Overall. We appreciate the remarks and note that a number of reviews recommended accepting the paper. Moreover, everyone seemed to understand our p-norm model, algorithmic contributions, and experiments. And we also appreciate the useful reviews and concrete statements about our paper! (Thanks!!)

Regarding the focus on theory vs. algorithms and experiments.. We focused the 8 page paper on the experiments and setting up the problem, model, and algorithm – without as much space devoted to the planted problem theory (Theorem 4.1). The particular assumptions underlying it were explained more in the supplementary materials in terms of why they are reasonable. After getting your feedback, we remain convinced this balance is the right call, although we would attempt to add just a few more statements on the rationale for assumptions 1 & 2 into the paper.

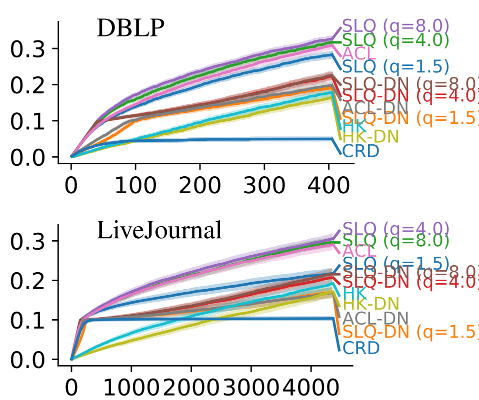
Reviewer 1.4 (Correctness). κ=0 means the L1 regularization term becomes zero, while the second sum is not part of the L1 regularization, but a part of the cut objective on the modified graph (see line 122-123 in main).

Reviewer 2.3 (Weaknesses). Regarding Figure 1, the target set in that figure is fully connected because each pixel is connected to others within distance 40 (so the cluster does have small diameter). That example, however, is not covered by the recovery theory (Theorem 4.1) because we assume unweighted graphs in a few places. We will also admit that other approaches may remove assumptions in Theorem 4.1 (but we don't know how yet).

Reviewer 3.3 (Weaknesses). Our apologies for not defining Gap. That was an oversight and we would make space for that definition in the final one if it were accepted and add some additional intuition (see above). We use the same definition as in the previous manuscript.

Reviewer 3.8 (Feedback). Regarding $q < 2$ or $q > 2$. In the conductance theory, we show that $q < 2$ is better. But to present a more rounded evaluation, we wanted to study a problem where conductance wasn't the objective. Kleinberg and Kloumann found that ACL/PageRank – with the standard degree normalization for conductance based sweepcuts performed WORSE than PageRank/ACL without degree normalization in this particular setting. So this experiment is a case where the algorithms behave differently from what we would expect based on conductance theory. (more precisely... conductance theory says you get better results with degree normalization and also $q < 2$). So what we wanted to show was that we ALSO find something different from conductance theory using the flexibility with q, which is what the figure shows. So yes, if you care about the best conductance bounds, use $1 < q < 2$. But if you care about something else – as in the Kleinberg Kloumann paper – then $q > 2$ can (and in this case does!) give better performance.

Reviewer 4.3 (Weaknesses). Regarding the note that our paper needs to be compared with more methods. We would like to point out that, in Table 1, we compare SLQ to CRD, ACL, FS, HK, NLD and GCN in both F1 scores and running time. In Figure 4, we had thought to focus this on SLQ vs ACL because ACL/PageRank was the point of the original and this experiment is not about getting better F1 scores or conductance but to show SLQ can also find something different from conductance theory as we explained in the previous answer. But your point is good! We will add more methods here, see the updated Figure 4 at right, where we added another two methods for comparison (CRD, HK).



In Figure 3, we compare SLQ to ACL, CRD and heat kernel because these are the methods that are in some sense similar to ours. ACL is a $q = 2$ special case of SLQ, heat kernel is another type of diffusion method and CRD is an algorithm combining flow and spectral ideas which often performs the best among existing methods in terms of conductance based on our previous experience.

Regarding the performance comparison of SLQ and CRD in Figure 3, the biggest improvement of SLQ is speed and simplicity. In our experiments, SLQ can achieve similar or better performance but running at least 20 to 30 times faster. Also, CRD has a lot of parameters that are not intuitive and often difficult to set, while the parameters of SLQ has the same or very similar meaning to the parameters of ACL/PageRank, which are much easier to set and understand.

Regarding Figure 4, see (3.8 above), our point is that conductance theory doesn't always explain real world performance. The difference is outside of two standard errors.

Reviewer 4.4 (Correctness). The objective we give is well-posed and the algorithm (in Sect 3) will work regardless of the assumptions of Theorem 4.1. Theorem 4.1 is simply a standard type of recovery result that shows a scenario when the algorithm will necessarily be sensitive to a particular and well-known aspect of the property (conductance).

Reviewer 4.11. We would appreciate any more insight you could provide in your review about dimensions where we could have discussed the broader impacts.

Typos. We thank the reviewers for the list of typos that unfortunately escaped our notice.