We would like to thank the five reviewers for their feedback. Upon acceptance, we will include in the final version (a) *improved comparison with prior work*, and (c) *missing references*. We first discuss a few common concerns shared by **Reviewer 1**, **Reviewer 5**, and **Reviewer 6**.

NUMERICAL EXPERIMENTS:

Reviewer 1: We thank the reviewer for the comments/remarks on our paper.

COMPARISON WITH [CHEN ET. AL, 2020]:

BIAS OF v, THE ESTIMATE OF THE SECOND ORDER MOMENT:

Reviewer 5: We thank the reviewer for valuable comments. We add the following:

COMPARISON WITH [CHEN ET. AL, 2019] AND [ZHOU, DONGRUO, ET AL., 2018]:

Reviewer 6: We thank you for the valuable comments on our submission. We are revising our paper and will update as soon as it is done. Following is our answer to your questions.

EXPLANATIONS ON THE ASSUMPTIONS: As rightly mentioned by the reviewer, the stepsize is in order $\alpha_t = 1/\sqrt{T}$. The dependence in d leads to a small learning rate in the presence of large networks but our theorem states that the rate would then be as fast as we present it. Hence, the bound in our Theorem prevails over the intuition that the convergence will be slow due to a small learning rate.

discussion on matrix W when number of nodes is large

Reviewer 8: We thank the reviewer for his/her interest in our paper. Below we address your concerns about our contribution.

DISCUSSION ON THE MATRIX W: