Scalable Durational Event Models: Application to Physical and Digital Interactions

Cornelius Fritz[1]*, Riccardo Rastelli[2]†, Michael Fop[2]†, Alberto Caimo[2]†

[1]School of Computer Science and Statistics, Trinity College Dublin
[2]School of Mathematics and Statistics, University College Dublin

Abstract

Durable interactions are ubiquitous in social network analysis and are increasingly observed with precise time stamps. Phone and video calls, for example, are events to which a specific duration can be assigned. We refer to this type of data encoding the start and end times of interactions as "*durational event data*". Recent advances in data collection have enabled the observation of such data over extended periods of time and between large populations of actors. Building on Relational Event Models, we propose the "*Durational Event Model*" as a framework for studying durational events by separately modeling event incidence and duration. To accommodate large-scale applications, we introduce a fast, memory-efficient, and exact block-coordinate ascent algorithm. Theoretical and numerical results demonstrate several advantages of this approach over traditional Newton-Raphson-based methods. We apply the model to physical and digital interactions among college students in Copenhagen. Our empirical findings reveal that past interactions are the main drivers of physical interactions, whereas digital interactions are more strongly influenced by friendship ties and prior dyadic contact.

Keywords: Block Coordinate Algorithms, Large Event Data, Relational Event Model, MM Algorithm

1 Introduction

Driven by the rapid process of digitization, the availability of large-scale online networks is growing at a fast pace (Lazer et al., 2009). Contrasting the few hundred actors in a

---

*Corresponding author: fritzc@tcd.ie.

†Riccardo Rastelli, Michael Fop, and Alberto Caimo contributed equally to the work. Alberto Caimo conceived the original idea of the model.

network measured previously, digitally measured networks encompass millions of actors and edges and are relatively easy to collect (Wagner et al., 2021). The raw observations underlying these large networks are often automatic logs with precise timestamps of events; for instance, the time an email was sent. Complementing these digital networks, wearable sociometric badges allow continuous measurements of physical interactions (Eagle and Pentland, 2006). The observed data encode massive networks that represent relations in various contexts, such as patient transfer between hospitals (Vu et al., 2017) or instant messages among people on the Microsoft Messenger platform (Leskovec and Horvitz, 2008).

These massive networks come with unique computational and inferential challenges. In the literature, most algorithms for estimating models for network data are based on Markov chain Monte Carlo (MCMC) simulation schemes (see, e.g. Jin and Liang, 2013; Hummel et al., 2012; Everitt, 2012), which scale poorly to higher parameter dimensions and larger networks (Roberts and Rosenthal, 2001). Since a growing network intuitively necessitates an increasingly complex model, this error-prone regime is expected for large networks. One way to mitigate these issues of larger networks is to base inference on an approximation of the full likelihood (e.g. Raftery et al., 2012). In settings with latent variables, variational approximations are often employed, turning sampling from a posterior via MCMC into an optimization problem (e.g. Agarwal et al., 2025). For the corresponding optimization problem, Minorization-Maximization (MM) algorithms, introduced by Lange et al. (2000), were devised to obtain robust and scalable algorithms (Vu et al., 2013; Fritz et al., 2024).

Often, one can represent these networks as sequences of events between entities, observed in almost real-time. Relational event models (REMs, Butts, 2008) provide a framework for analyzing such events. We refer to (Fritz et al., 2020) for a recent review. Given event data over the continuous time interval $\mathcal{T} := [0, T]$ between the actors in the set

2

$\mathcal{A} := \{1, \ldots, N\}$, the REM specifies a dyadic intensity for an event between the pair of actors $(i, j) \in \mathcal{B} := \{(i, j) : \{i, j\} \subset \mathcal{A} \text{ and } i \neq j\}$ at time $t \in \mathcal{T}$:

$$\lambda_{i,j}(t \mid \mathcal{H}_t, \boldsymbol{\theta}) = \exp\left(\boldsymbol{\theta}^\top \boldsymbol{s}_{i,j}(\mathcal{H}_t)\right). \tag{1}$$

The summary statistics $\boldsymbol{s}_{i,j}(\mathcal{H}_t) \in \mathbb{R}^p$ for actors i and j are general functions of past events occurring up to but not including time t, denoted by $\mathcal{H}_t$. For instance, one entry of $\boldsymbol{s}_{i,j}(\mathcal{H}_t)$ might be the count of common partners between actors i and j until t. These statistics are weighted by the parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, estimated from the observed data. This model class was subsequently extended to account for spurious events (Fritz et al., 2023) and uncover latent communities (Matias et al., 2018). An additive model where the exponential link function is replaced by the identity function was employed in Vu et al. (2011a). This model variant encompasses Hawkes processes as a special case (Fang et al., 2024; Cai et al., 2024). To manage the potentially large size of event data, caching algorithms (Vu et al., 2011b) and sampling-based procedures (Lerner and Lomi, 2020) have been introduced.

Although time stamps are often available for durable ties, relational event modeling has so far almost exclusively focused on events without any duration. Many time-dependent interactions, such as phone or Zoom calls, naturally have a temporal duration associated with each observed event. We coin the term "*durational event*" to describe this particular type of eveny. Differentiating between duration and incidence is crucial for understanding the strength and dynamics of interactions. While duration captures interaction depth and time investment, engagement and volume are reflected in frequency of occurrences, offering complementary insights. One way to model such data is to treat the duration of an event as an attached weight or mark and use a model for weighted events (Lerner et al., 2013). This representation, on the other hand, disregards that the duration of an interaction is an endogenous process, influenced by factors that arise during the interaction itself or by

external factors, rather than being determined at the start of the interaction.

Stadtfeld et al. (2017) took first steps toward analyzing a related data structure by proposing a model for coordination ties. These ties are a particular case of durational events, where the creation of a link between two actors is a two-sided decision process, in which both actors have to pick each other out of all possible actors. Still, their application focused solely on the incidence of events, disregarding their duration. Hoffman et al. (2020) introduced a model for group-based interactions, where actors can join and leave groups. Rastelli and Fop (2020) propose a stochastic block model for durational events.

Despite these initial approaches, there is a gap in the literature on how to model general durational events beyond the discussed special cases. We address this limitation by introducing the Durational Event Model (DEM) as a general framework for analyzing durational events (Section 2). The proposed model faces the aforementioned challenges of large-scale networks, since its number of parameters grows with the number of actors and length of the observed time-frame. Thus, standard techniques, such as Fisher scoring or the Newton-Raphson method, are impractical for estimation in most applications. We develop in Section 3 a block-coordinate ascent method based on separate minorization-maximization and closed-form steps to overcome this limitation. The proposed estimation algorithm naturally extends to REMs (as shown in the application in the Supplementary Material B.4). In Section 4, we assess the performance of our algorithm in a simulation study. Next, we apply our model class to physical and digital interaction data from the Copenhagen Networks Study (Sapiezynski et al., 2019) in Section 5. Finally, we discuss possible future extensions in Section 6. We provide the `DEM` package for R (R Core Team, 2024) as an implementation of our method.

2 Durational Event Model

A durational event between the pair of actors $(i, j) \in \mathcal{B}$, beginning at time $b \in \mathcal{T}$ and ending at time $e \in \mathcal{T}$, with $b < e$, is represented by the four-dimensional tuple $d = (i, j, b, e)$. Henceforth, we focus on undirected durational events, although extending the methodology to directed events is straightforward. We define by $\mathcal{U}^{0 \to 1}(t) \subseteq \mathcal{B}$ the set of possible actor pairs that may experience the start of a durational event at $t \in \mathcal{T}$ (i.e., change from a "not interacting" status, denoted with 0, to an "interacting" status, denoted with 1). Similarly, $\mathcal{U}^{1 \to 0}(t) \subseteq \mathcal{B}$ is the set of possible actor pairs that may experience the end of a durational event at $t \in \mathcal{T}$. If durational events correspond to phone calls, actors may only engage in one interaction at a time, hence $\mathcal{U}^{0 \to 1}(t) \neq \mathcal{B}$ and $\mathcal{U}^{1 \to 0}(t) \neq \mathcal{B}$ holds. We assume that these sets are known for $t \in \mathcal{T}$.

2.1 Model Specification

To model the dynamics of durational events, we follow previous research (Perry and Wolfe, 2013; Vu et al., 2011a; Fritz et al., 2023) and specify the durational event model as a multivariate counting process. Deviating from previous approaches for REMs, we define two separate counting processes: the *formation* process, $N_{i,j}^{0 \to 1}(t)$, counting the number of times that i and j have started an interaction up to time point t; and the *dissolution* process, $N_{i,j}^{1 \to 0}(t)$, counting the number of times that the actors have stopped interacting before and up to t. Together, the two stochastic processes count the frequency with which actor pairs transition between the states "not interacting" and "interacting".

The stochastic variation of $N_{i,j}^{0 \to 1}(t)$ and $N_{i,j}^{1 \to 0}(t)$ is characterized by their respective instantaneous probabilities of a jump. We define these two intensities as the *incidence intensity*, $\lambda_{i,j}^{0 \to 1}(t \mid \mathcal{H}_t)$, and the *duration intensity*, $\lambda_{i,j}^{1 \to 0}(t \mid \mathcal{H}_t)$. In this paper, we assume

that $N_{i,j}^{0\to1}(t)$ and $N_{i,j}^{1\to0}(t)$ with $(i,j) \in \mathcal{B}$ are non-homogeneous Poisson processes.

We expect the counting processes to be dependent on one another over time. For instance, if the actor pairs (i,h) and $(j,h) \in \mathcal{B}$ have previously interacted, the intensity $\lambda_{i,j}^{0\to1}(t)$ may increase due to the shared connection with actor h. The intensity of two actors i and j in $\mathcal{A}$ to interact may thus be influenced by the past in numerous ways:

(a) their previous interactions;

(b) past interactions involving common neighbors;

(c) ongoing interactions within the broader population.

Therefore, $\lambda_{i,j}^{0\to1}(t)$ and $\lambda_{i,j}^{1\to0}(t)$ are functions of the observed history of the data, up to but not including t:

$$
\begin{aligned}
\lambda_{i,j}^{0\to1}\left(t \mid \mathcal{H}_t, \boldsymbol{\theta}^{0\to1}\right) &= \exp\left(\boldsymbol{\alpha}^{0\to1}\, \boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_t) + \beta_i^{0\to1} + \beta_j^{0\to1} + f(t, \boldsymbol{\gamma}^{0\to1})\right), \\
\lambda_{i,j}^{1\to0}\left(t \mid \mathcal{H}_t, \boldsymbol{\theta}^{1\to0}\right) &= \exp\left(\boldsymbol{\alpha}^{1\to0}\, \boldsymbol{s}_{i,j}^{1\to0}(\mathcal{H}_t) + \beta_i^{1\to0} + \beta_j^{1\to0} + f(t, \boldsymbol{\gamma}^{1\to0})\right)
\end{aligned}
\tag{2}
$$

for $(i,j) \in \mathcal{U}^{0\to1}(t)$ or $\mathcal{U}^{1\to0}(t)$, respectively, where

- $\boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_t) = (s_{i,j,1}^{0\to1}(\mathcal{H}_t), \dots, s_{i,j,P}^{0\to1}(\mathcal{H}_t))^\top \in \mathbb{R}^P$ are summary statistics for the pair of actors (i,j) that are functions of $\mathcal{H}_t$. These summary statistics capture any dependence of the intensity on both endogenous and exogenous processes derived from the past. In this context, endogenous processes originate from events being modeled, such as the number of common partners actors i and j had up to t, while exogenous processes encompass any covariate-driven processes that, for example, capture homophily based on gender. A more detailed description of these statistics is provided in Section 2.2.

- $\boldsymbol{\alpha}^{0\to1} = (\alpha_1^{0\to1}, \dots, \alpha_P^{0\to1}) \in \mathbb{R}^{1\times P}$ is the parameter vector determining the respective effects of the statistics $\boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_t)$.

6

- $\boldsymbol{\beta}^{0\to1} = (\beta_1^{0\to1}, \ldots, \beta_N^{0\to1})^\top \in \mathbb{R}^N$ is the parameter vector capturing all degree-related characteristics, analogous to degree correction or popularity terms that are used in other network models (see, e.g., Chatterjee et al., 2011). We can interpret these parameters as a sociality effect, since they determine the overall activity level of each actor.

- $f : \mathcal{T} \times \mathbb{R}^Q \mapsto \mathbb{R}$ with $f(t, \boldsymbol{\gamma}^{0\to1}) = \sum_{q=1}^{Q} \gamma_q \, \mathbb{I}(c_{q-1} \le t < c_q)$ captures temporal variations in the data. The indicator function $\mathbb{I}(c_{q-1} \le t < c_q)$ takes the value 1 if $c_{q-1} \le t < c_q$ and 0 otherwise, while the parameter vector $\boldsymbol{\gamma}^{0\to1} = (\gamma_1^{0\to1}, \ldots, \gamma_Q^{0\to1}) \in \mathbb{R}^Q$ determines the value of f within the Q segments. To ensure identifiability, we impose $\gamma_1^{0\to1} = 0$. Sections 5 and Supplementary Material B.3 provide concrete examples and sensitivity analyses for the specification of this step-function.

