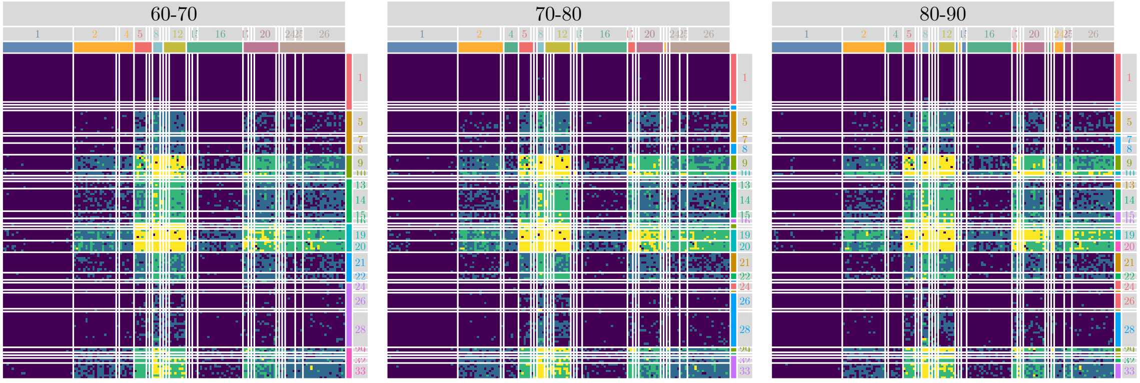


Figure 2.9: Adjacency matrices of the co-occurrence networks among underlying and contributing causes for USA female elder population. Rows and columns are re-ordered by estimated groups (colors) and *meet* clusters (lines-numbers).

adolescent/early-adulthood (10–20, 20–30, 30–40) ages, respectively. These analyses rely on the study of the meet clusters introduced in Section 2.3.3 (see Figures 2.9–2.10) and highlight yet-unexplored patterns that contribute to an improved understanding of the determinants underlying the recent growing mid-life mortality in the USA (e.g., Woolf and Schoomaker, 2019; Mehta et al., 2020; Case and Deaton, 2021). Focusing first on the female population, Figure 2.9 unveils a number of interesting meet clusters both in terms of composition and evolution. This is the case, for example, of the underlying meet clusters 9 (which includes respiratory and genitourinary conditions) and 10–11 (that comprise mental disorders and central nervous system neoplasms, respectively). While these three subsets of causes form a single group in the 60–70 class, at subsequent ages the meet cluster 9 eventually creates a separate group due to different co-occurrence strengths with the contributing causes in the meet cluster 20 (which consists of severe conditions such as lymphoid, breast and female genital neoplasms, and lung diseases caused by external agents). Similarly, the underlying meet clusters 19 and 20 (circulatory system causes and neoplasms, respectively) initially belong to the same group, due to similar interaction patterns with contributing causes. By age 80–90, however, the weakening of co-occurrences between the causes of the underlying meet cluster 20 and those of the contributing meet clusters 20 and 25 is in contrast with the strengthening of the interactions showcased by the underlying meet cluster 19, thereby yielding to a fragmentation in two separate groups. Notice that, although modest in size, the contributing meet cluster 25 (extrapyramidal and movement disorders, osteoarthritis and inflammatory polyarthropathies and other osteopathies), shows a very distinct pattern in the transition from the 70–80 to the 80–90 age class, as it separates from the adjacent meet clusters, primarily due to a strengthening of the co-occurrences with causes from the underlying meet cluster 19. Interestingly, such a shift might signal an onset of age-related mobility dysfunctions contributing to death by aggravating circulatory conditions.

We now turn our attention to the analysis of the meet clusters for the male population in those age classes that have witnessed peculiar mortality increments in USA over the

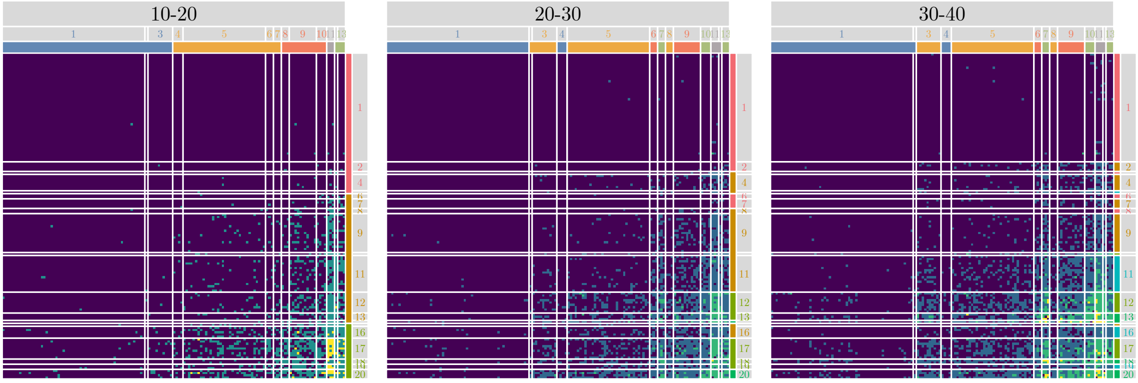


Figure 2.10: Adjacency matrices of the co-occurrence networks among underlying and contributing causes for USA male young population. Rows and columns are re-ordered by estimated groups (colors) and *meet* clusters (lines-numbers).

recent years. As clarified in Figure 2.10, although the determinants of such increments have generated a debate around different views (Woolf and Schoomaker, 2019; Mehta et al., 2020; Case and Deaton, 2021), a detailed analysis of the meet clusters and modules inferred by our model may lead to a consensus among such views.

Focusing first on the contributing causes, it is possible to recognize the meet cluster 11 as a highly-active one in the age class 10–20. This cluster mostly comprises various forms of heart and acute respiratory diseases that often arise from infectious phenomena common to young ages. In fact, at early-adulthood stages, the meet cluster 11 becomes progressively less active, whereas the contributing meet clusters 7 and 10 start displaying more intense interactions. Interestingly, these clusters include causes such as mental and behavioral disorders due to psychoactive substance use, diabetes mellitus, hypertensive diseases, diseases of liver, metabolic disorders, ischemic heart diseases, pulmonary heart disease and diseases of pulmonary circulation. This is an important finding which clarifies that, although current literature views cardiovascular diseases (Mehta et al., 2020) and “death of despair” type causes (Woolf and Schoomaker, 2019; Case and Deaton, 2021) as alternative explanations of the recent increments in USA mid–life mortality, in fact, such causes cluster together, and hence, should not be treated as alternatives to each other, but rather as closely related and jointly contributing to overall mortality patterns. Such a result is further reinforced by the analysis of the underlying meet clusters in Figure 2.10, which display high co-occurrences for the meet clusters from 16 to 20 in the age class 10–20. Similarly to the contributing meet cluster 11, these groups mainly encompass conditions related to respiratory diseases (e.g., flu, pneumonia, chronic lower respiratory diseases), and further contain nervous system disorder (e.g., episodic and paroxysmal disorders, cerebral palsy and other paralytic syndromes, other disorders of the nervous system) along with neoplasms (e.g., musculoskeletal, eye, brain, central nervous system and endocrine gland cancers). Conversely, in populations aged 20–30 and 30–40, a second group of meet clusters (12, 13, 14) emerges through a transition from a low-interactions group (orange) to the more active one (green). Interestingly, such a subset displays increasingly-strong co-occurrences with the contributing meet clusters 7 and 10, and is composed by a similar combination of cardiovascular and "death of despair" type causes. This result further strengthens the fact that the recent increments in USA mid–life mortality might arise from complex joint interactions among stochastically equivalent groups of underlying and contributing causes that appeared as alternatives to each other in previous analyses focused only on the underlying cause (see, e.g., Woolf and Schoomaker, 2019; Mehta et al., 2020; Case and Deaton, 2021), rather than on multiple interacting ones.

2.6 Conclusions and Future Research

Motivated by the importance of achieving an in-depth understanding of the core structures regulating cause-of-death networks across ages, and by the lack of a suitable statistical model capable of inferring such structures, we designed in Sections 2.2 and 2.3 an innovative stochastic block model for sequences of categorically-weighted directed networks. Consistent with empirical evidence from real-data applications, such a model accounts for modular interactions among groups of causes, and crucially allows for two separate partition structures on the rows and columns of the asymmetric adjacency matrices, corresponding to underlying and contributing causes of death, respectively. Under a Bayesian approach to inference, these partitions are assigned flexible priors that induce smooth transitions in the group structures across age classes, further informed by external macro-classifications of death causes.

The simulation studies in Section 2.4 confirm that the proposed model is superior to state-of-the-art alternatives not only in realistic settings aligned with the motivating cause-of-death application, but also in simpler undirected networks scenarios favorable to currently-available formulations. This motivates an extensive use of the proposed model in a wide variety of situations where multiple directed (or undirected) networks are observed across a temporal index. The application to USA causes-of-death networks in Section 2.5 further confirms this point by showcasing the potential of the proposed model in unveiling fundamental group structures among underlying and contributing causes that were hidden from classical MCOD and single-cause analyses. Besides the potential policy implications, the evolution and composition of such groups plays also an important role in the understanding of modern mortality trends, such as the recent increment of mid-life mortality in the USA.

Interesting directions for future research emerge from our contribution. A natural one is to include further dimensions within our model, such as calendar years, geographical units (e.g., USA states), and others. This would require the design of a joint prior for random partitions that induces dependence not only between consecutive age classes, but also across contiguous calendar years and spatially-close units. Representations of this type could be found in the ongoing literature on Bayesian nonparametric priors relying on notions of separate exchangeability (Rebaudo et al., 2025), conditional partial exchangeability (Franzolini et al., 2023) and combinations of these two notions. Any advancement along these lines could be inherited within our formulation and would also help in the design of a joint model for the male and female populations that borrows information among the associated partitions.

Besides the above methodological extensions, it shall be emphasized that the proposed model naturally applies also to (rectangular) bipartite networks encoding interactions (possibly varying with a temporal index) among two different sets of entities (e.g., users and items interactions across years in recommender systems' applications), thereby motivating further exploration of the potential of our construction also in these settings. Additional empirical analyses are also of interest within the motivating causes-of-death application to assess whether the inferred group structures are robust to changes in the way through which interactions are defined. For example, as anticipated in Section 2.1.1, one option could be to replace the discretized co-occurrence strengths we employ, with normalized versions of the pairwise co-occurrence counts accounting for the overall degree of appearance of the involved causes in death certificates (Hidalgo et al., 2009; Chmiel et al., 2014; Fotouhi et al., 2018). Preliminary analyses show that the group structures we infer align also with the block patterns induced by this measure.

Chapter 3

Bayesian Temporal Clustering of Age-Specific Leading Causes of Death Profiles across Multiple Countries

Joint work with R. Piccarreta and D. Durante

3.1 Introduction

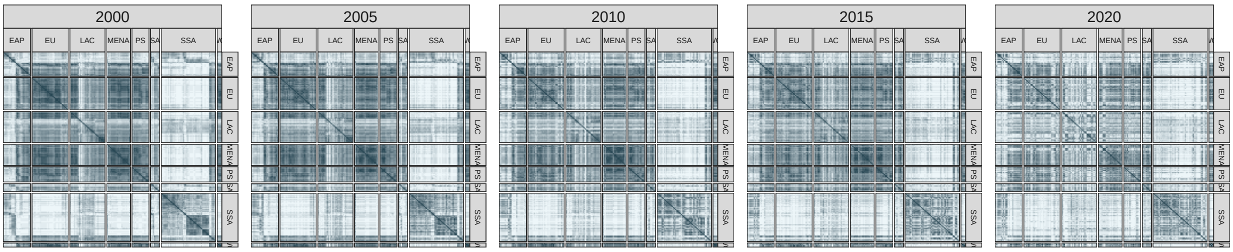
The comprehensive analysis of global mortality patterns represents a critical resource for health policymakers. Focusing exclusively on global lifespan-related quantities, such as life expectancy, or on population mortality risk measures, such as mortality rates, is insufficient to fully understand the mortality trends and identify critical points whose resolution could substantially improve health outcomes. A more complete understanding requires complementary investigation of causes of death patterns (Roser et al., 2021; Dattani et al., 2023), which constitute important determinants of longevity (Depaoli et al., 2024). It is acknowledged in the literature that differences in age- and cause-specific mortality rates significantly contribute to variations in life expectancy between populations. Notable examples include the sex gap observed in low-mortality countries (Meslé, 2004; Preston and Wang, 2006; Feraldi and Zarulli, 2022) and country-specific disparities (Canudas-Romo et al., 2020). As such, a thorough analysis of evolving causes-of-death landscapes offers crucial insights to guide targeted policy addressing country-, sex- and age-specific deficiencies in disease prevention and treatment. Consequently, age-specific leading causes of death serve as a fundamental and accessible indicator of national health status, proving instrumental for the rapid detection of emerging disease burdens and directing economic and medical resources toward mitigating mortality from such causes. It is also essential to this aim that analyses are conducted with sufficient granularity to properly capture the heterogeneity in mortality causes that naturally characterizes different population subsets defined by factors such as age, sex, and socio-economic characteristics. Previous studies of the leading cause of death have predominantly focused on descriptive analyses of trends within specific regions, often compressing or selecting subsets of the temporal and age dimensions, or examining single or limited groups of causes (Jemal et al., 2005; Saadat et al., 2015; Bray et al., 2021; Heron, 2021; Korenjak-Černe and Kejžar, 2022). While Mathers et al. (2009) provides a comprehensive overview stratified by age, geographic region, and sex, such a study still lacks a formal statistical model capable of providing inference across all these dimensions simultaneously.

The World Health Organization (WHO) maintains a comprehensive mortality database (World Health Organization, 2024a) encompassing detailed annual tabulations of the leading causes of death derived from the registration systems of member states, with causes classified according to the 11th revision of the International Classification of Diseases (ICD; World Health Organization, 2025). Specifically, for each combination of country (183 worldwide), sex (male, female), age class (0–1, 1–4, 16 groups of 5-year width up to 80–84, 85+), and calendar year (2000, 2005, 2010, 2015, 2020), we consider the leading cause of death, defined as the cause with the highest number of deaths in that specific subpopulation. Focusing temporarily on a specific year and sex, this yields 183 vectors of leading causes of death across 19 age classes. This focus on the leading cause of death is motivated by two main considerations. First, conducting a more granular cause-of-death analysis at the global level would require comprehensive and comparable data across all countries and time periods, which are not currently available. In practice, such detailed cause-specific data exist only for a limited set of countries with complete and reliable vital registration systems, restricting the scope for global comparisons. Second, focusing on the leading cause of death helps mitigate the widespread data quality issues affecting many developing countries, where incomplete or inconsistent death registration remains a challenge. For example, World Health Organization (2024b) reports that the WHO mortality database includes "high-quality" data for only 66 countries, a fact which makes a detailed global study infeasible.

Over the past century, the epidemiological debate has been dominated by the concept of epidemiological transition, which posits that countries follow a predictable trajectory of mortality patterns, transitioning from high mortality due to infectious diseases to other non-communicable conditions (Omran, 1971). However, this perspective has been increasingly challenged, as many countries have not followed the expected transitions, with leading causes of death demonstrating more complex and heterogeneous evolutionary trajectories (Caselli et al., 2002; Santosa et al., 2014; Mercer, 2018). For instance, in some countries, the leading causes of death have remained stable in certain age classes, while in others new causes have emerged or the existing ones have declined at different rates across age classes and calendar years.

