

# 小世界现象

# 小世界现象

- **Stanley Milgram, 1933-1984**
  - The Small-world Problem, Psychology Today, 1967
- 基本实验
- 两点启示
  - “小世界”现象
  - “自我寻的”现象
- 体现小世界现象的模型
- 体现自我寻的的现象的模型



- 模型在实践中的验证与推广

# Stanley Milgram 的实验

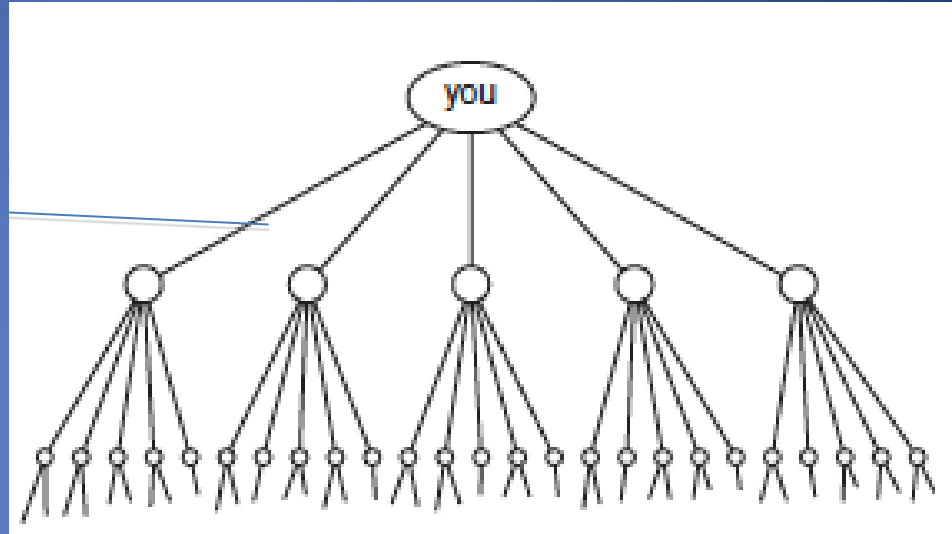
- 几百名“初始者”，要求每人尝试通过转发让一个指定的人收到一封信；
- 向每个初始者提供了目标收信人的姓名、地址、职业等个人信息；
- 规定：参与者只能将信件转发给能直呼其名的熟人，并请他继续转发。因此，如果一个参与者不认识目标收信人，则他不能直接将信寄给他；
- 要求参与者力争让这信件能尽早达到目的地；
- 结果，约三分之一的信件经过平均六次转发到达了目标

# 实验结果带给人们两点惊奇

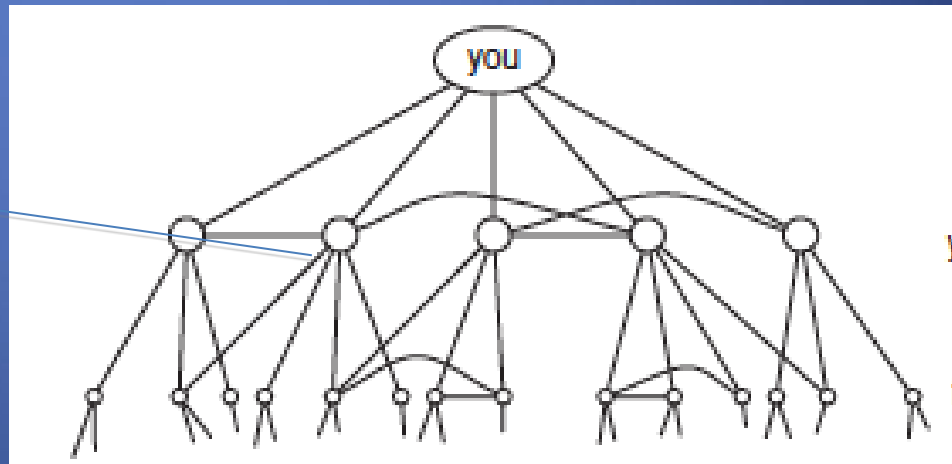
1. 社会网络中包含丰富的**短路径**
  2. 通过“有意识的转发”能够“自动地”找到这些短路径
- 问题
    - 为什么社会网络具有这样的性质？它们源于社会网络的哪些基本原理？
    - 反过来说，能否依据社会网络的某些基本原理，构建出反映这种性质的网络模型？
  - （完全随机的网络不具备这样的性质）

# 短路径的存在性初探

没有三元闭包  
的朋友关系。  
可通过很短路  
径到达任何人。



体现三元闭包  
的朋友关系。  
是否一定有短  
路径就不明显  
了。



# 形成社会网络的两种基本力量

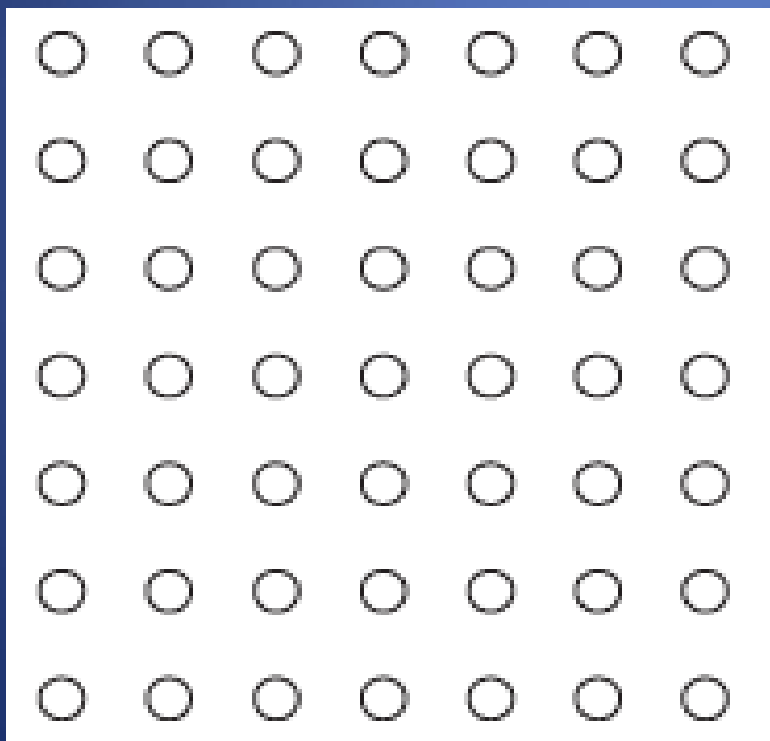
- 同质性（选择，社会影响），三元闭包
  - 家庭成员，邻里关系，同学，同事
  - 对应社会网络中的大量的“三角形”
  - 体现某种“亲近”（例如地理范围的）
- 弱联系
  - 偶然的原因，认识的“远程”朋友
  - 对其所在的圈子并不一定熟悉

