Computational Statistics and Data Analysis 224 (2026) 108428



Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda





Fully Bayesian estimation of temporal decay in ordinal relational event models

Philip Leifeld [a,1,*], Jackie Siaw Tze Wong [b,2]

[a] Department of Social Statistics, University of Manchester, Humanities Bridgeford Street Building, Oxford Road, Manchester, M13 9PL, UK
[b] School of Mathematics, Statistics and Actuarial Science, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

ARTICLE INFO

Keywords:
Relational event model
Bayesian estimation
Network science
Social network analysis
Complex social systems
Longitudinal

ABSTRACT

Relational event models (REMs) can infer the generative properties of longitudinally observed social networks with instantaneous edges. They assume conditional independence of edges given sufficient network statistics formed over the past event sequence. A popular specification in REMs is to subject these statistics to exponential temporal decay with a fixed half-life parameter to attribute higher importance to more recent edge events in the formation of network statistics. Assuming a fixed half-life parameter may cause biased estimates and obfuscates the temporal horizon over which network effects operate in empirical social systems. These limitations are addressed by proposing fully Bayesian estimation of REMs and designating the half-life parameter as an estimable quantity. A "pre-computation" strategy is devised to speed up calculations for practical feasibility of the sampling procedure. The approach is adapted to discourse network analysis, which models political actors' statements about their preferred policy beliefs as dynamic networks. An application to the policy debate on reforming the German public pension system illustrates how temporal decay for inertia, actor activity, belief popularity, and actor homophily can be estimated alongside the main coefficients. Convergence diagnostics and an illustration of bias correction relative to fixed parameters are provided, stabilisation using hyper-parameters and computational complexity are discussed, and the approach is extended to include both Breslow's and Efron's methods for breaking ties in the event sequence.

1. Introduction

Social systems often exhibit complex interdependent behaviour between units and are best modelled as networks (Ward et al., 2011). Network ties usually develop over time and in response to each other in a complex way. Relational event models (REMs) (Butts, 2008) are a popular way to model how ties form in a temporal sequence of network ties (Bianchi et al., 2023; Butts et al., 2023). REMs have been applied to animal behaviour (Tranmer et al., 2015), team work (Leenders et al., 2016; Pilny et al., 2016), inter-hospital patient transfers (Vu et al., 2017), radio communication (Butts, 2008), corporate board interlocks (Valeeva et al., 2020), legislative co-sponsorship (Brandenberger, 2018), parliamentary veto cascades (Malang et al., 2019), rivalries between street gangs (Gravel et al.,

* Corresponding author.
E-mail addresses: philip.leifeld@manchester.ac.uk (P. Leifeld), jw19203@essex.ac.uk (J.S.T. Wong).

[1] PL would like to acknowledge the assistance given by Research IT and the use of the Computational Shared Facility at The University of Manchester.

[2] JSTW acknowledges the use of the High Performance Computing Facility (Ceres) and its associated support services at the University of Essex in the completion of this work.

https://doi.org/10.1016/j.csda.2026.108428
Received 11 February 2025; Received in revised form 17 March 2026; Accepted 27 May 2026
Available online 28 May 2026
0167-9473/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2023), Wikipedia edits (Lerner and Lomi, 2020), open-source software collaboration (Quintane et al., 2014), international conflict (Brandes et al., 2009; Lerner et al., 2013a), and other complex social systems. Most recently, REMs have been extended by adding capabilities for polyadic hyper-events (Lerner and Lomi, 2023), multi-level event networks (Vieira et al., 2024), actor-based objective functions (Stadtfeld and Block, 2017), goodness-of-fit assessment (Brandenberger, 2019), change point detection (Kamalabad et al., 2023), temporal heterogeneity (Meijerink-Bosman et al., 2022; Mulder and Leenders, 2019), weighted and signed ties (Brandes et al., 2009; Lerner, 2016), latent space components (Artico and Wit, 2023), and large datasets (Lerner and Lomi, 2020; Overgoor et al., 2020).

In REMs, past interactions in the network contribute to the tie formation hazard through endogenously formed statistics over the past event sequence (Butts, 2008, Eq. 4). An early modification by (Brandes et al., 2009, p. 202) of the original model by Butts (Butts, 2008) scaled the past ties entering these network statistics by an exponential temporal decay factor with a fixed half-life parameter. With this decay function, recent ties have a stronger impact on current tie hazard than ties that occurred long ago. Many applications have adopted this temporal decay (Brandenberger, 2018; Leenders et al., 2016; Lerner et al., 2013a; Lerner and Lomi, 2023; Malang et al., 2019; Pilny et al., 2016; Valeeva et al., 2020).

These studies assume a fixed half-life parameter and therefore a specific functional form of the temporal decay. Fixing the half-life parameter, however, may bias inferences because the coefficient magnitude and uncertainty as well as covariance with other model terms may be impacted by the choice of the half-life parameter. Choosing the half-life parameter based on theory alone is usually not possible because theories are rarely explicit about the temporal horizon of network effects. In the words of Arena et al., "endogenous statistics are no more sufficient for the estimation of the corresponding vector of effects $\boldsymbol{\beta}$, because their value changes depending on the value of the memory parameter" (Arena et al., 2023, p. 273). Recent work has addressed this issue and suggested model selection for different half-life parameters using Bayesian model averaging and Bayes factors (Arena et al., 2022, 2023).

We build on this important work by incorporating half-life parameters into the model as Bayesian hyper-parameters and by estimating their posterior distributions along with the posterior distributions of the coefficients for the network statistics in a fully Bayesian estimation framework. Doing so allows us to treat temporal decay as an estimation problem, rather than a post-estimation model selection problem, by including one or more half-life parameters directly in the likelihood function and applying Markov chain Monte Carlo (MCMC) methods for estimation. Treating half-life parameters as quantities to be estimated in a Bayesian framework offers several advantages. It minimises bias, allows us to assess the temporality of network effects for theoretical gain, produces the posterior distribution of the temporal half-life of network effects, including the uncertainty around the mean or median half-life estimate, and makes post-estimation procedures, other than model summary, goodness-of-fit assessment, and model selection, obsolete.

The approach is demonstrated using discourse network analysis, the application of network science to the study of policy debates (Leifeld, 2013, 2017, 2020). Discourse network analysis has been applied as an exploratory or descriptive tool in more than 300 publications (Leifeld, 2024) and would benefit particularly from methods permitting insights into the temporal horizon of the unfolding network (Leifeld, 2017, 2020).

Discourse network analysis models policy debates as event sequences in which political actors, such as interest groups, legislators, and government agencies, make positive or negative public statements about normative policy beliefs. The event sequences are characterised by network dependence, whereby tie events display effects such as positive and negative reciprocity between actors, actor type homophily, and preferential attachment to policy beliefs (Brandenberger, 2019; Leifeld and Brandenberger, 2019). As journalists introduce uncertainty around the exact timing of political statements and binning into observable days, policy debates are best modelled in ordinal time, and we consequently focus on ordinal-time REMs in our exposition. The real-data example below re-analyses discourse network data on the German pension policy debate (Leifeld, 2013) using fully Bayesian REMs to shed light on the time horizon with which different network mechanisms operate.

2. REMs with temporal decay

The ordinal REM is both an event history model and a discrete choice model. We will first describe it as an event history model and then as a choice model to illustrate this point.

In survival models, the survivor function S expresses the probability that an event has not been experienced yet at time t (Cox, 1972, p. 187):

$$S(t) = P(T > t).$$

A parametric survival function, such as the exponential or Weibull function, can be used in lieu of the empirical survivor function.

The hazard is the risk to experience an event in a given time interval, given that it has not happened yet. It is often expressed as a rate, the "event rate" per time increment (Cox, 1972, Eq. (1)):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}.$$

The baseline hazard in a proportional hazard model, $h_0(t)$, denotes the probability that an event occurs when all covariates are zero. It is scaled upward ($\mathbf{x}_i^\top \boldsymbol{\beta} > 0$) or downward ($\mathbf{x}_i^\top \boldsymbol{\beta} < 0$) by the exponentiated linear predictor term to form the hazard of event i to occur at time t (Cox, 1972, Eq. (9)):

$$h_i(t) = h_0(t) \exp(\mathbf{x}_i^\top \boldsymbol{\beta}). \tag{1}$$

Cox proposed a version of the proportional hazard model for time-ordered data, where the probability of a particular event i to occur (Y_i) at time t_j is the hazard of the event over the hazards of all events still at risk of occurring, known as the "risk set" R (Cox, 1972, Eq. (12)):

$$P(Y_i = 1 \mid \mathbf{x}_i, \boldsymbol{\beta}) = \frac{h_i(t_j)}{\sum_{l \in R_j} h_l(t_j)} = \frac{h_0(t_j) \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}{\sum_{l \in R_j} h_0(t_j) \exp\left(\mathbf{x}_l^\top \boldsymbol{\beta}\right)} = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{\sum_{l \in R_j} \exp\left(\mathbf{x}_l^\top \boldsymbol{\beta}\right)}.$$

It becomes apparent that the constant baseline hazard is cancelled out of the fraction, leaving the exponentiated linear predictor term for the focal event over sum of the exponentiated linear predictor terms for the risk set, with exp ensuring positive scaling around 1 and preventing division by zero.

In a variant of this model, the risk set R at all times t_j can be specified to contain all possible events, rather than those that have not occurred yet. Doing so is equivalent to stratifying on the time variable and is the time-ordered cousin of continuous-time, parametric proportional hazard models with piece-wise constant hazards (Butts, 2008). We call this variant the "stratified Cox model" to indicate the stratification on the time variable.

In the (stratified) Cox model, the joint (partial) likelihood for all observed events is the product over the events in the sequence, with the (dynamic) risk set propped up with null events that could have occurred and $\mathbf{x}_i$ containing event-specific covariates:

$$\mathcal{L}(\boldsymbol{\beta} \mid \mathbf{X}) = \prod_{i=1}^{n} \left[\frac{\exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)}{\sum_{l \in R_i} \exp\left(\mathbf{x}_l^\top \boldsymbol{\beta}\right)} \right]. \tag{2}$$

The corresponding log likelihood is

$$\log \mathcal{L}(\boldsymbol{\beta} \mid \mathbf{X}) = \sum_{i=1}^{n} \left[\mathbf{x}_i^\top \boldsymbol{\beta} - \log\left(\sum_{l \in R_i} \exp\left(\mathbf{x}_l^\top \boldsymbol{\beta}\right) \right) \right].$$

With this discrete, ordered model, the risk set can be dynamically specified to permit repeated events and exclude forbidden events, and time points without events drop out of the equation because $\exp(0) = 1$.

The stratified Cox model in Eq. (2) is at the same time a discrete choice model, where the risk set R contains the possible alternatives (the "choice set"), $\mathbf{x}_i$ contains covariate values for alternative i, the alternatives are the different events (actual and null events) eligible to be selected to occur, and the chooser is constant. In the econometrics literature, this model is known as the conditional logit model and was proposed by McFadden (McFadden, 1973).

REMs replace $\mathbf{x}_i$ in the hazard (Eq. (1)) by functions of the past event sequence of the form $u(e_t, E_{t-1}, \mathbf{X})$, where $E = \{e_1, \ldots, e_n\}$ is the set of all events, also called event sequence, and $E_{t-1} = \{e_1, \ldots, e_{t-1}\}$ is the set of past events, with special case $E_0 = \emptyset$. Each event $e = (s_e, r_e, a_e, t_e)$ comprises event time t_e, a sender s_e, a receiver r_e, and a relationship type or weight a_e (cf. original exposition (Butts, 2008, Eq. 4)). $\mathbf{X} = (\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_n)^\top$ contains exogenous covariates and can, but need not, be processed by any particular u function. The sufficient statistics can represent exogenous covariates, network statistics, or combinations of both. Examples of network statistics are inertia, recency, activity, popularity, and other functions of the past event sequence acting as change statistics for the current tie event hazard (Butts, 2008). This replacement of covariates by network statistics formed over the past event sequence yields a reformulated likelihood

$$\mathcal{L}(\boldsymbol{\beta} \mid E, \mathbf{X}) = \prod_{t=1}^{n} \left[\frac{\exp\left(\mathbf{u}(e_t, E_{t-1}, \mathbf{X})^\top \boldsymbol{\beta}\right)}{\sum_{e \in R_t} \exp\left(\mathbf{u}(e, E_{t-1}, \mathbf{X})^\top \boldsymbol{\beta}\right)} \right] \tag{3}$$

and log likelihood

$$\log \mathcal{L}(\boldsymbol{\beta} \mid E, \mathbf{X}) = \sum_{t=1}^{n} \left[\mathbf{u}(e_t, E_{t-1}, \mathbf{X})^\top \boldsymbol{\beta} - \log\left(\sum_{e \in R_t} \exp\left(\mathbf{u}(e, E_{t-1}, \mathbf{X})^\top \boldsymbol{\beta}\right) \right) \right],$$

where R_t contains all possible events e that could have occurred at time point t. Because the events are conditionally independent given their prior event sequence and covariates (Lerner et al., 2013b), the statistics $\mathbf{u}$ can be inserted into the stratified Cox model as covariates, hence conditional logit estimation is feasible.

The original exposition of the REM includes both the continuous-time version based on piece-wise constant hazards (Butts, 2008, p. 162 and Eq. 1 and 2) and the stratified Cox model for ordered events in Eq. (3) (Butts, 2008, p. 165 and Eq. 3).

In subsequent applications, this ordinal model has become popular, and we focus our attention on this version in the methodological contribution and application below, without loss of generality. Exponentiated estimated coefficients are interpreted as factor changes of the hazard of tie event occurrence relative to the tie events that could have occurred.

Brandes, Lerner and Snijders proposed a modification of the network statistics in Eq. (3) of the form $u(e_t, E_{t-1}, \mathbf{X}, T_{1/2})$, where $T_{1/2}$ is a constant half-life parameter determining exponential decay of past events (Brandes et al., 2009) of the form

$$w(t_e, t, T_{1/2}) = \exp\left\{ -(t - t_e)\left(\frac{\log(2)}{T_{1/2}} \right) \right\} \frac{\log(2)}{T_{1/2}}.$$

With this change, $w(t_e, t, T_{1/2})$ down-weights past tie events in the calculation of network statistics more the longer ago they occurred. Fig. 1 illustrates the functional form with given half-life constants.

