

# 基于贝叶斯的社会性学习模型

## Social Learning Model based on Bayesian Theory

陈丽云

eBay Inc.

纪念贝叶斯定理250周年  
天津财经大学, 2013

# Outline

## 1 什么是社会性学习？

- 什么是社会性学习？
- 美国高中生性知识教育项目

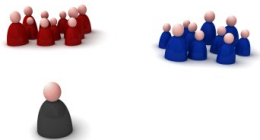
## 2 社会性学习模型：对知识传播的理解和数值模拟

- 网络形态可视化
- 基于贝叶斯公式的学习：信念更新
- 数值模拟

# 人是社会化的动物

- 影响别人：比如我们这群人今天站在麦克风旁边
- 被人影响：比如在场的所有人
  - 特别的，容易被周围的人影响。朋友，同学，亲人，等等。
- 社会学习：一个群体内部、个体之间相互学习的过程。

# 社会性学习模型



- 对于群体内部相互之间关系的数学描述：社会网络
- 一个圈子中间各个成员之间的相似性来源于什么？
  - 圈子形成过程的自选择（self-selection）：物以类聚、人以群分？  
→网络构成模型（Network Formation Models）
  - 圈子形成之后的同伴效应（peer effects）：近朱者赤、近墨者黑？  
→社会性学习模型（Social Learning Models Information Spread Models, Social Influencer Models）

# 豆瓣的例子

- 结果:

- 用户对于其现有好友评分为1星的书籍，相比于其未来好友评分为1星的书籍，会更倾向于给前者较低的评分；对于现有好友评分较高的如5星的书籍，会相比而言给予更高的评分。
- 朋友之间影响更强烈的情况：热门书籍；评分较晚的用户；使用时间较短、阅读经验较少的用户；小众圈子用户。

Source: Wang, Alex, Xiaoquan (Michael) Zhang and Il-Horn Hann, 2010, "Social Bias in Online Product Ratings" Workshop on Information Systems and Economics (WISE), December 2010, St. Louis, USA.

# Outline

## 1 什么是社会性学习？

- 什么是社会性学习？
- 美国高中生性知识教育项目

## 2 社会性学习模型：对知识传播的理解和数值模拟

- 网络形态可视化
- 基于贝叶斯公式的学习：信念更新
- 数值模拟

# 美国高中生性知识教育项目

- 美国青少年艾滋病传播情况：2006年到2009年期间，超过 8,500例新增艾滋病人为13-19岁的青少年。
- 传统性知识教育方式：课堂教育，受诸如宗教信仰、老师讲演水平限制很大。多数青少年不懂的如何利用正确的方式保护自己，不会主动和正确的使用安全套等基本预防手段。
- 欲尝试的新方式：通过青少年网络进行的知识传播——青少年之间性已经是不可避免的话题，相互之间的知识和态度共享非常普遍。
  - 对青少年朋友网络进行调查，选取其中的关键人物和活跃人物。
  - 对活跃人物进行性知识普及教育，引导其正确使用相关措施。
  - 鼓励活跃人物主动向朋友们传播知识。