一种什么样的形式化网络，既体现这两种力量的作用，也便于我们分析其中是否有小世界现象？

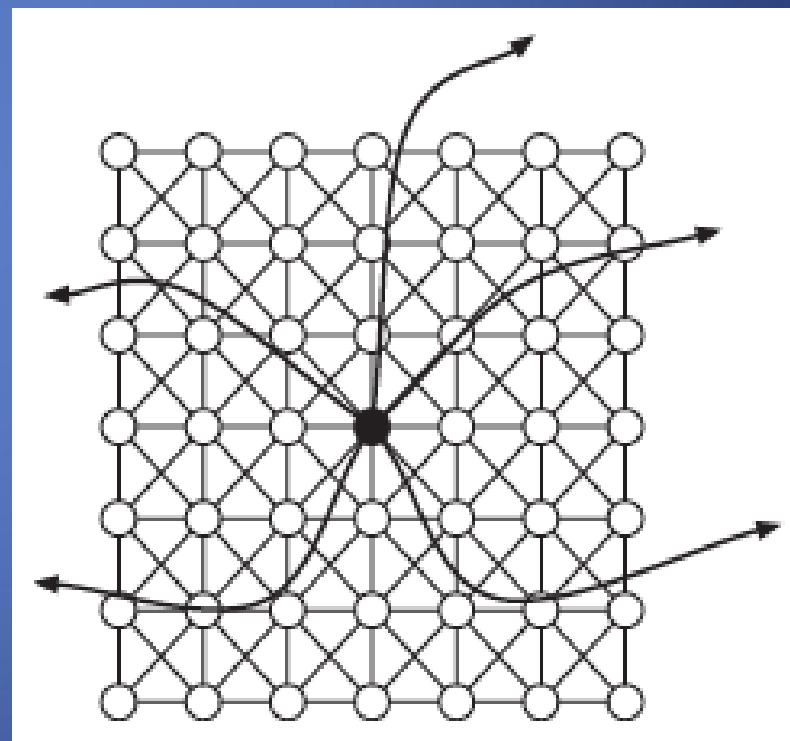
# Watts-Strogatz模型 [1998]



- 定义一种图（网络），它体现这两种力量  
– 存在许多“三角形”和少数随机的“远程边”



想象大量节点排布成均匀网格状



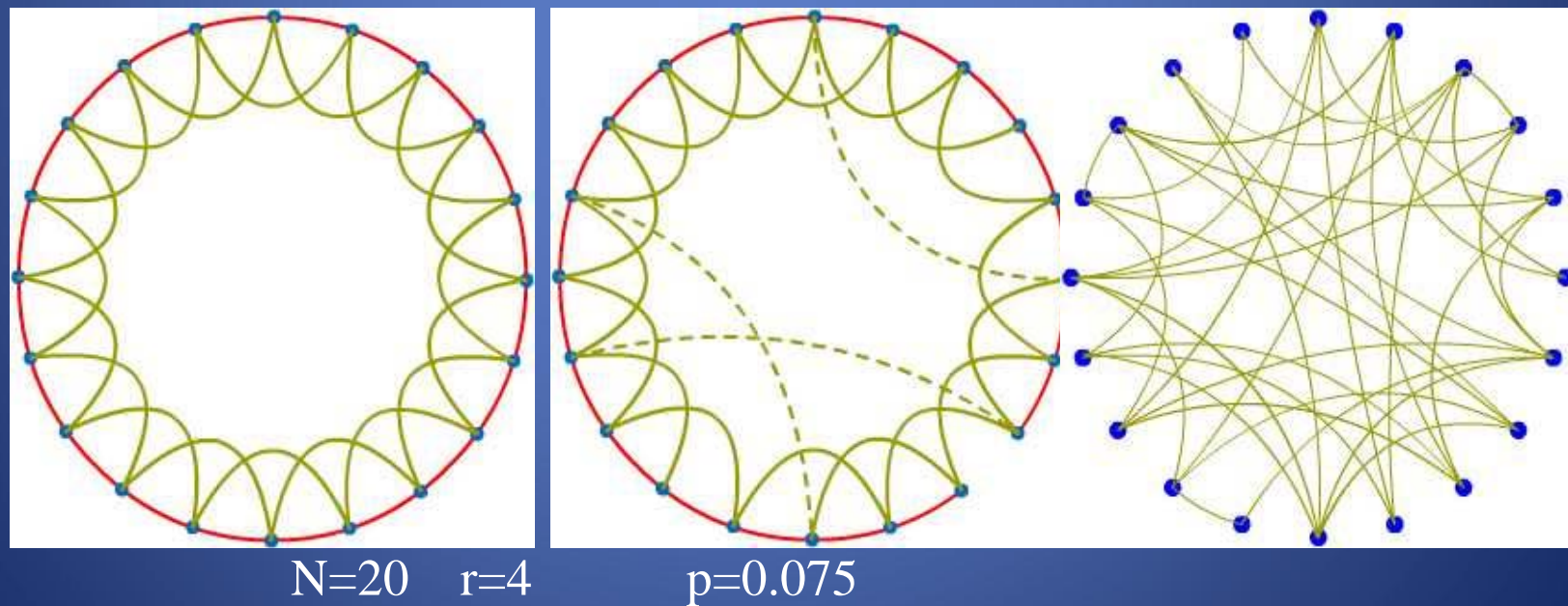
连接近邻:确定性, 连接远程:随机性



# WS小世界网络模型

小世界网络指聚类系数较大且平均路径长度较小的网络 Watts-Strogatz(W-S) 模型

W-S模型的构建： (1)从一个有 $N$ 个节点的一维规则圆环开始，每一个点与它的 $r$ 个最近邻居连结。(2)以概率 $P$ 按顺时针方向对规则圆环中每一条边进行重新连结。重连生成 $k$ 条边。在重连过程中，不能进行自我连结和重复连结。



