Survival Fully-Collapsed Copula Mixed Membership Blockmodel

Yanxing Song and **Richard Yi Da Xu**
Department of Mathematics, Hong Kong Baptist University
yanxingsong@life.hkbu.edu.hk, xuyida@hkbu.edu.hk

Abstract

Survival models on networks describe the time-evolving dynamics of complex systems, but existing frameworks suffer from two key limitations: First, many past approaches implicitly assume independent censoring, which can introduce substantial bias. Second, point-process-based models often emphasize only sending behavior and overlook shared latent structure between interacting nodes. Recent work has incorporated copulas into deep survival models to mitigate censoring bias, but these methods do not capture topological role structure. Another line of research combines survival analysis with the Mixed-Membership Stochastic Blockmodel (MMSB) to model heterogeneity via role-specific hazard rates; however, it typically assumes structural independence and therefore misses important correlations between roles. To bridge these gaps, we apply an improved Copula–MMSB mechanism and propose the first Survival Fully-Collapsed Copula Mixed Membership Blockmodel. Our framework jointly models time-to-event dynamics and complex dependencies among latent roles. On the Enron benchmark, it surpasses state-of-the-art survival blockmodels in both predictive likelihood and structural interpretability. Additionally, the model admits an exact closed-form discrete distribution over role pairs, which enables a fully collapsed Gibbs sampler and avoids the slow mixing and poor convergence common in non-conjugate inference. In simulations, this yields a 100× improvement in sampling efficiency over semi-collapsed baselines.

1 Introduction

Survival models in networks describe the time-evolving dynamics of interaction systems by treating each edge as a time-to-event outcome rather than a purely topological link. Unlike traditional graph approaches that focus on who connects to whom, these probabilistic frameworks explicitly model waiting times and event timing, offering a natural way to explain response delays, interaction durations, and future event occurrence [Perry and Wolfe, 2013; Trivedi *et al.*, 2019; Kumar *et al.*, 2019]. Recent advances have extended classical survival analysis to network data via multivariate point processes and hazard-based formulations, enabling continuous-time modeling of social and biological systems [Sit *et al.*, 2021; Rastelli and Fop, 2020]. Within this paradigm, interaction timing is often shaped by latent roles of the participating actors, where different role combinations can produce distinct temporal patterns.

However, existing network survival frameworks still suffer from two key limitations. First, many approaches implicitly assume that the censoring mechanism is independent of the underlying event process once observed covariates are accounted for. This assumption is difficult to justify in real interaction logs and can introduce substantial bias when censoring is informative [Gharari *et al.*, 2023; Kalbfleisch and Prentice, 2002]. Second, point-process-based models on networks commonly emphasize the initiation dynamics of senders and treat receivers primarily as context, which makes it hard to capture the shared latent structure that jointly governs pairwise behavior [Perry and Wolfe, 2013; Trivedi *et al.*, 2019; Song *et al.*, 2024]. These issues become especially pronounced at scale, where observation windows create pervasive censoring and interaction behavior is highly heterogeneous [Rastelli and Fop, 2020; Wang and Wong, 1987].

Recent work has begun to address the first limitation by explicitly relaxing the standard independent-censoring assumption. In particular, Gharari *et al.* [2023] propose a copula-based deep survival framework that couples the event-time and censoring-time mechanisms through a survival copula, allowing their dependence to be learned from right-censored observations, where the event time is only known to exceed the observation window. This approach builds on the classical role of copulas in separating marginal behavior from dependence structure [Nelsen, 2006]. However, this line of work is primarily developed for independent instances with covariates, and it does not explicitly model network topology or the latent roles that organize interaction patterns across node pairs [Perry and Wolfe, 2013; Sit *et al.*, 2021].

A separate line of research focuses on the network setting and addresses the sender-centric bias common in point-process formulations. Song *et al.* [2024] propose the Survival Mixed Membership Blockmodel (SMMB), which combines mixed-membership blockmodeling with survival analysis to model edge-level response processes through role-dependent hazards [Song *et al.*, 2024; Airoldi *et al.*, 2008]. This provides a structured way to account for heterogeneity across role pairs and to accommodate censoring in time-to-event network data [Kalbfleisch and Prentice, 2002; Song *et al.*, 2024]. SMMB follows the standard mixed-membership assumption that sender and receiver roles are drawn independently from their membership vectors [Airoldi *et al.*, 2008; Song *et al.*, 2024]. However, in many social and biological networks, structural patterns such as reciprocity, hierarchy, and assortativity imply that interacting roles are coupled rather than independent [Hoff, 2005; Karrer and Newman, 2011]. As a result, important role correlations can be missed when the prior cannot represent such dependence directly.

Copulas also provide a direct way to introduce dependence into mixed-membership modeling without changing the marginal membership interpretation [Nelsen, 2006]. Building on this idea, Fan *et al.* [2016] introduce the Copula–MMSB (CMMSB), which uses a copula layer to model the joint distribution of sender and receiver roles and thereby relaxes the conditional independence built into standard MMSB-style priors [Fan *et al.*, 2016]. The modeling gain is clear, but it comes with a practical cost: once a copula is introduced, the Dirichlet–multinomial conjugacy that makes role-assignment updates simple no longer holds, so inference becomes non-conjugate and more expensive. As a result, common strategies rely on augmentation or auxiliary-variable constructions to restore tractable updates [Damlen *et al.*, 1999; Neal, 2000]. In CMMSB settings this typically leads to semi-collapsed samplers that only partially marginalize latent parameters, leaving strongly coupled latent variables in the chain; in practice, this coupling can yield high autocorrelation and slow mixing as network size and model complexity grow [Fan *et al.*, 2016; Liu, 1994].

To bridge these gaps, we propose the first Survival Fully-Collapsed Copula Mixed Membership Blockmodel, which jointly models time-to-event dynamics on networks and complex dependencies among latent roles. By combining a survival blockmodel with an improved Copula–MMSB role mechanism, it captures dependent sender–receiver role pairs while preserving the mixed-membership interpretation. Another key implication of this construction is that it resolves the computational bottleneck of copula-based mixed-membership blockmodels. Such models are typically non-conjugate, and existing inference procedures are often only semi-collapsed, which leads to slow mixing and poor convergence in large networks. Under our formulation, however, we can analytically integrate out both the node membership variables and the copula latent variables, yielding an exact closed-form discrete distribution over role pairs. This makes it possible to use a fully collapsed Gibbs sampler that draws role assignments directly from the marginal distribution and avoids inner-loop Metropolis–Hastings updates or auxiliary-variable constructions. Empirically, on the Enron benchmark our approach improves predictive likelihood and yields clearer role-dependence structure than survival blockmodel baselines, and in simulations it achieves a $100\times$ gain in sampling efficiency over semi-collapsed alternatives.

