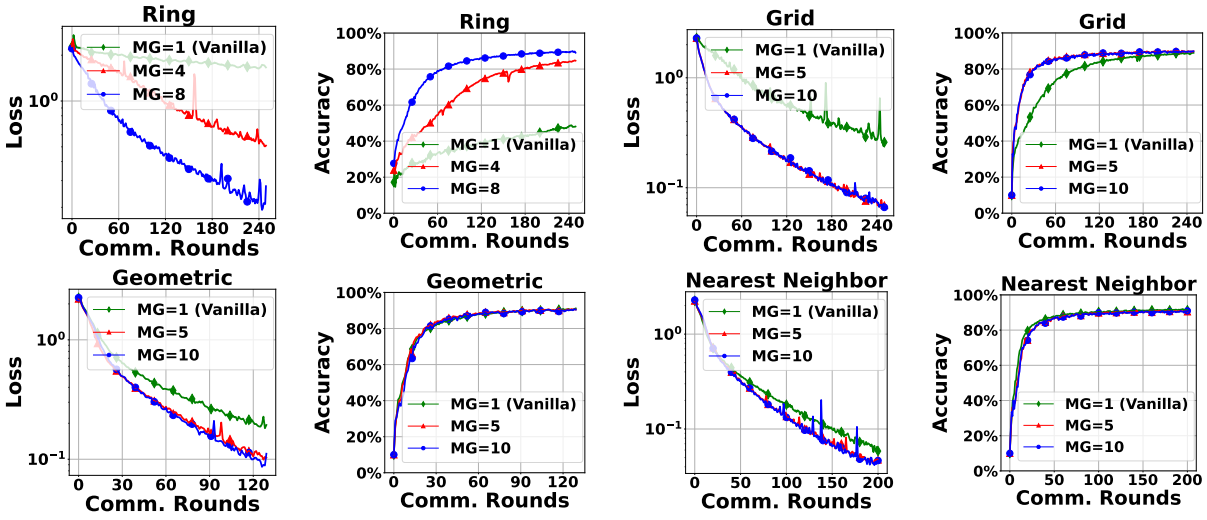

Figure 1: Training loss and test accuracy of ResNet-18 trained using MG-PullDiag-GT and vanilla PullDiag-GT on the CIFAR-10 dataset. MG-PullDiag-GT consistently outperforms the baseline across all network topologies.

