

网络级联

姜少峰



北京大学前沿计算研究中心

Center on Frontiers of Computing Studies, Peking University

新事物的传播：早期研究

- Diffusion of innovations
- 新的行为，方法，意见，标准，技术，如何在社会网络人与人之间传播？
- 早期研究：基于信息效应的传播（类似信息级联）
- 农场主何时、依据什么决定采用一种新高产作物品种 (Ryan, Gross 1943)
 - 多数农场主虽然最早从推销员知道的新品种，但是实际决定是看邻居/朋友农场主的经验做出的
- 类似情况发生在医生临床采用一种新药 (Coleman, Katz, Menzel 1966)

早期研究进一步的发现/讨论

- 初始使用因新事物的未知性存在使用风险，但最终会受益
- 早期使用者都具有很好的社会经济地位，并且善于联络交际
- 多数最终决定采用的人都基于他们的邻居/同事等社会网络结构

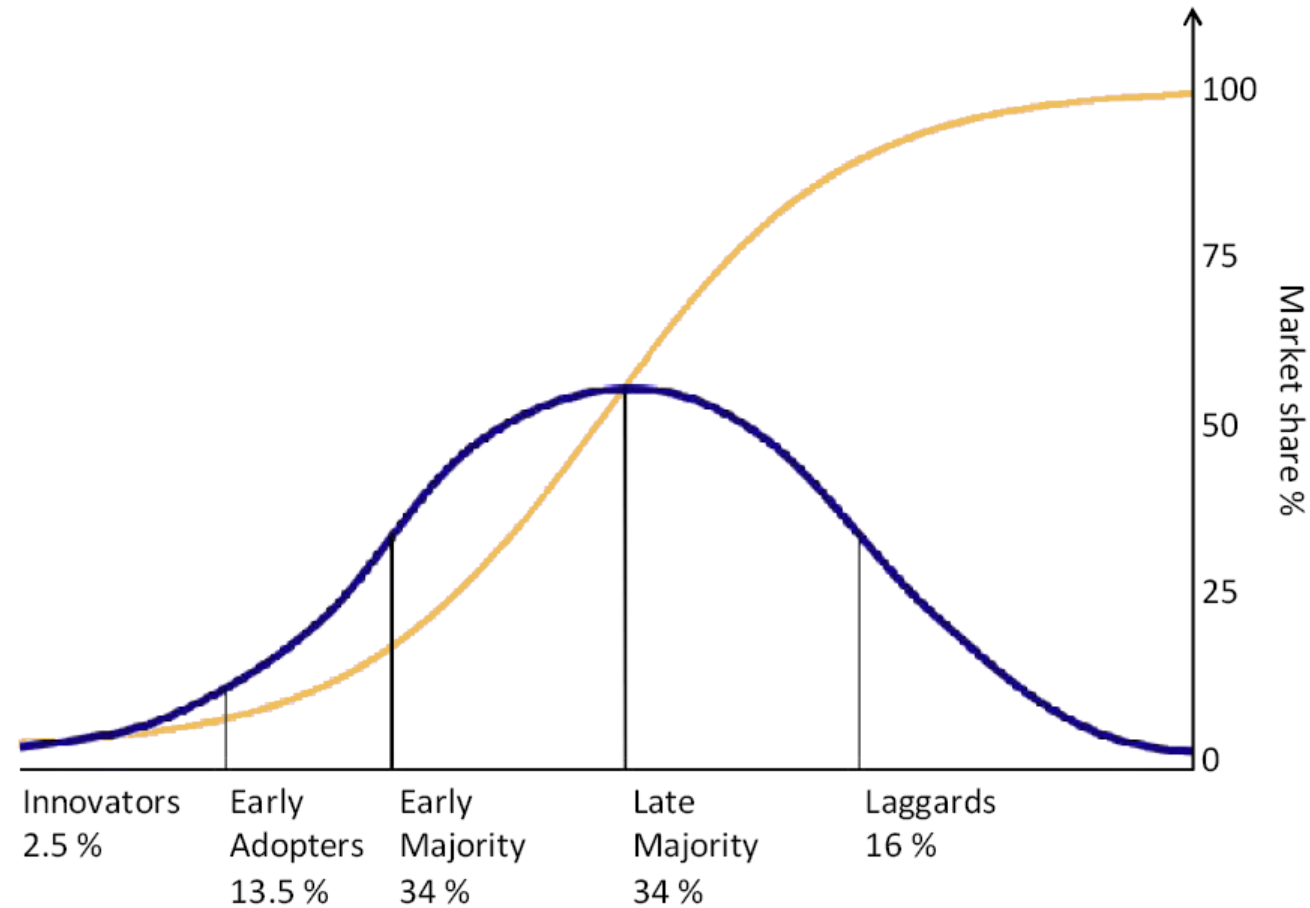
另一类研究

- 类似“**网络效应**”：采用新事物本身对自身是有**直接收益**的，不需要**参考别人的意见**来决定是否有收益
- 研究表明：例如电话，传真，电子邮件等的传播依赖于人们与其他使用这些技术的朋友的交流需要

- Everett Rogers: 《Diffusion of Innovations》
- 人们采用/理解新技术的“复杂性”
- 新技术的“可见性”，即人们是否容易知道别人是否采用
- “可处理性”，人们可在逐渐使用新技术的过程中处理带来的风险
- 与现有社会系统/网络的“兼容性”
 - 例如社会网络的“同质性”可能阻止新技术的传播：新技术都是从“外来”群里进来，而同质性的内部良好互联导致了巨大的传播壁垒

Everett Rogers的新事物曲线

Diffusion of innovations (Everett Rogers)



新事物传播的网络建模 (Morris 2000)

- 从个人的决策角度建模：每个人的决策基于他在网络中的邻居
- 两种个人决策机制：
 - 信息效应（是否有益取决于别人经验）
 - 直接受益（类似“网络效应”）
- 我们聚焦于“直接受益”

“只要自己足够多的邻居采用了新技术，我就有足够的益处采纳”

- 例如大家都用微信交流工作（不用邮件），我也该用微信

一个定义在网络上的协调博弈

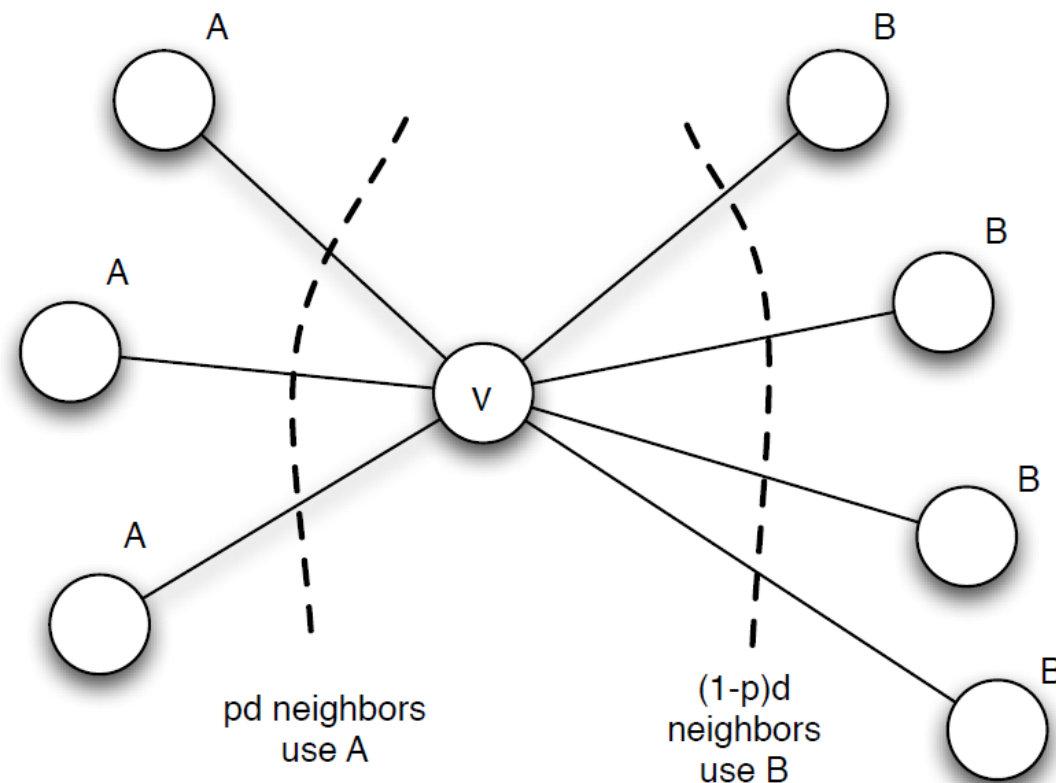
- 模型中每个节点/人需要从A和B两个行为中做出选择
- 如果节点v和w有边相连，则他们有意愿使行为一致
 - 若v和w都采用了A，则他们得到收益 $a > 0$
 - 若都采用B，则都得到收益 $b > 0$
 - 若采用了相反的行为，收益都是0
- 以上博弈是定义在一条边上的
- 对于整个网络，该博弈在每条边都独立进行
- 对于某个点v，考虑v在所有相邻边的博弈的总收益
 - v的策略应该取决于它所有邻居

	w选A	w选B
v选A	a, a	0, 0
v选B	0, 0	b, b

节点v的策略

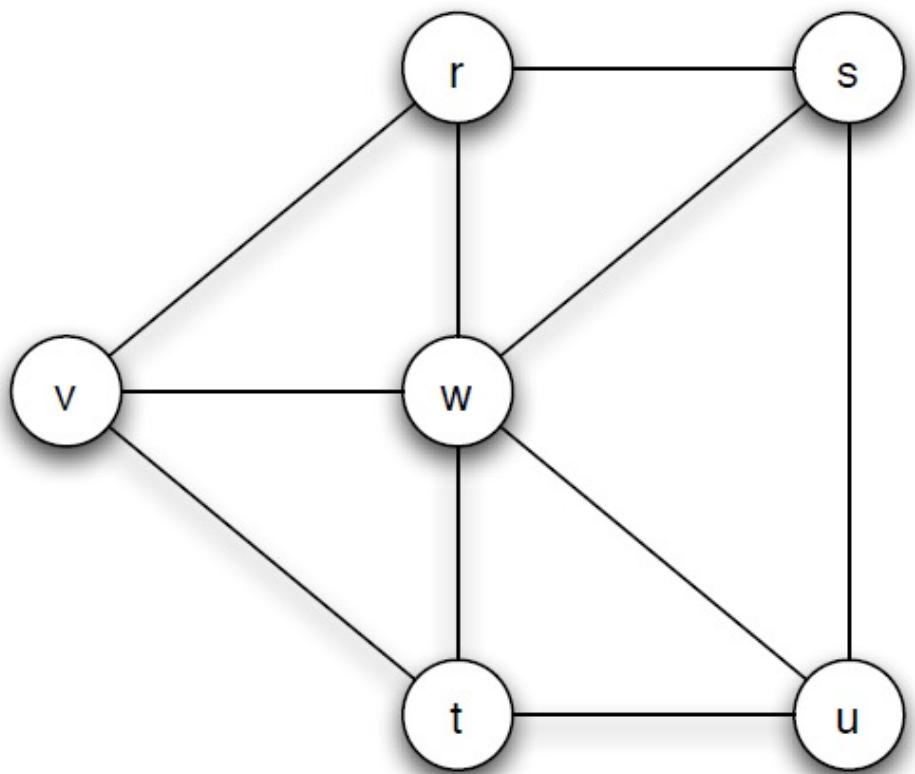
- 节点v为最大化自己的收益，需要考虑邻居中有多少比例采用A和B
- 设v有 d 个邻居，并有 p 比例采用A， $1 - p$ 比例采用B
- v采用A收益 pda
- v采用B收益 $(1 - p)db$
- $p \geq \frac{b}{a+b}$ 时选A，否则选B
- 设该阈值 $\frac{b}{a+b}$ 为 q

