

A Counting Process View of Relational Event Models: Practical Asymptotics

Cornelius Fritz ^{1*}† and Alexander Fuchs-Kreiß ^{2†}

¹*School of Computer Science and Statistics, Trinity College Dublin, Dublin, D02N8P5, Co. Dublin, Republic of Ireland.

²Institute for Mathematics, Mathematics Education and Computer Science Education - IMMI, University of Hildesheim, Samelsonplatz 1, Hildesheim, 31141, Lower Saxony, Germany.

*Corresponding author(s). E-mail(s): c.fritz@tcd.ie;
Contributing authors: fuchskreiss@uni-hildesheim.de;

†These authors contributed equally to this work.

Keywords: Asymptotic normality, Counting processes, Relational event models

Relational Event Models (REMs) provide a rigorous framework for analyzing dyadic interactions observed in continuous time, capturing history-dependent dynamics such as triadic closure and reciprocity. Framing REMs through the lens of counting processes embeds the model in a rich theoretical foundation, facilitating its mathematical analysis. While Maximum Likelihood Estimation (MLE) is standard practice for estimating these models, the underlying statistical guarantees rely on specific asymptotic regimes, namely, whether the network size (n), the observational period (T), or both approach infinity. We review the theoretical foundations of such counting-process-based models, formalizing the core assumptions required to achieve asymptotic normality across these different limits. With a specific focus on Cox-type multiplicative models, we detail the circumstances under which these assumptions hold. Supported by simulation studies, we illustrate how structural modeling choices, including temporal windowing and logarithmic transformations, affect empirical coverage and estimator convergence. We thereby derive several guiding principles for specifying such models in realistic contexts, bridging theory and practice.

1 Introduction

With the advent of computational social science, continuously collected interaction data at scale, measuring, for instance, e-mail traffic via logged data or face-to-face interactions via sociometric badges, are becoming widely available [1, 2]. Contrary to survey data [3], we can easily measure such event data at scale and with fine-grained temporal resolution. This data enables insights into micro behavior within organizations, such as hospitals [4], or in disrupted environments, such as the September 11 attacks on the World Trade Center [5].

The statistical unit in this context is a relational event, representing instantaneous interactions between entities at a given point in time. Denoting the set of observed entities by $\mathcal{P}_n := \{1, \dots, n\}$ and the observational period by $\mathcal{T}_T := [0, T]$ with $T > 0$, a relational event is denoted by a tuple $e = (i, j, t)$ for an event that occurred between sender $i \in \mathcal{P}_n$ and receiver $j \neq i \in \mathcal{P}_n$ at time $t \in \mathcal{T}_T$. In the examples mentioned in the opening paragraph, an event $e = (i, j, t)$ corresponds to a patient transfer between hospitals i and j at time t in Kitts et al. [4], while it corresponds to communication between two responder groups during the World Trade Center attack in Renshaw et al. [5]. We refer to the sender, receiver, and time of the m th event by i_m, j_m , and t_m , respectively. Events can more broadly be associated with a particular direction, duration, and weight [6] and even occur between sets of entities [7]. Most material developed in this manuscript generalizes to such settings, but for the sake of simplicity, we only cover undirected dyadic interactions in this article. The set of sender and receiver pairs (i, j) is then given by $\mathcal{D}_n := \{(i, j) \mid i, j \in \mathcal{P}_n, i < j\}$.

We specify generative models for these dyadic events via $n(n-1)/2$ -dimensional counting processes $\mathbf{N}(t) = (N_{i,j}(t))_{(i,j) \in \mathcal{D}_n}$. The (i, j) th coordinate of $\mathbf{N}(t)$ counts the number of observed events between entities i and j that were observed until and including any point in time $t \in \mathcal{T}_T$. To simplify later notation, we write $N_{i,j}(t) = N_{j,i}(t)$ for $i > j$. We characterize the instantaneous probability of a jump in any coordinate of these counting processes by the intensity $\boldsymbol{\lambda}(t) = (\lambda_{i,j}(t))_{(i,j) \in \mathcal{D}_n}$. Different parametrizations for this intensity are surveyed in Fritz et al. [8] and Bianchi et al. [9].

The common quantitative research pipeline starts by theorizing about potential statistics that inform interactions based on prior theoretical research, such as [10, 11], that operationalize testable hypotheses. Neighboring entities in interaction networks are unlikely to behave independently, therefore, any realistic theory for such complex behavior puts dependence between studied entities at its core [12]. For instance, if entities i and j have previously interacted with a common partner h , they might be more likely to interact with one another in the future [13]. In the words of Butts [3, p. 160], “history creates the context for present action”. An appropriate statistical model in this context can encode such dependencies in different model specifications. Both frequentist and Bayesian approaches can be used to draw inferential conclusions given the interaction data at hand. To our knowledge, most approaches within the Bayesian paradigm are methodological papers, e.g., [14, 15], while the entirety of substantive application papers relied on frequentist approaches, e.g., [4, 5, 16, 17]. We therefore here focus on frequentist approaches where the parameters are estimated via Maximum Likelihood.

To carry out statistical tests to decide between different possible data-generating mechanisms or obtain confidence intervals, we heavily rely on the sampling distribution of the Maximum Likelihood estimate. Since this distribution is generally unavailable for finite samples, we need a central limit theorem to guarantee that the sampling distribution can be approximated by a multivariate normal distribution under some asymptotic regime. For event data studied in this article, one can think of different asymptotic regimes:

1. $T \rightarrow \infty$ with fixed n : Assume we are studying interactions among a fixed set of entities over some time and intend to draw inferential conclusions on where the estimates would lie if we were to observe the set for longer;
2. $n \rightarrow \infty$ with fixed T : Assume we are studying a growing population of entities over a fixed time and intend to draw inferential conclusions based on a subsample where all sufficient statistics only depend on other entities also in the sample.
3. T and $n \rightarrow \infty$: Assume we are studying a growing population of entities over an increasing observational period. This setting is the least explored for counting-based models and we need to assume some growth rates of n and T .

While all asymptotic regimes can lead to a valid central limit theorem, different conditions must be met in each case. We argue that practitioners should carefully assess these while specifying their models and interpreting the results. We first describe in Section 2 how counting processes serve as a natural framework for modeling relational event data and then review in Section 3 theoretical results. We here focus on the necessary assumptions and what they imply in practical settings. Using simulations and theoretical results, we illustrate in Section 4 how particular choices affect inference results, drawing on the survival analysis literature on independent observations.

2 A Counting Process Framework for Modeling Relational Event Data

Relational event data is a distinct type of time-to-event data in which events occur between pairs of entities, such as an email sent between them [18], rather than a single entity, such as a patient passing away. The most general framework for time-to-event data is based on counting processes developed by Andersen and Gill [19], which we adapt to relational event data. Introducing advanced mathematical concepts such as martingales, filtration, and predictability would make the review of more involved mathematical concepts necessary. Thus, we refer to Andersen et al. [20] and Aalen et al. [21] for a thorough treatment of these foundations.

A counting process for pair $(i, j) \in \mathcal{D}_n$ on the interval \mathcal{T}_T is a stochastic process $N_{i,j} : \mathcal{T}_T \rightarrow \mathbb{N}_0 := \{0, 1, 2, \dots\}$, whose state at time $t \in \mathcal{T}_T$ is defined by

$$N_{i,j}(t) = \sum_{m \geq 1} \mathbb{I}(e_m = (i, j, t_m) \wedge t_m \leq t), \quad (1)$$

and counts the number of interactions between entities i and j with $(i, j) \in \mathcal{D}_n$. We distinguish between the entire stochastic process, $N_{i,j} := \{N_{i,j}(t) \mid t \geq 0\}$, and its status at $t \in \mathcal{T}_T$, $N_{i,j}(t)$, and collect the counting processes of all pairs of entities in

$\mathbf{N} := (N_{i,j})_{(i,j) \in \mathcal{D}_n}$. By Equation (1), each coordinate of this multivariate counting process is nonnegative, integer-valued, and adapted to the natural filtration $\mathcal{F}_t := \sigma(\{\mathbf{N}(u) : u \leq t\})$, which is a so-called σ -field growing with time. One can think of \mathcal{F}_t as the collection of all available information until and including time t . If additional information, such as covariate information for each unit or dyad, is available over time, we can augment the filtration \mathcal{F}_t accordingly.

Given these properties, one can show that \mathbf{N} is a sub-martingale, allowing us to write its (i,j) th coordinate as a superposition of a martingale and compensator process:

$$N_{i,j}(t) = M_{i,j}(t) + \Lambda_{i,j}(t), \quad (2)$$

for all $t \in \mathcal{T}_T$ and $(i,j) \in \mathcal{D}_n$. Equation (2) is called the Doob-Meyer decomposition. The martingale $M_{i,j} : \mathcal{T}_T \rightarrow \mathbb{R}$ is a stochastic process such that $\mathbb{E}(M_{i,j}(t) \mid \mathcal{F}_s) = M_{i,j}(s)$ for all $s \leq t$. One can understand the martingale component as a residual or noise process without any structure related to the available information \mathcal{F}_t . The compensator process, on the other hand, is the part of the counting process that is predictable with the available information.

Since the compensator is a non-decreasing predictable process, we assume that it can be written as an integral

$$\Lambda_{i,j}(t) = \int_0^t \lambda_{i,j}(r) dr \quad (3)$$

for some non-negative stochastic process $\lambda_{i,j} : \mathcal{T}_T \rightarrow [0, \infty)$ for each $(i,j) \in \mathcal{D}_n$, called intensity. Thus, we here implicitly assume that the process defined through Equation (3) is absolutely continuous. In case the compensator $\Lambda_{i,j}$ experienced a jump at t_0 when $N_{i,j}(t_0)$ experiences a jump with positive probability at t_0 , it would prohibit the integral representation in Equation (3). Therefore, we exclude such discrete behavior and focus on the continuous case. The matrix $\boldsymbol{\lambda} = (\lambda_{i,j})_{(i,j) \in \mathcal{D}_n}$ defines in its (i,j) th coordinate the instantaneous probability of a jump in the counting process $N_{i,j}(t)$:

$$\lambda_{i,j}(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}(N_{i,j}(t+h) - N_{i,j}(t) = 1 \mid \mathcal{F}_{t^-}), \quad (4)$$

where t^- denotes the time point right before $t \in \mathcal{T}_T$. The information from t is here excluded since what happens at instance t should not affect $\lambda_{i,j}(t)$.

