ASSIGNMENT 5

Chunyao Wang(chw132),
Wenxing Li(wel85),
Jie Rong(jir18)

Contents

1 Introduction

In this project we use the software GeNIe to verify the consequence and conclusions in the article Application of the TETRAD II Program to the Study of Student Retention (Marek J. Druzdzel and Clark Glymour, 1994). First, we analyze tendencies and trends. Then we use PC algorithm to generate learning networks of the data. Our aim is comparing our result with Druzdzel & Glymour's conclusions (causes of freshmen retention) and analyzing the result in our own view.

Questions we want to answer in this project:

- Do the 1993 data support Druzdzel & Glymour's conclusions? If not, why not.
- Can you find anything else going on in the data?
- What are the causal graphs suggested by GeNIe?
- What causes student retention?

Data description:

- spend: the average spending per student
- strat: student teacher ratio
- salar: faculty salary
- rejr: rejection rate
- pacc: percentage of who accept the university's offer
- tstsc: average test score.
- top10: class standing
- apret: The average freshmen retention rate

First we import data into GeNIe.

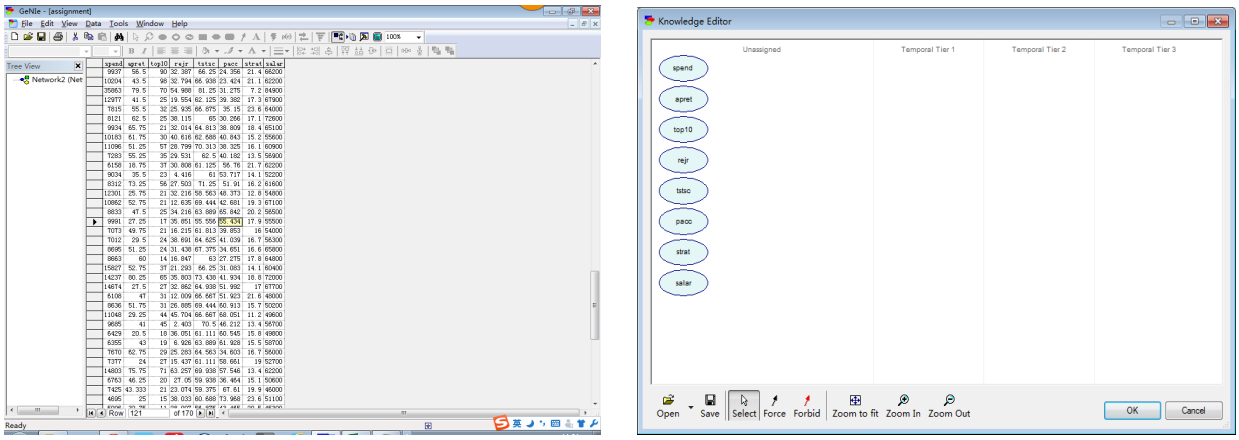


