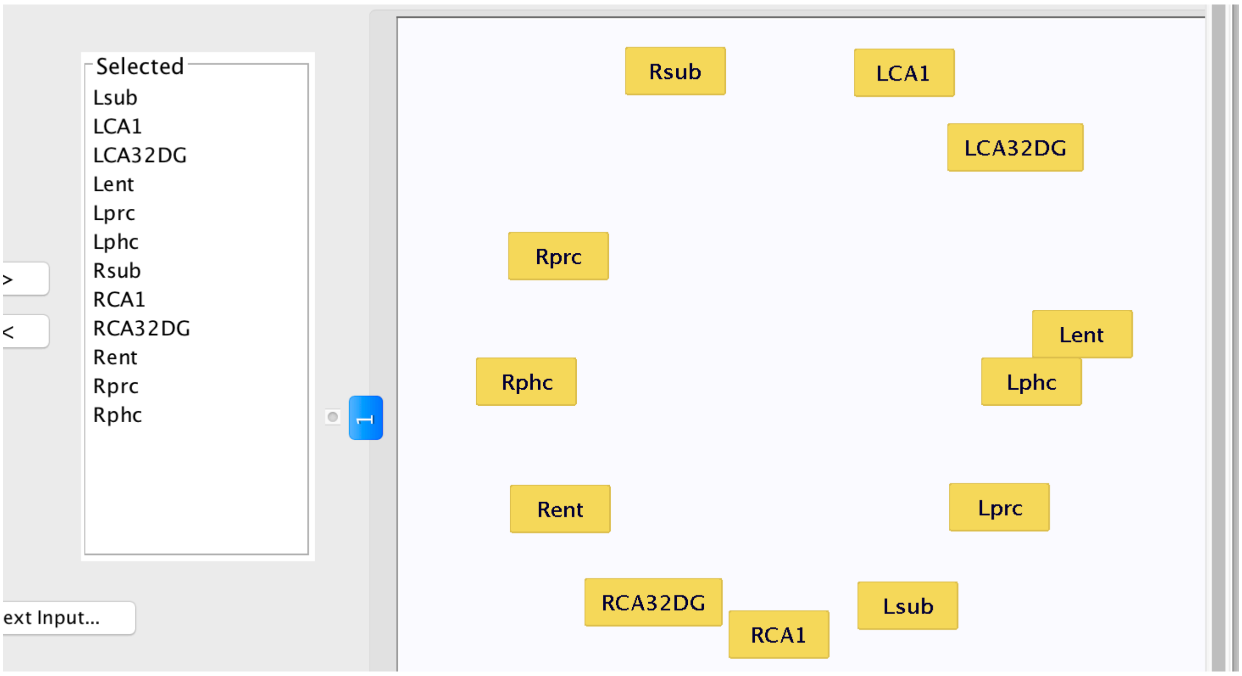
Tetrad各类算法概况

1552970 李向真

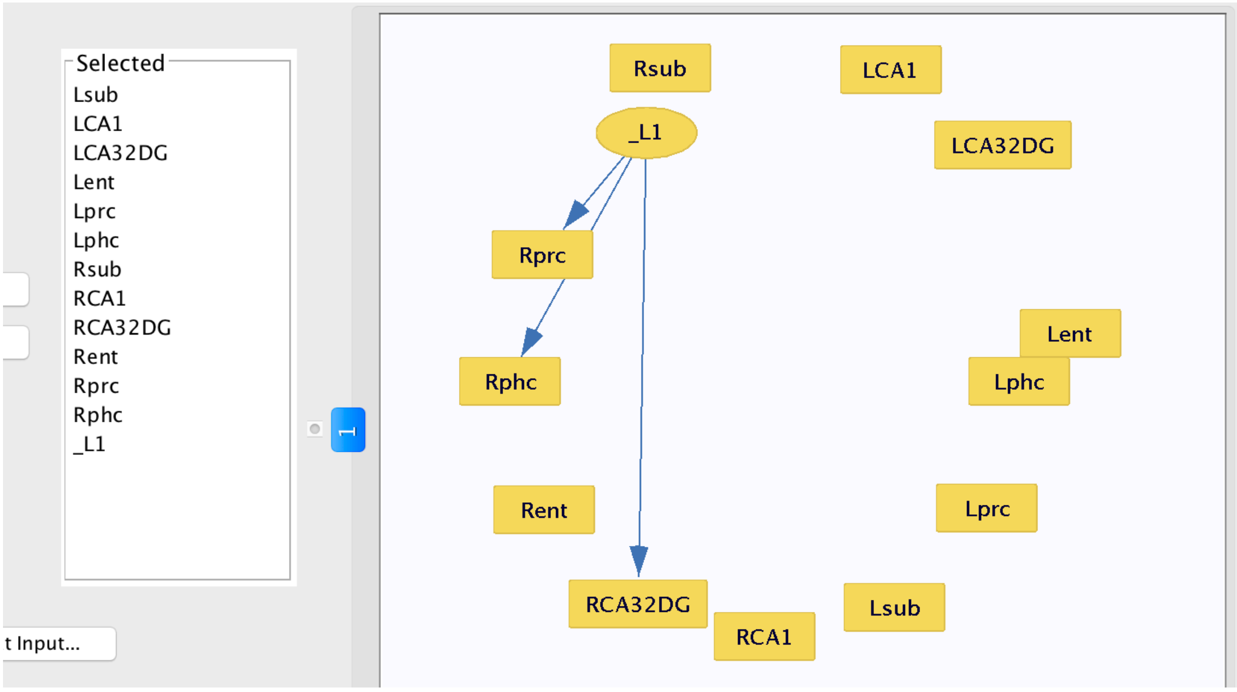
1.1 BPC

- 全名：Build Pure Clusters
- CMU的Richard *Scheines* 提出的算法，只在论文中出现，唯一的实现案例就是Tetrad。它会找出不同变量的共同隐藏根因
- 特点：detect latent variables，即会引入新的隐含变量
- alpha参数：p-values > alpha 则认为独立性成立。故而alpha越小，模型越倾向于认为节点间独立，因此最终生成图中的边越少。
- 实验

