We greatly thank the reviewers for their constructive comments. We would like to specifically emphasize that the proposed frequentist confidence interval method is different but complementary to a Bayesian UQ approach; please see our response to Reviewer 2. We will also include the suggested references for Bayesian UQ methods.

Response to Reviewer 1. Thank you for the valuable feedback about model misspecification and low-rank parameterization. In the neuronal connectivity application, while it is commonly agreed that Hawkes process is a good fit for the spike train data, such data are generated from a scientific model (i.e., **not** exactly Hawkes; in fact, the synthetic data are generated by solving a complex PDE system), so there is indeed a slight model mismatch in our numerical example. In future work, it will be very interesting to study the influence of model mismatch and the confidence sequences when the weight matrix A is low-rank or sparse.

Response to Reviewer 2. 1) The first-order Taylor approximation of $g_k(\boldsymbol{\alpha}_i^*)$ at $\widehat{\boldsymbol{\alpha}}_i$ is for the purpose of giving a tangible form (polyhedron) of the confidence set. However, we can definitely evaluate g_k at every $\boldsymbol{\alpha}_i$ and perform numerical inversion to find the confidence set. Moreover, we have empirically validated the performance of the confidence set from Algorithm 1 in the numerical example. 2) The integrals in Section 3.3 (which go into Algorithm 1) is a one-dimensional integral over t, and hence can be solved efficiently by standard quadrature techniques. 3) The UQ for the intensity can be derived from the UQ for the weights A; this will be an interesting future work.

4) **Comment on the comparison with Bayesian approach.** Given sufficient computing resources and little domain knowledge, we agree that a fully Bayesian nonparametric approach may provide a richer quantification of uncertainty. However, in certain applications, there is scientific evidence for a *parametric* form of the triggering kernel (e.g., [Beggs (2008)] for our neural spike data application). For large networks, the full posterior may also be computationally expensive to sample with HMC, since one needs to tune both the stepsize and number of leapfrog steps, among other settings. We will add further context on when our method may be more or less preferable to a fully Bayesian approach in practice.

5) Due to space limitation, we did not include in the paper a simulated example of confidence bands as a function of T. We show here a small example with 5 nodes in Fig. 1. We see that the proposed CIs are valid and becomes narrower as T increases, as desired. Moreover, when T is small, Fig. 1 also shows that the asymptotic CI can have poor coverage performance (i.e., it does not contain the true parameter), whereas our proposed method uniformly covers the true parameter even for small T. We have similar observations for recovering other edges in this example.

Response to Reviewer 3. We agree that the idea can be extended beyond Hawkes processes, and it would be an interesting future direction.

1) The reason for focusing on the Hawkes model is motivated by the neural connectivity application, and we would like to generalize this to broader models in future work. Moreover, we will adjust Section 2-3 as suggested.

2) The width of the main confidence band in the non-asymptotic case in Corollary 1 will be similar to that of the asymptotic confidence band width shown in Proposition 1, since the estimated Fisher information $\hat{I}_i$ will converge to the true I_i^* as $T \to \infty$.

3) Proof of Theorem 3.3 mirrors the proof of Theorem 3.2 by replacing the fixed vector $\boldsymbol{z}$ with the adapted measurable function $\boldsymbol{z} : (\mathcal{H}_t)_{t=0}^T \to \mathbb{R}^D$, and Corollary 1 is a straightforward consequence of Theorem 3.3 by considering multiple $\boldsymbol{z}$ functions, the choice of $\boldsymbol{z}$ in Corollary 1 is inspired by Proposition 1.

4) An example of the integral in Line 121 for the exponential kernel function is given in Appendix A; The $\boldsymbol{\alpha}_i$ in Lemma 3.1 actually can be an arbitrary $\boldsymbol{\alpha}_i$, not necessarily $\boldsymbol{\alpha}_i^*$; Equation (10) holds uniformly in $\boldsymbol{z}$; the departure set of the function $\boldsymbol{z}$ means that $\boldsymbol{z}$ is determined by past event times; as justified in Response to Reviewer 2 - point 1, the approximation aims to provide a tangible form of confidence set and is validated empirically, and we can also evaluate g_k numerically to find the exact confidence set; the analysis on dN_t is a standard technique in Hawkes literature and is commonly seen in related literatures, such as [Hawkes (1971)] [Bacry and Muzy (2016)].

5) Comment on Line 378: We appreciate reviewer's careful reading and suggestion; we believe the final conclusion is correct (as validated by numerical examples); we will make this more rigorous in the full paper.

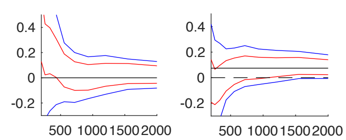

Figure 1: The CIs for two select edges at level $\varepsilon = 0.05$ over time T, for a Hawkes network with 5 nodes and influence functions are $\varphi_{ij}(t) = e^{-t}$. In this picture, the solid blackline is true α_{ij} and the dashed line is the zero level. In each picture, the two blue curves outline the proposed CIs, and the two red curves outline the asymptotic CI. Note that the proposed CI uniformly covers the true parameter even when T is small, whereas the asymptotic CI can have poor coverage when T is small. This shows the advantage of our method over the asymptotic approach, and that our method works well for small T.

Response to Reviewer 4.

1) In the revised paper, we will add a paragraph to discuss the difference between our parametric approach with Bayesian model and add all mentioned references. 2) Currently there is no sparsity structure imposed in the maximum likelihood estimate, since our main goal is to obtain confidence intervals rather than point estimators; this is an interesting direction and we will leave it for future work.