影响力最大化2：基于最优路径

覃雄派

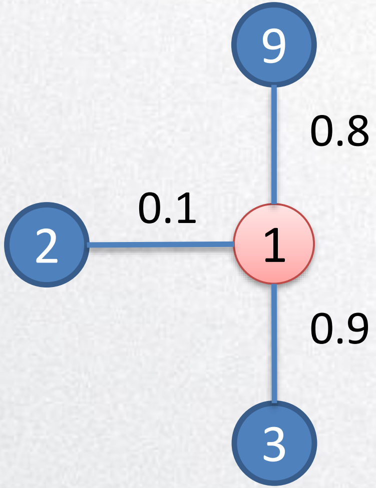
影响力最大化2：基于最优路径

提纲

- 从直接影响到间接影响
- 最短路径算法
- 基于最优路径的影响力最大化
- 实例分析

影响力最大化2：基于最优路径

- Degree Discount算法的局限性
- 基本想法
  - 假设：节点只对其 直接邻居 产生影响
  - 选择的种子节点的影响范围避免重叠
  - 节点还通过邻居对其他节点产生影响

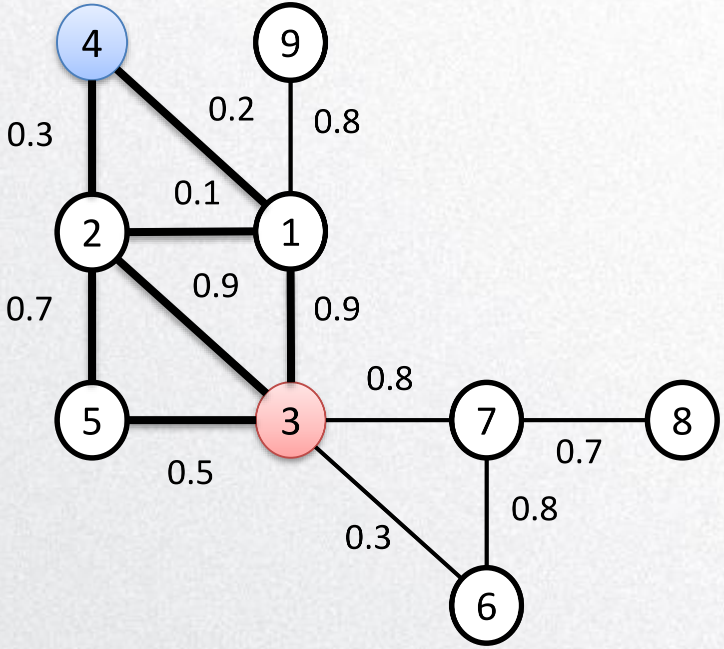


如何考虑间接影响?

选择1，$\Delta\sigma(S) = 1 + .8 + .9 + .1 = 2.8$

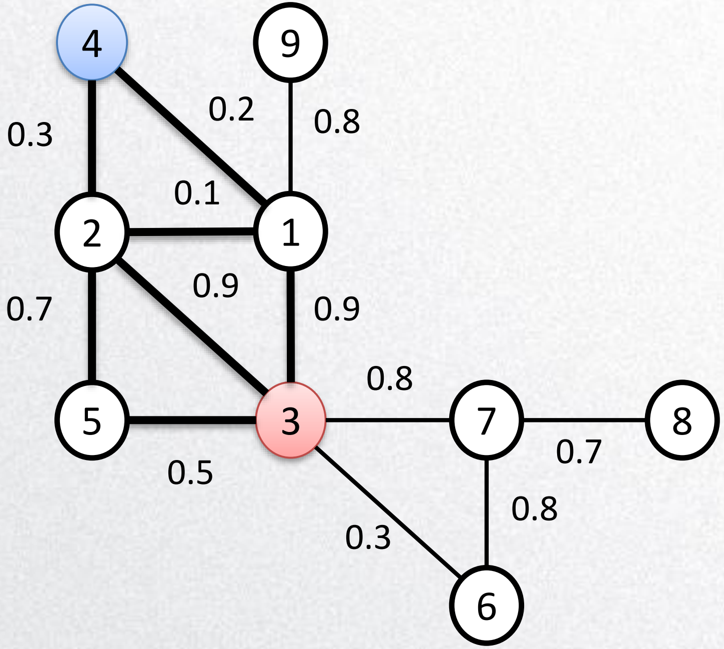
影响力最大化2：基于最优路径

- 节点的间接影响力
- 引入两点间路径的概念
  - 节点3激活节点4的路径可能有多条
    - 3 → 2 → 4
    - 3 → 1 → 4
    - 3 → 2 → 1 → 4
    - 3 → 1 → 2 → 4
    - 3 → 5 → 2 → 4
    - ……



影响力最大化2：基于最优路径

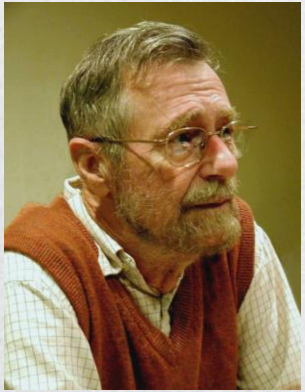
- 节点的间接影响力
- 引入两点间路径的概念
  - 节点3激活节点4的路径可能有多条
  - **只考虑激活概率最大的路径**
  - 其中激活概率为路径中边上概率的乘积
    - 3 → 2 → 4　　$\boxed{0.9 * 0.3 = 0.27}$
    - 3 → 1 → 4　　0.9*0.2 = 0.18
    - 3 → 2 → 1 → 4　　0.9*0.1*0.2=0.018
    - 3 → 1 → 2 → 4　　0.9*0.1*0.3=0.027
    - 3 → 5 → 2 → 4　　0.5*0.7*0.3=0.105
    - ……



影响力最大化2：基于最优路径

影响力最大化2：基于最优路径

- 最短路径算法： Dijkstra algorithm
  - 迪杰斯特拉算法是由荷兰计算机科学家狄克斯特拉于1959 年提出的，因此又叫狄克斯特拉算法
  - 是从一个顶点到其余各顶点的最短路径算法，解决的是有权图中最短路径问题：单源最短路径算法
    - 迪杰斯特拉算法的主要特点是以起始点为中心向外层层扩展，直到扩展到终点为止

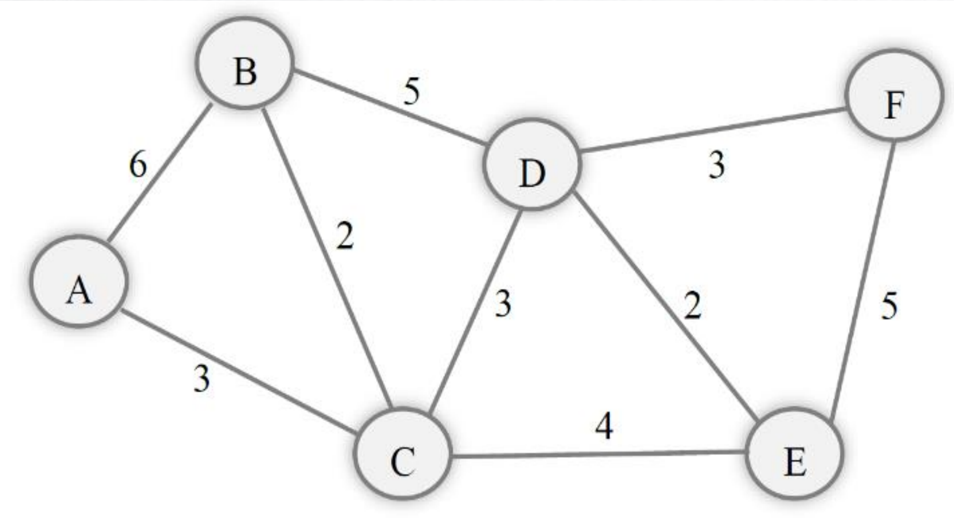


影响力最大化2：基于最优路径

- 算法的具体思路为
  - 设G=(V,E)是一个带权无向图，V为顶点集合，E为边的集合
  - 首先，把图中顶点集合V分成两组，第一组为已求出最短路径的顶点集合，用S表示；初始时S中只有一个源点，以后每求得一条最短路径，就将它加入到集合S中，直到全部顶点都加入到S中，算法就结束了
  - 第二组为其余未确定最短路径的顶点集合，用U表示，按最短路径长度的递增次序依次把第二组的顶点加入S中
  - 在加入的过程中，有一个约束条件，总保持从源点v到S中各顶点的最短路径长度，不大于从源点v到U中任何顶点的最短路径长度
    - 此外，每个顶点对应一个距离
    - S中的顶点的距离就是从v到此顶点的最短路径长度
    - U中的顶点的距离，是从v到此顶点只包括S中的顶点为中间顶点的当前最短路径长度