alpha <= 0.9



alpha >= 1

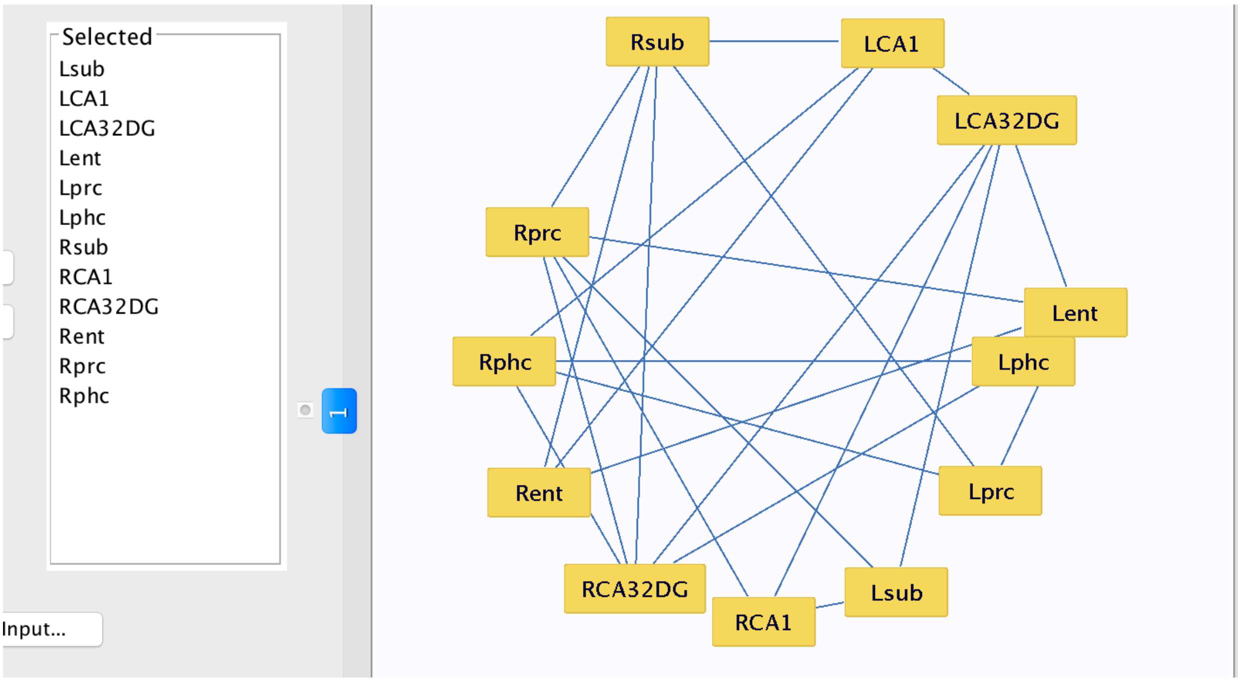


生成图的边少，引入共同的根因隐藏节点

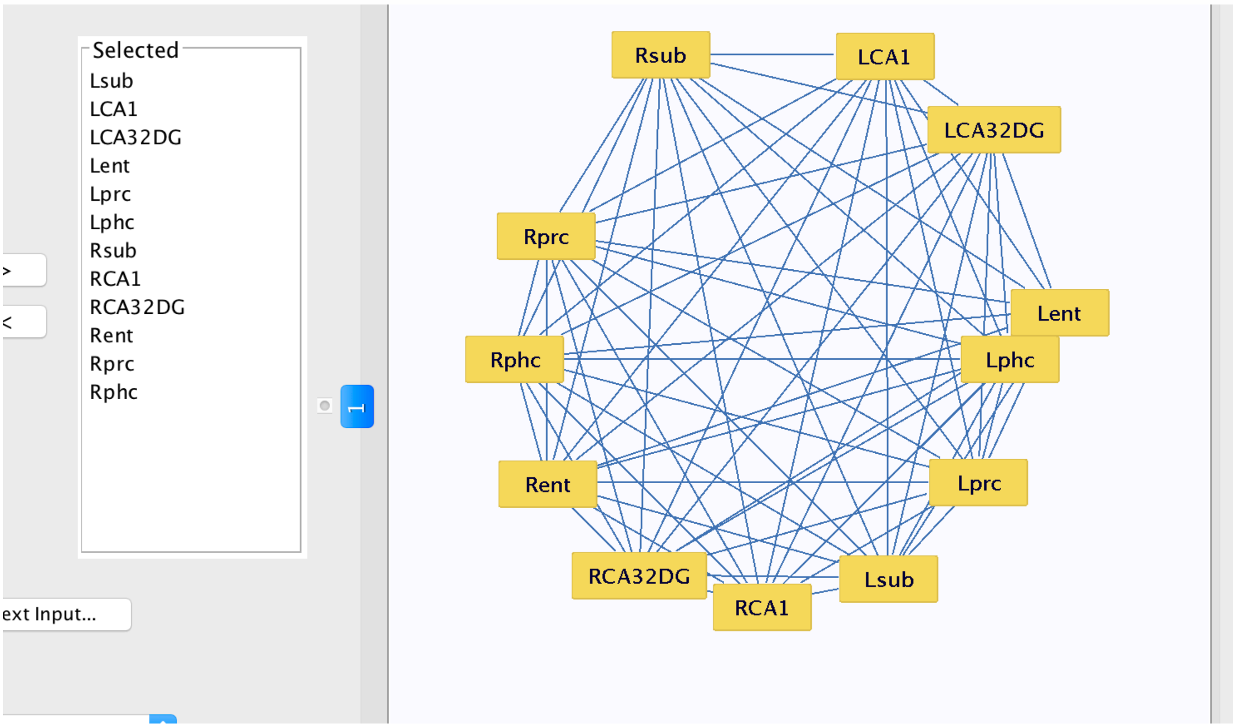
1.2 FAS

- 全名：fast adjacency search
- PC算法的变种，邻接搜索
- 特点：算法输出一个无向图
- 实验

alpha = 0.5



alpha = 1



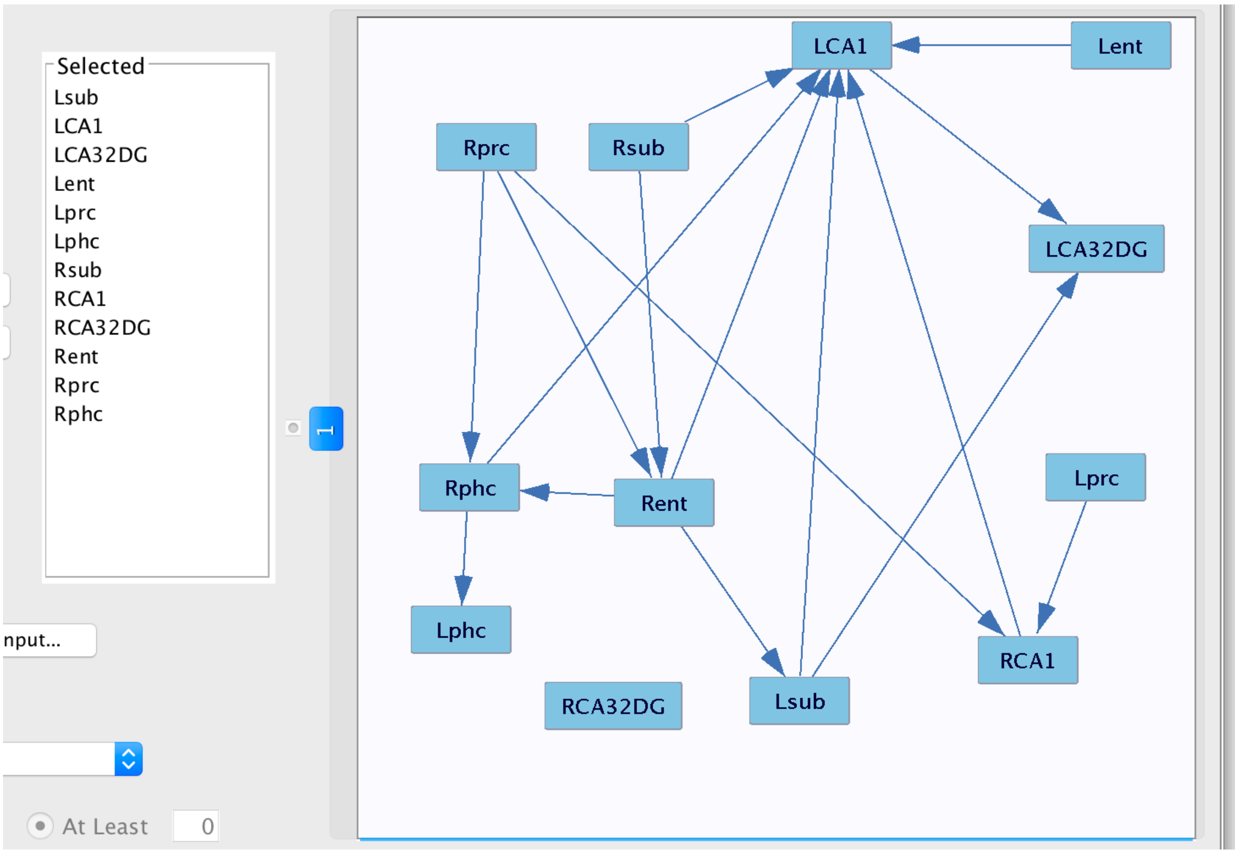
对条件独立不敏感，会保留较多的边

速度较慢，alpha = 1 时，有明显延迟

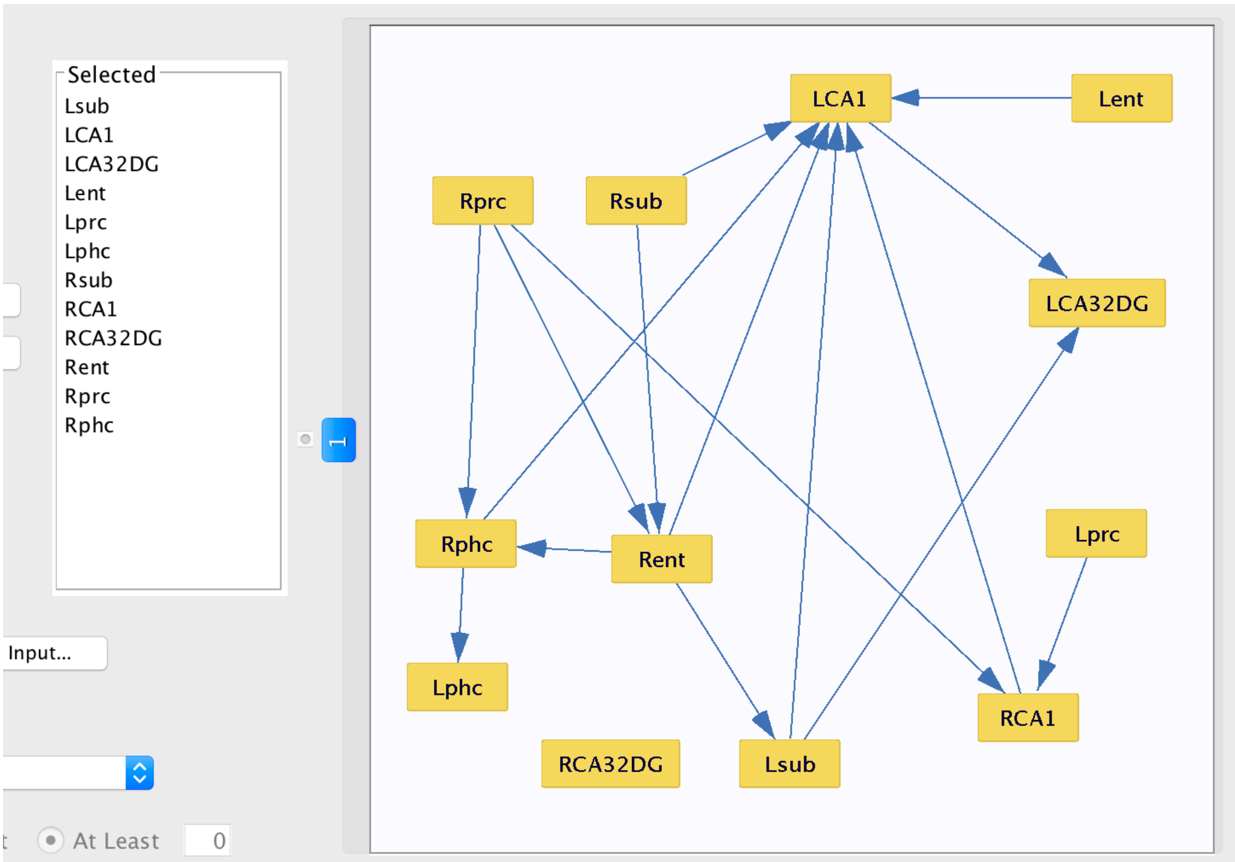
1.3 FASK and FASK Concatenated

- FASK: Fast adjacency search followed by robust skew orientation
