

Dependent Stochastic Block Models for Age-Indexed Sequences of Directed Causes-of-Death Networks

Giovanni Romanò[†], Cristian Castiglione^{*} and Daniele Durante^{*,†}

^{*}Bocconi Institute for Data Science and Analytics, Bocconi University

[†]Department of Decision Sciences, Bocconi University

Abstract

Death events commonly arise from complex interactions among interrelated causes, formally classified in reporting practices as *underlying* and *contributing*. Leveraging information from death certificates, these interactions can be naturally represented through a sequence of directed networks encoding co-occurrence strengths between pairs of underlying and contributing causes across ages. Although this perspective opens the avenues to learn informative age-specific block interactions among endogenous groups of underlying and contributing causes displaying similar co-occurrence patterns, there has been limited research along this direction in mortality modeling. This is mainly due to the lack of suitable stochastic block models for sequences of directed networks indexed by a predictor. We cover this gap through a novel Bayesian formulation which crucially learns two separate group structures for underlying and contributing causes, while allowing both structures to change smoothly across ages via dependent random partition priors. As illustrated in simulation studies, the proposed formulation outperforms state-of-the-art solutions that could be adapted to our motivating application. Moreover, when applied to USA mortality data, it unveils structures in the composition, evolution, and modular interactions among causes-of-death groups that were hidden to classical demographic studies. Such findings could have relevant policy implications and contribute to an improved understanding of the recent “death of despair” phenomena in the USA.

Keywords: Causes of Death; Dependent Random Partition Prior; Directed Network; Mortality Modeling; Stochastic Block Model

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](#); [Grippio 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”), have 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, 2014; 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., 2014; Trias-Llimós and Permanyer, 2023). As a consequence, MCODE 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, 2014), this network perspective has been mostly overlooked in MCODE 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 MCODE 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 MCODE 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 [1.1](#) and [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 [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 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 2–3, 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 4, where the model we propose in Sections 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 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 6.

1.1 USA Causes-of-Death Data

As anticipated in Section 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, 2014](#); [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 6.

Network sequence over selected age classes for the male population

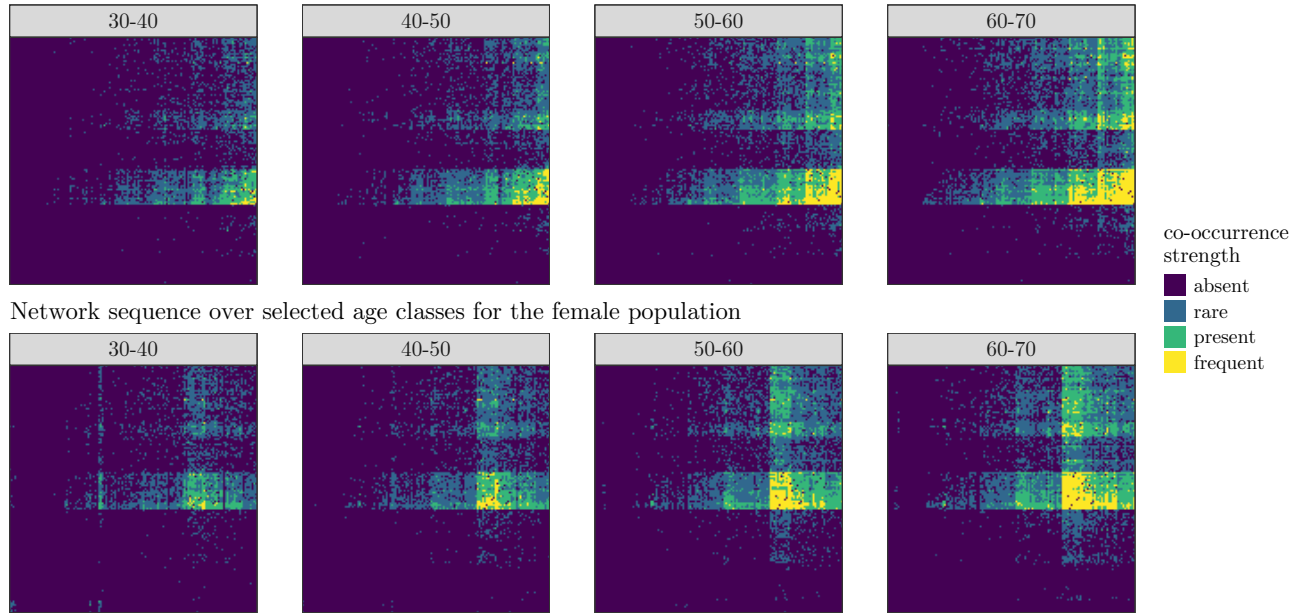


Figure 1: For the male and female populations, graphical representations of the adjacency matrices associated with the cause-of-death networks studied, at four selected consecutive age classes. In each matrix, rows (underlying causes) and columns (contributing causes) are re-ordered according to the 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.

Figure 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 2–3, 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 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 2–3. In fact, while the results in Figure 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 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, 2014; 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 and 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 Dependent Stochastic Block Models for Directed Networks

Define the four levels **absent**, **rare**, **present** and **frequent** introduced in Section 1.1 through the numerical label $c = 1, \dots, 4$. Moreover, recode the age classes $[0, 1]$, $(1, 10]$, $(10, 20]$, \dots , $(90, 100]$ into the ordered indexes $x = 1, 2, \dots, m$ (with $m = 11$ in our application). Then, the networks analyzed are available in the form of a sequence $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ of $n \times n$ categorically-weighted asymmetric adjacency matrices, where the generic \mathbf{Y}_x has entries $y_{ijx} = c$ 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 c .

In Section 2.1, we generalize state-of-the-art SBMs to define a probabilistic generative mechanism for $\mathbf{Y}_1, \dots, \mathbf{Y}_m$, which addresses points (i)–(v) discussed in Section 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.

2.1 Model Formulation

Guided by the empirical evidence in Figure 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 $\mathbf{Y}_1, \dots, \mathbf{Y}_m$. Recalling Section 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 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 $x = 1, \dots, m$, 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 $\mathbf{z}_{(1)x} = (z_{(1)1x}, \dots, z_{(1)nx})$ and $\mathbf{z}_{(2)x} = (z_{(2)1x}, \dots, z_{(2)K_x x})$ be the group membership vectors associated to the generic row and column partitions $\mathbb{Z}_{(1)x} = \{\mathbb{Z}_{(1)1x}, \dots, \mathbb{Z}_{(1)H_x x}\}$ and $\mathbb{Z}_{(2)x} = \{\mathbb{Z}_{(2)1x}, \dots, \mathbb{Z}_{(2)K_x x}\}$, such that $z_{(1)ix} = h$, for $h = 1, \dots, H_x$, if and only if $i \in \mathbb{Z}_{(1)hx}$ and, analogously, $z_{(2)jx} = k$, for $k = 1, \dots, K_x$, if and only if $j \in \mathbb{Z}_{(2)kx}$. Then, extending the classical Bernoulli likelihoods for binary edges

to the categorically-weighted interactions characterizing the networks under analysis, we assume that $(y_{ijx} \mid z_{(1)ix} = h, z_{(2)jx} = k, \boldsymbol{\theta}_{h k x}) \sim \text{Cat}_{1:4}(\boldsymbol{\theta}_{h k x})$, independently for every $i = 1, \dots, n$, $j = 1, \dots, n$ and $x = 1, \dots, m$, where $\text{Cat}_{1:4}(\boldsymbol{\theta}_{h k x})$ is a categorical variable indexed by the probability vector $\boldsymbol{\theta}_{h k x}$, with entries $\theta_{h k x c} = \text{pr}(y_{ijx} = c \mid z_{(1)ix} = h, z_{(2)jx} = k) \in (0, 1)$, for $c = 1, \dots, 4$. 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 $\mathbf{z}_{(1)x}$ and $\mathbf{z}_{(2)x}$, $x = 1, \dots, m$, together with the block-specific parameters $\boldsymbol{\theta}_{h k x}$, for $h = 1, \dots, H_x$, $k = 1, \dots, K_x$ and $x = 1, \dots, m$. 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 $\boldsymbol{\theta}_{h k x}$, $h = 1, \dots, H_x$, $k = 1, \dots, K_x$ and $x = 1, \dots, m$. 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 $\text{Cat}_{1:4}(\boldsymbol{\theta}_{h k x})$ 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 $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ conditioned on the group membership vectors $\mathbf{z}_{(1)x}$ and $\mathbf{z}_{(2)x}$, $x = 1, \dots, m$. In particular

