1.网络与图论

1.1 无处不在的网络

Summary

- 网络并不是今天才出现的现象，自然界原本就有，人类社会自古也有
- 对人类社会而言，凡是有人群的时间与空间，就有网络
- 网络不仅无处不在，而且形形色色
- ICT环境的发展， 使人群之间的网络现象及其意义更加凸显
- 计算能力的发展，使得针对网络现象的可计算性越来越强，进而使得对人群的行为及其与网络之间的互动的可预测性也越来越有意义

1.2 网络与图

Summary

- 图是网络结构信息的抽象，表达的是网络中各种事物之间的关系
- 这里所说的“图”，不同于日常生活中看到的“图像”，尽管图常常也可以被画出来，呈现出一种图像的形式
- 同一个图，可能有多种不同的画法，可能呈现出不同的图像形式 **(同构)**
- 信息可视化编码（节点大小，形状，颜色，标号等）
- 局部、微观 -> 全局、宏观 (整个网络)

图的数据表示

- 邻接矩阵
- 关联矩阵
- 邻接表
- 边列表

生活中的需求 ---- 图上的操作

- 找一个不认识的人帮忙，如何最有效 —— 找图的最短路径
- 新到一个单位，跟哪些人交朋友，可能获得最丰富多样的信息 —— 找到各个连通分量

1.3 路径与连通

连通分量：

- 节点之间存在路径
- 不包含在其他的连通分量中

1.4 二部图与广度优先搜索

二部图

- 顶点集V可分割为两个互不相交的子集，并且图中每条边依附的两个顶点都分属于这两个互不相交的子集，两个子集内的顶点不相邻

充要条件

- 一个图是二部图的充要条件是它没有**长度为奇数的圈**

BFS法判断

- 从任何节点开始，在BFS遍历过程中，一旦发现同一层的节点之间有边，则图中一定存在长度为奇数的圈

1.5 三元闭包与聚集系数

三元闭包

- 社会网络演化的基本结构性原因。
- 如果两个互不相识的人有了一个共同的朋友，则他们两人将来成为朋友的可能性提高
- 社会网络中边增加的主要原因
- 微信：推荐好友名片

聚集系数

- 节点A的聚集系数 = 与A相邻的任意两个朋友之间也是朋友的概率 = 与A相邻的朋友对的个数/总数
- 对一个节点属性的测度，表示"凝聚力的大小"

人际关系强度

- 接触频度
- 情感亲密度

Summary

- 节点之间的关系，会随着时间的变化而发生变化，有些之前没有关系的节点之间，也可能出现边
- 结构性机制：在三个点之间如果有两条边，则没有边的节点之间极有可能发展出边
- 在一个网络结构中，某些位置的点具有特殊意义
- 刻画一个网络结构中节点的属性，可采用聚集系数

1.6 三元闭包原理的大数据验证

用数据分析三元闭包原理

1. 将三元闭包原理最初的定性陈述转变成一中可以定量考察的表达
2. 找到一种合适的社会网络数据

三元闭包原理的两种表达

1. 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
2. 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

数据

- 电子邮件网络 ≈ 社会网络

可能性衡量

- 假设网络中有100对节点，某一时刻前没有边，但分别都恰好有5个共同的朋友
- 如果一个月内，其中有20对节点两两之间发生了通信，80对仍然没有
- 则两个不相识但有5个共同朋友的人，在一个月里将成为朋友的概率为0.2
- 要求
    - 网络数据足够大
    - 时间跨度足够长

Summary

- 以三元闭包原理的验证为例，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
- **两个关键**
    - 定性原理定量化
    - 占比≈概率

1.7 强关系与弱关系

边的一种属性类型

- **强联系**：对应朋友关系
- **弱联系**：对应熟人关系
- 简化的测度 (二分)

被动参与

- 三元闭包原理暗含了一个随时间推移的可能：有的人会被动地加入某些网络

嵌入性

- 一条边两端共同的邻里数
- 嵌入性越强的边，相互之间的信任就越强
- 嵌入性越强的边，社会资本也越多

结构洞

- 一个节点，移除该节点就会使网络变成多个连通分量的节点
- **意义**
    - 了解多方面的信息
    - 处于捷径的一端，对其“长处”有放大影响
    - 对与其相邻的节点甚至具有“权力”
    - 冗余(凝聚力冗余和结构等位冗余)越小的结构洞，社会资本就越多

Summary

- 对网络中某个时点结构中的边的属性，可以用强弱进行测度
- 边的属性可以用嵌入性来测度，嵌入性越强，社会资本越强
- 点的属性除了聚集系数外，还有特殊的社会意义，即结构洞意义，冗余越小的结构洞，社会资本也越多

1.8 弱关系与网络中的捷径

捷径

- 将两个原本相距较远的节点直接相连，没有共同邻居

强三元闭包原理

- 如果A和B、C之间的关系分别为强关系，则B和C之间形成边的可能性应该很高
- 若节点A有两个强关系邻居B和C，但B和C之间没有任何关系，则称节点A违背了强三元闭包原理
- 若节点A没有违背强三元闭包原理，则称节点A符合强三元闭包原理

捷径~弱关系

- 若节点A符合强三元闭包，且至少有两个强关系邻居，则与A相连的任何捷径必定意味着是弱关系
- 共同朋友的存在对强关系的维系很重要

大数据验证

- 强三元闭包原理的精神
  - 没有共同朋友->捷径->弱关系
- 定量化表述
  - 关系强度与共同朋友数正相关

Summary

- 在一定条件下，社交网络中的捷径意味着弱关系
- 将来源于两个不同领域的概念联系了起来

2.社会选择与社会影响

2.1 同质性与社交关系

同质性

- 身份同质性：相同身份的人会彼此相互联系
- 价值同质性：相同价值观的人会彼此相互联系

同质性机制

- 生态过程：场所的影响，同一个机构、社区，共同参加活动
- 关系过程：交叉关系的影响，一个人在不同的场所
- 网络过程：随时间的变化而变化的动态

人的特质

- 固有特质：性别、种族等，自然属性
- 可变特性：居住区、偏好，建构属性

同质性是社会网络结构形成的基本外部原因

2.2 社交网中同质性的一种测量方法

社交网中的同质性现象

朋友~相似

定量评估

- 给定社交网，只考虑两种不同特征 (两种颜色)
- 能得到的信息
  - 节点数 (n)， 边数 (e)
  - 不同颜色节点的占比: p, q = 1-p
  - 两端节点相同的边数 (s)
- 两端节点相同的边在总边数中的占比越高，同质性越明显
- 衡量基准
  - 给定不同颜色节点的占比 (红 p，白 q)，随机情况下，任何一条边两节点相同的概率为 p^2+q^2

2.3 物以类聚人以群分

社会选择的机制

- 人们通过自然属性或社会属性的相似性、或价值观的相似性进行选择
- 原因
  - 个体具有“加入”的主动性和自主性
  - 个体具有“退出”的主动性和自主性
- 网络的同质性是一个动态过程

社会交往

- 社会网络的建构，无论是加入还是退出，都可能是一个主动的过程
- 被动参与也是一种机制
- 个体的兴趣与能力，或许不限于既有，可能会被诱发
- 拓展可以使选择，更多或是受到影响的

形成网络同质性的机制之一，是个体的主动选择

2.4 近朱者赤近墨者黑

社会归属网：三类闭包

- 三元闭包

  因人际关系产生的人与人之间的关系
- 社团闭包
  - 因人与事之间的关系产生的人与人之间的关系
  - 共同参与的事情越多，建立联系的可能性越高
- 会员闭包

    - 因人际关系产生的人与事之间的关系
    - 参与某件事的朋友越多，其被影响而参与的可能性就越高

形成网络同质性的机制之一，是个体之间的相互影响

2.5 朋友与相似 (大数据实验)

朋友间相似的原因

- 相似->朋友
- 朋友->相似

结论支持

- 两人之间的社会活动和相似性随时间同时提高
- 两人认识后，相似性进一步提高

数据集

- 反应随时间变化的大规模社会归属网
- 大规模：人多，社交聚点多
- 随时间变化：人和人之间，人和社交聚点之间

2.6 谢林模型

- 自然属性相同，选择相同
- 相互认识，相互影响，进而趋同

谢林模型示意

- 节点为居住者 (O, X)
- 约束条件
    - 每一个居住者都要与一定数量 (t) 的同类相邻
- 动态
    - 如果一个居住者发现自己的邻居数小于t，他就有兴趣搬家，以满足邻居数

社会意义

- 如果同质性是一个自然现象，则促进或阻止不同社会背景下的同质性，将会对社会发展产生重要影响

3. 小世界

3.1 小世界实验及其惊奇

问题

- 两个互不相识的人，如果想认识，中间需要经过几个人

假设

- 假设每个人都有熟人，熟人间没有芥蒂，可以交往。因此不曾有连接的两个人之间，如果要建立联系，中间人的数量应该不多
- 每个人的确有熟人，不过，不同类型或阶层的熟人之间，不会有交往。因此不曾有连接的两个人之间，不可能建立连接

实验

- **设计**
  - 选择一个随机起点，观察需要经过几个中间人，能够到达目标点
- **规则**
  - 参与者只能将信件转发给能直呼其名的熟人，并请他继续转发；如果一个参与者不认识目标收信人，则他不能直接将信寄给他
  - 参与者需4力争让信件能尽早到达目的地
- **结果**
  - 平均中间人数为5

小世界问题

- 在Milgram的研究之前，人们感觉到世界很小，却没有证据

小世界现象

- Milgram的研究证明
  - 世界是小的，社会网络中包含丰富的短路径
  - “自动寻找”短路径，“有意识的转发”能“自动地”找到这些短路径

3.2 小世界现象的普遍性

后续研究

- 不少重复研究，包括Milgram自己，中间人数只有5-7人
- 情景拓展
  - 运用书信所做的研究具有重复性，电子邮件呢？
  - 发现人数依然只有5-7人

网页之间的“社交”

- 全球互联网超过几百亿网页，比人口总数多
- 问题：网页之间也有小世界现象吗？
- 没有关系的两个网页之间的直径为18.59次点击

Summary

- 通过熟人送文件，在不断地检验小世界现象的存在
- 加上种族因素，小世界现象依然存在
- 通过熟人转发电子邮件，依然有小世界
- 网页之间的社交，也有短路径

3.3 小世界的基本模型

人类社会的小世界现象

- 社会网络中两节点间包含丰富的短路径，任意两节点间存在短路径的概率很高
- 短视搜索能够有效地找到这些短路径
  - 短视搜索：在到达目标节点的过程中，每一步只能看到邻居节点
- 对于十分稀疏的社会网络来说，这并不是必然的

问题

- 为什么社会网络具有这样的性质，他们源于社会网络的哪些基本原理？
  - 可以证明，完全随机的网络没有这样的性质
- 能否依据社会网络的某些基本原理，构建出反应这种性质的网络模型，说明这种现象的必然性？

形成社会网络的基本力量

- **同质性**
  - 共同朋友、邻里关系、同学、同事、兴趣
  - 对应社会网络中大量的“三角形”
- **弱联系**
  - 偶然的原因，认识的“远程”朋友
  - 对其所在的圈子不一定熟悉

Watts-Strogatz模型

- 想象大量节点排布成均匀网格状，每两点之间有一个“网格距离”
- 有许多“三角形”和少数随机的“远程边”
- 连接近邻：确定性，连接远程：随机性
- 两个距离
  - 网格距离
  - 网络距离：最短路径长度
- 意义和局限
  - 体现了同质连接和弱关系连接，可以看成是现实社会网络的合理近似
  - 可以证明，任意两节点之间存在短路径的概率很高，但短视搜索的路径太长

3.4 小世界的精细模型

短视搜索

- 每一个节点有一个特征，任何两个节点之间的特征可以谈差别 (距离)
- 每个节点都知道目标节点的特征，也知道自己和自己邻居节点的特征
- 搜索过程可看成是信息传递的过程，节点将信息传给离目标节点距离较近(差别较小)的邻居节点

需要的结构特征

- 两个节点无论相距多远，都要有机会很快接近
- 两个节点的距离越近，存在直接连接的机会越大

Watts-Strogatz-Kleinberg模型

- 在Watts-Strogatz模型的基础上，让两个节点之间存在随机边的概率与它们的网格距离的某个幂次 (q, 聚集系数) 成反比

$$p \propto \frac{1}{d^q}$$

- q较小，随机边倾向于较远；q较大，随机边倾向于较近
- Watts-Strogatz模型对应于 q = 0
- 理论结果：当 q = 2时，分散搜索达到最佳效果

3.5 WSK模型中优化参数的大数据验证

社会网络中结合地理距离的节点相对排名

- 对社会网络的建立而言，地理范围内的人数比距离具有更实质性的意义
- 可以看成是节点在地理上均匀分布时区域范围概念的一中推广，"排名"与"距离"有对应关系
- 使我们可以一般性地处理节点在地理上分布不均匀的问题
- "在均匀地理分布情形，一个节点在任一距离上的朋友数量在同等距离节点总数中的占比随距离平方递减 ($\frac{1}{d^2}$)" 等价于 "一个节点在任一排名上的朋友数量在同等排名节点总数中的占比随排名递减 ($\frac{1}{r}$)

3.6 核心-外围结构

核心-边缘结构模型

- 地位较高的人，都连接在一个密集连接的核心
- 地位较低的人，都分散在网络的外围

理论与现实

- 理论上
  - 处在网络结构中的节点，不同的节点如果有相同的聚集系数，其被连接到的概率是一样的
- 现实中
  - 寻找地位较低的人会更加困难
  - 媒体寻人较个体寻人有更高的成功率
  - 处在结构洞位置上的人被找到的概率大

社会意义

- 网络结构的社会属性与网络结构同样重要
- 具有相似网络结构，却有着不同社会属性的网络，在现实社会中具有不同的"可连通性"

4. 万维网结构、链接分析与网络搜索

4.2 将Web看成有向图

领结

- 万维网包含一个超大连通分量SCC

    - 链入，链出，卷须，游离

获得有向图的组成部分

- 简化：只关心SCC, IN, OUT部分
- 假设某一个节点一定在SCC中
- 正图和逆图分别BFS，得节点集合FS、BS
- SCC = FS ∩ BS, IN = BS - SCC, OUT = FS - SCC

4.3 中枢与权威

搜索引擎基本问题

- 显示最重要的网页
- 快速

反复改进原理

与关键词字面上相关的网页 -> 它们指向的网页，得到的票数表示一定认可度 -> 反过来评估推荐者的分量 -> 重新评估网页

HITS算法

- 计算网页的权威值(auth)和中枢值(hub)
- 输入：一个有向图
- 初始化：对于每一个节点p, auth(p) = 1, hub(p) = 1
- 利用中枢值更新权威值
    - 对于每一个节点p，auth(p)等于指向p的所有节点q的hub(q)之和
- 利用权威值更新中枢值
    - 对于每一个节点p，hub(p)等于p指向的所有节点的q的auth(q)之和
- 重复上述两步k次

归一化与极限

- 数值随迭代次数递增，auth和hub值的意义在于相对大小
- 在每一轮结束后做归一化处理，归一化结果收敛

4.4 PageRank

基本要领

- 每一个节点将自己的值均分给出向邻居
- 每个节点将从入向邻居收到的值加起来
- 迭代，结果收敛

PageRank算法

- 输入：一个有n个节点的有向图，设所有节点的PageRank初始值为1/n
- 选择操作的步骤数k
- 迭代k次

    - 每个节点将自己当前的PageRank值通过出向链接均分传递给所指向的节点
        - 若没有出向链接，则认为传递给自己
    - 每个节点以入向链接获得的(包括自己)所有的值之和更新它的PageRank

4.5 同比缩减与等量补偿

同比缩减

- 每次运行基本的PageRank更新规则后，将每一节点的PageRank值都乘以一个小于1的比例因子s, 0<s<1, 经验值在0.8-0.9

统一补偿

- 在每一节点的PageRank值上统一加上(1-s)/n

即维持了$\sum PR = 1$，也防止PageRank集中到某几个网页上

随机游走

- 从一篇随机选择的网页开始，随机选择其中的链接浏览到下一篇网页，并不断如此运行，称为“随机游走”
- 经过k步到达X网页的概率等于运行PageRank基本算法k步得到的值

5. 博弈论

5.1 何为博弈

博弈三要素

- 参与人 (player)
- 策略集 (strategy)
    - 每个参与人有一个策略集
    - 策略组：每个参与人出一个策略构成的策略组合
    - 对应每个策略组，每个参与人有一个回报
- 回报 (payoff)

博弈的要义

- 相互影响

博弈论中的博弈不总是讲输赢

收益矩阵

- 表达博弈的一种直观方式

博弈论的关切

- 在“理性人”等基本假设下，博弈(作为一个整体)的结果、走向、发展趋势、哪些策略组会被人们采用

博弈推理的假定

- 自己的回报是每个参与人关心的唯一因素
- 参与人都是“理性人”，即只要可能，总是选择有更好回报的策略
- 每个参与人都对博弈结构完全了解

基本类型

- 囚徒困境
- 协调博弈
- 鹰鸽博弈
- 零和博弈

5.2 何为博弈的解

- 博弈的解释“稳定的策略组”，要求**其中任何参与人不可能通过单方面改变策略而获得更好回报**
- 不是所有博弈都有解
- 博弈均衡的概念

5.3 博弈的求解

严格占优策略

- 对一个参与人来说，若存在一个策略，无论另一个参与人选择何种策略，该策略都是严格最佳的选择，则这个策略就称为是前者的严格占优策略
- 按照博弈推理假设，参与人将选择严格占优策略

严格最佳应对

- 第一个参与人有严格占优策略，第二个参与人没有严格占优策略，则第二个参与人会选择对应于第一个参与人的严格占优策略的占优策略 (最佳应对)

解的可能情况

- 占优策略+占优策略
- 占优策略+最佳应对
- 最佳应对+最佳应对

5.4 纳什均衡

纳什均衡

- 互为最佳应对的策略组
- 有助于缩小预测范围，不一定给出唯一的预测
- 具有有限参与者和有限纯策略集的博弈一定存在纳什均衡 (包括混合策略均衡)

混合策略

- 引入随机性，考虑参与人将以一定的概率在不同策略间进行选择，一个概率对应一个策略 (混合策略)
- 此时，选择策略就是选择概率
- 博弈矩阵中给出的选项称为纯策略
- 收益为各个纯策略在概率意义下的期望

- 概率策略的纳什均衡满足无差异原理

5.5 博弈的解与社会福利

社会福利

- 一个策略组对应的回报的总合

社会最优

- 社会福利最大的策略组合
- 均衡是博弈的解，但不一定是社会最优的
- 均衡与社会最优一致的系统是理想系统

5.6 Question

mooc第7题

6.网络流量博弈、拍卖、匹配市场

6.1 交通网络上的博弈

- 结构对均衡的影响
- **布雷斯悖论**: 意在增加社会福利的行动，反而增加了社会成本
- 现实生活中参与一个博弈可能是无形中的

6.2 布雷斯悖论的一般性

- 布雷斯悖论是社会政策、公共服务中可以意会到的现象
- 对布雷斯悖论的科学论证与推广，还是一个可探索的领域

6.3 拍卖的意义及其形式

拍卖

- 普遍存在的经济互动形态
  - 模式简单，互动复杂
- 参与者：参加拍卖的人，买、卖双方
- 策略：出价
- 收益：对物品的估值，或为0
- 均衡：在该状态下所有参与者的策略互为最佳应对

拍卖的形式

- 英式拍卖
  - 竞买者逐步加价，直到最后只有一个人投标
  - 出价最高者获得拍卖品

    - 古董、艺术品
- 荷兰式拍卖
    - 出价者从一个较高价格开始逐步降价，直到有人愿意购买
    - 农产品的交易、荷兰的鲜花
- 首价密封拍卖
    - 在某一个约定的时间同时公开所有投标人的报价
    - 最高的、或最低的报价
    - 不知道他人的出价
    - 招标
- 次价密封拍卖
    - 在某一个约定时间同时公开所有投标人的报价
    - 支付价格：次高的、或次低的报价
    - 互联网门户或搜索引擎公司出售的网页上的广告位

6.4 拍卖中的博弈与占优策略

采用次价密封拍卖的规则，对参拍人来说，按照估值报价是占优策略

- 估值不等于"我愿意出的钱数"
- 估值等于"我绝不接受高于这个钱数

首次密封拍卖没有这个性质

6.5 匹配问题

- 二部图中的完美匹配：一组边覆盖了所有节点，没有节点冲突
- 供需方个数分别为N，则有N!个配置方案
- 社会最优：参与人收益总和最大

6.6 匹配市场问题的解

- 达到完美匹配前，卖方价格不断增加
- 最优配置对应估值矩阵中使求和最大的不同行不同列的那些元素
- 清仓价格 -> 最大的∑估值，∑收益(差价) = ∑估值 - ∑价格
- 清仓价格：每个需方都能无冲突地得到最大的收益

