We thank the reviewers for their thoughtful comments and will make the suggested changes to clarify figures and steps of the derivation.

Expanded empirical validation We will use most of the extra page in the revised version to expand the section on empirical results and in our plots will now show network dynamics as well as spectra.

We have empirically characterized noise-prune's performance on non-symmetric clustered networks (i.e., going beyond the theory) and find it works very well. The figure included below shows sample plots, comparing noise-prune to a control that prunes only based on weight. These empirical results thus extend beyond the theory and show that the guiding principle works in the right direction. We will include an expanded set of these results in the manuscript.

The current state of the art in pruning in machine learning prunes while preserving a task-dependent cost function and either uses multiple steps of training and pruning or complex non-local measures of synaptic importance. We will add a more detailed discussion of these algorithms. We have not yet characterized noise-prune's performance against these algorithms (we intend to in future work). But we would be surprised if noise-prune performed better given that it is so simple, uses only local information, and seeks to preserve all dynamics rather than performance on a specific task. To avoid any confusion, we will explicitly state that we do not expect noise-prune to perform better than these algorithms.

We believe our work is nevertheless of broad interest to researchers studying many types of neural networks and not just to researchers interested in unsupervised learning, dynamical systems and computational neuroscience. As the reviewers point out, we provide a novel perspective and strong theoretical results on the problem of pruning. These may contribute to future algorithms. Moreover, we provide a bridge between pruning and graph sparsification, noise-driven dynamical systems, and matrix concentration of measure techniques, which are all active areas of research with many more possibly fertile ideas for pruning.

We plan to follow this study (focused on introducing and developing the theory) with an extensive empirical characterization of noise-prune's performance on nonlinear and non-symmetric networks, but believe (and hope that the reviewers agree) that the inclusion of some numerical results on non-symmetric networks (in addition to expanded results on symmetric networks) in the current study will adequately support and point beyond the theory.

Biological applicability (R4): We agree that the computational role of synaptic pruning is likely not captured solely by preservation of dynamics. However, we would argue that a good synaptic pruning rule should at least preserve the broad pattern of dynamics. Building off of the Reviewer's language learning example, even here the dynamical patterns are likely to be similar between unpruned and pruned networks, because they need to carry out a similar set of input-output transformations, even if the pruned network is faster and more reliable. Thus dynamics preservation could be used as a building block for more complex pruning algorithms. For example, note in the model the non-pruned synapses are also strengthened. If the model included multiplicative fluctuations in synapse strengths (as suggested by spine head size fluctuations), then non-pruned strengthened synapses would be more reliable, making dynamics more precise and reproducible. Or, adding nonlinearities to downstream neurons could make these synapses faster or more efficient at driving downstream activity. We will now include some text on other benefits of sparsity in the Discussion.

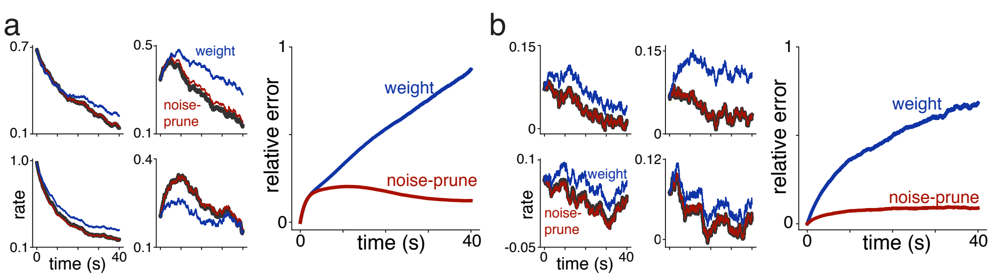

Figure 1: **Noise-prune on non-symmetric clustered networks**. Networks have dense within-cluster and sparse between-cluster connections. Black traces are original unpruned network; red traces are networks pruned to 20% sparsity with noise-prune; blue traces are pruned to 20% based only on weights. (a) Response to a random input at time 0. Left: response of 4 different network nodes, showing activity over time. Noise-prune (red) and the original network (black) are very close, with black sometimes covered by red. Right: Difference between network activity vectors before and after pruning. Red trace shows $||x_{orig}(t) - x_{np}(t)||_2/||x_{orig}(t)||_2$, where $x_{orig}(t)$ is the original network's activity vector in response to a random input and x_{np} is the equivalent for the noise-prune network. Blue shows equivalent curve for weight-based pruning. (b) Results in (a) were for a random input. This panel shows results for an input along one of the slow eigenmodes of the network. Such slow eigenmodes are thought to be important for maintaining information over time in biological networks. Noise-prune performs significantly better than pruning by weights.