$$p(\mathbf{Y}_1, \dots, \mathbf{Y}_m \mid (\mathbf{z}_{(1)x}, \mathbf{z}_{(2)x}), x = 1, \dots, m) = \prod_{x=1}^m \left(\prod_{h=1}^{H_x} \prod_{k=1}^{K_x} \left[\frac{\Gamma(a_\bullet^\theta)}{\Gamma(a_\bullet^\theta + n_{h k x \bullet})} \prod_{c=1}^4 \frac{\Gamma(a_c^\theta + n_{h k x c})}{\Gamma(a_c^\theta)} \right] \right), \quad (1)$$

where $n_{h k x c}$ represents the number of pairs (i, j) such that $z_{(1)ix} = h$, $z_{(2)jx} = k$ and $y_{ijx} = c$, while $n_{h k x \bullet} = \sum_{c=1}^4 n_{h k x c}$ and $a_\bullet^\theta = \sum_{c=1}^4 a_c^\theta$. The likelihood in (1) formally treats the quantities $\boldsymbol{\theta}_{h k x}$, $h = 1, \dots, H_x$, $k = 1, \dots, K_x$, $x = 1, \dots, m$ as nuisance parameters and focuses inference on the group structures encoded in $\mathbf{z}_{(1)x}$ and $\mathbf{z}_{(2)x}$, for $x = 1, \dots, m$. While learning the block-specific probability vectors $\boldsymbol{\theta}_{h k x}$ 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 $\mathbf{z}_{(1)x}$ and $\mathbf{z}_{(2)x}$, for $x = 1, \dots, m$, thereby motivating our main focus on these two vectors under the Dirichlet–categorical joint likelihood in (1). As discussed in Section 3, this choice further facilitates posterior computation, and does not prevent from obtaining ex–post sensible estimates of $\boldsymbol{\theta}_{h k x}$, $h = 1, \dots, H_x$, $k = 1, \dots, K_x$, $x = 1, \dots, m$.

Notice that the likelihood in (1) factorizes across the age–class indexes $x = 1, \dots, m$. Although this formulation is useful to avoid overly–sophisticated and intractable representations, as shown within Figure 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, we incorporate this dependence through carefully designed priors for the two sequences $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$ that further include external information on cause–of–death macro–classifications.

2.2 Prior for Dependent Sequences of Random Partitions

As anticipated in Section 2.1, we elicit priors for the two sequences $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$ 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 $\mathbf{w} = (w_1, \dots, w_n)$, whose generic categorical entry w_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 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

$$\begin{aligned} (\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m} \mid \mathbf{w}, \boldsymbol{\alpha}_{(1)}, \eta_{(1)}) &\sim \text{DRPM-w}(\boldsymbol{\alpha}_{(1)}, \eta_{(1)}), \\ (\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m} \mid \mathbf{w}, \boldsymbol{\alpha}_{(2)}, \eta_{(2)}) &\sim \text{DRPM-w}(\boldsymbol{\alpha}_{(2)}, \eta_{(2)}), \end{aligned} \tag{2}$$

where $\boldsymbol{\alpha}_{(1)}, \boldsymbol{\alpha}_{(2)}, \eta_{(1)}$ and $\eta_{(2)}$ are prior hyper–parameters whose meaning will be clarified in the following, whereas DRPM-w denotes the assumed dependent random partition prior informed by \mathbf{w} . Such a prior is presented in detail in Section 2.2.1 for the generic sequence $\mathbf{z}_1, \dots, \mathbf{z}_m$, where the indexes (1) and (2) associated with the two partitions under analysis are removed for the sake of generality.

2.2.1 Constructive Representation

The DRPM-w(α, η) prior for a generic sequence of group allocations $\mathbf{z}_1, \dots, \mathbf{z}_m$ is defined hierarchically through a Markovian mechanism, which combines two main constructions. The first one is a flexible, yet smooth, transition mechanism from \mathbf{z}_{x-1} to \mathbf{z}_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-w(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., \mathbf{z}_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 \mathbf{z}_{x-1} to \mathbf{z}_x , for each age $x = 2, \dots, m$.

Recalling classical results from Bayesian nonparametrics (see Gershman and Blei (2012) for an introductory overview) in connection with covariate-dependent product partition models (Müller et al., 2011), the assumed CRP-w(η) prior has probability mass function for the generic group membership vector \mathbf{z} defined as $p(\mathbf{z} \mid \eta, \mathbf{w}) \propto [\eta^H \Gamma(\eta) / \Gamma(\eta + n)] \prod_{h=1}^H \rho_{\text{DC}}(\mathbf{w}_h, \mathbf{a}^{\mathbf{w}}) (n_h - 1)!$. In this expression, $\Gamma(\cdot)$ corresponds to the Gamma function, n_h denotes the number of causes in group h , and \mathbf{w}_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 $p(\mathbf{z} \mid \eta, \mathbf{w})$ via the Dirichlet-categorical cohesion function $\rho_{\text{DC}}(\cdot, \mathbf{a}^{\mathbf{w}})$ with parameters $\mathbf{a}^{\mathbf{w}} = (a_1^{\mathbf{w}}, \dots, a_L^{\mathbf{w}})$ (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 $\rho_{\text{DC}}(\mathbf{w}_h, \mathbf{a}^{\mathbf{w}})$, the CRP-w(η) supervises the classical CRP(η) prior by the external covariate \mathbf{w} 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 z_i given $\mathbf{z}^{(-i)} = (z_1, \dots, z_{i-1}, z_{i+1}, z_n)$. Under the CRP-w(η) prior, this conditional distribution is

$$\text{pr}(z_i = h \mid \mathbf{z}^{(-i)}, \eta, \mathbf{w}) \propto \begin{cases} n_h^{(-i)} \frac{\bar{n}_{hw_i}^{(-i)} + a_{w_i}^{\mathbf{w}}}{\bar{n}_{h\bullet}^{(-i)} + a_{\bullet}^{\mathbf{w}}} & \text{for } h = 1, \dots, H^{(-i)}, \\ \eta \frac{a_{w_i}^{\mathbf{w}}}{a_{\bullet}^{\mathbf{w}}} & \text{for } h = H^{(-i)} + 1, \end{cases} \quad (3)$$

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 $\bar{n}_{hw_i}^{(-i)}$ corresponds to the number of

causes in group h with the same macro-category of the i -th one. Finally, $\bar{n}_{h\bullet}^{(-i)} = \sum_{l=1}^L \bar{n}_{hl}^{(-i)} = |\mathbf{w}_h^{(-i)}|$, $a_{\bullet}^{\mathbf{w}} = \sum_{l=1}^L a_l^{\mathbf{w}}$ and $a_{w_i}^{\mathbf{w}}$ is the entry of $\mathbf{a}^{\mathbf{w}}$ corresponding to the macro-category of cause i . According to (3), conditioned on $\mathbf{z}^{(-i)}$, 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., $\mathbf{a}^{\mathbf{w}} = (a_1^{\mathbf{w}}, \dots, a_L^{\mathbf{w}})$, with $a_l > 0$, $l = 1, \dots, L$). Consistent with (3), larger values for η yield a higher expected number of clusters in \mathbf{z} .

Besides illustrating the tractability of the CRP-w(η) prior, along with the meaning of its parameters, the scheme in (3) plays also a key role in the Gibbs sampling algorithm designed in Section 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-w(η) formulation with the previously-introduced Bernoulli transition mechanism yields the following hierarchical construction for the DRPM-w($\boldsymbol{\alpha}, \eta$) prior on the generic sequence $\mathbf{z}_1, \dots, \mathbf{z}_m$ of group membership vectors

$$\begin{aligned} (\gamma_{ix} \mid \boldsymbol{\alpha} = (\alpha_2, \dots, \alpha_m)) &\stackrel{\text{indep.}}{\sim} \text{Bernoulli}(\alpha_x), \quad i = 1, \dots, n, \quad x = 2, \dots, m, \\ (\mathbf{z}_1 \mid \eta, \mathbf{w}) &\sim \text{CRP-w}(\eta), \quad (\mathbf{z}_x \mid \mathbf{z}_{x-1}, \boldsymbol{\gamma}_x, \eta, \mathbf{w}) \sim \text{CRP-w}(\eta)_{|\mathcal{P}_{\mathbf{z}_{x-1}, \boldsymbol{\gamma}_x}}, \quad x = 2, \dots, m, \end{aligned} \tag{4}$$