2 Model

Symbol	Description
$N, \mathcal{U}$	Number of nodes and the node set $\{1, \ldots, N\}$.
$\mathcal{E}$	Set of observed directed edges (i, j).
K	Number of latent roles.
$\boldsymbol{\pi}_i$	Mixed-membership vector for node i.
$\alpha, \boldsymbol{\eta}$	Concentration parameter $\alpha > 0$ and base distribution $\boldsymbol{\eta}$ on the $(K-1)$-dimensional probability simplex, i.e., $\boldsymbol{\eta} \in \{\mathbf{v} \in \mathbb{R}^K : v_k \geq 0,\ \sum_{k=1}^{K} v_k = 1\}$, for the Dirichlet prior $\text{Dir}(\alpha\boldsymbol{\eta})$ on $\boldsymbol{\pi}_i$.
$\Pi_i(k)$	Cumulative membership probability $\sum_{m=1}^{k} \pi_{im}$.
$C_\theta(\cdot, \cdot)$	Bivariate copula with dependence parameter θ.
u_{ij}, v_{ij}	Latent copula variables for edge (i, j).
s_{ij}, r_{ij}	Latent sender and receiver roles for edge (i, j).
$p_{k\ell}^{(ij)}$	Collapsed probability mass of role pair (k, ℓ) for edge (i, j).
t_{ij}, δ_{ij}	Observed time and event indicator, where $\delta_{ij} = 1$ denotes an observed event and $\delta_{ij} = 0$ denotes right censoring.
$\lambda_g^{k\ell}$	Hazard for role pair (k, ℓ) in time segment g.
$\beta_{k\ell}$	Covariate coefficient vector for role pair (k, ℓ).

Table 1: Key Notations.

We consider a directed interaction network with N nodes, denoted by $\mathcal{U} = \{1, \ldots, N\}$. For each ordered pair $(i, j) \in \mathcal{E}$, we observe an observed time t_{ij} and an event indicator δ_{ij}, where $\delta_{ij} = 1$ indicates that an event occurred and $\delta_{ij} = 0$ indicates right censoring. Each edge may also have an associated covariate vector x_{ij}. To capture both temporal dynamics and structural dependence between interacting nodes, we define a generative model that couples latent sender and receiver roles through a copula while driving event times through a role-dependent survival process. Key notations are summarized in Table 1.

Generative process. The model is specified hierarchically as follows. Figure 1 shows the resulting fully collapsed dependency structure used by the Gibbs sampler.

1. **Node memberships.** For each node i, draw a mixed-membership vector
$$\boldsymbol{\pi}_i \sim \text{Dir}(\alpha\boldsymbol{\eta}),$$
where π_{ik} represents the propensity of node i to participate in role k.

2. **Coupled role assignments.** For each edge (i, j), draw a pair of latent uniform variables
$$(u_{ij}, v_{ij}) \sim C_\theta,$$
from a bivariate copula with dependence parameter θ. These uniforms are mapped to discrete sender and receiver roles via the inverse CDFs of the membership vectors:
$$s_{ij} = \min\{k : u_{ij} \leq \Pi_i(k)\},$$
$$r_{ij} = \min\{\ell : v_{ij} \leq \Pi_j(\ell)\}.$$
where $\Pi_i(k) = \sum_{m=1}^{k} \pi_{im}$. This construction ensures the correct marginals, $s_{ij} \sim \text{Cat}(\boldsymbol{\pi}_i)$ and $r_{ij} \sim \text{Cat}(\boldsymbol{\pi}_j)$, while inducing dependence between them through C_θ.

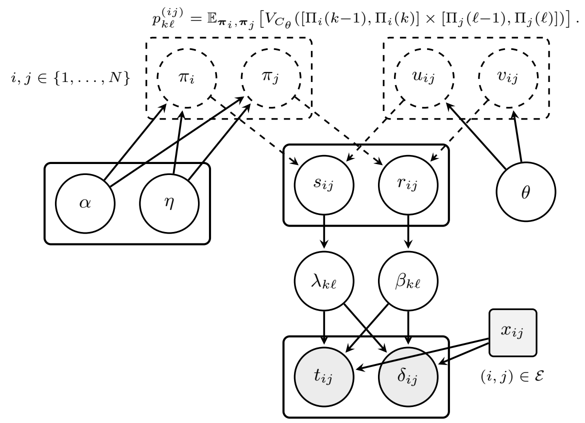

Figure 1: Graphical representation of the Survival Fully-Collapsed Copula Mixed Membership Blockmodel. Covariates x_{ij} are fixed observed inputs. (t_{ij}, δ_{ij}) are observed survival outcomes. Dashed nodes $(\boldsymbol{\pi}, u, v)$ are analytically integrated out, and solid arrows indicate collapsed conditional dependencies.

3. **Time-to-event outcome.** Conditional on the role pair $(s_{ij}, r_{ij}) = (k, \ell)$, the edge (i, j) follows a role-specific survival model. Let $\lambda_g^{k\ell}$ denote the baseline hazard for role pair (k, ℓ) in time segment g, and let $\beta_{k\ell}$ be the regression coefficients for covariates x_{ij}. The observed time t_{ij} and event indicator δ_{ij} are then generated according to this role-dependent hazard.

