课程群

2025社会计算课...

群号: 1035623172

扫一扫二维码，加入群聊

QQ

社会计算

第一节

社会网络基础

徐童

2025.2.24

教师信息

授课教师：徐童 特任教授

研究方向：数据挖掘与社交媒体分析

个人主页：http://staff.ustc.edu.cn/~tongxu/

课程主页：http://staff.ustc.edu.cn/~tongxu/socomp

课程邮箱：ustcweb2022@163.com

课程QQ群：1035623172

- **我学这门课能收获到什么？**

- 如果你希望了解人工智能/机器学习相关理论知识
  - 建议选修：统计学习、高级人工智能、深度学习、强化学习，etc.
- 如果你希望学习大数据分析的应用技术
  - 建议选修：数据科学导论、机器学习与知识发现、信息检索与数据挖掘，etc.
- 那么，《社会计算》讲什么？
  - 只谈Social ：与社交网络相关的各种主题，我们都略懂略懂
  - 重问题，轻方法：方法论只是工具，我们更乐于讨论问题与背后的逻辑

课程起源

CCF ADL25：《社会网络》讲习班开始报名

中国计算机学会《学科前沿讲习班》

The CCF Advanced Disciplines Lectures

主题　社会网络

2011年10月28-30日北京

社会网络（Social Networks）是指社会中个人之间、组织之间或者个人与组织之间比较持久、稳定的社会关系模式。传统上，针对社会网络的研究是社会学的一个重要分支，基于社会网络的分析方法作为一种关键技术在现代社会学、人类学、经济学，以及生物学等很多不同的领域已有广泛的应用。但，随着Web2.0的出现，社会网络又有了新的"用武之地"。基于社会网络的各类新型互联网应用得到了蓬勃地发展，既有用于个人之间在线通讯的多种应用程序（如即时通讯软件、聊天软件、在线游戏等），又有促进社会交互的各类网络站点（如论坛、社交网站、博客、微博等）。社会网络与互联网的结合，从多种角度改变了每个用户的工作和生活方式，因此针对社会网络的研究迅速引起了计算机领域专家学者的广泛关注。

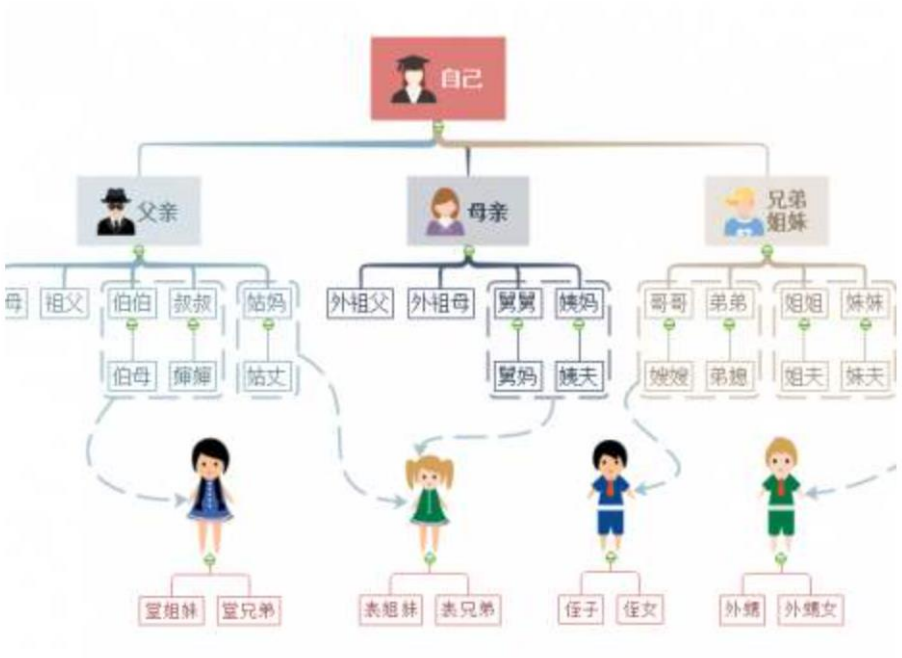
2012年，“社会计算”成为交叉学科（~~学科代码“99J2”~~）

十余年风雨，社会计算早已润物无声

- **无所不在的社会网络**

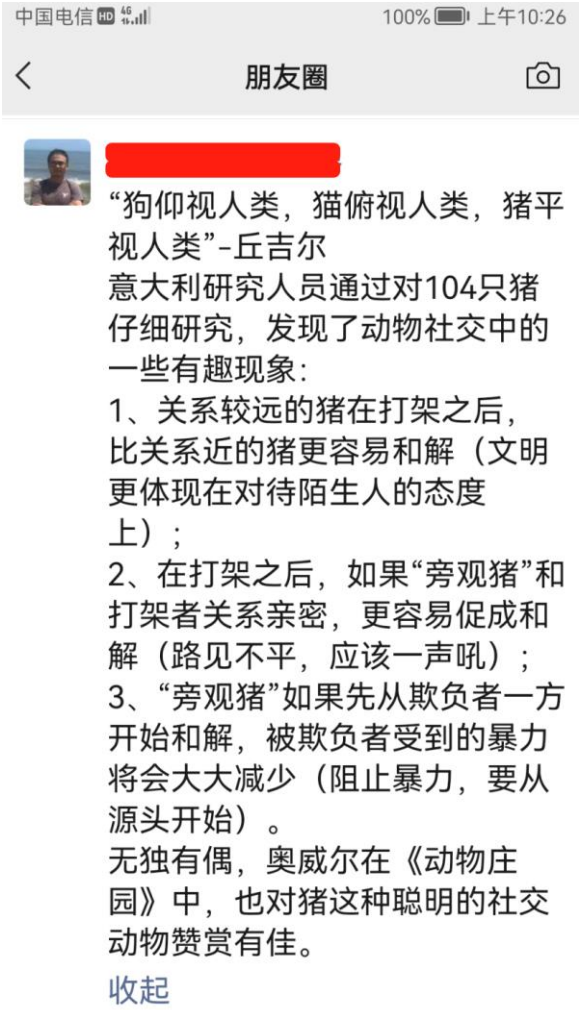
- 我们的格局不是一捆一捆扎清楚的柴，而是好像把一块石头丢在水面上所发生的一圈圈推出去的波纹。每个人都是他社会影响所推出去的圈子的中心。被圈子的波纹所推及的就发生联系。每个人在某一时间某一地点所动用的圈子是不一定相同的。

——费孝通《乡土中国》，1948



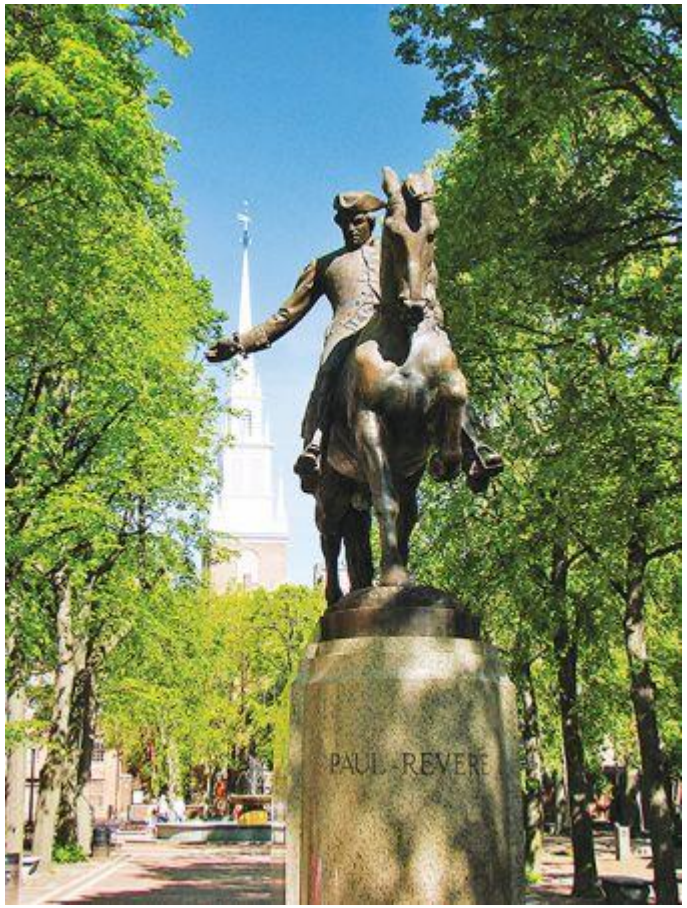
- **无所不在的社会网络**

- 甚至，动物世界中也存在着复杂的网络结构与社交互动行为
  - 族群与生存，例如：狼群、蜜蜂……
  - 复杂的交互行为



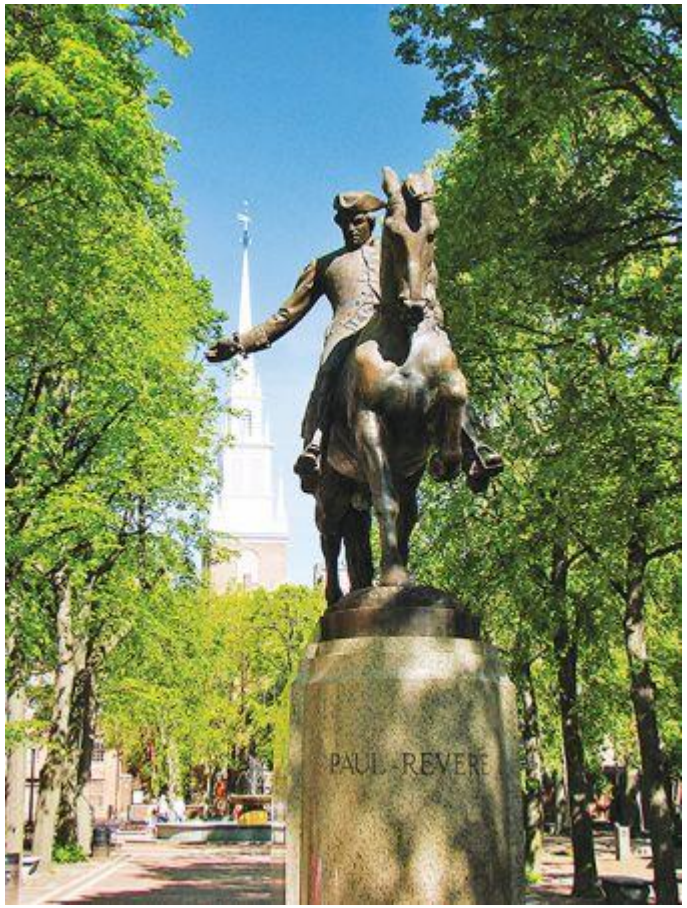
- **社会网络的巨大潜力**

- 过去：1775，列克星敦的枪声
  - 1775年4月18日夜，英军发起了搜查反英组织秘密军火库的军事行动
  - 由于消息走漏，两位勇士——保罗·里维尔与威廉·戴维斯提前向美国民兵告警
  - 然而，两者的结局却不尽相同
    - 前者有效地激发了民兵的有组织抵抗，而后者却未能引起民兵的足够警惕



- **社会网络的巨大潜力**

- 为什么两者的相似行动，却导致了截然不同的结果？
  - 起初，人们认为戴维斯途径的城镇有“亲英倾向”，但事实证明并非如此
  - 后来人们发现，两者在信息扩散渠道上存在不同
    - 里维尔“协助组建了一个对英军的情报与警报系统”“每来到一个城镇，都确切知道应该敲谁家的门，当地民兵组织的领导人是谁，谁是该镇的首要人物”……



- **社会网络的巨大潜力**

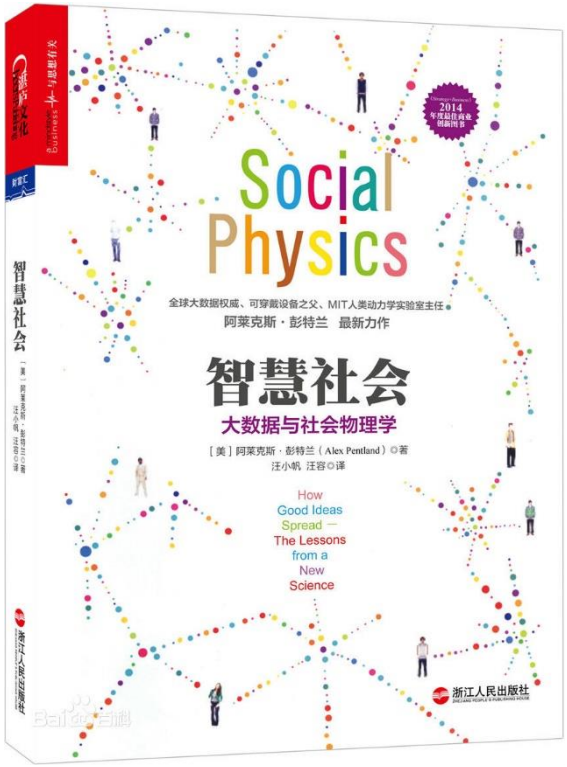
- 如今：2009年，“**寻找红气球**”挑战赛
  - 美国国防部高级研究计划署（DARPA）为了纪念互联网诞生40周年所举办。
  - 他们在全美各地布设了10个红气球，能用最短时间找到全部气球坐标的个人或组织，将获得4万美元的高额奖金。
  - 美国国家地理空间情报局（NGA）的一位高级分析员将之称为“传统的情报收集方法无法解决”的难题。

