Stochastic blockmodeling of relational event dynamics

Anonymous Author(s)
Affiliation
Address
email

Abstract

For continuous-time network data, several approaches have recently been proposed for modeling dyadic event rates conditioned on the observed history of events and nodal or dyadic covariates. In many cases, however, interaction propensities – and even the underlying mechanisms of interaction – vary systematically across subgroups whose identities are unobserved. For static networks, such heterogeneity has been treated via methods such as stochastic blockmodeling, which operate by assuming latent groups of individuals with similar tendencies in their group-wise interactions. Here, we combine these two approaches by positing a latent partition of the node set such that event dynamics within and between subsets evolve in potentially distinct ways. We illustrate the use of our model family by application to several forms of dyadic interaction data, including email communication and Twitter direct messages. Parameter estimates from the fitted models clearly reveal heterogeneity in the dynamics among groups of individuals. We also find that the fitted models have better predictive accuracy than either baseline models or relational event models without latent structure. Our approach illustrates the utility of latent structure methods based on detailed dynamics, which can succeed even in the absence of differences in marginal interaction rates across groups.

1 Introduction

Statistical methods for analyzing network data are becoming increasingly useful for studying phenomena ranging from online social behavior to protein interactions [1]. Recent work has expanded to include settings in which we observe events occurring between nodes over time (i.e., *relational events*, as opposed to static edge structures or ongoing relationships), with the common goal of modeling interaction dynamics in terms of both endogenous mechanisms and exogenous covariates. A key concern in this regard is the ability to detect differential behavioral tendencies on the part of subsets of nodes, the dynamic analog of *role structure* in classical network analysis [2].

In the cross-sectional case, *stochastic blockmodels* [3] have been proposed as a family of approaches that capture behavioral similarity by identifying subsets of nodes with similar patterns of ties to those in other sets. While it is natural to apply these ideas directly to relational event data via blockmodeling of the time-marginalized rates of interaction among dyads (effectively treating the event structure as a valued graph), there are limits to what this approach can detect. Consider, e.g., Figure 1. In the top left panel, we depict the time-marginalized frequencies of simulated interactions between members of two groups (A and B), with darker cells indicating higher interaction frequencies. As can be seen, a clear role structure is present, with members of each subgroup interacting at higher rates with co-members than out-group members; such a structure is easily detectable via conventional blockmodeling techniques. By contrast, consider the interaction patterns shown in the top right panel. Here, there is no systematic difference in marginal rates between the two groups, rendering them invisible under a standard blockmodeling approach.

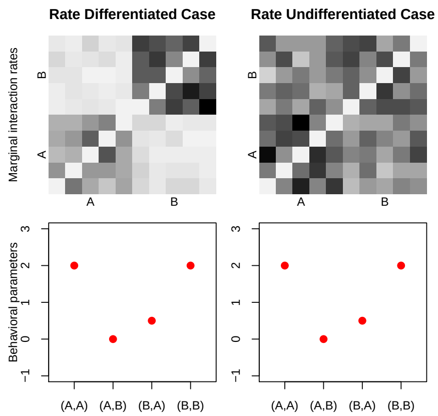


Figure 1: Illustration of differentiation by rate of interaction versus by dynamic behavior. The left column shows an example where within-block communication is large, and there is a higher tendency for reciprocity within a block. In the situation where the two groups are not differentiated by rate, as in the right column, a standard blockmodel is unable to distinguish between groups A and B. The proposed method can, however, learn such groups by employing more flexible definitions of shared dynamics.

There is, however, a difference to be detected. The bottom panels of Figure 1 show, for each respective simulation, parameters (as discussed in Section 2.1) governing the tendency towards reciprocation for members of each group vis a vis communications coming from in- or out-group members. For both simulated cases, the log-hazard for an event in which a member of group A or group B immediately responds to an incoming communication from an in-group member is increased by an increment of 2 (relative to a non-reciprocating event), and the log-hazard for an event in which a member of group B immediately responds to an incoming communication from a member of group A is increased by a corresponding increment of 0.5. Groups A and B are thus distinctive in terms of their dynamic behavior, even in the absence of marginal differences in propensity to communicate. A blockmodeling approach that classifies nodes based on shared dynamics (rather than merely shared marginal communication rates) can potentially identify such subtle distinctions; in this paper, we introduce such an approach.

Borrowing from the intuition of stochastic blockmodels, we propose a novel continuous-time model family for network-based event sequences where latent clusters of nodes share similar patterns of interaction. Our approach employs a flexible framework for specifying how the process depends on the previous history of events [4, 5]. In this way one can compare theories about underlying processes and make predictions about future data conditioned on the past, simultaneously adjusting for unobserved heterogeneity.