Collapsed Role Distribution. To enable efficient inference, we follow the principle of collapsed Gibbs sampling [Griffiths and Steyvers, 2004] and analytically integrate out the continuous latent variables $(\boldsymbol{\pi}, u_{ij}, v_{ij})$. Let $(u_{ij}, v_{ij}) \sim C_\theta$ denote the latent copula variables for edge (i, j). Conditioned on the membership vectors $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$, the probability that this edge is assigned to role pair (k, ℓ) is given by

$$\begin{aligned}
&\Pr(s_{ij} = k, r_{ij} = \ell \mid \boldsymbol{\pi}_i, \boldsymbol{\pi}_j, \theta) \\
&= \mathbb{E}_{(u,v)\sim C_\theta}\Big[\mathbf{1}\big\{\Pi_i^{-1}(u) = k,\ \Pi_j^{-1}(v) = \ell\big\}\Big] \\
&= \iint_{(0,1)^2} \mathbf{1}\big\{\Pi_i^{-1}(u) = k,\ \Pi_j^{-1}(v) = \ell\big\}\ dC_\theta(u, v).
\end{aligned} \tag{1}$$

Geometrically, this probability represents the volume under the copula density. The integration region is a rectangle defined by the membership boundaries. This leads to the *copula rectangle mass*:

$$\begin{aligned}
&q_{ij}(k, \ell \mid \boldsymbol{\pi}_i, \boldsymbol{\pi}_j, \theta) \\
&\quad = V_{C_\theta}([\Pi_i(k{-}1), \Pi_i(k)] \times [\Pi_j(\ell{-}1), \Pi_j(\ell)]) \\
&\quad = C_\theta(\Pi_i(k), \Pi_j(\ell)) - C_\theta(\Pi_i(k{-}1), \Pi_j(\ell)) \\
&\qquad - C_\theta(\Pi_i(k), \Pi_j(\ell{-}1)) + C_\theta(\Pi_i(k{-}1), \Pi_j(\ell{-}1)).
\end{aligned} \tag{2}$$

The core of our contribution is the full collapse. By integrating out the Dirichlet-distributed membership vectors $\boldsymbol{\pi}$, we obtain the marginal likelihood of the role pair (k, ℓ) in closed form:

$$\begin{aligned}
p_{k\ell}^{(ij)} &= \iint q_{ij}(k, \ell \mid \boldsymbol{\pi}_i, \boldsymbol{\pi}_j, \theta) \\
&\qquad \times \mathrm{Dir}(\boldsymbol{\pi}_i \mid \alpha\boldsymbol{\eta})\, \mathrm{Dir}(\boldsymbol{\pi}_j \mid \alpha\boldsymbol{\eta}) d\boldsymbol{\pi}_i\, d\boldsymbol{\pi}_j \\
&= \mathbb{E}_{\boldsymbol{\pi}_i,\boldsymbol{\pi}_j}\Bigg[V_{C_\theta}\bigg([\Pi_i(k{-}1), \Pi_i(k)] \\
&\qquad\qquad \times [\Pi_j(\ell{-}1), \Pi_j(\ell)]\bigg)\Bigg].
\end{aligned} \tag{3}$$

This quantity $p_{k\ell}^{(ij)}$ acts as a "collapsed prior" for the role pair, incorporating both the community structure and the dependency θ.

Survival Likelihood. Conditional on the latent role pair $(s_{ij}, r_{ij}) = (k, \ell)$, edge (i, j) follows a piecewise-constant survival model. Let the time axis be partitioned into g segments indexed by $g = 1, \dots, G$, and let exposure$_{ij,g}$ denote the time that edge (i, j) is at risk in segment g. The baseline hazard for role pair (k, ℓ) in segment g is $\lambda_g^{k\ell}$, giving a cumulative hazard contribution

$$H_{ij,g}^{k\ell} = \lambda_g^{k\ell}\, \mathrm{exposure}_{ij,g}.$$

Covariates enter multiplicatively through

$$\rho_{ij}^{k\ell} = \exp(x_{ij}^\top \beta_{k\ell}),$$

where x_{ij} is the covariate vector and $\beta_{k\ell}$ is the role-pair-specific coefficient. The survival likelihood is then

$$\begin{aligned}
&L_{\mathrm{surv}}^{k\ell}(t_{ij}, \delta_{ij} \mid x_{ij}, \{\lambda_g^{k\ell}\}_{g=1}^G, \beta_{k\ell}) \\
&= \exp\Big(-\rho_{ij}^{k\ell}\sum_{g=1}^{G} H_{ij,g}^{k\ell}\Big)\big(\rho_{ij}^{k\ell}\lambda_{g(t_{ij})}^{k\ell}\big)^{\delta_{ij}}.
\end{aligned} \tag{4}$$

Marginalizing over role pairs yields the collapsed edge likelihood

$$\begin{aligned}
&p(t_{ij}, \delta_{ij} \mid x_{ij}, \lambda, \beta) \\
&= \sum_{k,\ell} p_{k\ell}^{(ij)} L_{\mathrm{surv}}^{k\ell}(t_{ij}, \delta_{ij} \mid x_{ij}, \{\lambda_g^{k\ell}\}_{g=1}^G, \beta_{k\ell})
\end{aligned} \tag{5}$$

which provides the discrete weights for sampling (s_{ij}, r_{ij}) in the fully collapsed Gibbs sampler.

3 Model Identifiability

Identifiability is a prerequisite for interpretability in latent variable models: distinct parameter settings should not induce the same observed data distribution. In our setting, this means that the role-specific survival mechanisms and the copula-induced role dependence are uniquely determined from the observed pairwise survival data. We adopt the general identifiability framework for finite mixtures and latent variable models in Allman *et al.* [2009].

Recall that the collapsed edge likelihood in (5) can be written as a finite mixture over role pairs:

$$\begin{aligned}
&p(t, \delta \mid x; \theta, \alpha, \boldsymbol{\eta}, \lambda, \beta) \\
&= \sum_{k=1}^{K}\sum_{\ell=1}^{K} p_{k\ell}(\theta, \alpha, \boldsymbol{\eta})\ L_{\mathrm{surv}}^{k\ell}\big(t, \delta \mid x, \{\lambda_g^{k\ell}\}_{g=1}^G, \beta_{k\ell}\big).
\end{aligned} \tag{6}$$

where $L^{k\ell}_{\mathrm{surv}}$ is the role-pair-specific survival kernel defined in Section 2, G is the number of time segments in the piecewise-constant hazard specification and $p_{k\ell}(\theta,\alpha,\boldsymbol{\eta})$ denotes the collapsed role-pair mass induced by the copula parameter θ and the Dirichlet prior $\mathrm{Dir}(\alpha\boldsymbol{\eta})$.

