We thank all reviewers for appreciating the importance and the challenge of the problem we tackle in this paper. We are glad to hear that all reviewers found the paper to be well written.

Reviewer 1:

1. With increasing density in the graph, the experiment design part will have more flexibility to redistribute the exposure optimally, and the post-experiment adjustment part will have a larger sample size for density estimation (see Section A.5 in the supplement). We will add these to the paper.
2. The shadow population is not the same as the control population. The shadow population members are randomly assigned to consumer-side treatment or consumer-side control (as in classical A/B testing design). In the absence of interference (i.e., the classical setting), the shadow population is the same as the measurement population. However, these two populations are different in the presence of interference (i.e., non-zero producer-side effect), since the shadow population members do not get the correct producer-side experience.

Reviewer 2:
Thank you for sharing the papers. We will cite and discuss the relevant ones.

Reviewer 3:

1. In Theorem 1, we do not use $Y_i(T_E^{(r)})$ to compute the average treatment effect (since they are unobservable). Instead, we apply importance sampling adjustment to the observed response $Y_i(T_E^*)$. The importance weights are computed based on the observed $Z_i(T_E^*)$ values and estimated densities of $Z_i(T_E^*)$ and $Z_i(T_E^{(r)})$. For the density estimation of $Z_i(T_E^{(r)})$ from the experimental data, please see Section A.5 in the supplement. Note that we do not need to know the p_{ij} values at this point. However, we use the $p_{ij}^{(r)}$ and the p_{ij}^{base} values for the design of experiment (see eq. 2). To this end, we assume that the $p_{ij}^{(r)}$ and the p_{ij}^{base} values are known in the simulation settings, whereas we estimated them from the data in the real-world experiment (lines 285–293).
2. Thanks for catching this. It should be $Pa(i)$ (see line 105).
3. The presence of an intervening variable is indeed a strong assumption. However, it is a natural assumption in some situations. For example, in the content market place setting, it is a reasonable assumption that the producer-side experience of a member is driven by the total feedback received from the content-consumers. It is challenging to develop a sound method for verifying this assumption, and it is beyond the scope of this paper.
4. Thanks for catching the typo. It should be Figure 2 instead.
5. Thanks again, it should be eq. (2).
6. Note that the estimator in Theorem 1 is an empirical average when the densities are known, implying the consistency of the bootstrap variance estimators under mild assumptions. When the densities are unknown, the consistency of the bootstrap variance relies on the consistency of the density estimation (which holds under some regularity conditions). We will add references.

Reviewer 4:

1. Thank you for sharing the papers. We will cite and discuss the relevant ones.
2. In our real-world experiment, we need to have millions of observations to detect small improvements in metrics (1-2%). While OASIS experiments can have millions of observations, the state-of-the-art ego-cluster experiments would be, for example, limited to 100K i.i.d. observations if we fix the proportion of inter-cluster and intra-cluster edges to 1:4. In this particular setting, we do not have a reasonable cluster-based baseline to compare with. However, this is indeed a valuable future work to compare OASIS with a reasonable baseline in real-world experiments.
3. We appreciate your feedback, and we will try our best to improve the readability.