	w选A	w选B
v选A	a, a	0, 0
v选B	0, 0	b, b

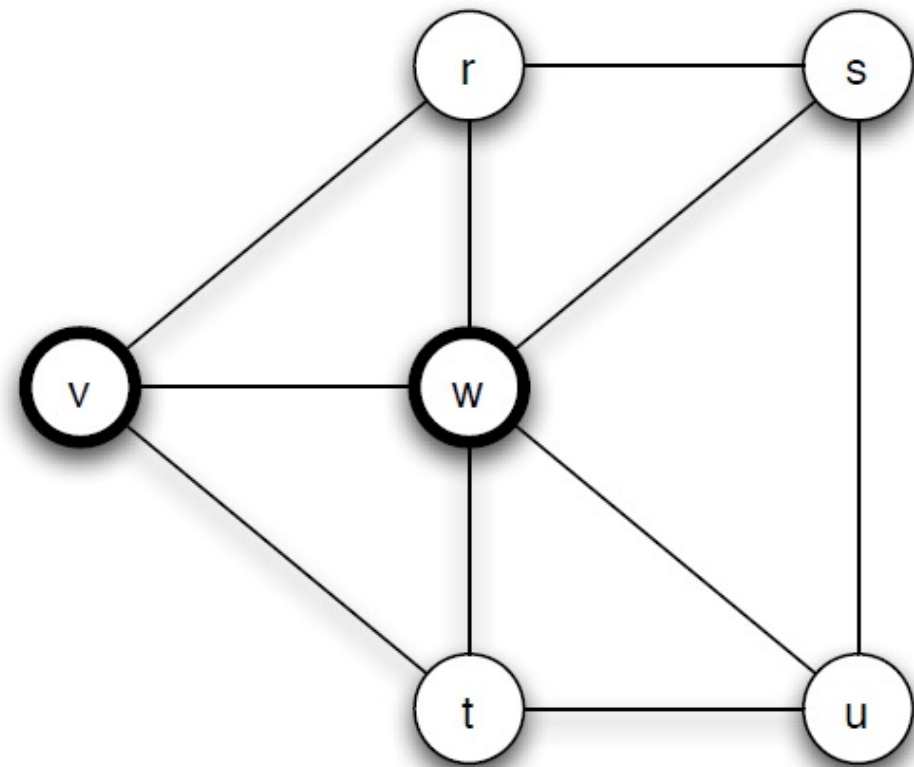


级联行为

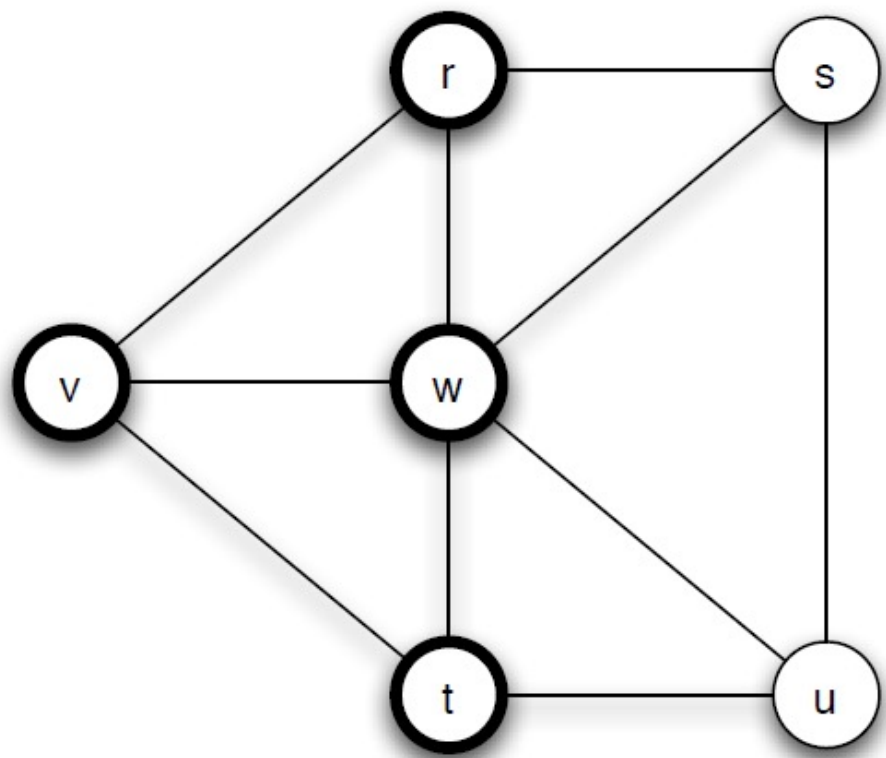
- 两个显然的均衡：所有人都用A，或者所有人都用B
- 考虑如何从一个态B “转化” 为另一个态A
- 初始态：所有人采用B
- 一小部分 “初始点集” S换到了A（独立于系统，不经过协调博弈）
- S的邻居进行博弈，决定是否采用A，产生S'
- 在S'继续，循环往复，直到采用A/B的情况不再变化
- 最后是否全部都改成了A？ 还是中间停止 “传播”？ 条件是什么？



