



A Bayesian actor-oriented multilevel relational event model with hypothesis testing procedures

Fabio Vieira¹ · Roger Leenders^{2,3} · Daniel McFarland⁴ · Joris Mulder¹

Received: 2 February 2023 / Accepted: 21 June 2023 / Published online: 17 July 2023
© The Author(s) 2023, corrected publication 2023

Abstract

Relational event network data are becoming increasingly available. Consequently, statistical models for such data have also surfaced. These models mainly focus on the analysis of single networks; while in many applications, multiple independent event sequences are observed, which are likely to display similar social interaction dynamics. Furthermore, statistical methods for testing hypotheses about social interaction behavior are underdeveloped. Therefore, the contribution of the current paper is twofold. First, we present a multilevel extension of the dynamic actor-oriented model, which allows researchers to model sender and receiver processes separately. The multilevel formulation enables principled probabilistic borrowing of information across networks to accurately estimate drivers of social dynamics. Second, a flexible methodology is proposed to test hypotheses about common and heterogeneous social interaction drivers across relational event sequences. Social interaction data between children and teachers in classrooms are used to showcase the methodology.

Keywords Relational event history · Bayesian inference · Multilevel analysis · Social networks

✉ Fabio Vieira
f.v.generosovieira@tilburguniversity.edu

Roger Leenders
r.t.a.j.leenders@tilburguniversity.edu

Daniel McFarland
dmcfarla@stanford.edu

Joris Mulder
j.mulder3@tilburguniversity.edu

¹ Department of Methodology and Statistics, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

² Department of Organization Studies, Tilburg University, Tilburg, The Netherlands

³ Jheronimus Academy of Data Science, 's-Hertogenbosch, The Netherlands

⁴ Department of Education, Stanford University, Stanford, CA, USA

1 Introduction

Current technological advancements, combined with constant development of new communication applications, has originated huge amounts of data on social interactions (Eagle and Pentland 2003). This has enabled researchers to build and test increasingly rich and complex models of human interaction using time-stamped data. Social network analysis is among the fields that have benefited the most from having access to rich temporal interaction data. In this paper, we specifically focus on the Relational Event network model (Butts and Marcum 2017; Stadtfeld and Block 2017; Mulder and Leenders 2019)

A relational event network is comprised of multiple interactions among a finite set of actors. Observation units are called “relational events”, which are defined as discrete instances of interactions among social entities along a timescale (Butts and Marcum 2017). In a *directed* network, events contain a clear indication of sender and receiver. Thus, every observation unit displays information on which actor was the sender, which actor was the receiver, and at which point in time the interaction occurred. The stream of events in a given network is often referred to as a relational event sequence.

One of the traditional approaches to social network analysis is based on Markovian random graphs theory and involves the aggregation of events into a graph (Van Der Hofstad 2009). Transitions on the graph structure are then modeled via sufficient statistics, which are functions of the observed network ((Frank and Strauss, 1986; Hanneke et al., 2010; Lusher et al., 2013)). A different approach was taken by Butts (2008), who employed survival analysis concepts and introduced the *relational event model* which profits from the temporal structure of relational events. In this framework, the main goal is to model the rates of communication between senders and receivers via a log-linear function, without requiring the data to be aggregated. This is done by employing sufficient statistics that capture important social patterns and actor-specific covariates.

Since then, the relational event model has gained increasing popularity and received multiple extensions. For instance, Vu et al. (2011) introduced a model with time-varying parameters, based on the additive Aalen model (Aalen 1989), and developed an algorithm for online inference in social networks. Perry and Wolfe (2013) used a partial likelihood approach to modeling who the receiver will be a given sender. Vu et al. (2015) implemented case–control sampling to decrease the number of computations in estimating relational event models. Later, the control–case sampling approach was further explored by Lerner and Lomi (2020) to estimate relational event models in large networks. Stadtfeld et al. (2017a) and Stadtfeld and Block (2017) introduced a two-step model, where the first step consists of modeling the activity rate of the sender and the second one features the choice of the receiver conditional on the sender. Mulder and Leenders (2019) proposed a way to analyze the temporal evolution of effects in the social network, by estimating effects in different subsets of the data determined by overlapping intervals, which they called moving windows, this approach was further developed by Meijerink et al. (2022).

These contributions have mainly focused on the statistical analysis of a single relational event sequence (with a notable exception of DuBois et al. (2013)); while in practice, we also often observe multiple independent relational event sequences which show similar but not identical social interaction behavior. For example, Blonder and Dornhaus (2011) analyses information flow in multiple relational event networks of ant colonies, DuBois et al. (2013) studies several relational event sequences of high-school students interactions and Kiti et al. (2016) examines social contact in numerous networks of Kenyan households. The gold standard for analyzing such clustered (or hierarchical) data is using a multilevel or mixed effects approach. In a relational event modeling approach, this implies that a probability distribution is specified for the network effects (or drivers) of social interaction behavior (such as inertia, reciprocity, transitivity, or group effects) across clusters. A multilevel approach will result in ‘pooled’ estimates of the cluster-specific network effects where information about social interaction behavior in other clusters is borrowed to improve the estimation (Gelman and Hill 2006). We achieve this by modeling the relational event sequences in a Bayesian hierarchical framework via Markov chain Monte Carlo (MCMC) methods.

Therefore, our first contribution is a multilevel extension for the dynamic actor-oriented model (Stadtfeld et al. 2017a) for independent relational event sequences. In this framework, a relational event between a sender and a receiver in a network is modeled by separately modeling (i) when an actor decides to initiate an event as a sender given the past history, and (ii) modeling the receiver that is chosen given the sender and the past event history between the actors. Thereby, this approach differs from the dyadic approach of Butts (2008), and the hierarchical extension by DuBois et al. (2013), where the time, sender, and receiver are jointly modeled using the same set of parameters (Stadtfeld et al. 2017b). The actor-oriented approach is therefore more flexible by separately modeling the behavior of actors to become a sender and the behavior of senders to choose a receiver in contrast to the dyadic approach. A second important difference is that we explicitly model the full unstructured covariance matrix of the network effects across sequences. Thereby, the dependency between interaction behavior across sequences, such as inertia and reciprocity (which is generally nonzero), is included in the analysis resulting in improved estimation using the concept of pooling. A final important difference between a dyadic relational event model and an actor-oriented approach is the computational burden, which is considerably larger in the case of a dyadic approach. This can be explained by the size of the risk set (i.e., the possible events that can be observed at every step), which is equal to $N(N - 1)$ in the case of a network consisting of N actors where $N(N - 1)$ dyads are at risk for each event, while the risk set of the sender model in an actor-oriented model consists of N actors, and the risk set of the receiver model consists of $N - 1$ actors (when assuming that the sender cannot be equal to the receiver). In the case of moderate to large networks, this difference can have a huge impact on the computational costs when using a dyadic approach.

Our second contribution is an extensive set of statistical tests using the Bayes factor (Jeffreys 1961; Kass and Raftery 1995) for evaluating hypotheses about social interaction behavior in the case of independent relational event sequences. The first test can be used to assess whether specific interaction behavior, such as inertia

(which quantifies the degree of habitual behavior between actors to keep sending messages of one actor towards another actor), is equal or differs across sequences (i.e., whether it should be modeled as a fixed effect or as a random effect). In the case of different interaction behavior across sequences, a second test is proposed to assess the degree of heterogeneity (or variability) of network effects on social interaction behavior, e.g., to assess whether inertia behavior varies more across clusters than reciprocity, or whether the impact of a group effect (e.g., whether the sender is the teacher in a class room or not) on social interaction dynamics is more or less heterogeneous than another group effect (e.g., whether the sender is male or female). Third, a test is proposed for evaluating hypotheses with equality and order constraints on the average relative importance across clusters (i.e., on the fixed effects) of drivers on social interaction dynamics. The methodology is flexible to also allow for testing the absolute values of network effects, which is useful when the exact sign is unknown before observing the data. These tests build on and further extends recent developments on Bayes factor hypothesis testing (Mulder et al. 2021).

Our methods have been implemented in R (R Core Team 2017), interfacing with Stan through the rstan package (Stan Development Team 2018). We provide the full code to facilitate the fitting of other relational event models (https://github.com/Fabio-Vieira/bayesian_dynamic_network).

The remainder of this paper is divided as follows: in Sect. 2, we describe the relational events framework and present the hierarchical actor-oriented relational event model. This Section also introduces the Bayesian specification of the model by presenting prior distributions and briefly discussing a reparametrization that allows more efficient sampling; we detail the hypothesis testing methods in Sect. 3 and conduct the empirical application in Sect. 4. The paper ends with a discussion in Sect. 5, where we go through some limitations of the model and possible routes for future research.

2 The relational event framework

The relational event model (REM) is used to model the rate of interactions in dynamic social networks (Butts 2008). In this framework, events happen among actors at particular points in time, being represented by tuples in the form $e_t = (s, r, t)$, where s is the sender, r is the receiver, and t is the time point of the interaction. Thus e_t is called a relational event. Therefore, we have K clusters (in this paper we will use cluster, groups, relational event sequence and network interchangeably), each of which with N_1, \dots, N_K social actors allowed to be senders or receivers at any given time. At time t , in cluster k , sender $s \in \{1, 2, \dots, N_k\}$ interacts with receiver $r \in \{1, 2, \dots, N_k \mid r \neq s\}$, forming a dyad $(s, r) \in \mathcal{R}_k(t)$, where $\mathcal{R}_k(t) = \{(s, r) : s, r \in \{1, 2, \dots, N_k\}, s \neq r\}$ is called the risk set, which comprises all possible dyads (s, r) at a particular point in time.

The main idea consists of assuming that, for each cluster k , for $k = 1, \dots, K$, we are able to observe an ordered sequence of M_k dyadic events among N_k individuals on the time window $[0, \tau_k) \in \mathbb{R}^+$. Thus, a relational event sequence for cluster k is formally defined as

$$\mathbf{E}_k = \{e_{t_m} = (s_m, r_m, t_m) : (s_m, r_m) \in \mathcal{R}_k(t_m), 0 < t_1 < \dots < t_{M_k} < \tau_k\},$$

where t_m is the time at which the m th event occurred, in cluster k . Then, following Butts (2008), who borrows concepts from survival analysis, the relational event history \mathbf{E}_k is modeled as a stochastic process, where the rate of events from sender s to receiver r , with $(s, r) \in \mathcal{R}_k(t)$, is given by the intensity $\lambda_{sr}(t|\mathbf{E}_k)$. This intensity function has the form of a Cox proportional hazards model (Cox 1972). Moreover, the intensity is assumed constant between subsequent events and the waiting times conditionally exponentially distributed. This amounts to the well-known piece-wise constant exponential model for survival data (Friedman 1982). Thus, the survival function is given by $S_{sr}(t_m - t_{m-1}|\mathbf{E}_k) = \exp\{- (t_m - t_{m-1})\lambda_{sr}(t_m|\mathbf{E}_k)\}$.

2.1 The actor-oriented multilevel relational event model

In this paper, we focus on an alternative relational event model that was proposed by Stadtfeld et al. (2017a) and Stadtfeld and Block (2017). Their approach models the receiver given the sender, similar to Perry and Wolfe (2013). Conceptually, it builds on the same tradition as the stochastic actor-oriented model (Snijders 1996), where the evolution of the network is assumed to be a product of actors’ individual behaviors as they constantly seek to maximize their own utilities. The basic framework consists of a two-step approach based on log-linear predictors. First, the waiting time until an actor becomes active is modeled. After that, a multinomial choice model (McFadden 1973) is employed to determine the choice of the receiver by the active sender actor. These two steps are assumed to be conditionally independent given the available information about the past up to that event.

The Bayesian hierarchical approach we will develop for multiple relational event sequences (denoted as "clusters") will build on this model. At each point in time, in cluster k , sender $s \in \{1, 2, \dots, N_k\}$ starts an interaction with intensity $\lambda_s(t|\mathbf{E}_k)$. This intensity is directly proportional to the probability of a given actor s to be the next sender. This probability is given by $P(s_m = s|\mathbf{E}_k) = \lambda_s(t|\mathbf{E}_k) / \sum_{h \in k} \lambda_h(t|\mathbf{E}_k)$, $\forall h \in \{1, 2, \dots, N_k\}$ and $m \in \{1, 2, \dots, M_k\}$, where M_k is the number of events in cluster k . Next, this sender chooses the receiver $r \in \{1, 2, \dots, N_k | r \neq s\}$, forming the dyad $(s, r) \in \mathcal{R}_k(t)$, with intensity $\lambda_{r|s}(t|s, \mathbf{E}_k)$. The receiver intensity represents the rate at which actor s chooses actor r to form a dyad, which is proportional to the probability of observing dyad $(s, r) \in \mathcal{R}_k(t)$ as the next one in the sequence. This probability is given by $P(r_m = r|s_m = s, \mathbf{E}_k) = \lambda_{r|s}(t|s, \mathbf{E}_k) / \sum_{h \in k} \lambda_{h|s}(t|s, \mathbf{E}_k)$, $\forall h \in \{1, 2, \dots, N_k | h \neq s\}$ and $m \in \{1, 2, \dots, M_k\}$. Then, these intensities for cluster k , for the sender and receiver steps of this model, will be given by the following log-linear functions:

$$\begin{aligned} \lambda_s(t|\mathbf{E}_k) &= \exp\{\boldsymbol{\phi}'\mathbf{z}_s(t) + \boldsymbol{\gamma}'_k\mathbf{x}_s(t)\}, \\ \lambda_{r|s}(t|s, \mathbf{E}_k) &= \exp\{\boldsymbol{\psi}'\mathbf{z}_{sr}(t) + \boldsymbol{\beta}'_k\mathbf{x}_{sr}(t)\}, \end{aligned} \tag{1}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ are a vectors of fixed-effect parameters, $\boldsymbol{\gamma}_k$ and $\boldsymbol{\beta}_k$ are a vectors of random-effect parameters for cluster k , with $k = 1, \dots, K$. The vectors of statistics

$\mathbf{z}_s(t)$ and $\mathbf{x}_s(t)$ are associated with actor $s \in \{1, 2, \dots, N_k\}$, whereas $\mathbf{z}_{sr}(t)$ and $\mathbf{x}_{sr}(t)$ are vectors of statistics associated with dyad $(s, r) \in \mathcal{R}_k(t)$.