where $\boldsymbol{\gamma}_x = (\gamma_{1x}, \dots, \gamma_{nx})$, while $\text{CRP-w}(\eta)_{|\mathcal{P}_{\mathbf{z}_{x-1}, \boldsymbol{\gamma}_x}}$ is the supervised Chinese restaurant process prior constrained to the set of partitions $\mathcal{P}_{\mathbf{z}_{x-1}, \boldsymbol{\gamma}_x}$ compatible with \mathbf{z}_{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 \mathbf{z}_{x-1} by reallocating only those causes for which $\gamma_{ix} = 0$. Recalling Page et al. (2022), under this constraint, $\text{pr}(\mathbf{z}_x = \mathbf{z} \mid \mathbf{z}_{x-1}, \boldsymbol{\gamma}_x, \eta, \mathbf{w}) \propto p(\mathbf{z} \mid \eta, \mathbf{w}) \mathbb{1}(\mathbb{Z} \in \mathcal{P}_{\mathbf{z}_{x-1}, \boldsymbol{\gamma}_x})$, where $p(\mathbf{z} \mid \eta, \mathbf{w})$ is the previously-

defined probability mass function of the CRP-w(η) prior, whereas \mathbb{Z} is the partition of the n causes associated with the generic group membership vector \mathbf{z} . As such, the prior on \mathbf{z}_x is anchored to \mathbf{z}_{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-w(η) prior, thereby removing the dependence on \mathbf{z}_{x-1} . Therefore, as α_x ranges from 0 to 1, the vectors \mathbf{z}_{x-1} and \mathbf{z}_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 $\alpha_2, \dots, \alpha_m$, which are assigned conjugate Beta hyperpriors, i.e., $\alpha_x \sim \text{Beta}(a_\alpha, b_\alpha)$, for $x = 2, \dots, m$. This allows smoothness to be inferred adaptively from the observed matrices $\mathbf{Y}_1, \dots, \mathbf{Y}_m$. 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 3, this prior elicitation facilitates the design of a tractable tempered Gibbs-sampler for posterior inference on the sequences of partitions $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$.

3 Posterior Computation and Inference

Posterior inference for the sequences of partitions $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$ that parameterize the Bayesian model presented in Section 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 3.1–3.2 by exploiting the tractability of the likelihood in (1) along with the hierarchical representation (4) of the priors in (2). Section 3.3, concludes by presenting useful inferential strategies to perform point estimation and uncertainty quantification on the group structures encoded in $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$ leveraging the Gibbs samples.

3.1 Collapsed Gibbs Sampler

The proposed Gibbs algorithm samples iteratively from the full-conditional distributions of the group membership vectors $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$, the persistency variables $\gamma_{(1)2}, \dots, \gamma_{(1)m}$ and $\gamma_{(2)2}, \dots, \gamma_{(2)m}$ appearing in the hierarchical representation (4) of the DRPM-w priors in (2), and finally,

the hyperparameters $\boldsymbol{\alpha}_{(1)}$, $\eta_{(1)}$, $\boldsymbol{\alpha}_{(2)}$, and $\eta_{(2)}$. As clarified in the following, by combining the priors presented in Section 2.2 with the collapsed likelihood in (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 $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$, $\boldsymbol{\gamma}_{(1)2}, \dots, \boldsymbol{\gamma}_{(1)m}$, $\boldsymbol{\alpha}_{(1)}$ and $\eta_{(1)}$, let us first introduce some useful notation. In particular, let $\Gamma_{(1)x} = \{i = 1, \dots, n : \gamma_{(1)ix} = 1\}$ be the set of underlying causes whose group allocation does not change from $x - 1$ to x . Moreover, denote with $\mathbb{Z}_{(1)x|\Gamma_{(1)x}}$ the partition induced by $\mathbf{z}_{(1)x}$ considering only those underlying causes with indexes in $\Gamma_{(1)x}$. Then, adapting the derivations in Page et al. (2022) to our construction, the full-conditional distributions for the binary persistency variables $\gamma_{(1)ix}$, $i = 1, \dots, n$, $x = 2, \dots, m$, are Bernoulli with probabilities

$$\text{pr}(\gamma_{(1)ix} = 1 \mid -) = \frac{\alpha_{(1)x}}{\alpha_{(1)x} + (1 - \alpha_{(1)x}) \frac{p(\mathbb{Z}_{(1)x|\Gamma_{(1)x}^{(+i)}} \mid \eta_{(1)}, \mathbf{w})}{p(\mathbb{Z}_{(1)x|\Gamma_{(1)x}^{(-i)}} \mid \eta_{(1)}, \mathbf{w})}} \mathbb{1}[\mathbb{Z}_{(1)x|\Gamma_{(1)x}^{(+i)}} = \mathbb{Z}_{(1)x-1|\Gamma_{(1)x}^{(+i)}}], \quad (5)$$

for each node $i = 1, \dots, n$ and age class $x = 2, \dots, m$, where $\mathbb{1}[\cdot]$ corresponds to the indicator function, whereas $\Gamma_{(1)x}^{(-i)} = \Gamma_{(1)x} \setminus \{i\}$ and $\Gamma_{(1)x}^{(+i)} = \Gamma_{(1)x} \cup \{i\}$. Notice that within the above expression the ratio $p(\mathbb{Z}_{(1)x|\Gamma_{(1)x}^{(+i)}} \mid \eta_{(1)}, \mathbf{w}) / p(\mathbb{Z}_{(1)x|\Gamma_{(1)x}^{(-i)}} \mid \eta_{(1)}, \mathbf{w})$ relates directly to the conditional distribution for $z_{(1)ix}$ given $\mathbf{z}_{(1)x}^{(-i)}$, $\eta_{(1)}$ and \mathbf{w} . As such, recalling also Page et al. (2022), it can be computed from (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 (3) are also useful to sample the allocations $z_{(1)ix}$ at each age class $x = 1, \dots, m$ for those underlying causes whose sampled $\gamma_{(1)ix}$ is equal to 0; if $\gamma_{(1)ix} = 1$, then $z_{(1)ix}$ is kept fixed to the group allocation drawn for cause i at $x - 1$. More specifically, denote with $\mathbb{Z}_{(1)x}^{z_{(1)ix}=h}$ the partition induced by the group membership vector $(z_{(1)1x}, \dots, z_{(1)ix} = h, \dots, z_{(1)nx})$ 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 $z_{(1)ix}$ with probabilities

$$\text{pr}(z_{(1)ix} = h \mid -) \propto \text{pr}(z_{(1)ix} = h \mid \mathbf{z}_{(1)x}^{(-i)}, \eta_{(1)}, \mathbf{w}) \mathbb{1}[\mathbb{Z}_{(1)x|\Gamma_{(1)x+1}}^{z_{(1)ix}=h} = \mathbb{Z}_{(1)x+1|\Gamma_{(1)x+1}}] \frac{p(\mathbf{Y}_x \mid z_{(1)ix} = h, \mathbf{z}_{(1)x}^{(-i)}, \mathbf{z}_{(2)x})}{p(\mathbf{Y}_x^{(-i)} \mid \mathbf{z}_{(1)x}^{(-i)}, \mathbf{z}_{(2)x})}, \quad (6)$$

for every $h = 1, \dots, H_x^{(-i)} + 1$, $i = 1, \dots, n$ and $x = 1, \dots, m$, where $\mathbf{Y}_x^{(-i)}$ corresponds to the ad-

jacency matrix at age class x without the i -th row. In the above expression, the prior probabilities $\text{pr}(z_{(1)ix} = h \mid \mathbf{z}_{(1)x}^{(-i)}, \eta_{(1)}, \mathbf{w})$ are available directly from (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 2.2.1 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 (1), also the likelihood factor within (6) admits the closed–form expression

$$\frac{p(\mathbf{Y}_x \mid z_{(1)ix} = h, \mathbf{z}_{(1)x}^{(-i)}, \mathbf{z}_{(2)x})}{p(\mathbf{Y}_x^{(-i)} \mid \mathbf{z}_{(1)x}^{(-i)}, \mathbf{z}_{(2)x})} = \prod_{k=1}^{K_x} \frac{\Gamma(a_{\bullet}^{\theta} + n_{h k x \bullet}^{(-i)})}{\Gamma(a_{\bullet}^{\theta} + n_{h k x \bullet})} \prod_{c=1}^4 \frac{\Gamma(a_c^{\theta} + n_{h k x c})}{\Gamma(a_c^{\theta} + n_{h k x c}^{(-i)}), \quad (7)$$

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

Given the persistency variables $\gamma_{(1)ix}$, $i = 1, \dots, n$, $x = 2, \dots, m$, also the prior hyperparameters $\alpha_{(1)x}$, $x = 2, \dots, m$ admit closed–form full conditional distributions. More specifically, combining the Bernoulli likelihood for each $\gamma_{(1)ix}$ with the Beta priors assumed in Section 2.2.1 for every $\alpha_{(1)x}$, yields

$$(\alpha_{(1)x} \mid -) \sim \text{Beta} \left(a_{\alpha} + \sum_{i=1}^n \gamma_{(1)ix}, b_{\alpha} + n - \sum_{i=1}^n \gamma_{(1)ix} \right), \quad (8)$$

independently for $x = 2, \dots, m$. Similarly–tractable full–conditionals are also available for the CRP concentration parameter $\eta_{(1)}$ via a direct application of the results in Escobar and West (1995).