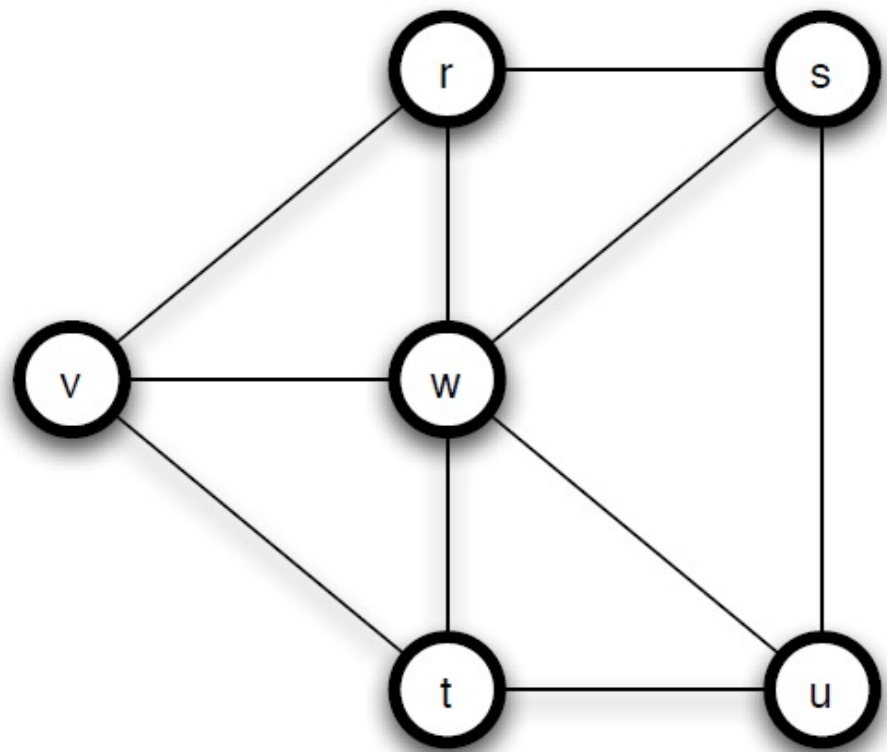
(a) *The underlying network*



(b) *Two nodes are the initial adopters*

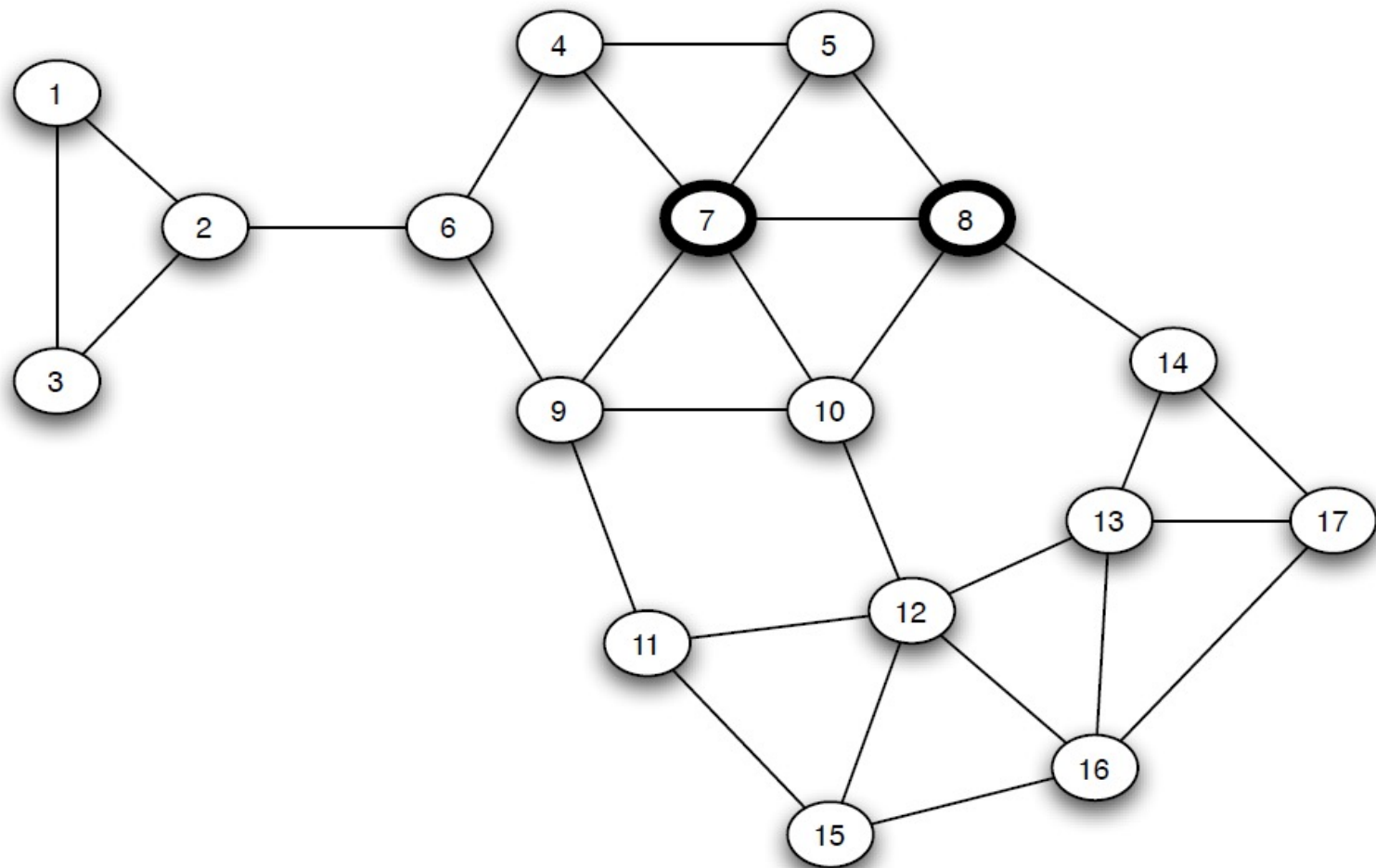


(c) *After one step, two more nodes have adopted*

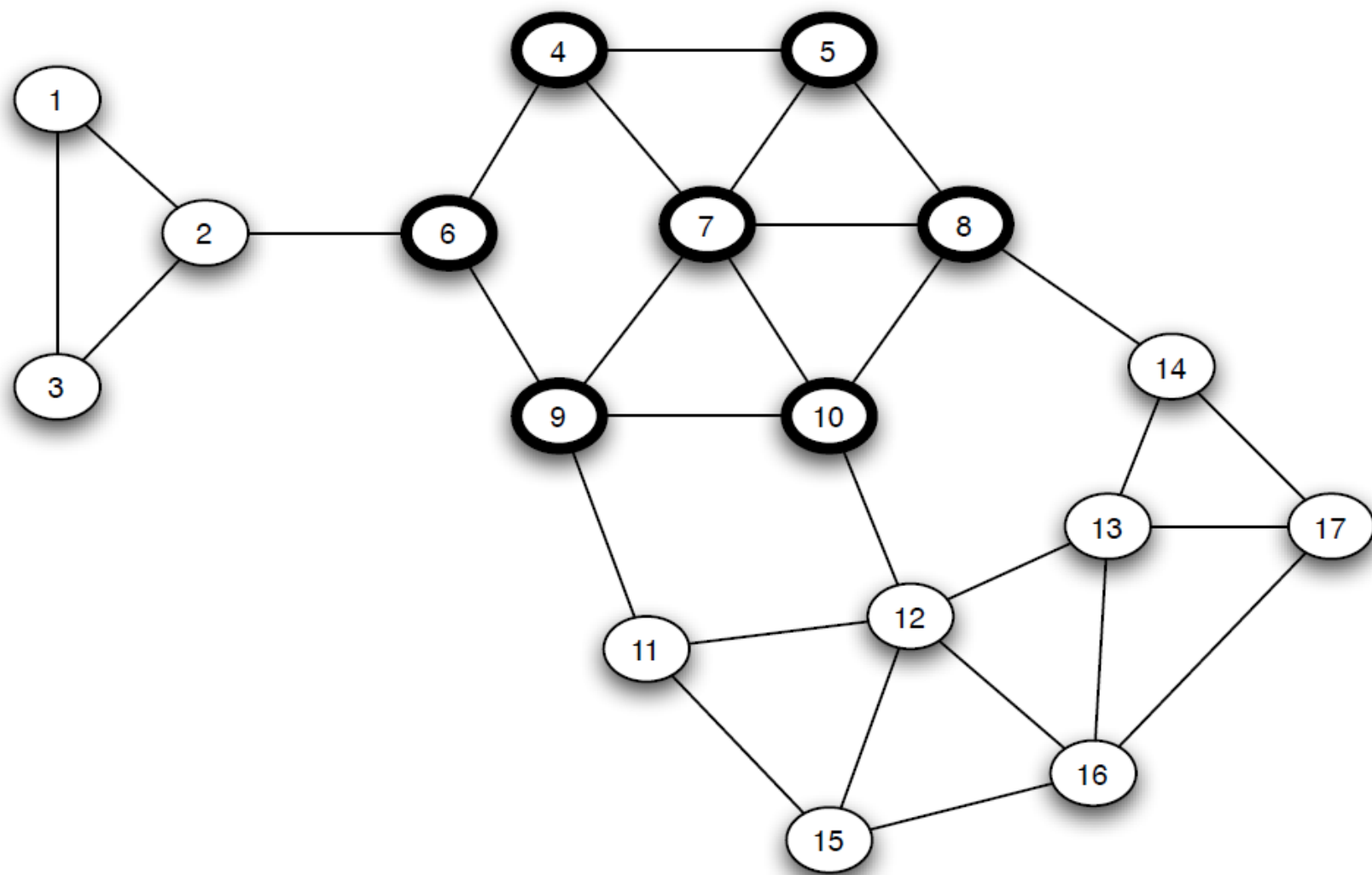


(d) *After a second step, everyone has adopted*

设 $a = 3, b = 2$, 那么阈值 $q = \frac{b}{a+b} = \frac{2}{5}$



$$a = 3, b = 2, \text{ 閾值 } q = \frac{2}{5}$$



- 称转换到A的连锁反应为A的“级联”
- 我们需要区分下面两种情况：
 - 运行足够长时间后停止，仍有部分人采用B
 - 运行足够长事件后停止，所有人都采用A

如果最后所有人采用A，则称初始使用者集合S及阈值 q 引起了一个完全级联

网络结构的影响

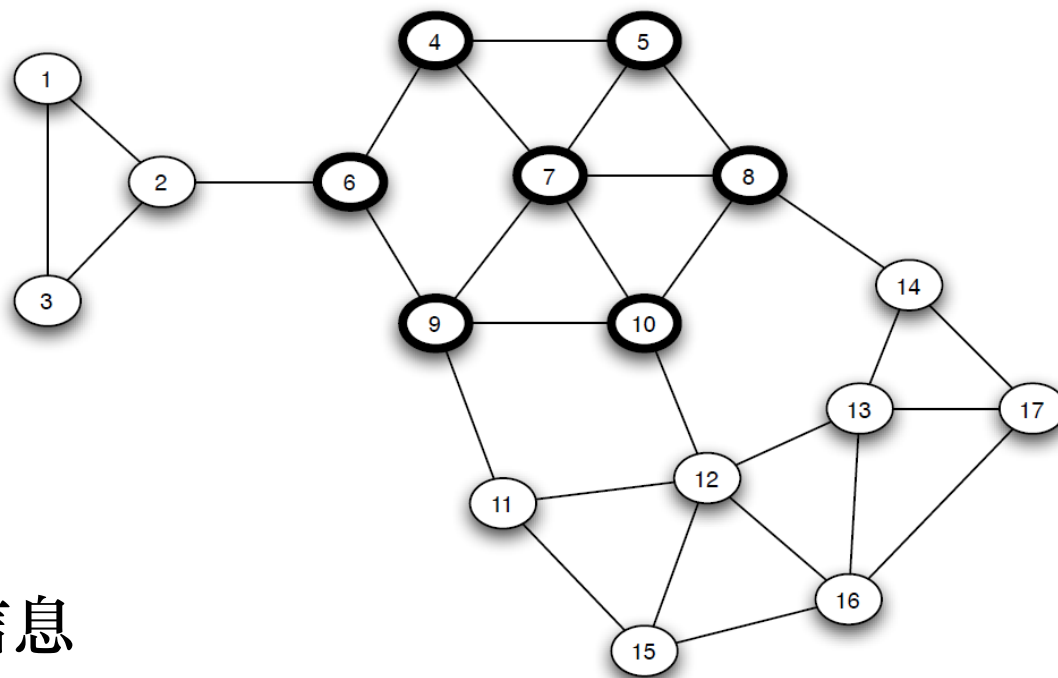
- 新事物容易在紧密相连的区域传播，但是很难突破“边界”
 - “同质性”的重要结构作用

如何突破壁垒/病毒营销？

- 提高A的“质量”：a从3变成4， $q=1/3$ ，所有人都用A
- 转化“关键”节点：12；11？

与网络效应的对比？

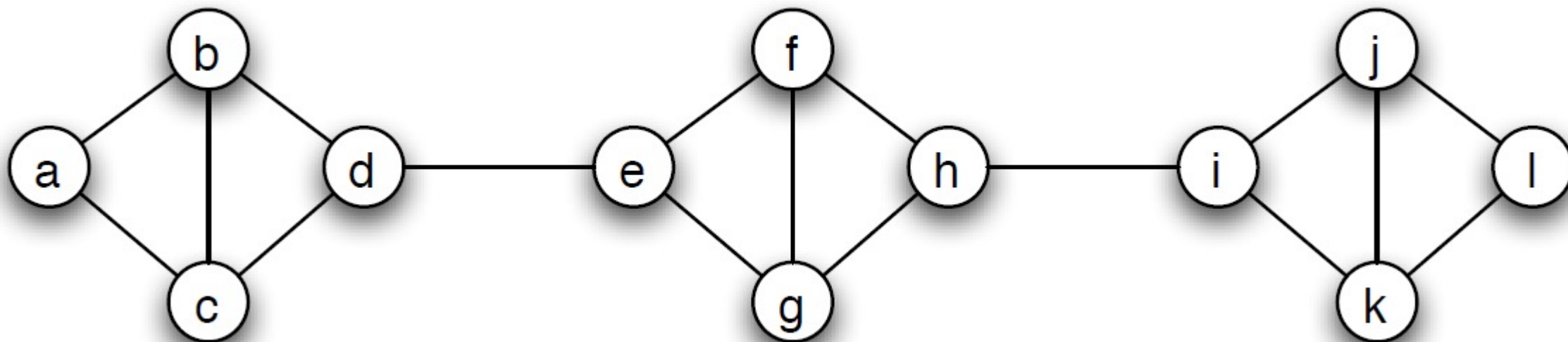
大家仅关注局部邻居，对总体没有任何预期/信息



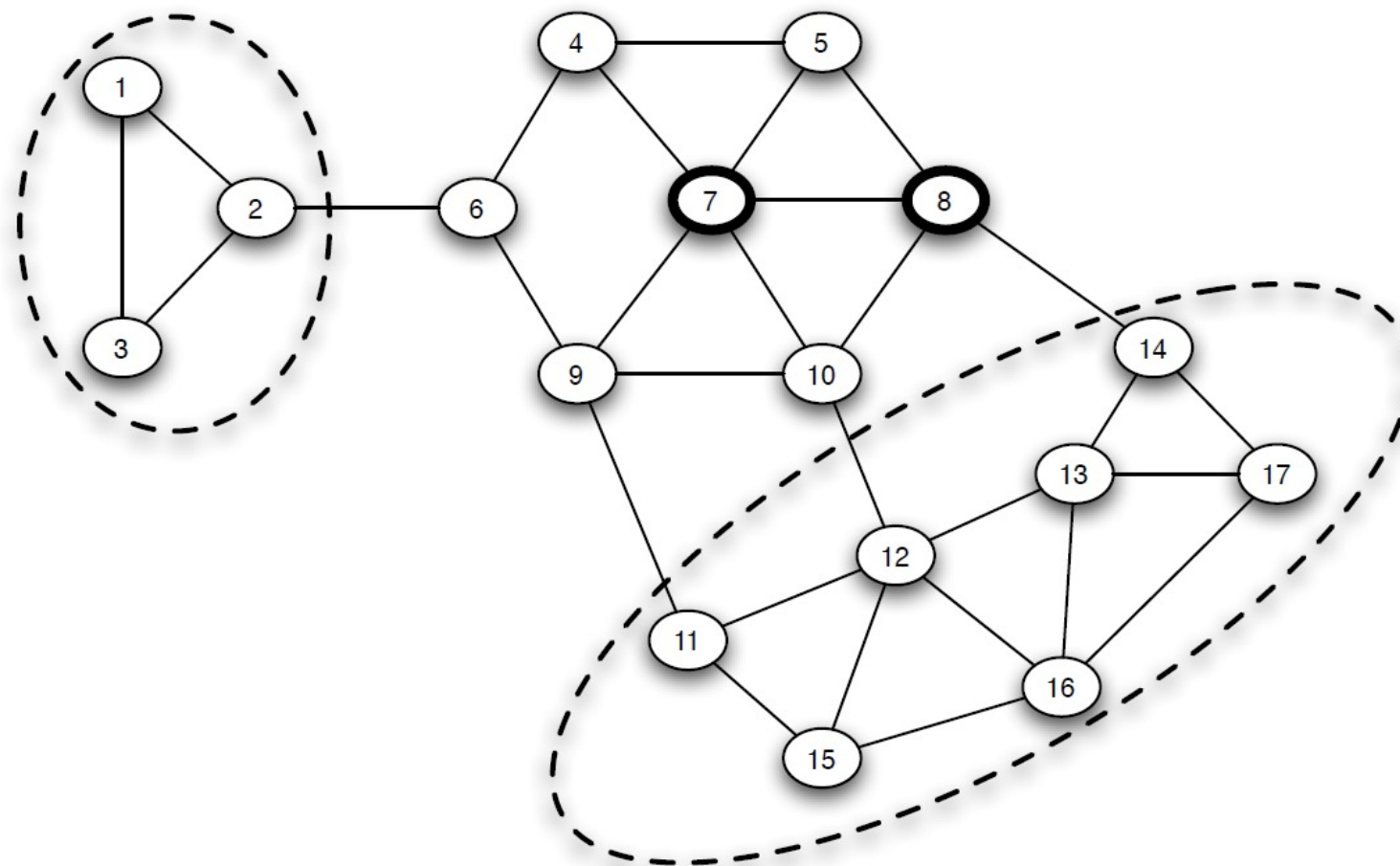
“紧密”连接的区域

- 定义“密集连接区域”

称一个节点集S为密度 p 的cluster，若S每个节点至少有比例 p 的邻居属于S



3组4点组成的、密度为 $\frac{2}{3}$ 的cluster



两个密度为 $2/3$ 的cluster，恰好对应于级联过程所不能到达的点集

定理：

设 S 是行为 A 的初始节点集，且剩余节点采用 A 的阈值是 q

1. 若剩余网络包含一个密度大于 $1 - q$ 的cluster，则 S 不能产生完全级联
2. 若 S 不能产生完全级联，则剩余网络必然包含一个密度大于 $1 - q$ 的cluster

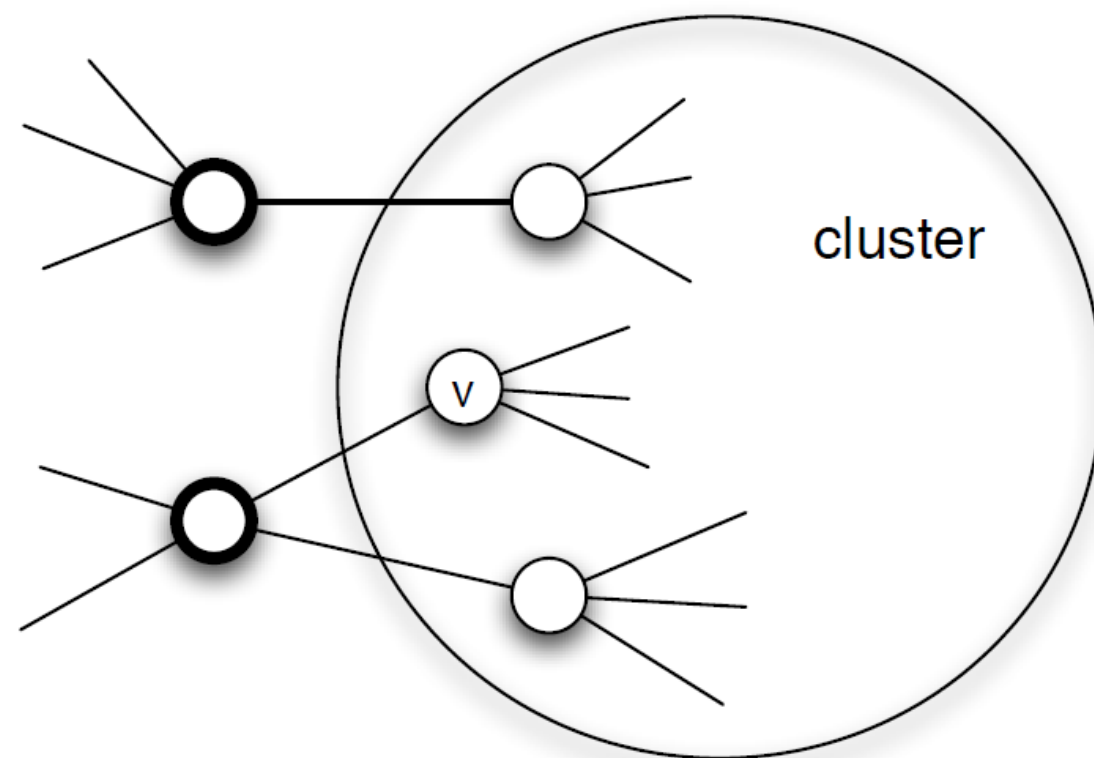
证明：cluster是级联的障碍

- 设行为A初始集合S阈值 q ，剩下网络中有一个密度 $> 1 - q$ 的cluster C
- 证：该cluster C的所有点都不能采用A

反证：考虑C中第一个采用A的点 v

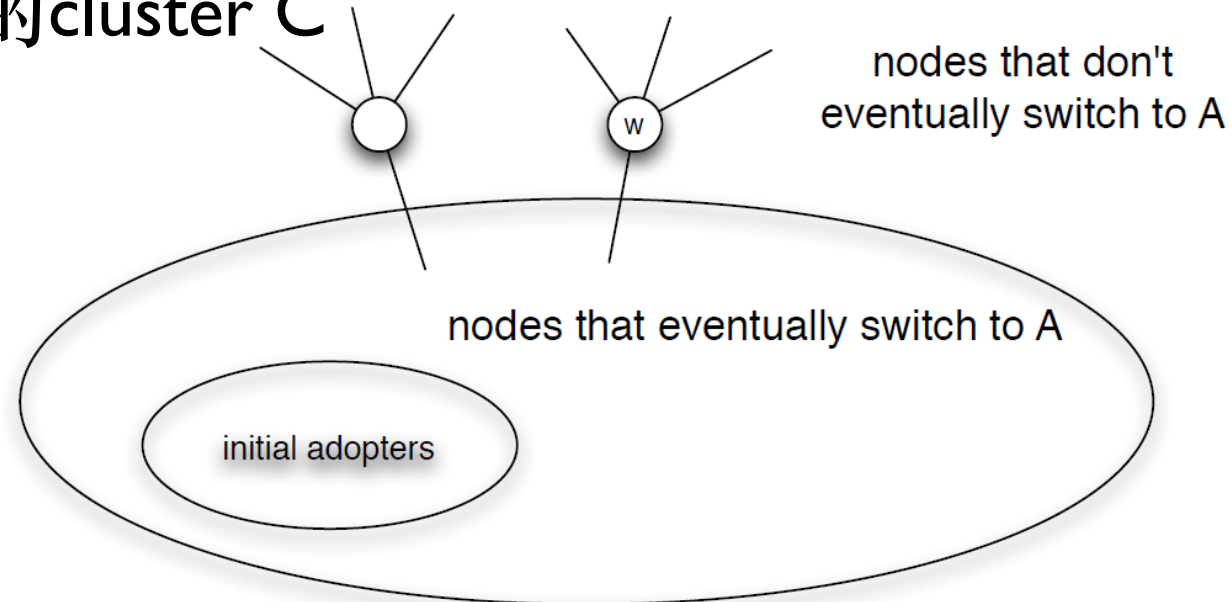
- v 采用A完全基于 $t - 1$ 时刻采用A的点集，而此时只有C以外的点可能采用A
- 因C的密度大于 $1 - q$ ，故 v 小于 q 比例的邻居不在C中