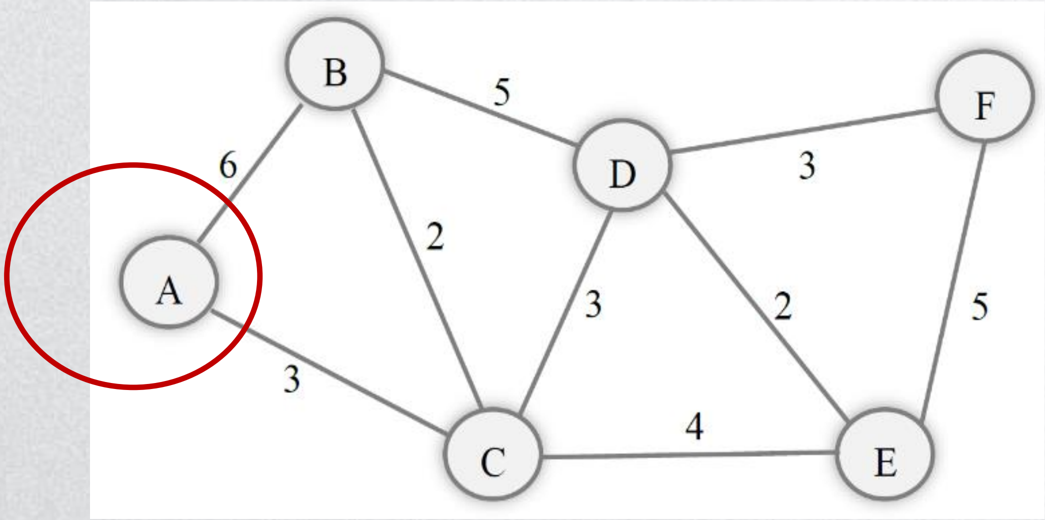
影响力最大化2：基于最优路径

- 一个实例
  - 下图为一个无向图
  - 图中每个边上标注的数字，表示边的长度



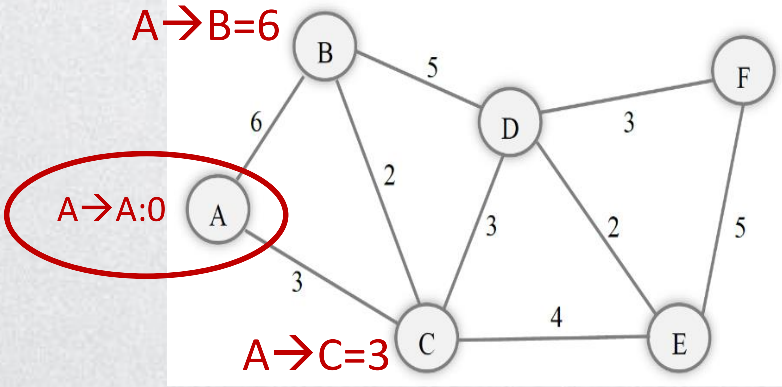
影响力最大化2：基于最优路径

- 一个实例
  - 下图为一个无向图，图中每个边上标注的数字，表示边的长度
  - 现在要求出从A到其他各个顶点的最短路径



影响力最大化2：基于最优路径

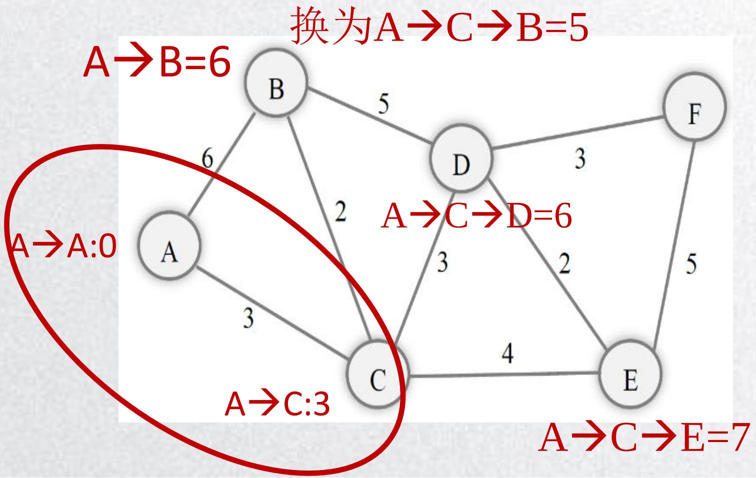
- 一个实例



步骤	S集合	U集合
1	选择A, S= <A> 最短路径A→A=0 以A为中间点，从节点A开始查找	U=<B,C,D,E,F> A→B=6 **A→C=3** A→其他U中的顶点=∞ 发现A→C=3，为最短

影响力最大化2：基于最优路径

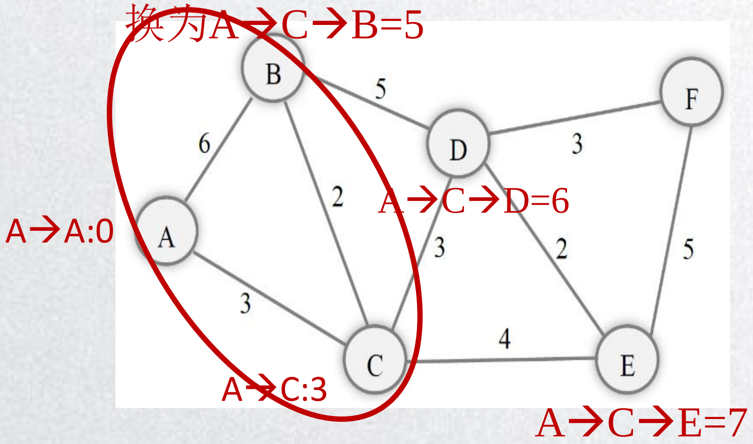
- 一个实例



步骤	S集合	U集合
2	选择C，这时S=<A,C> 最短路径A→A=0, A→C=3 以C为中间点，从A→C=3这条最短路径开始查找	U=<B,D,E,F> A→C→B=5(比第一步的A→B=6短，这时候，把B的权值改为A→C→B=5) A→C→D=6 A→C→E=7 A→C→其他U中的顶点=∞ **发现A→C→B=5，为最短**

影响力最大化2：基于最优路径

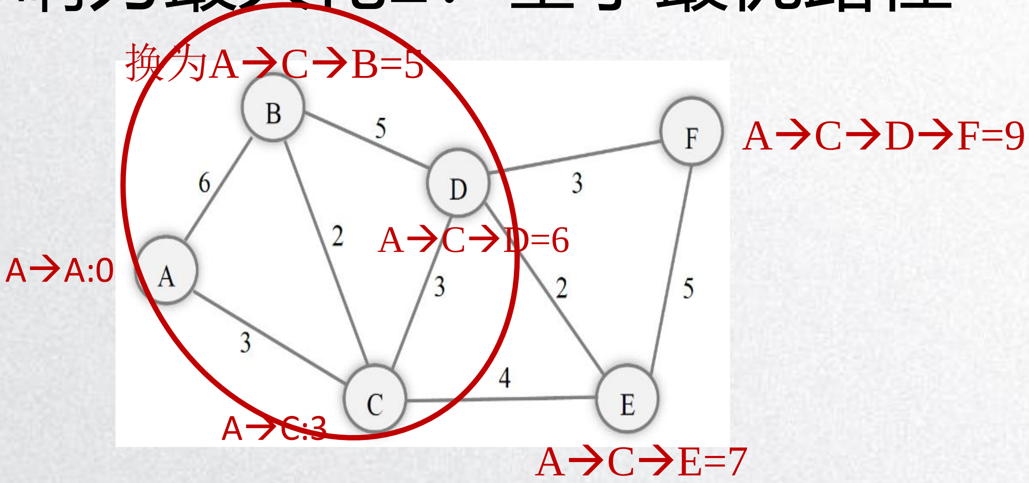
- 一个实例



步骤	S集合	U集合
3	选择B，这时S=<A,C,B> 最短路径A→A=0, A→C=3, A→C→B=5 以B为中间点，从A→C→B这条最短路径开始查找	U=<D,E,F> A→C→B→D=10(比第二步A→C→D=6要长，这时到D的权值，维持为A→C→D=6) A→C→B→其他U中的顶点=∞ **发现A→C→D=6，为最短**

影响力最大化2：基于最优路径

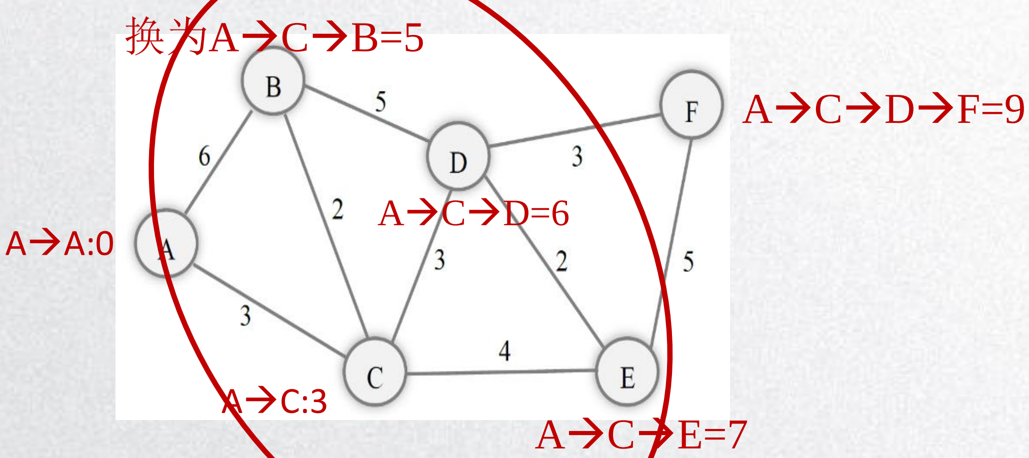
- 一个实例



步骤	S集合	U集合
4	选择D，这时S=<A,C,B,D> 最短路径A→A=0，A→C=3，A→C→B=5，A→C→D=6 以D为中间点，从A→C→D这条路径开始查找	U=<E,F> A→C→D→E=8(比第二步A→C→E=7要长，这时到E的权值，维持为A→C→E=7) A→C→D→F=9 **发现A→C→E=7为最短**

影响力最大化2：基于最优路径

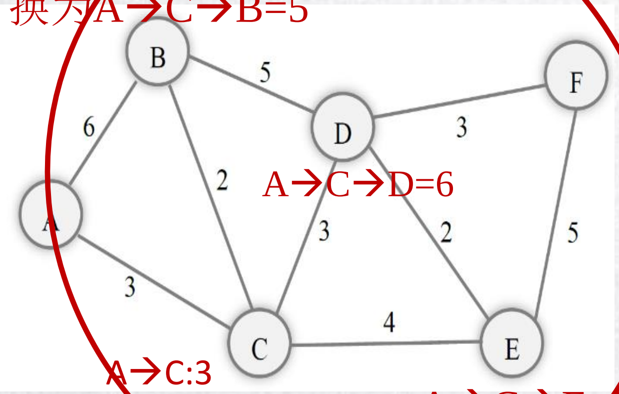
- 一个实例



步骤	S集合	U集合
5	选择E，这时S=<A,C,B,D,E> 最短路径A→A=0，A→C=3，A→C→B=5，A→C→D=6, A→C→E=7 以E为中间点，从A→C→E=7这条最短路径开始查找	U=<F> A→C→E→F=12(比以上第四步的A→C→D→F=9要长，这时F的权值维持为A→C→D→F=9) **发现A→C→D→F=9权值最短**

影响力最大化2：基于最优路径

- 一个实例

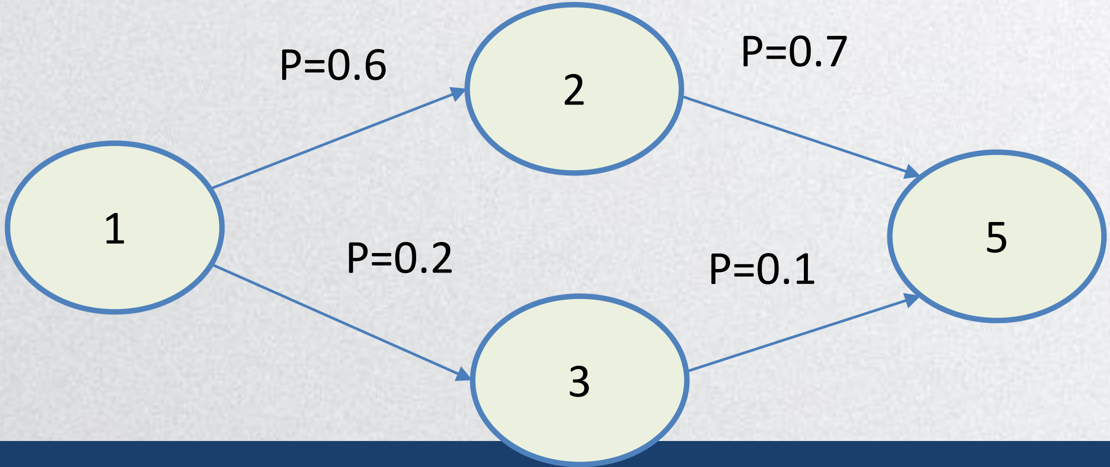


步骤	S集合	U集合
6	选择F，这时S=<A,C,B,D,E,F> 最短路径为 A→A=0，A→C=3, A→C→B=5，A→C→D=6，A→C→E=7, A→C→D→F=9	U集合已空，查找完毕

影响力最大化2：基于最优路径

影响力最大化2：基于最优路径

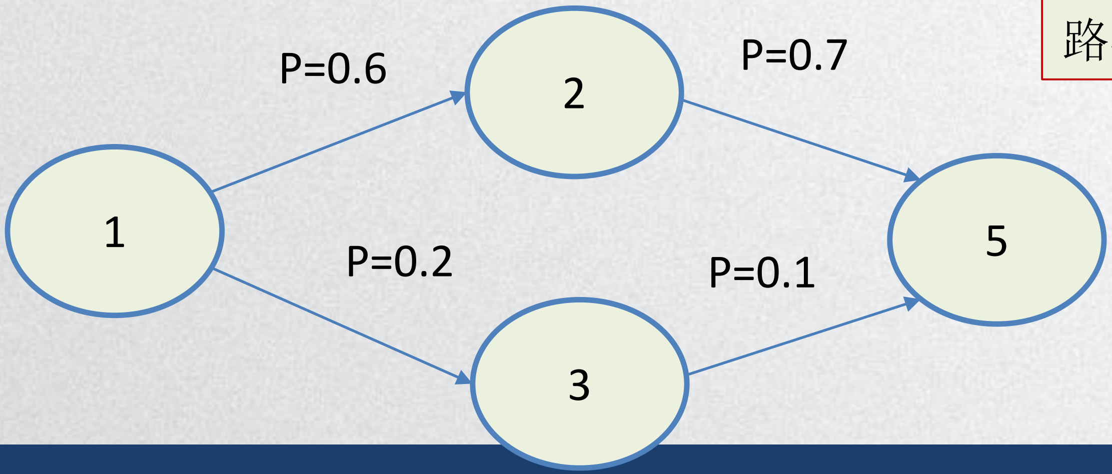
- 基于最优路径的影响力最大化
  - 最优传播路径
  - 从1到5有两条路径
    - 沿着1→2→5的路径传播，那么1影响5的概率为0.6 ×0.7
    - 沿着1→3→5的路径传播，那么1影响5的概率为0.2×0.1



影响力最大化2：基于最优路径

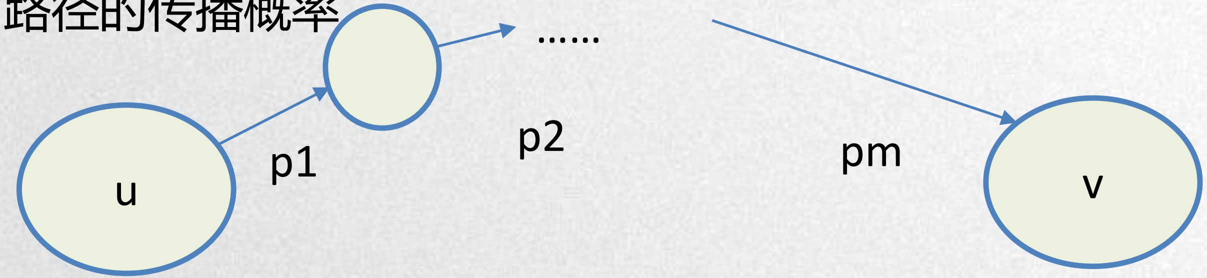
- 基于最优路径的影响力最大化
  - 最优传播路径
  - 从1到5有两条路径
    - 沿着1→2→5的路径传播，那么1影响5的概率为0.6 ×0.7
    - 沿着1→3→5的路径传播，那么1影响5的概率为0.2×0.1

1→2→5为最优传播路径，即概率最大



影响力最大化2：基于最优路径

- 最优传播路径的定义
  - 路径的传播概率



For a path $P = \langle u = p_1, p_2, \ldots, p_m = v \rangle$, we define the *propagation probability* of the path, $pp(P)$, as

$$pp(P) = \Pi_{i=1}^{m-1} pp(p_i, p_{i+1}).$$

P路径;
pp路径的传播概率

影响力最大化2：基于最优路径

- 最优传播路径的定义(最大影响路径)

Definition 1 (Maximum Influence Path) *For a graph* G*, we define the* maximum influence path $MIP_G(u,v)$ *from* u *to* v *in* G *as*

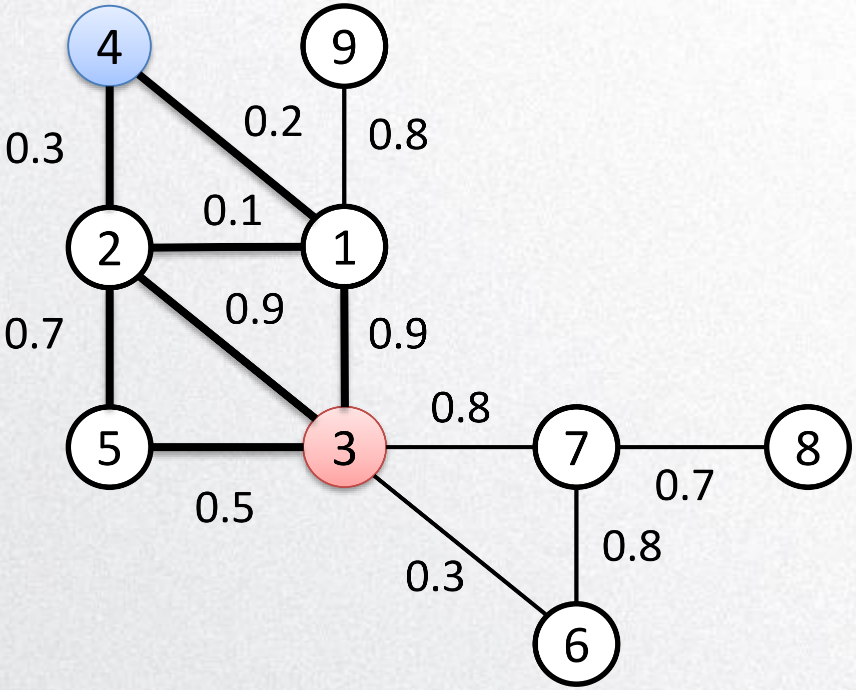
$$MIP_G(u,v) = \arg\max_P \{pp(P) \mid P \in \mathcal{P}(G,u,v)\}.$$

