

# 传染病的网络传播模型

(第21章)

# 疾病传播模型：基本要素

接触网络

传染概率

疾病特征

关心的问题：

- 多大范围（规模）？要不要采取措施？
  - 许多流行病会自生自灭，例如季节性的
- 多长时间？会不会反复（“死灰复燃”）
- 对患者影响的严重性（生命、损伤、不适）

# 流行病传播与社会行为传播的差别

- 已经讨论过的传播（影响）模型

- 总体模型——总体效应

- 网络模型——朋友关系网络结构

- 基于门槛值的模型

决策模型

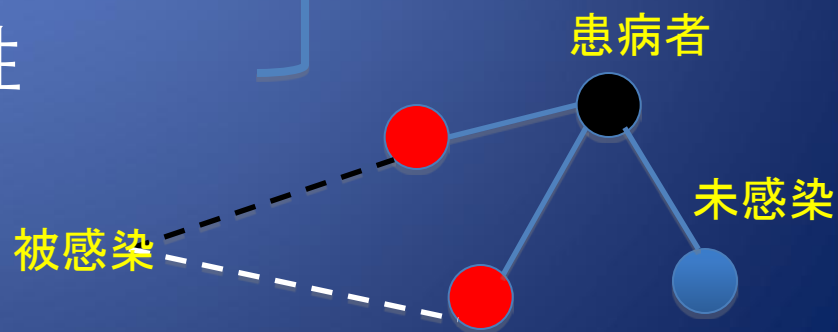
- 流行病传播特点

- 传播原因：传染

- 传播途径：接触关系网络结构

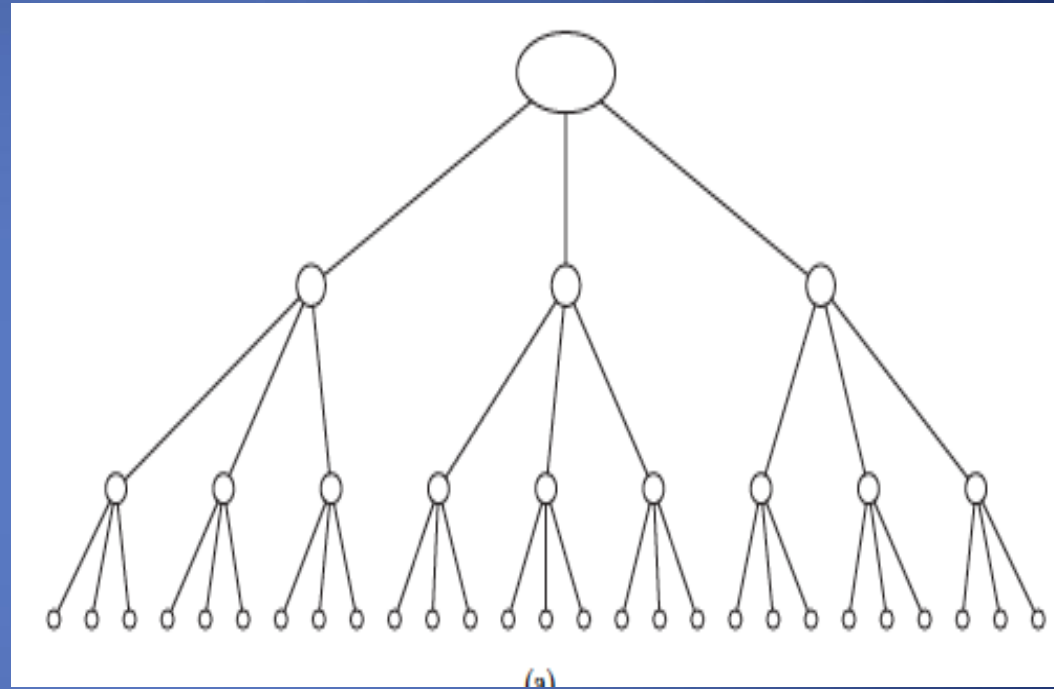
- 传播方式：具有随机特性

随机模型



# 分支过程：一个最简单的疾病传播模型

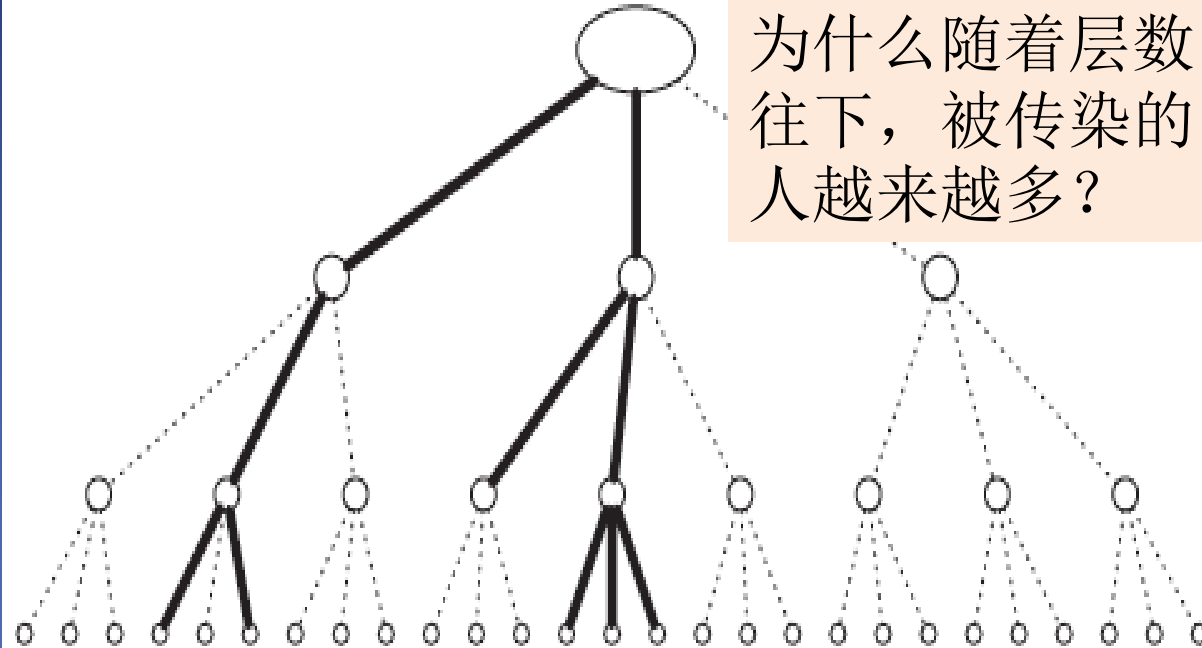
- 分支过程形成树结构接触网络，最初一个人携带病菌进入人群，以一个独立的概率  $p$  传染给遇到的每个人。假设疾病感染期间每个人遇到  $k$  个其他人