While the formulation in Equation (4) is nonparametric and potentially dependent on the complete filtration \mathcal{F}_{t^-} , in most real-life applications it is more practical to assume that some low-dimensional vector encodes all necessary information from the past. To describe this low-dimensional information, we introduce the sufficient statistics $\mathbf{s}_{i,j} : \mathcal{T}_T \rightarrow \mathbb{R}^p$ to map the high-dimensional historical information to a p -dimensional covariate space. We omit the specific dependence on \mathcal{F}_{t^-} and write $\mathbf{s}_{i,j}(t) = (s_{i,j,1}(t), \dots, s_{i,j,p}(t))^\top$ but keep in mind that $\mathbf{s}_{i,j}(t)$ may only depend on information that is observable before time t . These statistics summarize all dependence of $\lambda_{i,j}(t)$ on the past and facilitate the integration of substantive theory in model specification. For example, the theory of triadic clustering [13] can be assessed

by incorporating the count of common partners of pair $(i, j) \in \mathcal{D}_n$ in the q th coordinate of the sufficient statistics via $s_{i,j,q}(t) = \sum_{h \notin \{i,j\}} \mathbb{I}(N_{i,h}(t^-) > 0) \mathbb{I}(N_{h,j}(t^-) > 0)$ for some $q \in \{1, \dots, p\}$. Similarly, the conversational norm of reciprocal communication [22] can be captured in the statistic $s_{i,j,q}(t) = \mathbb{I}(\tilde{e}_t = (j, i))$, with \tilde{e}_t denoting the tuple of sender and receiver of the last event observed before time t . Consequently, sufficient statistics $\mathbf{s}_{i,j}(t)$ are the primary tool for encoding arbitrary dependence assumptions of past events on future events.

The final model component is the specification of the structural relationship between $\mathbf{s}_{i,j}(t)$ and $\lambda_{i,j}(t)$. First, we transform the sufficient statistics $\mathbf{s}_{i,j}(t)$ to a scalar value $\eta_{i,j,t} = \boldsymbol{\theta}^\top \mathbf{s}_{i,j}(t)$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$ is a p -dimensional unknown parameter vector. Second, we transform this value by a function $g : \mathbb{R} \rightarrow \mathbb{R}^+$ that is assumed to be strictly positive, bounded away from zero, and three times continuously differentiable. The intensity is then given by:

$$\lambda_{i,j}(t | \boldsymbol{\theta}) = g(\eta_{i,j,t}), \quad (5)$$

which generalizes several prominent model classes within the literature. In this paper, we assume a correctly specified model, such that there exists a true parameter $\boldsymbol{\theta}^*$ such that $(\lambda_{i,j}(t | \boldsymbol{\theta}^*))_{(i,j) \in \mathcal{D}_n}$ is equal to the intensity functions associated with \mathbf{N} .

Example 2.1 (Additive Models). Letting $g(x) = x$ and defining the sufficient statistic vector as $\mathbf{s}_{i,j}(t) = (1, \int_0^t \exp(-(t-u)) dN_{i,j}(u))^\top$ yields a Hawkes process for the interaction data [23]. In this example, only previous events between entities $(i, j) \in \mathcal{D}_n$ affect the intensity $\lambda_{i,j}(t | \boldsymbol{\theta})$ additively in the near future, where the past is exponentially down-weighted. We need to make sure that $\mathbf{s}_{i,j}(t)^\top \boldsymbol{\theta} > 0$ holds for all $(i, j) \in \mathcal{D}_n$ and $t \in \mathcal{T}_T$. Several comparable methods were proposed in the literature in this additive setting [24–27].

Example 2.2 (Multiplicative Models). Setting $g : x \mapsto \exp(x)$, yields the original Relational Event Models as discussed in Butts [28]. The effect of the q -th coordinate $s_{i,j,q}(t)$ of the sufficient statistics increasing by one unit, is then given by the multiplicative factor $\exp(\theta_q)$. Much literature on such Cox-type models has been developed [14, 29–31].

3 Practical Asymptotics

The choice of sufficient statistics reflects competing hypotheses regarding the underlying data-generating process. In applied research, selecting the most appropriate specification requires a principled statistical framework. Standard inferential tools, such as confidence intervals and hypothesis tests, typically rely on the asymptotic distribution of the estimator $\hat{\boldsymbol{\theta}}$. In this section, we will discuss, how the asymptotic regimes where either $T \rightarrow \infty$ or $n \rightarrow \infty$ or both, affect the mathematical analysis. To keep our notation general, we introduce the sample size $\mathcal{S} := \mathcal{S}(n, T)$ and write $\mathcal{S} \rightarrow \infty$ for either $n \rightarrow \infty$, $T \rightarrow \infty$ or both. In the classical setting of iid data, we would expect that T is fixed and $n \rightarrow \infty$. However, in an REM context, it is not

always reasonable to assume that units behave independently. But, as we will point out, this is not necessarily a problem as long as behavior of a new unit brings new information that can be leveraged for estimating the true data-generating parameter denoted by $\boldsymbol{\theta}^*$. On the other hand, from a time-series point of view, we would expect that n is fixed and $T \rightarrow \infty$. In such a setting, observing the units for longer times provides more information for estimating $\boldsymbol{\theta}^*$. In this context, dependence between units is natural, but a certain stationarity in time will be required.

Our focus lies on the maximum likelihood estimator

$$\hat{\boldsymbol{\theta}}_S := \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \ell_S(\boldsymbol{\theta}), \quad (6)$$

where the log-likelihood of the counting-process based models developed above is

$$\ell_S(\boldsymbol{\theta}) := \sum_{(i,j) \in \mathcal{D}_n} \left(\int_{\mathcal{T}_T} \log \lambda_{i,j}(t | \boldsymbol{\theta}) dN_{i,j}(t) - \int_{\mathcal{T}_T} \lambda_{i,j}(t | \boldsymbol{\theta}) dt \right).$$

In practice, the maximizer in Equation (6) can often be found via iterative methods, such as Newton-Raphson [30], Fisher scoring [28], or Expectation Maximization algorithms [24]. In Section 3.1, we will start to discuss the general mathematical starting point to obtain an asymptotic distribution of the resulting Maximum Likelihood estimator, independent of the exact regime of interest. In Section 3.2, we then emphasize for the case of a multiplicative Cox-type model, how the regimes affect the mathematical analysis.

To explain the relevance of the assumptions that will be mentioned below, we will use the following definitions

$$U_S(\boldsymbol{\theta}) := \partial_{\boldsymbol{\theta}} \ell_S(\boldsymbol{\theta}) = \sum_{(i,j) \in \mathcal{D}_n} \left(\int_{\mathcal{T}_T} \frac{\partial_{\boldsymbol{\theta}} \lambda_{i,j}(t | \boldsymbol{\theta})}{\lambda_{i,j}(t | \boldsymbol{\theta})} dN_{i,j}(t) - \int_{\mathcal{T}_T} \partial_{\boldsymbol{\theta}} \lambda_{i,j}(t | \boldsymbol{\theta}) dt \right), \quad (7)$$

$$H_S(\boldsymbol{\theta}) := - \partial_{\boldsymbol{\theta}}^2 \ell_S(\boldsymbol{\theta}),$$

$$F_S(\boldsymbol{\theta}) := \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} \left(\frac{\partial_{\boldsymbol{\theta}} \lambda_{i,j}(t | \boldsymbol{\theta})}{\lambda_{i,j}(t | \boldsymbol{\theta})} \right) \left(\frac{\partial_{\boldsymbol{\theta}} \lambda_{i,j}(t | \boldsymbol{\theta})}{\lambda_{i,j}(t | \boldsymbol{\theta})} \right)^{\top} \lambda_{i,j}(t | \boldsymbol{\theta}) dt, \quad (8)$$

where $\partial_{\boldsymbol{\theta}}$ and $\partial_{\boldsymbol{\theta}}^2$ denote the gradient and Hessian matrix with respect to $\boldsymbol{\theta}$, respectively¹. Note that $\log \lambda_{i,j}(t | \boldsymbol{\theta})$ appears as the derivative $\partial_{\boldsymbol{\theta}} \log \lambda_{i,j} = \partial_{\boldsymbol{\theta}} \lambda_{i,j} / \lambda_{i,j}$ in integrals with respect to $N_{i,j}$ or when multiplied with $\lambda_{i,j}(t | \boldsymbol{\theta}^*)$. Thus, we restrict all integrals to $\{t \in [0, T] : \lambda_{i,j}(t | \boldsymbol{\theta}^*) > 0\}$ to avoid taking the logarithm of 0. Typically, U_S and H_S are called score function and observed information, respectively. The random variable F_S is related to the variance of the score function and we will discuss later in which cases both H_S and F_S converge to the Fisher information.

¹Hence, we implicitly assume that g is sufficiently often differentiable.

3.1 Asymptotic normality for parametric models

We give now relatively straightforward assumptions that one has to check to obtain asymptotic normality of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ in a parametric REM.

Assumption 1 (Smoothness and bounded statistics). The function $g : \mathbb{R} \rightarrow (0, \infty)$ is three times continuously differentiable. The parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$ is bounded. Moreover, the sufficient statistics are uniformly bounded, that is,

$$\sup_{n \in \mathbb{N}} \sup_{t \in \mathcal{T}_T} \max_{(i,j) \in \mathcal{D}_n} \|\mathbf{s}_{i,j}(t)\| \leq M \text{ almost surely}$$

for some $M < \infty$.

Assumption 2 (Convergence of observed information). For $a_{\mathcal{S}} = \sqrt{Tn(n-1)/2} \rightarrow \infty$ it holds that

$$a_{\mathcal{S}}^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} (\partial_{\theta_k} \partial_{\theta_l} \log \lambda_{i,j}(t | \boldsymbol{\theta}^*))^2 \lambda_{i,j}(t; \boldsymbol{\theta}^*) dt$$

converges in probability to a finite constant for all $k, l \in \{1, \dots, p\}$. There is a positive definite matrix $\Sigma(\boldsymbol{\theta}^*) \in \mathbb{R}^{p \times p}$ such that

$$a_{\mathcal{S}}^{-2} F_{\mathcal{S}}(\boldsymbol{\theta}^*) \xrightarrow{\mathbb{P}} \Sigma(\boldsymbol{\theta}^*).$$

Assumption 3 (Scaling). The sequence

$$a_{\mathcal{S}}^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} \lambda_{i,j}(t | \boldsymbol{\theta}^*) dt$$

converges in probability to a finite constant.