Propagation probability最大

u到v的所有路径

影响力最大化2：基于最优路径

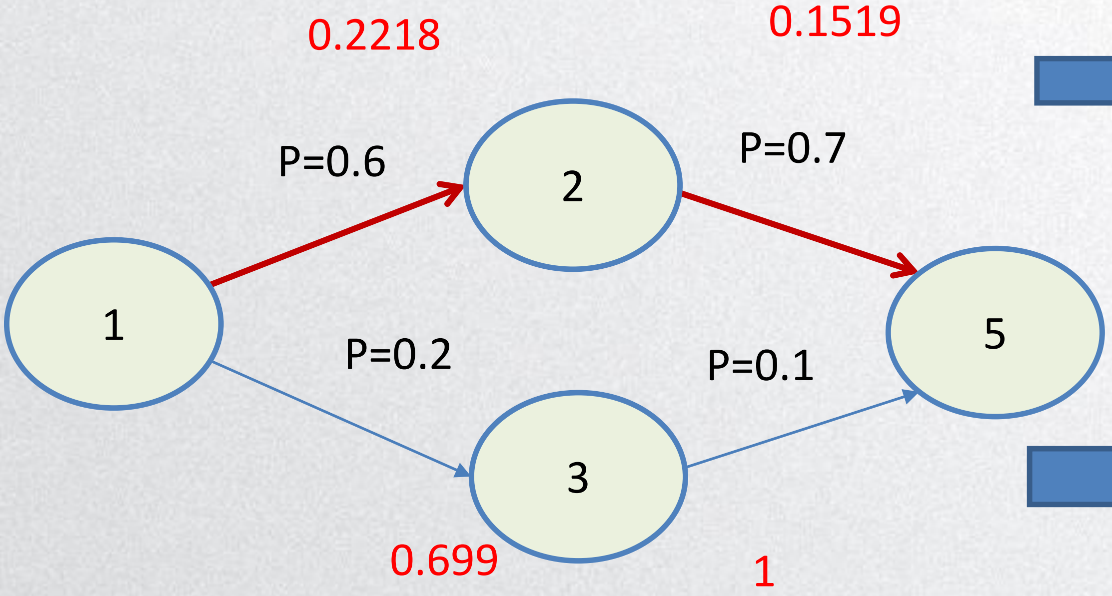
- 如何使用Dijkstra算法
  - 最大化∏
    - p_1 p_2… p_m
  - 等于于最大化∑
    - $logp_1$ $logp_2$… $logp_m$
  - 等价于最小化∑
    - $-logp_1$ $-logp_2$… $-logp_m$



影响力最大化2：基于最优路径

- 如何使用Dijkstra算法
  - 计算-log (p)

该路径的指标最小



累加得到0.3737

最优路径结果和前文的结果一致

累加得到1.699

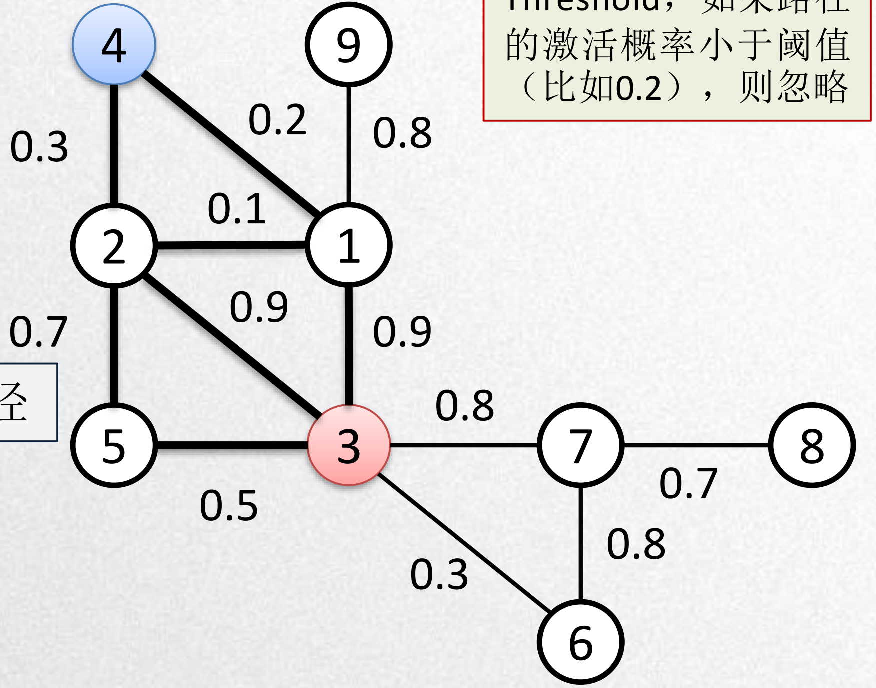
影响力最大化2：基于最优路径

影响力最大化2：基于最优路径

- 基于最优路径的影响力最大化
  - 实例：激活概率最大的路径

$$pp(s \leadsto t) = \prod_{e \in s \leadsto t} pp(e)$$

练习：请计算节点3到其它结点的最优路径



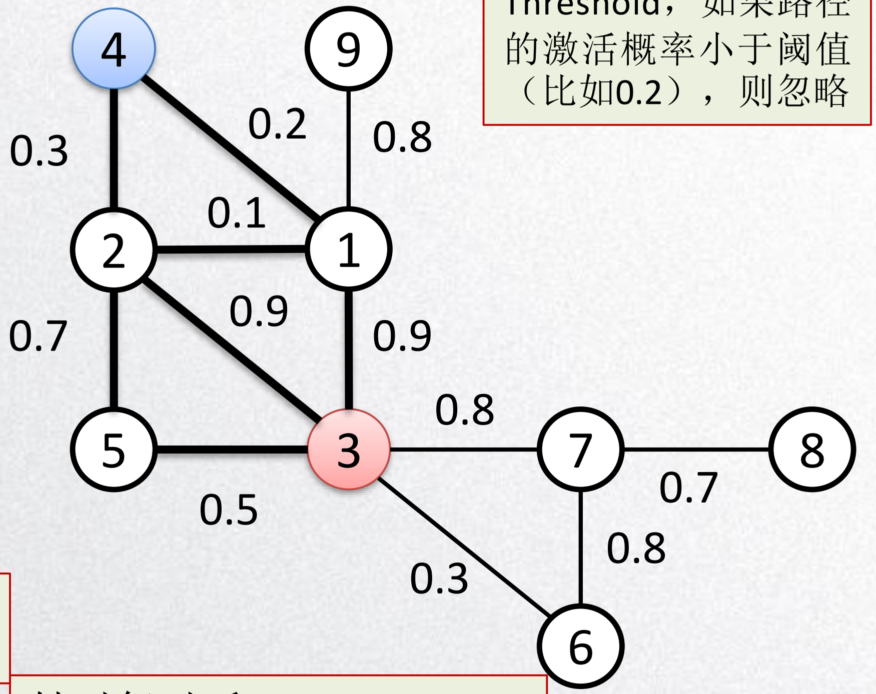
可以设定一个阈值Threshold，如果路径的激活概率小于阈值（比如0.2），则忽略

同学们练习，然后核对答案

影响力最大化2：基于最优路径

- 基于最优路径的影响力最大化
  - 实例：激活概率最大的路径

$$pp(s \rightsquigarrow t) = \prod_{e \in s \rightsquigarrow t} pp(e)$$



可以设定一个阈值Threshold，如果路径的激活概率小于阈值（比如0.2），则忽略

练习：请计算节点3到其它结点的最优路径

3到1: 3 →1　　0.9
3到2: 3 →2　　0.9
3到3: 3 →3　　1.0
3到4: 3 →2 →4　　0.9*0.3
3到5: 3 →2 →5　　0.9*0.7
3到6: 3 →7 →6　　0.8*0.8
3到7: 3 →7　　0.8
3到8: 3→7→8　　0.8*0.7
3到9: 3 →1 →9　　0.9*.08

累加得到
6.42

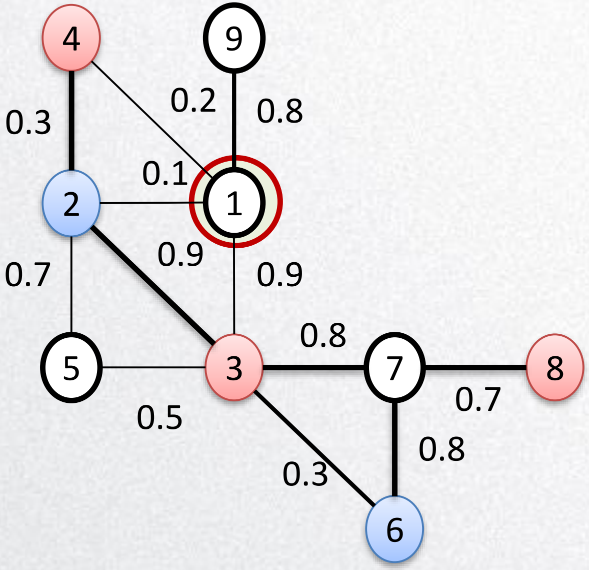
针对每对s和t：
从s到t的所有路径中，寻找激活概率最大的路径

影响力最大化2：基于最优路径

影响力最大化2：基于最优路径

- 假设种子集合S，$\sigma(S)$为图中所有节点激活概率之和
  - $\sigma(S)=\sum_{v\in V} pp(S\sim v)$

> - 练习
> S={1}，请计算对应的$\sigma(S)$



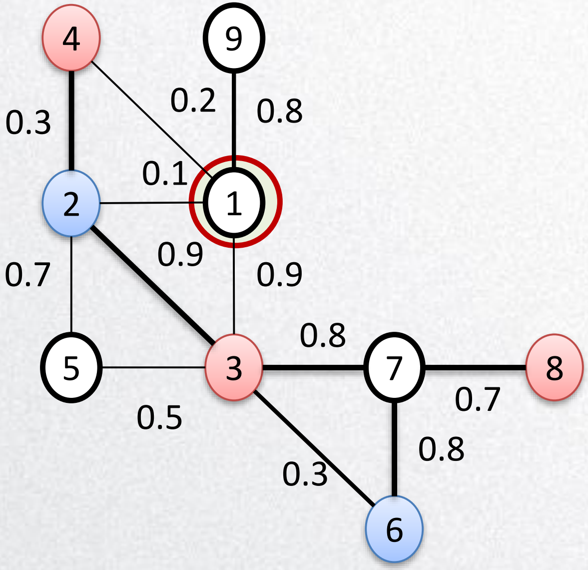
影响力最大化2：基于最优路径

- 假设种子集合S，$\sigma(S)$为图中所有节点激活概率之和
  - $\sigma(S)=\sum_{v\in V} pp(S \rightsquigarrow v)$

- 练习
  S={1}，请计算对应的$\sigma(S)$

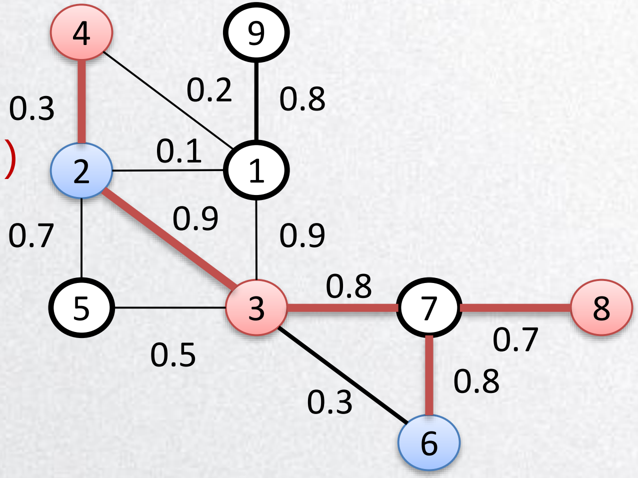
$pp(S \rightsquigarrow 1)$=1	1→1
$pp(S \rightsquigarrow 2)$=0.81	1→3→2
$pp(S \rightsquigarrow 3)$=0.9	1→3
$pp(S \rightsquigarrow 4)$=0.243	1→3→2→4
$pp(S \rightsquigarrow 5)$=0.567	1→3→2→5
$pp(S \rightsquigarrow 6)$=0.576	1→3→7→6
$pp(S \rightsquigarrow 7)$=0.72	1→3→7
$pp(S \rightsquigarrow 8)$=0.504	1→3→7→8
$pp(S \rightsquigarrow 9)$=0.8	1→9

累加得到
$\sigma(\{1\}) = \mathbf{6.12}$



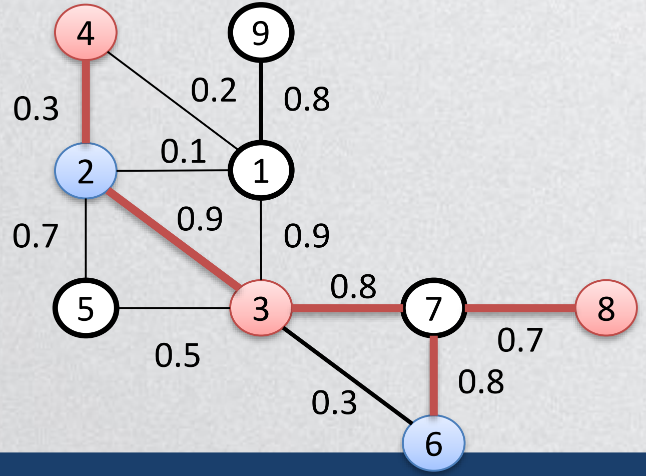
影响力最大化2：基于最优路径

- 前文的S只有一个节点
  - 计算其他节点被激活的概率比较简单
- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 假设种子集合$S = \{3,4,8\}$（红色）
  - 计算节点2和节点6被激活的概率（蓝色）



影响力最大化2：基于最优路径

- 前文的S只有一个节点
  - 计算其他节点被激活的概率比较简单
- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 假设种子集合$S = \{3,4,8\}$
  - 计算节点2和节点6激活的概率



先看看节点6

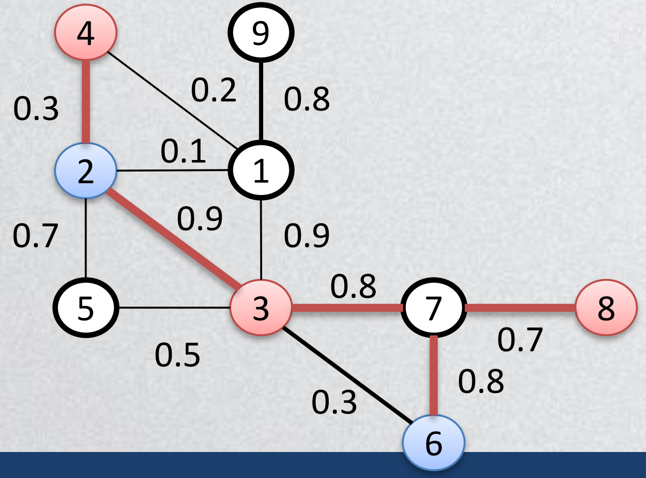
3到6的最优传播路径 3→7→6 0.8*.08=0.64
8到6的最优传播路径 8→7→6 0.7*0.8=0.56
4 到 6 的 最 优 传 播 路 径 4→2→3→7→6 0.3*0.9*0.8*0.8=0.1728

~~考虑阈值为0.2那么4→2→3→7→6的传播路径不予考虑~~

实际上4→2→3→7→6经过3→7→6，所以被覆盖了，请参考后文的继续推导

影响力最大化2：基于最优路径

- 前文的S只有一个节点
  - 计算其他节点被激活的概率比较简单
- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 假设种子集合$S = \{3,4,8\}$
  - 计算节点2和节点6激活的概率



先看看节点6

3到6的最优传播路径　3→7→6　0.8*.08=0.64
8到6的最优传播路径　8→7→6　0.7*0.8=0.56

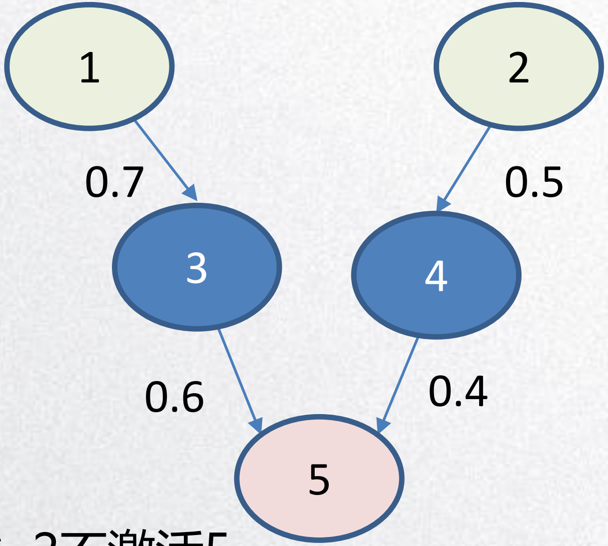
3→7→6和8→7→6两条路径有重叠

6被激活的概率为：7被激活的概率 * 0.8
7被激活的概率为：1 - 3不激活7的概率 * 8不激活7的概率
得到$\big(1-(1-0.8)\cdot(1-0.7)\big)\cdot 0.8 = \boxed{0.752}$

{3,4,8}影响节点6的概率

影响力最大化2：基于最优路径

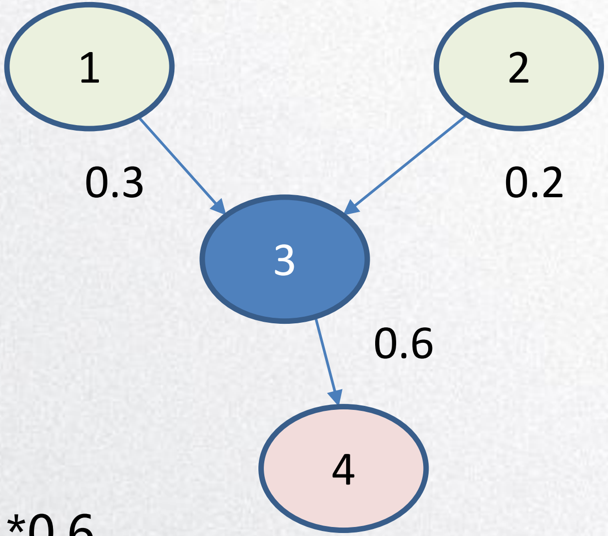
- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 一般化讨论
  - 种子节点为{1,2}
  - 目标节点为5
  - 1到5的最优路径为1→3→5
  - 2到5的最优路径为2→4→5
  - 路径没有重叠



  - 5被激活的概率为：1 - 1不激活5 * 2不激活5
  - =1 – (1-0.7*0.6)(1-0.5*0.4)
  - =1-0.58×0.8 = 0.536

影响力最大化2：基于最优路径

- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 一般化讨论
  - 种子节点为{1,2}
  - 目标节点为4
  - 1到4的最优路径为1→3→4
  - 2到4的最优路径为2→3→4
    - 路径有重叠
  - 4被激活的概率为：3被激活的概率 *0.6
  - 3被激活的概率为：1- 1不激活3 * 2不激活3

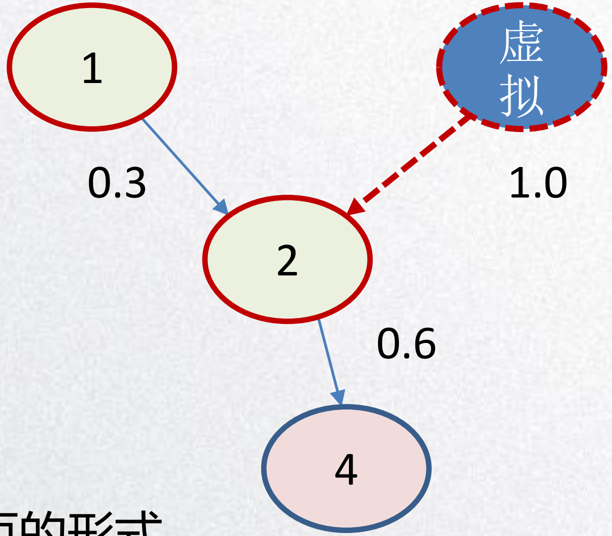


（1-(1-0.3)*(1-0.2)） * 0.6

=（1-0.7×0.8） ×0.6 = 0.44*0.6=0.264

影响力最大化2：基于最优路径

- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 一般化讨论
  - 种子节点为{1,2}
  - 目标节点为4
  - 1到4的最优路径为1→2→4
  - 2到4的最优路径为2→4
    - 路径有重叠
  - 构造虚拟节点，把问题转化为上一页的形式
    - 4被激活的概率为：2被激活的概率 *0.6
    - 2被激活的概率为：1- 1不激活2 * 虚拟不激活2

