**Authors' Responses to Associate Editor**

*The Hyperedge Event Model*

Thank you for acknowledging our efforts and contributions, and also for your constructive suggestions, which are very helpful to improve the quality of our paper.

- Presentation

  **Response:**

  - Unusual notations such as actor $A$ and covariates $y$. Also $u_{ie}$ is the $i$th line of matrix $u_e$, while $\tau_e = min_i(\tau_{ie})$. (done)
  - Change the order of Equation (2.2) and (2.1) and not use 'intensity'. (done)
  - Hard to understand 2.2 without 2.3. More explanation on $\tau_{ie}$ above Equation (2.5) and what $\mu$ represents in $V(\mu)$. (done)

- Discussion of relevant literature

  **Response:**

  - Perry and Wolfe (2013) arxiv version has a model for multicast. Any differences/advantages? (done)
  - Why cite Snijders (1996) in 2.3? Be specific. (done)

- Model, covariates and missing data

  **Response:**

  - Observations $(s_e, r_e, t_e)_{e=1,\dots,E}$ are not conditionaly independent since covariates depend on last 7 days. State this in Section 2 and modify out-of-sample algorithm using $(s_{e'}, r_{e'}, t_{e'})_{e':t_e<t'_e<t_e+l_e}$. (discuss)

- MCMC sampler

  **Response:**

  - Details on M-H proposals for $\boldsymbol{b}$ and $\boldsymbol{\eta}$ in Section 3.2. (done)
  - Inefficient sampler for $u_{iej}$ especially when most are one-to-one. Comment on this and the mixing of MCMC samplers. (discuss)

  - Move Geweke to appendix (done) and use larger number of nodes and events.
  - Computational complexity per iterations of the samplers.

- Typos (done)

  **Response:** We fixed all the typos identified by the reviewer as well as other writing issues, and we highly appreciate your considerable comments on these which were extremely helpful.

**Authors' Responses to Reviewer 1**

*The Hyperedge Event Model*

Thank you for acknowledging our efforts and contributions, and also for your constructive suggestions, which are very helpful to improve the quality of our paper.

- Presentation and writing

  **Response:** We fixed all the typos, unclear parts, and issues in the bibliography (bib not working) identified by the reviewer. We highly appreciate your considerable comments on these which were extremely helpful. (done)

- Literature review

  **Response:** We added a subsection following the description of the HEM in which we ground it in the structure of existing models for networks.

  - More comprehensive review including temporal ERGMS and dynamic latent variable models, and discuss contributions and novelties in the light of alternatives

- Section 2

  **Response:** Rewrite Section 2 to provide a much clear picture of the model.

- Prior specification

  **Response:**

  - Use weakly informative priors as generic priors instead of assuming $N(0, \infty)$ (done)
  - Sensitivity analyses to check how much posterior inference is affected by the hyperparameters' settings, and, possibly, suggest some default values. (discuss)

- Posterior computation (discuss)

  **Response:**

  - Type of MH, proposal distribution, acceptance rate, smart proposal
  - Comment on poor mixing on data augmentation
  - Extent of scaling, bigger dataset, information on computational time

- Application (discuss)

  **Response:**

  - Better baseline than random guess 1/18
  - Compare with SAOMs and extensions in PPE and PPC
  - Bad results in predicting timestamps (MdAPE) (done)
  - More conservative interpretations (done)

**Revision Plans**

1. Section 1: Literature review

   - Bomin: add literature review on Perry and Wolfe (2013) arxiv version and comment how our model differs from point process based models (AE1 bullet 2)
   - Bruce: add literature review on the general class of dynamic network inference including temporal ERGMS and dynamic latent variable models (R1 comment 2)

2. Section 2: Generative process

   - Bomin: change few notations and add notation table in Appendix (AE1 bullet 1)
   - Bruce: overall rewriting such as rephrasing or clarification (R1 comment 3 & AE1 bullet 1, minor ones already resolved)

3. Section 3: Inference

   - Bomin: add more MH details and add a subsection 3.2 for computational issues—e.g., complexity and limitations (R1 comment 4, 5 & AE1 bullet 4)
   - Bruce: check the added subsection 3.2 and revise

4. Section 4: Application

   - Bomin: re-run PPE considering conditional dependence (section 4.2) and update results & interpretations in 4.2 & 4.4 (R1 comment 6 & AE1 bullet 2)
   - Bruce: add why direct comparison with SAOMs not possible and come up with better idea than random guess...? (R1 comment 6)

5. Section 5: Conclusion

   - Bruce: possibly further discussing our contribution in the light of alternatives added in the literature review (R1 comment 2)

6. Bibliography

   - Bomin: tons of issues but somehow changes are not reflected...? Double check! (R1 comment 1)

7. Discuss: sensitivity analysis?

**Point-by-point discussion of Reviews and Response to Comments**

First, we thank the editor...