The smoothness requirements on g in Assumption 1 are often fulfilled in practice and ensure that some technical conditions are met. The boundedness of the sufficient statistics can, for example, be achieved by rescaling. We return to this point in remark 3.3. The choice of $a_{\mathcal{S}}$ in Assumption 2 is natural in view of the required convergences because $n(n-1)/2$ equals the number of unordered pairs and T is the length of the integral. It will turn out in Theorem 3.1 below that $a_{\mathcal{S}}$ determines the rate of convergence of $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ to $\boldsymbol{\theta}^*$. The limit $\Sigma(\boldsymbol{\theta}^*)$ in Assumption 2 plays the role of the Fisher information, and the two convergences in Assumption 2 together imply that

$$a_{\mathcal{S}}^{-2} H_{\mathcal{S}}(\boldsymbol{\theta}^*) \xrightarrow{\mathbb{P}} \Sigma(\boldsymbol{\theta}^*) \tag{9}$$

for $\mathcal{S} \rightarrow \infty$, which we would expect because we consider correctly specified models. Assumption 2 is the main requirement that one has to check for a specific model. We

provide more details in case of the Cox model in Section 3.2. The final Assumption 3 is posed for technical reasons. Given Assumption 2, it does not pose any substantial constraint because the convergence requirements of Assumption 2 are very similar to those of Assumption 3. For the Cox-type model discussed in Section 3.2, we even observe that Assumption 3 is not required.

Under these assumptions, the following theorem guarantees asymptotic normality, as a direct consequence of the general results in Andersen et al. [20].

Theorem 3.1 (Theorems VI.1.1 and VI.1.2 [20]). Let Assumptions 1-3 hold. Then, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$ is consistent and

$$a_{\mathcal{S}} \left(\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^* \right) \xrightarrow{d} \mathcal{N}(0, \Sigma(\boldsymbol{\theta}^*)^{-1}).$$

We discuss in Appendix A how our Assumptions 1-3 imply the conditions in Andersen et al. [20]. To give some insight into the proof, we provide a short discussion here. Assumptions 1 and 2 imply that $U_{\mathcal{S}}(\boldsymbol{\theta}^*)$ behaves asymptotically like a normal distribution. Plugging the Doob-Meyer decomposition from Equation (2) into Equation (7), we obtain

$$\begin{aligned} U_{\mathcal{S}}(\boldsymbol{\theta}^*) &= \sum_{(i,j) \in \mathcal{D}_n} \left(\int_{\mathcal{T}_T} \frac{\partial_{\boldsymbol{\theta}} \lambda_{i,j}(r | \boldsymbol{\theta}^*)}{\lambda_{i,j}(r | \boldsymbol{\theta}^*)} dM_{i,j}(r) + \int_{\mathcal{T}_T} \frac{\partial_{\boldsymbol{\theta}} \lambda_{i,j}(r | \boldsymbol{\theta}^*)}{\lambda_{i,j}(r | \boldsymbol{\theta}^*)} \lambda_{i,j}(r | \boldsymbol{\theta}^*) dr \right. \\ &\quad \left. - \int_{\mathcal{T}_T} \partial_{\boldsymbol{\theta}} \lambda_{i,j}(r | \boldsymbol{\theta}^*) dr \right) = \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} \frac{\partial_{\boldsymbol{\theta}} \lambda_{i,j}(r | \boldsymbol{\theta}^*)}{\lambda_{i,j}(r | \boldsymbol{\theta}^*)} dM_{i,j}(r). \end{aligned}$$

The last stochastic integral is a martingale because its integrator is a martingale itself. Therefore, Rebolledo's martingale central limit theorem (e.g., Theorem II.5.1 in [20]) implies under Assumptions 1 and 2 that $a_{\mathcal{S}}^{-1} U_{\mathcal{S}}(\boldsymbol{\theta}^*)$ converges to a p -dimensional normal distribution $\mathcal{N}(0, \Sigma(\boldsymbol{\theta}^*))$. To transfer this asymptotic normality to the estimator $\hat{\boldsymbol{\theta}}_{\mathcal{S}}$, one essentially applies a Taylor expansion and proves that the remainder term converges to zero as $\mathcal{S} \rightarrow \infty$. More details on this topic are provided in Appendix A.1.

The statement of Theorem 3.1 holds for all asymptotic regimes; however, the growth rate of $a_{\mathcal{S}}$ and the validity of the assumptions depend on it. The main question is if Assumption 2 holds. In principle, it can hold either if $n \rightarrow \infty$, $T \rightarrow \infty$, or both. Thus, estimation of the parameter $\boldsymbol{\theta}^*$ is possible in large networks that are observed over a short time period, or in small networks that are observed for a long time. From a substantive perspective, which asymptotic framework is more reasonable depends on the data at hand and what type of inferential conclusions should be drawn from it, see McCullagh [32] for a wider discussion. From a mathematical perspective, Assumption 2 can hold for $T \rightarrow \infty$ only if the averages over time of the intensities stabilize. If one wants to employ an $T \rightarrow \infty$ assumption, one has to believe that the behavior in different (far apart) time intervals is only weakly correlated. Conversely, in the case $n \rightarrow \infty$, spatial averages are required to stabilize, i.e., one has to be comfortable assuming that different (far apart) sub-networks are only weakly correlated.

Remark 3.2 (Feasible Set of Interactions). We assumed here that interactions are possible between all pairs $(i, j) \in \mathcal{D}_n$. One may exclude certain pairs from the interactions by studying intensities of the form $C_{i,j}(t) g(\boldsymbol{\theta}^\top \mathbf{s}_{i,j}(t))$, where $C_{i,j}(t) \in \{0, 1\}$ indicates if the pair (i, j) is at risk of an event at time t . The discussion in this section, can be extended to this case by choosing a different sequence $a_{\mathcal{S}}$. For details we refer to the discussion in Kreiß et al. [29].

Remark 3.3 (Scaling of Sufficient Statistics). When dealing with sufficient statistics that are not uniformly bounded (e.g., those growing with $\log n$ or even n), one may standardize the covariates prior to estimation to satisfy Assumption 1. However, scaling the sufficient statistics by a sequence $d_{\mathcal{S}} \rightarrow \infty$ requires the true parameter $\boldsymbol{\theta}^*$ to change with \mathcal{S} to absorb this scaling. Therefore, the unscaled coefficient must shrink to zero at the rate $\mathcal{O}(d_{\mathcal{S}}^{-1})$ as the network grows and $d_{\mathcal{S}} \times a_{\mathcal{S}}$ determines the convergence rate of $\boldsymbol{\theta}^*$. As an alternative, one may assume the true parameter $\boldsymbol{\theta}^*$ to remain fixed. In that case, the change in scale in $\mathbf{s}_{i,j}(t)$ can instead be controlled by standardizing the score and the observed information. We provide more details how this works in Appendix A.1.1 and give only a brief discussion here. Assume for simplicity that

$$v_{\mathcal{S},k} := \mathbb{E} \left((\partial_{\boldsymbol{\theta}_k} \lambda_{i,j}(t | \boldsymbol{\theta}^*))^2 \lambda_{i,j}(t | \boldsymbol{\theta}^*) \right)$$

is independent of t for $k = 1, \dots, p$. Let $\mathbf{D}_{\mathcal{S}}$ be a diagonal matrix with diagonal entries $(\sqrt{v_{\mathcal{S},k}})_{k=1}^p$. One can use this matrix to scale the score vector and the observed information (e.g., $\mathbf{D}_{\mathcal{S}}^{-1} H_{\mathcal{S}}(\boldsymbol{\theta}^*) \mathbf{D}_{\mathcal{S}}^{-1} \xrightarrow{\mathbb{P}} \Sigma(\boldsymbol{\theta}^*)$) after the derivatives have been computed. This preserves the Central Limit Theorem in the sense that²

$$a_{\mathcal{S}} \mathbf{D}_{\mathcal{S}} \left(\hat{\boldsymbol{\theta}}_{\mathcal{S}} - \boldsymbol{\theta}^* \right) \xrightarrow{d} \mathcal{N} \left(0, \Sigma(\boldsymbol{\theta}^*)^{-1} \right).$$

Note, however, that assuming that the unscaled coefficient stays fixed can imply explosive behavior of the counting processes. Simulating from such a model can imply so-called exploding counting processes. A non-explosive counting process must satisfy the Feller criterion for any finite sample size \mathcal{S} :

$$\sum_{k=1}^{\infty} \left(\sum_{(i,j) \in \mathcal{D}_n} g(\boldsymbol{\theta}^\top \mathbf{s}_{i,j}(t_k)) \right)^{-1} = \infty. \quad (10)$$

Intuitively, one can comprehend Equation (10) as guaranteeing that the waiting time until the ∞ th event is ∞ . For $\mathbf{s}_{i,j}(t) \in \mathbb{R}^p$, it is, in practice, easier to define an envelope $g(\boldsymbol{\theta}^\top \tilde{\mathbf{s}}(T, n)) \geq g(\boldsymbol{\theta}^\top \mathbf{s}_{i,j}(t_k))$ for all $(i, j) \in \mathcal{D}_n$ and $t \in \mathcal{T}_T$ and substitute

²We give more details about the necessary assumptions in Appendix A.1.1. But we note here that in particular Lemma A.1 is potentially no longer applicable to prove Assumptions IV and V in Appendix A and other tools have to be used instead.

Equation (10) with

$$\sum_{k=1}^{\infty} \left(\sum_{(i,j) \in \mathcal{D}_n} g(\boldsymbol{\theta}^\top \tilde{\mathbf{s}}(T, n)) \right)^{-1} = \infty,$$

which is called the Jacobsen condition. Else, the Doob-Meyer decomposition from Equation (2) and the foundation of this mathematical framework is invalid [33].

3.2 Application to Cox-type Models

In this section, we discuss the multiplicative case mentioned in Example 2.2 with $g(x) := \exp(x)$ and $\lambda_{i,j}(t | \boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^\top \mathbf{s}_{i,j}(t))$. To fulfill Assumption 1, we assume that $\boldsymbol{\Theta}$ is bounded and that $\mathbf{s}_{i,j}(t)$ is uniformly bounded. The differentiability of g is guaranteed by the choice $g(x) = \exp(x)$. For the first part of Assumption 2, note that $\log \lambda_{i,j}(t | \boldsymbol{\theta}) = \mathbf{s}_{i,j}(t)^\top \boldsymbol{\theta}$ with

$$\partial_{\boldsymbol{\theta}_k} \log \lambda_{i,j}(r | \boldsymbol{\theta}^*) = \mathbf{s}_{i,j}^{(k)}(t).$$

This implies that $\partial_{\boldsymbol{\theta}_k} \partial_{\boldsymbol{\theta}_l} \log \lambda_{i,j}(t | \boldsymbol{\theta}) = 0$ and that the first part of Assumption 2 holds. The convergence in Assumption 3 is actually not required in this case: Upon closer inspection of Appendix A one can observe that Assumption 3 is only required for proving that the condition on $H_{i,j}$ ³ from Lemma A.1 holds. However, in the Cox case, $H_{i,j} \equiv 0$, that is, we can directly check the condition on $H_{i,j}$ from Lemma A.1 without further assumptions. Note that the above discussion is valid for all asymptotic regimes.