To conclude, it remains to update the column–specific quantities $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$, $\gamma_{(2)2}, \dots, \gamma_{(2)m}$, $\alpha_{(2)}$ and $\eta_{(2)}$. 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 (5)–(8) by replacing the quantities and indexes related to the underlying causes with those of the contributing ones, and vice–versa.

3.2 Adaptive Parallel Tempering

Albeit tractable, the Gibbs routine presented in Section 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 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 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.

3.3 Posterior Inference

Posterior inference on the quantities of interest is performed via Monte Carlo, relying on the samples produced by the routine outlined in [Sections 3.1–3.2](#). As discussed in [Sections 1–2](#), our primary focus lies on the group membership structures encoded in the sequences $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)m}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)m}$, 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 $\hat{\mathbf{P}}_{(1)1}, \dots, \hat{\mathbf{P}}_{(1)m}$ and $\hat{\mathbf{P}}_{(2)1}, \dots, \hat{\mathbf{P}}_{(2)m}$, whose generic entries $\hat{\mathbf{P}}_{(1)x[i,i']}$ and $\hat{\mathbf{P}}_{(2)x[j,j']}$ estimate $\text{pr}(z_{(1)ix} = z_{(1)i'x} \mid \mathbf{Y}_1, \dots, \mathbf{Y}_m)$ and $\text{pr}(z_{(2)jx} = z_{(2)j'x} \mid \mathbf{Y}_1, \dots, \mathbf{Y}_m)$ via the relative frequency of Gibbs samples in which $z_{(1)ix} = z_{(1)i'x}$ and $z_{(2)jx} = z_{(2)j'x}$, 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 $x = 1, \dots, m$, via

$$\hat{\mathbf{z}}_{(1)x} = \text{argmin}_{\mathbf{z}} \mathbb{E}[\text{VI}(\mathbf{z}_{(1)x}, \mathbf{z}) \mid \mathbf{Y}_1, \dots, \mathbf{Y}_m], \quad \text{and} \quad \hat{\mathbf{z}}_{(2)x} = \text{argmin}_{\mathbf{z}} \mathbb{E}[\text{VI}(\mathbf{z}_{(2)x}, \mathbf{z}) \mid \mathbf{Y}_1, \dots, \mathbf{Y}_m],$$

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 $\hat{\mathbf{P}}_{(1)x}$ and $\hat{\mathbf{P}}_{(2)x}$, respectively.

Besides studying the age-specific group structures among underlying and contributing causes encoded within $\hat{\mathbf{z}}_{(1)x}$ and $\hat{\mathbf{z}}_{(2)x}$, 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 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 $\boldsymbol{\theta}_{hkc}$ are integrated out to obtain the likelihood in (1), when such quantities are also of interest a plug-in estimate can be readily derived. In particular, denote with \hat{n}_{hkxc} the total number of pairs (i, j) such that $\hat{z}_{(1)ix} = h$, $\hat{z}_{(2)jx} = k$ and $y_{ijx} = c$. Then, leveraging Dirichlet-categorical conjugacy, a sensible point estimate for each θ_{hkxc} is

$$\hat{\theta}_{hkxc} = \mathbb{E}(\theta_{hkxc} \mid \mathbf{Y}_1, \dots, \mathbf{Y}_m, \hat{\mathbf{z}}_{(1)x}, \hat{\mathbf{z}}_{(2)x}) = (a_c^\theta + \hat{n}_{hkxc}) / (a_\bullet^\theta + \hat{n}_{hkx\bullet}), \quad (9)$$

for every $h = 1, \dots, \hat{H}_x$, $k = 1, \dots, \hat{K}_x$, $x = 1, \dots, m$ and $c = 1, \dots, 4$.

4 Simulation Studies

In this section, we present extensive simulation studies that illustrate the performance of the model proposed in Section 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 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 $(y_{ijx} \mid z_{(1)ix}^{(0)} = h, z_{(2)jx}^{(0)} = k, \boldsymbol{\theta}_{h k x}^{(0)}) \sim \text{Cat}_{1:4}(\boldsymbol{\theta}_{h k x}^{(0)})$, independently for every $i = 1, \dots, 100$, $j = 1, \dots, 100$ and $x = 1, \dots, 10$, where each $\boldsymbol{\theta}_{h k x}^{(0)}$ 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 3), whereas $\mathbf{z}_{(1)x}^{(0)}$ and $\mathbf{z}_{(2)x}^{(0)}$ 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 for a graphical representation of $\mathbf{z}_{(1)x}^{(0)}$ and $\mathbf{z}_{(2)x}^{(0)}$, $x = 1, \dots, 10$). 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 $\mathbf{z}_x^{(0)}$, for $x = 1, \dots, 10$, that is common to

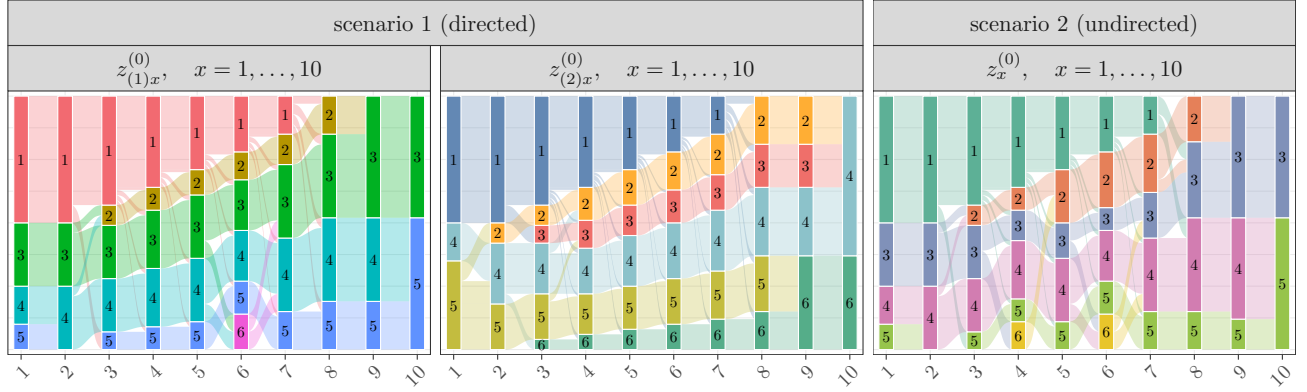


Figure 2: Riverplots representing the evolution of $\mathbf{z}_{(1)x}^{(0)}$ and $\mathbf{z}_{(2)x}^{(0)}$, $x = 1, \dots, 10$ in the first scenario, and $\mathbf{z}_x^{(0)}$, $x = 1, \dots, 10$ in the second scenario.

