

三元闭包原理的大数据验证

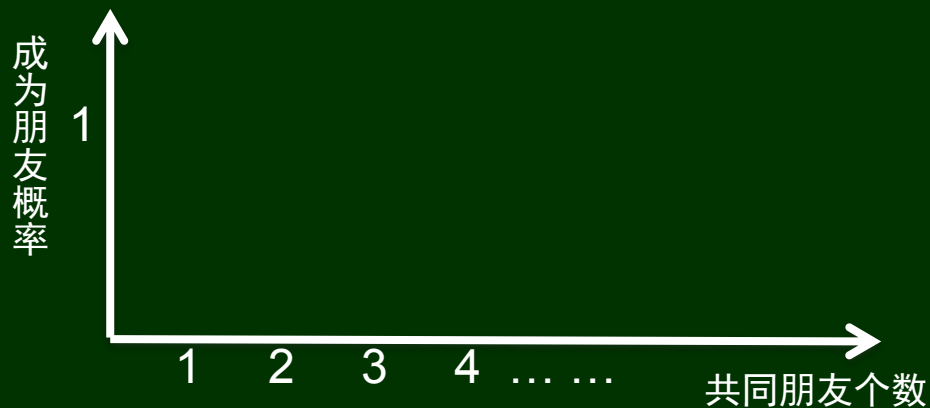
需要先解决两个问题

- 第一，将三元闭包原理最初的定性陈述转变成一种可以定量考察的表达；
- 第二，找到一种合适的社会网络数据。

三元闭包原理的两种表达

- 最初的表述：如果两个互不相识的人有了一个共同朋友，则他们俩在未来成为朋友的可能性增加。
- 可以转变成：如果两个互不相识的人的共同朋友数越多，则他们俩在未来成为朋友的可能性越大。

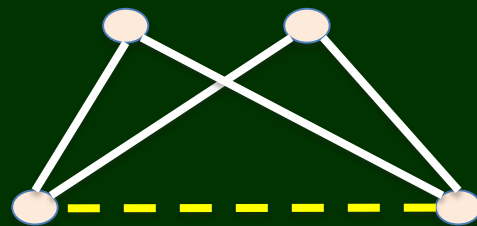
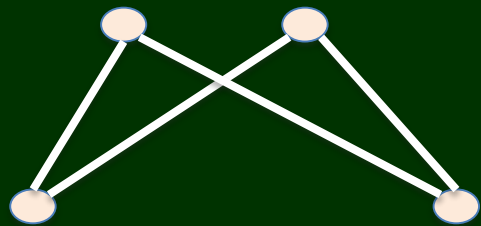
- 三元闭包原理：如果两个互不相识的人的共同朋友数越多，则他们俩在未来成为朋友的可能性越大



用什么数据验证？

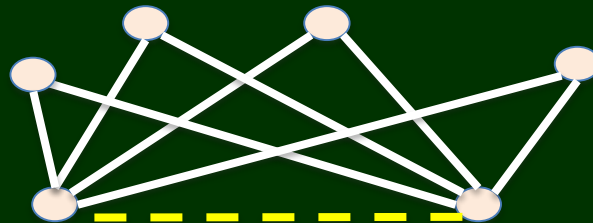
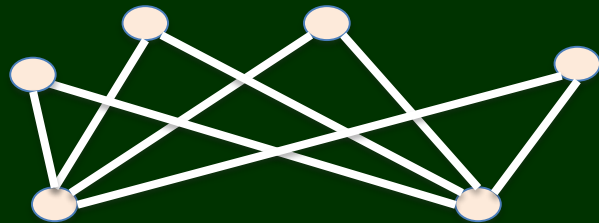
- 电子邮件网络~社会网络
 - 一所大学的2万多学生在一年里的通信关系数据
 - 只关心谁和谁何时有过通信，不关心内容

