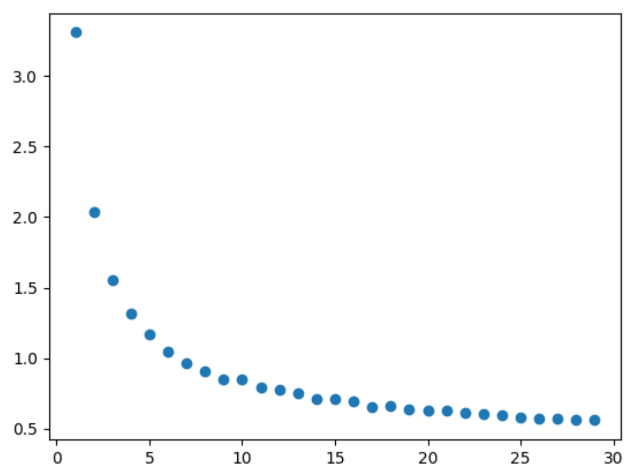
CS 461 Homework 3

Ben Garcia

December 2022

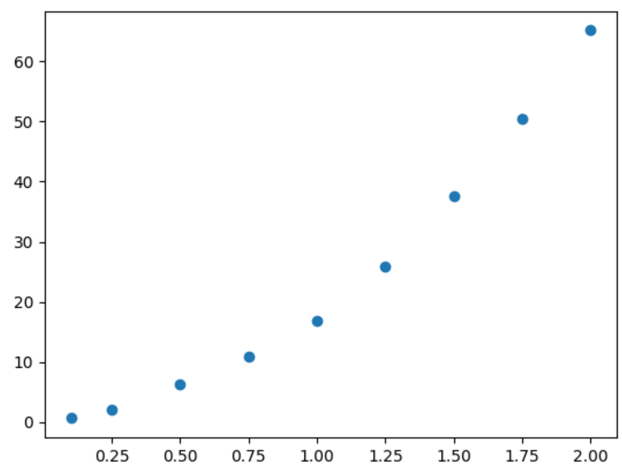
1 Autoencoder

Plot results as function of hidden layer size. What does it suggest about the dimensionality of the data?



The graph bellow shows the loss of the auto encoder as the number of hidden nodes increases from 1 to 30. The loss decreases with each increase in hidden layer size, but the decrease in loss starts to diminish rapidly after a few increases occur. Thus, the dimensionality of the data is small relative to it's total number of variables. This is suggested by the fact that even when the data is compressed into a hidden state with a small size it can still model the input data relatively well.

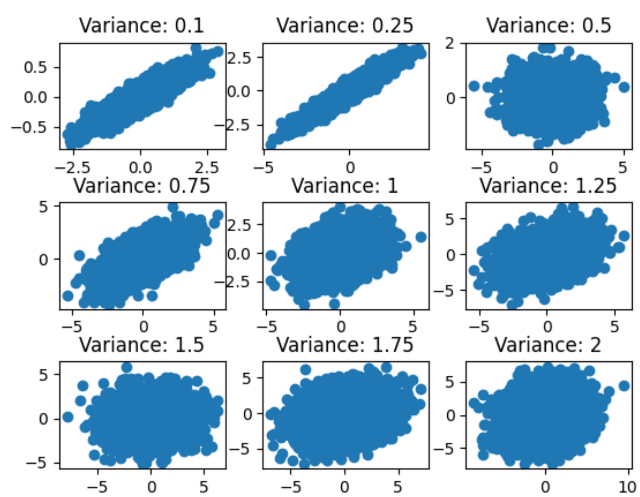
How does it change with variance?



As the variance of the data increases the loss for an auto-encoder with $k = 15$ increases. This makes sense because as variance in the data increases it's harder for the auto-encode to detect meaningful relationships between the variables.

2 PCA

What does the result suggest about the dimensionality of the data? How does it change with variance?

