Stochastic blockmodeling of relational event dynamics

Christopher DuBois
Department of Statistics
University of California, Irvine

Carter T. Butts
Department of Sociology
Department of Statistics
Institute for Mathematical and Behavioral Sciences
University of California, Irvine

Padhraic Smyth
Department of Computer Science
Department of Statistics
University of California, Irvine

Abstract

Several approaches have recently been proposed for modeling of continuous-time network data via 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 ideas from stochastic blockmodeling and continuous-time network models 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 both baseline models and relational event models that lack latent structure.

Appearing in Proceedings of the 16^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2013, Scottsdale, AZ, USA. Volume 31 of JMLR: W&CP 31.

1 Introduction

Statistical methods for analyzing network data are 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 [2, 3, 4]), with the common goals 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 social network analysis [5].

In the cross-sectional case, *stochastic blockmodels* [6] 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 Figure 1 for example. 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. 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 of Figure 1. Here, there is no systematic difference in marginal rates between the two groups.

There is, however, a difference to be detected in the top right panel: events from a member of group B to a member of group A tend to be reciprocated sooner

than events from group A's members to group B's members. These differences are depicted by the red dots in the lower two plots, where large y-values indicate a propensity for reciprocated events among a specific pair of blocks, as labeled on the x-axis. The fact that the two plots are identical shows this behavior is present in both the rate-differentiated and rate-undifferentiated examples. Even in the absence of marginal differences in propensity to communicate (i.e. the rate-undifferentiated case), groups A and B are distinctive in terms of their dynamic behavior. 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 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 [2, 7]. 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 one learns 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.

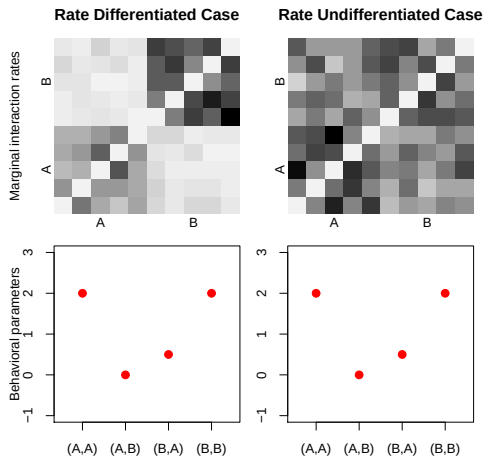


Figure 1: Illustration of differentiation by rate of interaction versus by dynamic behavior. The left column shows an example where within-block communication is large, and there is a higher tendency for reciprocity within a block. In the situation where the two groups are not differentiated by rate, 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.

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

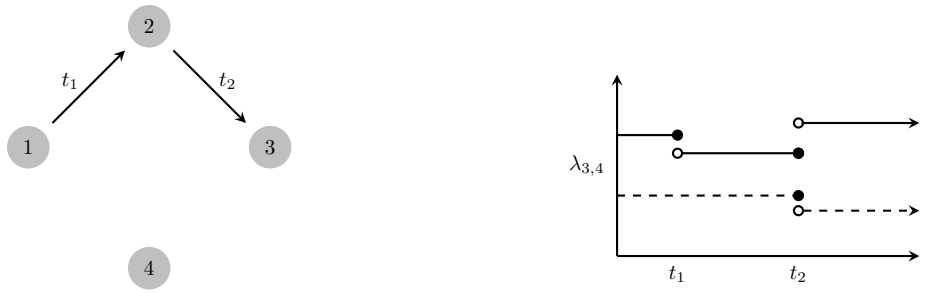
$$\mathcal{L}(\mathcal{A}|\theta) = \prod_{m=1}^{M} \lambda(t_m) \exp\left\{-\int_0^{t_M} \lambda(s)ds\right\} = \prod_{m=1}^{M} \lambda(t_m) \prod_{k=1}^{|\tau|} \exp\left\{-(\tau_k - \tau_{k-1})\lambda(\tau_k)\right\} \quad (1)$$

The above approach is easily extended to marked point processes where each event contains additional information. In Figure 2a we show a simple example of *relational events* occurring among N nodes, where 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* is defined as the set of all possible dyadic events. The study of human communication often involves such data (e.g., phone calls, online interaction, etc.) where the nodes represent people and each event represents one person communicating with another.