# Outline

## 1 什么是社会性学习？

- 什么是社会性学习？
- 美国高中生性知识教育项目

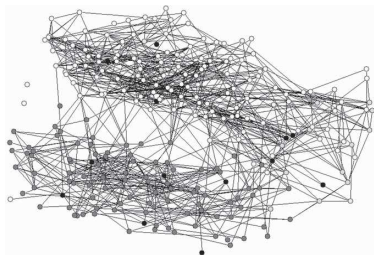
## 2 社会性学习模型：对知识传播的理解和数值模拟

- 网络形态可视化
- 基于贝叶斯公式的学习：信念更新
- 数值模拟



# 美国高中生朋友网络形态

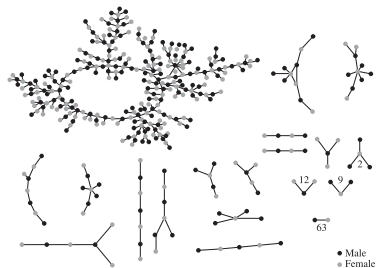
Figure: 美国高中生朋友网络形态



数据来源: ADD Health Database

# 美国高中生性网络形态

Figure: 美国高中生性关系网络形态



数据来源：ADD Health Database

# Outline

## 1 什么是社会性学习？

- 什么是社会性学习？
- 美国高中生性知识教育项目

## 2 社会性学习模型：对知识传播的理解和数值模拟

- 网络形态可视化
- 基于贝叶斯公式的学习：信念更新
- 数值模拟

# 基于贝叶斯公式的学习：信念更新

		收益	
		1 (约会)	$-\infty$ (感染HIV)
状态	安全 (概率 $1 - p$ )	1	0
	不安全 (概率 $p$ )	$1 - q$	$q$

# 期望最大化

- 最优决策原则：

$$action, S_B = \begin{cases} 1, & \text{如果风险预期 } r_B < \frac{1}{2} \\ 0, & \text{如果风险预期 } r_B \geq \frac{1}{2} \end{cases} \quad (1)$$

# 期望最大化： $p=1$

- 简化情形：  $p = 1$ , blind dating 永远是不安全的。

$$r_A^{post}(\text{state=unsafe}) = \begin{cases} \frac{q \cdot r_A^{pre}}{q \cdot r_A^{pre} + 0 \cdot (1 - r_A^{pre})} = 1 & \text{if } \pi_B = -\infty \\ \frac{(1-q) \cdot r_A^{pre}}{(1-q) \cdot r_A^{pre} + 1 \cdot (1 - r_A^{pre})} & \text{if } \pi_B = 1 \\ r_A^{pre} & \text{if } S_B = 0 \end{cases} \quad (2)$$

$$\Pr(\pi_B^t = 1, \forall t \leq T) = (1 - q)^T \rightarrow 0 \text{ as } T \rightarrow \infty \quad (3)$$

- 无论参与者一开始的风险预期是什么水平，随着时间的增加，此人的风险预期都会收敛到于1。

# 期望最大化: $p < 1$

- 一般情况: blind dating 以  $p$  的概率安全。

$$r_{A,m}^{t+1} = \frac{r_A^t(1-q)}{1-r_A^t q} = \frac{1-q}{1/r_A^t - q}, \quad \forall t \geq 1 \quad (4)$$

- when  $T \rightarrow \infty$ :

$$plim_{T \rightarrow \infty} r_{A,m}^T = 0 \quad (5)$$

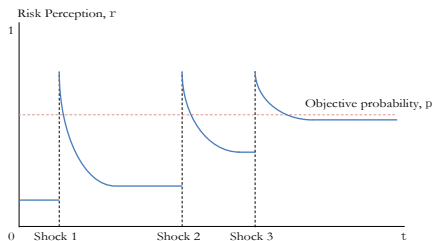
$$\Pr(\pi_B^t = 1, \forall t \leq T) = (1-q)^T \rightarrow 0 \text{ as } T \rightarrow \infty \quad (6)$$

$$plim_{t \rightarrow \infty} \bar{r}_A^t = \frac{pM \cdot 1 + (1-pM) \cdot 0}{M} = p$$

- 无论参与者一开始的风险预期是什么水平, 随着时间的增加, 此人的风险预期都会收敛到于真实概率  $p$ 。

# 更复杂的情形

- 不完美记忆：依旧会收敛到预期，但是收敛过程变长。





# 群体中的学习

- 群体中：同时观测到身边多个案例。

$$r_A^{post} = \frac{q^{s_A} (1 - q)^{\frac{1}{2} N_A - S_A} r_A^{pre}}{q^{s_A} (1 - q)^{\frac{1}{2} N_A - S_A} r_A^{pre} + (1 - q)^{s_A} q^{\frac{1}{2} N_A - S_A} (1 - r_A^{pre})} \quad (7)$$