- FASK Concatenated: Runs the search on the concatenated data
- 特点：禁止公共潜在根因变量
- 实验

FASK:



FASK Concatenated:



二者搜索所得结果相同

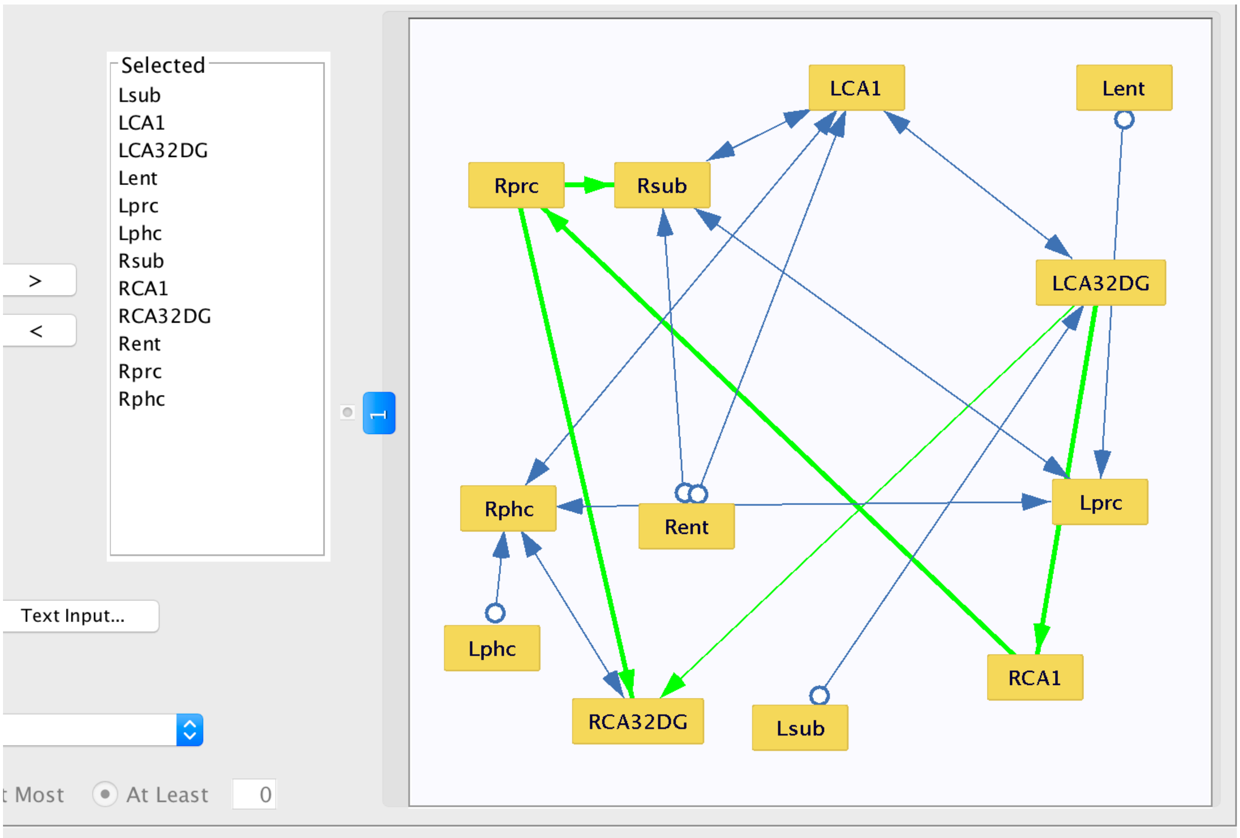
1.4 FCI

- 全名：Fast Causal Inference algorithm
- CMU 的 Jiji Zhang在《Causal Inference and Reasoning in Causally Insufficient Systems》中提出的增强放方法
- 特点：原理上与PC类似，产生PAG（Partial Ancestral Graph），图中包含多种边，能得到很多逆向结论,可以发现latent confounder（潜在混淆）
- FCI系列算法会得到双头边，待圆圈的边，提供变量间不存在因果的信息，而提供较少的因果信息
- 边的含义：