**Response to the Editor's Comments**

**E1** *Presentation: I found the presentation rather lacking; in particular, the description of the model in Section 2 is very difficult to follow, due to the numerous variables introduced, the order in which the different elements of the model are introduced, and the lack of definition for some of them. I think a major rewriting is needed here.*

**Addressed:** The Editor..

**E2** *Some notations are rather unusual, for example A for nodes. Although this is not indicated in the text, I assume this comes from the term "actor" in the stochastic actor-oriented model (SAOM) of Snijders, cited by the author? If this is the case, it would be worth mentioning it so that the reader recalls this later on. $y$ for the covariates is also rather unusual. Some notations are inconsistent. For example $u_{ie}$ denotes the first line of the matrix $u_e$, but $\tau_e = min_i(\tau_{ie})$.*

**Addressed:** The Editor...

**E3** *In section 2.1, I think it makes more sense to First introduce Equation (2.2) then (2.1). The term "intensity" used for $\lambda_{iej}$ is rather confusing. As the paper is concerned with a continuous-time model, one would expect that the term intensity refers to some point process, which is not the case here.*

**Addressed:** The Editor...

**E4** *It is difficult to understand Section 2.2 without reading section 2.3. The authors introduce the notation $\tau_{ie}$ above Equation (2.5), but do not explain what this represents (at this stage I thought it was a component of the vector $\tau_e$). I do not see what the variable $\mu$ represents in Equation (2.5), and if it is related to $\mu_{ie}$ in Equation (2.4)*

**Addressed:** The Editor...

**E5** *Some sentences are di cult to understand: Page 3: "we define a probability measure "MBG" motivated by the Gibbs measure"*

**Addressed:** The Editor...

**E6** *Discussion of relevant literature: The authors should provide a better discussion of how their model differs from other approaches, in particular the approach of Perry and Wolfe (2013), and Snijders (1996), based on point processes. Perry and Wolfe propose a specific model for multicast interaction (Equation (6) in the arxiv version of their paper). What are the differences/advantages between both constructions? In Section 2.3, the author cite Snijders (1996) when they introduce the model for the interaction arrival times. I am not sure what is meant here: is this part of the model already introduced in Snijders (1996)? The authors should be more specific here.*

**Addressed:** The Editor...

**E7** *Model, covariates and missing data: This is not explicitly mentioned in the definition of the model in Section 2, but the covariates $x_e$ and $y_e$ depend on the observed interaction data $(s_{e'}, r_{e'}, t_{e'})$ in the last 7 days, as written in Section 4.1. Hence the observations $(s_e, r_e, t_e)_{e=1,\dots,E}$ are not conditionally independent given the model parameters. This should be clearly stated in Section 2. For this reason, I do not think that the out-of-sample algorithm described in Algorithm 3 is correct. The conditional distribution for the missing observations $s_e$,$r_e$ and $t_e$ should depend on $(s_{e'}; r_{e'}; t_{e'})_{e':t_e<t_{e'}<t_e+l_e}$, which is not what is done.*

**Addressed:** The Editor...

**E8** *The authors should provide more details on the Metropolis-Hastings proposals for the parameters $b$ and $\eta$ in Section 3.2.*

**Addressed:** The Editor...

**E9** *The sampler usePs a blocked Gibbs sampler for the $u_{iej}$. Given the constraint that $\sum_j u_{iej} > 0$, this sampler may be rather inefficient in the case where there are many one-to-one interactions (as is the case in the application, where 83% of the interactions are dyadic). In order to go from the state $u_{ie}$ = (1,0,0,0,0,0) to $u_{ie}$ = (0,1,0,0,0,0) one needs to go through a state where two receivers are activated, which has low probability in this case. It would be good to comment on this, and in general on the mixing of the MCMC sampler.*

**Addressed:** The Editor...

**E10** *The authors provide in Section 3.2. some sanity checks on the sampler. While this is good practice to perform such checks, I am not sure it is very useful to include this in the main body (could be moved to the appendix), as this is done on a very small scale example with 5 nodes and 100 events, and does not really give an indication on the convergence properties of the algorithm in a more realistic scenario. I suggest the authors perform a simulation study with a larger number of nodes and events to demonstrate that the algorithm is able to approximate the posterior distribution well in that case.*