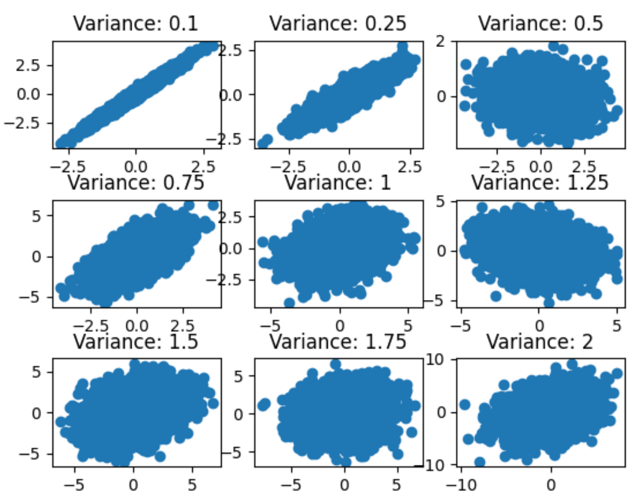


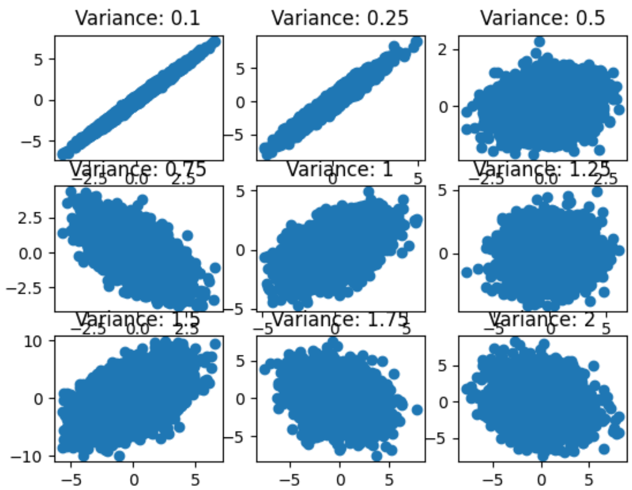
The plot above shows the 2 largest principal components plotted for data generated with different variances. The results of computing the principal components suggests that the dimensionality of the data is less than that of number of variables originally in the data. We can see this from the plots with low variance.

Notice, $X^T X$ is a symmetric matrix and should have orthogonal eigenvectors. The amount of correlation seen in the plots with low variance implies this principal components or eigenvectors are not orthogonal. This implies the dimensionality of the data is lower than the number of variables since orthogonal eigenvectors could not be found. This also matches with how the data was generated since all columns are essentially X_1 + Noise.

As the variance increases we can see the correlation disappear. Notice that this happens around when variance is greater than 1. This also makes sense since X_1 is sampled from a normal distribution with variance 1. The noises added to the other variables starts to drown out their relation to X_1 when the noise distributions have higher variance. This makes the data more random and more varied, thus less linearly dependent and harder to represent in few principal components.

How Robust is it?





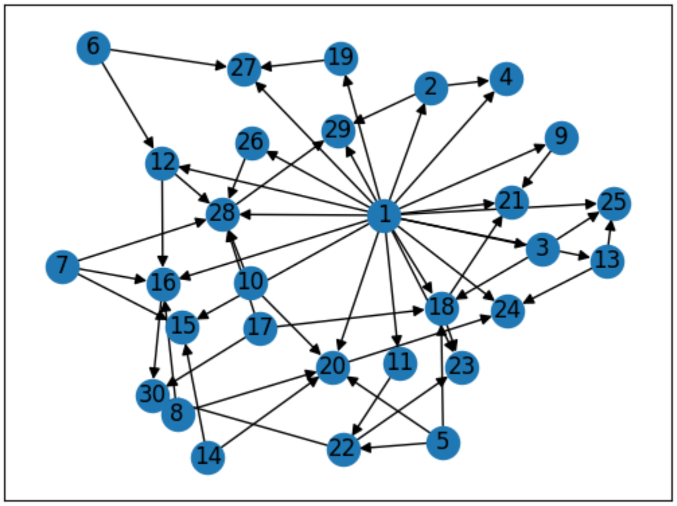
Similar results are acquired when PCA and this analysis are run on newly generated data.