Our identification strategy proceeds in two steps: (i) identify the set of mixture components $\{L^{k\ell}_{\mathrm{surv}}\}$ and their mixing weights $\{p_{k\ell}\}$ from the observed distribution; (ii) recover $(\theta,\alpha,\boldsymbol{\eta})$ from the identified weights.

Regularity conditions. We impose the following sufficient conditions [Teicher, 1963; Allman *et al.*, 2009; Kalbfleisch and Prentice, 2002]:

A1. The number of roles K is known and finite. The Dirichlet hyperparameters satisfy $\alpha > 0$ and $\boldsymbol{\eta} \in \{\mathbf{v} \in \mathbb{R}^K : v_k > 0,\ \sum_{k=1}^K v_k = 1\}$.

A2. The mapping

$$(\theta,\alpha,\boldsymbol{\eta}) \longmapsto \{p_{k\ell}(\theta,\alpha,\boldsymbol{\eta})\}_{k,\ell=1}^K$$

is locally one-to-one on the parameter space. Equivalently, the Jacobian of $\{p_{k\ell}\}$ with respect to $(\theta,\alpha,\boldsymbol{\eta})$ has full column rank in a neighborhood of the true parameter.

A3. The time grid contains at least two segments, and the baseline hazards $\{\lambda^{k\ell}_g\}$ are not identical across all role pairs. An identifiability constraint is imposed on $\{\beta_{k\ell}\}$ to fix the overall scale.

A4. The family of kernels $\{L^{k\ell}_{\mathrm{surv}}(\cdot)\}_{k,\ell}$ is linearly independent on the support of (t,δ,x) under the design.

A5. Both events and right censoring occur with positive probability.

Main result. Let $\Theta = (\theta,\alpha,\boldsymbol{\eta},\lambda,\beta)$ collect the model parameters, where $\lambda = \{\lambda^{k\ell}_g : k,\ell = 1,\ldots,K; g = 1,\ldots,G\}$ and $\beta = \{\beta_{k\ell} : k,\ell = 1,\ldots,K\}$.

Theorem 1. *Under assumptions A1–A5, if two parameter sets Θ and $\tilde\Theta$ induce the same observed distribution in* (6) *for all (t,δ,x), then they are equivalent up to a permutation σ of the role labels $\{1,\ldots,K\}$. In particular:*

1. *the survival parameters match up to relabeling, $\beta_{k\ell} = \tilde\beta_{\sigma(k)\sigma(\ell)}$ and $\lambda^{k\ell}_g = \tilde\lambda^{\sigma(k)\sigma(\ell)}_g$ for all $g = 1,\ldots,G$;*
2. *the mixing weights match up to the same relabeling, $p_{k\ell}(\theta,\alpha,\boldsymbol{\eta}) = p_{\sigma(k)\sigma(\ell)}(\tilde\theta,\tilde\alpha,\tilde{\boldsymbol{\eta}})$;*
3. *by A2, the hyperparameters governing the collapsed weights are locally identified, so $(\theta,\alpha,\boldsymbol{\eta}) = (\tilde\theta,\tilde\alpha,\tilde{\boldsymbol{\eta}})$.*

Proof sketch. By A3–A5, the component kernels $\{L^{k\ell}_{\mathrm{surv}}\}$ form a distinct and linearly independent family (A4), so standard identifiability results for finite mixtures imply that the observed distribution uniquely determines the set of components and their mixing weights $\{p_{k\ell}\}$ up to a permutation of labels [Teicher, 1963; Allman *et al.*, 2009]. Given the identified weights, A2 ensures that the system of equations linking $(\theta,\alpha,\boldsymbol{\eta})$ to $\{p_{k\ell}\}$ admits a unique (local) solution, completing the argument. □

For the FGM copula used in our experiments, A2 reduces to a concrete non-degeneracy requirement: the Jacobian of $\{p_{k\ell}(\theta,\alpha,\boldsymbol{\eta})\}$ with respect to $(\theta,\alpha,\boldsymbol{\eta})$ should have full column rank almost everywhere. In practice, we fix the label permutation by ordering roles according to an identifiable summary of their survival behavior, such as the average baseline hazard [Stephens, 2000].

4 Inference

We establish a coherent path from a variational lower bound to a fully collapsed Gibbs sampler. The formulation clarifies the origin of the bound, following standard variational arguments based on Jensen's inequality and KL divergences [Bishop, 2006; Wainwright and Jordan, 2008; Blei *et al.*, 2017]. It also shows how the closed form arises and why it improves mixing, in the spirit of collapsed sampling for latent discrete structures [Liu, 1994; Van Dyk and Park, 2008].

After analytically integrating out the node memberships $(\boldsymbol{\pi}_i,\boldsymbol{\pi}_j)$ and the copula uniforms (u_{ij},v_{ij}), each edge (i,j) induces a finite mixture over role pairs with collapsed masses

$$p^{(ij)}_{k\ell} = \mathbb{E}_{\boldsymbol{\pi}_i,\boldsymbol{\pi}_j}\Big[V_{C_\theta}([\Pi_i(k{-}1),\Pi_i(k)] \times [\Pi_j(\ell{-}1),\Pi_j(\ell)])\Big]. \tag{7}$$

Here $p^{(ij)}_{k\ell}$ defines a valid bivariate probability mass function over role pairs, satisfying $\sum_{k=1}^K\sum_{\ell=1}^K p^{(ij)}_{k\ell} = 1$.

Combining the collapsed role mechanism with the survival layer yields the marginal likelihood contribution of edge (i,j),

$$\begin{aligned} &p(t_{ij},\delta_{ij} \mid x_{ij};\theta,\alpha,\boldsymbol{\eta},\lambda,\beta) \\ &= \sum_{k=1}^K\sum_{\ell=1}^K p^{(ij)}_{k\ell}(\theta,\alpha,\boldsymbol{\eta})\, L^{k\ell}_{\mathrm{surv}}\big(t_{ij},\delta_{ij} \mid x_{ij},\{\lambda^{k\ell}_g\}_{g=1}^G,\beta_{k\ell}\big). \end{aligned} \tag{8}$$

Consequently, the role pair (s_{ij},r_{ij}) can be updated from a discrete distribution with unnormalized weights

$$w_{ij}(k,\ell) \propto p^{(ij)}_{k\ell}(\theta,\alpha,\boldsymbol{\eta})\, L^{k\ell}_{\mathrm{surv}}\big(t_{ij},\delta_{ij} \mid x_{ij},\{\lambda^{k\ell}_g\}_{g=1}^G,\beta_{k\ell}\big),$$

which eliminates the auxiliary-variable or inner-loop Metropolis–Hastings steps required by semi-collapsed copula blockmodels.