1969年，ARPANet诞生

2009年，DARPA的红气球

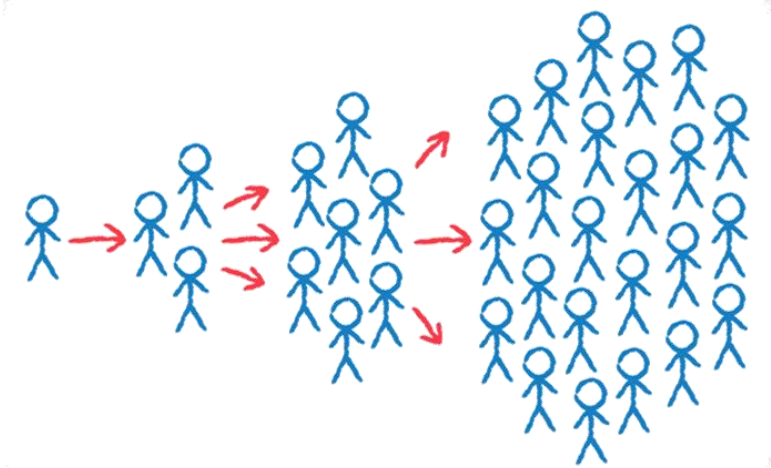
- **社会网络的巨大潜力**

- 然而，难题的快速解决出乎意料
  - 最后，来自MIT的团队仅用8小时52分41秒，就将10个红气球的坐标全部标示完毕。
    - 团队的领导阿莱克斯·彭特兰（Alex Pentland）是全球知名的计算机科学家，《智慧社会》一书的作者
  - 据说，该团队所动员的总人数大约为两百万！
    - 如何实现如此快速而大规模的动员？



- **社会网络的巨大潜力**

- 成功的秘诀：社会网络 + 巧妙的激励机制
  - 据称，彭特兰在短短数小时内便动态组建了一支成员多达5000人的团队，这5000名队员中的每个人又平均通知了400名朋友。
  - 如何实现动员激励？
    - 不仅奖励正确告知气球地点的人，还奖励那些把找到气球的人成功介绍给团队的人。
      - 例如，告知气球地点的人可以获得2000美元，而把这个人介绍给团队的人则可获得500美元，再前面的一个介绍者可以获得250美元，以此类推。

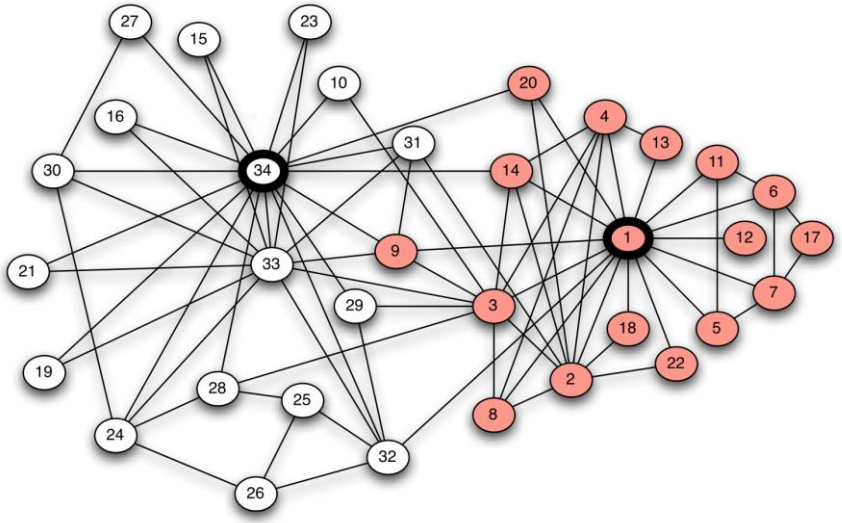


- **社会网络分析的学科交叉特性**

- 随着社会网络的应用日益广泛，海量社会数据催生了社会计算这一新兴学科
- 社会计算是一门集成了信息技术与诸多社会科学之间的交叉学科
  - 一方面，将信息技术、数学、物理学等学科技术运用于社会认知，从而提供更为量化、更为客观的视角以描述社会行为过程。
  - 另一方面，通过运用社会科学如社会学、经济学、心理学等学科理论，指导实现更为合理的量化描述，提升对于人类社会活动的认知水平。

• 社会计算的发展历程

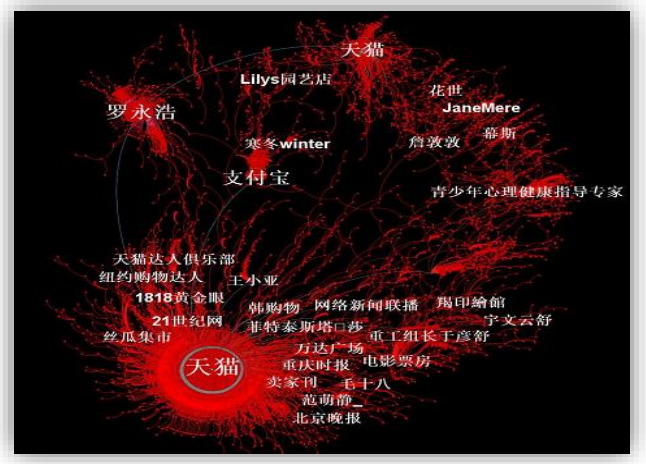
- 随着社会网络的应用日益广泛，海量社会数据催生了社会计算这一新兴学科
  - 现实社会的社会网络往往是隐性的，大规模社会数据难以观测与获取。
  - 早期的社会学学者只能以小规模社区人群为研究对象，依托问卷等基础调查方式，研究人际关系、人群结构、群体行为等问题（更关注孤立的成对目标）

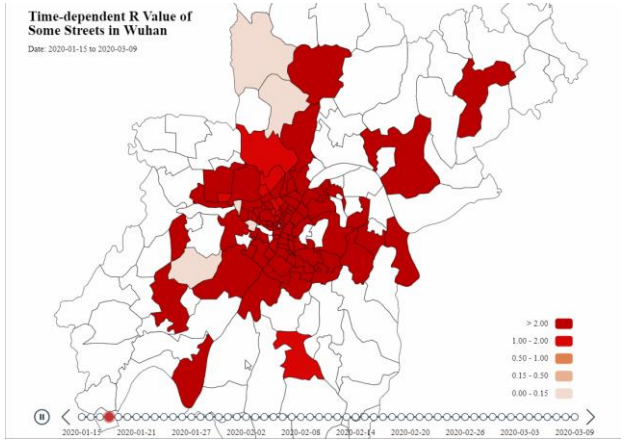


传说中的第一个社会网络图示，20世纪70年代

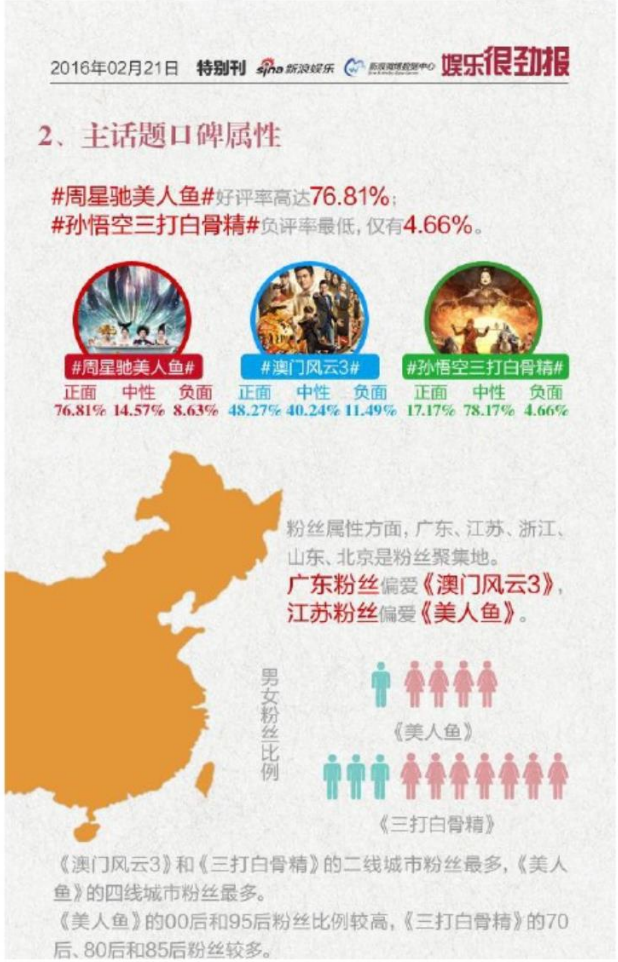
• 社会计算的发展历程

- 近二十年来，随着在线社交平台的涌现和社交元素与商务应用的深度结合，规模庞大的在线虚拟社会逐渐形成，并开始改变人们的在线社交方式
  - 虚拟社会的数据特点：显式（至少部分显式）、可观测、可量化、可追踪
  - 在线虚拟社会（Cyber network）与现实世界（Physical world）的结合成为研究热点





• 社会网络投影于现实问题



• 案例一：口碑营销问题

人们借助社会网络获取信息，也在潜移默化中受到来自社会网络的影响

美人鱼 (2016)

导演: 周星驰
编剧: 周星驰 / 李思臻 / 何妙祺 / 卢正雨 / 冯志强 / 陈庆嘉 / 江玉仪 / 曾瑾昌
主演: 邓超 / 罗志祥 / 张雨绮 / 林允 / 徐克 / 更多...
类型: 喜剧 / 爱情 / 奇幻
制片国家/地区: 中国大陆 / 中国香港
语言: 汉语普通话 / 英语 / 日语
上映日期: 2016-02-08(中国大陆/中国香港)
片长: 93分钟
又名: Mermaid
IMDb链接: tt4701660
官方小站: 《美人鱼》电影专题页面

豆瓣评分
6.7
674655人评价

5星 12.7%
4星 28.9%
3星 42.1%
2星 12.2%
1星 4.1%

好于 43% 喜剧片
好于 44% 奇幻片

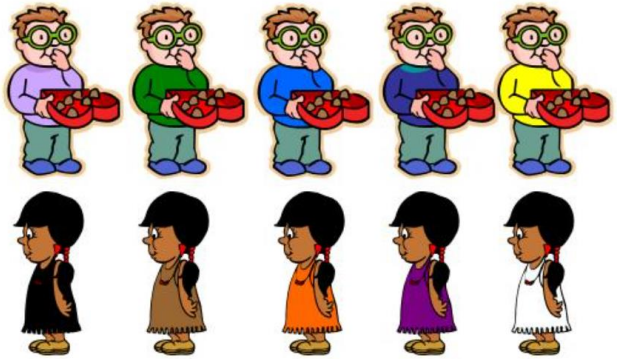
- **社会网络投影于现实问题**

- 案例二：稳定婚姻问题

谈一场稳定的恋爱，也可以借助网络与博弈

作为坚实的后盾

Stable Matching



假如这个世界上，有一个比你女朋友漂亮长的高比她可爱比她懂事的女生说要追你，你会放弃现在女朋友吗？

知乎 · 1,032 个回答 · 1364 关注 ›

沈江 ＋关注

我曾经在知乎上看过一个故事：

一个小伙子是成飞技术人员，小伙子已婚，个子矮，长得丑。有一个比他老婆漂亮长的高比她可爱比她懂事的女生一直追他。

这个小伙子分析：这个女的不可能图我帅，因为我又矮又丑；不可能图我工作稳定，因为我是已婚的，给不了人家稳定；不可能图我有钱，我的收入不够养她这个级别的小三。综上分析，图我的技术机密。

汇报领导.....间谍落网....小伙子个人三等功

- **社会网络投影于现实问题**

- 案例三：群体决策问题

如何化分歧为统一，需要在“集思广益”中寻找共同点

Order！！！！！ 心疼议长1秒钟哈哈哈哈哈哈哈哈

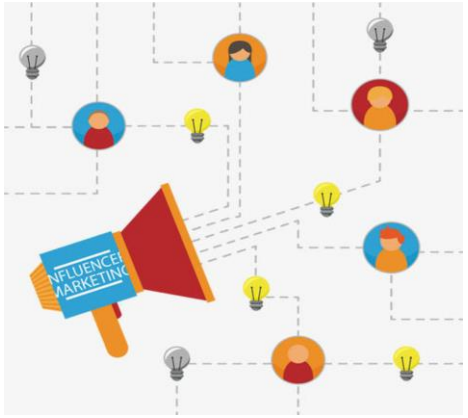
观察者网 · 2019-1-20

台湾“立法院”打架合辑（动物世界版）

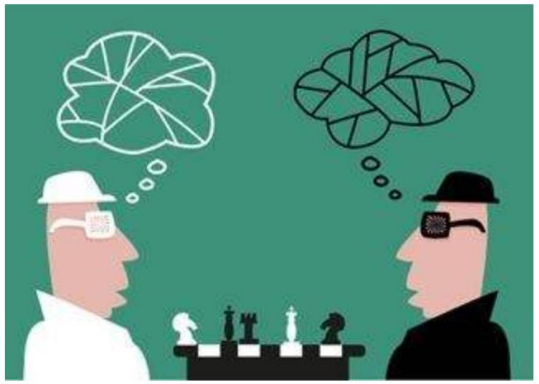
观察者网 · 2019-8-20

- **本门课程所要讨论的问题**

图论与社会网络



网络动力学



社会网络与博弈

社团与群体决策

• 本门课程的结构与安排

图论与社会网络

- 第1节　绪论与网络基础
- 第2节　链接的生成与演化

社会网络与博弈

- 第3节　社会博弈概述
- 第4节　网络中的地位与权力

网络动力学

- 第5节　信息级联现象
- 第6节信息传播与社交推荐

社团与群体决策

- 第7节　社团挖掘 & 群体智慧
- 第8节　双向选择与匹配市场

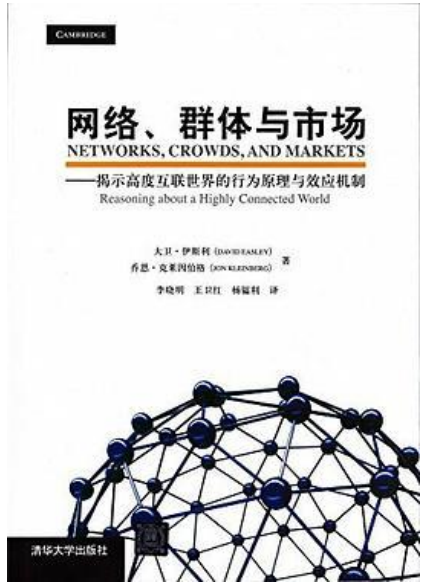
- **课程安排**
- 课程形式：授课+实验
  - 理论学时40学时 + 实验学时20学时
- 成绩组成
  - 30% 实验一 + 40% 实验二 + 30% 课程报告
  - 实验安排：社交博弈、社交游戏设计各一次实验
    - 分组实验，不多于 **8** 人一组，每组提交一份报告/汇报
    - 详细实验要求会在近期于课程群及课程主页公布

