

# The Hyperedge Event Model

Bomin Kim\*, Aaron Schein†, Bruce A. Desmarais‡, and Hanna Wallach§

**Abstract.** We introduce the hyperedge event model (HEM)—a generative model for events that can be represented as directed edges with one sender and one or more receivers or one receiver and one or more senders. We integrate a dynamic version of the exponential random graph model (ERGM) of edge structure with a survival model for event timing to jointly understand who interacts with whom, and when. The HEM offers three innovations with respect to the literature—first, it extends a growing class of dynamic network models to model hyperedges. The current state of the art approach to dealing with hyperedges is to inappropriately break them into separate edges/events. Second, the HEM involves a novel receiver-selection distribution that is based on established edge formation models, but assures non-empty receiver lists. Third, HEM integrates separate, but interacting, equations governing edge formation and event timing. We use the HEM to analyze emails sent among department managers in Montgomery County government in North Carolina. Our application demonstrates that the model is effective at predicting and explaining time-stamped network data involving edges with multiple receivers. We present an out-of-sample prediction experiment to illustrate how researchers can select between different specifications of the model.

**MSC 2010 subject classifications:** Primary 60K35, 60K35; secondary 60K35.

**Keywords:** dynamic network model, hyperedge, continuous time model, email data analysis.

## 1 Introduction

Processes that arise as time-stamped directed interactions are common in the social, natural, and physical sciences. The data produced by such processes can be represented as dynamic directed networks—an object that has given rise to the development of several statistical model families. For example, stochastic actor-oriented models (SAOMs) (Snijders, 1996; Snijders et al., 2007) characterize that networks are evolving as the senders of edges decide to add an edge to or remove an edge from a network one edge at a time, one sender at a time. Event-based network models (Butts, 2008; Vu et al., 2011; Hunter et al., 2011; Perry and Wolfe, 2013) provide a general framework for modeling the realization of edges that occur as instances in continuous time streams of events. These models are flexible enough to specify a generative model that accounts for the ways in which past network structure begets future network structure (e.g., nodes that sent a lot of edges in the past will send a lot in the future, if node  $i$  sent a tie to node  $j$  recently, then node  $j$  will send one to node  $i$  in the near future). Thus, dynamic net-

---

\*Department of Statistics, Pennsylvania State University [bzk147@psu.edu](mailto:bzk147@psu.edu)

†College of Information and Computer Sciences, UMass Amherst [aschein@cs.umass.edu](mailto:aschein@cs.umass.edu)

‡Department of Political Science, Pennsylvania State University [bdesmarais@psu.edu](mailto:bdesmarais@psu.edu)

§ Microsoft Research NYC [hanna@dirichlet.net](mailto:hanna@dirichlet.net)

work models are very useful to understand which traits and behaviours are predictive of interactions.

A major limitation of existing dynamic network models is that they apply to edges with one sender and one receiver. A lot of dynamic network data arise as “hyperedges” (Karypis et al., 1999; Ghoshal et al., 2009; Zlatić et al., 2009; Zhang and Liu, 2010) that include one sender and multiple receivers or one receiver and multiple senders. For example, email networks (Newman et al., 2002) arise with one sender and possibly multiple receivers; networks formed between neurons via axons and dendrites have one sender and multiple receivers (axons) or one receiver and multiple senders (dendrites) (Partzsch and Schüffny, 2012). Networks formed through the analysis of legislative bill cosponsorship (Fowler, 2006) arise with multiple senders (cosponsors) and one receiver (sponsor); and networks formed through the analysis of economic sanctions (Cranmer et al., 2014) between countries arise with multiple sending countries and one target country. Existing models require researchers to alter hyperedge data to fit with the pairwise edge structure of the model. For instance, Perry and Wolfe (2013) treat multicast interactions—one type of directed hyperedge which involves one sender and one or more receivers—via duplication (i.e., obtain pairwise interactions from the original multicast), to construct approximate likelihood function in their inferential framework for model parameters. Similarly, Fan and Shelton (2009) duplicates emails sent from one sender to one or more receivers and randomly jitter the sent times in order to avoid the violation of the assumption that two events cannot occur at the exact same time.

We develop a statistical dynamic network model, which we term the hyperedge event model (HEM), that integrates the two components that govern time-stamped event formation: 1) the formation of the vertices that are incident to the hyperedge, and 2) the timing of the event in which the edge manifests. In what follows, we introduce the HEM by defining a generative process for time and hyperedge attributed event data (Section 2), derive the conditional posteriors for Bayesian inference and present tests of our software implementation (Section 3). After defining the structure of and deriving Bayesian inference methods for HEM, we demonstrate the model’s applicability by analyzing data from a corpus of internal county government emails. In the application we illustrate model selection, posterior predictive checks, and exploratory analysis (Section 4). Finally, the paper finishes in Section 5 with conclusion and discussion.

## **2 The hyperedge event model**

In this section we define the HEM by describing a process according to which hyperedge events are generated in continuous time. Data generated under the HEM consists of  $D$  unique edges. A single edge indexed by  $d \in [D]$ , where  $[D]$  denotes a categorical set with  $D$  levels  $[D] = \{1, \dots, D\}$ , is represented by the three components: the sender  $a_d \in [A]$ , an indicator vector of receivers  $\mathbf{r}_d = (u_{dr})_{r=1}^A$ , such that  $u_{dr} = 1$  if  $r$  is a receiver of edge  $d$  and 0 otherwise, and the timestamp  $t_d \in (0, \infty)$ . For simplicity, we assume that edges are ordered by time such that  $t_d \leq t_{d+1}$ . While the model can be applied for two types of hyperedges—edges with (1) one sender and one or more receivers, and (2) one or more senders and one receiver—here we only present the generative process

for those involving one sender and one or more receivers (i.e., multicast). One notable feature of the HEM generative process is that we draw auxiliary variables that serve as candidate data. Data is generated from the HEM through a sampling process applied to the auxiliary variables. The auxiliary variables drawn for edge  $d$  include, for each sender  $a$ , a time increment from edge  $d - 1$  at which sender  $a$  would send edge  $d$ , and an  $A - 1$  length vector indicating which nodes would receive edge  $d$  if it were sent by sender  $a$ . The data generated for edge  $d$  under the HEM corresponds to the sender that would send edge  $d$  the soonest—at the smallest time increment from edge  $d - 1$ . The receivers of edge  $d$  generated under the HEM correspond to those receivers to which the sender with the minimum time increment would have sent edge  $d$ . We explain these steps in more detail below. For hyperedges that involve one receiver and one or more senders, we treat  $a_d$  to be an indicator vector of senders  $\mathbf{a}_d = (u_{dr})_{a=1}^A$  and  $r_d$  to be the single receiver, and then follow the alternative generative process provided in Appendix A, which we derive as a simple reversal of the process used for multicasts (i.e., one sender and multiple receivers).

## 2.1 Candidate receivers

For every possible sender–receiver pair  $(a, r)_{a \neq r}$ , we define the “receiver intensity”—the conditional logit of the probability that edge  $d$  is sent from sender  $a$  to receiver  $r$ —as a linear combination of statistics relevant to the receiver selection process:

$$\lambda_{adr} = \mathbf{b}^\top \mathbf{x}_{adr}, \quad (2.1)$$

where  $\mathbf{b}$  is a  $P$ -dimensional vector of coefficients and  $\mathbf{x}_{adr}$  is a set of receiver selection features.  $\lambda_{adr}$  is the logit of the probability that edge  $d$  is sent from sender  $a$  to receiver  $r$  conditional upon there being at least one other receiver to which sender  $a$  sends edge  $d$ . In the case that there are no other receivers to which sender  $a$  sends edge  $d$ , the logit of the probability that sender  $a$  sends edge  $d$  to receiver  $r$  is  $\infty$ . The features in Equation 2.1 can capture common network processes that include, but are not limited to, popularity, reciprocity, and transitivity, as well as the effects of attribute features defined on the sender, receivers, and/or sender-receiver pairs (e.g., the gender of the sender and/or receiver, an indicator of whether the sender is a supervisor of the receiver’s). In addition, we include an intercept term to account for the average (or baseline) number of receivers. We place a Normal prior  $\mathbf{b} \sim N(\boldsymbol{\mu}_b, \Sigma_b)$ .

As we describe in greater detail below, under the HEM the sender of each edge is selected to be the sender that would send the respective edge with the greatest urgency. As such, under the HEM we assume that each sender would plan on sending each edge to some set of recipients at some time point. For each edge  $d$  and each sender  $a$ , the HEM draws a list of candidate receivers that would be receivers on edge  $d$  if  $a$  were the sender. For an edge  $d$ , we first define an  $A \times A$  matrix  $\mathbf{u}_d$  where the  $a^{th}$  row denotes sender  $a$ ’s receiver vector of zeros and 1’s—i.e., 1’s indicate the nodes to which  $a$  intends to send edge  $d$ . We then assume that each receiver vector  $\mathbf{u}_{ad}$  comes from a modification of the multivariate Bernoulli (MB) distribution (Dai et al., 2013)—a model that has been used to model graphs in which the state of each edge indicator is drawn independently from an edge-specific Bernoulli distribution. In order to avoid drawing edges with no receivers,

we define a probability measure “ $\text{MB}_G$ ” motivated by the Gibbs measure (Fellows and Handcock, 2017). The probability measure we define amounts to a non-empty Gibbs measure, in which the all-zero vector is excluded from the support of the multivariate Bernoulli distribution. As a result, this measure helps us to 1) allow a sender to select multiple receivers for a single edge, 2) force the sender to select at least one receiver, and 3) ensure a tractable normalizing constant for the receiver selection distribution. To be specific, we draw a binary vector  $\mathbf{u}_{ad} = (u_{ad1}, \dots, u_{adA})$

$$\mathbf{u}_{ad} \sim \text{MB}_G(\boldsymbol{\lambda}_{ad}), \quad (2.2)$$

where  $\boldsymbol{\lambda}_{ad} = (\lambda_{ad1}, \dots, \lambda_{adA})$ . In particular, we define  $\text{MB}_G(\boldsymbol{\lambda}_{ad})$  as

$$\Pr(\mathbf{u}_{ad} | \mathbf{b}, \mathbf{x}_{ad}) = \frac{1}{Z(\boldsymbol{\lambda}_{ad})} \exp \left( \log(I(\|\mathbf{u}_{ad}\|_1 > 0)) + \sum_{r \neq a} \lambda_{adr} u_{adr} \right), \quad (2.3)$$

where  $Z(\boldsymbol{\lambda}_{ad}) = \prod_{r \neq a} (\exp(\lambda_{adr}) + 1) - 1$  is the normalizing constant and  $\|\cdot\|_1$  is the  $l_1$ -norm. Again, this is approximately equivalent to assuming that each  $u_{adr}$  is drawn with the probability of 1 being  $\text{logit}(\lambda_{adr})$ , excluding the case when  $u_{adr} = 0$  for all  $r \in [A]$ . This allows efficient inference since we can derive a closed-form normalizing constant which does not require brute-force summation over the support of  $\mathbf{u}_{ad}$  (i.e.,  $\forall \mathbf{u}_{ad} \in [0, 1]^A$ ). We provide detailed derivation steps for the normalizing constant  $Z(\boldsymbol{\lambda}_{ad})$  in Appendix B.