Fig1.1 import steps

2 Causal Discoveries –Tendencies and Trends

2.1 apret vs. tstsc

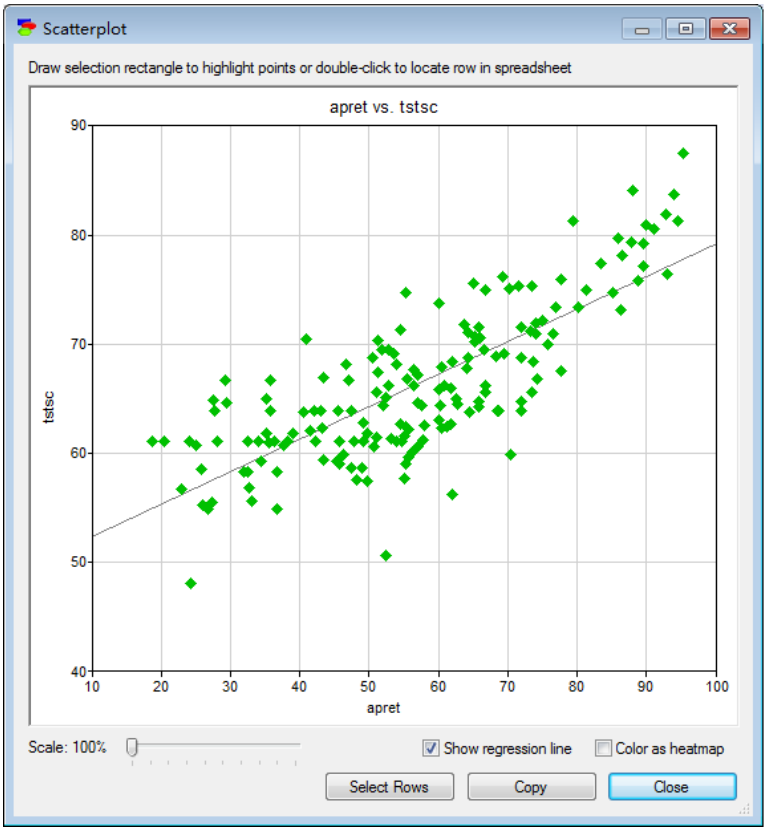


Fig2.1 apret – tstsc

According to Fig2.1, there is linear relationship between the average retention rate and the average test scores of incoming freshmen, which indicates there is a direct or positive association between them. As it can be seen, however, this relationship is not very strong. We can figure out that students doing well in college have higher possibility to also do well in high school. In contrast, students struggling in college are more likely to do poorly in high school.

2.2 salar vs. apret

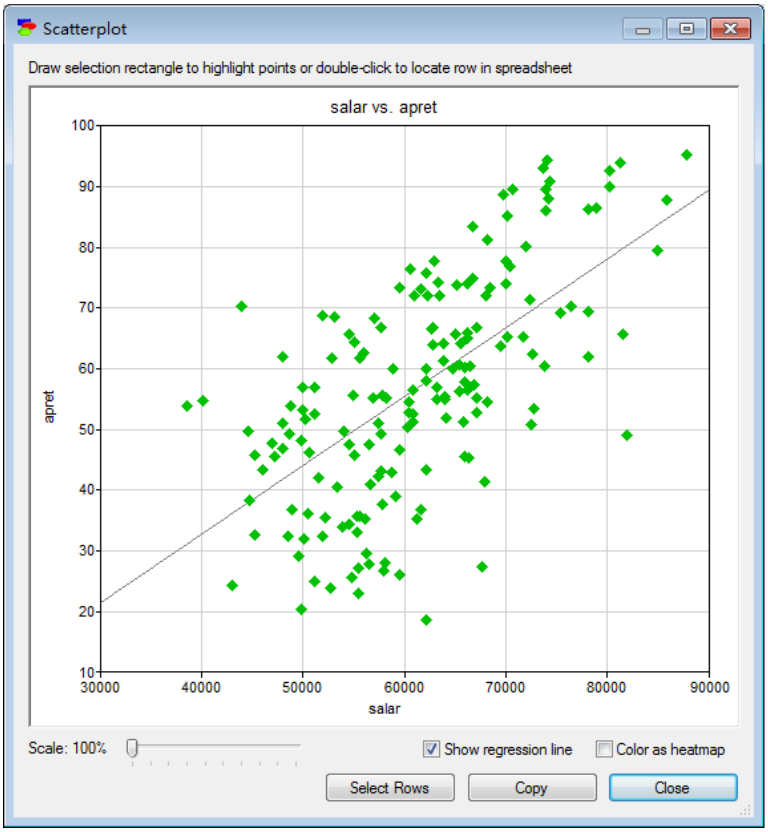


Fig2.2 salar - apret

According to Fig2.2, there is linear relationship between the average retention rate and the average faculty. Compared to Fig2.1, this relationship is weaker than that between the average retention rate and the average test scores of incoming freshmen. As it can be seen, there is a direct or positive association between both variables. As a result, conclusions can be made that

the higher income the faculty have, the more likely their students will become successful and do a good job.