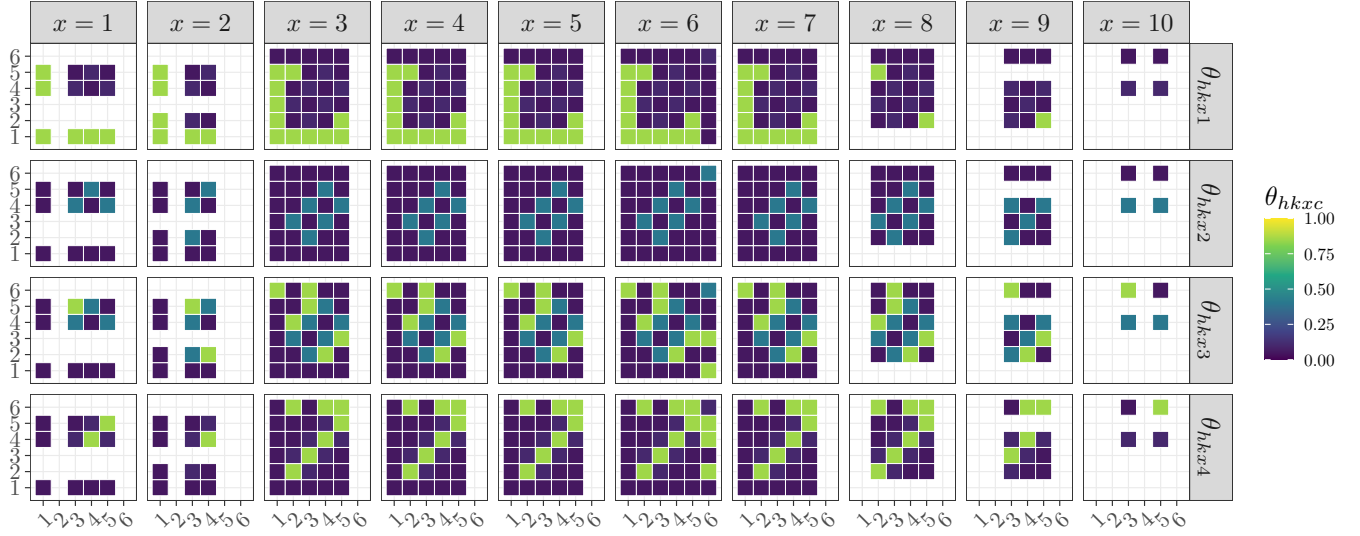


Figure 3: For the first simulation scenario (directed), graphical representation of the block-specific parameters, $\theta_{hkxc}^{(0)}$, evolving over age classes $x = 1, \dots, 10$ and stratified by category $c = 1, \dots, 4$. The numbered rows and columns correspond to row- and column-specific groups. White cells correspond to pairs of empty groups at the specific ages.

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 $\mathbf{z}_{(1)x}^{(0)} = \mathbf{z}_{(2)x}^{(0)} = \mathbf{z}_x^{(0)}$, $\theta_{hkc}^{(0)} = \theta_{khc}^{(0)}$, and $y_{ijx} = y_{jix}$ for each $x = 1, \dots, 10$ (see Figure 2 for a graphical illustration of $\mathbf{z}_x^{(0)}$, $x = 1, \dots, 10$). Notice that, in both scenarios, we do not consider supervision by external covariates \mathbf{w} 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{w} is not available, it suffices to remove $(\bar{n}_{hw_i}^{(-i)} + a_{w_i}^{\mathbf{w}})/(\bar{n}_{h\bullet}^{(-i)} + a_{\bullet}^{\mathbf{w}})$ and $a_{w_i}^{\mathbf{w}}/a_{\bullet}^{\mathbf{w}}$ in (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 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 param-

ters, respectively. Under such settings, posterior samples for $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)10}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)10}$ in the first simulation scenario can be obtained via the tempered Gibbs-sampler developed in Sections 3.1–3.2. Such a routine can be also adapted to the second scenario by sampling from the posterior of a single sequence $\mathbf{z}_1, \dots, \mathbf{z}_{10}$ (having the same prior as $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)10}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)10}$), while replacing the likelihood induced by directed networks with the one arising in undirected settings, where $y_{ijx} = y_{jix}$, and hence, $\boldsymbol{\theta}_{h k x} = \boldsymbol{\theta}_{k h x}$. As a consequence, in the directed case, $p(\mathbf{Y}_1, \dots, \mathbf{Y}_m \mid \mathbf{z}_x, x = 1, \dots, m) = \prod_{x=1}^m (\prod_{h=1}^{H_x} \prod_{k=1}^h [\{\Gamma(a_{\bullet}^{\theta})/\Gamma(a_{\bullet}^{\theta} + n_{h k x \bullet})\} \prod_{c=1}^4 \{\Gamma(a_c^{\theta} + n_{h k x c})/\Gamma(a_c^{\theta})\}])$.

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 $\mathbf{z}_{(1)1}, \dots, \mathbf{z}_{(1)10}$ and $\mathbf{z}_{(2)1}, \dots, \mathbf{z}_{(2)10}$ in the first scenario, and $\mathbf{z}_1, \dots, \mathbf{z}_{10}$ in the second, via the strategies in Section 3.3.

Figure 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., $\mathbf{z}_{(1)1}^{(0)}, \dots, \mathbf{z}_{(1)10}^{(0)}$ and $\mathbf{z}_{(2)1}^{(0)}, \dots, \mathbf{z}_{(2)10}^{(0)}$) and two (i.e., $\mathbf{z}_1^{(0)}, \dots, \mathbf{z}_{10}^{(0)}$). 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.

As shown in Figure 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-

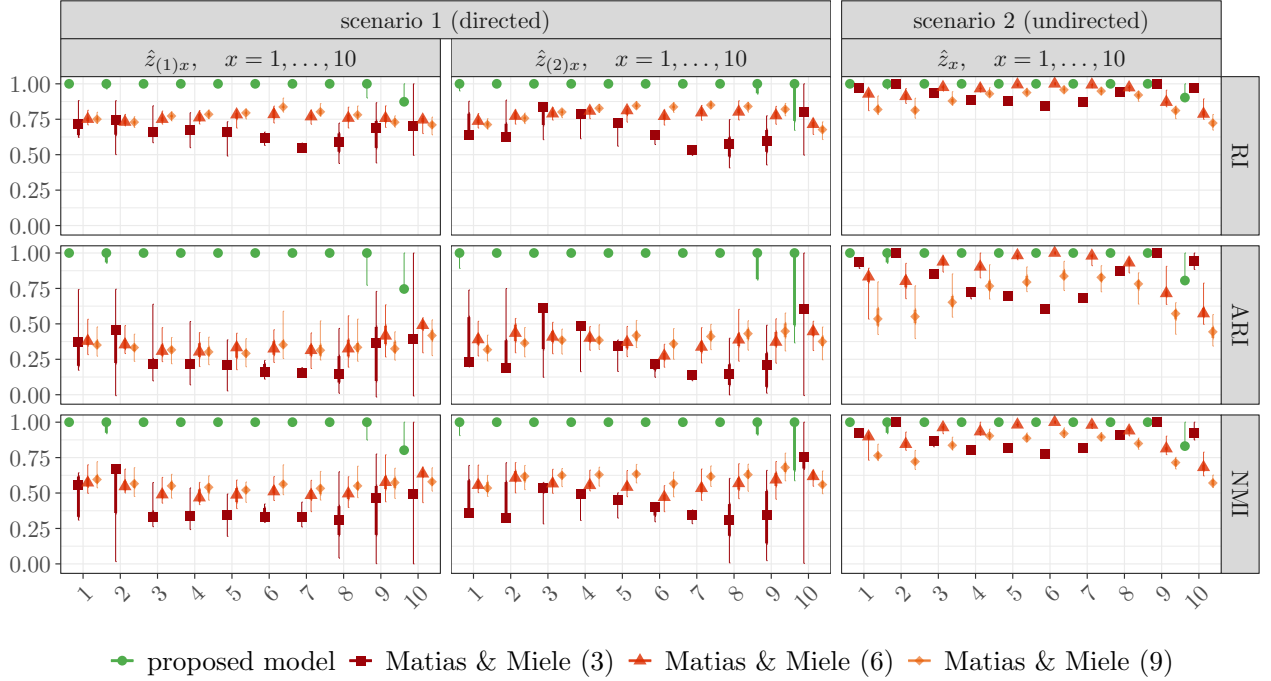


Figure 4: Accuracy of the proposed model and of the one designed by [Matias and Miele \(2017\)](#) (under three specifications for the total number of clusters) in estimating the true group membership vectors $\mathbf{z}_{(1)x}^{(0)}$ and $\mathbf{z}_{(2)x}^{(0)}$, $x = 1, \dots, 10$ in the first scenario, and $\mathbf{z}_x^{(0)}$, $x = 1, \dots, 10$ 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.

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 1.1 (see also Figure 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 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 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 ages 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 1. A detailed presentation of the important findings obtained from this analysis is provided in Section 5.