7. 搜索引擎广告位的定价

7.1 搜索引擎的广告位销售问题

付费方式

- 按展示计费
- 按行动计费
    - CPC广告 (Cost-per-click): 每次点击的费用
- 按销售计费

分配问题

市场匹配的基本原则: 价格可以使商品的分配分散化，产生最优的社会分配

每个广告位的价值

广告位价值 = 点击率 * 点击价值

定价问题

- 匹配市场: 前提是知道广告主的估值
- 竞价拍卖: 无从得知广告主的估值

7.2 多广告主、多广告位的匹配

匹配市场基本要素

- 商品有不同售价
- 买家对每个商品有估值
- 匹配原则: 最大化回报
- 完美匹配: 最优分配
- 市场清仓价: 完美匹配价格

按照点击价值的高低配置对应的广告位达到社会最优

7.3 GSP: 次价拍卖方式的直接推广

GSP

- 真实报价不是占优策略
- 存在多重均衡，一定存在
- 不一定是社会最优分配
- 实际中用得比较多
    - 广告主容易懂
    - 性质复杂

7.4 VCG: 次价拍卖方式的优化推广

1. **构建一个最优匹配**

    点击率r_i，估值v_j

    最大估值总和为$\sum v_{ij} = \sum r_i * v_j$
2. j为i支付的VCG价格

    其余人估值总和为$\sum_1$，去除j后最优匹配时其余人估值总和为$\sum_2$

    则j为i支付的VCG价格为$\sum_2 - \sum_1$

特性

- 社会最优、均衡
- 真实报价是占优策略
- 搜索引擎公司获得收入可能高于或低于GSP

7.5 VCG最优化证明

最高估值总和: $V_B^S = \sum v_{ij}$

i匹配给j后其余部分估值总和: V_{B-j}^{S-i}

除去j后，其余部分估值总和: V_{B-j}^{S}

VCG价格$p_{ij} = V_{B-j}^{S} - V_{B-j}^{S-i}$

设虚假报价为v'_j

则有$V_B^S >= V_B^{S'} = r_h * v'_j + V_{B-j}^{S-h}$

VCG占优策略前提

- 没有合谋、作弊等行为
- 广告位价格只与点击率有关

8. 关系的平衡

8.1 三节点敌友关系网络的结构平衡

关系属性与平衡

- 一种平衡状态
  - 如果实体间总是表现为一种关系状态
  - 或实体的所属实体也总是表现为一种关系状态
- 如果失衡
  - 则会出现促使平衡恢复的力量

三节点间的结构平衡

- 平衡：稳定，没有改变的力量

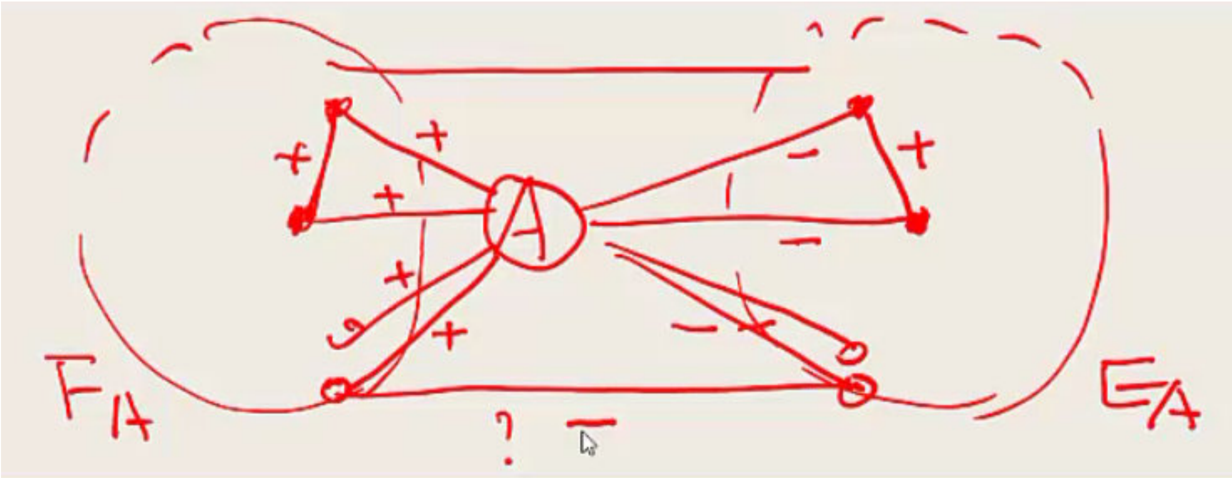
8.2 多节点网络的结构平衡——平衡定理

平衡结构

- 一个标注(+/-)完全图结构是平衡的，当且仅当它包含的所有三角关系都是平衡的
- n阶完全图的三角形个数: C_n^3

平衡定理

- 所有节点两两都是朋友关系
- 节点可以被分为两组，组内的节点两两都是朋友关系，而组间的节点两两都是敌人关系



弱平衡网络

- 标记的完全图中不存在(+, +, -)三角关系的网络
- 节点可分为若干组，组内均为朋友，组间均为敌人

8.3 社交网络结构对人际关系轻重的影响

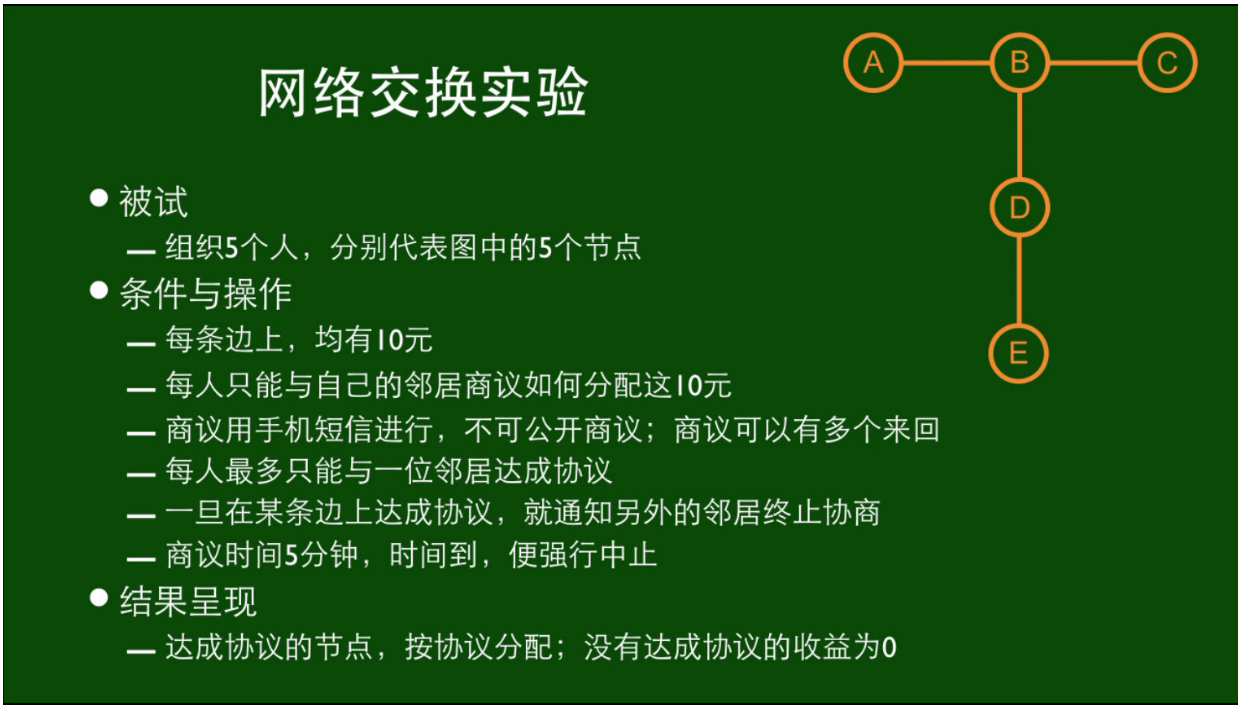
权力

- 迫使他人服从自己意志的能力
- 关系性、情景性

社会网络中的权力

- 其他节点对其具有依附性
- 具有排他性、饱和性、中心性
- 处于结构洞位置上的节点

8.4 网络交换实验



实验思路

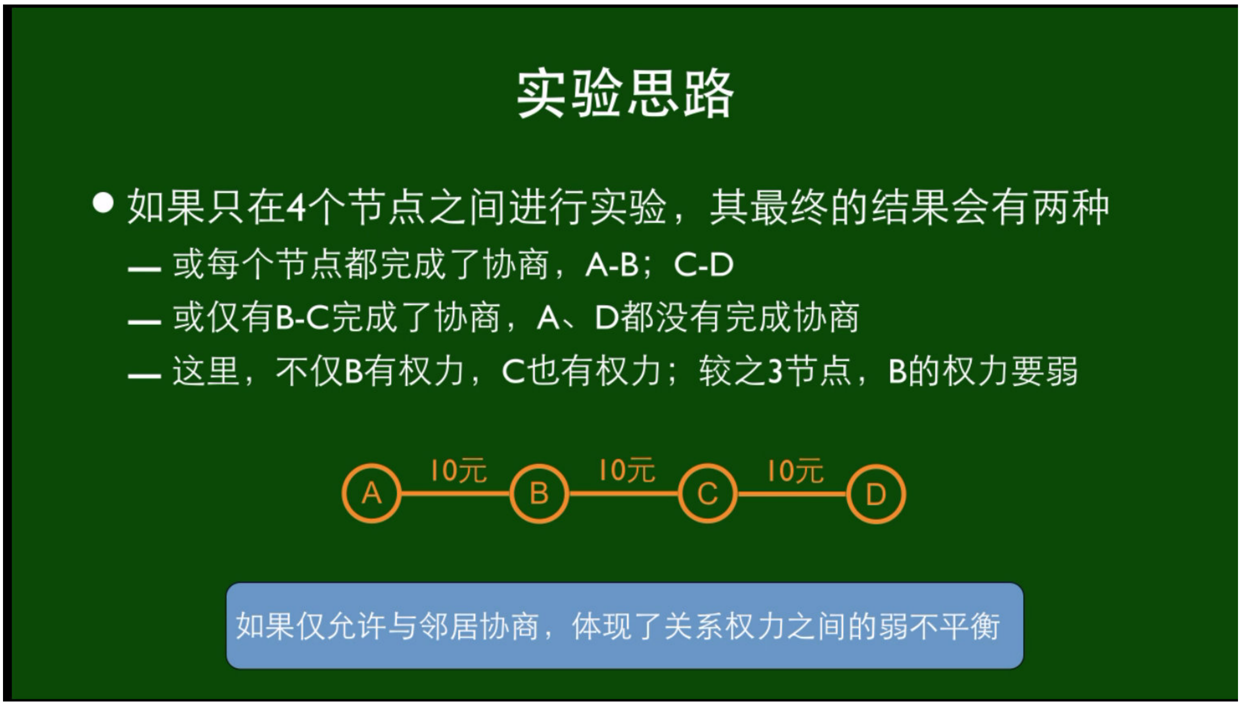
- 可以将前面的图，进行拆分，直至2节点路径
- 如果只在2个节点之间进行实验，其最终的结果会是
  — 两人平分边上的10元钱

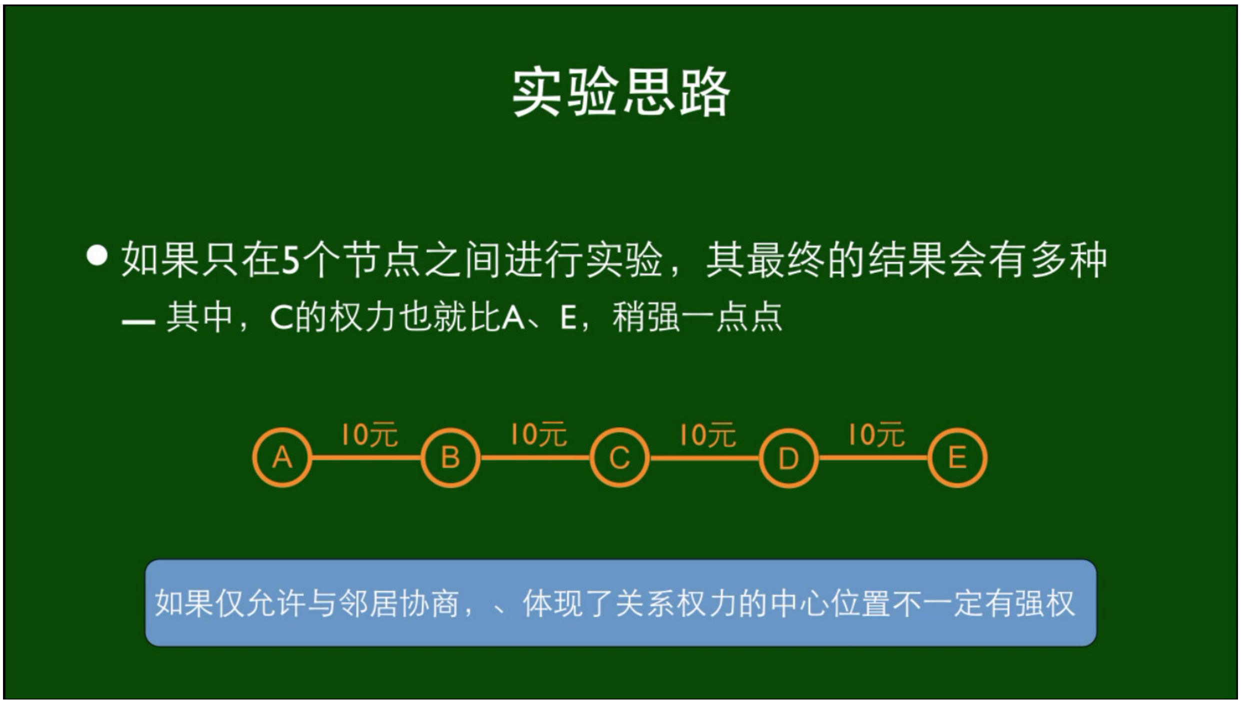
平分，就意味着两个人之间的关系权力是平等，或对称的

实验思路

- 如果只在3个节点之间进行实验，其最终的结果会是
  — B会得到绝大多数
  — 如果改变规则，允许B同时和两个协商，结果是B会与A、C 平分

如果仅允许与邻居协商，B就有垄断性权力，体现了权力间的极端不平衡





8.5 稳定结果

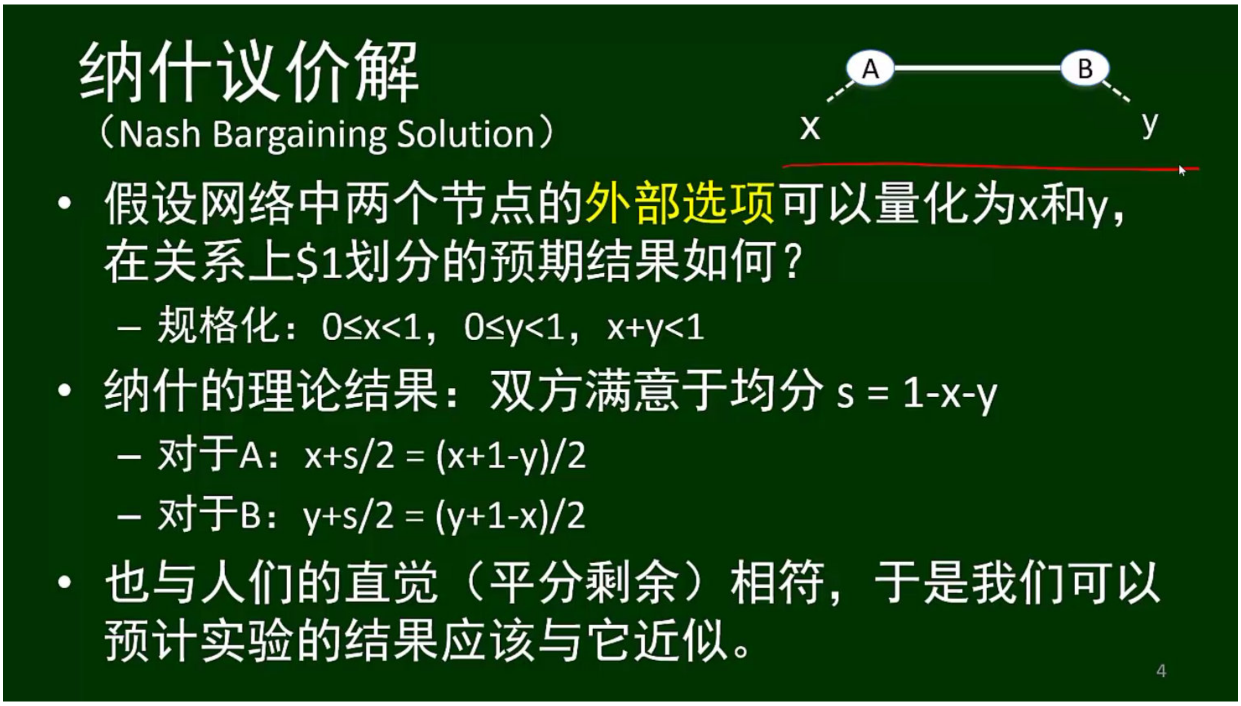
结果

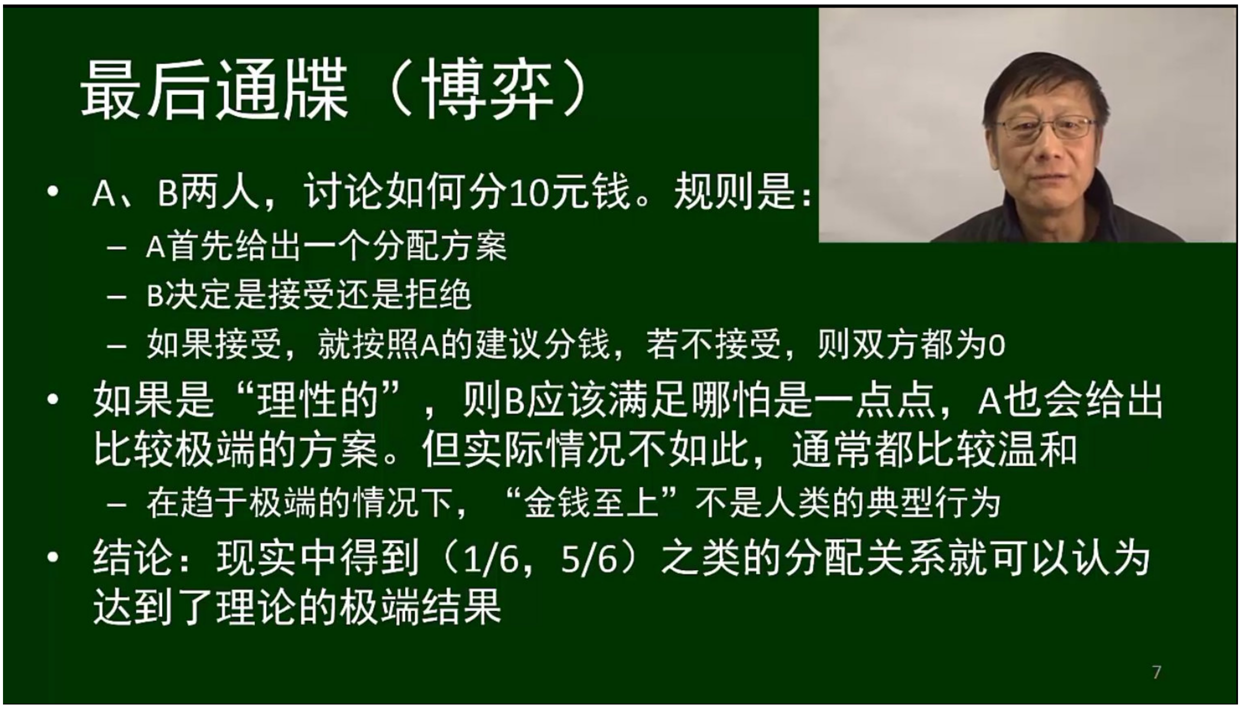
- 给定一个图，“结果”是其中的一个匹配(一个节点无冲突的边子集)，以及每个节点在[0, 1]区间的一个赋值，满足：
  - 如果节点u和v对应匹配中的一条边，则它们赋值之和为1
  - 如果节点u不涉及到匹配中的任何边，则它的赋值为0

稳定结果

- 不稳定因素：不在结果中的一条边，其两端节点的价值之和小于1
- 稳定结果：不存在不稳定因素的结果

8.6 纳什议价解





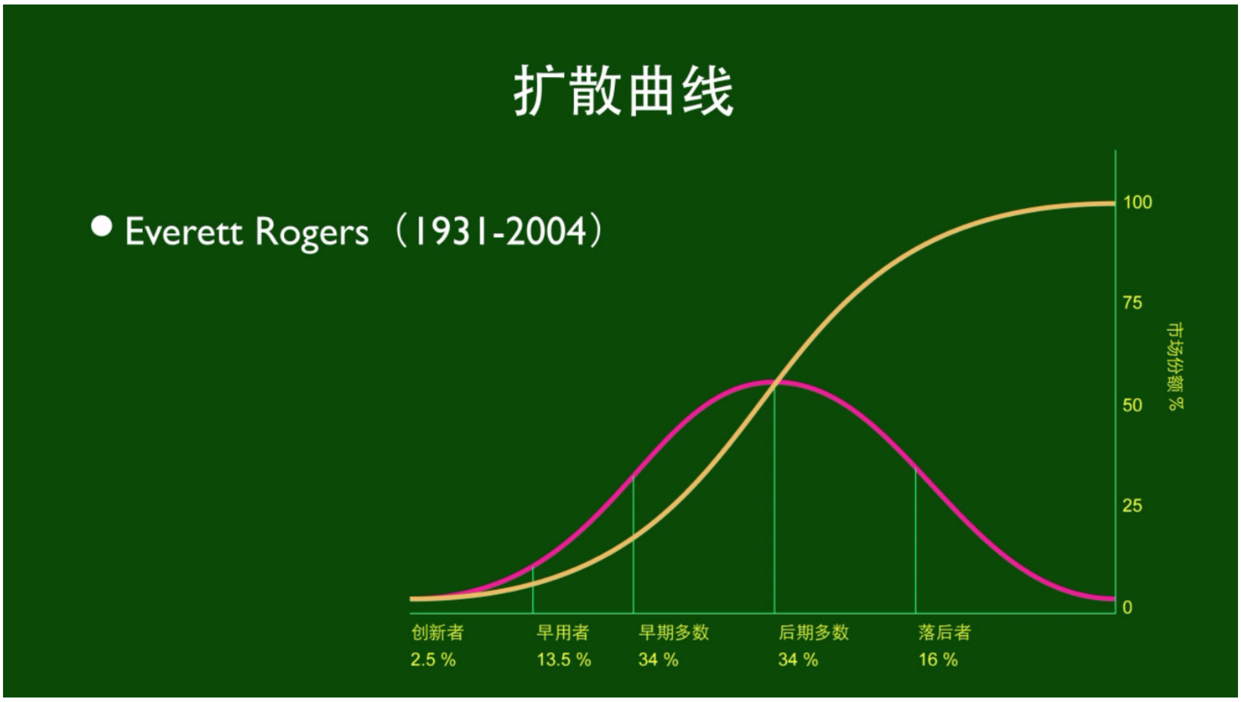
8.7 平衡结果

- 结果中匹配的每条边上的价值划分都满足纳什议价解
- 稳定结果不一定是平衡结果，平衡结果一定是稳定结果

9. 新事物在网络中的扩散

9.1 新生事物的社会传播

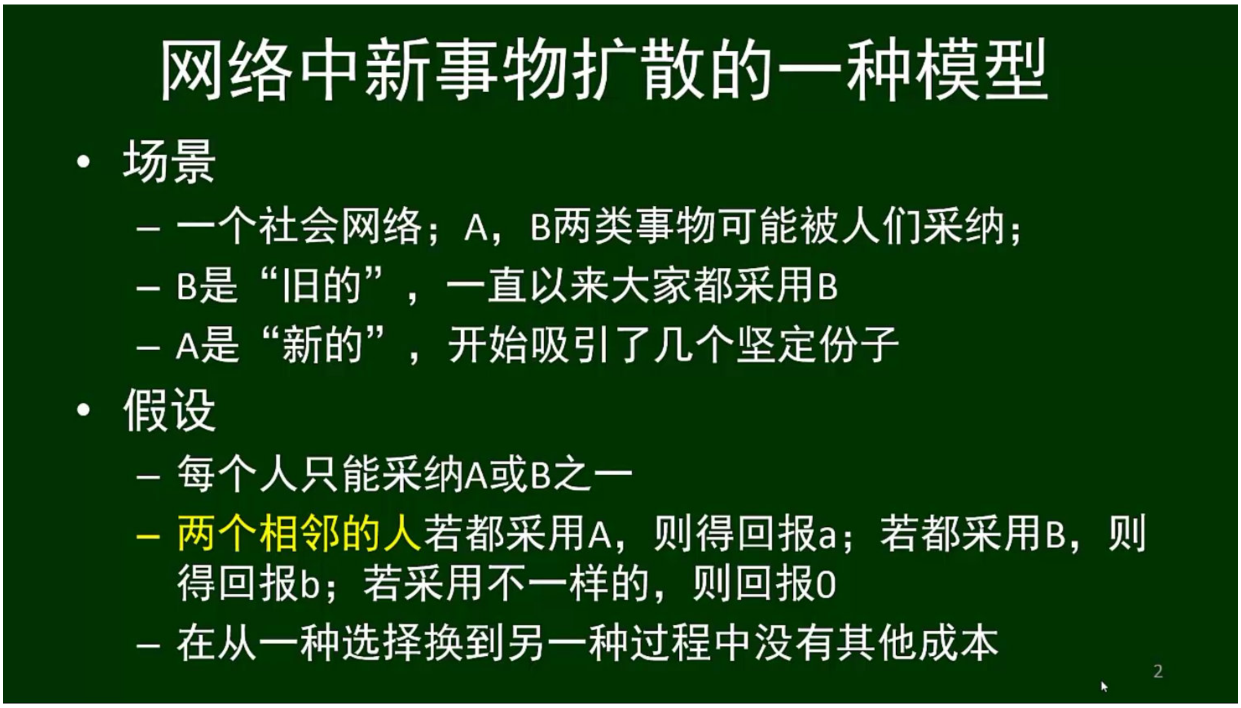
扩散曲线



特点

- 时间轴上呈现出S曲线
- 扩散与采用者在网络结构中的位置有关

9.2 一种网络级联扩散模型



表达为一个博弈

- 在一条边上的博弈
  - 如果v和w都选择A，它们分别得到回报a>0；
  - 如果它们都选择B，分别得到回报b>0；
  - 如果它们选择不同的选项，那么都得到回报为0