报告选题

- **课程报告**

- 围绕“如何让大模型成为社会人”，撰写一篇形式不限的课程报告。
- 报告内容需包含并不限于以下问题：
  - 如何让大模型的社交表现更接近人类
  - 如何让大模型的交互行为更具有策略性
  - 如何让大模型具备领导力和组织协调能力
  - 如何全面评估大模型的社交能力
- 课程报告同样采用分组形式，不多于 8 人一组
- 其他要求：不少于1万字，有真实参考文献，不说车轱辘话，没了

参考书目



网络、群体与市场，

David Esley / Jon Kleinberg 著，

李晓明 等译，清华大学出版社

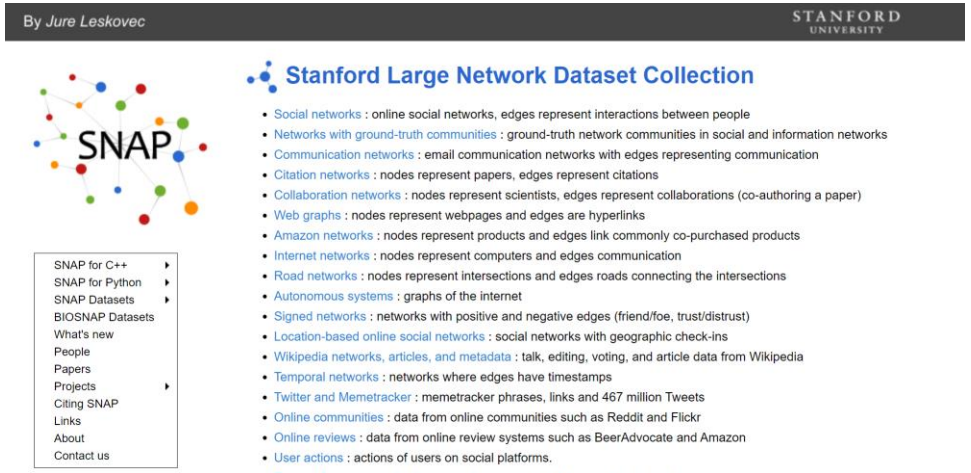
社会计算的基本方法与应用，

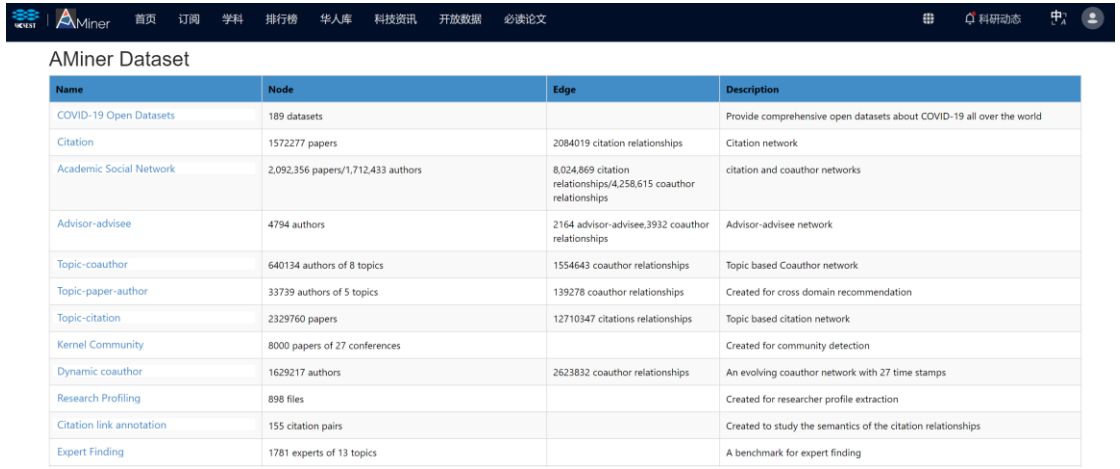
王飞跃 等著，

浙江大学出版社

参考书目

• 常用数据集





Name	Node	Edge	Description
COVID-19 Open Datasets	189 datasets		Provide comprehensive open datasets about COVID-19 all over the world
Citation	1572277 papers	2084019 citation relationships	Citation network
Academic Social Network	2,092,356 papers/1,712,433 authors	8,024,869 citation relationships/4,258,615 coauthor relationships	citation and coauthor networks
Advisor-advisee	4794 authors	2164 advisor-advisee,3932 coauthor relationships	Advisor-advisee network
Topic-coauthor	640134 authors of 8 topics	1554643 coauthor relationships	Topic based Coauthor network
Topic-paper-author	33739 authors of 5 topics	139278 coauthor relationships	Created for cross domain recommendation
Topic-citation	2329760 papers	12710347 citations relationships	Topic based citation network
Kernel Community	8000 papers of 27 conferences		Created for community detection
Dynamic coauthor	1629217 authors	2623832 coauthor relationships	An evolving coauthor network with 27 time stamps
Research Profiling	898 files		Created for researcher profile extraction
Citation link annotation	155 citation pairs		Created to study the semantics of the citation relationships
Expert Finding	1781 experts of 13 topics		A benchmark for expert finding

SNAP开放数据库

斯坦福大学Jure Leskovec教授团队维护

http://snap.stanford.edu/data/index.html

AMiner开放数据库

清华大学唐杰教授课题组维护

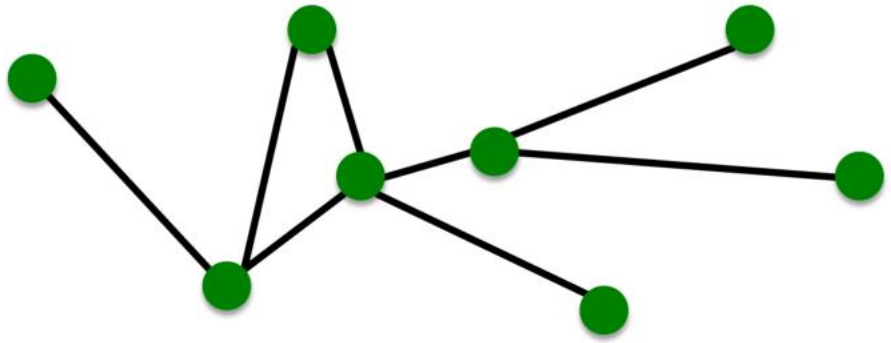
https://www.aminer.cn/data/?nav=openData

本节目录

- **社会网络基本概念**
  - 从节点、关系再到网络
  - 路径、连通性与连通组件
- 强关系与弱关系
  - 三元闭包
  - 弱关系与结构洞
- 小世界现象

- **社会网络的基本元素**

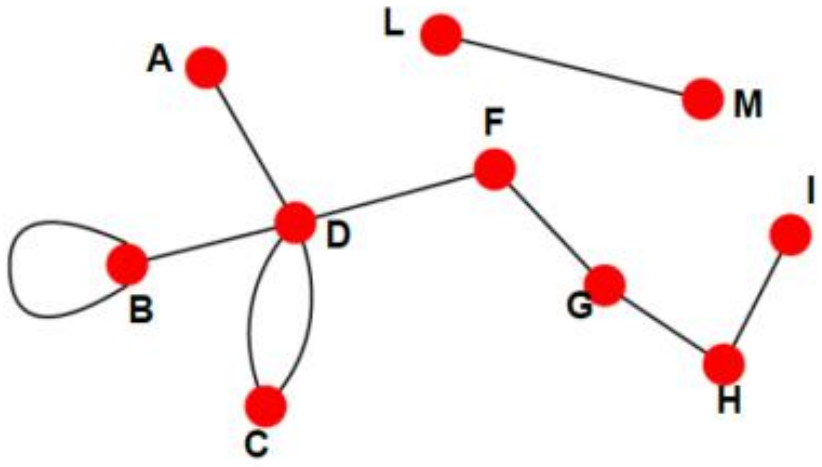
- 从数学抽象的角度，可以将社会网络表示为图（Graph）的结构
  - 节点（Node/Vertex）：用于表示网络中的实体，如社会网络中的人
  - 边（Link/Edge）：用于描述网络中的关系，如人们之间的社交关系



- **社会网络的基本元素：有向/无向边**

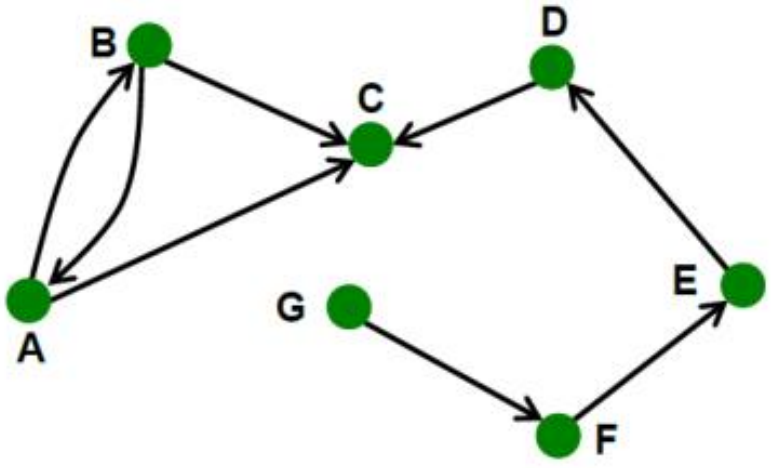
- 网络中的边可能有向，也可能无向，各自表达不同含义

无向边：或双向边、表示对称关系



例如：朋友、合作关系

有向边：表示非对称关系



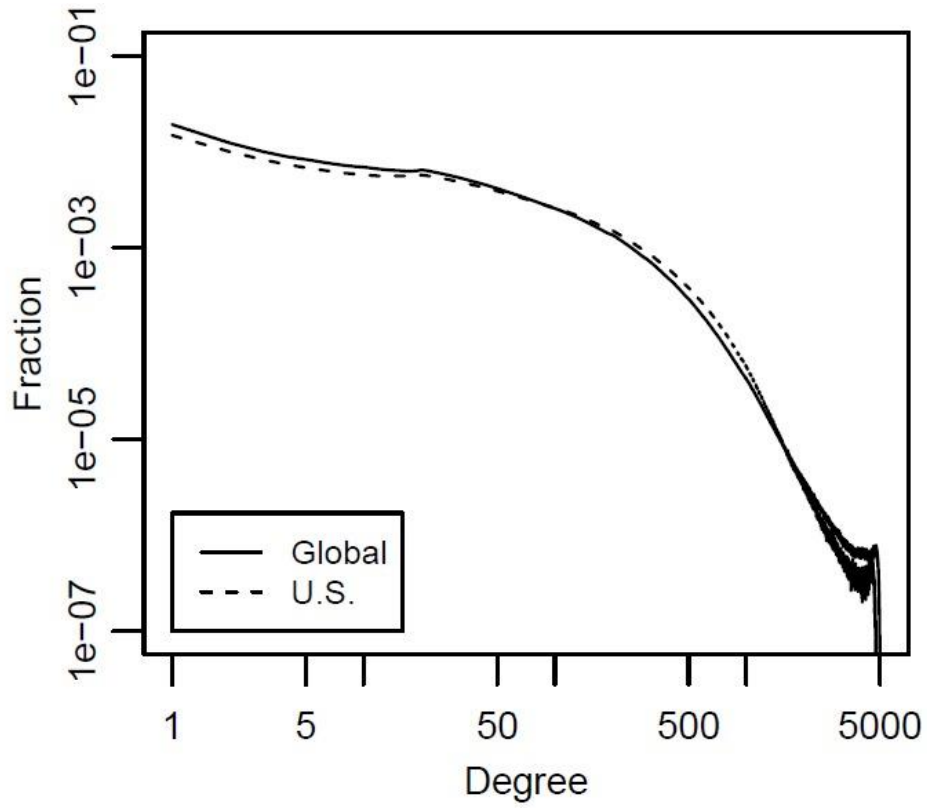
例如：通讯、关注关系

- **社会网络的基本元素：邻居、出入度**

- 对于任意节点而言，与其直接相连的节点被称作“邻居” （Neighbor）
  - 对节点 v 而言，往往采用 N(v) 表示其邻居集合
  - 在有向网络中，入边邻居和出边邻居集合应加以区分
- 同时，节点所连边的数量（或邻居的数量）被称作“度” （Degree）
  - 对节点 v 而言，往往采用 d_v 表示其度数
  - 同样，出度与入度应加以区分
    - 显然，对于一个网络而言，节点出度之和*等于*入度之和*等于*边数

- **社会网络的基本元素：邻居、出入度**

- 一般而言，真实网络中的节点度数往往符合**幂律**（Power-law）分布
  - 少数节点拥有大多数的边
  - 这些少数节点即形成了所谓“影响力节点”，也就是俗称的“大V”