5 Application to USA Cause-of-Death Networks

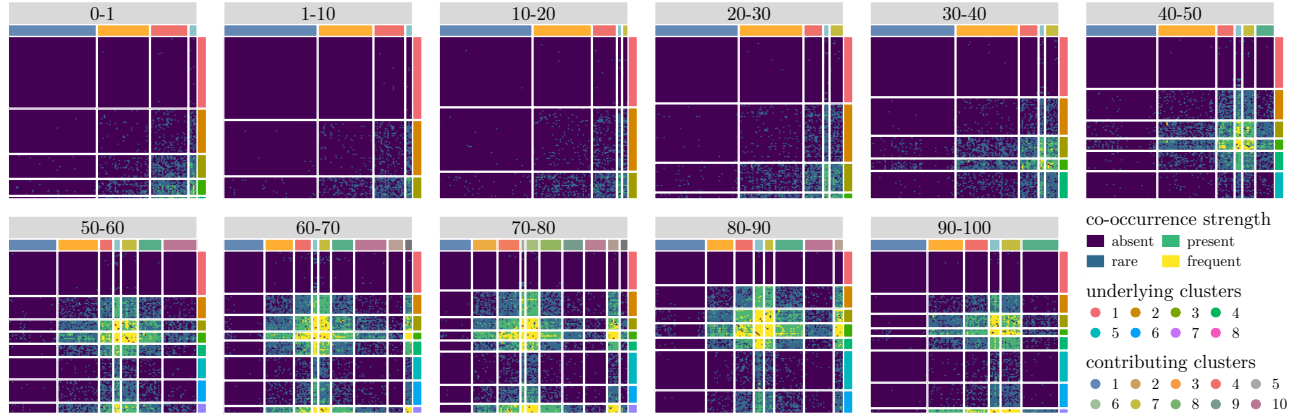
We conclude by illustrating the potential of the proposed model in unveiling group structures and modular interactions among causes of death that were hidden to previous studies of the USA data presented in Section 1.1. In line with standard practice in demography, we apply the model separately to male 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.

5.1 Empirical Results and Findings

As a first important assessment, Figure 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 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 5). As is evident from Figure 6, such an heterogeneity results from smooth fragmentations 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

Network sequence over age classes for the male population



Network sequence over age classes for the female population

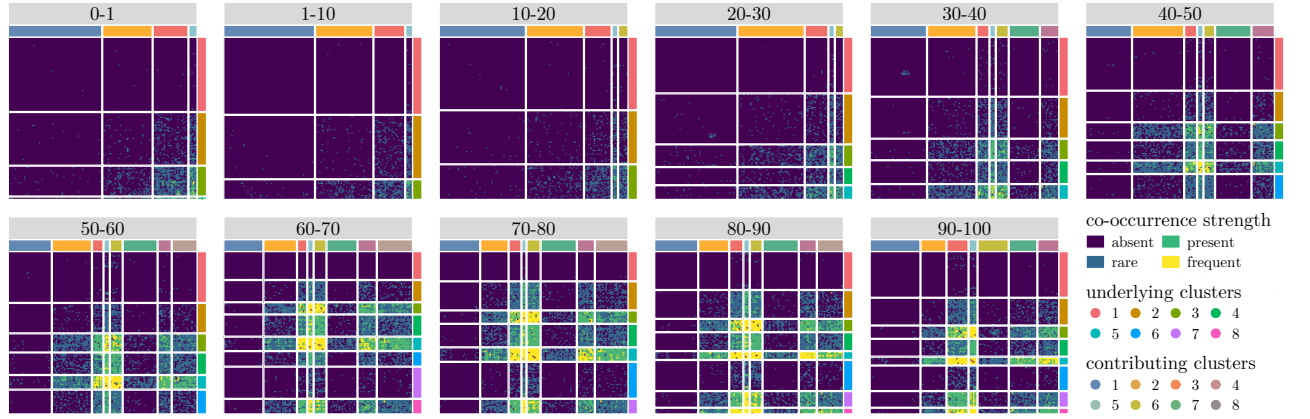


Figure 5: Observed co-occurrence networks among underlying and contributing causes of death at different age classes in the USA male (top) and female (bottom) populations. Rows and columns of the adjacency matrices are re-ordered according to the group structures estimated under the dependent stochastic block model proposed in Section 2.

(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 in Figures 5–6 for both the underlying and contributing causes. As illustrated within Figure 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.

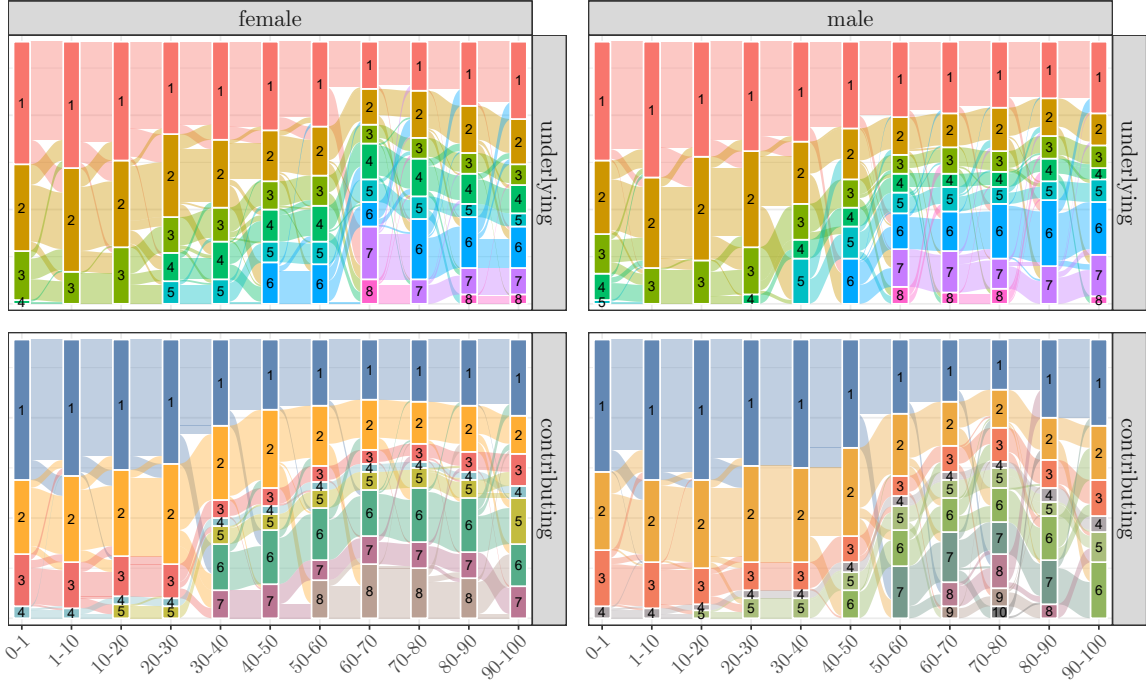


Figure 6: Riverplots representing the evolution of the estimated partitions $\hat{\mathbf{z}}_{(1)x}$ and $\hat{\mathbf{z}}_{(2)x}$ over age classes $x = 1, \dots, 11$ in the female and male populations. The size of each bar is proportional to the cardinality of the inferred group.

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 5–6, 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 6 highlights a slight increase in the size of group 1 at age classes 80–90 and 90–100. While this pattern deserves 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 5–6, we quantify in Figure 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 $\text{NMI}(\hat{\mathbf{z}}_{(1)x}, \hat{\mathbf{z}}_{(1)x-1})$, $\text{NMI}(\hat{\mathbf{z}}_{(2)x}, \hat{\mathbf{z}}_{(2)x-1})$ and $\text{NMI}(\hat{\mathbf{z}}_{(1)x}, \hat{\mathbf{z}}_{(2)x})$, respectively, for each x , and provide additional empirical support in favor of our model. In particular, the evolution of $\text{NMI}(\hat{\mathbf{z}}_{(1)x}, \hat{\mathbf{z}}_{(2)x})$ evidences marked and stable differences among the partition structures of underlying

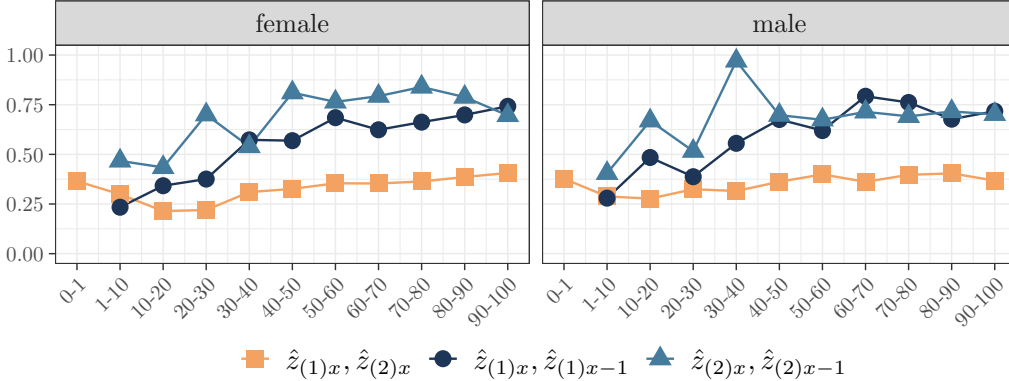


Figure 7: Normalized mutual information across ages between selected pairs (see the legend) of estimated partitions.