The quantities governing the duration intensity are defined analogously. We collect the parameters of the incidence and duration intensities in two parameter vectors: $\boldsymbol{\theta}^{0\to1} := (\boldsymbol{\alpha}^{0\to1}, \boldsymbol{\beta}^{0\to1}, \boldsymbol{\gamma}^{0\to1})$ and $\boldsymbol{\theta}^{1\to0} := (\boldsymbol{\alpha}^{1\to0}, \boldsymbol{\beta}^{1\to0}, \boldsymbol{\gamma}^{1\to0})$. Both intensities are piecewise constant functions, which can change at two types of time points: the time points where the baseline intensity changes, denoted by the set $\mathcal{D} := \{c_1, \ldots, c_Q\}$ and the time points where events occur, denoted by the set $\mathcal{C} := \{t_1, \ldots, t_M\}$. Here, Q is the number of time points where the baseline intensity changes, whereas M denotes the number of observed durational events. By setting $\lambda_{i,j}^{1\to0}(t|\mathcal{H}_t, \boldsymbol{\theta}^{1\to0}) = \infty$ for all pairs at all times, we obtain the REM of (1) as a special case of (2).

We refer to the model defined by (2) as the *Durational Event Model (DEM)*. This model is related to the Separable Temporal Exponential Random Graph Model (STERGM) introduced by Krivitsky and Handcock (2014), which similarly models tie formation and dissolution. While the STERGM relies only on discrete-time network snapshots, our frame-

work operates in a continuous-time setting, allowing for a more granular representation of relational dynamics.

2.2 Summary Statistics

The statistics $\boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_t)$ and $\boldsymbol{s}_{i,j}^{1\to0}(\mathcal{H}_t)$ characterize the influence of past interactions on the likelihood of future events. Careful consideration is necessary regarding how this influence is specified. Aalen and Gjessing (2007) note that a valid counting process must have an intensity function that remains finite at all times to prevent an explosion where the intensity diverges. Applying the Feller criterion (Aalen et al., 2008, Sect. 8.6.3), it becomes apparent that a counting process $N_{i,j}(t)$ is non-explosive for Cox-type models such as (2) if all past event counts are $\log(\,\cdot\,+1)$-transformed. Accordingly, we define all summary statistics that relate past interactions to future interactions on the $\log(\,\cdot\,+1)$ scale.

With two key exceptions, the summary statistics can then be defined similarly to their counterparts in REMs (see, Fritz et al., 2023; Butts and Marcum, 2017; Butts, 2008). First, we assume that the incidence and duration models may rely on distinct sets of statistics. Bellow, we introduce statistics which can be incorporated into either $\boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_t)$, $\boldsymbol{s}_{i,j}^{1\to0}(\mathcal{H}_t)$, or both. Second, the statistics can leverage additional information from durational events, such as which actors are currently interacting and the elapsed time since an interaction began. Let $u_{i,j}(t)$ be a binary indicator whether actors i and j have started an interaction that is still ongoing at time t, and let $v_{i,j}(t)$ indicate whether they have interacted before time t. The elapsed time since actors i and j last interacted is $\Delta_{i,j}(t) = t - t_{i,j}$ with $t_{i,j} < t$. Given these quantities, we provide a non-exhaustive list of summary statistics used in simulations and the application, accompanied by illustrations in Figure 1:

- $s_{i,j,CCP}(\mathcal{H}_t) = \log\left(\sum_{h\notin\{i,j\}} u_{i,h}(t)\,u_{h,j}(t) + 1\right)$: current common partner statistic
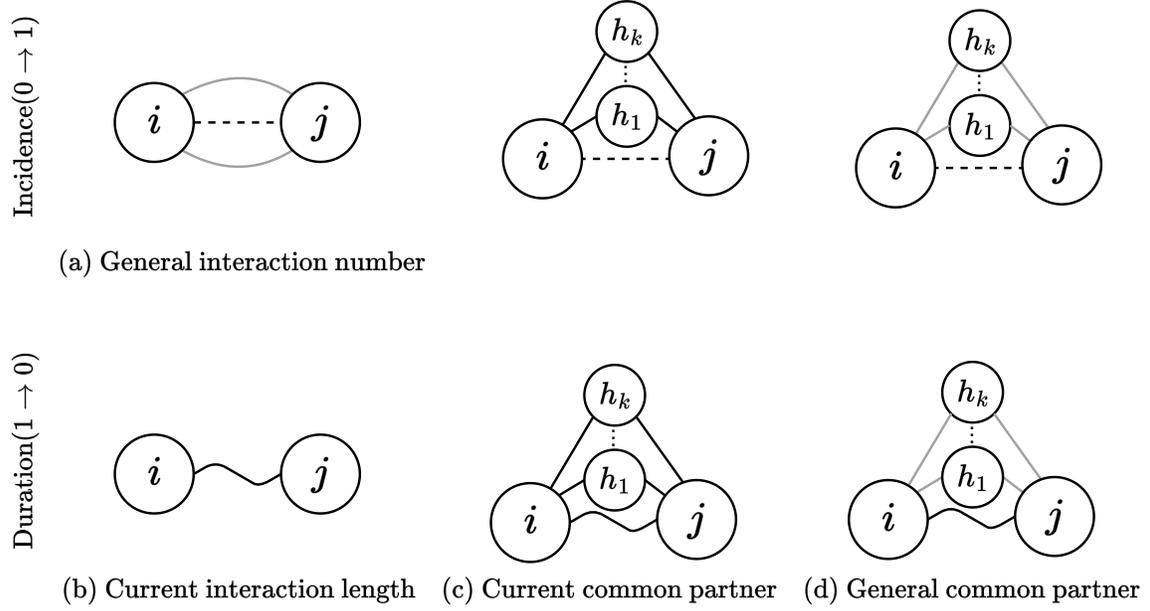
Figure 1: Graphs illustrating the proposed summary statistics. Dashed lines (- -) refer to transitions from $0 \to 1$, while the wiggly line ($\sim$) relates to $1 \to 0$. Other observed interactions are drawn as solid lines, in black if they are currently interacting actors (---) and gray if the event occurred sometime in the past (---).

representing the number of currently active partners shared by i and j at time t;

- $s_{i,j,GCP}(\mathcal{H}_t) = \log\left(\sum_{h\notin\{i,j\}} v_{i,h}(t)\,v_{h,j}(t) + 1\right)$: general common partner statistic representing the number of common partners of i and j up to time t;

- $s_{i,j,NI}(\mathcal{H}_t) = \log(N_{i,j}(t) + 1)$: number of interactions statistic representing the cumulative sum of interactions between i and j up to time t;

- $s_{i,j,NI}^{1\to0}(\mathcal{H}_t) = \log(\Delta_{i,j}(t) + 1)$: current interaction statistic representing the duration of the current interaction between actors i and j until time t;

- $s_{i,j,z}(\mathcal{H}_t) = z_{i,j}$: dyadic covariate effect, where, e.g., $z_{i,j} = |x_{i,1} - x_{j,1}|$ or $z_{i,j} =$

$\mathbb{I}(x_{i,2} = x_{j,2})$ depending on whether the effect is based on a categorical, $\boldsymbol{x}_1 = (x_{1,1}, \ldots, x_{N,1})$, or continuous, $\boldsymbol{x}_2 = (x_{1,2}, \ldots, x_{N,2})$, exogenous covariate.

Parallels between the DEM and proportional hazard model (Cox, 1972) facilitate the interpretation of the corresponding coefficients. Consider the incidence intensities of two durational events, i.e., $d_1 = (i, j, t, e)$ and $d_2 = (h, k, t, e)$. Assume that all summary statistics are identical with the exception that the pth entry of the summary statistics for d_1 is one unit higher than for d_2. In this case, the relationship between their intensities to start an interaction is given by:

$$\lambda_{i,j}^{0 \to 1}(t|\mathcal{H}_t, \boldsymbol{\theta}^{0 \to 1}) = \exp(\alpha_p^{0 \to 1})\lambda_{h,k}^{0 \to 1}(t|\mathcal{H}_t, \boldsymbol{\theta}^{0 \to 1}). \tag{3}$$

For $\alpha_p^{0 \to 1} > 0$, observing d_1 at time t is $\exp(\alpha_p^{0 \to 1})$ times more likely than d_2.

As stated in Section 2.2, statistics involving counts of past events, such as the number of current common partners between actors i and j until time t, are $\log(\cdot + 1)$-transformed. Similar to geometrically weighted statistics in the context of ERGMs (Hunter, 2007), this transformation formalizes the intuition that the initial change in a statistic has the greatest impact, while subsequent changes have diminishing returns. A change from l to $l+1$ on the original scale of a $\log(\cdot + 1)$-transformed statistic affects the intensity by the multiplicative factor $((l + 2)/(l + 1))^{\alpha_p^{0 \to 1}}$. Setting $l = 0$, the quantity $2^{\alpha_p^{0 \to 1}}$ represents the multiplicative effect of the first unit increase of a statistic on its original scale and is therefore a valuable tool to interpretat the model. For $s_{i,j,p}^{0 \to 1}(\mathcal{H}_t) = s_{i,j,GCP}^{0 \to 1}(\mathcal{H}_t)$, the first common partner has a the multiplicative effect $2^{\alpha_p^{0 \to 1}}$, while the 20th common partner has the multiplicative effect $(21/20)^{\alpha_p^{0 \to 1}} = 1.05^{\alpha_p^{0 \to 1}}$ on the incidence intensity between actors i and j.

10

3 Scalable Block-Coordinate Ascent Algorithm

Given the paths of $N_{i,j}^{0\to1}(t)$ and $N_{i,j}^{1\to0}(t)$ with $(i,j) \in \mathcal{B}$ and $t \in \mathcal{T}$, we propose a scalable method to estimate $\boldsymbol{\theta}^{0\to1}$ and $\boldsymbol{\theta}^{1\to0}$ maximizing the log-likelihood

$$\ell^\star(\boldsymbol{\theta}^{0\to1}, \boldsymbol{\theta}^{1\to0}) \quad \propto \quad \ell^{0\to1}(\boldsymbol{\theta}^{0\to1}) + \ell^{1\to0}(\boldsymbol{\theta}^{1\to0}). \tag{4}$$

The log-likelihood of each sub-model has the following form

$$\ell(\boldsymbol{\theta}) \quad = \quad \sum_{t \in \mathcal{D} \cup \mathcal{C}} \sum_{(i,j) \in \mathcal{U}(t)} y_{i,j,t} \log\left((t - t^\star)\lambda_{i,j}(t \mid \mathcal{H}_t, \boldsymbol{\theta})\right) - (t - t^\star)\lambda_{i,j}(t \mid \mathcal{H}_t, \boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{0\to1}$ or $\boldsymbol{\theta}^{1\to0}$ and $y_{i,j,t} \in \{0,1\}$ indicates whether a formation or dissolution event occurred between actors i and j in the time interval from $t^\star$ to t. Here, $t^\star$ denotes the most recent time point before t in the set $\mathcal{D} \cup \mathcal{C}$. For the chronologically first time point in this set, we set $t^\star = 0$. From (4), it follows that the log-likelihood is separable with respect to $\boldsymbol{\theta}^{0\to1}$ and $\boldsymbol{\theta}^{1\to0}$. We can thus independently estimate the incidence and duration model using the same estimation procedure. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, $\mathcal{U}(t)$, $\ell(\boldsymbol{\theta})$, and $\boldsymbol{s}_{i,j}(\mathcal{H}_t)$ correspond to either $\boldsymbol{\theta}^{0\to1}$ or $\boldsymbol{\theta}^{1\to0}$, $\ell^{0\to1}(\boldsymbol{\theta}^{0\to1})$ or $\ell^{1\to0}(\boldsymbol{\theta}^{1\to0})$, $\mathcal{U}^{0\to1}(t)$ or $\mathcal{U}^{1\to0}(t)$, and $\boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_t)$ or $\boldsymbol{s}_{i,j}^{1\to0}(\mathcal{H}_t)$, depending on estimating $\boldsymbol{\theta}^{0\to1}$ or $\boldsymbol{\theta}^{1\to0}$. The estimate of $\boldsymbol{\theta}$ in the kth iteration of $\boldsymbol{\theta}$ is denoted by $\boldsymbol{\theta}^{(k)}$.

The Newton-Raphson method is the state-of-the-art technique for estimating $\boldsymbol{\theta}$ in REMs (Butts, 2008; Stadtfeld et al., 2017). The algorithm updates $\boldsymbol{\theta}^{(k+1)}$ by the following rule:

$$\boldsymbol{\theta}^{(k+1)} \quad = \quad \boldsymbol{\theta}^{(k)} - \boldsymbol{g}(\boldsymbol{\theta}^{(k)})\boldsymbol{\Sigma}(\boldsymbol{\theta}^{(k)})^{-1}, \tag{6}$$

where $\boldsymbol{g}(\boldsymbol{\theta}^{(k+1)}) := \nabla_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k+1)}} \in \mathbb{R}^{P+N+Q}$ and

$\boldsymbol{\Sigma}(\boldsymbol{\theta}^{(k+1)}) := \nabla_{\boldsymbol{\theta}}^2\,\ell(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k+1)}} \in \mathbb{R}^{(P+N+Q)\times(P+N+Q)}$ denote the gradient and Hessian of (5)

evaluated at $\boldsymbol{\theta}^{(k+1)} \in \mathbb{R}^{P+N+Q}$, respectively. We partition $\Sigma(\boldsymbol{\theta})$ along $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\gamma}$:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} & \boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\beta}} & \boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\gamma}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{\top} & \boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\beta}} & \boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\gamma}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\gamma}}^{\top} & \boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\gamma}}^{\top} & \boldsymbol{\Sigma}_{\boldsymbol{\gamma},\boldsymbol{\gamma},} \end{pmatrix} \tag{7}$$

where, e.g., the matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \in \mathbb{R}^{|\boldsymbol{\alpha}| \times |\boldsymbol{\beta}|}$ defines the block of the Hessian matrix pertaining to the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

In most real-world settings, data is available across many actors (large N) and over an extensive time interval (large Q), making the estimation $\boldsymbol{\theta}$ via (6) impractical. Evaluating $\boldsymbol{g}(\boldsymbol{\theta}^{(k)})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ involves $O(N^2 \times Q \times M)$ summands, while inverting $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ has a computational complexity of $O((N+Q+M)^3)$, making it impractical for large-scale applications. To bypass this computational burden, we devise a block-coordinate ascent algorithm to update $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ in blocks:

Step 1: Set $\boldsymbol{\alpha}^{(k+1)}$ such that $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}) \geq \ell(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$ by a Newton-Raphson update.

