We thank the reviewers for their constructive feedback. We first address the general comments and then answer specific questions.

GENERAL COMMENTS

Reviewers 1 and 2 both enquired about the novelty of the proof techniques in this paper. In a nutshell, the original proof from Mitrovic et al. (2018) could not be extended to (weak) adaptivity because it was not linear enough to allow for the necessary and extensive use of the linearity of expectation. Our proof replaces several of their lemmas with tighter, more linear analyses (e.g., Lemma 5 in our paper replaces Lemma 3.3[1]). We agree that a high-level summary would be useful for clarity, so we will add a paragraph in the main paper summarizing the techniques we used to achieve the necessary linearity, and the resulting benefits.

Reviewers 1 and 3 brought up valid questions about the computational feasibility of this approach in practice. It is true that if the distribution of realizations is difficult to calculate, then our approach may struggle. Theoretically speaking, an approximation of the expected values which is in a $1 \pm \varepsilon$ range of the true value introduces only a $(1-\varepsilon)$ factor in our guarantees. Furthermore, many real-world applications can utilize efficient and accurate approximations. For example, in our first experiment, the "true" distribution over a user's product preferences is unknown. Instead, we approximate the conditional probabilities simply by counting the number of times a shopper has purchased product i after product j. In fact, as the attached code shows, the runtime of our approach is superior to the deep learning methods. In the revised version, we can add graphs comparing the runtimes. We also note that there are several works on speeding-up greedy algorithms (e.g., stochastic greedy), which can help us scale our approach to much larger datasets.

REVIEWER 1

"I think the hardness of approximation results should be $n^{\frac{-1}{(\log \log n)^c}}$, so that it is a number less than 1."
Thank you for noting this, we will fix it. We wanted to be consistent with the paper we were citing, but you are correct that this is inconsistent with the rest of our paper.

"What is the weak submodularity constant [of the probabilistic coverage utility function]?"
It depends on the data, but roughly speaking, it is inversely proportional to the smallest edge weight in the underlying graph.

"...this type of problem (no user features or item features) is not where LSTMs are known to shine."
This is a fair point and we will add it as a remark in the paper. We would like to note that our approach can also take advantage of user/product features (as a part of the distribution over realizations), but the data did not include any additional features.

"To improve clarity, the authors should formally write out the optimization problem that they want to solve."
"... authors were able to improve results from (Mitrovic et al., 2018) ... feature these results more prominently."
We appreciate the reviewer's suggestions for improving the clarity of the paper, and we will incorporate these changes.

"Improvements: If the authors could tighten the approximation guarantees..."
We have hope that it is possible to remove the 2 from the current approximation factor of $\frac{\gamma}{2d_{\text{in}}+\gamma}$, but we believe that achieving a sub-linear dependence on the degree would require a totally different proof approach.

REVIEWER 2

"In line 110, the authors explain how the state of an edge is decided, but this part is not clear to me."
The basic idea of our framework is that the state of each edge is determined entirely by the state of its endpoints. It may be easier to think about it as a function that takes in the states of the two endpoint vertices and outputs the state of the edge. One key advantage of our framework is that it works with *any* such function. For example, in our experiments, the state of each edge is assigned to be the same as the state of its start vertex. We will clarify this point further in the revised version.

REVIEWER 3

"I think the framework needs to be extended so that a set/sequence of elements is selected at each step ..."
Generally speaking, the point of adaptive submodularity is to select items one at a time and receive feedback, so our theoretical analysis followed this standard. That being said, we agree that in practice it is common to have to select multiple items before receiving feedback. Our approach can be easily adjusted to do so by simply delaying the observation of the states of the selected vertices (i.e., waiting to update the conditional distribution). This idea can also be linked to the recently studied problem of submodular maximization via "adaptive rounds", where in each round one picks multiple items in parallel [5, 6, 18]. It must be noted that new theoretical analysis would be required for this change, but we agree that batch selection for weakly adaptive sequence submodularity would be an interesting direction for future work.

"...it is not intuitively clear to me why the function is modeled as weakly submodular ..."
There are two main reasons for using weakly submodular functions. First, weakly submodular functions can model a much greater variety of problems (Khanna et al., 2017; Elenberg et al.,2017). Second, weak submodularity is strictly more general. Intuitively, γ tells us how close the problem is to being submodular and when $\gamma = 1$ we recover regular submodularity. We want to emphasize that we do not need to know γ to run our algorithm, but we do agree that γ can sometimes be infeasible to calculate, and thus, we may not know the guarantee we are getting in practice.

"The current method only outperforms DNN based methods when there are few training data ..."
Given enough data, neural network-based approaches show state-of-the-art performance for most tasks, so we did not expect our method to achieve superior accuracy in such scenarios. Instead, our approach provides numerous other advantages such as theoretical guarantees, interpretability, ease of implementation, and robustness (while still maintaining comparable accuracy). In the second experiment, we saw that even when we use all available data, it is still not enough for the deep learning-based approaches, which shows that the robustness against data scarcity is a practically significant aspect of our method.

[1] As numbered in the version of their paper appearing in `https://arxiv.org/pdf/1802.09110.pdf`.