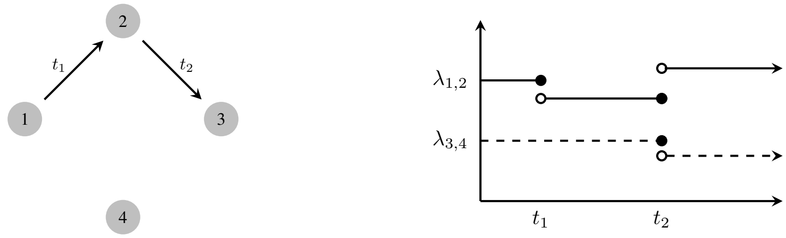
We describe how we learn the latent cluster assignments and model parameters via MCMC, and illustrate the behavior of the model with simulated data. Using several real-world social network data sets involving dyadic communication, we compare the predictive performance of the fitted models to standard baselines. We conclude by showing that the parameter estimates reveal interpretable structure in the event dynamics, enabling the study of a dynamic extension to stochastic equivalence.

2 Model

Consider a sequence of events $\mathcal{A} = (0, t_1, \ldots, t_M)$ arising from a nonhomogeneous Poisson process with intensity $\lambda(t)$. If the intensity is left continuous and piecewise constant with respect to a set of knots τ then the likelihood can be written

$$\begin{aligned}\mathcal{L}(\mathcal{A}|\theta) &= \prod_{m=1}^{M} \lambda(t_m) \exp\left\{-\int_0^{t_M} \lambda(s)ds\right\} \\ &= \prod_{m=1}^{M} \lambda(t_m) \prod_{k=1}^{|\tau|} \exp\left\{-(\tau_k - \tau_{k-1})\lambda(\tau_k)\right\} \qquad (1)\end{aligned}$$

One can extend the above to marked point processes where each event contains additional information. In the context of relational events occurring among N nodes, each event in the process contains both a sender i and a recipient j such that $(i, j) \in \mathcal{R}$, where $\mathcal{R}$ is the *risk set* comprised of all allowed dyads. We assume each event involving dyad $(i, j) \in \mathcal{R}$ occurs as a Poisson process with intensity $\lambda_{ij}(t|\cdot)$ that is piecewise constant and depends on the previous history of events



(a) Dynamic network data (b) Intensities for two dyads

Figure 2: Illustration of relational event data and the assumptions of the model. (a) A sequence of two events among four nodes: (1,2) occurs at time t_1 and (2,3) occurs at time t_2. (b) The intensity functions $\lambda_{1,2}(t)$ (solid) and $\lambda_{3,4}(t)$ (dashed). We assume $\lambda_{3,4}(t)$ can only change after events sent or received by node 3.

$\mathcal{A}_t = \{(t_m, i_m, j_m) : t_m \in [0, t)\}$. The likelihood of an observed event history $\mathcal{A}_{t_M}$ (extending from time 0 to the time of the final event and denoted $\mathcal{A}$ for convenience) is

$$\mathcal{L}(\mathcal{A}|\theta) = \prod_{m=1}^{M} \lambda_{i_m,j_m}(t_m|\cdot) \prod_{(i,j)\in\mathcal{R}} \exp\{-(t_m - t_{m-1})\lambda_{ij}(t_m|\cdot)\} \tag{2}$$

Here, we aim to learn about the dynamics within and between subsets of nodes. To facilitate this, we assume each node i belongs to latent cluster z_i and use a log linear model for the intensity functions

$$\log \lambda_{ij}(t|\mathcal{A}_t, \beta, \mathbf{z}) = \boldsymbol{\beta}'_{z_i,z_j} \mathbf{s}(t, i, j, \mathcal{A}_t)$$

where for each pair of clusters (k,l) we have a parameter vector $\boldsymbol{\beta}_{k,l} \sim \mathrm{N}_P(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 I)$ that corresponds to the P-dimensional vector of statistics $\mathbf{s}(t, i, j, \mathcal{A}_t)$ computed from the previous history $\mathcal{A}_t$. Thus, the rate of (i, j) events has the same parameters as other events occurring between group z_i and z_j.

As shown in Figure 2, we only allow each intensity function $\lambda_{ij}(t)$ to change following an event where i was the sender or the recipient. This is sensible in distributed settings where i has limited knowledge about interactions among other actors. In the Supplementary Material we show how this assumption also reduces the computational complexity of computing the above likelihood.