## 2.2 Candidate timestamps

To generate timing for each edge, the HEM first draws a candidate time at which the edge would be sent given the candidate sender and receiver combinations. The timing rate for sender  $a$  and edge  $d$  is

$$\mu_{ad} = g^{-1}(\boldsymbol{\eta}^\top \mathbf{y}_{ad}), \quad (2.4)$$

where  $\boldsymbol{\eta}$  is a  $Q$ -dimensional vector of coefficients with a Normal prior  $\boldsymbol{\eta} \sim N(\boldsymbol{\mu}_\eta, \Sigma_\eta)$ ,  $\mathbf{y}_{ad}$  is a set of event timing features—covariates that could affect timestamps of the edge, and  $g(\cdot)$  is the appropriate link function such as identity, log, or inverse.

In modeling “when,” we do not directly model the timestamp  $t_d$ . Instead, we assume that each sender’s “time increment”—i.e., waiting time to next interaction since  $t_{d-1}$ —is drawn from a specific exponential family distribution. We define the time increment from edge  $d-1$  to edge  $d$  as  $\tau_d$  (i.e.,  $\tau_d = t_d - t_{d-1}$ ) and specify the distribution of candidate timestamps with sender-specific mean  $\mu_{ad}$ . Following the generalized linear model (GLM) framework (Nelder and Baker, 1972), we assume the mean and variance of the  $\tau_{ad}$  satisfy

$$\begin{aligned} E(\tau_{ad}) &= \mu_{ad}, \\ V(\tau_{ad}) &= V(\mu_{ad}), \end{aligned} \quad (2.5)$$

where  $\tau_{ad}$  here is a positive real number. Possible choices of distribution include exponential, Weibull, gamma, and log-normal<sup>1</sup> distributions, which are commonly used in

---

<sup>1</sup>The log-normal distribution is not exponential family but can be used via modeling of  $\log(\tau_d)$ .

time-to-event modeling (Rao, 2000; Rizopoulos, 2012). Based on the specific distribution, we may need other latent variables to draw the time increment, to account for the variance of time increments, beyond the coefficients for the features used to model the rate.  $V(\mu)$ —e.g., the shape parameter  $k$  for the Weibull, the shape parameter  $\theta$  for the gamma, and the variance parameter  $\sigma_\tau^2$  for the log-normal. We use  $f_\tau(\cdot; \mu, V(\mu))$  and  $F_\tau(\cdot; \mu, V(\mu))$  to denote the probability density function (p.d.f) and cumulative density function (c.d.f), respectively, with mean  $\mu$  and variance  $V(\mu)$ .

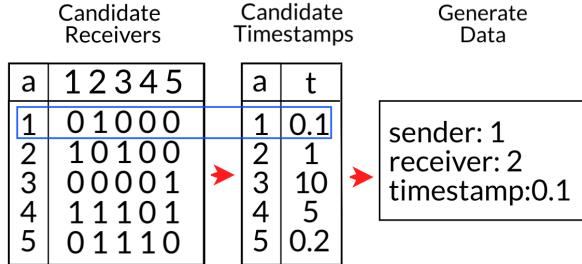


Figure 1: An illustrative example of the generative process of the HEM.

### 2.3 Senders, receivers, and timestamps

Finally, under the HEM the observed sender, receivers, and timestamp of edge  $d$  is generated by selecting the sender-receiver-set pair with the smallest time increment ([Snijders, 1996](#)):

$$\begin{aligned} a_d &= \operatorname{argmin}_a(\tau_{ad}), \\ \mathbf{r}_d &= \mathbf{u}_{a_d d}, \\ t_d &= t_{d-1} + \tau_{a_d d}. \end{aligned} \tag{2.6}$$

Therefore, it is a sender-driven process in that the receivers and timestamps of an edge are jointly determined by the sender’s urgency to send the edge to the selected receivers. Note that our generative process accounts for tied events such that in case of tied events (i.e., multiple senders draw exactly same candidate timestamps), we observe all of the tied events without assigning the orders of tied events. Algorithm 1 summarizes the entire generative process for directed edges with one sender and one or more receivers, and Figure 1 presents an illustrative example on how the  $d$ th edge is generated, assuming  $t_{d-1} = 0$  for simplicity.

## 3 Posterior inference

In this section we describe how we invert the generative process to obtain the posterior distribution over the latent variables—candidate receivers  $\{\mathbf{u}_d\}_{d=1}^D$ , coefficients for edge covariates  $\mathbf{b}$ , and coefficients for timestamp covariates  $\boldsymbol{\eta}$ —conditioned on the observed data  $\{(a_d, \mathbf{r}_d, t_d)\}_{d=1}^D$ , covariates  $\{(\mathbf{x}_d, \mathbf{y}_d)\}_{d=1}^D$ , and hyperparameters  $(\boldsymbol{\mu}_b, \Sigma_b, \boldsymbol{\mu}_\eta, \Sigma_\eta)$ . We draw the samples using Markov chain Monte Carlo (MCMC) methods, repeatedly resampling the value of each latent variable from its conditional posterior via a Metropolis-within-Gibbs sampling algorithm. In this section, we provide each latent variable’s conditional posterior, and demonstrate how we perform the prior-posterior simulator test of [Geweke \(2004\)](#), in which we evaluate the integrity of both our mathematical derivations and the sampling code, for the HEM. At the end of this section, we provide the pseudocode of our MCMC algorithm in Algorithm 2.

**Algorithm 1** Generative Process: one sender and one or more receivers

---

**Input:** number of edges and nodes ( $D, A$ ), covariates ( $\mathbf{x}, \mathbf{y}$ ), and the coefficients ( $\mathbf{b}, \boldsymbol{\eta}$ )

```

for d=1 to D do
  for a=1 to A do
    for r=1 to A (r ≠ a) do
      set  $\lambda_{adr} = \mathbf{b}^\top \mathbf{x}_{adr}$ 
    end for
    draw  $\mathbf{u}_{ad} \sim \text{MB}_G(\boldsymbol{\lambda}_{ad})$ 
    set  $\mu_{ad} = g^{-1}(\boldsymbol{\eta}^\top \mathbf{y}_{ad})$ 
    draw  $\tau_{ad} \sim f_\tau(\mu_{ad}, V(\mu_{ad}))$ 
  end for
  if  $n \geq 2$  tied events then
    set  $a_d, \dots, a_{d+n-1} = \text{argmin}_a(\tau_{ad})$ 
    set  $\mathbf{r}_d = \mathbf{u}_{a_dd}, \dots, \mathbf{r}_{d+n-1} = \mathbf{u}_{a_{d+n-1}d}$ 
    set  $t_d, \dots, t_{d+n-1} = t_{d-1} + \min_a \tau_{ad}$ 
    jump to  $d = d + n$ 
  else
    set  $a_d = \text{argmin}_a(\tau_{ad})$ 
    set  $\mathbf{r}_d = \mathbf{u}_{a_dd}$ 
    set  $t_d = t_{d-1} + \min_a \tau_{ad}$ 
  end if
end for

```

---

### 3.1 Conditional posteriors

#### Candidate receivers

In the HEM, direct computation of the posterior densities for the latent variables  $\mathbf{b}$  and  $\boldsymbol{\eta}$ —i.e.,  $P(\mathbf{b}|\mathbf{x}, \mathbf{a}, \mathbf{r}, \mathbf{t})$  and  $P(\boldsymbol{\eta}|\mathbf{y}, \mathbf{a}, \mathbf{r}, \mathbf{t})$ —is not possible. However, it is possible to augment the data by candidate receivers  $\mathbf{u}$  such that we can obtain their conditional posterior by conditioning on samples of  $\mathbf{u}$ . We use this common tool in Bayesian statistics called “data augmentation” (Tanner and Wong, 1987; Neal and Kypraios, 2015). Since  $u_{adr}$  is a binary random variable, its new value may be sampled directly from a Bernoulli distribution with probability  $p_{adr} = \frac{\exp(\lambda_{adr})}{\exp(\lambda_{adr}) + I(\|\mathbf{u}_{ad \setminus r}\|_1 > 0)}$ , since

$$\begin{aligned} \Pr(u_{adr} = 1 | \mathbf{u}_{ad \setminus r}, \mathbf{b}, \mathbf{x}, \mathbf{a}, \mathbf{r}, \mathbf{t}) &\propto \exp(\lambda_{adr}); \\ \Pr(u_{adr} = 0 | \mathbf{u}_{ad \setminus r}, \mathbf{b}, \mathbf{x}, \mathbf{a}, \mathbf{r}, \mathbf{t}) &\propto I(\|\mathbf{u}_{ad \setminus r}\|_1 > 0), \end{aligned} \tag{3.1}$$

where  $I(\cdot)$  is the indicator function that is used to prevent from the instances where a sender chooses zero number of receivers.