Figure 3.1 illustrates this dynamic landscape through the evolution of the pairwise Hamming similarities (Hamming, 1950) between profiles of leading causes of death for years 2000, 2005, 2010, 2015, 2020, computed as the proportion of age classes for which the leading cause

Hamming similarity matrices for male populations



Hamming similarity matrices for female populations

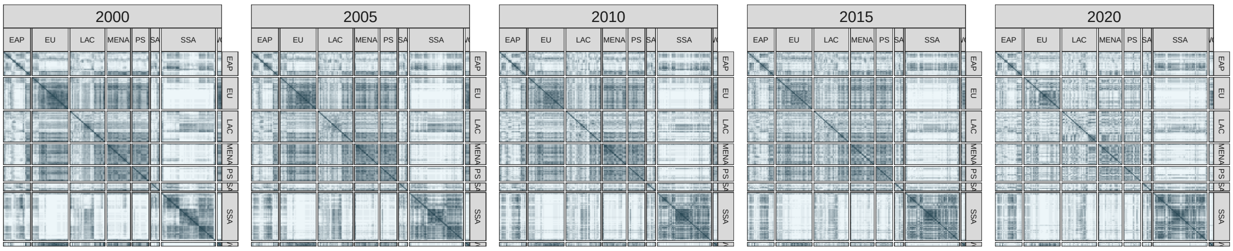


Figure 3.1: Pairwise (normalized) Hamming similarity between leading-cause profiles across ages for years 2000, 2005, 2010, 2015, 2020. The color scale ranges from pale cyan (minimum similarity) to blue (maximum similarity). Blocks refer to the following geographic regions (from left to right and top to bottom): East Asia & Pacific (EAP), Europe (non Post-Soviet) (EU), Latin America and Caribbean (LAC), Middle East & North Africa (MENA), Sub-Saharan Africa (SSA), Post-Soviet (P-S), South Asia (SA), Western Offshoots (WO).

of death is the same across two populations. Each matrix is further divided into blocks corresponding to combinations of 8 geo-cultural regions — Sub-Saharan Africa (SSA), Middle East & North Africa and (MENA), Latin America & Caribbean (LAC), Western Offshoots (WO), Europe (non Post-Soviet) (EU), Post-Soviet (P-S), South Asia (SA), East Asia & Pacific (EAP) — which can serve as proxies for the socio-economic characteristics of the populations (World Bank, 2025; Inglehart, 2006) and are displayed in Figure 3.2. This exploratory analysis already reveals some interesting patterns. First, it is evident that the similarity structures are not constant over the 20-year period, but smoothly evolve for both sexes. As such, this underscores the need for a dynamic clustering framework that takes into account in a principled manner the varying degree of smoothness of such a process. Additionally, while geographical features of countries partially account for similarities in leading causes of death, Figure 3.1 clearly shows that this classification does not fully capture the phenomenon under analysis. For example, the 2000 male panel shows that East Asia & Pacific countries have a block-diagonal pattern, indicating similarity only among certain countries in that region. A similar structure is present in Sub-Saharan female populations in 2000. Conversely, some between-region similarities stand out, such as the one in 2000 between European (non P-S) and Middle Eastern & North African male populations. Lastly, female population similarities tend to be weaker than male population ones, a fact that supports a separate analysis of the two groups, toward a deeper understanding of this trend.

These perspectives and the available data motivate a thorough multi-year analysis of

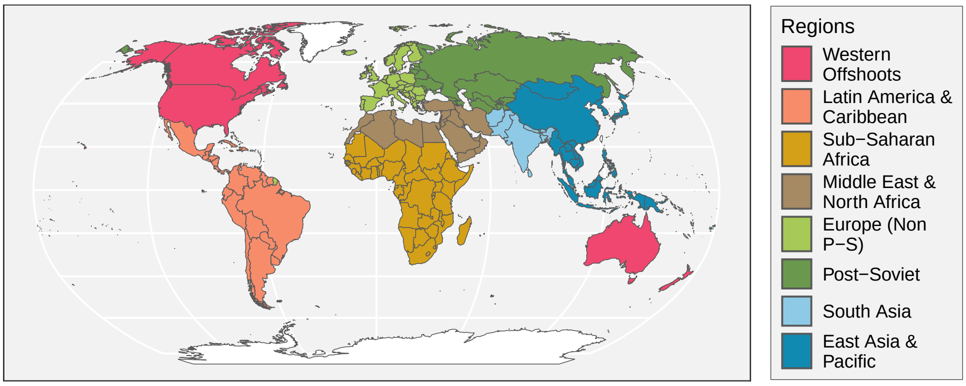


Figure 3.2: Geo-cultural regions of the world (World Bank, 2025; Inglehart, 2006).

the leading cause patterns across multiple sex-specific populations to investigate local similarities and differences over time. Such analysis can enable the identification of previously unexplored demographic and epidemiological trends, and enhance the understanding of mortality transitions potentially arising from targeted public health policies adopted by particular nations. Addressing the dual objectives of accurate mortality characterization and dynamic clustering discovery requires a principled formulation that accounts not only for the core structures of the country- and sex-specific leading-cause profiles over age classes, but also for the temporal clustering patterns among populations arising from overlaps in these vectors across particular calendar years. Recognizing the necessity for a principled statistical procedure, we propose a novel Bayesian formulation to model the evolution of age-period leading causes of death that accounts for the categorical nature of the leading causes across age classes while incorporating the clustering patterns among countries that are permitted to evolve dynamically over time. This is accomplished by modeling the single-year leading-cause profiles via a flexible Bayesian framework that relies on a mixture of exponential-distance components based on the Hamming distance for categorical vectors. The partitioning structure implicitly induced by this mixture formulation is then allowed to change over time, thereby enabling dynamic clustering of countries based on their leading cause profile. Model-based clustering of categorical vectors has been previously explored in the literature, mainly in the form of latent class models (Goodman, 1974), which consist of finite mixtures of multivariate multinomial distributions. While highly flexible, such models are known to have two major limitations: (i) the requirement of imposing constraints on an otherwise high-dimensional parameter space, and (ii) the obligation to specify and fix the number of clusters a priori, which is unsuitable for applications where the number of clusters is unknown and likely to vary over time. Our proposed model addresses these limitations by adapting to our setting the recently proposed exponential-distance model from Argiento et al. (2024), which extends the classical latent class model to a mixture with a random number of components and enables a natural interpretation of estimated clusters through a reparameterization of the multinomial parameters in terms of cluster-specific central profiles and scale parameters, closely related to the approach previously introduced in Celeux and Govaert (1991). To our knowledge, the literature on clustering of categorical vectors has not yet explored the possibility of dynamic clustering, which is a key feature of our model. In particular, we allow the clustering structure to evolve over time through a dependent random partition model (tRPM) (Page et al., 2022). The resulting model facilitates the identification of local clustering patterns among countries for specific years and structural changes in cause-specific mortality burdens that may concern only subsets of nations.

As clarified in Section 3.3, the proposed structured representation, while maintaining considerable flexibility, remains amenable to tractable posterior inference via a carefully designed collapsed Gibbs-sampling algorithm that automatically learns the unknown year-specific number of clusters and facilitates both point estimation and uncertainty quantification on the grouping structures and the resulting mortality patterns. As illustrated in Sections 3.4–3.5, the ability to identify locally converging or diverging trends in multi-country age-specific leading cause patterns, while preserving structured representations that account for the fundamental characteristics of these sequences, represents a significant advancement in epidemiological and demographic modeling. Such capabilities not only provide more realistic characterizations of mortality transitions across diverse populations but also open avenues for scientifically addressing important questions regarding epidemiological convergence, divergence, and the corresponding socio-economic and policy implications. Finally, in Section 3.6 we discuss some potential future directions of research and we underline that, although our methodological development is motivated by the analysis of age-specific leading causes of death, the proposed Bayesian temporal clustering framework is broadly applicable to any context where the goal is to cluster trajectories of categorical vectors observed across multiple units and time points.

3.2 Model Formulation

Let us denote with $y_{ixt} \in \mathcal{M}_x = \{1, \dots, m_x\}$ the leading cause of death within population $i \in \{1, \dots, n\}$, age class $x \in \mathcal{X} = \{1, \dots, X\}$, and period $t \in \{1, \dots, T\}$ (which in our case correspond to calendar years). Since the set of observed causes of death is not constant across the age classes, it is critical that we let the number of possible values assumed by y_{ixt}, namely m_x, depend on the age class x. In particular, these values range from 8 to 38 and from 9 to 37 respectively in male and female populations, with high values for younger populations, and lower values for the perinatal one and the elders, as shown in Table 3.1. This is not surprising since the leading causes of death for children are more heterogeneous, while for the perinatal and elder populations they tend to be more concentrated around a few causes, such as congenital anomalies and birth trauma for the former and cardiovascular diseases and cancers for the latter. Table 3.A.1 reports the complete list of causes of death, their corresponding abbreviations used in the figures, and the age ranges where each has

x	0–1	1–4	5–9	10–14	15–19	20–24	25–29	30–34	35-39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85+
m_x	8	20	38	31	21	19	16	18	19	16	18	17	17	18	14	14	12	12	11

(a) Male population.

x	0–1	1–4	5–9	10–14	15–19	20–24	25–29	30–34	35-39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85+
m_x	9	20	37	36	28	26	27	25	27	26	24	18	17	18	18	15	13	13	11

(b) Female population.

Table 3.1: Number of possible values for the leading causes of death y_{ixt} across age classes.

been registered as a leading cause.

Following the flexible construction in Argiento et al. (2024) for categorical vectors we posit that, for each year t, the vectors $\boldsymbol{y}_{it}$, $i = 1, \dots, n$, follow a mixture of K_t components, each corresponding to a cluster of homogeneous populations in year t. The mixture model is defined as follows:

$$\begin{aligned} &(\boldsymbol{y}_{it} \mid \boldsymbol{\theta}_{it}, \boldsymbol{\sigma}_{it}) \overset{\text{ind}}{\sim} \text{Hamming}(\boldsymbol{y}_{it} \mid \boldsymbol{\theta}_{it}, \boldsymbol{\sigma}_{it}) \\ &(\boldsymbol{\theta}_{it} \mid c_{it} = k, \boldsymbol{\theta}^\star_{kt}) = \boldsymbol{\theta}^\star_{kt} \qquad (\boldsymbol{\sigma}_{it} \mid c_{it} = k, \boldsymbol{\sigma}^\star_{kt}) = \boldsymbol{\sigma}^\star_{kt} \end{aligned} \qquad \text{for } i = 1, \dots, n$$

where $\boldsymbol{c}_t = (c_{1t}, \dots, c_{nt})$ is the vector of latent cluster assignment associated to a partition of the n countries $\mathbb{c}_t$, $\text{Hamming}(\cdot \mid \boldsymbol{\theta}, \boldsymbol{\sigma})$ is the Hamming distribution with center $\boldsymbol{\theta}$ and scale $\boldsymbol{\sigma}$ introduced in Argiento et al. (2024) and $\boldsymbol{\theta}^\star_{kt} = [\theta^\star_{k1t}, \dots, \theta^\star_{kXt}] \in \mathcal{M}_1 \times \cdots \times \mathcal{M}_X$ and $\boldsymbol{\sigma}^\star_{kt} = [\sigma^\star_{k1t}, \dots, \sigma^\star_{kXt}] \in (\mathbb{R}^+)^X$, $k = 1, \dots, K_t$, are the cluster-specific parameters for year t.

The component-specific $\text{Hamming}(\boldsymbol{y}_{it} \mid \boldsymbol{\theta}_{it}, \boldsymbol{\sigma}_{it})$ likelihood is defined as an exponential-distance distribution

$$\text{Hamming}(\boldsymbol{y}_{it} \mid \boldsymbol{\theta}_{it}, \boldsymbol{\sigma}_{it}) \propto \exp\left\{ -\sum_{x=1}^{X} \frac{\mathbb{1}[y_{ixt} \neq \theta_{ixt}]}{\sigma_{ixt}} \right\}, \tag{3.1}$$

where $\mathbb{1}[\cdot]$ denotes the indicator function. If common scale parameters $\sigma_{i1t} \equiv \cdots \equiv \sigma_{iXt} \equiv \sigma_{it}$ are used across all components of the vectors, then the log-density reduces to the rescaled Hamming distance $(\sigma_{it})^{-1} \sum_{x=1}^{X} \mathbb{1}[y_{ixt} \neq \theta_{ixt}]$ between the observed profile $\boldsymbol{y}_{it}$ and the cluster-specific central one $\boldsymbol{\theta}_{it} = \boldsymbol{\theta}^\star_{c_{it}t}$. A fundamental advantage of this distance choice is the availability of a closed-form expression for the normalization constant, which is crucial for efficient posterior inference (Argiento et al., 2024).

In contrast with standard multinomial mixture models, this representation allows for a natural interpretation of each cluster as a set of populations whose profile distribution is concentrated around a central vector $\boldsymbol{\theta}^\star_{kt}$, with variability around such a mode controlled by the scale parameters $\boldsymbol{\sigma}^\star_{kt}$ that allow for different levels of dispersion $\sigma^\star_{kxt}$ across clusters for different components of the profiles, each corresponding to an age class $x = 1, \ldots, X$. As such, this formulation also implicitly enables a system of importance weights, which can be interpreted as the relative importance of each component y_{ixt} in driving the homogeneity of the cluster, since $\sigma^{\star}_{kxt}$ can be interpreted as the contribution of the x–th component to the rescaled Hamming distance between the observed $\boldsymbol{y}_{it}$ and the cluster center $\boldsymbol{\theta}^{\star}_{c_{it}t}$.

To capture the temporal evolution in the clustering structure shown in Figure 3.1, we allow the partition of the countries, and the associated cluster parameters $\boldsymbol{\theta}^{\star}_{kt}$ and $\boldsymbol{\sigma}^{\star}_{kt}$, $k = 1, \ldots, K_t$, to evolve over time according to the dependent random partition model (tRPM) of Page et al. (2022), which permits the clusters to change composition across years while encouraging temporal smoothness. This is achieved through a Markovian process $\mathrm{pr}(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_T) = \mathrm{pr}(\boldsymbol{c}_1) \prod_{t=2}^{T} \mathrm{pr}(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1})$, where only a subset of the populations can change cluster assignment from one year to the next one, as we do not expect abrupt variations in the clustering structure involving all the populations, but rather that a limited number of countries experience changes in their leading-cause profiles that bring them closer to another group of country-sex pairs (in case of converging trends) or further away (in case of diverging trends). This is accomplished by combining a flexible prior $\mathrm{pr}(\boldsymbol{c}_1)$ on the initial memberships $\boldsymbol{c}_1$ with an intuitive two-step transition procedure $\mathrm{pr}(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1})$. In particular, we assume that the mechanism for the formation of the initial partition $\boldsymbol{c}_1$ is driven by the Gnedin process (Gnedin, 2010; De Blasi et al., 2013) with parameter η. This distribution belongs to the broader class of Gibbs-type priors (De Blasi et al., 2015), from which it inherits analytical tractability, and exhibits desirable properties and theoretical guarantees crucial to approach a clustering task. In particular, this process can be interpreted as a Dirichlet-Multinomial model with a finite random number of total components K_1 following distribution $\mathrm{pr}(K_1 = k) = \frac{\eta(1-\eta)_k}{k!}$, where $(a)_b$ is the ascending factorial $a(a+1)\cdots(a+b-1)$. This probability mass function is characterized by the mode at 1 and heavy tails that induce infinite expectation, which implies an appropriate balance between parsimony of the clustering structure and robustness to the total number of components. This is crucial in the demographic application, where we expect to observe a converging pattern of disappearing infectious conditions in favor of non-communicable diseases, such as cardiovascular diseases and cancer, which are expected to be the leading causes of death in most countries. Moreover, the availability of a closed-form expression also for the distribution of the number of allocated components enables a moment matching procedure to elicit the hyperparameter η (see, e.g Legramanti et al., 2022). Furthermore, the above equivalence with mixture of finite mixtures guarantees the consistency for the estimated grouping structure (Argiento et al., 2024), which will be adopted in Section 3.4 and Section 3.5.