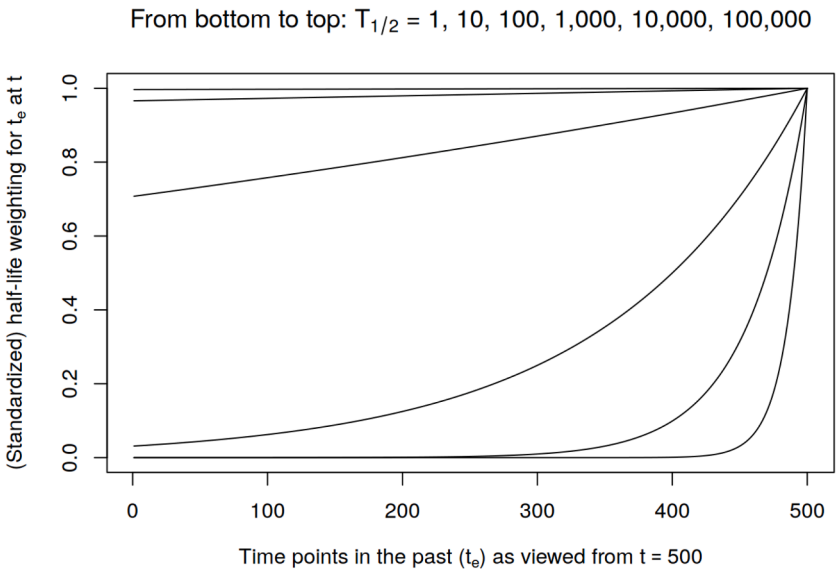


Fig. 1. Exponential temporal decay as a function of the half-life parameter $T_{1/2}$. The different curves illustrate the decay in importance of past statements the more time passes between a past event t_e and the current time point t, here $t = 500$ for illustration, under different given parametrisations of $T_{1/2}$.

An example of a network statistic with exponential temporal decay is inertia,

$$u_{\text{inertia}}(e, E_{t-1}, T_{1/2}) = \sum_{e^* \in E_{t-1}} [s_{e^*} = s_e][r_{e^*} = r_e][a_{e^*} = a_e] w(t_{e^*}, t_e, T_{1/2}), \tag{4}$$

which counts the number of past identical events e^* (on all elements except time), down-weighted by the difference in event time with a given half-life parameter. $[\cdot]$ denotes Iverson brackets, which return 1 if the condition in the brackets is true and 0 otherwise. The statistic captures the tendency for path dependence through repetition of sender–receiver–type combinations.

Another simple statistic is popularity,

$$u_{\text{popularity}}(e, E_{t-1}, T_{1/2}) = \sum_{e^* \in E_{t-1}} [r_{e^*} = r_e] w(t_{e^*}, t_e, T_{1/2}), \tag{5}$$

which counts the number of past receivers r_{e^*} that are identical to the receiver in the current candidate tie event r_e, down-weighted by the temporal distance between the current time t_e and the past event time t_{e^*}.

A similar statistic is activity,

$$u_{\text{activity}}(e, E_{t-1}, T_{1/2}) = \sum_{e^* \in E_{t-1}} [s_{e^*} = s_e] w(t_{e^*}, t_e, T_{1/2}), \tag{6}$$

which captures path dependence in senders.

A slightly more involved statistic is sender homophily (cf. Malang et al., 2019),

$$u_{\text{homophily}}(e, E_{t-1}, T_{1/2}, \eta) = \frac{\sum_{e^* \in E_{t-1}} [s_{e^*} \neq s_e][r_{e^*} = r_e][a_{e^*} = a_e] w(e^*, t_e, T_{1/2}) \eta(s_e, s_{e^*})}{\sum_{e^* \in E_{t-1}} [s_{e^*} \neq s_e][r_{e^*} = r_e][a_{e^*} = a_e] w(e^*, t_e, T_{1/2})} \tag{7}$$

with

$$\eta_{\text{categorical}}(s_e, s_{e^*}) = [v_{s_e} = v_{s_{e^*}}] \quad \text{or} \quad \eta_{\text{continuous}}(s_e, s_{e^*}) = \left| v_{s_e} - v_{s_{e^*}} \right|,$$

which attributes a greater hazard to tie events in which the sender shares categorical attributes v with past events ($\eta_{\text{categorical}}(s_e, s_{e^*})$) or has dissimilar values on continuous attributes v ($\eta_{\text{continuous}}(s_e, s_{e^*})$), down-weighted by temporal decay. Functions $\eta_{\text{categorical}}(s_e, s_{e^*})$ and $\eta_{\text{continuous}}(s_e, s_{e^*})$ are drop-in replacements for the general form of similarity or dissimilarity function $\eta(s_e, s_{e^*})$ in Eq. (7). For example, do political actors express a policy belief in a debate if the same belief was recently expressed by other actors who share a categorical attribute with the current actor, such as actor type or partisan leaning?

2.1. Tiebreaker

If there are no ties, the contribution to partial likelihood by an event at time point t is

$$\frac{\exp(\mathbf{u}(e_t, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta})}{\sum_{e \in R_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta})}. \tag{8}$$

We need to allow for the possibility of tied events. For example, an event sequence of normative ideological statements by political actors, which is considered as an example below, can be mostly time-ordered but will exhibit some simultaneity, which is an artefact of binning observations into days as newspapers are only published once per day. Ties in the ordering of events need to be carefully accounted for to mitigate estimation bias.

The idea of breaking ties involves adjusting the risk set according to the order of the tied events, which can include many possible sequences. We consider two common approaches from the literature. The first one is the approach by Breslow (1974), which is simply to ignore the issue and calculate the partial likelihood in the usual way, assuming the risk set is not affected by the ties. Suppose there are $m_t > 2$ event ties at time point t, and let M_t denote the set of tied events. The contribution to the partial likelihood under Breslow's approximation is given by

$$\frac{\prod_{e\in M_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta})}{\left\{\sum_{e\in R_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta})\right\}^{m_t}}.$$

The full partial log-likelihood is then

$$\begin{aligned}\log \mathcal{L}\left(\boldsymbol{\beta}, T_{1/2} \mid E, \mathbf{X}\right) = \sum_{t=1}^{n} \Bigg\{ &\sum_{e\in M_t} \mathbf{u}(e_t, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta} \\ &- m_t \log\left(\sum_{e\in R_t} \exp\left(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta}\right)\right)\Bigg\}.\end{aligned} \tag{9}$$

By construction, this corresponds to the sharing of the same risk set R_t across all tied events at time point t.

As noted in the literature, Breslow's approximation ignores any changes in risk set due to event ties, which will lead to a consistent underestimation of the partial likelihood. Hence, it might be better to consider Efron's approximation (Efron, 1977). Efron's approximation is more accurate by adjusting the partial likelihood with a correction term to account for the influence of tied events. Mathematically, the contribution to partial likelihood becomes

$$\frac{\prod_{e\in M_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta})}{\prod_{k=1}^{m_t}\left(\sum_{e\in R_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta}) - \underbrace{\frac{k-1}{m_t}\sum_{e\in M_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta})}_{\text{Efron's correction}}\right)},$$

where $k = 1, 2, \ldots, m_t$ is the k^{th} tied event.

The full partial log-likelihood under Efron's approximation is

$$\begin{aligned}\log \mathcal{L}\left(\boldsymbol{\beta}, T_{1/2} \mid E, \mathbf{X}\right) = \sum_{t=1}^{n} \Bigg\{ &\sum_{e\in M_t} \mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta} \\ &- \sum_{k=1}^{m_t} \log\Bigg(\sum_{e\in R_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta}) \\ &- \frac{k-1}{m_t}\sum_{e\in M_t} \exp(\mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2})^\top \boldsymbol{\beta})\Bigg)\Bigg\}.\end{aligned} \tag{10}$$

3. Bayesian estimation of exponential decay

Arena et al. noted that assuming a fixed half-life parameter $T_{1/2}$ leads to biased inference because the sufficient statistics are not sufficient (Arena et al., 2023, p. 273). I. e., the assumed constant half-life may be an accurate or inaccurate reflection of the true temporal decay, and the more it diverges from the true temporal decay, the greater the expected bias is in the estimated coefficient for the u statistic. Misspecification of $T_{1/2}$ can lead to biased inferences and may lead to false substantive conclusions about the data-generating process because the coefficient magnitude and uncertainty as well as covariance with other coefficients are partial functions of $T_{1/2}$.

Arena et al. proposed Bayesian model averaging over different decay parameterisations and the use of Bayes factors for model selection to overcome the reliance on assumed constant parameters (Arena et al., 2022, 2023). This is an important development but relies on post-estimation model selection steps. Here, we propose estimating the half-life parameters along with the main effects for network statistics in a fully Bayesian setting.

This development is akin to the concept of curved exponential family models, where the optimal decay of non-linear higher-order network statistics, such as geometrically weighted edge-wise shared partners, over additional third-party nodes is estimated from the network data along with the main effect to fix the problem of degeneracy in exponential random graph models (Hunter and Handcock, 2006).

3.1. Bayesian formulation

Bayesian inference is a methodology by which our prior belief for the joint parameter, $\theta = \{\beta, T_{1/2}\}$, is updated in light of the observed data to yield the joint posterior distribution, $\pi(\theta \mid E, \mathbf{X})$ (see, for example, Gelman et al. (1995)). Mathematically, this is achieved using Bayes' theorem, i. e.,

$$\pi(\theta \mid E, \mathbf{X}) \propto f(E \mid \theta, \mathbf{X}) \times \pi(\theta), \tag{11}$$

where $\pi(\theta)$ is the prior distribution quantifying our prior belief and $f(E \mid \theta, \mathbf{X}) = \mathcal{L}(\theta \mid E, \mathbf{X})$ is essentially the likelihood function reformulated to include the half-life parameter $T_{1/2}$, as given by either Eq. (9) or Eq. (10).

A crucial difference between frequentist and Bayesian paradigms lies in the subjective notion of probability. Fundamentally, the model parameter θ is treated as unknown but fixed under the existing frequentist framework, whereas in the Bayesian paradigm θ is treated as a random variable with a probability distribution characterising the uncertainty. The Bayesian approach then works by simultaneously estimating θ by computing the joint posterior distribution, producing results that encapsulate various sources of uncertainty. An advantage in our context is that $T_{1/2}$ can be formulated as an explicit parameter to be estimated through the Bayesian procedure alongside uncertainties in the form of a probabilistic distribution. This is pivotal because an uninformed specification of $T_{1/2}$ inherently leads to biased inference and over-confident intervals, as will be demonstrated numerically in Section 7.5.

3.2. Prior specification

Prior distributions can be specified to reflect our prior knowledge about the model parameters. But in most applications, we do not possess prior domain knowledge, in which case Bayesian methods allow us to learn about θ from the data. For example, one could consider arbitrarily diffused priors in the absence of prior information as in Daziano (2013). Alternatively, the class of conjugate priors proposed by Koop and Poirier (1993) for conditional logit models could be adopted. A conjugate prior is a special class of prior distributions that leads to posterior distributions of the same family of distributions. Conjugate priors possess desirable computational properties for posterior sampling. Conjugate priors have the advantage that it is easy to mathematically quantify the level of information contained in the priors relative to that in the data. Future research should investigate the use of diffuse and conjugate priors for REMs. We choose a different strategy here.

Our initial strategy is to specify (improper) flat priors for all parameters, i. e.,

$$\pi(\theta) = \pi(\beta) \times \pi(T_{1/2}) \propto 1, \tag{12}$$

such that results derived from our Bayesian analysis are directly comparable with those from the frequentist fixed-decay framework. This is in the spirit of data-driven inferences, where we allow the data to "speak for themselves". Specifying (improper) flat priors also serves as an indication if there is sufficient information in the data to estimate both β and $T_{1/2}$. We note that the use of improper flat priors is common and convenient and typically leads to proper posteriors, as pointed out by Gelfand and Sahu (1999).

After establishing that β and $T_{1/2}$ are estimable, it is also of interest to determine the influence of the prior specification on the results. We continue to impose $\pi(\beta) \propto 1$ but explore different choices of $\pi(T_{1/2})$. For instance, the flat prior $\pi(T_{1/2}) \propto 1$ assigns equal prior mass to arbitrarily large half-life values and thus places no a priori penalty on extremely slow decay, corresponding to an infinite-memory regime, which may be unrealistic in some applications. We therefore focus on priors of the form $\pi(T_{1/2}) \sim \text{Gamma}(a, b)$, where a and b are hyper-parameters to be specified as follows:

1. $a = 2$ and $b = 0.1$.
2. $a = 20$ and $b = 1$.
3. $a = 200$ and $b = 10$.
4. $a = 2000$ and $b = 100$.

All these gamma priors possess the same mean of 20 but different variances, quantifying varying degrees of uncertainty. The mean of 20 was chosen as our prior belief because it represents the fixed choice from a previous analysis using non-Bayesian REMs on the same data (Leifeld and Brandenberger, 2019), where the value was determined naïvely by comparing several model BIC values with seemingly plausible $T_{1/2}$ values.

Moving from 1 to 4, the prior variance decreases from 200 to 0.2, implying that we place increasingly more emphasis on the role of the prior distribution in the estimation process. The prior becomes more informative, and we are increasingly more confident of the prior knowledge. Informative priors can be particularly useful when there is insufficient information in the data to estimate $T_{1/2}$.

Conversely, the larger the variance, the more uncertain we are and thus the more the estimation is dominated by the data. In terms of the hyper-parameter b, this corresponds to values tending to 0. Indeed, the flat prior in Eq. (12) is the limiting case when $b \to 0$. In extreme cases, both ends of the spectrum will, respectively, produce results that tend to the frequentist estimation with fixed $T_{1/2} = 20$ (infinitesimally small variance) and Bayesian estimation using the flat prior in Eq. (12) (infinitesimally large variance).

The practice of hyper-parameter tuning in this context is the essence of Bayesian elicitation of subject-specific knowledge, where the proposed methodology allows users/experts to specify the mean value of $T_{1/2}$ (perceived as the "average response time") as well as the variance (quantifying how confident they are). This corresponds to the calibration of one's prior belief about $T_{1/2}$ against the observed data in the form of hyper-parameters, where a and b are tuned to reflect personal degree of uncertainty. By contrast, there is no such option in the frequentist fixed-decay framework, which requires deterministic specification of the value of $T_{1/2}$ and inherently leads to biased inferences.

This exercise can also be viewed as a sensitivity analysis with respect to prior uncertainty/variance, where we test how different the results are relative to the initial result based on improper flat priors (see Section 7.6).