To compute the collapsed masses in (7), we consider polynomial copulas admitting an expansion

$$C(u,v \mid \theta) = \sum_{a,b\geq 1}\gamma_{ab}(\theta)\, u^a v^b. \tag{9}$$

Under this representation, the copula rectangle volume factorizes term-wise,

$$\begin{aligned} &V_{C_\theta}([\Pi_i(k{-}1),\Pi_i(k)] \times [\Pi_j(\ell{-}1),\Pi_j(\ell)]) \\ &= \sum_{a,b}\gamma_{ab}(\theta)\big(\Pi_i(k)^a - \Pi_i(k{-}1)^a\big)\cdot\big(\Pi_j(\ell)^b - \Pi_j(\ell{-}1)^b\big). \end{aligned}$$

Since $\boldsymbol{\pi}_i \sim \mathrm{Dir}(\alpha\boldsymbol{\eta})$, the cumulative membership $\Pi_i(k) = \sum_{m\leq k}\pi_{im}$ follows a Beta distribution,

$$\Pi_i(k) \sim \mathrm{Beta}(A_k,B_k), \quad A_k = \alpha\sum_{m=1}^k \eta_m,\ B_k = \alpha - A_k.$$

Standard Dirichlet–Beta identities give the moments

$$\mathbb{E}\big[\Pi_i(k)^a\big] = \frac{(A_k)_{(a)}}{(\alpha)_{(a)}}, \quad \Delta M_i^{(a)}(k) = \frac{(A_k)_{(a)} - (A_{k-1})_{(a)}}{(\alpha)_{(a)}},$$

where $(x)_{(a)}$ denotes the rising Pochhammer symbol. Substituting these expressions into (7) yields the closed form

$$p_{k\ell}^{(ij)} = \sum_{a,b} \gamma_{ab}(\theta)\, \Delta M_i^{(a)}(k)\, \Delta M_j^{(b)}(\ell). \tag{10}$$

For the FGM copula [Nelsen, 2006], whose expansion is $C(u,v) = uv - \theta u^2 v - \theta u v^2 + \theta u^2 v^2$, this reduces to

$$\begin{aligned} p_{k\ell}^{(ij)} &= \Delta M_i^{(1)}(k)\Delta M_j^{(1)}(\ell) - \theta\, \Delta M_i^{(2)}(k)\Delta M_j^{(1)}(\ell) \\ &\quad - \theta\, \Delta M_i^{(1)}(k)\Delta M_j^{(2)}(\ell) + \theta\, \Delta M_i^{(2)}(k)\Delta M_j^{(2)}(\ell). \end{aligned} \tag{11}$$

To define a valid copula, the function $C(u,v)$ must satisfy the standard boundary conditions and be 2-increasing on $[0,1]^2$ [Joe, 2014]. These constraints ensure that the induced copula rectangle volumes $V_{C_\theta}(\cdot)$ define a proper bivariate distribution over role indices.

Although arbitrary polynomials do not in general satisfy these conditions, several well-established constructions yield polynomial expansions of the form (9) while preserving copula validity. In particular, Bernstein copulas provide a basis in which the copula constraints reduce to linear inequalities on the coefficients $\{\gamma_{ab}(\theta)\}$ [Sancetta and Satchell, 2004; Choudhuri *et al.*, 2004]. More generally, copulas defined through valid densities or spline representations admit equivalent polynomial or series expansions after integration [Joe, 2014]. Crucially, regardless of the construction, any copula admitting a separable expansion of the form (9) yields collapsed role-pair masses through (10). Hence, enforcing copula validity does not interfere with the closed-form marginalization underlying our fully collapsed sampler.

We construct a fully collapsed Markov chain Monte Carlo algorithm that integrates out the node memberships $(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)$ and copula uniforms (u_{ij}, v_{ij}), and updates role pairs and survival parameters directly from their marginal posteriors. The number of roles K, the number of time segments G, and the segment cutpoints are fixed.

At each iteration, the sampler proceeds as follows.

1. **Update role pairs** (s_{ij}, r_{ij})**.** For each edge (i,j), remove its current contribution from the global sufficient statistics and compute the collapsed role-pair masses $p_{k\ell}^{(ij)}$ using (10) or (11). The joint posterior update is then

$$\begin{aligned} &\Pr(s_{ij}=k, r_{ij}=\ell \mid -) \\ &\propto\; p_{k\ell}^{(ij)}\, L_{\text{surv}}^{k\ell}\big(t_{ij}, \delta_{ij} \mid x_{ij}, \{\lambda_g^{k\ell}\}_{g=1}^G, \beta_{k\ell}\big)\,. \end{aligned}$$

   This joint update avoids the auxiliary-variable or inner-loop Metropolis–Hastings steps required by semi-collapsed copula blockmodels.

2. **Update baseline hazards** $\{\lambda_g^{k\ell}\}$**.** With independent Gamma priors $\lambda_g^{k\ell} \sim \text{Gamma}(\kappa_\lambda, \nu_\lambda)$ for all (k,ℓ,g), where $\kappa_\lambda > 0$ is the shape and $\nu_\lambda > 0$ is the rate,

$$\lambda_g^{k\ell} \mid - \;\sim\; \text{Gamma}\Big(\kappa_\lambda + \text{fail}_g^{k\ell},\; \nu_\lambda + \text{exposure}_g^{k\ell}\Big),$$

   where

$$\begin{aligned} \text{fail}_g^{k\ell} &= \sum_{(i,j)\in\mathcal{E}:\, s_{ij}=k,\, r_{ij}=\ell} \delta_{ij}\, \mathbf{1}\{g(t_{ij}) = g\}, \\ \text{exposure}_g^{k\ell} &= \sum_{(i,j)\in\mathcal{E}:\, s_{ij}=k,\, r_{ij}=\ell} \rho_{ij}^{k\ell}\, \text{exposure}_{ij,g}, \\ \rho_{ij}^{k\ell} &= \exp(x_{ij}^\top \beta_{k\ell}). \end{aligned}$$

   Here $g(t)$ denotes the time-segment index induced by fixed cutpoints $0 = \tau_0 < \tau_1 < \cdots < \tau_G$, and $\text{exposure}_{ij,g}$ is the time-at-risk contributed by edge (i,j) within segment g.