Let $\boldsymbol{c}_1^{(-i)}$ be the membership vector $\boldsymbol{c}_1$ after the removal of the i–th observation, and $K_1^{(-i)}$ the corresponding number of clusters. Finally, denote with $n_{k1}^{(-i)}$ the number of populations in the k–th cluster of $\boldsymbol{c}_1^{(-i)}$. Then, under the Gnedin assumption, the predictive distribution of allocation c_{i1} of population i in year 1, given the other $n-1$ allocations, coincides with

$\mathrm{pr}(c_{i1} \mid \boldsymbol{c}_1^{(-i)}) \propto n_{k1}^{(-i)}(n-1-K_1^{(-i)}+\eta)$ for clusters $k=1,\dots,K_1^{(-i)}$ already occupied by the other $n-1$ populations, and $\mathrm{pr}(c_{i1}=K_1^{(-i)}+1 \mid \boldsymbol{c}_1^{(-i)}) \propto K_1^{(-i)}(K_1^{(-i)}-\eta)$, for a new cluster $K_1^{(-i)}+1$. This urn scheme clarifies the role of the parameter η as *discount* factor in cluster formation and will be crucial in the definition of the transition distribution $\mathrm{pr}(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1})$ and in the construction of the sampling scheme in Section 3.3.

The prior on $\boldsymbol{c}_1$ is then combined with a tractable yet flexible transition mechanism $\mathrm{pr}(\boldsymbol{c}_t \mid \boldsymbol{c}_{t-1})$ for $t=2,\dots,T$, as introduced by Page et al. (2022) and detailed in Introduction. Specifically, through a system of latent binary indicators $\gamma_{it} \overset{\text{iid}}{\sim} \mathrm{Ber}(\alpha_t)$, $i=1,\dots,n$, at each time t a subset of populations is randomly selected to be eligible for cluster reassignment ($\gamma_{it}=0$), while the remaining are kept in the same cluster as in the previous period ($\gamma_{it}=1$). Then, conditionally on $\boldsymbol{\gamma}_t=[\gamma_{1t},\dots,\gamma_{nt}]$, only a subset of all possible partitions of n individuals is *compatible* with $\boldsymbol{c}_{t-1}$, in particular those that can be obtained from $\boldsymbol{c}_{t-1}$ by changing allocations only for the populations i for which $\gamma_{it}=0$, and we denote this set with $\mathbb{C}(\boldsymbol{c}_{t-1},\boldsymbol{\gamma}_t)$. To conclude, $\boldsymbol{c}_t$ is sampled from this set of compatible partitions with probability proportional to the one assigned by the Gnedin process assumed for $t=1$. While simple in its construction, this procedure is highly flexible, allowing for a data-driven estimate of the period-specific number of clusters K_t, thanks to the Gnedin process prior on $\boldsymbol{c}_1$, and a varying degree of temporal smoothness in the clustering structure, controlled by the parameter $\alpha_t \in (0,1)$, which can be interpreted as the expected proportion of populations that are allowed to change cluster assignment from $t-1$ to t.

The Bayesian formulation is concluded by specifying priors on the cluster-specific parameters $\boldsymbol{\theta}^\star_{kt}$, $\boldsymbol{\sigma}^\star_{kt}$ and the partition persistency parameter α_t, for $t=1,\dots,T$ and $k=1,\dots,K_t$. Following Argiento et al. (2024), we elicit independent uniform priors on $\{1,\dots,m_x\}$ for the components of the cluster centers $\theta^\star_{kxt}$ and a semi-conjugate Hypergeometric Inverse Gamma $\sigma^\star_{kxt} \sim \mathrm{HIG}(a^\sigma_x, b^\sigma_x)$ prior on the scale parameters $\sigma^\star_{kxt}$. The latter is defined as

$$p(\sigma^\star_{kxt}) \propto \left[1+(m_x-1)\exp\left(-\frac{1}{\sigma^\star_{kxt}}\right)\right]^{-(a^\sigma_x+b^\sigma_x)} \exp\left(-\frac{b^\sigma_x+1}{\sigma^\star_{kxt}}\right)\frac{1}{(\sigma^\star_{kxt})^2}, \tag{3.2}$$

for $\sigma^\star_{kxt}>0$, where $a^\sigma_x>0$ and $b^\sigma_x>0$ are hyperparameters that control the importance of each component of the vector in determining the cluster homogeneity (see Argiento et al., 2024, for a thorough analysis). Lastly, we assume independent conjugate Beta priors $\alpha_t \sim \mathrm{Beta}(a_\alpha, b_\alpha)$ on the transition parameters α_t, for $t=1,\dots,T$, with hyperparameters a_α and b_α controlling the prior mean $\frac{a_\alpha}{a_\alpha+b_\alpha}$ for the expected proportion of populations that are allowed to change cluster assignment in the transitions.

3.3 Bayesian Computation and Inference

Posterior inference on the Bayesian model parameters in Section 3.2 is performed through Monte Carlo procedures leveraging draws from a collapsed Gibbs sampler targeting the posterior distribution of the parameters given the observed leading causes of death $(y_{i1t}, \dots, y_{iXt})$, $i = 1, \dots, n$, observed in the different periods $t = 1, \dots, T$. In Section 3.3.1 we describe such a Gibbs sampler, while in Section 3.3.3 we discuss the Monte Carlo inference.

3.3.1 Collapsed Gibbs sampler

The proposed scheme alternates between two main steps, corresponding to the two main components of the model: (i) sampling the tRPM parameters, and in particular the dynamic partitions of the populations, and (ii) sampling the cluster-specific centers $\boldsymbol{\theta}^\star_{kt}$, for $k = 1, \dots, K_t$ and $t = 1, \dots, T$. In contrast, the scale parameters $\boldsymbol{\sigma}^\star_{kt}$ are marginalized out from the posterior distribution, by considering the marginal likelihood formulation detailed in Argiento et al. (2024, Proposition 3). First, we draw the cluster assignments and the related tRPM hyperparameters by adapting the scheme in Page et al. (2022, section B, Supplementary Materials) to our setting. Then, conditionally on the sampled partitions, we draw the cluster-specific centers from their full-conditionals, by adapting the procedure in Argiento et al. (2024).

As for the tRPM block of the model, we alternate sampling of the latent variables $\boldsymbol{\gamma}_t$ and $\boldsymbol{c}_t$, for every time period t. We start by sampling the latent variables γ_{it}, for $t = 2, \dots, T$, independently for $i = 1, \dots, n$, from their full-conditional distribution presented in Page et al. (2022)

$$\mathrm{pr}(\gamma_{it} = 1 \mid -) = \frac{\alpha_t}{\alpha_t + (1 - \alpha_t)\, p\Big(\mathbb{c}_t^{\Gamma_t^{(+i)}}\Big) / p\Big(\mathbb{c}_t^{\Gamma_t^{(-i)}}\Big)} \mathbb{1}\left[\mathbb{c}_{t-1}^{\Gamma_t^{(+i)}} = \mathbb{c}_t^{\Gamma_t^{(+i)}}\right],$$

where $\Gamma_t = \{i \colon \gamma_{it} = 1\}$, $\Gamma_t^{(-i)} = \Gamma_t \setminus \{i\}$, $\Gamma_t^{(+i)} = \Gamma_t^{(-i)} \cup \{i\}$. Moreover, $\mathbb{c}_t^{\Gamma_t^{(-i)}}$ and $\mathbb{c}_t^{\Gamma_t^{(+i)}}$ are the partition $\mathbb{c}_t$ restricted to populations in $\Gamma_t^{(-i)}$ and $\Gamma_t^{(+i)}$, respectively. The quantity $p\Big(\mathbb{c}_t^{\Gamma_t^{(+i)}}\Big) / p\Big(\mathbb{c}_t^{\Gamma_t^{(-i)}}\Big)$ at the denominator is the ratio of the Gnedin process probabilities of the two partitions and it is equivalent to the conditional probability $p(c_{it} \mid \mathbb{c}^{\Gamma_t^{(-i)}})$, which can be computed in closed form as discussed in Section 3.2.

Then, to satisfy the compatibility constraint, we sample the variables c_{it} only for populations $i \notin \Gamma_t$, that is, those populations that are eligible for cluster reassignment, while we set c_{it} to the allocation at the previous time step $c_{i,t-1}$ for $i \in \Gamma_t$. Let us denote with $\boldsymbol{c}_t^{(c_{it}=k)}$ the vector $\boldsymbol{c}_t$ with the i–th entry set to k, and recall that $\boldsymbol{c}_t^{(-i)}$ and $K_t^{(-i)}$ are the membership vector $\boldsymbol{c}_t$ after removal of the i–th entry and the corresponding number of clusters. Leveraging Bayes rule and the predictive scheme of the Gnedin prior, the full-conditional distribution of c_{it} for $k = 1, \ldots, K_t^{(-i)} + 1$ is given by

$$\operatorname{pr}(c_{it} = k \mid -) \propto \operatorname{pr}(c_{it} = k \mid \boldsymbol{c}_t^{(-i)}) \mathbb{1}[\mathbb{c}_{t+1} \in \mathbb{C}(\boldsymbol{c}_t^{(c_{it}=k)}, \boldsymbol{\gamma}_{t+1})] p(\boldsymbol{y}_{it} \mid \boldsymbol{\theta}_{kt}^{\star}),$$

where $\mathbb{C}(\boldsymbol{c}_t^{(c_{it}=k)}, \boldsymbol{\gamma}_{t+1})$ is the set of all partitions at time $t + 1$ that are compatible with the current allocation $\boldsymbol{c}_t^{(c_{it}=k)}$ and the latent variables $\boldsymbol{\gamma}_{t+1}$, which are sampled in the previous step. The first term at the right-hand side is the Gnedin prior probability of the allocation $c_{it} = k$, while the second term is the marginal Hamming likelihood of $\boldsymbol{y}_{it}$ given the cluster-specific parameters $\boldsymbol{\theta}_{kt}^{\star}$, presented in Argiento et al. (2024, Proposition 3) . The compatibility constraint ensures that only those allocations that are compatible with the future allocations $\mathbb{c}_{t+1}$ are considered, as required by the tRPM construction. Notice that only compatibility with the future allocations $\mathbb{c}_{t+1}$ is required, as the one with the past $\mathbb{c}_{t-1}$ is already satisfied by construction.

We conclude the tRPM block by sampling the hyperparameters α_t from their full-conditional distribution, which is obtained through the conjugate update of the Beta prior as

$$(\alpha_t \mid \boldsymbol{\gamma}_t) \sim \operatorname{Beta}\left(a_\alpha + \sum_{i=1}^{n} \gamma_{it},\ b_\alpha + n - \sum_{i=1}^{n} \gamma_{it}\right), \quad t = 2, \ldots, T.$$

The second step of the Gibbs scheme consists in sampling the cluster-specific parameters $\boldsymbol{\theta}_{kt}^{\star}$, for $k = 1, \ldots, K_t$ and $t = 1, \ldots, T$, conditionally on the sampled partitions $\boldsymbol{c}_t$ (and latent variables $\boldsymbol{\gamma}_t$), from the full-conditional distribution derived in Argiento et al. (2024). In particular, the components $\theta_{kxt}^{\star}$ of cluster centers are sampled from categorical distributions with probabilities $p(\theta_{kxt}^{\star} \mid -) = p(\{\boldsymbol{y}_{ixt} \colon c_{it} = k\} \mid \theta_{kxt}^{\star}, \boldsymbol{c}_t)$ for $\theta_{kxt}^{\star} \in \{1, \ldots, m_x\}$.

3.3.2 Adaptive Parallel Tempering

Although the Gibbs sampling scheme detailed in Section 3.3.1 is theoretically valid, the posterior distribution involves a complex, high-dimensional discrete space defined by the sequence of dependent random partitions. To prevent the sampler from becoming trapped in local modes, a common risk in such settings, we improve the standard Gibbs routine with an adaptive Parallel Tempering (PT) framework, a method proven to facilitate the exploration of challenging probability landscapes (Earl and Deem, 2005; Syed et al., 2022).

The core mechanism involves simulating multiple chains at varying temperatures: high-temperature chains explore the space broadly by sampling from flattened distributions, while lower-temperature chains focus on local refinement of the true posterior, gradually approaching the chain with the minimum temperature (equal to 1), which targets the true posterior. Periodically, these chains exchange their most recent states via Metropolis-Hastings moves, which preserves detailed balance. Furthermore, to avoid manual prior tuning of the temperature schedule, we employ the stochastic optimization approach of Miasojedow et al. (2013), which recursively adjusts the temperature grid to target specific acceptance rates via Robbins-Monro updates. Combined with non-reversible deterministic swaps (Syed et al., 2021), this setup guarantees that the global exploration achieved at high temperatures is efficiently transferred to the chains of interest, ensuring reliable inference.

3.3.3 Monte Carlo Inference

Posterior inference on quantities of interest is performed via Monte Carlo, relying on samples drawn by the routine outlined in Section 3.3.1. Our primary focus is to uncover clustering structures among populations, based on similarity between trajectories of age-specific leading causes of death. We summarize the posterior samples of the membership vectors c_{it}, for $i = 1, \dots, n$ and $t = 1, \dots, T$, through the $n \times n$ posterior *co-clustering* matrices $\hat{\mathbf{P}}_t$, for $t = 1, \dots, T$. The generic element $\hat{\mathbf{P}}_{t[i,i']}$ corresponds to the proportion of posterior samples in which populations i and i' are assigned to the same cluster at time t. This yields an estimate of the posterior probability $\mathrm{pr}(c_{it} = c_{i't} \mid \mathbf{Y})$, given the three-dimensional array of leading causes of death across all age classes and periods $\mathbf{Y}$, serving not only as a point estimate but also as quantification of uncertainty in the clustering patterns.

To obtain a point estimate $\hat{\boldsymbol{c}}_t$ of $\boldsymbol{c}_t$, for each $t = 1, \dots, T$, we rely on the decision-theoretic framework of Dahl et al. (2022). Given the above posterior co-clustering matrices and the posterior samples, the estimate is obtained by solving the optimization problem

$$\hat{\boldsymbol{c}}_t = \underset{\boldsymbol{c}_t'}{\mathrm{argmin}} \; \mathbb{E}[\, \textsc{vi}(\boldsymbol{c}_t, \boldsymbol{c}_t') \mid \mathbf{Y}],$$

where VI is the variation of information distance, which has been introduced by Meilă (2007) as an information-theoretic measure for differences between two partitions of the same set of entities. We find the minimizer of such a problem, for each $t = 1, \dots, T$, via the `R` package `salso` (Dahl et al., 2022).