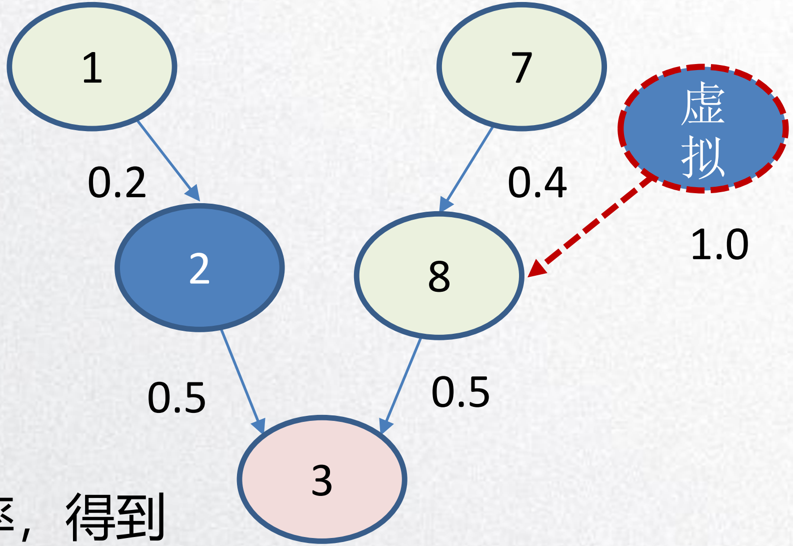


（1-(1-0.3)*(1-1)） * 0.6
=（1-0.7×0）×0.6 = 1*0.6=0.6

影响力最大化2：基于最优路径

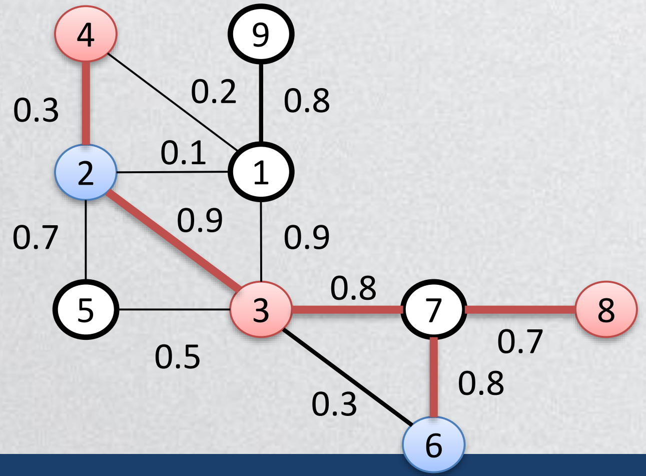
- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 一般化讨论
  - 种子节点为{1，7，8}
  - 目标节点为3
  - 1到3的最优路径为1→2→3
  - 7到3的最优路径为7→8→3
  - 8到3的最优路径为8→3
    - 第二条和第三条路径有重叠
- 构造虚拟节点，计算7和8影响3的概率，得到
  - (1-(1-0.4)(1-1))*0.5 = 0.5
  - 在考虑123路径，节点3被影响的概率
  - 1-(1-0.1)(1-0.5) = 0.55

0.2*0.5



影响力最大化2：基于最优路径

- 前文的S只有一个节点
  - 计算其他节点被激活的概率比较简单
- 如果节点被多条路径激活
  - 如何计算节点被激活的概率
- 假设种子集合$S = \{3,4,8\}$
  - 计算节点2和节点6激活的概率



现在看看2
3到2的最优传播路径 3→2 0.9
4到2的最优传播路径 4→2 0.3
8 到 2 的 最 优 传 播 路 径 8→7 →3 →2
0.7*0.8*0.9=0.504

8→7 →3 →2 和3→2两条路径有重叠
根据上一个页面计算方法，有8和3影响2的概率为0.9
~~这时候，感觉8没有影响~~

2被激活的概率为：1 - 4不激活它的概率 * 3不激活它的概率
得到$\left(1-(1-0.3)\cdot(1-0.9)\right) = 0.93$

{3,4,8}影响节点2的概率

影响力最大化2：基于最优路径

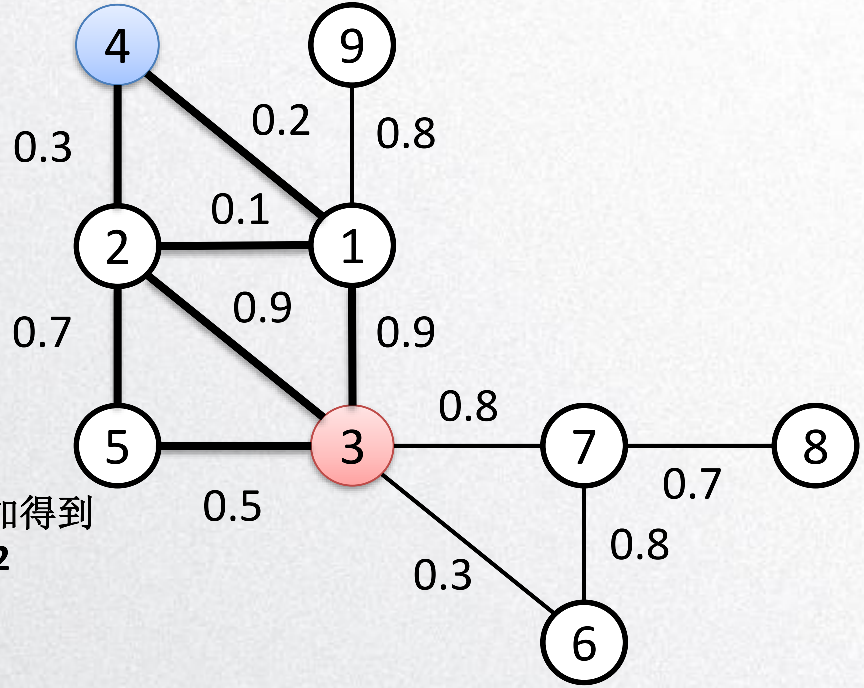
影响力最大化2：基于最优路径

- 回顾：先选了3
  - 作为种子

.请计算节点3到其它结点的最优路径

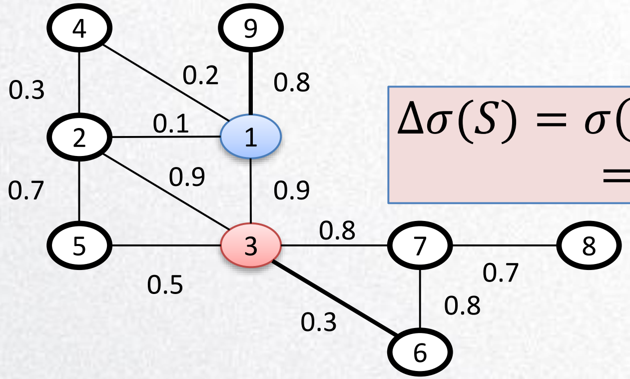
{3}到1: 3 →1　　0.9
{3}到2: 3 →2　　0.9
{3}到3: 3 →3　　1.0
{3}到4: 3 →2 →4　　0.9*0.3
{3}到5: 3 →2 →5　　0.9*0.7
{3}到6: 3 →7 →6　　0.8*0.8
{3}到7: 3 →7　　0.8
{3}到8: 3→7→8　　0.8*0.7
{3}到9: 3 →1 →9　　0.9*.08

累加得到
6.42



影响力最大化2：基于最优路径

- 在选到3的基础上
  - 尝试增加选择1
  - 看看影响范围的变化，需要计算
    - {3,1}的影响概率- {3}的影响概率
    - 即$\sigma(\{1,3\}) - \sigma(\{3\})$



$\Delta\sigma(S) = \sigma(\{3,1\}) - \sigma(\{$
$= 0.18$

.请出节点{3,1}到其它结点的最优路径

{3,1}到1: 3 →1/1→1 1.0
{3,1}到2: 3 →2 /1→3→2 0.9
{3,1}到3: 3 →3/1→3 1.0
{3,1}到4: 3 →2 →4 /1→3→2→4 0.9*0.3
{3,1}到5: 3 →2 →5/1→3→2→5 0.9*0.7
{3,1}到6: 3 →7 →6/1→3→7→6 0.8*0.8
{3,1}到7: 3 →7/1→3→7 0.8
{3,1}到8: 3→7→8/1→3→7→8 0.8*0.7
{3,1}到9: 3 →1 →9/1→9 0.8

.请出{3}到其它结点的最优路径

{3}到1: 3 →1 0.9
{3}到2: 3 →2 0.9
{3}到3: 3 →3 1.0
{3}到4: 3 →2 →4 0.9*0.3
{3}到5: 3 →2 →5 0.9*0.7
{3}到6: 3 →7 →6 0.8*0.8
{3}到7: 3 →7 0.8
{3}到8: 3→7→8 0.8*0.7
{3}到9: 3 →1 →9 0.9*.08

$\Delta pp(S \leadsto 1)$=0.1
$\Delta pp(S \leadsto 2)$=0
$\Delta pp(S \leadsto 3)$=0
$\Delta pp(S \leadsto 4)$=0
$\Delta pp(S \leadsto 5)$=0
$\Delta pp(S \leadsto 6)$=0
$\Delta pp(S \leadsto 7)$=0
$\Delta pp(S \leadsto 8)$=0
$\Delta pp(S \leadsto 9)$=0.08