Assuming this log-linear form will allow us to conduct inference at the actor level, unveiling effects that make actors more (or less) prone to start interactions or more (or less) likely to be chosen as the next receiver. Also, this idea helps us to limit the size of the set of possible dyads that need to be analyzed at each point in time: if actor s is the one starting an interaction at time t , then all dyads, where s is not the sender, become impossible to happen. Thus, for the actor-oriented model, the risk set, $\mathcal{R}_k(t)$, will have size $N_k - 1$ (given the sender), whereas for a dyadic model, such as in DuBois et al. (2013), the risk set has size $(N_k - 1)N_k$, which can easily become massive for medium to large networks.

The likelihood for the actor-oriented model is given by

$$p(\mathbf{E}|\boldsymbol{\theta}, \mathbf{Z}, \mathbf{X}) = \prod_{k=1}^K \prod_{m=1}^{M_k} \left[\left(\lambda_{s_m}(t_m|\mathbf{E}_k) \prod_{s \in k} S_s(t_m - t_{m-1}|\mathbf{E}_k) \right) \times \frac{\lambda_{r_m|s_m}(t_m|s_m, \mathbf{E}_k)}{\sum_{r \in k} \lambda_{r|s_m}(t_m|s_m, \mathbf{E}_k)} \right] \times \prod_{s \in k} S_s(\tau_k - t_{M_k}|\mathbf{E}_k), \quad (2)$$

where $\boldsymbol{\theta}$ is the vector containing all parameters and \mathbf{Z} and \mathbf{X} are matrices with fixed and random-effects covariates, respectively. The time of the last observed event in cluster k is denoted by t_{M_k} and τ_k the end of the observation period. In most empirical applications, it is assumed that $t_{M_k} = \tau_k$. This way the last part of the likelihood is equal to 1, since due to the piece-wise constant exponential assumption we have $S_{s_m}(t_m - t_{m-1}|\mathbf{E}_k) = \exp\{- (t_m - t_{m-1})\lambda_s(t_m|\mathbf{E}_k)\}$. In this setting, at time t_m , $P(s = s_m|\mathbf{E}_k) = \frac{\lambda_{s_m}}{\sum_{h \in k} \lambda_h}$, $\forall s \in k$ is the probability of sender s being active. The probability of actor r being the receiver given that s is the sender is given by $P(r_m = r|s_m = s, \mathbf{E}_k) = \frac{\lambda_{r|s}}{\sum_{h \in k} \lambda_{h|s}}$, $\forall r \in k$.

The likelihood in Eq. (2) is a product of the two pieces. The first one, representing the sender model, is a piece-wise constant exponential likelihood

$$p_{\text{sender}}(\mathbf{E}|\boldsymbol{\theta}, \mathbf{Z}, \mathbf{X}) = \prod_{k=1}^K \prod_{m=1}^{M_k} \left[\left(\lambda_{s_m}(t_m|\mathbf{E}_k) \prod_{s \in k} S_s(t_m - t_{m-1}|\mathbf{E}_k) \right) \right],$$

and the second one, representing the receiver model, is a multinomial likelihood

$$p_{\text{receiver}}(\mathbf{E}|\boldsymbol{\theta}, \mathbf{Z}, \mathbf{X}) = \prod_{k=1}^K \prod_{m=1}^{M_k} \left[\frac{\lambda_{r_m|s_m}(t_m|s_m, \mathbf{E}_k)}{\sum_{r \in k} \lambda_{r|s_m}(t_m|s_m, \mathbf{E}_k)} \right].$$

In the literature, it has been shown that both of these models are special cases of the Poisson model. Holford (1980) and Laird and Olivier (1981) are examples of cases where the Poisson representation of the piece-wise exponential model is discussed.

Baker (1994) extensively discusses the multinomial-Poisson transformation and how their likelihoods yield identical estimates. For proofs, see Appendix A. The Poisson regression model has been extensively studied, and it is well understood in the statistical literature (Frome 1983; Consul and Famoye 1992; Hayat and Higgins 2014).

The second level of the hierarchical actor-oriented relational event model specifies the multivariate distributions of network effects across the K sequences. We follow the standard approach in multilevel modeling by assuming multivariate normal distributions for β_k under the sender model and for γ_k under the receiver model, i.e.,

$$\begin{aligned} \beta_k &\sim \mathcal{N}(\mu_\beta, \Sigma_\beta), \\ \gamma_k &\sim \mathcal{N}(\zeta_\gamma, \Sigma_\gamma), \end{aligned}$$

for $k = 1, \dots, K$. The mean vectors quantify the overall, global effect of the network effects and the unstructured covariance matrices quantify the variability of the effects across clusters and the dependency structure between the effects across clusters. Thus, when estimating these distributions when fitting the multilevel model using independent relational event sequences, the estimated means and (co)variances of the second level are used to improve the estimates of the specific parameters in the separate sequences (especially in the case of short sequences).

2.2 Prior specification

In a Bayesian approach, prior distributions have to be specified which reflect our uncertainty about the model parameters before observing the data. Throughout this paper, we shall work with vague, noninformative priors (which are completely dominated by the data).

The prior choice of the random-effects covariance matrices is most important. We decompose the random-effects covariance matrix under the receiver model according to

$$\Sigma_\beta = \sigma_\beta \times \Omega_\beta \times \sigma_\beta, \tag{3}$$

where $\sigma_\beta := \text{diag}(\sigma_{\beta,1}, \dots, \sigma_{\beta,P})$ is a diagonal matrix of standard deviations and Ω_β is a correlation matrix (Gelman and Hill 2006). Following Carpenter et al. (2017), Ω will have a Lewandowski–Kuwowicka–Joe (LKJ) prior and τ a half-Cauchy prior as follows

- $\sigma_{\beta,p} \sim \text{half-Cauchy}(0, \tau_\beta)$, $p = 1, \dots, P$,
- $\Omega_\beta \sim \text{LKJCorr}(\eta_\beta)$.

It has been shown that a half-Cauchy prior for the random-effects standard deviation results in desirable estimates in the case of multilevel data with few clusters (Gelman and Hill 2006). Furthermore, by setting very large prior scale parameters, we can obtain approximately flat priors for the standard deviations. Further note that a half-Cauchy prior for the standard deviation corresponds to a F distribution on the

variance (Mulder and Pericchi 2018), which is a common distribution for modeling variance components.

The LKJ prior is defined as $\text{LKJCorr}(\mathbf{\Omega}|\eta) \propto \det(\mathbf{\Omega})^{\eta-1}$, with $\eta \in \mathbb{R}^+$. This distribution allows us to sample uniformly from the space of positive definite correlation matrices and has a behavior similar to the beta distribution (Wang et al. 2018; Lewandowski et al. 2009). For example, when $\eta = 1$, it has a uniform behavior, when $\eta < 1$ it favors stronger correlation; whereas when $\eta > 1$, it favors weaker correlation. For the random-effects covariance matrix under the sender $\mathbf{\Sigma}_\gamma$, the same prior is specified.

Finally, multivariate normal priors are specified for the fixed effects and random effects, i.e.,

$$\begin{aligned}\boldsymbol{\phi} &\sim N(\mathbf{0}, \mathbf{\Psi}_\phi), \\ \boldsymbol{\psi} &\sim N(\mathbf{0}, \mathbf{\Psi}_\psi), \\ \boldsymbol{\mu}_\beta &\sim N(\mathbf{0}, \mathbf{\Psi}_\mu), \\ \boldsymbol{\zeta}_\gamma &\sim N(\mathbf{0}, \mathbf{\Psi}_\zeta).\end{aligned}$$

For the prior covariance matrices, diagonal matrices are specified with very large variances. This builds upon existing knowledge of weakly informative prior specification for mixed effects generalized linear models, which have become the standard approach under this class of mixed effects models (Gelman et al. 2013; Gelman and Hill 2006).

2.3 Reparameterization to improve Bayesian computation

Due to the hierarchical structure of the data, the random-effects parameters $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_k$ are highly correlated with the population parameters $\boldsymbol{\mu}$, $\boldsymbol{\zeta}$ and $\mathbf{\Sigma}$. This introduces severe computational inefficiencies in the sampling process. When the data are sparse, which is a characteristic of most social network data, the geometry of the posterior distribution makes it very difficult to sample from the highest posterior density areas. Betancourt and Girolami (2015) called these issues *pathologies of the hierarchical model*. Therefore, to ease the burden on the sampler, we take advantage of the multivariate normal structure of the random effects and apply a non-centered linear transformation to those parameters.

Lemma 1 *Let $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, where $\boldsymbol{\beta} \in \mathbb{R}^p$. Then, with $\boldsymbol{\mu} \in \mathbb{R}^p$ and \mathbf{A} being a $p \times p$ matrix, such that $\mathbf{A}\mathbf{A}' = \mathbf{\Sigma}$, one can write $\boldsymbol{\beta} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$, where $\mathbf{Z} \in \mathbb{R}^p$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where \mathbf{I} is a $p \times p$ identity matrix.*

This transformation can be applied to both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. A natural candidate for matrix \mathbf{A} is the Cholesky factorization of the covariance matrix $\mathbf{\Sigma}$. For details see appendix B.

This reparameterization is more efficient for two reasons. First, it reduces the dependency between the random-effects parameters and the population parameters by sampling from independent standard normal distributions. This simplifies the geometry

of the posterior and avoids inverting Σ at every evaluation of the multivariate normal density (Carpenter et al. 2017). Therefore, we can safely and efficiently transform the random-effects parameters without causing any change in the prior specification of population parameters.

The model has been implemented in Stan, a probabilistic programming language that employs Hamiltonian Monte Carlo (HMC) algorithms to sample from posterior distributions. The advantage of HMC methods is that they avoid the random walk behavior and the sensitivity to posterior correlations that plague many Bayesian applications (Hoffman and Gelman 2014), including hierarchical models. This particularities allow HMC to generally converge to high dimensional distributions much faster than Metropolis–Hastings or Gibbs sampler methods (Hoffman and Gelman 2014; Betancourt and Girolami 2015).

3 Bayesian hypothesis testing under the actor-oriented multilevel relational event model

In multilevel analysis, researchers are usually interested in testing which theories receive the most support from the observed data, so that inferences about the population can be conducted. This kind of analysis is carried out through a process called hypothesis testing. From a Bayesian perspective, those tests are usually performed by the computation of Bayes factors (BF) (Jeffreys 1961). The BF is given by the ratio of marginal likelihoods under the parameter space of competing hypotheses, which quantifies the probability of observing the data under one hypothesis relative to another hypotheses, and thereby providing a quantification of the relative evidence in the data between the hypotheses. Let \mathbf{E} be the observed data, θ a vector of parameters in the space Θ , and $H_0 \in \Theta_0$ be a hypothesis that will be tested against $H_1 \in \Theta_1$, then BF_{01} is expressed as

$$BF_{01} = \frac{m(\mathbf{E}|H_0)}{m(\mathbf{E}|H_1)}, \tag{4}$$

where $m(\mathbf{E}|H_i) = \int_{\theta_i \in \Theta_i} p(\mathbf{E}|\theta_i)p(\theta_i)d\theta_i$, for $i = 0, 1$, with $p(\mathbf{E}|\theta_i)$ being the likelihood and $p(\theta_i)$ the prior. Also, $\Theta_0 \cap \Theta_1 = \emptyset$, with both Θ_0 and Θ_1 being subsets of Θ . Kass and Raftery (1995) provide a rule-of-thumb for Bayes Factors interpretation. In their setting, the evidence provided by the BF_{01} in favor H_0 can be seen as “insufficient” if $1 < BF_{01} < 3$, “positive” if $BF_{01} > 3$, “strong” if $BF_{01} > 20$, and “very strong” if $BF_{01} > 150$. These are rough guidelines to aid the interpretation of Bayes factors and should not be used as strict cut-off values.

Furthermore, when prior probabilities of the hypotheses have been formulated before observing the data, the prior odds can be updated with the Bayes factor to obtain the posterior odds of the hypotheses, i.e.,

$$\frac{P(H_0|\mathbf{E})}{P(H_1|\mathbf{E})} = BF_{01} \times \frac{P(H_0)}{P(H_1)}.$$

For example, when both hypotheses are equally likely a priori, i.e., $P(H_0) = P(H_1) = \frac{1}{2}$, the posterior probabilities of the hypotheses are given by $P(H_0|\mathbf{E}) = \frac{BF_{01}}{BF_{01}+1}$ and $P(H_1|\mathbf{E}) = \frac{1}{BF_{01}+1}$. These posterior probabilities quantify the probabilities that each of these hypotheses are true after observing the data (when assuming that one of these hypotheses is true). Because Bayes factors are consistent under mild conditions, the evidence for the true hypothesis goes to infinity and the posterior probability of the true hypothesis goes to 1 as the sample goes to infinity.

Below, we introduce Bayes factors for three types of hypothesis testing problems for the multilevel relational event modeling literature. Test I is useful for testing the homogeneity of network effects across independent sequences. Test II is useful for testing the degree of heterogeneity of network effects across sequences using order constraints on random-effects variances. Test III is useful to test equality and order constraints on fixed effects. Prior specification and numerical computation, which are important aspects when computing Bayes factors and posterior probabilities of the hypotheses, are discussed for each test separately.

3.1 Test I: testing for homogeneous social interaction behavior across sequences

When building multilevel relational event models, a central question is whether a driver of specific social interaction behavior (as quantified via the network effects) is constant over the sequences or whether it varies across the sequences. Testing this is important from a statistical point of view (i.e., to keep the model as parsimonious as possible, and thus to avoid an enormous overparameterization using different effects across all K sequences) but also from a substantive point of view (i.e., to understand which drivers of social interaction behavior in relational event data are constant across sequences and which (highly) differ). Thereby, testing this contributes to a better understanding of the heterogeneity of social interaction behavior.