3. **Update regression coefficients** $\{\beta_{k\ell}\}$**.** Each $\beta_{k\ell}$ is updated using a Metropolis–Hastings step targeting its conditional posterior.

4. **Update concentration** α**.** We propose α on the log scale and accept or reject using the collapsed Dirichlet–multinomial marginal likelihood.

5. **Update base weights** $\boldsymbol{\eta}$**.** We employ a mixture of local Dirichlet proposals and global independence moves.

6. **Update copula parameter** θ**.** For low-dimensional copula families, θ is updated by grid-based sampling or treated as fixed.

The survival component follows a piecewise-constant hazard model, which admits conjugate Gamma updates and is standard in Bayesian event-history analysis [Sinha and Dey, 1997]. By analytically integrating out the node memberships and copula latents, the sampler operates on a discrete marginal over role pairs, dramatically reducing posterior dependence and enabling rapid traversal of the multimodal state space [Gelman *et al.*, 1997].

4.1 Computational Complexity

The dominant computational cost of the proposed algorithm arises from the joint update of the role pair (s_{ij}, r_{ij}) for each edge. For a network with E observed edges and K latent roles, evaluating the fully collapsed conditional distribution requires computing K^2 role-pair weights $p_{k\ell}^{(ij)} L_{\text{surv}}^{k\ell}$. As a result, the per-sweep complexity of the collapsed sampler is $O(EK^2)$. In contrast, the SMMB baseline updates sender and receiver roles through conditionals that scale linearly in K, leading to an $O(EK)$ per-sweep cost. However, such semi-collapsed or uncollapsed updates typically retain stronger posterior dependence between latent variables, which can translate into slower mixing and higher autocorrelation in large networks. The fully collapsed sampler makes a deliberate trade-off: it replaces separate conditional updates by a single joint update over (k,ℓ), thereby reducing posterior coupling between sender and receiver roles at the level of each edge. This trade-off is favorable in the regime most common for blockmodels, where K is small (e.g., $K \le 10$) while E may be large. In this setting, the extra factor of K has limited impact on runtime, whereas the gain in mixing can be substantial.

As shown in Section 5.1, the fully collapsed sampler yields markedly lower autocorrelation, which reduces the number of sweeps required to reach a target Monte Carlo precision. Consequently, despite the higher per-sweep complexity, the total wall-clock time to effective convergence can be lower than that of semi-collapsed baselines. Moreover, the algorithm scales linearly in the number of edges E, making it suitable for large sparse networks. This effect is reflected in the large ESS/s gains reported in Section 5.1.

5 Experiments

5.1 Simulation: MCMC Sampling Efficiency

To evaluate the practical benefits of the proposed fully collapsed inference scheme, we compare its sampling efficiency with two semi-collapsed baselines under the FGM copula specification. The goal is to isolate the effect of marginalizing out both the membership vectors $(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)$ and the copula uniforms (U, V) on the mixing behavior of the Markov chain.

We consider three samplers that differ only in the set of latent variables that are analytically integrated out. The first two correspond to the partially collapsed schemes studied in Fan *et al.* [2016], while the third is the fully collapsed method proposed in this paper:

1. *CMMSB*$_\pi$ (node-membership marginalization). The copula uniforms (U, V) are integrated out, but the node membership vectors $\boldsymbol{\pi}_i$ are sampled explicitly. Each $\boldsymbol{\pi}_i$ is updated by Metropolis–Hastings based on the product of likelihood contributions from all edges incident to node i, after which (s_{ij}, r_{ij}) is sampled from the conditional rectangle probabilities $p_{k\ell}^{(ij)}(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)$.
2. *CMMSB*$_{uv}$ (copula-uniform marginalization). The membership vectors $(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)$ are integrated out, while the copula uniforms (U, V) are sampled. Each pair (u_{ij}, v_{ij}) is updated by Metropolis–Hastings given (s_{ij}, r_{ij}), after which (s_{ij}, r_{ij}) is resampled from the induced prior distribution with leave-one-out counts.
3. *CMMSB*$_{Full}$ (full marginalization). Both $(\boldsymbol{\pi}_i, \boldsymbol{\pi}_j)$ and (U, V) are integrated out, and each role pair (s_{ij}, r_{ij}) is sampled directly from the closed-form discrete distribution $p_{k\ell}^{(ij)}$ derived in Section 4.

All samplers are run under identical settings. We simulate networks with $K = 4$ latent roles and $N = 36$ nodes, yielding $E = 1260$ directed edges. The Dirichlet concentration is set to $\alpha_0 = 1$, and the FGM copula parameter is fixed at $\theta = 0.5$. Each chain is run for 5000 sweeps with 500 burn-in iterations, and three independent chains are generated for each method.

To assess mixing, we monitor the match rate

$$\mathrm{MR} = |\mathcal{D}|^{-1} \sum_{(i,j)\in\mathcal{D}} \mathbf{1}\{s_{ij} = r_{ij}\},$$

together with the log prior. Such low-dimensional summaries of the latent configuration are commonly used as diagnostics for convergence and mixing in latent variable models [Brooks and Gelman, 1998]. Sampling efficiency is quantified using the Integrated Autocorrelation Time (IACT) [Geyer, 1992], the Effective Sample Size (ESS) [Kass and Raftery, 1995],

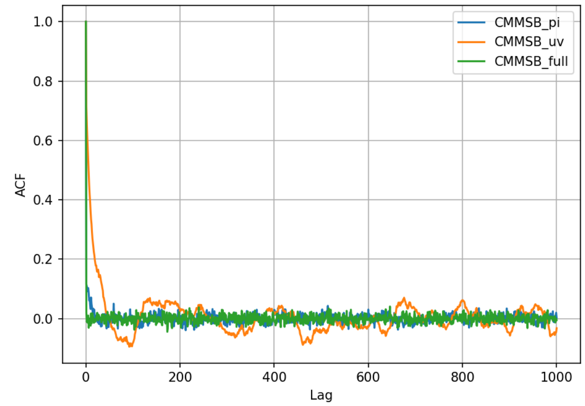

Figure 2: Autocorrelation functions of the match rate. The fully collapsed sampler (CMMSB$_{\mathrm{Full}}$, green) exhibits near-zero autocorrelation beyond lag one, whereas the semi-collapsed baselines show persistent serial dependence. Settings: $K = 4$, $N = 36$, FGM copula $\theta = 0.5$.