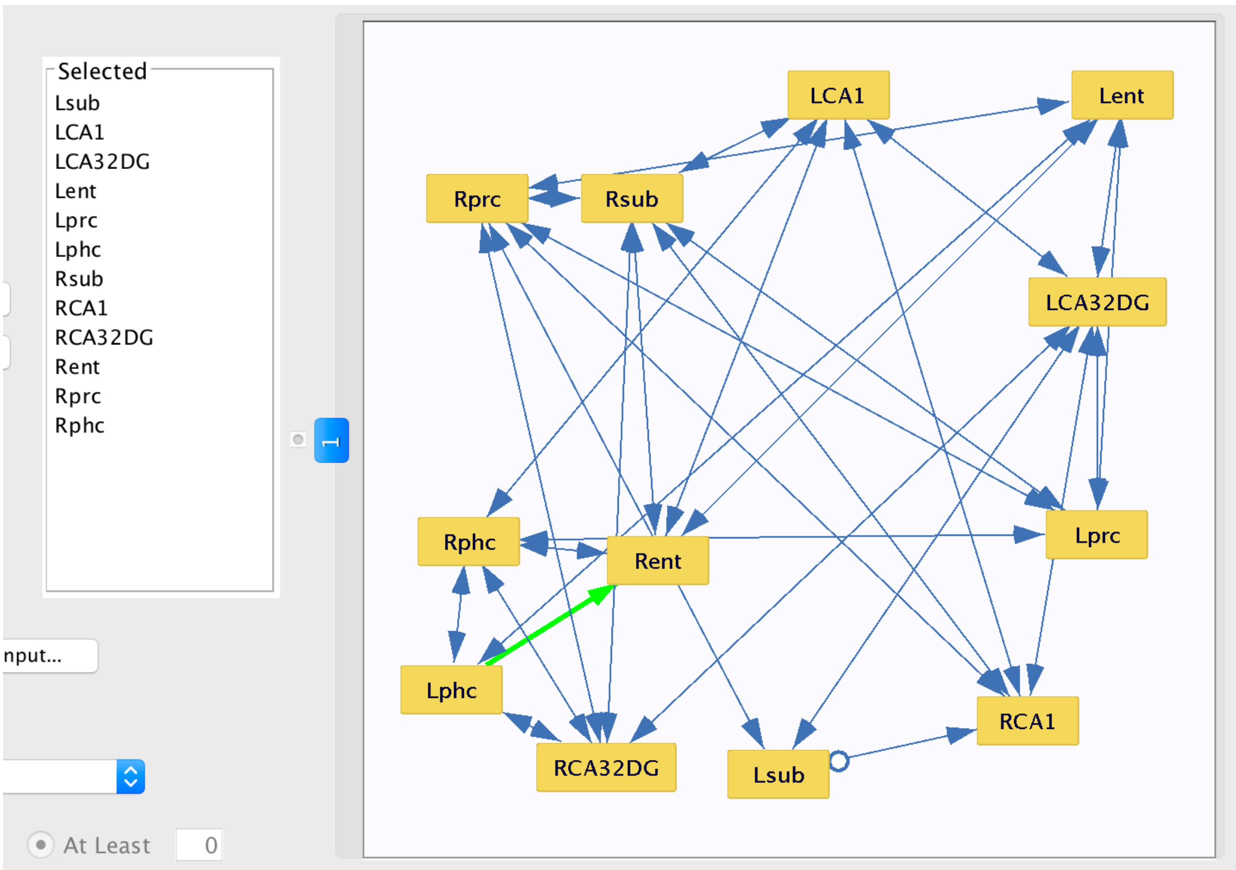
<table>
<tr><th>Edge Types</th><th>Present Relationships</th><th>Absent Relationships</th></tr>
<tr><td>A --> B</td><td>A is a cause of B.
It may be a direct or indirect cause that may include other measured variables. Also, there may be an unmeasured confounder of A and B.</td><td>B is not a cause of A</td></tr>
<tr><td>A <-> B</td><td>There is an unmeasured confounder (call it L) of A and B. There may be measured variables along the causal pathway from L to A or from L to B.</td><td>A is not a cause of B.
B is not a cause of A.</td></tr>
<tr><td>A o-> B</td><td>Either A is a cause of B (i.e, A --> B) or there is an unmeasured confounder of A and B (i.e, A <-> B) or both.</td><td>B is not a cause of A.</td></tr>
<tr><td>A o-o B</td><td>Exactly one of the following holds:
1. A is a cause of B
2. B is a cause of A
3. there is an unmeasured confounder of A and B
4. both a and c
5. both b and c</td><td></td></tr>
<tr><td colspan="3">If an edge is green that means there is no latent confounder. Otherwise, there is possibly latent confounder.</td></tr>
<tr><td colspan="3">If an edge is bold (thickened) that means it is definitely direct. Otherwise, it is possibly direct.</td></tr>
</table>