三元闭包：哪一种情形更有可能



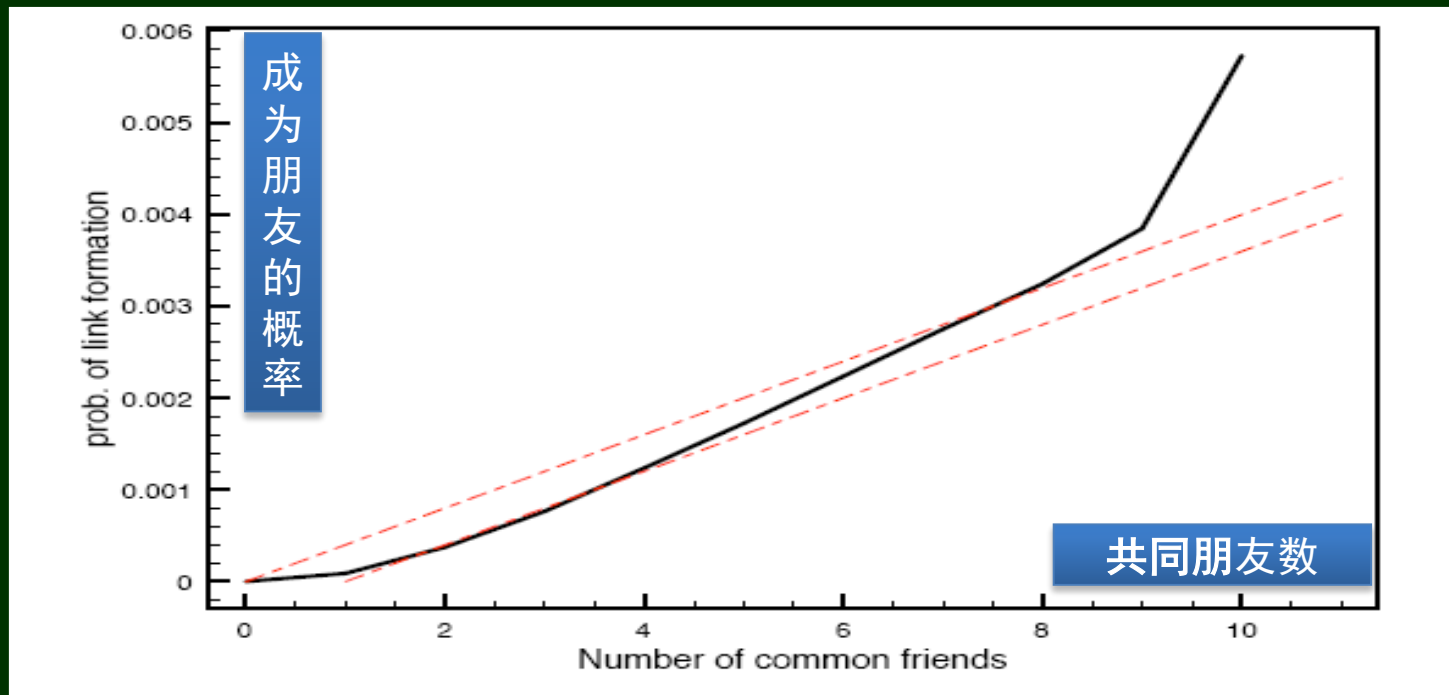
现在

一段时间后



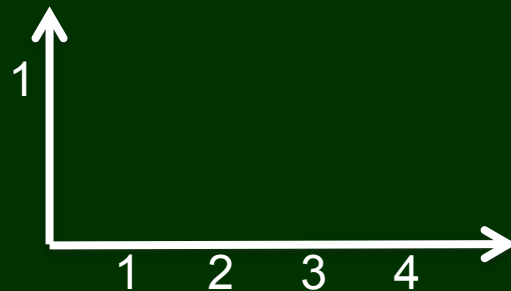
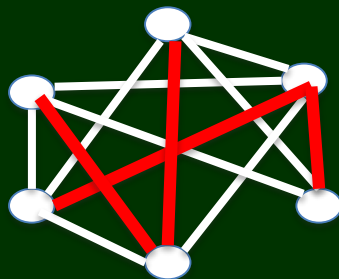
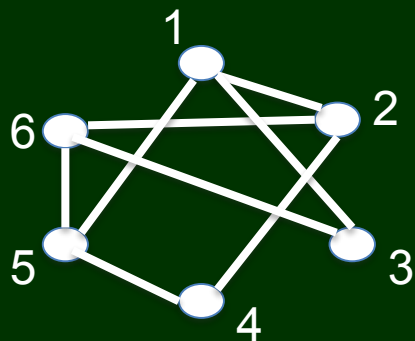
“可能性”怎么衡量？





“Empirical Analysis of an evolving social network,” *Science* 2006

数据验证过程示意



| | | |
|-------|---|---|
| (1,4) | 2 | * |
| (1,6) | 3 | |
| (2,3) | 2 | * |
| (2,5) | 3 | * |
| (3,4) | 0 | |
| (3,5) | 2 | |
| (4,6) | 2 | * |

| 共同朋友数 | 0 | 1 | 2 | 3 | 4 |
|--------|---|---|------|-----|---|
| 成为朋友概率 | 0 | - | 0.75 | 0.5 | - |

小结

- 以三元闭包原理的验证为例，我们看到了一种利用大数据分析，定量考察某些社会科学定性认识的方法
- 其中有两个关键
 - 将社会科学原理的定性描述，转化为便于定量分析的表述，形成数据指标（与共同朋友数对应的概率）
 - 选择合适的数据，以及从原始数据中提炼出指标数据的方法