2.3 spend vs. apret

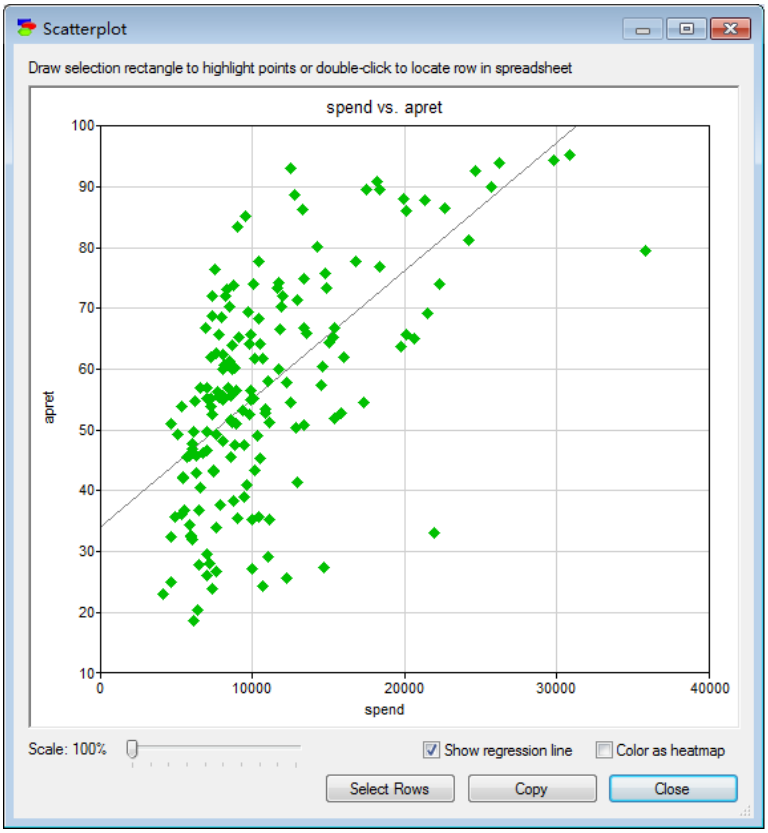


Fig2.3 spend - apret

According to Fig2.3, there is slightly linear relationship between the average retention rate and spending per student. What is more, there is a direct or positive association between them. Therefore, spending more money on student contributes to students' better performing in college. As it can be seen, most of the universities are in the range of 500-1500 for spending and receives high retention rate while keeping the cost relatively low. Besides it is apparent that the retention rate is 79.5% in the college that spends the most per student while the retention rate has been able to reach 80% in five universities which do not spend more than 5 thousand dollars for each student.

2.4 tstsc

According to Fig2.4, central tendency indicates that average value and the exact middle value of the set are almost identical. The sample variance and standard deviation show the difference of the salary among schools. Mode, 48 000 dollars per year, is more closed to the minimum than maximum.