如此简单、“不实际”  
的模型有什么用？

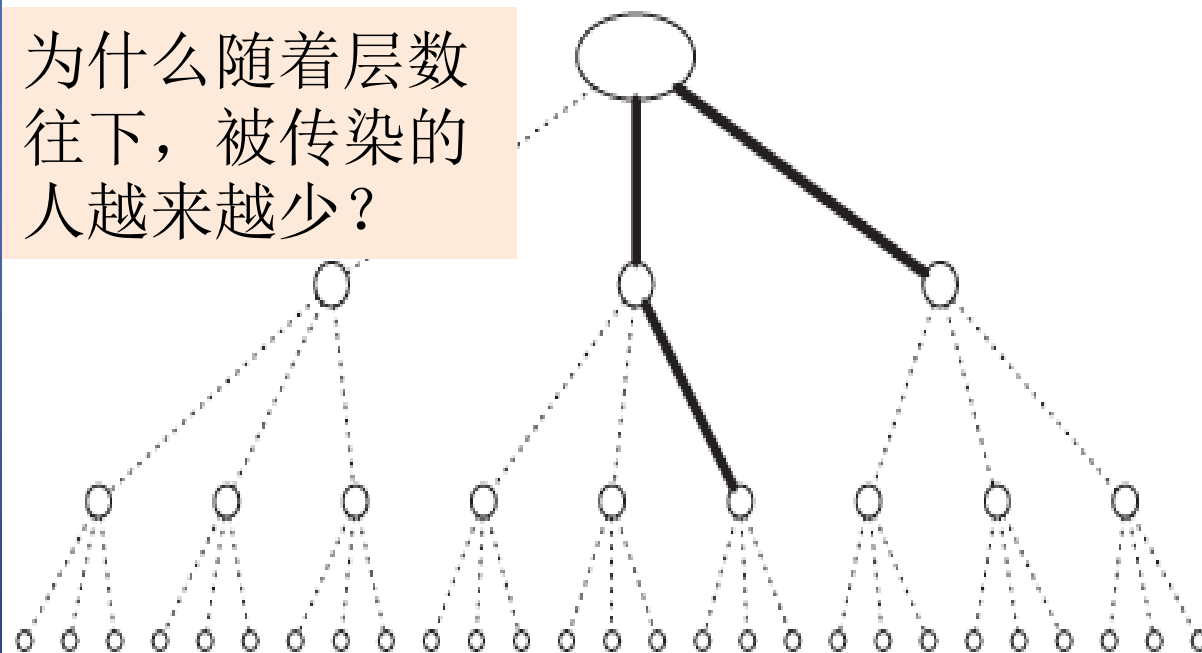
# 传染过程停止的条件？

为什么随着层数往下，被传染的人越来越多？



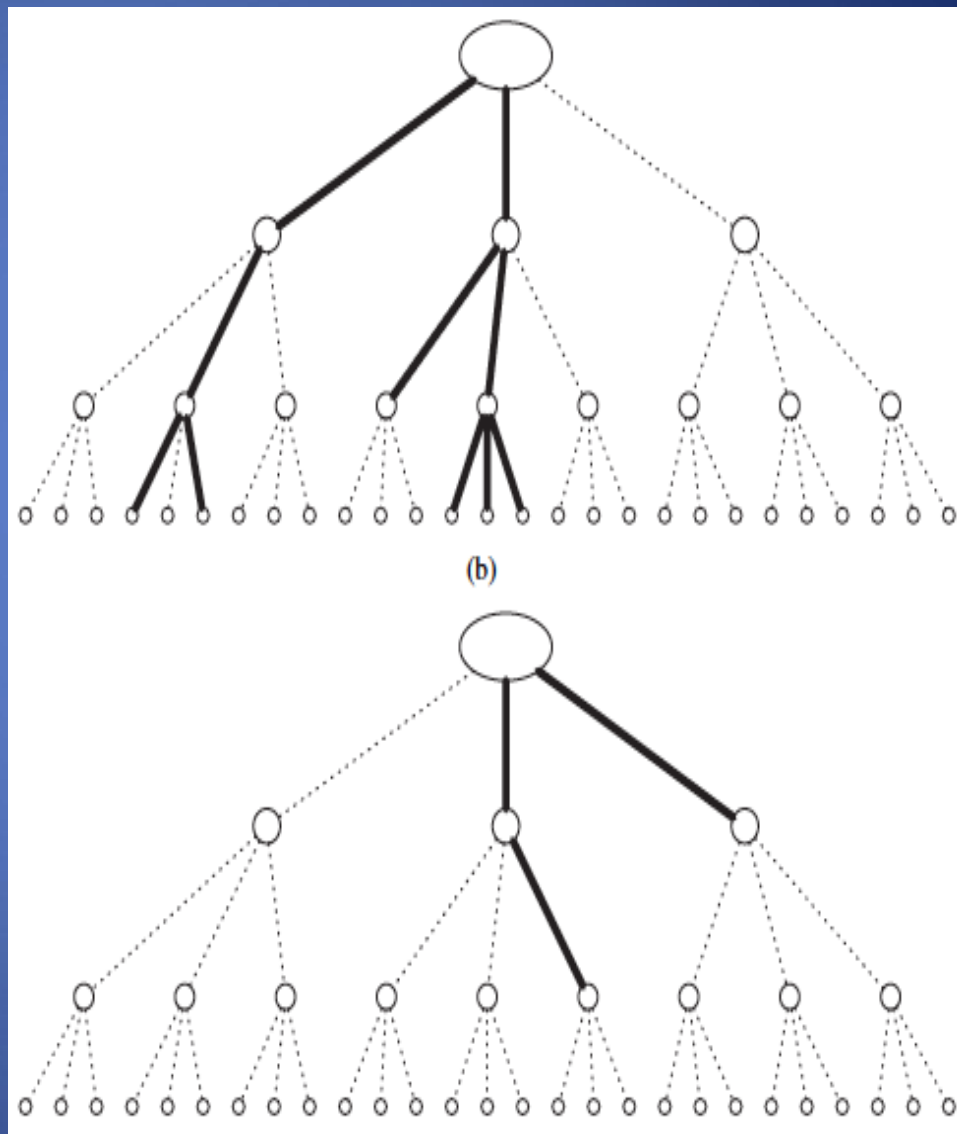
(b)

为什么随着层数往下，被传染的人越来越少？



# 基本再生数 (reproductive number)

- **基本再生数**：由单一个体引起的新发病例数期望值，记为 $R_0$ ,  $R_0 = pk$
- **断言**：如果 $R_0 < 1$ ，则疾病将以概率1在有限的过程后消失。如果 $R_0 > 1$ ，则疾病持续在每一波以 $>0$ 的概率至少感染一个人



# 基本再生数的特点

- 因为 $R_0 = p \cdot k$ ，当 $R_0$ 接近于1时，稍微改变接触人数 $k$ 或传染概率会对结果产生很大影响。 $R_0$ 接近于1时有一个“刀刃”特性
  - 如果 $R_0$ 略低于1，稍微增加传染概率 $p$ ；结果可能会使 $R_0$ 最终高于1，造成一个突然的疾病爆发
  - 略微减少疾病的传染性，可以将 $R_0$ 减小到1以下，消除疾病大范围流行的风险
- **结论：**当门槛值 $R_0 = 1$ 附近，社会应该付出努力减小基本再生数（隔离，减少接触就是措施之一）
- 实际的疾病传播要复杂得多，分支结构没有考虑三角关系，弱连接……

# SIR疾病传播模型

- SIR模型可以描述更一般性网络结构的传播情况，接触网络中的节点在流行病传播过程中经历三个状态(susceptible,infectious,removed)：
  - 易感状态(S)：易感者，容易被一个患病邻居节点感染
  - 传染状态(I)：患病者，以一定的概率将疾病传染给处于易感状态的邻居
  - 移出状态(R)：痊愈并免疫者，当一个节点经历了完整的病期康复后，就不再会被传染，也不会再对其他节点造成威胁，相当与从接触网络中移出了



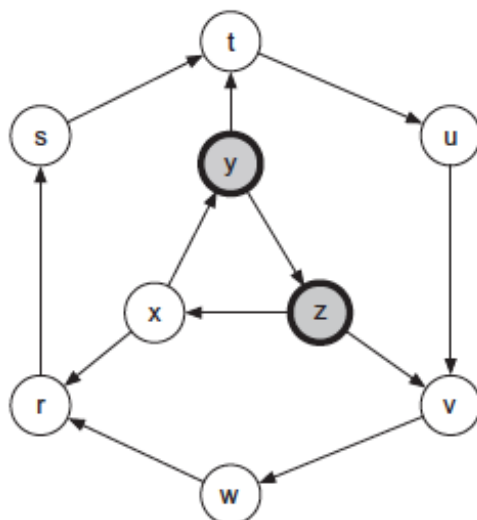
# SIR模型描述

- SIR模型流行病的传播过程取决于接触网络结构，以及两个量值： $p$ （传染的概率）和  $t_i$ （传染期的长度）
  - 最初，一些节点处于状态 I，所有其他节点处在状态 S。
  - 每个进入状态 I 的节点  $v$  在固定的步骤  $t_i$  期间具有传染性。在  $t_i$  的每一步， $v$  以概率  $p$  将疾病传染给它的处于易感状态的邻居
  - 经过  $t_i$  步后，节点  $v$  进入移出状态（R），成为网络中的惰性节点，不会再被传染，也不会向其他人传播

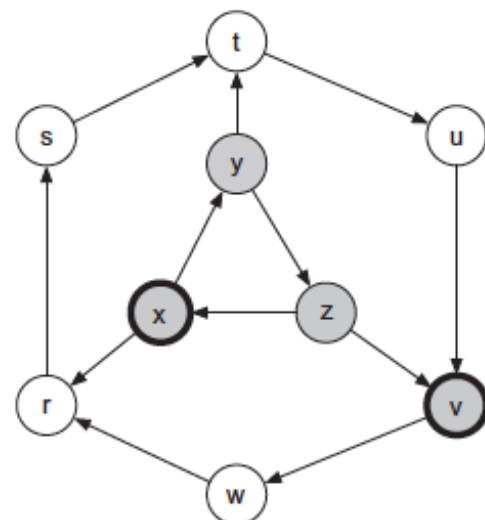


# 一个 $\tau_i = 1$ 的SIR模型示例

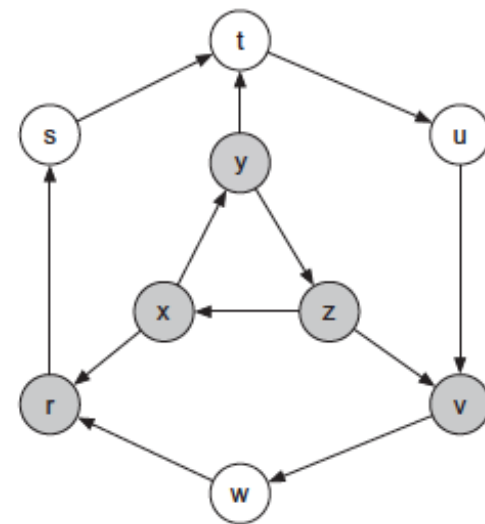
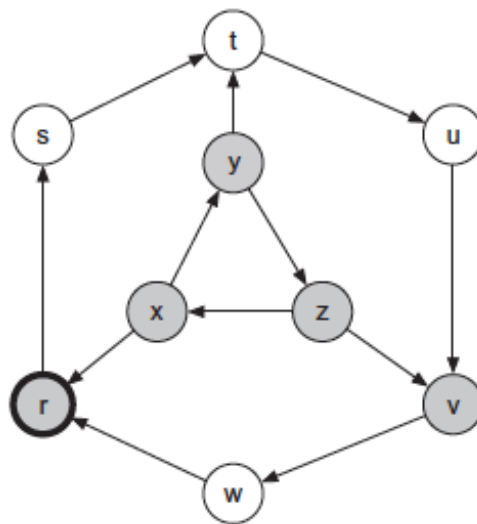
- 接触网是一个有向图
- 图中有三种不同的节点
  - 白色（易感）
  - 粗边界的灰色（传染）
  - 细边界的灰色（移出）