We are left to discuss the second part of Assumption 2, that is, the convergence of $a_S^{-2} F_S(\boldsymbol{\theta}^*) \rightarrow \Sigma(\boldsymbol{\theta}^*)$ for a positive definite matrix $\Sigma(\boldsymbol{\theta}^*)$. We rewrite

$$F_S(\boldsymbol{\theta}^*) = \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} \mathbf{s}_{i,j}(t)^\top \mathbf{s}_{i,j}(t) \exp(\mathbf{s}_{i,j}(t)^\top \boldsymbol{\theta}^*) dt = \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} F_{i,j}(t) dt,$$

where $F_{i,j}(t) := \mathbf{s}_{i,j}(t)^\top \mathbf{s}_{i,j}(t) \exp(\mathbf{s}_{i,j}(t)^\top \boldsymbol{\theta}^*)$. Consider first the case where n is fixed and $T \rightarrow \infty$. In this case, $\bar{F}_n(t) := 2/(n(n-1)) \sum_{(i,j) \in \mathcal{D}_n} F_{i,j}(t)$ can be understood as a single stochastic process evolving over $t \in \mathcal{T}_T$. We require convergence of $1/T \int_{\mathcal{T}_T} \bar{F}_n(t) dt$, which follows if the process \bar{F}_n has certain ergodic properties⁴. Such a property is, for instance, implied by a fading memory in the sense that characteristics of \bar{F}_n on a certain time interval $I_1 \subset \mathcal{T}_T$ behave almost independently of characteristics of \bar{F}_n on another time interval $I_2 \subset \mathcal{T}_T$ when I_1 and I_2 drift further apart. In applied research, two ways are common implemented to guarantee such a fading memory: windowed statistics are employed in Stadtfeld and Block [35] where events can only affect the intensity for a specific amount of time, whereas Brandes

³ $H_{i,j}$ bounds $\partial_{\boldsymbol{\theta}_k} \partial_{\boldsymbol{\theta}_l} \partial_{\boldsymbol{\theta}_m} \log \lambda_{i,j}(t | \boldsymbol{\theta}) = 0$.

⁴See Chapter 25 of [34] for an in-depth treatment.

et al. [36] assumed that the effect of interactions at time s for the intensity at time $t > s$ is given by the exponential function with a particular half time.

For fixed n , the dependence structure of $(F_{i,j})_{(i,j) \in \mathcal{D}_n}$ is not irrelevant but plays a less prominent role. On the other hand, if $n \rightarrow \infty$, justifying ergodicity purely from the temporal behavior becomes more difficult as the summation over $(i,j) \in \mathcal{D}_n$ in Equation (8) can introduce an additional drift. In fact, if T is finite, such an argument becomes even impossible, and one has to rely on a certain ergodicity property in location. Writing

$$a_S^{-2} F_S(\boldsymbol{\theta}^*) = \frac{2}{n(n-1)} \sum_{(i,j) \in \mathcal{D}_n} \tilde{F}_{i,j}$$

with $\tilde{F}_{i,j} := \frac{1}{T} \int_0^T F_{i,j}(t) dt$, we now require the average over dyads, $\sum_{(i,j) \in \mathcal{D}_n}$, instead of the integral, $\int_{\mathcal{J}_T}$, to stabilize. We illustrate how such a property can be derived from a weak correlation property for disjoint dyads in the case $n \rightarrow \infty$ and T fixed, using ideas from [29], and then motivate weak correlation by introducing a fading memory in location.

Suppose that the whole process is vertex exchangeable, that is, the joint distributions of $(N_{i,j}, \mathbf{s}_{i,j})_{(i,j) \in \mathcal{D}_n}$ and $(N_{\pi(i), \pi(j)}, \mathbf{s}_{\pi(i)\pi(j)})_{(i,j) \in \mathcal{D}_n}$ are the same for any permutation π of $\{1, \dots, n\}$. In particular, this assumption implies that $(N_{i,j}, \mathbf{s}_{i,j})_{(i,j) \in \mathcal{D}_n}$ is a collection of identically distributed random variables. A simple, and in this case sufficient, strategy to prove convergence of $a_S^{-2} F_S(\boldsymbol{\theta}^*)$ relies on the Markov inequality to check the definition of convergence in probability. For illustration, we assume that $\mathbf{s}_{i,j}(t) \in \mathbb{R}$, but, on the expense of more involved notation, the argument transfers to the multivariate case. For arbitrary $\epsilon > 0$, it holds that

$$\begin{aligned} & \mathbb{P} \left(\left| a_S^{-2} F_S(\boldsymbol{\theta}^*) - \frac{1}{T} \int_{\mathcal{J}_T} \mathbb{E}(F_{1,2}(t)) dt \right| > \epsilon \right) \\ & \leq \frac{1}{\epsilon^2} \mathbb{E} \left(\left| a_S^{-2} F_S(\boldsymbol{\theta}^*) - \frac{1}{T} \int_{\mathcal{J}_T} \mathbb{E}(F_{1,2}(t)) dt \right|^2 \right) \\ & = \frac{1}{\epsilon^2} \mathbb{E} \left(\left| a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{J}_T} (F_{i,j}(t) - \mathbb{E}(F_{i,j}(t))) dt \right|^2 \right) \\ & = \frac{1}{\epsilon^2 a_S^4} \sum_{(i,j) \in \mathcal{D}_n} \text{Var} \left(\int_{\mathcal{J}_T} F_{i,j}(t) dt \right) \\ & \quad + \frac{1}{\epsilon^2 a_S^4} \sum_{(i,j) \in \mathcal{D}_n} \sum_{(k,l) \in \mathcal{D}_n \setminus \{(i,j)\}} \text{Cov} \left(\int_{\mathcal{J}_T} F_{i,j}(t) dt, \int_{\mathcal{J}_T} F_{k,l}(t) dt \right) \\ & = \frac{2}{\epsilon^2 \cdot n(n-1)} \text{Var} \left(\frac{1}{T} \int_{\mathcal{J}_T} F_{1,2}(t) dt \right) \end{aligned} \tag{11}$$

$$+ \frac{2(n-2)}{\epsilon^2 \cdot n(n-1)} \text{Cov} \left(\frac{1}{T} \int_{\mathcal{J}_T} F_{1,2}(t) dt, \frac{1}{T} \int_{\mathcal{J}_T} F_{1,3}(t) dt \right) \tag{12}$$

$$+ \frac{(n-2)(n-3)}{\epsilon^2 \cdot n(n-1)} \text{Cov} \left(\frac{1}{T} \int_{\mathcal{T}_T} F_{1,2}(t) dt, \frac{1}{T} \int_{\mathcal{T}_T} F_{3,4}(t) dt \right), \quad (13)$$

where the last equality follows because, by vertex exchangeability, the covariance terms depend only on the amount of overlap between the pairs (i, j) and (k, l) ⁵.

If we can show that Equations (11)-(13) converge to 0 for all $\epsilon > 0$ when $n \rightarrow \infty$, and that $\Sigma(\boldsymbol{\theta}^*) := 1/T \int_{\mathcal{T}_T} \mathbb{E}(F_{1,2}(t)) dt$ is invertible⁶, we obtain Assumption 2. Recall that $F_{i,j}(t)$ is a function of $\mathbf{s}_{i,j}(t)$, thus, such requirements are effectively concerning the statistics $(\mathbf{s}_{i,j})_{(i,j) \in \mathcal{D}_n}$. We see that Equations (11) and (12) converge to zero under mild assumptions because the factor in front of the variance (covariance) converges to zero. We provide a detailed discussion of these assumptions in Appendix B.

In Equation (13), the factor in front of the covariance converges to a non-zero constant, i.e., we require that the covariance over disjoint dyads converges to zero. To give a heuristic motivation for such an assumption, suppose that there is an underlying, potentially unknown, sparse network G that contains all dyads $(i, j) \in \mathcal{D}_n$ as vertices. We assume that the amount of dependence between $\mathbf{s}_{i,j}$ and $\mathbf{s}_{k,l}$, and hence $F_{i,j}$ and $F_{k,l}$, decays as the shortest path length distance in G , denoted by $d((i, j), (k, l))$, increases. The amount of dependence could, e.g., be measured in terms of correlation of $F_{i,j}(t)$ and $F_{k,l}(t)$. In such scenarios, it is plausible that

$$\mathbb{E}(F_{1,2}(t)F_{3,4}(s) \mid d((1, 2), (3, 4)) = k) \approx \mathbb{E}(F_{1,2}(t))\mathbb{E}(F_{3,4}(s)),$$

when k is large. Since, in sparse networks, it is much more likely that $d((1, 2), (3, 4))$ is large, these cases dominate, and we may expect the covariance in Equation (13) to converge to zero. We give more details in Appendix B. Thus, we may expect that $a_S F_S(\boldsymbol{\theta}^*)$ converges and Assumption 2 can be expected to hold.

In summary, we have shown that Assumptions 1-3 are satisfied as long as the sufficient statistics are defined with caution: For $T \rightarrow \infty$, they should have a fading memory with passing time, for $n \rightarrow \infty$, they should not induce global dependencies.

4 Simulation Study

We assess the finite-sample validity of the theoretical results from Section 3 through Monte Carlo simulations. The R package `redeem` facilitates both data generation and model estimation [37]. We adopt the multiplicative Cox-type specification detailed in Section 3.2, where the vector of sufficient statistics comprises an intercept alongside selected endogenous sufficient statistics. To test the boundaries of our asymptotic framework, the study includes model specifications that intentionally violate the stability conditions derived in Section 3.2. To model varying degrees of temporal dependence, we construct both windowed and cumulative formulations of these statistics. We apply identity and $\log(\cdot + 1)$ transformations to these statistics to evaluate

⁵There are $n(n-1)/2$ possibilities to chose (i, j) , $n(n-1)(n-2)/2$ possibilities to choose (i, j) and (k, l) with one overlapping vertex, and $n(n-1)(n-2)(n-3)/4$ possibilities to choose (i, j) and (k, l) without overlap.