Step 2: Set $\boldsymbol{\beta}^{(k+1)}$ such that $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)}) \geq \ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$ by a Minorize-Maximization update.

Step 3: Set $\boldsymbol{\gamma}^{(k+1)}$ such that $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k+1)}) \geq \ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)})$ by a closed form update.

This blockwise algorithm comes with several advantages:

1. The complexity per iteration compared to the Newton-Raphson update in (6) is reduced from $O((N+Q+M)^3)$ to $O(N^2+M \times N+M)$. This improvement is achieved by restricting matrix inversions to the first step, leveraging the low dimensionality

of $\boldsymbol{\alpha}$ to ensure scalability with respect to N and Q. In that step, we also employ coaching algorithms, akin to those introduced in Vu et al. (2011b).

2. There is no need to store large matrices at any step, reducing memory usage.

3. Any stepwise algorithm iterating according to the scheme detailed above will be exact and converge to the maximum likelihood estimator $\boldsymbol{\theta}^{\star}$, defined as the argument of the maximum of $\ell(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \mathbb{R}^{P+N+Q}$. This property holds since the likelihood function (5) can be decomposed into sums of likelihoods of Poisson-distributed random variables (Fritz et al., 2023). According to standard theory for exponential families (Barndorff-Nielsen, 1978), $\ell(\boldsymbol{\theta})$ is then a concave function. Therefore, we obtain global convergence, $\boldsymbol{\theta}^{(k)} \to \boldsymbol{\theta}^{\star}$ with $k \to \infty$ for any starting value $\boldsymbol{\theta}^{(0)}$. This result aligns with corollaries from more general results in Razaviyayn et al. (2013).

4. Contrasting the Newton-Raphson method, our algorithm enjoys the ascent property, $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k+1)}) \geq \ell(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$. This property makes the algorithm robust and reliable.

5. Our algorithm naturally extends to standard REMs.

Since the convergence of the algorithm is independent of its initialization, we seed our algorithm by setting $\boldsymbol{\theta}^{(0)} = \mathbf{0}_{P+N+Q}$ without loss of generality. The vector $\mathbf{0}_n \in \mathbb{R}^n$ with $n \in \{1, 2, \ldots\}$ is defined as a n-dimensional vector filled with zeros. We declare convergence once both $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|_2$ and $|(\ell(\boldsymbol{\theta}^{(k+1)}) - \ell(\boldsymbol{\theta}^{(k)})) / \ell(\boldsymbol{\theta}^{(k)})|$ are below 10^{-3}. We describe each step of the algorithm in the ensuing paragraphs.

Step 1: Update of $\boldsymbol{\alpha}$. To employ a Newton-Raphson update for $\boldsymbol{\alpha}$ as in (6), only slight adaptations are required. Specifically, we evaluate $\nabla_{\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})\big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(k)}} \in \mathbb{R}^P$ and

$\nabla^2_{\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})\big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(k)}} \in \mathbb{R}^{P \times P}$ given by

$$
\begin{aligned}
\nabla_{\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})\big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(k)}} &= \sum_{(i,j) \in \mathcal{B}} \sum_{t \in \mathcal{C}} \boldsymbol{s}_{i,j}(\mathcal{H}_t) \left(y_{i,j,t} - \int_{t^\star}^{t} \lambda_{i,j}(u \mid \mathcal{H}_u, \boldsymbol{\theta}^{(k)}) \, \mathrm{d}u \right) \\
\nabla^2_{\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})\big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(k)}} &= - \sum_{(i,j) \in \mathcal{B}} \sum_{t \in \mathcal{C}} \boldsymbol{s}_{i,j}(\mathcal{H}_t)^{\otimes 2} \left(\int_{t^\star}^{t} \lambda_{i,j}(u \mid \mathcal{H}_u, \boldsymbol{\theta}^{(k)}) \, \mathrm{d}u \right),
\end{aligned}
\tag{8}
$$

where $\boldsymbol{s}_{i,j}(\mathcal{H}_t)^{\otimes 2} := \boldsymbol{s}_{i,j}(\mathcal{H}_t)\, \boldsymbol{s}_{i,j}(\mathcal{H}_t)^\top$. Since the intensity function is piecewise-constant, the integral $\int_{t^\star}^{t} \lambda_{i,j}(u \mid \mathcal{H}_u, \boldsymbol{\theta}) \, \mathrm{d}u$ can be computed exactly. Both terms in (8) contain $O(N^2 \times M)$ summands, eliminating the dependence on the number of change points of the baseline intensity of (5). Similar to the coaching algorithms introduced in Vu et al. (2011b), we only update the summary statistics corresponding to pairs affected by d after observing the event d at time t. Since all summary statistics in Section 2.2 are defined locally, we only need to update a small number of pairs, typically of order $O(1)$ each time an event occurs. In most real-world applications $\max(N^2 + M, P^3) = N^2 + M$ holds, reducing the complexity of this step to $O(N^2 + M)$.

Step 2: Update of $\boldsymbol{\beta}$. A Newton-Raphson update for $\boldsymbol{\beta}$ is computationally infeasible due to the dimension of $\boldsymbol{\beta}$ being N. To avoid this bottleneck, we derive a surrogate function that is easier to optimize and whose maximizer guarantees $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)}) \geq \ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$. This strategy is in the general framework of Minorize-Maximization (MM) algorithms pushed forward in Lange et al. (2000) and Hunter and Lange (2004).

First, we restate $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}, \boldsymbol{\gamma}^{(k)})$, which is a function of $\boldsymbol{\beta}$ with fixed $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(k+1)}$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(k)}$, by

$$
\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}, \boldsymbol{\gamma}^{(k)}) \propto \sum_{(i,j) \in \mathcal{B}} (\log p_i + \log p_j) \left(\sum_{t \in \mathcal{C}} y_{i,j,t} \right) - p_i \, p_j \left(\sum_{t \in \mathcal{C}} p_{i,j,t} \right), \tag{9}
$$

where

$$
p_{i,j,t} := \left(\int_{t^\star}^{t} \exp\left(f(u, \boldsymbol{\gamma}^{(k)}) \right) \, \mathrm{d}u \right) \exp\left(\boldsymbol{\alpha}^{(k+1)} \, \boldsymbol{s}_{i,j}(\mathcal{H}_t) \right)
$$

and $p_i := \exp(\beta_i)$ for $t \in \mathcal{C}$ and $(i,j) \in \mathcal{B}$. By the inequality of arithmetic and geometric means, we get

$$p_i \, p_j \;\leq\; \frac{p_j^{(k)}}{2\,p_i^{(k)}}\,p_i^2 + \frac{p_i^{(k)}}{2\,p_j^{(k)}}\,p_j^2, \tag{10}$$

which enables us define a surrogate function $m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$:

$$\begin{aligned}
\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}, \boldsymbol{\gamma}^{(k)}) \;\geq\;& \sum_{(i,j) \in \mathcal{B}} (\log p_i + \log p_j)\left(\sum_{t \in \mathcal{C}} y_{i,j,t}\right) \\
&- \left(\frac{p_j^{(k)}}{2\,p_i^{(k)}}p_i^2 + \frac{p_i^{(k)}}{2\,p_j^{(k)}}p_j^2\right)\left(\sum_{t \in \mathcal{C}} p_{i,j,t}\right) \\
=:\;& m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}).
\end{aligned}$$

Since equality holds in (10) with $p = p^{(k)}$ and $p^{(k)} := \exp(\boldsymbol{\beta}^{(k)})$, the following two properties hold for $m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$:

$$\begin{aligned}
m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}) \;&\leq\; \ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}, \boldsymbol{\gamma}^{(k)}) \qquad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^P \\
m(\boldsymbol{\beta}^{(k)} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}) \;&=\; \ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}).
\end{aligned} \tag{11}$$

Therefore, $m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$ is a *minorizer* of $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}, \boldsymbol{\gamma}^{(k)})$ in $\boldsymbol{\beta}^{(k)}$. Applying (11) demonstrates that setting $\boldsymbol{\beta}^{(k+1)}$ to the value maximizing $m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$ with respect to $\boldsymbol{\beta}$, guarantees $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)}) \geq \ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$.

The surrogate function $m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$ is separable with respect to all entries of $\boldsymbol{\beta}$. Setting $\nabla_{\boldsymbol{\beta}}\, m(\boldsymbol{\beta} \mid \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k+1)}} = 0$ then yields for $i \in \mathcal{A}$ the following updates:

$$\beta_i^{(k+1)} \;=\; \log\left(\sqrt{p_i^{(k)}\,\frac{\displaystyle\sum_{j \neq i}\sum_{t \in \mathcal{C}} y_{i,j,t}}{\displaystyle\sum_{j \neq i}\sum_{t \in \mathcal{C}} p_{i,j,t}\,p_j^{(k+1)}}}\right). \tag{12}$$

Applying (12) for all N actors has the algorithmic complexity $O(N^2 + M \times N)$.

Step 3: Update of $\boldsymbol{\gamma}$. With $q_{i,j,t} := \int_{t^\star}^t \exp(\boldsymbol{\alpha}^{(k+1)}\, \boldsymbol{s}_{i,j}(\mathcal{H}_u) + \beta_i + \beta_j)\,\mathrm{d}u$ and $q_t := \exp(f(t, \boldsymbol{\gamma}))$ for $t \in \mathcal{D}$ and $(i,j) \in \mathcal{B}$, we can sort the summands of $\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma})$

15

according the intervals defined through the timepoints where the baseline intensity changes, $0 = c_0 < c_1 < \ldots < c_Q$:

$$\ell(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}) \;\propto\; \sum_{(i,j) \in \mathcal{B}} \sum_{q=1}^{Q} \sum_{c_{q-1} \le t < c_q} y_{i,j,t} \log q_t - q_t \, q_{i,j,t}. \tag{13}$$

This function is separable with respect to all coordinates of $\boldsymbol{\gamma}$ and the closed-form solution for its qth entry is

$$\gamma_q^{(k+1)} \;=\; \log\left(\frac{\displaystyle\sum_{(i,j) \in \mathcal{B}} \sum_{c_{q-1} \le t < c_q} y_{i,j,t}}{\displaystyle\sum_{(i,j) \in \mathcal{B}} \sum_{c_{q-1} \le t < c_q} q_{i,j,t}} \right). \tag{14}$$

Employing update (14) for $q = 1, \ldots, Q$ has the algorithmic complexity of $O(N^2 + M)$.

Uncertainty Quantification. Inference is possible in our model for fixed N and Q by standard results based on the inverse Fisher information evaluated at the converged estimate of $\boldsymbol{\theta}$. Note that the Fisher information corresponds to the Hessian given in (7) for the DEM. Under suitable regularity conditions similar results hold also in the regime that Q or N grow (see, Portnoy, 1988; He and Shao, 2000). In most application, the interest mainly lies in quantifying the uncertainty of $\boldsymbol{\alpha}$. Therefore, we consider $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to be nuisance parameters and only evaluate the top left $P \times P$ submatrix of the inverse Hessian, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, which we denote by $\boldsymbol{\Lambda}(\boldsymbol{\theta}) \in \mathbb{R}^{P \times P}$. Applying Theorem 8.5.11 in Harville (1997), $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ is given by:

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) \;=\; \left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} - \boldsymbol{X}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}} \, \boldsymbol{Y}_{\boldsymbol{\beta},\boldsymbol{\gamma}}^{-1} \, \boldsymbol{X}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}}^{\top} \right)^{-1}, \tag{15}$$

with

$$\boldsymbol{X}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}} \;=\; \left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\gamma}} \;\; \boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{\top} \right) \text{ and } \boldsymbol{Y}_{\boldsymbol{\beta},\boldsymbol{\gamma}} \;=\; \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\gamma},\boldsymbol{\gamma}}^{-1} & \boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\gamma}}^{\top} \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\gamma}} & \boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{\beta}} \end{pmatrix},$$

where all needed terms are defined in (7).