Method	ESS	IACT	MCSE	Sweeps/s	ESS/s
CMMSB$_\pi$	1474	3.05	3.3×10^{-4}	3.1	0.91
CMMSB$_{uv}$	214	21.01	9.2×10^{-4}	1.8	0.08
CMMSB$_{\mathrm{Full}}$	**4336**	**1.04**	$\mathbf{1.9 \times 10^{-4}}$	**400.2**	**347.1**

Table 2: Comparison of MCMC sampling efficiency across three inference schemes.

and the Monte Carlo Standard Error (MCSE) computed via batch means [Flegal *et al.*, 2008; Jones *et al.*, 2006].

Figure 2 highlights a substantial difference in mixing behavior. For CMMSB$_{\mathrm{Full}}$, the autocorrelation of the match rate drops essentially to zero after the first lag, indicating that successive draws are nearly independent. By contrast, both semi-collapsed samplers display pronounced autocorrelation over many lags, reflecting slow exploration of the posterior.

These differences are quantified in Table 2. The fully collapsed sampler attains the largest effective sample size (ESS $= 4336$) and an IACT close to one (1.04), indicating near-ideal mixing. This improvement directly translates into higher estimation precision, with CMMSB$_{\mathrm{Full}}$ achieving the smallest Monte Carlo standard error.

In terms of computational throughput, CMMSB$_{\mathrm{Full}}$ executes 400.2 sweeps per second, compared with 3.1 and 1.8 for the two semi-collapsed schemes. When combined with the gains in mixing, this yields more than two orders of magnitude improvement in effective samples per second (ESS/s).

Taken together, these results show that analytically integrating out both membership vectors and copula coordinates removes the auxiliary dependencies that dominate the posterior geometry in semi-collapsed copula blockmodels. In this experiment, the resulting fully collapsed sampler attains more than a 10^2-fold increase in effective samples per second relative to the two semi-collapsed baselines, reflecting both substantially faster mixing and lower per-sweep overhead.

Model	Time (s)	$\mathrm{ESS}_{\mathrm{ll}}$	ESS/s	$\hat{R}$	IACT
MMSB	1092	22.9	0.024	4.39	5.11
$\mathrm{CMMSB}_{\mathrm{Full}}$	1622	**37.2**	**0.069**	**1.26**	**1.45**

Table 3: Sampling efficiency and convergence diagnostics on Enron ($K = 4$, $G = 5$). ESS/s denotes effective sample size per second. Reported values are medians over three independent chains.

Model	DIC (p_D)	Δ DIC	Hold-out	Δ Hold-out
MMSB	244,494	—	−2.1987	—
$\mathrm{CMMSB}_{\mathrm{Full}}$	**241,284**	**−3,210**	**−2.1766**	**+0.0221**

Table 4: Model fit (DIC) and predictive log-likelihood on held-out data ($K = 4$).

Metric	$\mathrm{CMMSB}_{\mathrm{Full}}$	MMSB	Δ
Prior MI	2.75×10^{-3}	0.000	$+2.75 \times 10^{-3}$
Prior Lift	1.133	1.000	+0.133
Post Entropy H	**0.384**	0.412	−0.028
Post Top-1 Mass	**0.850**	0.839	+0.011
Pair-cond. MI	**0.039**	0.112	−0.073
Pair-cond. $\mathrm{Lift}_{0.95}$	**2.463**	4.711	−2.248

Table 5: Structural diagnostics of prior assumptions and posterior role assignments.

5.2 Real Data Analysis: Enron Corpus

We apply the proposed model to the Enron e-mail corpus, a standard benchmark for temporal and relational network analysis [Klimt and Yang, 2004]. Following the preprocessing protocol of Song *et al.* [2024], each directed edge (i, j) is associated with an observed event time t_{ij}, a right-censoring indicator δ_{ij}, and a vector of covariates x_{ij}. Event times are linked to latent role pairs through the piecewise-constant hazard specification described in Section 2.

We fix the number of roles to $K = 4$ and the number of time segments to $G = 5$, using the same cutpoints as in Song *et al.* [2024]. For the copula family, we use the FGM copula with parameter $\theta = 0.30$, chosen to reflect moderate positive dependence. The fully collapsed $\mathrm{CMMSB}_{\mathrm{Full}}$ sampler is compared against the standard MMSB baseline in terms of both sampling behavior and statistical fit.

Convergence and Mixing. We first evaluate robustness of posterior exploration using short MCMC runs $(1000/500/10)$, where the notation denotes total iterations, burn-in, and thinning. Table 3 reports standard multi-chain diagnostics. The MMSB baseline exhibits severe lack of convergence, with a Gelman–Rubin statistic $\hat{R} = 4.39$ [Gelman and Rubin, 1992], indicating that chains remain trapped in distinct posterior modes. In contrast, the collapsed $\mathrm{CMMSB}_{\mathrm{Full}}$ sampler achieves $\hat{R} = 1.26$, reflecting substantially improved mixing. Although $\mathrm{CMMSB}_{\mathrm{Full}}$ has a higher per-iteration cost, its effective sampling rate is nearly three times larger (median ESS/s 0.069 versus 0.024), showing that the cost of collapsing is more than offset by the reduction in autocorrelation.

Model Fit and Predictive Performance. Using longer chains $(2400/1200/10)$, we evaluate model fit and out-of-sample predictive performance. Table 4 reports the Deviance Information Criterion (DIC) [Spiegelhalter *et al.*, 2002] and the posterior-averaged log-likelihood on a held-out test set. The fully collapsed copula model attains a substantially lower DIC than MMSB, with a difference exceeding 3000, indicating a markedly better tradeoff between goodness of fit and model complexity. The hold-out log-likelihood improves by $+0.0221$ nats per edge, which accumulates to approximately 280 nats over the full test set. This improvement reflects the fact that allowing dependence between sender and receiver roles captures systematic structure that the independence assumption of MMSB cannot represent.