因此 $t - 1$ 时刻采用A的 v 邻居小于 q 比例，矛盾！



Cluster是级联的唯一障碍

- 设行为A初始集合S阈值 q ，S不能产生完全级联
- 证：必然存在一个密度大于 $1 - q$ 的cluster C



考虑级联过程停止的状态

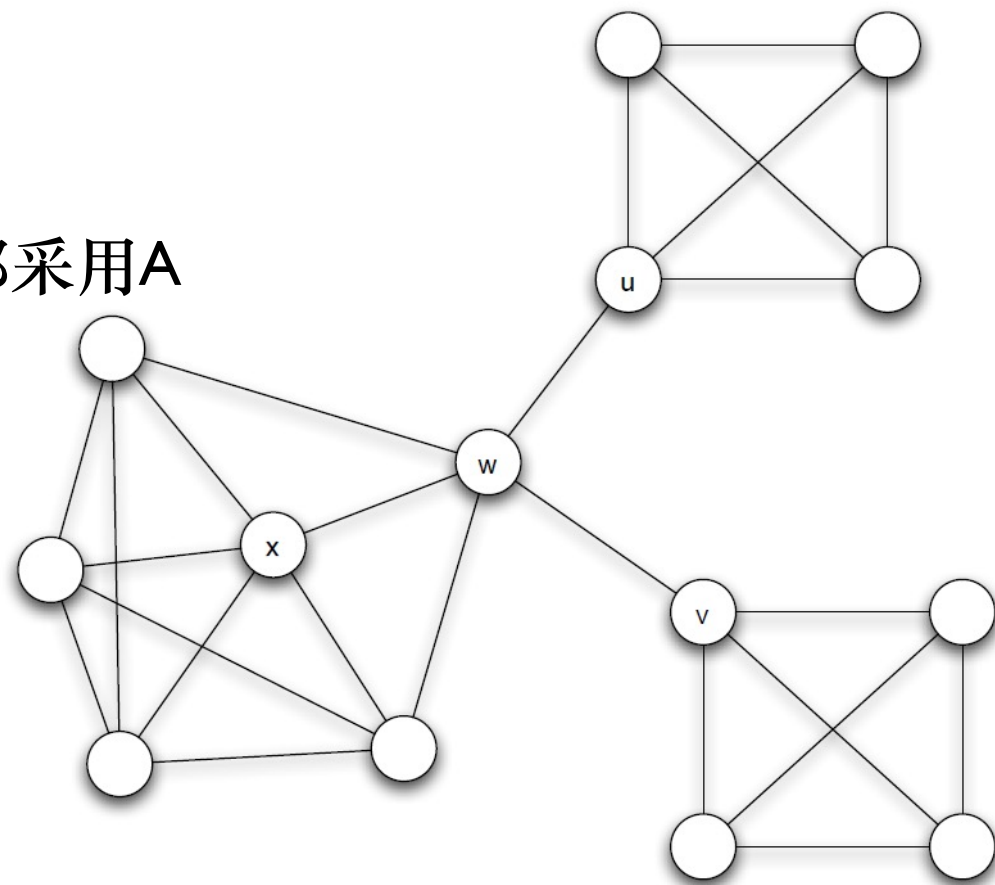
令C为此时仍使用B的人的集合

考虑C中任意点 w

- 必然有采用A的比例小于 q
 - 因此 w 采用B的邻居比例大于 $1 - q$ ，并且全在C里
- 因此C就是一个密度大于 $1 - q$ 的cluster!

结构的重要性 vs 阈值的重要性

- 结构洞：处于几个cluster交界处的点具有信息上的优势
- 考虑 w x 是初始点
- v 可能通过 w 获知 w 的邻居都采用了A
- 然而，如果加入 $q = 0.5$ 的阈值
 - u 不会采用A，即使 w 告诉 u 自己的3个邻居都采用A
- 说明：
 - 低阈值信息特别容易扩散（视频）
 - 高阈值信息难以跨cluster扩散（政治观点）



级联模型的推广

- **异质阈值**：对于点 v ， v 参与协调博弈的收益是 a_v, b_v

- 如果节点 v 和 w 有边相连

- 若 v 和 w 都采用了A，则他们得到收益 $a_v, a_w > 0$

- 若都采用B，则都得到收益 $b_v, b_w > 0$

- 若采用了相反的行为，收益都是0

	w选A	w选B
v选A	a_v, a_w	0, 0
v选B	0, 0	b_v, b_w

- 对 v 来说，若 v 有 d 个邻居且 p 的比例采用了A， $1 - p$ 比例采用B

- v 采用A仅当 $p \geq \frac{b_v}{a_v + b_v}$ ，这里 $q_v := \frac{b_v}{a_v + b_v}$ 是 v 的**个人**阈值

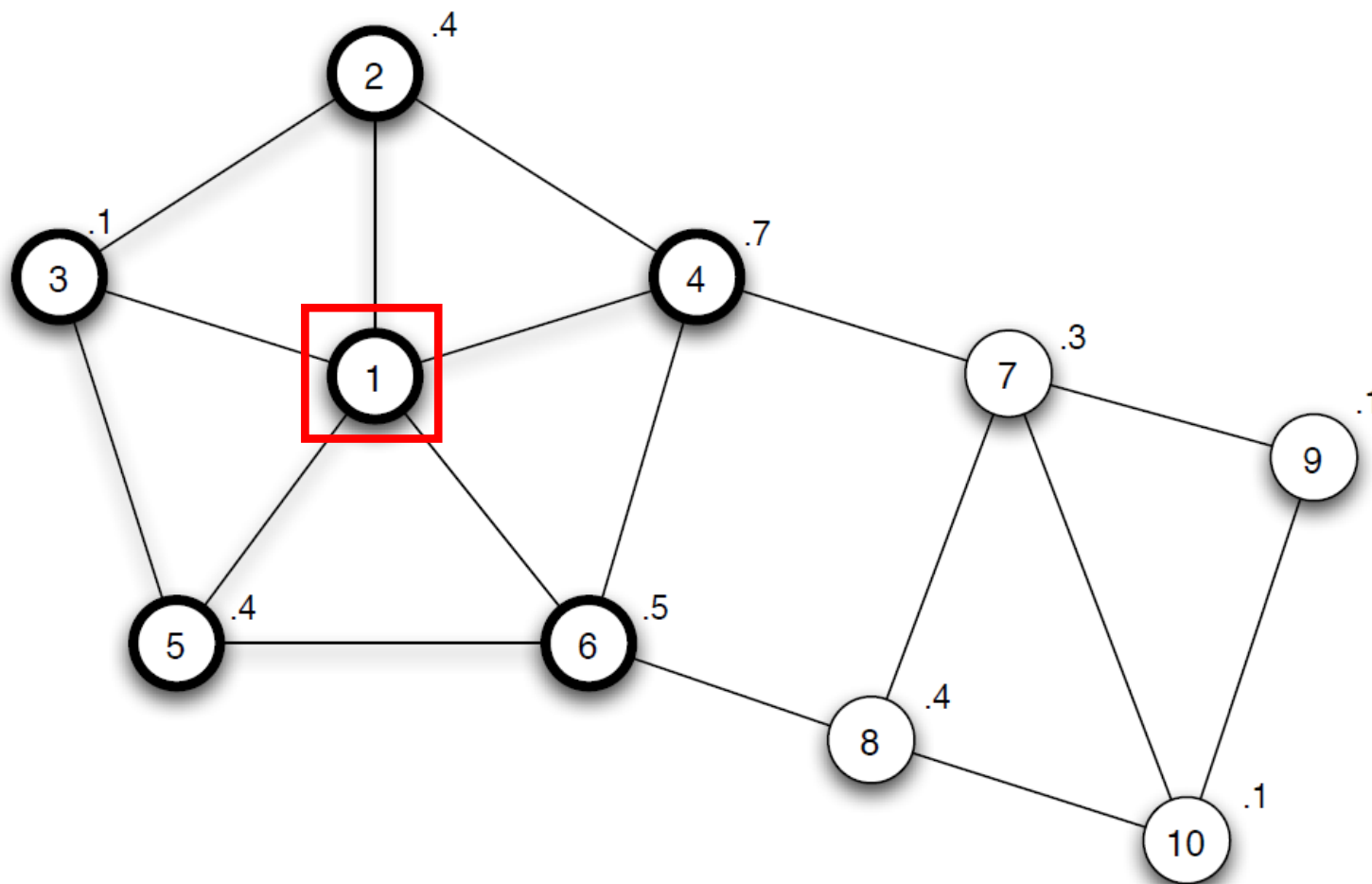
异质阈值模型下的传播过程

- 整个过程依然从一组初始（使用A）节点S开始，且每个 v 有自己的 q_v
- 迭代：每个 v 若没使用A，则独立判断使用A的邻居比例是否超过阈值
- 关心：过程终止后，是否达成了完全级联？

阈值 vs 网络结构

虽然1是“中心”点，但阈值的设定可弱化该结构优势

阈值与结构同时决定级联



推广的完全级联的充要条件

- **blocking cluster**: 点集 C 对任何 $v \in C$ 有多于 $1 - q_v$ 比例/密度邻居在 C 里

定理:

设 S 是行为 A 的初始节点集

1. 若剩余网络包含一个blocking cluster, 则 S 不能产生完全级联
2. 若 S 不能产生完全级联, 则剩余网络必然包含一个blocking cluster

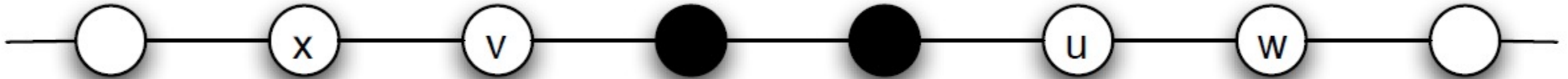
级联能力

- 给定一个网络，使用一个“小”的初始节点集S并产生完全级联的最高阈值是什么？
- 阈值： $q = \frac{b}{a+b}$
- 当采用A的邻居占比至少q就会采用A
- 阈值理解： $1 - q$ 是a相对于b的优势
 - 如果q远低于0.5，那么说明a需要比b有很大的优势
 - 如果q接近0.5，说明a相对b无明显优势
- 确定最高阈值的意义：网络结构的度量；越高越“容易”级联；如果很低，那么网络很“保守”，需要很高的相对优势才能形成级联

	w选A	w选B
v选A	a, a	0, 0
v选B	0, 0	b, b

无限图的级联能力

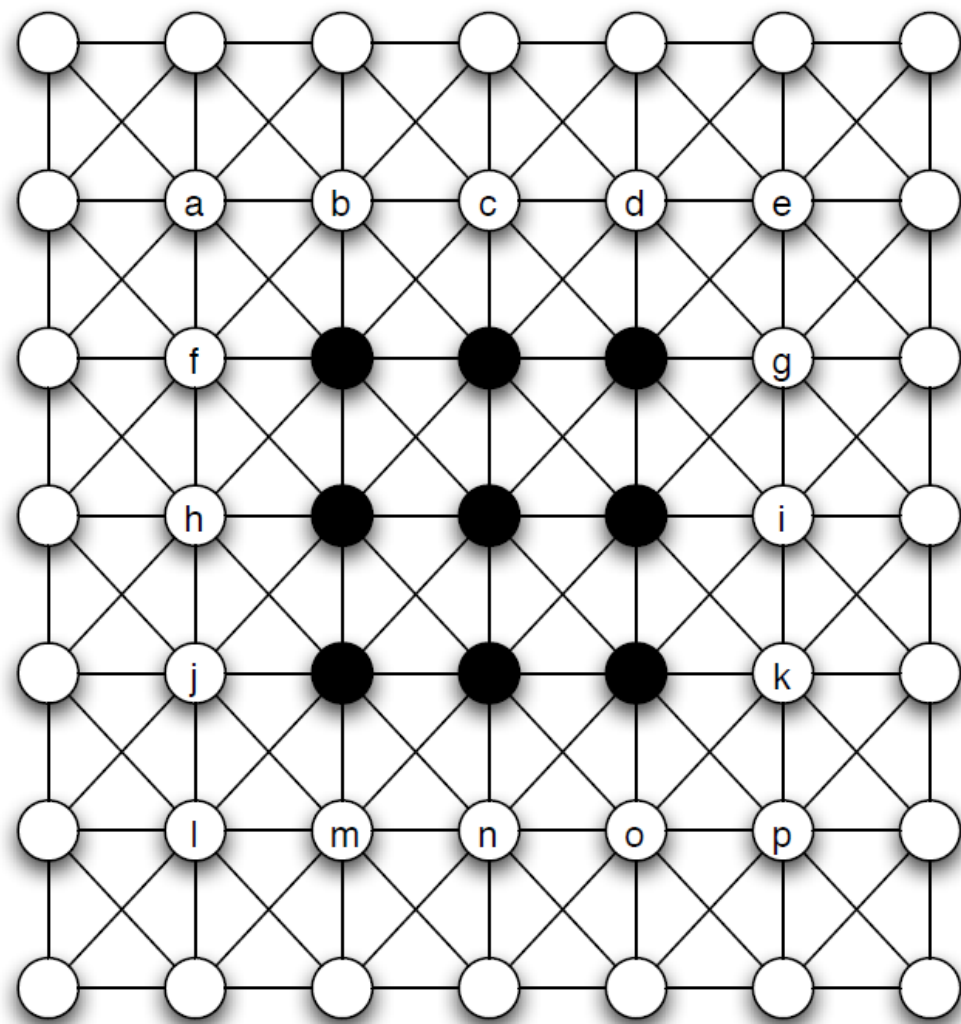
- 如何建模“小”的初始集合？
- 考虑无限图，并考虑任何一个有限的初始集合S
- 定义完全级联：对任意节点v，存在一个时刻t，t步之后v采用A