We note that gamma priors are not the only viable choice in this setting; for example, log-normal distributions could serve a similar role. However, gamma priors offer several practical advantages: they define a smooth and proper density on the positive real line and admit highly interpretable hyper-parameters (expressed directly in terms of moments), which facilitates the elicitation of prior beliefs. For these reasons, we focus on gamma priors in this paper.

4. Computational strategy

4.1. Pre-computation

Bayesian estimation in our context poses a computation challenge. Repetitive evaluations of the temporal statistics, $\mathbf{u}(e_t, E_{t-1}, \mathbf{X}, T_{1/2})$, at various values of the half-life parameter are expensive because one needs to iterate through all instances of true events and null events. To aid our computations, we propose a re-parametrisation strategy, without which a Bayesian implementation of the REM with temporal decay may be infeasible in empirical contexts with medium-sized or large datasets.

At time t_e, we first define the temporally ordered subset of eligible events $\tilde{E} \subseteq E_{t-1}$ such that only those events form the subset that are required by the conditions in the respective statistic. For example, for inertia, we define

$$\tilde{E}_t^{\text{inertia}} = \{e^* \mid e^* \in E_{t-1} \wedge [s_{e^*} = s_e] \wedge [r_{e^*} = r_e] \wedge [a_{e^*} = a_e]\} \subseteq E_{t-1} \tag{13}$$

such that the elements are ordered by $t_{e_1^*} < t_{e_2^*} < \ldots < t_{e^*_{|\tilde{E}_t^{\text{inertia}}|}}$. Next, we define a vector $\boldsymbol{\psi}_t \in \mathbb{R}^{|\tilde{E}_t|}$ containing elements $\psi_i = t_e - t_{e_i^*}$, $\forall i \in \{1, 2, \ldots, |\tilde{E}_t|\}$, for example $\boldsymbol{\psi}_t^{\text{inertia}}$. If the conditions in Eq. (13) are not all satisfied, then $\tilde{E}_t = \emptyset \Rightarrow \boldsymbol{\psi}_t \in \mathbb{R}^0$. $\boldsymbol{\psi}_t$ is usually small because the conditions are not all satisfied for the majority of events. This is vital for computation time and memory footprint as we can pre-compute and store $\boldsymbol{\psi}_t$ for every event e along the event sequence E where $|\tilde{E}_t| > 0$ for re-use at each iteration of the MCMC chain when the respective statistic needs to be evaluated after $T_{1/2}$ has been updated.

The respective $u(e, E_{t-1}, T_{1/2})$ can then be calculated for all e at a given value of $T_{1/2}$ as

$$u(e, E_{t-1}, T_{1/2}) = \begin{cases} \exp\left(-\boldsymbol{\psi}_t^\top \times \left(\frac{\log(2)}{T_{1/2}}\right)\right) \times \mathbf{1}_{|\tilde{E}_t|} \times \frac{\log(2)}{T_{1/2}}, & \text{if } |\tilde{E}_t| > 0 \\ 0, & \text{if } |\tilde{E}_t| = 0 \end{cases},$$

where $\mathbf{1}_{|\tilde{E}_t|}$ is a vector of ones of the same length as $\tilde{E}_t$ and $\boldsymbol{\psi}_t$.

The advantage of re-parametrising using $\boldsymbol{\psi}_t$ is that these vectors can be pre-computed and stored before undertaking MCMC computations. Doing so does not occupy substantial amounts of memory space because the events in the past sequence that meet the conditions are relatively "sparse", as explained above. Thus, repetitive evaluations of the temporal statistics $u(e, E_{t-1}, T_{1/2})$ at various values of $T_{1/2}$ (within the MCMC sampler) becomes computationally feasible without having to scan through the whole iteration of true and null events. This strategy allows us to perform exact Bayesian computation without the need for approximate inference (see Section 8).

Similar procedures can be followed for computing other temporal statistics, for example $u_{\text{activity}}(e, E_{t-1}, T_{1/2}), u_{\text{popularity}}(e, E_{t-1}, T_{1/2})$, or $u_{\text{homophily}}(e, E_{t-1}, T_{1/2})$. Specifically for the homophily statistic, we require two such quantities, one each for the numerator and denominator of Eq. (7), respectively given by

$$\tilde{E}_t^{\text{homophily},N} = \{e^* \mid e^* \in E_{t-1} \wedge [s_{e^*} \neq s_e] \wedge [r_{e^*} = r_e] \wedge [a_{e^*} = a_e] \wedge [v_{s_{e^*}} = v_{s_e}]\}$$

with $\boldsymbol{\psi}_t^{\text{homophily},N} \in \mathbb{R}^{|\tilde{E}_t^{\text{homophily},N}|}$ for the numerator and

$$\tilde{E}_t^{\text{homophily},D} = \{e^* \mid e^* \in E_{t-1} \wedge [s_{e^*} \neq s_e] \wedge [r_{e^*} = r_e] \wedge [a_{e^*} = a_e]\}$$

with $\boldsymbol{\psi}_t^{\text{homophily},D} \in \mathbb{R}^{|\tilde{E}_t^{\text{homophily},D}|}$ for the denominator. It is easy to see that

$$\tilde{E}_t^{\text{homophily},N} \subseteq \tilde{E}_t^{\text{homophily},D}.$$

Then,

$$u_{\text{homophily}}(e, E_{t-1}, T_{1/2}, \eta)$$

$$= \begin{cases} \frac{\exp\left(-\left(\boldsymbol{\psi}^{\text{homophily},N}\right)^\top \times \left(\frac{\log(2)}{T_{1/2}}\right)\right) \times \mathbf{1}_{\left|\tilde{E}_t^{\text{homophily},N}\right|} \times \frac{\log(2)}{T_{1/2}}}{\exp\left(-\left(\boldsymbol{\psi}^{\text{homophily},D}\right)^\top \times \left(\frac{\log(2)}{T_{1/2}}\right)\right) \times \mathbf{1}_{\left|\tilde{E}_t^{\text{homophily},D}\right|} \times \frac{\log(2)}{T_{1/2}}}, & \text{if } \left|\tilde{E}_t^{\text{homophily},N}\right| > 0 \\ 0, & \text{if } \left|\tilde{E}_t^{\text{homophily},N}\right| = 0 \end{cases}$$

$$= \begin{cases} \frac{\exp\left(-\left(\boldsymbol{\psi}^{\text{homophily},N}\right)^\top \times \left(\frac{\log(2)}{T_{1/2}}\right)\right) \times \mathbf{1}_{\left|\tilde{E}_t^{\text{homophily},N}\right|}}{\exp\left(-\left(\boldsymbol{\psi}^{\text{homophily},D}\right)^\top \times \left(\frac{\log(2)}{T_{1/2}}\right)\right) \times \mathbf{1}_{\left|\tilde{E}_t^{\text{homophily},D}\right|}}, & \text{if } \left|\tilde{E}_t^{\text{homophily},N}\right| > 0 \\ 0, & \text{if } \left|\tilde{E}_t^{\text{homophily},N}\right| = 0 \end{cases}$$

As an illustration, without pre-computation, the time required to run one full iteration of the MCMC sampling algorithm in the empirical illustration below is around 5266 s on a contemporary computing core. With pre-computation, an iteration requires only around 130 s, approximately a 40 times speed-up.

4.2. MCMC scheme

Since the priors for $\theta = \{\beta, T_{1/2}\}$ are not conjugate, we propose to apply the variable-at-a-time random walk Metropolis-Hastings (MH) algorithm as described in O'Hagan and Forster (2004). In addition, we also suggest updating β and $T_{1/2}$ separately in blocks, rather than all in one block, to improve the efficiency of the MCMC scheme (see Roberts and Sahu, 1997).

First, we note that the conditional posterior distributions of both β and $T_{1/2}$ are proportional to the likelihood function multiplied by their prior distributions, i. e.,

$$\pi(\beta \mid T_{1/2}, E, \mathbf{X}) \propto \mathcal{L}(\theta \mid E, \mathbf{X}) \times \pi(\beta), \tag{14}$$

and

$$\pi(T_{1/2} \mid \beta, E, \mathbf{X}) \propto \mathcal{L}(\theta \mid E, \mathbf{X}) \times \pi(T_{1/2}). \tag{15}$$

MCMC updating schemes are then applied iteratively to sample from both Eqs. (14) and (15), which can be combined to form samples of the joint posterior distribution, $\pi(\theta \mid E, \mathbf{X})$.

4.2.1. Random walk MH step for β

To sample from Eq. (14) using a random walk MH updating scheme, we first propose a candidate randomly based on the current iteration. Mathematically, a candidate β^* is proposed at the ith iteration as

$$\beta^* \sim N(\beta^{(i)}, \Sigma_\beta),$$

where $\beta^{(i)}$ is the current value and Σ_β is the proposal variance matrix. The proposal β^* is then accepted according to the acceptance ratio, which can be derived as the ratio of conditional posterior distributions, given as

$$\alpha(\beta^* \mid \beta^{(i)}, T_{1/2}^{(i)}) = \min\left\{1, \frac{\mathcal{L}\left(\beta^*, T_{1/2}^{(i)} \mid E, \mathbf{X}\right) \times \pi(\beta^*)}{\mathcal{L}\left(\beta^{(i)}, T_{1/2}^{(i)} \mid E, \mathbf{X}\right) \times \pi(\beta^{(i)})}\right\}, \tag{16}$$

where $\beta^{(i)}, T_{1/2}^{(i)}$ are values at the current iteration.

Although the choice of Σ_β is arbitrary, we recommend following the suggestion by Roberts and Rosenthal (2001) to optimise posterior explorations. This corresponds to setting Σ_β to be proportional to the negative inverse of the Hessian matrix of $\mathcal{L}\left(\beta \mid E, \mathbf{X}, T_{1/2}\right)$, which is easy to obtain in practice, e. g., from a standard optimisation routine in `R`, with a naïve initial guess of $T_{1/2}$. The constant of proportionality can be obtained from pilot runs using the numerical search as described in Wong et al. (2018). For example, the `clogit` function in the `R` package `survival` (Therneau, 2024) produces a useful approximation. The resulting MCMC chains have been found to converge more quickly due to a higher effective sample size per unit time.

4.2.2. Random walk MH step for $T_{1/2}$

In a similar way, the half-life parameter $T_{1/2}$ is sampled univariately from Eq. (15) using a random walk MH scheme. To avoid proposing negative values, we perform the updating scheme on a log scale. We propose

$$\log(T_{1/2}^*) \sim N(\log(T_{1/2}^{(i)}), \sigma_T^2),$$

where $T_{1/2}^{(i)}$ is the current iterate, σ_T^2 is the proposal variance chosen such that the resulting acceptance rate is between 0.15 and 0.45 Roberts and Rosenthal (2001).

The acceptance ratio for $T_{1/2}^*$ is

$$\alpha(T_{1/2}^* \mid \beta^{(i)}, T_{1/2}^{(i)}) = \min\left\{1, \frac{\mathcal{L}\left(\beta^{(i)}, T_{1/2}^* \mid E, \mathbf{X}\right) \times \pi(T_{1/2}^*)}{\mathcal{L}\left(\beta^{(i)}, T_{1/2}^{(i)} \mid E, \mathbf{X}\right) \times \pi(T_{1/2}^{(i)})}\right\}. \tag{17}$$

When (improper) flat priors are used, then the acceptance ratios in Eqs. (16) and (17) simply reduce to ratios of likelihood functions.

4.2.3. MCMC algorithm

A generic MCMC scheme then proceeds as follows:

1. For a given dataset, pre-compute the ψ_t vectors for every temporal statistic, such as inertia, activity, popularity, and homophily, and for every e using the pre-computation procedure described in Section 4.1. The ψ_t vectors have varying lengths and can be stored in lists per statistic.
2. Set $i = 0$ and get initial values $\theta^{(0)} = \{\beta^{(0)}, T_{1/2}^{(0)}\}$ according to Section 4.2.4.

Table 1
The PSRF for each parameter under both Breslow's and Efron's likelihoods. See Section 7.2 for an explanation of the model specification and data.

	Breslow	Upper C.I.	Efron	Upper C.I.
β_{gov}	1.00363	1.00525	1.00309	1.00447
β_{sci}	1.00266	1.00387	1.00215	1.00313
$\beta_{\text{agreement}}$	1.00240	1.00349	1.00247	1.00359
β_{inertia}	1.01449	1.02076	1.01167	1.01670
β_{activity}	1.00736	1.01055	1.00656	1.00943
$\beta_{\text{popularity}}$	1.01552	1.02221	1.01446	1.02069
$\beta_{\text{homophily}}$	1.00379	1.00549	1.00303	1.00439
$T_{1/2}$	1.03982	1.05689	1.02969	1.04243

Table 2
Effective sample sizes (rounded to nearest integers) of the 10000 posterior MCMC samples for all parameters under Breslow's and Efron's approximations.

