We thank all of the reviewers for their comments and suggestions.

Reviewers 2, 3, and 5 asked how well our proposed algorithms address the identified challenges. For the challenges (i) and (ii) involving reviewer manipulation of the assignment, our problem formulations *consider the worst-case manipulation*; that is, we want to effectively address the challenges even if a malicious reviewer could somehow always get themselves assigned to their desired paper under a standard (deterministic) assignment algorithm. In the worst case, the maximum probability of that reviewer-paper assignment given as input to our algorithm *exactly specifies how well we address the manipulation*, since it is the probability that the manipulation works (reduced from 100% in the deterministic case). This approach allows us to *avoid making assumptions* on what reviewers are willing to do that may not hold in practice; for example, reviewers have been known to fully falsify reviewer profiles to get a desired paper assignment. For the de-anonymization challenge (iii), the maximum probability of a reviewer-paper assignment given as input to our algorithm is exactly the highest certainty that an author can have that a specific reviewer reviewed their paper (assuming the review contains no identifying information), *specifying how well we address this challenge*.

Reviewers 1 and 5 raised similar points regarding alternate approaches to the problem that more explicitly model reviewers' manipulation of similarity scores (e.g., reducing the influence of bids in the similarity computation). Modeling reviewers' manipulation of similarities more explicitly *would require making some assumptions* on reviewer behavior/incentives (which may be unlikely to hold in practice) as well as on the similarity computation used (which varies across conferences). Our approach to the problem **does not make any such assumptions** and instead provides worst-case guarantees as described above.

Reviewer 1:

- *Regarding surveying reviewers to determine the ground truth assignment quality:* We agree with you that similarities are a noisy approximation. However, similarities are the standard measure of assignment quality used in past work as well as in practice (including at NeurIPS 2020!), and the question of better evaluating assignment quality even in the absence of malicious behavior is open. Additionally, since we compare our proposed assignment to the standard deterministic assignment and the similarities are used to approximate the assignment quality for both assignments, any noise in the similarities should not impact our evaluation significantly.
- *Regarding conflict-of-interest constraints:* As we mention in Line 115 of the paper, conflicts-of-interest can be incorporated into the similarity matrix by assigning a similarity of $-\infty$ to any reviewer-paper pair with a conflict of interest, so conflict-of-interest constraints do not change the problem and all results go through.
- *Regarding the "simpler version of the sampling algorithm" (Line 174):* The algorithm does ensure that the load balance constraints are met every time.
- *Regarding the additional line of related work:* We appreciate your suggestion, and will expand our coverage of related work in any revisions of the paper.

Reviewer 2:

- *Regarding an empirical evaluation of our algorithm's effectiveness at addressing reviewer manipulation:* In real-world conferences, reviewers do not reveal the ground truth of whether they engaged in manipulation, so it is not possible to empirically evaluate the extent of reduction in manipulation. This motivates our "worst-case" view on manipulation described above where we aim to optimally mitigate manipulations no matter what the adversary does.
- *Regarding comparing our algorithm's performance against prior work:* To clarify, we theoretically show that our algorithms **optimally** solve the problems they are designed for. Furthermore, there is no prior literature in peer review addressing the challenges we identify, and the literature in other fields such as recommender systems requires additional data or assumptions that are unavailable or inapplicable here. Hence, our empirical evaluation compares our algorithm against the current method used in practice (including in NeurIPS 2020), a deterministic assignment algorithm.

Reviewer 5:

- *Regarding reviewer incentives for manipulation:* It is impossible to entirely remove incentives for reviewers to behave maliciously unless we make strong additional assumptions or have access to some information exogeneous to the conference (e.g., external communication between reviewer and author). Since we cannot entirely prevent manipulation in our setting, we aim to mitigate it as much as possible.
- *Regarding the use of the algorithm from [4]:* We would like to clarify that we optimize and simplify the algorithm from [4] significantly for our setting.