4 Simulation Study

Through three simulation studies, we illustrate the performance of our methodology with exogenous continuous and discrete covariates. The continuous covariate $\boldsymbol{x}_1 = (x_{1,1}, \ldots, x_{N,1})$ is drawn from a standard normal distribution, i.e., $X_{i,1} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$ independently for $i = 1, \ldots, N$. The discrete covariate $\boldsymbol{x}_2 = (x_{1,2}, \ldots, x_{N,2})$ is sampled independently from a categorical distribution with three equally probable outcomes. Three summary statistics are chosen for the incidence and duration: $\boldsymbol{s}_{i,j}^{0 \to 1}(\mathcal{H}_t) = (s_{i,j,CCP}(\mathcal{H}_t),\ s_{i,j,\boldsymbol{z}_1}(\mathcal{H}_t),\ s_{i,j,\boldsymbol{z}_2}(\mathcal{H}_t))^\top$ and $\boldsymbol{s}_{i,j}^{1 \to 0}(\mathcal{H}_t) = (s_{i,j,NI}(\mathcal{H}_t),\ s_{i,j,\boldsymbol{z}_1}(\mathcal{H}_t),\ s_{i,j,\boldsymbol{z}_2}(\mathcal{H}_t))^\top$, where $z_{i,j,1} = |x_{i,1} - x_{j,1}|$ and $z_{i,j,2} = \mathbb{I}(x_{i,2} = x_{j,2})$. The true parameters are fixed as follows $\boldsymbol{\alpha}^{0 \to 1} = (-1/2, 1, 1/2)$ and $\boldsymbol{\alpha}^{1 \to 0} = (1/2, 1/2, 1/2)$. These parameter values guarantee a balance between the lengths of interactions and non-interactions for all pairs of nodes, making the dataset realistic and suitable for inference. The true popularity parameters for each actor $i = 1, \ldots, N$ are sampled from Gaussians:

$$\beta_i^{0 \to 1} \overset{\text{i.i.d}}{\sim} \mathcal{N}(-6 - 1/10 \log(N), 1)$$
$$\beta_i^{1 \to 0} \overset{\text{i.i.d}}{\sim} \mathcal{N}(8/5 - 1/10 \log(N), 1).$$

The popularity parameters decrease on average with increasing N, implying sparsity of observed events as the number of actors N grows. We set the time interval as $\mathcal{T} = [0, 10.000]$ with nine equally spaced change points $c_1, \ldots, c_9$. The true values of the baseline function, $\boldsymbol{\gamma}^{0 \to 1} = \boldsymbol{\gamma}^{1 \to 0}$, are linearly decreasing from zero to $-1/10$.

We assess the finite-sample performance of our estimators using three measures: the average point estimate (AVE), the root-mean-squared error (RMSE), and the coverage probabilities (CP), defined in the Supplementary Material A.1. Once $S = 1.000$ datasets are generated according to the algorithm detailed in Supplementary Material A.2, we run our inference algorithm of Section 3 independently on each dataset to obtain S experiments.

Table 1: Simulation study 1: for each effect, we report the AVE, RMSE, and CP.

Summary Statistic	α	AVE	RMSE	CP
Incidence ($\alpha^{0\to1}$)				
Current Common Partner	$-.5$	$-.510$	.074	.951
Continuous Cov.	1.0	.998	.010	.944
Categorical Cov.	.5	.499	.010	.957
Duration ($\alpha^{1\to0}$)				
Number Interaction	.5	.500	.010	.945
Continuous Cov.	.5	.502	.011	.939
Categorical Cov.	.5	.501	.011	.944

Simulation Study 1: Parameter Estimation and Model Selection. In the first simulation study, we assess how accurately our novel algorithm recovers the correct model for $N = 500$ actors. Once the model structure is inferred, we check if the inference on the parameter values is accurate.

We apply the following greedy model selection procedure: first, we fit a simple model where the incidence model only includes the current common partner, and the duration model has no covariates. Then, we calculate the Akaike Information Criterion for this initial model. We proceed by adding covariates to the incidence structure in a step-wise fashion, one at a time. For each fitted model, we again calculate the AIC and retain the new model as optimal if a better criterion value is obtained. Once we converge to an optimal incidence model structure, we apply an analogous procedure to the duration model. The model achieving the highest value of the criterion, overall, is retained as the
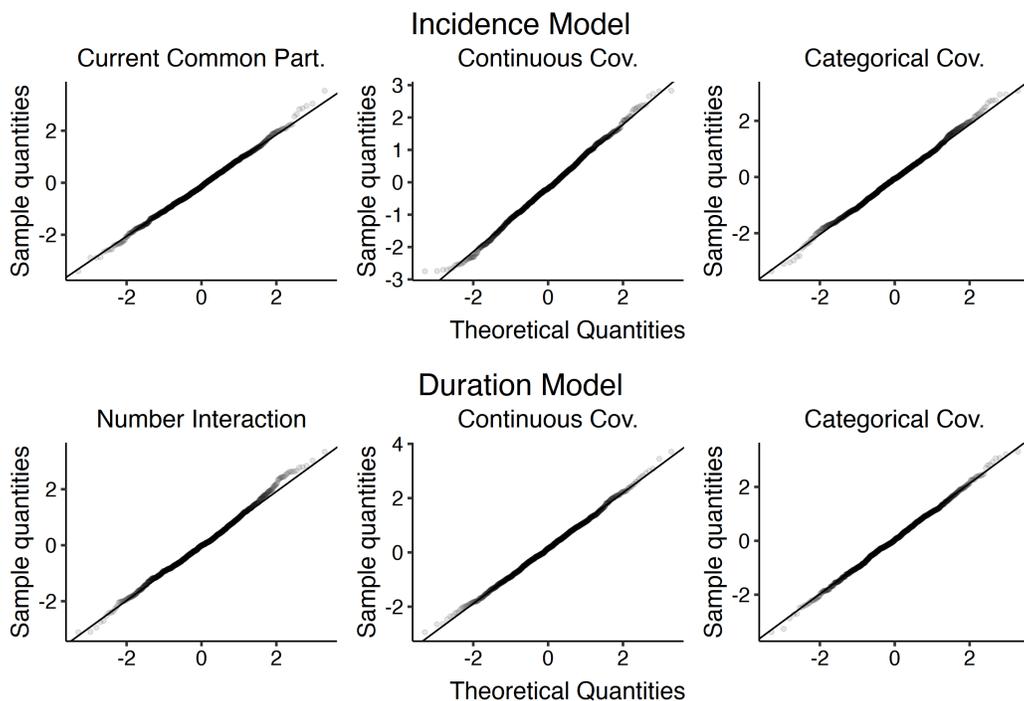
Figure 2: Simulation Study 1: Quantile-quantile plots for the incidence (top row) and duration model (bottom row) comparing the theoretical quantiles of a standard normal distribution with the sample quantiles of $z_{(s)} = \mathbf{\Lambda}(\hat{\boldsymbol{\theta}}_{(s)})^{1/2}(\hat{\boldsymbol{\alpha}}_{(s)} - \boldsymbol{\alpha})$ for $s = 1, ..., 1.000$.

optimal inferred model. The inferred optimal model coincides with the data-generating model in all of the datasets generated in this experiment.

The results in Table 1 demonstrate that all parameters encompassed in $\boldsymbol{\alpha}$ are estimated with high accuracy. The empirical coverage probability closely matches the nominal 0.95 level, indicating reliable uncertainty quantification. This suggests accurate and robust inference. This finding is further supported by the quantile-quantile plots in Figure 2, which illustrate a close alignment of sample quantiles with theoretical normal quantiles. For the current common partner statistic, we note a slightly higher RMSE compared to the other statistics. This phenomenon may stem from the statistic using less data by only relying on the current status of the network rather than its entire history.
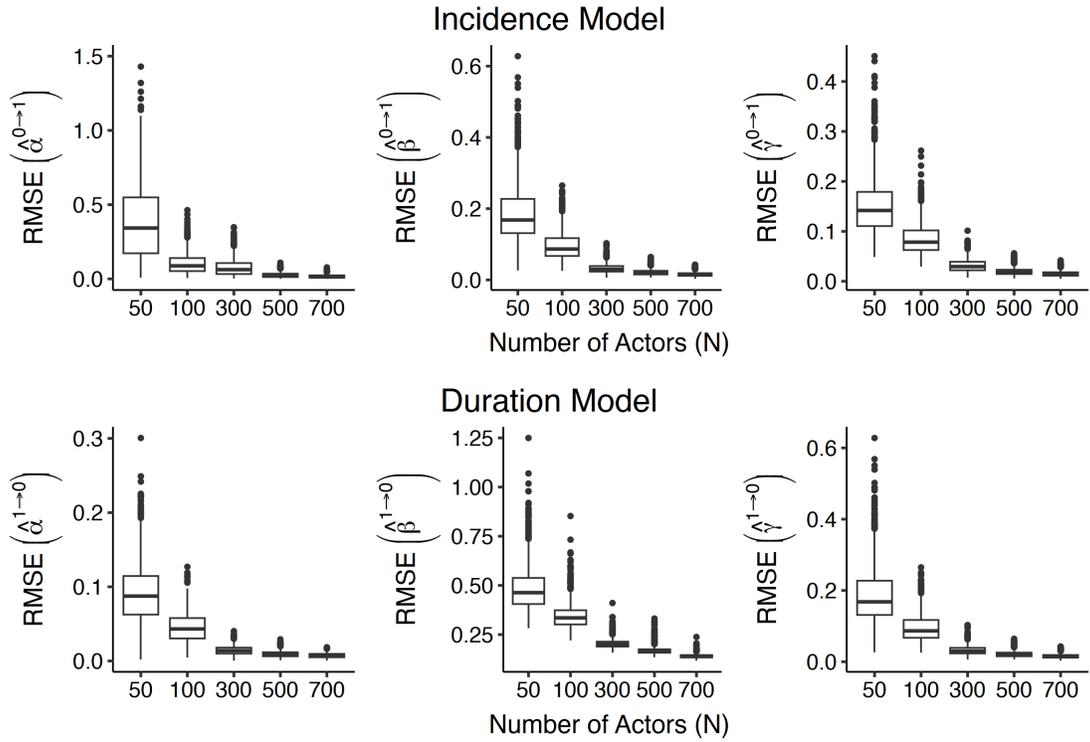
Figure 3: Simulation Study 2: RMSE of the estimates of $\boldsymbol{\alpha}$ (first column), $\boldsymbol{\beta}$ (second column), and $\boldsymbol{\gamma}$ (third column) over all simulations for the incidence (top row) and duration model (bottom row).

Simulation Study 2: Estimation Error under Increasing Actors. We rely on an analogous setup as in Simulation Study 1, with the only difference being that the number of actor varies from 50 to 700. We assess the estimation error of all parameters.

For both the incidence and duration models, Figure 3 shows that the RMSE of all estimated parameters decreases as the number of actors N increases. The baseline intensity and summary statistics coefficients exhibit lower estimation error than the popularity parameters, whose dimensionality grows with N. This behavior is expected due to the increasing dimension of the popularity effects. These results provide empirical evidence for the consistency of our estimators in regimes where the number of parameters is a function of the number of actors.
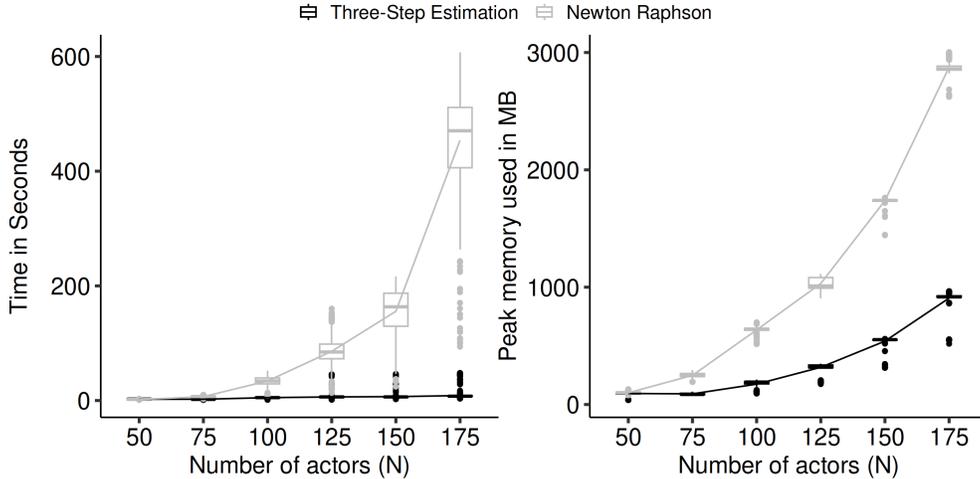
Figure 4: Simulation Study 3: Comparison of computational time (left) and memory needed for estimation (right) for different number of actors of our proposed three-step estimator and the common Newton-Raphson estimator.

Simulation Study 3: Computational Improvement. Last, we compare the computational efficiency of our proposed block-coordinate algorithm (Section 3) to the Newton-Raphson approach. We measure computational speed by the execution time (in seconds) for estimating the model on a single dataset. Memory allocation is assessed by each algorithm's peak random access memory usage. The peak memory usage reflects the largest dataset a system can handle.

Figure 4 visualizes the execution time and memory usage for both algorithms under increasing number of actors. The largest dataset we consider has only up to 175 nodes, beyond which the Newton-Raphson algorithm becomes impractical. The results provide empirical evidence that our block-coordinate ascent algorithm improves upon state-of-the-art methods by several orders of magnitude. From the right plot in Figure 4, it is evident that our proposed algorithm scales more efficiently than the Newton-Raphson method. Additionally, our approach exhibits lower variability in computing time and memory usage. Both findings align with the theoretical complexity reduction stated in Section 3.

