Graph creation algorithm

Data processing

1. MSstats stuff?

Extract INDRA network

1. Determine Gene sets/pathways with high correlation and coverage
2. Pull statements out of INDRA for genes in step 1
   1. Only increase or decrease events
   2. Filter based on
      1. Organism
      2. Disease
      3. Cell type

Input Indra dataframe of statements **stmts**, experiment data **df**

1. Filter Indra dataframe for evidence\_count > 1
2. Create network DiGraph
3. For edge in range(len(stmts)):

G.add(stmts[“out”], stmts[“in”])

1. Get cycles: cycles = [i for I in nx.simplecycles(G)]
2. For c in cycles:

If len(c) == 2:

If stmt[c[0]][“evidence\_count”] > stmt[c[1]][“evidence\_count”]:

G.remove(c[0], c[1])

Else if stmt[c[1]][“evidence\_count”] > stmt[c[0]][“evidence\_count”]:

G.remove(c[1], c[0])

Else:

Directionality (i.e. does indra say up or down?)

Else:

Loop over edges and drop them if they have lowest evidence count

If equal drop based on correlation and directionality

1. Look at edges with evidence == 1
   1. If edge shows high correlation (maybe like some percentile?) and edge does not create cycle and edge matches correlation directionality:
      1. Add edge

Fit causal model