To characterize the estimated clusters we obtain straightforward plug-in estimates for the centers and scales, conditionally on the computed partitions. Let $\hat{K}_t$ be the number of estimated clusters at period t, then, for every $x = 1, \dots, X$, and $k = 1, \dots, \hat{K}_t$, a sensible proposal for the center components is the marginal posterior mode

$$\hat{\theta}^\star_{kxt} = \underset{\theta}{\mathrm{argmax}} \; \mathrm{pr}(\theta^\star_{kxt} = \theta \mid \mathbf{Y}, \boldsymbol{c}_t = \hat{\boldsymbol{c}}_t),$$

while for the scale parameters we consider the posterior median

$$\hat{\sigma}^\star_{kxt} = \mathrm{median}\left(\sigma^\star_{kxt} \mid \mathbf{Y}, \boldsymbol{c}_t = \hat{\boldsymbol{c}}_t\right).$$

Monte Carlo estimates of $\hat{\theta}^\star_{kxt}$ and $\hat{\sigma}^\star_{kxt}$ can be obtained relying on additional samples from the full-conditional distributions of the cluster-specific parameters, given the estimated partitions $\hat{\boldsymbol{c}}_t$. In order to obtain draws from both the centers and the scale parameters, we employ in this phase a standard Gibbs sampler, without marginalizing out $\boldsymbol{\sigma}^\star_{kt}$, $k = 1, \dots, \hat{K}_t$ and $t = 1, \dots, T$. In particular, the components $\theta^\star_{kxt}$ of cluster centers are sampled from categorical distributions with probabilities proportional to $\exp\{-(n_{kt} - \sum_{i:\, c_{it}=k} \mathbb{1}[y_{ixt} = \cdot])\}$ for $\cdot \in \{1, \dots, m_x\}$, where n_{kt} is the number of populations in cluster k at time t. The scale parameters $\sigma^\star_{kxt}$ are sampled from Hypergeometric Inverse Gamma distributions with parameters $a^\sigma_x + \sum_{i:\, c_{it}=k} \mathbb{1}[y_{ixt} = \theta^\star_{kxt}]$ and $b^\sigma_x + n - \sum_{i:\, c_{it}=k} \mathbb{1}[y_{ixt} = \theta^\star_{kxt}]$.

It is important to underline that we must use quantile-based summaries of the full-conditional distributions of the scale parameters, as the HIG distribution has infinite mean for every possible value of the hyperparameters, and so does the full-conditional distribution (Argiento et al., 2024). We refer to Argiento et al. (2024) for an alternative parameterization of the scale parameters, as $\omega^\star_{kxt} = \exp(-1/\sigma^\star_{kxt})$, for which analytical expressions of the finite mean and mode are available, both for the prior and the full-conditional distributions.

Finally, we summarize the posterior samples of the transition parameters α_t, for $t = 2, \dots, T$, through their posterior means and highest posterior density credible intervals, which are obtained from the samples of the full-conditional distributions, drawn during the main Gibbs scheme detailed in Section 3.3.1.

3.4 Simulation Study

To evaluate the performance of the proposed model, we conduct a simulation study based on synthetic data generated as follows. We consider a scenario with $n = 100$ populations, $T = 8$ years, and $X = 5$ age classes, with $m_x = \{3, 7, 7, 7, 4\}$ possible categories for each

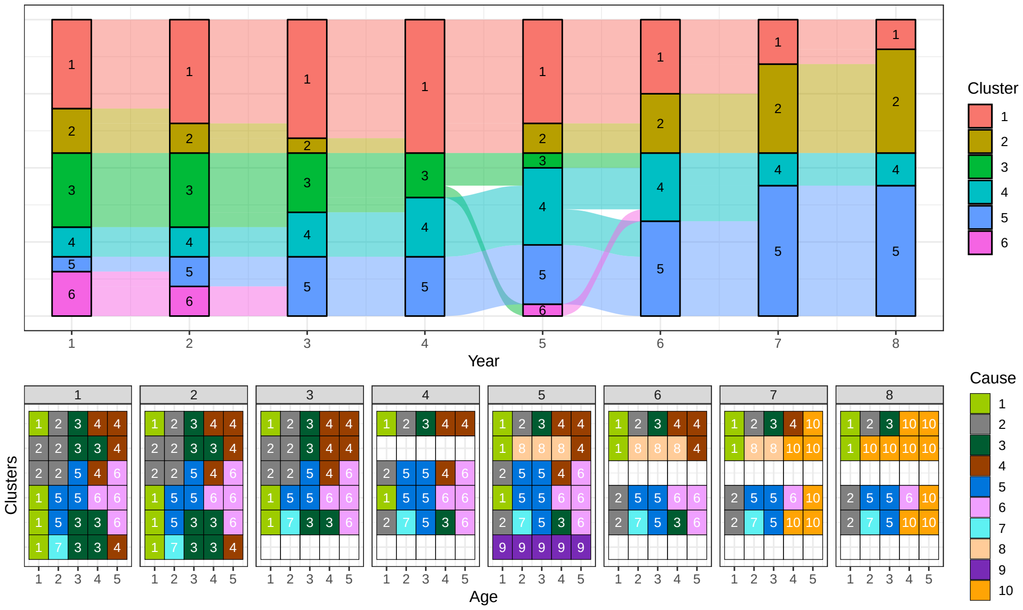


Figure 3.3: Realization of the simulated data: (top) the evolution of the partitions, (bottom) cluster-specific centers.

age class. To check the robustness of the results, we replicate the simulation 100 times, fixing the number of clusters for every year, their sizes and the number of populations transitioning between clusters across years, while randomly generating the populations that change allocations. For each combination of cluster and year (k, t), we fix a modal profile $\boldsymbol{\theta}^\star_{kt}$ and we sample the observed components y_{ixt} from categorical random variables that assign 0.95 probability to the modal value $\theta_{ixt} = \theta^\star_{c_{it}xt}$ and spread the remaining 0.05 probability uniformly across the other possible values for each age class x. The top panel of Figure 3.3 illustrates a realization of the simulated dynamic partition across years, which shows smooth transitions between configurations with varying numbers and sizes of clusters. Moreover, the bottom panel displays the cluster-specific modal profiles, which highlight the differences in leading causes of death across clusters, age classes, and years. Reflecting the realistic scenario of leading causes of death, we impute at the fifth year a sudden change in the partition structure, with a new small cluster emerging, which is characterized by a center that is significantly different from the others. This change is designed to mimic the impact of external shocks, such as natural disasters or wars. Additionally, we also induce a significant change in the modal profiles in the seventh and eigth years, which is designed to resemble the impact of a global COVID-19 pandemic.

The model in Section 3.2 is fully specified by setting the hyperparameters as follows. We fix $a^\sigma_x = 5, 4.5$ and 4 respectively for $m_x = 3, 4$ and 7, while $b^\sigma_x = 0.25$ for every age class. This choice allows to achieve an approximate uniform prior on the normalized Gini index (Gini, 1912) of each categorical variable, as suggested by Argiento et al. (2024) to reflect the absence of knowledge on the variability of the phenomenon. We fix the Gnedin parameter $\eta = 0.5$, which implies a prior expected number of clusters close to 9, which is significantly larger than the maximum number of *true* clusters across years. The hyperparameters of the Beta priors on the transition parameters α_t, $t = 2, \dots, T$, are set to $a_\alpha = b_\alpha = 1$, which correspond to a uniform prior on $[0, 1]$, reflecting our lack of prior knowledge on the expected proportion of populations that are allowed to change cluster assignment from one year to the next. We fit the proposed Bayesian model to the simulated data using the PT scheme relying on a collapsed Gibbs sampler described in Section 3.3.1, running 30,000 iterations and discarding the first 5,000 as burn-in, and thinning by 10 iterations, obtaining a sample of size 2,500. Posterior inference is performed as in Section 3.3.3, focusing on the recovery of the *true* clustering structure and cluster-specific parameters.

Consistent with our overarching goal, we first assess the model's ability to recover the *true* clusters by computing the normalized mutual information (NMI) (Kvalseth, 1987) between the simulated and estimated partitions for each year. The NMI is an *information-theoretic* measure of similarity between two partitions, that can be computed as the normalized difference between the joint entropy and the VI of the two partitions. As such, it assumes values close to 1 when the two partitions are similar (low VI), and close to 0 when they are dissimilar (high VI). Table 3.2 reports the quartiles of the NMI, across the 100 replicates,

Cluster recovery

Year	1	2	3	4	5	6	7	8
Min	0.85	0.88	0.85	0.93	0.87	0.92	0.96	0.92
Q25	0.95	0.94	0.89	0.97	0.97	0.97	1.00	0.96
Median	0.97	0.97	0.89	1.00	0.99	1.00	1.00	1.00
Q75	0.98	1.00	0.89	1.00	1.00	1.00	1.00	1.00
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Central profile recovery

Year	1	2	3	4	5	6	7	8
Min	0.92	0.90	0.92	0.92	0.92	0.93	0.92	0.93
Q25	0.94	0.94	0.93	0.94	0.94	0.94	0.95	0.94
Median	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95
Q75	0.96	0.95	0.95	0.96	0.96	0.96	0.96	0.95
Max	0.97	0.98	0.96	0.98	0.97	0.97	0.97	0.97

Table 3.2: Summary statistics of simulation performance metrics across 100 replications.

between the *true* and estimated partitions, illustrating the ability of the proposed Bayesian temporal clustering model to capture the dynamic partition patterns in multivariate categorical data, with the NMI values close to 1 for most replicates and years. We proceed with the analysis by verifying the model's ability to recover the cluster-specific central profiles $\boldsymbol{\theta}^\star_{kt}$, for $k = 1, \dots, K_t$ and $t = 1, \dots, T$. In particular, for each population $i = 1, \dots, n$ and year $t = 1, \dots, T$, we compare the observed data $\boldsymbol{y}_{it}$ with $\hat{\boldsymbol{\theta}}_{it} = \hat{\boldsymbol{\theta}}^\star_{\hat{c}_{it}t}$, computing the normalized Hamming similarity $\hat{s}_{it} = (1/X)\sum_{x=1}^{X} \mathbb{1}[y_{ixt} = \hat{\theta}_{ixt}]$, which corresponds to the proportion of correctly recovered components. We then take the mean across populations, $\hat{s}_t = n^{-1}\sum_{i=1}^{n} \hat{s}_{it}$, obtaining a global measure of accurate recovery for each year t. The quartiles, across the 100 replicates, of such a measure in Table 3.2 show that the model is not only able to recover the clustering structure, but also to accurately estimate the cluster-specific central profiles, with no values below 0.92 for every possible combination of year and replicate.

These results demonstrate that the proposed Bayesian temporal clustering model can effectively identify dynamic clustering patterns and characterize cluster-specific center structures in time-indexed multivariate categorical data.

3.5 Evolution of Leading-Cause Profiles in 183 Countries in the 21$^{\text{st}}$ century

In this section, we illustrate the ability of the proposed model to capture the evolution of the leading causes of death profiles across multiple countries and age classes, focusing on the period from 2000 to 2020. The analysis is conducted on the WHO mortality database (World Health Organization, 2024a), described in Section 3.1, which contains the leading cause of death for each combination of 183 countries, 2 sexes, 19 age classes, and 5 calendar years. To fit the model, we elicit priors following a similar approach to the one described in Section 3.4, but with some modifications to account for the specific characteristics of the data. In particular, we fix $a^\sigma_x = 4$ and $b^\sigma_x = 0.25$, for every $x \in \mathcal{X}$ and the Gnedin parameter $\eta = 0.55$, implying an *a priori* expected number of clusters around 9, which is a reasonable assumption that aligns with the results of preliminary explorative analyses performed separately for each year. Finally, we opt again for a uniform prior on α_t, for every

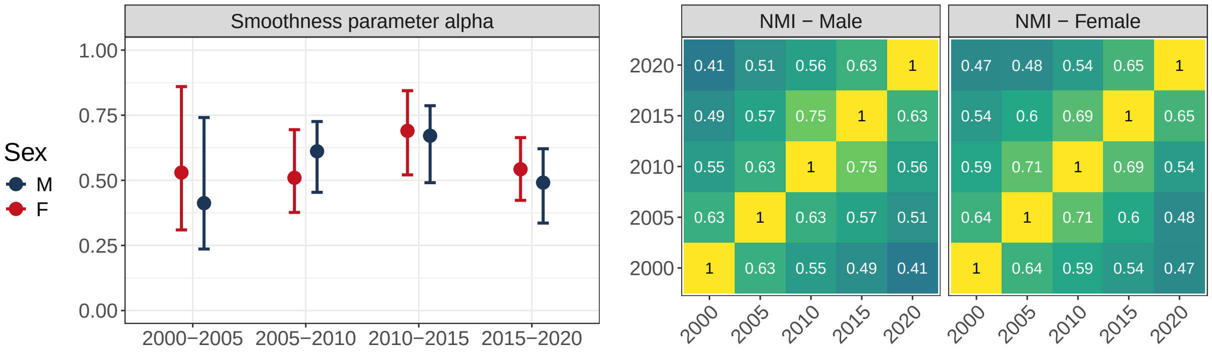


Figure 3.4: Left: Posterior means (dot), 95% highest posterior density credible intervals (whiskers) for estimated transition parameters α_t. Right: pairwise NMI between the clustering structures estimated for each year. Colors range from blue (minimum NMI, 0) to yellow (maximum NMI, 1).

$t = 2, \dots, T$. As in the simulation study, the sampling scheme is run for 30,000 iterations, with the first 5,000 discarded as burn-in, and thinned by 10 iterations, yielding a posterior sample of size 2,500.

The left panel of Figure 3.4 displays the posterior means and 95% highest posterior density credible intervals of the transition parameters α_t, which describe the estimated smoothness in the evolution of countries' partitions over time. The posterior means fluctuate around 0.55 for both male and female populations across the four transitions, showing a slightly increasing trend from 2005 to 2015. This is followed by a mild decline in the transition to 2020, likely associated with the COVID-19 pandemic, which interrupted the convergence process of the leading causes of death landscape. These findings are supported by the pairwise NMI between the estimated partitions for each year, illustrated in the right panel of Figure 3.4. These quantities extend the analysis of partition similarity beyond consecutive time points, highlighting a substantial change from 2000 to 2015, with overlaps of 49% and 54%, in NMI terms, for male and female populations, respectively.