In the proposed approach we assume that events involving dyad $(i, j) \in \mathcal{R}$ follow a Poisson process with intensity $\lambda_{i,j}(t|\cdot)$ that is piecewise constant with knots $\tau = \{t_0, \ldots, t_M\}$ at each event. The intensity $\lambda_{i,j}(t|\cdot)$ depends on the previous history of events $\mathcal{A}_t = \{(t_m, i_m, j_m) : t_m \in [0, t)\}$. The likelihood of an observed event history $\mathcal{A}_{t_M}$ (extending from time 0 to the time t_m of the final event and denoted $\mathcal{A}$ for convenience) is

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

Here we aim to learn about the dynamics within and between subsets of nodes. Specifically, we propose allowing each node i to belong to a latent cluster z_i while



(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) Examples of intensity functions consistent with (a): $\lambda_{1,2}(t)$ (solid) and $\lambda_{3,4}(t)$ (dashed).

using a log linear model for the intensity functions

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

where for each pair of clusters (k, l) we have a parameter vector $\boldsymbol{\beta}_{k,l} \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 I)$ 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 $\lambda_{i,j}(t)$ 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.

We allow the blocks to share information by placing a hierarchical prior on the collection of $\boldsymbol{\beta}_{k.l}$ parameters where $\mu_p \propto N(0, 1)$ and $\sigma_p^2 \sim \text{Inv-Gamma}(\alpha_\sigma, \beta_\sigma)$, where $1 \leq p \leq P$. The cluster assignments are given a nonparametric prior $z_i \sim \text{CRP}(\alpha)$, 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, when a node's out-degree is indicative of future activity, one may wish to include s_4 (sender out-degree) in the specification of $\mathbf{s}$. We use $\mathbf{1}$ to denote the indicator function where $\mathbf{1}(A) = 1$ if A is true, and 0 otherwise. We normalize the counts associated with the statistics 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 [2, 8].

The rest of the statistics in Table 1 aim to capture various types of *participation shift* that play a role in conversational norms [9]. For example, an `ab-ba` effect indicates an increased propensity for reciprocity where the event (a, b) is followed by (b, a). Another example is a *turn-taking* effect where an event (a, b) is followed by b initiating an event with an individual other than a, denoted `ab-by`. These statistics are binary and simply indicate whether or not the event in question can be classified as an example of that particular transition. Although in this paper we only use the statistics in Table 1, in general one can use any quantity that is computed using the previous history of events or known covariates about nodes or dyads. The only restriction within the proposed framework 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 [6] for static networks which models the probability of a dyad as $p(y_{i,j}) = \text{logit}^{-1}(\eta_{z_i,z_j})$ where η_{z_i,z_j} is interpreted as a mixing rate between group z_i and group z_j. In our 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

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

Alternatively, if one models only the order of the events (ignoring the times at which they occur), we obtain a

conditional logit model

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

The functional form is similar to conditional logit models used for discrete choice data [10], though here the possible choices are all the dyads in $\mathcal{R}$.

Stochastic blockmodels have been extended to longitudinal network data involving a sequence of networks occurring at discrete times [11, 12]. The temporal extensions to block models, ERGMs, etc., are only applicable to longitudinal settings where an entire network is observed at discrete time steps and where relationships between nodes are temporally extensive. These longitudinal extensions to stochastic blockmodels are not directly applicable to the data that we model, namely sequences of timestamped, dyadic events reflecting temporally non-extensive interaction.

The proposed model is an extension of recent work modeling event-based social network data using an event history approach [2, 3, 4, 13, 8, 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, using decision trees to 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. Other approaches include using Hawkes processes to model reciprocation among clusters of individuals [16], allowing previous events among those individuals to alter the current intensity of interactions—but this approach cannot flexibly incorporate other types of effects as we do here.

3 Inference

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

3.1 Sampling z

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

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

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

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.

3.2 Sampling $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}^2$

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

$$p(\boldsymbol{\beta}_{k,l}|\mathcal{A}, \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \propto p(\boldsymbol{\beta}_{k,l}|\boldsymbol{\mu}, \boldsymbol{\sigma}) p(\mathcal{A}|\mathbf{z}, \boldsymbol{\beta})$$

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

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

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. We use a conjugate prior $(\mu_p, \sigma_p^2) \sim NIG(0, \nu, \alpha_\sigma, \beta_\sigma)$ that allows for straightforward sampling from the posterior of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ conditioned on $\boldsymbol{\beta}$, though we omit the details here due to space constraints.