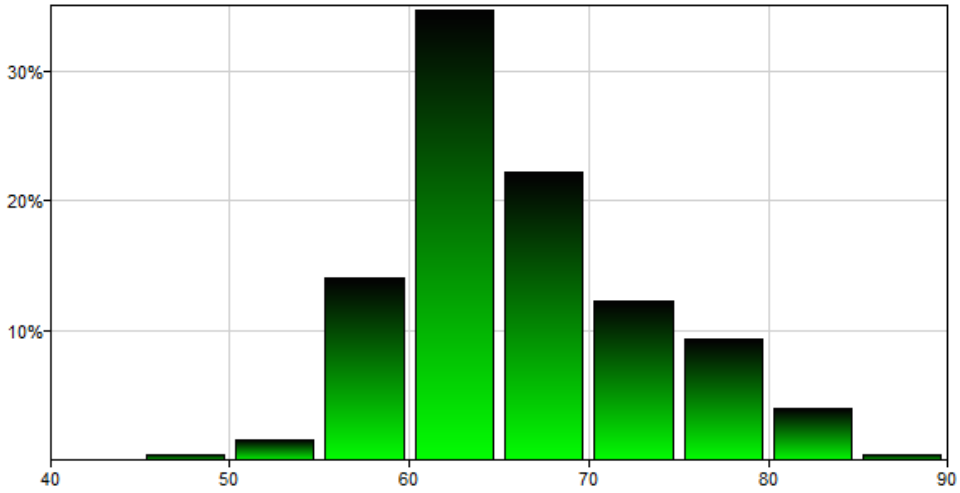
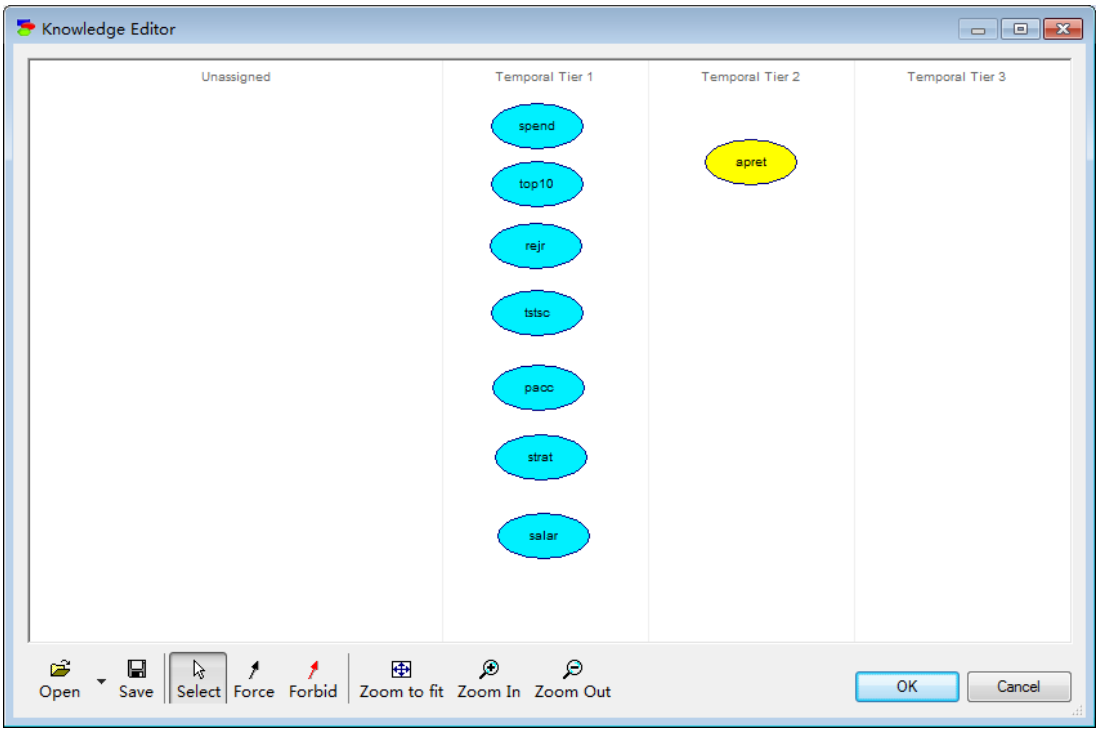