(a)

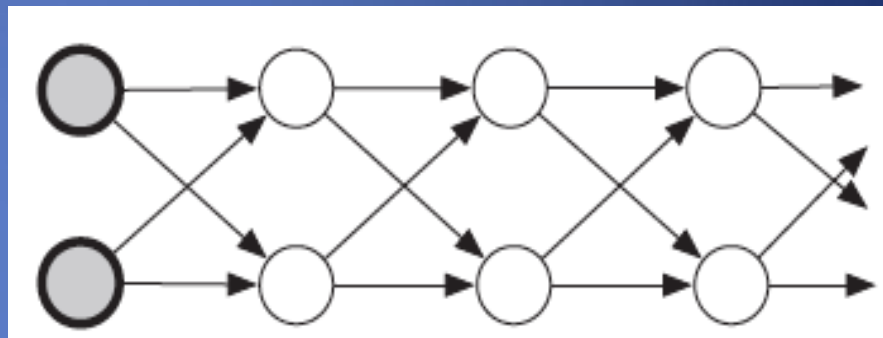


(b)



# 非树型网络的 $R_0$ 不再有“刀刃”结论

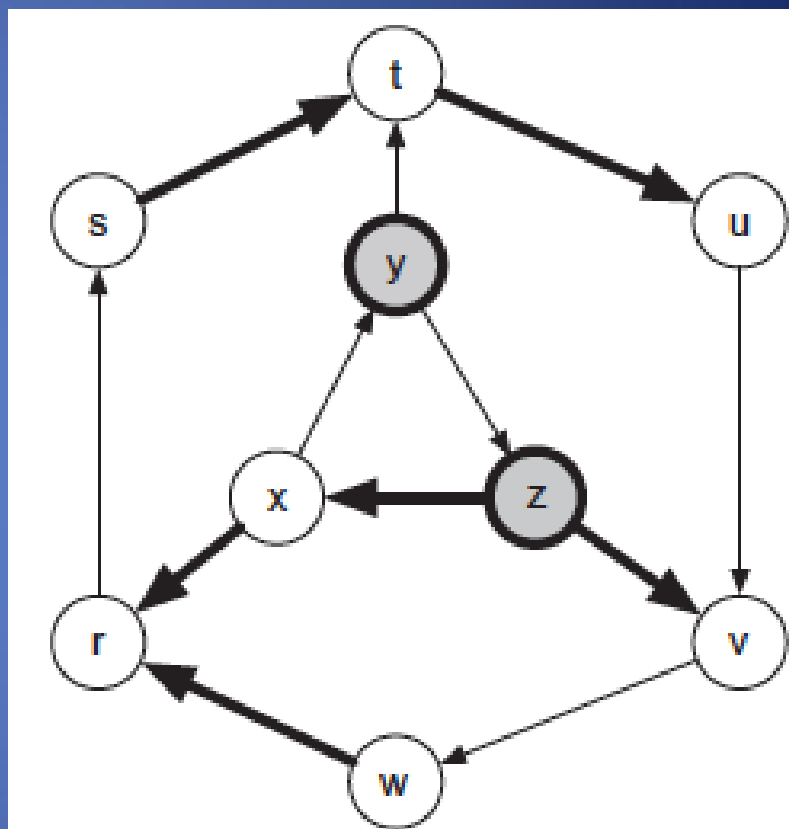
- 网络由每层两个节点组成向右无限伸延， $t_i = 1$ ，传染概率 $p$ 为 $2/3$ ，最左边的两个节点是最初传染者
- $R_0$ 是由一个节点造成的新病例，预期为 $4/3 > 1$
- 每条边不传播疾病的概率为 $1/3$ ，所有四条边都不会传播疾病的概率为 $(1/3)^4$ ，每一层成为最后一层（疾病传染终止层）的概率至少是 $1/81$



这种结构相当于形成了一个路障，疾病无法超越它向远方传播

# SIR模型（运行）的静态图示

- 接触网络每条边预先投掷硬币（按照给定传染概率 $p$ ），投掷硬币成功的边定义为 **开放边**；其余的边定义为 **阻塞边**
- 利用开放边和阻塞边描述流行病传播过程：最终受感染的节点正是那些能够从最初感染节点沿着在网络中随机选定的 **开放边**能够达到的节点
- 意义:有利于对随机现象的观察



图中，粗黑线为开放边。  
注意到不是所有开放边都起作用的（例如  $w \rightarrow r$ ）

# SIR模型的形式化描述

- 易感人数、传染人数、移出人数分别为  $S(t), I(t), R(t)$ ，占比分别为  $s(t), i(t), r(t)$
- $\beta$ : 个人单位时间与其他人平均有效接触次数
- $\gamma$ : 单位时间治愈人数
- SIR模型描述为:

$$\frac{ds(t)}{dt} = -\beta \cdot i(t) \cdot s(t)$$

$$\frac{di(t)}{dt} = \beta \cdot i(t) \cdot s(t) - \gamma \cdot i(t)$$

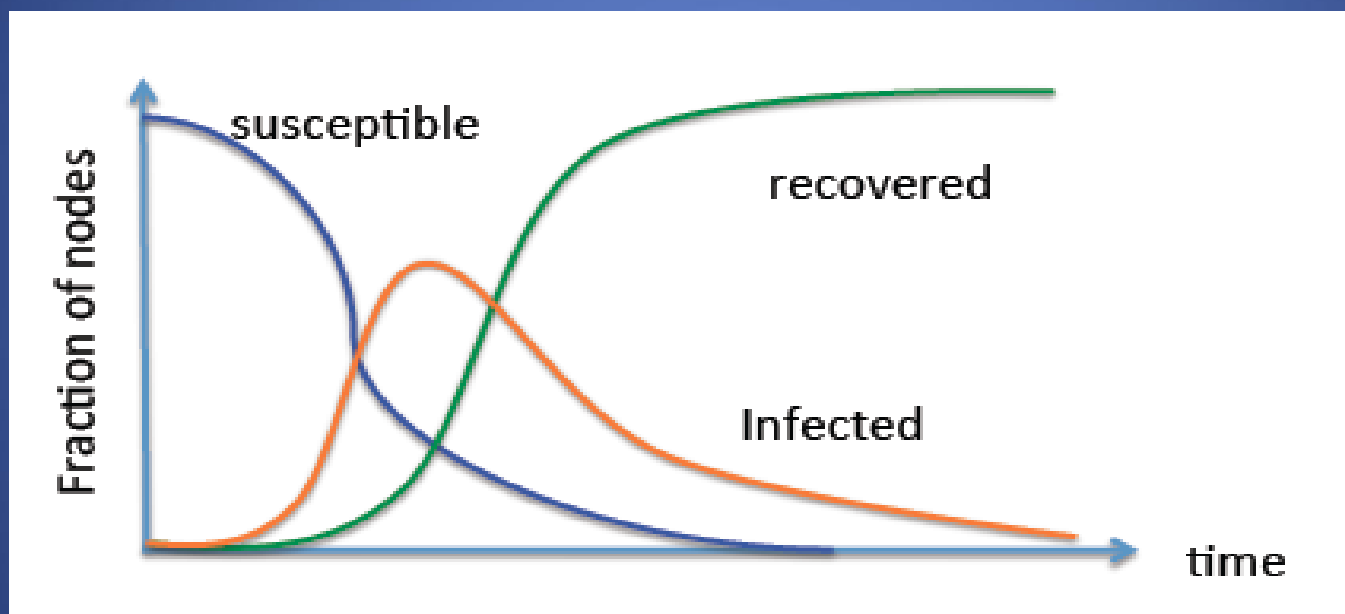
$$\frac{dr(t)}{dt} = \gamma \cdot i(t)$$

$$s(t) + i(t) + r(t) = 1$$

来自最早提出SIR模型的文献:  
[Kermack W O, McKendrick A G.  
Contributions to the mathematical  
theory of epidemics, part I[J].  
Proceedings of the Royal Society  
of London A, 1927, 115: 700–721]

# 一个给定 $\beta$ 和 $\gamma$ 的SIR时间曲线

- 设 $\beta=1, \gamma=0.4, s(\text{at start})=0.99, i(\text{at start})=0.01, r(\text{at start})=0$