⁶Recall that T is fixed. However, if $T \rightarrow \infty$, assuming the convergence of this expression is the only extra assumption. All of the following arguments will remain true. Note furthermore that $\Sigma(\boldsymbol{\theta}^*)$ is always positive semi-definite.

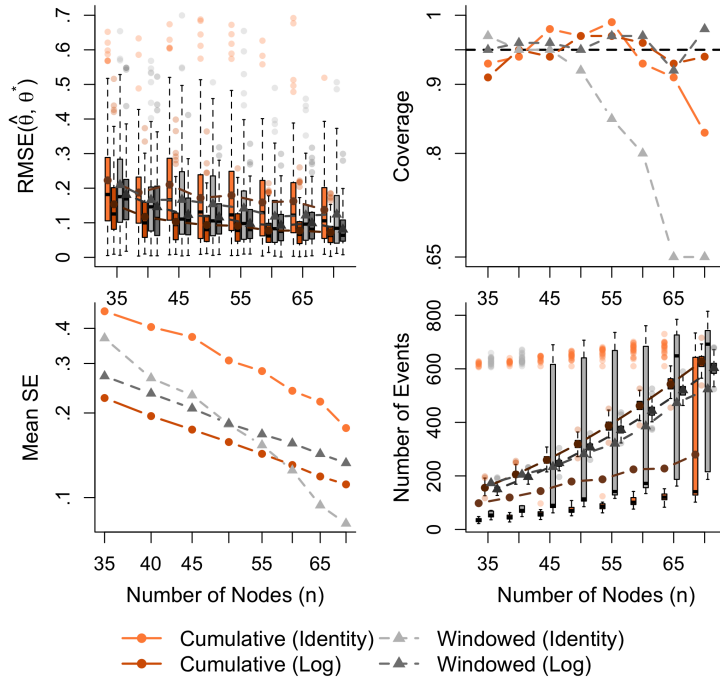


Fig. 1: Simulation Study with $n \rightarrow \infty$: Left panels display the estimator RMSE and Mean Standard Error; right panels display empirical coverage probabilities and event counts.

the sensitivity of estimator stability. This yields four distinct model specifications per simulation setting: cumulative versus windowed statistics (orange and grey in subsequent visualizations) with identity versus log transformations (bright and dark). Because the scales across those statistics are not directly comparable, we restrict the true parameter values to the positive real line, \mathbb{R}_+ , and provide their exact numerical specifications in Appendix C.

The Root Mean Square Error (RMSE) measures consistency of the point estimates. The mean standard error serves as a proxy for how fast the model accumulates information, while empirical coverage probabilities validate the variance estimator and confirm the adequacy of the Gaussian approximation in finite samples. Finally, we monitor the total number of events to detect explosive behavior in the counting processes.

Scenario: $n \rightarrow \infty$: We employ the inertia statistic as the endogenous statistic, counting the number of prior interactions for each pair. In Figure 1, we evaluate the asymptotic behavior of the estimator as the cardinality of \mathcal{P}_n expands from 35 to 70 entities for a fixed T . As $|\mathcal{P}_n|$ increases, both the Root Mean Square Error (RMSE) and the theoretical standard errors decrease for all specifications. In line with Remark 3.3, the model involving the raw count of prior interactions accumulates information

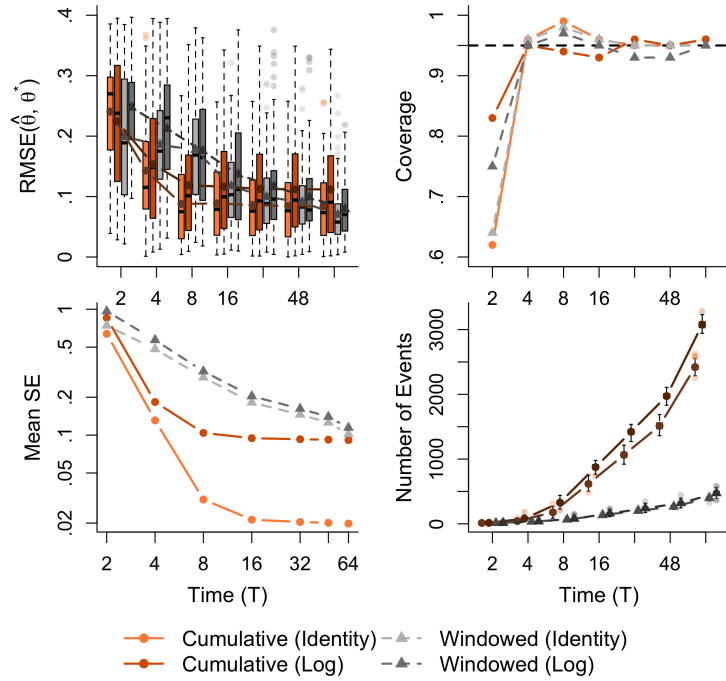


Fig. 2: Simulation Study with $T \rightarrow \infty$: Left panels display the estimator RMSE and Mean Standard Error; right panels display empirical coverage probabilities and event counts.

faster than the log-transformed specification. Approximately 30% of these simulations display explosive counting processes⁷, leading to improper empirical coverage and extreme variance in the total event count across simulation runs. The $\log(\cdot + 1)$ transformation yields well-behaved, stable stochastic processes, highlighting the need for transforming raw counts.

Scenario: $T \rightarrow \infty$: For the setting with $T \rightarrow \infty$, we utilize the common partner statistic. This statistic evaluates the number of common partners for a given dyad $(i, j) \in \mathcal{D}_n$, being the number of entities $h \in \mathcal{P}_n$ satisfying both $N_{i,h} > 0$ and $N_{h,j} > 0$. Unlike its windowed counterpart, the cumulative specification of this effect lacks temporal ergodicity and ceases to learn anything new at some point in time. This phenomenon is illustrated in the lower-left plane of Figure 2 by the flat trajectory of the Mean SE of the cumulative specifications (orange lines), rendering the estimators inconsistent. Windowed specifications, on the other hand, learn continuously and stabilize the counting processes, as shown by the steadily decreasing mean standard error and the linearly increasing number of events.

⁷It is straightforward to see that here, Equation (10) is not satisfied. In our simulations, we define a counting process to be exploding whenever any dyadic intensity exceeds 1.79769×10^{308} , indicating overflow.

5 Discussion

The theory discussed in Section 3.1 covers the case of a finite-dimensional parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ with p fixed. Worthwhile generalizations of this framework would be to high-dimensional parametric, semiparametric, and nonparametric settings.

For high-dimensional models with $p \rightarrow \infty$ or even $p > n$, penalized estimation is a classical approach, and inference methods have been discussed in the context of classical counting-process. Tibshirani [38] provides a general discussion of penalized estimation in Cox models, while Yu et al. [39] provide an inferential framework.

If one is primarily interested in a low-dimensional subset of coefficients, another possibility is to consider a semiparametric model incorporating a potentially infinite-dimensional nuisance parameter $\boldsymbol{\eta}$ substituting Equation (5) with:

$$\lambda_{i,j}(t \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = g(\boldsymbol{\theta}^\top \mathbf{s}_{i,j}(t) + m(\boldsymbol{\eta}, \mathbf{r}_{i,j}(t))),$$

where $\mathbf{r}_{i,j}$ describe a new set of low-dimensional statistics, and m is a known link function. In such a model, the focus remains on estimation of $\boldsymbol{\theta}$. In such cases, a partial likelihood approach as described in [40] is a promising starting point for inference.

Relaxing the assumption of linear covariate effects introduces nonparametric specifications, such as $\lambda_{i,j}(t \mid \varphi) = \varphi(t, \mathbf{s}_{i,j}(t))$. To bypass the curse of dimensionality plaguing such models, additive or multiplicative intensities could be specified through

$$\lambda_{i,j}(t \mid \varphi) = \varphi_0(t) \prod_{q=1}^p \varphi_q(\mathbf{s}_{i,j,q}(t)) \text{ and } \lambda_{i,j}(t \mid \varphi) = \varphi_0(t) + \sum_{q=1}^p \varphi_q(\mathbf{s}_{i,j,1}(t)),$$

respectively. Robust estimation of the functions $\varphi_0, \dots, \varphi_p$ could be potentially achieved by adapting smooth backfitting algorithms [41, 42] to the REM framework.

Beyond this, the mathematical framework and empirical simulations presented here offer practical guidance for researchers applying REMs to real-world data. The lessons learned from this asymptotic theory for the applied model are:

1. Relying on raw counts for endogenous statistics, such as the number of prior events of some pair of entities, can lead to explosive counting processes and improper empirical coverage. Applying transformations, such as $\log(\cdot + 1)$, is then critical to stabilize the stochastic processes and ensure reliable inference. Alternative transformations are bounded functions such as $1/(\cdot + 1)$, $\cdot/(\cdot + K)$, or the indicator $\mathbb{I}(\cdot > 0)$ with some $K \in \mathbb{N}_0$, all available in the R package `redeem` [37].
2. In asymptotic regimes where $T \rightarrow \infty$, cumulative statistics lack temporal ergodicity and eventually cease to learn. Employing some form of windowed statistics ensures continuous learning and prevents estimator inconsistency over long time horizons.
3. When analyzing growing networks where the number of entities approaches infinity, sufficient statistics must be specified to avoid inducing global dependencies. Ensuring that the behavior of disjoint dyads remains asymptotically uncorrelated is necessary for the convergence of the observed information.