3.3 Hyperparameter settings

In our experiments we use $\alpha = 1$ and use Algorithm 8 from [17] with 5 extra clusters drawn from the prior.We set $\alpha_\sigma = 5$, $\beta_\sigma = 1$, and $\nu = 1$ so that in the presence of little data we encourage shrinkage towards the upper level parameters $\boldsymbol{\mu}$. As with other Bayesian latent variable models, we note that the predictive accuracy of the model can depend on the hyperparameters.

3.4 Scalability

Likelihood computation depends on the number of changepoints (i.e. knots) in all of the intensity functions being modeled. We take advantage of our restriction on the types of statistics $\mathbf{s}$ to reduce the computational complexity of computing our likelihood as

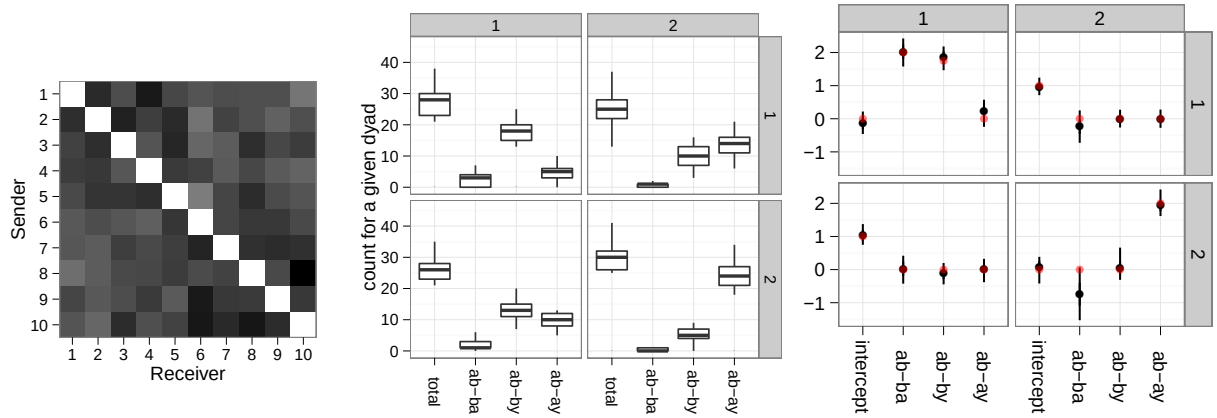
$$\mathcal{L}(\mathcal{A}_{t_M}|\theta) = \prod_{m=1}^{M} \lambda_{i_m,j_m}(t_m|\cdot) \times \prod_{(i,j)\in\mathcal{R}_{i_m,j_m}} \exp\{-(t_m - \tau_{m,i,j})\lambda_{i,j}(t_m|\cdot)\}$$

where event m is the dyad (i_m, j_m), $\tau_{m,i,j}$ is the time of the changepoint for $\lambda_{i,j}(t|\cdot)$ prior to the mth event, and $\mathcal{R}_{i,j}$ is the set of dyads whose intensity changes if (i,j) occurs.

By limiting the number of changepoints, computing the likelihood $p(\mathcal{A}|\mathbf{z}, \boldsymbol{\beta})$ for Gibbs sampling z_r is

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

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.



(a) (b) (c)

Figure 3: Illustration of 2500 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).

$O(|\mathcal{U}_r| \cdot P \cdot N)$, avoiding a factor of N^2. In practice, we precompute $\tau_{m,i,j}$ and $\mathbf{s}(t_m, i, j, \mathcal{A}_t)$ for all m, i, and j. This can be done in one pass through the data set.