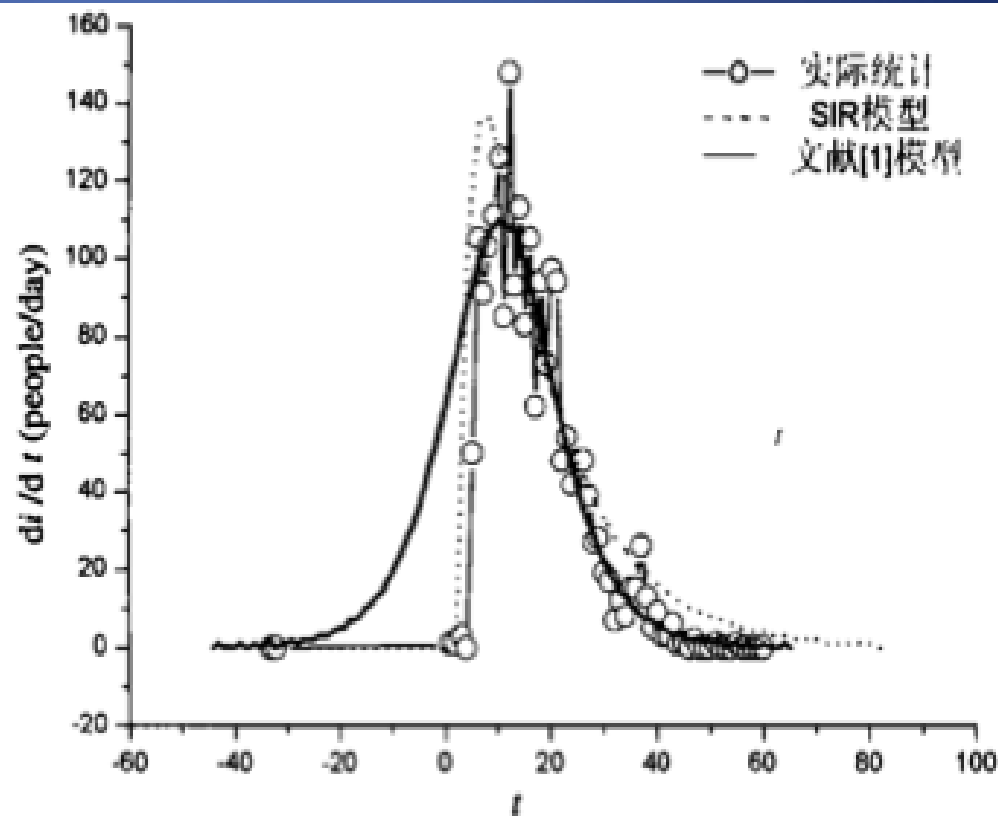


- 疾病什么时候消失呢？

取决于： $\beta/\gamma$ 。如果发病率高于痊愈率，则疾病会持续，否则疾病将消失

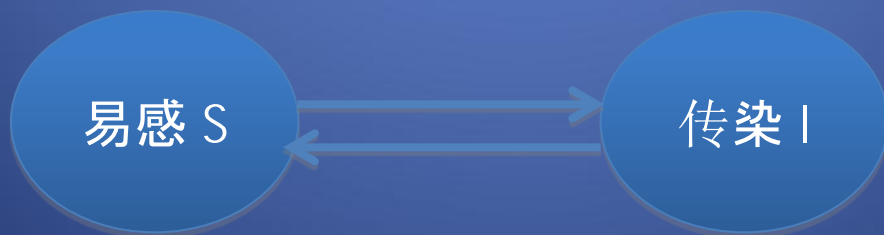
# 利用SIR模型分析SARS传播

- 利用SIR模型中 $di/dt$ 方程模拟北京在2003年4月18日到6月16日的每天新增病例
- 模拟中使用的参数是： $\beta=0.148$ ,  $\gamma=0.85$ , 初始值为： $s(0)=1199$ ,  $i(0)=1$ （这里总人数应该看作是接触到病源的易感、传染、免疫的人数之和,而不是北京市总人口）



# SIS流行病传播模型

- SIR模型只适用于一次患病终身免疫的情况。SIS模型描述了更一般的疾病传播。节点结束传染状态后再回到易感状态，节点在 S 和 I 两种状态中交替，因此得名**SIS模型**
  - 最初，一些节点处于状态 I，其余节点处于状态 S
  - 每个进入状态 I 的节点  $v$  在  $t_i$  期间内具有传染性，在  $t_i$  的每一步，节点以概率  $p$  将疾病传染给其处在易感状态的邻居
  - 经过  $t_i$  步后，感染节点不再具有传染性，返回状态 S



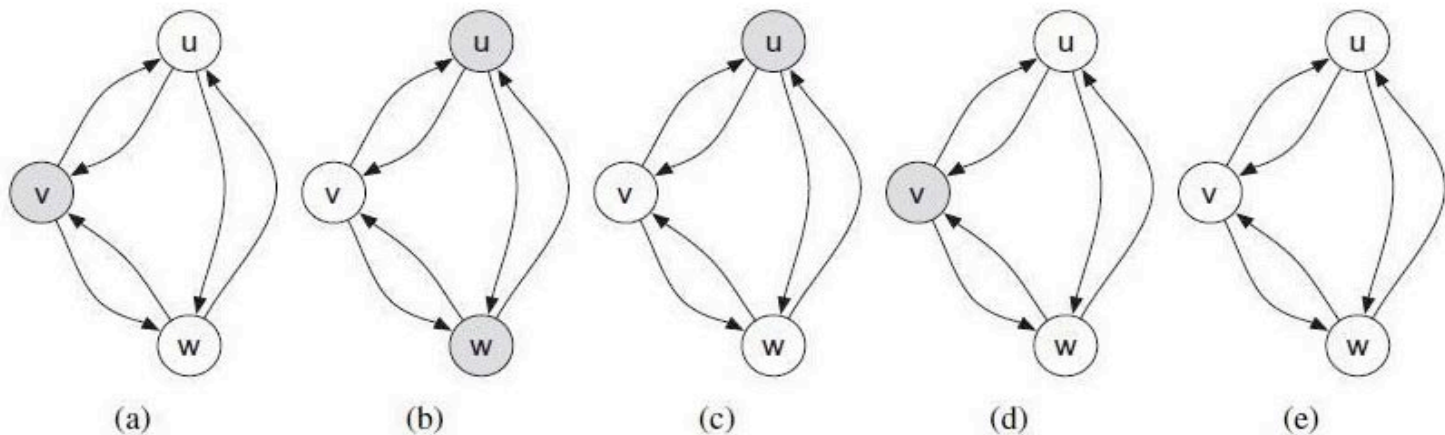


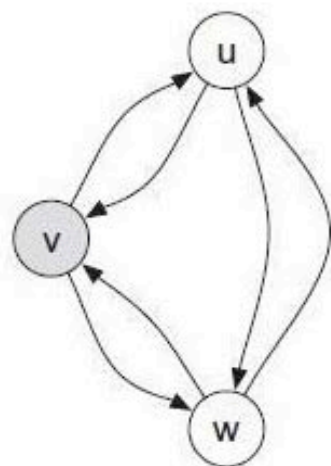
# SIS模型的特点

- SIS模型刻画的流行病的生命周期会比SIR模型的要长
- 基本再生数具有刀刃特点：
  - 可以证明，对于某些类型的接触网络，存在特定传染概率 $p$ ，在其附近，网络中的SIS流行病在一边表现为“快速消失”，在另一边则会“持续一个很长的时间”。该传染概率临界值以微妙的方式依赖于网络结构。

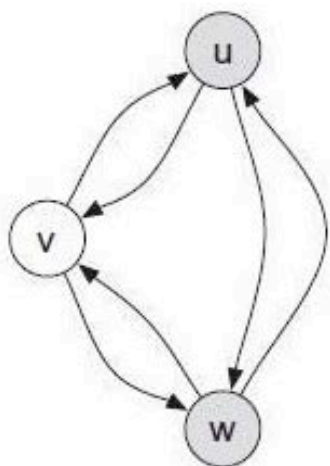
# SIR与SIS的关系

- SIS模型可以看成是SIR模型的特例
- 设  $t_1=1$ ，将患病节点 $v$ 在每一个时间步视为“不同的个体”，可以将SIS模型表示为SIR模型
- 在每个时间周期  $t=0、1、2、3\cdots$ ，为每个节点分别创建一个节点副本，对于原网络中 $v$ 连接到 $w$ 的边，复制  $t$  时刻扩展网络中  $v$  到  $t+1$  时刻 $w$ 的边

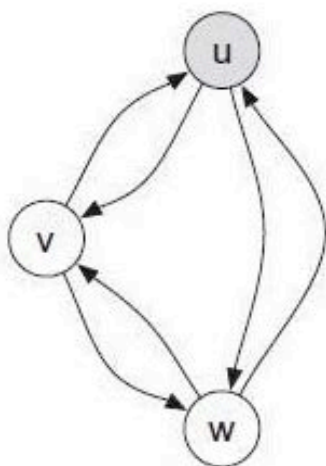




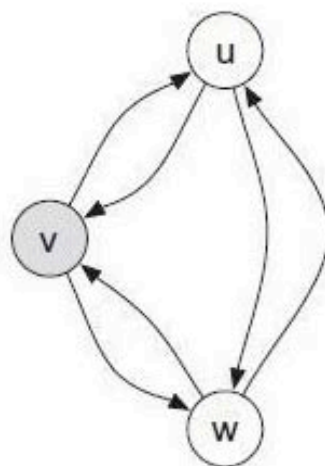
(a)



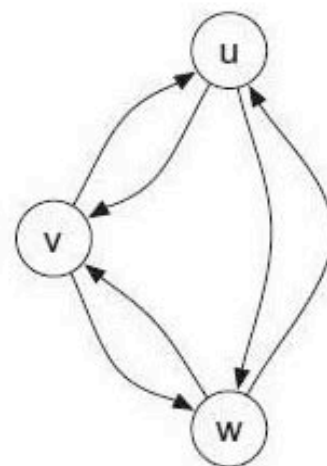
(b)



