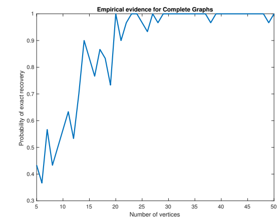
We thank the reviewers for their comments and feedback. For all reviewers, we plan to add some experiments to show empirical correctness of our results. Fig. 1 is a plot for complete graphs, where for each number of vertices, 30 runs were computed in order to estimate the probability of success. The ground truth y^* was generated uniformly at random and we used $p = 0.2$ and $q = 0.2$. Standard CVX code was used for SDP.



R1: • We appreciate the reference of B. Mohar (1989), regarding tightness, our bound is tight up to a factor of $1/2$. This can be achieved by considering a signed Laplacian matrix $\mathbf{M}$ with first eigenvector $y = \mathbf{1}$ and first eigenvalue 0, then we obtain a bound for the typical Laplacian matrix, whose known tight lower bound for λ_2 is $\frac{\phi_{\mathcal{G}}^2}{2\Delta_{\max}}$. See for instance section 3 in F. R. K. Chung (1996), "Laplacians of graphs and Cheeger inequalities" (where $\Delta_{\max}$ does not appear in their bound due to their different definition of Laplacian matrix, which is already weighted by $1/\Delta_{\max}$). Comparing to Theorem 4.2 in B. Mohar (1989), "Isoperimetric numbers of graphs", their bound is tighter but has a dependency on λ_2, i.e., the lower bound is $\frac{\phi_{\mathcal{G}}^2+\lambda_2^2}{2\Delta_{\max}}$, thus, while interesting, the proof requires a lower bound independent of λ_2. We propose to add these brief comments in Section 5.

• For better presentation we plan to replace the start of Section 3 with these lines: "Our approach consists of two stages, similar in spirit to (Globerson et al., 2015). We first use only the quadratic term from (2), which will give us two possible solutions with high probability, as stated in Theorem 2. Then as a second stage, the linear term is used to decide with high probability the best between these two solutions, as stated in Theorem 3. Combining both theorems, we find the sufficient graph properties to achieve exact recovery."

R2: • We think it is unfair to say that the model is toyish and makes results less interesting. While the model is simple, it deserves attention as it provides further theoretical understanding to structure prediction. Additionally, Globerson et al. (2015) and Foster et al. (2018) use the same model and have been published in venues such as ICML and AISTATS, conferences that are similar in spirit to NeurIPS.

• Reviewer 1 also suggested to improve the start of Section 3. Please see bullet 2 in answers to R1.

• We appreciate the new reference (Heidari et al., 2019). The generalization from binary to multiclass, under our approach, remains as an open question for us. We believe it deserves careful thinking and might even be a conference paper on its own. The references of Weller et al. (2016), and Meshi et al. (2016), while not very close to our work, will be added to have a more complete overview of results in approximate inference.

• Line 167: Yes, we will fix this typo, it should say βw_i^+.

• Line 175: In this case we want to know the probability of $i \in \mathcal{S}$ and $j \in \mathcal{S}^C$, or **viceversa** (note this), for $(i, j) \in \mathcal{E}$. Then, by construction of $\mathcal{S}$, for this to happen, one w^+ should be greater than t and the other less than t. Since t is a uniform random variable in $[0, 1]$, the probability is just the difference of the squared values, i.e., $(w_i^+)^2 - (w_j^+)^2$ (see line 173-174). Then, the absolute value makes sure that if $(w_j^+)^2 > (w_i^+)^2$ then the difference is positive. This would count the case in which $i \in \mathcal{S}^C$ and $j \in \mathcal{S}$.

• Expansion in graphs is related to algebraic connectivity. The Cheeger constant is one notion of expansion, where a low Cheeger constant means that the graph has poor algebraic connectivity. One example of graphs that can be considered bad expanders are grid graphs (as stated in lines 284-285). We will add this notion of expansion before using it in order to make our arguments more clear.

R3: • The model proposed by R3 would indeed be simpler resulting in no need for SDP or other sophisticated methods/algorithms. This is because one would only have the noisy information from the nodes, and if edges are derived from these noisy nodes then the edge values are still consistent with the noisy nodes. Thus, edges provide no further information in R3's suggested model. The hard part of our setting is the inconsistency between noisy edges and noisy nodes, creating the need for solving a complex combinatorial problem.

• We mention in Section 2, lines 89-101, as well as in the Theorem statements that the graphs are undirected. Therefore, the node degrees are non-oriented.

• It is hard to talk about $\Delta_{\max}$ or $\phi_{\mathcal{G}}$ in isolation and might not be the way to look at the bounds. Thus we focus on family of graphs and provide examples in Section 4. We also mention cases where the guarantee does not work, for example, for grid graphs (lines 284-285). This is the reason why we discuss smoothed complexity.

• We appreciate the reference suggestion "Statistical physics of inference: Thresholds and algorithms" by Zdeborova and Krzakala 2018, it contains a comprehensive review of inference problems and we plan to add it in our paper.