影响力最大化2：基于最优路径

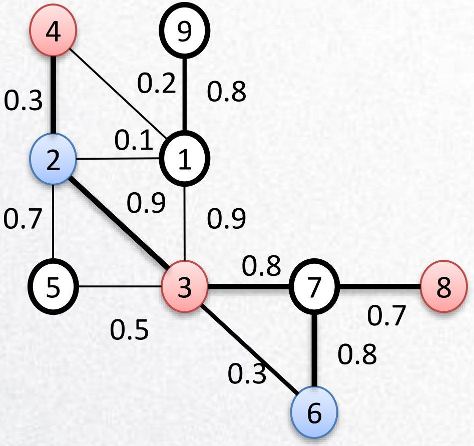
- 在选到3的基础上
  - 尝试增加选择其他节点u
  - 看看影响范围的变化，需要计算
    - {3,u}的影响概率 - {3}的影响概率
    - 即$\sigma(\{u,3\})-\sigma(\{3\})$
  - 挑选最大的u进入{3}这个集合作为新的种子集合

尝试一批u

影响力最大化2：基于最优路径

影响力最大化2：基于最优路径

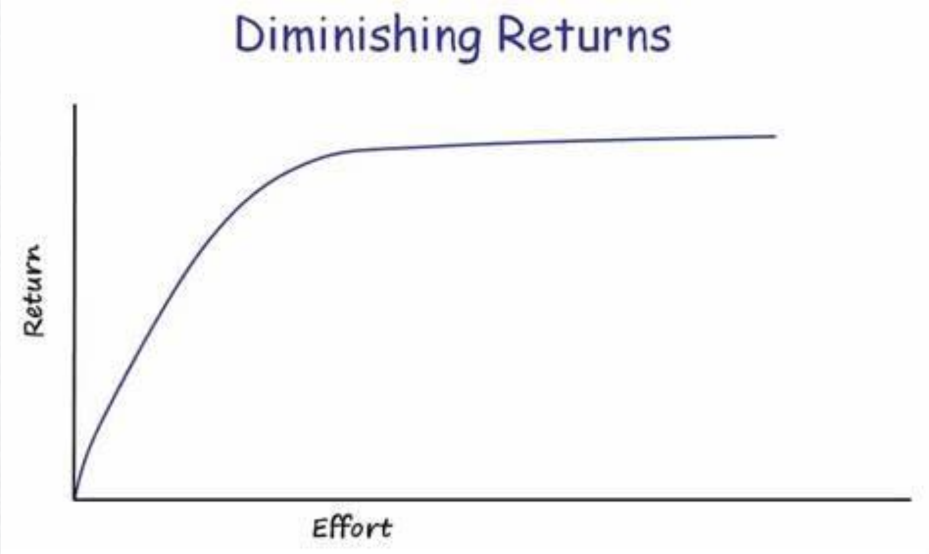
- 假设种子集合S，$\sigma(S)$为图中所有节点激活概率之和
- $\sigma(S) = \sum_{v \in V} pp(S \rightsquigarrow v)$



- 请证明$\sigma(S)$的以下性质
  - 单调性（monotonicity），即$\forall S_1, S_2, S_1 \subseteq S_2$,
    - 有$\sigma(S_1) \leq \sigma(S_2)$
  - 次模性（submodularity），即$\forall S_1, S_2, S_1 \subseteq S_2, \forall u \in V$,
    - 有$\sigma(S_1 \cup \{u\}) - \sigma(S_1) \geq \sigma(S_2 \cup \{u\}) - \sigma(S_2)$

影响力最大化2：基于最优路径

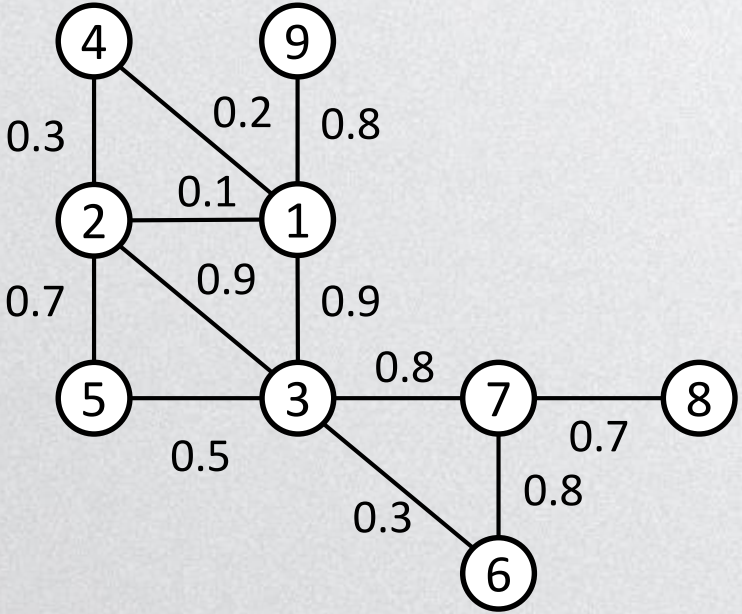
- 什么是次模性（Sub modularity）
  - 次模函数（Submodular Function）
    - 边际收益递减：diminishing returns



影响力最大化2：基于最优路径

影响力最大化2：基于最优路径

- 课后练习
  - 请使用贪心算法（结合最优路径）计算下图在$k = 2$时的最优种子集合



- 第一趟迭代

1.种子集合S为空
2.从9个节点，挑选某个节点称为u1
3.计算{u1}的激活概率
4.挑选最大的那个进入S

- 第二趟迭代

5.在剩下的节点中，挑选某个称为u2，加入S
6.计算{u1,u2}的影响概率
6.计算影响概率的变化量，即{u1,u2}的影响概率减去{u1}的激活概率
7.挑选影响概率的变化量最大的那个进入S

请参考前文$\sigma(\{1,3\}) - \sigma(\{3\})$的计算

影响力最大化2：基于最优路径

影响力最大化2：基于最优路径

- 扩展阅读

RESEARCH-ARTICLE

Scalable influence maximization for prevalent viral marketing in large-scale social networks

Authors: Wei Chen, Chi Wang, Yajun Wang

Authors Info & Claims

KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining • July 2010 • Pages 1029–1038 • https://doi.org/10.1145/1835804.1835934

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.180.4943&rep=rep1&type=pdf

影响力最大化2：基于最优路径

- 扩展阅读

Wei Chen
Microsoft Research Asia
Beijing, China
weic@microsoft.com

Chi Wang
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801 USA
chiwang1@illinois.edu

Yajun Wang
Microsoft Research Asia
Beijing, China
yajunw@microsoft.com

Abstract

Influence maximization, defined by Kempe, Kleinberg, and Tardos (2003), is the problem of finding a small set of seed nodes in a social network that maximizes the spread of influence under certain influence cascade models. The scalability of influence maximization is a key factor for enabling prevalent viral marketing in large-scale online social networks. Prior solutions, such as the greedy algorithm of Kempe et al. (2003) and its improvements are slow and not scalable, while other heuristic algorithms do not provide consistently good performance on influence spreads. In this paper, we design a new heuristic algorithm that is easily scalable to millions of nodes and edges in our experiments. Our algorithm has a simple tunable parameter for users to control the balance between the running time and the influence spread of the algorithm. Our results from extensive simulations on several real-world and synthetic networks demonstrate that our algorithm is currently the best scalable solution to the influence maximization problem: (a) our algorithm scales beyond million-sized graphs where the greedy algorithm becomes infeasible, and (b) in all size ranges, our algorithm performs consistently well in influence spread — it is always among the best algorithms, and in most cases it significantly outperforms all other scalable heuristics to as much as 100%–260% increase in influence spread.

Keywords: influence maximization, social networks, viral marketing

1 Introduction

Word-of-mouth or viral marketing differentiates itself from other marketing strategies because it is based on trust among individuals' close social circle of families, friends, and co-workers. Research shows that people trust the information obtained from their close social circle far more than the information obtained from general advertisement channels such as TV, newspaper and online advertisements [15]. Thus many people believe that word-of-mouth marketing is the most effective marketing strategy (e.g. [14]).

The increasing popularity of many online social network sites, such as Facebook, Myspace, and Twitter, presents new opportunities for enabling large-scale and prevalent viral marketing online. Consider the following hypothetical scenario as a motivating example. A small company develops a cool online application and wants to market it through an online social network. It has a limited budget such that it can only select a small number of initial users in the network to use it (by giving them gifts or payments). The company wishes that these initial users would love the application and start influencing their friends on the social network to use it, and their friends would influence their friends' friends and so on, and thus through the word-of-mouth effect a large population in the social network would adopt the application. The problem is whom to select as the initial users so that they eventually influence the largest number of people in the network.