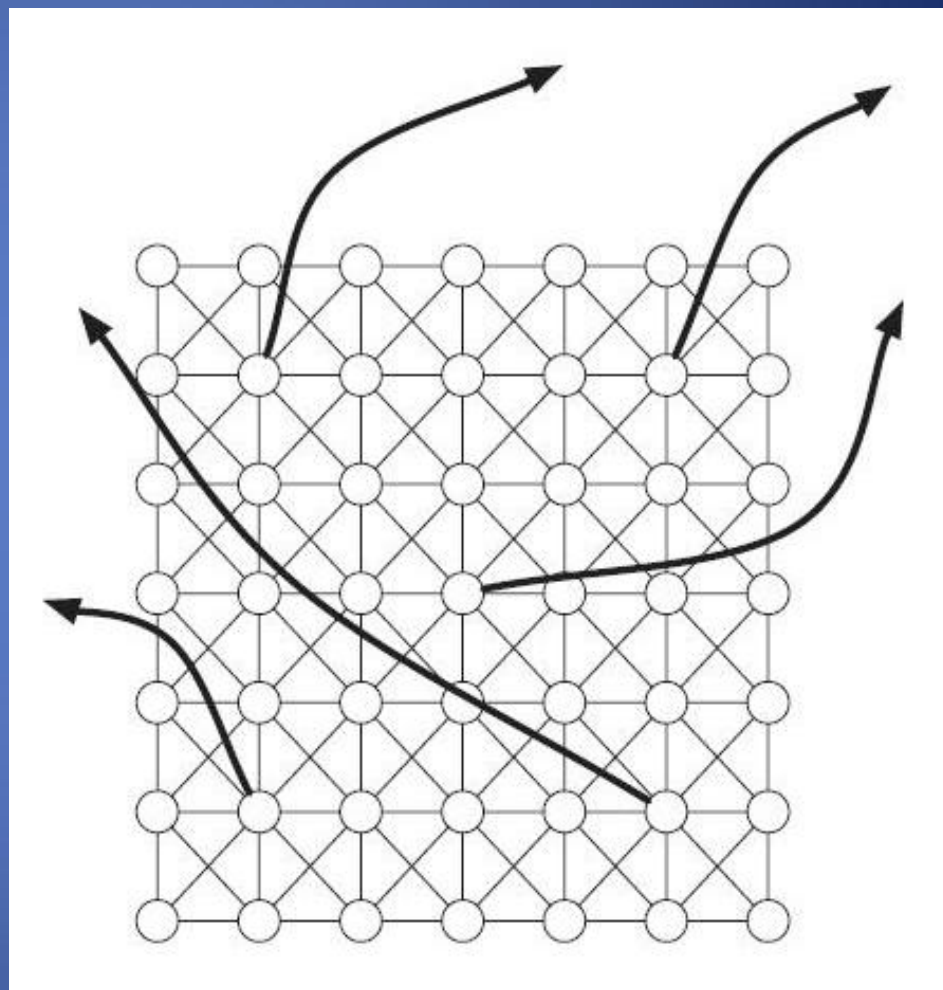


# Watts-Strogatz模型 ( $r, k$ )

- 体现了同质连接和弱关系连接的概念（**利用 $r$ 和 $k$ 控制相关的强度**），于是可以看成是现实社会网络的一个合理近似
- Watts-Strogatz证明了：在这样的网络中，任意两点之间存在短路径的概率很高
  - 基本思路：弱连接的随机性使其形成闭包的概率很小
- Ballobas和Chung进一步从数学上确定了典型路径的长度

# 古代的社会网络是不是也是小世界？

- 那时许多人实际上没有“远程联系”
- 回答是肯定的
  - 设 $k \times k$ 的子网格中只有 $k$ 个远程随机连接
  - 将这子网格看成是一个节点就回到了最初的Watts-Strogatz模型形式
  - 所谓“短路径”此时只不过增加了在子网格中的长度 ( $2k$ )



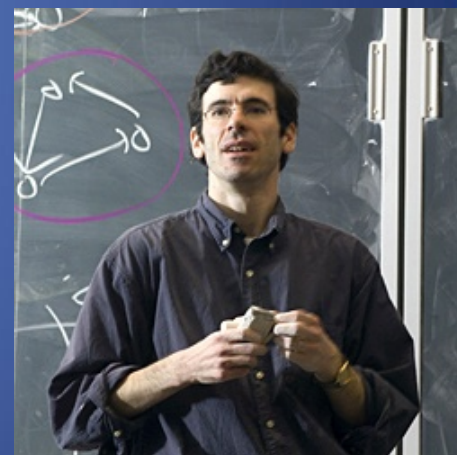
亦即，只要在基础网络结构上存在很少的全局随机性，就可以看到小世界现象

# 分散搜索（decentralized search）

- 如果当年Milgram的实验经费很充足
  - 让每个参与者可以将信件复印任意多份，寄给所有能直呼其名的朋友（flooding, BFS）
- 就不会给我们带来第二个惊奇了：那些信怎么就能真的到达目的地？
  - 虽然人们是有意识地希望信能送到，但她若不认识目标人（他），就只能“估计”她的哪个朋友可能“更接近于”认识他。
  - 因此，我们没有理由认为信件会以很高的概率送到，更没有理由认为它会走一条短路径

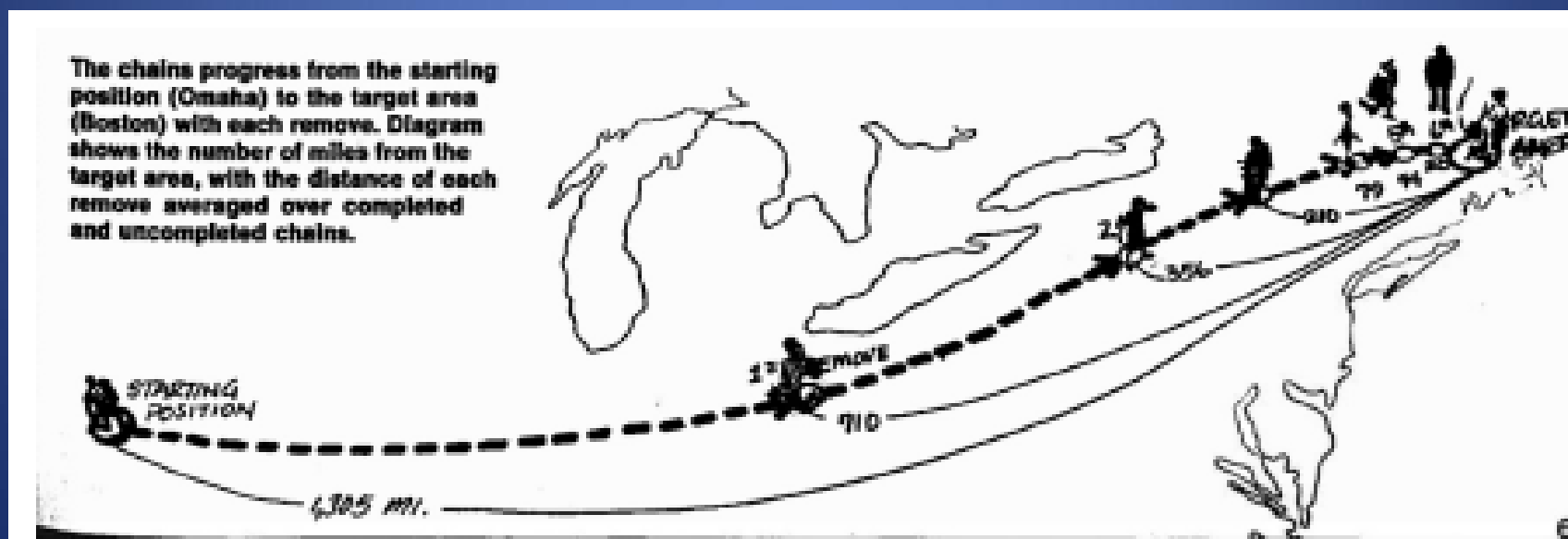
# 探索一种社会网络模型

- 既反映任何**节点对**之间短路径的存在性，也反映这种转发方式达到目的的可能性
- 可以证明，**Watts-Strogatz模型**不能很好地体现第二个要求（搜索路径太长）
- 重新构造一个模型？
- **扩展Watts-Strogatz模型**？
  - Jon Kleinberg, Navigation in a small world. Nature, 2000