We allow the blocks to share information by placing a hierarchical prior on the collection of $\boldsymbol{\beta}_{k.l}$ where $\mu_p \propto N(0,1)$ and $\sigma_p^2 \sim$ Inv-Gamma$(\alpha_\sigma, \beta_\sigma)$. The cluster assignments are given a non-parameteric prior $z_i \sim$ CRP(α), allowing for a flexible number of clusters.

2.1 Model specification

Table 1 lists the statistics $\mathbf{s}(t, i, j, \mathcal{A}_t)$ used in Section 5. For example, a particular node sending often may indicate they will continue to be a sender in the future. We normalize these counts by the number of events up until a dyad's prior changepoint by using $f(x) = \log \frac{x+1}{m+N(N-1)}$. Other statistics could be of interest for particular substantive questions [5, 6].

The next set of statistics are *participation shift* effects inspired by research in conversational norms [7]. For example, an `ab-ba` effect indicates an increased propensity for reciprocity where the event (a, b) is followed by (b, a). Though only the statistics in Table 1 are used in our experiments, one may use any quantity that is computed using the previous history of events or known covariates about nodes or dyads. The only restriction is that the statistic may not change in value between each observed event.

2.2 Relation to other models

Our formulation is reminiscent of the stochastic blockmodel [3] for static networks which models the probability of a dyad as $p(y_{ij}) = \text{logit}^{-1}(\eta_{z_i,z_j})$ where η_{z_i,z_j} is interpreted as a mixing rate between

Statistic	Formula
Intercept	$s_0(t,i,j,\mathcal{A}_t) = 1$
Reciprocity (AB-BA)	$s_1(t,i,j,\mathcal{A}_t) = I(i_m = i, j_m = j, i_{v_{mij}} = j, j_{v_{mij}} = i)$
Turn-continuing (AB-AY)	$s_2(t,i,j,\mathcal{A}_t) = I(i_m = i, j_m = j, i_{v_{mij}} = i, j_{v_{mij}} \neq j)$
Turn-taking (AB-BY)	$s_3(t,i,j,\mathcal{A}_t) = I(i_m = i, j_m = j, i_{v_{mij}} = j, j_{v_{mij}} \neq i)$
Sender out-degree	$s_4(t,i,j,\mathcal{A}_t) = f(\sum_{m:t_m<t} I(i_m = i))$
Sender in-degree	$s_5(t,i,j,\mathcal{A}_t) = f(\sum_{m:t_m<t} I(j_m = i))$
Dyad count	$s_6(t,i,j,\mathcal{A}_t) = f(\sum_{m:t_m<t} I(i_m = i, j_m = j))$

Table 1: Statistics used to specify intensity functions using the previous history $\mathcal{A}_t$, where v_{mij} is the index of the changepoint for $\lambda_{i,j}(t)$ previous to event m.

group z_i and group z_j. In the proposed method, however, the blockmodel structure facilitates the study of intra-group and inter-group dynamics via a continuous-time network model.

The proposed family of models generalizes several important special cases. For example, using only the intercept statistic $s_0(t,i,j,\mathcal{A}_t) = 1$ is analogous to the stochastic block model for static networks. Under this model each dyad is a homogeneous Poisson process and all dyad intensities $\lambda_{i,j}$ within block (z_i, z_j) have the same intensity, $\exp\{\boldsymbol{\beta}_{z_i,z_j}\}$. As these intensities do not change under this specification, the likelihood simplifies to

$$\mathcal{L}(\mathcal{A}|\beta) = \prod_{m=1}^{M} \lambda_{i_m,j_m} \prod_{(i,j)\in\mathcal{R}} \exp\{-t_M \lambda_{i,j}\} \tag{3}$$

Alternatively, if one models only the order of the events (ignoring the times at which they occur), the likelihood can be written as

$$\mathcal{L}_{\text{mult}}(\mathcal{A}|\beta) = \prod_{m=1}^{M} \frac{\lambda_{i_m,j_m}(t_m|\cdot)}{\sum_{(i,j)\in\mathcal{R}} \lambda_{i,j}(t_m|\cdot)}. \tag{4}$$

The functional form is similar to conditional logit models used for discrete choice data [8], though here the possible choices are all the dyads in $\mathcal{R}$.

