We thank the reviewers for their work. We ask Reviewer 2 and Reviewer 3 to reconsider their decision in the light of our feedback.

Reviewer 1

1. The assumptions used in cluster recovery are fairly strong in my opinion. It is not clear to what extent they are required. The assumptions are necessary, at least in part. For $\varepsilon \geq \frac{1}{2}$, changing the sign of εC edges of a node in a cluster C makes C not optimal anymore, so no cluster recovery algorithm should output it. Moreover, this holds for exact recovery, while for recovery *in expectation* our assumptions are weaker.

Reviewer 2

[W1] [...] The authors assume that the similarity function has binary output. This means that their work is highly focussed towards entity resolution based setting where the ground truth consists of a collection of entities and the goal is to resolve the entities. If this is the case, the authors should perform a thorough job of comparing with most recent works on entity resolution. [...]
[W3] The data sets are really small with no real crowd experiments validating if similarity queries are easy to answer for a user or not. The authors have used Yes/No answers from entity resolution crowdsourcing experiments as 1/0 similarity values and it does not seem like a fair justification of the proposed querying oracle. Our paper is not focussed on entity resolution and does not propose a query oracle. It is a theoretical study of correlation clustering under the lenses of active learning theory, with a standard query model that has been used for decades. This model uses a binary similarity function exactly because this captures entity resolution (ER) and many other problems, as we say clearly in the introduction. Thus, we don't think our approach is "focused on" ER. It is the CC framework which *comes from* ER-type problems. Because CC has been studied for over 15 years, we do not feel compelled to provide any further justification. Similarly, we do not see any reason to compare against ER: our paper focuses on the fundamental algorithmic mechanisms without referring to any specific applications domain,

[W2] Thm 1 does not seem tight e.g. when $Q = n^2$, we get expected error less than $3\mathrm{OPT} + O(n)$, whereas we know that the expected error should be $3\mathrm{OPT}$ as it is same as KwikCluster. We do not understand the reviewer's comment. The last three lines of the theorem state that, in the special case $Q = \binom{n}{2}$, our algorithm *exactly* achieves error $3\mathrm{OPT}$ like KwikCluster does, with no additive error term. We will clarify this should the paper be accepted.

[W3] There is no comparison with any baseline (Not even kwikcluster). We do not understand the reviewer's comment. The experiments report *precisely* a comparison with KwikCluster (caption of Figure 1, last line).

[W3] In the plots, authors do not mention OPT (or $3\mathrm{OPT}$ if calculating OPT was not possible). We do not know OPT since computing it is NP-hard, and the same obviously holds for $3\mathrm{OPT}$. However, the plots show the cost achieved by KwikCluster, which in expectation is precisely between OPT and $3\mathrm{OPT}$.

[W4] Authors should analyze the performance if the edge labels (binary similarities) are flipped with some probability. We are unsure about what the reviewer means. If one takes an instance and flips the signs of some edges, then one simply obtains another instance, whose optimum will be in general different, and to which our results still apply. We provide cluster recovery results even under adversarial edge flips, if this is what the reviewer meant.

Reviewer 3

[...] the queries [...] are prohibitively large close to n or n^2 in most cases. I believe this is a crucial problem with the model considered here. The model of Ailon [...] is much more reasonable and there they get guarantees compared to k the number of clusters in the optimum solution that is usually much fewer than n. [...] the originality of the query model is decreased and the significance of the results as well. We strongly disagree with this comment. Counting the number of similarity queries is the standard model for the analysis of query complexity, sublinear algorithms, and active learning on graphs, and we do not claim any originality on it. It captures the inherent effort of obtaining similarities, as this is a major bottleneck in practice (as Reviewer 2 points out, too). Ailon's model is about "computational complexity with advice": it gives all the $\binom{n}{2}$ similarities for free, and then equips the algorithm with an oracle so powerful to turn an APX-hard problem into a polynomial one! Only the queries to this oracle are counted in the analysis. We do not find this "much more reasonable" than our model; it is simply different. And, to be fair, Alon's algorithm needs *more* queries than us: $\binom{n}{2}$ similarity queries plus $O(k^{14}/\varepsilon^6)$ oracle queries, even fixing k. We will include this discussion in the related work section should the paper be accepted.

When $Q = n^2$ that the suggested algorithm matches the guarantees of KwikCluster but is not there an additional $O(n)$ term? Please see [W2] of Reviewer 2.