The hypothesis test for the p th random network effect in the receiver model can be formulated as $H_0 : \sigma_{\beta,p} = 0$ versus $H_1 : \sigma_{\beta,p} > 0$, or equivalently as, $H_0 : \beta_{p,1} = \dots = \beta_{p,K}$ versus $H_1 : \text{not } H_0$. The second formulation of the hypothesis is simpler as we avoid the need of testing whether the variance is equal to the boundary of 0, but instead we only need to test whether the random effects are equal across the K clusters. Using the fact that the distribution of the random effects effectively serves as a prior for the random effects on the first level, we can simply compute a Savage–Dickey density ratio using the estimates of the random-effects distribution from the data, i.e.,

$$BF_{01} = \frac{\pi(\xi_1, \dots, \xi_{K-1} = 0|\mathbf{E})}{\pi(\xi_1, \dots, \xi_{K-1} = 0)},$$

where $\xi_{p,k} = \beta_{p,k} - \beta_{p,k-1}$, for $k = 2, \dots, K$, are the differences between the random effects.

To compute the posterior and prior densities at the null value, we use the fact that our actor-oriented multilevel relational event model can be written as two

independent mixed effects Poisson regression models having specific forms (Appendix A). Thus, the two independent components of the model belong to the family of generalized linear mixed effects models, which is well understood in the statistical literature (e.g., Gelman and Hill 2006; Bürkner 2017). Moreover, the model results in unimodal posteriors when using weakly informative priors, as done in this paper. Additionally, we use the general property that posterior distributions converge to Gaussian distributions in well-behaved problems following large sample theory (Gelman et al. 2004, Ch. 4; Rafter 1995; Kim and Ibrahim 2000). We illustrate the accuracy of this Gaussian approximation later in this paper. The Gaussian approximation of the posterior is then used to compute the numerator (then having an analytic expression). Moreover, the prior distribution, which is needed to compute the denominator, follows a multivariate Gaussian distribution by definition because it is the second level of our multilevel model, $\beta_k \sim N(\mu_\beta, \Sigma_\beta)$, and thus the joint prior distribution of the contrasts (ξ) also follows a multivariate Gaussian distribution.

One interesting aspect of this Bayes factor is that the prior is fully determined from the data, similar to empirical Bayesian estimation of hierarchical models. This property is especially useful here as Bayes factors are known to be sensitive to the choice of the prior. Note that empirical Bayesian approaches to obtain Bayes factors have been proposed for testing regression coefficients using *g* priors (e.g., see Liang et al. 2008, and the references therein), but (to our knowledge) not for testing variance components as we do here. The advantage of this approach is that no external prior information is required and no other ad hoc choices for prior distributions are needed. The outcome of the test can be used to quantify the relative evidence in the data of whether a hierarchical structure for the network effects is applicable or not given the observed data.

3.2 Test II: testing the degree of heterogeneity of network effects across sequences

After establishing which network effects are heterogeneous across sequences (using the test from the previous subsection), it is useful to investigate the relative degree of heterogeneity of network effects across sequences, again with the goal to better understand the heterogeneity of social interaction behavior across independent relational event sequences. This comes down to testing whether a specific random effect varies more across sequences than another random effect, or, equivalently, whether one random-effect variance is larger than another. When generalizing this further, we would test specific orderings of random-effect variances (see also Böing-Messing et al. 2020, who consider testing order constraints on variances in a nonhierarchical setting).

Two different tests are proposed for this purpose. First, a confirmatory test is proposed of whether an anticipated ordering regarding the degree of heterogeneity of network effects across sequences is present:

$$H_0 : \sigma_1 < \sigma_2 < \dots < \sigma_P$$

$$H_1 : \text{not } H_0.$$

Second, an exploratory test is proposed to determine which ordering out of all $P!$ possible orderings receives most evidence from the data, i.e.,

$$H_1 : \quad \sigma_1 < \sigma_2 < \dots < \sigma_P$$

$$H_2 : \quad \sigma_2 < \sigma_1 < \dots < \sigma_P$$

$$\vdots$$

$$H_{P!} : \quad \sigma_P < \sigma_{P-1} < \dots < \sigma_1.$$

Following the existing Bayesian literature on order-constrained hypothesis testing (e.g., Klugkist et al. 2005), we specify the prior under an order-constrained hypothesis as a truncated version of an unconstrained prior truncated in the order-constrained subspace. The Bayes factor of an order-constrained hypothesis against the unconstrained hypothesis, H_u , can then be expressed as the ratio of the posterior and prior probability that the constraints hold (Appendix C):

$$\text{BF}_{1u} = \frac{P(\sigma_1 < \sigma_2 < \dots < \sigma_P | \mathbf{E}, H_u)}{P(\sigma_1 < \sigma_2 < \dots < \sigma_P | H_u)}. \quad (5)$$

The Bayes factors between constrained hypotheses of interest can then be obtained using the transitive property of the Bayes factors, e.g., $B_{12} = B_{1u}/B_{2u}$.

The Bayes factor will not be sensitive to the prior as long as very vague identical priors are specified for the variances. In this case, the prior probability in the denominator will be equal to $(P!)^{-1}$, and the posterior probability will be fully determined by the information in the data (similar as in Bayesian estimation). Bartlett's paradox is, thus, not an issue in Bayesian order or one-sided hypothesis testing (Jeffreys 1961; Klugkist and Hoijtink 2007; Liang et al. 2008; Mulder 2014a).

To compute the posterior probability that the order constraints hold under the unconstrained model, we can simply fit the unconstrained model (using independent vague priors) and compute the proportion of draws satisfying the order constraints. The current test extends the use of Bayes factors for testing order hypotheses on variance components (Böing-Messing et al. 2017; Mulder and Fox 2019; Böing-Messing et al. 2020, e.g.,) to multilevel relational event models.

3.3 Test III: testing common and average network effects over all sequences

In most applications, hypothesis tests are formulated on the common network parameters across clusters (i.e., the fixed effects) and on the average of the random-effects across clusters (i.e., the global random-effects means). Since both parameters have the same role, hypothesis tests of these parameters are both discussed here using the same methodology.

First, a common test is of whether a network statistic has no effect, a negative effect, or a positive effect on the interaction behavior, which could be formulated as

$$H_0 : \psi_1 = 0 \text{ versus } H_1 : \psi_1 < 0 \text{ versus } H_2 : \psi_1 > 0$$

for the first fixed effect of the receiver model, for instance. Second, hypotheses are often formulated on combinations of parameters using equality and/or order constraints on multiple parameters of interest based on existing theories or scientific expectations (Hojtink et al. 2019), e.g.,

$$H_1 : \psi_1 = \psi_2 < \psi_3 \text{ versus } H_2 : \text{‘not}H_1\text{’},$$

where hypothesis H_1 assumes that the first effect (ψ_1) is equal to the second effect (ψ_2) and larger than the third effect (ψ_3), and hypothesis H_2 assumes the complement is true. Third, when testing the effects of categorical (dummy) variables on the event rate, scientific expectations may be formulated which categorical variable has the largest impact on the event rate. A challenging testing problem is when the interest is in assessing which categorical variable has the largest impact when it is not known which category of each categorical variable results in the largest event rate. For example, it may be of interest whether the dichotomous gender variable (0 or 1) has a larger, equal, or smaller impact on the event rate than a dichotomous race variable (0 or 1), but it is not of particular interest which category results in the largest rate. This could be translated to constrained hypotheses on the absolute values of the effects of these two categorical variables according to

$$H_0 : |\psi_1| = |\psi_2| \text{ versus } H_1 : |\psi_1| < |\psi_2| \text{ versus } H_2 : |\psi_1| > |\psi_2|.$$

Finally, it is possible to test parameters between the sender model and the receiver model. This allows researchers to assess whether variable has a larger/equal/smaller to predict the next sender than to predict the next receiver.

The procedure for Test III builds on default Bayes factor methodology where the information in the data is split between a minimal subset, which is used for default prior specification (in combination with a noninformative prior) and a maximal subset which is used for hypothesis testing (O’Hagan 1995; Pérez and Berger 2002, among others). When using these methods, the resulting Bayes factors do not depend on the undefined normalizing constants of the improper priors (O’Hagan 1995), which practically means that arbitrarily vague priors can be used, as we do in this paper. These Bayes factors can be computed in an automatic fashion and manual prior specification can be avoided.¹ Here, we use fractional Bayes factors where a (minimal) fraction of the data, denoted by b , is used to construct a (default) fractional prior, given by

$$\pi(\phi, \psi, \mu, \zeta, \Sigma_\beta, \Sigma_\gamma | E^b) \propto p(E | \phi, \psi, \mu, \zeta, \Sigma_\beta, \Sigma_\gamma)^b \pi(\phi, \psi, \mu, \zeta, \Sigma_\beta, \Sigma_\gamma), \quad (6)$$

¹ Alternatively, JZS priors (Bayarri and García-Donato 2007; Rouder et al. 2009; Wetzels et al. 2012) could also be considered for this test but, to our knowledge, this class of priors has not yet been proposed for generalized linear mixed effects models. Moreover, note that the prior scale of the key parameter that is tested still needs to be manually specified, unlike the proposed fractional Bayes factor using a minimally informative fractional prior.

and the remaining fraction is used for hypothesis testing. It has been shown that fractional Bayes factors follow many important properties which are not shared by all default Bayes factors, such as large sample consistency, satisfying the likelihood principle, invariance to transformations of the data (O'Hagan 1997), and information consistency (Mulder 2014b). To properly incorporate the model complexity of order-constrained hypotheses, we apply a prior adjustment of the fractional prior (Mulder 2014b), and Gaussian approximations are applied to the posterior and fractional prior to simplify the computation (Kim and Ibrahim 2000; Gu et al. 2017). The accuracy of the Gaussian approximations will be discussed in the next section. The approximated fractional Bayes factor of a constrained hypothesis (e.g., with any set of equality and/or constraints on certain parameters) against an unconstrained model is then computed as the integral over the unconstrained posterior over the constrained subspace divided by the integral over the unconstrained fractional prior over the constrained subspace as a Savage–Dickey density ratio (Mulder 2014b). Existing functions in R using the 'mvtnorm' package (Genz et al. 2021) can be used to compute the Bayes factors (Mulder et al. 2021).

The choice of the fraction 'b' is based on the recommendation by Mulder and Fox (2019), which implies selecting a minimal sample that is based on the ratio between the total number of parameters (excluding group specific effects) and the total number of observations. Since sender and receiver models are two different processes assumed to be conditionally independent given the past, we have to define two difference fractions, one for each model (Hojtink et al. 2019, see also). In the sender model, we have P random effects and Q fixed effects, hence the fraction is $b_{\text{Snd}} = \frac{(P(P+1)/2)+(P+Q)}{\sum_{k=1}^K N_k}$ and in the receiver model we have V random effects and U fixed effects, resulting in the fraction $b_{\text{Rec}} = \frac{(V(V+1)/2)+(V+U)}{\sum_{k=1}^K N_k}$.

4 Estimating and testing classroom dynamics using the multilevel relational event model

The data that are considered here were collected by McFarland (2001) in a study to investigate student rebellion in the classroom. The data feature observations of interactions among high-school students in two different schools in the United States. For this illustration, we consider 15 independent classrooms from Magnet High School during the 1996–1997 school year. The student body of this high school can be considered academically homogeneous. The data were collected through classroom observations in which the conversations within the classroom were coded. In each of these fifteen classes, a teacher is present in addition to the students. The number of events (the number of times one person said something to another person) ranges from 86 to 628, and the number of persons (the students plus the teacher) ranges between 19 and 30 across the event sequences. The conversations happened in an orderly lecture-like fashion, so only one person was speaking at each time. The aim of this application is twofold. At first, we present a sequential analysis that is used as proof of concept for

the proposed model and statistical testing procedures. Second, we aim to provide insights into classroom dynamics by showing the evolution of network effects over time and extensively applying our tests to showcase multiple substantively interesting hypothesis tests. Below, we first discuss the actor-oriented multilevel relational event model that we use to analyze the data.

4.1 Model specification

We fit the hierarchical actor-oriented relational event model (described in Sect. 2) to the data. Initially, all effects are considered random, so the need for a hierarchical structure can be evaluated using the Bayes factor presented in Sect. 3. This first step will allow us to determine whether some effects can be treated as fixed and the model can hence be simplified.

We set the prior parameters to $\sigma_{\phi}^2 = \sigma_{\mu_{\rho}}^2 = \sigma_{\psi}^2 = \sigma_{\zeta_{\tau}}^2 = \sigma_{\tau} = 10$, making the priors for the fixed and random effects relatively vague, and $\eta = 2$, slightly favoring smaller correlations. For our illustration of the proposed methods, we will include the following covariates into the model:

Teacher: A dummy variable that indicates whether the actor is the teacher (one if the actor is the teacher and zero otherwise).

Gender: A dummy variable that indicates the gender of the actor (one if the actor is male and zero otherwise).

Race: A dummy variable that indicates the race of the actor. McFarland (2001) notes that 50 percent of the *Magnet High* population is Caucasian. Hence, the variable is one if the actor is Caucasian and zero otherwise.

Inertia: Inertia captures the persistence of the communication, where past interaction is likely to be repeated (Leenders et al. 2016). It is computed as the accumulated volume of past communication from a specific sender to a specific receiver. Because this statistic is dyadic, it will be included only in the receiver model.

Participation shifts: These statistics are used to reflect expectations of adherence of communication norms in small groups (Butts 2008). Gibson (2005) describes the framework for several types of participation shifts. In this application two distinct types of participation shifts will be included. The first belongs to the group of “*turn-receiving*” and is represented by the event pattern **ABBA**: an interaction from person A to person B is immediately followed by an interaction from B to A. The second is **ABAB**, which can be considered as a special case of “*turn-continuing*”: an interaction from person A to B is immediately followed by another interaction event from A to B.