2.3 Related work

Stochastic blockmodels have been extended to longitudinal network data involving a sequence of networks occurring at discrete times [9, 10], while we focus on methods for analyzing a stream of relational events. The proposed model is an extension of recent work on modeling event-based social network data using an event history approach [5, 11, 12, 13, 6, 14]. Relational event models such as these require a knot at each observed event, while other approaches such as [15] learn the regions where an intensity is constant. By using decision trees, [15] also allow for a nonlinear relationship between statistics and intensity functions. We instead use latent variables to allow for heterogeneity in intensities across the set of possible events.

3 Inference

We use Markov chain Monte Carlo to sample from the posterior distribution of our parameters.

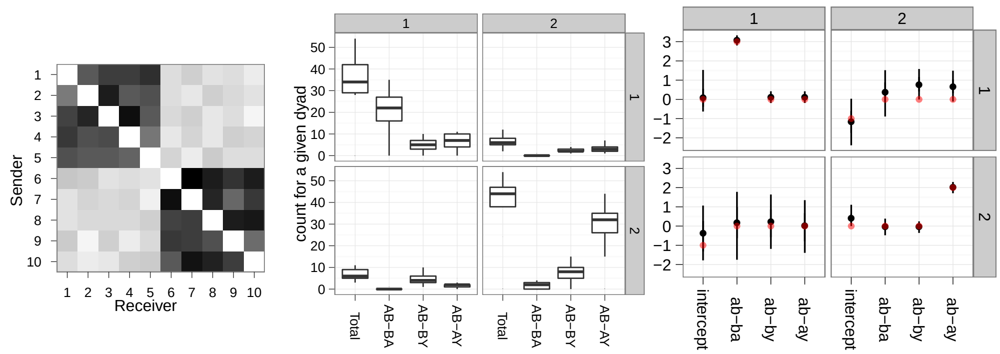
3.1 Sampling z given β and $\mathcal{A}$

We use Gibbs sampling to sample the latent class assignments $\mathbf{z}$ from the conditional distribution

$$p(z_r|\mathbf{z}_{-r}, \mathcal{A}, \alpha, \boldsymbol{\beta}) \propto p(\mathcal{A}|\mathbf{z}, \boldsymbol{\beta}) p(z_r|\mathbf{z}_{-r}, \alpha)$$

$$p(\mathcal{A}|\mathbf{z}, \boldsymbol{\beta}) \propto \prod_{m=1}^{M} \lambda_{i_m,j_m}(t_m|\cdot)^{\mathbf{1}[r\in\{i_m,j_m\}]} \prod_{(i,j)\in\mathcal{U}_r} \exp\{-(t_m - t_{m-1})\lambda_{ij}(t_m|\cdot)\}$$

where $\mathcal{U}_r = \{(i,j) \in \mathcal{R} : r \in \{i,j\}\}$ is the set of dyads involving node r. Under a CRP(α) prior, we have $p(z_i = k|z_{-i}, \alpha) = n_k$ and $p(z_i = K+1|z_{-i}, \alpha) = \alpha$ where n_k is the number of nodes assigned to cluster k.



(a) (b) (c)

Figure 3: Illustration of 2000 simulated events, as described in text. (a) Counts of each dyad. (b) Boxplot of distribution of participation counts across dyads. The top left shows an increased propensity for reciprocity within cluster 1; bottom right shows more AB-AY events within cluster 2. (c) Parameters (in red) and posterior credible intervals (in black).

3.2 Sampling β given z and $\mathcal{A}$

For each block (k, l) we need to sample the vector of parameters $\boldsymbol{\beta}_{k,l}$ from its posterior

$$p(\boldsymbol{\beta}_{k,l}|\mathcal{A}, \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \propto p(\boldsymbol{\beta}_{k,l}|\boldsymbol{\mu}, \boldsymbol{\sigma})p(\mathcal{A}|\mathbf{z}, \boldsymbol{\beta})$$

$$p(\boldsymbol{\beta}_{k,l}|\mu, \sigma) = \prod_{p=1}^{P} p(\beta_{k,l,p}|\mu_p, \sigma_p^2)$$

$$p(\mathcal{A}|\mathbf{z}, \boldsymbol{\beta}) \propto \prod_{m=1}^{M} \lambda_{i_m,j_m}(t_m|\cdot)^{\mathbf{1}[(i_m,j_m)\in\mathcal{V}_{k,l}]} \prod_{(i,j)\in\mathcal{V}_{k,l}} \exp\{-(t_m - t_{m-1})\lambda_{ij}(t_m|\cdot)\}$$

where $\mathcal{V}_{k,l} = \{(i, j) : z_i \in \{k, l\} \text{ or } z_j \in \{k, l\}\}$ is the set of dyads with a sender in group k and a recipient in l. We sample each $\beta_{k,l,p}$ via slice sampling.