	Breslow	Efron
β_{gov}	10408	9685
β_{sci}	10033	8938
$\beta_{\text{agreement}}$	7316	6322
β_{inertia}	2208	2731
β_{activity}	4321	4747
$\beta_{\text{popularity}}$	2239	2866
$\beta_{\text{homophily}}$	7926	8155
$T_{1/2}$	1091	1195

3. Propose

$$\boldsymbol{\beta}^* \sim N(\boldsymbol{\beta}^{(i)}, \Sigma_\beta).$$

4. Calculate the acceptance ratio $\alpha(\boldsymbol{\beta}^* \mid \boldsymbol{\beta}^{(i)}, T_{1/2}^{(i)})$ as given in Eq. (16).
5. Generate a uniform value $u_1 \sim U(0,1)$. Then set

$$\boldsymbol{\beta}^{(i+1)} = \begin{cases} \boldsymbol{\beta}^* & \text{if } u_1 \leq \alpha(\boldsymbol{\beta}^* \mid \boldsymbol{\beta}^{(i)}, T_{1/2}^{(i)}) \\ \boldsymbol{\beta}^{(i)} & \text{otherwise.} \end{cases}$$

6. Propose

$$\log(T_{1/2}^*) \sim N(\log(T_{1/2}^{(i)}), \sigma_T^2).$$

7. Compute the "proposed" temporal network statistics, e. g., $u_{\text{inertia}}(e, E_{t-1}, T_{1/2}^*), u_{\text{activity}}(e, E_{t-1}, T_{1/2}^*), u_{\text{popularity}}(e, E_{t-1}, T_{1/2}^*)$, and/or $u_{\text{homophily}}(e, E_{t-1}, T_{1/2}^*, \eta)$, using the procedure described in Section 4.1.
8. Calculate the acceptance ratio $\alpha(T_{1/2}^* \mid \boldsymbol{\beta}^{(i+1)}, T_{1/2}^{(i)})$ as given in Eq. (17).
9. Generate another uniform value $u_2 \sim U(0,1)$. Then,

$$\begin{cases} \text{set } T_{1/2}^{(i+1)} = T_{1/2}^* \text{ and accept} \\ \quad \mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2}) = \mathbf{u}(e, E_{t-1}, \mathbf{X}, T_{1/2}^*) & \text{if } u_2 < \alpha(T_{1/2}^* \mid \boldsymbol{\beta}^{(i+1)}, T_{1/2}^{(i)}). \\ \text{set } T_{1/2}^{(i+1)} = T_{1/2}^{(i)} & \text{otherwise.} \end{cases}$$

10. Set $i := i + 1$ then repeat Steps 3–10 until a desired number of posterior samples has been obtained.

4.2.4. *Initialisation, burn-in and thinning*

To assess the robustness of our MCMC sampling scheme with respect to initial values and ensure that we arrive at the same target posterior distribution irrespective of the initialisation used, we conducted the following robustness check. For initialisation of the MCMC chains, we followed the procedure by Gelman and Rubin (1992) in generating overdispersed initial values but adopted a simplified version for ease of implementation. Specifically, for $\boldsymbol{\beta}$, we generated random samples from a uniform distribution in the interval $[-5, 5]$, i. e., $\boldsymbol{\beta}^{(0)} \sim U(-5, 5)$, while $T_{1/2}$ was initialised randomly from the set $T_{1/2}^{(0)} \in \{20, 40, \ldots, 200\}$. We ran the MCMC sampling scheme described above in 50 parallel chains of 5100 iterations each. A burn-in phase of 1100 samples was applied followed by a thinning by 20 to mitigate effects of both initialisation and auto-correlation, resulting in 200 posterior samples per chain.

The convergence of these chains was assessed based on their within-chain and between-chain variances using the potential scale reduction factor (PSRF) developed by Gelman and Rubin (1992), where a value of close to 1 indicates convergence. Based on Table 1,

it can be concluded that convergence was attained. After checking that convergence was attained for each batch, the batches were combined to form a total of 10000 samples from the joint posterior distribution, $\pi(\theta \mid E, \mathbf{X})$. The corresponding effective sample sizes of the posterior samples are presented in Table 2.

5. Simulation results

Simulation results with plausible parameter configurations demonstrate that the model with a fixed half-life parameter $T_{1/2}$ indeed incurs bias in the β estimates when $T_{1/2}$ is misspecified and that the hyper-parameter approach outlined here can recover the true population decay.

For each experimental condition, we simulated 100 bipartite event sequences with a length of 3000 events with 20 sender nodes, 10 receiver nodes, parameters $\beta_{\text{inertia}} = 6.0, \beta_{\text{activity}} = 5.0, \beta_{\text{popularity}} = 4$, and a given half-life from the set $T_{1/2} = \{25, 50, 75, 100\}$. The sequences were simulated forward in time, evaluating the probabilities of all possible events (i. e., the risk set) using Eq. (8) at each new time step and sampling one event probabilistically given the vector of event probabilities. The first 1000 of the 3000 events were discarded as simulation burn-in.

Fig. 2 shows the frequentist maximum-likelihood estimates for inertia, activity, and popularity (boxplots) relative to the true simulated parameters (blue, horizontal lines) when the estimation erroneously assumes fixed $T_{1/2} = 20$ (top row) or $T_{1/2} = 120$ (bottom row). As the difference between simulated half-life and the fixed estimation parameter for half-life increases, the magnitude of the bias also increases, and vice versa.

However, the half-life parameter can be recovered from simulated relational event sequences using the Bayesian hyper-parameter approach. Fig. 3 shows the distributions of the recovered point estimates using the Bayesian approach as boxplots and the simulated parameters as blue lines. The original parameters are located within the interquartile range for all parameters. The estimation picks up the varied half-life parameters. We note that some inefficiency was introduced into the estimation by not making the previous history before the burn-in cut-off available to the network statistics, increasing noise in the results; however, the results are sufficiently precise to see how the estimation picks up the modified decay setting. Similar results were obtained for unipartite (one-mode) relational event sequences, but we report only the bipartite case here to mirror the empirical setup presented below.

6. Computational complexity

We assessed the computational time of the estimation as a function of the number of events n. Let b index the network statistics included in the model. For statistic b, let a_b denote the number of past-event indices that must be considered when evaluating the statistic, and let c_b denote the constant time required to evaluate 1 admissible configuration of current and past event(s).

At event i, only the $i-1$ preceding events are available in E_{t-1}. If statistic b requires a_b past-event indices, then the number of combinations that must be evaluated at event i is proportional to $(i-1)^{a_b}$. Hence, the cost of recomputing statistic b once across all events at a single MCMC iteration is

$$\sum_{i=2}^{n} (i-1)^{a_b} c_b.$$

Summing over all statistics yields the per-iteration cost

$$\sum_b \sum_{i=2}^{n} (i-1)^{a_b} c_b.$$

Let p_b be the statistic-specific pre-computation time. If the statistic evaluation is repeated for g MCMC iterations and pre-computation requires $p_b n$ operations for each statistic, the total computational cost is

$$\sum_b p_b n + g \sum_b \sum_{i=2}^{n} (i-1)^{a_b} c_b.$$

To study the asymptotic growth with respect to n, we treat the number of statistics, the number of MCMC iterations, and the constants c_b and p_b as fixed. Using the standard power-sum result that

$$\sum_{r=1}^{n-1} r^a = \Theta(n^{a+1})$$

(as a function of n), it follows that the computational complexity of recomputing statistic b is $\Theta(n^{a_b+1})$, and therefore the overall complexity, dominated by the statistic with largest a_b, is

$$O(n^{a+1})$$

per iteration or

$$O(gn^{a+1})$$

for all iterations, where $a = \max_b a_b$.

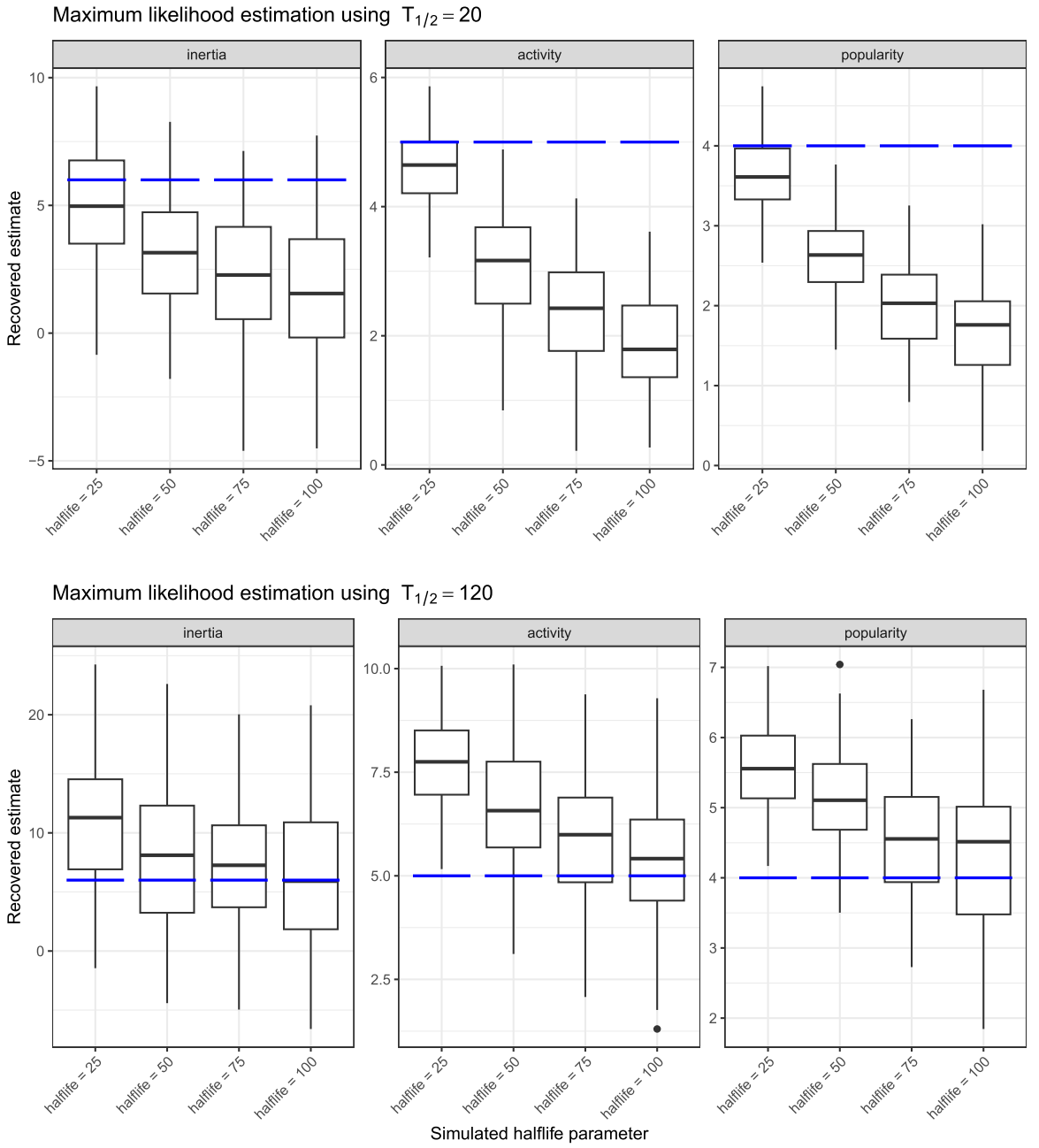


Fig. 2. Distributions of estimated parameters based on simulated sequences with constant inertia, activity, and popularity and varying half-life. Maximum likelihood estimation using half-life of $T_{1/2} = 20$ (top row) and $T_{1/2} = 120$ (bottom row). The β coefficients for inertia, activity, and popularity are biased.

In the network statistics considered in this paper, each statistic depends on a single past event index ($a_b = 1$), implying quadratic growth of the statistic recomputation cost. Since the pre-computation term $\sum_b p_b n$ is linear in n, it is asymptotically dominated by the quadratic term and does not affect the overall order of growth. Hence, the per-iteration complexity is $O(n^2)$.

The exact computation count corresponds to triangular sums of the form

$$\sum_{i=2}^{n} (i-1),$$

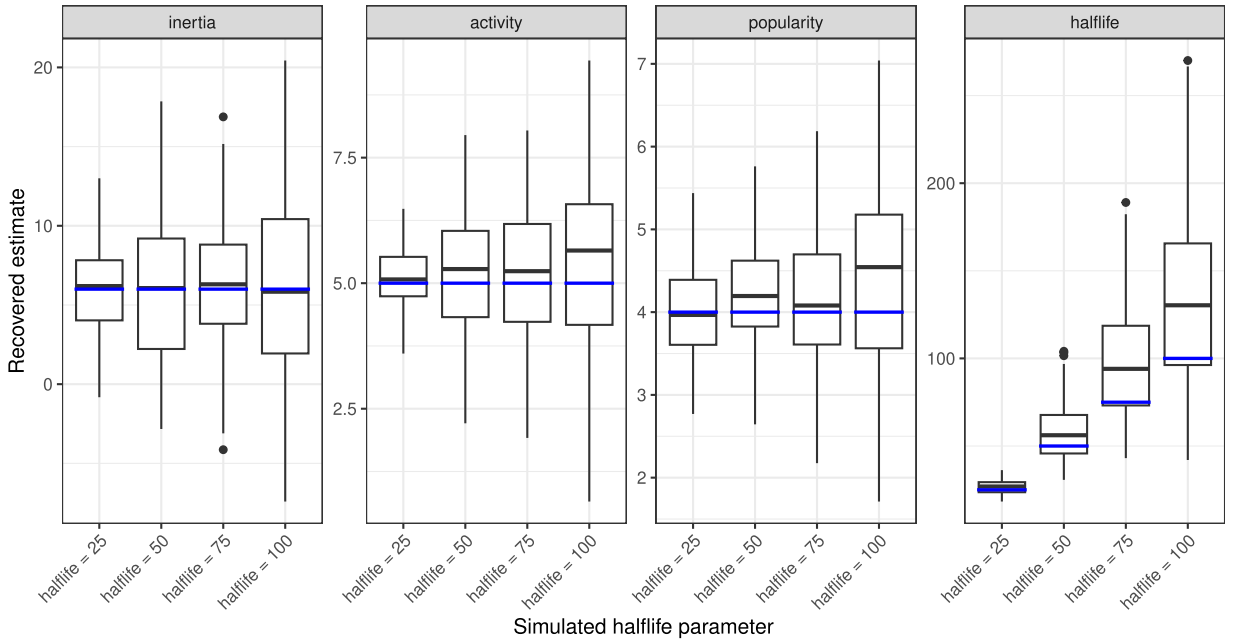


Fig. 3. Distributions of estimated parameters based on simulated sequences with constant inertia, activity, and popularity and different half-life. The estimation is sensitive to differences in half-life in the simulated sequences.