**Algorithm 2** MCMC Algorithm

---

**Input:** number of outer and inner iterations ( $O, I_1, I_2$ ) and initial values of  $(\mathbf{u}, \mathbf{b}, \boldsymbol{\eta})$

```

for o=1 to O do
  for d=1 to D do
    for a = 1 to A do
      for r = 1 to A ( $r \neq a$ ) do
        update  $u_{adr}$  using Gibbs update—equation (3.1)
      end for
    end for
    end for
    for n=1 to  $I_1$  do
      update  $\mathbf{b}$  using M-H algorithm—equation (3.2)
    end for
    for n=1 to  $I_2$  do
      update  $\boldsymbol{\eta}$  using M-H algorithm—equation (3.3) or (3.4)
    end for
    if extra parameter for  $V(\mu)$  then
      update the variance parameter using M-H algorithm—equation (3.3) or (3.4)
    end if
  end for
  summarize the results with the last chain of  $\mathbf{b}$  and  $\boldsymbol{\eta}$ 

```

---

**Coefficients for edge covariates**

Unlike the candidate receivers above, new values for  $\mathbf{b}$  cannot be sampled directly from its conditional posterior, but may instead be obtained using the Metropolis–Hastings (M-H) algorithm. Assuming an uninformative prior (i.e.,  $N(0, \infty)$ ), the conditional posterior over  $\mathbf{b}$  is

$$\Pr(\mathbf{b}|\mathbf{u}, \mathbf{x}, \mathbf{a}, \mathbf{r}, \mathbf{t}) \propto \prod_{d=1}^D \prod_{a=1}^A \frac{1}{Z(\boldsymbol{\lambda}_{ad})} \exp \left( \log(I(\|\mathbf{u}_{ad}\|_1 > 0)) + \sum_{r \neq a} \lambda_{adr} u_{adr} \right). \quad (3.2)$$

**Coefficients for timestamp covariates**

Likewise, we use the M-H algorithm to update the latent variable  $\boldsymbol{\eta}$ . Assuming an uninformative prior  $\boldsymbol{\eta}$  (i.e.,  $N(0, \infty)$ ), the conditional posterior for an untied event case is

$$\Pr(\boldsymbol{\eta}|\mathbf{u}, \mathbf{y}, \mathbf{a}, \mathbf{r}, \mathbf{t}) \propto \prod_{d=1}^D \left( f_\tau(\tau_d; \mu_{a_{dd}}, V(\mu_{a_{dd}})) \times \prod_{a \neq a_{dd}} (1 - F_\tau(\tau_d; \mu_{ad}, V(\mu_{ad}))) \right), \quad (3.3)$$

where  $f_\tau(\tau_d; \mu_{a_{dd}}, V(\mu_{a_{dd}}))$  is the probability that the  $d^{th}$  observed time increment comes from the specified distribution  $f_\tau(\cdot)$  with the observed sender's mean  $\mu_{a_{dd}}$ , and

$\prod_{a \neq a_d} (1 - F_\tau(\tau_d; \mu_{ad}, V(\mu_{ad})))$  is the probability that the rest of (unobserved) senders for event  $d$  all draw time increments greater than  $\tau_d$ . Moreover, under the existence of tied events, the conditional posterior of  $\boldsymbol{\eta}$  is written as

$$\begin{aligned} \Pr(\boldsymbol{\eta} | \mathbf{u}, \mathbf{y}, \mathbf{a}, \mathbf{r}, \mathbf{t}) &\propto \prod_{m=1}^M \left( \prod_{d: t_d = t_m^*} f_\tau(t_m^* - t_{m-1}^*; \mu_{add}, V(\mu_{add})) \right. \\ &\quad \times \left. \prod_{a \notin \{a_d\}_{d: t_d = t_m^*}} (1 - F_\tau(t_m^* - t_{m-1}^*; \mu_{ad}, V(\mu_{ad}))) \right), \end{aligned} \quad (3.4)$$

where  $t_1^*, \dots, t_M^*$  are the unique timepoints across  $D$  events ( $M \leq D$ ). If  $M = D$  (i.e., no tied events), equation (3.4) reduces to equation (3.3). Note that when we have the latent variable to quantify the variance in time increments  $V(\mu)$  (based on the choice of timestamp distribution in Section 2.2), we also use equation (3.3) (or equation (3.4) in case there exist tied events) for the additional M-H update—e.g.,  $\Pr(k|\boldsymbol{\eta}, \mathbf{u}, \mathbf{y}, \mathbf{a}, \mathbf{r}, \mathbf{t})$  for Weibull,  $\Pr(\theta|\boldsymbol{\eta}, \mathbf{u}, \mathbf{y}, \mathbf{a}, \mathbf{r}, \mathbf{t})$  for gamma, and  $\Pr(\sigma_\tau^2|\boldsymbol{\eta}, \mathbf{u}, \mathbf{y}, \mathbf{a}, \mathbf{r}, \mathbf{t})$  for log-normal distribution.

### 3.2 Getting it Right (GiR) test

Software development is integral to the objective of applying the HEM to real world data. Code review is a valuable process in any research computing context, and the prevalence of software bugs in statistical software is well documented (e.g., Altman et al., 2004; McCullough, 2009). With highly complex models such as the HEM, there are many ways in which software bugs can be introduced and go unnoticed. As such, we present a joint analysis of the integrity of our generative model, sampling equations, and software implementation.

Geweke (2004) introduced the “Getting it Right” (GiR) test—a joint distribution test of posterior simulators which can detect errors in sampling equations as well as software bugs—and it has been used to test the implementation of Bayesian inference algorithms (Zhao et al., 2016). The test involves comparing the distributions of variables simulated from two joint distribution samplers, which we call “forward” and “backward” samplers. The “forward” sampler draws joint samples of the latent and observable variables from the prior. The “backward” sampler begins by first drawing a joint sample of the latent and observed variables from the prior. It then alternates between re-sampling the latent variables, conditioned on the observable variables, from the MCMC transition operator, and then re-sampling the observable variable, conditioned on the latent variables, from the model likelihood. If the MCMC transition operator is correctly derived and implemented, this process should asymptotically generate joint samples of the latent and observable variables from the prior, like the forward sampler.

In the forward samples, both observable and unobservable variables are generated using Algorithm 1. In the backward samples, unobservable variables are generated using the sampling equations for inference, which we derived in Section 3.1. For each forward and backward sample that consists of  $D$  number of edges, we save these statistics:

1. Mean of observed receiver sizes  $\|\mathbf{r}_d\|_1$  across  $d = 1, \dots, D$ ,
2. Variance of observed receiver sizes  $\|\mathbf{r}_d\|_1$  across  $d = 1, \dots, D$ ,
3. Mean of time increments  $\tau_d$  across  $d = 1, \dots, D$ ,
4. Variance of time increments  $\tau_d$  across  $d = 1, \dots, D$ ,
5.  $b_p$  value used to generate the samples  $p = 1, \dots, P$ ,
6.  $\eta_q$  value used to generate the samples  $q = 1, \dots, Q$ ,
7.  $\sigma_\tau^2$  value used to generate the samples in log-normal distribution

To keep the computational burden of re-running thousands of rounds of inference manageable, we run the GiR using a relatively small artificial sample, consisting of  $D = 100$  edges,  $A = 5$  actors,  $P = 4$  number of edge covariates, and  $Q = 3$  number of timestamp covariates per each forward or backward samples, using log-normal distribution for the time increments  $f_\tau$ . We generated  $10^5$  sets of forward and backward samples, and then calculated 1,000 quantiles for each of the statistics. We also calculated t-test and Mann-Whitney test p-values in order to test for differences in the distributions generated in the forward and backward samples. Before we calculated these statistics, we thinned our samples by taking every 9th sample starting at the 10,000th sample for a resulting sample size of 10,000, in order to reduce the autocorrelation in the Markov chains. In each case, if we observe a large p-value, this gives us evidence that the distributions generated under forward and backward sampling have the same locations. We depict the GiR results using probability-probability (PP) plots, in which the empirical CDF values of the forward and backward samples are plotted on the  $x$  and  $y$  axes, respectively. If the two samples are from equivalent distributions, the empirical CDF values should line up on a line with zero  $y$ -intercept, and unit slope (i.e., a 45-degree line). The GiR test results are depicted in Figure 2. These results indicate that our sampling equations and software implementation pass the test on every statistic.

## 4 Application to email data

We now present a case study applying our method to Montgomery county government email data. Our data come from the North Carolina county government email dataset collected by [ben Aaron et al. \(2017\)](#) that includes internal email corpora covering the inboxes and outboxes of managerial-level employees of North Carolina county governments. Out of over twenty counties, we chose Montgomery County to 1) test our model using data with a large proportion of hyperedges (16.76%), all of which are emails sent from one sender to two or more receivers, and 2) limit the scope of this initial application. The Montgomery County email network contains 680 emails, sent and received by 18 department managers over a period of 3 months (March–May) in 2012. For this case study, we formulate a HEM specification through definitions of the edge covariates  $\mathbf{x}$  and timestamp covariates  $\mathbf{y}$ . We then report a suite of experiments—out-of-sample prediction for model selection and posterior predictive checks—that illustrate how alternative