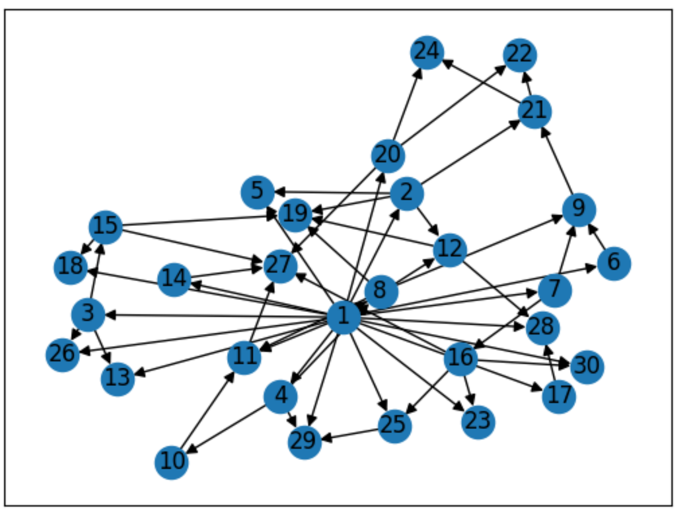
3 Correlation Graph

Build a graph with F=2, F=3, F= 4. Does this graph reproduce the 'true' dependency structure of this data? Is this robust (do you get the same resulting it again on a different data set)?

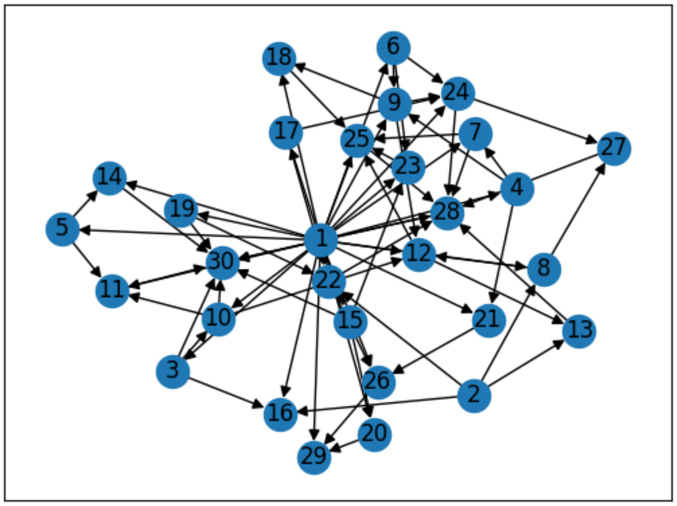
Bellow are graphs with the 2, 3, and 4 edges per variable showing their highest correlation. An edge pointing towards an variable represents a high correlation with the source variable. The results do not appear to be robust. Each generated graph shows different dependencies even with the same initial parameters. Furthermore, when analyzing the total number of accurate dependency relation the results are also not robust. All F=2,3,4 show near an average of 30% of the correct relations. This also showed high variance, individual graphs across all F values varied from 20% to 50%. This does not seem to reproduce the true relations.