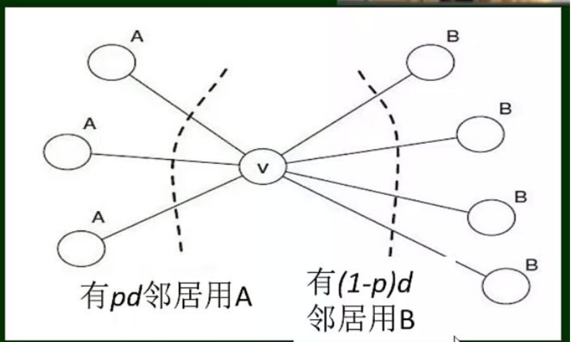
选择A或B？ 选择A或B？

		w A	w B
v	A	a, a	$0, 0$
	B	$0, 0$	b, b

网络节点v的决策门槛

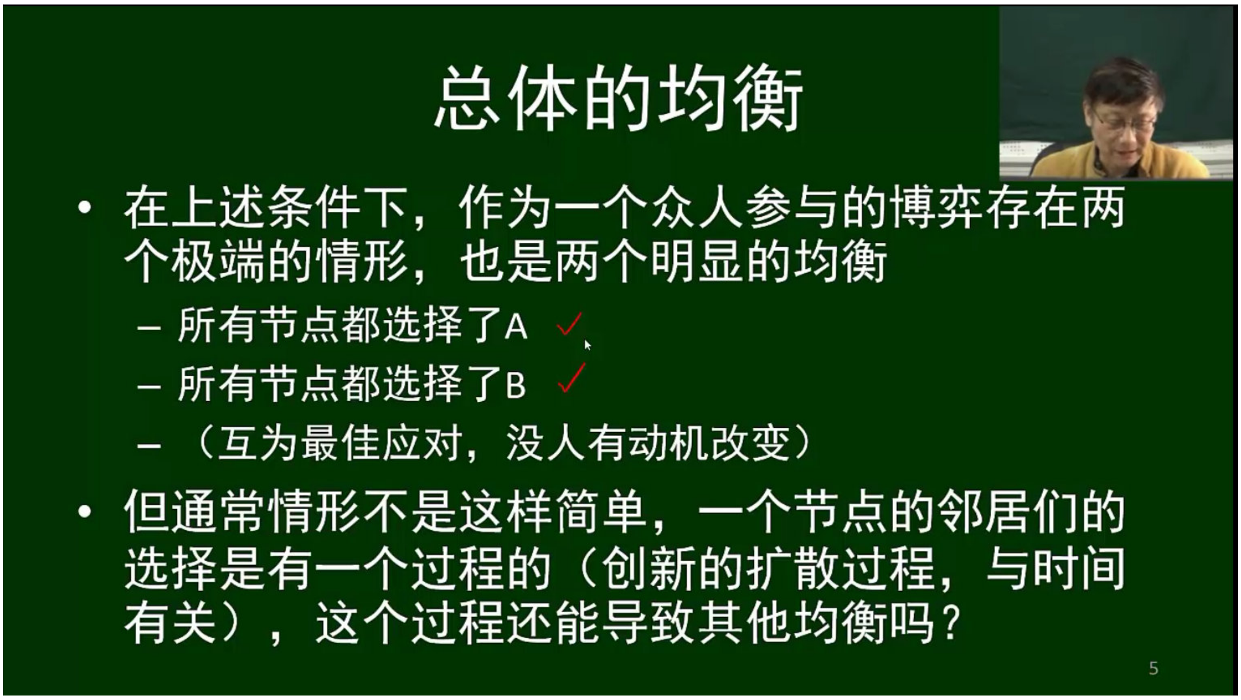
- 设v有d个邻居，在某一时刻，若占比p的邻居选A，占比1-p的邻居选B
- v选A的回报：*pda*

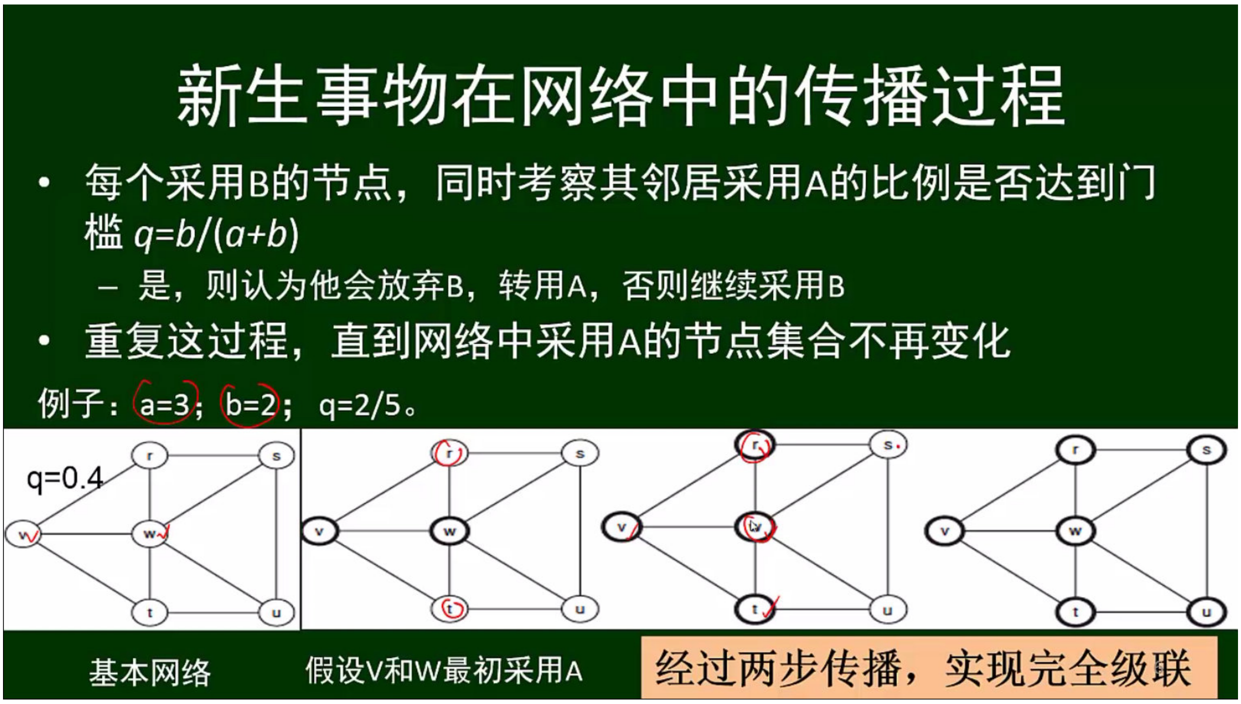
  选B回报：*(1-p)db*
- 如果 *pda* ≥ *(1-p)db*，即若 *p*≥*b*/(*a*+*b*)，则选A好；否则，选B更好。



$$\frac{b}{a+b} = q$$

门槛

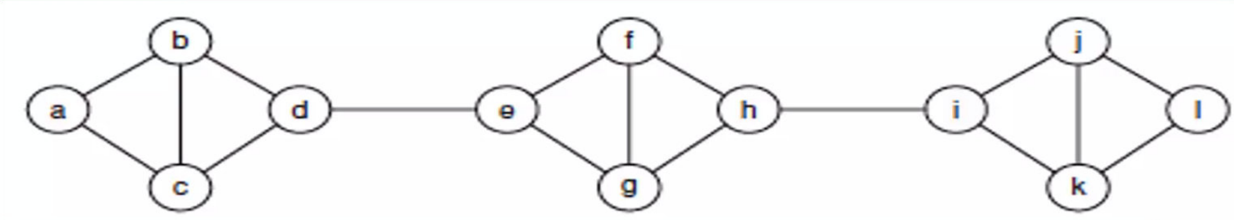

总体的均衡
• 在上述条件下，作为一个众人参与的博弈存在两个极端的情形，也是两个明显的均衡
– 所有节点都选择了A
– 所有节点都选择了B
– （互为最佳应对，没人有动机改变）
• 但通常情形不是这样简单，一个节点的邻居们的选择是有一个过程的（创新的扩散过程，与时间有关），这个过程还能导致其他均衡吗？
5


新生事物在网络中的传播过程
• 每个采用B的节点，同时考察其邻居采用A的比例是否达到门槛 q=b/(a+b)
– 是，则认为他会放弃B，转用A，否则继续采用B
• 重复这过程，直到网络中采用A的节点集合不再变化
例子：a=3；b=2；q=2/5。
q=0.4
r
s
v
w
t
u
基本网络
假设V和W最初采用A
经过两步传播，实现完全级联

9.3 何时形成完全级联

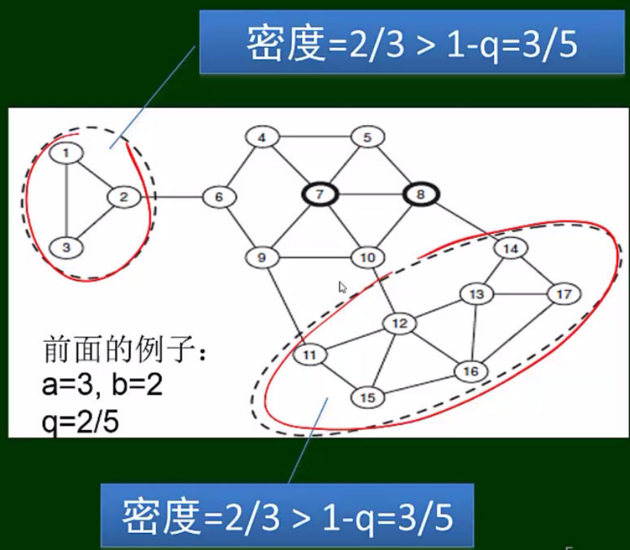
刻画阻挡级联的因素——聚簇

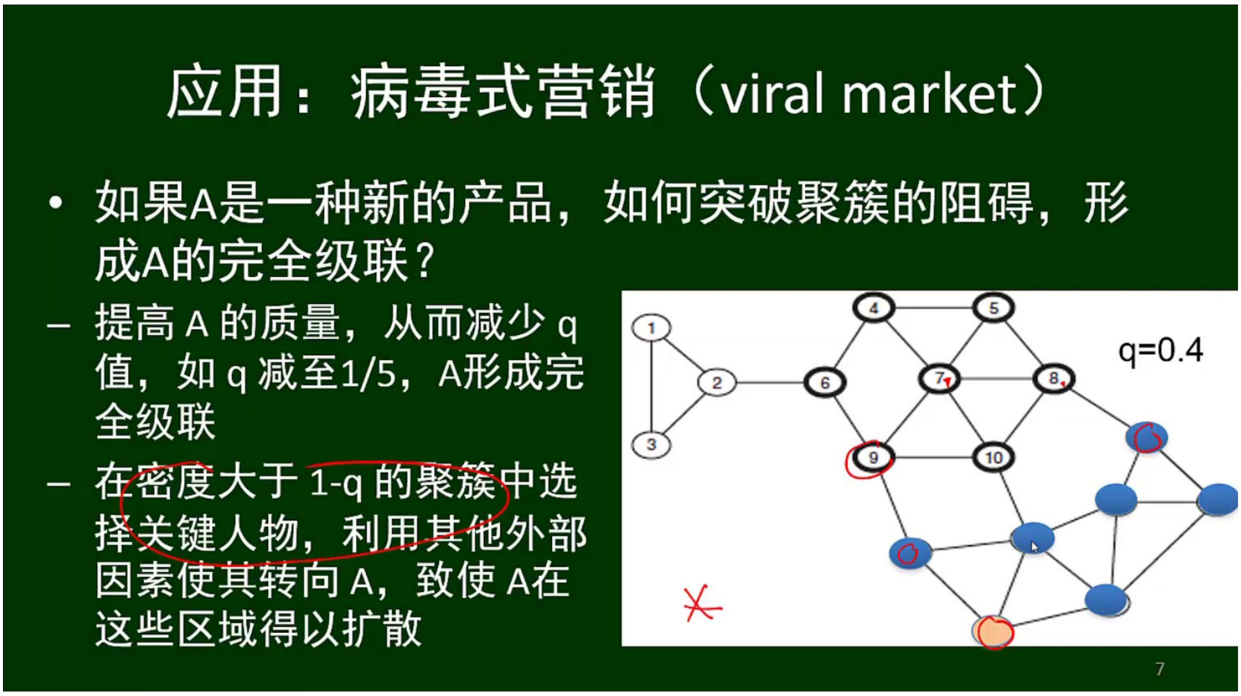
- 聚簇（“抱团”）：称一个节点集合为密度为 r 的聚簇，若其中每个节点至少有占比为 r 的网络邻居也属于这个节点集合
  - 每个节点的好友出现在这个集合中的比例

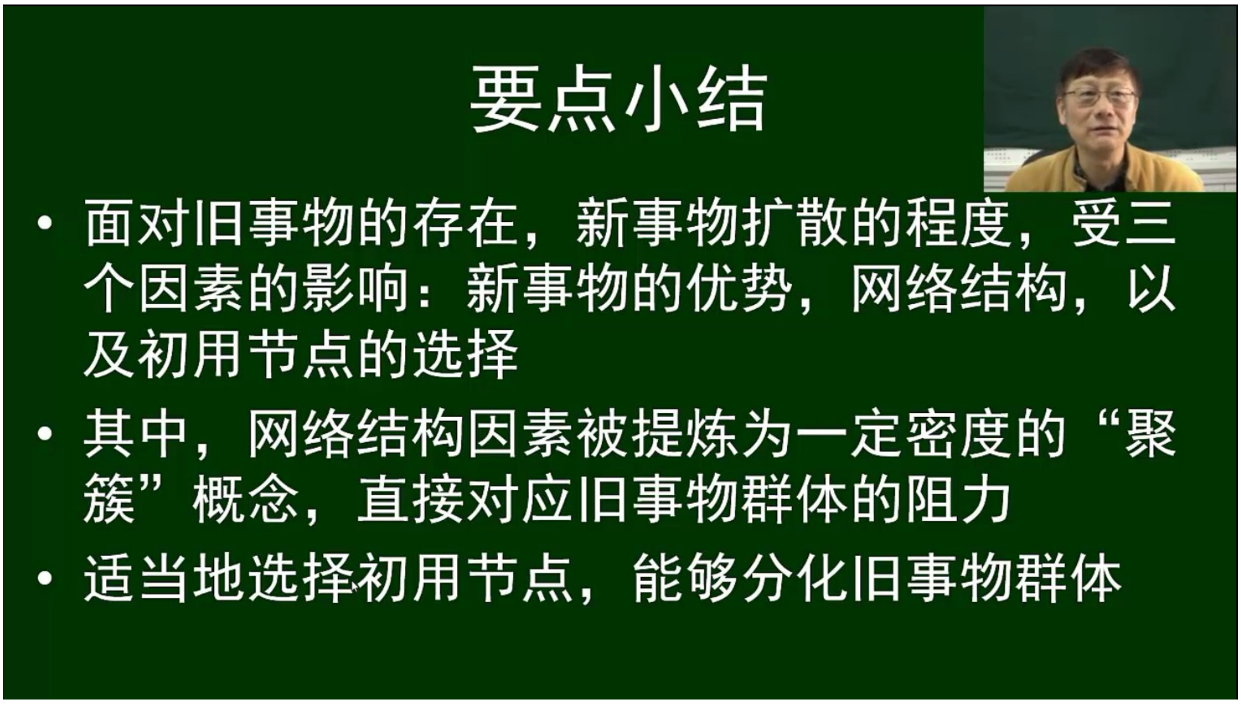


聚簇和级联的关系（定理）

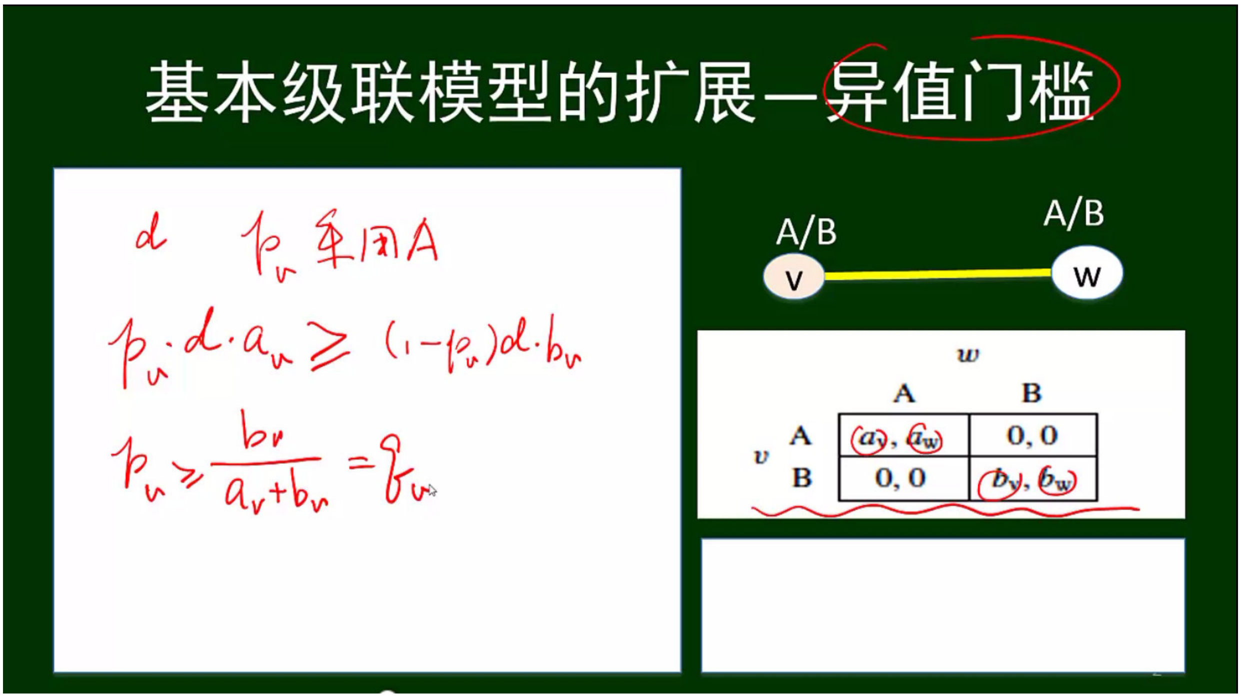
- 设网络中一个初用节点集采用 A，剩余网络的其他节点采用 B，且它们改用A的门槛值为 q
- 如果剩余网络中包含一个密度大于 1-q 的聚簇，则这个初用节点集不能形成A的完全级联
- 而且，如果一个初用节点集不能形成一个完全级联，则剩余网络一定包含一个密度大于 1-q 的聚簇

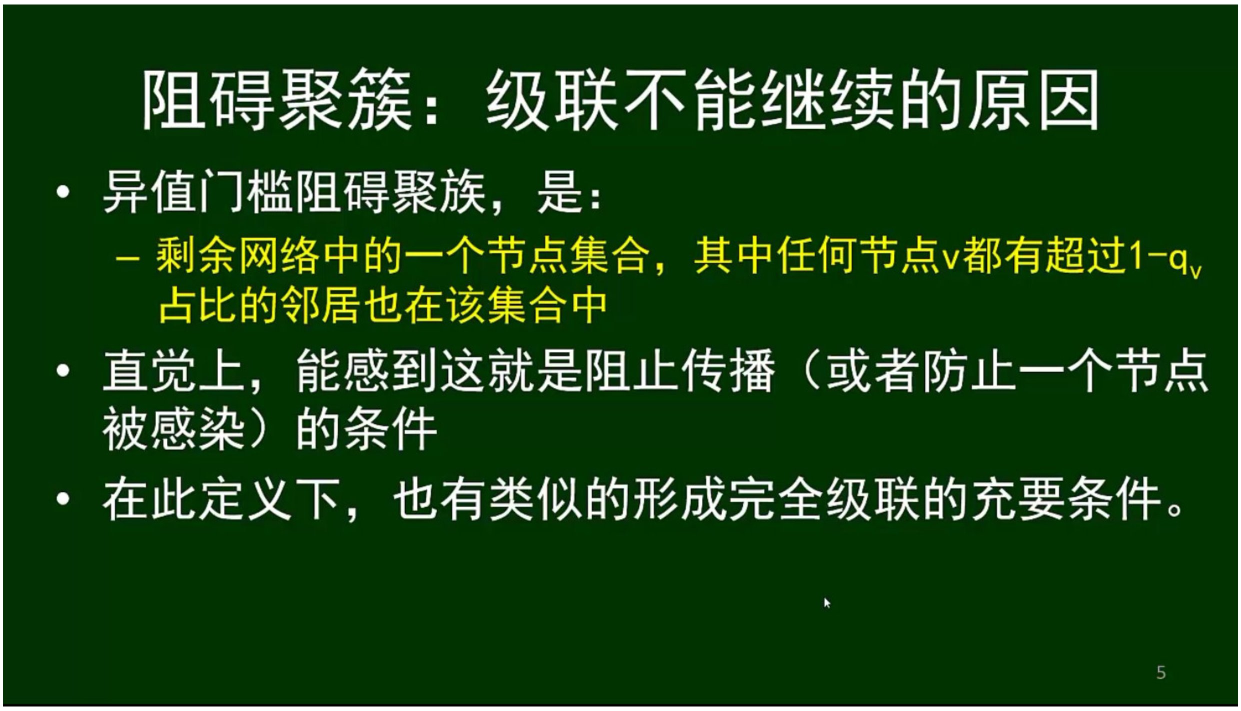



应用：病毒式营销（viral market）
• 如果A是一种新的产品，如何突破聚簇的阻碍，形成A的完全级联？
– 提高 A 的质量，从而减少 q 值，如 q 减至1/5，A形成完全级联
– 在密度大于 1-q 的聚簇中选择关键人物，利用其他外部因素使其转向 A，致使 A在这些区域得以扩散
q=0.4
1
2
3
4
5
6
7
8
9
10
7