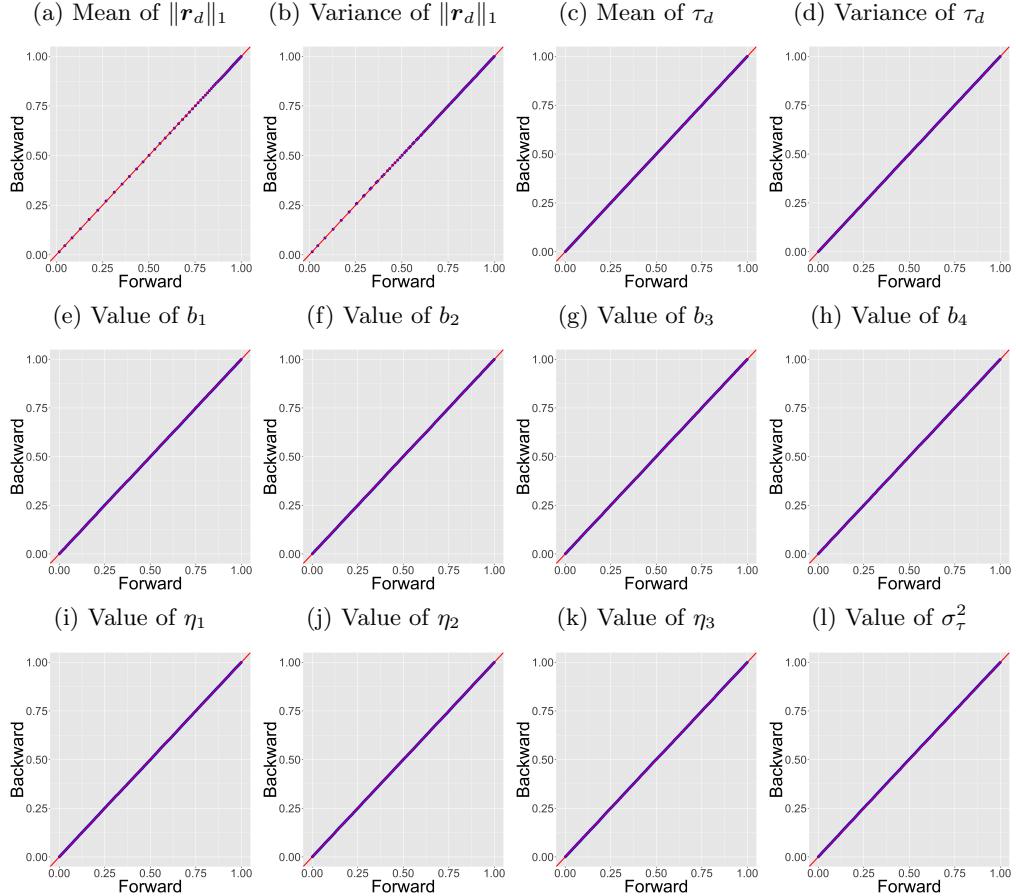


Figure 2: Probability-probability (PP) plots for the GiR test statistics.

formulations of the HEM can be compared, and evaluate how well the HEM recovers the distribution of the observed data. Finally, we demonstrate an exploratory analysis of Montgomery County email data using the model estimates to discover substantively meaningful patterns in organizational communication networks.

## 4.1 Covariates

### Edge covariates

A primary purpose of any network model is to use the posterior distributions to learn which features predict and/or explain edge formation (e.g., is edge formation reciprocal, are edges more likely to be formed among nodes with the same gender). This email application specifically gives rise to the following question: “To what extent are nodal,

dyadic or triadic network effects relevant to predicting future emails?” As an illustrative example, we form the receiver selection features  $\mathbf{x}$  for Montgomery County email data using nodal, dyadic, and triadic covariates. First, as we want to test whether gender plays a role in receiver selection process, we include three nodal covariates—the gender information of sender and receiver, and their homophily indicator (i.e., an indicator of whether the sender and receiver are of the same gender). Additionally, we include four interval-based nodal network covariates—outdegree of sender (i.e., the number of edges sent), indegree of receiver (the number of edges received), hyperedge size of sender (i.e., the number of total receivers of edges from the sender), and the interaction between (i.e., scalar product of) outdegree and hyperedge size—to study the effect of nodal behaviors on future interactions. For dyadic and triadic network effects, we employ the network statistics in Perry and Wolfe (2013) and summarize past interaction behaviors based on the time interval prior to and including  $t_{d-1}$ . Specifically, our time interval tracks 7 days prior to the last email was sent  $l_d = (t_{d-1} - 7 \text{ days}, t_{d-1}]$ . For  $a \in [A]$ ,  $r \in [A]$ , and  $d \in [D]$ , we define 14 covariates for  $\mathbf{x}_{adr}$ :

1. intercept:  $x_{adr1} = 1$ ;
2. gender of sender:  $x_{adr2} = I(\text{gender of sender } a = \text{female})$ ;
3. gender of receiver:  $x_{adr3} = I(\text{gender of receiver } r = \text{female})$ ;
4. gender homophily:  $x_{adr4} = I(x_{adr2} = x_{adr3})$ ;
5. outdegree of sender:  $x_{adr5} = \sum_{d':t_{d'} \in l_d} I(a_{d'} = a)$ ;
6. indegree of receiver:  $x_{adr6} = \sum_{d':t_{d'} \in l_d} I(u_{d'r} = 1)$ ;
7. hyperedge size of sender:  $x_{adr7} = \sum_{d':t_{d'} \in l_d} \sum_{r=1}^A I(a_{d'} = a) I(u_{d'r} = 1)$ ;
8. interaction between outdegree and hyperedge size:  $x_{adr8} = x_{adr5} \times x_{adr7}$ ;
9. send:  $x_{adr9} = \sum_{d':t_{d'} \in l_d} I(a_{d'} = a) I(u_{d'r} = 1)$ ;
10. receive:  $x_{adr10} = \text{send}(r, a)$ ;
11. 2-send:  $x_{adr11} = \sum_{h \neq a, r} \text{send}(a, h) \text{send}(h, r)$ ;
12. 2-receive:  $x_{adr12} = \sum_{h \neq a, r} \text{send}(h, a) \text{send}(r, h)$ ;
13. sibling:  $x_{adr13} = \sum_{h \neq a, r} \text{send}(h, a) \text{send}(h, r)$ ;
14. cosibling:  $x_{adr14} = \sum_{h \neq a, r} \text{send}(a, h) \text{send}(r, h)$ ;

where  $I(\cdot)$  is an indicator function. The network statistics (5–14) are designed so that their coefficients have a straightforward interpretations. The statistics “outdegree” and “indegree” measure the gregariousness and popularity effects of the node by counting the number of emails sent from  $a$  and received by  $r$ , respectively, within the last 7 days. The gregariousness effect refers to the tendency for nodes that sent many edges in the

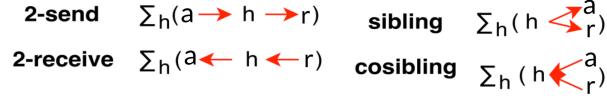


Figure 3: Visualization of triadic statistics: 2-send, 2-receive, sibling, and cosibling.

past to continue to do so in the future. The popularity effect refers to the tendency for nodes that received many edges in the past to continue to do so in the future. Moreover, in order to capture the individual tendency of sent emails to two or more receivers, we include the statistic “hyperedge size”—the number of emails sent from  $a$  within last 7 days where emails with  $k$  number of receivers are counted as  $k$  separate emails—as a variant of outdegree statistic, accounting for hyperedges. We also include the interaction term between outdegree and hyperedge size. This interaction allows us to model a possible tradeoff between the hyperedge size and the total number of edges sent. Dyadic statistics “send” and “receive” are defined as above such that these covariates measure the number of emails sent from  $a$  to  $r$  and  $r$  to  $a$ , respectively, within the last 7 days. In the example of triadic statistics, the covariate “2-send” counts the pairs of emails involving some node  $h$  distinct from  $a$  and  $r$  such that emails from  $a$  to  $h$  and  $h$  to  $r$  are both observed within the last 7 days. The 2-send statistic captures the tendency for emails to close transitive triads (i.e., triads in which  $a$  sends to  $r$  and  $h$ , and  $r$  sends to  $h$ ). We include other triadic covariates that behave similarly and exhibit analogous interpretations, which are illustrated in Figure 3.

### Timestamp covariates

For the event timing features  $\mathbf{y}$ , introduced in Section 2.2, we identify a set of covariates which may affect “time to the next email.” Similar to the edge covariates, we include nodal statistics which are time-invariant (such as gender or manager status) or time-dependent (such as the network statistics used for  $\mathbf{x}$ ). In addition, we select some edge-specific covariates based on the temporal aspect of the  $(d - 1)^{th}$  email—e.g., whether the previous email was sent (1) during the weekend and (2) before or past midday (AM/PM)—since we expect the email interactions within county government to be less active during the weekend and in the evening. To be specific, the timestamp statistics are defined as

1. intercept:  $y_{ad1} = 1$ ;
2. gender of sender:  $y_{ad2} = I(\text{gender of sender } a = \text{female})$ ;
3. manager status of sender:  $y_{ad3} = I(\text{sender } a \text{ is the County Manager})$ ;
4. outdegree of sender:  $y_{ad4} = \sum_{d':t_{d'} \in l_d} I(a_{d'} = a)$ ;
5. indegree of sender:  $y_{ad5} = \sum_{d':t_{d'} \in l_d} I(u_{d'a} = 1)$ ;
6. weekend indicator of previous email:  $y_{ad6} = I(t_{d-1} \text{ is during the weekend})$ ;

7. PM indicator of previous email:  $y_{ad7} = I(t_{d-1} \text{ in PM})$ .

Note that our generative process for timestamps in Section 2.2 is sender-oriented where the sender determines when to send the email; thus we incorporate network statistics that depend only on  $a$ —specifically, the in and outdegrees of sender  $a$ .

## 4.2 Model selection

The HEM defines a flexible family of models, each of which is indexed/defined by a set of features (i.e., the receiver selection features  $\mathbf{x}$ , the selection of the event timing features  $\mathbf{y}$ , and the distribution of time increments  $f$ ). Many of these components will be specified based on user expertise (e.g., regarding which features would drive receiver selection), but some decisions may require a data-driven approach to model specification. For example, though theoretical considerations may inform the specification of features, subject-matter expertise is unlikely to inform the decision regarding the family of the event time distribution. Furthermore, since different distribution families (and model specifications more generally) may involve different size parameter spaces, any data-driven approach to model comparison must guard against over-fitting the data. In this section we present a general-purpose approach to evaluating the HEM specification using out-of-sample prediction. We illustrate this approach by comparing alternative distributional families for the event timing component of the model. Here, we specifically compare the predictive performance from two distributions—log-normal and exponential. We particularly choose the log-normal distribution based on some exploratory analysis (e.g., histogram and simple regressions) on raw time increments data, and take exponential distribution as an alternative since exponential is the most commonly specified distribution for time-to-event data, and is also used in the stochastic actor-oriented models (SAOMs) (Snijders, 1996) as well as their extensions (Snijders et al., 2007).

