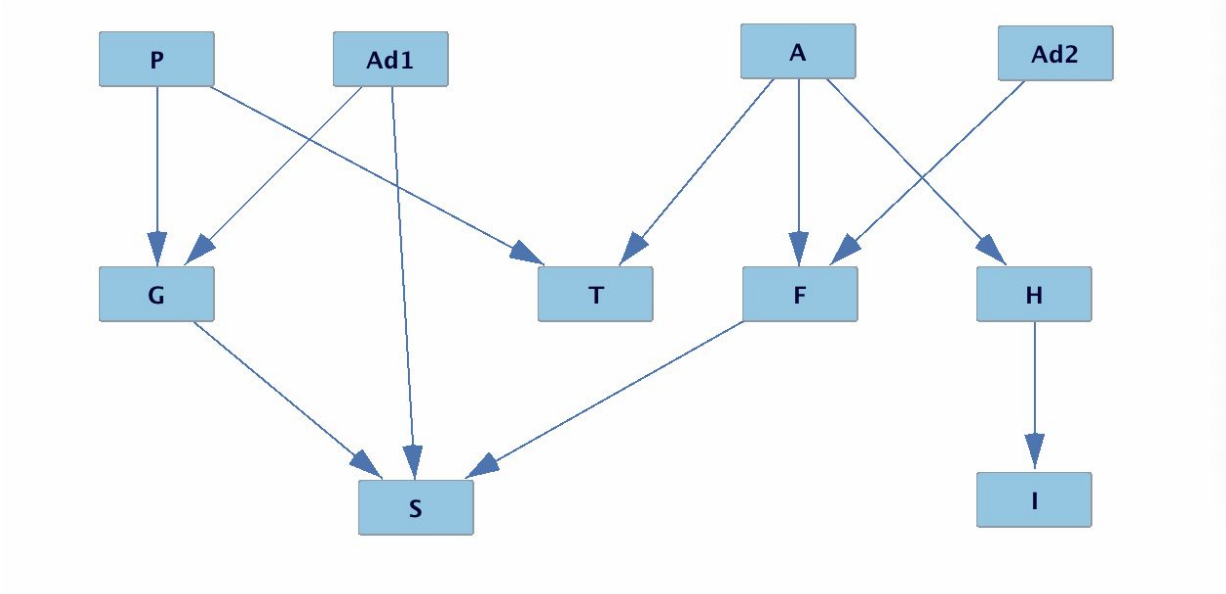
Maya Pegler-Gordon
Shanaya Nagendran
Lauren Lindsey

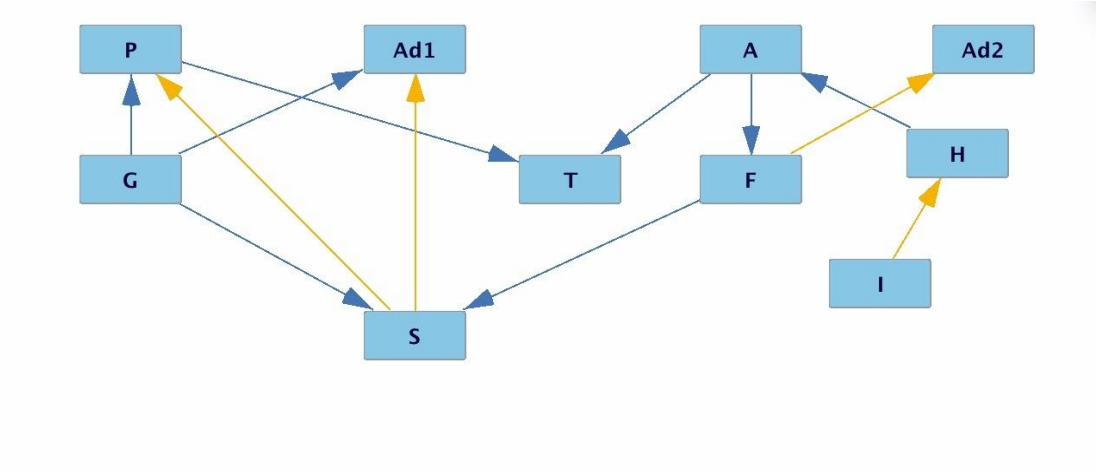
# Problem 1: Report

1. Take a screenshot of a topologically sorted arrangement of the nodes in the resulting network structure for your report (seeing nodes in their parent-child topologies is useful from a causal perspective).




2. Without having to look up the PC Structure Learning algorithm, examine and then explain why the system was able to orient the direction of some edges, but not others.
    a. We have a few theories as to how it is able to determine the DAG. One theory is that the order that the variables are listed might influence how the DAG is drawn. For example, Ad 1, Ad 2, P, and A are the first four variables listed, and they happen to be the "top" nodes in the topological picture. Another theory is some version of a BN sampling strategy, but reversed. We learned in class that you can get a list of samples by going through the BN, but perhaps the computer is able to get a BN by going through the existing samples. It is able to be accurate because the law of large numbers says that data becomes more accurate and the number of samples approaches infinity.
3. Repeat Step 3 (in the same Tetrad Session / Document, but using a new Pipeline) above using only the first 10,000 samples in the dataset, then answer the following: Is your model the same or not? Explain why it is the same or different.
    a. The graph is not the same since the dataset is smaller, so there is more uncertainty when drawing relationships between nodes. Many of the arrows pointed the opposite direction of the original dataset. In our original dataset, the cause and

effect relationship between home ownership and possession of home/renter's insurance looks like: H → I. This makes logical sense since whether or not someone owns a home would indicate if they have renter's insurance. In the smaller dataset, this relationship is reversed which does not make logical sense.


P
Ad1
A
Ad2
G
T
F
H
I
S