The ABBA and ABAB participation shifts are concerned with dyads, therefore they will be included in the receiver model only. However, we do include adapted versions in the sender model. Here, the statistics become **ABB** (after A has spoken to B, the next event is B starting the conversation) and **ABA** (after A has spoken to B, A speaks again).

Activity: This set of covariates captures the effect of actor activity as a sender or as a receiver (Vu et al. 2017). The first is the *outgoingness* of a person, defined as

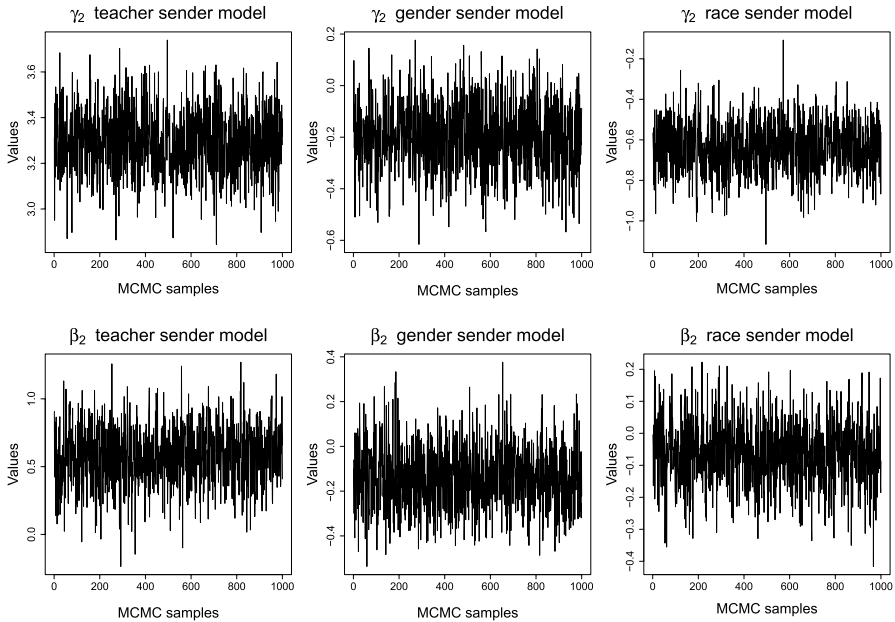


Fig. 1 Trace plots of some of the random-effect parameters. The number 2 in the subscript shows that the parameters belong to cluster number two. The top (bottom) line shows parameters in the sender (receiver) model

the number of events sent by one actor up until a specific point in time. This captures the tendency of the person to start conversations (or just to talk). The second is the *popularity* of a person, given by the number of events received by one actor up until a specific point in time. This captures the popularity of an individual as a receiver of the conversation.

Relational event models are vulnerable to process explosion, which happens when $\lambda(t) \rightarrow \infty$ often due to a feedback loop that may be caused by using statistics that are computed as cumulative sums (Aalen et al. 2008). This is particularly a problem for inertia and activity statistics. One way to alleviate this problem is via z -score standardization at every time point, which is defined as $z(t) = (x(t) - \bar{x}(t))/S_x(t)$, where $x(t)$ is the value of the statistic, $\bar{x}(t)$ is the sample mean, and $S_x(t)$ is the sample standard deviation at time t . We use standardized statistics in this application.

4.2 Exploring social interaction dynamics as class time progresses

By analyzing the relational event sequences as the time during class progresses, we can explore how statistical certainty increases over time using the proposed multi-level relational event model, and how interaction dynamics between children and the teacher evolves as class time progresses. To study this, we analyze the sequences

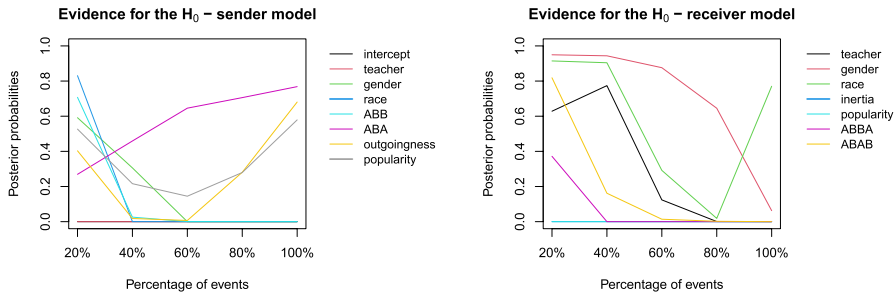


Fig. 2 Posterior probabilities of the fixed-effect hypothesis H_0 in the homogeneity test for different effects

with increasing batches of 20% of the total number of events in each sequence (i.e., 20%, 40%, 60%, 80%, and 100%).

4.2.1 Testing for homogeneous social interaction behavior across school classes

The first step is to fit the model with all effects considered random. We ran 2000 MCMC iterations and discarded the first 1000 samples as burn-in. We used \hat{R} as a convergence diagnostic measure. For all models, the metric values were fairly close to 1 all, falling all below the 1.05 value recommended by Carpenter et al. (2017). As an example, Fig. 1 shows several trace plots for the model run with 100% of the events. The chains show convergence. The number of posterior draws is not high because the Hamiltonian dynamics in Stan’s algorithm allow for faster convergence than standard MCMC methods (Hoffman and Gelman 2014). To check this, we investigated the bulk effective sample size and the tail effective sample size, which fell between 800 and 1100 for all parameters (Carpenter et al. (2017) recommends these values to be both above 100 for the sample to be considered reliable). This confirms the computational efficiency of the sampler. For other sample size recommendations, see Hecht and Zitzmann (2021). Since a more parsimonious model is preferred, we consider an effect to vary across classrooms via a random effect when the posterior probability for the hypothesis that an effect is fixed, H_0 , is less than 0.25, corresponding to $BF_{01} < 1/3$ (which can be interpreted as positive evidence that the effect varies across clusters; Kass & Raftery, 1995).

Figure 2 shows the evolution of the posterior probabilities of H_0 as we increase the number of events. In the beginning, when the number of events is small (varying from 17 to 125 events across school classes), the uncertainty in the posterior is relatively large leading to considerable overlap in the posterior distributions of the random effects. Therefore, the posterior probabilities for a fixed effect tend to start out large and eventually decrease to zero when the full sequences are considered for most effects, providing strong evidence for their heterogeneous nature. This is the case for most effects that are clearly heterogeneous if the sample sizes are large enough. On the other hand, for some covariates, the posterior probability gets larger as the sample sizes near 100% of the events. This is the case for **ABA**,

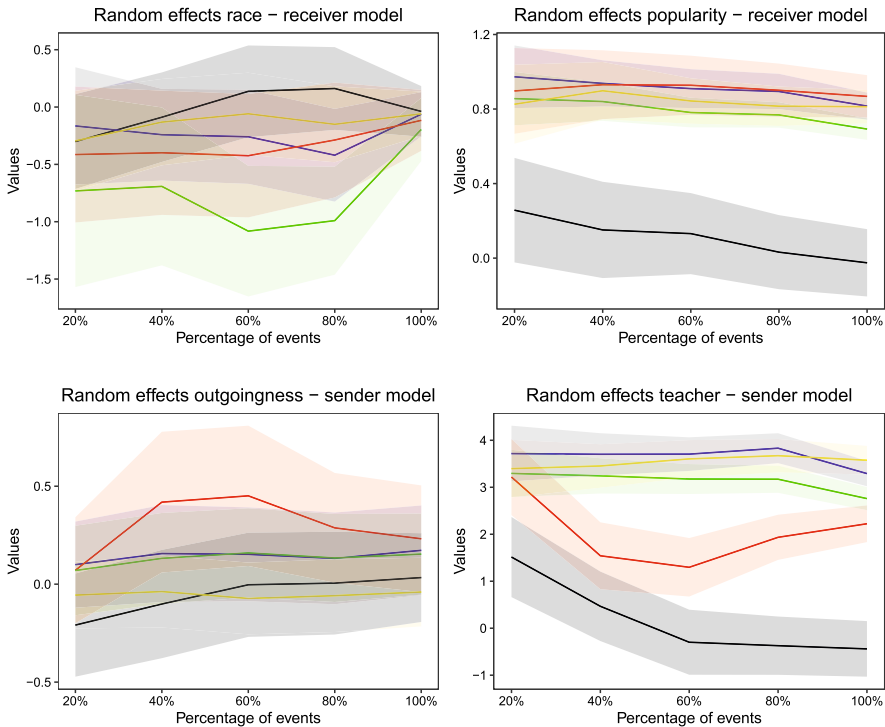


Fig. 3 Random-effect estimates for receiver (top) and sender (bottom) models for 5 groups

outgoingness and *popularity* in the sender model and for *race* in the receiver model, resulting in posterior probabilities for a fixed effect of 0.768, 0.681 and 0.579, 0.777, respectively. Thus, the parameters corresponding to those effects will be fixed in our model. Therefore, we continue with a mixed-effect model, where the sender model has *ABA*, *outgoingness*, and *popularity* as fixed effects, and *intercept*, *teacher*, *gender*, *race*, and *ABB* as random effects, and a receiver model where only *race* is a fixed effect and all other effects are considered random.

To understand the nonmonotonic development of the posterior probabilities where the lines first go down and eventually increase, it is useful to check the corresponding estimates of the random-effects. As an example, the top left panel of Fig. 3 shows random-effect estimates of the *race* covariate in the receiver model, where each line represents a five different classrooms (to keep the plot clear; the other classrooms showed similar patterns). As the lines show we see that the random effects become more heterogeneous until 80% of the events are included in the analysis but when all 100% of the events are included the estimates become more homogeneous which explains the evidence for a fixed effect for the race effect in the receiver model in Fig. 2 when all 100% of the events are considered.

In addition, in the bottom right panel of Fig. 3, which presents the random-effect estimates for *outgoingness* in the sender model, a similar pattern is observed. The random effects becomes heterogeneous between 40% and 60% of the events, but they rapidly converge displaying a pattern similar to *race* in the sender model. This behavior is reflected in the posterior probabilities, with the line going down for 40% and 60% of the events and then getting back up again. On the other hand, the estimates on the left-hand side panels of Fig. 3, which display *popularity* in the receiver model and *teacher* in the sender model, are consistently heterogeneous as class time progresses. This is corroborated by the posterior probabilities with both effects presenting small evidence for the null.

4.2.2 Evaluating the shrinkage behavior of random effects

Given that there is asymmetry in the distribution of the number of events across classrooms (varying from 86 to 628 events over classrooms), there will be a variation regarding the statistical uncertainty of the estimated random effects across classrooms. Due to the multilevel structure of the model, estimated effects with a relatively large uncertainty which also deviate much from the other estimates will then be (slightly) pulled towards the global mean of the effect over all class rooms. This statistical behavior is also known as shrinkage, or borrowing information across groups, which is an intrinsic property of multilevel models. The current sequential analysis where the sample sizes grows allow us to investigate this behavior, since with fewer events we should have less information and, therefore, less accurate classroom-specific estimates.

We use the R package *remestimate* (Arena et al. 2022) to obtain the estimates of the classroom-specific effects which ignores the multilevel structure of the data. Figure 4 shows the shrinkage for several effects in the sender and receiver models. The lines were plotted using a gray scale where lighter shades mean smaller sample sizes. In each panel, we have the estimates based on the estimation of the separate classrooms at the top, and the multilevel estimates at the bottom.

First, we see that there is considerably more variation of the estimated effects in the separate analyses (top of each panel), which ignore the multilevel structure, in comparison to the multilevel estimates (bottom of each panel). This confirms that the proposed multilevel model shows the anticipated shrinkage behavior. Second, we see that when using subsets of only 20% complete event sequences in the classrooms, we obtain more extreme estimates than when using the proposed multilevel model. Finally, we see that estimates for classrooms with shorter event sequences (represented with light gray lines) are on average more extreme than estimates for classrooms for longer event sequences. This illustrates that the proposed multilevel relational event models is particularly useful when the event sequences are relatively short to avoid extreme estimates.

