

1 We thank all three reviewers for their insightful feedback! We appreciate the positive comments regarding our main
 2 theoretical results (R2, R3, R4) and also our application case-study (R2, R3). We have addressed all reviewer comments
 3 in the following and will also incorporate the feedback in our paper. **@R2, restating bound of Theorem 3.1:** Yes.
 4 For instance, setting $\epsilon = \min\{\epsilon_1, \epsilon_2\}$, the bound can be further upper-bounded by $(1 + 2\sqrt{2}k)/(8\sqrt{k} - 1)$ exactly,
 5 provided $k \geq 5$. However, we think that doing so does not demonstrate the full potential of that theorem. In particular,
 6 we interpret ϵ as the desired recovery error level, and therefore not having ϵ in the RHS limits the error bound only
 7 in terms of problem parameter k as already noted by the reviewer. **@R2, reference for the statement on line 119:**
 8 Please see Theorem 2.2 in Mao et al. [2020]. We shall also include this reference in our submitted paper. **@R2, matrix**
 9 **multiplication in step 5 of algorithm 2 in $O(n^2)$ time:** The matrix product in step 5 is $R - \frac{\mathbf{p}_{s'} \mathbf{p}_{s'}^T R}{\|\mathbf{p}_{s'}\|^2}$. Computing
 10 $\mathbf{p}_{s'} \mathbf{p}_{s'}^T R$ takes $O(n^2)$ time provided we first compute the matrix-vector product $\mathbf{p}_{s'}^T R$ and then right-multiply the
 11 resulting vector to $\mathbf{p}_{s'}$ (which is a vector outer product). Then computing $\|\mathbf{p}_{s'}\|^2$ and $\frac{\mathbf{p}_{s'} \mathbf{p}_{s'}^T R}{\|\mathbf{p}_{s'}\|^2}$ take $O(n)$ and $O(n^2)$
 12 time respectively. Lastly, subtracting the matrix obtained in the previous sentence from R also takes $O(n^2)$ time since
 13 we perform exactly n^2 subtractions. **@R2, mildness of assumption $c_{\min}/c_{\max} > 1/2$ in Theorem 3.4:** Corollary
 14 B.2 in the supplemental shows that this assumption is satisfied by a vast margin with very high probability when all
 15 k Dirichlet parameters are equal. In fact, an analogous result can also be shown using the same techniques when
 16 some Dirichlet parameters are not equal. We highlight that this is an improvement to the recovery guarantee of
 17 competing GeoNMF which works only in the equal Dirichlet parameter setting. **@R2, broader impact:** We shall add
 18 more discussion on the broader impact. **@R3, comparison with modularity models for protein complex detection:**
 19 Indeed modularity-based methods may potentially be applied to protein complex detection. However, these methods are
 20 mainly heuristics and they suffer from the lack of recovery guarantees. Our main objective in this work is to advance
 21 a general-purpose mathematical theory for overlapping community detection. **@R3, LP in linear time claim (line**
 22 **134-135):** The LP effectively requires \mathbf{x} to be in the column range of P . However, since $P = \Theta B \Theta^T$ (according to
 23 MMSB), it has rank k with high probability. Consequently, the column range of P is same as the column range of Q
 24 where Q is an $n \times k$ matrix whose columns denote the k eigenvectors of P corresponding to its nonzero eigenvalues.
 25 This implies that the LP in step 3 of algorithm 1 is equivalent to $\{\min \mathbf{e}^T(Q\mathbf{y}) \text{ subject to } Q\mathbf{y} \geq 0, (Q\mathbf{y})_{\mathcal{J}(i)} \geq 1\}$
 26 which contains $n + 1$ constraints and whose dimension is k . Hence the Megiddo result implies that such LP can be
 27 solved in $O(n)$ time. Moreover, computing the k largest eigenvectors is a step that appears in many algorithms in
 28 literature (including the competing GeoNMF) and can be achieved in $O(n^2)$ time using randomized SVD techniques
 29 (Halko et al. [2011]). We shall include these clarifying remarks and references in even more detail in our submitted paper.
 30 **@R4, using different real-world network, and poor performance against state-of-the-art algorithms:** To our best
 31 knowledge, there does not exist a real-world dataset of weighted graphs for which an independently derived ground-truth
 32 validation set also exists. If the R4 can point us to one, we would be very happy to use it in our future works. Regarding
 33 the performance against other methods, as mentioned in the paper, protein complex detection is a well-studied problem
 34 in computational biology for which many tailor-made heuristics exist. However, these heuristics do not come with any
 35 theoretical guarantees and as such the focus of our work is to advance a general-purpose mathematical theory which is
 36 not restricted to only protein complex detection. It may well be possible to tweak our algorithm so that better results
 37 for protein complex detection are obtained, but that is not the scope of this work. **@R4, time complexity $O(n^2)$ of**
 38 **presented algorithm is prohibitive, and largest synthetic network has only 5000 nodes:** We would like to clarify
 39 some factual misunderstandings: our largest synthetic network has 12000 nodes, and also, to our best knowledge, there
 40 does not exist a competing algorithm whose time complexity is under $O(n^2)$. We shall update our paper so that these
 41 aspects are more clear. Indeed in practice our algorithm runs slower compared to competing GeoNMF, and we mention
 42 improving this as a possible future direction. This increase in running time is traded-off with improved recovery
 43 guarantees; our recovery algorithm is effectively parameter-free, does not require pure nodes, allows for different
 44 community sizes and any general full-rank community interaction matrix B , and provides smaller recovery error.
 45 **@R4, adding more experiments since algorithms behave differently on different datasets, and more discussion**
 46 **of algorithm and results:** We shall update the paper as suggested to include a discussion of the algorithm and the
 47 results! While we agree that algorithms sometimes perform differently on different datasets, analyzing such patterns
 48 entails a much more comprehensive experimental comparison across diverse datasets and algorithms. Such a study
 49 is beyond the scope of this work since our main contribution is on the theoretical side. Therefore we compare our
 50 algorithm with a highly-cited domain-specific algorithm (ClusterONE). However this is a useful future work direction.

51 References

- 52 Xueyu Mao, Purnamrita Sarkar, and Deepayan Chakrabarti. Estimating mixed memberships with sharp eigenvector
 53 deviations. *Journal of the American Statistical Association*, 0(just-accepted):1–24, 2020.
- 54 Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms
 55 for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.