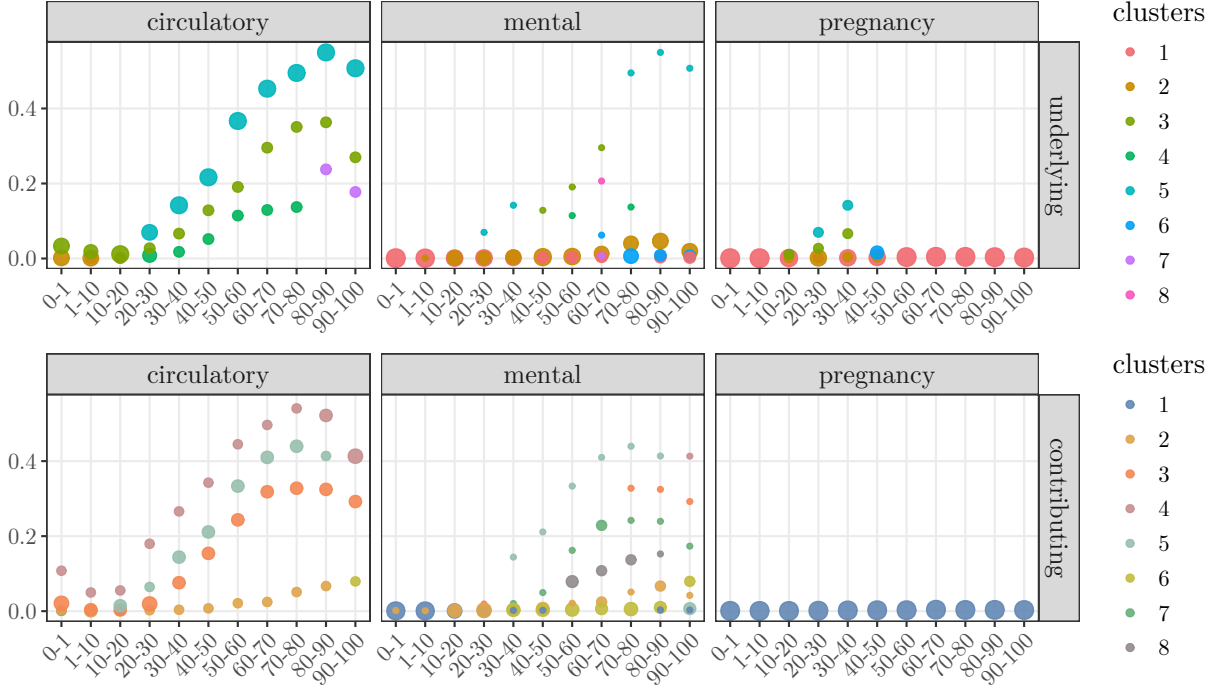


Figure 8: Group-specific marginal connectivity scores (MCS) over age classes for underlying (top) and contributing (bottom) causes in the female population, stratified by three macro-categories (circulatory, mental, and pregnancy). Point size is proportional to the number of causes in each macro-category belonging to a given group.

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 $\text{NMI}(\hat{\mathbf{Z}}_{(1)x}, \hat{\mathbf{Z}}_{(1)x-1})$ and $\text{NMI}(\hat{\mathbf{Z}}_{(2)x}, \hat{\mathbf{Z}}_{(2)x-1})$ 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 5, 6 and 7 inform on the overall structure and transitions of the inferred groups, Figure 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 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** interaction between the causes of a given underlying (contributing) group h (k) and those belonging to a generic contributing (underlying) cluster $k = 1, \dots, K_x$ ($h = 1, \dots, H_x$). More specifically, at each age class we compute $\text{MCS}_{(1)hx} = (1/\hat{K}_x) \sum_{k=1}^{\hat{K}_x} (\hat{\theta}_{h k x 3} + \hat{\theta}_{h k x 4})$, for $h = 1, \dots, \hat{H}_x$, and $\text{MCS}_{(2)kx} = (1/\hat{H}_x) \sum_{h=1}^{\hat{H}_x} (\hat{\theta}_{h k x 3} + \hat{\theta}_{h k x 4})$, for $k = 1, \dots, \hat{K}_x$, where $\hat{\theta}_{h k x 3}$ and $\hat{\theta}_{h k x 4}$ are obtained as in (9). The analysis of such measures in Figure 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 $\text{MCS}_{(1)hx}$. In contrast, contributing roles are more fragmented, with these conditions spreading across groups of varying $\text{MCS}_{(2)hx}$.

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 adolescent/early-adulthood (10–20, 20–30, 30–40) ages, respectively. These analyses rely on the study of the meet clusters introduced in Section 3.3 (see Figures 9–10) 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 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-

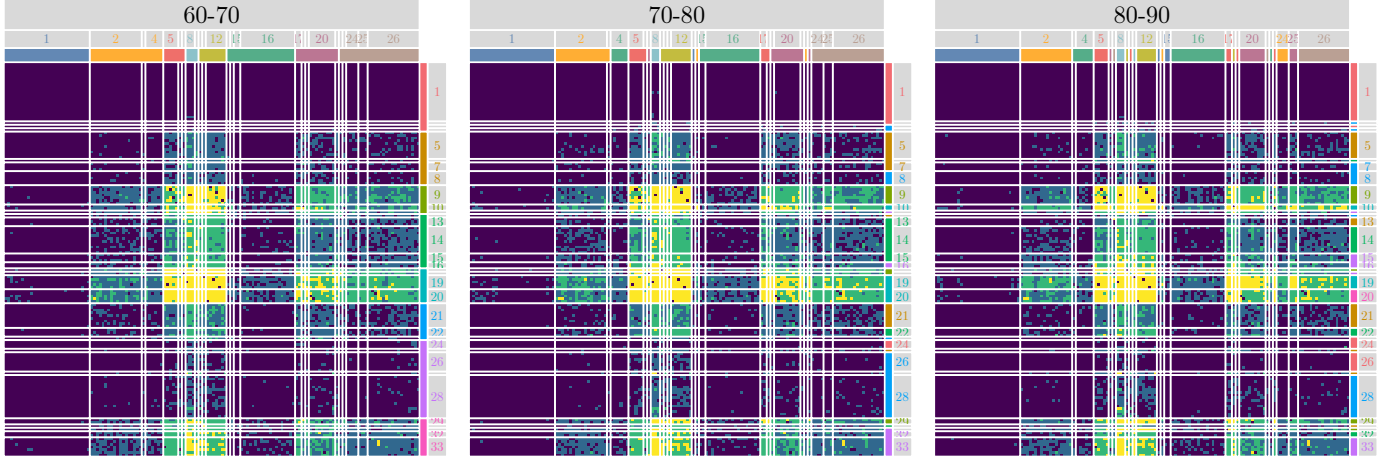


Figure 9: Adjacency matrices of the co-occurrence networks among underlying and contributing causes for USA female elder population. Rows and columns are re-ordered by estimated groups (colors) and *meet* clusters (lines-numbers).

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 recent years. As clarified in Figure 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-

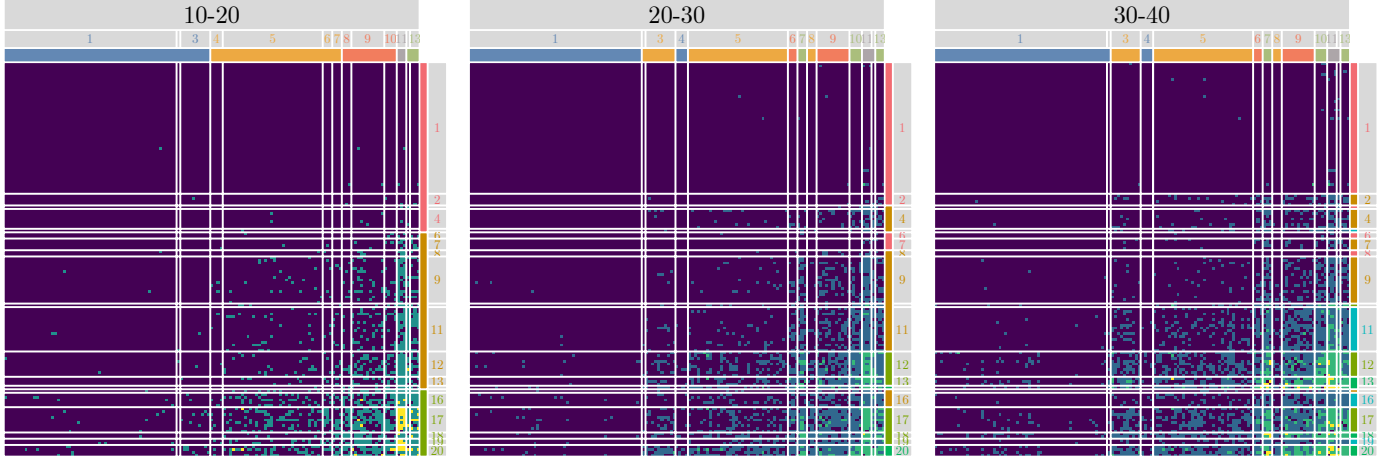


Figure 10: Adjacency matrices of the co-occurrence networks among underlying and contributing causes for USA male young population. Rows and columns are re-ordered by estimated groups (colors) and *meet* clusters (lines-numbers).

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 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.

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 and 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 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 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., 2025) 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 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.

Acknowledgments

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

References

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, 831.

- Ahmad, F. B., Cisewski, J. A., and Anderson, R. N. (2024), “Leading causes of death in the US, 2019–2023,” *JAMA*, 332, 957–958.
- 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, 1–12.
- 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, 333–344.
- 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, 650.
- Canudas-Romo, V. (2010), “Three measures of longevity: time trends and record values,” *Demography*, 47, 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, 1712–1718.
- Case, A., and Deaton, A. (2021), *Deaths of Despair and the Future of Capitalism*, Princeton University Press.
- Chmiel, A., Klimek, P., and Thurner, S. (2014), “Spreading of diseases through comorbidity networks across life and gender,” *New Journal of Physics*, 16, 115013.
- 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, 1668–1685.
- Désesquelles, A., Demuru, E., Egidi, V., Frova, L., Meslé, F., Pappagallo, M., and Salvatore, M. A. (2014), “Cause-specific mortality analysis: is the underlying cause of death sufficient?” *Quetelet Journal*, 2, 119–135.
- 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.— (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, 467–498.
- 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*, In press.
- Earl, D. J., and Deem, M. W. (2005), “Parallel tempering: theory, applications, and new perspectives,” *Physical Chemistry Chemical Physics*, 7, 3910–3916.
- 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.
- Escobar, M. D., and West, M. (1995), “Bayesian density estimation and inference using mixtures,” *Journal of the American Statistical Association*, 90, 577–588.
- 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, 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, 46.

- Franzolini, B., De Iorio, M., and Eriksson, J. (2025), “Conditional partial exchangeability: a probabilistic framework for multi-view clustering,” *arXiv preprint arXiv:2307.01152*.
- 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, 1–12.
- Goto, S., Takagishi, M., and Yadohisa, H. (2021), “Clustering for time-varying relational count data,” *Computational Statistics & Data Analysis*, 156, 107123.
- 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, 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, 36.
- Hamming, R. W. (1950), “Error detecting and error correcting codes,” *The Bell System Technical Journal*, 29, 147–160.
- 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, e1000353.
- Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983), “Stochastic blockmodels: First steps,” *Social Networks*, 5, 109–137.
- Hubert, L., and Arabie, P. (1985), “Comparing partitions,” *Journal of Classification*, 2, 193–218.
- Huynh, N., and Ludkovski, M. (2024), “Joint models for cause-of-death mortality in multiple populations,” *Annals of Actuarial Science*, 18, 51–77.
- 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, 161–179.
- Jeong, E., Ko, K., Oh, S., and Han, H. W. (2017), “Network-based analysis of diagnosis progression patterns using claims data,” *Scientific Reports*, 7, 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, 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*, pp. 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.
- Lee, C., and Wilkinson, D. J. (2019), “A review of stochastic block models and extensions for graph clustering,” *Applied Network Science*, 4, 1–50.
- 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, 2369–2395.
- 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, 2095–2128.
- Mariadassou, M., Robin, S., and Vacher, C. (2010), “Uncovering latent structure in valued graphs: a variational approach,” *The Annals of Applied Statistics*, 4, 715–742.
- Matias, C., and Miele, V. (2017), “Statistical clustering of temporal networks through a dynamic stochastic block model,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79, 1119–1141.
- 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, 6998–7000.
- Meilä, M. (2007), “Comparing clusterings — an information based distance,” *Journal of Multivariate Analysis*, 98, 873–895.
- Miasojedow, B., Moulines, E., and Vihola, M. (2013), “An adaptive parallel tempering algorithm,” *Journal of Computational and Graphical Statistics*, 22, 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, 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, 260–278.
- Newman, M. E., and Leicht, E. A. (2007), “Mixture models and exploratory analysis in networks,” *Proceedings of the National Academy of Sciences*, 104, 9564–9569.
- Nowicki, K., and Snijders, T. A. B. (2001), “Estimation and prediction for stochastic blockstructures,” *Journal of the American Statistical Association*, 96, 1077–1087.
- 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, 614–627.
- Peixoto, T. P. (2018), “Nonparametric weighted stochastic block models,” *Physical Review E*, 97, 012306.
- (2022), “Ordered community detection in directed networks,” *Physical Review E*, 106, 024305.
- Pensky, M., and Zhang, T. (2019), “Spectral clustering in the dynamic stochastic block model,” *Electronic Journal of Statistics*, 13, 678–709.
- Rand, W. M. (1971), “Objective criteria for the evaluation of clustering methods,” *Journal of the American Statistical Association*, 66, 846–850.
- Rebaudo, G., Lin, Q., and Mueller, P. (2025), “Separate exchangeability as modeling principle in Bayesian nonparametrics,” *arXiv preprint 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, 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, 104–108.
- 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, 12679–12684.

- Schmidt, M. N., and Morup, M. (2013), “Nonparametric Bayesian modeling of complex networks: An introduction,” *IEEE Signal Processing Magazine*, 30, 110–128.
- Stefanucci, M., and Mazzucco, S. (2022), “Analysing cause-specific mortality trends using compositional functional data analysis,” *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185, 61–83.
- Strehl, A., and Ghosh, J. (2002), “Cluster ensembles—a knowledge reuse framework for combining multiple partitions,” *Journal of Machine Learning Research*, 3, 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, 321–350.
- Syed, S., Romaniello, V., Campbell, T., and Bouchard-Côté, A. (2021), “Parallel tempering on optimized paths,” in *International Conference on Machine Learning*, PMLR, pp. 10033–10042.
- Tallberg, C. (2004), “A Bayesian approach to modeling stochastic blockstructures with covariates,” *Journal of Mathematical Sociology*, 29, 1–23.
- Trias-Llimós, S., and Permanyer, I. (2023), “Cause-of-death diversity from a multiple-cause perspective in the United States,” *Demography*, 60, 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, 149–165.
- Van Raalte, A. A. (2021), “What have we learned about mortality patterns over the past 25 years?” *Population Studies*, 75, 105–132.
- Wade, S., and Ghahramani, Z. (2018), “Bayesian cluster analysis: point estimation and credible balls,” *Bayesian Analysis*, 13, 559–626.
- Wang, Y. J., and Wong, G. Y. (1987), “Stochastic blockmodels for directed graphs,” *Journal of the American Statistical Association*, 82, 8–19.
- Woolf, S. H., and Schoomaker, H. (2019), “Life expectancy and mortality rates in the United States, 1959-2017,” *JAMA*, 322, 1996–2016.
- World Health Organization (2016), *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*, vol. 2, World Health Organization.
- Wyse, J., and Friel, N. (2012), “Block clustering with collapsed latent block models,” *Statistics and Computing*, 22, 415–428.
- Xu, K. (2015), “Stochastic block transition models for dynamic networks,” in *Artificial Intelligence and Statistics*, PMLR, pp. 1079–1087.
- 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, 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, 157–189.
- Zhang, J., He, X., and Wang, J. (2022), “Directed community detection with network embedding,” *Journal of the American Statistical Association*, 117, 1809–1819.
- Zhang, Y., Levina, E., and Zhu, J. (2016), “Community detection in networks with node features,” *Electronic Journal of Statistics*, 10, 3153–3178.