Figure 3.5 offers a comprehensive visualization of the evolution of estimated partitions by mapping cluster memberships from 2000 to 2020. Despite the dynamic landscape presented in Figure 3.4, some stable relationships persist across the years. Notably, a substantial group of 17 female populations demonstrates consistent co-clustering. This includes females from all Western Offshoots but the USA, Singapore, the UK and Ireland, the four Scandinavian nations, all major continental Western European countries (the Netherlands, Belgium, France, Germany, Switzerland, Austria), and Spain; all of these remain in the dark red cluster throughout the observed period. Consistent co-clustering across the five years is also evident in 9 female populations from Southern Africa, as well as 13 countries from the Middle East & North Africa and the Pacific Islands, which always belong to the dark blue and red cluster, respectively. It is noteworthy that the strong convergence trend toward fewer clusters found in female Sub-Saharan African populations is not observed in males, whose structure is characterized by a greater volatility. Nevertheless, we identify two large groups of male populations that share cluster memberships across all five years, too. The first set

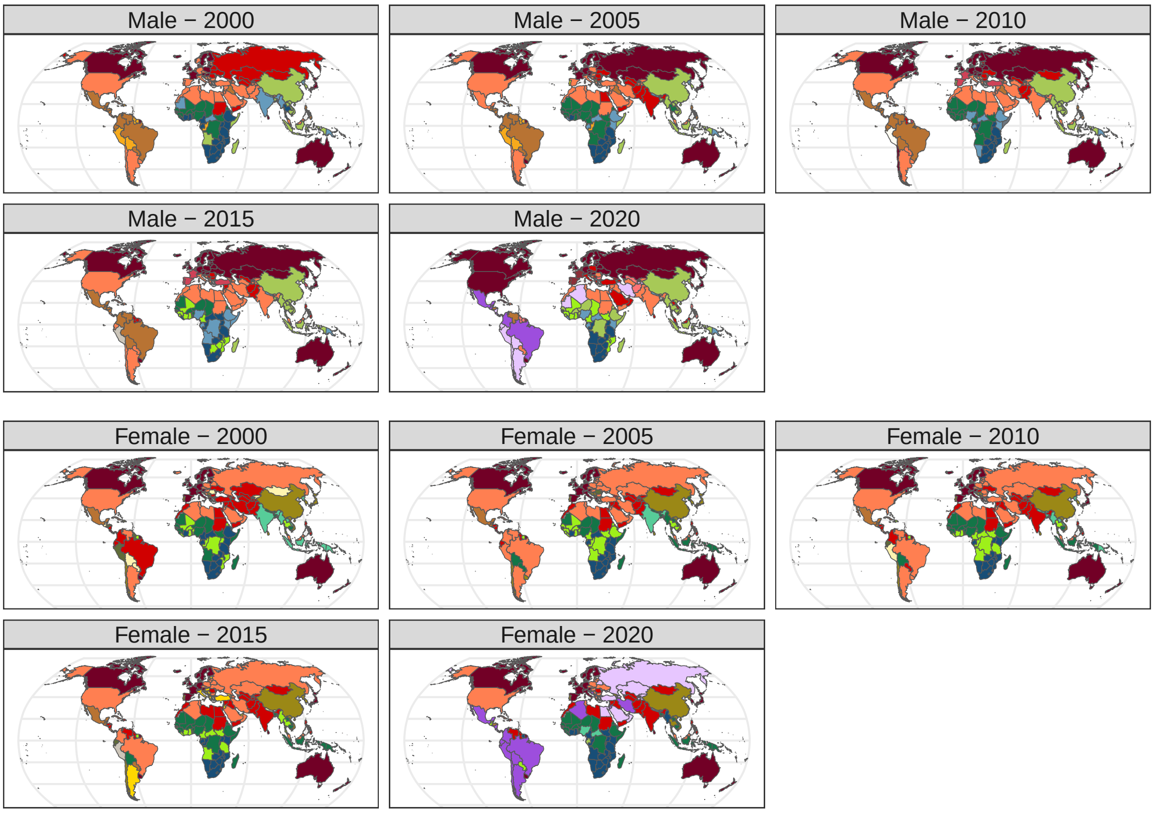


Figure 3.5: Countries colored according to the estimated cluster memberships, for both male (top) and female (bottom) populations.

includes 14 countries, consistently belonging to the dark red cluster, and partially overlaps with the aforementioned 17 female populations from developed countries, differing only in the exclusion of continental Western Europe and the inclusion of Iceland, Lithuania, and Japan. The final example of continuous co-clustering involves of 12 countries from Southern Europe (Italy and Greece), the Middle East (Lebanon, Azerbaijan, Bahrain, Kuwait, Oman, Saudi Arabia, and the UAE), and the Indian Ocean region (Brunei, Malaysia, and Mauritius). These nations always fall within the orange cluster, a group frequently joined by other Northern African countries, Argentina, and the USA.

We extend the analysis by examining the estimated clusters in terms of the corresponding central profiles, $\hat{\boldsymbol{\theta}}^{\star}_{kt}$, for $k = 1, \dots, \hat{K}_t$ and $t = 1, \dots, T$, as displayed in Figure 3.6. The opacity of the cell colors indicates the strength of the representativeness of the estimated central component $\hat{\theta}^{\star}_{kxt}$. Specifically, this is defined as the proportion of countries in a cluster with the same age-specific leading cause as the one displayed, with more opaque cells corresponding to higher representativeness. First, there is clear overall stability in the estimated profiles for both sexes over the observed period, with a notable absence of abrupt year-to-year changes. The only sudden shifts are associated with the emergence of respiratory system cancers (C2) in the 2005 panels, which will be discussed in more detail below, and the onset of the COVID-19 pandemic (I7) in the 2020 panels. This consistency

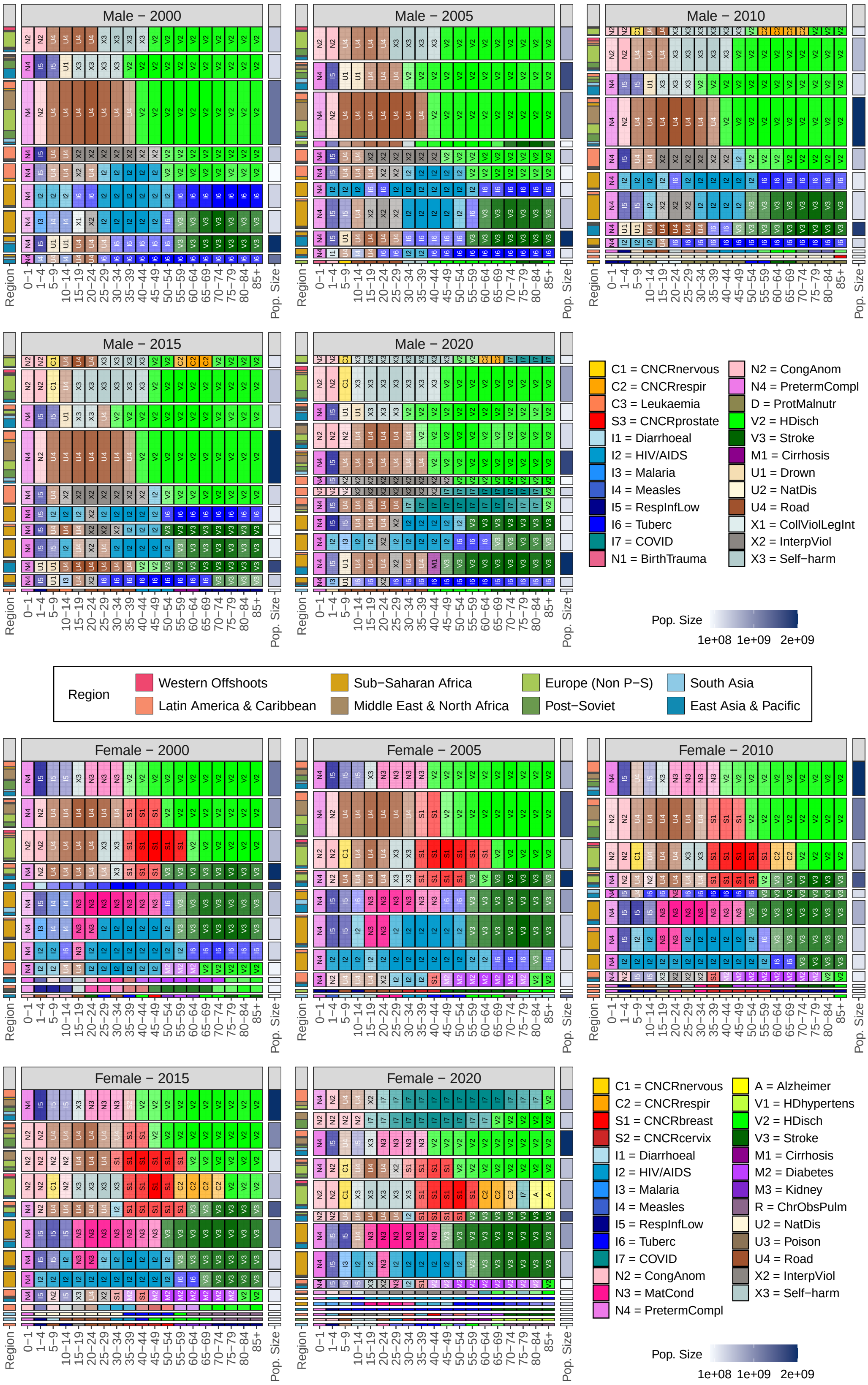


Figure 3.6: Estimated cluster centers $\hat{\theta}_{kxt}$ for male (top) and females (bottom) populations. The height of each row is proportional to the cluster size. Each cell's opacity indicates the share of cluster members whose age-specific leading cause of death matches the one shown (full opacity denotes 100%).

in the estimated profiles allows us to identify recurring patterns. Both male and female populations can be divided into two macro-groups based on the leading cause of death among individuals aged 50 and over: ischaemic heart disease (V2) in the first group, and tuberculosis (I6) or stroke (V3) in the second. Interestingly, this division is closely linked to the geography, particularly for males, where the second group consists almost entirely of populations from Sub-Saharan Africa and the East Asia & Pacific region.

Notably, in 2000, several clusters (two among males and three among females) are composed almost entirely of countries located in Sub-Saharan Africa. The two male clusters share HIV/AIDS (I2) as the predominant cause of death among individuals aged 25–49 but differ in mortality at younger and older ages, where tuberculosis and stroke become dominant. Among females, two clusters exhibit mortality profiles similar to those observed among males, while a third cluster emerges, distinguished by the incidence of maternal conditions (N3) during women's reproductive years. The evolution of these clusters is strongly driven by the declining burden of tuberculosis (I6), a pattern particularly noticeable among female populations. Indeed, the eighth cluster in the 2000 female panel of Figure 3.6 already displays a moderate prevalence of tuberculosis, indicated by the cells' transparency, which steadily decreases until 2020, when the Sub-Saharan tuberculosis-related cluster merges with its contiguous one. African male populations are affected by this trend to a lesser degree, possibly delayed by some years, as we notice a reduced scope of the tuberculosis-related Sub-Saharan cluster only in 2020. Conversely, some East Asian & Pacific male populations show a more consistent reduction of the tuberculosis burden. As such, the eighth cluster in the 2000 male panel of Figure 3.6 is only weakly characterized by tuberculosis, which is later replaced by stroke. Interestingly, this scenario differs from the female one as it affects not only different populations but also different age groups. These findings may support a further investigation into the dynamics of the tuberculosis burden and the disparities between male and female populations.

Latin American males present a distinct and evolving phenomenon. At the beginning of the century, they show similarities with both developed and developing countries, but subsequently converge toward a distinctive profile that diverged from other nations. More specifically, in 2000, the majority of countries from Central and South America formed two distinct clusters (fourth and fifth rows in the top-left panel of Figure 3.6). Both are characterized by ischaemic heart disease as the main cause of death among people aged 50 years or older, a characteristic shared with clusters comprising middle- and highly-developed countries from Europe, Western Offshoots, and the MENA region (first three rows in the top-left panel of Figure 3.6). The first cluster includes many of the continent's largest countries (Brazil, Colombia, Ecuador, Guatemala, Honduras, Mexico, Nicaragua, Paraguay, El Salvador and Venezuela) and is highly distinguishable due to the peculiar extended influence of interpersonal violence (X2) as the leading cause of death spanning the 15–19 to 45–49 age classes. On the other hand, the second Latin American cluster consists mainly of smaller insular nations (Antigua and Barbuda, Barbados, Bolivia, Dominican Republic, Grenada, Guyana, Haiti, Panama, Peru, Saint Vincent and the Grenadines, and Suriname) and, while similar to developed countries regarding the older population, this cluster's younger population are more closely related to Sub-Saharan and East Asian & Pacific countries, as it is characterized by a large window (ages from 25 to 49) of HIV/AIDS (I2) influence. The multi-faceted profile of this cluster evolved smoothly in 2005, when interpersonal violence emerges as the representative leading cause among males aged between 20 and 29 years. This evolution continued more strongly in 2010, when the two clusters merge into a single one and HIV/AIDS almost completely disappears from the estimated cluster profile. This transitional nature of the cluster can also be observed in the more transparent cells corresponding to adolescents and young adults in the aforementioned Latin American cluster, which depicts the mixed and unstable landscape of these countries in the early 2000s.

This pattern is only partially mirrored in Central and South American female populations, which are more dispersed across different clusters than their male counterparts. Nevertheless, in 2000 we notice a distinct, strong influence of diabetes among adults in some clusters including nations from Latin America; this persists in more recent years, though these clusters also include some East Asian & Pacific countries. The burden of diabetes is a well-acknowledged issue in this region. Its high prevalence and mortality are often ascribed to low awareness and unhealthy dietary habits, as well as late identification and limited access to insulin (Aschner et al., 2014; Sinisterra et al., 2019; Lopez-Jaramillo et al., 2021). Mexico, in particular, constitutes a worrying scenario (Meza et al., 2015), not only due to strong genetic susceptibility but also to the scarcity of clean water accessibility, which often reflects in the high consumption of cheap, sugary beverages (Stern et al., 2019). In the last decades, awareness surrounding this increasing burden has grown, leading to tailored policies expected to reduce the impact of diabetes on the Mexican population (Barrientos-Gutierrez et al., 2017).

On the other hand, we identified two additional clusters within the female populations that serve as a bridge between the European, MENA, and Western Offshoots nations, characterized by profiles involving road accidents, breast cancer, and ischaemic disease, and the Sub-Saharan clusters, which are dominated by maternal conditions, HIV/AIDS, and stroke. The first of these groups (first row in the first four female panels of Figure 3.6) is geographically diverse and is homogeneous mainly due to ischaemic disease being the leading cause of death in older populations. In contrast, the leading causes in early life and among young adults (lower respiratory system infections and maternal conditions, respectively) are similar to those of the first cluster with a Sub-Saharan majority (sixth row in the 2000, 2010, and 2015 panels and fifth row in the 2005 panel). However, it is important to note that these causes do not strongly represent the populations in the cluster, reflecting the varied scenario for these age groups. In contrast, the second cluster of this type resembles European and MENA clusters regarding leading causes among young people, such as road accidents, self-harm, and breast cancer, while resembling Sub-Saharan clusters regarding stroke in older populations. This latter cluster comprises countries from Europe, East Asia, and South America, with Albania, Chile, China, Montenegro, North Macedonia, Portugal, South Korea, and Vietnam consistently belonging to it across the analyzed period.

Shifting the focus to clusters with a European majority, we observe in Figure 3.6 the emergence, for both males and females, of a cluster whose estimated central leading cause of death in late adulthood (55–74) is cancer of the respiratory system. Although this trend is present in both sexes, the cluster members differ. For female in 2010, we observe a more heterogenous cluster of larger size, where only some of the countries are primarily affected by respiratory cancers, and the homogeneity is instead still more strongly driven by widespread self-harm, breast cancer, and ischaemic heart disease as the leading cause of death in the 25–34, 25–54, and 70–74 age classes, respectively. This cluster includes Nordic countries (Scandinavia, Iceland, the UK, and Ireland), the Benelux, German-speaking countries (Germany, Austria, Switzerland), and several Mediterranean nations (Spain, France, Italy, and Greece), as well as all Western Offshoots except the USA. The composition of this cluster remains largely unchanged in 2015, with 21 of its 23 members already present in 2010, but the homogeneity around respiratory cancers increases, particularly among individuals aged 60–69. In contrast, among males, a smaller but more homogeneous cluster already emerges in 2010, comprising Belgium, France, the Netherlands, Slovenia, South Korea, Spain, and Turkey. Interestingly, all of these countries except Turkey also share self-harm as the leading cause of death in adults. It is worth noting that only a few of these countries also appear in the corresponding female cluster, which instead overlaps more with the second male cluster in 2010 panel in Figure 3.6, including 18 of the 25 countries from the female one. As stated in Tanday (2016), it must be stressed that the appearance of this trend around 2010 should not be mistaken for an increasing, and therefore alarming, trend in mortality related to respiratory cancers. In fact, despite a decrease in the incidence of such cancers, the observed phenomenon results from a faster decline in deaths due to cardiovascular diseases (Townsend et al., 2016; European Heart Journal, 2016), which consequently cease to dominate as the leading cause of death. The inferred differences between male and female patterns align with recent findings (e.g. Jani et al., 2021), which highlight a faster decrease in male lung cancer mortality rates compared with the steadier decline observed among females.