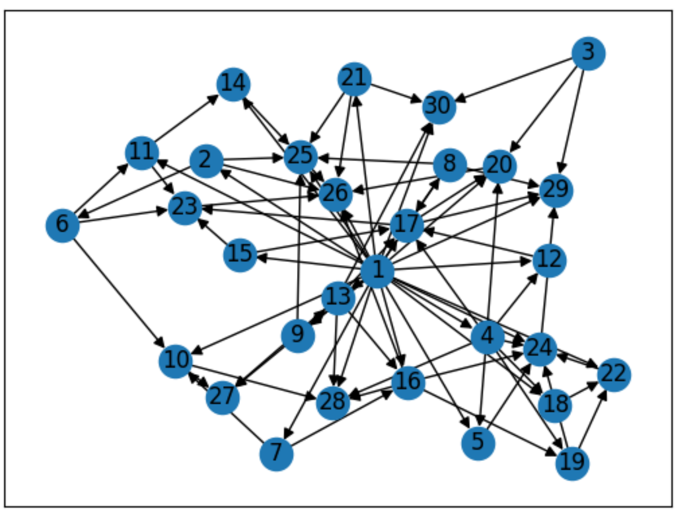
F=2


6
19
27
2
4
29
9
26
12
28
1
21
25
7
3
13
16
10
18
24
15
17
11
23
20
30
8
5
22
14

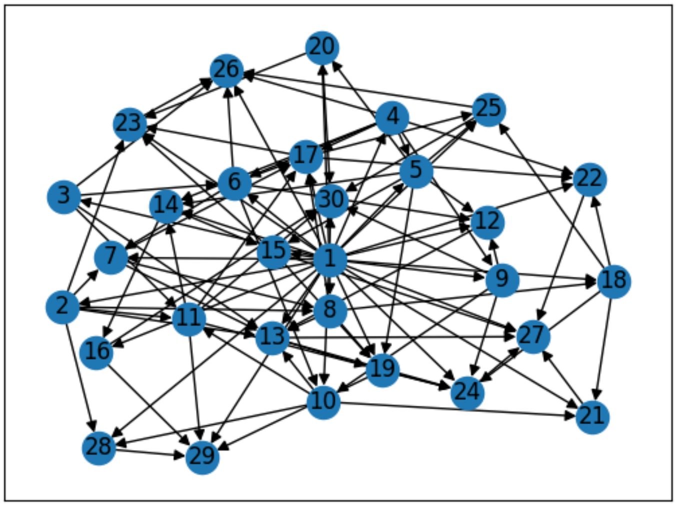

24
22
21
20
5
2
9
19
15
12
6
18
14
27
8
7
3
1
28
26
11
16
30
13
4
17
25
23
29
10

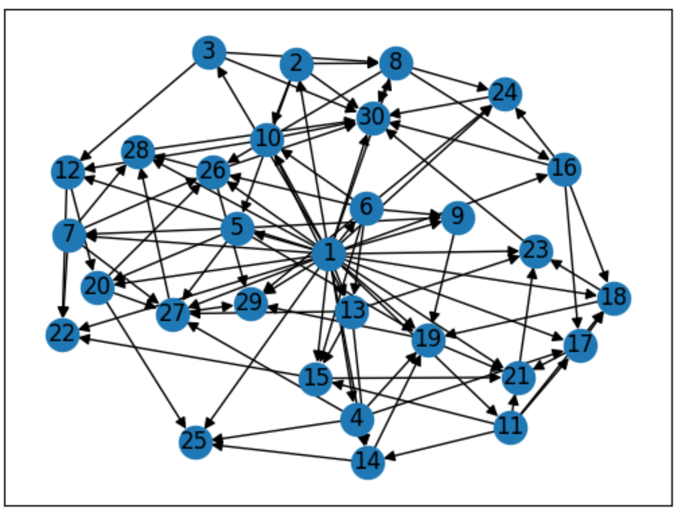
F=3





F=4

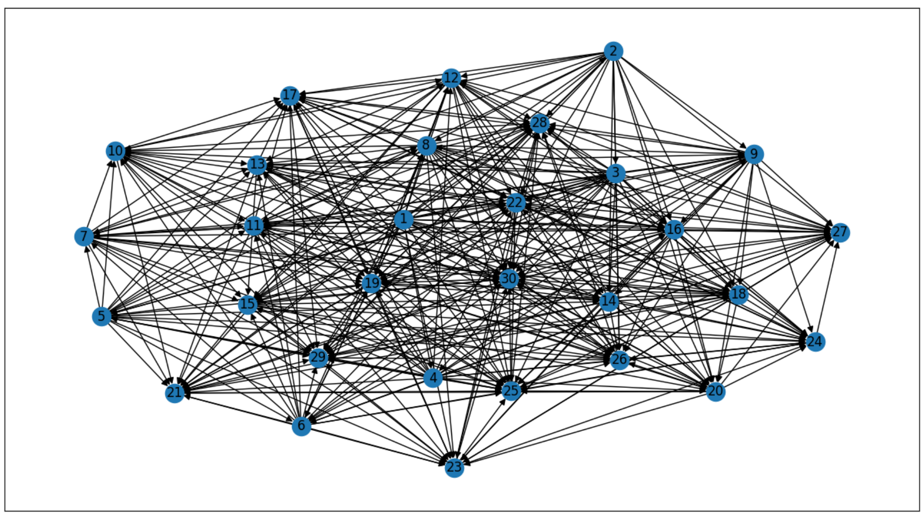


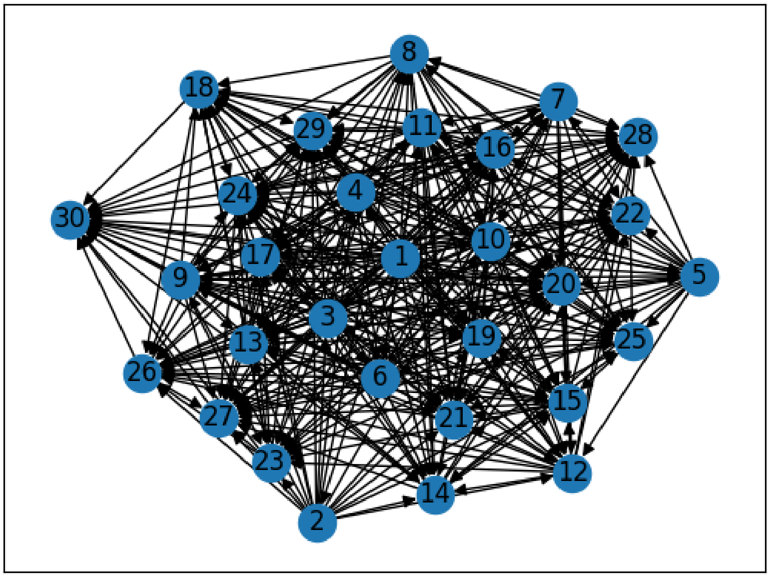


Consider building this graph by connecting each feature to the other features that have a weight in the model larger than some threshold. How does the resulting graph depend on the threshold taken? Are you able to reconstruct the true dependency graph?