3.3 Sampling μ and σ given β

The inverse gamma distribution is a conjugate prior to a Gaussian distribution with known location parameter, thus σ can be sampled from the conditional distribution

$$\sigma_p^2|\boldsymbol{\beta}, \mu_p, \alpha_\sigma, \beta_\sigma \sim \text{Inv-Gamma}\left(\alpha_\sigma + \frac{K^2}{2}, \beta_\sigma + \frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{K}(\beta_{k,l,p} - \mu_p)^2\right). \quad (5)$$

Each parameter μ_p can be sampled from its conditional distribution given σ_p^2 and the parameters $\boldsymbol{\beta}$,

$$\mu_p|\boldsymbol{\beta}, \sigma_p^2 \sim \text{Normal}\left(\frac{1}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\beta_{k,l,p}, \frac{\sigma_p^2}{\sqrt{K^2}}\right). \quad (6)$$

3.4 Hyperparameter settings

In our experiments we use $\alpha = 1$ and use Algorithm 8 from [16] with 5 extra clusters drawn from the prior.We set $\alpha_\sigma = 5$ and $\beta_\sigma = 1$ so that in the presence of little data we encourage shrinkage towards the upper level parameters μ. As with other models, we note that the predictive accuracy of the model can depend on the hyperparameters.

4 Simulation

We check our model fitting procedure using a small synthetic data set involving 10 nodes from 2 groups where 1) within group communication is more likely, 2) events among members in the first group are more likely to be reciprocated (i.e. a positive `ab-ba` effect), and 3) events among members of the second group are more likely to be followed by an event with the same sender (i.e. a positive `ab-ay` effect). The specification of **s** is therefore $\mathbf{s}(t, i, j, \mathcal{A}_t) = [s_0, s_1(t, i, j, \mathcal{A}_t), s_2(t, i, j, \mathcal{A}_t)]$. For the synthetic data set we use parameter vectors $\boldsymbol{\beta}_{1,1} = (0, 3, 0)$, $\boldsymbol{\beta}_{1,2} = \boldsymbol{\beta}_{2,1} = (-1, 0, 0)$, and $\boldsymbol{\beta}_{2,2} = (0, 2, 0)$. Data is generated by sequentially computing $\lambda_{ij}(t_m|\cdot)$ for all $(i, j) \in \mathcal{R}$, drawing $t_{m+1} - t_m \sim \text{Exp}(\sum_{ij} \lambda_{ij}(t_m|\cdot))$, and drawing the dyad $(i, j) \sim \text{Categorical}(\lambda_{ij}(t_m|\cdot)/\sum_{ij} \lambda_{ij}(t_m|\cdot))$.

Though the dyad counts for the synthetic data set suggest a stochastic blockmodel (as seen in Figure 3), the center plot shows each block has empirical differences in their dynamics. Intensities for reciprocal actions among nodes in block 1 are $e^3 \approx 20$ times greater, intensities for turn-taking actions among nodes in group 2 are e^2 times greater, and intensities for dyadic interactions between the two groups have a multiplicative effect of e^{-1} and thus occur less often. Fitting the model with $K = 2$ has similar predictive accuracy as the true model (see Table 2), recovers the true latent classes, and the posterior credible intervals of the parameters cover the true parameter values (see Figure 3c).

5 Model fitting and experiments

A variety of real world data sets are used to explore the efficacy of the model. The following data sets are sequences of dyadic events, where each event has a sender, recipient, and timestamp. For each data set, we hold out the final M_{test} events for evaluation of the model.

- Classroom [17]: 445 directed communication among 27 people a high school classroom collected via participant observation ($M_{test} = 145$).
- University email [18]: 3300 dyadic emails among 88 users with at least 30 emails ($M_{test} = 1300$).
- Enron email [19]: 4000 dyadic emails among 141 individuals between July 2001 and August 2001 ($M_{test} = 1000$).
- Twitter direct messages: Tweets from Twitter.com occurring between from May 11, 2009 to January 26, 2012 that contained the hashtag `#rstats`. This hashtag is used to denote messages pertaining to the R statistical computing environment and sometimes statistical discussion more generally. We collect dyadic events by selecting tweets beginning with the `@` symbol (called a *mention*), and mark the first mentioned user as the recipient. Of 28337 total tweets in this time period, 3926 were directed events among a total of 1079 users. We use a subset 4330 events among 487 users who participated in more than one event ($M_{test} = 1330$).[1]
- MIT Reality Mining [20]: 2000 phone calls among the 89 recipients between October 2001 and February 2002 ($M_{test} = 1000$).