# 思路

- 为什么Watts-Strogatz模型不奏效？
  - 体现弱关系的边太过于随机，不支持现实中人们会有意识向比较接近目标的朋友转发信件的行为的表达



每经过一步就向目标更接近

# 扩展的Watts-Strogatz网络模型

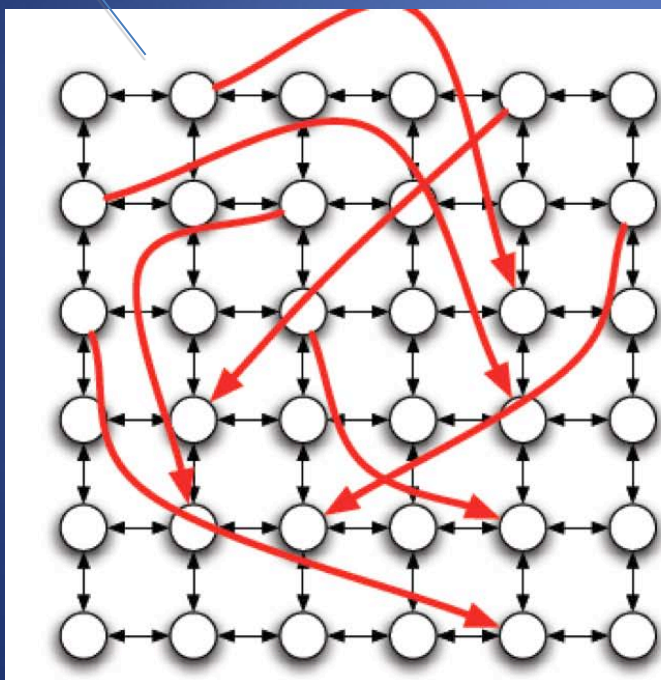
- 节点依然在网格点上排布
- 同质性局部连接相同：每个节点与在 $r$ 个网格步内的节点都有直接的边
- 对随机远程弱连接加以一定的控制：让两个节点之间连接的概率与它们的网格距离的幂次成反比关系（即随距离增加而衰减）
  - 记  $d(v,w)$  为 $v$ 到 $w$ 的距离（网格步数），则产生一条从 $v$ 到 $w$ 的随机边的概率与  $d(v,w)^{-q}$  成正比。
- 理论上可以证明，对于适当的 $q$ ，在如此形成的网络中的分散搜索有很高的效率（平均步数）
- “适当的 $q$ ”是多少？



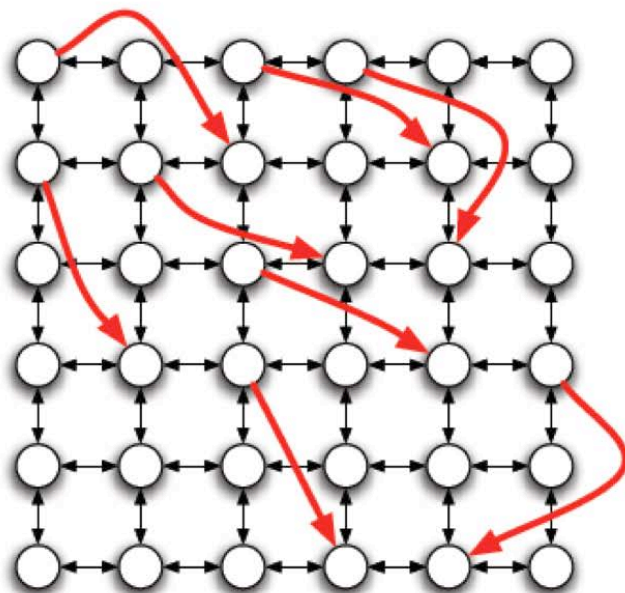
# 不同q值对随机连接长度的影响

q值较小，随机边倾向于较远

q值较大，随机边倾向于较近



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

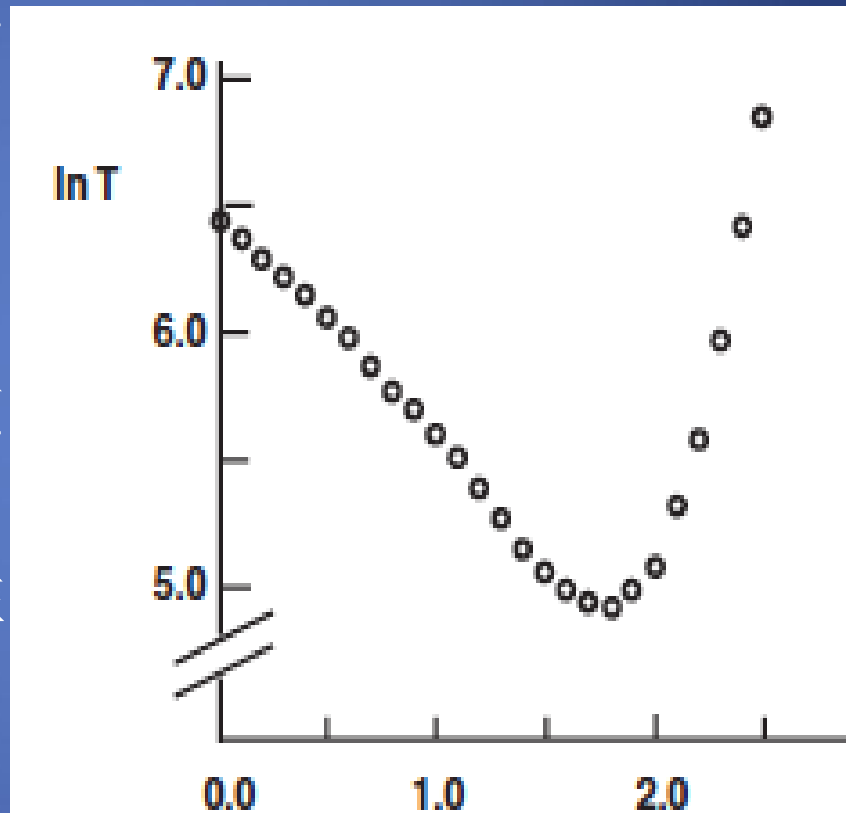


- Watts-Strogatz模型对应于q=0



# 该模型的最佳工作参数

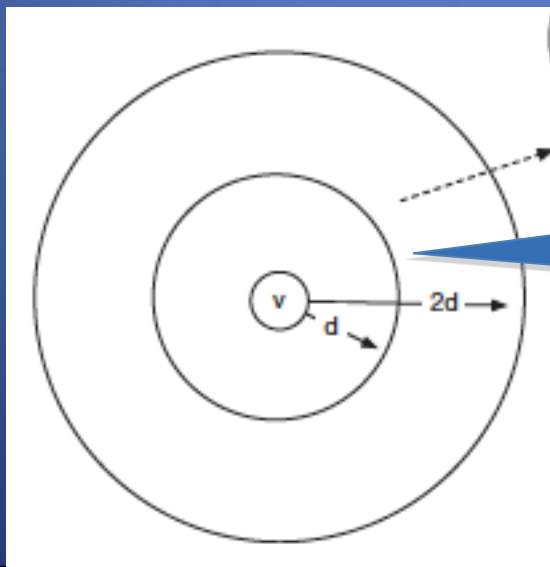
- 理论结果：当 $q=2$ 时，分散搜索达到最佳效果
- 仿真实验：由几亿个节点组成的网络中，针对不同的 $q$ 值而实施的分散搜索
  - 对于这种规模的网络，在指数 $q$ 介于1.5和2.0之间时搜索效果最佳
  - 随着网络规模的扩大，最佳的性能发生在指数 $q$ 越来越接近2