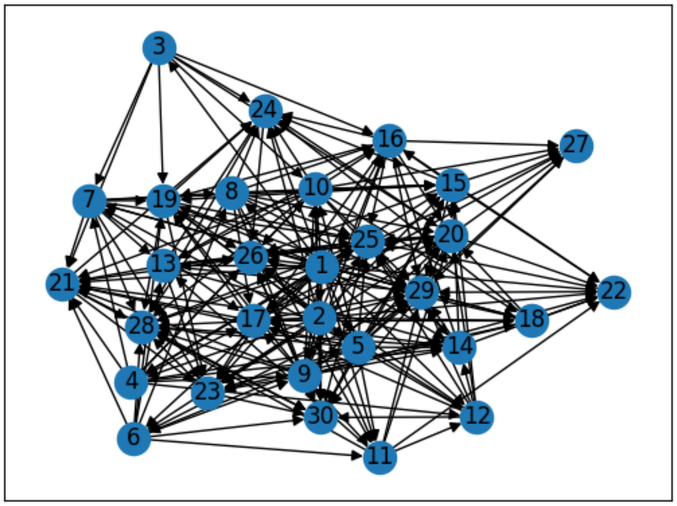
Here are a couple graphs and their average % of true dependency relations generated under different thresholds. They also do not seem to build the true relationships, but a low threshold gets relatively close. It seems like low F values or high thresholds are not likely to find the real dependency, but it's likely that the true dependency has a high correlation relative to other variables. Thus, a lower threshold increases the chance the true dependency is included.