要点小结
• 面对旧事物的存在，新事物扩散的程度，受三个因素的影响：新事物的优势，网络结构，以及初用节点的选择
• 其中，网络结构因素被提炼为一定密度的“聚簇”概念，直接对应旧事物群体的阻力
• 适当地选择初用节点，能够分化旧事物群体

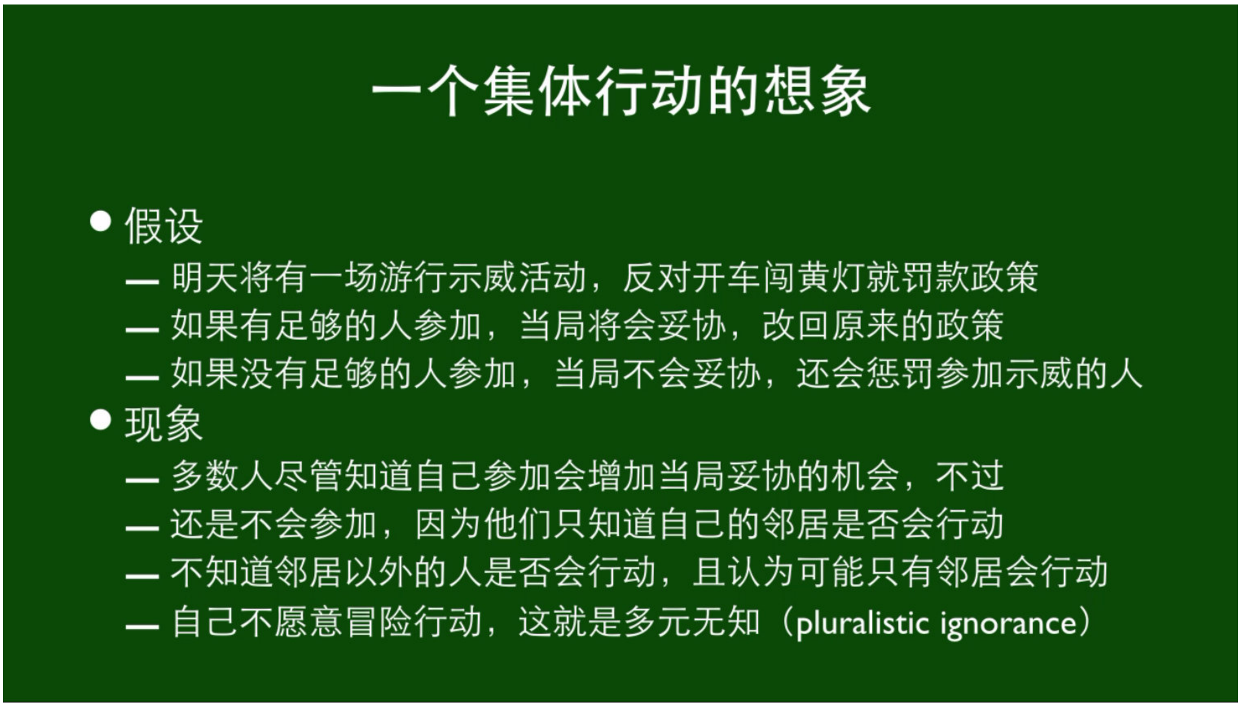
9.4 异值门槛下的扩散





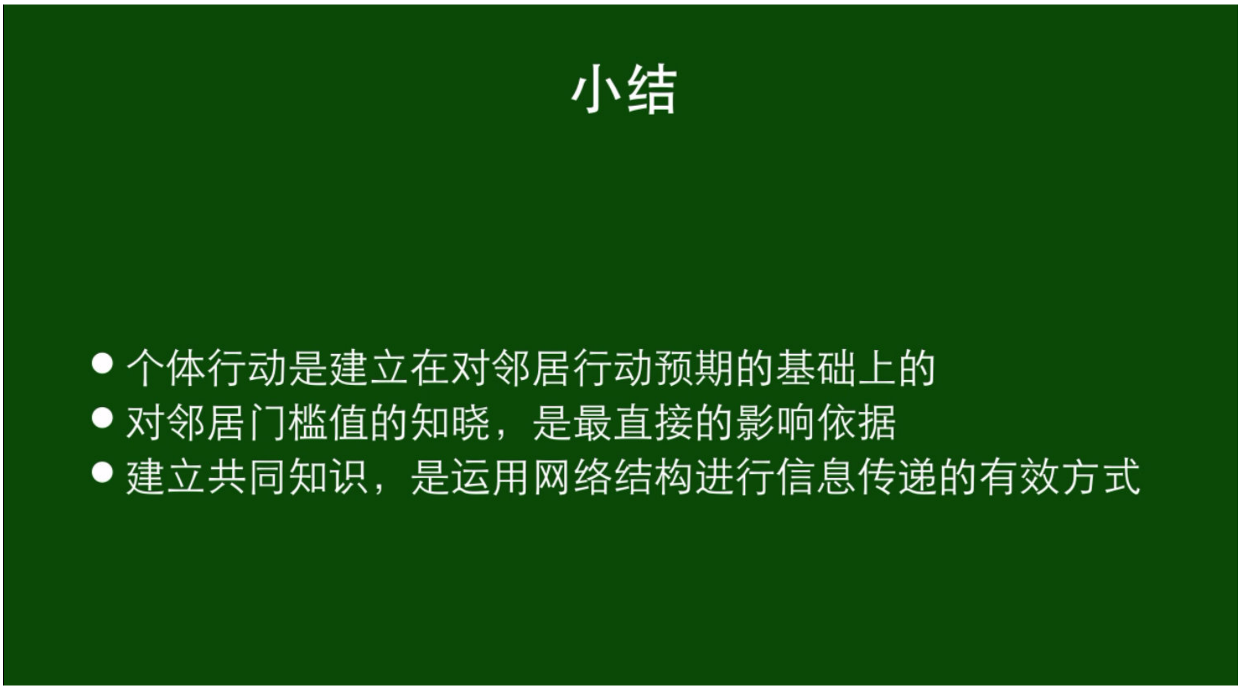
9.5 公共知识与集体行动

多元无知



沉默的螺旋

- 如果人们觉得自己的观点是公众中的少数派，他们将不愿意传播自己的看法；而如果他们觉得自己的看法与多数人一致，他们会说出来



10. 从众行为和事物的流行性

概率基础

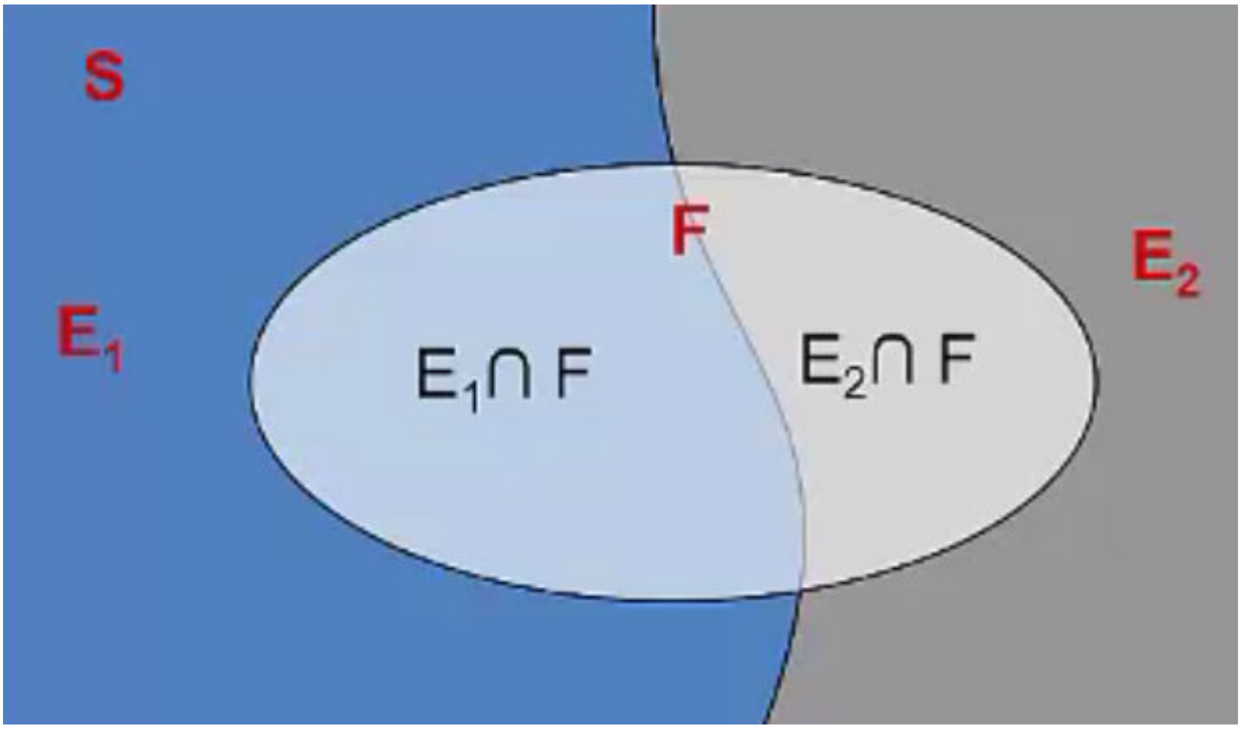
条件概率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Bayes'定理

$$P(E_1|F) = \frac{P(E_1 \cap F)}{P(F)} = \frac{P(E_1) \cdot P(F|E_1)}{P(E_1) \cdot P(F|E_1) + P(E_2) \cdot P(F|E_2)}$$

其中S为样本空间，$E_1 \cap E_2 = 0, E_1 \cup E_2 = S$, F为S中任意事件



幂律分布

极端不平衡性

- 概率最高的事件不反应平均行为
- 容易看到偏离均值很多的事件

无标度性

10.1 集群

信息效应

- 个体对信息不确定而选择与群体保持一致的模仿行为

直接受益效应

- 模仿别人直接影响个体回报和利益

信息级联产生的环境

- 个体的私有信息
- 依次决策

贝叶斯定律验证

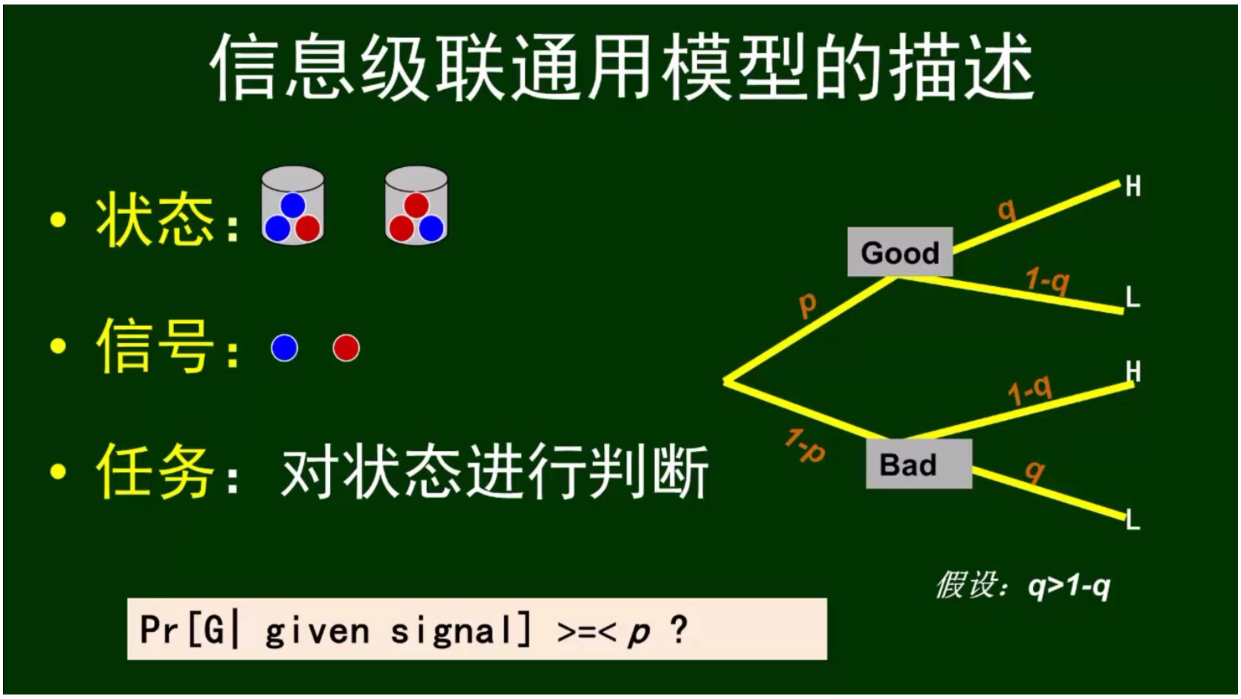
- 信号个数相差为2时产生级联

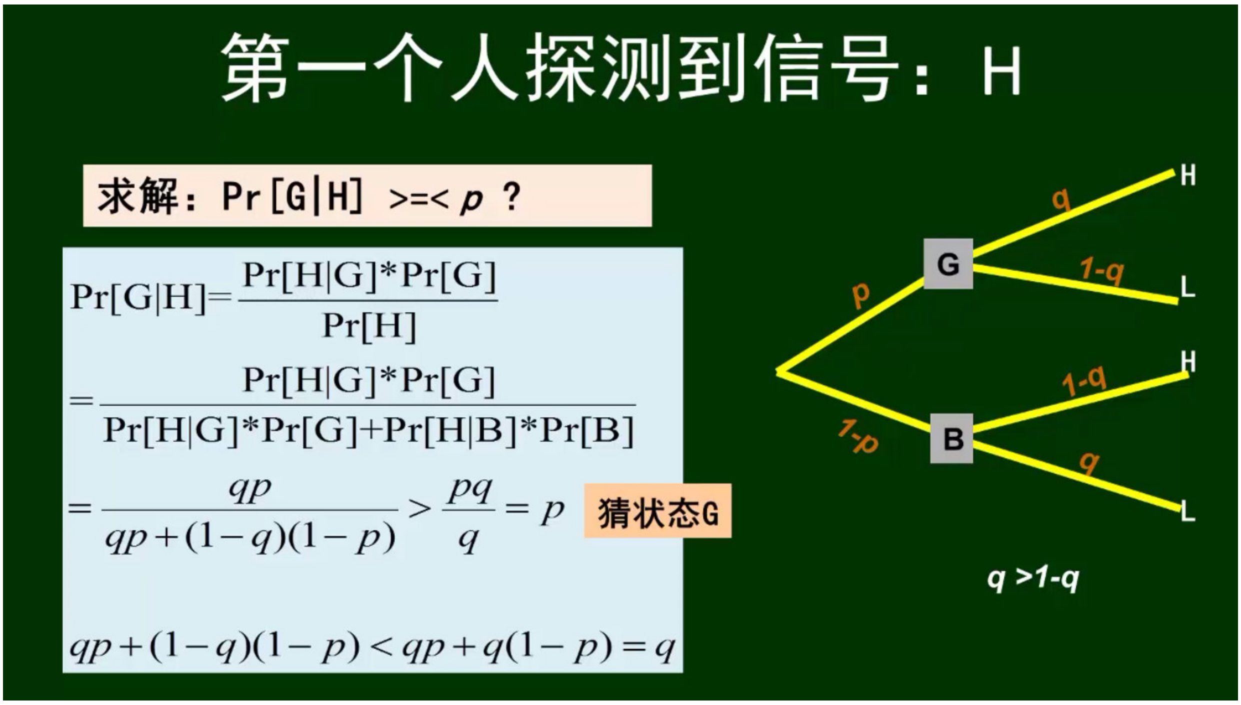
级联特征

- 基于少量信息，结构可能不是最优

10.2 级联的通用模型

信息级联通用模型的描述





决策模型推理（1）

假定：条件是前面的人的“私有信号”序列

- 利用贝叶斯公式，证明：

  $\Pr[G|s_1, s_2, \ldots, s_N] > p$，若S序列中H的个数大于L

  $\Pr[G|s_1, s_2, \ldots, s_N] = p$，若S序列中H的个数等于L

  $\Pr[G|s_1, s_2, \ldots, s_N] < p$，若S序列中H的个数小于L

设n_1, n_2分别为H、L信号个数
已知：q>1-q

$$\Pr[G|s_1,s_2,\ldots,s_n]=\frac{\Pr[s_1,s_2,\ldots,s_n|G]\cdot\Pr[G]}{\Pr[s_1,s_2,\ldots,s_n]}$$

$$=\frac{\Pr[s_1,s_2,\ldots,s_n|G]\cdot p}{\Pr[s_1,s_2,\ldots,s_n|G]\cdot p+\Pr[s_1,s_2,\ldots,s_n|B]\cdot(1-p)}$$

$$=\frac{\Pr[s_1|G]\cdot\Pr[s_2|G]\cdot\ldots\cdot\Pr[s_n|G]\cdot p}{\Pr[s_1|G]\cdot\Pr[s_2|G]\cdot\ldots\cdot\Pr[s_n|G]p+\Pr[s_1|B]\cdot\Pr[s_2|B]\cdot\ldots\cdot\Pr[s_n|B]\cdot(1-p)}$$

$$=\frac{q^{n_1}(1-q)^{n_2}p}{q^{n_1}(1-q)^{n_2}p+(1-q)^{n_1}q^{n_2}(1-p)}=\frac{p}{p+\left(\frac{1-q}{q}\right)^{n_1}\left(\frac{q}{1-q}\right)^{n_2}(1-p)}$$

$$=\frac{p}{p+\left(\frac{1-q}{q}\right)^{n_1-n_2}(1-p)}$$

决策模型推理（2）

实际情况：条件是前面的人的“判断结果”序列：

$R=\{r_1, r_2, \ldots, r_N\}$

- 假设到第k个人之前还没有形成级联，根据前面的分析，$\{s_1, s_2, \ldots, s_k\}=\{r_1, r_2, \ldots, r_k\}$，第k+1个人的选择取决于：
  $Pr[G|s_1, s_2, \ldots, s_k, s_{k+1}]$
- 比较$s_1 \sim s_k$中H、L信号个数
  - 个数相同，第k+1个人按s_{k+1}选择，$r_{k+1}=s_{k+1}$
  - 个数相差1，第k+1个人按s_{k+1}选择，$r_{k+1}=s_{k+1}$
  - 个数相差2，第k+1个人忽略s_{k+1}，从此r_{k+}与s_{k+}脱离关系，形成级联

证明：N→∞，产生级联的概率为1

- 证明：当N足够大时，存在连续3个相同信号的概率为1
  - N个信号序列3个一组分组
  - 一组3个信号相同的概率：$q^3+(1-q)^3$
  - 没有一组3个信号相同的概率：$(1-(q^3+(1-q)^3))^{N/3}$
  - 容易证明，N→∞，$(1-(q^3+(1-q)^3))^{N/3}$趋于→0
  - 即，N足够大时，信息级联的概率→1

更一般化信息级联通用模型

- 实际中的从众情况可能更复杂：
  - 个体探测到的信号可能多于2个，这些信号在不同状态中出现的概率也可能完全不同，等等
- Bikhchandani,等在其1992年发表的文章中，创建了更一般化的级联模型，并证明了只要：
  - 事物对个体具有不确定性，个体对事物的状态能探测到相应的信号，个体能够看到别人的决策结果，并且能够量化这些概率
  - 个体数足够大时产生信息级联概率为1

特殊个体对级联的影响

- Pr[Good]=Pr[Bad]=1/2, Pr[H|Good]=q, Pr[L|Good]=1-q
  Pr[H|Bad]=1-q, Pr[L|Bad]=q
- 假设第一个个体概率值q_1大于其他个体q值，
  即，$q_1>q_i$， 其他个体概率值相同$q_i=q$，（i>1）

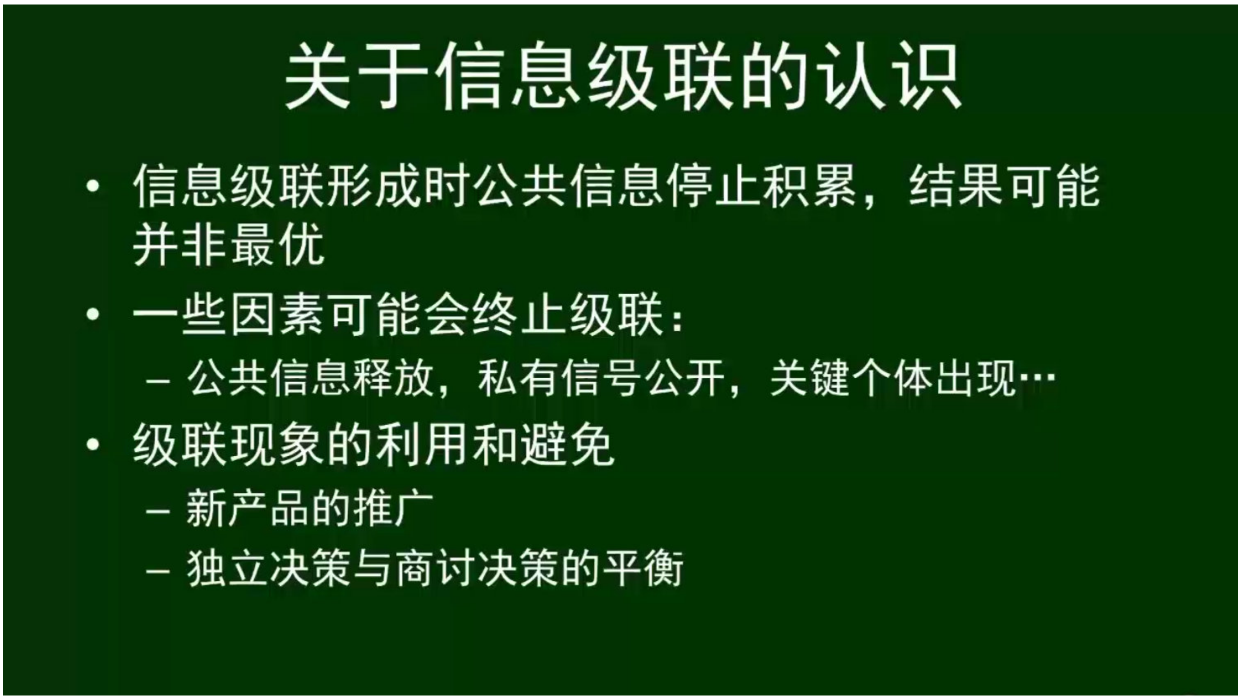
设第一个人探测到H，选G，第二个人探测到L，如何选择？

$$\Pr[G|H,L]=\frac{\Pr[H,L|G]*\Pr[G]}{\Pr[H,L|G]*\Pr[G]+\Pr[H,L|B]*\Pr[B]}$$

$$=\frac{q_1(1-q)}{q_1(1-q)+(1-q_1)q}>\frac{q_1-q_1q}{q_1+q_1-2q_1q}=\frac{1}{2}$$

$$(q_1+q-2q_1q<q_1+q_1-2q_1q, q_1>q)$$

从第二个人开始就形成级联



10.3 流行

事物的流行性

- 个体：一种特定的事物的流行性随着时间的演变
- 全体：一类事物中不同品种流行性的差异

10.4 幂律: 不同类事物流行的一般规律

流行性的定量观察

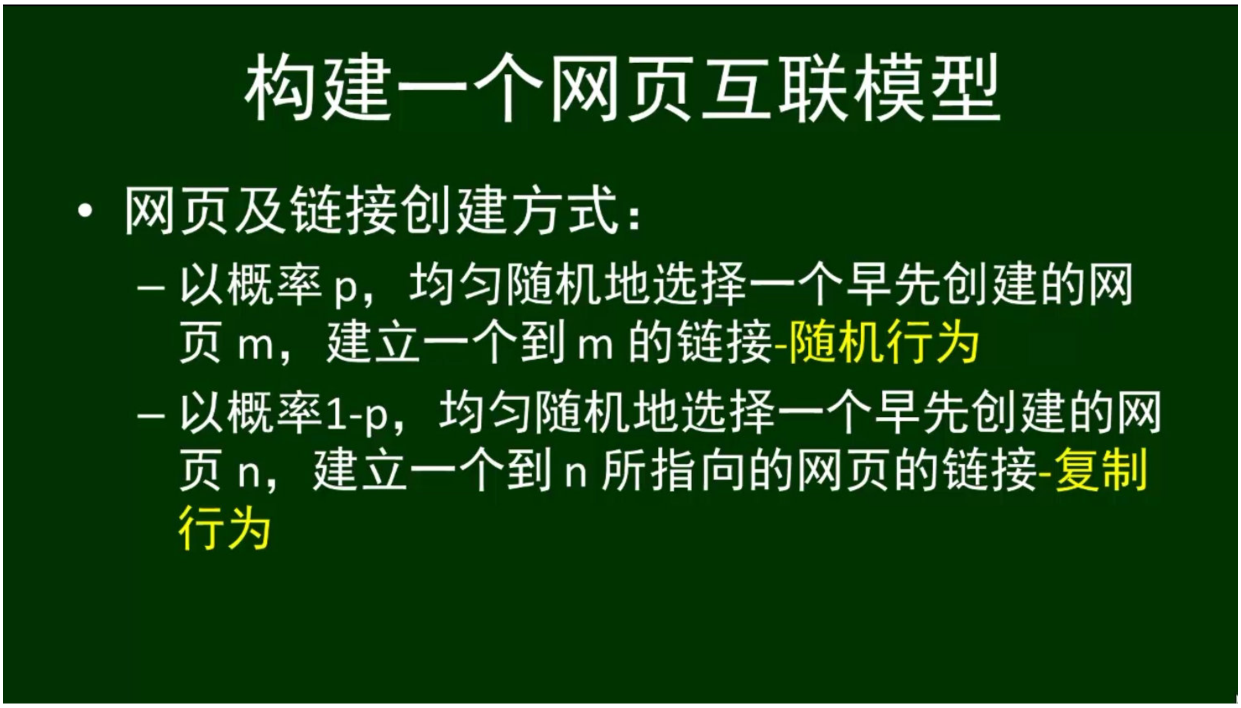
- 给定一个网页集合S，考察一个网页的入向链接数为k的概率f(k)
- 实验数据表明，网页流行度近似幂律分布

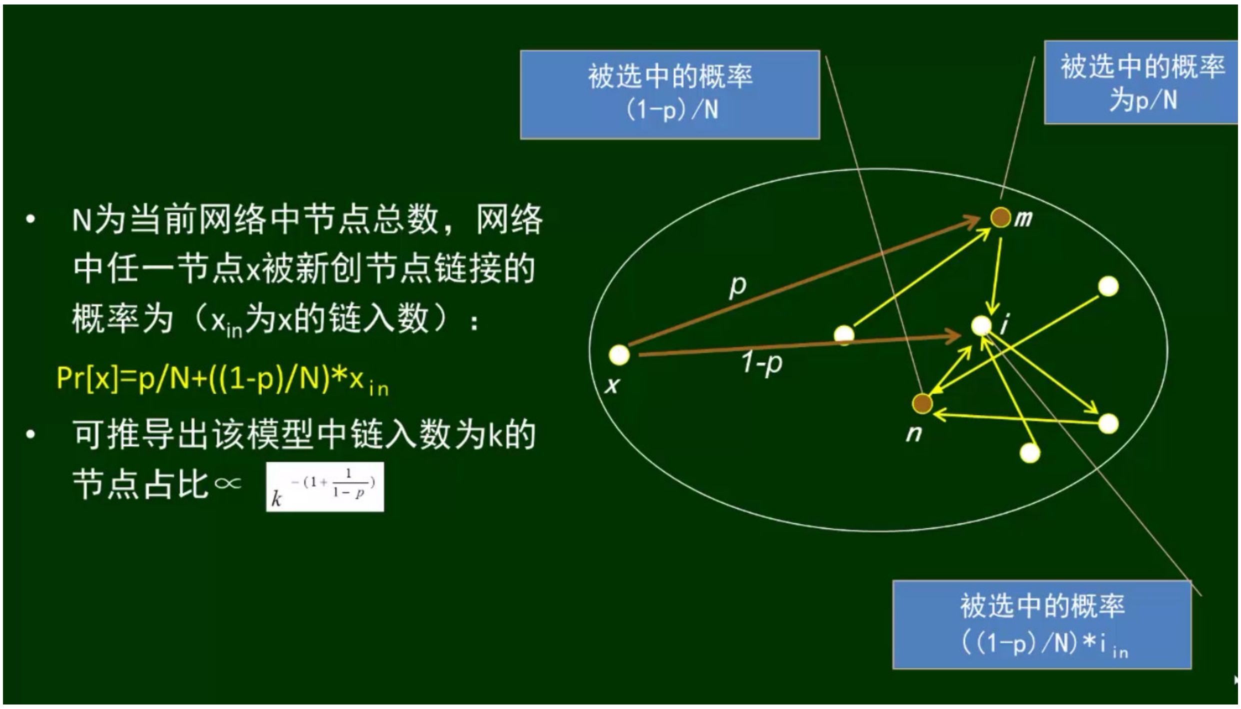
结论

- 偏离均值的网页占比下降较慢
- 少量网页链入数极高

10.5 幂律的成因

网页互联模型





Summary

- 幂律的成因：增长性和择优链接
- 引发事物的流行具有一定偶然性

10.5 看事物流行性的不同视角

帕累托分布(累计概率分布)

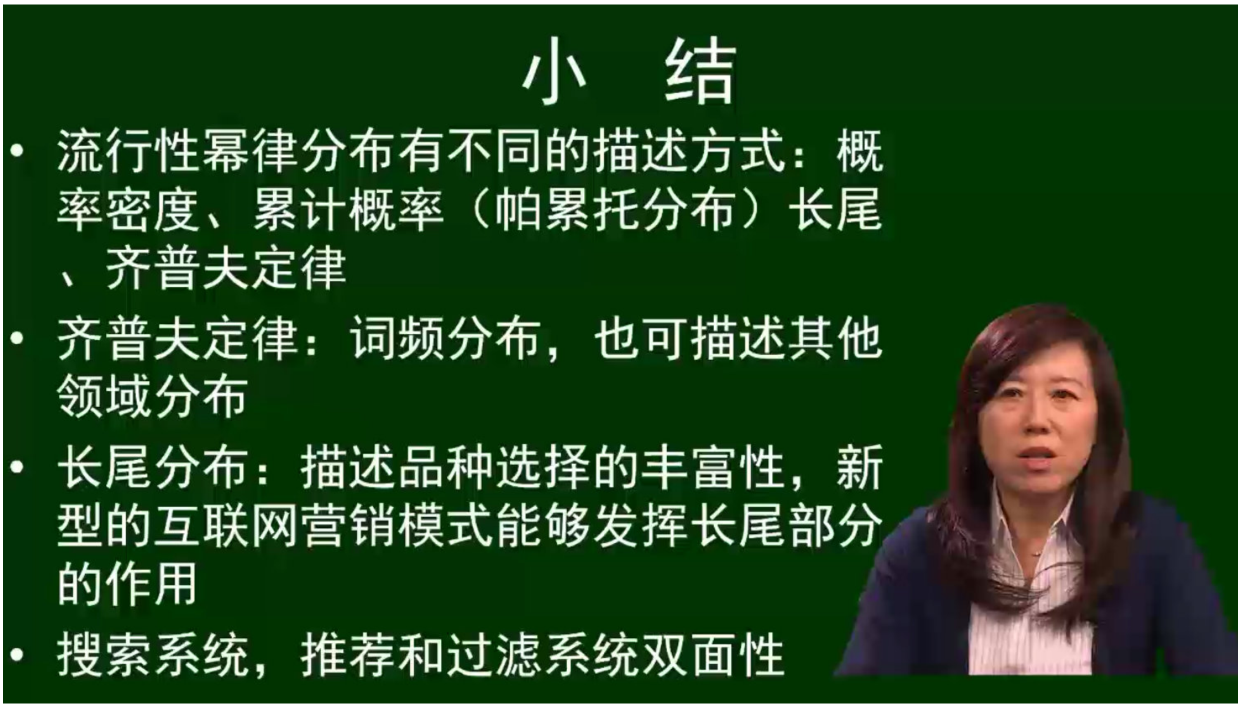
$$P(X > x) = \frac{a}{x^k}, k \geq 1$$

齐普夫定律

$$f(r) = \frac{a}{r^c}, r = 1, 2, 3...$$

r为流行度排名，$f(r)$为排名第r位实体的规模

2/8律



11. 信息不对称对市场的影响

11.1 内生事件与外生事件

制度

- 任何约束人们“社会行为”的规则

作为制度的市场

- 通过建立针对所有参与者的基本约定，以保证社会秩序
- 通过获得社会最优来保障社会公平
- 人们参与市场的行动，就是制度约束下的行动

外生事件市场

- 事件发生的概率不受行动者参与行为影响
- 彩票、选举

内生事件市场

- 事件发生的概率受到行动者参与行为影响
- 交通、二手车市场

11.2 简单预测市场

赛马问题

- 两匹马：A和B，即将比赛
- 你有钱数w用于下注
- 分别投放多少最好？

- 求r，0≤r≤1
- 用rw赌A，(1-r)w赌B
- 使得回报最大

如果“回报＝返回钱数的期望”

- 还假设
  - A取胜概率：a；B取胜概率：b=1-a
  - A取胜赔付率：o_A；B取胜赔付率：o_B
- 于是，要确定r，使下式最大

$$
\begin{aligned}
E[r] &= a\cdot(r\cdot w\cdot o_A) + b\cdot((1-r)\cdot w\cdot o_B) \\
&= r\cdot w(a\cdot o_A - b\cdot o_B) + w(1-a)o_B
\end{aligned}
$$

- 也就是，若 $ao_A > bo_B$，则r=1，否则r=0。

风险意识

- 担心坏事出现带来的损失，宁可放弃一些好事出现带来的利益

为风险意识下的回报建模

- 让财富增长带来的“好的感受”随财富量的增加而减少

这里，我们看到有可能对不同于金钱的回报建模

如何刻画“好的感受”的增长随财富量的增加而减小

- 效用函数（utility function）

$$U(w)-U(w-\Delta w)>U(w+\Delta w)-U(w)$$

$$U(w)>\frac{U(w+\Delta w)+U(w-\Delta w)}{2}$$

- 斜率递减的函数，$w^{1/2}$，$\ln(w)$，…
- 用财富效用函数（而不是财富）的期望作为评价投注策略回报的指标，以反映人们感受财富或投资风险的心理

取 $U(w)=\ln(w)$

体现财富倍增带来的感觉一样，与最初多少无关

$$\ln(2w)-\ln(w)=\ln(2)$$

于是，效用期望

$$E(r)=a\ln(r\cdot w\cdot o_A)+(1-a)\ln((1-r)\cdot w\cdot o_B)$$
$$=a\ln(r)+(1-a)\ln(1-r)+a\ln(wo_A)+(1-a)\ln(wo_B)$$

$$\frac{dE}{dr}=\frac{a}{r}-\frac{1-a}{1-r}$$ 令为0，得 r = a

11.3 市场价格

- 赔付率：赌场为赌客正确下注的一块钱需要支付的钱数
- 赔付率的倒数相当于“单位价格”，即赌客为了得到一块钱回报需要正确下注的钱数

赛马场的赔付率（无陪赚运行）

- 假设对数效用函数，有N个赌客（1,2,…,N）参与，有如下财富和信念参数，并按照各自的信念下注
  - w_n，$w=w_1+w_2+\ldots+w_N$
  - a_n，$b_n=1-a_n$
- 设赛马场按照“无赔赚”运行，即无论哪匹马取胜，收到的所有赌注（w）都给出去
- 如何确定赔付率（o_A和o_B）？

让财富在参与人之间流转

赛马场：如何确定赔付率o_A和o_B

- 马场总共收的钱数为w。其中在A和B上的投注量分别是

A: 10, 20, 25
B: 10, 15, 30

$$\sum_{n=1}^{N} a_n w_n \qquad \sum_{n=1}^{N} b_n w_n$$

- 如果A取胜，马场要把所有收的钱（w）按照一定的赔付率（O_A）都给出去，即

$$o_A \sum_{n=1}^{N} a_n w_n = w$$

类似地有

$$o_B \sum_{n=1}^{N} b_n w_n = w$$

赔付率、状态价格、信心的聚合

$$o_A = w \cdot \left(\sum_{n=1}^{N} a_n w_n \right)^{-1}$$

$$o_B == w \cdot \left(\sum_{n=1}^{N} b_n w_n \right)^{-1}$$

$$\sum_{n=1}^{N} a_n \boxed{\frac{w_n}{w}} = \sum_{n=1}^{N} a_n f_n = o_A^{-1} = \rho_A$$

$$\sum_{n=1}^{N} b_n \frac{w_n}{w} = \sum_{n=1}^{N} b_n f_n = o_B^{-1} = \rho_B$$

加权平均　“权”，人们各自的财富在总财富中的占比

谁的下注策略肯定不亏本？

- 若大家都按信念下注，且赌场按上述方式计算赔付率，一个人有没有可能，无论哪匹马赢，都拿回自己下注的财富？
- 答案：可能，若她对A的信念是ρ_A，对B的信念是ρ_B。
  - 例如，A若赢，她就得到（$w_n\rho_A$）$o_A = w_n$

如果你的智慧与“群众智慧”一致，则不会吃亏！

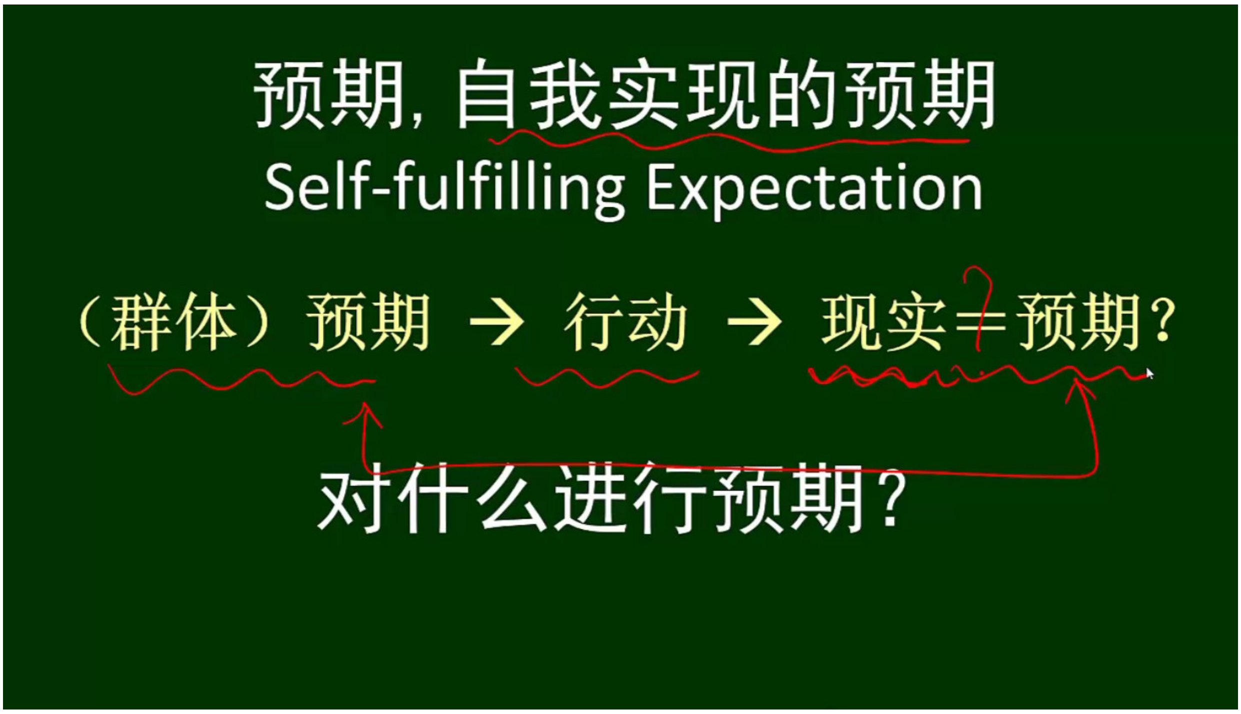
要点小结

- 通过简单预测市场的例子，讨论了市场价格的形成，尤其是价格的含义
- 市场决定出来的价格，对应着人们的信心，价格是人们信心的聚合
- 一个人的信心在这个聚合中的分量取决于他的财富量

11.4 内生事件与市场预期

内生事件

- 是否出现与人们的聚合行为直接相关
  - 根据对交通是否会拥堵的判断决定自己的出行
    - “拥堵”事件是否出现取决于人们的出行行为
  - 加入一个新出现的社交网站是否能获益
    - “获益”这件事是否成真取决于人们加入的情况
- 二手车市场
  - 冷淡还是火爆，取决于人们对待出售车的性价比的判断



11.5 柠檬市场: 信息不对称对市场的影响

柠檬市场的要点

1. 市场中的商品有多个质量等级
2. 买家和卖家对每一等级商品有不同的底线价格（设同一等级中买家估值>卖家底价）
3. 买卖双方对每一具体商品的质量信息不对称
4. 因此买家只可能出一个<u>期望价格</u>，卖家按照所持有具体商品的底价与买家给出的价格的关系决定是否出售

信息不对称：内生事件市场的复杂性

- 二手车市场：卖车的比买车的更了解车的状态
- 淘宝市场：卖家比买家更了解货物的质量
- 医疗保险市场：投保人（买家）比保险公司（卖家）更了解自己健康状况
- 人力资源市场：找工作的人（卖家）比雇主（买家）更了解自己的工作能力

信息弱势的一方会在考虑到另一方有充分信息的情形下形成对交易商品质量的预期，并按照预期决定自己愿意支付（接受）的期望价格。

二手车市场的分析与推理

- 假设有两种二手车“好车”与“破车”
  - 买家分别愿出不高于12和6买它们
  - 卖家的价格底线分别是10和4
  - 假设买家比车多（不挑车，值就行）

如果信息是对称的，即买卖双方都能对每一辆车况有相同的了解，市场会如何？
（车能否卖出去，价格？）

- 由于车少于买家，将分别在12和6全部卖出

二手车市场（不对称信息）

- 假设有两种二手车“好车”与“破车”
  - 买家分别愿出不高于12和6买它们，但不能识别
  - 卖家的价格底线分别是10和4，了解车的好坏
  - 鱼目混珠，于是只有统一的销售价。设买家比车多。

随机看见的一辆车，买家愿出多少钱？
出这个钱能成交吗？整个市场会是什么情形？

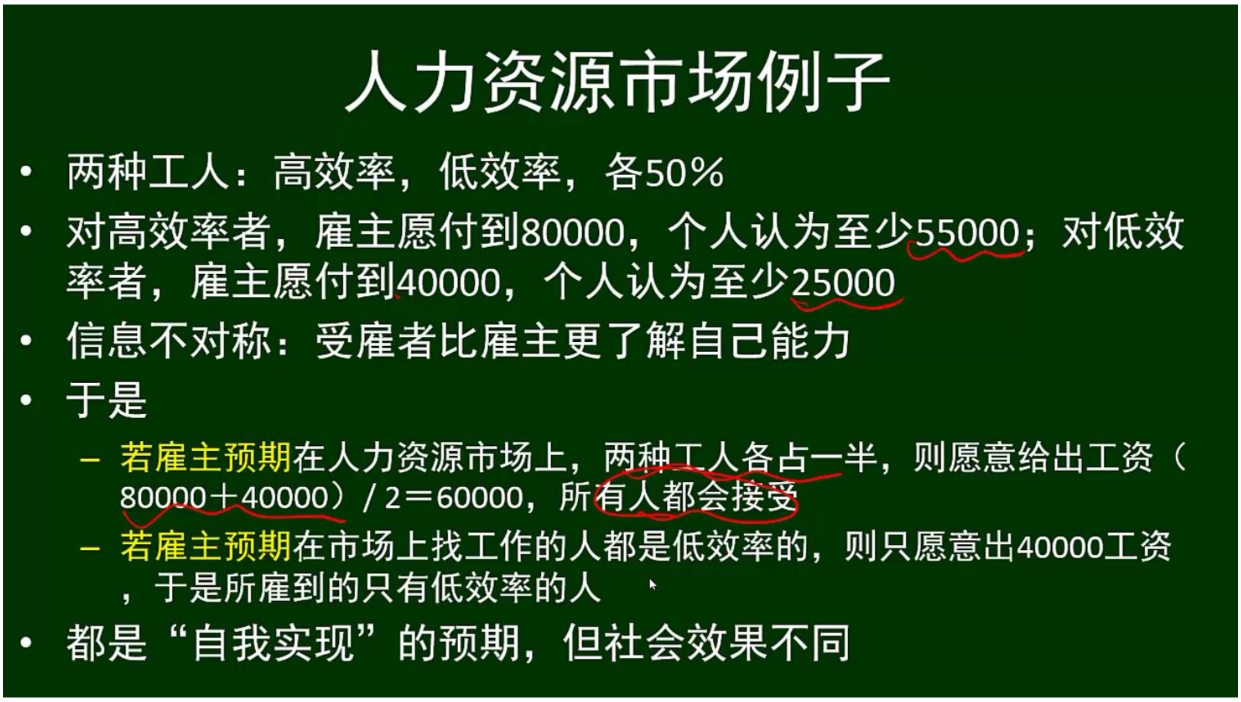
- 显然，如果大家感到市场上的破车偏多，则出价会较低，否则，出价会较高 好车占比
- 而出价如果太低，拥有“好车”的卖家就不会卖