5 Application to Physical and Digital Interactions

We demonstrate the DEM in an application to the Copenhagen Networks Study (Sapiezyn-ski et al., 2019). The study, conducted between 2012 and 2013, tracked 682 first-year students at the Technical University of Denmark over approximately 27 days. This data includes various forms of interactions, such as co-location (measured via Bluetooth), social media interactions (friendships on Facebook), and other communication channels (phone calls and text messages). The available data enables us to compare the evolution of physical and digital interactions between the same actors over time. This enables the identification of how individuals prioritize different relationships and the circumstances under which they rely on digital communication and face-to-face interactions. By doing that, we are able to obtain novel insights into how human behavior unfolds simultaneously over multiple channels (Stopczynski et al., 2014). In particular, we study digital communication through call data, while co-location events act as a proxy for physical activities. By distinguishing when interactions begin and end, we can (i) fully utilize the available data and (ii) assess whether similar factors drive both the incidence and duration of events. Text messaging events, in contrast, are instantaneous events without duration that naturally fit into the REM described in (1). We showcase how the computational techniques developed in Section 3 apply to this setting in the Supplementary Materials B.4.

5.1 Data

Several studies have used data from the Copenhagen Networks Study for descriptive analyses. Sekara et al. (2016) leverage this dataset to demonstrate how interaction data with fine-grained temporal resolution can be used to examine group formation. Building on the foundational work of Granovetter (1973), Ureña-Carrion et al. (2020) investigate the

strength of interactions with the data. Contrary to the standard practice to quantify the strength of ties by the number or duration of interactions, the authors employed descriptive measures, such as dynamic topological overlap between actors, to study the *strength of weak ties*. Using similar data from the Telefonica telecommunications company, Miritello et al. (2013) explore how interaction patterns change as personal networks expand. That said, no probabilistic models, such as REMs, have been applied to this dataset.

Physical Interactions via Bluetooth Scans. Physical interactions are quantified by co-location events, which, in turn, are measured via Bluetooth communication between mobile devices. Bluetooth scans run every five minutes determine when people are close to one another. We define a co-location event as as the presence of two individuals in at least two consecutive scans. Therefore, a physical interaction event $d_P = (i, j, b, e)$ indicates that student i was in the same place as student j from time b to e.

Digital Interactions via Call Data. The call data between students serves as a proxy for digital communication. A call event $d_D = (i, j, b, e)$ encodes a call between student i and student j between the time points b and e. The sets $\mathcal{U}^{0 \to 1}(t)$ and $\mathcal{U}^{1 \to 0}(t)$ are defined to ensure that no student can be active in multiple simultaneous calls.

Covariate Information. Participants provided time-invariant information on their Facebook friends and gender. However, this information is missing for a subset we exclude from the study. Assuming these values are missing completely at random, this exclusion does not affect the validity of our inferential conclusions. Additionally, the popularity estimates $\beta_i^{0 \to 1}$ and $\beta_i^{1 \to 0}$ of actor i for the physical and digital communication are only identifiable if actor i participated in at least one durational event in both communication modes (physical and digital). Applying these inclusion criteria results in a final dataset of 155,316 physical
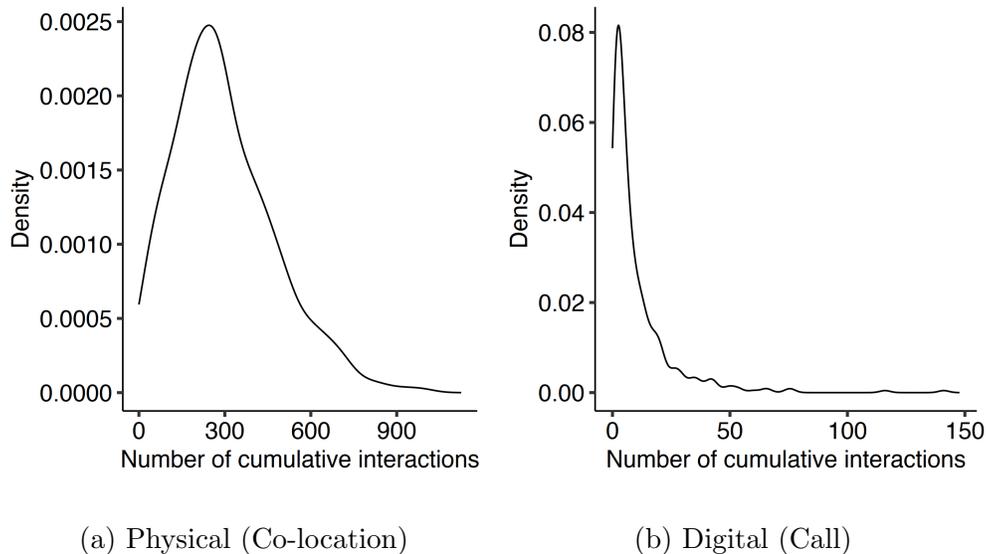
(a) Physical (Co-location)　　　　　　(b) Digital (Call)

Figure 5: Application: Kernel-density estimator of cumulative durational events per actors in the of physical (A) and digital interactions (B).

interactions and 4,152 digital interactions among $N = 400$ actors. Text messages were also collected between study participants.

5.2　Model Specification

To examine endogenous effects, we use summary statistics from Section 2.2. We evaluate whether actors i and j currently share and have previously shared a common partner, and we account for pairwise inertia by including the number of past interactions as statistics. For call events, having a common partner at the time of the event is impossible, as an individual can participate in only one phone call at a time. Accordingly, we exclude the current common partner terms from the corresponding incidence and duration models. We incorporate covariates related to Facebook friendships ($\boldsymbol{z}_1 = (z_{i,j,1})$, where $x_{i,j,1} = 1$ if actors i and j are Facebook friends) and gender ($\boldsymbol{x}_2 = (x_{i,2})$, where $x_{i,2} = 1$ if actor i is female) as dyadic and monadic statistics, respectively: $z_{i,j,1} = x_{i,j,1}$ for dyadic relationships

and $z_{i,j,2} = \mathbb{I}(x_{i,2} = x_{j,2})$ for gender homophily.

Figure 5 presents density estimators of the number of interactions individual actors are involved in for each event network. These distributions indicate substantial heterogeneity among actors in both settings. Including popularity estimates allows us to account for this in a flexible manner. Interaction frequencies are expected to fluctuate considerably throughout the day since students are more likely to interact during the daytime hours. To accommodate this, we allow the baseline function $f(t, \boldsymbol{\gamma})$ to change hourly, resulting in approximately $Q = 672$ equidistant change points[1]. We conduct a sensitivity analysis in Supplementary Material B.3 to show that allowing $f(t, \boldsymbol{\gamma})$ to change every two hours does not substantially affect the results. In total, $1,083 = 11(P) + 400(N) + 672(Q)$ parameters characterize our model. While solving this problem is infeasible with state-of-the-art techniques, our novel scalable block-coordinate ascent algorithm enables efficient model estimation within a few minutes on a standard laptop.

5.3 Results

The parameter estimates from the DEM reveal distinct patterns governing the incidence and duration of physical and digital interaction events. The estimates of $\boldsymbol{\alpha}^{0 \to 1}$ and $\boldsymbol{\alpha}^{1 \to 0}$ are provided in Table 2, whereas the popularity estimates $\boldsymbol{\beta}^{0 \to 1}$ and $\boldsymbol{\beta}^{1 \to 0}$ are shown in Figure 6. We conduct a post hoc analysis in Supplementary Material B.2 examining temporal trends and seasonal variations in the baseline intensities.

---

[1]If no interaction is observed during a particular hour, we assume that the baseline did not change from the previous hour. Therefore, the number of parameters in $\boldsymbol{\gamma}$ fluctuates slightly between each estimated model.

Table 2: Parameter estimates ($\hat{\boldsymbol{\alpha}}$) and corresponding standard errors (SE) of the Durational Event Model applied to physical (co-location, columns 2-4) and digital (call, columns 5-7) interaction data. The column $2^{\hat{\alpha}}$ shown for statistics transformed by $\log(\cdot + 1)$ represents the effect of the first change in the respective statistic.

Summary Statistic	Physical (Co-location)			Digital (Call)		
	$\hat{\alpha}$	SE	$2^{\hat{\alpha}}$	$\hat{\alpha}$	SE	$2^{\hat{\alpha}}$
Incidence ($\hat{\boldsymbol{\alpha}}^{0\to1}$)						
Current Common Partner	2.867	.006	7.295			
General Common Partner	.726	.007	1.654	.224	.125	1.168
Number Interaction	1.129	.005	2.187	1.631	.039	3.097
Friendship Match	.383	.010		5.687	.116	
Both Female	−.021	.010		.203	.085	
Duration ($\hat{\boldsymbol{\alpha}}^{1\to0}$)						
Number Interaction	−.158	.005	.896	−.274	.073	.827
Current Interaction	−.102	.002	.932	−.053	.024	.964
Current Common Partner	−.312	.006	.806			
General Common Partner	.080	.009	1.057	.518	.186	1.432
Friendship Match	−.535	.010		−.337	.464	
Both Female	−.018	.010		−.227	.320	

Summary Statistics. In the incidence model, shared current common partners are a key determinant of physical interactions. The first shared partner between actors i and j increases the corresponding event intensity by a multiplicative factor of 7.295. This

coefficient confirms that joining an existing group of people is much more likely than starting a physical interaction with a single actor. General common partners also affect the event intensities. However, their effects are moderate: they increase the incidence intensity of physical interactions by a multiplicative factor of 1.654 and of digital interactions by 1.168 in the same setting. Besides these triadic closure effects, repeated interactions affect the incidence intensity. The first interaction between arbitrary actors i and j increases $\lambda_{i,j}^{0\to1}\left(t\mid\mathcal{H}_t,\boldsymbol{\theta}^{0\to1}\right)$ of physical and digital interactions by multiplicative factors of 2.187 and 3.097, respectively. By the interpretation shown in (3), a shared Facebook friend affects the incidence intensity of digital communication by the multiplicative factor $\exp(5.687)\approx295$. The incidence intensity of physical meetings is increased only by the factor $\exp(.383)\approx1.46$ in the same setting. We estimate divergent effects for gender homophily. The incidence intensity of digital interaction between two female actors is significantly lower than for pairs where both actors are not female. However, for physical interactions, we do not observe this significant difference. In general, the results of the model for physical interactions based on co-location align with the results obtained in Hoffman et al. (2020), where similar data was studied from a group-level perspective.

Both the number of current common partners and prior interactions significantly prolong the duration of physical interactions. The first common partner and prior interactions reduce the duration intensity by multiplicative factors of 0.896 and 0.932, respectively. Similarly, shared friendship status approximately halves the duration intensity of physical interactions. This finding suggests that strong relational ties stabilize over time. In contrast, gender homophily does not have a statistically significant effect on the duration intensity of physical interactions. The durations of digital interactions, measured via call events, are driven by similar factors. However, in this case, all exogenous covariate effects
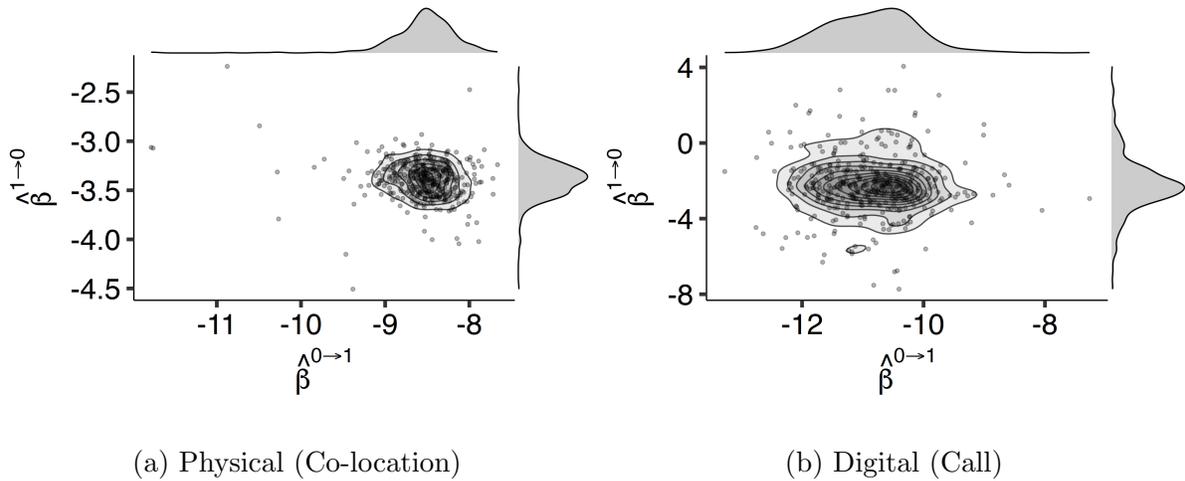
(a) Physical (Co-location) (b) Digital (Call)

Figure 6: Estimates of the popularity parameters for physical (A) and digital interactions (B).

are not significantly different from zero.

Popularity Estimates. The number of potential actor pairs to begin interacting is typically orders of magnitude larger than the pairs that may end them, while the number of started and ended events is roughly the same. This imbalance is reflected in the overall level of estimates shown in Figure 6, where the average of $\hat{\beta}^{0\to1}$ is lower than the average of $\hat{\beta}^{1\to0}$ for both interaction modes. Consistent with Figure 5, the popularity parameters in the incidence model are generally higher for co-location than for call interactions. The contour lines in Figure 5 enable a comparison of individual actors' popularity estimates between the incidence and duration models. There is no clear correlation between the two estimates for physical interactions. However, a pattern emerges for digital interactions where low to average popularity coefficients in the incidence model often correspond to average popularity effects in the duration model.

Overall, our results suggest that physical interactions are mainly driven by immediate social context and co-presence (e.g., group settings). In contrast, digital interactions are

primarily shaped by pre-existing social ties represented by Facebook friendships. The popularity coefficients in the duration model are higher for physical interactions, suggesting that digital interactions typically last longer than physical ones. An interpretation of the baseline intensity is provided in Supplementary Material B.2.