- 实验：

alpha = 0.01



alpha = 0.5

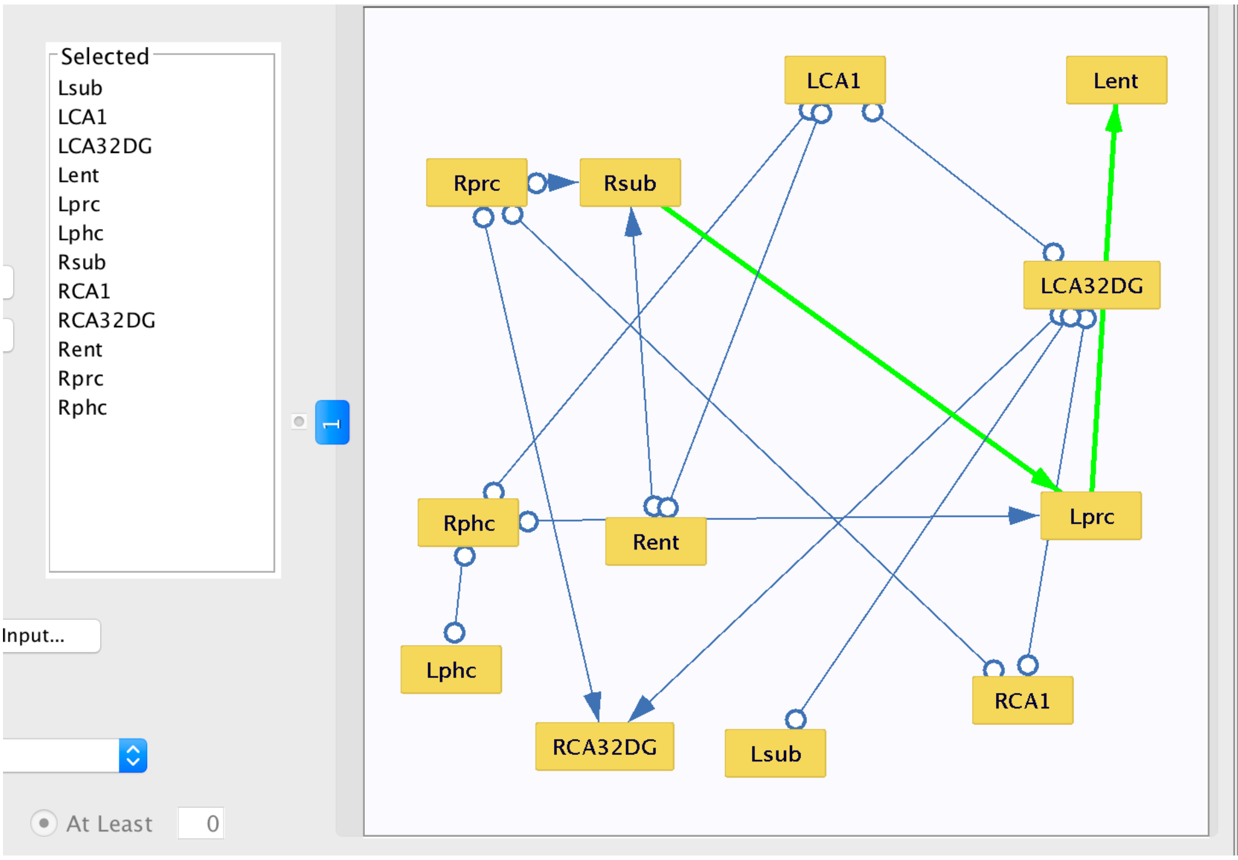


1.5 GFCI, RFCI, TsFCI, TsGFCI

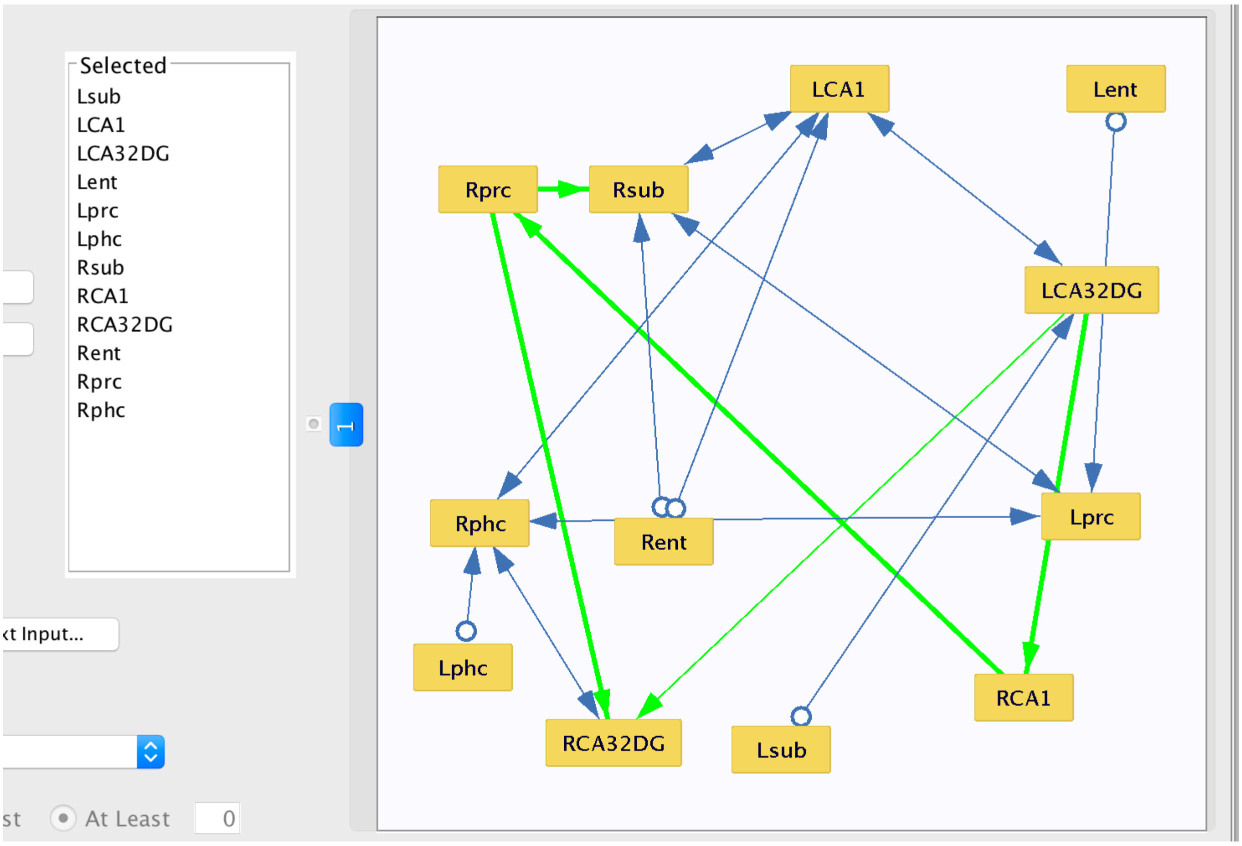
- 这些算法都是基于FCI算法的变种与改进

- 在算法性能，搜索细节上有所不同，例如tsGFCI使用BIC分数来搜索skeleton
- 以Ts起头的两个算法用于处理time series data 时间序列
- 算法返回的图和FCI一样，包含多种边以及其可能的latent confounder
- 实验