We evaluated the models ability to predict out-of-sample edges and timestamps on Montgomery County email data. We generated a train/test split of the data by randomly selecting 10% of sender, receiver, and time-stamp variables to be held out. Our model then imputed these missing variables during inference, sampling them from their conditional posterior along with the other latent variables. Algorithm 3 outlines this procedure in detail. We compare the predictive performance of two versions of the model, each with a different timing distribution over the time-increments using  $N = 500$ . We summarize the results of prediction experiments for missing senders, receivers, and timestamps in Figure 4. First, we compare the posterior probability of correct senders for each of the missing emails  $\{d : a_d = \text{NA}\}$ , which corresponds to  $\pi_{ad} = P(a_d = a_d^{obs} | \cdot)$  in Algorithm 3. We call this measure the “correct sender posterior probability.” On Figure 4 (a), we draw boxplots for the distribution of mean correct sender posterior probability—i.e.,  $\hat{\pi}_{ad} = \frac{1}{N} \sum_{n=1}^N \pi_{ad}^n$ —across the missing emails. The results show that both log-normal and exponential distributions achieve better predictive performance for missing senders compared to what is expected under random guess (i.e., choose one out of  $A$  possible senders = 1/18), with log-normal distribution showing better performance than exponential distribution. Secondly, since the receiver vector is binary, we compute  $F_1$  scores

**Algorithm 3** Out-of-Sample Predictions

---

**Input:** data  $\{(a_d, \mathbf{r}_d, t_d)\}_{d=1}^D$ , number of new data to generate  $R$ , and initial values of  $(\mathbf{b}, \boldsymbol{\eta}, \mathbf{u}, \sigma_\tau^2)$

**Test splits** (with  $p = 0.1$ ):  
draw test senders (out of  $D$  senders)  
draw test receivers (out of  $D \times (A - 1)$  receiver indicators  $\{\{\mathbf{r}_{dr}\}_{r \in [A] \setminus a_d}\}_{d=1}^D$ )  
draw test timestamps (out of  $D$  timestamps)  
set the test data as “missing” (NA)

**Imputation and inference:**

```

for  $r = 1$  to  $R$  do
    for  $d = 1$  to  $D$  do
        if  $a_d = \text{NA}$  then
            draw  $a_d \sim \text{Categorical}(\boldsymbol{\pi}_d)$ , where  $\pi_{da} = f_\tau(\tau_d; \mu_{ad}, V(\mu_{ad})) \times \prod_{a' \neq a} (1 - F_\tau(\tau_d; \mu_{a'd}, V(\mu_{a'd})))$ 
        end if
        for  $r \in [A] \setminus a_d$  do
            if  $r_{dr} = \text{NA}$  then
                draw  $r_{dr} \sim \text{Bernoulli}(p_{dr})$ , where  $p_{dr} = \frac{\exp(\lambda_{adr})}{\exp(\lambda_{adr}) + I(\|\mathbf{u}_{ad \setminus r}\|_1 > 0)}$ 
            end if
        end for
        if  $t_d = \text{NA}$  then
            draw  $\boldsymbol{\tau}_d^{new} \sim f_\tau(\tau_d; \mu_{ad}, V(\mu_{ad})) \times \prod_{a \neq a_d} (1 - F_\tau(\tau_d; \mu_{ad}, V(\mu_{ad})))$ , where  

            we sample from this distribution using importance sampling2
        end if
        run inference and update  $(\mathbf{u}, \mathbf{b}, \boldsymbol{\eta})$  given the imputed and observed data
    end for
    store the estimates for test data
end for

```

---

for missing receiver indicators (i.e., all  $d$  and  $r$  with  $r_{dr}=\text{NA}$ ) by taking the harmonic mean of precision and recall:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \text{ where} \\ \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{ and precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4.1)$$

with TP denoting true positive (i.e.,  $\mathbf{r}_{dr}^{obs} = \mathbf{r}_{dr}^{pred} = 1$ ), FN denoting false negative (i.e.,  $\mathbf{r}_{dr}^{obs} = 1$  but  $\mathbf{r}_{dr}^{pred} = 0$ ), and FP denoting false positive (i.e.,  $\mathbf{r}_{dr}^{obs} = 0$  but  $\mathbf{r}_{dr}^{pred} = 1$ ). Although the generative process for edges (Section 2.1) is not directly affected by the choice of timestamp distribution, Figure 4 (b) reveals slight difference between log-normal and exponential in their performance in predicting missing receiver indicators, where log-normal on average outperforms exponential. Finally, the prediction error for

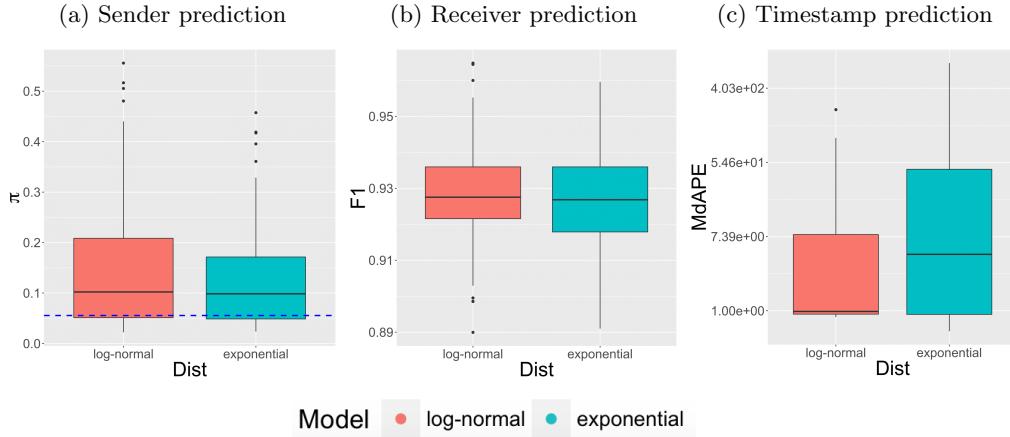


Figure 4: Comparison of predictive performance between log-normal and exponential distributions: (a) correct sender posterior probability from sender predictions, (b)  $F_1$  scores from receiver predictions, and (c) median absolute relative error from timestamp predictions. Blue line in (a) represents the correct sender probability expected by random guess—i.e.,  $1/A = 1/18 \approx 0.056$ .

$d^{th}$  missing timestamp is estimated using the median of absolute errors<sup>3</sup> across  $N = 500$  predictions:

$$e_{\tau_d} = \text{median}\left(\left|\frac{\tau_d^{obs} - \tau_d^{pred_1}}{\tau_d^{obs}}\right|, \dots, \left|\frac{\tau_d^{obs} - \tau_d^{pred_N}}{\tau_d^{obs}}\right|\right). \quad (4.2)$$

Figure 4 (c) presents boxplots for the median absolute relative errors, where plot the estimates in a log-scale. Surprisingly, we have a huge benefit in the performance of timestamp prediction when we assume log-normal distribution for time-increments compared to exponential distribution. We speculate that this difference can be explained by overdispersion in exponential distribution, because there exists greater variability in the time increments of emails than would be expected under exponential distribution. As illustrated above, we can use this out-of-sample prediction task for two uses—1) to provide an effective answer to the question “how does the HEM perform at filling in the missing components of time-stamped network data?” and 2) to offer one standard way to determine the distribution of time increments in Section 2.2.

### 4.3 Posterior predictive checks

In this section, we perform posterior predictive checks (PPC) (Rubin et al., 1984) to evaluate the appropriateness of our model specification for Montgomery County email data. We formally generated entirely new data by simulating  $N = 500$  synthetic email

<sup>3</sup>Hyndman and Koehler (2006) refer to this as median absolute percentage error (MdAPE).

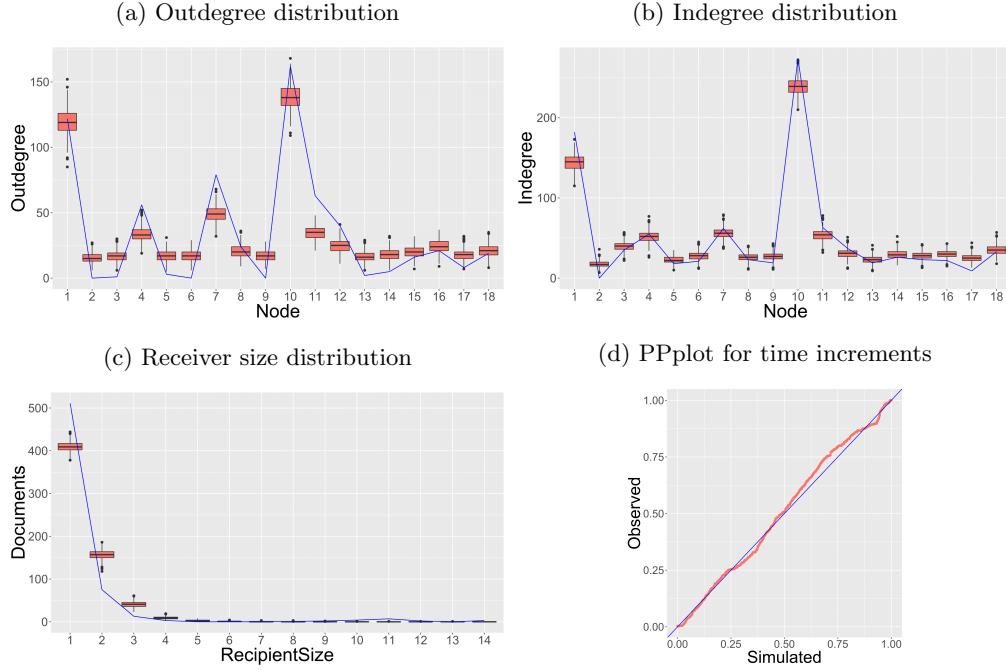


Figure 5: PPC results from log-normal distribution. Blue lines denote the observed statistics in (a)–(c) and denotes the diagonal line in (d).

datasets  $\{(a_d, \mathbf{r}_d, t_d)\}_{d=1}^D$  from the generative process in Section 2, conditional upon a set of inferred latent variables from the inference in Section 4.4. For the test of goodness-of-fit in terms of network dynamics, we use multiple statistics that summarize meaningful aspects of the data: outdegree distribution—the number of edges sent by each node, indegree distribution—the number of edges received by each node, receiver size distribution—the number of receivers on each edge, and a probability-probability (PP) plot for time increments.