- 不完美的信息传递：对网络进行加权，

$$r_A^{post} = \frac{\prod_{j \in G} q^{s_A \theta_{Aj}} (1 - q)^{(\frac{1}{2} N_A - S_A) \theta_{Aj}} r_A^{pre}}{\prod_{j \in G} q^{s_A \theta_{Aj}} (1 - q)^{(\frac{1}{2} N_A - S_A) \theta_{Aj}} r_A^{pre} + \prod_{j \in G} (1 - q)^{s_A \theta_{Aj}} q^{(\frac{1}{2} N_A - S_A) \theta_{Aj}} (1 - r_A^{pre})}$$

# 美国高中生性知识教育项目

- 数值模拟思路：按照社会网络学习模型，模拟没有外界干涉情况下网络自发的艾滋病传播情况。

# Outline

## 1 什么是社会性学习？

- 什么是社会性学习？
- 美国高中生性知识教育项目

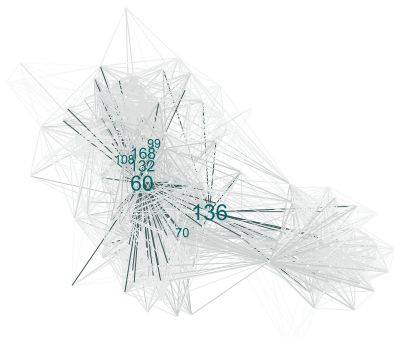
## 2 社会性学习模型：对知识传播的理解和数值模拟

- 网络形态可视化
- 基于贝叶斯公式的学习：信念更新
- 数值模拟

# 美国高中生性知识教育项目：数值模拟

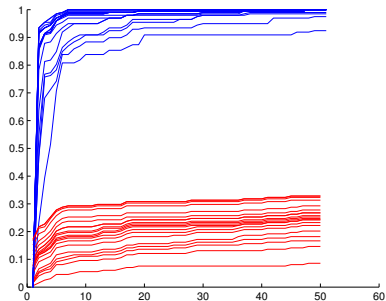
- 数值模拟思路：按照社会网络学习模型，模拟没有外界干涉情况下网络自发的艾滋病传播情况。
- 网络：爵士音乐家

Figure: 社会网络学习数值模拟网络



# 美国高中生性知识教育项目：数值模拟

Figure: 社会网络学习数值模拟结果

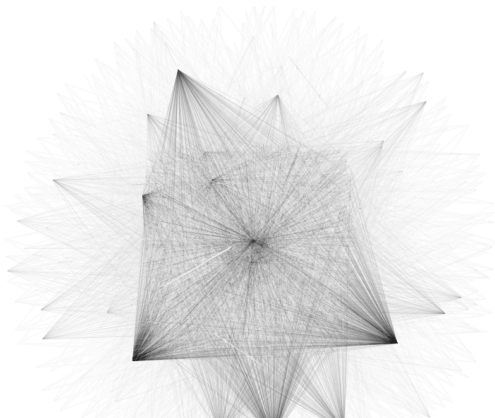


Time periods,  $t = 50$ ; *Repetitions* = 20; Number of nodes,  $N = 198$   
Red: proportion of nodes that are being infected  
Blue: average risk perception in the network

# 美国高中生性知识教育项目：数值模拟

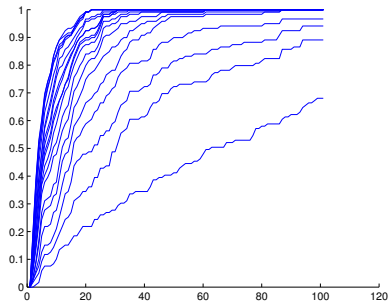
- 数值模拟思路：按照社会网络学习模型，模拟没有外界干涉情况下网络自发的艾滋病传播情况。
- 网络：短信通讯

Figure: 社会网络学习数值模拟网络



# 美国高中生性知识教育项目：数值模拟

Figure: 社会网络学习数值模拟结果



Time periods,  $t = 100$ ; Repetitions = 20; Number of nodes,  $N = 119$   
(network core)

Blue: average risk perception in the network

谢谢！