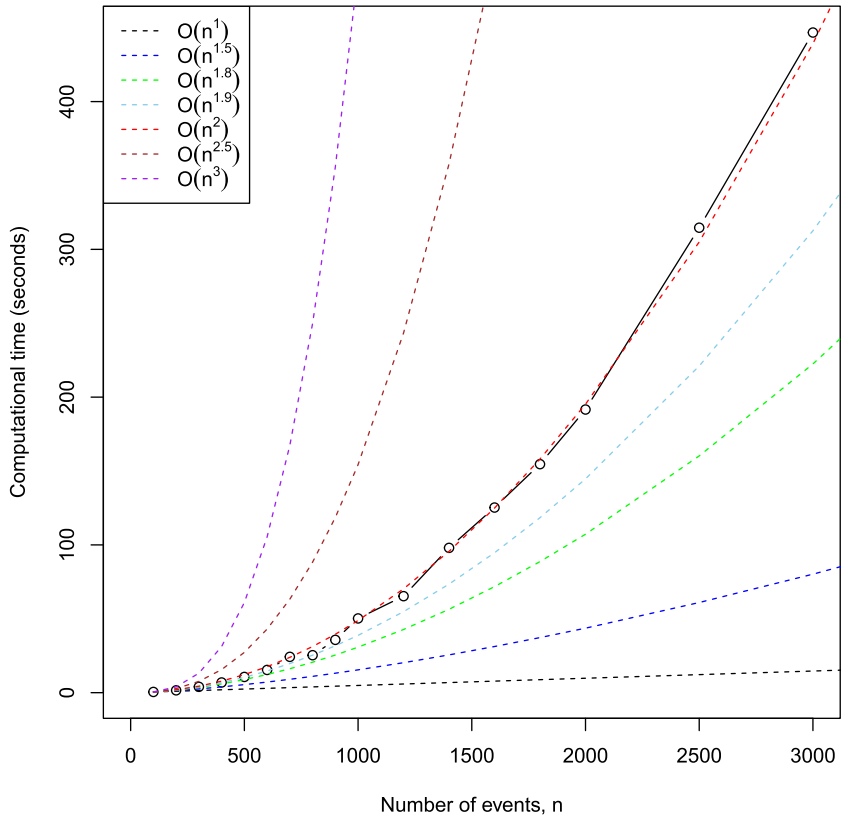


Fig. 4. Computational complexity, based on the simulation setup. Computational time is quadratic in the number of events.

which evaluates to

$$\frac{n(n-1)}{2}.$$

Thus, the exact number of operations grows proportionally to

$$\frac{1}{2}n^2 - \frac{1}{2}n.$$

Although this is smaller than the rectangular upper bound n^2 by a constant factor, both expressions have quadratic asymptotic growth. Consequently, the triangular structure of the loops reduces the constant factor but does not change the order of growth. This explains why the empirical timing results observed in the simulation as shown in Fig. 4 closely follow $O(n^2)$ behaviour.

For the empirical assessment in Fig. 4, we used simulated sequences with the same parameters as in the preceding section, 110 MCMC iterations including 10 burn-in iterations without thinning, and different sequence lengths of up to 3000 events.

7. Empirical case study

Below, we will illustrate the approach using a real-data case study on discourse networks in public policy and evaluate model convergence, substantive results, the bias arising from the difference between frequentist analysis with assumed half-life and the Bayesian approach with estimated half-life, and the sensitivity to priors.

7.1. Discourse networks as relational event sequences

Public policy is a subfield of political science that seeks to understand how policies are made and why certain policies are adopted rather than their possible alternatives. To explain policy outcomes in a subsystem, such as climate or pension policy, one needs to consider the beliefs, constraints, and behaviour of relevant political actors, such as interest groups, political parties, government agencies, and firms. These political actors try to shape the political agenda by participating in public debates through the news media, legislative speech, social media, and other arenas by making public statements announcing their beliefs. Each statement is an event. For instance, actor s may argue at time t that the solution r to the problem being debated is unable (a) to solve the problem, and this statement is recorded as an event $e = (s_e, r_e, a_e, t_e)$. Each statement event e contains the name of the political actor who sponsors the belief as the sender s_e, an abstract formulation of the normative policy belief as the receiver r_e, an agreement pattern, such as (positive) support or (negative) rejection of the belief by the actor, as the type a_e, and the day on which the statement appeared in the news media as the event time t_e. The collection of all statements is called a "discourse network" (Leifeld, 2017, 2020). Discourse networks are complex social systems because actors react to other actors' prior statements in the past event sequence E_{t-1}. With this setup, we can model the sequence of statements in a discourse network as a combination of complex functions of the past statement sequence using a REM.

7.2. Case study background, data, and hypotheses

The data re-analysed here were first collected by Leifeld (2013) and re-analysed in Brandenberger (2019), Leifeld (2017), Leifeld and Brandenberger (2019). They comprise a sequence of 6704 statement events by 245 political actors on 69 normative policy beliefs over nine years (1993–2001) in the German pension policy subsystem. Data were collected by manual annotation of statements as text portions in 1842 articles from the Frankfurter Allgemeine Zeitung, a centre-right newspaper with daily national circulation and a large audience, using the software Discourse Network Analyzer (Leifeld, 2023).

In this discourse network, actors debated how to fix the public pension system, the financing of which was being undermined by population ageing. Beliefs on how to solve the pension funding problem were related to ideas as diverse as system change (e. g., adding capital cover elements to the pay-as-you-go pension system by introducing tax-incentivised private savings schemes), fertility (e. g., tying one's pension level to one's number of children to reward the creation of new pension contributors), migration (e. g., increasing the number of pension contributors by boosting immigration), the labour market (e. g., increasing female labour force participation), or incremental change within the existing system (e. g., increasing contribution levels or the minimum retirement age). Actors included banks and insurance companies and their interest groups, industry and employers' associations, trade unions and social interest groups, scientific actors like universities and research institutes, and government actors, including statutory agencies, federal ministries, and political parties. The debate showed signs of multiple competing coalitions (Leifeld and Brandenberger, 2019). More details on the case and data collection can be found in Leifeld (2013).

We model this dataset using a set of seven u statistics. The first three are atemporal and do not require a half-life parameter. `gov` is a dummy variable indicating 1 if the sender is a government actor and 0 otherwise. `sci` is a dummy variable indicating 1 if the sender is a scientific actor and 0 otherwise. Due to their functional roles, government actors can be expected to be more active than other actors while scientific actors should be less likely than other actors to contribute statements to the debate. `agreement` is a fixed effect for stance, where 1 indicates a positive and 0 a negative agreement pattern in an event, which effectively controls for higher numbers of supporting statements than rejecting statements.

The remaining model terms are temporal network statistics and share a common half-life parameter $T_{1/2}$. `inertia` (Eq. (4)) captures path dependence by counting (temporally weighted) past statements with the same actor, belief, and agreement pattern. I. e., are actor–belief–agreement combinations likely repeated? `popularity` (Eq. (5)) is the effect of preferential attachment, whereby

Table 3
Estimated regression coefficients using Bayesian methods (posterior means) under Breslow's approximation, as well as the frequentist method using naïve guess ($T_{1/2} = 20$) and adjusted guess ($T_{1/2} = 111$). Standard errors are included in parentheses.

	Posterior mean (Breslow)	Frequentist $T_{1/2} = 20$	Frequentist $T_{1/2} = 111$
β_{gov}	1.23358	1.63444	1.23287
	(0.04105)	(0.0357)	(0.04021)
β_{sci}	-0.49603	-0.52381	-0.49504
	(0.05826)	(0.05869)	(0.05872)
$\beta_{\text{agreement}}$	0.56485	0.60523	0.56463
	(0.02697)	(0.02673)	(0.02681)
β_{inertia}	17.67374	9.25534	17.65294
	(0.73728)	(0.37565)	(0.64046)
β_{activity}	5.13205	3.26533	5.13558
	(0.08956)	(0.05931)	(0.08522)
$\beta_{\text{popularity}}$	4.81199	3.14784	4.82489
	(0.09163)	(0.04963)	(0.0772)
$\beta_{\text{homophily}}$	1.08851	0.86247	1.08829
	(0.04358)	(0.03703)	(0.04428)

statements about beliefs that have been frequently mentioned recently are more likely. `activity` (Eq. (6)) estimates the tendency for statements to cluster within actors such that actors who were recently active are more likely to make the next statement. `homophily` (Eq. (7)) evaluates the extent to which actors of the same type (e. g., financial market actor, trade union, government actor etc) tend to make the same statements, i. e., whether a statement becomes more likely if other actors of the same type recently referred to the same belief in the same way. Together, these seven statistics explain the statement sequence in the discourse network.

7.3. Convergence diagnostics

Before making any inferential comparisons, trace plots (see, for example, (Lunn et al., 2013)) can be examined to detect anomalies in the MCMC-generated posterior samples. The trajectories of the posterior samples shown in Fig. A.1 in the Appendix indicate that convergence has been achieved for all parameters, with proper mixing and no apparent anomaly. Multiple other convergence diagnostics are also all satisfied by the posterior samples: Geweke's diagnostics proposed by Geweke (1991) suggest that the test of equality of means has been fulfilled. Both the stationary test (Heidelberger and Welch, 1981) and half-width test (Heidelberger and Welch, 1983) also demonstrate that the posterior samples are sufficiently long for inferences to be drawn. See Appendix A for more details. In summary, the MCMC generated posterior samples seem to be well-behaved and, thus, are ready to be used to perform subsequent computations for accurate inferences to be drawn.

7.4. Model interpretation

Kernel estimates of the posterior density of all parameters, including $T_{1/2}$, are presented in Fig. 5. Substantively, the results confirm that being a government actor, a positive stance, higher values for inertia, prior actor activity, prior belief popularity, and actor type homophily increase the log odds of a statement relative to all other statements that could occur and that being a scientific actor makes statements less likely. Most interestingly for policy researchers, actor types display temporal clustering in their shared statements such that actors of a specific type are triggered to make a statement about a belief if the same statement was recently made by other actors sharing the same type. This finding is substantively significant as it demonstrates how political echo chambers are created along the lines of actor types. Financial firms and interest groups parrot each others solutions, government actors segregate into a distinct part of the discourse, and social actors and trade unions in turn pick up and promote each other's policy beliefs and preferred policy instruments in brief cascades.

Interestingly, it can be inferred that the marginal posterior distributions of all parameters are rather similar under Breslow's and Efron's approximations. Exceptions are the regression coefficient of the inertia statistic (β_{inertia}) and $T_{1/2}$. The posterior distribution of β_{inertia} under Efron's approximation is located to the right of that under Breslow, indicating that the estimated effect of the inertia statistic is larger for the former. The reverse is true for $T_{1/2}$, where the posterior of $T_{1/2}$ under Efron's approximation is located to the left of that under Breslow, indicating that the estimated half-life is longer for the latter.

The result has significant implications in terms of the model interpretation. As before, we can plot the temporal decay of $w(t_e, t, T_{1/2})$ using the estimated half-life parameters, shown in Fig. 6. Leifeld and Brandenberger (2019) guessed a fixed $T_{1/2} = 20$ using the same data but a different model specification for frequentist estimation of coefficients. Under Bayesian estimation, a larger half-life with a mode of around $T_{1/2} \approx 107$ (Efron) or $T_{1/2} \approx 111$ (Breslow) was estimated. We see in Fig. 6 that this larger estimated half-life implies that information is retained for a longer period of time relative to the naïve guess. The 95% credible intervals are also relatively narrow, meaning that the data are "confident" of the estimated $T_{1/2}$. Between Breslow and Efron, the rate of temporal decay of information is marginally quicker for Efron than its counterpart.

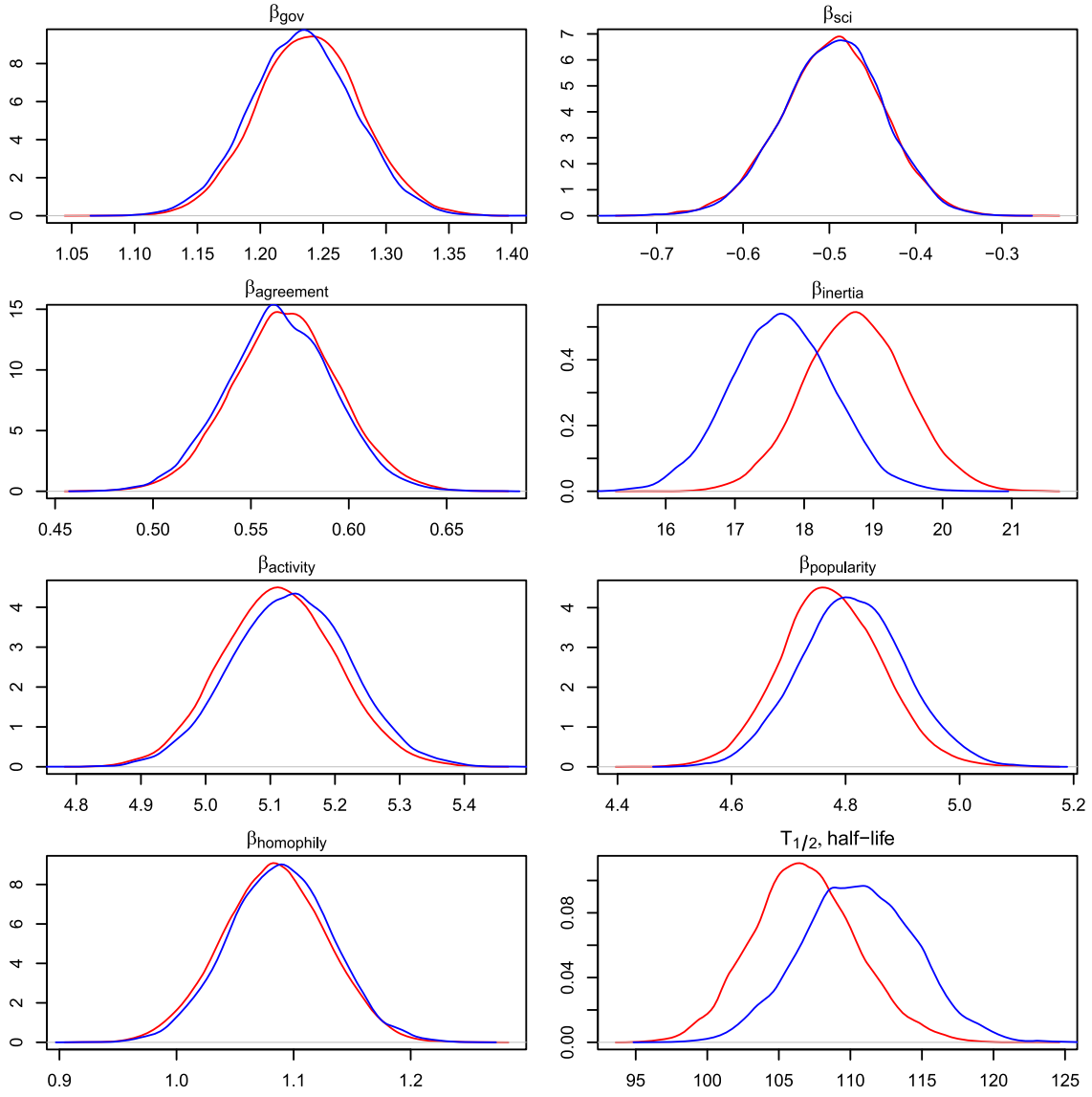


Fig. 5. Kernel density plots of θ for Bayesian REMs under Breslow's (blue) and Efron's (red) approximations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

7.5. Bias analysis

The motivation for the Bayesian approach to REM estimation was the alleged bias caused by fixing $T_{1/2}$. In this section, we assess this bias by presenting results for Bayesian REMs using the flat prior in Eq. (12) under both Breslow's and Efron's approximations and comparing these results to a frequentist approach with a naïve guess of fixed $T_{1/2} = 20$ (Leifeld and Brandenberger, 2019). Frequentist estimation results were obtained by running the `clogit` function of the `R` package `survival` (Therneau, 2024).