Fig2.4 tstsc

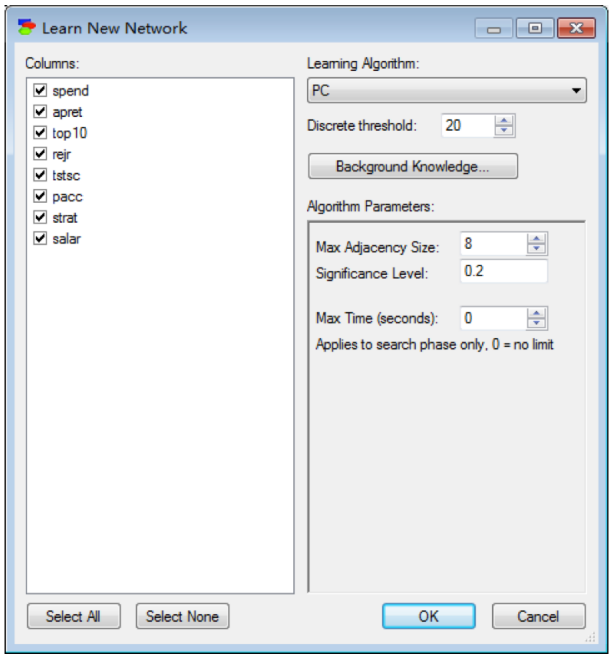
3 Causal discovery

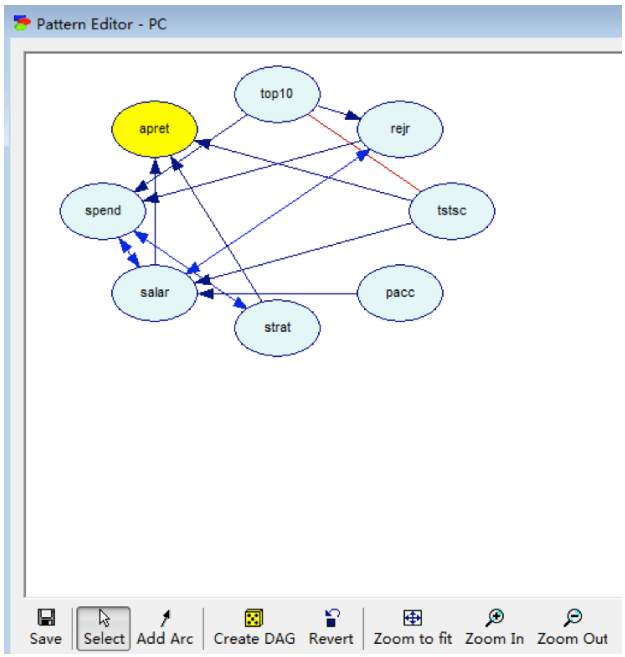
As preparation, we first imported data assignment6.txt into GeNie and then selected learn new network→PC algorithm. Because we need to compare the results with 1992, we choose the same significance levels as which in Druzdzel & Glymour's paper. They are: p = 0.2, 0.15, 0.1, 0.05, 0.01, and 0.001. Since we need to figure out the causes of low freshmen retention, we set background knowledge to separate possible causes and result (freshmen retention) to different temporal tier.



3.1 Set significance level p=0.2

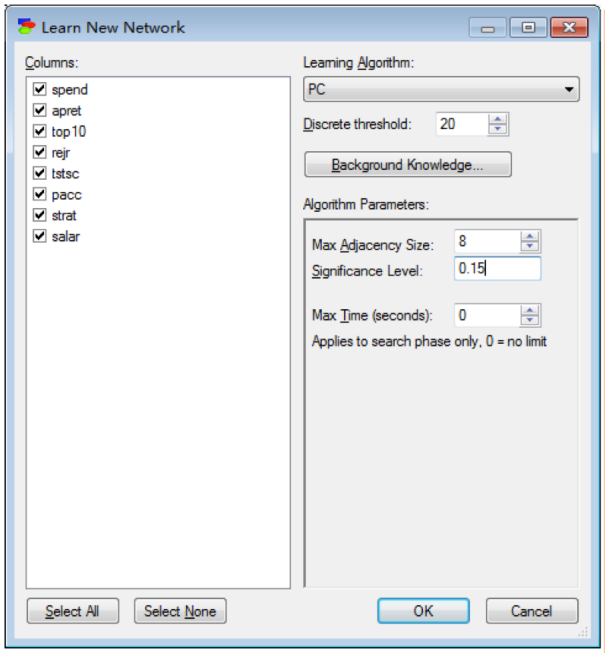
As Druzdzel & Glymour's explain in their paper, a single headed arrow means there is a direct causal relation between two variables and a double headed means there is a latent common cause between two variables. We can see there are three directly causes of freshmen retention: faculty salary, student teacher ratio and average test score. Our comprehension is: high faculty salary means the university has good financial status. And good financial status might give students better environment and facility support to their study. Thus, the are rarely leave the university. High student teacher ratio might bring students more social activity， and student will be happy about that. High test score means students there have good academic performance, they have lease opportunity to drop out school because of low GPA. Furthermore, there are three latent common causes: the average spending per student and faculty salary, the average spending per student and student teacher ratio, faculty salary and rejection rate. Because of this, other variables are also indirectly cause the freshmen retention.