Figure 5 illustrates the results of posterior predictive checks using the log-normal distribution, which shows better performance in Section 4.2. The upper two plots show node-specific posterior predictive degree distributions across  $N = 500$  synthetic samples, where the left one for outdegree statistic and the right plot is for indegree statistic. For both plots, the x-axis represents the nodes ( $A = 1, \dots, 18$ ), and the y-axis represents the number of emails sent or received by the node. When compared with the observed outdegree and indegree statistics (red lines), our model recovers the overall distribution of sending and receiving activities across the nodes. For example, node 1 and 10 show significantly higher level of both sending and receiving activities relative to the rest, and the model-simulated data captures those big jumps, showing acceptable fit to the data. Outdegree distribution of some low-activity nodes are not precisely recovered, however, indegree distribution looks much better. Since we use more information in

the receiver selection process (i.e., network effects) while we rely solely on minimum time increments when choosing the observed sender, these results are expected. The lower left plot is the distribution of receiver sizes, where the x-axis spans over the size of receivers 1 to 14 (which is the maximum size of observed receivers) and the y-axis denotes the number of emails with x-number of receivers. The result shows that our model is underestimating emails with one receiver while overestimating emails with two, three, and four receivers. One explanation behind what we observe is that the model is trying to recover so-called “broadcast” emails, which are the emails with  $\geq 10$  number of receivers, so that the intercept estimate  $b_1$  is slightly moved toward right. It would be an interesting problem in future research to consider how the hyperedge size distribution can be further modified to capture this distribution more accurately. The plot on the lower right is the PP plot for time increments, which depicts the two cumulative distribution functions—one for simulated time increments and another for observed time increments—against each other in order to assess how closely two data sets agree. Here, the closeness to the diagonal line connecting  $(0, 0)$  and  $(1, 1)$  gives a measure of difference between the simulated and observed time increments, and our PP plot shows that we have great performance in reproducing the observed time distribution. Our findings from the predictive experiments in Section 4.2 are further revealed in the PPC from exponential distribution, where the PPC plots comparing log-normal and exponential distributions are presented in Appendix C.

#### 4.4 Exploratory analysis

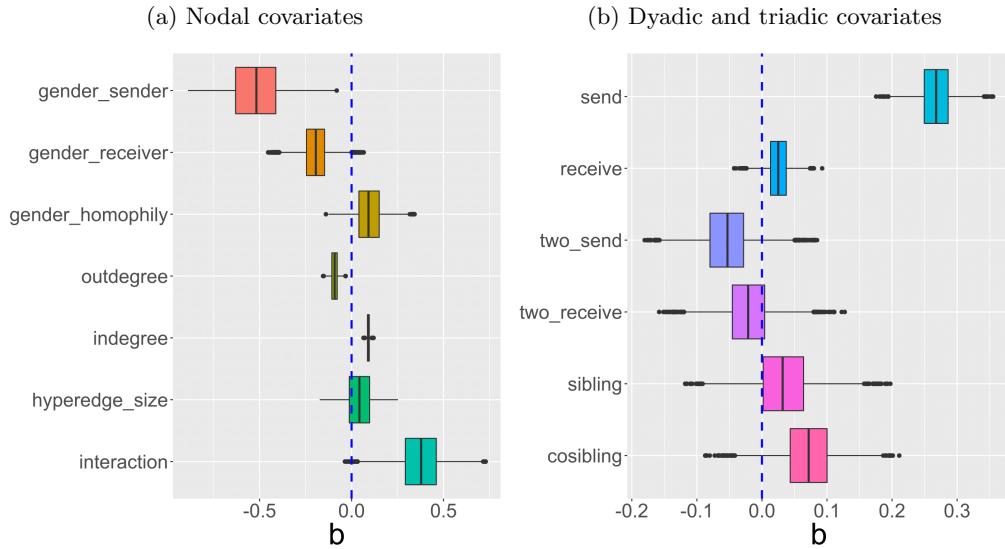
Based on the prediction experiments in Section 4.2, we interpret the results from the HEM using the log-normal distribution. We assume weakly informative priors for latent variables such as  $\boldsymbol{b} \sim N(\boldsymbol{\mu}_b = \mathbf{0}, \Sigma_b = 2 \times I_P)$ ,  $\boldsymbol{\eta} \sim N(\boldsymbol{\mu}_\eta = \mathbf{0}, \Sigma_\eta = 2 \times I_Q)$ , and  $\sigma_\tau^2 \sim \text{inverse-Gamma}(a = 2, b = 1)$ , and run MCMC algorithm in Algorithm 2 with  $O = 55,000$  outer iterations with a burn-in of 15,000, where we thin by keeping every 40th sample. While the inner iterations for  $\sigma_\tau^2$  is fixed as 1, we specify the inner iterations  $I_1 = 20$  for  $\boldsymbol{b}$  and  $I_2 = 10$  for  $\boldsymbol{\eta}$  to adjust for slower convergence rates. Convergence diagnostics including the traceplots and Geweke diagnostics (Geweke et al., 1991) are provided in Appendix D.

##### Coefficients for edge covariates

Figure 6 shows the boxplots summarizing posterior samples of  $\boldsymbol{b}$ , where Figure 6 (a) displays the coefficients for nodal covariates and 6 (b) displays the coefficients for dyadic and triadic covariates. Since we use the logit functional form

$$\text{logit}(\lambda_{adr}) = \log \left( \frac{\lambda_{adr}}{1 - \lambda_{adr}} \right) = b_1 + b_2 x_{adr2} \dots + b_{14} x_{adr14},$$

and can interpret the  $\boldsymbol{b}$  estimates in terms of odds ratios  $\frac{\lambda_{adr}}{1 - \lambda_{adr}} = \exp(b_1 + b_2 x_{adr2} \dots + b_{14} x_{adr14})$ . We find the effects of nodal covariates “gender of sender” and “gender of receiver” are both nearly always negative in the posterior samples. The log odds that any other node will be added as a receiver of an email is approximately 0.5 less if the

Figure 6: Posterior distribution of  $\boldsymbol{b}$  estimates.

sender is a woman. The posterior distribution of the statistic “outdegree” is mostly negative, if sender  $a$  sent  $n$  number of emails to anyone last week, then sender  $a$  is approximately  $\exp(-0.109 \times n) \approx (0.897)^n$  times less likely to send an email to  $r$ . However, this straightforward interpretation of the outdegree statistic only applies when the hyperedge size is low. The scenario in which a sender sends a lot of low-hyperedge-size emails may arise due to the use of email for a one-on-one conversation. The large positive estimates of the interaction between hyperedge size and outdegree indicate that those who have recently sent many emails with many receivers on each email are likely to continue sending these “broadcast” style emails. This scenario may arise from someone being responsible for distributing timely announcements. When we look at the effect of “indegree,” we see a clear popularity effect—those who have received a lot of emails a lot recently are likely to continue receiving a lot of emails. If the receiver  $r$  received  $n$  number of emails over the last week, sender  $a$  is  $\exp(0.086 \times n) \approx (1.091)^n$  times more likely to send an email to  $r$ .

When we look at the effects of dyadic and triadic covariates, one thing that stands out is the large and positive posterior distribution of the statistic “send” (i.e., number of times sender  $a$  sent emails to receiver  $r$  over the last week) with the posterior mean  $\hat{b}_9 = 0.274$ , implying that if sender  $a$  sent  $n$  number of emails to  $r$  last week, then sender  $a$  is approximately  $\exp(0.274 \times n) \approx (1.315)^n$  times more likely to send an email to  $r$ . The posterior distributions for the reciprocity effect (receive), and the four triadic effects, are all fairly evenly spread around zero, so our results do not justify conclusions regarding the nature of these effects in the Montgomery county email network.

### Coefficients for timestamp covariates

For timestamp covariates, Figure 7 shows the boxplots summarizing posterior samples of  $\boldsymbol{\eta}$ . Note that interpretations of the estimated coefficients for  $\hat{\boldsymbol{\eta}}$  should be based on the specified timeunit of the dataset, which we use “hour” for Montgomery county email data. Moreover, since we assume log-normal distribution for time increments, the coefficients are interpreted in terms of the change in the average log time.

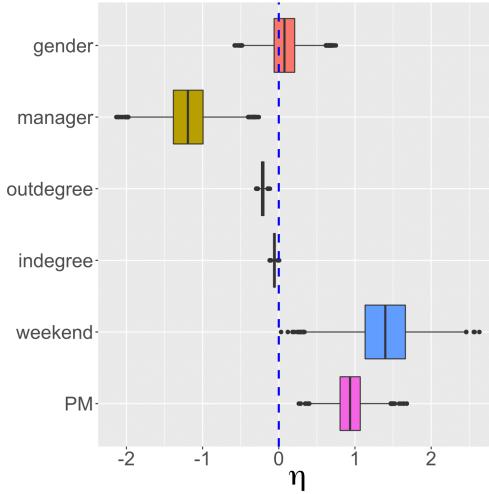
$$\log(\tau_{ad}) \sim N(\mu_{ad}, \sigma_\tau^2), \text{ with} \\ \mu_{ad} = \eta_1 + \eta_2 y_{ad2} \dots + \eta_7 y_{ad7}.$$

The posterior estimates of two temporal effects—“weekend” and “PM”—indicate that if the  $(d - 1)^{th}$  email was sent during the weekend or after midday, then the time to  $d^{th}$  email is expected to take  $\exp(1.552) \approx 4.722$  hours and  $\exp(0.980) \approx 2.665$  hours longer, respectively, compared to their counterparts (i.e., weekdays and am). On the contrary, the covariates “manager”, “outdegree”, and “indegree” shorten the amount of time to next email. For example, being a county manager (i.e., the lead county administrator) lowers the expected value of  $\log(\tau_{ad})$  by  $\hat{\eta}_3 = -1.070$ , since the manager in general sends or receives a lot more emails which may shorten the response time and many of those emails. The posterior estimates for the “outdegree” and “indegree” statistics, where the posterior means are approximately  $\hat{\eta}_4 = -0.206$  and  $\hat{\eta}_5 = -0.060$ , respectively. These effects indicate that those who are involved in either sending or receiving a lot of emails recently are likely to send emails with greater speed. The posterior distribution for the effect of the gender of the manager is pretty evenly spread around zero. In addition, the posterior mean estimates for variance parameter  $\sigma_\tau^2$  in the lognormal distribution is approximately  $\hat{\sigma}_\tau^2 = 14.093$  with its 95% credible interval (12.709, 15.555), indicating that there exists large variability in the time increments of emails.