References

- [1] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational social science. *Science* **323**, 721–723 (2009) <https://doi.org/10.1126/science.1167742>
- [2] Wagner, C., Strohmaier, M., Olteanu, A., Kicman, E., Contractor, N., Eliassi-Rad, T.: Measuring algorithmically infused societies. *Nature* **595**(7866), 197–204 (2021) <https://doi.org/10.1038/s41586-021-03666-1>
- [3] Butts, C.T.: Social network analysis: A methodological introduction. *Asian Journal of Social Psychology* **11**, 13–41 (2008) <https://doi.org/10.1111/j.1467-839X.2007.00241.x>
- [4] Kitts, J.A., Lomi, A., Mascia, D., Pallotti, F., Quintane, E.: Investigating the temporal dynamics of interorganizational exchange: Patient transfers among Italian hospitals. *American Journal of Sociology* **123**(3), 850–910 (2017) <https://doi.org/10.1086/693704>
- [5] Renshaw, S.L., Livas, S.M., Petrescu-Prahova, M.G., Butts, C.T.: Modeling complex interactions in a disrupted environment: Relational events in the WTC response. *Network Science* **11**(2), 295–323 (2023) <https://doi.org/10.1017/nws.2023.4>
- [6] Fritz, C., Rastelli, R., Fop, M., Alberto, Caimo.: Scalable durational event models: Application to physical and digital interactions (2025) <https://doi.org/10.48550/arXiv.2504.00049>
- [7] Espinosa-Rada, A., Lerner, J., Fritz, C.: Socio-cognitive networks between researchers: Investigating scientific dualities with the group-oriented relational hyperevent model. *Social Networks* **83**, 1–13 (2025) <https://doi.org/10.1016/j.socnet.2025.04.005>
- [8] Fritz, C., Lebacher, M., Kauermann, G.: Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time. *Statistica Neerlandica* **74**(3), 275–299 (2020) <https://doi.org/10.1111/stan.12198>
- [9] Bianchi, F., Filippi-Mazzola, E., Lomi, A., Wit, E.C.: Relational Event Modeling. *Annual Review of Statistics and Its Application* **11**, 297–319 (2024) <https://doi.org/10.1146/annurev-statistics-040722-060248>
- [10] Uzzi, B., Lancaster, R.: Embeddedness and Price Formation in the Corporate Law Market. *American Sociological Review* **69**(3), 319–344 (2004) <https://doi.org/10.1177/000312240406900301>

- [11] Uzzi, B.: The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect. *American Sociological Review* **61**(4), 674–698 (1996) <https://doi.org/10.2307/2096399>
- [12] Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge (1994)
- [13] Holland, P.W., Leinhardt, S.: Some evidence on the transitivity of positive interpersonal sentiment. *American Journal of Sociology* **77**, 1205–1209 (1972)
- [14] Fritz, C., Mehrl, M., Thurner, P.W., Kauermann, G.: All that glitters is not gold: Relational events models with spurious events. *Network Science* **11**(Special Issue 2) (2023) <https://doi.org/10.1017/nws.2022.22>
- [15] Leifeld, P., Wong, J.S.T.: Fully bayesian estimation of temporal decay in ordinal relational event models. *Computational Statistics & Data Analysis*, 108428 (2026) <https://doi.org/10.1016/j.csda.2026.108428>
- [16] Quintane, E., Pattison, P.E., Robins, G.L., Mol, J.M.: Short- and long-term stability in organizational networks: Temporal structures of project teams. *Social Networks* **35**(4), 528–540 (2013) <https://doi.org/10.1016/j.socnet.2013.07.001>
- [17] Tranmer, M., Marcum, C.S., Morton, F.B., Croft, D.P., De Kort, S.R.: Using the relational event model (REM) to investigate the temporal dynamics of animal social networks. *Animal Behaviour* **101**, 99–105 (2015) <https://doi.org/10.1016/j.anbehav.2014.12.005>
- [18] Perry, P.O., Wolfe, P.J.: Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **75**(5), 821–849 (2013) <https://doi.org/10.1111/rssb.12013>
- [19] Andersen, P.K., Gill, R.D.: Cox’s regression model for counting processes: A large sample study. *The Annals of Statistics* **10**(4), 1100–1120 (1982) <https://doi.org/10.1214/aos/1176345976>
- [20] Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N.: *Statistical Models Based on Counting Processes*. Springer, New York (1993)
- [21] Aalen, O.O., Borgan, Ø., Gjessing, H.K.: *Survival and Event History Analysis*. Springer, New York, NY (2008). <https://doi.org/10.1007/978-0-387-68560-1>
- [22] Gibson, D.R.: Taking Turns and Talking Ties: Networks and Conversational Interaction1. *American Journal of Sociology* **110**(6), 1561–1597 (2005) <https://doi.org/10.1086/428689>
- [23] Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**(1), 83–90 (1971) <https://doi.org/10.2307/2334319>

- [24] Passino, F.S., Heard, N.A.: Mutually exciting point process graphs for modeling dynamic networks. *Journal of Computational and Graphical Statistics* **32**(1), 116–130 (2023) <https://doi.org/10.1080/10618600.2022.2096048>
- [25] Vu, D.Q., Asuncion, A., Hunter, D.R., Smyth, P.: Continuous-Time Regression Models for Longitudinal Networks. In: *Neural Information Processing Systems* (2011)
- [26] Soliman, H., Zhao, L., Huang, Z., Paul, S., Xu, K.S.: The Multivariate Community Hawkes Model for Dependent Relational Events in Continuous-time Networks. In: *International Conference on Machine Learning*, pp. 20329–20346 (2022)
- [27] Fang, G., Xu, G., Xu, H., Zhu, X., Guan, Y.: Group Network Hawkes Process. *Journal of the American Statistical Association* **119**(547), 2328–2344 (2024) <https://doi.org/10.1080/01621459.2023.2257889>
- [28] Butts, C.T.: A Relational Event Framework for Social Action. *Sociological Methodology* **3**, 155–200 (2008) <https://doi.org/10.1111/j.1467-9531.2008.00203.x>
- [29] Kreiß, A., Mammen, E., Polonik, W.: Nonparametric inference for continuous-time event counting and link-based dynamic network models. *Electronic Journal of Statistics* **13**(2), 2764–2829 (2019) <https://doi.org/10.1214/19-EJS1588>
- [30] Vu, D.Q., Asuncion, A., Hunter, D.R., Smyth, P.: Dynamic egocentric models for citation networks. In: *Proceedings of the 28th International Conference on Machine Learning* (2011)
- [31] Fritz, C., Thurner, P.W., Kauermann, G.: Separable and semiparametric network-based counting processes applied to the international combat aircraft trades. *Network Science* **9**(3), 291–311 (2021) <https://doi.org/10.1017/nws.2021.9>
- [32] McCullagh, P.: What is a statistical model? *The Annals of Statistics* **30**(5), 1225–1310 (2002) <https://doi.org/10.1214/aos/1035844977>
- [33] Gjessing, H.K., Røysland, K., Pena, E.A., Aalen, O.O.: Recurrent events and the exploding Cox model. *Lifetime Data Analysis* **16**(4), 525–546 (2010) <https://doi.org/10.1007/s10985-010-9180-y>
- [34] Kallenberg, O.: *Foundations of Modern Probability*, 3rd edn. *Probability Theory and Stochastic Modelling*. Springer, Switzerland (2021). <https://doi.org/10.1007/978-3-030-61871-1>
- [35] Stadtfeld, C., Block, P.: Interactions, actors, and time: Dynamic network actor models for relational events. *Sociological Science* **4**, 318–352 (2017) <https://doi.org/10.15195/v4.a14>

- [36] Brandes, U., Lerner, J., Snijders, T.A.B.: Networks evolving step by step: Statistical analysis of dyadic event data. In: *Advances in Social Network Analysis and Mining*, pp. 200–205. IEEE, ??? (2009)
- [37] Fritz, C.: Redeem: Relational Event and Durational Event Models. (2026). R package version 1.0.0. <https://corneliusfritz.github.io/redeem/>
- [38] Tibshirani, R.J.: Univariate shrinkage in the cox model for high dimensional data. *Statistical Applications in Genetics and Molecular Biology* **8**(1), 21 (2009) <https://doi.org/10.2202/1544-6115.1438>
- [39] Yu, Y., Bradic, J., Samworth, R.J.: Confidence intervals for high-dimensional cox models. *Statistica Sinica* **31**(1), 243–267 (2021) <https://doi.org/10.5705/ss.202018.0247>
- [40] Murphy, S.A., Vaart, A.W.: On profile likelihood. *Journal of the American Statistical Association* **95**(450), 449–465 (2000). Accessed 2026-05-28
- [41] Hiabu, M., Mammen, E., Martínez-Miranda, M.D., Nielsen, J.P.: Smooth back-fitting of proportional hazards with multiplicative components. *Journal of the American Statistical Association* **116**(536), 1983–1993 (2021) <https://doi.org/10.1080/01621459.2020.1753520> <https://doi.org/10.1080/01621459.2020.1753520>
- [42] Bischofberger, S.M., Hiabu, M., Mammen, E., Nielsen, J.P.: Smooth backfitting for additive hazard rates. *Scandinavian Journal of Statistics* **52**(4), 1625–1669 (2025) <https://doi.org/10.1111/sjos.70004> <https://onlinelibrary.wiley.com/doi/pdf/10.1111/sjos.70004>

Acknowledgements. We thank the other organizers and attendees of the workshop New Trends in Statistical Network Analysis, Carsten Jentsch and Göran Kauermann, where this work originated for fostering a stimulating and productive scientific exchange.

Appendix A Justification for Theorem 3.1

Theorem 3.1 is a direct consequence of Theorems VI.1.1 and VI.1.2 from [20] if the following general Assumptions I-V hold true. While these assumptions are general, their exact meaning can be different for different models. We will argue afterwards why our Assumptions 1-3 imply them. Note that Theorems VI.1.1 and VI.1.2 from [20] are formulated for general specifications of $\lambda_{i,j}(t \mid \boldsymbol{\theta})$ while we focus here on $\lambda_{i,j}(t \mid \boldsymbol{\theta}) = g(\mathbf{s}_{i,j}(t)^\top \boldsymbol{\theta})$. We therefore reformulate the assumptions in [20] for this special case when it shortens the formulas.

Assumption I (Smoothness of the model). The function g is continuously differentiable up to third order. Moreover, $\ell_{n,T}(\boldsymbol{\theta})$ may be differentiated with respect to $\boldsymbol{\theta}$ up to third order by interchanging differential and integral.

Assumption II (Convergence of observed information). There is an increasing sequence $(a_S)_{S \in \mathbb{N}} \subseteq [0, \infty)$ with $a_S \rightarrow \infty$ such that

$$a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} (\partial_{\boldsymbol{\theta}_k} \partial_{\boldsymbol{\theta}_l} \log \lambda_{i,j}(t; \boldsymbol{\theta}^*))^2 \lambda_{i,j}(t; \boldsymbol{\theta}^*) dt.$$

converges in probability to a finite constant for all $k, l \in \{1, \dots, p\}$. There is a positive definite matrix $\Sigma(\boldsymbol{\theta}^*) \in \mathbb{R}^{p \times p}$ such that $a_S^{-2} F_S(\boldsymbol{\theta}^*) \xrightarrow{\mathbb{P}} \Sigma(\boldsymbol{\theta}^*)$.

