Node Expansion Effects

This is based on the analysis performed in:

CommunityDetection/Node-Expansion/node\_expansion.py

Figures can be found in the same directory as the script.

Conclusion:

Compared to the total number of known gene – HPO connections found in Jenkins, MyGene2 contains on average half the number of connections (MyGene2 average = 19.94, Jenkins average = 38.55 HPOs per gene). After expanding genes to include their interacting neighbors in String DB averages increase to 80.25 in MyGene2 and 560.77 in Jenkins. From this it is evident that MyGene2 contains only a small portion of the known gene-HPO connections.

Based on the results of Figure 5 and Figure 6 it is clear that most of the genes interacting with genes in MyGene2 and Jenkins have no connection to HPOs. We understand how genes interact with each other better than how genes effect phenotypes. This lack of understanding could be exaggerated due to double (or more) counting of neighbors connected to multiple genes in MyGene2 and Jenkins and being repeatedly counted as having no connection. But the double counting is a two way street and could also occurs in the number of neighbors that are found.

Results:

A picture containing screenshot

Description automatically generated

Figure 1 (hpos-per-gene-boxplot.png): Box Plots of the number of HPO terms found per gene in MyGene2 and Jenkins (The Gene – HPOs known connection network) and the number of HPOs per Gene in those two networks when the gene of interest is expanded to include all of it’s neighboring genes in String DB.

A screenshot of a cell phone

Description automatically generated

Figure 2 (hpos-per-gene-histogram.png): Contains the same information as Figure 1, note the loss of information.

A screenshot of a video game

Description automatically generated

Figure 3 (log-hpos-per-gene-boxplot.png): Same information as Figures 1-4 but using a log scale on the y-axis.

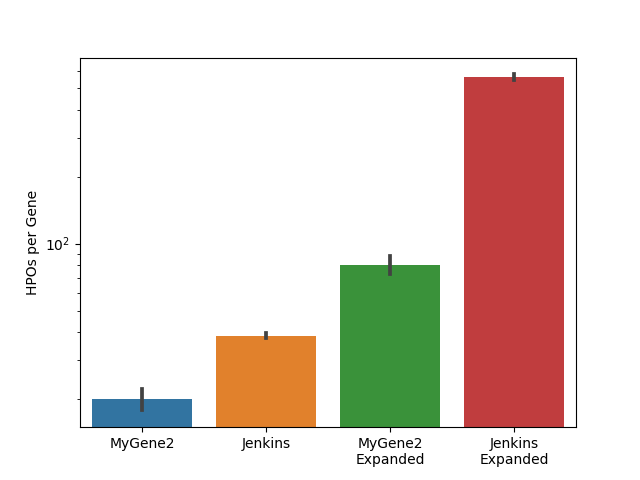


Figure 4 (log-hpos-per-gene-histogram.png): Same information as in Figures 1-4 but using a log scaled y-axis.

A screenshot of a cell phone

Description automatically generated

Figure 5 (genes-not-found-in-StringDB-histogram.png): Depicted is the number of genes from MyGene2 and Jenkins that are or are not found in String DB (genes that are lost and not represented in Figure 1-4)

A screenshot of a cell phone

Description automatically generated

Figure 6 (neighbors-not-found-histogram.png): For a given gene in MyGene2 and Jenkins, this the number of that gene’s neighbors in StringDB that are not present in MyGene2 and Jenkins.