MNIST MG high kappa

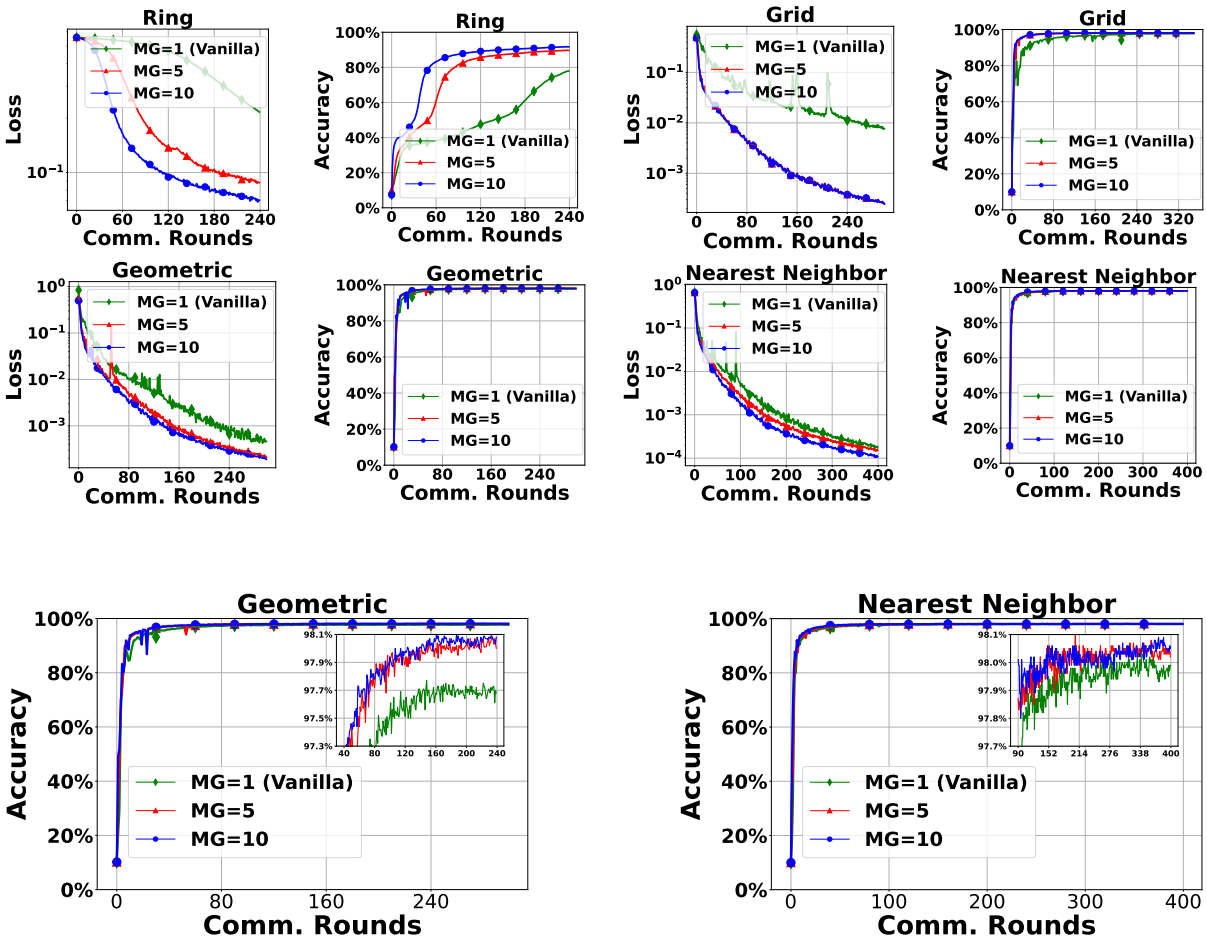

Figure 2: Training loss and test accuracy of a 4-layer neural network trained using MG-PullDiag-GT and vanilla PullDiag-GT on the MNIST dataset. The data is distributed in a heterogeneous manner (see Figure 3).

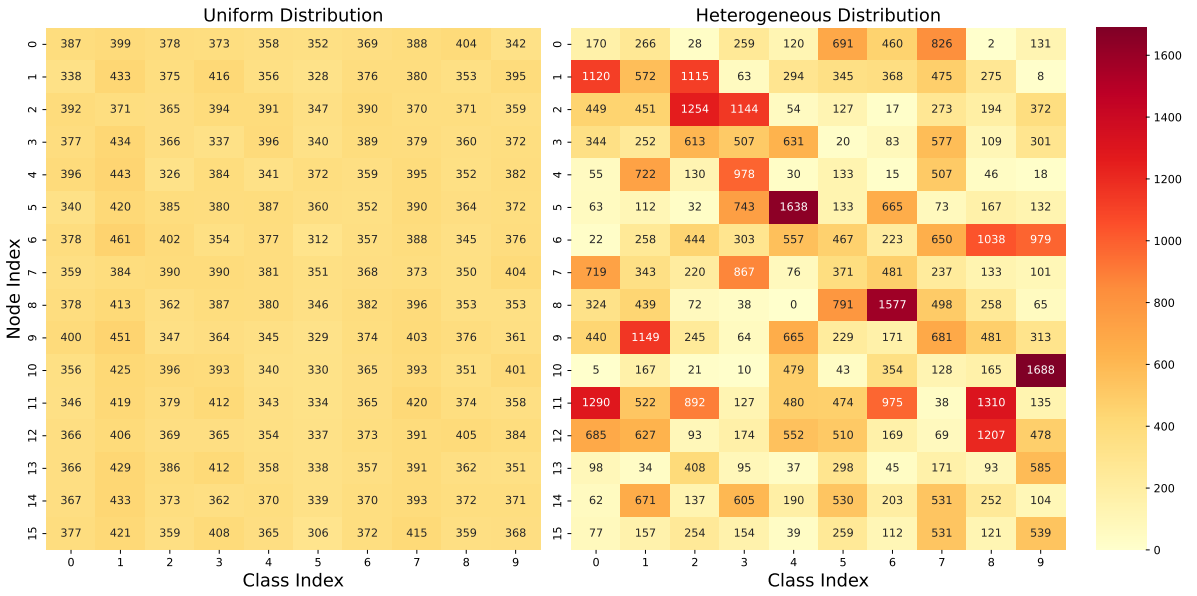

Figure 3: Left: Uniform MNIST data distribution used in the experiments in Section 6. Right: Heterogeneous MNIST data distribution used in the additional experiment shown in Figure 2.