Assumption III (Lindeberg condition). For all $k \in \{1, \dots, p\}$, and $\epsilon > 0$

$$a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} \left(\frac{\partial_{\boldsymbol{\theta}_k} \lambda_{i,j}(r \mid \boldsymbol{\theta}^*)}{\lambda_{i,j}(r \mid \boldsymbol{\theta}^*)} \right)^2 \mathbb{1} \left(\left| a_S^{-1} \frac{\partial_{\boldsymbol{\theta}_k} \lambda_{i,j}(r \mid \boldsymbol{\theta}^*)}{\lambda_{i,j}(r \mid \boldsymbol{\theta}^*)} \right| > \epsilon \right) \lambda_{i,j}(r \mid \boldsymbol{\theta}^*) dr \xrightarrow{\mathbb{P}} 0.$$

Assumption IV (Uniform convergence of observed information). For any random sequence $\boldsymbol{\theta}_S \xrightarrow{\mathbb{P}} \boldsymbol{\theta}^*$, it holds that $a_S^{-2} H_S(\boldsymbol{\theta}_S) \xrightarrow{\mathbb{P}} -\Sigma(\boldsymbol{\theta}^*)$.

Assumption V (Consistency). The estimator $\hat{\boldsymbol{\theta}}_S$ is consistent, that is, $\hat{\boldsymbol{\theta}}_S \xrightarrow{\mathbb{P}} \boldsymbol{\theta}^*$.

Assumptions IV and V are high-level and can be proven under different assumptions for different models. For our case, the following Lemma A.1, which is also a consequence of Theorems VI.1.1 and VI.1.2 in [20], gives more accessible conditions under which Assumptions IV and V hold; but there are other ways of establishing them. They are essentially required to prove that the remainder term in a Taylor expansion converges to zero.

Lemma A.1. Let Assumptions I and II hold. Suppose in addition that there are predictable processes $G_{i,j} : \mathcal{T}_T \rightarrow [0, \infty)$ and $H_{i,j} : \mathcal{T}_T \rightarrow [0, \infty)$ for all $i, j \in \{1, \dots, n\}$

and all $n \in \mathbb{N}$ such that

$$\begin{aligned} \max_{k,l,m \in \{1, \dots, p\}} \sup_{\boldsymbol{\theta} \in \bar{\Theta}^*} |\partial_{\boldsymbol{\theta}_k} \partial_{\boldsymbol{\theta}_l} \partial_{\boldsymbol{\theta}_m} \lambda_{i,j}(t | \boldsymbol{\theta})| &\leq G_{i,j}(t), \\ \max_{k,l,m \in \{1, \dots, p\}} \sup_{\boldsymbol{\theta} \in \bar{\Theta}^*} |\partial_{\boldsymbol{\theta}_k} \partial_{\boldsymbol{\theta}_l} \partial_{\boldsymbol{\theta}_m} \log \lambda_{i,j}(t | \boldsymbol{\theta})| &\leq H_{i,j}(t), \end{aligned}$$

such that

$$a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} G_{i,j}(t) dt, \quad a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} H_{i,j}(t) \lambda_{i,j}(t | \boldsymbol{\theta}^*) dt,$$

all converge in probability to finite constants. Finally, suppose that, for any $\epsilon > 0$,

$$a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{T}_T} H_{i,j}(t) \mathbb{1} \left(a_S^{-1} \sqrt{H_{i,j}(t)} > \epsilon \right) \lambda_{i,j}(t | \boldsymbol{\theta}^*) dt \xrightarrow{\mathbb{P}} 0.$$

Then, Assumptions IV and V hold.

We argue next that Assumptions I-V hold under our Assumptions 1-3. Note firstly that, by Assumption 1, the parameter space $\Theta \subseteq \mathbb{R}^p$ is bounded, and, hence, it a subset of a compact set $\bar{\Theta}$.

Assumption I clearly holds as we assume differentiability of g in Assumption 1. Differentiability of $\ell_{n,T}(\boldsymbol{\theta})$ is then also a direct consequence: Note that the first integral with respect to $N_{i,j}(t)$ can be rewritten as a sum. Therefore, differentiability of it follows by the classical rules of differentiation. The second integral is a parameter integral and can be differentiated under the integral sign if all derivatives of $\lambda_{i,j}(t | \boldsymbol{\theta})$ are uniformly bounded on the compact set $\bar{\Theta}$. This is the case because $\mathbf{s}_{i,j}(t)$ are uniformly bounded by Assumption 1, and g is continuously differentiable.

Assumption II and Assumption 2 are identical.

To argue that Assumption III holds, we firstly note that $\lambda_{i,j}(t | \boldsymbol{\theta})$ is uniformly bounded away from zero on $\bar{\Theta}$. More specifically, by Assumption 1, we have that

$$\inf_{\boldsymbol{\theta} \in \bar{\Theta}} \inf_{n \in \mathbb{N}} \inf_{t \in \mathcal{T}_T} \min_{i,j=1, \dots, n} g(\boldsymbol{\theta}^\top \mathbf{s}_{i,j}(t)) \geq \epsilon_0 > 0 \text{ almost surely}$$

for a suitable $\epsilon_0 > 0$ because $\mathbf{s}_{i,j}(t)$ is uniformly bounded by Assumption 1. Therefore,

$$\inf_{n \in \mathbb{N}} \inf_{t \in \mathcal{T}_T} \inf_{\boldsymbol{\theta} \in \bar{\Theta}} \min_{i,j=1, \dots, n} \lambda_{i,j}(t | \boldsymbol{\theta}) \geq \epsilon_0 \text{ almost surely.}$$

Furthermore

$$\begin{aligned} \sup_{n \in \mathbb{N}} \sup_{t \in \mathcal{T}_T} \max_{i,j=1, \dots, n} |\lambda_{i,j}(t; \boldsymbol{\theta}^*)| &= |g(\mathbf{s}_{i,j}(t)^\top \boldsymbol{\theta}^*)| \leq M_0 \text{ almost surely, and} \\ \sup_{n \in \mathbb{N}} \sup_{t \in \mathcal{T}_T} \max_{i,j=1, \dots, n} |\partial_{\boldsymbol{\theta}_k} \lambda_{i,j}(t; \boldsymbol{\theta}^*)| &= |\mathbf{s}_{i,j,k}(t) g(\mathbf{s}_{i,j}(t)^\top \boldsymbol{\theta}^*)| \leq M_1 \text{ almost surely (A1)} \end{aligned}$$

for suitable constants $M_0, M_1 \in (0, \infty)$, by the boundedness assumption in Assumption 1. We conclude that

$$\begin{aligned} & a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{J}_T} \left(\frac{\partial_{\theta_k} \lambda_{i,j}(r; \boldsymbol{\theta}^*)}{\lambda_{i,j}(r; \boldsymbol{\theta}^*)} \right)^2 \mathbb{1} \left(\left| a_S^{-1} \frac{\partial_{\theta_k} \lambda_{i,j}(r; \boldsymbol{\theta}^*)}{\lambda_{i,j}(r; \boldsymbol{\theta}^*)} \right| > \epsilon \right) \lambda_{i,j}(r; \boldsymbol{\theta}^*) dr \\ & \leq a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{J}_T} \left(\frac{M_1}{\epsilon_0} \right)^2 \mathbb{1} \left(a_S^{-1} \frac{M_1}{\epsilon_0} > \epsilon \right) M_0 dr \\ & = a_S^{-2} \frac{n(n-1)}{2} T \frac{M_1^2 M_0}{\epsilon_0^2} \mathbb{1} \left(\frac{M_1}{\epsilon_0 \epsilon} > a_S \right). \end{aligned}$$

Since we assume $a_S \rightarrow \infty$, the indicator on the right will eventually be equal to zero, and therefore Assumption III holds true.

We argue that Assumptions IV and V hold true by showing the conditions of Lemma A.1. Note firstly that similarly to Equation (A1), we have

$$\begin{aligned} & \max_{k,l,m \in \{1, \dots, p\}} \sup_{n \in \mathbb{N}} \sup_{t \in \mathcal{J}_T} |\partial_{\theta_k} \partial_{\theta_l} \partial_{\theta_m} \lambda_{i,j}(t; \boldsymbol{\theta}^*)| \leq M_3 \text{ almost surely, and} \\ & \max_{k,l,m \in \{1, \dots, p\}} \sup_{n \in \mathbb{N}} \sup_{t \in \mathcal{J}_T} |\partial_{\theta_k} \partial_{\theta_l} \partial_{\theta_m} \log \lambda_{i,j}(t; \boldsymbol{\theta}^*)| \leq \bar{M}_3 \text{ almost surely} \end{aligned}$$

for suitable constants $M_3, \bar{M}_3 \in (0, \infty)$. Then, we may choose $G_{i,j}(t) = M_3$ and $H_{i,j}(t) = \bar{M}_3$ in Lemma A.1. We conclude that

$$a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{J}_T} G_{i,j}(t) dt = a_S^{-2} \frac{n(n-1)}{2} T M_3 = M_3$$

by our choice of a_S in Assumption 2. As a constant, the above expression converges to a finite constant. Moreover,

$$a_S^{-2} \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{J}_T} H_{i,j}(t) \lambda_{i,j}(t | \boldsymbol{\theta}^*) dt = a_S^{-2} \bar{M}_3 \sum_{(i,j) \in \mathcal{D}_n} \int_{\mathcal{J}_T} \lambda_{i,j}(t | \boldsymbol{\theta}^*) dt,$$

which converges by Assumption 3 to a finite constant. The Lindeberg type condition in Lemma A.1 can be proven by similar arguments as we used for Assumption III.

A.1 Sketch of Theorem 3.1 proof

We emphasize that Theorem 3.1 is a direct consequence of Theorems VI.1.1 and VI.1.2 of [20]. However, to explain the main ideas, we give here a very short outline of the proof. By differentiability, guaranteed by Assumption I, we have by an application of Taylor's Theorem that

$$0 = \partial \ell_S(\hat{\boldsymbol{\theta}}_S) = U_S(\boldsymbol{\theta}^*) - H_S(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}^*) - (H_S(\tilde{\boldsymbol{\theta}}_S) - H_S(\boldsymbol{\theta}^*))(\hat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}^*),$$

where we abuse notation as follows: By $H_S(\tilde{\theta}_S)$ we mean a matrix, where the i -th row equals the i -th row of $H_S(\theta^{(i)})$ with $\theta^{(i)}$ lying on the connecting line between θ^* and $\hat{\theta}_S$. In other words, each row of $H_S(\tilde{\theta}_S)$ has its own intermediate parameter. Rearranging the above equation and multiplication with a_S yields

$$a_S(\hat{\theta}_S - \theta^*) = \left(a_S^{-2} H_S(\theta^*) + a_S^{-2} (H_S(\tilde{\theta}_S) - H_S(\theta^*)) \right)^{-1} a_S^{-1} U_S(\theta^*). \quad (\text{A2})$$

Assumptions IV and V imply that

$$\left(a_S^{-2} H_S(\theta^*) + a_S^{-2} (H_S(\tilde{\theta}_S) - H_S(\theta^*)) \right)^{-1} \xrightarrow{\mathbb{P}} \Sigma(\theta^*)^{-1}.$$

We have argued below Theorem 3.1 that, by Assumptions II and III, $a_S^{-1} U_S(\theta^*)$ converges to a normal distribution $\mathcal{N}(0, \Sigma(\theta^*))$. Hence, by Equation (A2) we obtain the convergence of $a_S(\hat{\theta}_S - \theta^*)$ to $\mathcal{N}(0, \Sigma(\theta^*)^{-1})$.