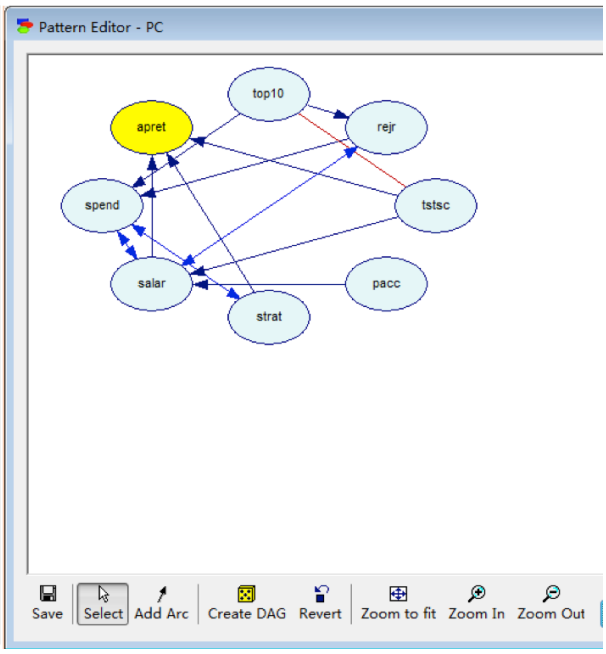




3.2 Set significance level p=0.15

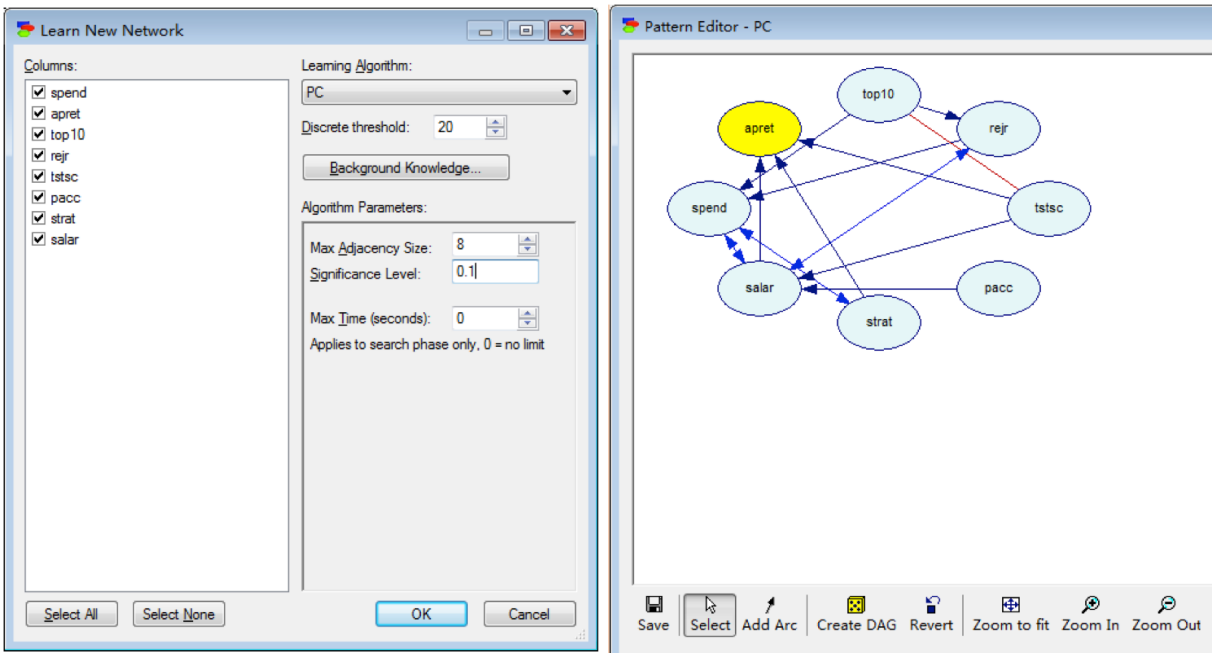
The result is as same as p=0.2





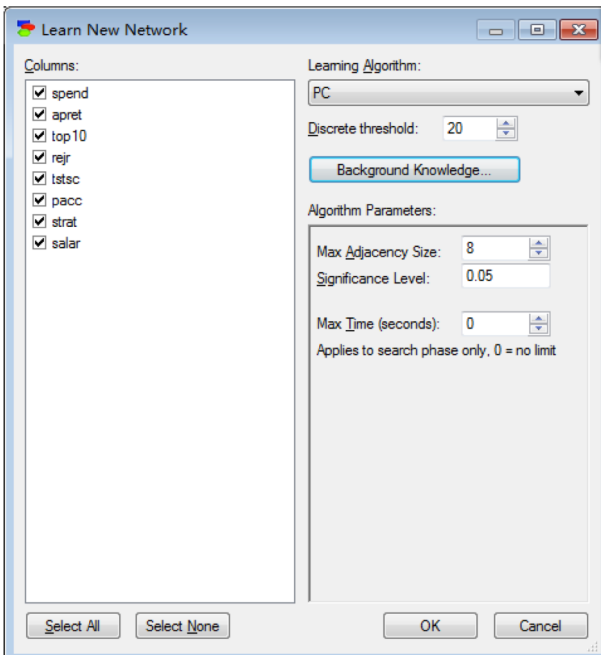
3.3 Set significance level p=0.1

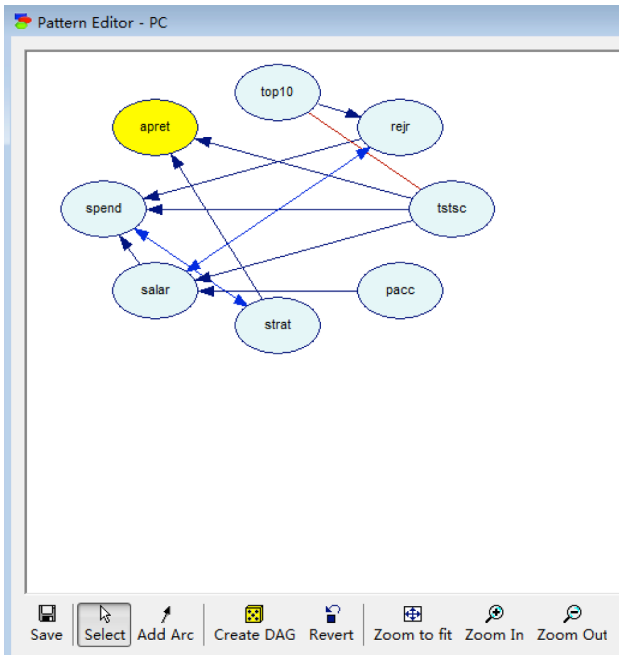
The result is as same as p=0.2



3.4 Set significance level p=0.05

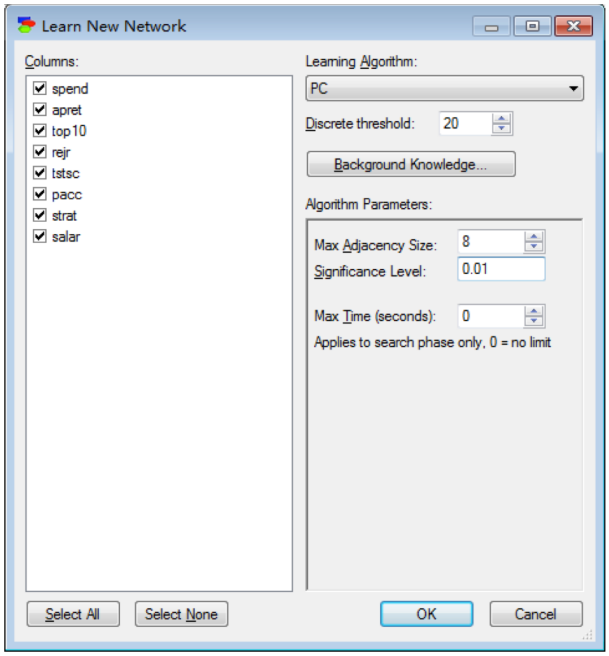
The difference between this and p=0.2 are: class standing doesn't directly cause the average spending per student, faculty salary doesn't directly cause freshmen retention, and the latent common cause relation between the average spending per student and student teacher ratio. The student teacher ratio and average test score become two main causes of freshmen retention. In this case, the number of indirectly causes decreased. Percentage of who accept the university's offer doesn't effect freshmen retention.