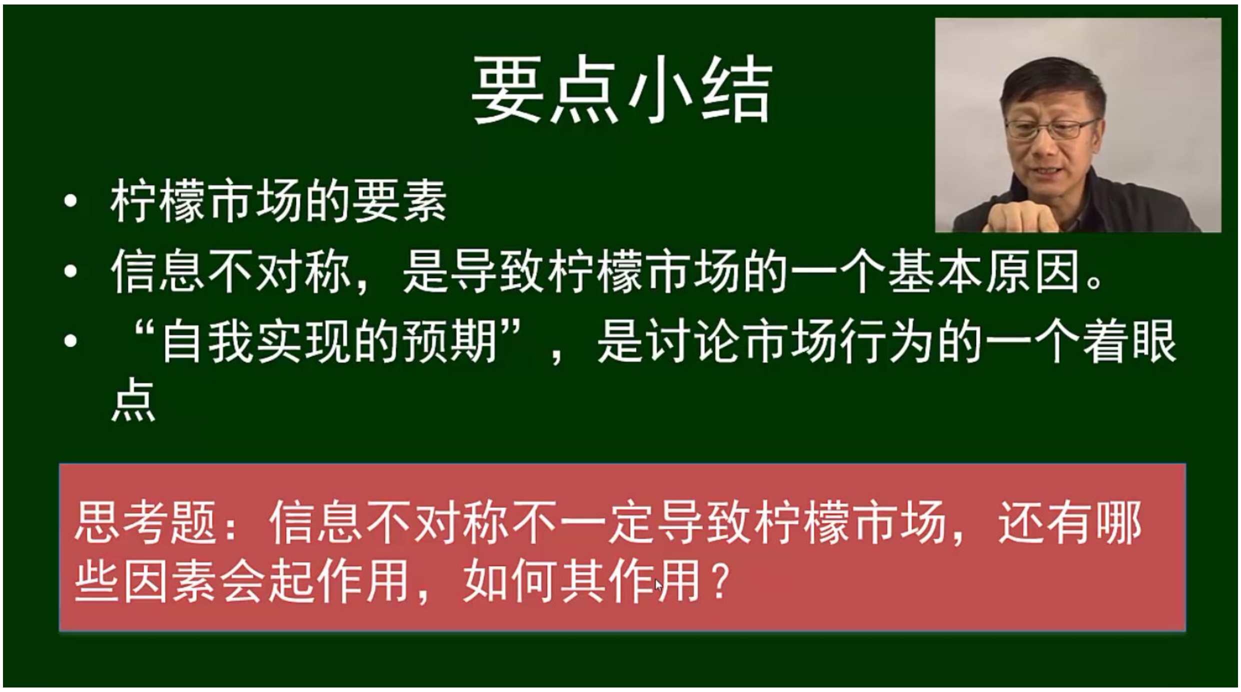
预期的对象：市场上好车的占比

- 设买家都预期好车占比h，则破车1-h，那么买家愿意付的价钱是：$12h + 6(1-h) = 6h + 6$
- 现在要问，对于哪些h，这个预期将“自我实现”？（也就是真有h占比的好车卖掉了）
- h=0？（相当于认为好车都没拿出来卖）
  - 此时买家们愿意付6，那些拥有好车的人（底价10）不会卖，破车（底价4）则统统卖掉了。即，真的就是h＝0，预期实现了。
- h=½？（假设好车的确有1/2）
  - 买家愿意支付的价格=9，拥有好车的人（底价10）不会卖，破车（底价4）则统统卖掉。于是人们在市场上看到的好车占比为0，意味着这个预期没有自我实现。
- h=⅔？（假设好车的确有2/3）
  - 买家愿意支付的价格=10，好车（底价10）和破车（底价4）统统卖掉了。于是人们在市场上看到的好车占比为2/3，这个预期自我实现了。

三种车的情形（好车、破车、柠檬）

- 假设
  - 三种旧车各占1/3。
  - 卖方底价分别为10，4，0
  - 买方估值分别为12，6，0。买家多于车。
- 若预期为“市场上三种车各占1/3”
  - 出价（12＋6＋0）／3＝6，于是
  - 于是只有破车和柠檬在卖，即市场没有实现预期
- 会导致进一步预期“市场上只有破车和柠檬”
  - 出价（6＋0）／2＝3，于是只有柠檬卖掉了，市场也没实现预期。市场失效。





11.6 减少信息不对称影响的措施

对市场失灵的再讨论

- 对内生事件的市场，市场失灵的影响来自于信息不对称
- 信息不对称的本质，就是交易中的一方
  - 拥有对交易品的优势信息，故
  - 要么使得优势信息公开，使其不再具有优势性
  - 要么降低优势信息方对市场可能带来的负面影响
- 不可能的任务
  - 如果要求拥有优势信息的一方共享信息，在市场条件下
  - 就是不可能的任务

降低信息不对称的负面影响

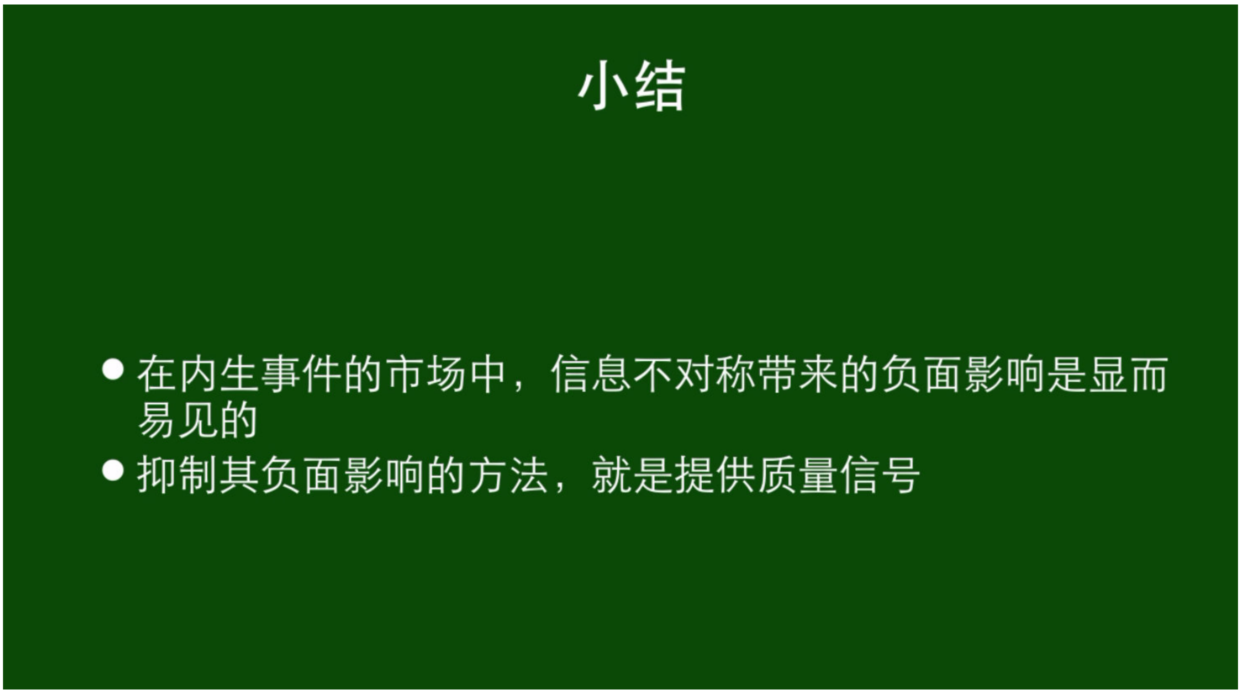
- 唯一途径，就是约束由信息优势方对市场带来的负面影响
- 单件交易品的价格与交易品的品质和稀缺性呈正相关，则
  - 提供包括交易品品质在内的稀缺性信息，将会让交易品物有所值
  - 如：第三方认证（certificate）
  - 再如：稀缺性（质量）保证，如承诺三包（wrranty）
  - 还有如：品牌（brand）等，都是对信息劣势方带来的补偿

降低信息不对称的负面影响

- 如果是劳动力
  - — 同样可以采用稀缺性（质量）信号方法
    - 受教育程度：质量信号
    - 毕业学校：第三方担保
- 稀缺性（质量）信号的意义
  - — 提供了交易品的直接价值
  - — 提供了针对交易品的信心价值

降低信息不对称的负面影响

- 如果是在互联网上呢？例如
  - — 淘宝，赶集，同城等
  - — 同样面对由信息不对称所带来的市场失灵问题，最常见的是欺诈
- 公共信息的影响
  - — 为此，互联网交易平台普遍采用了由信息劣势方评价的方法
  - — 所产生信息优势方信誉，实际就是社会担保
- 问题
  - — 合谋、恶意欺诈，依然是难以解决的难题



12. 表决

12.1 表决在社会中的作用及其多种形式

市场与表决

- 都是形成集体决策的方式
- 市场的议题不一定很明确，表决的议题较明确
- 市场的时间约束不一定很强，表决的时间约束较强
- 市场的场合通常是间接的，表决的场合通常是直接的

表决的影响

- 表决总是一部分人的共意的表达，同时也是社会共意形成的基本机制
- 表决是人类社会形成的集体决策

表决的形式

- 议决：通过协商、讨论，达成多数人同意的决策
    - 每个人公开表达自己的意见
- 票决：通过投票的方式，达成多数人同意的决策
    - 公开或不公开

表决的本质

- 多数决，由多数人的意见来代表集体的决策

合理的表决制度

- 是否真正体现了集体意志或真实地反映了社会事实

小结

- 表决，是人类社会形成集体决策最直接的、也是最古老的形式
- 表决所产生的结果，实际是多数人（票）的决策
- 影响表决行为的是信息，因此，表决实际是信息聚合的一种形式

12.2 偏好关系

偏好

- 对两个需要表决的候选项 X 和 Y，如果表决者 i 选择 X（也称为她偏好X），则记为

  $X \succ_i Y$ 称“X优于Y”或“X大于Y”

- 通常，表决者可能面对多于两个候选项，用集合｛X, Y, Z,...,W, U,...,V｝表示，我们可以问她对其中任何两个元素的偏好态度，或者说请她给出<u>偏好</u>。

偏好关系

- 给定一个候选项集合 A＝｛A_1, A_2, ..., A_k｝，一个偏好关系指的是对其中某些元素对之间偏好的集合。
- 例如，X＝｛A, B, C, D, E｝
  - R1＝｛A>B, B>C, C>D, D>E｝
  - R2＝｛B>A, B>C, C>E｝
  - R3＝｛A>B, B>A, C>D, E>D｝
  - R4＝｛A>B, B>C, C>D, D>E, A>C, A>D, A>E, B>D, B>E, C>E｝
  - R5＝｛A>E, C>A, C>B, C>D, C>E, B>D, B>E, B>A, A>D, E>D｝
  - R6＝｛A>B, B>C, C>D, D>E, E>A｝

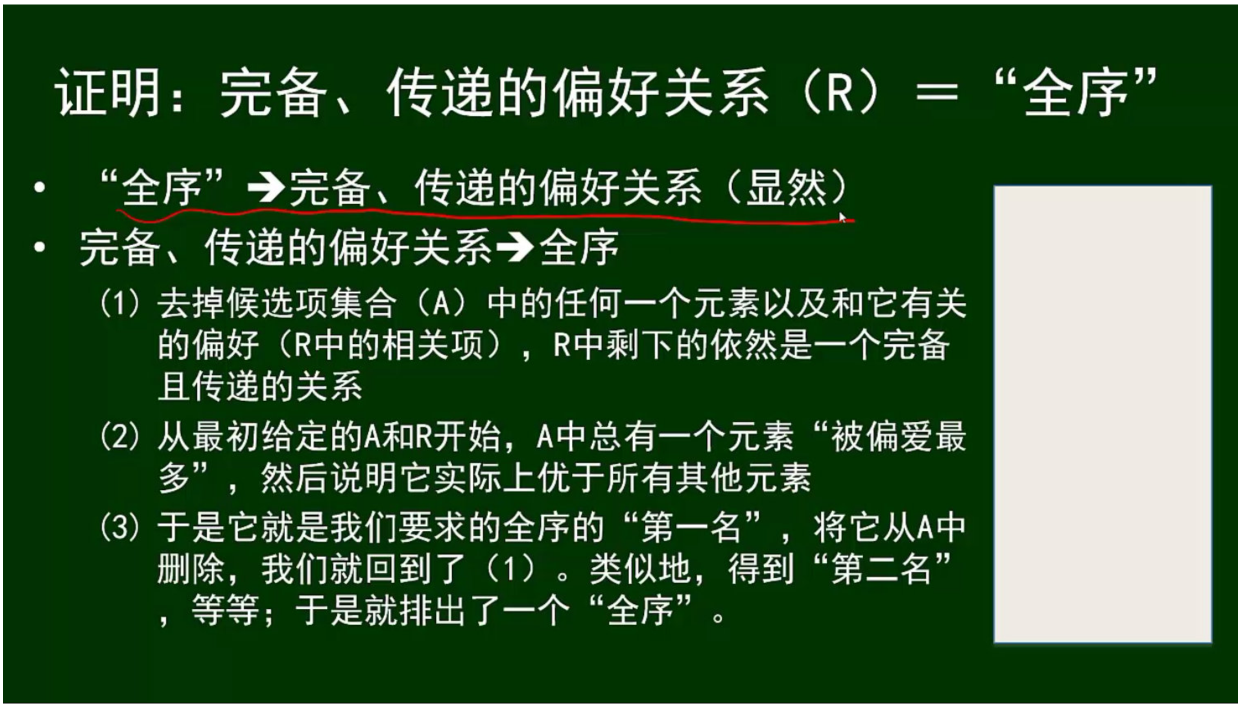
针对表决问题对偏好关系的（合理）假设

- 反对称性（anti-symmetric）

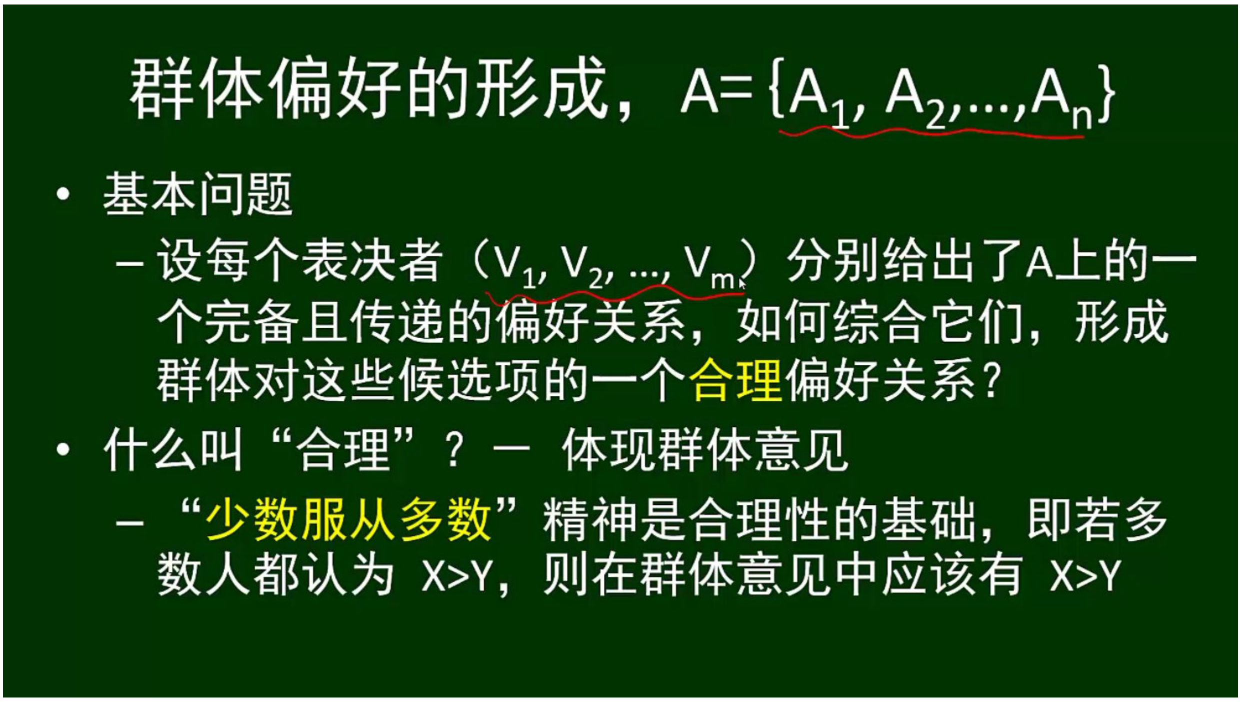
可以看成是传递性要求的特殊情况

- 完备性（complete）
  - 对于给定的选项（X, Y），要么偏好X，要么偏好Y；不能“无可奉告”
- 传递性（transitive）
  - 假定有3个选项（X, Y, Z），如果在（X, Y）比较中，偏好X；在（Y, Z）比较中，偏好Y；则在（X, Z）比较中，应该偏好X

这等价于在集合元素之间有一个全序（全部包含的顺序）


证明：完备、传递的偏好关系（R）＝“全序”
• “全序”→完备、传递的偏好关系（显然）
• 完备、传递的偏好关系→全序
（1）去掉候选项集合（A）中的任何一个元素以及和它有关的偏好（R中的相关项），R中剩下的依然是一个完备且传递的关系
（2）从最初给定的A和R开始，A中总有一个元素“被偏爱最多”，然后说明它实际上优于所有其他元素
（3）于是它就是我们要求的全序的“第一名”，将它从A中删除，我们就回到了（1）。类似地，得到“第二名”，等等；于是就排出了一个“全序”。

12.3 少数服从多数


群体偏好的形成，A={A1, A2,…,An}
• 基本问题
– 设每个表决者（V1, V2, …, Vm）分别给出了A上的一个完备且传递的偏好关系，如何综合它们，形成群体对这些候选项的一个合理偏好关系？
• 什么叫“合理”？ — 体现群体意见
– “少数服从多数”精神是合理性的基础，即若多数人都认为 X>Y，则在群体意见中应该有 X>Y

当只有两个候选项（X, Y）

- 设$V_1, V_2, \ldots, V_m$为奇数个表决者，每个人给出$X>_i Y$ 或 $Y>_i X$。如果多数人偏好为X>Y，则群体偏好为 X>Y，否则群体偏好为 Y>X

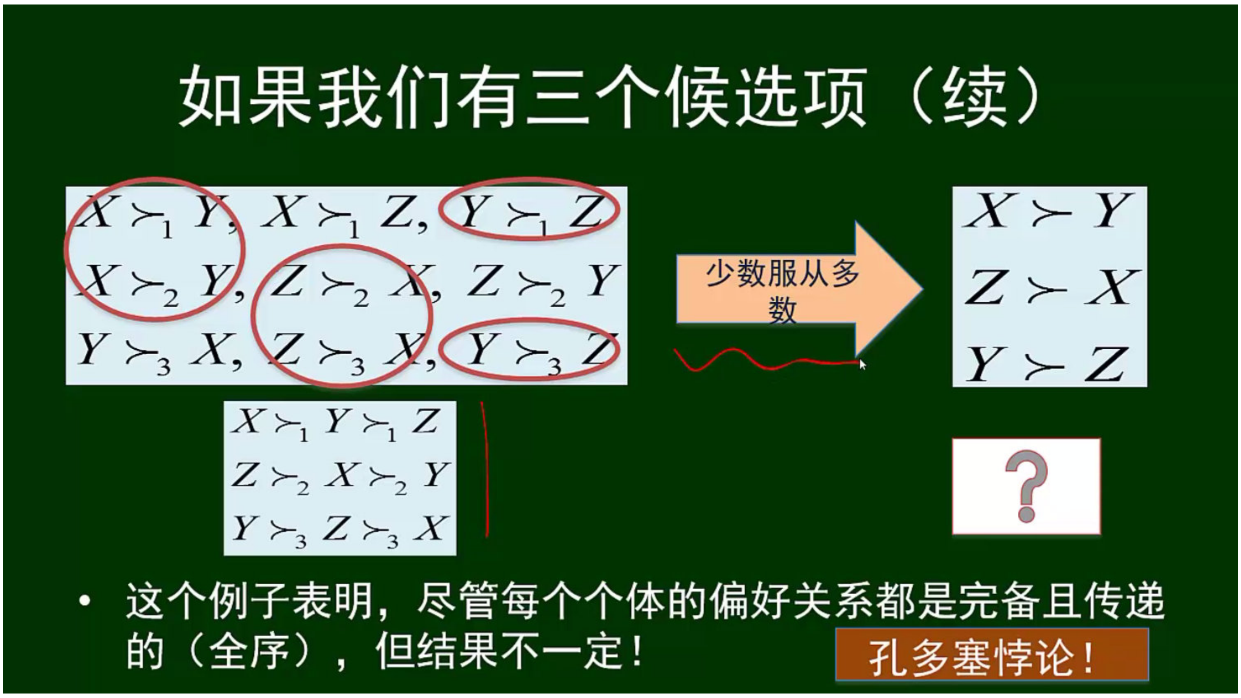
$$\begin{array}{c} X \succ_1 Y \\ X \succ_2 Y \\ Y \succ_3 X \end{array} \xrightarrow{\text{少数服从多数}} X \succ Y$$

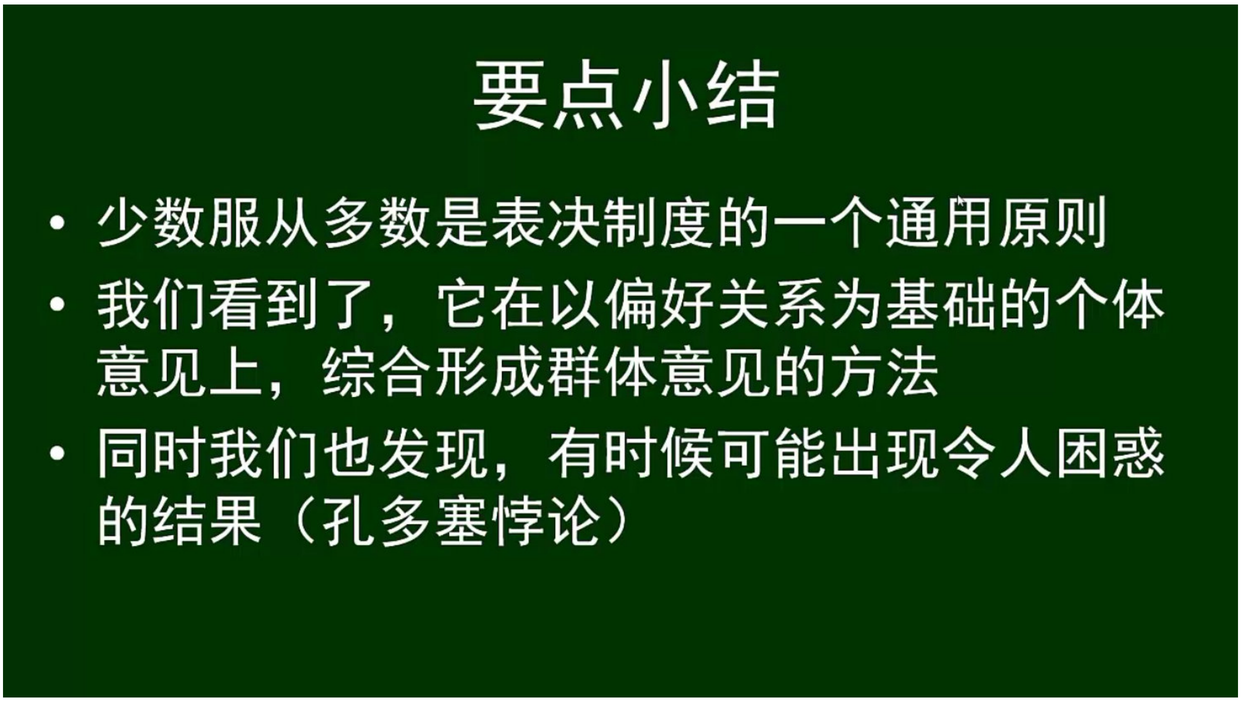
如果我们有三个候选项(X, Y, Z)

$$\begin{array}{c} X \succ_1 Y \succ_1 Z \\ X \succ_2 Z \succ_2 Y \\ Y \succ_3 X \succ_3 Z \end{array}$$

$$\begin{array}{ccc} X \succ_1 Y, & X \succ_1 Z, & Y \succ_1 Z \\ X \succ_2 Y, & X \succ_2 Z, & Z \succ_2 Y \\ Y \succ_3 X, & X \succ_3 Z, & Y \succ_3 Z \end{array} \xrightarrow{\text{少数服从多数}} \begin{array}{c} X \succ Y \\ X \succ Z \\ Y \succ Z \end{array}$$

$$X \succ Y \succ Z$$


如果我们有三个候选项（续）
$X \succ_1 Y,\ X \succ_1 Z,\ Y \succ_1 Z$
$X \succ_2 Y,\ Z \succ_2 X,\ Z \succ_2 Y$
$Y \succ_3 X,\ Z \succ_3 X,\ Y \succ_3 Z$
少数服从多数
$X \succ Y$
$Z \succ X$
$Y \succ Z$
$X \succ_1 Y \succ_1 Z$
$Z \succ_2 X \succ_2 Y$
$Y \succ_3 Z \succ_3 X$
?
• 这个例子表明，尽管每个个体的偏好关系都是完备且传递的（全序），但结果不一定！
孔多塞悖论！


要点小结
• 少数服从多数是表决制度的一个通用原则
• 我们看到了，它在以偏好关系为基础的个体意见上，综合形成群体意见的方法
• 同时我们也发现，有时候可能出现令人困惑的结果（孔多塞悖论）

12.4 孔多塞悖论

孔多塞悖论

- 孔多塞（Condorcet, 1700）遇到的问题
  - 前面的讨论已经说明，当只有两个备选项（X,Y）时，以选举为例
    - 如果投票人为奇数，且每个投票人的偏好满足完备性和传递性
    - 则一定会产生一个投票结果，即两个备选项中，总有一个在第一位
  - 可在3个人对3个备选项（X,Y, Z）进行表决时
    - 可能每个投票人的偏好都满足完备性和传递性
    - 按少数服从多数原则，得到的群体偏好却不一定满足传递性
    - 从传递性个体偏好，按少数服从多数聚合方式，得出非传递性群体偏好，称为孔多塞悖论

合理的个体行为 ＋ 合理的聚合方式 ➔ 不合理的群体结论！个体理性，群体无理性

思考讨论题

- 假设
  - 有m（奇数）个表决者，n个候选项
  - 每人给出候选项集合上的一个完备且传递的偏好关系
- 问题
  - 这些关系总共包含多少候选项对?
  - 按少数服从多数原则形成的偏好关系中一共包含多少候选项对?
  - 它是否一定完备?
  - 它是否一定传递?

