We would like to thank the reviewers for their insightful and helpful comments. We address some specific points raised below, and would like to thank the reviewers in advance for considering our responses.

Reviewer $\#3$

Thank you for your comments.

Reviewer $\#4$

To my opinion, the optimality and the tightness of the results should be discussed - We fully agree, a discussion was initially omitted due to space constraints. First note, in terms of the prior literature for graph labeling without switching lower bounds were proved in references [6,7]. Lower bounds for switching in the experts model are given in [1,21]. These are for the "log" and "mix" loss respectively, these are both unbounded loss functions compared to "mistake" counting. For these unbounded loss functions there is a $\Omega(\Phi \log(T/\Phi))$ term in their lower bound but as we speculate in line 328 in our model the dependence on T may be an artifact. A sketch of a lower bound on a line graph is provided below if desired this can be summarised into a theorem and included in the appendix if suggested by the reviewers.

Observe that if we have a single graph-labeling problem on an n-vertex line graph with a cut size of 1 it is not difficult to force $O(\log n)$ mistakes; likewise if we have a uniformly labeled line graph hence no cuts and a single cut is introduced we can force $O(\log n)$ mistakes. On the other hand if we have a line graph with a cut size Φ we can force $\Phi/4$ mistakes by "removing" $\Phi/2$ cuts. Now for a switching sequence of graph labeling problems, $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T$, let $\Phi(\boldsymbol{\mu}_t) \ll n$ for all t. For a labeling $\boldsymbol{\mu}$ observe that we can divide the line graph into $\Phi(\boldsymbol{\mu}) + 1$ segments, of length $\frac{n-1}{\Phi(\boldsymbol{\mu})+1}$, where each segment can be made independent of one another by fixing the boundary vertices between segments. We therefore have $\Phi(\boldsymbol{\mu}) + 1$ independent learning problems and can force $\Theta(\log \frac{n}{\Phi(\boldsymbol{\mu})})$ mistakes for every cut introduced and 1 mistake whenever we remove 2 cuts. Note beside the $\log \log T$ dependence there still remains some gap between this sketched lower bound and Corollary 7.

Might seem as incremental combinations of graph prediction and online learning - We believe that a logarithmic-time adaptive online algorithm for graph prediction is a significant improvement over the state of the art. Furthermore the bound in Theorem 2 is a new result for switching in the specialists setting with the asymmetric J() appearing in the bound.

How was α tuned in the experiments? - Lines 286/622: α was tuned using exhaustive search over the ranges given in the appendices over 24 hours of training data, rather than optimally from the bound.

Adaptive vs. oblivious adversary - All formally numbered theorems, corollaries and lemmas as stated are correct with an **adaptive** adversary. However, if as discussed in lines 152-157, one would like to convert the *deterministic* mistake bounds with respect to the *cut* to *expected* mistake bounds with respect to the *resistance weighted cut* then it is sufficient for that conversion to assume an **oblivious** adversary. We can add this additional comment to lines 152-157.

Would it be possible to get high-probability bounds? We suspect the technique in lines 152-157 could be adapted to a high probability mistake bound.

268: the sum should start at 1 - Thank you for pointing that out.

Could be interesting to plot the upper bound on the experiments - We are happy to plot the computed upper bounds for the given data on the experiments, thank you for the suggestion.

Measure of robustness (error bars/standard deviations) - Error bars were initially omitted for 'neatness' in the plot. The standard deviations given in Table 2 in the appendix do imply an increasing robustness with larger ensemble sizes. We are happy to include a discussion on this and include error bars in the plot.

More explanations on Alg. 1 with brief sentences defining A_t, Y_t. The notations A_t, Y_t are defined within the Algorithm, at their first appearance. Resistance distance (effective resistance) is defined lines 53-54, we can also supply the equation,

$$r(i,j) = \frac{1}{\min_{\boldsymbol{u} \in \Re^n} \sum_{(p,q)\in E} (u_p - u_q)^2 : u_i - u_j = 1} .$$

Reviewer $\#5$

Thank you for your comments. In particular thank you for the additional reference we will add to line 239.