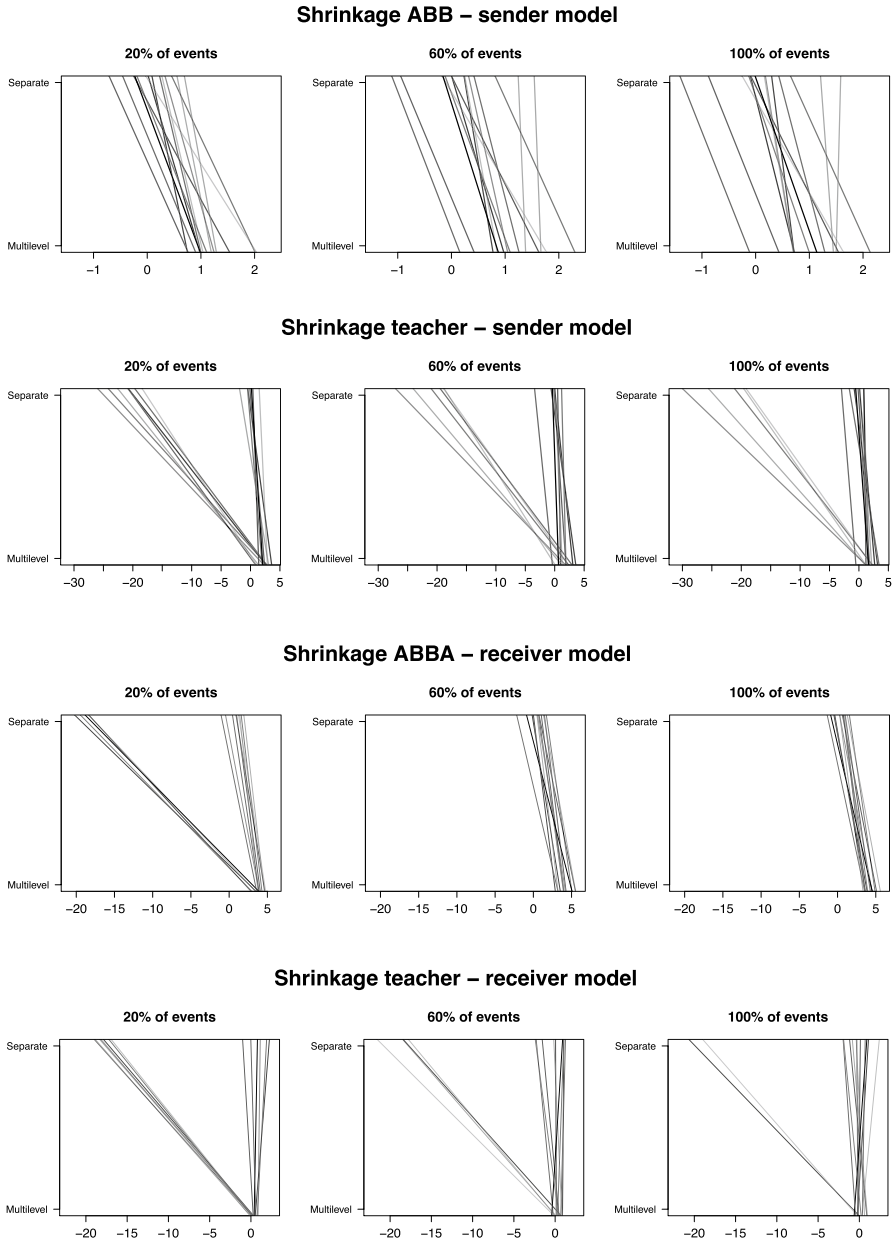


Fig. 4 Plots of classroom-specific effects when fitting a model separately per classroom and when jointly fitting the multilevel model over all classrooms. Darker (lighter) lines indicate classrooms with relatively many (few) events

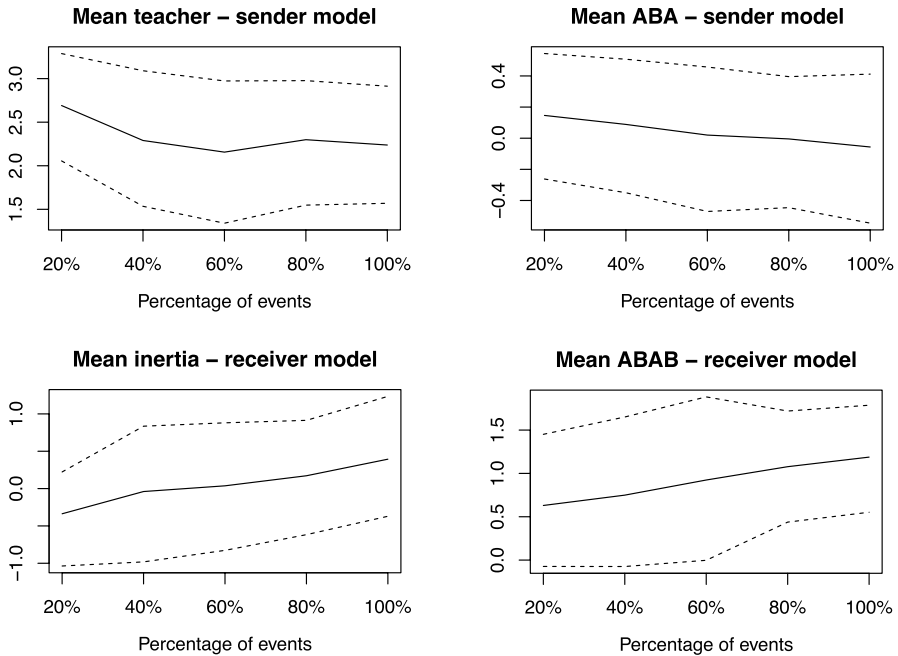


Fig. 5 The development of different drivers of interaction behavior as class time progresses from 20% to 100% of the observed events in each classroom

4.2.3 Evolution of classroom behavior

Another interesting aspect to look at in the sequential analysis is the potential change of the estimated of the effects as class time progresses. This will give us an insight into the social interaction dynamics in classrooms. Figure 5 shows the evolution of some mean effects in the sender and receiver models. Here, we restrict ourselves to some trends of average network behavior across all classrooms. Note that the intervals are relatively wide because the limited lengths of the event sequences in a class period.

First, we see that the teacher effect in the sender model is positive and large in the sender model. This was expected given the dominant role of the teacher during lectures. The plot in Fig. 5 (upper left panel) also shows a slight decrease as class time progresses. This suggests that the dominant role of the teacher slightly decreases as the lectures approach the end. This indicates that teachers may lose some of the control towards the end or that more students become engaged or activated towards the end. This is also confirmed by the ABA effect in the sender model (Fig. 5, upper right panel) which is initially positive and decrease to just below zero towards the end of the class period. Thus, it becomes less likely that the sender of the previous messages (typically the teacher because these are lectures) becomes the sender of the next event.

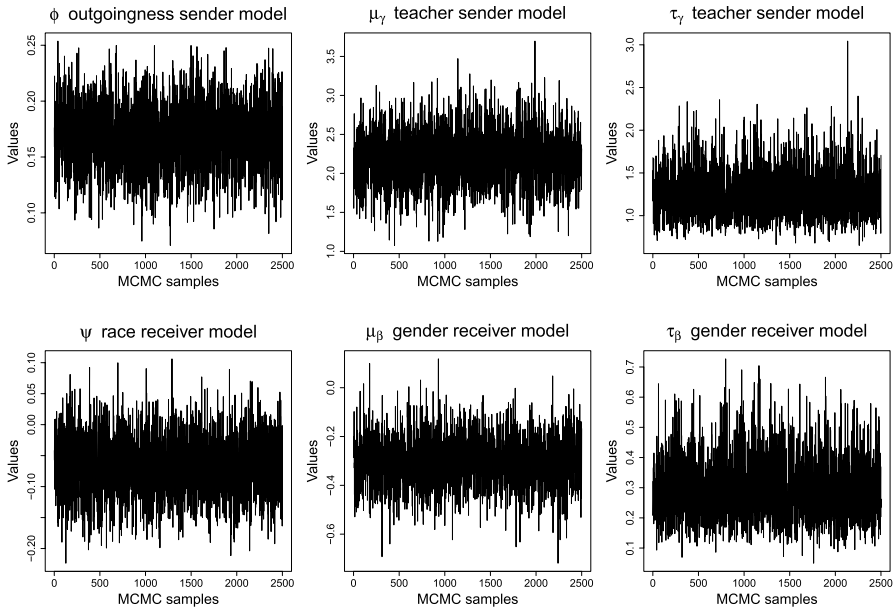


Fig. 6 Trace plots of some of the population parameters in the mixed-effects model. The top (bottom) line show parameters in the sender (receiver) model. The figure display one fixed effect and one random-effect mean with its variance for each model

Another interesting trend can be observed for the global inertia effect across classrooms in the receiver model (Fig. 5, lower left panel) In the beginning of the class period, inertia starts out negative which indicates that it is less likely that actors who were the receiver of relational events in the past (typically a question or remark by the teacher towards a student because lectures are considered) become a receiver of the next. This indicates that teachers aim to address different students in the beginning of the lecture with the goal to get more students engaged or activated during lectures. The effect, however, gradually increases as the class time progressed. As the class period nears the end, the inertia effect becomes positive which implies that it has become more likely that a student is addressed by the teacher who was also addressed by the teacher in the past. This suggests that teachers tend to focus more on the same students (possibly the more motivated students) rather than trying to involve other less active students in the discussion. This is also confirmed by the positive and increasing ABAB effect in the receiver model during class time (Fig. 5, lower right panel).

4.3 Mixed-effects analysis on the full sequences

In this subsection, we carry out the other proposed tests under the mixed-effects model obtained from the homogeneity tests that were carried out in the previous

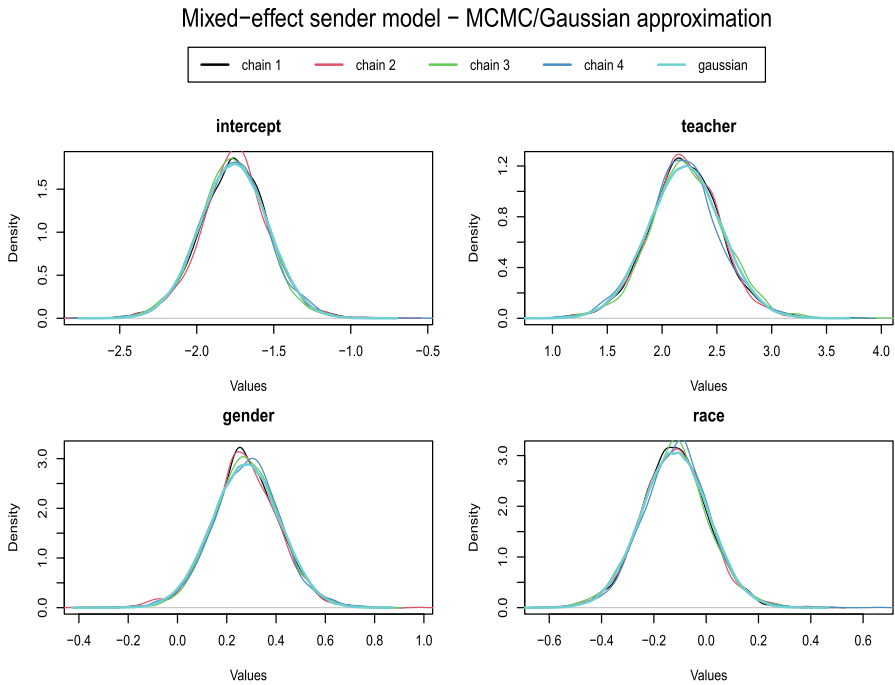


Fig. 7 Estimated posterior density and Gaussian approximation for a some fixed effects in the sender model

section. We only restrict ourselves to the results for the full event sequences to keep the discussion of the results as concise as possible. We ran 4 chains, each with 5000 MCMC iterations with a burn-in of 2500 draws. Figure 6 shows the trace plots of a few parameters, all chains show convergence. Here, we also used the \hat{R} as a convergence measure, they were essentially 1 in all chains for all parameters. In addition, the bulk and tail effective sample sizes were all close to around 2000 samples.

Moreover, Figs. 7 and 8 display the density estimated from each of the 4 chains and a Gaussian approximation on top. As can be seen, the Gaussian approximation seems to be acceptable in the bulk as well as in the tails of the distribution. Therefore, empirically, the approach of conducting the tests using these approximations as proxies for the posterior distributions seems reasonable. From now on, given that the results for the four chains are virtually identical, we proceed to analyze the results of a single chain.

We can see that the proposed testing procedure to determine which effects are random and which are fixed works correctly because a similar fit to the data is obtained in the more parsimonious model (where some effects are fixed) compared to the larger (i.e., all random) model. We can see this by computing the point-wise deviance residuals for both models under the full multilevel model where all effects are random across classrooms and the mixed effects model where certain effects are assumed fixed across classrooms (according to the proposed Bayes factor test).

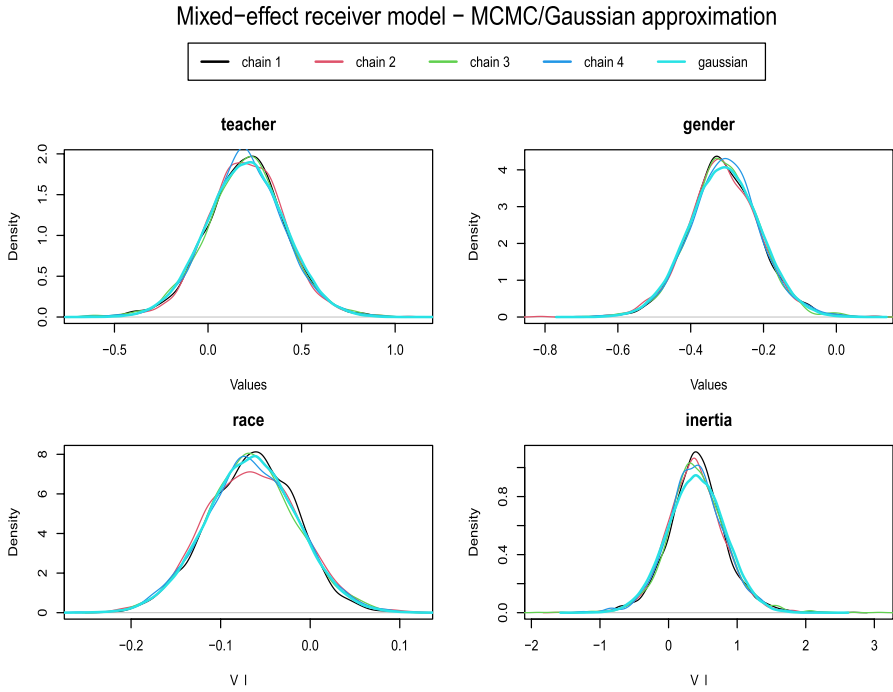
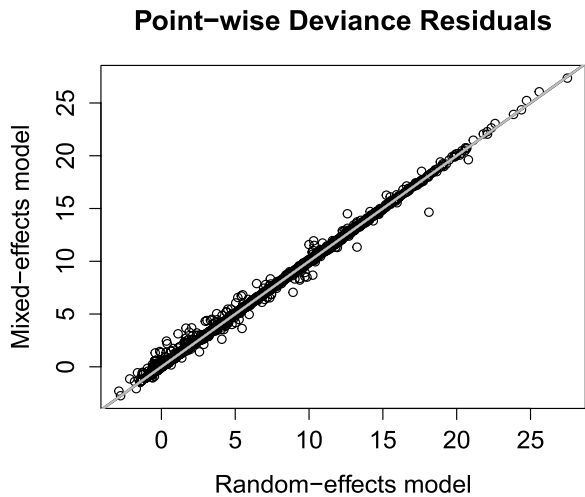


Fig. 8 Estimated posterior density and Gaussian approximation for a some fixed effects in the receiver model

