We thank the reviewers for the valuable comments and discussions. Please find our clarifications below. We use [Narasimhan et al.'15] for Narasimhan H et al., Learnability of influence in networks, NeurIPS'2015.

- Reviewer 3: About the setting of online linear threshold model

Recall that in setting of the (offline) linear threshold (LT) model, the weight $w(e)$ associated with each edge $e \in E$ is fixed and known and the threshold for each node is uniformly drawn from $[0, 1]$. The weights are model parameters while the thresholds are not model parameters. This setting was originally proposed in the seminal work by Kempe et al. [19], and most follow-up studies adopt this particular setup (e.g. [9,12,16]). The particular setting of using fixed and nonnegative weights on the edges (with the sum of weights of the incoming edges of each node at most 1) plus the uniform sampled threshold from $[0, 1]$ enables the LT model to have an equivalent live-edge graph formulation, and unifies the LT and IC under the more general triggering model [19], which is in turn important for deriving a number of properties (such as submodularity) and algorithmic solutions (such as the reverse influence sampling approach [38]) for the LT model. Therefore, our online influence maximization (OIM) is directly on the classical LT model, turning model parameters $w(e)$ to be unknown (but fixed) and to be learned in an iterative manner. This is in parallel to the OIM for IC model [11,43,45], which also learns unknown edge probability parameters.

It is interesting that the reviewer brought up the frequentist versus Bayesian view on OIM-LT. Per our above discussion, we first want to clarify that the threshold on each node is not a model parameter of the classical LT model and our work is a frequentist approach for the online setting. Alternative Bayesian approach, such as Thompson sampling algorithm, is one of the future directions we plan to explore next. It is also possible to analyse Bayesian regret under Bayesian setting where the weights follow some prior distributions. The offline problem of including thresholds as model parameters, where both weights and thresholds are fixed and known, is the fixed threshold model [9,19] and has very different property and behavior, e.g. it is not submodular and is NP-hard to approximate to any nontrivial factor [9,19]. Also its diffusion process is deterministic for each seed set, while under IC and LT models the diffusions are random. So how to design its online setting and the corresponding Bayesian setting would be interesting future directions.

- Reviewer 1&2: About node-level feedback

First, we will clarify our naming by saying "full node-level feedback" as knowing the set of nodes activated at each time step, and "partial node-level feedback" as knowing only the set of nodes activated by the end of the diffusion process. This naming is consistent with [Narasimhan et al.'15], which also shows that these two types of feedback give different PAC-learnability results — the full feedback allows polynomial-time learning algorithm while the partial feedback may require an exponential-time learning algorithm.

Our paper studies the full node-level feedback. As reviewer 1 pointed out, it is possible to apply LinUCB-type algorithm to partial node-level feedback, but the difficulty is to analyze its regret. The key to bound the regret is to prove a similar GOM property (our Theorem 1) using the information that can be observed. Such a Lipschitz-type property is essential since we always use estimated weights to select seeds in the online setting and need to bound the difference caused by the estimation error. Thus, Theorem 1 is one of our main technical contributions, and extending it to the partial node-level feedback is unclear at the moment and is part of future research work. The probability 1/2 (line 12 of Algorithm 1) comes from the key GOM property (Theorem 1) and also addresses the correlation between $E_{t,1}$ and $E_{t,2}$.

- Reviewer 1: The idea of using LinUCB and the importance of LT model

(a) Yes, the idea of using LinUCB is natural, since the diffusion process involves linear structure. As we stated above, the main difficulty is to analyze the regret for such a specific setting and feedback. (b) Although IC may be studied in the literature more than LT, LT is still a fundamental diffusion model and various aspects of LT has been studied (e.g. [9,12,16,19,20,38] and many other studies). For the online setting, existing work does focus on IC with edge-level feedback [11,43,45], and this is because the independence on edge-level propagation and edge-level feedback make the setting easier to analyze.

- Reviewer 2: Experiments and the definitions of $E_{t,1}, E_{t,2}$

(a) We have shown experiments to compare LT-LinUCB and OIM-ETC in Appendix F. Since the algorithm of randomly selecting seeds does not learn good seed set, we did not include it. (b) For each diffusion, $E_{t,2}(v)$ denotes the set of incoming edges of active in-neighbors for node v at the time when v is activated; and $E_{t,1}(v)$ denotes the set of incoming edges of active in-neighbors for node v just one time step before v is activated, which are the edges just failing to influence v; if v is not activated in the end, $E_{t,1}(v)$ is the set of incoming edges of all its active in-neighbors, which is the largest edge set failing to influence v. We will add more descriptions for better understanding of these two terms.