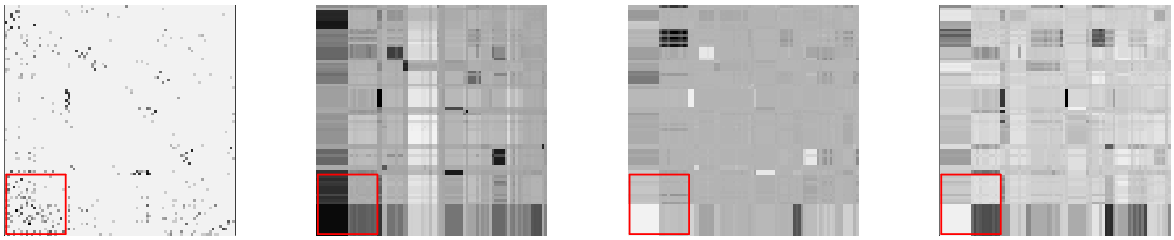
4 Simulation

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 $\mathbf{s}$ is $\mathbf{s}(t,i,j,\mathcal{A}_t) = [s_0, s_1(t,i,j,\mathcal{A}_t), s_2(t,i,j,\mathcal{A}_t)]$. 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 $\lambda_{i,j}(t_m|\cdot)$ for all $(i,j) \in \mathcal{R}$, drawing $t_{m+1} - t_m \sim \text{Exp}(\sum_{i,j} \lambda_{i,j}(t_m|\cdot))$, and drawing the dyad $(i,j) \sim \text{Categorical}(\lambda_{i,j}(t_m|\cdot)/\sum_{i,j} \lambda_{i,j}(t_m|\cdot))$.

Though the dyad counts for the synthetic data set do not reveal any block structure (as seen in Figure 3), the center plot shows each block has empirical differences in their dynamics. Intensities for reciprocal actions among nodes in group 1 are $e^2 \approx 7.4$ times greater, intensities for turn-taking actions among nodes in group 1 are $e^{1.75}$ times greater, intensities for turn-continuing 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. Fitting the model with $K = 2$ recovers the true latent classes, and the posterior credible intervals of the parameters cover the true parameter values (see Figure 3c). As discussed in Figure 1, a standard stochastic block model fit to the aggregated counts in Figure 3a is unable to identify the latent group structure that is present.

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



(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 for the University data set. 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) show a magnified view of the first two blocks; while the overall rate of (1,2) and (2,1) events is similar, the tendency for `ab-by` transitions differs.

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.

- Classroom [18]: 445 directed communication among 27 people a high school classroom collected via participant observation ($M_{test} = 145$).
- University email [19]: 3300 dyadic emails among 88 users with at least 30 emails ($M_{test} = 1300$).
- Enron email [20]: 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 [21]: 2000 phone calls among the 89 recipients between October 2001 and February 2002 ($M_{test} = 1000$).

[1]Data availability limited by Twitter terms of service.

5.1 Model-based exploratory analysis

Parameter estimates of the proposed model can reveal the presence of both symmetries and asymmetries in the block structure of communication rates among the nodes of the network. The asymmetries that the model reveals are often of interest. For example, an asymmetry in intercept estimates for a particular pair of groups implies that one group tends to be the sender and the other tends to be the recipient. Asymmetries for other effects are similarly interpretable. Consider interactions initiated by a member of block B_1 and directed to a member of block B_2. If the estimate for the `ab-ba` parameter is larger for the set of (B_1, B_2) interactions than for (B_2, B_1) interactions, then the model suggests that those in block B_2 tend to respond to emails from those in B_1 than vice versa.

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. In this particular example we used the University data set and fit the proposed model with an intercept s_0, `ab-ba` effects s_1, and `ab-by` effects s_3. In Figure 4a we show the observed number of times each dyadic event occurred over the course of the data set, while the intercept estimates in Figure 4b reveal the general propensity at which two blocks tend to communicate. The rows and columns have been sorted according to block membership.

An interesting example of asymmetric dynamics exists in the University data between block 1 and block 2. Figure 4e-h focuses on the interactions between members of block 1 to members of block 2, providing a close-up version of the respective areas surrounded by a red square in Figure 4a-d. Under the model we find that communication from block 2 to block 1 has a higher propensity for `ab-by` transitions under the model. Though such asymmetries are unsurprising from a sociological point of view, our model has allowed us to examine this tendency while adjusting for the overall propensity for interaction.

5.2 Prediction experiments

We evaluate the predictive ability of the fitted models by comparing models based on both a) the loglikelihood and b) recall, for held-out data. Throughout the experiments we use all of the statistics found in Table 1. 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.

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 dyads,
- `online` ranks events at time t by the number of times the dyad has occurred previously $r_{online}(m,i,j) = \sum_{m:t_m<t} \mathbf{1}(i_m = i, j_m = j)$,
- `marginal` uses the product of the observed marginal frequencies $r_{marg}(m,i,j) = \sum_{m:t_m<t} \mathbf{1}(i_m = i)\sum_{m:t_m<t} \mathbf{1}(j_m = j)$.
- `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. However, neither `marginal` nor `BM` can capture temporal structure such as that seen in Figure 4c and Figure 4d. These baselines can be quite competitive when evaluated predictively and can also provide insight into the performance of more complex models on the same data sets.