Fig. 9 Comparison of the point-wise deviance residuals between the random-effects (x-axis) and the mixed-effects (y-axis) models. The gray line is the equality line



If a similar fit is obtained for both models, the point-wise deviance residuals for both models are similar meaning that none of the two models dominates the other in terms of fit. Figure 9 clearly shows that that is the case: most points lie around

Table 1 Parameter estimates of the mean effects

Parameter Estimates of Mean Effects				
Effect	Random-Effects Model		Mixed-Effects Model	
	Sender	Receiver	Sender	Receiver
Intercept	-1.75 (-2.08, -1.37)	-	-1.77 (-2.19, -1.32)	-
Teacher	2.22 (1.56, 2.83)	0.21 (-0.25, 0.67)	2.22 (1.60, 2.97)	0.20 (-0.21, 0.65)
Gender	0.27 (0.01, 0.53)	-0.30 (-0.51, -0.12)	0.29 (0.02, 0.57)	-0.31 (-0.49, -0.09)
Race	-0.12 (-0.36, 0.11)	-0.07 (-0.18, 0.05)	-0.12 (-0.37, 0.14)	-0.06 (-0.17, 0.04)
Inertia	-	0.42 (-0.38, 1.15)	-	0.36 (-0.46, 1.17)
ABB (ABBA)	1.06 (0.67, 1.46)	4.30 (3.88, 4.72)	1.08 (0.67, 1.52)	4.27 (3.79, 4.70)
ABA (ABAB)	-0.06 (-0.52, 0.33)	1.29 (0.57, 1.84)	-0.05 (-0.42, 0.32)	1.18 (0.53, 1.81)
Outgoingness	0.05 (-0.08, 0.20)	-	0.07 (0.00, 0.12)	-
Popularity	0.25 (0.12, 0.39)	0.58 (0.40, 0.77)	0.17 (0.11, 0.23)	0.59 (0.39, 0.78)

Numbers in bold are effects that were treated as fixed, either ϕ or ψ , in the mixed-effects model. Not-bolded numbers represent random-effect means, μ . Numbers between parentheses are 95% credible intervals. Dashes indicate that the covariate was not included in the respective model

the gray diagonal where the values are equal. This indicates that the fit provided by both models is similar. Details of the point-wise deviance computation are provided in appendix E. Finally, Table 1 shows a comparison of the posterior estimates for both models, where the bold ones are effects treated as fixed in the mixed-effects model. In sum, the mixed effects model results in a comparable fit but with much less parameters to estimate, resulting in a more parsimonious model and a simpler explanation of social interaction behavior across classrooms.

4.3.1 Testing the degree of heterogeneity of network effects across school classes

By testing order constraints on the random-effects variance parameters, we can assess which characteristics cause most variation in the social interaction behavior across school classes. In particular, we focus on the random effects of the *teacher*, *gender*, and *race* variables in the sender model (which are all dummy variables). The same test can also be applied to the variance parameters in the receiver model in a similar manner. The objective is to determine the amount of evidence regarding the level of heterogeneity among these effects. Let $\sigma^2_{teacher}$, σ^2_{gender} , and σ^2_{race} be the between classrooms variances of the *teacher*, *gender*, and *race* effects, respectively, where these parameters are extracted from the diagonal of Σ_γ . Given the special (dominant) role of the teacher in a classroom, and the fact that different teachers have different teaching styles, it is expected that the variance of the teacher effect is largest. We tested all six possible order hypotheses for the random-effects variances:

$$\begin{aligned}
 H_1 &: \sigma_{teacher}^2 > \sigma_{gender}^2 > \sigma_{race}^2 \\
 H_2 &: \sigma_{teacher}^2 > \sigma_{race}^2 > \sigma_{gender}^2 \\
 H_3 &: \sigma_{race}^2 > \sigma_{teacher}^2 > \sigma_{gender}^2 \\
 H_4 &: \sigma_{race}^2 > \sigma_{gender}^2 > \sigma_{teacher}^2 \\
 H_5 &: \sigma_{gender}^2 > \sigma_{teacher}^2 > \sigma_{race}^2 \\
 H_6 &: \sigma_{gender}^2 > \sigma_{race}^2 > \sigma_{teacher}^2.
 \end{aligned}$$

The goal is to determine which hypothesis receives most evidence from the data. Below, the Bayes factors for each one of these hypotheses against one another is represented in the form of an evidence matrix:

	H ₁	H ₂	H ₃	H ₄	H ₅	H ₆
H ₁	1.00	1.47	5788.40	615.79	5788.40	275.64
H ₂	0.73	1.00	4200.00	446.81	4200.00	200.00
H ₃	0.00	0.00	1.00	0.00	0.00	0.00
H ₄	0.00	0.00	8.40	1.00	8.40	0.40
H ₅	0.00	0.00	0.00	0.00	1.00	0.00
H ₆	0.00	0.00	20.00	2.13	20.00	1.0

Each cell of the matrix represent the comparison of the hypothesis in the rows against the hypothesis in the columns. So, for example, the evidence for H₁ against H₂ is equal to 1.5. The evidence for H₂ against H₁ is then 1/1.50 = 0.67.

The evidence matrix clearly shows that the evidence for hypotheses H₁ and H₂ is much larger than the evidence for any of the other hypotheses, which indicates that the *teacher* effect shows indeed most heterogeneity across the fifteen classrooms. When inspecting the estimates of the variances, this is also confirmed: $\bar{\sigma}_{teacher}^2 = 1.65$, $\bar{\sigma}_{gender}^2 = 0.25$, and $\bar{\sigma}_{race}^2 = 0.22$. Note that the added value of the Bayes factor complementary to eyeballing the estimates is that the effect sizes and their uncertainty are combined in a principled manner to determine which effects are most heterogeneous across classrooms. Finally note that there is no clear evidence that either the gender effect or the race effect is more heterogeneous across classrooms. Thus, we can conclude that teachers have the biggest impact on the variability of social interaction behavior across classrooms.

4.3.2 Testing the impact of nodal characteristics on classroom dynamics

In this subsection, we test the effects of nodal characteristics how they affect classroom dynamics. In the sender model, due to the special role of the teacher to control interaction behavior during lectures, the *teacher* variable is expected to have the strongest impact on the rate at which an actor starts an interaction when compared to the other two personal-trait variables *gender* and *race*. No expectations are formulated about which of these latter two variables has the largest effect in the model, or about which category of these dichotomous variables results in a positive effect.

This results in the following hypotheses where constraints are formulated on the absolute values of the respective effects of these variables

$$\begin{aligned}
 H_1 &: |\zeta_{teacher}| > |\zeta_{gender}| > |\zeta_{race}| \\
 H_2 &: |\zeta_{teacher}| > |\zeta_{gender}| = |\zeta_{race}| \\
 H_3 &: |\zeta_{teacher}| > |\zeta_{race}| > |\zeta_{gender}| \\
 H_4 &: \text{neither } H_1, H_2, \text{ nor } H_3.
 \end{aligned}$$

The complement hypothesis H_4 is included as a “safety net” in case our expectations about the dominance of the teacher variable would not be supported by the data. Once again, we present the resulting fractional Bayes factors in an evidence matrix form

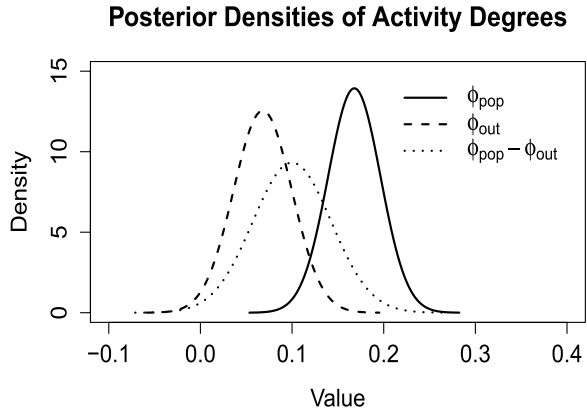
$$\begin{matrix}
 & H_1 & H_2 & H_3 & H_4 \\
 \begin{matrix} H_1 \\ H_2 \\ H_3 \\ H_4 \end{matrix} & \begin{pmatrix} 1.00 & 0.21 & 4.74 & 2.39e8 \\ 4.56 & 1.00 & 21.63 & 1.09e9 \\ 0.21 & 0.05 & 1.00 & 5.05e7 \\ 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix}
 \end{matrix}$$

Again, we translate these Bayes factors to posterior probabilities when assuming that each hypothesis is equally likely a priori, which results in posterior probabilities of 0.235, 0.723, 0.042, and 0.000, for H_1 , H_2 , H_3 , and H_4 , respectively. Based on these results we can safely rule out the complement hypothesis H_4 . Moreover, hypothesis H_2 (which assumes that the teacher variable has the largest effect, and the gender effect and the race effect are equal) receives most evidence; but only approximately 3 times more evidence than the hypothesis H_1 (which assumes that gender plays a larger role than race after the teacher effect). In order to draw more decisive conclusions of whether H_1 or H_2 is true more data would be required. These results are confirmed when looking at the posterior estimates, i.e., 2.22, 0.289, 0.143, and their posterior standard deviations, i.e., 0.344, 0.131, 0.103, of the absolute values of the teacher, gender, and race effect. Note again that the Bayes factors and posterior probabilities are a principled probabilistic methodology to summarize these findings without requiring subjective eyeballing of the estimates.

Under the receiver model, we consider a more exploratory approach by considering all possible order hypotheses on the absolute effects of these nodal characteristics. Note that even though the teacher still has a dominant role, this may not necessarily be the case in the receiver model. The following hypotheses are considered

$$\begin{aligned}
 H_1 &: |\mu_{teacher}| > |\mu_{gender}| > |\mu_{race}| \\
 H_2 &: |\mu_{teacher}| > |\mu_{race}| > |\mu_{gender}| \\
 H_3 &: |\mu_{race}| > |\mu_{teacher}| > |\mu_{gender}| \\
 H_4 &: |\mu_{race}| > |\mu_{gender}| > |\mu_{teacher}| \\
 H_5 &: |\mu_{gender}| > |\mu_{race}| > |\mu_{teacher}| \\
 H_6 &: |\mu_{gender}| > |\mu_{teacher}| > |\mu_{race}|,
 \end{aligned}$$

Fig. 10 Posterior distribution of $\phi_{\text{popularity}}$ (solid line) and $\phi_{\text{outgoingness}}$ (traced line) and their difference (dotted line)



where $|\mu|$ is the absolute value of μ , which represents the global mean of the random effects in the receiver model. We also display the results in the form of an evidence matrix which yields:

	H ₁	H ₂	H ₃	H ₄	H ₅	H ₆
H ₁	1.00	45.52	354.37	254.63	3.82	0.44
H ₂	0.02	1.00	7.78	5.39	0.08	0.01
H ₃	0.00	0.12	1.00	0.69	0.01	0.00
H ₄	0.00	0.18	1.44	1.00	0.02	0.00
H ₅	0.26	11.91	92.69	64.25	1.00	0.12
H ₆	2.28	103.89	888.80	560.61	8.73	1.00

and translating these to posterior probabilities result in the following posterior probabilities for the respective order hypotheses: 0.291, 0.006, 0.001, 0.001, 0.068, 0.633. These results indicate that almost all probabilities mass goes to H_1 and H_6 , which suggests that the race variable has the smallest impact. Furthermore, the gender variable is expected to play the largest role in the receiver model among these nodal characteristics. These results are confirmed when looking at the posterior estimates, which equal 0.243, 0.310, 0.070, and posterior standard deviations 0.172, 0.100, and 0.045, for the absolute teacher, gender, and race effect, respective.

4.3.3 Testing the impact of network statistics on classroom dynamics

Next, we focus on testing the effects of endogenous (network) characteristics of actors. Specifically, we evaluate whether the fixed effects of popularity or outgoingness makes actors more likely to be the next sender. Let $\phi_{\text{popularity}}$ be the mean effect of *popularity* and $\phi_{\text{outgoingness}}$ be the mean *outgoingness* effect, both in the sender model. We test the following hypotheses against one another:

$$\begin{aligned}
 H_1 &: \phi_{\text{popularity}} = \phi_{\text{outgoingness}} \\
 H_2 &: \phi_{\text{popularity}} < \phi_{\text{outgoingness}} \\
 H_3 &: \phi_{\text{popularity}} > \phi_{\text{outgoingness}}
 \end{aligned}$$

Having more than two hypotheses, we show the Bayes factors in an evidence matrix form, which yields

$$\begin{matrix}
 & H_1 & H_2 & H_3 \\
 \begin{matrix} H_1 \\ H_2 \\ H_3 \end{matrix} & \left(\begin{array}{ccc} * 20/1.00 & 48.55 & 0.43 \\ 0.02 & 1.00 & 0.01 \\ 2.29 & 110.99 & 1.00 \end{array} \right)
 \end{matrix}$$

Assuming equal prior probabilities of the three hypotheses, the posterior probabilities are equal to $P(H_1|\mathbf{E}) = 0.384$, $P(H_2|\mathbf{E}) = 0.006$, and $P(H_3|\mathbf{E}) = 0.610$. The results indicate that H_3 is most likely to be true after observing the data, being approximately twice more likely than H_1 , and approximately 100 times more likely than H_2 . For this reason, we can with almost complete certainty state that the popularity effect is not smaller than the outgoingness effect. This is confirmed by the posterior estimates of the popularity and the outgoingness effect which are equal to 0.17 and 0.07, respectively; see also the posteriors in Fig. 10. More data are required to obtain more decisive evidence of whether H_1 or H_3 is true.

