For the creation of the input dataset we used:

1. 25,438 positive PPIs from iRefindex

We applied the same filters we used in the curation of the positive training dataset. The exception was that we included all the PPI detection methods recorded in iRefindex, with the exception of 'yeast to hybrid' and 'tandem affinity purification', which were the filters used for the creation of the positive training dataset. We thus reached 25,439 PPIs in comparison to the 1,678 positive PPIs of the training ds.

The Positive PPIs are indicated with '1' in the 'PPI\_type' column

2. 25,432 negative PPIs from 'Russel's Negative Dataset'

First we filtered the original Russel's Negative Dataset, in order to rule out the negative interactions from the training dataset. The original dataset was reduced from 894,213 PPIs to 890,857 PPIs (exactly the 3,356 PPIs used in the training ds).

Secondly, we randomly selected from 'test\_dataset\_negative' 25,438 interactions to match the number of positive ones. Then we removed potential duplicates that exist in a) the positive test dataset, and b) in the positive training dataset. Indeed, we found 6 duplicates that were, then, removed, reducing the negative PPIs to a total of 25,432.