6   Discussion

The proposed framework, accompanied by a scalable estimation algorithm, offers a robust foundation for modeling the dynamics of durational events. Nonetheless, there are still several exciting avenues for future extensions. Concerning the model specification, the integration of multiple states beyond two states and higher-order events such as meetings between multiple actors lasting a particular time (see, e.g., Espinosa-Rada et al., 2024) would further broaden the model's applicability. For large populations, it is reasonable to assume that durational events between actors i and j are primarily influenced by their respective local neighborhoods, rather than by the entire network. This assumption is grounded in the concept of local dependence, introduced by Schweinberger and Handcock (2015), which can be extended to the context of durational events. This mounts to a combination of the algorithm introduced in this paper with model-based clustering for durational events (Rastelli and Fop, 2020).

The modular structure of the proposed block-coordinate ascent algorithm from Section 3 allows for the independent application of any approximation or acceleration technique in (8), (12), and (14) for the three updates separately. In particular, it will be worthwhile to adapt sample-based approximations, such as case-control sampling (Lerner and Lomi, 2020) or Horvitz–Thompson-estimators (Raftery et al., 2012), to scale estimation to even larger networks. For Step 3, acceleration techniques tailored to MM algorithms, discussed

in Agarwal and Xu (2024), can be employed to improve convergence speed.

We provide the R package `DEM`, which implements our method and the suite of summary statistics introduced in Section 2.2. To accommodate problem-specific requirements and different data sources, users can define custom summary statistics and transformations using `C++` code. Our approach parallels the functionality of the `ergm.userterms` package, which enables the inclusion of user-defined statistics for estimating Exponential Random Graph Models (Hunter et al., 2013). Similarly, `DEM` is a flexible and extensible toolbox, allowing researchers to tailor the statistical framework to meet their specific modeling needs.

References

O. O. Aalen and H. K. Gjessing. Stochastic processes in survival analysis. In *Advances in Statistical Modeling and Inference*, Series in Biostatistics, pages 23–43. World Scientific Publishing, Singapore, 2007.

O. O. Aalen, Ø. Borgan, and H. K. Gjessing. Marginal and dynamic models for recurrent events and clustered survival data. In *Survival and Event History Analysis: A Process Point of View*, pages 301–346. Springer, New York, NY, 2008.

A. Agarwal, L. , Kevin H., and L. and Xue. Clustering time-evolving networks using temporal exponential-family random graph models with conditional dyadic independence and dynamic latent blocks. *Journal of Computational and Graphical Statistics*, (OnlineFirst): 1–21, 2025. doi: 10.1080/10618600.2025.2484011.

M. Agarwal and J. Xu. Quasi-Newton acceleration of EM and MM algorithms via Broyden's

method. *Journal of Computational and Graphical Statistics*, 33(2):393–406, 2024. doi: 10.1080/10618600.2023.2257261.

O. E. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. John Wiley & Sons, New York, NY, 1978.

C. T. Butts. A relational event framework for social action. *Sociological Methodology*, 38 (1):155–200, 2008. doi: 10.1111/j.1467-9531.2008.00203.x.

C. T. Butts and C. S. Marcum. A relational event approach to modeling behavioral dynamics. In A. Pilny and M. S. Poole, editors, *Group Processes*, pages 51–92. Springer, Cham, 2017. doi: https://doi.org/10.1007/978-3-319-48941-4_4.

B. Cai, J. Zhang, and Y. Guan. Latent network structure learning from high-dimensional multivariate point pq rocesses. *Journal of the American Statistical Association*, 119(545): 95–108, 2024. doi: 10.1080/01621459.2022.2102019.

S. Chatterjee, P. Diaconis, and A. Sly. Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435, 2011. doi: 10.1214/10-AAP728.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 34(2):187–220, 1972. doi: 10.1111/j.2517-6161.1972. tb00899.x.

N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006. doi: 10.1007/s00779-005-0046-3.

A. Espinosa-Rada, J. Lerner, and C. Fritz. Socio-cognitive networks between researchers: investigating scientific dualities with the group-oriented relational hyperevent model. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2407.21067.

R. G. Everitt. Bayesian parameter estimation for latent Markov random fields and social networks. *Journal of Computational and Graphical Statistics*, 21:940–960, 2012. doi: https://doi.org/10.1080/10618600.2012.687493.

G. Fang, G. Xu, H. Xu, X. Zhu, and Y. Guan. Group network Hawkes process. *Journal of the American Statistical Association*, 119(547):2328–2344, 2024. doi: 10.1080/01621459. 2023.2257889.

C. Fritz, M. Lebacher, and G. Kauermann. Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time. *Statistica Neerlandica*, 74(3):275–299, 2020. doi: 10.1111/stan.12198.

C. Fritz, M. Mehrl, P. W. Thurner, and G. Kauermann. All that glitters is not gold: Relational events models with spurious events. *Network Science*, 11(2):184–204, 2023. doi: 10.1017/nws.2022.22.

C. Fritz, C.-P. Georg, A. Mele, and M. Schweinberger. Vulnerability webs: Systemic risk in software networks. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2402.13375.

M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

D. A. Harville. *Matrix algebra from a statistician's perspective*. Springer, New York, NY, 1997.

X. He and Q.-M. Shao. On parameters of increasing dimensions. *Journal of Multivariate Analysis*, 73(1):120–135, 2000. doi: 10.1006/jmva.1999.1873.

M. Hoffman, P. Block, T. Elmer, and C. Stadtfeld. A model for the dynamics of face-to-face

interactions in social groups. *Network Science*, 8(S1):1–22, 2020. doi: 10.1017/nws.2020. 3.

R. M. Hummel, D. R. Hunter, and M. S. Handcock. Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics*, 21:920–939, 2012. doi: https://doi.org/10.1080/10618600.2012.679224.

D. R. Hunter. Curved exponential family models for social networks. *Social Networks*, 29 (2):216–230, 2007. doi: 10.1016/j.socnet.2006.08.005.

D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58 (1):30–37, 2004. doi: https://doi.org/10.1198/0003130042836.

D. R. Hunter, S. M. Goodreau, and M. S. Handcock. Ergm.userterms: A template package for extending statnet. *Journal of Statistical Software*, 52:1–25, 2013. doi: 10.18637/jss. v052.i02.

I. H. Jin and F. Liang. Fitting social network models using varying truncation stochastic approximation MCMC algorithm. *Journal of Computational and Graphical Statistics*, 22:927–952, 2013. doi: https://doi.org/10.1080/10618600.2012.680851.

P. N. Krivitsky and M. S. Handcock. A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46, 2014. doi: 10.1111/rssb.12014.

K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9:1–20, 2000. doi: https: //doi.org/10.2307/1390605.

D. Lazer, A. Pentland, L. Adamic, S. Aral, A. L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2009. doi: 10.1126/science.1167742.

J. Lerner and A. Lomi. Reliability of relational event model estimates under sampling: How to fit a relational event model to 360 million dyadic events. *Network Science*, 8(1): 97–135, 2020. doi: 10.1017/nws.2019.57.

J. Lerner, M. Bussmann, T. A. B. Snijders, and U. Brandes. Modeling frequency and type of interaction in event networks. *Corvinus Journal of Sociology and Social Policy*, 4(1): 3–32, 2013. doi: 10.14267/cjssp.2013.01.01.

J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *Proceedings of the 17th International Conference on World Wide Web*, pages 915–924, 2008. doi: 10.1145/1367497.1367620.

C. Matias, T. Rebafka, and F. Villers. A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika*, 105(3):665–680, 2018. doi: 10.1093/biomet/asy016.

G. Miritello, E. Moro, R. Lara, R. Martínez-López, J. Belchamber, S. G. B. Roberts, and R. I. M. Dunbar. Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks*, 35(1):89–95, 2013. doi: 10.1016/j.socnet.2013.01.003.

P. O. Perry and P. J. Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 75(5):821–849, 2013. doi: 10.1111/rssb.12013.

S. Portnoy. Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *The Annals of Statistics*, 16(1):356–366, 1988. doi: 10.1214/aos/1176350710.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL `https://www.R-project.org/`.

A. E. Raftery, X. Niu, P. D. Hoff, and K. Y. Yeung. Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21:901–919, 2012. doi: 10.1080/10618600.2012.679240.

R. Rastelli and M. Fop. A stochastic block model for interaction lengths. *Advances in Data Analysis and Classification*, 14(2):485–512, 2020. doi: 10.1007/s11634-020-00403-w.

M. Razaviyayn, M. Hong, and Z.-Q. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23 (2):1126–1153, 2013. doi: 10.1137/120891009.

G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001. doi: 10.1214/ss/1015346320.

P. Sapiezynski, A. Stopczynski, D. D. Lassen, and S. Lehmann. Interaction data from the Copenhagen Networks Study. *Scientific Data*, 6(1):315, 2019. doi: 10.1038/s41597-019-0325-x.

M. Schweinberger and M. S. Handcock. Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 77(3):647–676, 2015. doi: 10.1111/rssb.12081.

V. Sekara, A. Stopczynski, and S. Lehmann. Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences*, 113(36):9977–9982, 2016. doi: 10.1073/pnas.1602803113.

C. Stadtfeld, J. Hollway, and P. Block. Dynamic network actor models: Investigating coordination ties through time. *Sociological Methodology*, 47(1):1–40, 2017. doi: 10. 1177/0081175017709295.

A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann. Measuring Large-Scale Social Networks with High Resolution. *PLOS ONE*, 9(4):e95978, 2014. doi: 10.1371/journal.pone.0095978.

J. Ureña-Carrion, J. Saramäki, and M. Kivelä. Estimating tie strength in social networks using temporal communication data. *EPJ Data Science*, 9(1):1–20, 2020. doi: 10.1140/ epjds/s13688-020-00256-5.

D. Vu, A. Lomi, D. Mascia, and F. Pallotti. Relational event models for longitudinal network data with an application to interhospital patient transfers. *Statistics in Medicine*, 36(14):2265–2287, 2017. doi: 10.1002/sim.7247.

D. Q. Vu, A. U. Asuncion, D. R. Hunter, and P. Smyth. Continuous-time regression models for longitudinal networks. *Advances in Neural Information Processing Systems*, pages 2492–2500, 2011a.

D. Q. Vu, A. U. Asuncion, D. R. Hunter, and P. Smyth. Dynamic egocentric models for citation networks. *International Conference on Machine Learning*, pages 857–864, 2011b.

D. Q. Vu, D. R. Hunter, and M. Schweinberger. Model-based clustering of large networks. *The Annals of Applied Statistics*, 7(2):1010–1039, 2013. doi: 10.1214/12-AOAS617.

C. Wagner, M. Strohmaier, A. Olteanu, E. Kıcıman, N. Contractor, and T. Eliassi-Rad. Measuring algorithmically infused societies. *Nature*, 595(7866):197–204, 2021. doi: 10.1038/s41586-021-03666-1.

Supplementary material

Scalable Durational Event Models: Application to Physical and Digital Interactions

Cornelius Fritz[1*], Riccardo Rastelli[2†], Michael Fop[2†], Alberto Caimo[2†]

[1]School of Computer Science and Statistics, Trinity College Dublin

[2]School of Mathematics and Statistics, University College Dublin

---

[*]**Corresponding author:** fritzc@tcd.ie.

[†]Riccardo Rastelli, Michael Fop, and Alberto Caimo contributed equally to the work. Alberto Caimo conceived the original idea of the model.

A Further Information on the Simulation Study

A.1 Definition of Performance Measures

The AVE of a specific estimate $\hat{\boldsymbol{\theta}}$ is the empirical mean of the estimates over the S datasets:

$$\text{AVE}(\hat{\boldsymbol{\theta}}) \;=\; \frac{1}{S}\sum_{s=1}^{S}\hat{\boldsymbol{\theta}}_{(s)},$$

where $\hat{\boldsymbol{\theta}}_{(s)} = (\hat{\beta}_{(s)}, \hat{\beta}_{(s)}, \hat{\gamma}_{(s)})$ is the estimate of $\boldsymbol{\theta}$ in the sth simulation run. Given that all datasets are generated under the same parameter values, this statistic provides a meaningful assessment of estimation bias. To gauge both bias and variance, we report the RMSE of $\boldsymbol{\theta}$:

$$\text{RMSE}(\hat{\boldsymbol{\theta}}) \;=\; \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\hat{\boldsymbol{\theta}}_{(s)} - \boldsymbol{\theta}\right)^{\top}\left(\hat{\boldsymbol{\theta}}_{(s)} - \boldsymbol{\theta}\right)}.$$

To assess the uncertainty quantification, we examine coverage probabilities. The coverage probability is the percentage of simulations in which the true parameter falls within the confidence interval based on the normal approximation:

$$Z_{(s)} \;=\; \Lambda(\hat{\boldsymbol{\theta}}_{(s)})^{1/2}(\hat{\alpha}_{(s)} - \alpha) \;\sim\; \mathcal{N}(0,1). \tag{A.1}$$

To validate this normal approximation, we compare the observed quantiles of $Z_{(s)}$ to those we expect from standard normal random variables.