A.1.1 Different scaling

Note that a_S primarily guarantees that $a_S^{-1} U_S(\theta^*)$ converges to a normal distribution and that $a_S^{-2} H_S(\theta^*) \rightarrow \Sigma(\theta^*)$. We would hence typically expect that a_S equals the square root of $\text{Var}(U_{S,k}(\theta^*))$, where $U_{S,k}$ denotes the k -th entry of the vector U_S . A problem arises if different entries of $U_S(\theta^*)$ have different variances. We suppose that $s_{i,j}$ are identically distributed but scale differently and that, e.g.,

$$\begin{aligned} \text{Var}(U_{S,k}(\theta^*)) &= \sum_{(i,j) \in \mathcal{D}_n} \text{Var} \left(\int_{\mathcal{J}_T} \partial_{\theta_k} \log \lambda_{i,j}(r \mid \theta^*) dM_{i,j}(r) \right) \\ &= \frac{n(n-1)}{2} \int_{\mathcal{J}_T} \mathbb{E} \left((\partial_{\theta_k} \log \lambda_{i,j}(r \mid \theta^*))^2 \lambda_{i,j}(r \mid \theta^*) \right) dr \\ &= \frac{Tn(n-1)}{2} v_{S,k}, \end{aligned}$$

where we have assumed that the variance is the same for all pairs (i,j) and that the integrand in the second line is bounded away from zero. The last equality serves as a definition of $v_{S,k}$. Let $(v_{S,k})_S$ denote sequences that potentially have different growth rates for different k . Let $\mathbf{D}_S \in \mathbb{R}^{p \times p}$ be a sequence of diagonal matrices with diagonal entries $(\sqrt{v_{S,k}})_{k=1}^p$. It is then plausible that for $a_S = \sqrt{Tn(n-1)/2}$, we have that $a_S^{-1} \mathbf{D}_S^{-1} U_S(\theta^*)$ converges to a normal distribution. It follows, in turn, that $a_S^{-2} \mathbf{D}_S^{-1} F_S(\theta^*) \mathbf{D}_S^{-1}$ plausibly converges to some constant matrix $\Sigma(\theta^*)$. In a Cox model, one can directly conclude that $a_S^{-2} \mathbf{D}_S^{-1} H_S(\theta^*) \mathbf{D}_S^{-1}$ converges to $\Sigma(\theta^*)$. In other models, an equivalent to the first part of Assumption II is required. Moreover, we have to prove that this convergence holds uniformly as in Assumption IV. If we can show this, we can multiply Equation (A2) from the left with \mathbf{D}_S and obtain

$$a_S \mathbf{D}_S (\hat{\theta}_S - \theta^*) = \left(a_S^{-2} \mathbf{D}_S^{-1} H_S(\theta^*) \mathbf{D}_S^{-1} + a_S^{-2} \mathbf{D}_S^{-1} (H_S(\tilde{\theta}_S) - H_S(\theta^*)) \mathbf{D}_S^{-1} \right)^{-1}$$

$$\times a_S^{-1} \mathbf{D}_S^{-1} U_S(\boldsymbol{\theta}^*).$$

If we believe the assumptions that we have just discussed, we can now conclude as in Section A.1 that

$$a_S \mathbf{D}_S \left(\hat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}^* \right) \xrightarrow{d} \mathcal{N} \left(0, \Sigma(\boldsymbol{\theta}^*)^{-1} \right).$$

Appendix B More details to Section 3.2

To give more details on the heuristic discussion in Section 3.2, note that

$$\begin{aligned} & \text{Cov} \left(\int_{\mathcal{T}_T} F_{i,j}(t) dt, \int_{\mathcal{T}_T} F_{k,l}(t) dt \right) \\ &= \int_{\mathcal{T}_T} \int_{\mathcal{T}_T} \underbrace{\mathbb{E} (F_{i,j}(t) F_{k,l}(s))}_{=: \phi_{ij,kl}(t,s)} dt ds - \left(\int_{\mathcal{T}_T} \underbrace{\mathbb{E} (F_{i,j}(t))}_{=: \psi_{ik}(t)} dt \right)^2. \end{aligned}$$

Applying the above in Equation (11) yields

$$\begin{aligned} (11) &= \frac{n(n-1)}{2\epsilon^2 a_S^4} \text{Var} \left(\int_{\mathcal{T}_T} F_{1,2}(t) dt \right) \\ &= \frac{2}{\epsilon^2 \cdot n(n-1)} \left(\frac{1}{T^2} \int_{\mathcal{T}_T} \int_{\mathcal{T}_T} \phi_{12,12}(t,s) dt ds - \left(\frac{1}{T} \int_{\mathcal{T}_T} \Psi_{12}(t) dt \right)^2 \right), \end{aligned}$$

which converges to zero, when assuming that $n \rightarrow \infty$ and

$$\frac{1}{T^2} \int_{\mathcal{T}_T} \int_{\mathcal{T}_T} \mathbb{E} (F_{1,2}(t) F_{1,2}(s)) dt ds = O(1) \text{ and } \frac{1}{T} \int_{\mathcal{T}_T} \mathbb{E} (F_{1,2}(t)) dt = O(1). \quad (\text{B3})$$

These assumptions seem plausible for $T \rightarrow \infty$ or T fixed, and many choices of sufficient statistics. For Equation (12), we obtain in a similar fashion that

$$(12) = \frac{2(n-2)}{\epsilon^2 n(n-1)} \left(\frac{1}{T^2} \int_{\mathcal{T}_T} \int_{\mathcal{T}_T} \phi_{12,13}(t,s) dt ds - \left(\frac{1}{T} \int_{\mathcal{T}_T} \psi_{12}(t) dt \right)^2 \right).$$

The above expression converges to zero when $n \rightarrow \infty$ and under a boundedness assumption on the integrals similar to Equation (B3).

Finally, Equation (13) can be rewritten by the same logic to

$$(13) = \frac{(n-2)(n-3)}{\epsilon^2 n(n-1)} \left(\frac{1}{T^2} \int_{\mathcal{T}_T} \int_{\mathcal{T}_T} \phi_{12,34}(t,s) dt ds - \left(\frac{1}{T} \int_{\mathcal{T}_T} \psi_{12}(t) dt \right)^2 \right)$$

The above converges to zero for $n \rightarrow \infty$ when one assumes that the behavior on disjoint edges becomes asymptotically uncorrelated. Specifically, we require that (recall that

$$\psi_{12}(s) = \psi_{34}(t)$$

$$\phi_{12,13}(t, s) = \mathbb{E}(F_{1,2}(t)F_{3,4}(s)) \approx \mathbb{E}(F_{1,2}(t))\mathbb{E}(F_{3,4}(s)) = \psi_{12}(t)\psi_{12}(s)$$

by concretely assuming that

$$\frac{1}{T^2} \int_{\mathcal{T}_T} \int_{\mathcal{T}_T} \phi_{12,34}(t, s) dt ds - \left(\frac{1}{T} \int_{\mathcal{T}_T} \psi_{12}(t) dt \right)^2 \rightarrow 0.$$

To give a heuristic motivation for the last assumption, consider the following set-up. Suppose that there is an underlying, potentially unknown, sparse network G that contains all dyads $(i, j) \in \mathcal{D}_n$ as vertices. Let $d((i, j), (k, l))$ denote the length of the shortest path between dyads (i, j) and (k, l) within this network. We assume that the amount of dependence between $\mathbf{s}_{i,j}$ and $\mathbf{s}_{k,l}$ decreases when $d((i, j), (k, l))$ increases. Then, the amount of dependence of $F_{i,j}$ and $F_{k,l}$ is also decaying when distance between (i, j) and (k, l) in the network G increases. Assume finally that $d((i, j), (k, l)) \geq 1$ when $\{i, j\} \cap \{k, l\} = \emptyset$. In such scenarios it is plausible that

$$\mathbb{E}(F_{1,2}(t)F_{3,4}(s) \mid d((1, 2), (3, 4)) = k) \approx \mathbb{E}(F_{1,2}(t))\mathbb{E}(F_{3,4}(s)),$$

when k is large. Moreover, in sparse networks it is much more likely to observe a large $d((1, 2), (3, 4))$. Hence, the large distance expectations play an increasingly important role in

$$\begin{aligned} & \mathbb{E}(F_{1,2}(t)F_{3,4}(s)) \\ &= \sum_{d=1}^{\infty} \mathbb{E}(F_{1,2}(t)F_{3,4}(s) \mid d((1, 2), (3, 4)) = d) \mathbb{P}(d((1, 2), (3, 4)) = d), \end{aligned}$$

and, hence, we may expect asymptotically that disjoint edges are uncorrelated.

Appendix C Simulation Study Setup

Table C1 summarizes the configuration of the simulation study designed to evaluate the finite-sample behavior of the estimator under both spatial ($n \rightarrow \infty$) and temporal ($T \rightarrow \infty$) asymptotic regimes. The table details the chosen sufficient statistics, observation grids, and true parameter vectors across different functional transformations.

Table C1: Configuration of simulation studies. The number of independent replications is set to 100 for all scenarios. For the windowed statistics, the window is fixed at 0.5 for Simulation Study 1 and at 1 for Simulation Study 2.

	Simulation 1	Simulation 2
Regime	$n \rightarrow \infty$	$T \rightarrow \infty$
Statistics	(Intercept, Inertia)	(Intercept, Common Partner)
Fixed	$T = 1$	$n = 10$
Grid	$n \in \{35, 40, \dots, 70\}$	$T \in \{2, 5, \dots, 60\}$
True Parameters		
Raw	$(-2.95, 1.2)$	$(-2.0, 0.25)$
Log	$(-1.5, 1.2)$	$(-2.0, 1.0)$
Windowed Raw	$(-2.5, 1.2)$	$(-2.0, 0.45)$
Windowed Log	$(-1.5, 1.2)$	$(-2.0, 1.2)$