Over the last decade, many countries, particularly high-income and low-mortality ones, have experienced an increase in the burden of premature mortality due to “deaths of despair”, especially those caused by self-harm (Hawton et al., 2012; Case and Deaton, 2021, 2020). This trend is also evident in the evolution of cluster profiles, as shown in Figure 3.6, which highlights the growing impact of self-harm, evidenced by both the increased number of affected male populations (corresponding to taller tiles in Figure 3.6) and the younger age groups involved. Notably, self-harm became the leading cause of death for the 10–14 age group in three clusters in 2020. To gain a deeper understanding of this scenario we show

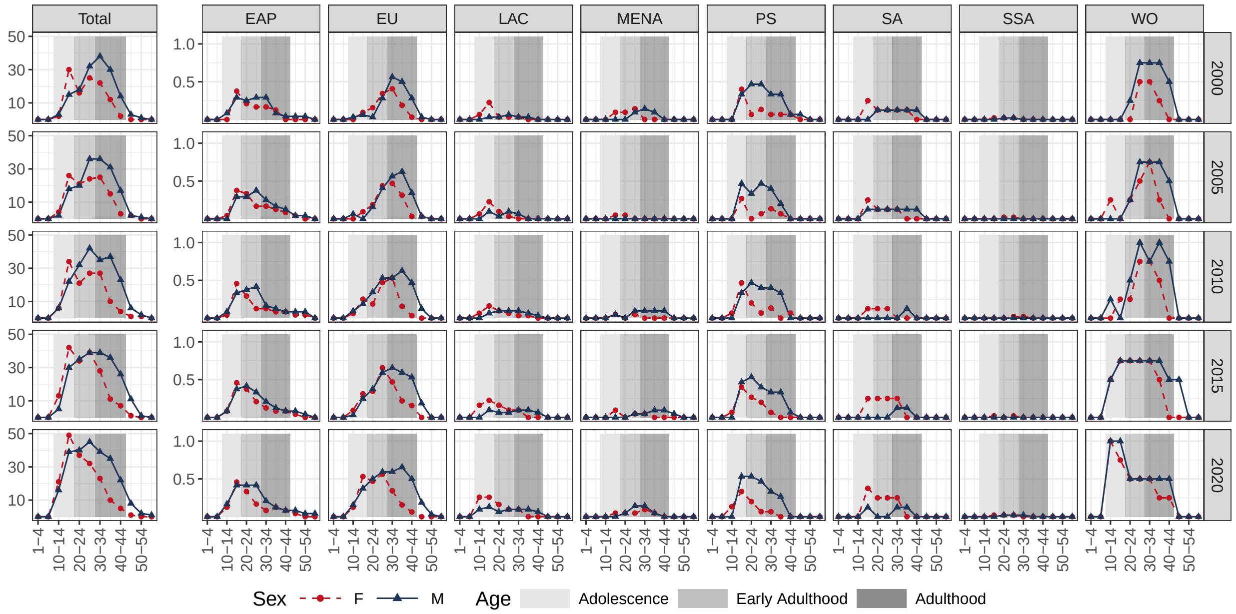


Figure 3.7: Left: number of populations with self-harm as the leading cause for male (blue lines) and female (red lines) populations across age classes and years. Right: proportion of populations from each region with self-harm as the leading cause for male and female populations across age classes and years.

in Figure 3.7 how the number of age-specific sub-populations with self-harm as the leading cause of death has changed across years, regions and sexes. While there are some differences between the male and female populations, we highlight some persistent trends. First, the total number of such sub-populations has increased over the last twenty years, from an observed minimum of 263 (154 male, 109 female) in 2000 to 425 (247 male, 178 female) in 2020. Moreover, we initially observe a bimodal distribution of self-harm cases in the female population, with one peak in adolescence and another in adulthood, but has since shifted toward a unimodal, right-skewed distribution with a single peak at younger ages. On the other hand, the male population has never exhibited a bimodal pattern, but shares with the female population the shift to younger ages, although the modal age never reaches the early adolescent ages. Figure 3.7 raises a particular concern for Western Offshoots, whose observed leading causes are displayed in Figure 3.8. The shift toward an earlier onset of self-harm is evident and applies to both females and males; moreover, the latter still show a wider impact range, reaching adults up to age 49. Furthermore, the pciture highlights that the USA is also affected by conditions related to deaths of despair, even though this group appears in the high-intensity clusters (dark red in Figure 3.5) only in 2020 for males. While self-harm is not as widespread a leading cause as it is in Australia, Canada, and New Zealand, there is a strong presence of drug use disorders (X4). This trend began in 2005 with females aged 35–39 and later spread to the 20–24 and 45–49 age groups. Males are affected by a similar pattern, but with a five-year lag. This analysis demonstrates that while the self-harm trend is clear, a complete evaluation of the "deaths of despair" phenomenon requires a classification of causes tailored to the specific problem, which goes beyond the

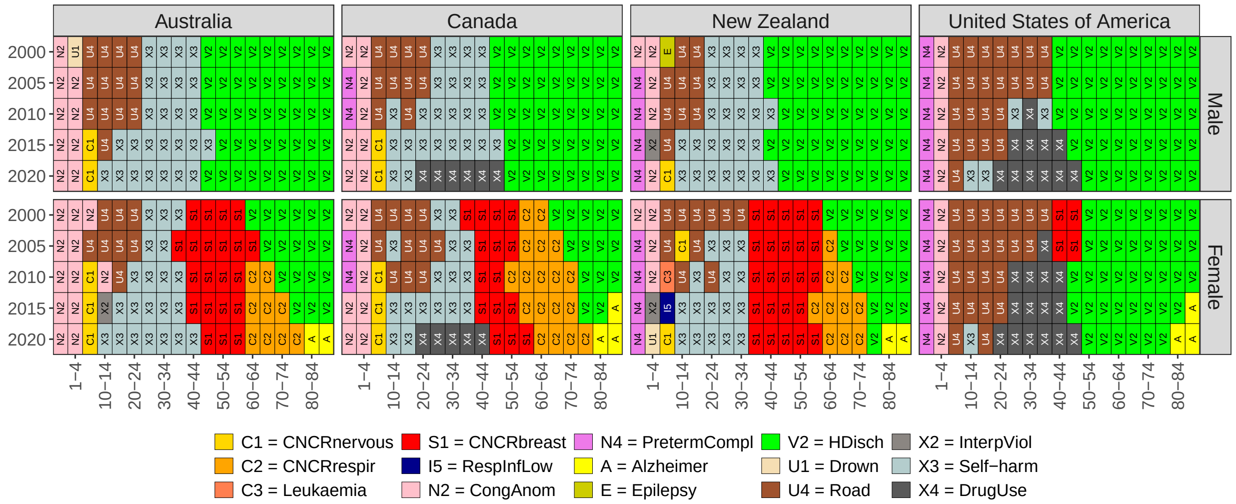


Figure 3.8: Observed leading causes of deaths in Western Offshoots over the analyzed period.

scope of this study.

The composition of the self-harm-related male clusters is also noteworthy. In 2000, one cluster was mainly composed of European (non P-S) and Western Offshoot countries, characterized by a much later onset of self-harm than the second cluster, which primarily included post-Soviet and East Asia & Pacific countries. This clear division gradually faded due to the earlier onset of self-harm as leading cause in countries from the first cluster (columns 3 and 9 of Figure 3.7), aligning them with those in the second cluster (columns 2 and 6 of Figure 3.7). By 2020, the main difference between the two clusters only lies in the more extended duration of the self-harm window in the one mainly composed by European countries. Moreover, it is noteworthy that, despite the emergence of COVID-19 in 2020, self-harm continues to be a leading cause of death among adolescents and young adults.

Although a similar trend is partially present among females (bottom half of Figure 3.6), there are clear differences. The main one concerns the number of affected age groups, which is significantly smaller for women due to the strong influence of female-specific causes of death, such as breast cancer and maternal conditions, which contribute to the earlier disappearance of self-harm. This difference also impacts the geographical composition of these clusters. Indeed, a Europe (non P-S) and Western Offshoots cluster is present across all five years and remains unaffected by countries from other regions, unlike the male population. The other cluster of female populations has self-harm as the leading cause only in the 15–19 age group and includes populations from Central and South America, North Africa, and the Middle East. These sex-specific differences are well established in the literature (see, e.g., Knipe et al., 2022) and are clearly illustrated in Figure 3.7, which shows a stronger impact among females in younger age classes (10–14 and 15–19), reversing in older age classes (20–24 to 45–49), where self-harm is more prevalent among males. Although this pattern has long been observed in post-Soviet countries, it has notably widened in recent years across other European countries.

To conclude the analysis of the estimated clusters, we stress that the proposed model successfully captures potential shocks induced by sudden external and unexpected events. For instance, we observe in 2010 and 2015 the emergence of clusters related to earthquakes in Haiti (in males) and in Nepal (in females), respectively (DesRoches et al., 2011; Hall et al., 2017). The model also captures the impact of the 2010 East African drought, which led to a severe famine in Somalia (Maxwell and Fitzpatrick, 2012). This event is reflected as a cluster characterized by protein-energy malnutrition as the leading cause of death across almost all age classes in the male population.

3.6 Conclusions and Future Research Directions

Leading causes of death are a straightforward yet fundamental indicator of population health, providing an interpretable summary of mortality patterns. Analyzing age-, period-, and sex-specific leading causes of death across multiple populations is crucial for understanding the evolution of health conditions and identifying emerging health threats. In this work, we have introduced a Bayesian model for clustering profiles of leading causes of death across multiple populations, leveraging the temporal structure of the data through a time-varying random partition model (tRPM). Profiles of leading causes of death within each year-specific cluster are modeled as vectors of categorical variables through a highly interpretable model that parameterizes the profile distribution in terms of cluster centers and scale parameters. The resulting model is suitable for tractable posterior inference through a collapsed Gibbs sampling scheme embedded in a PT framework. This approach not only allows for clustering estimation but also provides estimates of the cluster centers and the scale parameters, which are crucial for understanding the underlying health patterns. We have applied the proposed model to a comprehensive dataset of leading causes of death from 183 countries stratified by two sexes and 19 age classes, covering the period from 2000 to 2020. The results reveal interesting stability across years and sexes, with some differences mainly attributable to the occurrence of sex-specific causes of death. Moreover, we have identified a clear connection between the estimated population clusters and geographic regions, highlighting the importance of socio-economic and health conditions in shaping mortality patterns. We concluded by identifying two major transitions in the last twenty years. First, we observed the replacement of ischaemic heart disease with cancer of the respiratory system as leading cause of death among older adults in lower-mortality countries. This shift is ascribable to the rapid decline in mortality rates of the former, driven by focused and effective preventive measures. Second, we highlighted the worsening trends of self-harm which involve an ever-increasing number of countries across different regions and worryingly extend toward younger age groups.

An important future research direction involves improving the distance evaluation between the observed vectors of leading causes of death and the cluster centers. The Hamming distance used in this work is a simple yet effective measure, but it does not account for the intrinsic ordering of the age classes. While this has not resulted in significant issues in the application presented here, it is worth exploring more sophisticated distance measures. Optimal matching (Abbott and Forrest, 1986) is a well-known method for comparing sequences of categorical variables, which could be adapted to our setting. However, its use is often limited by computational complexity, due to the intractability of the normalization constant in the likelihood of the exponential-distance model. A possible solution could be to leverage approximate Bayesian computation (Sisson et al., 2018) to estimate the model parameters, which would allow for more flexible distance measures without the need for exact likelihood computation. A less computationally intensive alternative approach could be to use a distance measure that accounts for the age proximity without losing the ability to compute the likelihood in closed form.

We conclude the discussion by highlighting the potential of the proposed model for further applications beyond the analysis of leading causes of death. The model can be adapted to other types of categorical vectors and can be applied to domains where the dynamic nature of clustering structures is crucial.

Appendix

3.A Table of causes

Table 3.A.1: Causes of Death

Abbreviation	**Cause**	**Age Range**
AfrTrypa	African trypanosomiasis	10–14 to 10–14
Alcohol	Alcohol use disorders	25–29 to 50–54
Alzheimer	Alzheimer disease and other dementias	5–9 to 85+
Asthma	Asthma	5–9 to 35–39
BirthTrauma	Birth asphyxia and birth trauma	0–1 to 0–1
ChrObsPulm	Chronic obstructive pulmonary disease	10–14 to 85+
Cirrhosis	Cirrhosis of the liver	5–9 to 60–64
CNCRbreast	Breast cancer	25–29 to 70–74
CNCRcervix	Cervix uteri cancer	25–29 to 70–74
CNCRcolon	Colon and rectum cancers	10–14 to 65–69
CNCRkidney	Kidney cancer	1–4 to 10–14
CNCRlarynx	Larynx cancer	55–59 to 55–59
CNCRliver	Liver cancer	5–9 to 70–74
CNCRmouth	Mouth and oropharynx cancers	10–14 to 15–19
CNCRnervous	Brain and nervous system cancers	1-4 to 40–44
CNCRovary	Ovary cancer	5–9 to 25–29
CNCRprostate	Prostate cancer	60–64 to 85+
CNCRrespir	Trachea, bronchus, lung cancers	10–14 to 75–79
CNCRskin	Melanoma and other skin cancers	40–44 to 40–44
CNCRstomach	Stomach cancer	20–24 to 65–69
CollViolLegInt	Collective violence and legal intervention	1-4 to 45–49
CongAnom	Congenital anomalies	0–1 to 25–29
COVID	COVID-19	1–4 to 85+
Dengue	Dengue	5–9 to 5–9
Diabetes	Diabetes mellitus	5–9 to 80–84
Diarrhoeal	Diarrhoeal diseases	0–1 to 85+
Drown	Drowning	1–4 to 45–49
DrugUse	Drug use disorders	15–19 to 45–49
Epilepsy	Epilepsy	5–9 to 25–29
Falls	Falls	5–9 to 20–24
Fire	Fire, heat and hot substances	1–4 to 10–14
Gastritis	Gastritis and duodenitis	10–14 to 10–14

HDhypertens	Hypertensive heart disease	5–9 to 85+
HDisch	Ischaemic heart disease	5–9 to 85+
HDmusclevalve	Cardiomyopathy, myocarditis, endocarditis	5–9 to 85+
HDrheum	Rheumatic heart disease	5–9 to 20–24
HepB	Acute hepatitis B	5–9 to 5–9
HIV/AIDS	HIV/AIDS	0–1 to 70–74
InterpViol	Interpersonal violence	1-4 to 70–74
IntestStop	Paralytic ileus and intestinal obstruction	10–14 to 10–14
Kidney	Kidney diseases	5–9 to 65–69
Leish	Leishmaniasis	5–9 to 10–14
Leukaemia	Leukaemia	1–4 to 30–34
Lymphomas	Lymphomas, multiple myeloma	1–4 to 35–39
Malaria	Malaria	0–1 to 65–69
MatCond	Maternal conditions	15–19 to 45–49
Measles	Measles	0–1 to 20–24
MechForces	Exposure to mechanical forces	5–9 to 40–44
Mening	Meningitis	1–4 to 10–14
NatDis	Natural disasters	0–1 to 80–84
NeonatSepInf	Neonatal sepsis and infections	5–9 to 10–14
Oral	Oral conditions	5–9 to 5–9
Poison	Poisonings	5–9 to 15–19
PretermCompl	Preterm birth complications	0–1 to 0–1
ProtMalnutr	Protein-energy malnutrition	1–4 to 85+
Rabies	Rabies	5–9 to 5–9
RespInfLow	Lower respiratory infections	0–1 to 85+
Road	Road injury	1–4 to 60–64
Self-harm	Self-harm	10–14 to 55–59
Sickle	Sickle cell disorders and trait	1–4 to 20–24
Stroke	Stroke	10–14 to 85+
Syphilis	Syphilis	5–9 to 10–14
Thalass	Thalassaemias	5–9 to 5–9
Tuberc	Tuberculosis	1–4 to 85+
WhoopCough	Whooping cough	1–4 to 5–9

Discussion

This thesis has developed new Bayesian methodologies for modeling the complex dependence structures characterizing demographic data, with a focus on dynamic clustering across time, age, and population groups. The central theme has been the identification of evolving communities of homogeneous statistical units, such as countries or causes of death, through probabilistic models that allow clusters to change smoothly along ordered dimensions. While the empirical focus has been on mortality-related phenomena, each methodological contribution addresses modeling challenges that are equally relevant in other areas of demography. For example, the models introduced in the three chapters could be applied respectively to fertility rate trajectories, comorbidity networks, and life-course sequences

The methodological foundation of this work lies in the temporal random partition model (tRPM) of Page et al. (2022), which introduces temporal dependence among partitions while preserving the marginal properties of classical random partition models. This framework enables a coherent treatment of evolving group structures and provides a flexible tool for studying how similarities among demographic units persist or change over time.