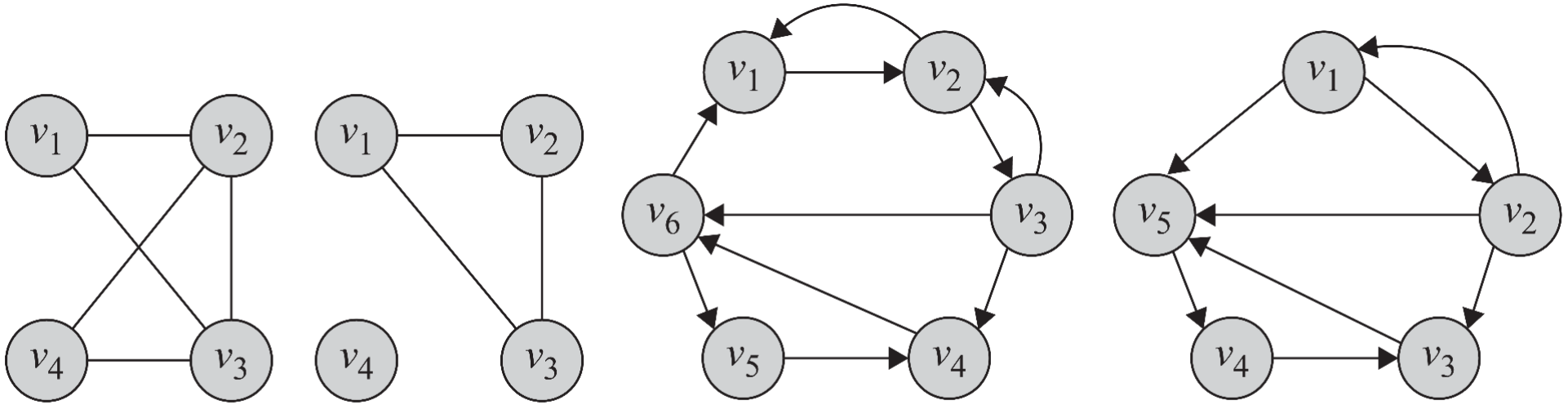


- **社会网络的基本元素：连通性**

- 两个节点是连通的（Connected），当且仅当节点之间存在一条路径（Path）
  - 注意：这并不意味着两个节点之间是直接相连的
- 进而，一个图是连通的，当且仅当图中任意两个节点都是连通的
  - 对于有向图而言，有强 / 弱连通的区别
    - 任意两个节点之间存在双向的连通路径，即为强连通
    - 忽略方向的前提下任意两个节点之间存在一条路径，即为弱连通
  - 如果存在两个节点无法连通，则图是不连通的（Disconnected）

- **社会网络的基本元素：连通性**

- 图的连通性（Connectivity）的实例



(a) Connected (b) Disconnected (c) Strongly connected (d) Weakly connected

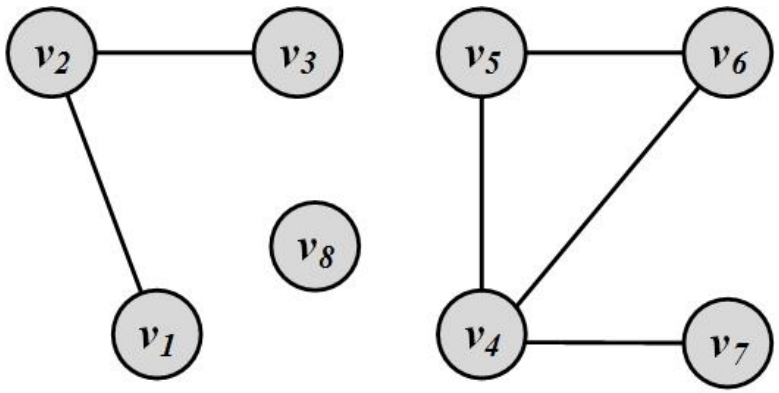
(d)中仅存在单向连通路径

- **社会网络的基本元素：连通组件**

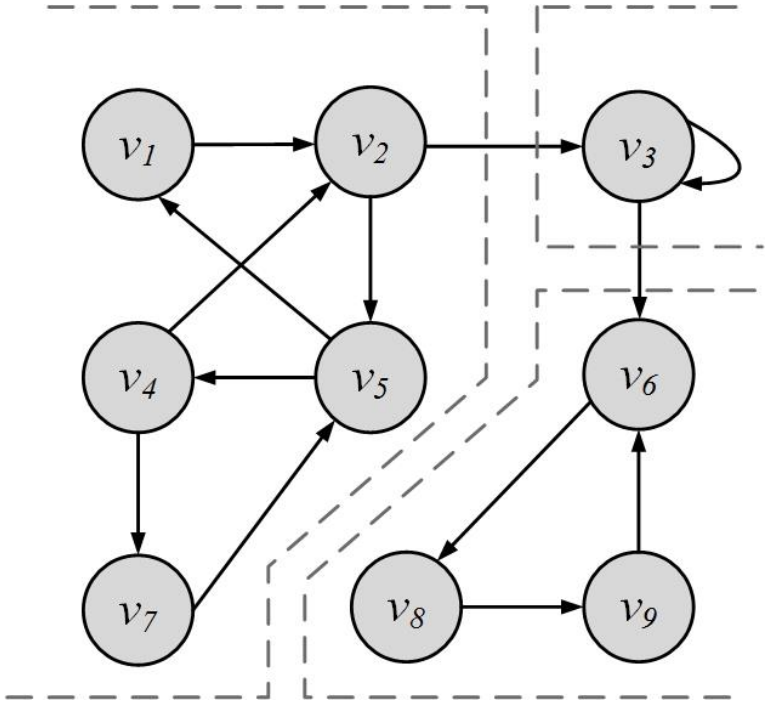
- 对于一个图而言，其中的一个连通组件（Component），即一个连通的子图
  - 换言之，该子图的任意两个节点之间都是连通的
  - 在有向图中，我们将强连通组件定义为该有向图的一个强连通子图
    - 即任意两个节点之间存在双向连通路径
    - 相应的，弱连通子图对应着弱连通路径

- **社会网络的基本元素：连通组件**

- 连通组件（Component）的实例



3 components



3 Strongly-connected components

- **社会网络的基本元素：连通组件**

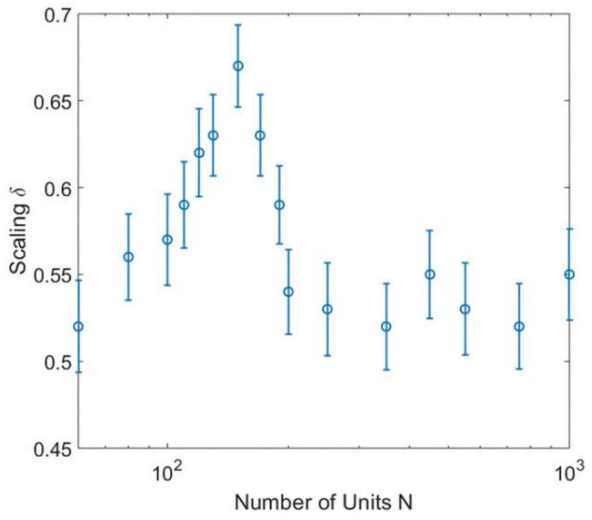
- 有关连通组件的拓展阅读：邓巴数与150原则
  - 150原则：人类智力所允许拥有稳定社交网络的规模大约是150人。
  - 该原则由英国牛津大学的人类学家罗宾·邓巴（Robin Dunbar）在20世纪90年代提出，源于以下观察：
    - 古代：兄弟会（Anabaptist）的不成文规定：每当聚居人数超过150人的规模，他们就把聚居点变成两个，再各自发展
    - 动物社会中也有类似的社群与分裂现象，例如：蜘蛛猿通常形成2只到17只的子群，但每个子群往往只持续一两个小时

- **社会网络的基本元素：连通组件**

- 有关连通组件的拓展阅读：邓巴数与150原则
  - 基于观察，邓巴教授估算出社群规模与灵长类脑容量的大致相关性，如下：

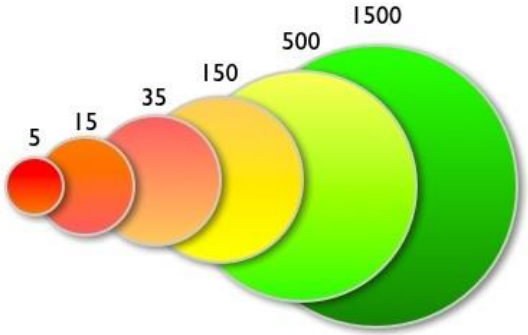
$$\mathrm{Log10(N)} = 0.093 + 3.389\ \mathrm{log10(CR)}$$

  - 得到的结论是，不论是何种文化及语言，是否使用诸如互联网等现代工具，每个人能够维持社交关系是150人左右
    - 即使现代通讯工具极大地便利了人类的通讯方式，但维持关系所需要的通讯时间成本大体上是不变的。想要与更多人进行交流，就会削弱交流的效果。



- **社会网络的基本元素：连通组件**

- 有关连通组件的拓展阅读：邓巴数与150原则
  - 邓巴数的成因：人们需要通过合作来发挥潜能，但过大规模的网络将导致沟通效率的下降，最终导致团队的分裂
    - 这个过程，无疑是连通组件拆分为更多子连通组件的过程
    - 某种意义上说，150可视作一个“维持社交关系”的人数上限
      - 例如，早期手机通讯录与社交软件的好友上限往往即为150
      - 现在，人数/群用户数虽然更为扩大，但仍有上限

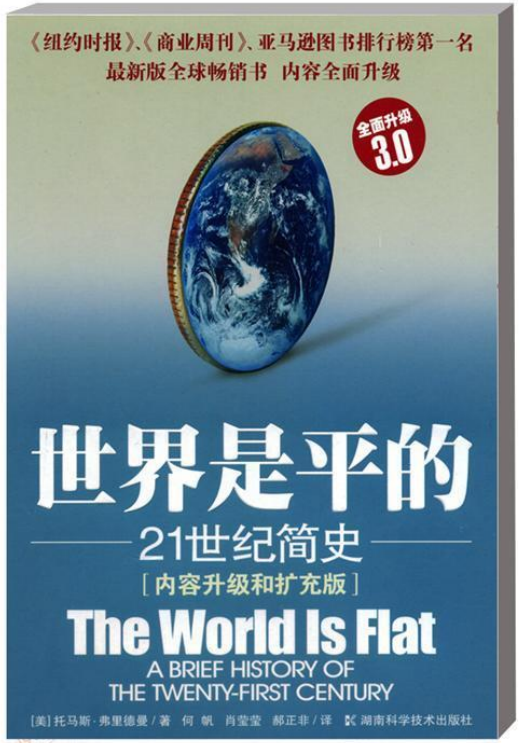


本节目录

- 社会网络基本概念
  - 从节点、关系再到网络
  - 路径、连通性与连通组件
- **强关系与弱关系**
  - 三元闭包
  - 弱关系与结构洞
- 小世界现象

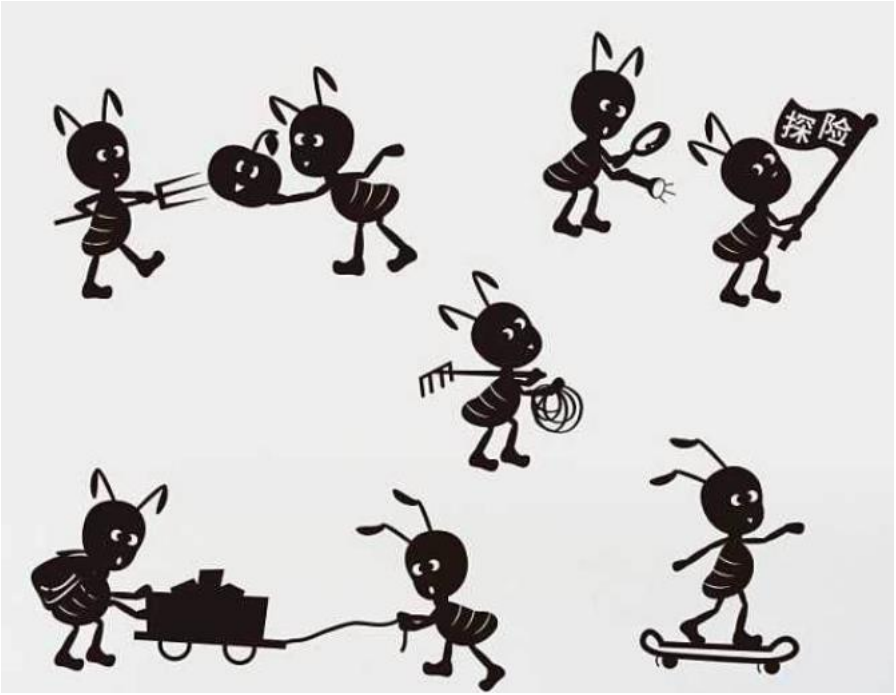
- **节点之间是平等的吗?**

- 托马斯·弗里德曼的知名著作《世界是平的》曾经指出："个人透过全球化进程获得权力"，并指出这一过程与科技发展如网络密切相关
- 然而，这是否意味着网络中的节点是平等的?
  - 回顾开头的例子：保罗·里维尔的成功秘诀
    - "每来到一个城镇，都确切知道……当地民兵组织的领导人是谁，谁是该镇的首要人物"