级联能力/最大阈值 $q = 0.5$

q = 0.5不形成完全级联

级联能力是 $q = \frac{3}{8}$

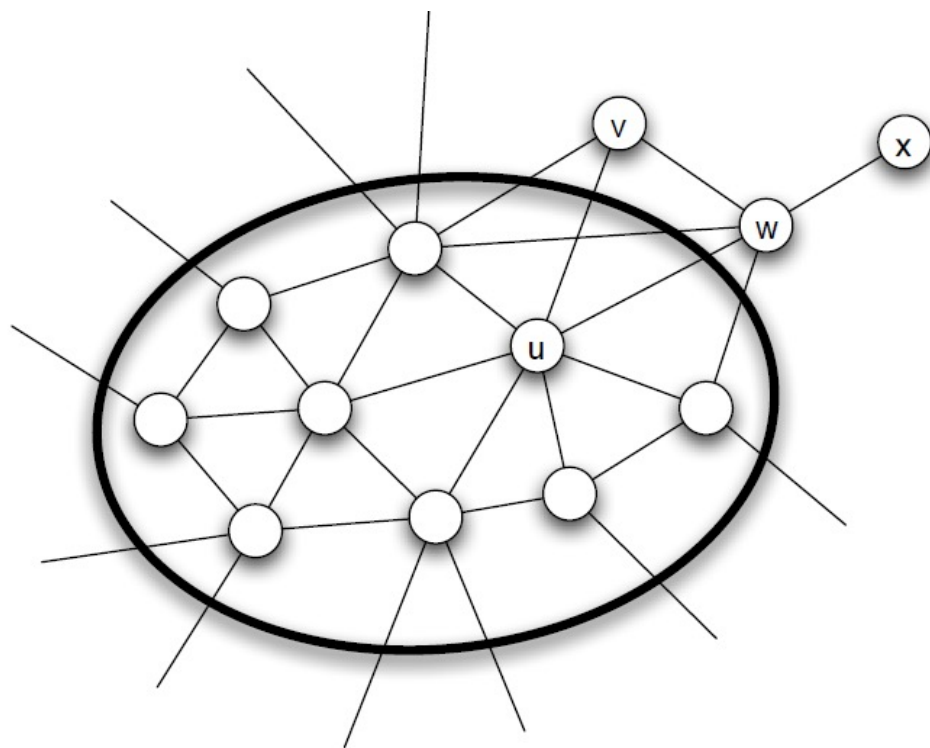


级联能力不会超过0.5

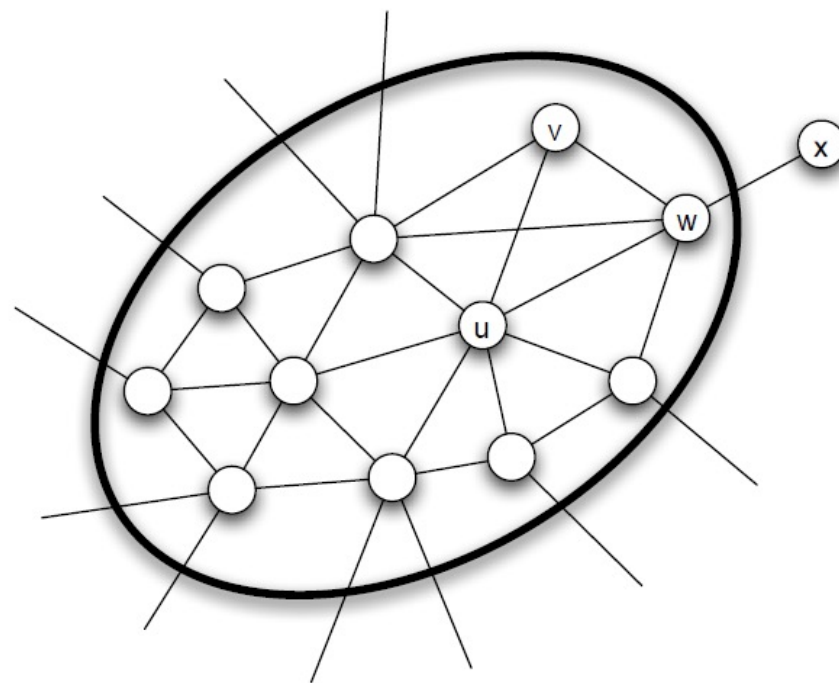
定理：没有网络的级联能力可以超过0.5

- 因此：如果一个新事物要求 $q = 0.5 + \epsilon$ ，也就是若A逊于B，则一定不会完全级联，无论网络的结构如何
- 证明思路：
 - 设A-B边为“跨越”边
 - 跨越边数起始是一个有限的正数，并且永远是非负的
 - 证：跨越边数单调递减

跨越边数单调递减



(a) Before v and w adopt A



(b) After v and w adopt A

考虑一个 w : t 轮还用 B , $t + 1$ 轮改用 A

设 t 轮时 w 有 m 个 A 邻居, n 个 B 邻居, 那么必有 $m > n$

w 改用 A 后跨越边数的变化量 $= n - m \leq -1$

在级联中考虑“兼容性”

- 实际中未必是AB二选一，可能是**AB共存**
 - 同时使用两套社交软件（腾讯会议/Zoom；Whatsapp/微信）
 - 同时运行两种操作系统（一台Windows PC，一台Mac）
- 如何建模这种兼容性？这种兼容性对级联行为有什么影响？
 - **增加一种策略：AB都选**
 - AB vs A的收益是a，AB vs B的收益是b，但AB vs AB的收益是 $\max(a, b)$
 - 另外：**选AB需要付出一个c的代价**（只需要付出一次）

	w选A	w选B	w选AB
v选A	a, a	0, 0	a, a
v选B	0, 0	b, b	b, b
v选AB	a, a	b, b	$\max(a, b)$, $\max(a, b)$

“兼容性”级联的例子

$a = 2, b = 3, c = 1$
r和s是初始点



u和v应该采用AB: $2 + 3 - 1$

w和x应该采用B, 之后邻居也都采用B

	w选A	w选B	w选AB
v选A	a, a	0, 0	a, a
v选B	0, 0	b, b	b, b
v选AB	a, a	b, b	$\max(a, b), \max(a, b)$

注: 选AB需要付出一个c的代价 (只需要付出一次)

- 调高a, 使得 $a = 5$, $b = 3$, $c = 1$?



Start	B	B	B	A	A	B	B	B
Step 1	B	B	AB	A	A	AB	B	B
Step 2	B	AB	AB	A	A	AB	AB	B
Step 3	AB	AB	A	A	A	A	AB	AB
Step 4	AB	A	A	A	A	A	A	AB

无穷路径图上的级联

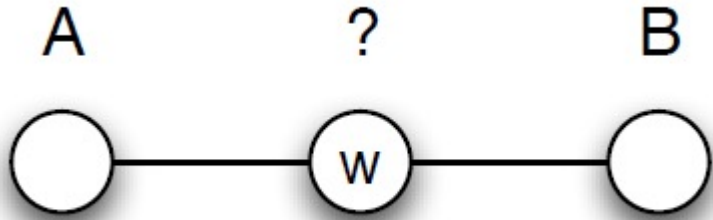
- 无穷路径图：



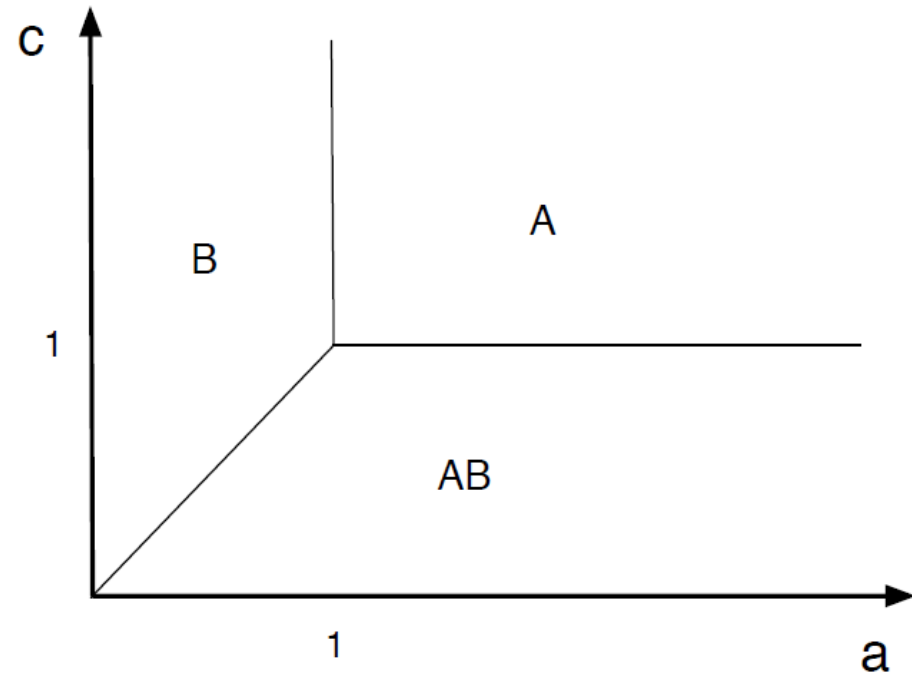
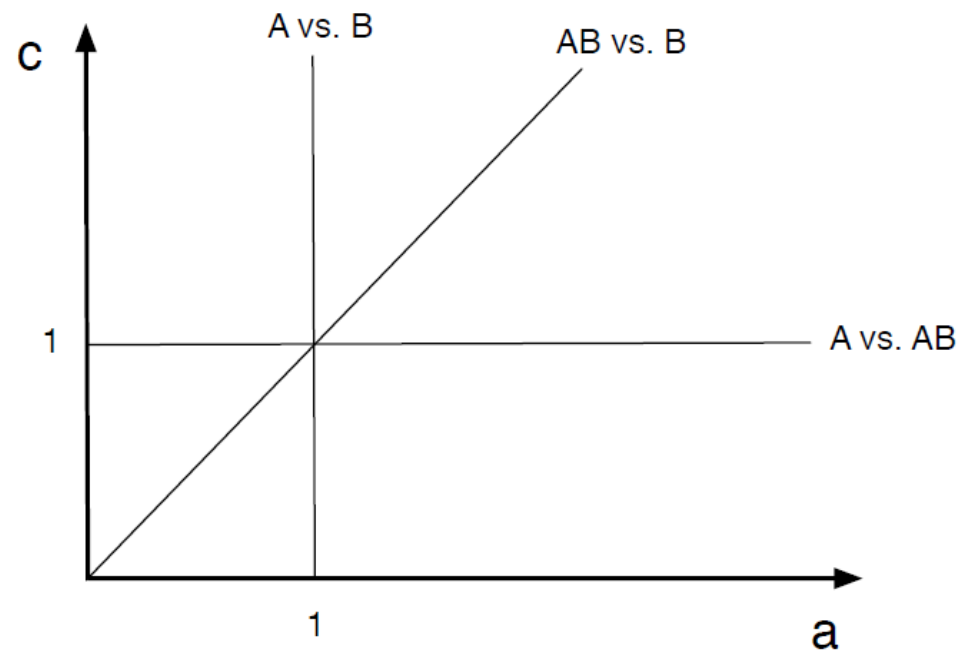
- 完全级联：左右有无穷多点采用A；
- 考虑该图上级联发生的条件：取决于 a ， b ， c 取值
- 不失一般性假设 $b = 1$ （可以通过放缩 a 和 c ）

	w选A	w选B	w选AB
v选A	a, a	0, 0	a, a
v选B	0, 0	b, b	b, b
v选AB	a, a	b, b	$\max(a, b)$, $\max(a, b)$