In addition to testing the fixed effects as above, we can also test the global means of random effects. Here, we test the participation shift **ABBA** is more likely than the participation shift **ABAB** (both are random effects) after controlling for the other effects in the receiver model. Let μ_{abba} be the global mean effect of **ABBA** and μ_{abab} be the global mean effect of **ABAB**. This comparison captures the tension between an individual’s tendency to continue to speak to the same individual or the societal norm of reciprocity where the previous receiver because the sender of the next event and the previous sender because the receiver. In polite conversation norms, we would expect to see more turn-switches between two individuals (A speaks to B and then B responds to A) than turn-continuing (where A directs a comment to B and then continues to speak without giving B the opportunity to respond first). The estimates in Table 1 suggests that this is indeed the case since the mean effect of **ABBA** is almost four times as large as the mean effect of **ABAB**, 4.27 and 1.18, respectively. Here, we formally test this by considering the following three hypotheses:

$$\begin{aligned}
 H_1 &: \mu_{\text{abba}} = \mu_{\text{abab}} \\
 H_2 &: \mu_{\text{abba}} < \mu_{\text{abab}} \\
 H_3 &: \mu_{\text{abba}} > \mu_{\text{abab}}
 \end{aligned}$$

This yields the following evidence matrix:

$$\begin{matrix} & H_1 & H_2 & H_3 \\ \begin{matrix} H_1 \\ H_2 \\ H_3 \end{matrix} & \left(\begin{array}{ccc} * 20/1.00 & 143.26 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 4.19 \times 10^{12} & 6.00 \times 10^{14} & 1.00 \end{array} \right) \end{matrix}$$

The evidence favoring H_3 is overwhelmingly larger than the evidence for any other hypothesis, strongly supporting the hypothesis that the effect of immediate reciprocity is larger than that of immediate inertia across these fifteen classrooms at Magnet High school. This is also confirmed by the posterior probabilities $P(H_1|\mathbf{E}) = 0.000$, $P(H_2|\mathbf{E}) = 0.000$, and $P(H_3|\mathbf{E}) = 1.000$, when assuming equal prior probabilities.

Finally, we showcase a hypothesis test of network effects between the receiver model and the sender model. Specifically, we test whether the *teacher* variable has a larger effect on the rate of being a sender or on the rate of being chosen as receiver. Given the dominant role of the teacher as a sender, we expect the teacher effect is larger in the sender model. We formulate the hypotheses as follows:

$$\begin{aligned} H_1 &: \zeta_{teacher} = \mu_{teacher} \\ H_2 &: \zeta_{teacher} < \mu_{teacher} \\ H_3 &: \zeta_{teacher} > \mu_{teacher}, \end{aligned}$$

where hypothesis H_3 corresponds to our scientific expectation.

The resulting evidence matrix for the Bayes factors is:

$$\begin{matrix} & H_1 & H_2 & H_3 \\ \begin{matrix} H_1 \\ H_2 \\ H_3 \end{matrix} & \left(\begin{array}{ccc} 20/1.00 & 92.15 & 0.00 \\ 0.01 & 1.00 & 0.00 \\ 5.84 \times 10^4 & 5.38 \times 10^6 & 0.00 \end{array} \right) \end{matrix}$$

Indeed, the evidence matrix shows convincing evidence that the teacher effect is larger in the sender model than in the receiver model. The posterior probabilities of the three hypotheses are equal to $P(H_1|\mathbf{E}) = 0.000$, $P(H_2|\mathbf{E}) = 0.000$, and $P(H_3|\mathbf{E}) = 1.000$, when assuming equal prior probabilities, which results in the same conclusion. The results are also confirmed by the estimates which are equal to $\bar{\zeta}_{teacher} = 2.22$ and $\bar{\mu}_{teacher} = 0.20$.

5 Discussion

In this paper, we presented a Bayesian actor-oriented multilevel relational event model for studying social interaction behavior from independent relational event sequences. This model allows for inferences at the actor level, thus opening the possibility of unveiling effects that make an actor more prone to send or receive an interaction in the population under study. Our results show that the model is able to capture the effects that have the largest impacts in actors preferences, even when those effects are different in the sender activity and receiver choice rates. The models can be estimated using the Stan programming language, the Stan code for the actor-oriented relational event model is available on Github (https://github.com/Fabio-Vieira/bayesian_dynamic_network).

Furthermore, a flexible set of hypothesis testing procedures was proposed for this class of models, facilitating inferences on population parameters of the estimated effects. These tests can be used for testing whether effects should be treated as fixed or random across sequences, for testing the relative heterogeneity of different network effects across sequences, and for testing equality and order constraints on the fixed effects, even on their absolute values. These tests are useful to get a better understanding about the heterogeneity of social interaction behavior across independent relational event sequences. These novel testing procedures can also be applied to other mixed effects models to better understand the heterogeneity in other types of multilevel or clustered data.

The computational issues originating from the inefficiencies that result from the hierarchical structure of the model are the main limitations in the estimation process. We have taken advantage of the multivariate normal structure of the hierarchical prior to induce a non-centered transformation in the random-effects parameters, which is more efficient in practice for the reasons aforementioned. An issue that we have not directly addressed in our paper is the sparsity of relational event data which may cause the geometry of the posterior distribution to be highly complex (Betancourt and Girolami (2015)). The alien form of the likelihood of the relational event model also presents a challenge to the estimation of hierarchical relational event models. An especially attractive next step would be to construct alternative representations of the model so that posterior distributions in closed form could be derived, improving efficiency in the estimation process.

Another promising direction to improve computational feasibility when fitting multilevel relational event models to the large clustered relational event sequences is using meta-analytic approximations (Borenstein et al. 2010). A Gibbs sampler could be derived using the point estimates from the independent relational event sequences as data observations. The question would then be how large the relational event sequences should be in order for these meta-analytic approximation to be accurate enough to make reliable statistical inferences. Our flexible Bayes factor tests could then be built on top of that. Finally, another important direction for future research would be to model time-varying coefficients across sequences in order to discover complex temporal changes of the network drivers of social interactions over time.

A Actor-oriented relational event model as a Poisson regression model

The likelihood of the actor-oriented relational event model is a product of a piece-wise constant exponential likelihood and a multinomial likelihood. Where the former corresponds to the sender model and the latter to the receiver model. Both of these models are special cases of the Poisson model.

Sender model as a Poisson regression

Proof Assuming M events are observed, and the risk set contains N actors,

$$\begin{aligned}
P(\mathbf{E} | \boldsymbol{\gamma}) &= \prod_{m=1}^M P\left((s_m, t_m) | \boldsymbol{\gamma}, (e_1, \dots, e_{m-1})\right) = \\
&= \prod_{m=1}^M \lambda_{s_m} \times \exp\left\{-\left(t_m - t_{m-1}\right) \sum_{s \in \mathcal{R}} \lambda_s\right\} \\
&= \prod_{m=1}^M \exp\left\{\boldsymbol{\gamma} \mathbf{x}'_{s_m}\right\} \times \exp\left\{-\exp\left\{v_m\right\} \sum_{s \in \mathcal{R}} \exp\left\{\boldsymbol{\gamma} \mathbf{x}'_{s_m}\right\}\right\} \quad (v_m = \log(t_m - t_{m-1})) \\
&\propto \prod_{m=1}^M \exp\{v_m + \boldsymbol{\gamma} \mathbf{x}'_{s_m}\} \times \prod_{s \in \mathcal{R}} \exp\{-\exp\{v_m + \boldsymbol{\gamma} \mathbf{x}'_{s_m}\}\} \\
&= \prod_{m=1}^M \prod_{s \in \mathcal{R}} \exp\{v_m + \boldsymbol{\gamma} \mathbf{x}'_{s_m}\}^{y_s} \times \exp\{-\exp\{v_m + \boldsymbol{\gamma} \mathbf{x}'_{s_m}\}\} \\
&= \prod_{m=1}^M \prod_{s \in \mathcal{R}} P\left(y_{s_m} | \lambda_{s_m} = \exp\{v_m + \boldsymbol{\gamma} \mathbf{x}'_{s_m}\}\right),
\end{aligned}$$

where $y_{s_m} = 1$ if actor s is observed and zero otherwise. As a result, the factorial term in the Poisson likelihood will always be equal to one, since $1! = 1$. \square

Receiver model as a Poisson regression:

Proof Once again, assuming M events are observed, and the risk set contains N actors, the likelihood function of the receiver model has the following form

$$P(\mathbf{E} | \boldsymbol{\beta}) = \prod_{m=1}^M \frac{\lambda_{r_m | s_m}}{\sum_{r \in \mathcal{R}} \lambda_{r | s_m}} = \prod_{m=1}^M \frac{\exp\{\boldsymbol{\beta} \mathbf{x}'_{s_m r_m}\}}{\sum_{r \in \mathcal{R}} \exp\{\boldsymbol{\beta} \mathbf{x}'_{s_m r}\}}.$$

Taking the logarithm of $p(\mathbf{E} | \boldsymbol{\beta})$ yields

$$\log(p(\mathbf{E} | \boldsymbol{\beta})) = \sum_{m=1}^M \boldsymbol{\beta} \mathbf{x}'_{s_m r_m} - \left(\sum_{r \in \mathcal{R}} \exp\{\boldsymbol{\beta} \mathbf{x}'_{s_m r}\} \right) \quad (\star)$$

Then, when considering an alternative formulation using a Poisson model, an equivalent likelihood can be written as

$$\tilde{Y}_{r_m | s_m} \sim \text{Poisson}\left(\exp\{\boldsymbol{\beta} \mathbf{x}'_{s_m r_m} + \alpha_m\}\right),$$

where α_m is an event-specific intercept and $\tilde{Y}_{r | s} = 1$ if the dyad (s, r) was observed and zero otherwise. The likelihood of the Poisson model can then be written

$$P(\tilde{Y}_{r|s} | \beta, \alpha) = \prod_{m=1}^M \prod_{r \in \mathcal{R}} \left(\exp\{\beta x'_{s_m r} + \alpha_m\} \right)^{\tilde{y}_{r_m|s_m}} \times \exp\{-\exp\{\beta x'_{s_m r} + \alpha_m\}\} \times \tilde{y}_{r_m|s_m}^{-1},$$

where $\tilde{y}_{r_m|s_m} = 1$. To write the likelihood as a function of β , we need to derive the maximum likelihood estimate of α_m . Take the logarithm and deriving with respect to α_m yields the following MLE expression

$$\hat{\alpha}_m = -\log \left(\sum_{r \in \mathcal{R}} \exp\{\beta x'_{s_m r}\} \right).$$

Plug in the MLE into the likelihood

which is equal to (★). □

B Linear transformation of multivariate normal distribution

Here, we show that the transformation of the random-effects parameters is valid. Let X be a p -dimensional random variable, with $X \sim \mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^p$ is the mean vector and Σ is a $p \times p$ covariance matrix. Thus, the moment generating function of X is given by

$$M_X(t) = \exp \left\{ t' \mu + \frac{1}{2} t' \Sigma t \right\}.$$

Now, assuming we can write $X = \mu + AZ$, where A is a $p \times p$ matrix, with $\Sigma = A'A$, and $Z = (Z_1, Z_2, \dots, Z_p)$ is an independent normal random vector, with $Z_i \sim \mathcal{N}(0, 1)$, for $i = 1, \dots, p$. Therefore, if we can show that $M_X(t) = M_{(\mu+AZ)}(t)$, then the transformation holds. So, we derive the moment generating function of the transformation as

Proof

$$\begin{aligned} M_{(\mu+AZ)}(t) &= E(e^{t'X}) = E(e^{t'(\mu+AZ)}) \\ &= e^{t' \mu} E(e^{t'Z}), \text{ where } l' = t'A \\ &= e^{t' \mu} E\left(e^{\sum_{i=1}^p l_i Z_i}\right) \\ &= e^{t' \mu} \prod_{i=1}^p E(e^{l_i Z_i}) \\ &= e^{t' \mu} \prod_{i=1}^p e^{l_i^2 / 2} \\ &= \exp \left\{ t' \mu + \sum_{i=1}^p \frac{l_i^2}{2} \right\} \end{aligned}$$

□

Thus, $M_X(t) = M_{(\mu+AZ)}(t)$. Hence, we can transform the random-effects parameters in an MCMC algorithm by following the steps,

1. Obtain posterior samples of mean vector $\mu \in \mathbb{R}^p$;
2. Obtain posterior samples of the $p \times p$ covariance matrix Σ ;
3. Compute the Cholesky factor A of Σ ;
4. Sample a p -dimensional vector of independent and identically distributed standard normal variables Z ;
5. Compute the transformation $\beta = \mu + AZ$.

C Bayes factor of constrained against unconstrained hypotheses

Let us suppose that we wish to test an informative hypothesis $H_i : \theta \in \Theta_i$, where Θ_i can be seen as a truncation of the space of $\theta \in \Theta$, with $\Theta_i \subset \Theta$. Thus, if we define the prior in the space Θ_i as a truncation of the prior in Θ , such as $\pi_i(\theta) = c^{-1} \pi(\theta) \mathbb{I}\{\theta \in \Theta_i\}$, where c is a normalizing constant of the form $c = \int_{\theta \in \Theta_i} \pi(\theta) d\theta$ and $\pi(\theta)$ is the prior under the unconstrained space. Then, the Bayes factor of H_i against H_u will be given by

$$\begin{aligned} \text{BF}_{iu} &= \frac{m(\mathbf{E}|H_i)}{m(\mathbf{E}|H_u)} = \frac{\int_{\theta \in \Theta_i} \pi_i(\theta) \pi(\mathbf{E}|\theta) d\theta}{\int \pi(\theta) \pi(\mathbf{E}|\theta) d\theta} = c^{-1} \int_{\theta \in \Theta_i} \frac{\pi(\theta) \pi(\mathbf{E}|\theta) d\theta}{\int \pi(\theta) \pi(\mathbf{E}|\theta) d\theta} = \\ &= c^{-1} \int_{\theta \in \Theta_i} \pi(\theta|\mathbf{E}) d\theta = \frac{\int_{\theta \in \Theta_i} \pi(\theta|\mathbf{E}) d\theta}{\int_{\theta \in \Theta} \pi(\theta) d\theta} = \frac{P(\theta \in \Theta_i|\mathbf{E})}{P(\theta \in \Theta_i)}. \end{aligned}$$

Therefore, the Bayes factor of a constrained against an unconstrained hypothesis can be reduced to a ratio of posterior and prior probabilities in the space of the constrained hypothesis. In case of an exact hypothesis, such as $H : \theta = \mathbf{0}$, the ratio of probabilities becomes the Savage–Dickey density ratio (Dickey 1971).