GFCI

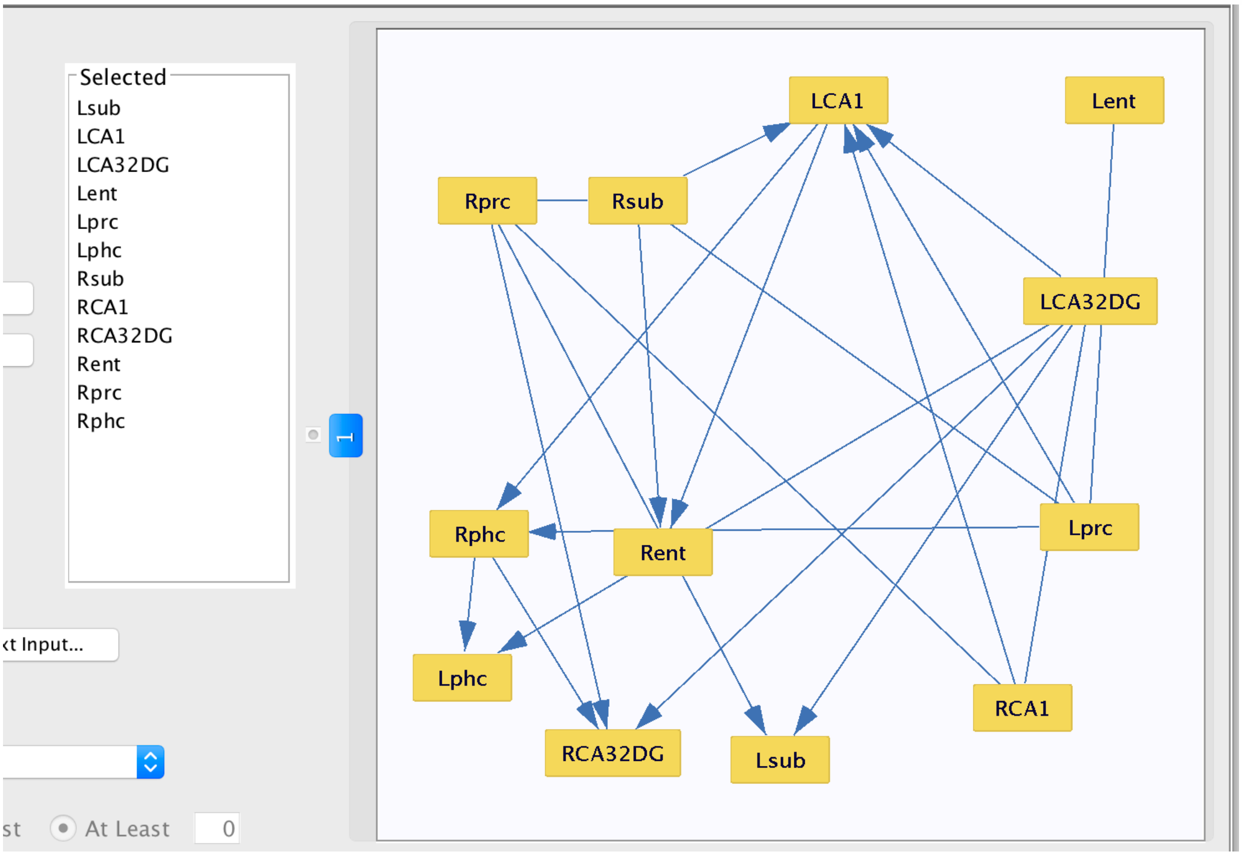


RFCI



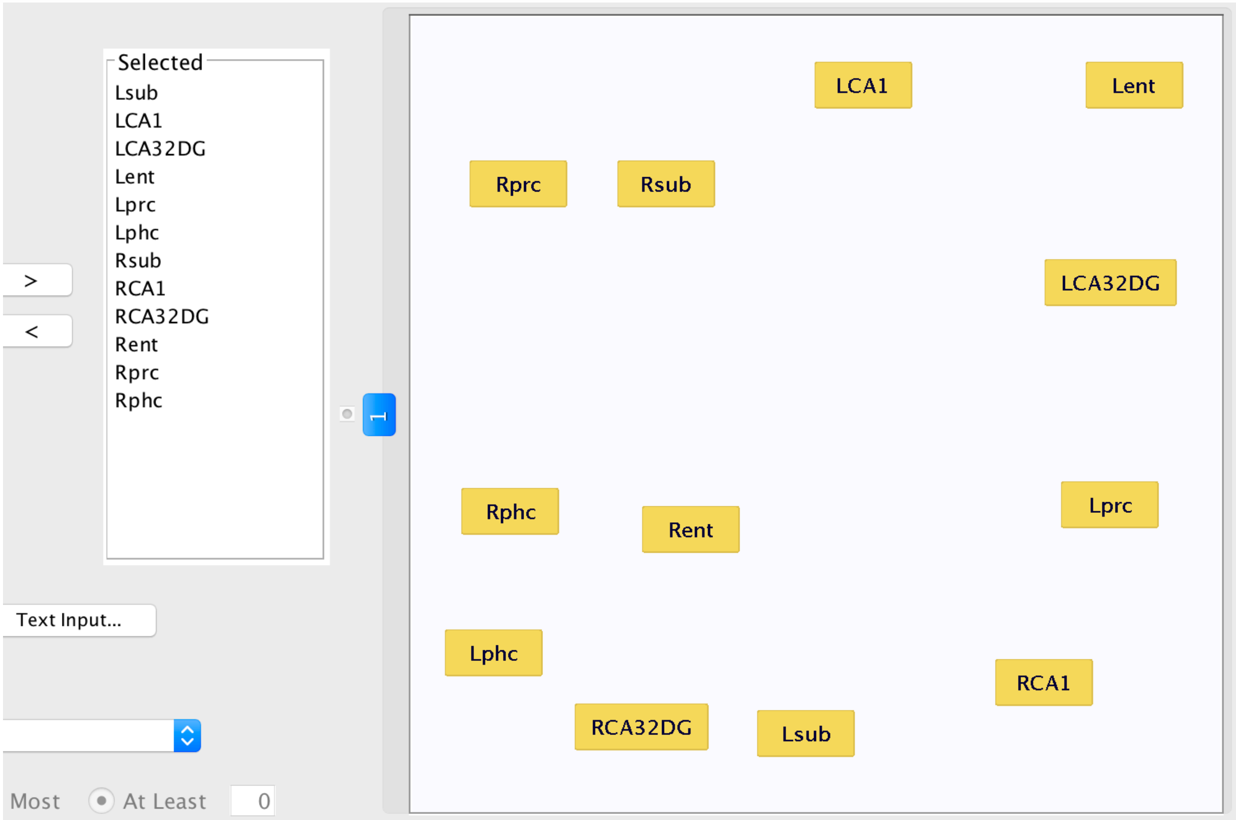
1.6 FGES, FGES-MB, IMaGES-Continunous, IMaGES-Discrete

- Greedy Equivalence Search
- 这类算法即为《Causal Discovery Algorithms and Real World Systems》中提到的基于分数的贪婪搜索算法，与之形成对比的是PC一族算法（基于约束constraint）
- 算法返回的是CPDAG图，这类图中只有一部分遍是被指定方向的，其余边均为无向边
- 实验



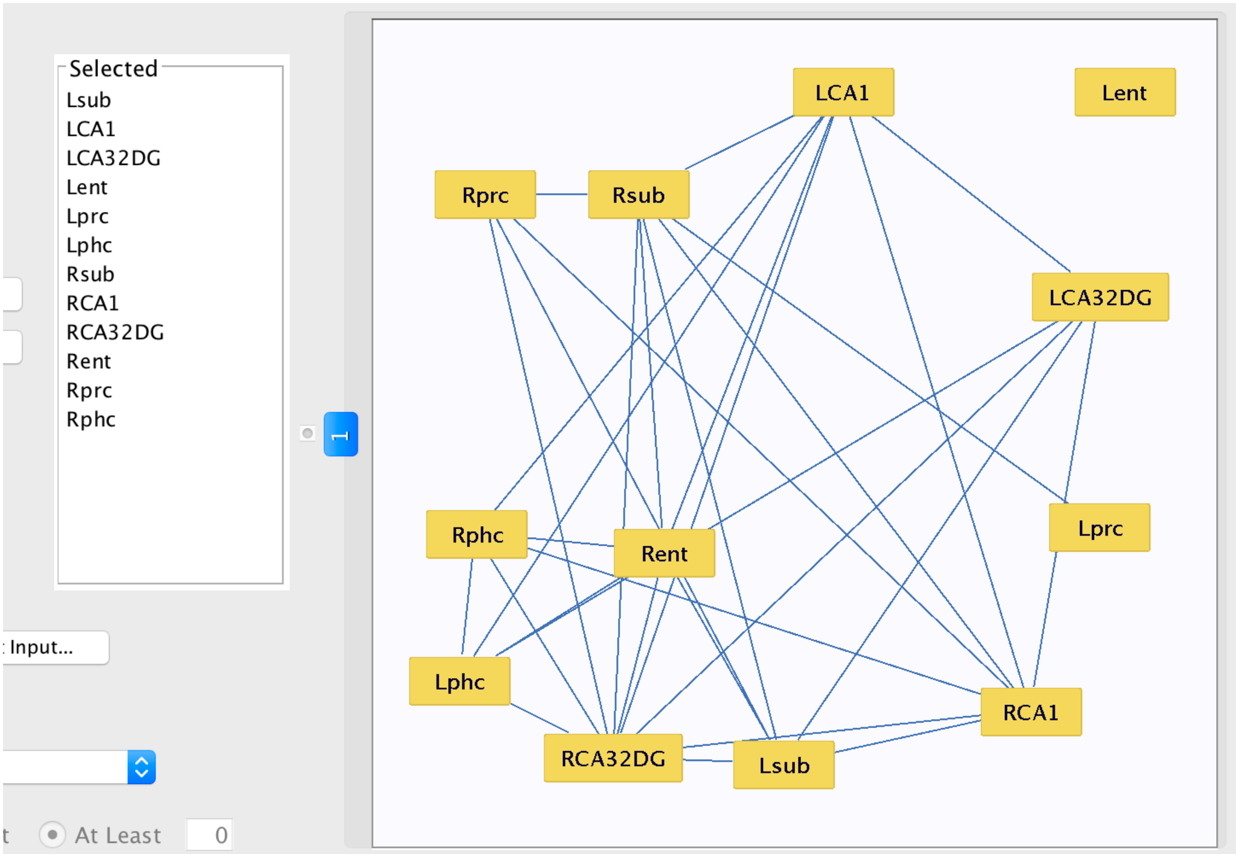
1.7 FOFC, FTFC, MIMBUILD

- FOFC和FTFC都是对BPC的变种
- MIMBUILD需要以BPC, FOFC或FTFC模型为前置条件
- 这些模型共同特点是他们会发掘已知变量共同的潜在根因变量
- 实验