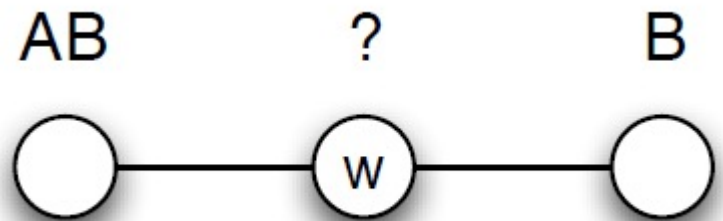
情况一：左右分别是A和B



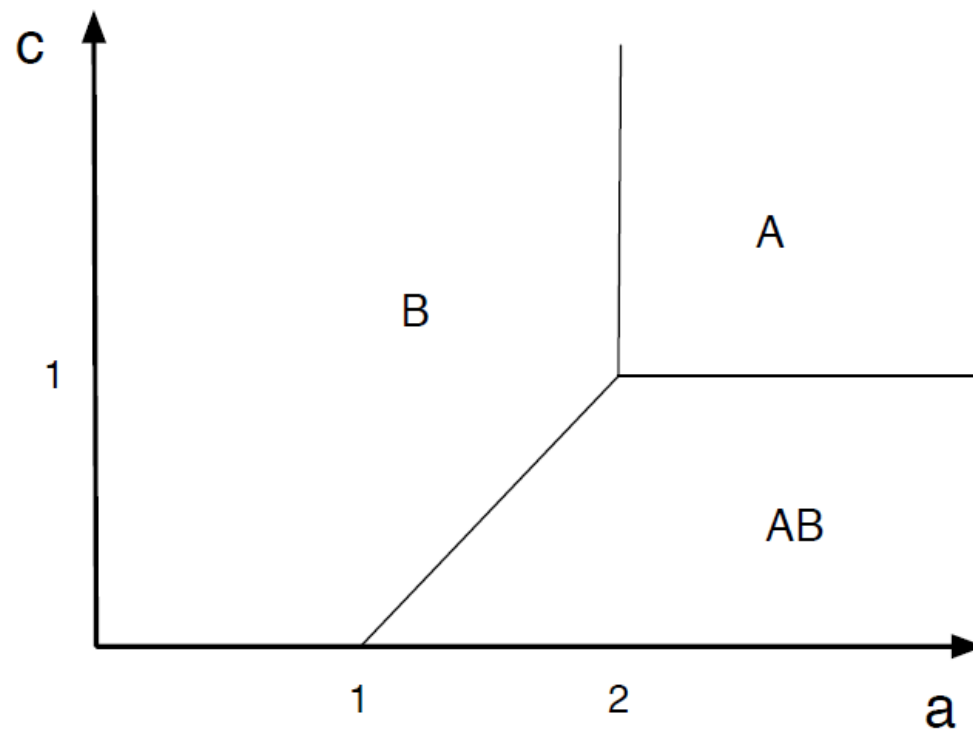
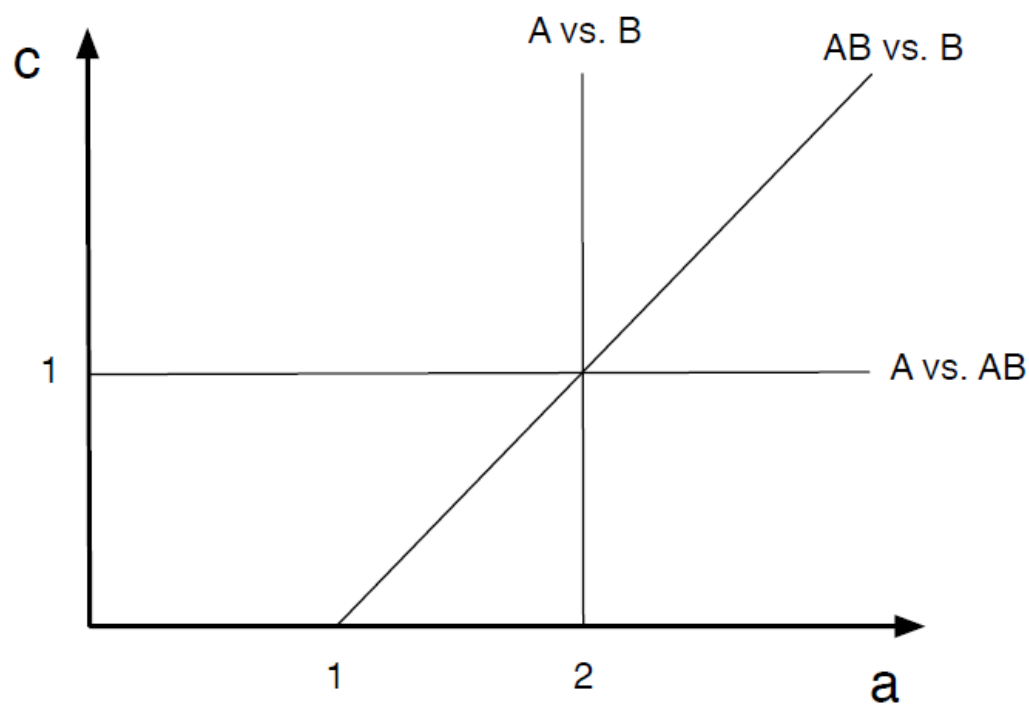
收益分析：w选A得到a；选B得到1；选AB得到 $a + 1 - c$



情况二：左右两边分别是AB和B

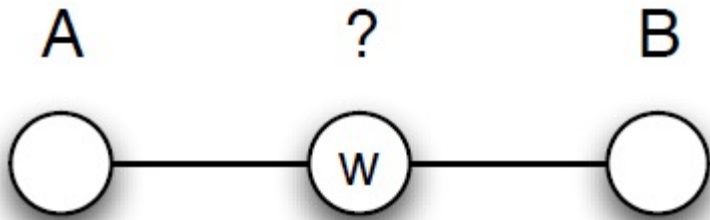


收益分析：w选A得到a；选B得到2；
选AB得到 $a + 1 - c$

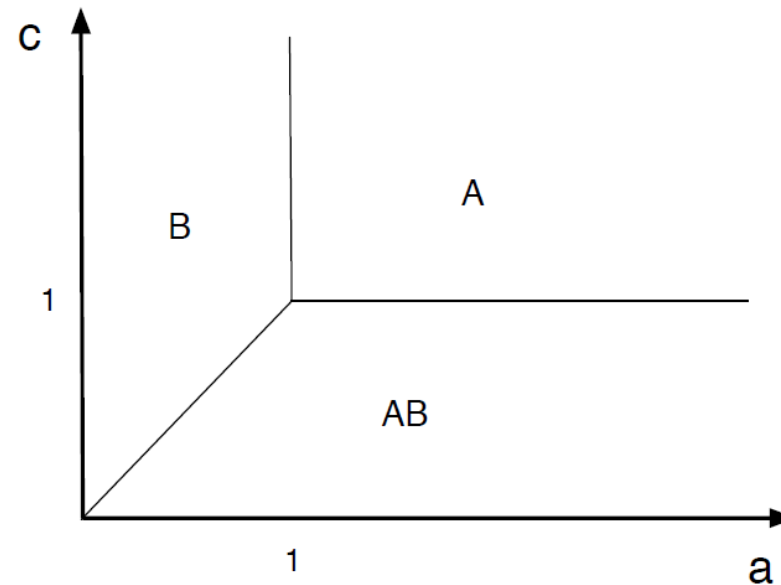


何时产生级联?

- 设有一段连续的初始点集 S 采用 A ，设 w 紧邻 S 右侧（左侧对称分析）

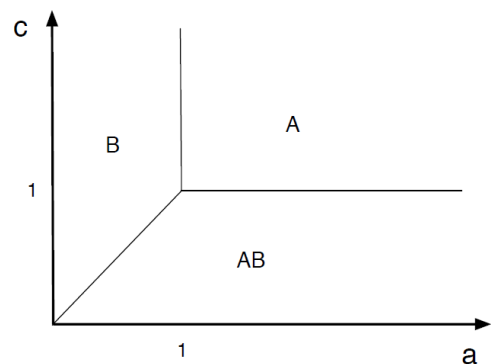
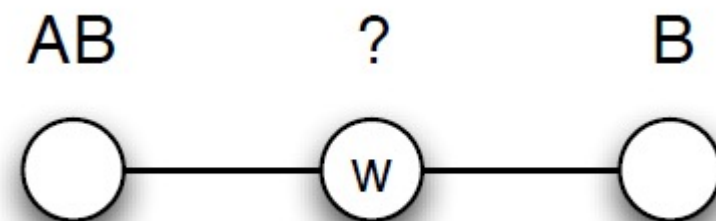


- 若在 B 区，则不会产生级联
- 若在 A 区，则产生级联
- 若在 AB 区，需要继续分析

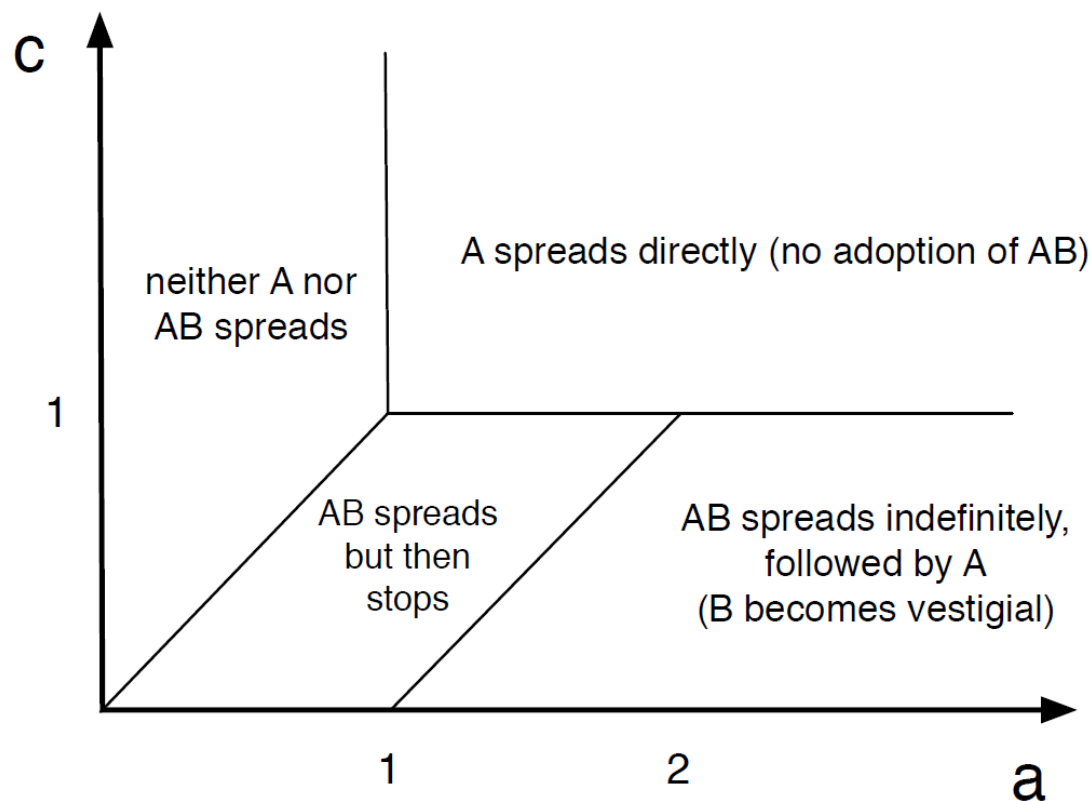
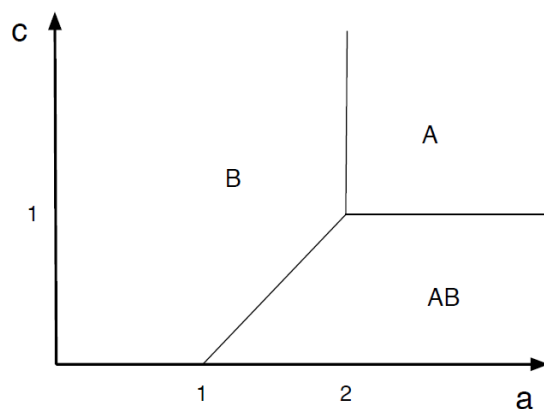


AB区的情况

- 已知：AB在第一步决策中是最优的
- 因此只需要在第一次决策的AB区继续考虑



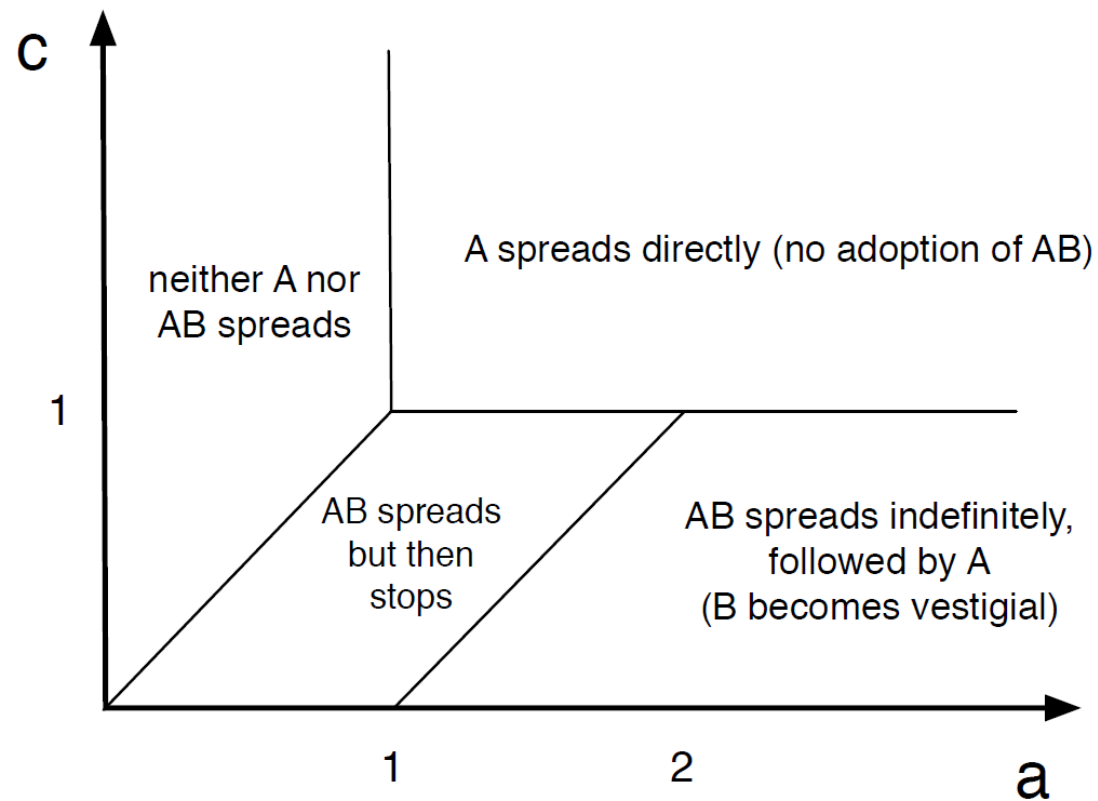
叠合



最终的产生级联的区域

直观解释

- 设A提供更好的技术， $a = 1.5$ ， $b = 1$
- 考虑 c 的取值的影响
- 若 c 很小，则同时维护AB非常容易，则一旦AB广泛传播，人们会过渡到A
- 若 c 很大，则A需要很强，达到直接征服的效果
- 问题出在 c 在中间的情况，会出现有利于B的情况：存在A和B之间的“缓冲地带”，B通过部分与A兼容阻止A传播

