We thank the reviewers for their nice and helpful comments. We are happy that the reviewers see value and potential in our main claim that spectral modification can circumvent a fundamental weakness of spectral clustering.

Our modification algorithm is intended as a proof of concept of what we view as a general framework. It is by design very fast, and it always outputs a graph that spectrally dominates the input graph while preserving its cuts. It is indeed heuristic in the sense that the output is not provably spectrally close to the maximimizer. However its derivation is not ad-hoc; it is based on theoretical intuition that we briefly discuss in page 7. In fact, we believe that a still-practical variant of the algorithm should be analyzable.

There have been several recent works on graph embedding methods. These are quite successful, but appear to lack any theoretical justification of their empirical performance gains. These methods effectively compute modified linear operators from the input graph via ‘deep walks’ that associate nodes at longer ranges, followed by eigenvector-based embeddings (e.g. see [22]). This is very much in the spirit of our work and we believe that our findings can shed theoretical light to their success. With this work we want to lay the theoretical groundwork towards that direction, which we believe merits a separate treatment, as we explain below.

Comments on the choice of experiments.

• Our experimental examples are well known and appreciated in spectral graph theory. They are understood to capture a fundamental weakness of spectral partitioning, and for that reason we would not consider them as “somewhat arbitrary”. We can report that the unsupervised versions of recent graph embedding methods (e.g. NetMF [22]) **fail** to compute the correct solutions on these examples . Getting the correct output requires a higher dimension in the embedding, **and** a significant amount of supervision, whereas our result uses no supervision; we will add these experiments in the supplementary material and arXiv version. So, up to our knowledge, there is no other graph embedding algorithm that correctly computes a good solution for these difficult instances.

• The **size** of the graphs we use for the visualization had to be small for the visualization to work properly. The same experiments have been repeated for much larger versions of the same graphs (with millions of nodes), with the same outcome. That is why Figures (2) and (3) refer to an **asymptotic** improvement (with respect to n) of the value of the cut. Notably, this improvement is with respect to a precise optimization problem (i.e. conductance), and the obtained solution is optimal. So, with the exception of the image examples, we would not consider these results as “qualitative”. Clearly, demonstrating asymptotic improvements cannot be done via empirical evaluations on fixed-size networks.

• We are aware of the fact that recent graph embedding works provide extensive experimental evaluations. However, in the absence of any theoretical justification, extensive evaluation is necessary to demonstrate their value; we believe that this is not the case with our work. Importantly, these evaluations are empirical, i.e. against given ‘ground truths’ and not with respect to well-defined optimization problems as in our case. They pertain to the **practical** utility of the algorithms, which is a markedly distinct topic relative to the more theoretical orientation of our work.

• We do however believe in the applicability of spectral modification at least for certain classes of problems, and we plan to conduct comparisons with other methods. It should be noted though that these recent works have been successful mostly in the **supervised** setting. Our algorithm has a supervised version –implied by Theorem 3.3, where an appropriately formulated graph B can encode supervision information. This supervised extension is also a distinct task with a host of technical issues that require a separate focus in order to attain the full potential of our approach. In this context, we plan to also study if and to what extent other embedding methods are realizations of spectral modification.

Answer to specific comments by reviewer #2.

• In Theorem 3.3 we wrote “there exists a graph H...” in order make the theorem self-contained, without reference to the preceding definitions. But H is indeed the spectral maximizer; a clarification will be added. Note that Theorem 3.1 essentially states that the maximizer has the maximum possible eigenvalues for the given cuts, but it does not quantify that maximum. With Theorem 3.1 alone, the second eigenvalue can still be lower than the conductance. Theorem 3.3 states that the second eigenvalue is actually within $\widetilde{O}(1)$ from ϕ, which produces a tight gap in the Cheeger inequality. This is in fact explained in lines 144-147, but we realize that a clearer explanation is needed.

• The rigorous justification about the diameter of the tree is given in Lemma 3.3, which states that it is $O(\log n)$.

• **Minor Points:** The quality of the spectral maximizer indeed depends on α and the diameter of the tree. We omit them from the definition for the sake of conceptual simplicity and brevity, as we remark in lines 124-125. We will add symbol ϕ in line 25, right after ‘conductance’. Line 131 appears problematic due to the fact that the two paragraphs in lines 127–129 and 130–134 were unintentionally swapped. This was just meant as a minor technical point that spectral dominance does not require H to be connected. We will fix it. In Theorem 3.2 we will add “for all $S \subseteq V$”, which specifies what S is. We will fix the typo in Definition 3.2. In Definition 3.3 we will add a concrete statement that H is the maximizer. [ ] will be used for references.