5.1 Model-based exploratory analysis

Figure 4 uses a fitted model to provide an example of the differences in the dynamics that can exist between two groups of dyads with similar rate of occurrence. We focus on the interactions between members of block 1 to members of block 3; 179 events occurred originating from a member of group 3 and 178 originated from group 1. For this example we fit a model with an intercept s_0, `ab-ba` effects s_1, and `ab-by` effects s_3. The two panels on the left show the variation in $s_0(t_M, i, j, \mathcal{A}_t)$, $s_1(t_M, i, j, \mathcal{A}_t)$ and $s_3(t_M, i, j, \mathcal{A}_t)$ across the dyads (i, j) in blocks $(1, 3)$ and $(3, 1)$, respectively. The boxplots on the right describe the posterior samples of the corresponding parameters $\beta_{1,3,p}$ and $\beta_{3,1,p}$. Under the model, events from group 3 to group 1 occur roughly e^1 times as often. Similarly, block (3,1) has a higher propensity for `ab-by` transitions under the model. The model has learned that this structure exists in the data, and furthermore our inferences from the parameter estimates are sensible given the observed statistics.

[1]This dataset will be made publicly available.

Dataset	unif	marg	online	BM	K^*=1	K^*=2	K^*=3	K^*=10
Synthetic	-0.741	-0.741	-0.400	-0.391	-0.211	0.190	**0.194**	0.192
Classroom	-5.379	-3.837	-3.320	-3.404	-3.023	**-2.960**	-3.087	-3.203
University Email	-8.764	-7.729	-6.661	-7.594	-6.013	-6.029	-5.995	**-5.977**
Enron Email	-9.355	-9.657	-7.593	-8.425	-7.025	-6.860	**-6.835**	-7.264
Mobile Phone Calls	-9.612	-7.756	-6.607	-9.619	-6.783	-7.417	**-6.107**	-6.605
Twitter Dir. Messages	-5.106	-3.662	-4.216	-2.962	-3.170	-2.266	**-2.016**	-4.432

Table 2: Comparing mean loglikelihood for each event across methods for each dataset. Larger values are better. See text for details.

κ	Dataset	unif	marg	online	BM	K^*=1	K^*=2	K^*=3	K^*=10
5	Synthetic	0.000	0.228	0.228	0.228	0.207	**0.234**	0.069	0.186
	Classroom	0.000	0.000	0.020	0.002	0.047	0.047	0.047	0.026
	University Email	0.000	0.029	0.029	0.007	0.029	0.030	0.044	**0.047**
	Enron Email	0.002	0.086	0.085	0.000	0.053	0.065	0.088	0.088
	Mobile Phone Calls	0.010	0.023	0.020	0.024	0.028	**0.163**	0.162	0.157
	Twitter Dir. Messages	0.000	0.000	0.000	0.000	0.030	**0.007**	0.006	0.000
20	Synthetic	0.000	0.310	0.283	0.303	0.331	**0.338**	0.234	0.317
	Classroom	0.000	0.034	0.042	0.003	0.062	0.063	**0.077**	0.058
	University Email	0.000	0.029	0.029	0.016	0.036	0.038	**0.069**	0.060
	Enron Email	0.002	0.116	0.135	0.000	0.127	0.116	0.138	**0.182**
	Mobile Phone Calls	0.024	0.027	0.042	0.046	0.056	0.260	0.261	**0.262**
	Twitter Dir. Messages	0.000	0.000	0.020	0.000	**0.047**	0.041	0.045	0.031

Table 3: Comparing recall at cutoff 5 and 20 across methods for each test data set. Larger values are better. See text for details.