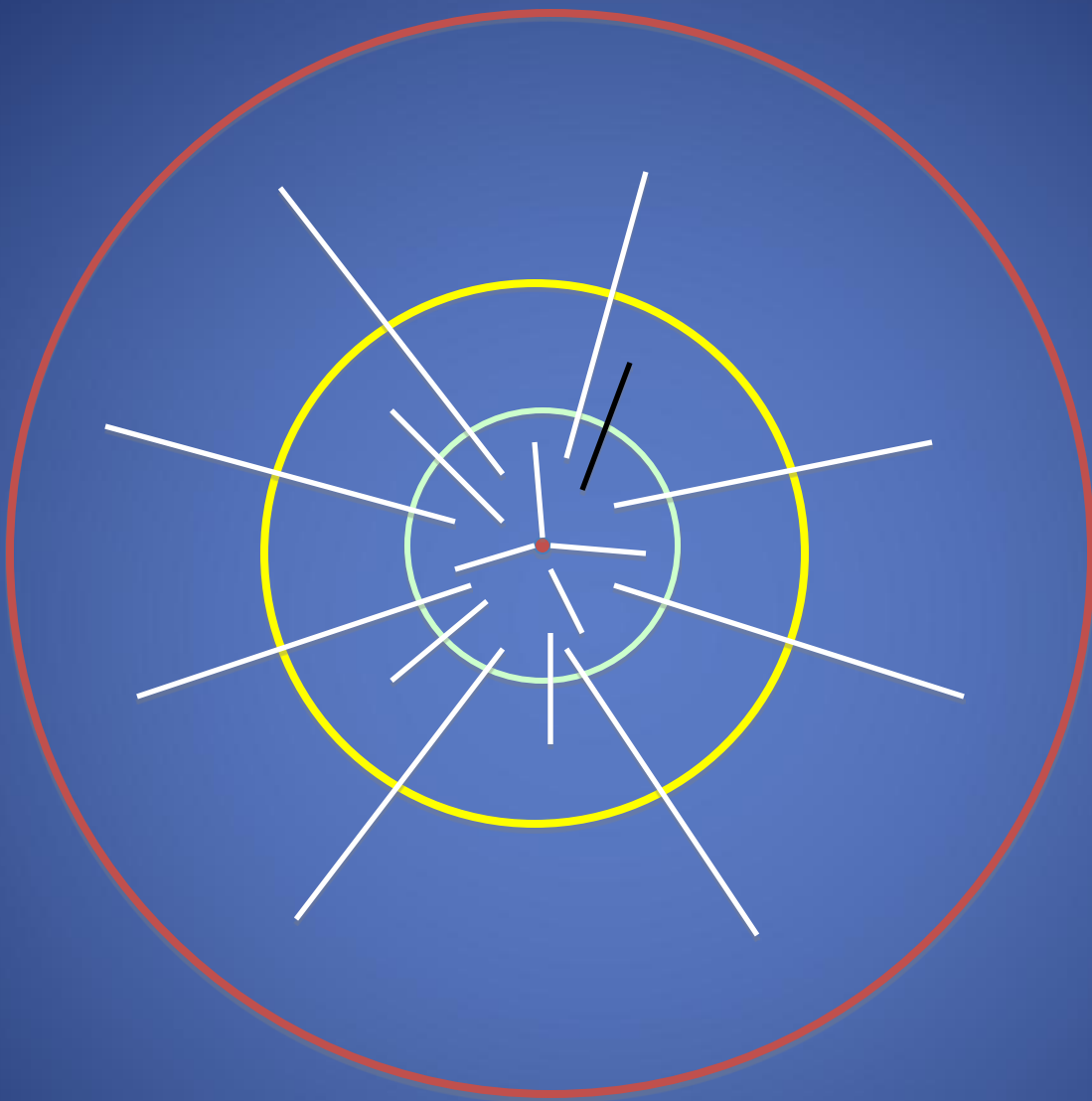
横轴为参数  $q$ ，纵轴为从一个节点到达另一个节点所需的平均时间（跳步）

# $d^{-q}$ : 为什么 $q=2$ 时效果最佳?

- 考虑节点  $v$  随机连接到距离在  $d$  到  $2d$  环内的概率
  - 节点数与  $d^2$  成正比 (均匀分布, 节点数与面积成比例)
  - 于是随机连接到其中一个节点的概率与  $d^2$  成比例
- 因此从  $v$  发出的随机连接落到该区域的概率与  $d$  无关
  - 意味着 (1) 无论一个转发节点与目标节点相距多远, 他都很可能有一个离目标距离近一半的朋友; (2) 信件离目标越近, 与目标有边的可能性就越大



节点数与  $d^2$  成正比,  
连接到每个节点的概率  
与  $d^2$  成正比



# 模型的验证与推广

- 大规模在线社会网络是否（如何）体现了这个（Watts-Strogatz-Kleinberg）网络的特点？
  - 如果没有体现，则可能说明这个网络作为社会网络的模型有问题
- 什么是需要体现的特点？
  - 从地理上看，一个节点在任一距离上的朋友数量在同等距离上网络节点总数中的占比（对应于概率）随距离按反平方递减
- 在线社会网络如何体现地理位置关系？
  - 同一“距离”所“包含”的节点数相同

# LiveJournal (LJ), 1999年建立的社交网站


LiveJournal.com - Start a Free Blog / Journal Today

http://www.livejournal.com/

Google

Apple中国 中国雅虎 维基百科 新闻 热门

LiveJournal.com - Start a Free Blo...



Home Create an account Explore Shop LJ Extras Games




用户名:

密码:

☐ 记住我

登录


[开设一个账户](#)  
[忘记了您的登录?](#)



[English](#) • [Español](#) • [Deutsch](#) • [Русский...](#)

Writer's Block

Forum | Suggest a Question | Subscribe | Unplugged Questions



Frozen delights

What is your favorite ice cream flavor?


Answer

Nov, 29(1/2)

» View Answers


Sign up for LiveJournal now: It's fast, fun, and free!

Spotlight





anythingdisney


Miley Cyrus, Selena Gomez, the Jonas Brothers, and the current crop of Disney Channel stars... Love 'em or hate 'em, you can still get in on all their latest news and gossip!




aramatheydidnt

Take one part  ohnotheydidnt and one part j-pop and what do you get?  aramatheydidnt is the premiere community to fulfill all your japanese pop culture needs.


vaginapagina

 vaginapagina is an online community that offers a supportive, progressive, body- and sex-positive environment in which to discuss issues related to female sexual and reproductive health and wellness. It is a unique empowerment based safe space that is GLBTQIA+ and kink friendly. And boy, there really is no


LiveJournal Shop



Snow Globe  
\$0.99



\$0.99



Colorful Lights  
\$0.99

About LiveJournal


44.3 million  
journals and  
communities

130.2  
thousand  
posts in the last 24 hours

Open Minded and Open Source.  
Volunteer supported by people who love LJ.