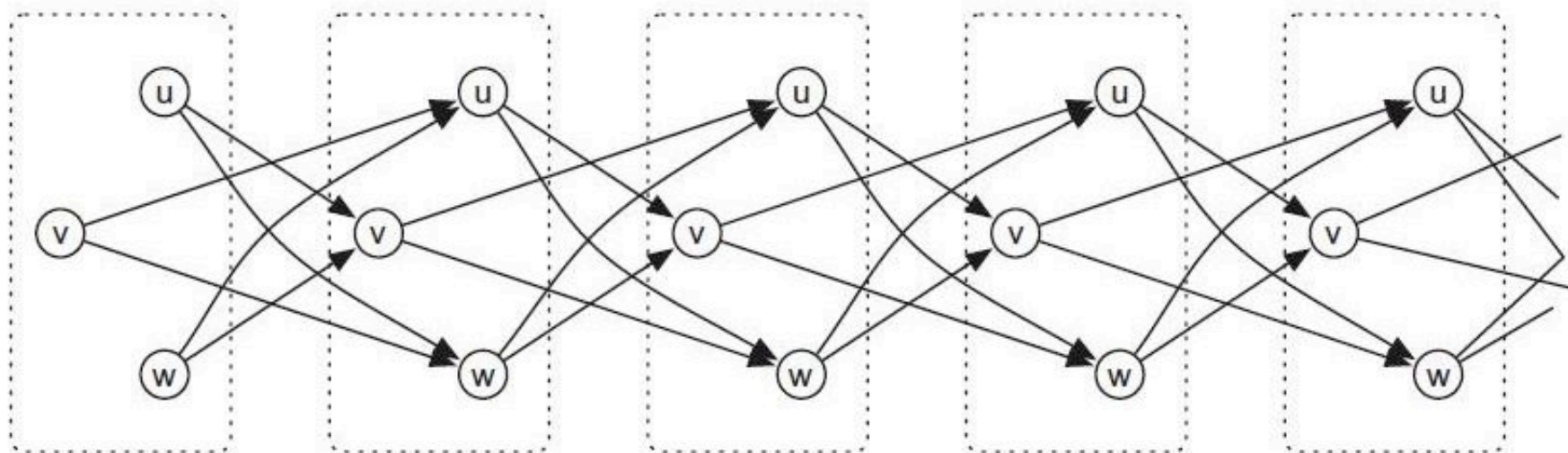
(c)



(d)



(e)



step 0

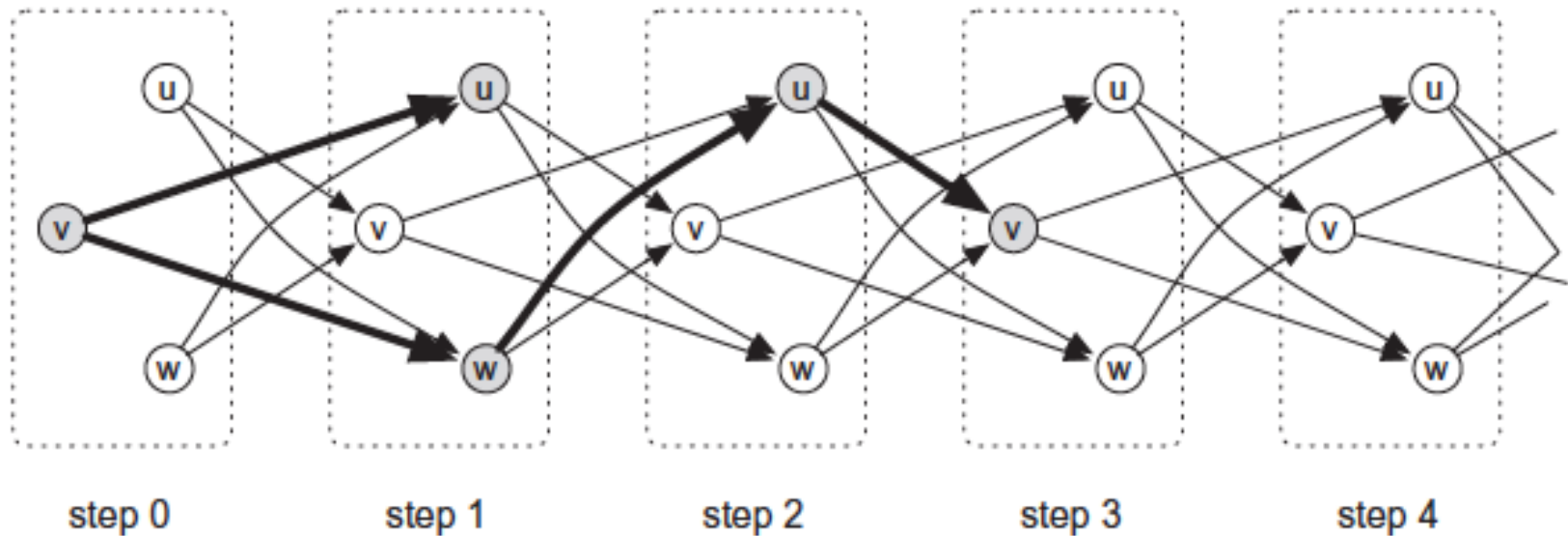
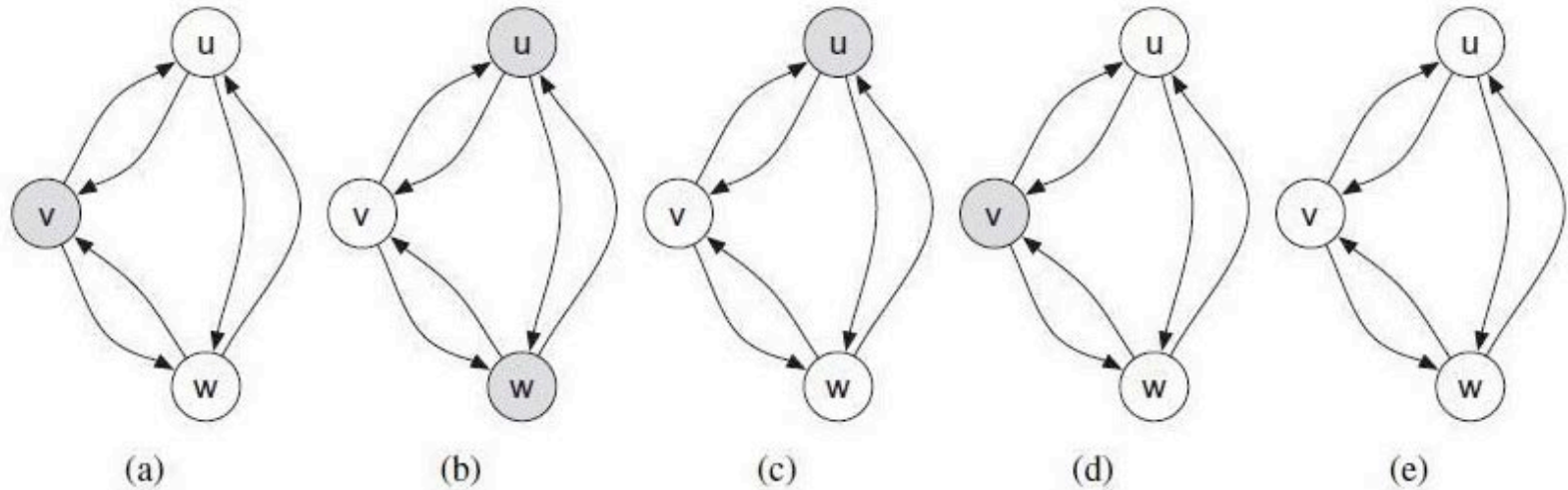
step 1