## 5 Conclusion

Motivated by a growing class of dynamic network models which deal with edges exchanged in continuous time, the hyperedge event model (HEM) can effectively learn the underlying dynamics in edge and timestamp formations, providing novel insights to the literature. The HEM explicitly models hyperedges through a receiver-selection distribution that forces the sender to select at least one receiver, which is a more realistic approach for events with one sender and one or more receivers and one or more sender and one receiver compared to treating them as pure duplicates. In modeling the timestamps (more precisely time increments) of events, our generalized linear model (GLM) based formulation offers new innovations by eliminating the need to stick with one parameter distribution (e.g., exponential distribution). To our knowledge, the HEM is the only existing model that can be used to generate the sender, recipients, and time stamp of interactions in real time. To make better use of the proposed model, we provide an algorithm for predictive experiments that help to learn which specification of HEM provides a better fit to the data.

We have demonstrated the effectiveness of our model by analyzing Montgomery

Figure 7: Posterior distribution of  $\eta$  estimates.

County government emails, where emails serve as a canonical example of directed edges with one sender and one or more receivers. The estimated effects for edge covariates reveal that the HEM is able to understand the structural dynamics similar to those used in the exponential random graph model (ERGM), but we also learn about the effects of timestamp covariates by integrating a survival model for event timing. Although we illustrate the entire framework and application in the context of one type of hyperedge, one sender and one or more receivers, our model can be easily extended to allow the opposite case, one or more sender and one receiver, by slight modification of the generative process (shown in Appendix A). This extension involves promising applications to socio-political networks such as international sanctions and co-sponsorship of bills, and biological networks such as those formed through neural dendrites.

### Acknowledgments

This work was supported in part by the University of Massachusetts Amherst Center for Intelligent Information Retrieval and in part by National Science Foundation grants DGE-1144860, SES-1619644, and CISE-1320219. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect those of the sponsors.

## Appendix

### Appendix A: Alternative generative process

---

**Algorithm 4** Generative Process: one receiver and one or more senders

---

```

Input: number of edges and nodes ( $D, A$ ), covariates ( $\mathbf{x}, \mathbf{y}$ ), and the coefficients ( $\mathbf{b}, \boldsymbol{\eta}$ )
for d=1 to D do
    for r=1 to A do
        for a=1 to A ( $a \neq r$ ) do
            set  $\lambda_{adr} = \mathbf{b}^\top \mathbf{x}_{adr}$ 
        end for
        draw  $\mathbf{u}_{rd} \sim \text{MB}_G(\boldsymbol{\lambda}_{rd})$ 
        set  $\mu_{rd} = g^{-1}(\boldsymbol{\eta}^\top \mathbf{y}_{rd})$ 
        draw  $\tau_{rd} \sim f_\tau(\mu_{rd}, \sigma_\tau^2)$ 
    end for
    if  $n \geq 2$  tied events then
        set  $r_d = \text{argmin}_r(\tau_{rd})$ 
        set  $\mathbf{a}_d = \mathbf{u}_{rdd}, \dots, \mathbf{a}_{d+n-1} = \mathbf{u}_{r_{d+n-1}d}$ 
        set  $t_d, \dots, t_{d+n-1} = t_{d-1} + \min_r \tau_{rd}$ 
        jump to  $d = d + n$ 
    else
        set  $r_d = \text{argmin}_r(\tau_{rd})$ 
        set  $\mathbf{a}_d = \mathbf{u}_{rdd}$ 
        set  $t_d = t_{d-1} + \min_r \tau_{rd}$ 
    end if
    end for

```

---

### Appendix B: Normalizing constant of $\text{MB}_G$

Our probability measure “ $\text{MB}_G$ ”—the multivariate Bernoulli distribution with non-empty Gibbs measure—defines the probability of sender  $a$  selecting the binary receiver vector  $\mathbf{u}_{ad}$  as

$$\Pr(\mathbf{u}_{ad} | \mathbf{b}, \mathbf{x}_{ad}) = \frac{1}{Z(\boldsymbol{\lambda}_{ad})} \exp \left( \log(I(\|\mathbf{u}_{ad}\|_1 > 0)) + \sum_{r \neq a} \lambda_{adr} u_{adr} \right),$$

where the receiver intensity is a linear combination of edge covariates—i.e.,  $\lambda_{adr} = \mathbf{b}^\top \mathbf{x}_{adr}$ —as defined in Section 2.1.

To use this distribution efficiently, we derive a closed-form expression for  $Z(\boldsymbol{\lambda}_{ad})$  that does not require brute-force summation over the support of  $\mathbf{u}_{ad}$  (i.e.,  $\forall \mathbf{u}_{ad} \in [0, 1]^A$ ). We recognize that if  $\mathbf{u}_{ad}$  were drawn via independent Bernoulli distributions in which  $\Pr(u_{adr} = 1 | \mathbf{b}, \mathbf{x}_{ad})$  was given by  $\text{logit}(\lambda_{adr})$ , then

$$\Pr(\mathbf{u}_{ad} | \mathbf{b}, \mathbf{x}_{ad}) \propto \exp \left( \sum_{r \neq a} \lambda_{adr} u_{adr} \right).$$

This is straightforward to verify by looking at

$$\Pr(u_{adr} = 1 | \mathbf{u}_{ad[-r]}, \mathbf{b}, \mathbf{x}_{ad}) = \frac{\exp(\lambda_{adr})}{\exp(\lambda_{adr}) + 1}.$$

We denote the logistic-Bernoulli normalizing constant as  $Z^l(\boldsymbol{\lambda}_{ad})$ , which is defined as

$$Z^l(\boldsymbol{\lambda}_{ad}) = \sum_{\mathbf{u}_{ad} \in [0,1]^A} \exp\left(\sum_{r \neq a} \lambda_{adr} u_{adr}\right).$$

Now, since

$$\exp\left(\log\left(\mathbb{I}(\|\mathbf{u}_{ad}\|_1 > 0)\right) + \sum_{r \neq a} \lambda_{adr} u_{adr}\right) = \exp\left(\sum_{r \neq a} \lambda_{adr} u_{adr}\right),$$

except when  $\|\mathbf{u}_{ad}\|_1 = 0$ , we note that

$$\begin{aligned} Z(\boldsymbol{\lambda}_{ad}) &= Z^l(\boldsymbol{\lambda}_{ad}) - \exp\left(\sum_{\forall u_{adr}=0} \lambda_{adr} u_{adr}\right) \\ &= Z^l(\boldsymbol{\lambda}_{ad}) - 1. \end{aligned}$$

We can therefore derive a closed form expression for  $Z(\boldsymbol{\lambda}_{ad})$  via a closed form expression for  $Z^l(\boldsymbol{\lambda}_{ad})$ . This can be done by looking at the probability of the zero vector under the logistic-Bernoulli model:

$$\frac{1}{Z^l(\boldsymbol{\lambda}_{ad})} \exp\left(\sum_{\forall u_{adr}=0} \lambda_{adr} u_{adr}\right) = \prod_{r \neq a} \left(1 - \frac{\exp(\lambda_{adr})}{\exp(\lambda_{adr}) + 1}\right).$$

Then, we have

$$\frac{1}{Z^l(\boldsymbol{\lambda}_{ad})} = \prod_{r \neq a} \frac{1}{\exp(\lambda_{adr}) + 1}.$$

Finally, the closed form expression for the normalizing constant is

$$Z(\boldsymbol{\lambda}_{ad}) = \prod_{r \neq a} (\exp(\lambda_{adr}) + 1) - 1.$$

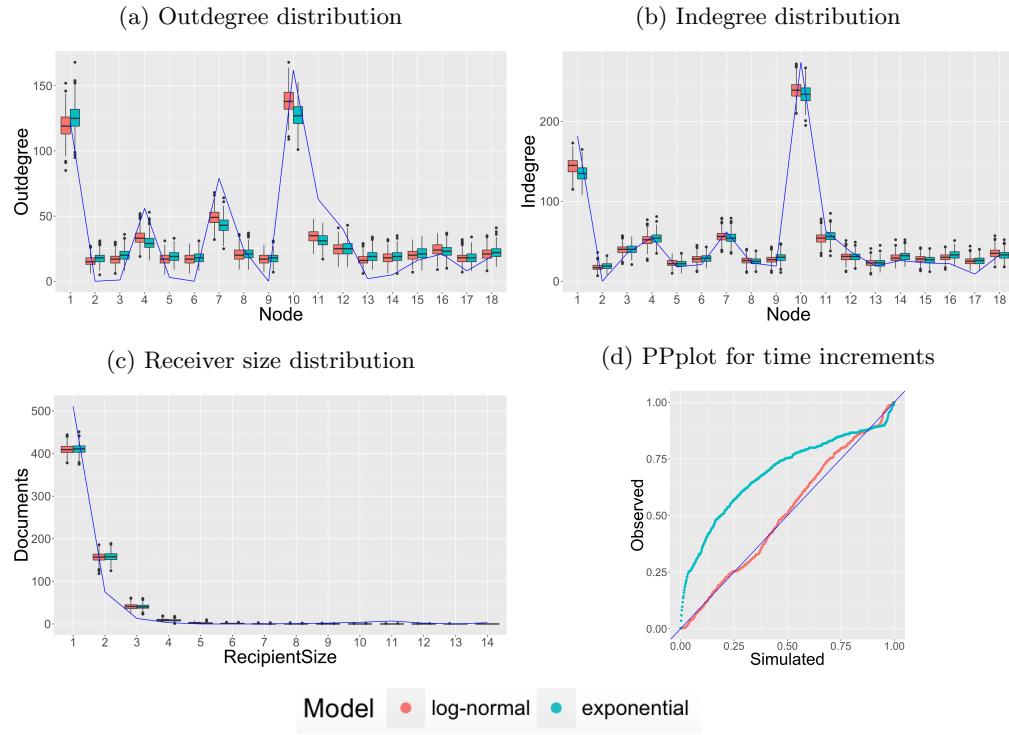
**Appendix C: Comparison of PPC results: log-normal vs. exponential**


Figure 8: Comparison of PPC results between log-normal (red) and exponential (green) distributions. Blue lines denote the observed statistics in (a)–(c) and denotes the diagonal line in (d).