Read about our latest features in [News](#).

Creative Expression

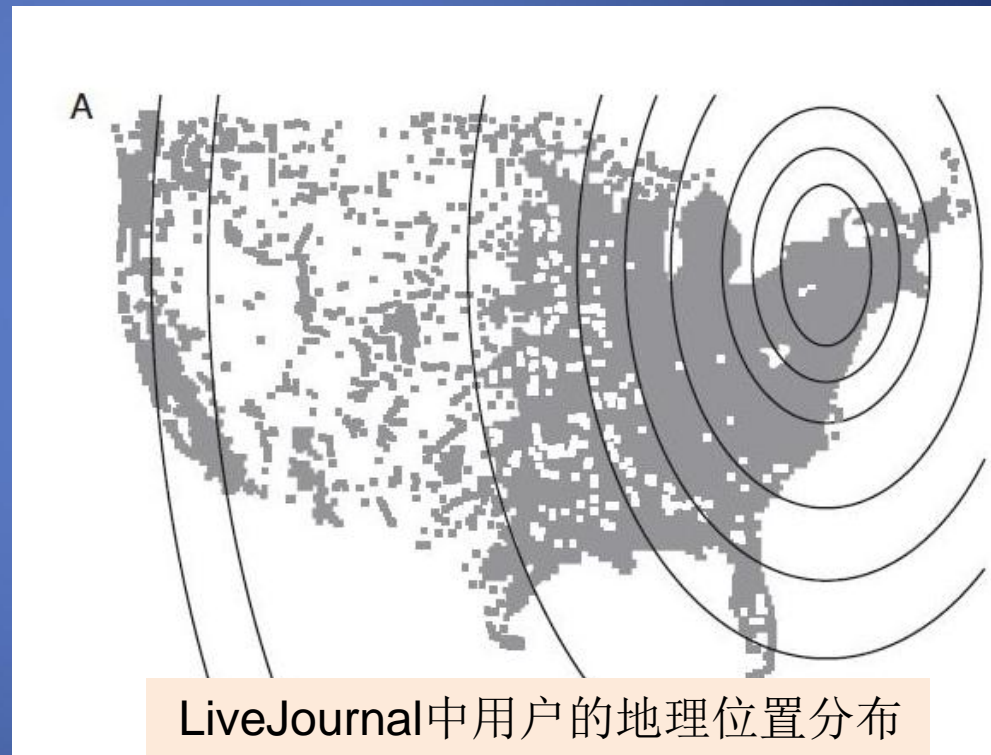


ashleybouvier for  
lj\_photophile

正在载入 "http://www.livejournal.com/", 已完成 89 项 (共 93 项, 1 个错误)

# 来自LiveJournal的实验数据

- 50万用户，含邮政编码信息（地理信息）
- 但他们是不均匀分布的，不符合模型的假设，需要做一些“适配性”工作
- 依据地理信息定义一个节点相对于另一个节点的**排名**
  - 节点 $w$ 在节点 $v$ 眼里的排名， $\text{rank}(w)$

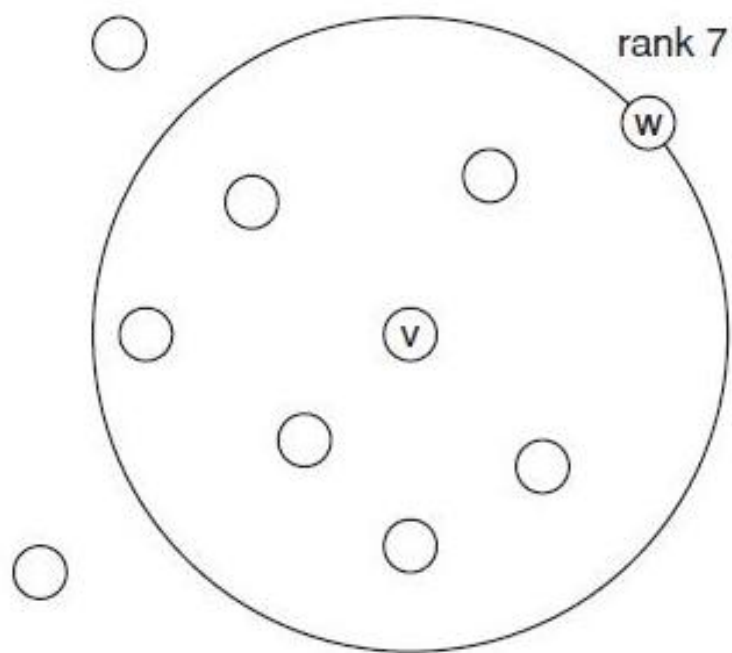


等于网络中**比 $w$ 离 $v$ 近**的节点的个数。

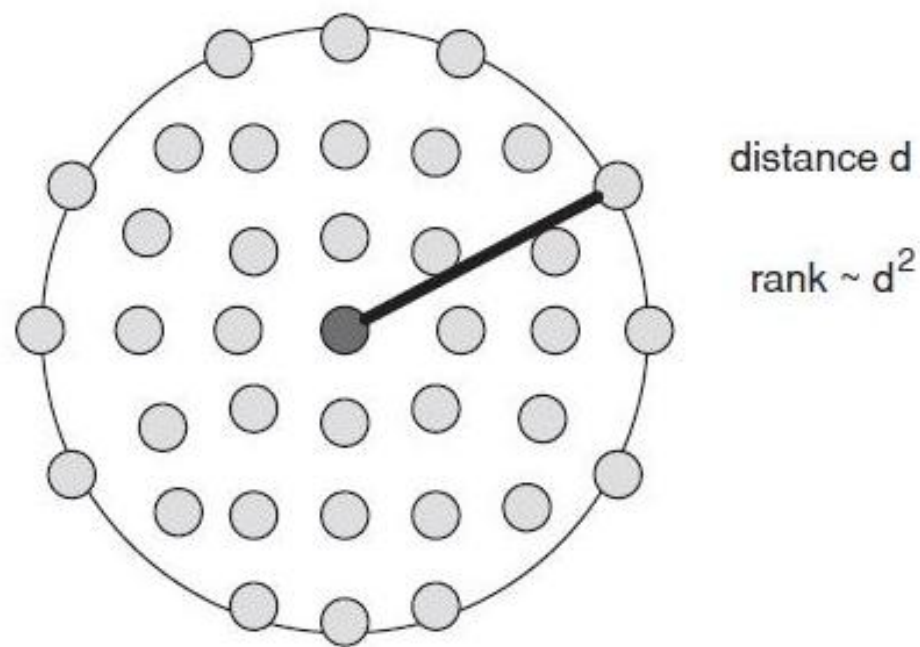
同一排名含同样多节点，  
统一了不同密度的区域



# 社会网络中结合地理位置信息的节点排名



(a)