step 2

step 3

step 4

# SIR与SIS的关系

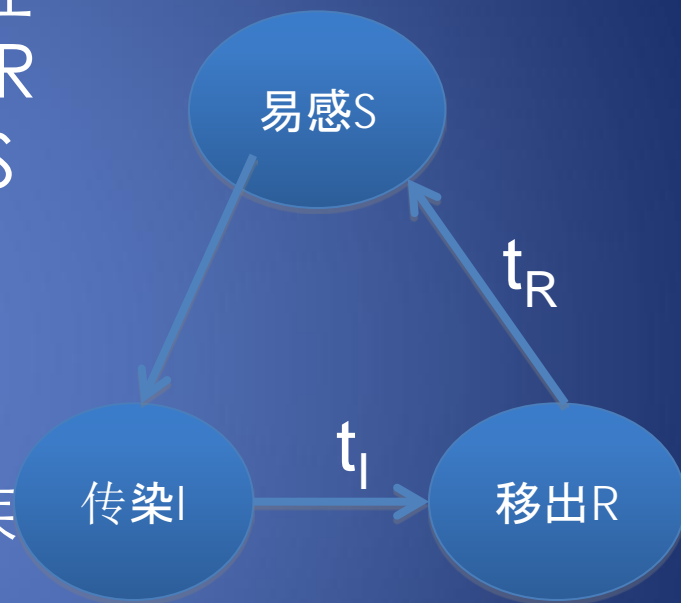


# SIRS 流行病传播模型

- 结合SIR和SIS，疾病有暂时免疫特性：  
受传染节点恢复后短暂地经过R状态，然后再次进入S状态——SIRS模型

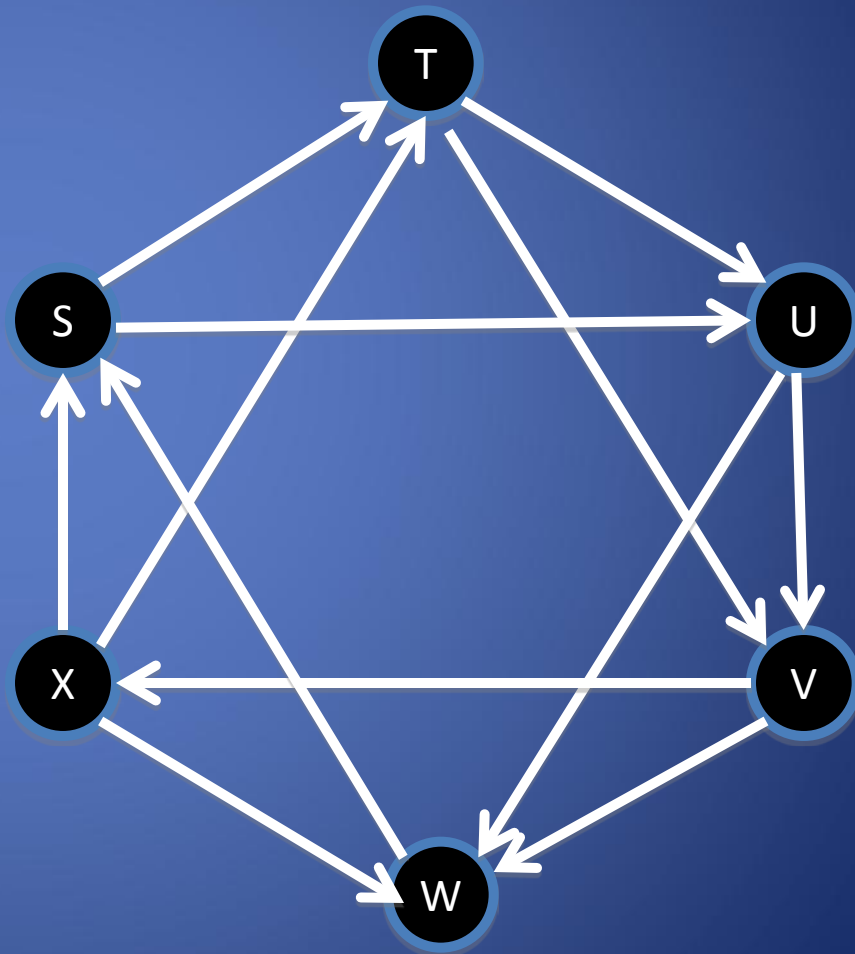
- 开始某些节点在I状态，其他在S状态
- 进入I状态的节点v在固定时间  $t_I$  内具有传染性，在  $t_I$  的每一步以概率  $p$  将疾病传播给易感状态的邻居
- 经过  $t_I$  步后，节点进入状态R，经过  $t_R$  后再次回到S状态

疾病传播过程受量值  $p$ 、 $t_I$  以及  $t_R$  的影响，同时受到网络结构的影响



# SIRS模型用于研究疾病传播的震荡效应

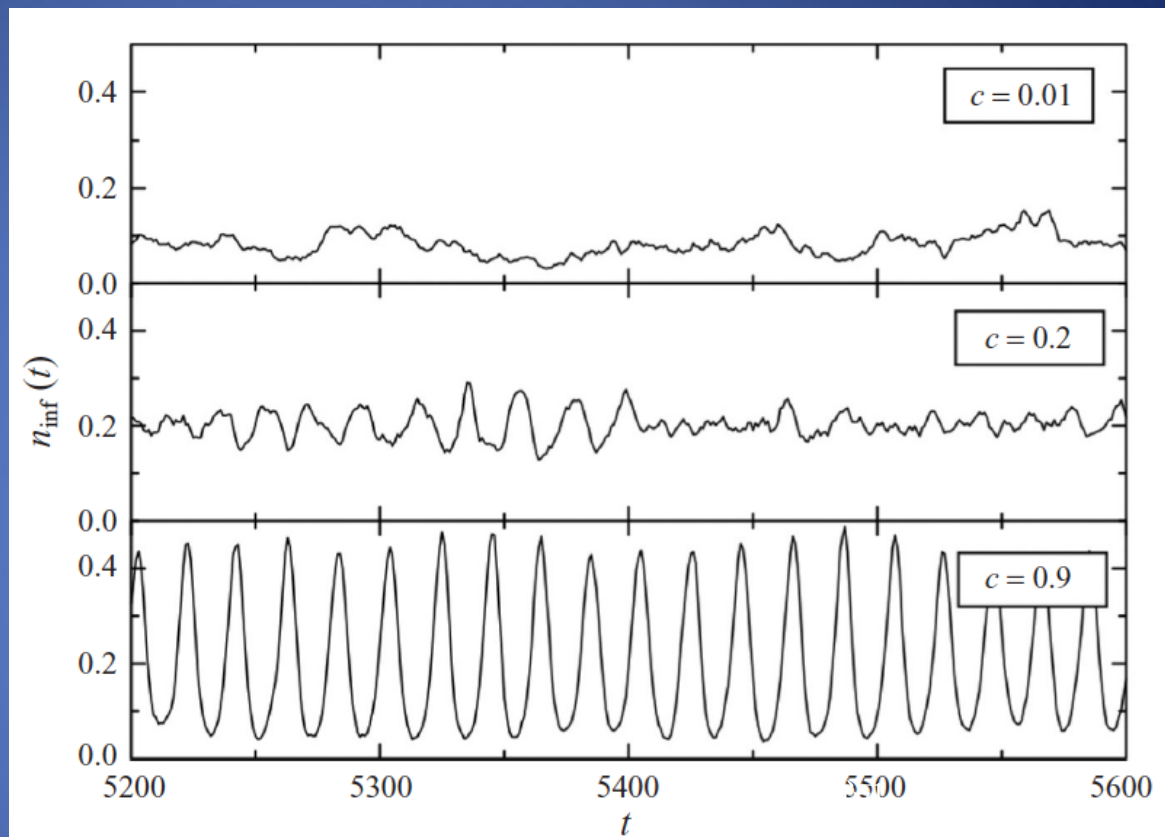
| S  | T  | U  | V  | W  | X  |
|----|----|----|----|----|----|
| S  | I  | S  | S  | S  | S  |
| .. | R  | I  | .. | .. | .. |
| .. | S  | R  | I  | .. | .. |
| .. | .. | S  | R  | I  | .. |
| .. | .. | .. | S  | R  | I  |
| I  | .. | .. | .. | .. | .. |
| R  | I  | .. | .. | .. | .. |
| S  | R  | I  | .. | .. | .. |
| .. | S  | R  | I  | .. | .. |
| .. | .. | S  | R  | I  | .. |
| .. | .. | .. | S  | R  | I  |





# Watts-Strogatz小世界模型上运行SIRS模型

- $\tau_I=4$ ,  $\tau_R=9$
- $N_{inf}(0)=0.1$
- 模拟  $c=0.01$ ,  $0.2, 0.9$  三种情况



结论：当弱连接概率较高时，所有节点的感染周期趋于同步。同样的时间点被感染，持续相同的时间，进入隔离，之后又同时被感染

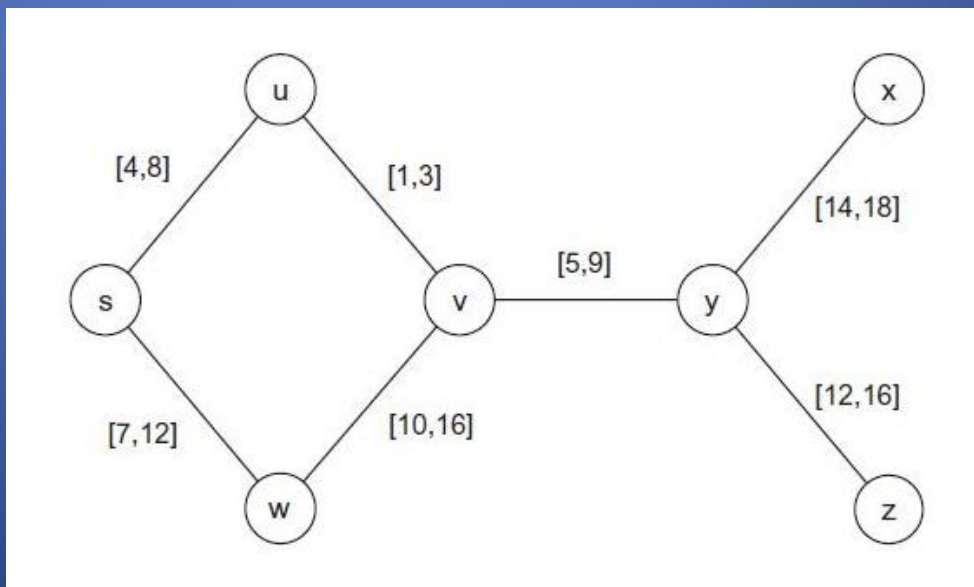
# 为什么会有同步震荡？

- 当 $p$ 较小时：0~0.5，网络呈现较高的有序的聚集特点，疾病周期会在一个局部区域内完成，之后进入免疫期（不活跃期）
- 当 $p$ 值较大时： $>0.5$ ，弱连接打破了有序的聚集区域，刚刚痊愈的个体可能又通过一个弱连接被再次感染。因此整个网络成同步状态



# 模型中可以考虑的一些其他因素

- 感染期不同时段传染性可以不同
- 不同个体传染性的不同
- 患病者可能在传染期间不同的阶段恢复
- 接触网络的边具有时间特性
- .....