The Akaike Information Criterion (AIC, Akaike, 1973) for the DEM with converged estimate

$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}^{0\rightarrow1}, \hat{\boldsymbol{\theta}}^{1\rightarrow0})$ under M events is given by the sum of the AIC of the incidence and duration model:

$$AIC^{0\rightarrow1} \;\coloneqq\; 2\log M - \ell(\hat{\boldsymbol{\theta}}^{0\rightarrow1}) \text{ and } AIC^{1\rightarrow0} \;\coloneqq\; 2\log M - \ell(\hat{\boldsymbol{\theta}}^{1\rightarrow0}).$$

Result: Durational Event Matrix $E \in \mathbb{R}^{L \times 4}$, $\quad E_{i,1}, E_{i,2} \in \mathcal{A}$, $\quad E_{i,3} \in \mathcal{T}$,
$\qquad E_{i,4} \in \{0, 1\}$ for $i \in \mathcal{A}$, where L is the maximal number of events.

Set $t_{\mathrm{curr}} = 0$ and $n_{\mathrm{ev}} = 0$.

while $t_{curr} < T$ *or* $n_{ev} < L$ **do**

    1. Step: Compute Intensities

    Set

$$\lambda_{i,j}^{0\to1}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}\right) = \exp\left(\boldsymbol{\alpha}^{0\to1}\, \boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_{t_{\mathrm{curr}}}) + \beta_i^{0\to1} + \beta_j^{0\to1} + o^{0\to1}\right)$$

    and

$$\lambda_{i,j}^{1\to0}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{1\to0}\right) = \exp\left(\boldsymbol{\alpha}^{1\to0}\, \boldsymbol{s}_{i,j}^{1\to0}(\mathcal{H}_{t_{\mathrm{curr}}}) + \beta_i^{1\to0} + \beta_j^{1\to0} + o^{1\to0}\right)$$

    for all $(i, j) \in \mathcal{B}$.

    2. Step: Select Intensities

    Set $\boldsymbol{\lambda}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}, \boldsymbol{\theta}^{1\to0}\right) = \left(\lambda_{i,j}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}, \boldsymbol{\theta}^{1\to0}\right)\right)$

$$\begin{aligned} \lambda_{i,j}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}, \boldsymbol{\theta}^{1\to0}\right) = \; & (1 - f_{i,j}(t_{\mathrm{curr}}))\, \lambda_{i,j}^{0\to1}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}\right) \\ + \; & f_{i,j}(t_{\mathrm{curr}})\, \lambda_{i,j}^{1\to0}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{1\to0}\right) \end{aligned}$$

    for all $i, j) \in \mathcal{B}$.

    3. Step: Sample Time Increment

    Sample time between successive events:

$$T^{\star} \sim \mathrm{Exp}(\|\boldsymbol{\lambda}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}, \boldsymbol{\theta}^{1\to0}\right)\|_0)$$

    with observation $t^{\star}$.

    4. Step: Sample Event

    Sample which pair will experience the event:

$$(I^{\star}, J^{\star}) \sim \mathrm{Multinomial}\left(\frac{\boldsymbol{\lambda}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}, \boldsymbol{\theta}^{1\to0}\right)}{\|\boldsymbol{\lambda}\left(t_{\mathrm{curr}} \mid \mathcal{H}_{t_{\mathrm{curr}}}, \boldsymbol{\theta}^{0\to1}, \boldsymbol{\theta}^{1\to0}\right)\|_0}, n = 1\right)$$

    with observation $(i^{\star}, j^{\star})$.

    5. Step: Save Sampled Event

    Set $E_{n_{\mathrm{ev}},1} = i^{\star}$, $E_{n_{\mathrm{ev}},2} = j^{\star}$, $E_{n_{\mathrm{ev}},3} = t^{\star} + t_{\mathrm{curr}}$, and $E_{n_{\mathrm{ev}},4} = f_{i^{\star},j^{\star}}(t_{\mathrm{curr}})$.

    6. Step: Update History and Counters

    Update $\boldsymbol{s}_{i,j}^{0\to1}(\mathcal{H}_{t+t_{\mathrm{curr}}})$ and $\boldsymbol{s}_{i,j}^{1\to0}(\mathcal{H}_{t+t_{\mathrm{curr}}})$ for all $(i, j) \in \mathcal{B}$.

    Set $f_{i^{\star},j^{\star}}(t_{\mathrm{curr}} + t^{\star}) = 1 - f_{i^{\star},j^{\star}}(t_{\mathrm{curr}})$, $t_{\mathrm{curr}} = t^{\star} + t_{\mathrm{curr}}$, and $n_{\mathrm{ev}} = n_{\mathrm{ev}} + 1$.

end

Return $E[, 0 : (n_{\mathrm{ev}} - 1)]$

Algorithm 1: Pseudo-Code to sample durational events in $(0, T]$.

3

A.2 Simulation of Durational Events

We can reformulate our data representation in an equivalent form, as follows: a durational event is characterized by two tuples in the form (i, j, t, r), where t is a time stamp, and r is a binary indicator that determines if the tuple refers to the beginning $(r = 1)$ or end $(r = 0)$ of an interaction. Then we can equivalently represent durational event $d = (i, j, b, e)$ by $\tilde{d}_1 = (i, j, b, 1)$ and $\tilde{d}_1 = (i, j, e, 0)$. This is the format used for the provided R package DEM and in the following description of the pseudo-code to sample durational events from 2.

We present algorithm 1 for sampling durational events within an arbitrary time frame $(0, T]$ under the assumption that $f(t, \boldsymbol{\gamma}^{0 \to 1})$ and $f(t, \boldsymbol{\gamma}^{1 \to 0})$ remain constant in that interval. When these baseline step functions vary over time as step functions, we apply Algorithm 1 separately to each interval and concatenate the results. To accommodate for a flexible specification of $\mathcal{U}^{0 \to 1}(t)$ and $\mathcal{U}^{1 \to 0}(t)$, we include for each pair $(i, j) \in \mathcal{B}$ an indicator $p_{i,j}(t)$ of whether any type of event, start or end, is possible between the actors at time $t \in \mathcal{T}$. Our algorithm, therefore, assumes the specification of the following terms:

1. Number of actors N in the durational event network.

2. Sets of summary statistics for the incidence and duration model, $\boldsymbol{s}_{i,j}^{0 \to 1}(\mathcal{H}_t)$ and $\boldsymbol{s}_{i,j}^{1 \to 0}(\mathcal{H}_t)$.

3. Parameter vectors $\boldsymbol{\alpha}^{0 \to 1}$ and $\boldsymbol{\alpha}^{1 \to 0}$ for the summary statistics.

4. Parameter vectors $\boldsymbol{\beta}^{0 \to 1} \in \mathbb{R}^N$ and $\boldsymbol{\alpha}^{1 \to 0}$ for the popularity effects.

5. Parameters $o^{0 \to 1}$ and $o^{1 \to 0}$ defining the constant value of $f(0, \boldsymbol{\gamma}^{0 \to 1})$ and $f(0, \boldsymbol{\gamma}^{1 \to 0})$, respectively.

Together, these parameters are defining $\boldsymbol{\theta}^{0 \to 1}$ and $\boldsymbol{\theta}^{1 \to 0}$. Denoting all durational events that start before t by $\mathcal{D}(t)$, we can define the indicator function $f_{i,j}(t)$ whether actor pair

$(i, j) \in \mathcal{B}$ are at time $t \in \mathcal{T}$ are in an ongoing duration

$$f_{i,j}(t) := \mathbb{I}(\exists\, d = (i, j, b, e) \in \mathcal{D}(t) \text{ such that } t \in [b, e]).$$

B Further Information on the Application

B.1 Initial Censoring

For our applications, we assume that the observations start with the first event. Chronologically ordering the observed events by $d_1, \ldots, d_M$, we thus set $d_1 = (i_1, j_1, 0, e_1)$. We have no information on when exactly the observational process starts and, therefore, condition on the start of the first event without modeling it. This practice disregards the initial censoring of observations. However, the effect is negligible since 155.316 and 4,152 for the application to the proximity and call data, respectively.

B.2 Interpretation of Baseline Step Function

The estimates of the baseline step functions $f(t, \hat{\boldsymbol{\gamma}}^{0 \to 1})$ and $f(t, \hat{\boldsymbol{\gamma}}^{1 \to 0})$ for the co-location and call data are presented in the first rows of Figures 7 and 8, respectively. As described in Section 5.1, each step in these step functions corresponds to a one-hour interval. A clear daily cyclical pattern is evident in Figures 7 and 8. To separate long-term trends from cyclical variations, we apply the local smoothing technique developed by Cleveland (1979) to each baseline step function independently. The resulting trend estimates are represented by black smooth lines in all plots.

From the first row of Figures 7 and 8, we observe that controlling for all other effects, the baseline step function for the initiation of a durational event shows a consistent decline over the observed period. In contrast, the baseline step function for terminating a dura-
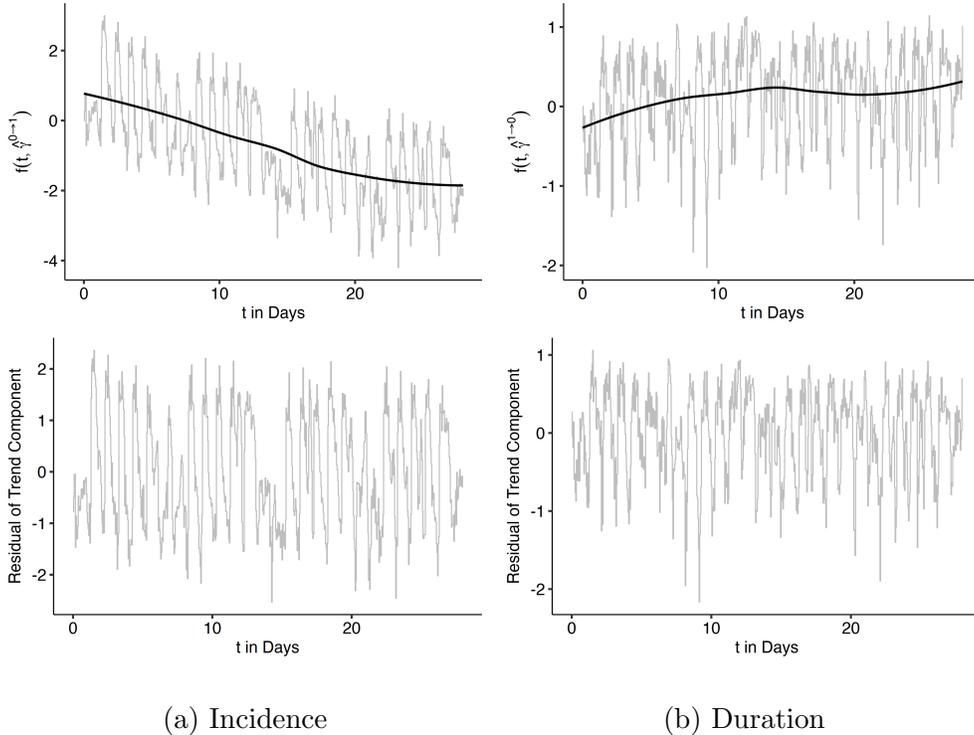
(a) Incidence   (b) Duration

Figure 7: Estimates of baseline intensity of the co-location model.

tional event exhibits a slight increase or remains stable. The second row of these figures displays the residuals from the local regression, which capture cyclical effects. Notably, the step function residuals for the incidence models reveal strong daily patterns, where high intensities align with daylight hours and low intensities correspond to nighttime. Additionally, in the first plot of the second row in Figure 7, an apparent disruption in this pattern around day 15 may be associated with weekend activities.

B.3   Sensitivity Analysis of Change-Point Specification

As detailed in the main text, the user must specify the set of change points $0 = c_0 < c_1 < \cdots < c_Q$, at which the baseline step function $f(t, \boldsymbol{\gamma}) = \sum_{q=1}^{Q} \gamma_q \, \mathbb{I}(c_{q-1} \leq t < c_q)$ is allowed to change. In the application presented in Section 5, we construct an hourly grid spanning the entire 27-day observation period. As a sensitivity check, we re-estimate the
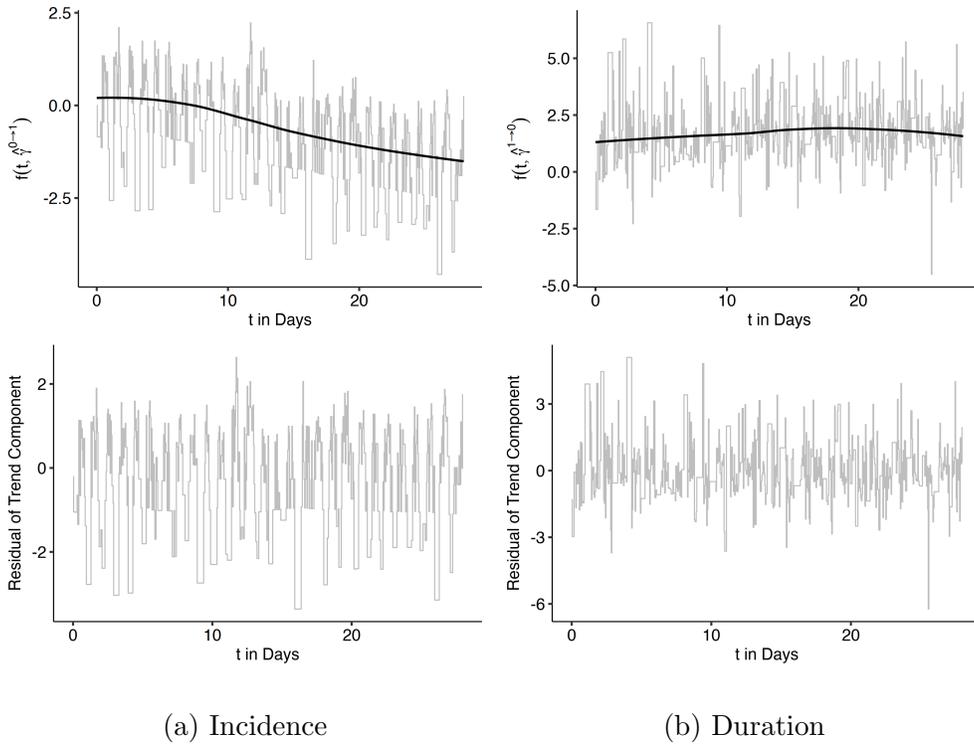
(a) Incidence   (b) Duration

Figure 8: Estimates of baseline step function of the call model.

same models reported in Table 2, but under a coarser temporal resolution, allowing for a change point every two hours instead of one. The corresponding estimates of $\hat{\boldsymbol{\alpha}}^{0\to1}$ and $\hat{\boldsymbol{\alpha}}^{1\to0}$ under this specification are presented in Tables 3 and 4, while the estimated baseline step function $f(t, \hat{\boldsymbol{\gamma}}^{0\to1})$ and $f(t, \hat{\boldsymbol{\gamma}}^{1\to0})$ are visualized in Figure 9. All those results confirm the robustness of our results concerning the choice of grid resolution for change points. The estimated effects remain qualitatively the same, demonstrating that the model's conclusions are not sensitive to the specific granularity of the baseline step function.