(b)

- 可以看成是节点在地理上均匀分布时区域范围的一种推广，“排名”与“距离”有对应关系
- 这就使我们能一般性地处理节点分布不均匀的问题了



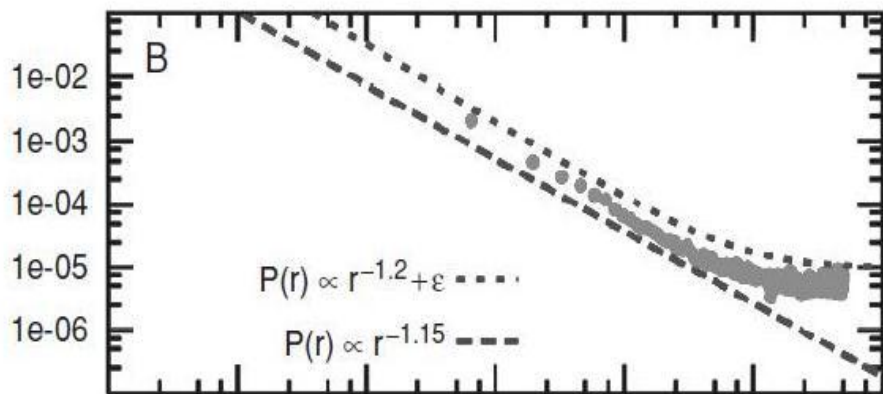
# 这样

- 要验证的是

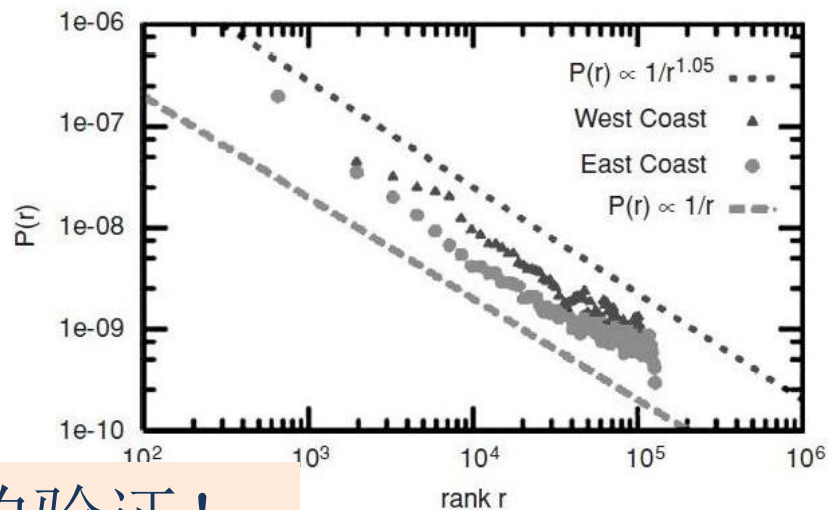
- 从地理上看，一个节点在任一距离上的朋友数量在同等距离节点总数中的占比随距离按反平方递减 ( $1/d^2$ )

- 此时等价于看

- 一个节点在任一排名上的朋友（即有连接）数量在同等排名节点总数中的占比随排名成反比 ( $1/r$ )



(a)



(b)

近乎完美的验证！

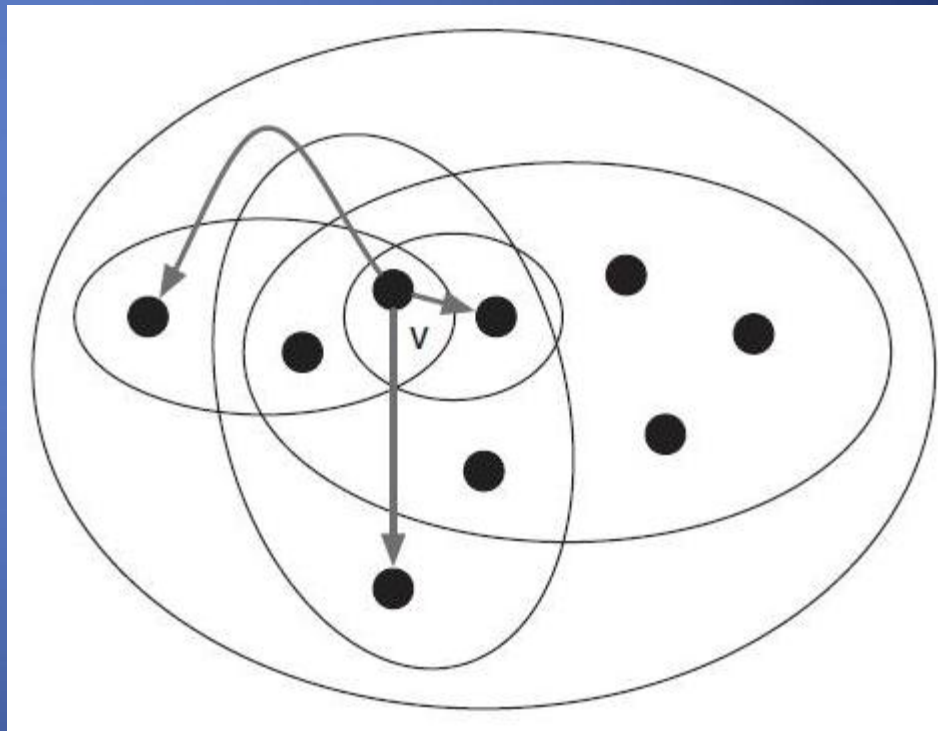
真实社会网络的测量参数与模型最优参数相当吻合！

# 进一步考察网络模型与社会网络的关系

- Watts-Strogatz-Kleinberg网络模型试图反映的社会网络特征
  - 由某种“亲近”（相似性）形成的近距离边
  - 随机的远距离弱联系边
  - 远程相邻节点的数量在同距离节点数中的占比随距离的平方减少
- 距离，是其中一个关键概念
  - 地理空间位置的关系体现最直接的距离概念
  - 节点的相对排名也体现了一种距离概念，克服了处理节点在地理空间上分布不均匀情形的困难
  - 在社会网络中，还有什么有意义的距离概念？

# 社团（social foci, 社交聚点）与距离

- 一个人可能参加多种**社团**（组织机构，兴趣爱好群体等），社团是两人建立关系的一个基础（社团闭包）
- 可以想象，两人的亲近程度（“距离”）与社团的规模有关，越小越近
- 定义“**社会距离**”：两人同属最小社团的规模。