# 一个生物学基因遗传的研究

- 线粒体夏娃理论：1987年，Rebecca Cann、Mark Stoneking和Allan Wilson在《自然》杂志上发表的一篇论文。
- 每个人都有一条母系祖先的线索，称为**母亲的血统**。所有这些血统源自10万至20万年前（14万年）的一个单一女子，可能是在非洲，她是我们所有人的母系祖先。

# 一点生物遗传知识

基本单位

特定的脱氧核苷酸序列代表遗传信息，每个基因含多个脱氧核苷酸

脱氧核苷酸 a

基因 b

具有遗传效应的 DNA 片段，每个 DNA 分子含多个基因

DNA c

主要载体，每条染色体含有 1 个或 2 个 DNA 分子

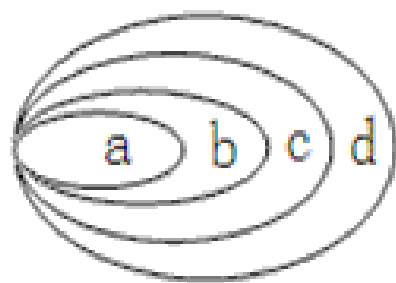
染色体 d

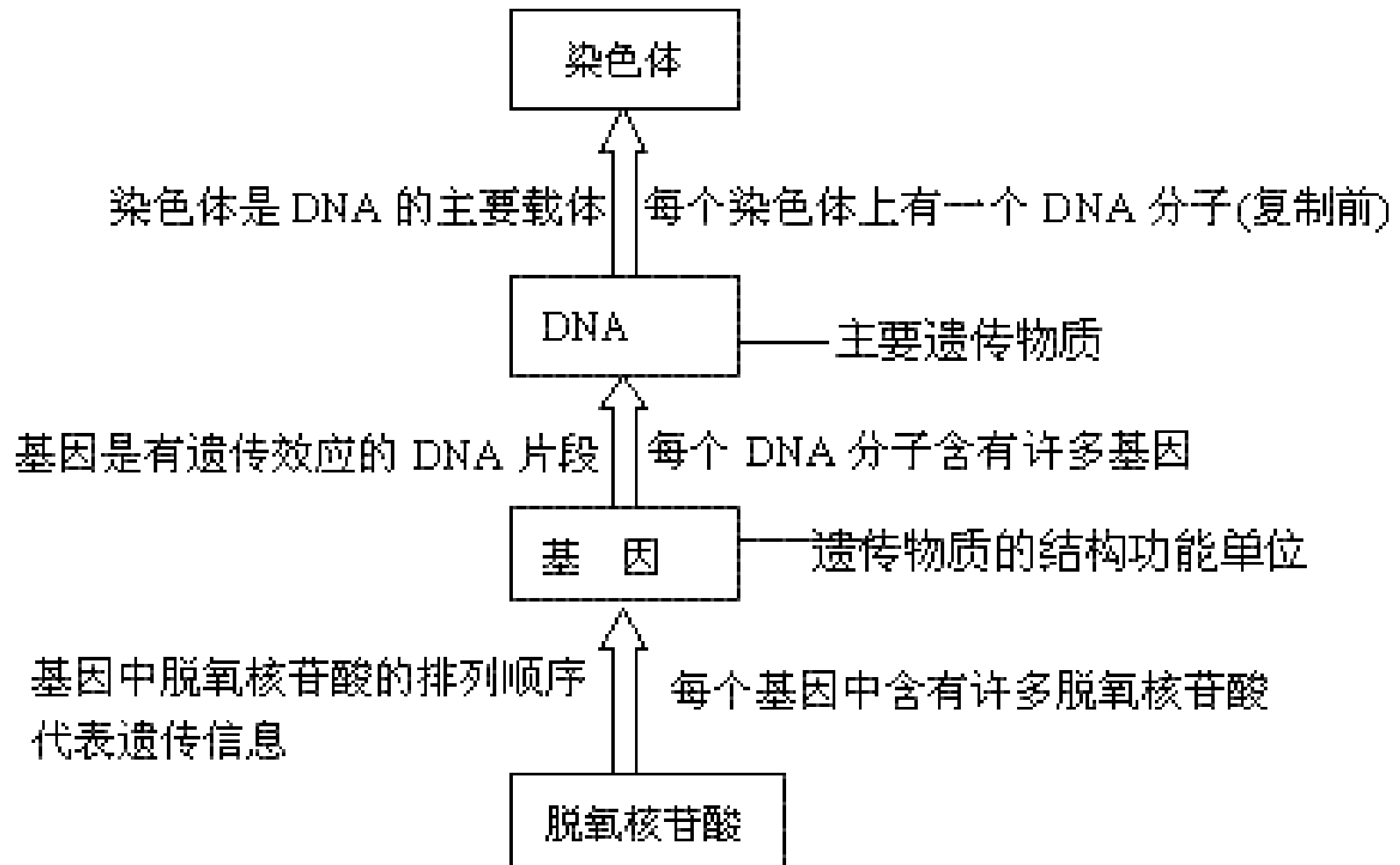
解旋

高度螺旋化，缩短变粗

染色质

基因在染色体上呈直线排列





核

可综合细胞内其他各个结构的运行，通过核孔运输物质。

染色质

分裂时，可浓缩成染色体。

高尔基体

动植物都有它。与储存、排泄有关。

仁

含有蛋白质和核糖酸。

细胞膜

中心粒

在动物细胞分裂中起固定中心的作用。

线粒体

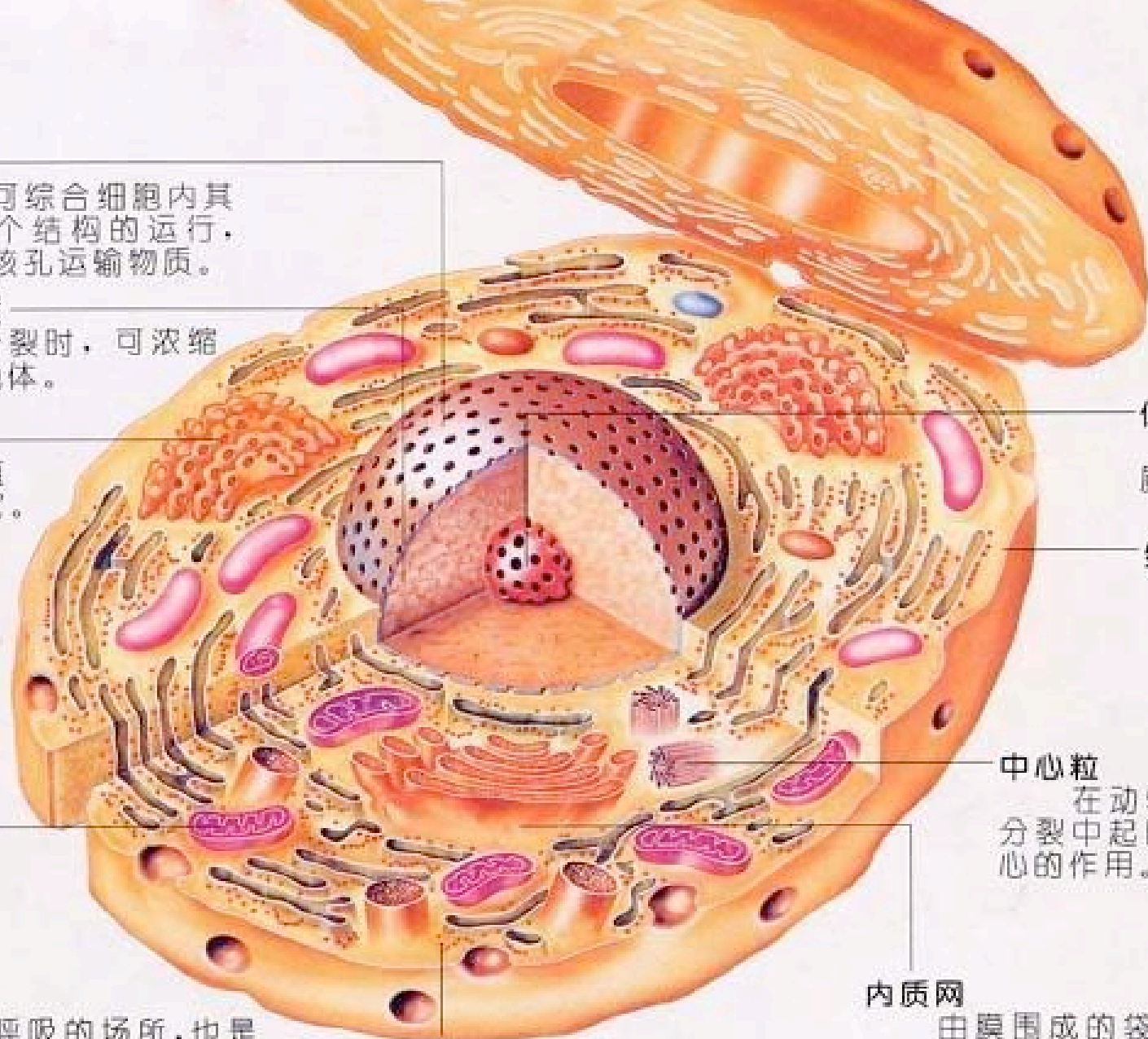
进行呼吸的场所，也是制造三磷酸腺苷的工厂。氧气和食物分子发生反应，生出三磷酸腺苷。