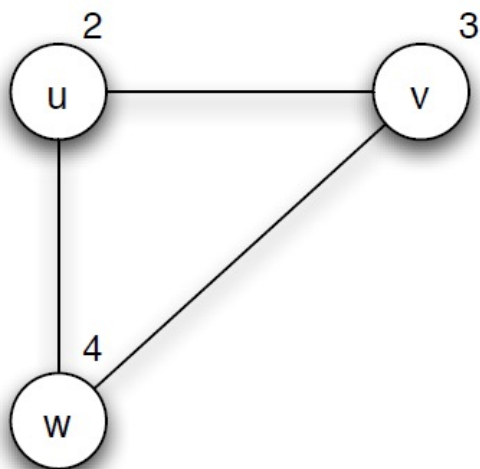


多元无知现象

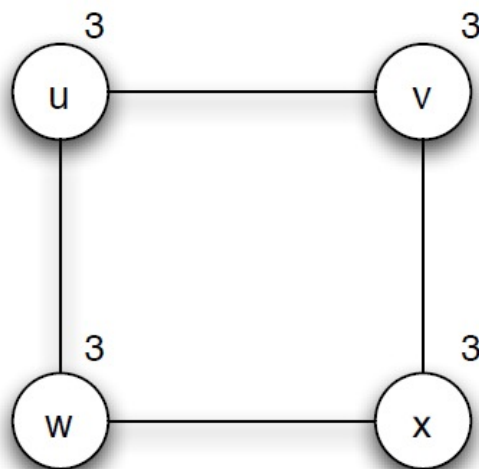
- 考虑一项需要集体参与的活动，**只有召集足够多的人才能产生利益**
- 类似“网络效应”：只有直到足够多人都会买，自己才会买
- 不同点：**没有共识**性的知识，**只能**看到自己的**邻居**的情况
- 结果：尽管很多人都愿意参加，但是最后却几乎无人真正去参加（因为人们觉得别人都不会去参加），或者相反，总之“一叶障目不见泰山”
- 1970年美国的一项调查：只有少数美国白人真正主张种族隔离，但有超过50%的人认为他们所在地区大多数美国白人支持这种主张

集体行动的知识效应 (Chwe 1999,2000)

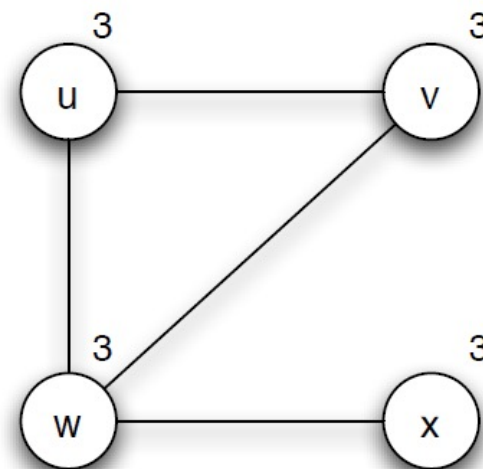
- 每个人 v 有一个阈值 k ，代表如果有至少 k 人参加，则 v 参加
 - 想象如果不到阈值自己贸然参加，可能会带来危害的活动，但又想去
- 只可以得到直接邻居的阈值，但不允许交流（故更远的阈值未知）
- 给定一个网络和对应阈值，推测会发生什么情况



都不参加



?



?

共同知识

- 覆盖广泛的社会体系的一大作用是让人们**实现共同认知**
 - 一个收视率高的公开演讲、流行报纸/网站/赛事的广告，不止传递了信息，还让人们认识到许多其他人也有同样的信息
- 苹果电脑1984年“超级碗”广告（Ridley-Scott执导）
 - “有史以来最伟大的电视广告”
 - 确实解决Mac系统的痛点：小众操作系统与其他系统不完全兼容，需要大量用户采用才能有效推广开；全球用户真正知道了苹果这个品牌
 - <https://www.youtube.com/watch?v=dgZFvCMqITw>

其他广泛采用的传播模型

Linear threshold模型

- 推广的Morris模型（Morris模型是我们刚刚研究的模型）
- 点 v 的每个邻居 w 有一个“影响力”值 $b_{v,w}$ ，满足
 - $\sum_{w \text{ is neighbor of } v} b_{v,w} \leq 1$
- 每个点 v 有一个 $[0, 1]$ 上均匀随机的阈值 θ_v
- 从初始点集 S 开始，每轮每个点采用 A 的条件是：
 - $\sum_{w: \text{neighbor of } v \text{ and use } A} b_{v,w} \geq \theta_v$
- 与Morris相比：Morris中threshold是确定的某个 q ， $b_{v,w}$ 也都相等

Independent cascade模型

- (Goldenberg, Libai, Muller 2001)
- 初始集合S
- 当一个点v采用A时，对每个v的B邻居w，独立以 $p_{v,w}$ 概率使w采用A
 - 一开始v是S的点；每个v仅在第一次采用A时影响邻居
 - 每个点w可能被影响多次

计算问题：Influence Maximization/Submodular Optimization

计算问题

- 对于linear threshold或者independent cascade模型，设 S^* 是最后采用A的人的集合
- 考虑“病毒营销”场景：公司预算有限，起初只可推广到k个人使用A
- 问题：如何设计初始集合S满足 $|S| \leq k$ ，使得 $|S^*|$ 最大？
- 暴力算法： n^k 时间穷举点集的所有k点子集，然后poly(n)时间计算 $|S^*|$

近似算法：总体思路

- Notation $[n] = \{1, \dots, n\}$
- 将整个传播过程看作一个函数 $f: 2^{[n]} \rightarrow \mathbb{R}_+$
 - $2^{[n]}$ 上的元素是 $[n]$ 的子集
 - f 输入初始点集 S ，返回级联过程结果大小 $|S^*|$
 - 这种定义在子集上的函数叫做 set function
- 证明 f 具有某些特殊性质
 - 具体来说，我们将定义 + 证明 f 是 **submodular set function**
- 套用求解最大化特殊性质函数 f 的一般近似算法

Submodular functions

- 称一个set function $f: 2^{[n]} \rightarrow \mathbb{R}_+$ 是submodular的, 若对于任何子集 $X \subseteq Y$ 和 $e \notin Y$, 有

$$f(X + e) - f(X) \geq f(Y + e) - f(Y)$$

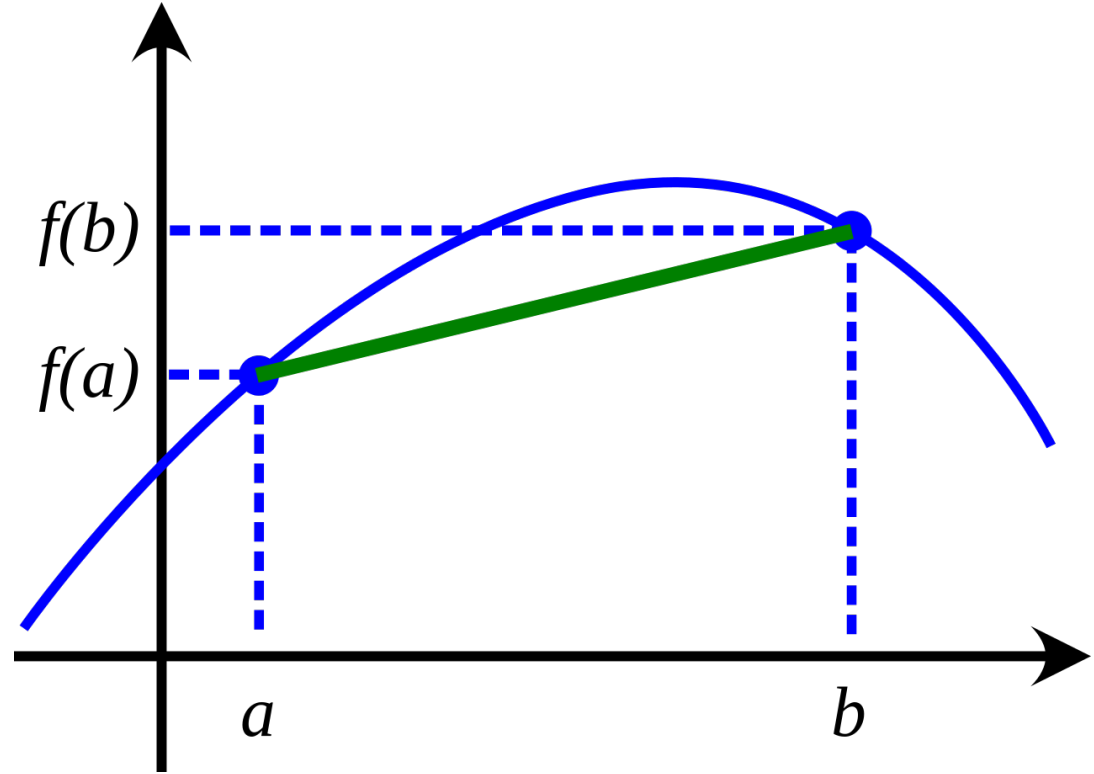
- 此处 “+” 运算指代集合求并 $\cup$
- 定义 $f_e(X) := f(X + e) - f(X)$, 并称之为 “marginal increase”
- 定义等价于 $f_e(X) \geq f_e(Y)$; 如何理解?
- “针对加入新元素 e 的marginal increase是递减的” (注意到 $X \subseteq Y$)

Marginal increase: 现实意义

- 批发商品：买量越大，多买一件需要的价格越低
- 同样的事情见的越多，再次看到类似事情的惊喜/惊讶程度越低
- 同样topic的paper越多，同样“技术含量”的贡献得到的评价越低

Marginal increase: 类比凹函数

- 凹函数 f : $f'(x)$ 递减



- Submodular set function可看作“离散”版本的“凹”函数

Submodular functions

- 举例：线性函数
 - 设有 n 个实数 $w_1, \dots, w_n$, $f(S) = \sum_{i \in S} w_i$
 - 验证:
 - 对于 $S \subseteq T, e \notin T$, $f(S \cup \{e\}) - f(S) = w_e = f(T \cup \{e\}) - f(T)$

- 称 f 是monotone的, 若对任何 $S \subseteq T$, 有 $f(S) \leq f(T)$
- 举例: coverage function
 - 设 $\Omega = \{E_1, \dots, E_n\}$ 每个 E_i 都是某个集合, $f(S) = |\bigcup_{i \in S} E_i|$
 - 验证:
 - 对于 $S \subseteq T, e \notin T$: 设 $A = \bigcup_{i \in S} E_i$, $B = \bigcup_{i \in T} E_i$, $A \subseteq B$
 - $f(S \cup \{e\}) - f(S) = |A \cup E_e| - |A| = |E_e \setminus A|$
 - $f(T \cup \{e\}) - f(T) = |B \cup E_e| - |B| = |E_e \setminus B|$
 - 因此 $f(S \cup \{e\}) - f(S) \geq f(T \cup \{e\}) - f(T)$

Submodular functions

- 称 f 是symmetric的若对任意 S : $f(S) = f(\bar{S})$
- 举例: cut function
 - 对于一个 n 点无向图 $G(V, E)$, 对 $S \subseteq V$, $f(S)$ 代表 $E \cap (S \times (V \setminus S))$
 - 验证 (对于完全图):
 - 对于 $S \subseteq T, e \notin T$; 注: e 是一个点
 - $f(S \cup \{e\}) - f(S) = (|S| + 1)(n - |S| - 1) - |S|(n - |S|) = n - 2|S| - 1$
 - $f(T \cup \{e\}) - f(T) = n - 2|T| - 1$
 - 因此 $f(S \cup \{e\}) - f(S) \geq f(T \cup \{e\}) - f(T)$
 - 带权非完全图的cut function也是submodular的

Submodular functions 基本性质

- f 和 g 是submodular且 $\alpha, \beta \geq 0$ 则 $\alpha f + \beta g$ submodular
- 等价定义: $\forall S, T, f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$
- 性质: 若 f 是submodular的, 则对任意 S , f_S 是submodular的 (why?)