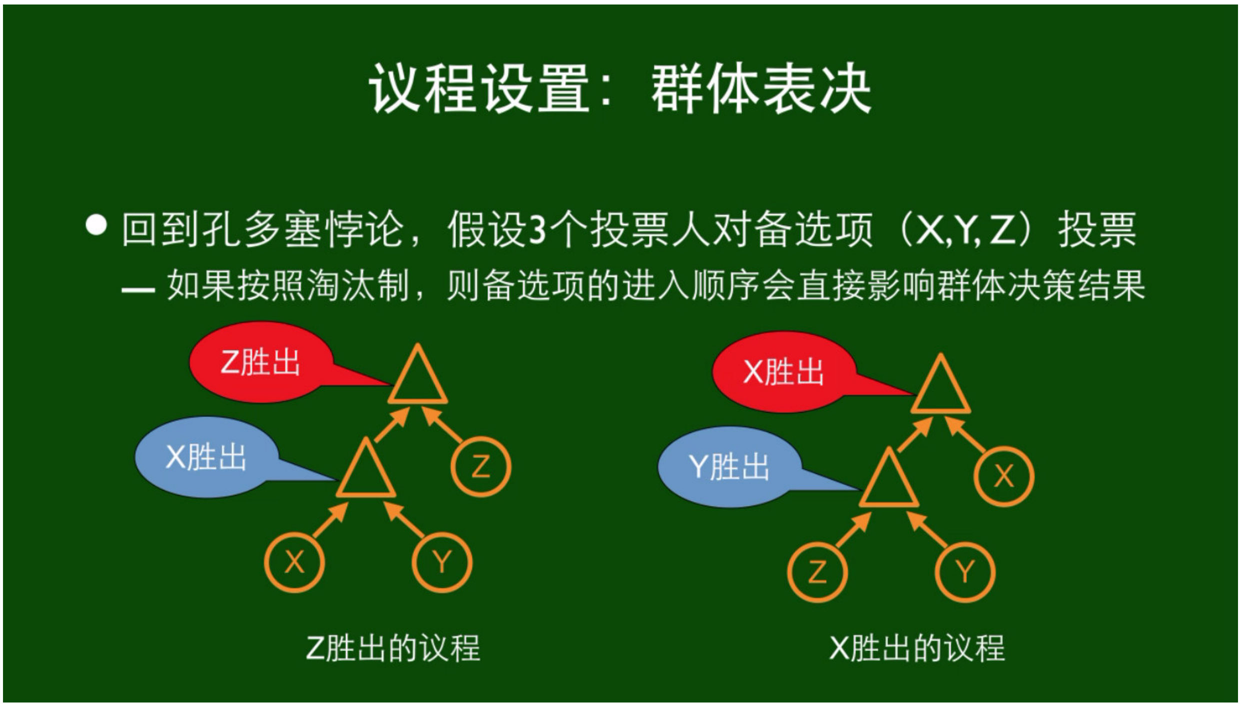
孔多塞悖论的解决思路

- 合理的个体行为 + 合理的聚合方式 ➔ 不合理群体结论
  - — 或者个体理性导致的群体无理性
- 解决思路
  - — 或调整个体行为假设
  - — 或调整聚合方式

12.5 议程设置

改变聚合方式

- 基于完备且传递的个体偏好
  - — 对备选项两两进行“少数服从多数”对比，是聚合的一种方式
  - — “逐一胜出（淘汰）”是另外一种可能
    - 假定备选项的任意一个序列，X,Y, Z, …
    - 沿着这个序列，开始比较 X 和 Y（依照少数服从多数原则），然后胜者再和 Z 相比，…，即保证“传递性”，就可以得到一个“最大的”
    - 对剩下的再进行这个过程，得到“次大的”…
    - 由此，就可以在整体上群体层次的保证传递性
- 如果将这个设计做一般化处理，就
  - — 形成了表决中议程设置对表决结果的影响



议程设置：个人决策

大学	全国排名	班级平均规模	奖学金
X	4	40	￥30000
Y	8	18	￥10000
Z	12	24	￥80000
	X>Y>Z	Y>Z>X	Z>X>Y

- 假设
  - 依据通知书到达的时间，来决定放弃某所大学
    - 且如果到达的顺序是 X,Y, Z，则当Y到达时，就放弃Y而保留X
    - 当Z到达时，就放弃Y，而保留Z；如果是另一种顺序，结果可能不一样

12.6 波达计数法

波达记数法

- 是由Jean-Charles de Borda 1770年提出的
  - 每个备选项依据其在个体排序中的位置，得到相应权重（积分）
  - 将每个备选项从个体获得的积分汇集，就得到了群体层次的排序
- 假设
  - 有 N 个备选项，个体 *i* 对备选项的排序对应一种赋值
  - 偏好排在第一的，赋值 N-1；以此类推，最后一个赋值为 0
  - 依据每个备选项得到的赋值和（积分），由高到低排序，从而形成群体偏好排序

积分制

- 假设有2个个体（1, 2）面对4个备选项（A, B, C, D）；个体的排序以及群体的积分如下

个体	第一偏好	第二偏好	第三偏好	第四偏好
1	A	B	C	D
2	B	C	A	D
积分	A=3+1=4	B=2+3=5	C=1+2=3	D=0+0=0

积分制的弊病

- 先看一个例子，假设有5个影评家（1, 2, 3, 4, 5）对两部影片（A, B）的排序

影评家	第一部影片	第二部影片
1	1	0
2	1	0
3	1	0
4	0	1
5	0	1
积分	3	2

- 看起来结果不错，哈？！

积分制的弊病

- 如果备选项由2项变为3项
- 如果按照真实想法投票

影评家	第一部影片	第二部影片	第三部影片
1	2	1	0
2	2	1	0
3	2	1	0
4	1	2	0
5	1	2	0
积分	8	7	0

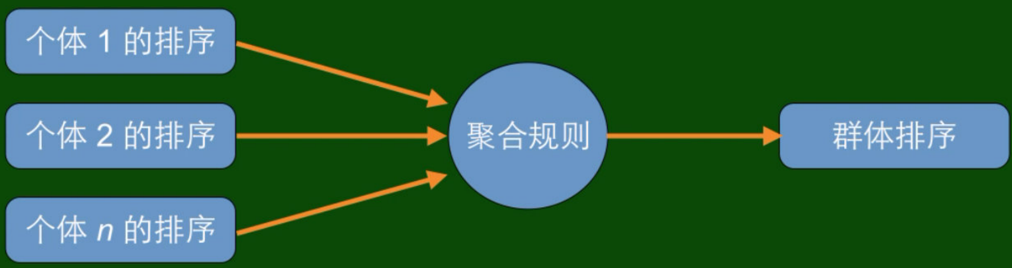
- 如果希望第二部影片当选
- 尽管公认第三部最差

影评家	第一部影片	第二部影片	第三部影片
1	2	1	0
2	2	1	0
3	2	1	0
4	0	2	1
5	0	2	1
积分	6	7	2

12.7 阿罗不可能定理

表决系统

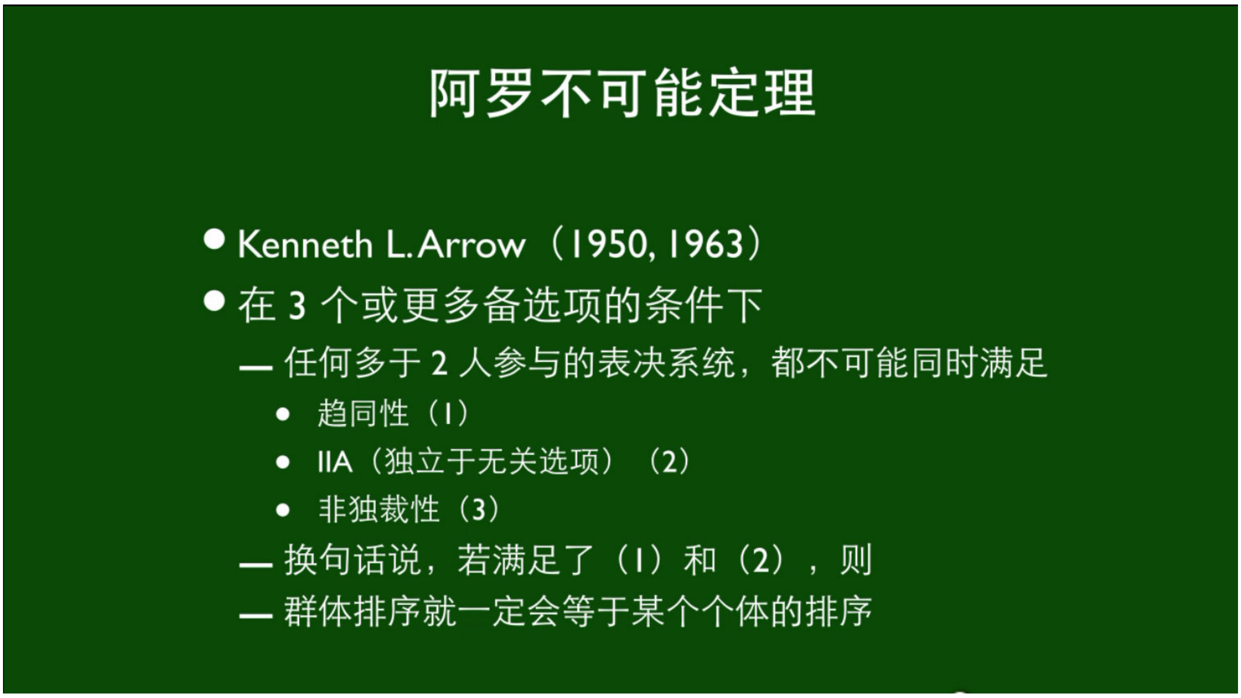
- 到目前位置，我们都是在现有的表决系统内思考解决方案

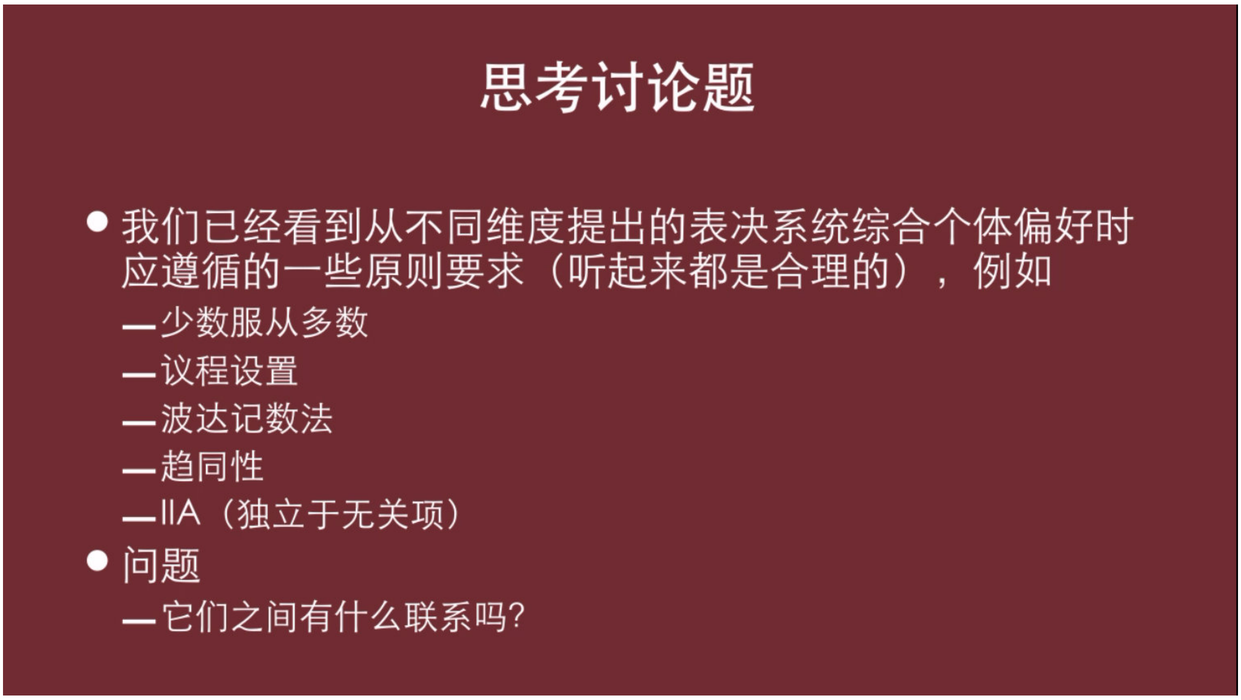


- 结论：在聚合规则上寻找解决方案，似乎比较困难了

跳出聚合规则

- 从整体上把握
  - 表决系统：对固定数量（k）的投票者，表决系统是一个函数，即
    - 以 k 个个人排序为依据，产生一个群体排序，即集体决策
  - 表决系统没有弊病，是指表决系统满足
- 帕累托原则（Pareto Principle）：给定两个备选项（X,Y），如果所有投票者给出的排序均为 $X >_i Y$，则群体排序就等于 $X > Y$，这就是趋同性（unanimity principle），即群体排序最少要反应个体排序
- 独立于无关项（independence of irrelevant alternatives, IIA）：群体对备选项（X, Y）的排序，仅取决于个体对它们的偏好，与个体对其他备选项的看法无关，即X 和 Y 在群体排序中的结果，不能因为某个个体调整了某个 备选项的相对位置而改变


阿罗不可能定理
Kenneth L. Arrow（1950, 1963）
在 3 个或更多备选项的条件下
— 任何多于 2 人参与的表决系统，都不可能同时满足
趋同性（1）
IIA（独立于无关选项）（2）
非独裁性（3）
— 换句话说，若满足了（1）和（2），则
— 群体排序就一定会等于某个个体的排序


思考讨论题
我们已经看到从不同维度提出的表决系统综合个体偏好时应遵循的一些原则要求（听起来都是合理的），例如
—少数服从多数
—议程设置
—波达记数法
—趋同性
—IIA（独立于无关项）
问题
—它们之间有什么联系吗?

12.8 单峰偏好

理解“合理的个体行为”

	偏好1	偏好2	偏好3
个体1	X	Y	Z
个体2	Y	Z	X
个体3	Z	X	Y

- 假设（X, Y, Z）分别代表财政支出由低到高，我们来理解三个人表态的逻辑
  - 个体1：钱花得越少越好
  - 个体2：少了可能不够，多了可能浪费，中间较好；如果还不够，就多花点
  - 个体3：？（行为很难解释，尽管在形式上也有个全序）

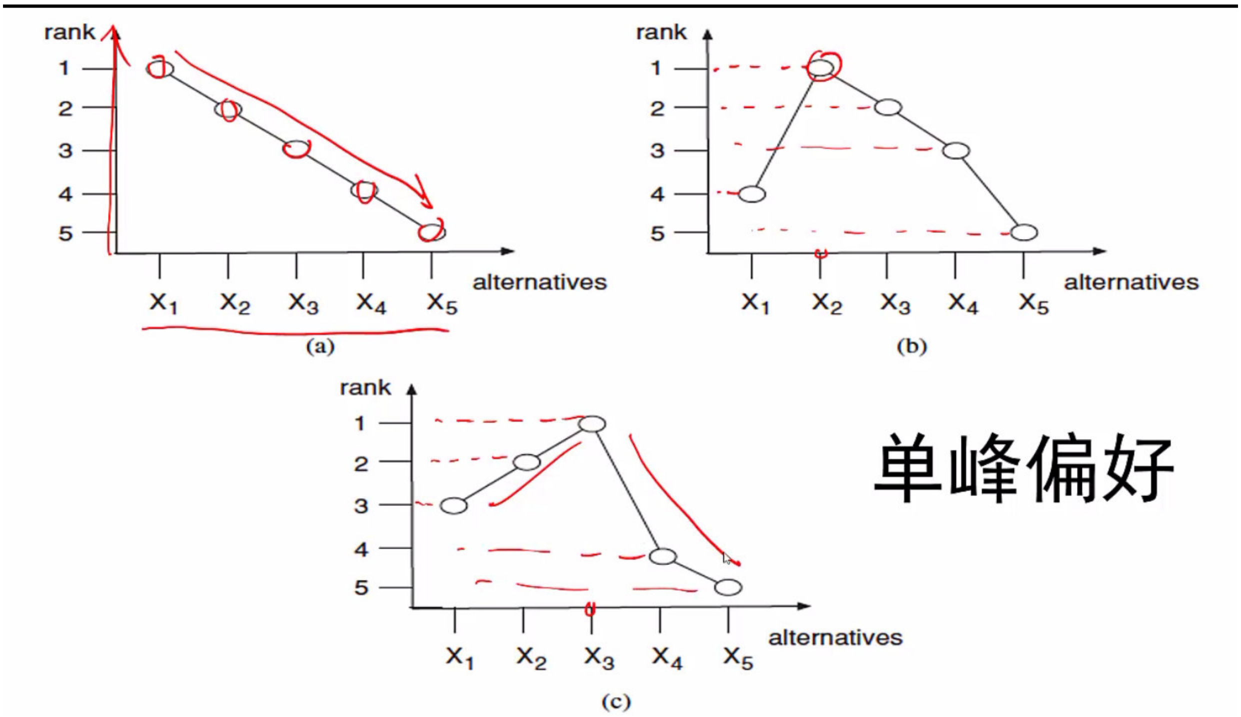
导致以“两两比较”＋“少数服从多数”原则得到的群体偏好不传递：X>Y, Y>Z, Z>X，即出现孔多塞悖论

单峰偏好 — 表决者的合理行为

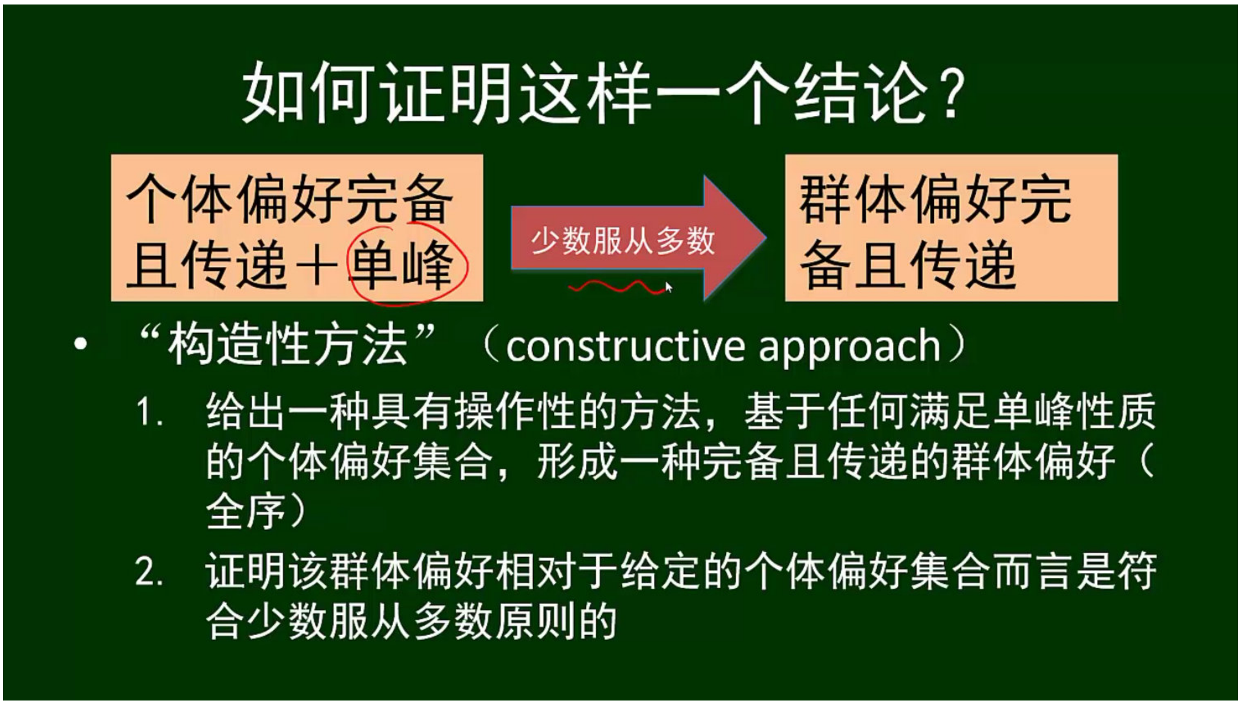
- 设想侯选项集合｛$X_2, X_1, X_5, \ldots\ldots X_N$｝的某种性质（P）有一种隐含的顺序（如支出预算数，个子高低，财富多少，考试分数高低等），如下所列

$$P_{X1} > P_{X2} > \ldots\ldots > P_{XN}$$

- 所谓表决者的态度应该满足单峰偏好指的是
  - 若X_i被他排在了第一位，则对于k<j<i，X_j要排在X_k的前面；且对于i<j<k，X_j要排在X_k的前面
- 即选择了一个“最爱”后，对其他选项的偏爱程度应参照所关注的性质在两边随与这个最爱的距离下降
  - （对左右两边相互之间没有要求）



