We thank all three reviewers for your thorough and thoughtful comments. We have already incorporated them into our revised version of the paper, and are very appreciative of the improvements. Due to space, we group concerns together.

1. Section 3 R1,R2,R3: We significantly restructured this section to clarify the following issues.

1.1. Simultaneous reduction of edges and nodes R1: When comparing with sparsification, we restrict our method to only deletion/reweight. Otherwise, we allow both deletion and contraction, reducing both edges and nodes. We now emphasize that we simply *prioritize* reduction of edges or nodes (as opposed to reducing only one or the other).

1.2. Section 3.1 (now 3.2) R3: Indeed, deleting an edge with $w_e\Omega_e = 1$ would disconnect the graph and invalidate the use of the S-M-W formula, but this is precluded by the requirement that $\mathbb{E}[f] = 0$. We added this to the text.

1.3. $\mathbb{E}[\|\sum \boldsymbol{\Delta L}^\dagger\|_{\mathbf{F}}^2] \approx \sum \mathbb{E}[\|\boldsymbol{\Delta L}^\dagger\|_{\mathbf{F}}^2]$ **approximation** R3: The randomness lies in probabilistic choice of edges. R3,R2: Our "reasonable assumption" comment was indeed imprecise. We removed it, and added an study using ER and real-world networks, showing that our approximation is either nearly exact or a conservative estimate.

1.4. Section 3.4 and choice of cost function R1,R2: This choice of cost function naturally arises when minimizing the expected squared error for a given expected number of reductions. R2: Moreover, as the squared Frobenius error empirically sums, minimization for each edge acted upon can be seen as a probabilistic greedy algorithm for minimizing the cost function of the final reduced graph. R1,R2,R3: A detailed derivation of the single-edge minimization is now in the appendix. In the main text, we better explain the regimes for β. We also fixed the sign error, thank you so much, R3!

1.5. Why unbiased (ie, $\mathbb{E}[\boldsymbol{\Delta L}^\dagger] = \boldsymbol{0}$**)?** R1: This condition is necessary for the answers 1.2–1.4. Additionally, the desire of an unbiased algorithm is quite standard, eg, Spielman et al. reweight the edges so as to ensure $\mathbb{E}[\boldsymbol{\Delta L}] = \boldsymbol{0}$.

1.6. Section 3.5 R1,R2: This choice ensures that contraction yields a node-weighted $\boldsymbol{L}^\dagger$ consistent with the $w_e \to \infty$ limit. We now focus the section on the definitions needed for our algorithm, also including how to convert the coarsened $\boldsymbol{L}$ to a matrix of the original size. In the appendix, we give a detailed rationale for them and explicit numerical examples.

2. Algorithm 2 and solving issues with multiple simultaneous contractions and deletions R1,R2,R3: We removed Algorithm 1 (as it was not used), and incorporated a modification (previously in the appendix) to Algorithm 2 that ameliorates these issues. In particular, we change our edge sampling method from uniform to a random independent edge set. This explicitly solves the issue with multiple contractions, and empirically solves the problem of multiple deletions. R1: We also refined our experimental studies of the hyperparameters and included a detailed discussion.

3. Algorithmic complexity and approximating $\boldsymbol{L}^\dagger$ R2: The S-M-W formula is needed to derive the optimal probabilistic actions to an edge, but it is actually not necessary for the algorithm itself; the more scalable implementation directly constructs approximations to Ω_e and m_e at each iteration, taking $\tilde{\mathcal{O}}(\langle k\rangle|E|)$ time. We add more in-depth analysis and a study suggesting that such an approximation does not notably decrease the quality of reduction (appendix). As stated, our framework cannot start with the empty graph (however, even algorithms that start with an empty graph often need to store the full graph at some point). We now have a discussion section where we clarify these points.

4. Frobenius norm, error bounds, and preferential preservation of global eigenvalues R1: Using a perturbative analysis, we now bound the change in the eigenvalues in terms of our cost function. We also show that the bound on the relative error of the global (top) eigenvalues of $\boldsymbol{L}^\dagger$ is smaller, hence preferentially preserving the portion of the spectrum associated with large-scale structure.

5. Tradeoff between error and reduction and when to stop R1: As mentioned in answer 1.3, our approximation holds well. As it is trivial to keep a running total of this error estimate, one may simply terminate the algorithm when the Frobenius error reaches the desired limit. We included this in our discussion of Algorithm 2 (now 1).

6. Our proposed metric and its relationship with standard measures R3: We slightly modified the definition of our metric (giving additional desirable properties, eg, triangle inequality). We added a proof of its relationship with the standard notion of spectral approximation (appendix). R2: We also reproduce our figures using standard measures.

7. More comparisons with other methods and synthetic graphs R2,R3: We added to all of our coarsening comparisons a recent method by Loukas, which is similar in spirit to our approach, and show that our method more accurately preserves the global structure. R3: Most of these citations were about *using* coarsening for other tasks, as opposed to proposing new coarsening schemes. R2,R3: Indeed, it would be good to compare with several sparsification methods. However, in our literature search, we found that they either do not explicitly preserve properties associated with the Laplacian (and hence, would be an unfair comparison), or are very similar to Spielman's algorithm, with the goal of computational efficiency (at the expense of some accuracy). As speed is not the primary focus of our paper, we compared with the exact implementation. We added this discussion to the text. We are working on a comparison with a method that sequentially combines sparsification and coarsening (the closest we found to our unifying approach). R2: and on experiments comparing our method with other coarsening methods using SBMs. All these will be included if our paper is accepted! For either sparsification or coarsening, if you know about methods that we are unaware of, we would be happy to add them to our comparisons. R1,R2,R3: Please zoom for some new figures: