October 25 - Nov 8 Summary

- To detect drug-target interaction pairs, we could potentially use a force-directed graph; i.e., initialize the graph randomly, and for every known interacting pair, decrease the repulsive forces, and for every known non-interacting pair, increase the repulsive forces. After this training phase, you could potentially look at nearby drug-target pairs. This probably will not work, but it might be worth trying.

- The BioBERT code still does not work with the authors' recommendations. There are other people who have raised essentially the same issue on their GitHub.

- I've started looking at other language models as alternatives, and found [FLAIR](#), which has a cleaner API (and works). The NER with FLAIR is pretty useless in a medical context (detects nothing), so we could

  - use POS tagging with this model, pick only nouns, and form pairs, discarding terms that are not in some dictionary of known medical terms + colloquialisms.
  - fine-tune this on medical data. There seems to be an attempt at doing this, called [BioFLAIR](#), and I'm currently trying to put together the two pieces (API use for fine-tuning + how the BioFLAIR repository code works). I will also try emailing the authors to see if they have trained models ready to use.

  At present, though, we could use this crappy model, and finish working on the pipeline, and later replace this part.

- An early draft of the documentation of all the work and code done so far is ready.