

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

First, we thank the editor for acknowledging our efforts and contributions, and also for your constructive suggestions, which are very helpful to improve the quality of our paper.

### 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 pointed out the lack of detailed and coherent explanation on the model in Section 2, suggesting better definition of variables and improvement on writing. We have significantly revised Section 2 by changing the order of subsections along with general rewriting—see E2 to E5 for details.

**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 identified some unusual notations which are not consistent with the literature. Instead of  $A$  for nodes (which we got idea from SAOM as the Editor assumed) we now use  $V$  for nodes (referring to vertex in network literature), and we also replaced  $y$  by  $z$  for time-related covariates—which are clearly stated in Section 2. We also added detailed definition on the variables and notations, such as the usage of subscripts  $i, e, j$  right before Section 2.1.

**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:** Following the Editor’s suggestion, we changed the order of Equation (2.2) and (2.1) as well as their following texts to help readers’ understandability. Moreover, we agree that the term “intensity” could be a confusing term so we revised the sentence without using the term.

**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 expressed difficulty in understanding Section 2.2 and 2.3, and suggested further clarification on the variables  $\tau$  and  $\mu$ . We moved Fig. 1 to earlier parts of Section 2 and briefly summarized the entire Section 2, such that the readers can understand Section 2.2 without reading Section 2.3. We also further explained the concepts and notations related to timestamp variables,  $\tau$  and  $\mu$ , to avoid any confusion.

**E5** *Some sentences are difficult to understand: Page 3: “we define a probability measure “MBG” motivated by the Gibbs measure”*

**Addressed:** According to the Editor’s comment, we revised the sentence mentioned above and also went through entire draft to revise sentences that were not clearly written before.

**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 indicated lack of discussion on the relevant literature, including specific comparison between other papers we mentioned (e.g., Perry and Wolfe (2013) and Snijders (1996)). We created new section 2.4 “Basis in existing models” which provides detailed summaries and comparisons about existing models that affected the HEM. Furthermore, we added more comments when we introduce/cite the papers with emphasis on what are the motivation and how our model uses differently.

**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 called attention to a very important point about conditional dependency that we were missing. Our out-of-sample algorithm (Algorithm 3) was not correct, and the conditional distribution for missing observations had to be changed as the Editor illustrated. We modified our algorithm to consider those dependencies, and Section 4.2 is fully updated based on new conditional distribution and experiment results.

**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 asked to provide more details on the proposals we used for the Metropolis-Hastings algorithm, thus we added further explanations for both parameters  $b$  and  $\eta$  in Section 3.1.

**E9** *The sampler uses 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 in-*

teractions (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 identified some inefficiency issue in our inference algorithm, which we did not fully commented. Considering our general need to discuss more about our model’s computational complexity, we generated new section 3.2 about “computaional complexity” and left various comments including the usage of a blocked Gibbs sampler and the mixing of our MCMC sampler.

**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:** Previously we presented Section 3.2. as an effective way to check the validity of our derivation and code using small scale example, which is slightly different from traditional simulation studies that are designed to provide realistic scenario under the model and/or show convergence properties. Although we believe that GiR test is a good practice and should be highly recommended, we still admit that Section 3.2. does not provide significant amount of additional information on the HEM itself (other than sanity checks) and thus moved this section to Appendix C. For convergence properties of the algorithm, we created new section 3.2 about “computaional complexity” and left some comments on convergence using our application data.

**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 suggested more detailed explanation on the computational complexity, especially since our application data is rather small, so (as mentioned in E9) we generated new section 3.2 about “computational complexity” and left comments on our computational limitation. As the Editor pointed out in E9, we have some inefficiency issue in our inference algorithm which limits the applicability of our model for large scale dataset (with huge number of nodes/events). We admitted this limitation in new section 3.2.

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

**Addressed:** The Editor identified numerous typos, and we properly fixed all those. We greatly appreciate the Editor’s help on these.

### **Response to Comments by the Reviewer**

**R1** *TYPOS...*

**Addressed:** The reviewer identified numerous typos, and we properly fixed all those. We greatly appreciate the Editor’s help on these.

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

**Addressed:** The reviewer pointed out multiple unclear parts such as notation issues and sentences that needs more clarification. We went through each item and revised our draft as suggested by the reviewer.

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

**Addressed:** The reviewer asked the authors to carefully review the bibliography and correct all the misspecifications (e.g., uncapitalized book title, wrong upper/lower cased letters). We revisited each bibliography and addressed the existing issues. We are very thankful for your comment.

**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 indicated lack of discussion on our model’s contribution on the general class of dynamic network literature, specifically in lights of temporal ERGMs and dynamic latent variable models. We created new section 2.4 “Basis in existing models” which provides detailed summaries on our contribution and comparisons with other existing models in wider framework.

**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:** As suggested by the reviewer (as well as the Editor), we found out the lack of detailed and coherent explanation on the model in Section 2. We have significantly revised Section 2 by changing the order of subsections along with the general rewriting, and Figure 1 is also moved to earlier parts when summarizing the HEM at the beginning of Section 2.

**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 suggested to use generic Gaussian priors in page 7, instead of

introducing the full conditionals assuming uninformative  $N(0, \infty)$  priors. We addressed this issue by assuming generic Normal priors  $N(\mu, \Sigma)$  for both  $b$  and  $\eta$  and revised our equations and corresponding explanations in page 7.

**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 suggested more sensitivity analyses on hyperparameter specification. We tried two alternative specification of prior distribution—one with half size of current variance and another with double size of current variance—and summarized the results in Appendix. Also, we left some comments on these settings in Section 3.2.

**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 suggested more detailed explanation on the inference scheme such as the type of MH, proposal distribution, acceptance rate, and any recommendation on proposals. We not only provided additional explanation on some generic proposal distribution in Section 3.1, but also generated new section 3.2 about “computational complexity” for other information on our inference scheme in general.

**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 suggested more detailed explanation on the computational complexity, especially because we had slow mixing of MCMC samples and thus end up with large thinning. In our new section 3.2 about “computational complexity”, we left comments on our computational limitations, including poor mixing of data augmentation.

**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:** Again, we created a new section 3.2 to discuss more about the HEM’s computational complexity. Here we acknowledged the HEM’s limitation on its applicability to a bigger dataset, by outlining the scalability of the model and computational time. We also mentioned some possible ways to improve the model’s scalability as a future direction of research.

**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 suggested to consider alternative measure to compare our performance in Section 4.2., instead of our current comparison (i.e., random guess 1/18). We came up with the idea to use Dirichlet Multinomial as a baseline to sender prediction, and updated our comparison in Section 4.2. with some explanations.

**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 misunderstood Figure 4c due to our lack of explanation on using the logarithmic scale. We used the logarithmic scale to draw the boxplot (since original scale was highly right-skewed, making the plot less informative), however, we maintained the original scale on the axis. As a result, the third quartile provides an MdAPE of 7.36 (not  $\exp(7.36)$ ). We revised the draft such that the readers are not confused with log-scale.



**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 commented on our misinterpretation of coefficients. We changed the way we summarize the estimated value of  $\eta_6$  and  $\eta_6$  (as we did with other  $\eta$ 's) so that we no longer provide wrong interpretation.

**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  $X \sim \text{log-normal}(\mu, \sigma^2)$ , 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 also pointed out additional mistake when we provided the interpretation of coefficients on the original time scale. As suggested, we simply commented on the effects on the log-time (as done immediately after) instead of using the original time scale.