Assignment 5: Causal Discovery

Rebecca Ly and Don Stuart

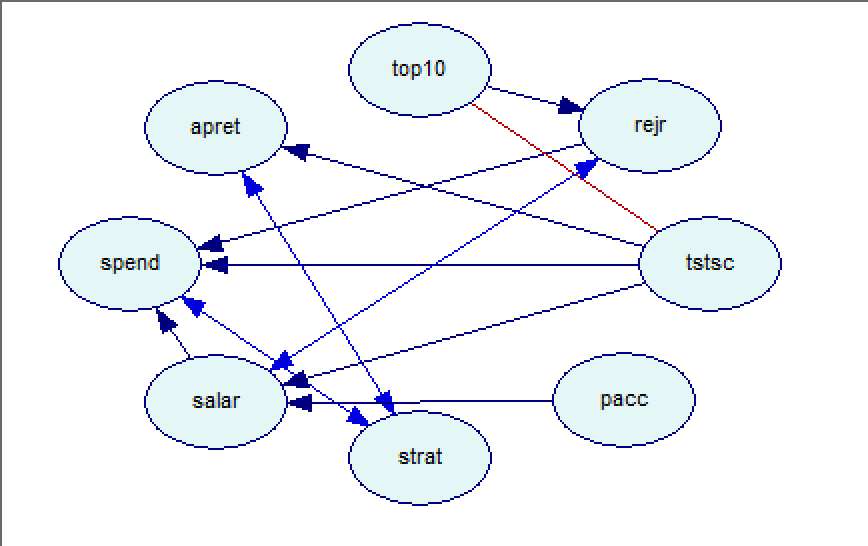
**Main Conclusion from Druzdzel & Glymour 1994 Study**

According to Druzdzel and Glymour, student retention is directly linked to the average test scores and high school class standing of the incoming freshman.

**Node Definitions/Properties**  
spend - average spending per student (in dollars)  
apret - average retention rate (i.e., percentage of students  
 making it through the studies)  
top10 - percentage of incoming freshmen who were among the top  
 10% students in their high schools  
rejr - school's rejection rate (percentage of applicants  
 denied admission)  
tstsc - average test scores of incoming freshmen  
pacc - percent of admitted applicants who accept  
 university's offer  
strat - student-teacher ratio  
salar - average faculty salary (in dollars)

**Performing Data Analysis**

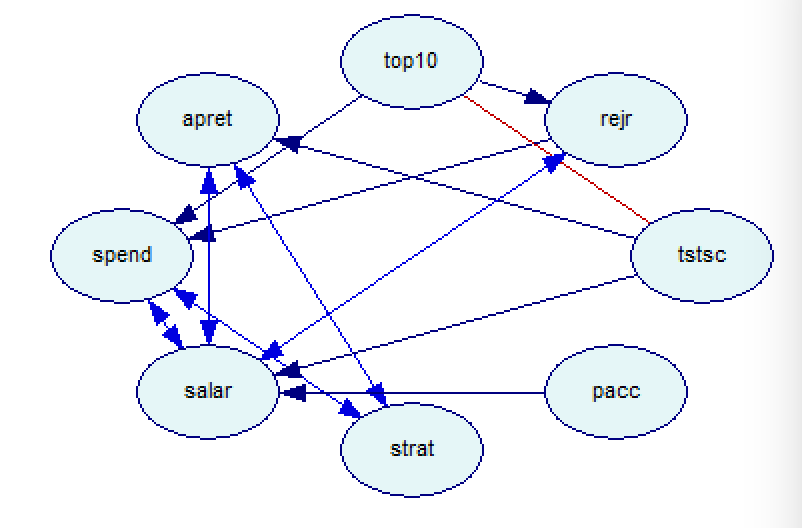
**Running the PC Algorithm with a 0.05 significance level**



**Observations/Conclusions**

This is the initial run of the PC algorithm. According to this figure, the percentage of incoming freshman who were among the top 10% students in their high schools affects the school’s rejection rate. There are many factors that affect the average spending per student such as the rejection rate, average test scores of incoming freshman, student-teacher ratio, and average faculty salary.

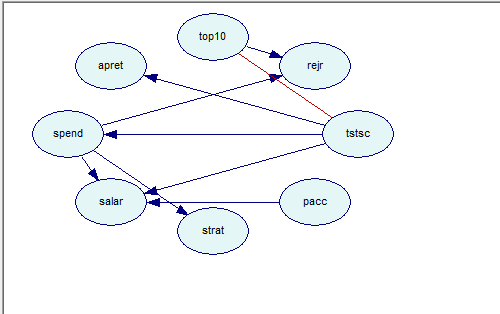
**Running the PC Algorithm with a 0.1 significance level**



**Observations/Conclusions**

When the significance level is improved, you can see that the types of relationships appear to vary greatly with the increased significance level. There is a mutual relationship that forms between apret and strat, apret and salar, and spend and salar. Essentially this means that both variables in the relationship depend on each other for causation.

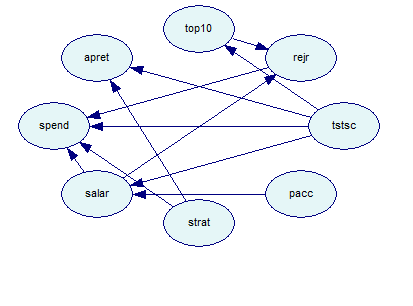
**Running the PC Algorithm with a 0.2 significance level**



**Observations/Conclusions**

This significance level affects the diagram in a few different ways. The top10 node, which represents all incoming freshman who were in the top ten percentile for academic performance in their respective high schools, continues to stand alone as a parent node. Nothing appears to cause this, although it does play a direct role in causing the rejection rate. Of course this makes sense logically, as the higher ranked students, will correlate with a lower rejection rate by the colleges. The average test scores of incoming freshmen once again have a direct causal relationship with the retention rate. This was the main conclusion from the earlier 90s study as well.

**Running the PC Algorithm with a 0.4 significance level**



**Observations/Conclusions**

Setting the significance level to .4 makes the top10 node a child mode and creates a causal relationship between the incoming freshmen average test scores and their top 10 percentile ranking in high school. This makes perfect sense, that students who have higher average test scores also will generally be ranked in the top 10 percentile of their high school and vice versa. The student to teacher ratio appears to directly cause the retention rate. Many students will be more likely to stay at their current school given a low student-teacher ratio. They are likely to seek another school if they feel that they cannot get the intimate help they need which is harder to obtain in a setting where students greatly outnumber the teachers. Another observation is that several variables appear to have a causal relationship with the average spending per student. This is also most often true as there tends to be many factors which cause a student’s spending to be higher as well as lower.