核糖体

分布在核周围和内质网上，是蛋白质的合成场所。

内质网

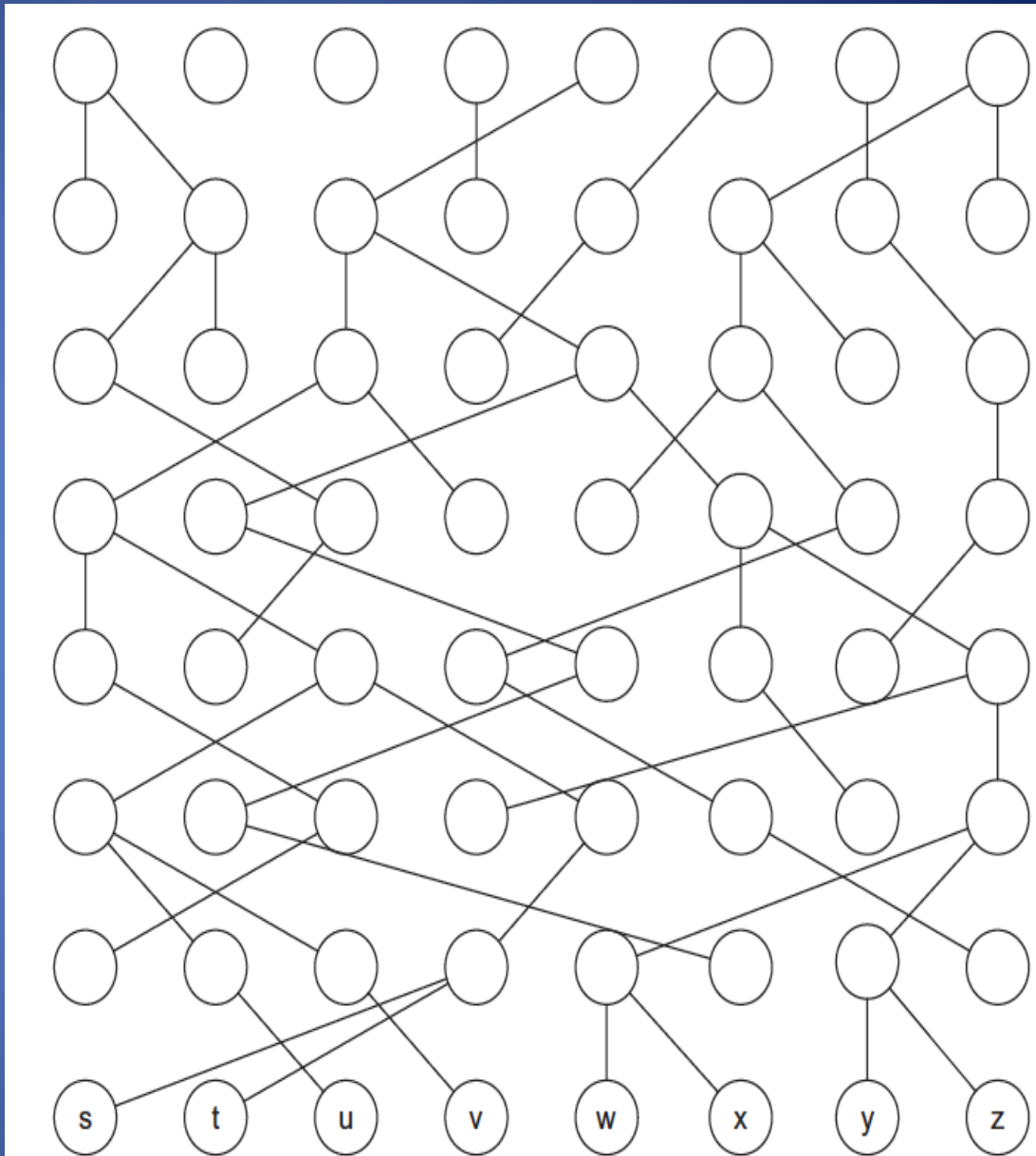
由膜围成的袋状或管状系统。它的一端与外膜连着，起物质输送作用。



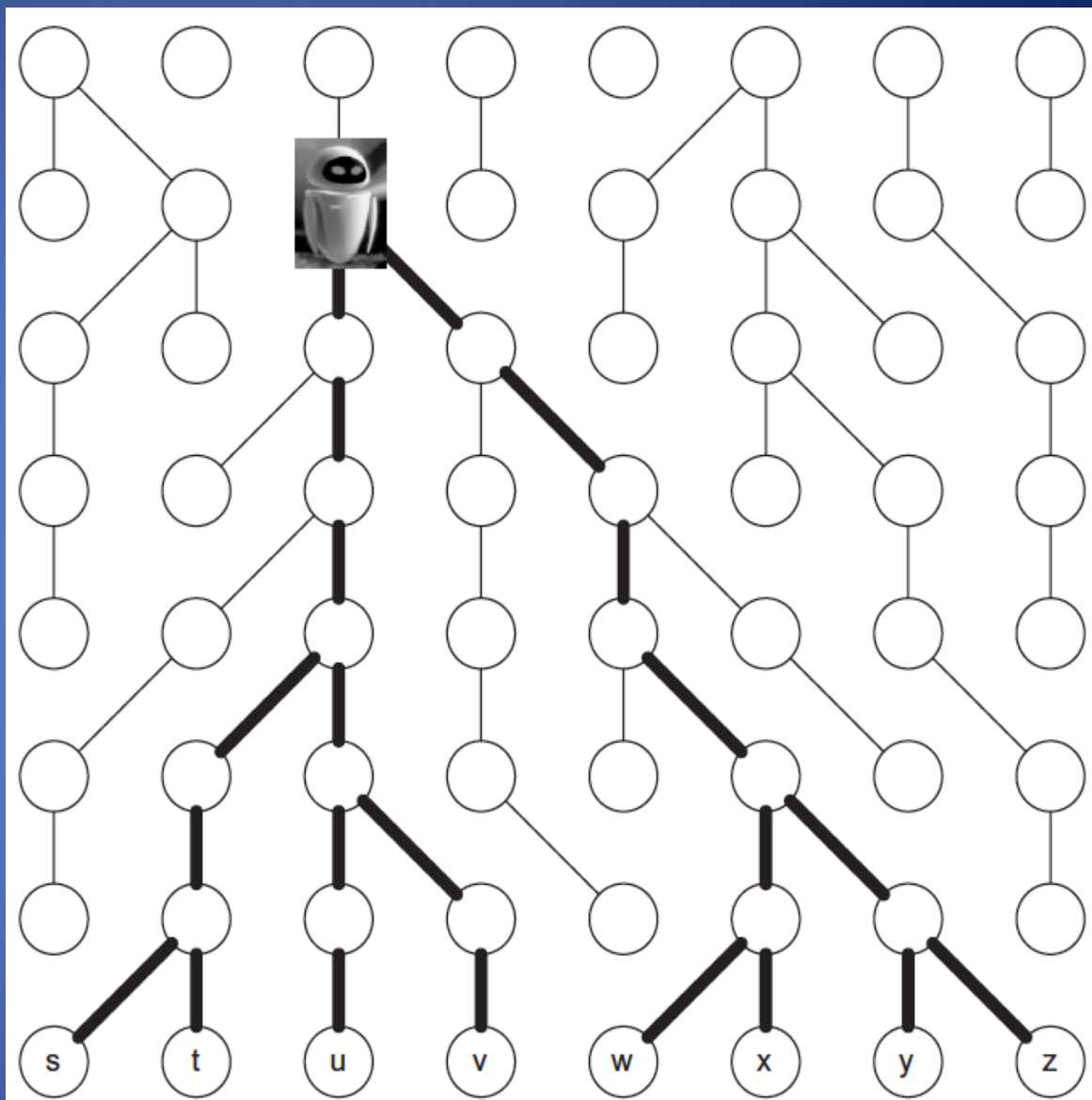
# 线粒体遗传的网络结构动力学解释

- Wright-Fisher单亲祖先模型：
  - 每一代的人口数量维持在固定的 $N$
  - 新的一代是由当前一代 $N$ 个个体产生总共 $N$ 个后代
  - 每个后代由一个单亲产生，由当前一代中独立地均匀随机选择，当前一代特定的个体可能会有多个孩子

- 每个个体都向上连接到自己的母亲；
- 时间自上向下运行，最底层表示当前有N个个体
- 从底部开始，任何一个个体都可以沿着向上的边追溯他／她的单亲血统
- 能否说明这些边最终将汇聚到一点？



- 简单的概率分析，可以发现当前一定数量的个体在向上追溯单亲祖先的过程中会不断发生汇聚（发现同一个母亲），于是独立的线索越来越少，最终汇聚到一点
- 这个距当代最近的汇聚点就是线粒体夏娃





# 基因遗传学的解释

- 当时肯定同时生活着许多女人和许多男人，只不过她们的线粒体基因没有遗传到现在。
- 不仅是线粒体基因存在这种情形，任何一种基因，只要把它们突变过程追溯得足够远，总能找到一个共同祖先，这些祖先生活的时间可能各不相同，甚至未必是人，有的可能要追溯到人进化出来之前的某个动物

# Wright-Fisher单亲祖先模型的意义

- 可以分析单亲物种的遗传特点
- 可以分析双亲物种中那些只由单亲遗传的现象，如线粒体DNA
- 分析社会继承行为（包括学术继承等）
- .....

# 小结

- 疾病传染的简单网络模型：
  - 分支结构、SIR、SIS、SIRS
  - 不同模型下的不同现象与问题
- 生物遗传过程也可以自然看成是一个网络，从中可能发现出人意料的现象

# 作业

- 第21章，练习第1题。