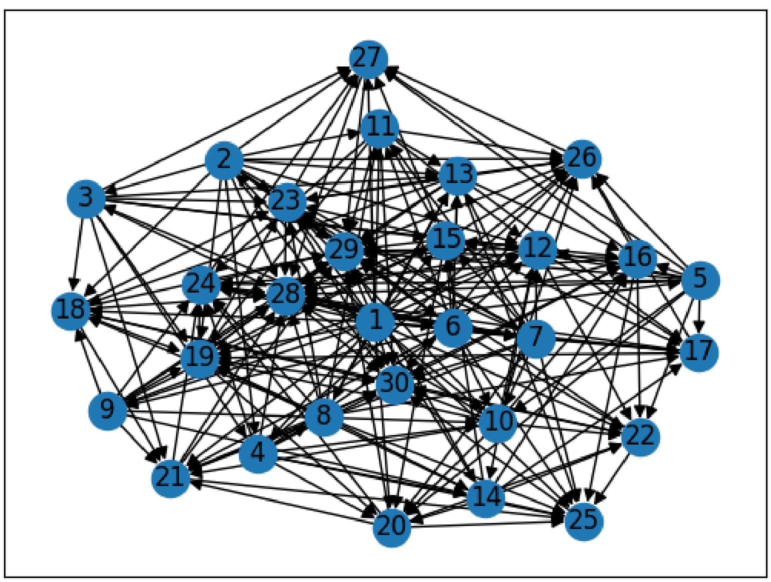
Threshold = .033, Average % true dependencies = 53.1%:



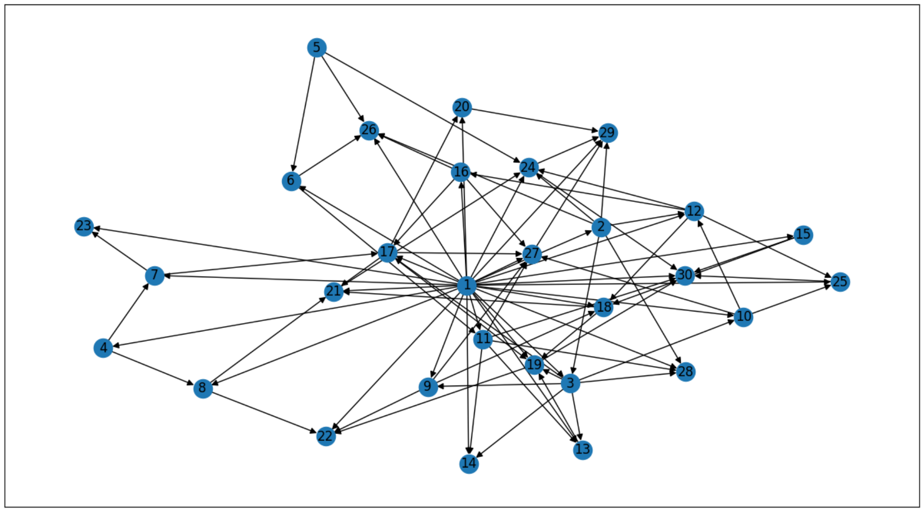


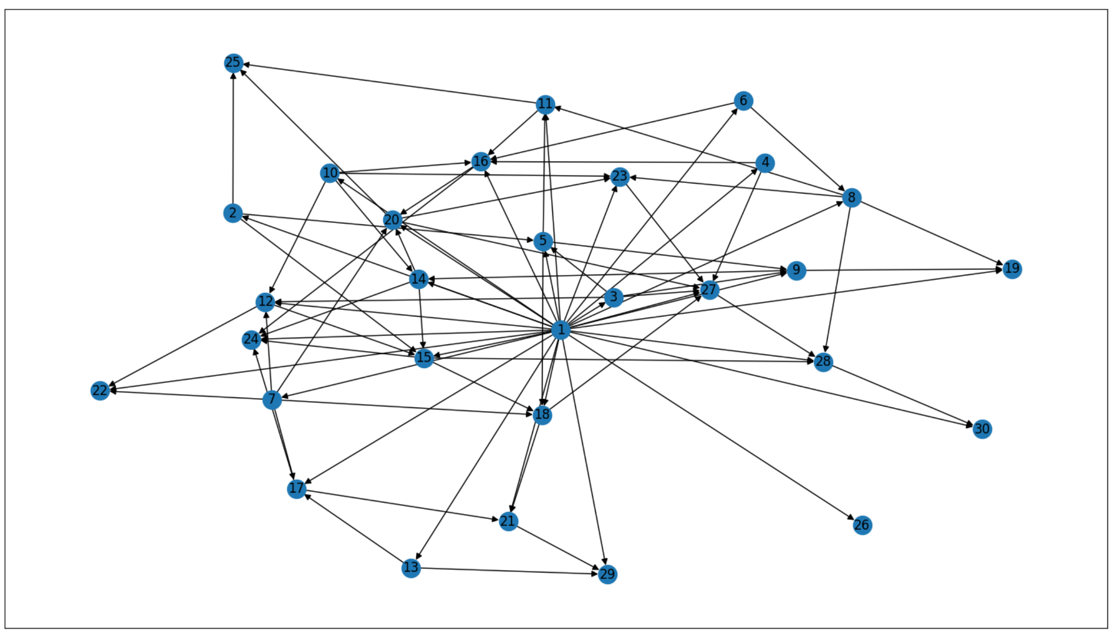
Threshold = .036, Average % true dependencies = 45.8%:





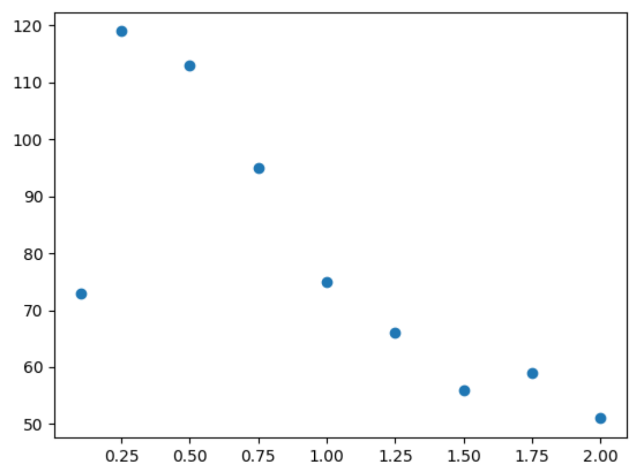
Threshold = .04, Average % true dependencies = 37.5%:



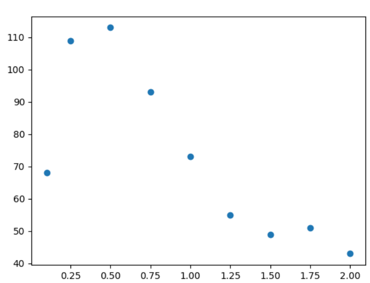


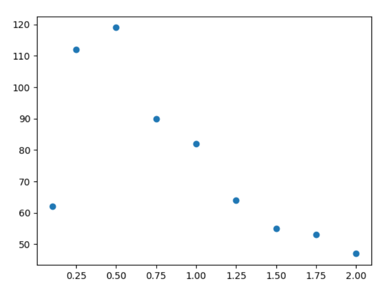
As the threshold increases the number of connections, and percent of true dependencies appear in the constructed graph.

How does the result change as $\sigma 2$ changes, between 0 and 2?



This plot shows the number of dependency relationships in the constructed graph that actually represent true relationships, over the variance on the x-axis. As the variance increases the number of true dependency connection in the correlation graph maximises around $\sigma^2 = .25 - .75$. This relationship is robust and can be seen when this is run multiple times:





I assume this is because a slightly higher variance makes it less likely to add an inaccurate dependency relationship in the graph. As multiple rounds of noise are added to successive variables become skewed away from other previous variables. A variance value that's extremely low may not skewed these variables enough, resulting in the graph be constructed with dependency relationships that "skip over" the actually dependencies. When the variance is two high it also becomes hard to find the correct dependencies because each successive round of added noise drowns out the true values generated from previous variables. For example, with low variance, close to or at 0, all the variables show dependencies with x_1 instead of X_i they are defined from.