Paper 3 [Nov 7 2022]

**What was the paper doing?**

This paper developed a deep learning based GNN architecture called DeepFRI to predict the locations within proteins, by also leveraging the use of a LSTM based language model to outperform current (2021 a time of publishing) protein function predictions.

**What new thing did I learn from the paper?**

I had learnt abundant information about GO terms and EC count, which are used in classifying different protein subsections into different categories. Moreover, I learnt about how by combining the results of another model (of a completely different type), with the base data a more powerful model with far higher accuracy can be created. More specifically, how the utilisation of an LSTM upon the protein sequence to create an embedding combined with a graphical representation of the protein can further help improve the accuracy in isolating the significance of certain parts of the protein.

**Was there anything they could improve on?**

There was one single thing that I thought they could try to improve on (it took me a while to find it). I found that within the Graph neural network part of their model they hadn’t tested the use of a Graph Injective Network (GIN), and had only used GraphSage, GAT, and ChebConv. I thought that GIN may have been more useful here considering that the shape of a protein sequence typically has great effect upon its function, which would have been better identified by a GIN. Although, this is only a speculation as I am no means an expert within the field of proteins, and we would need to actually test this theory empirically to test to what extent my point holds.

**What applications are there for this paper that I can think of?**

The application of this paper is quite clear. It is immeasurably helpful in the research and development of better medication of viruses and diseases. By gaining a better understanding of the functions of proteins, we may develop proteins of better functions and those that prove more benefit to humanity as a whole.

**Does this paper share ideas/similarities to other concepts/papers I have considered before?**

This paper shares an ever so slight similarity to the first paper I read about COVID-19 case predictions within the US. In the sense that they utilise a graph neural network to isolate and propagate information. However, this argument can technically be made for all applications of Graph Neural Networks.

[Below are my personal notations whilst reading the paper, and contains more information that I could not provide whilst answering my preset questions I have given myself]

[Protein\_Folding.pdf](https://drive.google.com/file/d/1i58H3cdO-boUZqn9IPC0IlcmdSXLHnkj/view?usp=sharing)