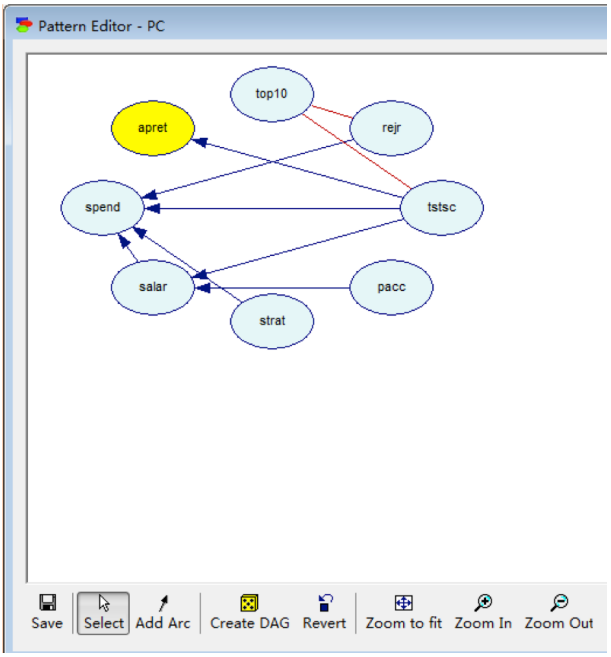




3.5 Set significance level p=0.01

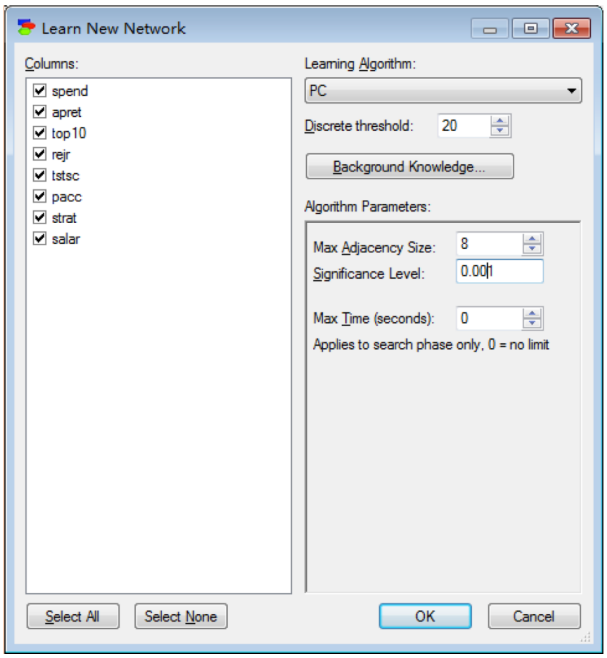
In this case, there is no latent common cause between any variables. The only cause of freshmen retention is students' average test score.

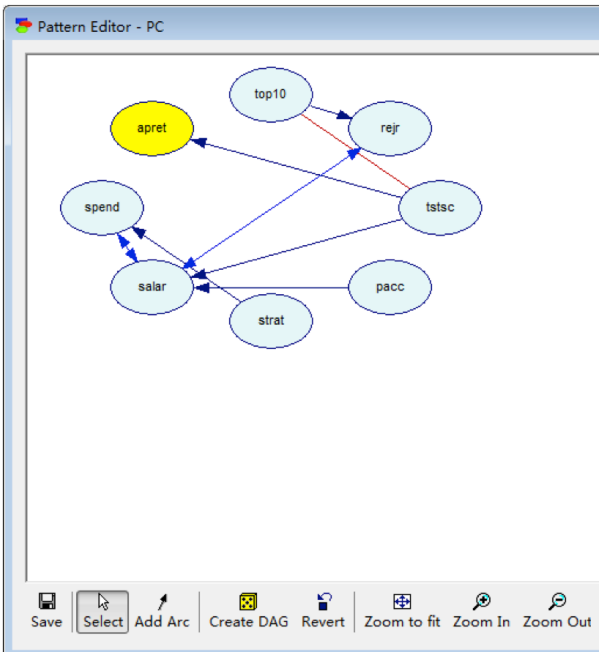




3.6 Set significance level p=0.001

The only cause of freshmen retention still is students' average test score, and there are two latent common causes reappear.





3.7 Summary

Druzdzel & Glymour's conclusion is that Student retention is directly related to the average test scores and high school class standing of the in coming freshmen, none of other variables directly related to freshmen retention. However, in our analysis of data in 1993, we found that only average test score directly causes freshmen retention whatever the significance level change. The high school class standing only has indirectly relation and the relation become less important while significance level decreasing.