**Addressed:** The Editor...

**E11** *The authors should provide some indication of the computational complexity per iterations of their sampler. The application to email interaction data is rather small (18 nodes and 680 emails), so it would be good to know how many nodes/events the proposed approach can handle.*

**Addressed:** The Editor...

**E12** *Typos: The article contains numerous typos. Here are some of them:*

**Addressed:** The Editor...

**Response to Comments by the Reviewer**

**R1** *TYPOS...*

**Addressed:** The reviewer...

**R2** *UNCLEAR PARTS...*

**Addressed:** The reviewer...

**R3** *BIBLIOGRAPHY:...*

**Addressed:** The reviewer...

**R4** *As already mentioned, the authors focus on the somewhat less explored area of continuous?time relational event models where each edge has its own different time index?instead of considering time?varying models for snap- shots of networks collected on a pre?specified time grid. However, the contribution is still within the general class of dynamic network inference. In this respect, the literature review provides a poor picture for the state-of-the-art in this wider framework. I think the authors should provide a more comprehensive literature review including also temporal ERGMs and dynamic latent variable models (e.g. dynamic stochastic block models, dynamic mixed membership stochastic block models, dynamic latent space models, . . .). Discussing your contribution in the light of these alternative (and quite different) models would further clarify the key novelties our the proposed methods.*

**Addressed:** The reviewer...

**R5** *I fully understood the first paragraph in page 3 (summarizing HEMs) after reading the subsections 2.1, 2.2 and 2.3. This part should provide a much clear picture of your model instead of creating confusion. I suggest to improve it, leveraging also some intuitive illustrative figure. For example you could place Figure 1 much early and comment it while summarizing the HEM at the beginning of Section 2.*

**Addressed:** The reviewer...

**R6** *You present the full conditionals in page 7 assuming uninformative* $N(0,\infty)$ *priors, but then you rely on weakly informative priors in the application (see page 17). I found this confusing. I?d present results in page 7 for generic Gaussian priors.*

**Addressed:** The reviewer...

**R7** *Given that the model is relatively complex, some sensitivity analyses should be carried out to check how much posterior inference is affected by the hyperparameters? settings, and, possibly, suggest some default values.*

**Addressed:** The reviewer...

**R8** *There are many MH routines in the literature. Which type of MH do you consider? What is the proposal distribution? What about the acceptance rate? Is there any smart proposal in this case which helps in increasing the acceptance rate.*

**Addressed:** The reviewer...

**R9** *You rely on data augmentation MCMC, which has been shown to mix quite poorly (also in recent theoretical papers). Indeed, as expected, you end up thinning the chains every 40 samples in the application. However, there is no comment on this in the paper. I think it should be highlighted and not just hidden in the thinning.*

**Addressed:** The reviewer...

**R9** *Your quantitative assessments are based on 5 nodes in the simulation and 18 nodes in the application. These are quite small networks compared to those one would expect in real?world settings (such as those you list in the introduction). To what extent are your computational methods able to scale to much larger networks? An application to a bigger dataset would be useful. Moreover, more information on computational time should be provided.*

**Addressed:** The reviewer...

**R10** *You compare performance in predicting missing senders with random guess 1/18. This is a quite naive competitor and I am sure the authors can find much better ones?still relatively simple.*

**Addressed:** The reviewer...

**R11** *In Figure 4c the choice of the logarithmic scale for MdAPE seems to hide a quite poor performance of your model in predicting the timestamps. It is true that the log?normal improves over the exponential, but the log?normal boxplot has still a third quartile providing an MdAPE of exp(7.36) which looks quite big. This is to me a negative result, which should be discussed and addressed in the paper.*

**Addressed:** The reviewer...

**R12** *Based on your comments in the paper you are interpreting the posterior expectations of exp($\eta_6$) and exp($\eta_7$)–i.e. E[exp($\eta_6$) — data] and E[exp($\eta_7$) — data]. However you seem to compute this as exp[E(exp($\eta_6$) — data)] = exp(1.552) and exp[E(exp($\eta_7$) — data)] = exp(0.980). By Jensen inequality this is wrong.*

**Addressed:** The reviewer...

**R13** *It is also not correct to claim that the $e^{th}$ email is expected to take E[exp($\eta$) — data] longer compared to their counterpart. The term longer seems to refer to an additive effect on the event timing, which is not the case here since you assume a log?normal for the event timing and hence the effect is on the scale of the timing and not on the location?note that, if $X \sim$ log-normal$(\mu, \sigma^2)$, then $E(X) = exp(\mu + \sigma^2/2)$. The authors should be more careful in the interpretations on the original time scale, or, more conservatively, they could simply comment on the effects on the log-time (as done immediately after).*

**Addressed:** The reviewer...