5.2 Prediction experiments

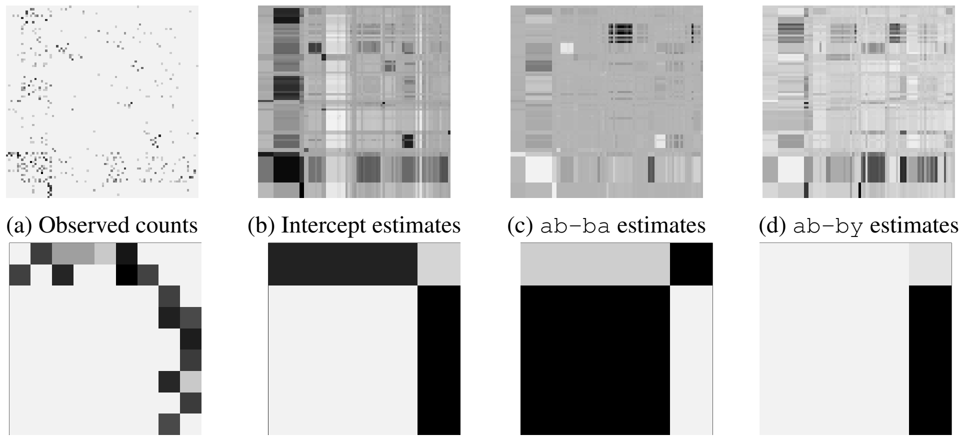
We evaluate the predictive ability of the fitted models by comparing models based on the loglikelihood and recall for held-out data. The loglikelihood of the test set is computed sequentially using Equation 2 where $\hat{\lambda}_{i,j}(t_m) = \frac{1}{L}\sum_l \lambda_{i,j}(t_m|\boldsymbol{\beta}^{(l)}, \mathbf{z}^{(l)}, \mathcal{A}_t)$ is averaged using L posterior samples given the training data.

In addition, we compute recall to evaluate whether the next observed event is among the most likely according to the model. At each event m we sort the predicted intensities of all possible events in decreasing order, find the rank of the observed event in the list of predicted intensities, and compute the mean number of events ranking above cutoffs $\kappa = 5$ and $\kappa = 20$.

Several baselines are included for comparison: `uniform` places uniform probability on all possible dyads, `online` ranks events at time t by the number of times the dyad has occurred previously $r_{online}(m,i,j) = \sum_{m:t_m<t} I(i_m = i, j_m = j)$, and `marginal` uses the product of the observed marginal frequencies $r_{marg}(m,i,j) = \sum_{m:t_m<t} I(i_m = i)\sum_{m:t_m<t} I(j_m = j)$. Finally, `BM` is a stochastic blockmodel (i.e. our model with only an intercept term). Note for processes that are homogeneous over time, `online` should do well with large amounts of data while `marginal` and `BM` can capture individual heterogeneity and group level heterogeneity in overall activity.

Our method jointly models *which* dyads occur and *when* they occur. To compare the above baselines to our model using the likelihood of observed data, we assume each dyad is a Poisson process with estimated rate $\hat{\lambda}_{i,j}(t_m) = \frac{M}{t_M}\frac{r_b(i,j)+\xi}{\sum_{ij} r_b(i,j)+\xi}$, where $r_b(i,j)$ is the statistic a baseline (described above) and $\xi = 1$ is a smoothing parameter.

In order to investigate the role of the number of clusters, we introduce an upper bound K^* on the number of clusters during the fitting process. In Table 2 we compare the loglikelihood on held-out test data while varying K^*. A relational event model (i.e. our model with $K = 1$) outperforms the baselines on these data sets. For the synthetic data example introduced in Section 4 the fitted models with $K > 1$ have predictive performance comparable to the true model; a standard relational event model does not perform as well because it does not have the flexibility to model the dynamics of each block separately. For the classroom data set we see only a slight improvement with $K > 1$,



(a) Observed counts (b) Intercept estimates (c) ab−ba estimates (d) ab−by estimates

(e) Observed counts (f) Intercept estimates (g) ab−ba estimates (h) ab−by estimates

Figure 4: Comparing observed counts and parameter estimates. Darker values are larger. Estimates are rescaled posterior means $(\hat{\beta}_{z_i,z_j,p} - \hat{\mu}_p)/\hat{\sigma}_p$ for each dyad (i, j). Learned parameters suggest heterogeneity exists in both total activity (b) as well as dynamics, as seen in (c) and (d). Figures (e-h) give a zoomed-in view of groups 1 and 3, showing that while the overall rate of (1,3) and (3,1) events is similar, the tendency for ab−by transitions differs.

while with the mobile phone calls the $K = 3$ model performs best. In Table 3 we include the corresponding results for the recall experiment at a cutoff of 5 and 20, respectively. The results there are similar.

6 Discussion

We have here introduced a family of relational event models that can flexibly capture heterogeneity in the underlying interaction dynamics. Our approach generalizes traditional, static notions of stochastic equivalence on nodes (such as stochastic blockmodels) to the dynamic context. The proposed model family posits the existence of groups of nodes, such that all members of a group are governed by the same dynamic process, and groups are differentiated by having different dynamic processes.