- **网络中的节点存在着不同**

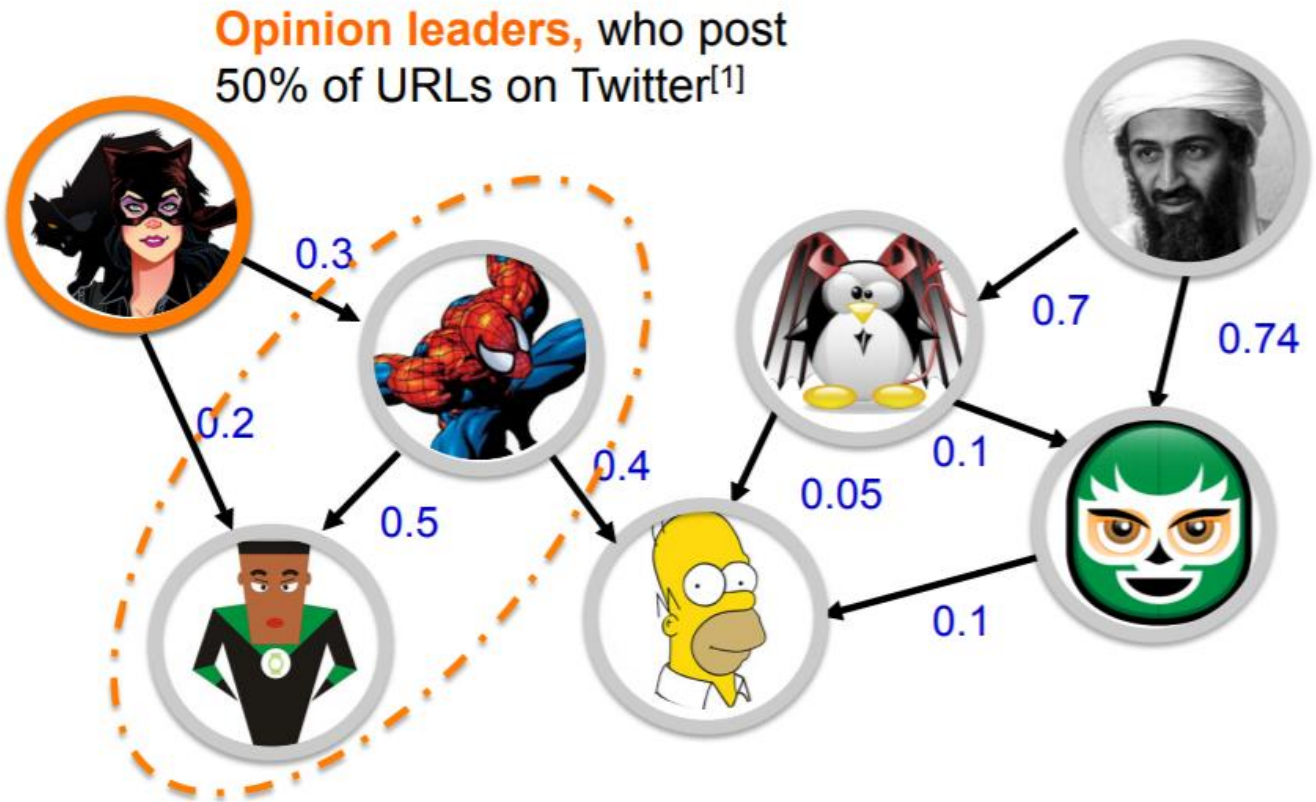
- 另一方面，研究者也发现，在同一个社会网络中，不同节点往往也扮演不同角色
  - 例如，一个部门，往往有以下分工：
    - 部门经理：负责领导部门
    - 技术专家：负责提供技术指导
    - 项目经理：负责外部需求沟通
    - ……



强弱关系

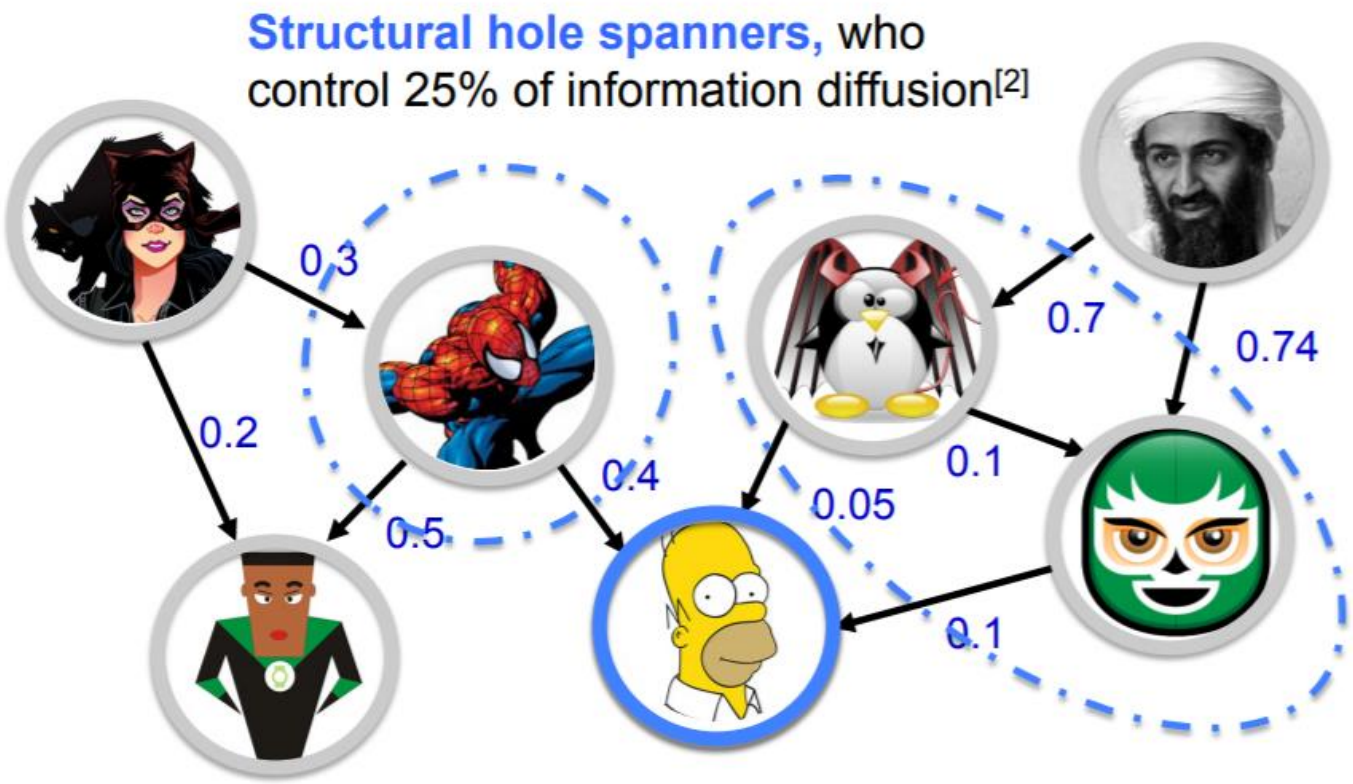
- **网络中的节点存在着不同**

- （1）网络中节点的地位不同（详情将在第4节：网络中的地位与权力介绍）



- **网络中的节点存在着不同**

- （2）网络中节点 / 关系的分工不同



强关系与弱关系

- 1973年，美国社会学家格兰诺维特（Granovetter）提出了强/弱关系理论，对关系功能进行了区分
  - 这里的强弱，并不限于关系的强度（边的权重）
  - 强关系：更偏重现实世界中的熟人关系，有较强的同质性和情感维系
  - 弱关系：类似于所谓“泛泛之交”，关系的异质性较强，联系并不紧密，但可作为信息来源

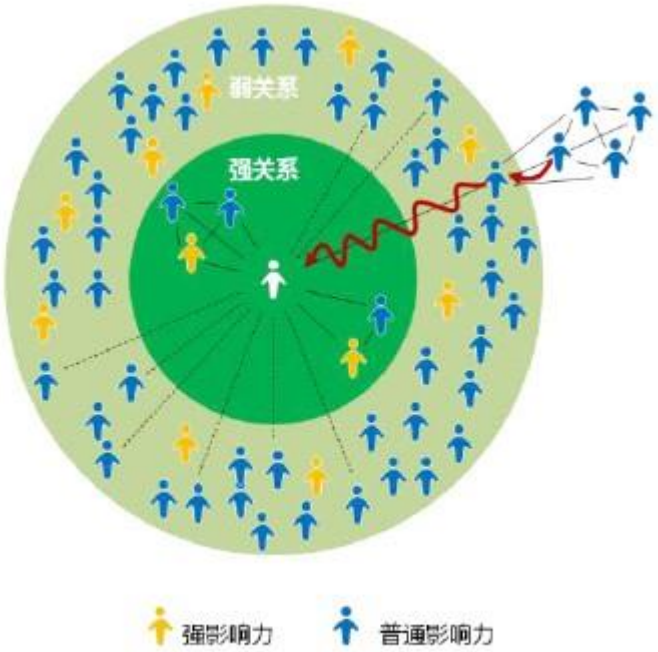


图 1 关系图谱中的强弱关系及其影响力示意图

弱关系的力量

- 格兰诺维特（Granovetter）理论的来源
  - 20世纪60年代末期，格兰诺维特采访了一批最近更换工作的人，询问他们如何找到工作
  - 采访中发现，多数人通过私人介绍找到工作，但：
    - 仅有16.7%的人与求助者经常来往
    - 55.6%的人只不过偶尔与求助者联系
    - 28%的人很少见到帮助自己的人

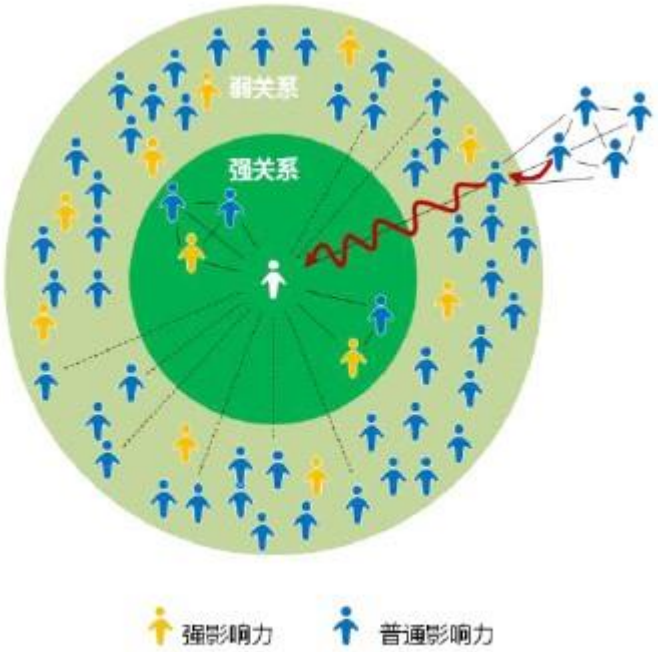
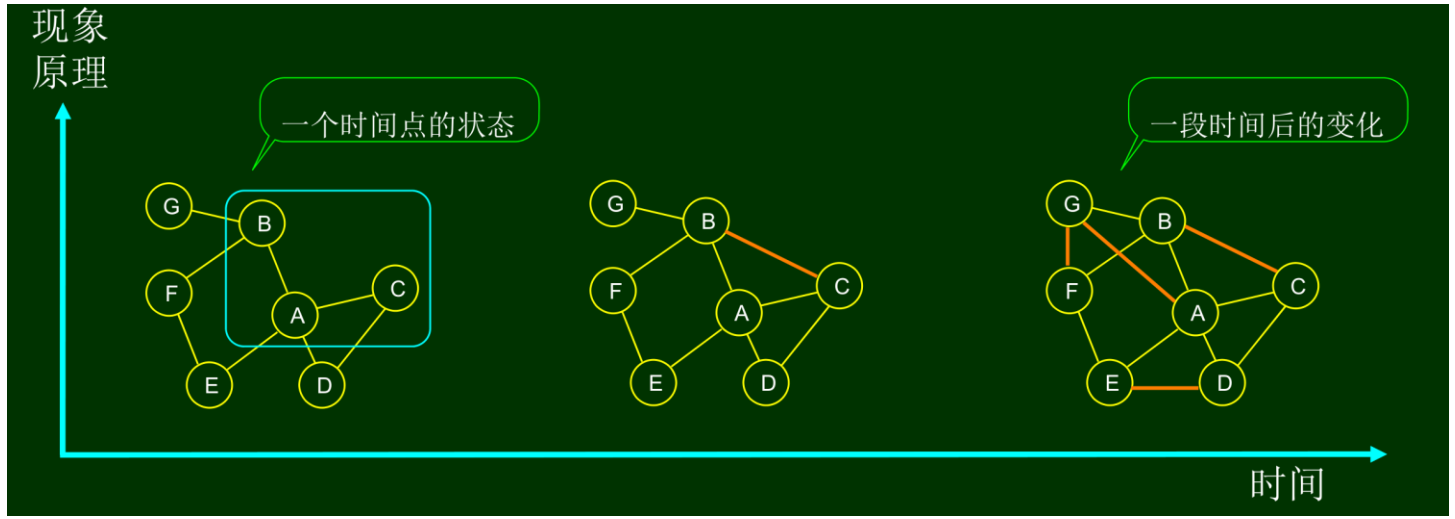


图 1 关系图谱中的强弱关系及其影响力示意图

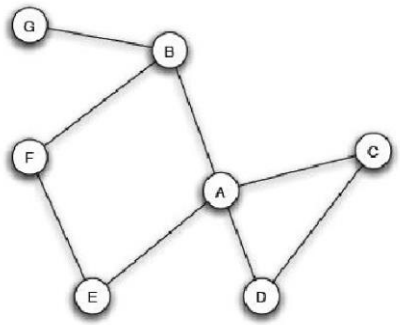
- **弱关系从何而来**

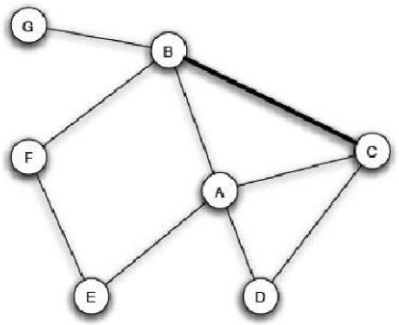
- 为什么“弱关系”发生了实际作用，却并不意味着强有力的联系？
  - 日常观察经常可见的一个现象：在一个社交圈内，如果两个人有共同的朋友，则两个人在未来很可能成为好友
    - 共同好友越多，两个人成为好友的概率越高（将在下节详细介绍）