D Bayes factor for testing random-effect structures

In this appendix, we show details on the derivation of the parameters in the Bayes factor for testing random-effect structures. Assuming we have K social networks and each of them has its own β_k , for $i = 1, \dots, K$, regression parameter. Where $\beta_k \sim \mathcal{N}(\mu, \sigma^2)$. Thus, let $\bar{\mu}$, $\bar{\sigma}^2$, $\bar{\beta}_k$ be posterior estimates for μ , σ^2 and β_k , $\forall k$. Also, let $\bar{\tau}_k^2$ be the point estimate for the variance of the posterior distribution of β_k . Then, we can write

$$\begin{aligned} \text{posterior: } \beta_k | \mathbf{E}_k &\sim \mathcal{N}(\bar{\beta}_k, \bar{\tau}_k^2) \\ \text{prior: } \beta_k &\sim \mathcal{N}(\bar{\mu}, \bar{\sigma}^2). \end{aligned} \tag{7}$$

Assuming we want to test $H_0 : \beta_1 = \dots = \beta_K$ against $H_1 : \text{"at least one is different"}$. Thus we can do so by writing $\xi_j = \beta_{j+1} - \beta_j$, for $j = 1, \dots, K - 1$. Therefore, we can approximate the joint distribution of $\xi = (\xi_1, \dots, \xi_{K-1})$ by a multivariate normal

$$\begin{aligned} \text{posterior: } \xi | E &\sim \mathcal{N}(\bar{\mu}_\xi, \bar{\Sigma}_\xi) \\ \text{prior: } \xi &\sim \mathcal{N}(\mathbf{0}, \bar{\Lambda}_\xi). \end{aligned}$$

Thus, we can derive the parameters of the distributions of every component $j = 1, \dots, K - 1$ of ξ as follows

$$\begin{aligned} E(\xi_j | E) &= E(\beta_{j+1} - \beta_j) \\ &= E(\beta_{j+1}) - E(\beta_j) = \bar{\beta}_{j+1} - \bar{\beta}_j, \\ \text{Var}(\xi_j | E) &= \text{Var}(\beta_{j+1} - \beta_j) \\ &= \text{Var}(\beta_{j+1}) + \text{Var}(\beta_j) = \bar{\tau}_{j+1}^2 + \bar{\tau}_j^2, \\ \text{Cov}(\xi_j, \xi_{j+h} | E) &= E(\xi_j \xi_{j+h}) - E(\xi_j)E(\xi_{j+h}) \\ &= E\left((\beta_{j+1} - \beta_j)(\beta_{j+h+1} - \beta_{j+h})\right) - E\left((\beta_{j+1} - \beta_j)\right)E\left((\beta_{j+h+1} - \beta_{j+h})\right) \\ &= E(\beta_{j+1}\beta_{j+h+1}) - E(\beta_{j+1}\beta_{j+h}) - E(\beta_j\beta_{j+h+1}) + E(\beta_j\beta_{j+h}) - \\ &\quad - E(\beta_{j+1})E(\beta_{j+h+1}) + E(\beta_{j+1})E(\beta_{j+h}) + E(\beta_j)E(\beta_{j+h+1}) - E(\beta_j)E(\beta_{j+h}) \\ &= \begin{cases} -\text{Var}(\beta_{j+1}) = -\bar{\tau}_{j+1}^2, & \text{if } h = 1 \\ 0, & \text{if } h > 1. \end{cases} \end{aligned}$$

The parameters in the prior are derived in the same way with $E(\xi_j) = 0$, $\text{Var}(\xi_j) = 2\bar{\sigma}^2$ and $\text{Cov}(\xi_j, \xi_{j+h}) = -\bar{\sigma}^2$, if $h = 1$, and zero otherwise.

E Point-wise deviance residuals

After estimation, it is common practice to check how the proposed model fits the data under study. In DuBois et al. (2013), they used the posterior draws to evaluate the log likelihood at every data point to determine the dynamic adequacy of the model to every data point. Thus, for the actor-oriented relational event model, for $k = 1, \dots, K$ and $m = 1, \dots, M_k$, this quantity would be computed as

$$\begin{aligned} \text{Res}_m &= -2 \left\{ \log \left(\lambda_{s_m} \left(t_m | E_k \right) \right) + \log \left(\lambda_{r_m | s_m} \left(t_m | s_m, E_k \right) \right) - \right. \\ &\quad \left. - \log \left(\sum_{s,r} \lambda_{r | s_m} \left(t_m | s_m, E_k \right) \right) - (t_m - t_{m-1}) \left(\sum_i \lambda_s(t_m | E_k) \right) \right\}. \end{aligned} \tag{8}$$

This measure is usually called residual deviance (Collett 2015), and it is useful to compare models and see which one fits better each data point by having smaller values of Res_m .

Funding This research was supported by the Netherlands Organization for Scientific Research (NWO) Grant to JM and FV (452-17-006), as well as the ERC Starting Grant “TIMEISNOW” to JM and RL (758791).

Data availability Requests for data access should be directed to Daniel McFarland.

Competing Interests The corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aalen O (1989) A linear regression model for the analysis of life times. *Stat Med* 8(8):907–925
- Aalen O, Borgan O, Gjessing H (2008) *Survival and event history analysis: a process point of view*. Springer, New York
- Arena G, Lakdawala R, Meijerink-Bosman M, Karimova D, Shafiee Kamalabad M, Generoso Vieira F (2022) *remstimate: optimization tools for tie-oriented and actor-oriented relational event models* [Computer software manual]. Retrieved from <https://github.com/TilburgNetworkGroup/remstimate> (R package version 1.0)
- Baker SG (1994) The multinomial-Poisson transformation. *J Royal Stat Soc: Ser D (The Statistician)* 43(4):495–504
- Bayarri MJ, García-Donato G (2007) Extending conventional priors for testing general hypotheses in linear models. *Biometrika* 94(1):135–152
- Betancourt M, Girolami M (2015) Hamiltonian Monte Carlo for hierarchical models. *Curr Trends Bayesian Methodol Appl* 79(30):2–4
- Blonder B, Dornhaus A (2011) Time-ordered networks reveal limitations to information flow in ant colonies. *PLoS one* 6(5):e20298
- Böing-Messing F, Mulder J (2020) Bayes factors for testing order constraints on variances of dependent outcomes. *Am Stat* 75(2):152–61
- Böing-Messing F, van Assen MA, Hofman AD, Hoijtink H, Mulder J (2017) Bayesian evaluation of constrained hypotheses on variances of multiple independent groups. *Psychol Methods* 22(2):262
- Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 1(2):97–111
- Bürkner P-C (2017) *brms: an r package for Bayesian multilevel models using Stan*. *J Stat Softw* 80:1–28
- Butts CT (2008) A relational event framework for social action. *Sociol Methodol* 38(1):155–200
- Butts CT, Marcum CS (2017) A relational event approach to modeling behavioral dynamics. *Group processes*. Springer, Cham, pp 51–92
- Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Riddell A (2017) Stan: a probabilistic programming language. *J Stat Softw*. <https://doi.org/10.18637/jss.v076.i01>
- Collett D (2015) *Modelling survival data in medical research*. CRC Press
- Consul P, Famoye F (1992) Generalized Poisson regression model. *Commun Stat-Theory Methods* 21(1):89–109
- Cox DR (1972) Regression models and life-tables. *J Royal Stat Soc: Ser B (Methodological)* 34(2):187–202
- Dickey JM (1971) The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann Math Stat* 42(1):204–223

- DuBois C, Butts CT, McFarland D, Smyth P (2013) Hierarchical models for relational event sequences. *J Math Psychol* 57(6):297–309
- Eagle N, Pentland AS (2003) Social network computing. In: *International conference on ubiquitous computing*, pp 289–296
- Frank O, Strauss D (1986) Markov graphs. *J Am Stat Assoc* 81(395):832–842
- Friedman M (1982) Piecewise exponential models for survival data with covariates. *Ann Stat* 10(1):101–113
- Frome EL (1983) The analysis of rates using Poisson regression models. *Biometrics* 39(3):665–674
- Gelman A, Hill J (2006) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian data analysis*. CRC Press
- Gibson DR (2005) Taking turns and talking ties: networks and conversational interaction. *Am J Sociol* 110(6):1561–1597
- Hanneke S, Fu W, Xing EP (2010) Discrete temporal models of social networks. *Electron J Stat* 4:585–605
- Hayat MJ, Higgins M (2014) Understanding Poisson regression. *J Nurs Educ* 53(4):207–215
- Hecht M, Zitzmann S (2021) Sample size recommendations for continuous-time models: compensating shorter time series with larger numbers of persons and vice versa. *Struct Equ Model: Multidiscip J* 28(2):229–236
- Hoffman MD, Gelman A (2014) The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 15(1):1593–1623
- Hojtink H, Gu X, Mulder J (2019) Bayesian evaluation of informative hypotheses for multiple populations. *Br J Math Stat Psychol* 72(2):219–243
- Holford TR (1980) The analysis of rates and of survivorship using log-linear models. *Biometrics* 36(2):299–305
- Jeffreys H (1961) *Theory of probability*, 3rd edn. Oxford University Press, Oxford
- Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90(430):773–795
- Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M et al (2016) Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Sci* 5(1):1–21
- Klugkist I, Hoijtink H (2007) The Bayes factor for inequality and about equality constrained models. *Comput Stat Data Anal* 51(12):6367–6379
- Klugkist I, Laudy O, Hoijtink H (2005) Inequality constrained analysis of variance: a Bayesian approach. *Psychol Methods* 10(4):477
- Laird N, Olivier D (1981) Covariance analysis of censored survival data using log-linear analysis techniques. *J Am Stat Assoc* 76(374):231–240
- Leenders RTA, Contractor NS, DeChurch LA (2016) Once upon a time: understanding team processes as relational event networks. *Organ Psychol Rev* 6(1):92–115
- 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
- Lewandowski D, Kurowicka D, Joe H (2009) Generating random correlation matrices based on vines and extended onion method. *J Multivar Anal* 100(9):1989–2001
- Liang F, Paulo R, Molina G, Clyde MA, Berger JO (2008) Mixtures of g priors for Bayesian variable selection. *J Am Stat Assoc* 103(481):410–423
- Lusher D, Koskinen J, Robins G (2013) *Exponential random graph models for social networks: theory, methods, and applications*. Cambridge University Press
- McFadden D (1973) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) *Frontiers in econometrics*. Academic Press, pp 105–142
- McFarland DA (2001) Student resistance: how the formal and informal organization of classrooms facilitate everyday forms of student defiance. *Am J Sociol* 107(3):612–678
- Mulder J (2014) Bayes factors for testing inequality constrained hypotheses: issues with prior specification. *Br J Math Stat Psychol* 67(1):153–171
- Mulder J (2014) Prior adjusted default Bayes factors for testing (in) equality constrained hypotheses. *Comput Stat Data Anal* 71:448–463
- Mulder J, Fox J-P (2019) Bayes factor testing of multiple intraclass correlations. *Bayesian Anal* 14(2):521–552

- Mulder J, Leenders RTA (2019) Modeling the evolution of interaction behavior in social networks: a dynamic relational event approach for real-time analysis. *Chaos, Solitons Fractals* 119:73–85
- Mulder J, Pericchi LR (2018) The matrix-f prior for estimating and testing covariance matrices. *Bayesian Anal* 13(4):1193–1214
- Mulder J, Williams DR, Gu X, Tomarken A, Böing-Messing F, Olsson-Collentine A et al (2021) Bfpack: flexible Bayes factor testing of scientific theories in r. *J Stat Softw* 100:1–63
- O’Hagan A (1995) Fractional Bayes factors for model comparison. *J Royal Stat Soc: Ser B (Methodological)* 57(1):99–118
- O’Hagan A (1997) Properties of intrinsic and fractional Bayes factors. *Test* 6(1):101–118
- Pérez JM, Berger JO (2002) Expected-posterior prior distributions for model selection. *Biometrika* 89(3):491–512
- Perry PO, Wolfe PJ (2013) Point process modelling for directed interaction networks. *J Royal Stat Soc: Ser B (Stat Methodol)* 75(5):821–849
- R Core Team (2017) R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev* 16:225–237
- Snijders TA (1996) Stochastic actor-oriented models for network change. *J Math Sociol* 21(1–2):149–172
- Stadtfeld C, Block P (2017) Interactions, actors, and time: dynamic network actor models for relational events. *Sociol Sci* 4:318–352
- Stadtfeld C, Hollway J, Block P (2017) Dynamic network actor models: investigating coordination ties through time. *Sociol Methodol* 47(1):1–40
- Stadtfeld C, Hollway J, Block P (2017) Rejoinder: dynamics and the grounds for actor-oriented network event models. *Sociol Methodol* 47(1):56–67
- Stan Development Team (2018) RStan: the R interface to Stan. Retrieved from <http://mc-stan.org/> <http://mc-stan.org/> (R package version 2.17.3)
- Van Der Hofstad R (2009) Random graphs and complex networks. Available on <http://www.win.tue.nl/rhofstad/NotesRGCN.pdf> <http://www.win.tue.nl/rhofstad/NotesRGCN.pdf>
- Vu D, Hunter D, Smyth P, Asuncion AU (2011) Continuous-time regression models for longitudinal networks. In: *Advances in neural information processing systems*, pp 2492–2500
- Vu D, Pattison P, Robins G (2015) Relational event models for social learning in MOOCs. *Soc Netw* 43:121–135
- 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
- Wang Z, Wu Y, Chu H (2018) On equivalence of the lkj distribution and the restricted Wishart distribution. arXiv preprint [arXiv:1809.04746](https://arxiv.org/abs/1809.04746)
- Wetzels R, Grasman RP, Wagenmakers E-J (2012) A default Bayesian hypothesis test for ANOVA designs. *Am Stat* 66(2):104–111

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.