Our approach has the ability to uncover systematic differences in dynamic behaviors among subsets of nodes, even in the absence of differences in marginal interaction rates. The analyses of Section 5 show that this model family has improved predictive accuracy over baseline methods on real data with respect to ranking tasks and the likelihood of unobserved data. In particular, our proposed approach leads to improved predictive accuracy when compared to both (a) relational event models that lack latent clusters, and (b) stochastic blockmodels for count data, both of which are special cases of our model family. Though prediction is not the main focus of our effort, these results provide evidence that having latent structure (i.e. $K > 1$) can lead to improvements in predictive power. However, our prediction results also show that it is easy to overfit with this model family. As K increases beyond 2 (e.g., for the unrestrained CRP prior), these models are prone to overfit. This is not surprising, given that the number of fitted parameters scales as the square of the number of components K in the model. A natural direction worth exploring for this family is a class of priors that are more resistant to overfitting (e.g., by imposing more structure on the parameters within each group).

Another direction to explore is that of including different statistics in the specification of $\mathbf{s}(t, i, j, \mathcal{A}_t)$, and using our approach to study how the roles of these statistics vary across nodes. One potentially interesting extension, analogous to [21], would be to allow the latent class z_i to be drawn from node-specific membership vectors π_i after each change point. Allowing underlying dynamics to be governed by a latent membership structure opens the door to a wide range of possibilities, with many opportunities for further elaboration.

References

[1] Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi. A Survey of Statistical Network Models. *Foundations and Trends in Machine Learning*, 2(2), 2009.

[2] Stanley Wasserman and Katherine Faust. *Social network analysis*. Cambridge University Press, 1994.

[3] Krzysztof Nowicki and Tom A B Snijders. Estimation and Prediction for Stochastic Blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

[4] Aalen, Odd O., Ornulf Borgan, and Hakon K. Gjessing. *Survival and Event History Analysis: A Process Point of View*. Springer, 2008.

[5] Carter T. Butts. A Relational Event Framework for Social Action. *Sociological Methodology*, 38(1):155–200, July 2008.

[6] Duy Q Vu, Arthur U. Asuncion, David R. Hunter, and Padhraic Smyth. Continuous-Time Regression Models for Longitudinal Networks. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems*, pages 1–9, 2011.

[7] D. R. Gibson. Participation Shifts: Order and Differentiation in Group Conversation. *Social Forces*, 81(4):1335–1380, June 2003.

[8] D McFadden. Conditional logit analysis of qualitative choice behavior. 1973.

[9] Katsuhiko Ishiguro, Tomoharu Iwata, and Joshua Tenenbaum. Dynamic Infinite Relational Model for Time-varying Relational Data Analysis. In *Advances in Neural Information Processing Systems 22 (NIPS)*, pages 1–9.

[10] Abel Rodriguez. Modeling the dynamics of social networks using Bayesian hierarchical blockmodels. *Statistical Analysis and Data Mining*, pages 1–23, 2011.

[11] Ulrik Brandes, Jürgen Lerner, and Tom A.B. Snijders. Networks Evolving Step by Step: Statistical Analysis of Dyadic Event Data. *2009 International Conference on Advances in Social Network Analysis and Mining*, pages 200–205, July 2009.

[12] Patrick O Perry and Patrick J Wolfe. Point process modeling for directed interaction networks. *New York*, 2011.

[13] Christoph Stadtfeld and Andreas Geyer-Schulz. Analyzing event stream dynamics in two-mode networks: An exploratory analysis of private communication in a question and answer community. *Social Networks*, pages 1–15, October 2011.

[14] Duy Q Vu, Arthur U. Asuncion, David R. Hunter, and Padhraic Smyth. Dynamic Egocentric Models for Citation Networks. *Proceedings of the 28th International Conference on Machine Learning*, pages 857–864, 2011.

[15] Asela Gunawardana, Steve Meek, and Francis Morris. A Model for Temporal Dependencies in Event Streams. *NIPS*, March 2011.

[16] RM Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

[17] Daniel A. McFarland. Student Resistance: How the Formal and Informal Organization of Classrooms Facilitate Everyday Forms of Student Defiance. *American Journal of Sociology*, 107(3):612–678, November 2001.

[18] Jean-Pierre Eckmann, Elisha Moses, and Danilo Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences of the United States of America*, 101(40):14333–7, October 2004.

[19] Bryan Klimt. Introducing the Enron corpus: a new data set for email classification research. *European Conference on Machine Learning*, pages 217–226, 2004.

[20] Nathan Eagle and AS Pentland. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36):15274–15278, 2009.

[21] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed Membership Stochastic Blockmodels. *JMLR*, 9:1981–2014, September 2008.