Fig. 7 depicts the posterior means of all the estimated regression coefficients under both Breslow's and Efron's approximations, as accompanied by their associated 95% credible intervals. Their corresponding numerical values are provided in Tables 3 and 4. Evidently, the frequentist approach with naïvely specified $T_{1/2} = 20$ underestimated the regression coefficients for the temporal statistics β_{inertia}, β_{activity}, and $\beta_{\text{popularity}}$ considerably. These differences are mainly driven by $T_{1/2}$, highlighting the significance of our approach for drawing correct inferences and theoretical conclusions, i. e., the derivation of the appropriate $T_{1/2}$ from the data.

For a better visualisation of the bias, the plot also displays frequentist estimates using adjusted fixed half-life parameters corresponding to the Bayesian posterior means (Breslow: $T_{1/2} \approx 111$, Efron: $T_{1/2} \approx 107$). The Bayesian point estimates are very close to those from the frequentist fixed-decay approach with adjusted parameters since we imposed non-informative flat priors for the model parameters, meaning that estimates were derived purely from information contained in the data.

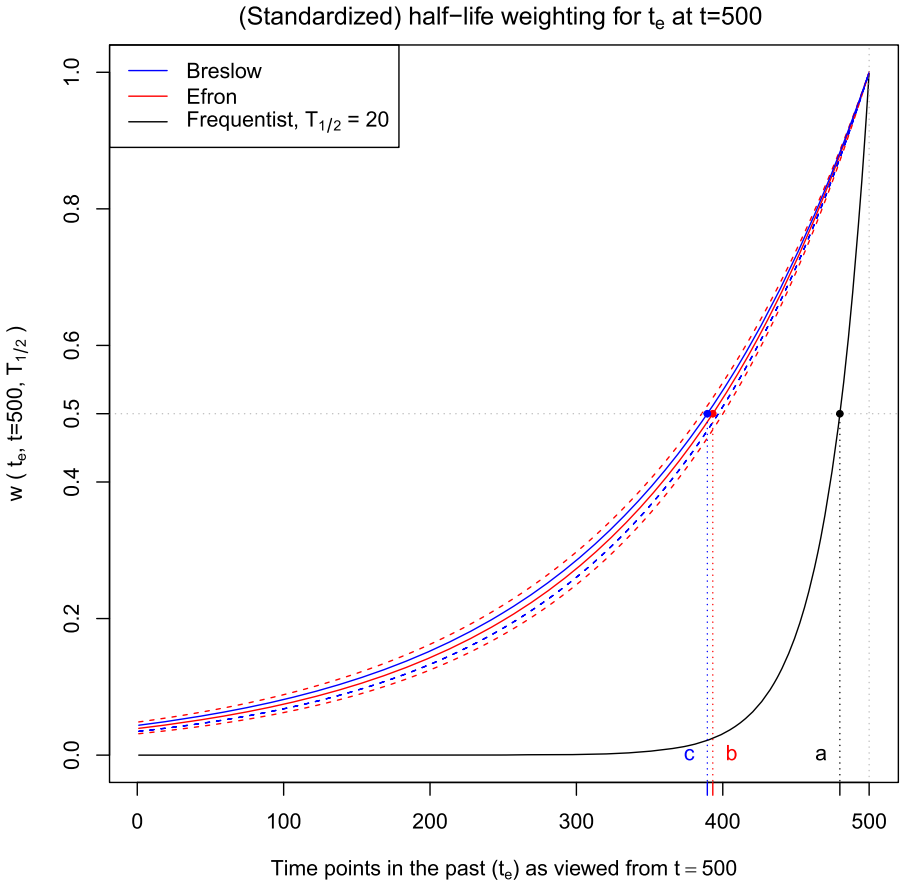


Fig. 6. Exponential temporal decay, $w(t_e, t = 500, T_{1/2})$, estimated using the Bayesian approach under Breslow's and Efron's approximations, along with 95% credible intervals (dashed lines). Also included are the values of $t_e = t - T_{1/2}$ such that past tie events are down-weighted by one half. Specifically, $a = 500 - 20 = 480$, $b = 500 - 107 = 393$, and $c = 500 - 111 = 389$. The flatter curves, governed by the greater half-life parameters of 107 and 111 as opposed to the assumed 20, indicate a greater relative importance of earlier events in the sequence, such that the importance of past events halves after 107 or 111 days rather than after 20 days (c.f. Fig. 1).

Table 4
Estimated regression coefficients using Bayesian methods (posterior means) under Efron's approximation, as well as the frequentist method using naïve guess ($T_{1/2} = 20$) and adjusted guess ($T_{1/2} = 107$). Standard errors are included in parentheses.

	Posterior mean (Efron)	Frequentist $T_{1/2} = 20$	Frequentist $T_{1/2} = 107$
β_{gov}	1.23874	1.63405	1.23842
	(0.04096)	(0.03571)	(0.04013)
β_{sci}	-0.49655	-0.52343	-0.49534
	(0.05902)	(0.05869)	(0.05871)
$\beta_{\text{agreement}}$	0.56795	0.6067	0.56779
	(0.02714)	(0.02672)	(0.02681)
β_{inertia}	18.7381	10.49073	18.69286
	(0.7253)	(0.37654)	(0.63066)
β_{activity}	5.11455	3.26888	5.11829
	(0.08731)	(0.0594)	(0.08463)
$\beta_{\text{popularity}}$	4.77476	3.14888	4.78336
	(0.08842)	(0.0495)	(0.07611)
$\beta_{\text{homophily}}$	1.08361	0.86413	1.08263
	(0.04402)	(0.03703)	(0.0441)

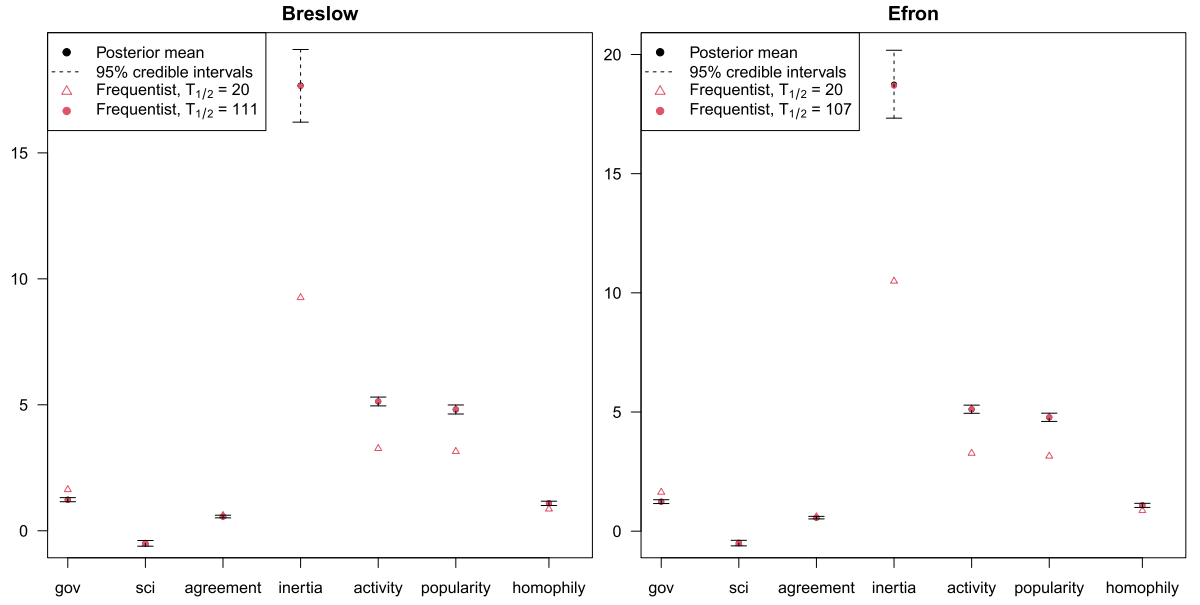


Fig. 7. Estimated coefficients for Bayesian REMs under both Breslow's (left) and Efron's (right) approximations, and frequentist REMs using both naïve ($T_{1/2} = 20$) and adjusted (Breslow: $T_{1/2} \approx 111$, Efron: $T_{1/2} \approx 107$) guesses.

Table 5
Model summaries for Bayesian REMs (posterior means) under both Breslow's and Efron's approximations, as well as the frequentist approach (fixing $T_{1/2} = 20$). Standard errors are included in parentheses.

	Posterior mean (Breslow)	Frequentist (Breslow)	Posterior mean (Efron)	Frequentist (Efron)
β_{gov}	1.23358	1.63444	1.23874	1.63405
	(0.04105)	(0.0357)	(0.04096)	(0.03571)
β_{sci}	−0.49603	−0.52381	−0.49655	−0.52343
	(0.05826)	(0.05869)	(0.05902)	(0.05869)
$\beta_{\text{agreement}}$	0.56485	0.60523	0.56795	0.6067
	(0.02697)	(0.02673)	(0.02714)	(0.02672)
β_{inertia}	17.67374	9.25534	18.7381	10.49073
	(0.73728)	(0.37565)	(0.7253)	(0.37654)
β_{activity}	5.13205	3.26533	5.11455	3.26888
	(0.08956)	(0.05931)	(0.08731)	(0.0594)
$\beta_{\text{popularity}}$	4.81199	3.14784	4.77476	3.14888
	(0.09163)	(0.04963)	(0.08842)	(0.0495)
$\beta_{\text{homophily}}$	1.08851	0.86247	1.08361	0.86413
	(0.04358)	(0.03703)	(0.04402)	(0.03703)
$T_{1/2}$	110.55224	20	106.81214	20
	(3.91312)	(NA)	(3.61755)	(NA)
Log likelihood	−55,940.11	−56,835.06	−55,796.67	−56,659.85
BIC	309,541,000.21	309,542,790.11	309,540,713.34	309,542,439.71

While the point estimates under both frameworks are very close in this particular case, they are intrinsically different since the Bayesian approach naturally integrates all uncertainties coherently (primarily uncertainties with respect to $T_{1/2}$, but also parameter uncertainty due to $\boldsymbol{\beta}$) through simultaneous computation of the joint posterior distribution. Tables 3 and 4 show that standard errors generated under Bayesian approaches produce greater uncertainty relative to frequentist fixed-decay approaches. They are considerably larger compared to the frequentist approach with the naïve guess and mostly larger compared to frequentist results with adjusted half-life, particularly for the temporal network statistics ($\beta_{\text{inertia}}, \beta_{\text{activity}}, \beta_{\text{popularity}}, \beta_{\text{homophily}}$). This is expected as they incorporate uncertainties due to $T_{1/2}$.

Judging from Table 5, both Bayesian approaches performed significantly better than frequentist fixed-decay approaches (based on BIC). Between the Bayesian approaches, Efron's approximation yields larger likelihood and hence lower BIC than Breslow's. Hence, it can be deduced that the Bayesian REM using Efron's approximation is the best model (lowest BIC).

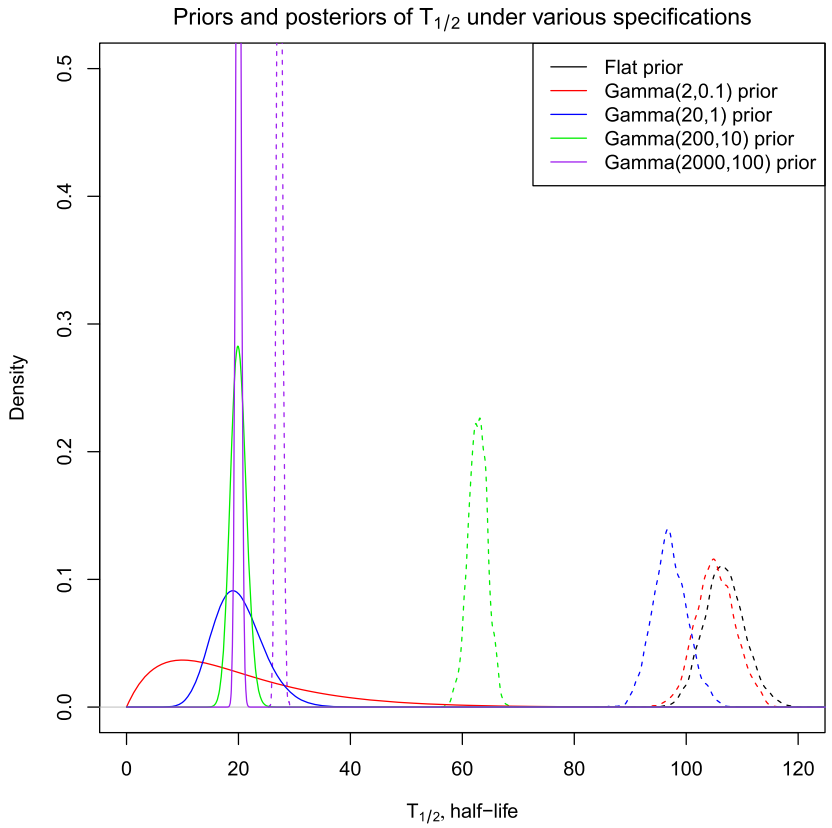


Fig. 8. Density plots of the prior distribution of $T_{1/2}$ (solid lines) with various specifications of hyperparameters, and the corresponding posterior distributions under Efron's approximation (dashed lines).

7.6. Sensitivity analysis

Finally, we examine the role of priors in the estimation of $T_{1/2}$ and assess the sensitivity of results with respect to the hyper-parameters. Fig. 8 illustrates the prior distributions of $T_{1/2}$ under various specifications as described in Section 3.2, along with the corresponding posteriors. For illustrative purposes, only results from Efron's approximation are included, but the same idea applies to Breslow's.