## Appendix D: Convergence diagnostics

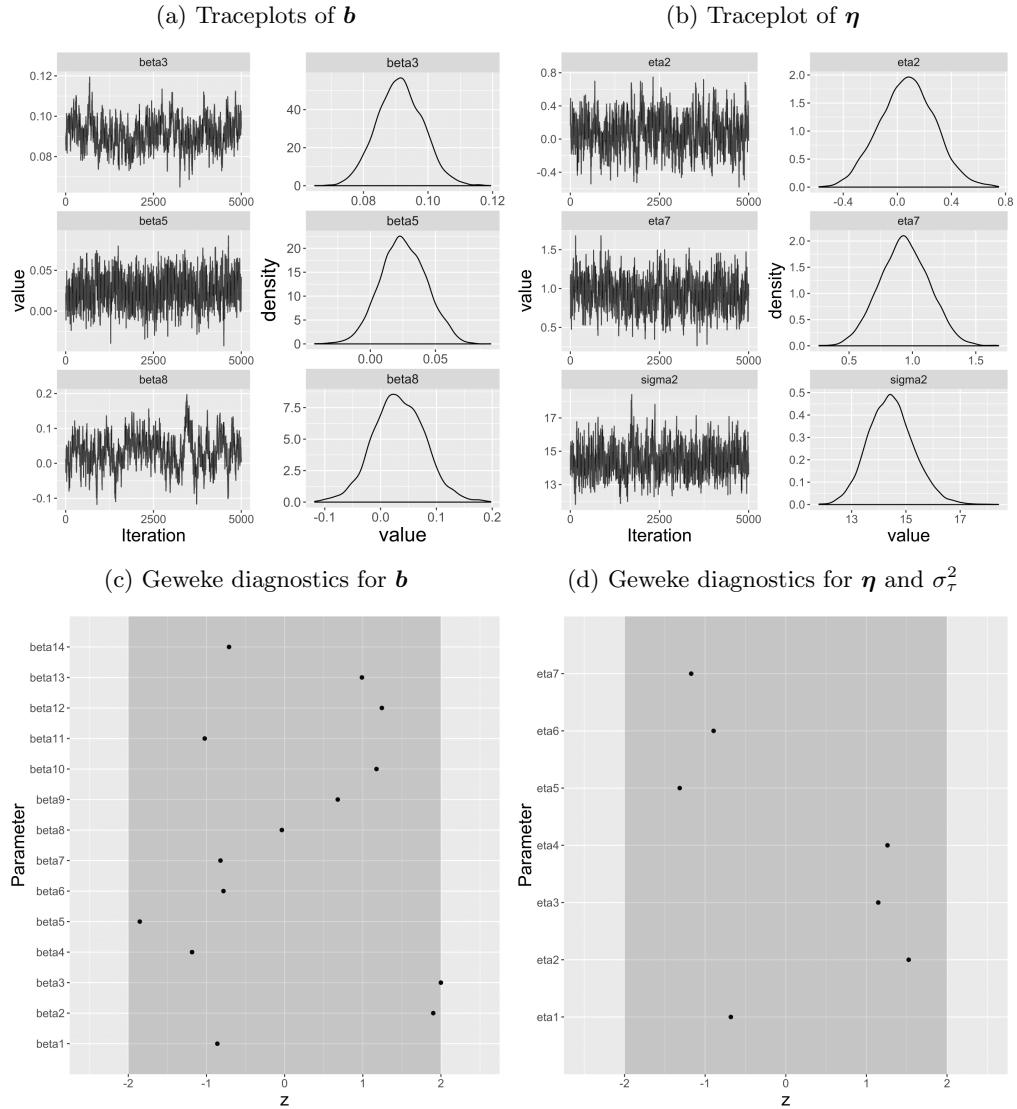


Figure 9: Convergence diagnostics from log-normal distribution.

## References

- Altman, M., Gill, J., and McDonald, M. P. (2004). *Numerical issues in statistical computing for the social scientist*, volume 508. John Wiley & Sons. 9

- ben Aaron, J., Denny, M., Desmarais, B., and Wallach, H. (2017). “Transparency by Conformity: A Field Experiment Evaluating Openness in Local Governments.” *Public Administration Review*, 77(1): 68–77. [10](#)
- Butts, C. T. (2008). “A RELATIONAL EVENT FRAMEWORK FOR SOCIAL ACTION.” *Sociological Methodology*, 38(1): 155–200. [1](#)
- Cranmer, S. J., Heinrich, T., and Desmarais, B. A. (2014). “Reciprocity and the structural determinants of the international sanctions network.” *Social Networks*, 36: 5–22. [2](#)
- Dai, B., Ding, S., Wahba, G., et al. (2013). “Multivariate bernoulli distribution.” *Bernoulli*, 19(4): 1465–1483. [3](#)
- Fan, Y. and Shelton, C. R. (2009). “Learning continuous-time social network dynamics.” In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 161–168. AUAI Press. [2](#)
- Fellows, I. and Handcock, M. (2017). “Removing Phase Transitions from Gibbs Measures.” In *Artificial Intelligence and Statistics*, 289–297. [3](#)
- Fowler, J. H. (2006). “Legislative cosponsorship networks in the US House and Senate.” *Social Networks*, 28(4): 454–465. [2](#)
- Geweke, J. (2004). “Getting it right: Joint distribution tests of posterior simulators.” *Journal of the American Statistical Association*, 99(467): 799–804. [6, 9](#)
- Geweke, J. et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA. [18](#)
- Ghoshal, G., Zlatić, V., Caldarelli, G., and Newman, M. (2009). “Random hypergraphs and their applications.” *Physical Review E*, 79(6): 066118. [2](#)
- Hunter, D., Smyth, P., Vu, D. Q., and Asuncion, A. U. (2011). “Dynamic egocentric models for citation networks.” In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 857–864. [1](#)
- Hyndman, R. J. and Koehler, A. B. (2006). “Another look at measures of forecast accuracy.” *International journal of forecasting*, 22(4): 679–688. [16](#)
- Karypis, G., Aggarwal, R., Kumar, V., and Shekhar, S. (1999). “Multilevel hypergraph partitioning: applications in VLSI domain.” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 7(1): 69–79. [2](#)
- McCullough, B. D. (2009). “The accuracy of econometric software.” *Handbook of computational econometrics*, 55–79. [9](#)
- Neal, P. and Kypraios, T. (2015). “Exact Bayesian inference via data augmentation.” *Statistics and Computing*, 25(2): 333–347. [7](#)
- Nelder, J. A. and Baker, R. J. (1972). *Generalized linear models*. Wiley Online Library. [4](#)

- Newman, M. E., Forrest, S., and Balthrop, J. (2002). "Email networks and the spread of computer viruses." *Physical Review E*, 66(3): 035101. [2](#)
- Partzsch, J. and Schüffny, R. (2012). "Developing structural constraints on connectivity for biologically embedded neural networks." *Biological cybernetics*, 106(3): 191–200. [2](#)
- Perry, P. O. and Wolfe, P. J. (2013). "Point process modelling for directed interaction networks." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5): 821–849. [1](#), [2](#), [12](#)
- Rao, P. (2000). "Applied survival analysis: regression modeling of time to event data." *Journal of the American Statistical Association*, 95(450): 681–681. [4](#)
- Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC Press. [4](#)
- Rubin, D. B. et al. (1984). "Bayesianly justifiable and relevant frequency calculations for the applied statistician." *The Annals of Statistics*, 12(4): 1151–1172. [16](#)
- Snijders, T., Steglich, C., and Schweinberger, M. (2007). *Modeling the coevolution of networks and behavior*. na. [1](#), [14](#)
- Snijders, T. A. (1996). "Stochastic actor-oriented models for network change." *Journal of mathematical sociology*, 21(1-2): 149–172. [1](#), [6](#), [14](#)
- Tanner, M. A. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *Journal of the American statistical Association*, 82(398): 528–540. [7](#)
- Vu, D. Q., Hunter, D., Smyth, P., and Asuncion, A. U. (2011). "Continuous-Time Regression Models for Longitudinal Networks." In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems 24*, 2492–2500. Curran Associates, Inc. [1](#)
- Zhang, Z.-K. and Liu, C. (2010). "A hypergraph model of social tagging networks." *Journal of Statistical Mechanics: Theory and Experiment*, 2010(10): P10005. [2](#)
- Zhao, T., Wang, Z., Cumberworth, A., Gsponer, J., de Freitas, N., Bouchard-Côté, A., et al. (2016). "Bayesian analysis of continuous time Markov chains with application to phylogenetic modelling." *Bayesian Analysis*, 11(4): 1203–1237. [9](#)
- Zlatić, V., Ghoshal, G., and Caldarelli, G. (2009). "Hypergraph topological quantities for tagged social networks." *Physical Review E*, 80(3): 036118. [2](#)