In Chapter 1, we proposed a dynamic clustering model for log-mortality rates that captures local similarities among countries across ages and calendar years. By representing mortality surfaces through B-spline expansions and clustering their coefficients dynamically, the model handles the functional nature and high dimensionality of the data while uncovering persistent and transient regional patterns in mortality evolution.

Chapter 2 generalized the class of extended stochastic block models (ESBM) to analyze sequences of directed, categorically weighted networks. Applied to the dynamic co-occurrence patterns of underlying and contributing causes of death in the US, this model captured the asymmetric and age-dependent relationships among causes.

In Chapter 3, we introduced a Bayesian framework for clustering countries based on their profiles of leading causes of death over time. Modeling these categorical vectors through mixtures with exponential-distance components and embedding them in a dynamic partition process enabled the identification of stable and evolving clusters of countries, revealing global epidemiological convergence and divergence from 2000 to 2020.

These projects open several avenues for future research. First, the discussions in the three chapters highlight the potential impact of limited high-quality data on the resulting analyses. Improvements in data availability and granularity would allow for a deeper understanding of specific aspects of mortality trends. For example, future work could investigate the relationship between mortality rates and socioeconomic indicators in Chapter 1, or provide a more detailed decomposition of deaths by cause in Chapter 3. Another promising direction would be to incorporate additional sources of dependence when modeling clustering patterns among countries and causes of death. For instance, one could examine the dependence between the partitions of male and female populations in Chapters 1 and 3, or extend the model for sequences of directed networks in Chapter 2 to include temporal dependence as well.

Lastly, scalability is often a challenge for Bayesian procedures, particularly for nonparametric models used to infer partitions. Implementing these models requires substantial computational effort via principled Markov chain Monte Carlo (MCMC) algorithms. Improving the sampling schemes would certainly open avenues for applying these models to larger problems, whether due to increased sample sizes or an higher number of parameters, which would further enhance the flexibility of such models. Future directions could consider either improved MCMC sampling schemes targeting the exact posterior distribution or approximations of the posterior distributions, such as those provided by variational inference or divide-and-conquer procedures. Recent research moved into this directions, for instance deriving block-update schemes for the cluster assignments (Franzolini et al., 2023; Ricci et al., 2024; Argiento et al., 2024). Nevertheless, the current approaches require nontrivial adaptations to be applied to the models proposed in the thesis. One promising option consists in adopting split-and-merge procedures (e.g., Jain and Neal, 2004), which improve the exploration of the domain of the posterior distribution of the partitions. This is achieved by introducing in the MCMC updates some splits and merges of clusters, which attempt to modify the partition draw of the last iteration through global moves rather than individual cluster assignments. The detailed-balance equation is preserved through Metropolis-Hastings steps, which can accept or reject the proposed split/merge. Extending this strategy to a dynamic partition is challenging due to the dependence between partitions at consecutive time points. In particular, the *compatibility constraints* underlying the tRPM by Page et al. (2022) would be violated by most split/merge proposals, rendering them highly unlikely to be accepted.

Bibliography

Abbott, A. and Forrest, J. (1986). Optimal matching methods for historical sequences. *The Journal of Interdisciplinary History*, 16(3):471–494.

Aburto, J. M., Wensink, M., van Raalte, A., and Lindahl-Jacobsen, R. (2018). Potential gains in life expectancy by reducing inequality of lifespans in Denmark: an international comparison and cause-of-death analysis. *BMC Public Health*, 18(1):831.

Ahmad, F. B., Cisewski, J. A., and Anderson, R. N. (2024). Leading causes of death in the US, 2019-2023. *JAMA*, 332(12):957–958.

Alexopoulos, A., Dellaportas, P., and Forster, J. J. (2019). Bayesian forecasting of mortality rates by using latent Gaussian models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):689–711.

Aliverti, E., Mazzuco, S., and Scarpa, B. (2022). Dynamic modelling of mortality via mixtures of skewed distribution functions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3):1030–1048.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174.

Antonio, K., Bardoutsos, A., and Ouburg, W. (2015). Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal*, 5:245–281.

Argiento, R., Filippi-Mazzola, E., and and, L. P. (2024). Model-based clustering of categorical data based on the Hamming distance. *Journal of the American Statistical Association*, 0(0):1–23.

Arnold (-Gaille), S. and Sherris, M. (2013). Forecasting mortality trends allowing for cause-of-death mortality dependence. *North American Actuarial Journal*, 17(4):273–282.

Aschner, P., Aguilar-Salinas, C., Aguirre, L., Franco, L., Gagliardino, J. J., de Lapertosa, S. G., Seclen, S., and Vinocour, M. (2014). Diabetes in South and Central America: An update. *Diabetes Research and Clinical Practice*, 103(2):238–243.

Ascolani, F., Lijoi, A., Rebaudo, G., and Zanella, G. (2023). Clustering consistency with Dirichlet process mixtures. *Biometrika*, 110(2):551–558.

Barrientos-Gutierrez, T., Zepeda-Tello, R., Rodrigues, E. R., Colchero-Aragonés, A., Rojas-Martínez, R., Lazcano-Ponce, E., Hernández-Ávila, M., Rivera-Dommarco, J., and Meza, R. (2017). Expected population weight and diabetes impact of the 1-peso-per-litre tax to sugar sweetened beverages in Mexico. *PLOS ONE*, 12(5):1–15.

Bergeron-Boucher, M.-P., Aburto, J. M., and van Raalte, A. (2020). Diversification in causes of death in low-mortality countries: emerging patterns and implications. *BMJ Global Health*, 5(7):1–12.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika*, 65(1):31–38.

Bishop, K., Balogun, S., Eynstone-Hinkins, J., Moran, L., Martin, M., Banks, E., Rao, C., and Joshy, G. (2023). Analysis of multiple causes of death: review of methods and practices. *Epidemiology*, 34(3):333–344.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, 3(1-2):3–43.

Bray, F., Laversanne, M., Weiderpass, E., and Soerjomataram, I. (2021). The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*, 127(16):3029–3030.

Calazans, J. A. and Permanyer, I. (2023). Levels, trends, and determinants of cause-of-death diversity in a global perspective: 1990–2019. *BMC Public Health*, 23(1):650.

Camarda, C. G. (2019). Smooth constrained mortality forecasting. *Demographic Research*, 41:1091–1130.

Canudas-Romo, V. (2010). Three measures of longevity: time trends and record values. *Demography*, 47(2):299–312.

Canudas-Romo, V., Adair, T., and Mazzuco, S. (2020). Reflection on modern methods: cause of death decomposition of cohort survival comparisons. *International Journal of Epidemiology*, 49(5):1712–1718.

Case, A. and Deaton, A. (2020). The epidemic of despair: Will America's mortality crisis spread to the rest of the world? *Foreign Affairs*, 99(2):pp. 92–102.

Case, A. and Deaton, A. (2021). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

Caselli, G., Meslé, F., and Vallin, J. (2002). Epidemiologic transition theory exceptions. *Genus*, 58(1):9–51.

Celeux, G. and Govaert, G. (1991). Clustering criteria for discrete data and latent class models. *Journal of Classification*, 8:157–176.

Chmiel, A., Klimek, P., and Thurner, S. (2014). Spreading of diseases through comorbidity networks across life and gender. *New Journal of Physics*, 16(11):115013.

Currie, I. D. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4):356–383.

Currie, I. D., Durban, M., and Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298.

Dahl, D. B., Johnson, D. J., and and, P. M. (2022). Search algorithms and loss functions for Bayesian clustering. *Journal of Computational and Graphical Statistics*, 31(4):1189–1201.

Dattani, S., Spooner, F., Ritchie, H., and Roser, M. (2023). Causes of death. *Our World in Data*. https://ourworldindata.org/causes-of-death.

Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*, 56:463–474.

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):212–229.

De Blasi, P., Lijoi, A., and Prünster, I. (2013). An asymptotic analysis of a class of discrete nonparametric priors. *Statistica Sinica*, 23(3):1299–1321.

De Jong, P. and Tickle, L. (2006). Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, 13(1):1–18.

Debon, A., Haberman, S., and Piscopo, G. (2023). Multipopulation mortality analysis: bringing out the unobservable with latent clustering. *Quality & Quantity*, 58:1–17.

Depaoli, E. G., Stefanucci, M., and Mazzuco, S. (2024). Functional concurrent regression with compositional covariates and its application to the time-varying effect of causes of death on human longevity. *The Annals of Applied Statistics*, 18(2):1668–1685.

Désesquelles, A., Demuru, E., Egidi, V., Frova, L., Meslé, F., Pappagallo, M., and Salvatore, M. A. (2014a). Cause-specific mortality analysis: is the underlying cause of death sufficient? *Quetelet Journal*, 2(1):119–135.

Désesquelles, A., Demuru, E., Salvatore, M. A., Pappagallo, M., Frova, L., Meslé, F., and Egidi, V. (2014b). Mortality from Alzheimer's disease, Parkinson's disease, and dementias in France and Italy: a comparison using the multiple cause-of-death approach. *Journal of Aging and Health*, 26(2):283–315.

Désesquelles, A., Salvatore, M. A., Frova, L., Pace, M., Pappagallo, M., Meslé, F., and Egidi, V. (2010). Revisiting the mortality of France and Italy with the multiple-cause-of-death approach. *Demographic Research*, 23:771–806.

Désesquelles, A. F., Salvatore, M. A., Pappagallo, M., Frova, L., Pace, M., Meslé, F., and Egidi, V. (2012). Analysing multiple causes of death: which methods for which data? an application to the cancer-related mortality in France and Italy. *European Journal of Population*, 28(4):467–498.

DesRoches, R., Comerio, M., Eberhard, M., Mooney, W., and Rix, G. J. (2011). Overview of the 2010 Haiti earthquake. *Earthquake Spectra*, 27(1_suppl1):1–21.

Dharamshi, A., Alexander, M., Barbieri, M., and Winant, C. (2025). Jointly estimating subnational mortality for multiple populations. *Demographic Research*, 52(3):71–110.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2):363–375.

Dimai, M. (2025). Clustering of mortality paths with the Hellinger distance and visualization through the distatis technique. *Statistical Methods & Applications*, pages 1–40. In press.

Dong, Y., Huang, F., Yu, H., and Haberman, S. (2020). Multi-population mortality forecasting using tensor decomposition. *Scandinavian Actuarial Journal*, 2020(8):754–775.

Durante, D., Gaffi, F., Lijoi, A., and Prünster, I. (2025). Partially exchangeable stochastic block models for (node-colored) multilayer networks. *Journal of the American Statistical Association*, page In press.

Earl, D. J. and Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916.

Eberstein, I. W., Nam, C. B., and Hummer, R. A. (1990). Infant mortality by cause of death: main and interaction effects. *Demography*, 27:413–430.

Egidi, V., Salvatore, M. A., Rivellini, G., and D'Angelo, S. (2018). A network approach to studying cause-of-death interrelations. *Demographic Research*, 38:373–400.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121.

Enchev, V., Kleinow, T., and Cairns, A. J. (2017). Multi-population mortality models: fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, 2017(4):319–342.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588.

European Commission and Executive Agency for Health and Consumers (2013). *Health Inequalities in the EU – Final Report of a Consortium – Consortium Lead: Sir Michael Marmot*. European Commission.

European Heart Journal (2016). Cancer overtakes cardiovascular disease as the main cause of death in 12 European Union countries. *European Heart Journal*, 37(42):3183–3184.

Feraldi, A. and Zarulli, V. (2022). Patterns in age and cause of death contribution to the sex gap in life expectancy: a comparison among ten countries. *Genus*, 78(1):23.

Foreman, K. J., Marquez, N., Dolgert, A., Fukutaki, K., Fullman, N., McGaughey, M., Pletcher, M. A., Smith, A. E., Tang, K., Yuan, C.-W., et al. (2018). Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet*, 392(10159):2052–2090.

Fotouhi, B., Momeni, N., Riolo, M. A., and Buckeridge, D. L. (2018). Statistical methods for constructing disease comorbidity networks from longitudinal inpatient data. *Applied Network Science*, 3(1):46.

Franzolini, B., De Iorio, M., and Eriksson, J. (2023). Conditional partial exchangeability: a probabilistic framework for multi-view clustering. *arXiv:2307.01152*.

Fruhwirth-Schnatter, S., Celeux, G., and Robert, C. P. (2015). *Handbook of Mixture Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis.

Geng, J., Bhattacharya, A., and Pati, D. (2019). Probabilistic community detection with unknown number of communities. *Journal of the American Statistical Association*, 114:893–905.

Gershman, S. J. and Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1–12.

Gini, C. (1912). *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche.[Fasc. I.]*. Tipogr. di P. Cuppini.

Glei, D. A. (2022). The US midlife mortality crisis continues: excess cause-specific mortality during 2020. *American Journal of Epidemiology*, 191(10):1677–1686.

Gnedin, A. (2010). A species sampling model with finitely many types. *Electronic Communications in Probability*, 15:79–88.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2):215–231.

Goto, S., Takagishi, M., and Yadohisa, H. (2021). Clustering for time-varying relational count data. *Computational Statistics & Data Analysis*, 156:107123.

Grazian, C. (2025). Advances in Bayesian random partition models: A comprehensive review. *arXiv:2303.17182*.

Grippo, F., Désesquelles, A., Pappagallo, M., Frova, L., Egidi, V., and Meslé, F. (2020). Multi-morbidity and frailty at death: a new classification of death records for an ageing world. *Population Studies*, 74(3):437–449.

Grippo, F., Frova, L., Pappagallo, M., Barbieri, M., Trias-Llimós, S., Egidi, V., Meslé, F., and Désesquelles, A. (2024). Beyond the underlying cause of death: an algorithm to study multi-morbidity at death. *Population Health Metrics*, 22(1):36.

Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1):35–55.

Hall, M., Lee, A., Cartwright, C., Marahatta, S., Karki, J., and Simkhada, P. (2017). The 2015 Nepal earthquake disaster: lessons learned one year on. *Public Health*, 145:39–44.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160.

Hartigan, J. A. (1990). Partition models. *Communications in Statistics – Theory and methods*, 19(8):2745–2756.

Hatzopoulos, P. and Haberman, S. (2013). Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics*, 52(2):320–337.

Hawton, K., Saunders, K. E., and O'Connor, R. C. (2012). Self-harm and suicide in adolescents. *The Lancet*, 379(9834):2373–2382.

Hennig, C., Meila, M., Murtagh, F., and Rocci, R. (2015). *Handbook of Cluster Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis.

Heron, M. P. (2021). Deaths: Leading causes for 2018. Technical Report vol. 70 no. 4, National Center for Health Statistics (U.S.).

Hidalgo, C. A., Blumm, N., Barabási, A.-L., and Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. *PLoS Computational Biology*, 5(4):e1000353.

Ho, J. Y. and Hendi, A. S. (2018). Recent trends in life expectancy across high income countries: retrospective observational study. *BMJ*, 362.

Højstrup, S., Thomsen, J. H., and Prescott, E. (2023). Disparities in cardiovascular disease and treatment in the nordic countries. *The Lancet Regional Health – Europe*, 33.