Submodular functions基本性质

- 若 $f(\emptyset) \geq 0$, 则 f 是subadditive的, 即 $f(S \cup T) \leq f(S) + f(T)$
 - 证明: 设 $T = \{e_1, \dots, e_m\}, T_i = \{e_1, \dots, e_i\}, T_0 = \emptyset$
 - $f(S \cup T) = f(S) + \sum_{i=1}^m f(S \cup T_i) - f(S \cup T_{i-1})$
 - $\leq f(S) + \sum_{i=1}^m f(T_i) - f(T_{i-1})$
 - $= f(S) + f(T) - f(\emptyset) \leq f(S) + f(T)$
- 推论: $\forall S \subseteq T, f(T) \leq f(S) + \sum_{e \in T \setminus S} f_S(e)$
 - 证明: $f(T) - f(S) = f_S(T) = f_S(T \setminus S) \leq \sum_{e \in T \setminus S} f_S(e)$
- 另外推论: 如果 f 是monotone的, 那么若 S 不是 T 的子集结论依然成立
 - 证明: $f(T) \leq f(T \cup S)$, 然后利用上述结论 with $T' = T \cup S, S' = S$

Independent cascade is submodular

- Set function f : 输入初始点集 S , 返回级联过程结果大小 $|S^*|$

初始集合 S ; 当一个点 v 采用 A 时, 对每个 v 的 B 邻居 w , 独立以 $p_{v,w}$ 概率使 w 采用 A

- 对于 $S \subseteq T, e \notin T$, 证 $f_e(S) \geq f_e(T)$
- $f_e(S) = f(S + e) - f(S)$ 如何分析?
- 困难点: S^* 是由复杂迭代过程定义的

转化成“静态”性质

初始集合 S ；当一个点 v 采用 A 时，对每个 v 的 B 邻居 w ，独立以 $p_{v,w}$ 概率使 w 采用 A

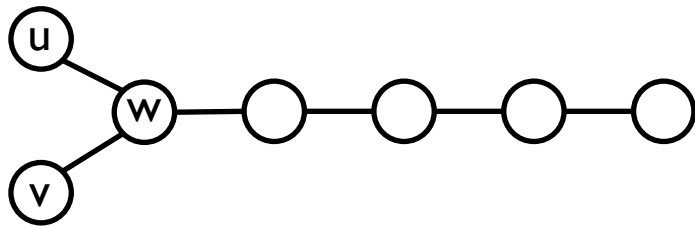
- 注意到：每条（无向）边 (v, w) 都有一个 $p_{v,w}$ ，仅会据此随机决定一次是否“影响”
- 因此可以预先决定每条边的随机outcome，然后分析确定性的过程
 - 设所有边的所有随机outcome为 X
 - 构造一个辅助图 H （取决于 X ）：图 G 中 X 决定选的边组成

事实：对任意 X ，一个节点 w 选用 A ，当且仅当在 H 中有一条自初始集合 S 到 w 的路径

- 设 $R_X(v)$ 代表在 X 下， v 能到达的点集
- 则 $S^* = \bigcup_{v \in S} R_X(v)$ ；令 $f^X(S) = |\bigcup_{v \in S} R_X(v)|$
 - 可将 $f^X(\cdot)$ 看作 $\Omega = \{R_X(v) : v \in V\}$ 的coverage function
 - 因此 f^X 是submodular的；所以 $f = \mathbb{E}_X[f^X]$ 是submodular的

Linear threshold?

- 结论: linear threshold model也是submodular的
- 但是, Morris model未必是submodular的
 - 每个阈值都是 $[0, 1]$ 的随机数很重要



考虑Morris模型: $q = 0.5$

- $S = \emptyset, e = u, T = v$
- $f(S + e) - f(S) = 1$
- $f(T + e) - f(T) = 6$

点 v 的每个邻居 w 有一个“影响力”值 $b_{v,w}$, 满足

$$\sum_{w \text{ is neighbor of } v} b_{v,w} \leq 1$$

每个点 v 有一个 $[0, 1]$ 上均匀随机的阈值 θ_v
从初始点集 S 开始, 每轮每个点采用 A 的条件是:

$$\sum_{w: \text{neighbor of } v \text{ and use } A} b_{v,w} \geq \theta_v$$

Submodular maximization

- 设 $f: 2^{[n]} \rightarrow \mathbb{R}_+$ 是单调submodular的
- $[n]$ 是ground set, 找k点 $S \subseteq [n]$ 使得 $f(S)$ 最大
- 贪心算法:
 1. $S \leftarrow \emptyset$
 2. for $i = 1, 2, \dots, k$
 3. $e \leftarrow \arg \max_{e'} f_S(e')$
 4. $S \leftarrow S + e$
 5. return S

贪心算法近似比 $1 - e^{-1} \approx 0.632$

- Nemhauser, Wolsey, Fisher 78
- 设 S_i 为前 i 个被算法选择的元素，设 C 为最优解
- 对 i 归纳证明： $f(C) - f(S_i) \leq \left(1 - \frac{1}{k}\right)^i f(C)$ 推出 $f(C) - f(S_k) \leq e^{-1} f(C)$
- Base case: $i = 0$ trivial
- 设对 $i - 1$ 成立并考虑证明 i
- 在第 i 步，设算法选择了 e_i ，则 $f_{S_{i-1}}(e_i) \geq f_{S_{i-1}}(e'), \forall e'$
- $f(C) - f(S_{i-1}) \leq \sum_{e \in C \setminus S_{i-1}} f_{S_{i-1}}(e)$ ，推出
- $f_{S_{i-1}}(e_i) \geq \frac{1}{|C \setminus S_{i-1}|} \sum_{e \in C \setminus S_{i-1}} f_{S_{i-1}}(e) \geq \frac{1}{k} (f(C) - f(S_{i-1}))$

- 已经得到: $f_{S_{i-1}}(e_i) \geq \frac{1}{|C \setminus S_{i-1}|} \sum_{e \in C \setminus S_{i-1}} f_{S_{i-1}}(e) \geq \frac{1}{k} (f(C) - f(S_{i-1}))$
- 归纳假设: $f(C) - f(S_{i-1}) \leq \left(1 - \frac{1}{k}\right)^{i-1} f(C)$
- $f(C) - f(S_i) = f(C) - f(S_{i-1} + e_i) = f(C) - f(S_{i-1}) - f_{S_{i-1}}(e_i)$
- $\leq f(C) - f(S_{i-1}) - \frac{1}{k} (f(C) - f(S_{i-1}))$
- $\leq \left(1 - \frac{1}{k}\right) (f(C) - f(S_{i-1}))$
- $\leq \left(1 - \frac{1}{k}\right)^i f(C)$
- Uriel Feige证明 $1 - 1/e$ 是紧的 (assuming $P \neq NP$)

一般（非单调）情况？

- 对于一般submodular我们只考虑无约束的情况
 - 注：若 f 是monotone的且无约束，那么最大值取所有item时取到
- 考虑特例：max cut
 - 回忆：cut function是submodular
- 经典随机算法：每个点以0.5概率分到左边，否则右边
- 这样定义的S-T割的大小的期望是 $1/2 \text{ OPT}$
 - 每条边参与S-T割的概率是 $1/2$
 - 期望总共有 $1/2$ 的边数参与了S-T割
 - 而OPT至多是所有边数

(Feige, Mirrokni, Vondrák, 2009)

- 类比算法：每个element以0.5概率取，得到集合S
- 结论：若 f 非负，则 $\mathbb{E}[f(S)] \geq 0.25 \cdot \text{OPT}$
- 引理：设 $g: 2^{[n]} \rightarrow \mathbb{R}$ 是submodular function，对任意子集 $A \subseteq [n]$ ，设 $A(p)$ 代表每个A中元素以 p 概率独立选取得到的集合，则 $\mathbb{E}[g(A(p))] \geq (1 - p)g(\emptyset) + pg(A)$
- 对A的大小 $|A|$ 进行归纳。Base case $A = \emptyset$ 是直接的

引理证明

- 假设 $A = A' \cup \{x\}, x \notin A'$
- 考虑 $A(p) \cap A'$, 那么这个 (随机) 集合是 A' 上每个元素 p 独立选取
- 可以定义 $A'(p) = A(p) \cap A'$
- $g(A(p)) - g(A'(p)) \geq g(A(p) \cup A') - g(A')$
 - 利用 $f(S) + f(T) \geq f(S \cap T) + f(S \cup T)$
- $\mathbb{E}[g(A(p))] \geq \mathbb{E}[g(A'(p)) + g(A(p) \cup A') - g(A')]$
- $\quad = \mathbb{E}[g(A'(p))] + \mathbb{E}[g(A(p) \cup A') - g(A')]$
- $\quad = \mathbb{E}[g(A'(p))] + p(g(A) - g(A'))$
- 利用 hypothesis: $\mathbb{E}[g(A'(p))] \geq (1 - p)g(\emptyset) + pg(A')$

推出结论

- 主引理：对于任意子集 A, B ， $0 \leq p, q \leq 1$ ，有 $\mathbb{E}[f(A(p) \cup B(q))] \geq$
- $(1 - p)(1 - q)f(\emptyset) + p(1 - q)f(A) + (1 - p)qf(B) + pqf(A \cup B)$
- 引理如何推出算法是1/4近似的：
 - 设 S 是最优解的集合，设 $\bar{S}$ 是 S 的补集
 - 那么算法选取的集合 $R = S\left(\frac{1}{2}\right) \cup \bar{S}\left(\frac{1}{2}\right)$
 - 利用引理，有 $\mathbb{E}[f(R)] \geq \frac{1}{4}f(\emptyset) + \frac{1}{4}f(S) + \frac{1}{4}f(\bar{S}) + \frac{1}{4}f([n])$
 - 根据 f 非负的条件，我们有 $f(R)$ 期望意义下是0.25近似的
 - 若 f 还是symmetric的，则近似比是0.5

主引理的证明

- 主引理：对于任意子集 A, B ， $0 \leq p, q \leq 1$ ，有 $\mathbb{E}[f(A(p) \cup B(q))] \geq$
- $(1 - p)(1 - q)f(\emptyset) + p(1 - q)f(A) + (1 - p)qf(B) + pqf(A \cup B)$

引理：设 $g: 2^{[n]} \rightarrow \mathbb{R}$ 是submodular function，对任意子集 $A \subseteq [n]$ ，设 $A(p)$ 代表每个 $[n]$ 中元素以 p 概率独立选取得到的集合，则 $\mathbb{E}[g(A(p))] \geq (1 - p)g(\emptyset) + pg(A)$

- 设 $R := A(p)$ ，condition on这个 R ，定义 $g(T) := f(R \cup T)$
 - 注意到 g 是submodular
- 用引理，得到 $\mathbb{E}[g(B(q))] \geq (1 - q)f(R) + qf(R \cup B)$

主引理的证明

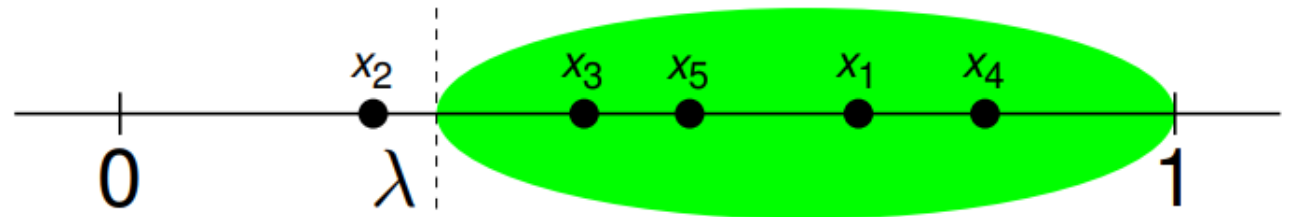
- $\mathbb{E}[g(B(q))] \geq (1 - q)f(R) + qf(R \cup B)$
- 另外, $\mathbb{E}[g(B(q))] = \mathbb{E}[f(A(p) \cup B(q)) \mid A(p) = R]$
- 利用上述不等式, 去掉condition并且代入R
- $\mathbb{E}_R [\mathbb{E}[g(B(q))]] \geq \mathbb{E}[(1 - q)f(A(p)) + qf(A(p) \cup B)]$

引理: 设 $g: 2^{[n]} \rightarrow \mathbb{R}$ 是 submodular function, 对任意子集 $A \subseteq [n]$, 设 $A(p)$ 代表每个 $[n]$ 中元素以 p 概率独立选取得到的集合, 则 $\mathbb{E}[g(A(p))] \geq (1 - p)g(\emptyset) + pg(A)$

- 对 f 使用引理, 得 $\mathbb{E}[f(A(p))] \geq (1 - p)f(\emptyset) + pf(A)$
- 对 $h(S) = f(S \cup B)$ 使用, $\mathbb{E}[f(A(p) \cup B)] \geq (1 - p)f(B) + pf(A \cup B)$

Minimization?

- 考虑一般的submodular set function $f: 2^{[n]} \rightarrow \mathbb{R}$
- 求 $S \subseteq [n]$ 使得 $f(S)$ 最小
- Lovász extension: 定义 $f^L: [0,1]^n \rightarrow \mathbb{R}$
 - $f^L(x) := \mathbb{E}_\lambda[f(T_\lambda(x))]$, where $T_\lambda(x) := \{i : x_i > \lambda\}$, $\lambda \in [0,1]$ 是均匀分布变量
- 性质: $f^L(x) = f(x)$ 对任意 $x \in \{0,1\}^n$ (注: $f(x)$ 中 x 理解成子集)
- 性质: f is submodular 等价于 f^L is convex
- 因此 f^L 可以用例如椭圆法求最小值
- 设最小值 x , 则
- $f^L(x) = \sum_{i=1}^n \alpha_i f(T_i)$
- 其中必然有 T_i 是 f 的最优解



submodular optimization的计算复杂性

- 最小化一般submodular function: $\text{poly}(n)$ 时间
 - (Grötschel-Lovász-Schrijver, 1981; Iwata-Fleischer-Fujishige / Schrijver 2000)
 - 利用“Lovász extension”来转化成连续凸优化问题，最后证明极值点是“整”集合
 - 类比min-cut来理解
- 最小化symmetric submodular, subject to downward-closed constraint
 - $\text{poly}(n)$; 但是NP hard如果是covering constraint, 比如说找至少 k 元子集
- 最大化一般submodular function: NP hard
- 最大化monotone submodular function subject to cardinality constraint: NP hard
- 该领域活跃学者: Jan Vondrák