Firstly, we can view the posterior based on the flat prior, $\pi(T_{1/2}) \propto 1$, as predominantly data-driven. When the prior variance is 200 (i. e., Gamma$(2, 0.1)$), the posterior of $T_{1/2}$ is very close to the flat prior. This is sensible because when the prior variance is substantially larger relative to the posterior variance (around 13–15 from Table 5, derived from the squared standard error for $T_{1/2}$), the prior is given less priority relative to the data. In other words, the prior contributes very little in the Bayesian estimation, implying that the results are data-dominated.

Conversely, we observe a gradual shift of the posterior distributions towards the prior mean of 20 as the prior variance becomes smaller (i. e., as b decreases). When the prior variance decreases and, thus, precision increases, the prior contributes increasingly more to the estimation, pulling the posterior of $T_{1/2}$ towards the prior mean of 20 and causing a gradual movement towards the left. In our extreme case (prior variance of 0.2; Gamma$(2000, 100)$), the prior is given most priority relative to the data (i. e., the prior variance is considerably smaller than 13–15) compared to other specifications, producing results resembling the frequentist approach with fixed decay (setting $T_{1/2} = 20$ a priori).

As $b \to 0$ (increasingly vague prior), we retrieve the same results as when the flat prior is used. By contrast, as b becomes larger, i. e., the more informative the prior is, the more the posterior distribution shifts towards the prior mean, and we allow for the prior distribution to take control of the estimation process with less emphasis on the data. In other words, hyper-parameters can be tuned to devise informative priors that can be used to supply additional information to the data in order to aid the estimation of $T_{1/2}$, though that is not necessary here. This is noteworthy as the data may not always contain sufficient information for $T_{1/2}$ to be estimable. For example, underpowered studies, where $T_{1/2}$ may not be estimable, can occur with temporal clustering of many events, short sequences, or increasing model complexity. If $T_{1/2}$ is not estimable given the data, we may leverage the use of hyper-parameters to stabilise the estimation procedure.

8. Conclusion

We presented a fully Bayesian approach to parameter estimation for the time-ordered REM, including computational and methodological advances, and showcased its use in an empirical application to the complex dynamics of political discourse.

As suggested in recent work (Arena et al., 2022, 2023), the analysis demonstrated that estimates for temporal statistics with fixed half-life parameters can indeed be biased. Estimating the temporal decay of network effects as a parameter from the data eliminates this bias.

Estimating the temporal horizon over which theoretical mechanisms play out can also be theoretically informative. For example, we know now that the importance of past news reporting for actors in policy debates who decide what statements to make halves after around 107 days. We can think of many applications in which estimating the time horizon of network dependence endogenously may lead to significant theoretical advances. For example, many published models explaining the occurrence of militarised interstate disputes in international relations employ year-to-year lagged variables or binary indicators of past conflicts within a certain time frame (e. g., Cranmer et al., 2012). Estimating over how many years past conflicts affect current conflicts, whether within conflict dyads or through network mechanisms like structural balance, should improve our understanding and prediction of conflicts substantially.

The analysis furthermore illustrated that the fully Bayesian approach comes with all the advantages typically afforded by Bayesian estimation, such as the possibility of sampling from the full posterior distribution, factoring in prior beliefs, and model diagnostics. Future research should build on these initial results and explore some of the advantages more thoroughly, for example improved prior selection, model selection, or forecasting of discourse networks or other complex dynamic systems by taking into account the time horizon and uncertainty revealed by Bayesian analysis.

The approach presented here also has several limitations, which should be addressed in future work. First, it is computationally demanding. We presented a pre-computation approach that reduces computation time and memory footprint significantly and renders the estimation for moderately sized datasets like discourse networks with several thousand events feasible. Future work should explore building negative sampling or other approaches to reduce the computational burden (Lerner and Lomi, 2020; Overgoor et al., 2020) right into the MCMC chain to permit parameter estimation from very large event sequences, along with an analysis of the tuning of such approaches in a Bayesian estimation context.

Second, future work should explore conjugate and diffuse priors and prior selection in more detail to factor prior beliefs into the analysis in a more principled way and exploit the computational advantages associated with them.

Third, the analysis considered only a single, shared half-life parameter for all temporal network statistics in the model. Permitting separate half-lives for different theoretical mechanisms may increase statistical and theoretical fit. First results conducted separately by the authors reveal that estimation of separate, variable-specific half-life parameters is possible in principle. But whether separate half-life parameters fit better and if the model is estimable in any given application will depend on the data and is an empirical question that should be addressed. One should explore the possibility of letting several temporal statistics share a half-life parameter and other mechanisms share a different parameter. Doing so may also involve Bayesian model selection to evaluate the heterogeneity of time horizons across mechanisms, and one may need to make use of priors to stabilise the estimation as shown in Section 7.6.

Fourth, we used exponential decay here, but other functional forms are generally plausible (Arena et al., 2023), such as power law decay (Amati et al., 2019). The jury is out on which functional form fits which context best.

Fifth, future research should investigate if and how the results extend to other specifications, such as the full-likelihood continuous-time REM (Butts, 2008), multiplex REMs with distinct relationship types, and heterogenous settings where groups of nodes display differential decay.

Sixth, one should explore whether temporal decay can be estimated using maximum likelihood estimation (MLE) and compare bias, efficiency, and convergence to the Bayesian approach under both simple specifications and with multiple, statistic-specific half-life parameters and other complex specifications.

Finally, software implementations should facilitate applications to applied research in the social sciences and lead to more widespread use of these promising techniques.

Appendix A. Convergence diagnostics

Fig. A.1 shows convergence in the MCMC trace for both Breslow's and Efron's tiebreakers.

Geweke's diagnostics proposed by Geweke (1991) is a way of assessing the equality of means of the initial and last portions of a Markov chain. If convergence has been attained, Geweke's statistic is asymptotically normally distributed. As illustrated in Table A.1, since most z-scores computed are within $[-1.96, 1.96]$, it can be concluded that convergence has been attained.

Additionally, a Geweke's plot can be used to detect non-equality of means by discarding successively increasing numbers of samples from the beginning of the chain and reassessing the Geweke statistic. According to Figs. A.2 and A.3, all posterior MCMC chains demonstrate convergence, except a slight divergence for $\beta_{\text{popularity}}$ in the initial part (but nothing concerning since they return to normality in the later part).

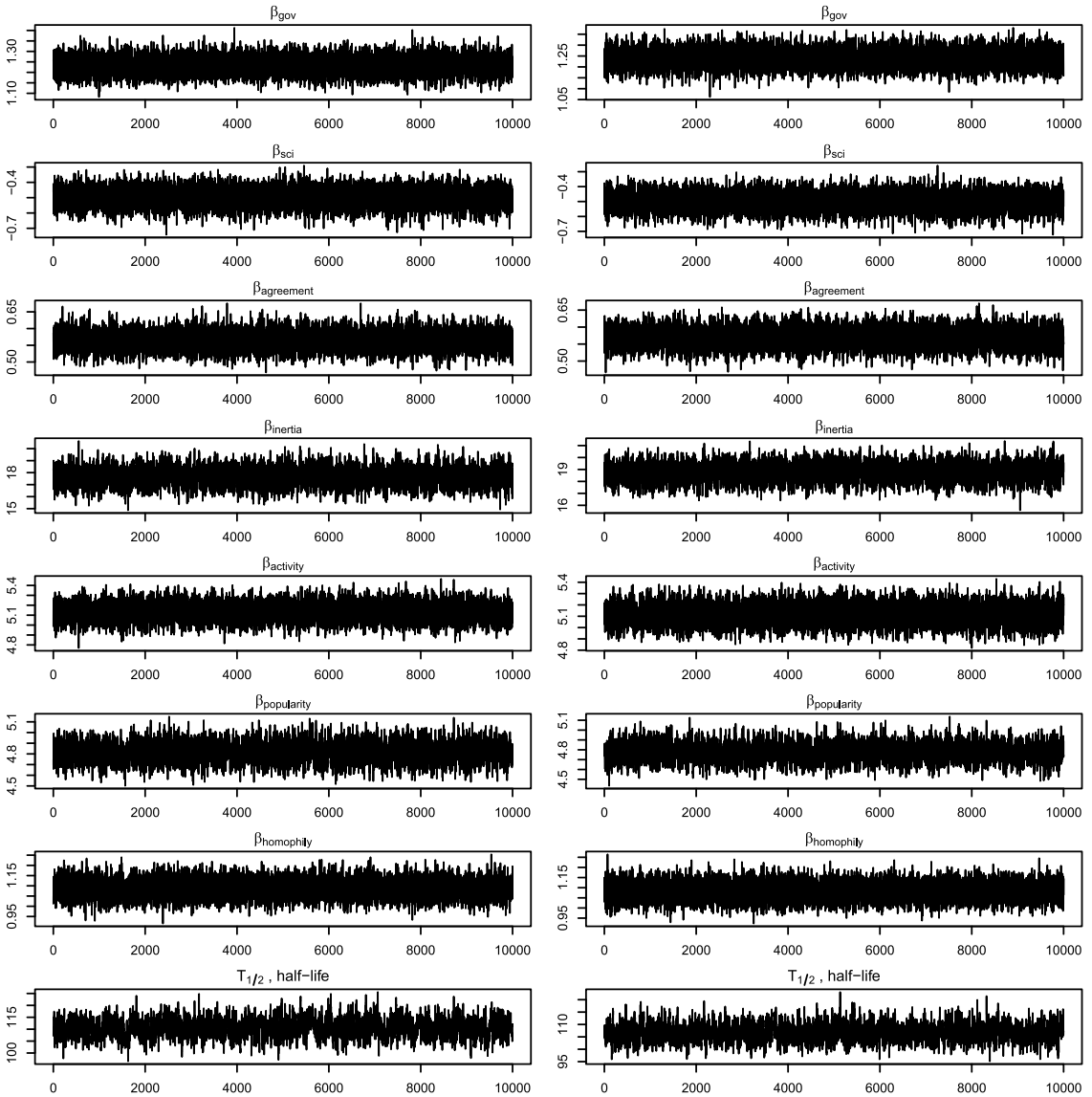


Fig. A.1. Trace plots of posterior samples of θ for Bayesian REMs under Breslow's (left panels) and Efron's (right panels) approximations.

Table A.1
Geweke's statistics/z-scores for all posterior samples under Breslow's and Efron's approximations.

	Breslow	Efron
β_{gov}	−0.68395	−0.91178
β_{sci}	−0.05287	0.49162
$\beta_{\text{agreement}}$	0.25835	0.08317
β_{inertia}	−2.10022	−2.15241
β_{activity}	−0.48175	−0.27483
$\beta_{\text{popularity}}$	−0.31407	1.48268
$\beta_{\text{homophily}}$	−0.49837	0.02294
$T_{1/2}$	−1.66290	−0.43194

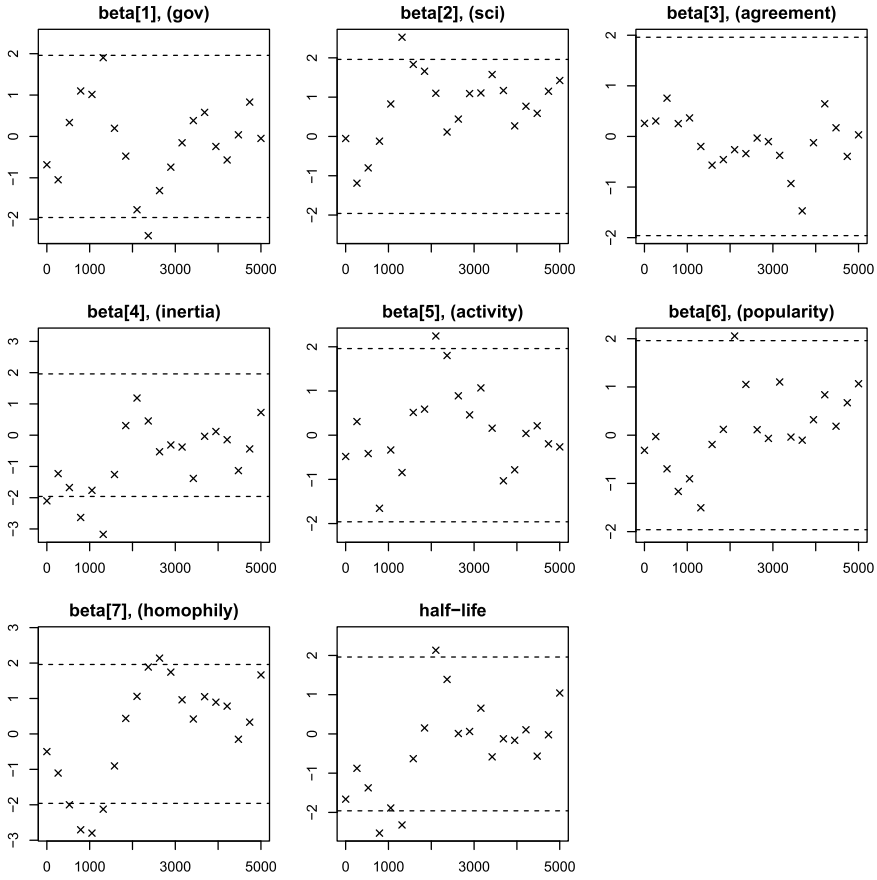


Fig. A.2. Geweke's plot of z-scores for posterior MCMC samples under Breslow's approximation.

Table A.2
Heidel's Stationary and Halfwidth tests for posterior MCMC samples under Breslow's (top) and Efron's (bottom) approximations.

	Stationary test	p-value	Halfwidth test	Mean	Halfwidth
β_{gov}	passed	0.5234	passed	1.23358	0.00079
β_{sci}	passed	0.3062	passed	−0.49603	0.00114
$\beta_{\text{agreement}}$	passed	0.99254	passed	0.56485	0.00062
β_{inertia}	passed	0.05529	passed	17.67374	0.03075
β_{activity}	passed	0.94719	passed	5.13205	0.00267
$\beta_{\text{popularity}}$	passed	0.56294	passed	4.81199	0.0038
$\beta_{\text{homophily}}$	passed	0.21208	passed	1.08851	0.00096
$T_{1/2}$	passed	0.1802	passed	110.55224	0.23216
β_{gov}	passed	0.92788	passed	1.23874	0.00082
β_{sci}	passed	0.99432	passed	−0.49655	0.00122
$\beta_{\text{agreement}}$	passed	0.49736	passed	0.56795	0.00067
β_{inertia}	passed	0.17061	passed	18.74657	0.02712
β_{activity}	passed	0.84117	passed	5.11455	0.00248
$\beta_{\text{popularity}}$	passed	0.09233	passed	4.77476	0.00324
$\beta_{\text{homophily}}$	passed	0.34171	passed	1.08361	0.00096
$T_{1/2}$	passed	0.71159	passed	106.81214	0.20512

Furthermore, Table A.2 shows that all posterior samples satisfied both the stationary (Heidelberger and Welch, 1981) and halfwidth test (Heidelberger and Welch, 1983), indicating that they are sufficiently long for inferences to be drawn.