影响力最大化2：基于最优路径

- MIIA最大影响概率的入树/MIOA最大影响概率的出树

Definition 2 (MAXIMUM INFLUENCE IN(OUT)-ARBORESCENCE) *For an influence threshold θ, the* maximum influence in-arborescence *of a node $v \in V$, $MIIA(v,\theta)$, is*

$$MIIA(v,\theta) = \cup_{u \in V, pp(MIP_G(u,v)) \geq \theta} MIP_G(u,v).$$

The maximum influence out-arborescence $MIOA(v,\theta)$ *is:*

$$MIOA(v,\theta) = \cup_{u \in V, pp(MIP_G(v,u)) \geq \theta} MIP_G(v,u).$$

u1/u2/u3/... 到 v 的最大影响路径，从中挑选影响概率大于θ的

从 v 到 u1/u2/u3/... 的最大影响路径，从中挑选影响概率大于θ的

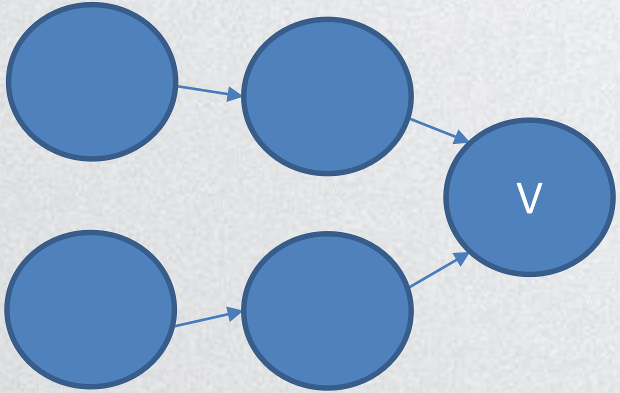
arborescence

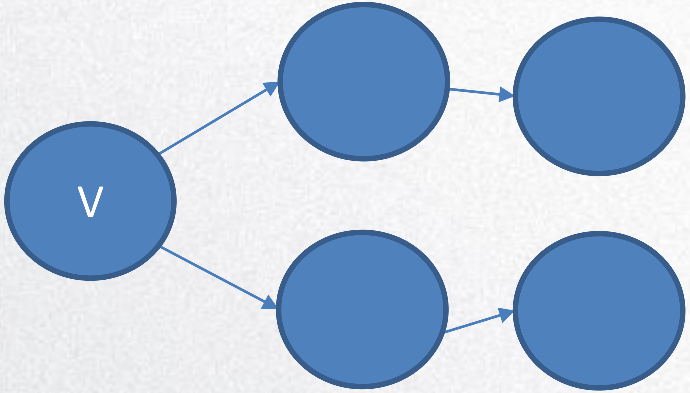
n. 树（枝，乔木）状；树质;

[例句] New Thought of "Arborescence Structure" about Teaching
树状结构教学新思路

影响力最大化2：基于最优路径

- MIIA最大影响概率的入树/MIOA最大影响概率的出树





Intuitively, $MIIA(v, \theta)$ and $MIOA(v, \theta)$ give the local influence regions of v, and different values of θ controls the size of these local influence regions.

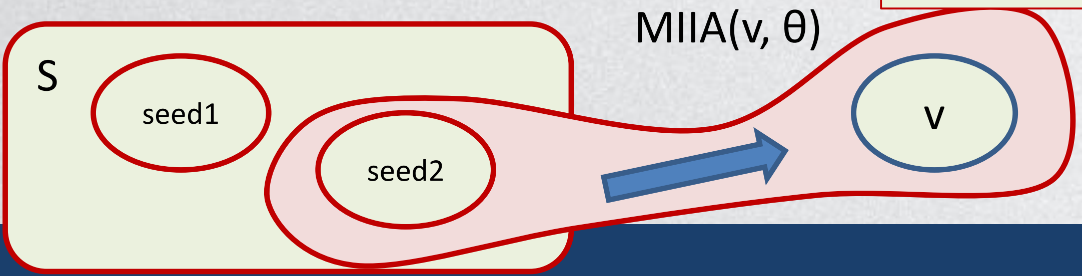
影响力最大化2：基于最优路径

- 基于最优路径的影响力最大化

Algorithm 1 $ap(u, S, MIIA(v, \theta))$

```
if u ∈ S then
    ap(u) = 1
else if N^in(u) = ∅ then
    ap(u) = 0
else
    ap(u) = 1 − Π_{w∈N^in(u)} (1 − ap(w) · pp(w, u))
end if
```

- 1. a set of seeds S
- 2. MIIA(v, θ) for some v ∉ S
- 3. we approximate the IC model by assuming that the influence from S to v is only propagated through edges in MIIA(v, θ). With this approximation, we can calculate the probability that v is activated given S exactly
- 4. Let the activation probability of any node u in MIIA(v, θ), denoted as ap(u, S, MIIA(v, θ)), be the probability that u is activated when the seed set is S and influence is propagated in MIIA(v, θ)



AP为activate probability

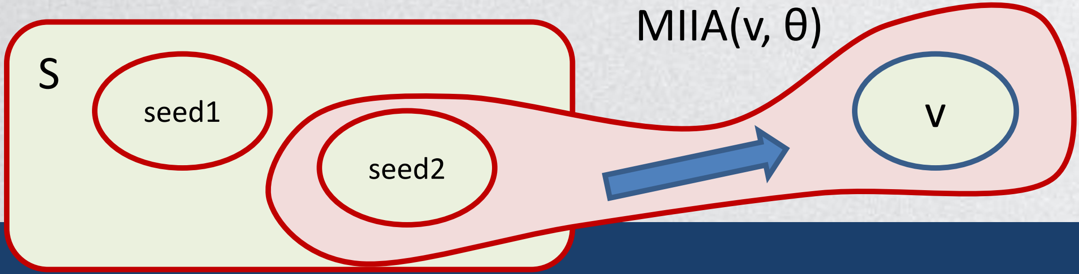
影响力最大化2：基于最优路径

- 基于最优路径的影响力最大化

Algorithm 1 $ap(u, S, MIIA(v, \theta))$

if $u \in S$ **then**
  $ap(u) = 1$
else if $N^{in}(u) = \emptyset$ **then**
  $ap(u) = 0$
else
  $ap(u) = 1 - \Pi_{w \in N^{in}(u)}(1 - ap(w) \cdot pp(w, u))$
end if

如果u已经在S里面，那么它被自己激活，那么active probability为1



影响力最大化2：基于最优路径

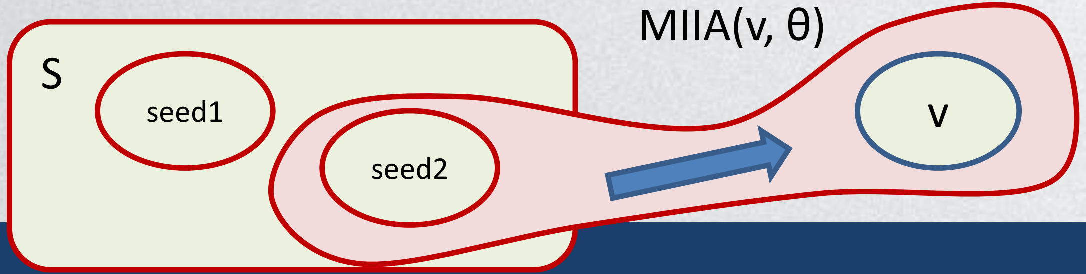
- 基于最优路径的影响力最大化

Algorithm 1 $ap(u, S, MIIA(v, \theta))$

if $u \in S$ **then**
  $ap(u) = 1$
else if $N^{in}(u) = \emptyset$ **then**
  $ap(u) = 0$
else
  $ap(u) = 1 - \Pi_{w \in N^{in}(u)}(1 - ap(w) \cdot pp(w, u))$
end if

U的In neighbors，和MIIA(v, θ)的交集为空，那么active probability为0

- let $N^{in}(u, MIIA(v, \theta))$ be the set of in-neighbors of u in MIIA(v, θ)



影响力最大化2：基于最优路径

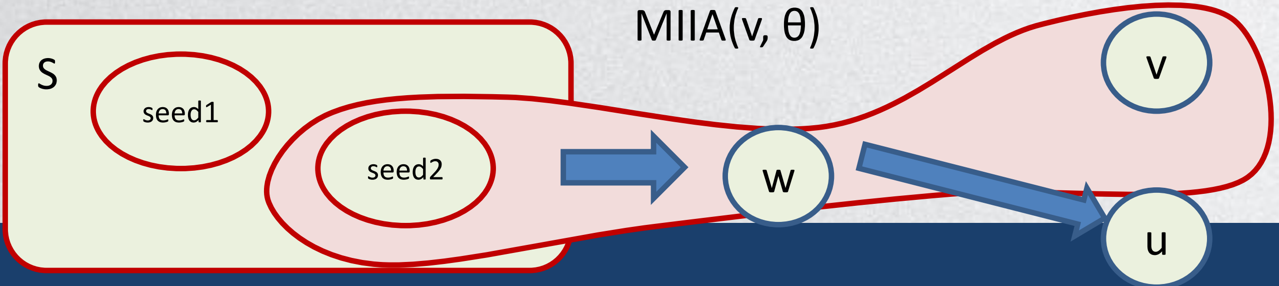
- 基于最优路径的影响力最大化

Algorithm 1 $ap(u, S, MIIA(v, \theta))$

if $u \in S$ **then**
  $ap(u) = 1$
else if $N^{in}(u) = \emptyset$ **then**
  $ap(u) = 0$
else
  $ap(u) = 1 - \Pi_{w \in N^{in}(u)}(1 - ap(w) \cdot pp(w, u))$
end if

1.考虑所有w in N^{in}
2.先考虑S激活w
2.再考虑w到u的最大激活路径概率
3.最后假设哪个w都不激活u，最后用1减去，即得到u的ap

- let $N^{in}(u, MIIA(v, \theta))$ be the set of in-neighbors of u in MIIA(v, θ)



影响力最大化2：基于最优路径

- S的影响范围

In our MIA model we assume that seeds in S influence every individual node v in G through its $MIIA(v, \theta)$. Let $\sigma_M(S)$ denote the influence spread of S in our MIA model, then we have

$$\sigma_M(S) = \sum_{v \in V} ap(v, S, MIIA(v, \theta)). \qquad (1)$$

影响力最大化2：基于最优路径

- 递推计算各个节点的ap

Note that the recursive computation of $ap(u)$ in Algorithm 1 can be transformed into an iterative form such that all $ap(u)$'s with u in $MIIA(v, \theta)$ can be computed by one traverse of the arborescence $MIIA(v, \theta)$ from leaves to the root. Thus, computing $\sigma_M(S)$ using Equation (1) and Algorithm 1 is polynomial-time.

影响力最大化2：基于最优路径

- 递推计算各个节点的ap
  - 结合标准greedy算法

Together with standard greedy Algorithm , we already have a polynomial-time approximation algorithm.

Algorithm standard Greedy(k, f)

1: initialize $S = \emptyset$
2: **for** $i = 1$ to k **do**
3: select $u = \arg\max_{w \in V \setminus S}(f(S \cup \{w\}) - f(S))$
4: $S = S \cup \{u\}$
5: **end for**
6: output S