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*, 2:193–218.

Human Mortality Database (2024). University of California Berkeley, USA and Max Planck Institute for Demographic Research, Germany.

Hunt, A. and Blake, D. (2021). On the structure and classification of mortality models. *North American Actuarial Journal*, 25(sup1):S215–S234.

Huynh, N. and Ludkovski, M. (2024). Joint models for cause-of-death mortality in multiple populations. *Annals of Actuarial Science*, 18(1):51–77.

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50:261–283.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.

Inglehart, R. (2006). Mapping global values. *Comparative Sociology*, 5(2-3):115–136.

Ishiguro, K., Iwata, T., Ueda, N., and Tenenbaum, J. (2010). Dynamic infinite relational model for time-varying relational data analysis. *Advances in Neural Information Processing Systems*, 23:1–9.

Israel, R. A., Rosenberg, H. M., and Curtin, L. R. (1986). Analytical potential for multiple cause-of-death data. *American Journal of Epidemiology*, 124(2):161–179.

Jain, S. and Neal, R. M. (2004). A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1):158–182.

Jani, C., Marshall, D. C., Singh, H., Goodall, R., Shalhoub, J., Al Omari, O., Salciccioli, J. D., and Thomson, C. C. (2021). Lung cancer mortality in Europe and the USA between 2000 and 2017: an observational analysis. *ERJ Open Research*, 7(4).

Jemal, A., Ward, E., Hao, Y., and Thun, M. (2005). Trends in the leading causes of death in the United States, 1970-2002. *JAMA*, 294(10):1255–1259.

Jeong, E., Ko, K., Oh, S., and Han, H. W. (2017). Network-based analysis of diagnosis progression patterns using claims data. *Scientific Reports*, 7(1):15561.

Jones, I., Cocker, F., Jose, M., Charleston, M., and Neil, A. L. (2023). Methods of analysing patterns of multimorbidity using network analysis: a scoping review. *Journal of Public Health*, 31(8):1217–1223.

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the 21st National Conference on Artificial Intelligence – Volume 1*, pages 381–388.

Kim, B., Lee, K. H., Xue, L., and Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, 12:105.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63:147–152.

Knipe, D., Padmanathan, P., Newton-Howes, G., Chan, L. F., and Kapur, N. (2022). Suicide and self-harm. *The Lancet*, 399(10338):1903–1916.

Korenjak-Černe, S. and Kejžar, N. (2022). Identifying patterns of main causes of death in the young EU population. *arXiv:2210.04469*.

Kvalseth, T. O. (1987). Entropy and correlation: Some comments. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(3):517–519.

Lam, K. K. and Wang, B. (2023). Multipopulation mortality modelling and forecasting: the weighted multivariate functional principal component approaches. *Journal of Applied Statistics*, 50(15):3177–3198.

Lavine, M. and West, M. (1992). A Bayesian method for classification and discrimination. *The Canadian Journal of Statistics*, 20(4):451–461.

Lee, C. and Wilkinson, D. J. (2019). A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1):1–50.

Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4):537–549.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Léger, A.-E. and Mazzuco, S. (2021). What can we learn from the functional clustering of mortality data? An application to the Human Mortality Database. *European Journal of Population*, 37:769–798.

Legramanti, S., Rigon, T., Durante, D., and Dunson, D. B. (2022). Extended stochastic block models with application to criminal networks. *The Annals of Applied Statistics*, 16(4):2369–2395.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42:575–594.

Li, N., Lee, R., and Gerland, P. (2013). Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, 50(6):2037–2051.

Lindahl-Jacobsen, R., Rau, R., Jeune, B., Canudas-Romo, V., Lenart, A., Christensen, K., and Vaupel, J. W. (2016). Rise, stagnation, and rise of Danish women's life expectancy. *Proceedings of the National Academy of Sciences*, 113(15):4015–4020.

Lopez-Jaramillo, P., Lopez-Lopez, J., Cohen, D., Alarcon-Ariza, N., and Mogollon-Zehr, M. (2021). Epidemiology of hypertension and diabetes mellitus in Latin America. *Current Hypertension Reviews*, 17(2):112–120.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S. Y., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380(9859):2095–2128.

Lutz, W. and Kc, S. (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554):2779–2791.

MacDorman, M. F., Hoyert, D. L., and Mathews, T. (2013). *Recent Declines in Infant Mortality in the United States, 2005-2011.* US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Mackenbach, J. P. (2017). Nordic paradox, Southern miracle, Eastern disaster: persistence of inequalities in mortality in Europe. *The European Journal of Public Health*, 27(suppl_4):14–17.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297. University of California press.

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.

Marmot, M. (2005). Social determinants of health inequalities. *The Lancet*, 365(9464):1099–1104.

Mathers, C. D., Boerma, T., and Ma Fat, D. (2009). Global and regional causes of death. *British Medical Bulletin*, 92:7–32.

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.

Maxwell, D. and Fitzpatrick, M. (2012). The 2011 Somalia famine: Context, causes, and complications. *Global Food Security*, 1(1):5–12. Special Issue on the Somalia Famine of 2011-2012.

Mazzuco, S., Scarpa, B., and Zanotto, L. (2018). A mortality model based on a mixture distribution function. *Population Studies*, 72(2):191–200.

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

Mehta, N. K., Abrams, L. R., and Myrskylä, M. (2020). US life expectancy stalls due to cardiovascular disease, not drug deaths. *Proceedings of the National Academy of Sciences*, 117(13):6998–7000.

Meilă, M. (2007). Comparing clusterings – an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895.

Mercer, A. J. (2018). Updating the epidemiological transition model. *Epidemiology and Infection*, 146(6):680–687.

Meslé, F. (2004). Écart d'espérance de vie entre les sexes: les raisons du recul de l'avantage féminin. *Revue d'Épidémiologie et de Santé Publique*, 52(4):333–352. "Genres et Santé".

Meza, R., Barrientos-Gutierrez, T., Rojas-Martinez, R., Reynoso-Noveron, N., Palacio-Mejia, L. S., Lazcano-Ponce, E., and Hernandez-Avila, M. (2015). Burden of type 2 diabetes in Mexico: past, current and future prevalence and incidence rates. *Preventive Medicine*, 81:445–450.

Miasojedow, B., Moulines, E., and Vihola, M. (2013). An adaptive parallel tempering algorithm. *Journal of Computational and Graphical Statistics*, 22(3):649–664.

Moreno-Betancur, M., Sadaoui, H., Piffaretti, C., and Rey, G. (2017). Survival analysis with multiple causes of death: extending the competing risks model. *Epidemiology*, 28(1):12–19.

Müller, P., Quintana, F., and Rosner, G. L. (2011). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics*, 20(1):260–278.

Murray, C. J. L. and Evans, D. B. & World Health Organization. Global Programme on Evidence for Health Policy. (2003). *Health systems performance assessment : debates, methods and empiricism / edited by Christopher J. L. Murray, David B. Evans*. World Health Organization.

Newman, M. E. and Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23):9564–9569.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087.

OECD (2024). Data warehouse. https://data-explorer.oecd.org/.

Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570):1029–1031.

Omran, A. R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. *The Milbank Memorial Fund Quarterly*, 49(4):509–538.

O'Neill, T. J. (1978). Normal discrimination with unclassified observations. *Journal of the American Statistical Association*, 73(364):821–826.

Page, G. L., Quintana, F. A., and Dahl, D. B. (2022). Dependent modeling of temporal sequences of random partitions. *Journal of Computational and Graphical Statistics*, 31(2):614–627.

Papanicolas, I., Rajan, D., Karanikolos, M., Soucat, A., and Figueras, J., editors (2022). *Health System Performance Assessment: a Framework for Policy Analysis.* European Observatory on Health Systems and Policies.

Pavone, F., Legramanti, S., and Durante, D. (2024). Learning and forecasting of age-specific period mortality via B-spline processes with locally-adaptive dynamic coefficients. *The Annals of Applied Statistics*, 18(3):1965 – 1987.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110.

Peixoto, T. P. (2018). Nonparametric weighted stochastic block models. *Physical Review E*, 97(1):012306.

Peixoto, T. P. (2022). Ordered community detection in directed networks. *Physical Review E*, 106(2):024305.

Pensky, M. and Zhang, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics*, 13(1):678–709.

Preston, S. H. and Wang, H. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography*, 43(4):631–646.

Raftery, A. E., Chunn, J. L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50(3):777–801.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850.

Rebaudo, G., Lin, Q., and Mueller, P. (2025). Separate exchangeability as modeling principle in Bayesian nonparametrics. *arXiv:2112.07755.*

Redelings, M. D., Sorvillo, F., and Simon, P. (2006). A comparison of underlying cause and multiple causes of death: US vital statistics, 2000–2001. *Epidemiology*, 17(1):100–103.

Redelings, M. D., Wise, M., and Sorvillo, F. (2007). Using multiple cause-of-death data to investigate associations and causality between conditions listed on the death certificate. *American Journal of Epidemiology*, 166(1):104–108.

Ricci, F. Z., Sudderth, E. B., Lee, J., Peters, M. A., Vannucci, M., and Guindani, M. (2024). Bayesian temporal biclustering with applications to multi-subject neuroscience studies. *arXiv:2406.17131.*

Rohe, K., Qin, T., and Yu, B. (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences*, 113(45):12679–12684.

Romanò, G., Castiglione, C., and Durante, D. (2025). Dependent stochastic block models for age-indexed sequences of directed causes-of-death networks. *arXiv:2510.01806*.

Romanò, G., Aliverti, E., and Durante, D. (2025). Bayesian local clustering of age-period mortality surfaces across multiple countries. *arXiv:2504.05240*.

Roser, M., Ritchie, H., and Spooner, F. (2021). Burden of disease. *Our World in Data*. https://ourworldindata.org/burden-of-disease.

Saadat, S., Yousefifard, M., Asady, H., Moghadas Jafari, A., Fayaz, M., and Hosseini, M. (2015). The most important causes of death in Iranian population: A retrospective cohort study. *Emergency (Tehran)*, 3(1):16–21. PMID: 26512364; PMCID: PMC4614603.

Santosa, A., Wall, S., Fottrell, E., Högberg, U., and and, P. B. (2014). The development and experience of epidemiological transition theory over four decades: a systematic review. *Global Health Action*, 7(1):23574. PMID: 24848657.

Schmidt, M. N. and Morup, M. (2013). Nonparametric Bayesian modeling of complex networks: An introduction. *IEEE Signal Processing Magazine*, 30:110–128.

Schnürch, S., Kleinow, T., and Korn, R. (2021). Clustering-based extensions of the common age effect multi-population mortality model. *Risks*, 9(3):45.

Scognamiglio, S. (2022). A multi-population locally-coherent mortality model. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pages 423–428. Springer.

Sinisterra, L. I., Cardelle, A., Abraham, A., MN, C., MA, E., gonzález olivares, L., Gotteland, M., A, P.-S., Martin, S., A, C., JJ, G., and IBEROBDIA, P. (2019). Diabetes in Latin America: Prevalence, complications, and socio-economic impact. *International Journal of Diabetes and Clinical Research*, 6.

Sisson, S. A., Fan, Y., and Beaumont, M., editors (2018). *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC, 1 edition.

Stefanucci, M. and Mazzuco, S. (2022). Analysing cause-specific mortality trends using compositional functional data analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(1):61–83.

Stern, D., Mazariegos, M., Ortiz-Panozo, E., Campos, H., Malik, V. S., Lajous, M., and López-Ridaura, R. (2019). Sugar-sweetened soda consumption increases diabetes risk among Mexican women. *The Journal of Nutrition*, 149(5):795–803.

Strehl, A. and Ghosh, J. (2002). Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(Dec):583–617.

Syed, S., Bouchard-Côté, A., Deligiannidis, G., and Doucet, A. (2022). Non-reversible parallel tempering: a scalable highly parallel MCMC scheme. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 84(2):321–350.

Syed, S., Romaniello, V., Campbell, T., and Bouchard-Côté, A. (2021). Parallel tempering on optimized paths. In *International Conference on Machine Learning*, pages 10033–10042. PMLR.

Tallberg, C. (2004). A Bayesian approach to modeling stochastic blockstructures with covariates. *Journal of Mathematical Sociology*, 29(1):1–23.

Tanday, S. (2016). Cancer overtakes heart disease as biggest killer in 12 countries. *The Lancet Oncology*, 17(9):e382.

Tang, C., Shang, H. L., and Yang, Y. (2022). Clustering and forecasting multiple functional time series. *The Annals of Applied Statistics*, 16(4):2523–2553.

Townsend, N., Wilson, L., Bhatnagar, P., Wickramasinghe, K., Rayner, M., and Nichols, M. (2016). Cardiovascular disease in Europe: epidemiological update 2016. *European Heart Journal*, 37(42):3232–3245.

Trias-Llimós, S. and Permanyer, I. (2023). Cause-of-death diversity from a multiple-cause perspective in the United States. *Demography*, 60(1):73–98.

Ukolova, E. and Burcin, B. (2023). Racial/ethnic disparities in the chains of morbid events leading to death: network analysis of us multiple cause of death data. *Biodemography and Social Biology*, 68(4):149–165.

Vallin, J. and Meslé, F. (2004). Convergences and divergences in mortality: a new approach of health transition. *Demographic Research*, 2:11–44.

Van Raalte, A. A. (2021). What have we learned about mortality patterns over the past 25 years? *Population Studies*, 75(1):105–132.

Varga, L. (2025). Forecasting cause-of-death mortality with single- and multi-population models in Hungary. *Scandinavian Actuarial Journal*, 0(0):1–32.

Vaupel, J. W., Zhang, Z., and van Raalte, A. A. (2011). Life expectancy and disparity: an international comparison of life table data. *BMJ Open*, 1(1):e000128.

Wade, S. and Ghahramani, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis*, 13(2):559–626.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*. MIT Press Cambridge, MA.

Wilson, C. (2011). Understanding global demographic convergence since 1950. *Population and Development Review*, 37(2):375–388.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5(3):329–350.

Wong, J. S., Forster, J. J., and Smith, P. W. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, 83:206–221.

Woolf, S. H. and Schoomaker, H. (2019). Life expectancy and mortality rates in the United States, 1959-2017. *JAMA*, 322(20):1996–2016.

World Bank (2025). The world by income and region. https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html. Accessed: 2025-06-07.

World Health Organization (2016). *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*, volume 2. World Health Organization.

World Health Organization (2024a). Global Health Estimates 2021: Deaths by cause, age, sex, by country and by region, 2000-2021. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death.

World Health Organization (2024b). WHO methods and data sources for country-level causes of death 2000–2021. Global Health Estimates Technical Paper WHO/DDI/DNA/GHE/2024.2, World Health Organization, Department of Data and Analytics, Division of Data, Analytics and Delivery for Impact, Geneva.

World Health Organization (2025). *International Classification of Diseases, Eleventh Revision (ICD-11)*. World Health Organization, Geneva. License: CC BY-ND 3.0 IGO.

Wyse, J. and Friel, N. (2012). Block clustering with collapsed latent block models. *Statistics and Computing*, 22(2):415–428.

Xu, K. (2015). Stochastic block transition models for dynamic networks. In *Artificial Intelligence and Statistics*, pages 1079–1087. PMLR.

Xu, K. S. and Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):552–562.

Yang, T., Chi, Y., Zhu, S., Gong, Y., and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks – a Bayesian approach. *Machine Learning*, 82(2):157–189.

Zhang, J., He, X., and Wang, J. (2022). Directed community detection with network embedding. *Journal of the American Statistical Association*, 117(540):1809–1819.

Zhang, Y., Levina, E., and Zhu, J. (2016). Community detection in networks with node features. *Electronic Journal of Statistics*, 10(2):3153–3178.