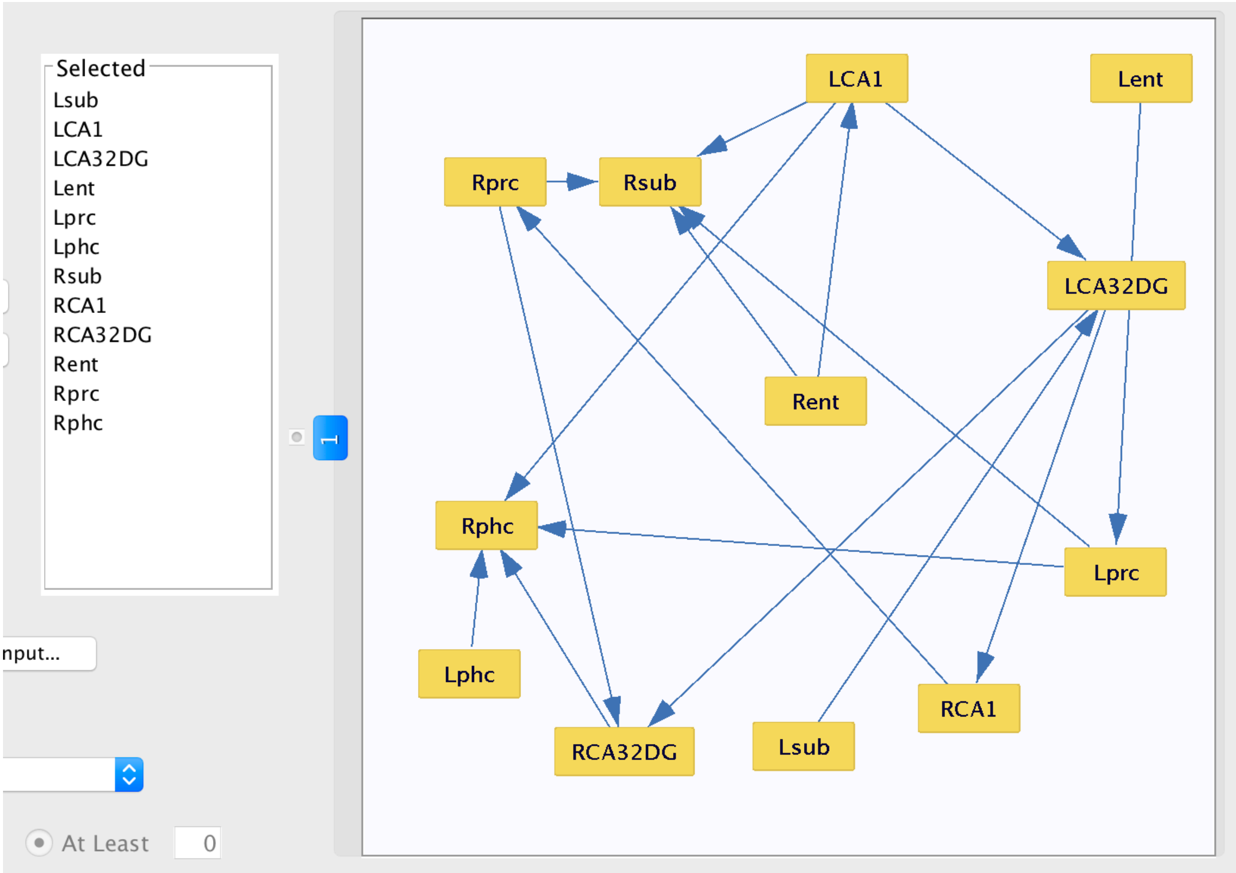
1.8 GLASSO and MGM

- 这两个算法功能相同，都是产生马尔可夫随机场，也就是一个无向图
- 执行MGM模型需要一个连续变量以及一个随机变量
- 实验



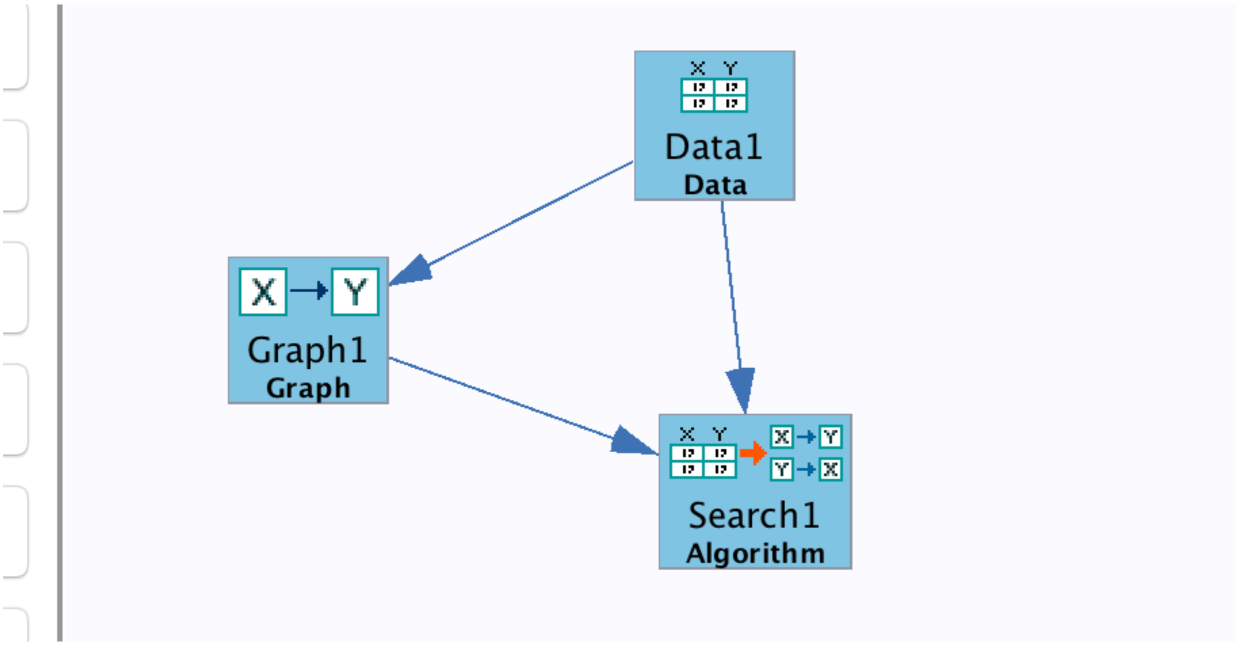
1.9 PC

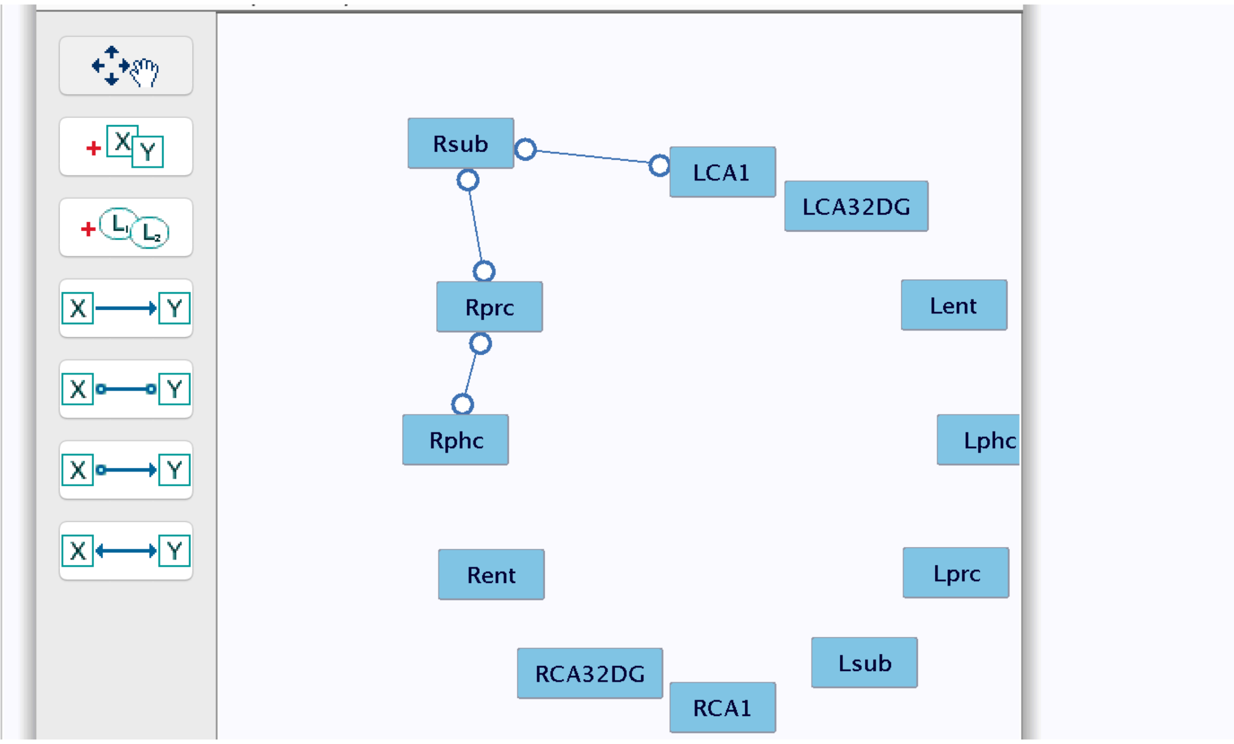
- PC算法，基于约束（constraint）
- 生成一个（DAG）
- 源代码有多种变种的PC算法，Tetrad版本迭代过程中抛弃了一部分，只留了PC all
- 实验

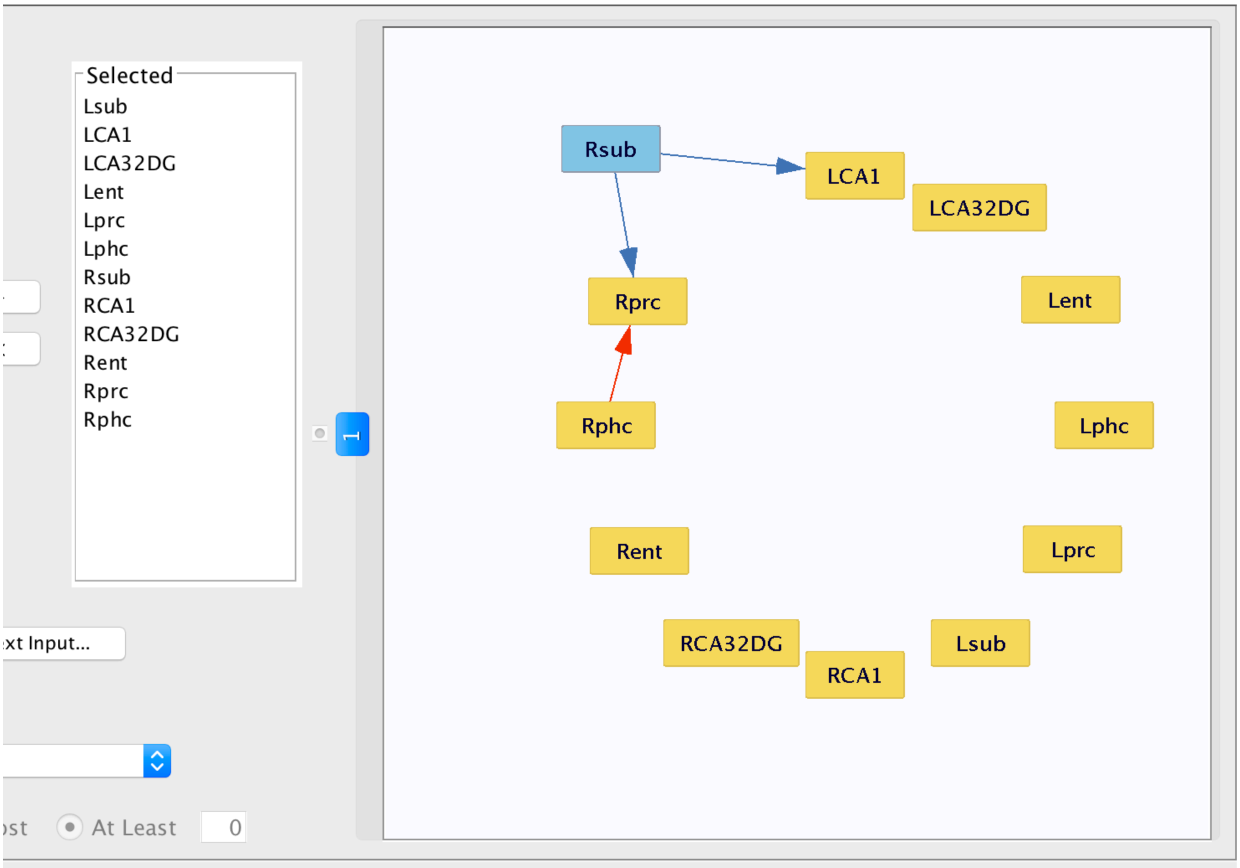


1.10 R1, R2, R3, Skew, SkewE, RSkew, RSkewE

- 这是一组相同功能的算法套件，用于实现无向图中边的定向
- 算法需要两项输入：图以及相关联的数据集
- 实验


Data1
Data
Graph1
Graph
Search1
Algorithm


Rsub
LCA1
LCA32DG
Rprc
Lent
Rphc
Lphc
Rent
Lprc
Lsub
RCA32DG
RCA1



2 总结

2.1 不适用的算法

MIMBUILD，FOFC，FTFC，BPC 这四个算法会求解现有变量的潜在根因，在图中引入新变量。

我们的问题是在确定的服务集中寻找相互潜在关系。

2.2 待商榷的算法

FCI一族的算法的运行结果提供较少的因果信息，相反，提供较多不存在因果的结论。

此外，还提供潜在的latent confounder。

这样的信息用起来不是很方便

2.3 一个可能的策略

以FAS为代表的几类算法会输出无向图，GEM类机遇分数的算法会输出只有部分边定向的CPDAG图。

而R系列以及Skew系列的算法实现边的定向化，二者理论上可以结合使用，单在算法原理上是否是正确的还有待讨论

2.4 符合要求的

PC 系列算法会输出一个DAG图，符合我们问题的需求