Posterior Structure. To examine how the copula affects posterior uncertainty, we summarize the latent role distributions using entropy, top-1 posterior mass, and pair-conditional dependence measures, reported in Table 5. The MMSB baseline yields higher posterior entropy and lower top-1 mass, indicating greater uncertainty in role assignments. By contrast, $\mathrm{CMMSB}_{\mathrm{Full}}$ produces more concentrated posteriors, with lower entropy and a higher probability assigned to the dominant role. The copula also reduces pair-conditional mutual information, showing that role assignments rely less on idiosyncratic pair-specific interactions and more on the global dependence structure captured by θ. This leads to a more stable and interpretable representation of the network.

6 Discussion

We proposed a fully collapsed copula mixed-membership blockmodel that integrates dependence between sender and receiver roles into a dynamic survival network while retaining exact inference. By analytically marginalizing both node memberships and copula coordinates, the resulting Gibbs sampler operates on a purely discrete state space, eliminating the auxiliary-variable and inner-loop sampling that typically dominate copula-based blockmodels. This structural simplification empirically demonstrates several strengths. Across both simulations and the Enron benchmark, full marginalization yields orders-of-magnitude gains in effective sampling rate and substantially improves posterior stability and predictive fit. These results indicate that much of the apparent uncertainty and poor mixing in standard blockmodels is induced by the independence assumption rather than by intrinsic ambiguity in the data. The current formulation assumes a fixed node set and a parametric copula family. Extending the fully collapsed construction to settings with evolving node populations or more flexible dependence structures is an important direction for future work. More broadly, the framework illustrates how non-conjugate dependence in dynamic networks can be handled through exact marginalization rather than approximation, providing a principled route to scalable Bayesian inference in complex relational models.

References

[Airoldi *et al.*, 2008] Edo M Airoldi, David Blei, Stephen Fienberg, and Eric Xing. Mixed membership stochastic blockmodels. *Advances in neural information processing systems*, 21, 2008.

[Allman *et al.*, 2009] Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37(6A), December 2009.

[Bishop, 2006] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[Blei *et al.*, 2017] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.

[Brooks and Gelman, 1998] Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

[Choudhuri *et al.*, 2004] Nidhan Choudhuri, Subhashis Ghosal, and Anindya Roy. Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059, 2004.

[Damlen *et al.*, 1999] Paul Damlen, John Wakefield, and Stephen Walker. Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):331–344, 1999.

[Fan *et al.*, 2016] Xuhui Fan, Richard Yi Da Xu, and Longbing Cao. Copula mixed-membership stochastic block model. In *IJCAI International Joint Conference on Artificial Intelligence*, 2016.

[Flegal *et al.*, 2008] James M Flegal, Murali Haran, and Galin L Jones. Markov chain monte carlo: Can we trust the third significant figure? *Statistical Science*, pages 250–260, 2008.

[Gelman and Rubin, 1992] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[Gelman *et al.*, 1997] Andrew Gelman, Walter R Gilks, and Gareth O Roberts. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.

[Geyer, 1992] Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pages 473–483, 1992.

[Gharari *et al.*, 2023] Ali Hossein Foomani Gharari, Michael Cooper, Russell Greiner, and Rahul G Krishnan. Copula-based deep survival models for dependent censoring. In *Uncertainty in Artificial Intelligence*, pages 669–680. PMLR, 2023.

[Griffiths and Steyvers, 2004] Thomas L Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1):5228–5235, 2004.

[Hoff, 2005] Peter D Hoff. Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association*, 100(469):286–295, 2005.

[Joe, 2014] Harry Joe. *Dependence modeling with copulas*. CRC press, 2014.

[Jones *et al.*, 2006] Galin L Jones, Murali Haran, Brian S Caffo, and Ronald Neath. Fixed-width output analysis for markov chain monte carlo. *Journal of the American Statistical Association*, 101(476):1537–1547, 2006.

[Kalbfleisch and Prentice, 2002] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*. John Wiley & Sons, 2002.

[Karrer and Newman, 2011] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107, 2011.

[Kass and Raftery, 1995] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[Klimt and Yang, 2004] Bryan Klimt and Yiming Yang. Introducing the enron corpus. In *CEAS*, volume 45, pages 92–96, 2004.

[Kumar *et al.*, 2019] Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1269–1278, 2019.

[Liu, 1994] Jun S Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966, 1994.

[Neal, 2000] Radford M Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265, 2000.

[Nelsen, 2006] Roger B Nelsen. *An introduction to copulas*. Springer, 2006.

[Perry and Wolfe, 2013] Patrick O Perry and Patrick J Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(5):821–849, 2013.

[Rastelli and Fop, 2020] Riccardo Rastelli and Michael Fop. A stochastic block model for interaction lengths. *Advances in Data Analysis and Classification*, 14(2):485–512, 2020.

[Sancetta and Satchell, 2004] Alessio Sancetta and Stephen Satchell. The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric theory*, 20(3):535–562, 2004.

[Sinha and Dey, 1997] Debajyoti Sinha and Dipak K Dey. Semiparametric bayesian analysis of survival data. *Journal of the American Statistical Association*, 92(439):1195–1212, 1997.

[Sit *et al.*, 2021] T Sit, Z Ying, and Yi Yu. Event history analysis of dynamic networks. *Biometrika*, 108(1):223–230, 2021.

[Song *et al.*, 2024] Fangda Song, Jing Chu, Shuangge Ma, and Yingying Wei. Survival mixed membership blockmodel. *Journal of the American Statistical Association*, 119(546):1647–1656, 2024.

[Spiegelhalter *et al.*, 2002] David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639, 2002.

[Stephens, 2000] Matthew Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809, 2000.

[Teicher, 1963] Henry Teicher. Identifiability of finite mixtures. *The annals of Mathematical statistics*, pages 1265–1269, 1963.

[Trivedi *et al.*, 2019] Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. Dyrep: Learning representations over dynamic graphs. In *International conference on learning representations*, 2019.

[Van Dyk and Park, 2008] David A Van Dyk and Taeyoung Park. Partially collapsed gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103(482):790–796, 2008.

[Wainwright and Jordan, 2008] Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2):1–305, 2008.

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