We additionally perform experiments that evaluate the likelihood of future data under our model. As our method jointly models *which* dyads occur and *when* they occur, any alternative baseline method needs to also model these two aspects. In this vein, we extend the above baselines by assuming each dyad is a Poisson process with estimated rate

$$\hat{\lambda}_{i,j}(t_m) = \frac{M}{t_M}\frac{r_b(i,j)+\xi}{\sum_{i,j} r_b(i,j)+\xi}$$

where $r_b(i,j)$ is the statistic for a given baseline (described above) and $\xi = 1$ is a smoothing parameter. There are a large number of other possible baselines that one might use, but most are not directly applicable as they do not work with events in continuous time.

In order to investigate the role of the number of clusters, we use an upper bound K^* on the number of clusters during the fitting process. Note that the model with $K^* = 1$ is an important baseline — it is simply a relational event model with no groups.

5.3 Results

In Table 2 we compare the loglikelihood of held-out test data under our proposed model to each of these baselines. There are two broad conclusions that are evident across all of the data sets: (1) the relational event models outperform all of the baselines and (2) the proposed method (i.e. $K^* > 1$) outperforms a relational event model lacking block structure (i.e. $K^* = 1$), indicating the presence of interaction heterogeneity in the datasets.

In Table 3 we include the corresponding results for the recall experiment at a cutoff of 5 and 20, respectively, with the same broad conclusions as for the loglikelihood experiment. As an example of these types of predictions, using the Mobile data we are making a prediction of the next event among 89^2 possible events—the next event is among the 5 highest-ranked edges 16.3 percent of the time by the K^*=2 model. This is 1291 times more often than predicting events with our uniform baseline. Note that the absolute numbers for the baselines (e.g., zero in the 1st 3 decimal places for uniform) are very low because of the very large number of outcomes (e.g, 89^2). For the ranking task on Enron there are small systematic improvements relative to all baselines (including $K^* = 1$), while on the phone and University data there are large systematic improvements. Although prediction of the next relational event in a large group is a difficult task, our models show respectable performance despite a lack of covariates or other information on the actors or groups. Such information can also be employed in our framework, where available, by appropriate choice of s.

	Baseline				Relational event model			
Dataset	`unif`	`marg`	`online`	`BM`	K^*=1	K^*=2	K^*=3	K^*=10
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	-7.374	-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.

		Baseline				Relational event model			
κ	Dataset	`unif`	`marg`	`online`	`BM`	K^*=1	K^*=2	K^*=3	K^*=10
5	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	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.

6 Discussion

In this paper we introduced a family of relational event models that can flexibly capture heterogeneity in underlying interaction dynamics of network data over time. 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.

The proposed 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 can show 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 work, 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 (e.g., see the non-bold entries in the rightmost column of Tables 2 and 3). 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 $\mathbf{s}(t, i, j, \mathcal{A}_t)$, modifying the prior on each block's β that induces some sparsity in the parameter estimates, and using our approach to study how the roles of these statistics vary across nodes. One potentially interesting extension, analogous to [22], 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 for further exploration.

Acknowledgements

This work was supported in part by a NDSEG Fellowship (CLD), a Google Research Award (PS), and by the Office of Naval Research/Multidisciplinary University Research Initiative under grant number N00014-08-1-1015.

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

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

[4] Patrick O Perry and Patrick J Wolfe. Point process modeling for directed interaction networks. *arXiv: 1011.1703*, pages 1–23, 2010.

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

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

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

[8] 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*, 2011.

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

[10] D McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 1(2):105–142, 1973.

[11] Katsuhiko Ishiguro, Tomoharu Iwata, and Joshua Tenenbaum. Dynamic Infinite Relational Model for Time-varying Relational Data Analysis. In *Advances in Neural Information Processing Systems*, 2010.

[12] Abel Rodriguez. Modeling the dynamics of social networks using Bayesian hierarchical blockmodels. *Statistical Analysis and Data Mining*, pages 1–23, 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 International Conference on Machine Learning*, 2011.

[15] Asela Gunawardana, Christopher Meek, and Puyang Xu. A Model for Temporal Dependencies in Event Streams. *Advances in Neural Information Processing Systems*, March 2011.

[16] Charles Blundell, K Heller, and J Beck. Modelling Reciprocating Relationships with Hawkes Processes. *Advances in Neural Information Processing Systems*, 2012.

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

[18] 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.

[19] 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.

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

[21] 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.

[22] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed Membership Stochastic Blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, September 2008.