B.4 REM Application to Texting

For the application to the texting data between the students, we apply similar inclusion criteria as in the applications shown in the main analysis. However, we apply the condition that all included actors need to have at least one relational event, i.e., texting event.

Table 3: Parameter estimates ($\hat{\alpha}$) and standard errors (SE) of the Durational Event Model applied to the co-location data assuming a baseline step function that can change every hour (left column) and every second hour(right column).

Summary Statistic	Every Hour		Every two hours	
	$\hat{\alpha}$	SE	$\hat{\alpha}$	SE
Incidence ($\hat{\alpha}^{0 \to 1}$)				
Current Common Partner	2.867	.006	2.834	.006
General Common Partner	.726	.007	.726	.007
Number Interaction	1.129	.005	1.127	.005
Friendship Match	.383	.010	.383	.010
Both Female	$-.021$	.010	$-.020$	.010
Duration ($\hat{\alpha}^{1 \to 0}$)				
Current Interaction	$-.102$	.002	$-.101$	.002
Number Interaction	$-.158$	.005	$-.165$	.005
Current Common Partner	$-.312$	.006	$-.319$	.006
General Common Partner	.080	.009	.086	.009
Friendship Match	$-.535$	.010	$-.536$	.010
Both Female	$-.018$	.010	$-.017$	.010

Thereby, we obtain 38,286 relational events between $N = 426$ students. Each relational event $d = (i, j, t)$ is a tuple representing a text message between students i and j at time t.

In accordance with (2) , the intensity of an event at time $t \in \mathcal{T}$ for pair $(i, j) \in \mathcal{B}$ is:

$$\lambda_{i,j}\left(t \mid \mathcal{H}_t, \boldsymbol{\theta}\right) = \exp\left(\boldsymbol{\alpha}^\top \boldsymbol{s}_{i,j}(\mathcal{H}_t) + \beta_i + \beta_j + f(t, \boldsymbol{\gamma})\right). \tag{B.1}$$

Table 4: Parameter estimates ($\hat{\boldsymbol{\alpha}}$) and standard errors (SE) of the Durational Event Model applied to the call data assuming a baseline step function that can change every hour (left column) and every second hour(right column).

Summary Statistic	Every Hour		Every two hours	
	$\hat{\boldsymbol{\alpha}}$	SE	$\hat{\boldsymbol{\alpha}}$	SE
Incidence ($\hat{\boldsymbol{\alpha}}^{0 \to 1}$)				
General Common Partner	1.631	.039	1.631	.039
Number Interaction	.224	.125	.227	.121
Friendship Match	5.687	.116	5.681	.116
Both Female	.203	.085	.201	.084
Duration ($\hat{\boldsymbol{\alpha}}^{1 \to 0}$)				
Current Interaction	$-.053$	.024	$-.060$	.023
Number Interaction	$-.274$	.073	$-.162$	.065
General Common Partner	.518	.186	.579	.170
Friendship Match	$-.337$	.464	$-.670$	.390
Both Female	$-.227$	.320	$-.211$	.287

The summary statistics are

$$
\boldsymbol{s}_{i,j}(\mathcal{H}_t) = \begin{pmatrix} \log\left(\sum_{h \notin \{i,j\}} v_{i,h}(t)\, v_{h,j}(t) + 1\right) \\ \log(N_{i,j}(t) + 1) \\ \mathbb{I}(x_{i,j,1} = 1) \\ \mathbb{I}(x_{i,2} = x_{j,2}) \end{pmatrix},
$$

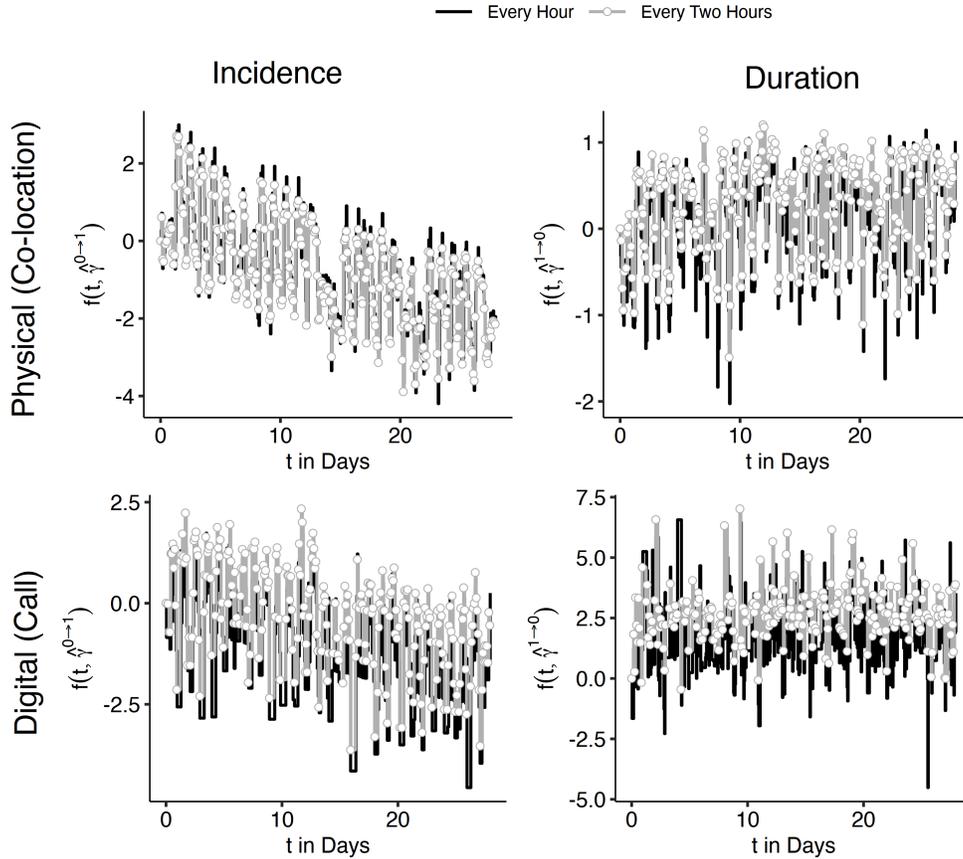where $v_{i,j}(t)$ indicates whether actors i and j have ever interacted before time t. The

Figure 9: Comparison of the estimated baseline step function under the assumption that it can change every hour and every two hours for durational events representing co-location (first row) and call events (second row) for the incidence (first column) and duration model (second column).

covariates are defined in Section 5.1.

Summary Statistics The estimates of α are given in Table 5 and can be interpret as detailed in Section 2.2. The first common partner and joint interaction have a multiplicative effect of 1.249 and 4.678 on the intensity between students i and j. These two coefficients demonstrate that sharing common text partner incentives triadic closure and that students repeatedly exchange texts between them. Contrary to these findings, being friends on Facebook has a small but significant negative effect on the intensity. Finally, we observe a

Table 5: Parameter estimates ($\hat{\alpha}$) and standard errors (SE) of the Relational Event Model applied to the Texting data.

Summary Statistic	$\hat{\alpha}$	SE	$2^{\hat{\alpha}}$
Common Partner	.321	.052	1.249
Number Interaction	2.226	.011	4.678
Friendship Match	$-.419$	.133	
Both Female	.229	.039	

positive gender-specific homophily effect, whereby two female students are more likely to exchange text messages than two males and mixed gender pairs.

Popularity Estimates. We visualize the observed dyadic text message exchanges in Figure 10 as a weighted network, where the weight of each edge represents the logarithm of the number of SMS events between the corresponding pair of students. We used the Fruchterman-Reingold algorithm (Fruchterman and Reingold, 1991) for this visualization. The size of each node is scaled relative to the corresponding popularity estimate $\hat{\beta}_i$. Figure 10 reveals that students with higher popularity estimates do not necessarily occupy central positions in the network. One possible explanation is that the interaction patterns between actors in central regions of the network are well captured by the summary statistics without the need for additional popularity effects.

Baseline Step Function. Finally, we visualize the estimated baseline step function $f(t, \hat{\gamma})$ in Figure 11. Consistent with the patterns observed in Figures 7 and 8, the function exhibits distinct daily cycles alongside a gradual downward trend over time.
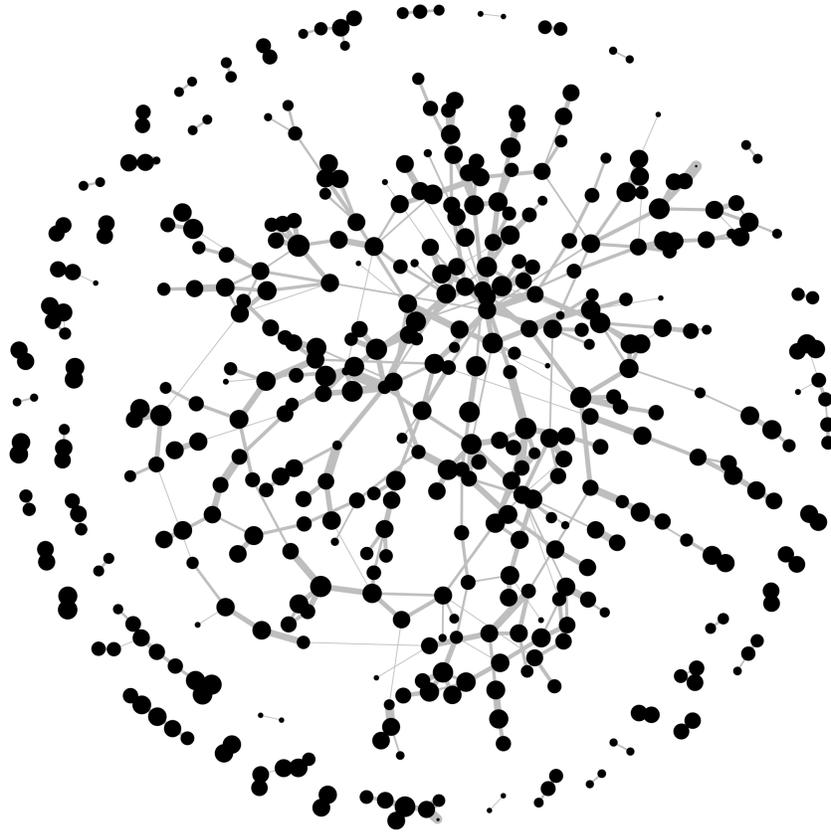
Figure 10: Observed texting events over the studied 28 days. The number of text messages between particular students is represented by the thickness of the edge and the size of the nodes represents the relative popularity estimate.
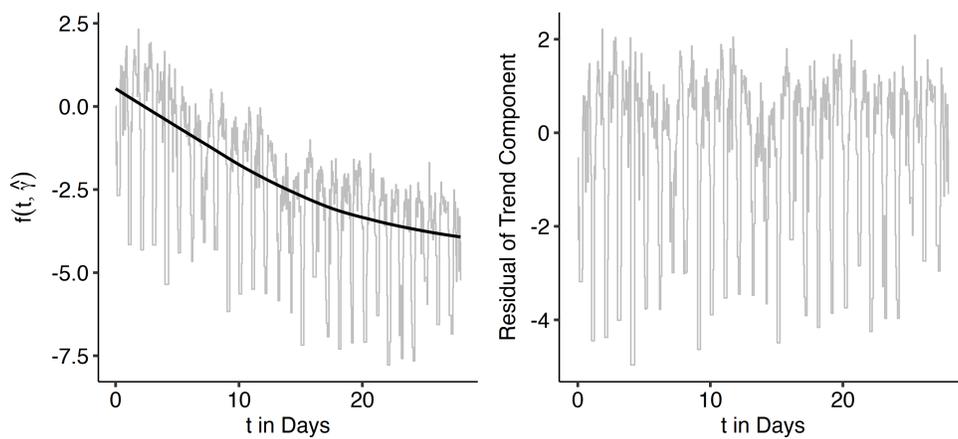

Figure 11: Estimated baseline step function for the REM applied to the Texting Data.

12

Supplementary Material References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, 267–281.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association 74*(368), 829–836.

Fruchterman, T. M. J. and E. M. Reingold (1991). Graph drawing by force-directed placement. *Software: Practice and Experience 21*(11), 1129–1164.