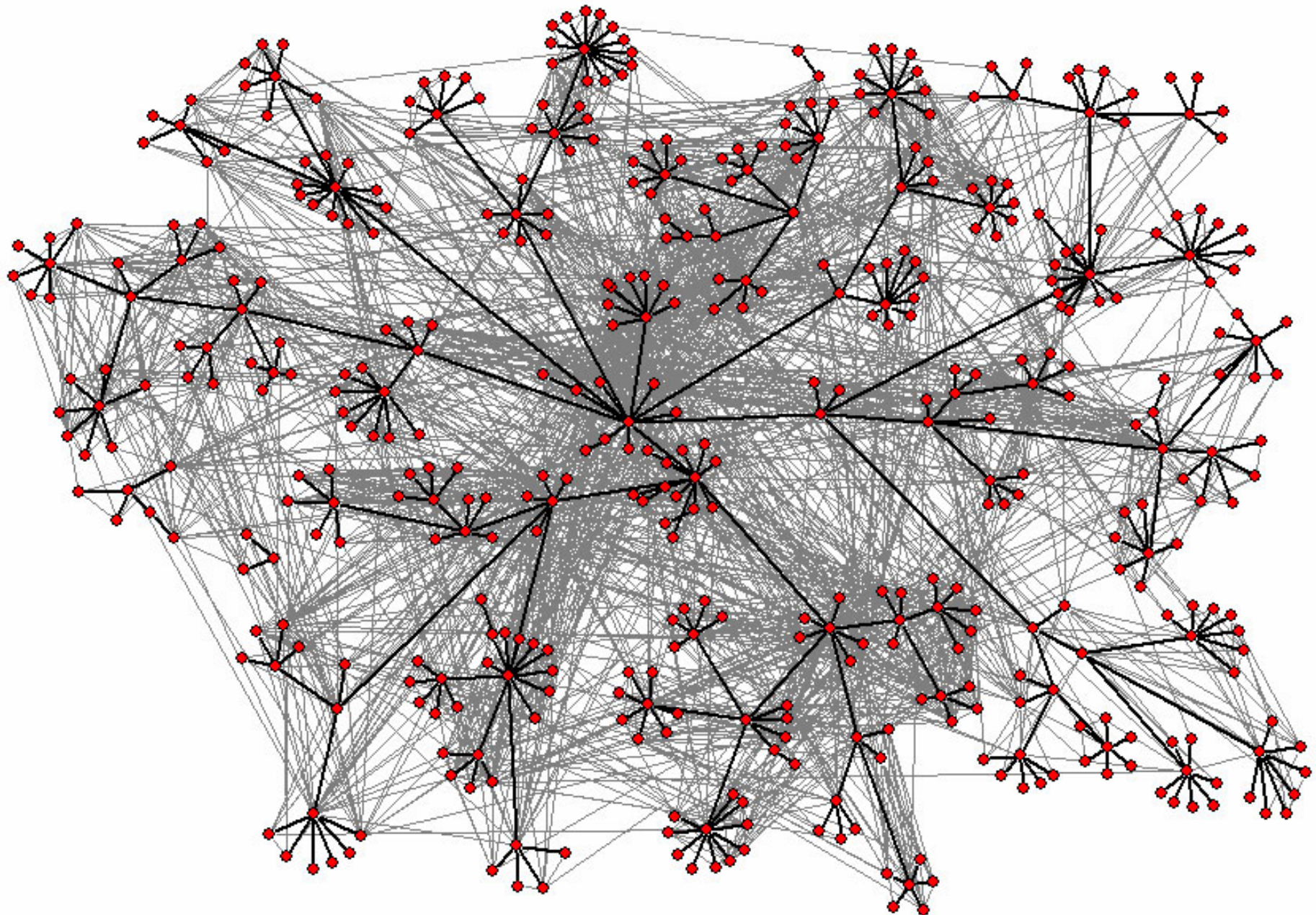
V属于5个大小分别为2, 3, 5, 7, 9的社团

# 进一步

- 如同排名，可以建立（现实社会网络）中的社会距离与Watts-Strogatz-Kleinbeig网络空间距离之间的对应关系
- 从而我们可以来考察以社会距离度量人们之间亲近关系的真实社会网络中是否也体现了Watts-Strogatz-Kleinbeig网络的第三条性质
- Adamic和Adar以惠普实验室人员的email数据为对象进行的分析表明，该性质体现明显：**社会距离为 $d$ 的员工之间产生连接的概率与 $d^{-3/4}$ 成比例**
  - 节点：员工；边：一段时间内交换过邮件
  - 社团：惠普实验室的管理层次结构

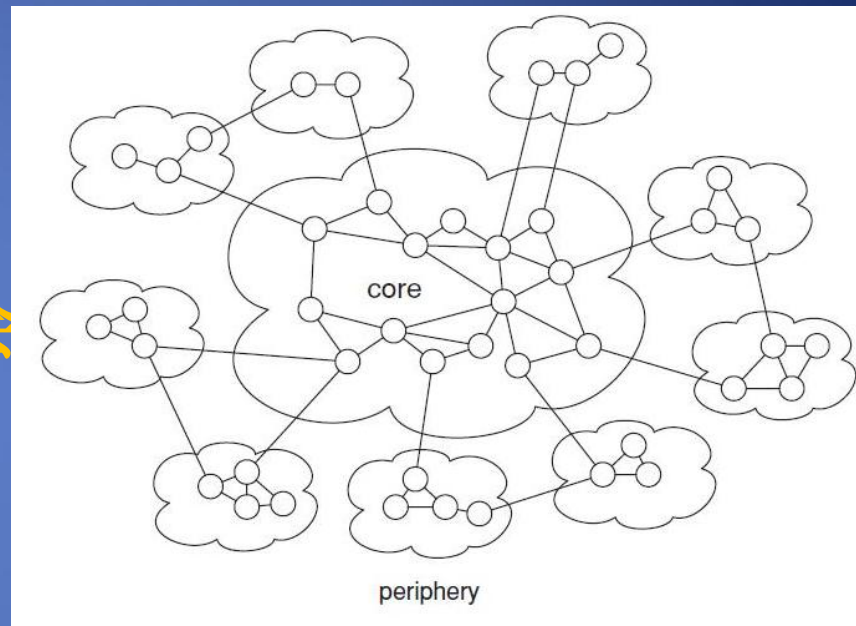


# Email联系与组织机构层次的叠加分析



# 社会网络的“核心—外围”结构

- 大量类似于Milgram的实验表明，找到路径并不容易
  - 人们的参与度是原因之一
- 发现相对容易找到通向社会地位状态高的人的路径
  - 不仅是人们可能更加愿意参与找通向他们的路径
- 还有社会网络的结构性原因！（他们相对容易找到）
  - 同质性：人们大致上按照类似社会地位状态相关联
  - 社会地位高的人倾向于有更多、更广的连接（关系）
- 这就促成了所谓“核心—外围”结构（研究前沿）





# 本章要点

- 围绕小世界问题体现出的研究范式
  - 实验现象 → 理论模型+完善 → 实际验证与推广  
(这可能需要几十年的时间!)
- “图20.10以及随后的相关结果标志着典型科学研究中一系列环节的终结，(1) 实验观察 (Milgram)，(2) 建立基于观察到的现象的数学模型，(3) 基于这个模型进行预测 (最优参数)，(4) 根据实际数据验证这个预测 (模型概念的“适应性处理”)。这就是人们所希望看到的实验、理论、测量相结合的研究方法论的实践。”



# 作业

- 第20章 1,3题