- **弱关系从何而来**

- 这一现象，催生了所谓“三元闭包”的形成
- 三元闭包的形成原因?
  - 新生成的边可能与已有的边是同质的
  - 但存在另一种可能，新生成的边是在已有的边影响下“被动”形成的
    - 已有边/共同好友的“中介”作用



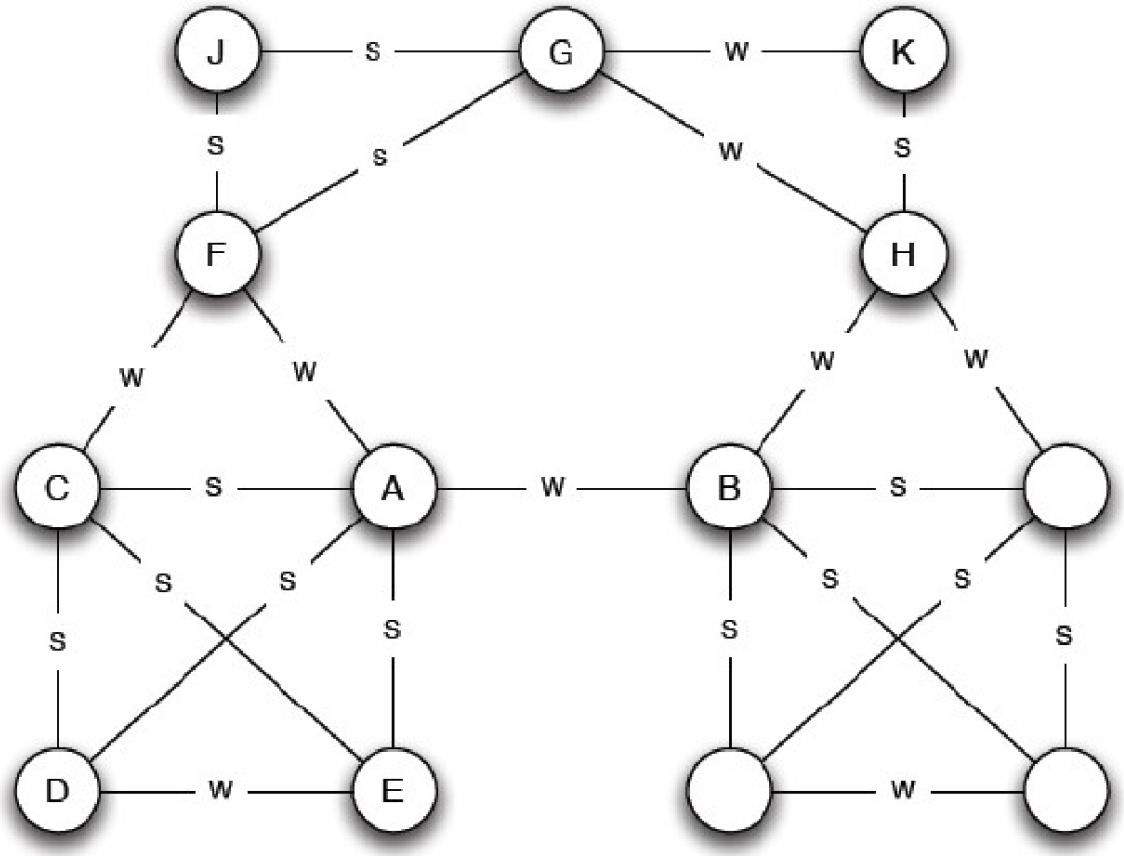


- **弱关系从何而来**

- 强三元闭包原理
  - 如果A与B、C之间的关系均为强关系，则B、C之间建立联系的概率很高
    - 即使暂时没有显式连接，也认为存在联系，或早晚将要存在联系
  - 因此，有“强三元闭包原理”如下：
    - 如果A有两个强关系邻居B、C，但B、C之间却没有任何关系（无论强弱），则节点A违背了强三元闭包原理
    - 换言之，如果A没有违背强三元闭包原理，就称A符合该原理

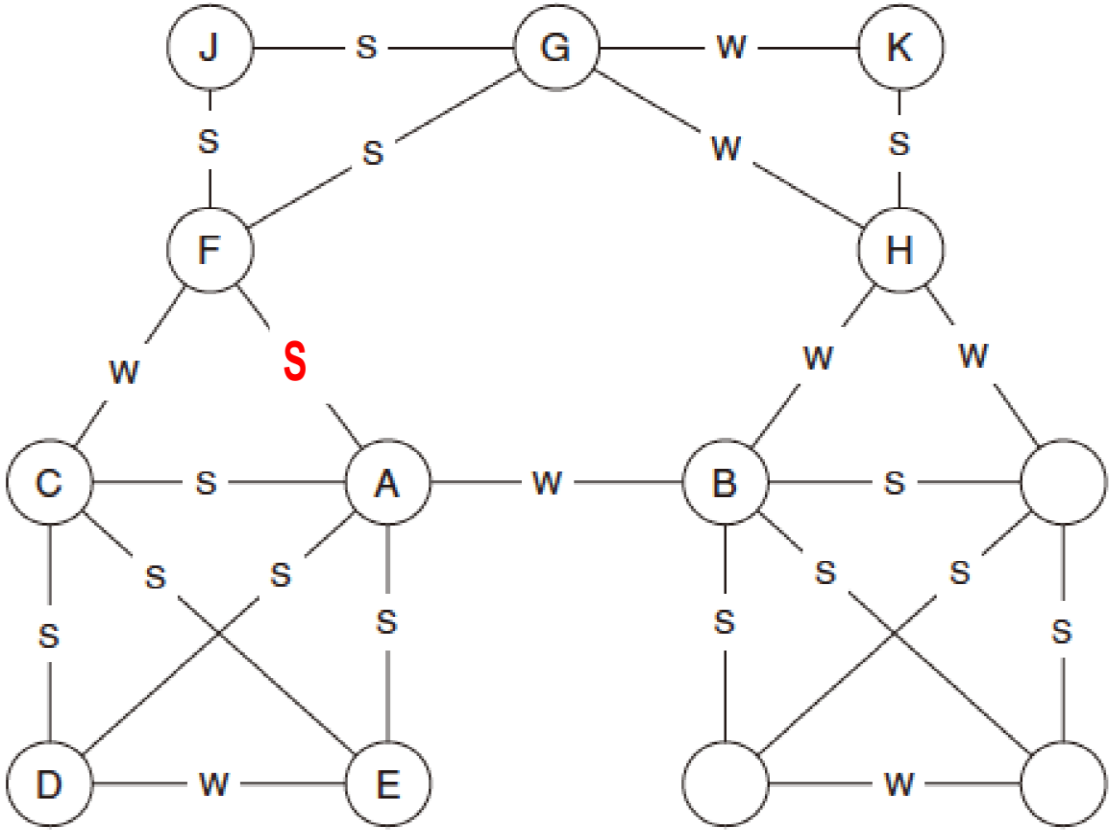
- **弱关系从何而来**

- 下面的例子中，哪些节点违背了强三元闭包原理?



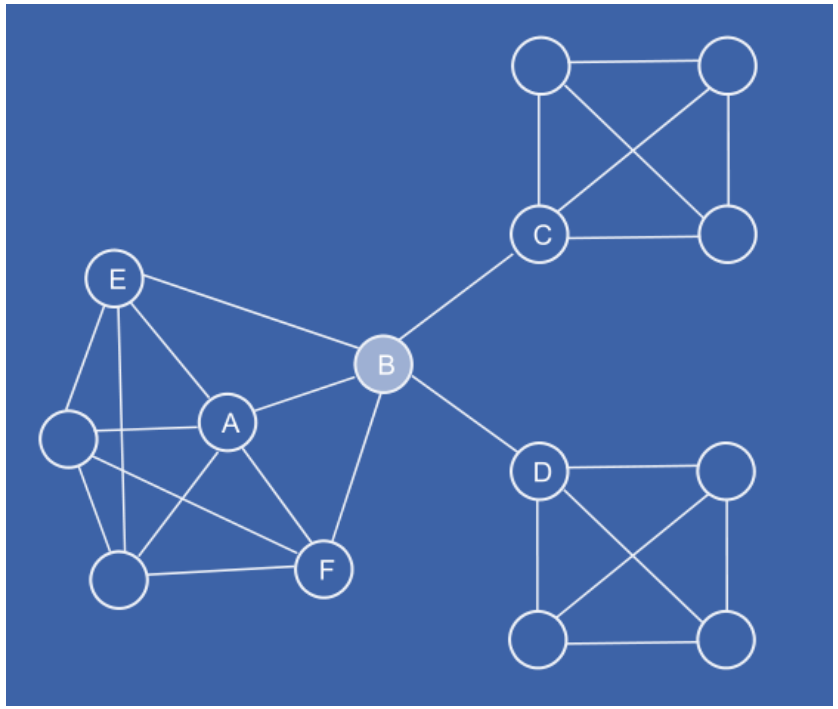
- **弱关系从何而来**

- 下面的例子中，哪些节点违背了强三元闭包原理?



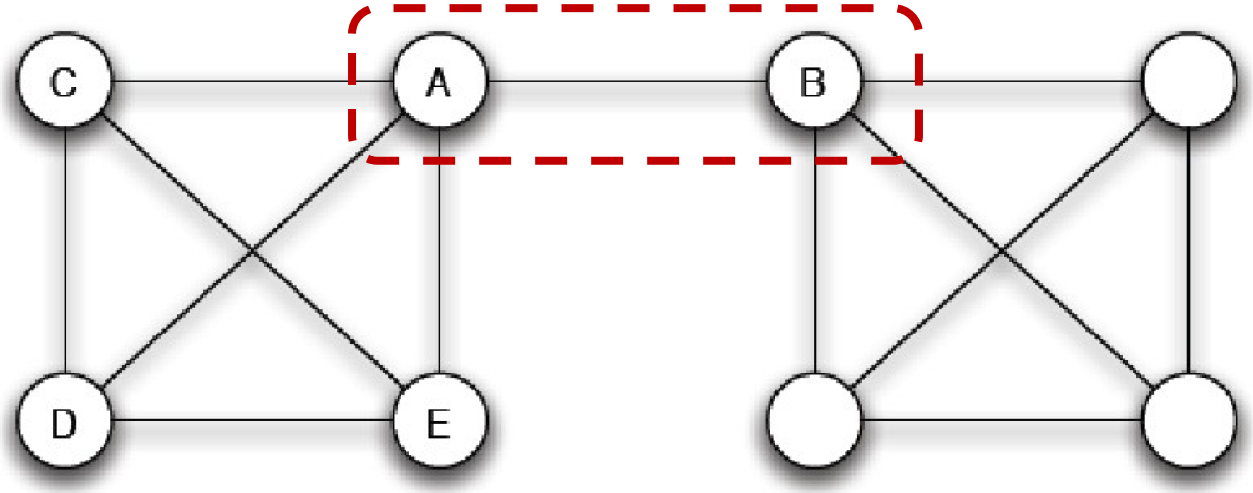
- **弱关系的细分**

- 笼统地说，弱关系为节点提供了更多的信息和更丰富的渠道
  - 例如，不同部门、不同社团之间的信息沟通：“项目经理”的角色
  - 然而，弱关系在网络中的位置/结构不同，其作用也不尽相同
    - 核心区别在于这条弱关系是不是唯一的信息渠道



- **弱关系的细分**

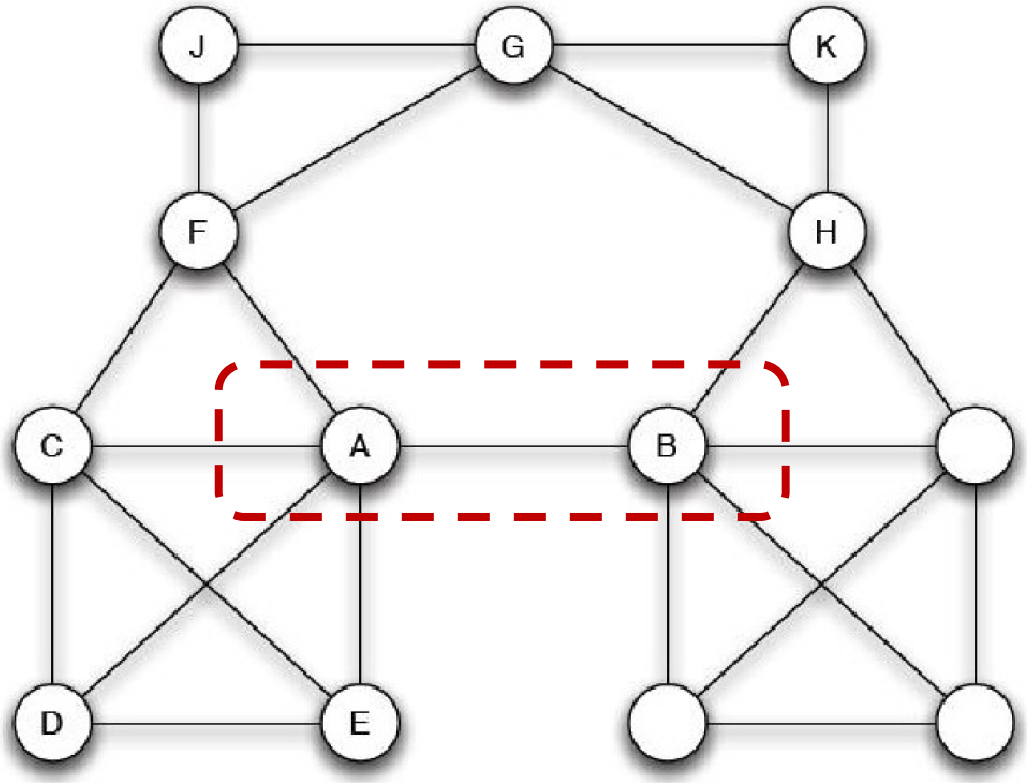
- **第一类：桥/结构洞**
  - 桥的主要作用，在于为组织引入外部的信息
  - 一种直观的判定方法为：如果移除某一条弱关系，就会使网络变成多个连通组件，则该关系为桥，两端的节点即为结构洞