12.9 单峰偏好下的表决结果

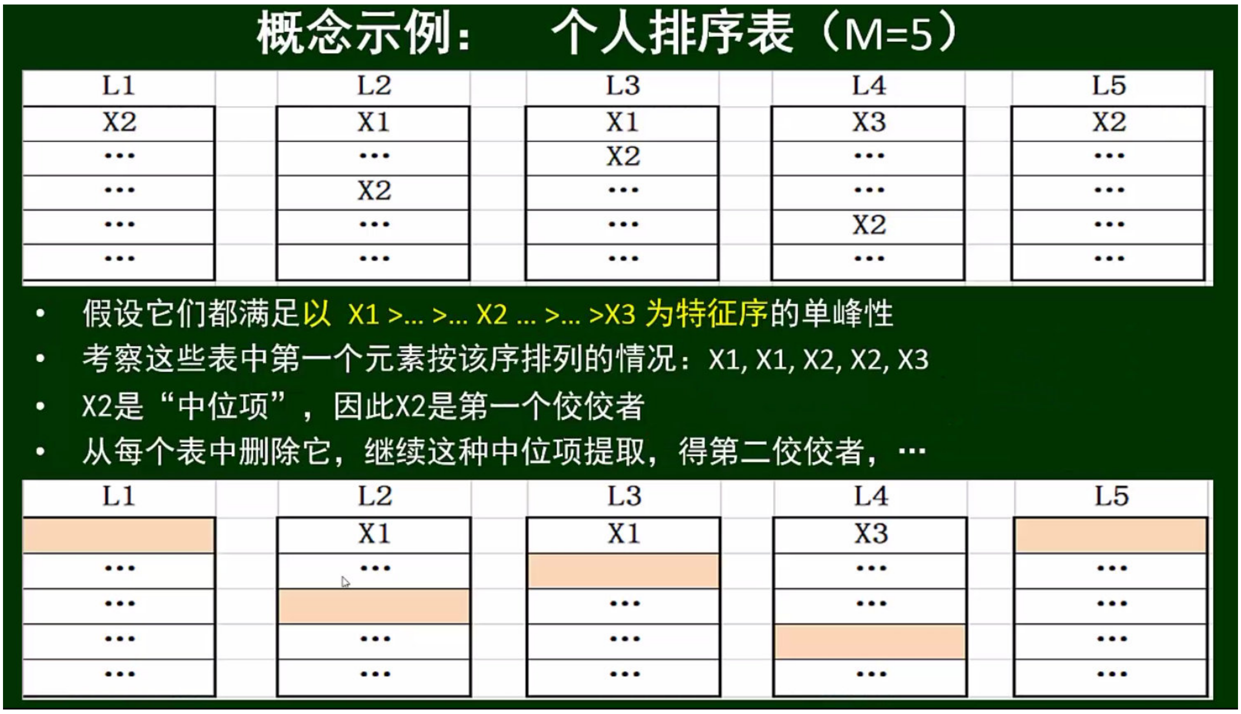


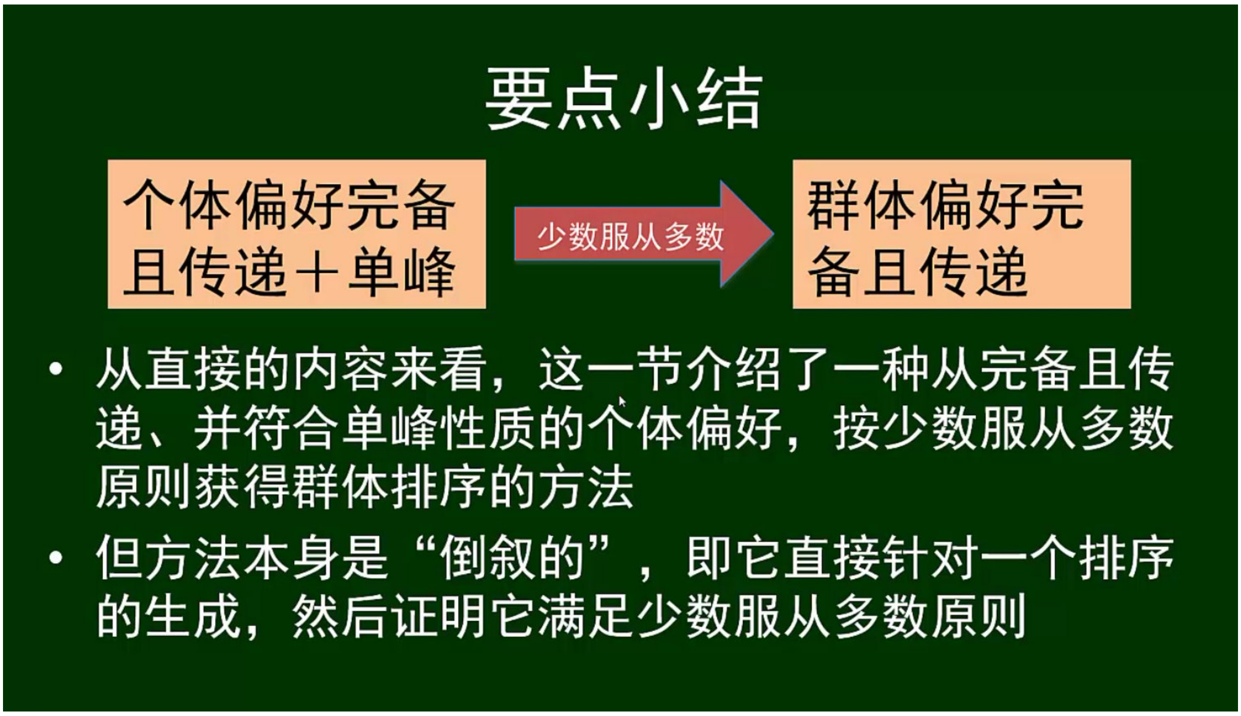
从单峰个体排序形成群体排序的一种方法

- 设
  - N个候选项，$X_1, X_2, \ldots\ldots X_N$，其中，下标序与某种特征属性序相同
  - M个（奇数）表决人，他们在上述候选项序上的偏好是（按该特征序）单峰的
- 构造一个群体排序表（即一个完备且传递的关系），满足：若其中若$X_i>X_j$，则在M个个体排序中的多数都有$X_i>X_j$。
  - 即两两都满足“少数服从多数”原则
- 下面是一个流程，说明存在性和具体结果

单峰偏好下群体排序的形成

- 要点：逐次找出“最大的”（群体意义）
- 记$L_1, L_2, \ldots, L_M$为个体排序表，$L_i(1)$为对应个体表中第一个（最大的）元素
- 将$L_i(1)$, $i=1,2,\ldots,M$按照$X_1, X_2, \ldots\ldots X_N$的特征序排列（一共M个，有的X可能有多次出现）
- 从如此排列的M个元素中取中间项为群体排序的第一项（最大的元素）
- 从$L_1, L_2, \ldots, L_M$中删除该元素，留下的依然是单峰排序表，接着可以取出第二个，等等


概念示例：个人排序表（M=5）
L1 L2 L3 L4 L5
X2 X1 X1 X3 X2
… … X2 … …
… X2 … … …
… … … X2 …
… … … … …
假设它们都满足以 X1 >… >… X2 … >… >X3 为特征序的单峰性
考察这些表中第一个元素按该序排列的情况：X1, X1, X2, X2, X3
X2是“中位项”，因此X2是第一个佼佼者
从每个表中删除它，继续这种中位项提取，得第二佼佼者，…
L1 L2 L3 L4 L5
X1 X1 X3
… … … …
… … … …
… … … …
… … … … …


要点小结
个体偏好完备且传递＋单峰
少数服从多数
群体偏好完备且传递
从直接的内容来看，这一节介绍了一种从完备且传递、并符合单峰性质的个体偏好，按少数服从多数原则获得群体排序的方法
但方法本身是“倒叙的”，即它直接针对一个排序的生成，然后证明它满足少数服从多数原则

12.10 中位项定理

中位项定理

对候选项集合A，给定一组（M奇数个）满足单峰偏好的排序，取其中每个排序的第一名，将它们按照单峰所依照的性质高低排列，那么，中间那一个（X），与A中任何其他元素（Y）相比，在M个排序表中一定有多数给出了X>Y。

中位项定理

- 即要说明，相继取出的那些“中间项”，在少数服从多数原则下，比其他所有还剩下的候选项都要大
  - 记L_1, L_2, …, L_M为个体排序表，$L_i(1)$为对应个体表中第一个（最大的）元素
  - 将$L_i(1)$, i=1,2,…,M按照候选项X_1, X_2, …… X_N的特征序排列（一共M个，有的X可能有多次出现）
  - 从如此排列的M个元素中取中间项为群体排序的第一项（最大的元素）
  - …

中位项定理的证明

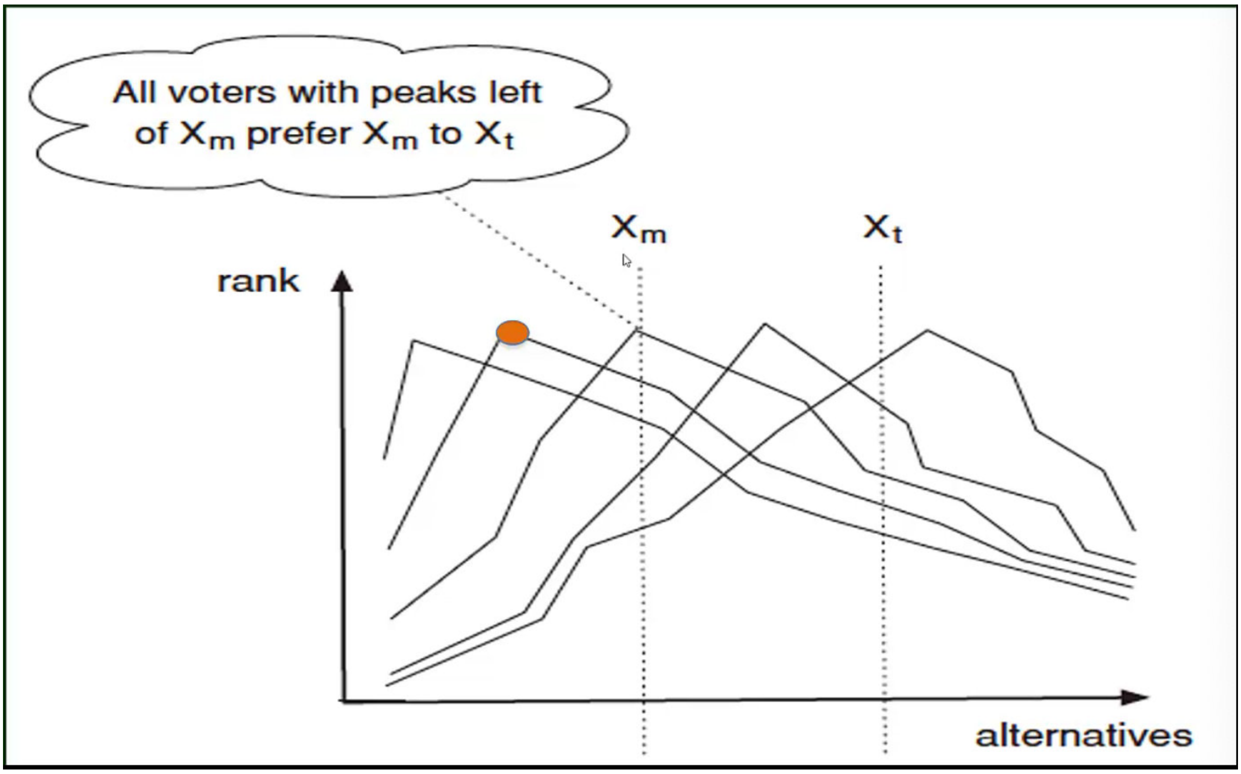
- 记$L_1, L_2, \ldots, L_M$为个体排序表，$L_i(1)$为对应个体表中第一名
- 将$L_i(1)$, $i=1,2,\ldots,M$按照候选项$X_1, X_2, \ldots\ldots X_N$的特征序排列（一共M个，有的X可能有多次出现）
- 只需说明，当$L_i(1)$, $i=1,2,\ldots,M$如此排列后，其中位项与其他M-1项中的不同元素在两两比较中均能基于M个个体排序中的情形，以少数服从多数原则胜出
- 从一个例子看就清楚了，若排列情况如下：

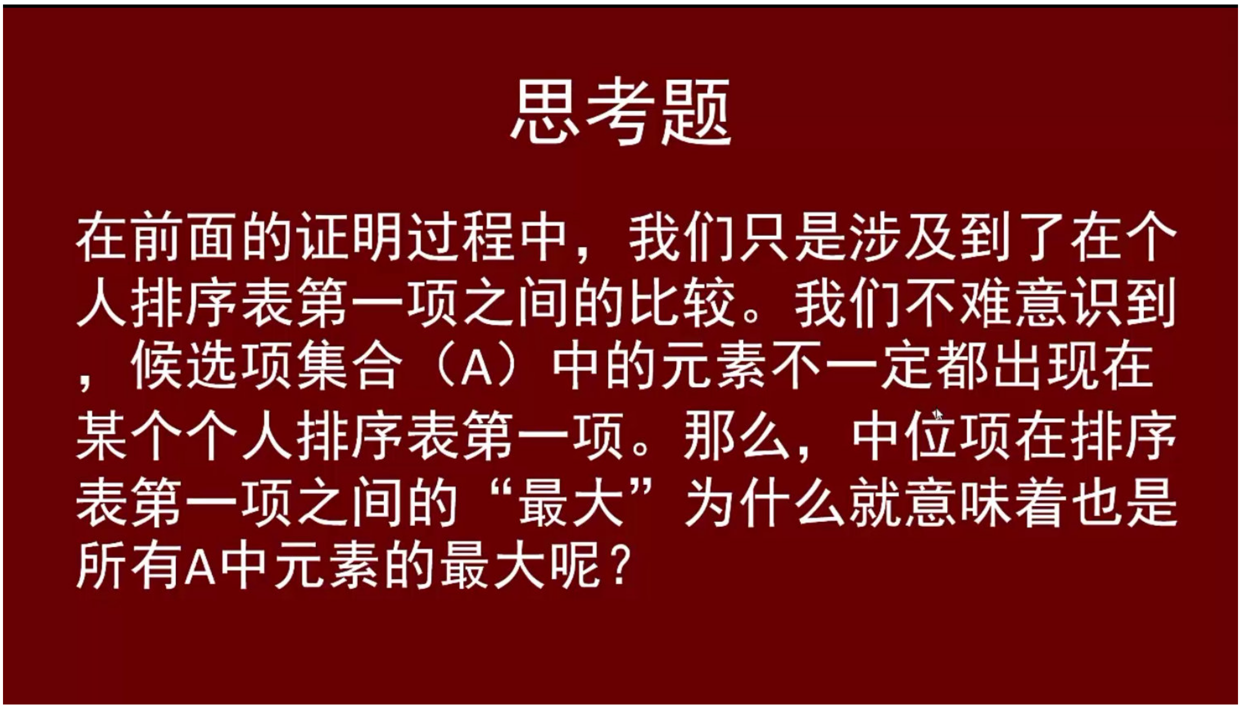
  X_1, X_1, X_2, X_2, X_3 （注：它们是5个个体排序的头）

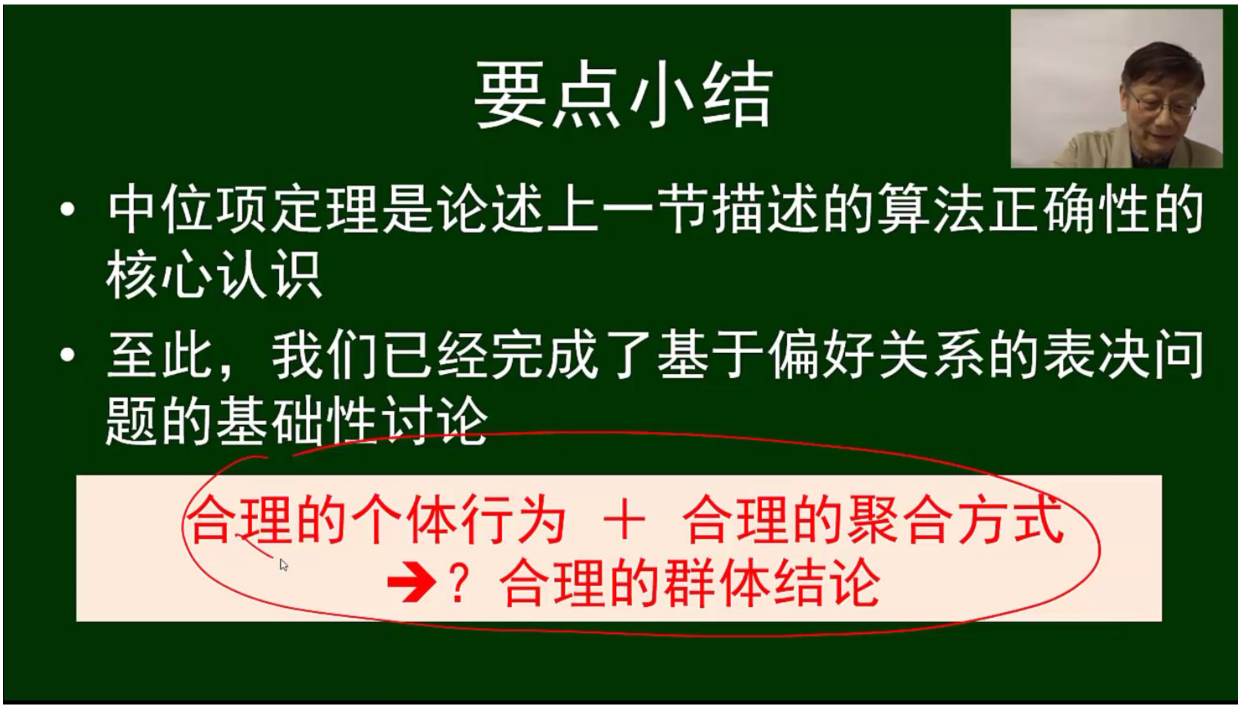
- 为什么说5人中至少3人认为$X_2>X_1$?

  中间那位如何认为？第4位如何认为？第5位如何认为？

  对称的，我们看到中间和左边的人都会认为$X_2>X_3$。



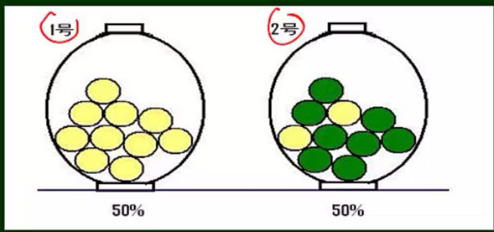


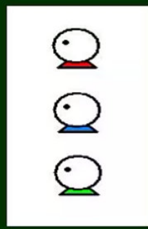


12.11 结果驱动的表决

考虑下面这个例子

普通思维





- 以50%概率拿出其中一个坛子供三人表决用
- 三人依次，随机取一个看看，放回；不交换意见
- 每人给出关于坛子是1号还是2号的判断
- 若多数对了，3人都得奖；否则，3人都受惩罚

有些场合，不一定要按照对“信号”的判断进行表决！

考虑投票问题的两种思路

- 信号驱动
  - 根据得到的信号，我该如何投票？
  - （判断在给定信号下不同选项结果的概率）
- 结果驱动
  - 我的一票在什么情况下起作用？
  - 我该如何投票，以使得那种情况发生时达到正确结果的可能性大些？
  - （从而可能应该忽略信号）

陪审团裁决制度

- 美国的陪审团制度
  - 无罪假设，有罪推定
  - 对刑事案件，需要陪审团所有人都认为被告“有罪”才能定罪（一票否决）
- 法院给陪审员的指导意见
  - 根据所得到的证据，只有“相当程度上怀疑被告有罪”才认为他有罪
  - 而不是“相比无辜而言有罪的可能性较大”

$$\Pr[guilty \mid all - information] > z$$

$$\Pr[guilty \mid all - information] > 0.5$$

你的一票在什么情况下有作用？

- **设**
  - Pr[guilty] = Pr[innocent] = 0.5
  - Gsignal：有罪（G）信号；Isignal：无罪（I）信号
  - Pr[guilty|Gsignal]=q>0.5, Pr[innocent|Isignal]=q>0.5
  - 其他人都诚实投票

你的一票有作用的情形：其他陪审员都得到有罪（G）信号，就你自己得到无罪（I）信号。

下面我们来说明，无论你得到什么信号，总投“有罪”票可能利于集体形成正确裁决

设陪审团共有K个成员

$$\Pr[guilty \mid \text{only I see Isignal}] = \frac{\Pr[guilty]\cdot\Pr[\text{only I see Isignal}\mid guilty]}{\Pr[\text{only I see Isignal}]}$$

$$= \frac{0.5\cdot q^{k-1}\cdot(1-q)}{\Pr[guilty]\cdot\Pr[\text{only I see Isignal}\mid guilty]+\Pr[innocent]\cdot\Pr[\text{only I see Isignal}\mid innocent]}$$

$$= \frac{0.5\cdot q^{k-1}\cdot(1-q)}{0.5\cdot q^{k-1}\cdot(1-q)+0.5\cdot(1-q)^{k-1}\cdot q} = \frac{q^{k-2}}{q^{k-2}+(1-q)^{k-2}} = \boxed{\frac{1}{1+[(1-q)/q]^{k-2}}}$$

由于q>0.5，最后这式子随k→∞而趋向1。

这也暗示，为了避免影响“大局”，人们会倾向于忽略得到的无辜信号，但这样就可能使无辜者获罪了。

要点小结

- 这一节，我们根据表决的两种不同用途（形成共识，找到真相），指出了行使表决的两种不同策略：基于信号，面向结果
- 为了得到较好的结果，人们可能会刻意忽略自己得到的信号
- 但当人们都倾向于这样做的时候，产生正确结果的可能性也许反而会降低