We can see from the results that several models performed well above 90% in average DDI accuracy. The paper reported that the 9-layer model with a hidden dimension size of 2,048 performed the best, however we saw that the 4-layer model with a hidden dimension size of 2,048 performed best. We can also see that all models with a hidden dimension size of 2,048 achieved 90%+ mean accuracy. Therefore, it is likely that even with our small subset of the full dataset that we can accurately predict unseen DDIs.

For the ablations, we first modified the complexity of the model by scaling the number of layers and hidden dimension size. We can see the trend that with the paper's 50 principal component dimensionality reduction and holding hidden dimension size constant, there is a negative correlation between the number of layers and performance. It is possible that with our reduced dataset that the model is overfitting, and the shallower models have better generalization.

Additionally, there is a clear positive correlation between model performance and the hidden dimension size, however intermediate hidden dimension sizes between 128 and 2048 were not tested.

Finally, we can see with our ablation study determining the impact of the PCA dimensionality reduction of the molecular fingerprint, there is no clear trend. Holding the number of layers (2) and the hidden dimension size (128) constant, we can there may be a negative correlation between the PCA dimension size and performance, however this contradicts the original paper's findings.

Discussion

The paper was fairly difficult to reproduce without the refrence to ChemicalX [2]. The amount of compute given the dataset size also made it difficult to reproduce at the same scale. The original paper [1] does include a link to their repository containing "source code," however it does not contain any code for training the models nor does it contain any model weights. Moreover, the dataset was fairly intensive for preprocessing, and the paper does not provide a clean version of their data for immediate use. That being said, the model itself is very simple being a standard feed-forward network, so model implementation was straightforward.

Overall, the findings and takeaways from the study still hold true in that predicting unseen drug-drug interactions is possible using molecular fingerprints based on structural similarity.

I would recommend to the authors that the model definition and trianing code be included in the repository, and the data used with SMILES strings be released as well as opposed to only listing the Drug Bank IDs.

References

1. Ryu, J. Y., Kim, H. U., & Lee, S. Y. (2018). Deep learning improves prediction of drug–drug and drug–food interactions. Proceedings of the national academy of sciences, 115(18), E4304-E4311.

2. Rozemberczki, B., Hoyt, C. T., Gogleva, A., Grabowski, P., Karis, K., Lamov, A., ... & Gyori, B. M. (2022, August). Chemicalx: A deep learning library for drug pair scoring. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 3819-