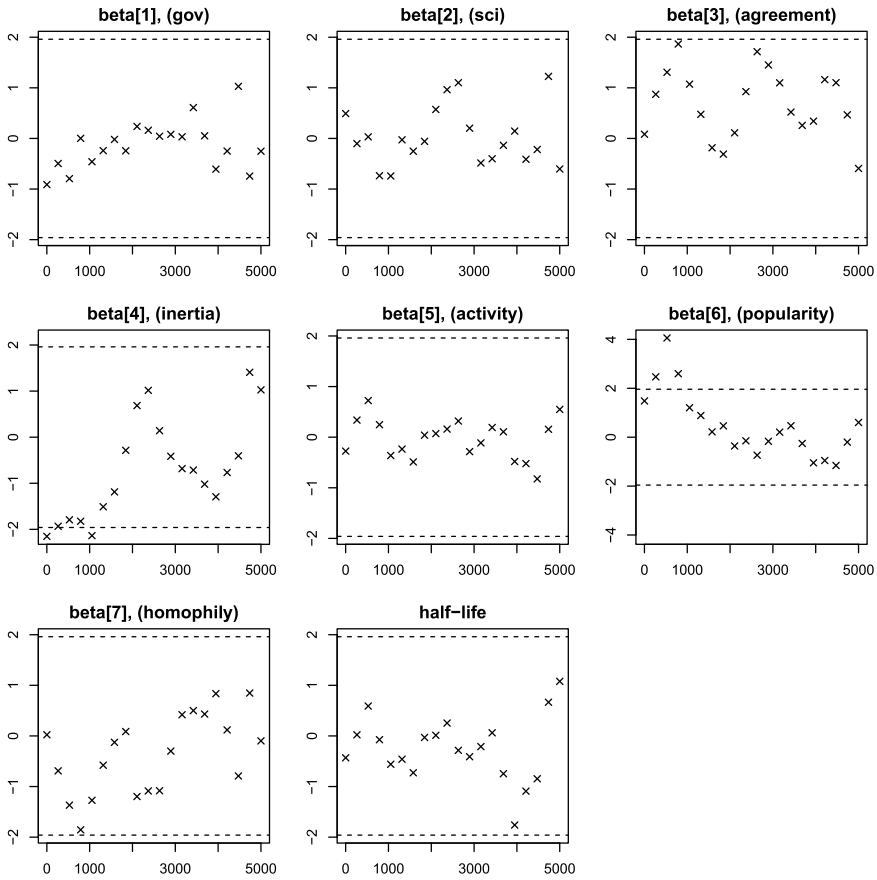


Fig. A.3. Geweke's plot of z-scores for posterior MCMC samples under Efron's approximation.

References

Amati, V., Lomi, A., Mascia, D., 2019. Some days are better than others: examining time-specific variation in the structuring of interorganizational relations. Soc. Netw. 57, 18–33. https://doi.org/10.1016/j.socnet.2018.10.001

Arena, G., Mulder, J., Leenders, R. T.A.J., 2022. A Bayesian semi-parametric approach for modeling memory decay in dynamic social networks. Sociol. Methods Res. 53 (3), 1201–1251. https://doi.org/10.1177/00491241221113875

Arena, G., Mulder, J., Leenders, R. T.A.J., 2023. How fast do we forget our past social interactions? understanding memory retention with parametric decays in relational event models. Netw. Sci. 11 (2), 267–294. https://doi.org/10.1017/nws.2023.5

Artico, I., Wit, E.C., 2023. Dynamic latent space relational event model. J. R. Stat. Soc. A 186 (3), 508–529. https://doi.org/10.1093/jrsssa/qnad042

Bianchi, F., Filippi-Mazzola, E., Lomi, A., Wit, E.C., 2023. Relational event modeling. Annu. Rev. Stat. Appl. 11. https://doi.org/10.1146/annurev-statistics-040722-060248

Brandenberger, L., 2019. Predicting network events to assess goodness of fit of relational event models. Polit. Anal. 27 (4), 556–571. https://doi.org/10.1017/pan.2019.10

Brandenberger, L.M., 2018. Trading favors – examining the temporal dynamics of reciprocity in congressional collaborations using relational event models. Soc. Netw. 54, 238–253. https://doi.org/10.1016/j.socnet.2018.02.001

Brandes, U., Lerner, J., Snijders, T. A.B., 2009. Networks evolving step by step: statistical analysis of dyadic event data. In: 2009 International Conference on Advances in Social Network Analysis and Mining. IEEE, pp. 200–205. https://doi.org/10.1109/ASONAM.2009.28

Breslow, N., 1974. Covariance analysis of censored survival data. Biometrics 30 (1), 89–99. https://doi.org/10.2307/2529620

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

Butts, C.T., Lomi, A., Snijders, T. A.B., Stadtfeld, C., 2023. Relational event models in network science. Netw. Sci. 11 (2), 175–183. https://doi.org/10.1017/nws.2023.9

Cox, D.R., 1972. Regression models and life-tables. J. R. Stat. Soc. B 34 (2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Cranmer, S.J., Desmarais, B.A., Menninga, E.J., 2012. Complex dependencies in the alliance network. Confl. Manag. Peace Sci. 29 (3), 279–313. https://doi.org/10.1177/0738894212443446

Daziano, R.A., 2013. Conditional-logit Bayes estimators for consumer valuation of electric vehicle driving range. Resour. Energy Econ. 35 (3), 429–450. https://doi.org/10.1016/j.reseneeco.2013.05.001

Efron, B., 1977. The efficiency of Cox's likelihood function for censored data. J. Am. Stat. Assoc. 72 (359), 557–565. https://doi.org/10.1080/01621459.1977.10480613

Gelfand, A.E., Sahu, S.K., 1999. Identifiability, improper priors, and Gibbs sampling for generalized linear models. J. Am. Stat. Assoc. 94 (445), 247–253. https://doi.org/10.1080/01621459.1999.10473840

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7 (4), 457–472. https://doi.org/10.1214/ss/1177011136

Gelman, A., Rubin, D.B., Carlin, J.B., Stern, H.S., 1995. Bayesian Data Analysis. Chapman and Hall, London.

Geweke, J., 1991. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Staff Report 148. Federal Reserve Bank of Minneapolis. https://doi.org/10.21034/sr.148
Gravel, J., Valasik, M., Mulder, J., Leenders, R., Butts, C., Brantingham, P.J., Tita, G.E., 2023. Rivalries, reputation, retaliation, and repetition: testing plausible mechanisms for the contagion of violence between street gangs using relational event models. Netw. Sci. 11 (2), 324–350. https://doi.org/10.1017/nws.2023.8
Heidelberger, P., Welch, P.D., 1981. A spectral method for confidence interval generation and run length control in simulations. Commun. ACM 24 (4), 233–245. https://doi.org/10.1145/358598.358630
Heidelberger, P., Welch, P.D., 1983. Simulation run length control in the presence of an initial transient. Oper. Res. 31 (6), 1109–1144. https://doi.org/10.1287/opre.31.6.1109
Hunter, D.R., Handcock, M.S., 2006. Inference in curved exponential family models for networks. J. Comput. Graph. Stat. 15 (3), 565–583. https://doi.org/10.1198/106186006X133069
Kamalabad, M.S., Leenders, R., Mulder, J., 2023. What is the point of change? Change point detection in relational event models. Soc. Netw. 74, 166–181. https://doi.org/10.1016/j.socnet.2023.03.004
Koop, G., Poirier, D.J., 1993. Bayesian analysis of logit models using natural conjugate priors. J. Econom. 56 (3), 323–340. https://doi.org/10.1016/0304-4076(93)90124-N
Leenders, R. T.A.J., Contractor, N.S., DeChurch, L.A., 2016. Once upon a time: understanding team processes as relational event networks. Organ. Psychol. Rev. 6 (1), 92–115. https://doi.org/10.1177/2041386615578312
Leifeld, P., 2013. Reconceptualizing major policy change in the advocacy coalition framework: a discourse network analysis of German pension politics. Policy Stud. J. 41 (1), 169–198. https://doi.org/10.1111/psj.12007
Leifeld, P., 2017. Discourse network analysis: policy debates as dynamic networks. In: Victor, J.N., Montgomery, A.H., Lubell, M.N. (Eds.), The Oxford Handbook of Political Networks. Oxford University Press, Oxford, pp. 301–326. https://doi.org/10.1093/oxfordhb/9780190228217.013.25
Leifeld, P., 2020. Policy debates and discourse network analysis: a research agenda. Polit. Gov. 8 (2), 180–183. https://doi.org/10.17645/pag.v8i2.3249
Leifeld, P., 2023. Discourse network analyzer (DNA). Software available from: https://github.com/leifeld/dna/.
Leifeld, P., 2024. Discourse network analysis: bibliography. https://github.com/leifeld/dna/blob/master/build/bibliography.md (accessed 18 December 2025).
Leifeld, P., Brandenberger, L., 2019. Endogenous coalition formation in policy debates. arXiv:1904.05327v1, available from https://arxiv.org/abs/1904.05327v1 (accessed 18 December 2025).
Lerner, J., 2016. Structural balance in signed networks: separating the probability to interact from the tendency to fight. Soc. Netw. 45, 66–77. https://doi.org/10.1016/j.socnet.2015.12.002
Lerner, J., Bussmann, M., Snijders, T. A.B., Brandes, U., 2013a. Modeling frequency and type of interaction in event networks. Corvinus J. Sociol. Soc. Policy 4 (1), 3–32. https://doi.org/10.14267/cjssp.2013.01.01
Lerner, J., Indlekofer, N., Nick, B., Brandes, U., 2013b. Conditional independence in dynamic networks. J. Math. Psychol. 57 (6), 275–283. https://doi.org/10.1016/j.jmp.2012.03.002
Lerner, J., Lomi, A., 2020. Reliability of relational event model estimates under sampling: how to fit a relational event model to 360 million dyadic events. Netw. Sci. 8 (1), 97–135. https://doi.org/10.1017/nws.2019.57
Lerner, J., Lomi, A., 2023. Relational hyperevent models for polyadic interaction networks. J. R. Stat. Soc. A 186 (3), 577–600. https://doi.org/10.1093/jrsssa/qnac012
Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2013. The BUGS Book: A Practical Introduction to Bayesian Analysis. Chapman and Hall, London.
Malang, T., Brandenberger, L.M., Leifeld, P., 2019. Networks and social influence in European legislative politics. Br. J. Polit. Sci. 49 (4), 1475–1498. https://doi.org/10.1017/S0007123417000217
McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.
Meijerink-Bosman, M., Leenders, R. T.A.J., Mulder, J., 2022. Dynamic relational event modeling: testing, exploring, and applying. PLOS One 17 (8), e0272309. https://doi.org/10.1371/journal.pone.0272309
Mulder, J., Leenders, R. T.A.J., 2019. Modeling the evolution of interaction behavior in social networks: a dynamic relational event approach for real-rime analysis. Chaos Solit. Fractals 119, 73–85. https://doi.org/10.1016/j.chaos.2018.11.027
O'Hagan, A., Forster, J.J., 2004. Kendall's Advanced Theory of Statistics. Vol. 2B. Arnold, London. 2nd edition.
Overgoor, J., Pakapol Supaniratisai, G., Ugander, J., 2020. Scaling choice models of relational social data. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1990–1998. https://doi.org/10.1145/3394486.3403250
Pilny, A., Schecter, A., Poole, M.S., Contractor, N., 2016. An illustration of the relational event model to analyze group interaction processes. Group Dyn. 20 (3), 181. https://doi.org/10.1037/gdn0000042
Quintane, E., Conaldi, G., Tonellato, M., Lomi, A., 2014. Modeling relational events: a case study on an open source software project. Organ. Res. Methods 17 (1), 23–50. https://doi.org/10.1177/1094428113517007
Roberts, G.O., Rosenthal, J.S., 2001. Optimal scaling for various Metropolis–Hastings algorithms. Stat. Sci. 16 (4), 351–367. https://doi.org/10.1214/ss/1015346320
Roberts, G.O., Sahu, S.K., 1997. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. J. R. Stat. Soc. B 59 (2), 291–317. https://doi.org/10.1111/1467-9868.00070
Stadtfeld, C., Block, P., 2017. Interactions, actors, and time: dynamic network actor models for relational events. Sociol. Sci. 4, 318–352. https://doi.org/10.15195/v4.a14
Therneau, T.M., 2024. A Package for Survival Analysis in R. R package version 3.6-4. https://CRAN.R-project.org/package=survival.
Tranmer, M., Marcum, C.S., Morton, F.B., Croft, D.P., de Kort, S.R., 2015. Using the relational event model (REM) to investigate the temporal dynamics of animal social networks. Anim. Behav. 101, 99–105. https://doi.org/10.1016/j.anbehav.2014.12.005
Valeeva, D., Heemskerk, E.M., Takes, F.W., 2020. The duality of firms and directors in board interlock networks: a relational event modeling approach. Soc. Netw. 62, 68–79. https://doi.org/10.1016/j.socnet.2020.02.009
Vieira, F., Leenders, R., McFarland, D., Mulder, J., 2024. A Bayesian actor-oriented multilevel relational event model with hypothesis testing procedures. Behaviormetrika 51 (1), 37–74. https://doi.org/10.1007/s41237-023-00203-4
Vu, D., Lomi, A., Mascia, D., Pallotti, F., 2017. Relational event models for longitudinal network data with an application to interhospital patient transfers. Stat. Med. 36 (14), 2265–2287. https://doi.org/10.1002/sim.7247
Ward, M.D., Stovel, K., Sacks, A., 2011. Network analysis and political science. Annu. Rev. Polit. Sci. 14, 245–264. https://doi.org/10.1146/annurev.polisci.12.040907.115949
Wong, J. S.T., Forster, J.J., Smith, P. W.F., 2018. Bayesian mortality forecasting with overdispersion. Insur. 83, 206–221. https://doi.org/10.1016/j.insmatheco.2017.09.023