- **弱关系的细分**

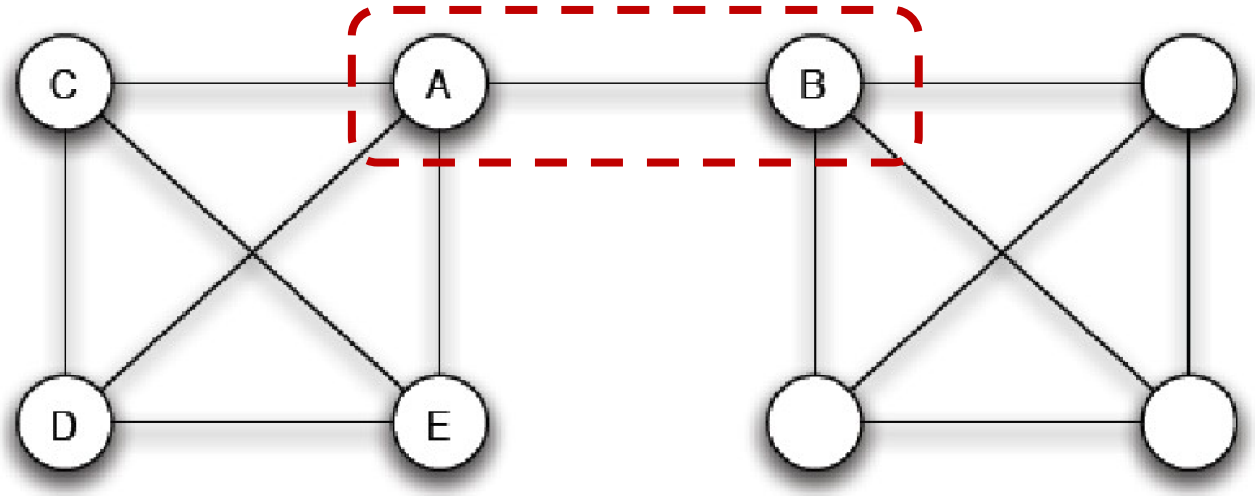
- **第二类：捷径**
  - 没有捷径，两端节点依然是连通的，但通讯的距离更远，代价更高

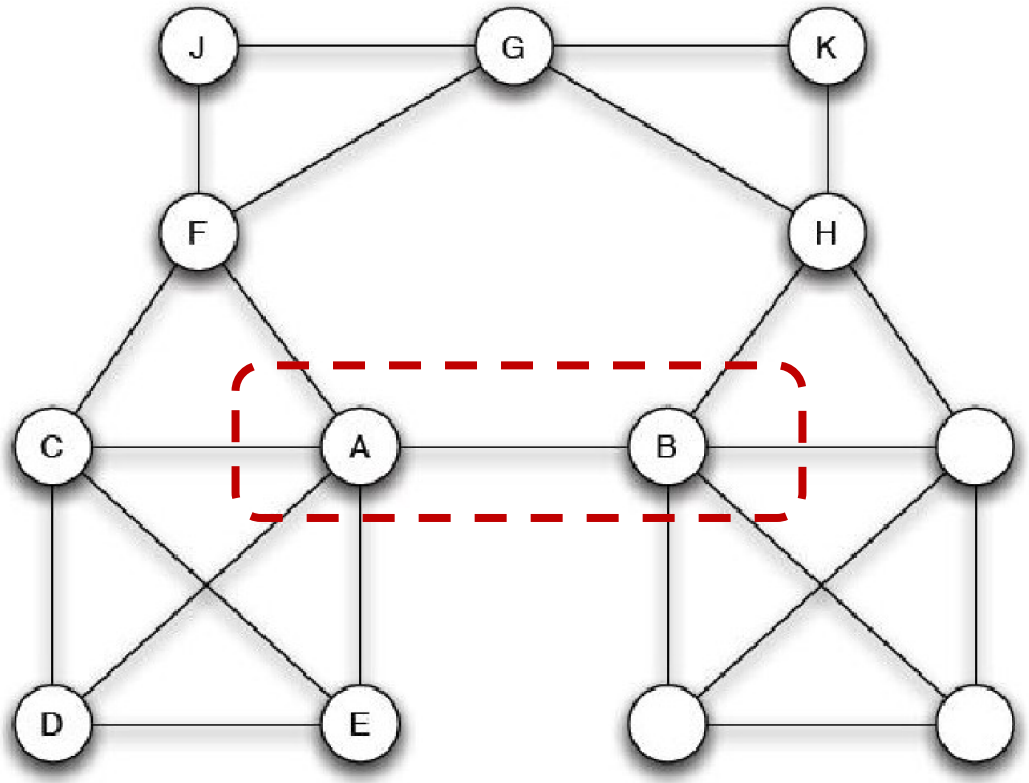
注意：捷径两端同样是没有共同好友的
因此，失去捷径意味着更高的沟通代价
（距离≥3）



- **弱关系的细分**

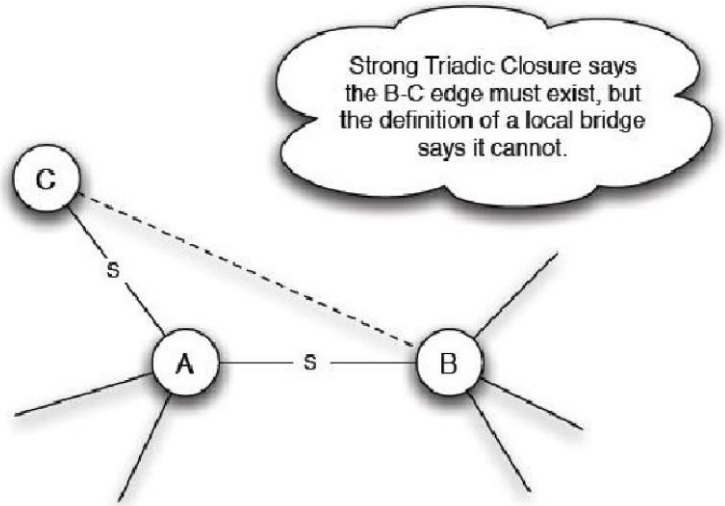
- **桥与捷径的关联?**
  - 在距离较远，通讯代价较高的背景下，桥与捷径本质上没有区别





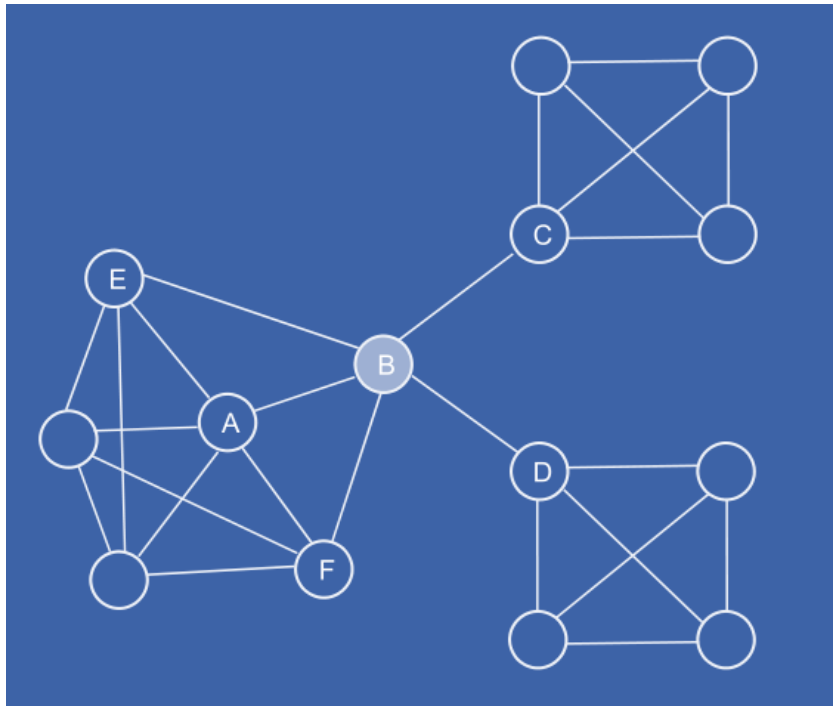
- **弱关系的细分**

- **关于捷径的断言**
  - 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
    - 假设A与B是强关系的捷径，由于A符合强三元闭包，必然还有与C的强关系
      - 那么，C和B之间就应该有关系（无论强弱）
      - 这样一来，A和B之间就不能算是捷径了



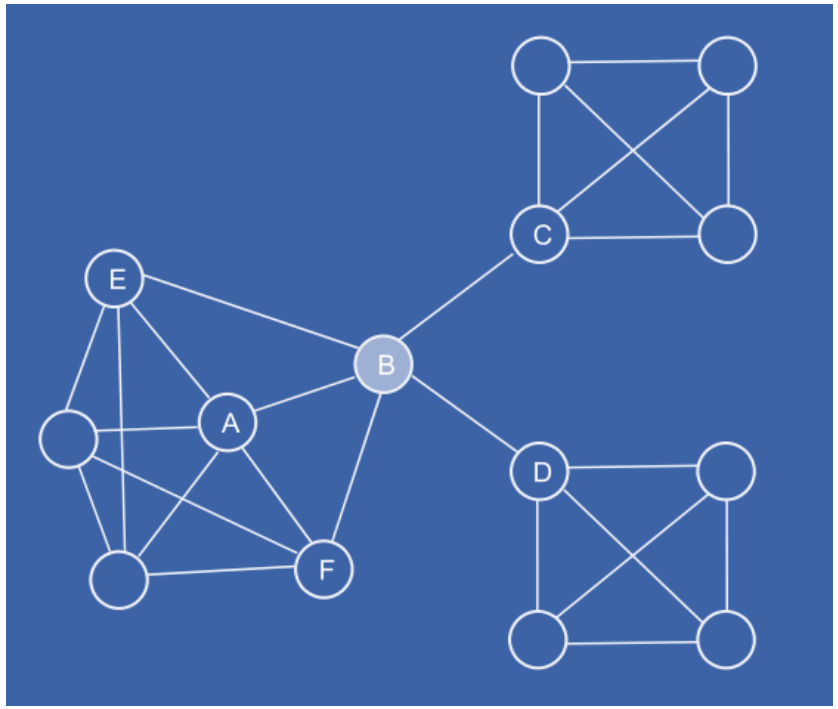
- **弱关系作用的度量**

- 如何量化衡量弱关系的强度?
  - 一种直观的结构洞的判定方法为:
    - 如果一个节点，移除该节点就会使网络变成多个连通组件，则该节点即为一个结构洞



- **弱关系作用的度量**

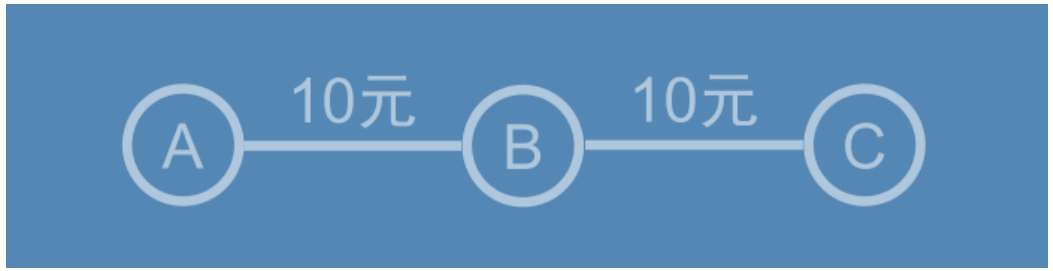
- 如何量化衡量弱关系的强度?
  - 通过连通组件判定，假设可能过强，计算也较为不便
  - 另一种衡量方式：聚集系数
    - 某个节点的聚集系数为：它的任意两个好友也互为好友的概率（比重）
    - 显然，聚集系数越低，该节点作为中介的作用也就越大



- **题外话：弱关系带来“中介”的议价能力**

- 弱关系形成的过程往往依赖于中介的“引荐”
- 相应的，这种“掮客”行为也成为了中介的“社会资本”
    - 例如，网络交换实验中，中介可能获得更大收益

往往以平分收场



如果仅可与邻居交涉，则B收益更大

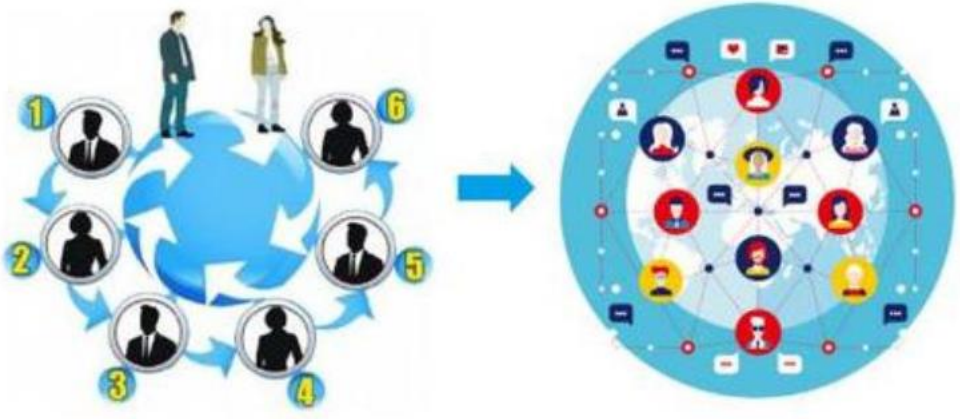
第四节课将详细介绍该部分内容

本节目录

- 社会网络基本概念
  - 从节点、关系再到网络
  - 路径、连通性与连通组件
- 强关系与弱关系
  - 三元闭包
  - 弱关系与结构洞
- **小世界现象**

- **小世界现象**

- 有关连通性的拓展阅读：小世界网络（Small World）
  - 茫茫人海，两个互不相识的人是否相互连通？
  - 如果想要互相认识，中间需要经过几个人？



六度空间理论　　六度人脉社交系统

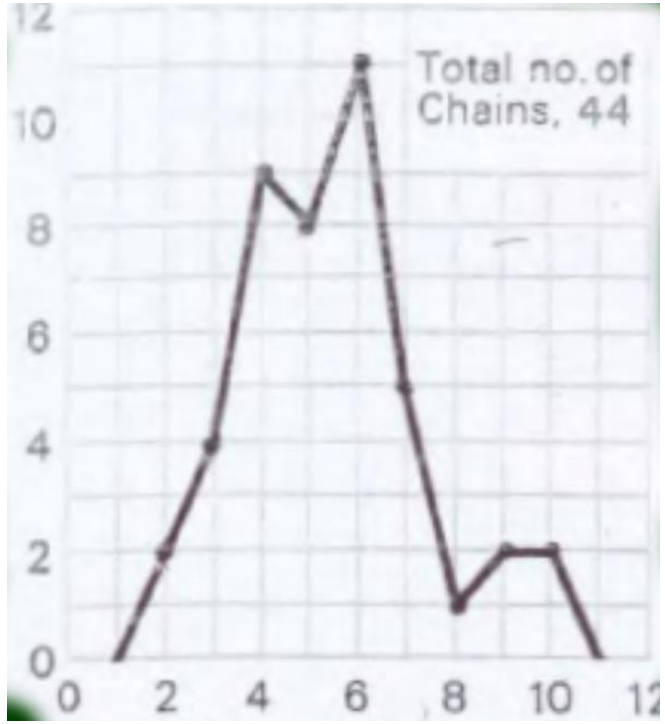
- **小世界现象**

- 六度空间理论，源自Stanley Milgram于1967年的著名实验
  - 设计：观察需要经过多少中间人，才能使信息从随机起点到达特定终点
  - 实验规则如下：
    - 参与人只能向信件转发给熟人，并请他继续转发
      - 换言之，不能通过陌生人，也不能直接送达目标
    - 参与者需力争让信件尽快达到目的地（有选择性地挑选转发人！）

- **小世界现象**

- 六度空间理论，源自Stanley Milgram于1967年的著名实验
  - 一共经历两次尝试（中部大农场 —— 东海岸）
    - 第一次：从Kansas的Wichita到哈佛大学神学院某学生的妻子
    - 第二次：从Nebraska的Omaha到Boston的Shanron的某位股票经纪人
  - 结果：平均通过 5 位中间人的转发可以抵达目标
    - 这就是 六 度空间的由来

右图为第二次实验所涉及44位被试的中介数分布→



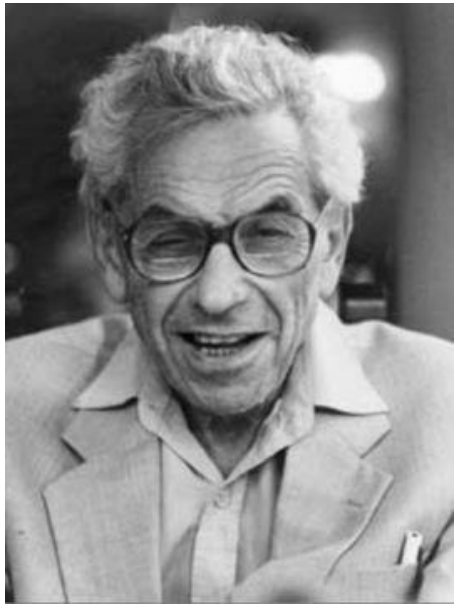
六度空间

- **小世界现象**

- 拓展阅读：Milgram实验在更大规模数据集上是否仍具有指导意义？
  - 2006年，微软的研究人员基于MSN数据，进行了类似的尝试
    - 结果发现，在将近2亿用户的网络中，平均6.6人即可使将近2000亿的配对产生关联
    - 超过87%的配对在7次以内即可产生关联

- **小世界现象**

- 拓展阅读：专业领域（更严格的关系筛选）是否同样存在小世界现象？
  - 保罗·埃尔德什（Paul Erdős），匈牙利数学家
    - 一生发表论文1475篇，与511人合作，被誉为“最高产的数学家”
  - 埃尔德什数（Erdős Number）
    - 描述学者与埃尔德什“合作距离”的一种方式
      - 埃尔德什本人，Erdős Number = 0
      - 直接与本人合作，Erdős Number = 1
      - 与其合作者合作， Erdős Number = 2，以此类推



- **小世界现象**

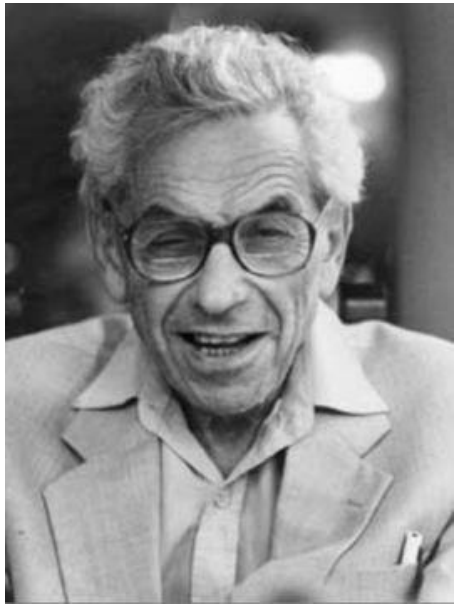
- 拓展阅读：专业领域（更严格的关系筛选）是否同样存在小世界现象？
  - 埃尔德什数（Erdős Number）
    - 统计显示，菲尔茨奖得主的Erdős Number中位数约为 3
    - 由于跨领域合作，许多非数学家也具有Erdős Number
      - 顺便提一句，本人的Erdős Number = 4

+1

Tong Xu

MR Erdos Number = 3

En-Hong Chen	coauthored with	Minming Li	MR2539922
Minming Li	coauthored with	Ronald Lewis Graham	MR2318683
Ronald Lewis Graham	coauthored with	Paul Erdős[1]	MR0592420



- **小世界现象**

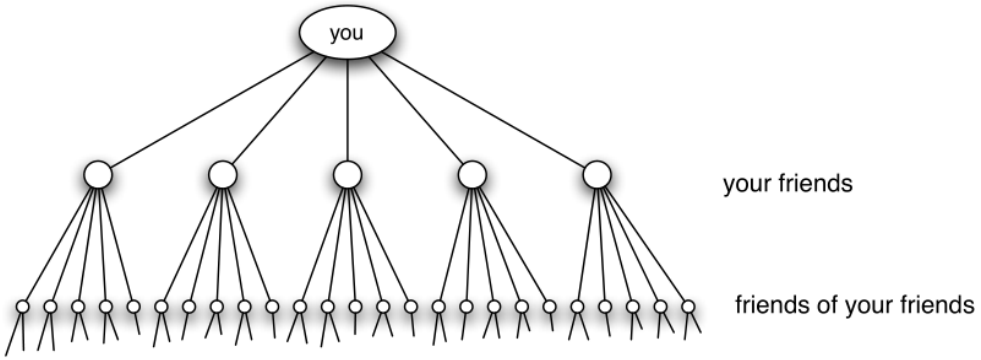
- 六度空间实验的意义
  - 在Milgram实验之前，人们意识到世界很小，但没有切实的证据
  - Milgran实验至少证明了“小世界现象”对应的两个惊奇
    - ✓ 世界是小的：社会网络中包含着丰富的短路径
    - ✓ “有意识的转发”能够“自动”找到这些短路径

➢ 问题来了：有没有模型能够描述和验证这两个性质？

- **从六度分隔到W-S模型**

- 有关短路径存在的直观化验证
  - 一个理想化的假设：我们的间接朋友数量惊人
    - 假设每个人平均有100个好朋友，那么两步之遥，你就可以认识10000个人
    - 换言之，经过以100的指数增长，6步以内，你可以覆盖全世界的所有人（>70亿）

现实真的如此“理想”？

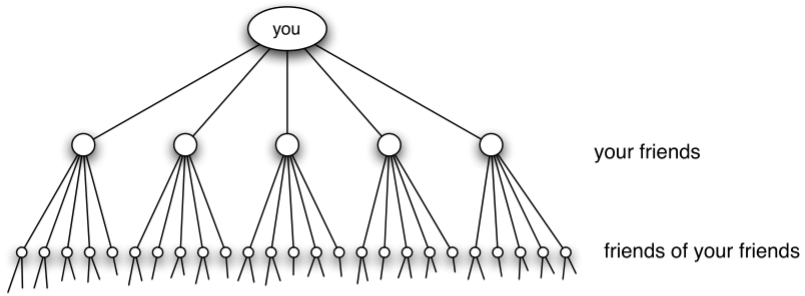


(a) *Pure exponential growth produces a small world*

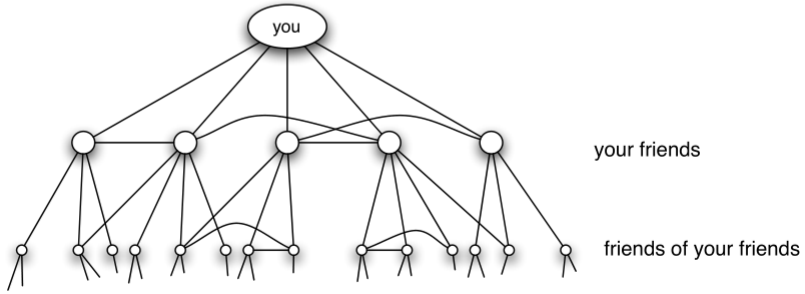
从六度分隔到W-S模型

- 有关短路径存在的直观化验证
  - 现实是：基于三元闭包理论，你和你好友的好友之间有着大量的重叠
    - 因此，多跳所能连接的好友人数非常有限
  - 注意：强弱连接在此时作用的差异
    - 强连接有更强局部性，重叠更多
    - 而弱连接会将相隔较远的节点相连，重叠更少

理想很丰满，现实很骨感→



(a) *Pure exponential growth produces a small world*



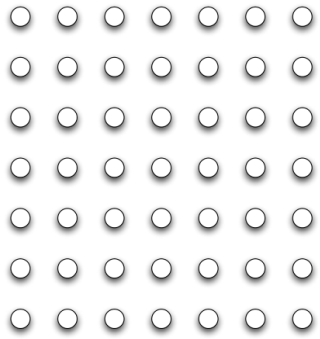
(b) *Triadic closure reduces the growth rate*

W-S模型

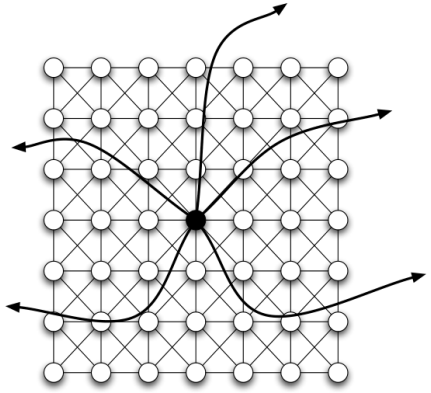
- **从六度分隔到W-S模型**

- 弱连接激发的瓦茨-斯特罗加兹（Watts-Strogatz，W-S）模型
  - 同质性：朋友、同学、同事等关系
    - 对应社会网络中的大量的“三角形”（圈子）
  - 弱连接：认识的“远程”朋友
    - 形成距离较远的“短连接”/捷径

我们需要一种模型，同时体现这两种力量！



(a) *Nodes arranged in a grid*

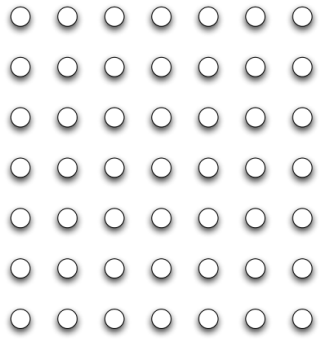


(b) *A network built from local structure and random edges*

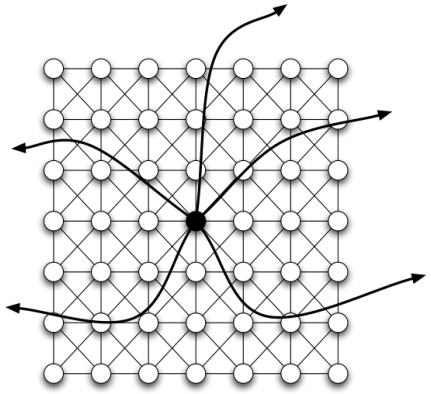
从六度分隔到W-S模型

- 弱连接激发的瓦茨-斯特罗加兹（Watts-Strogatz，W-S）模型
  - W-S模型同时包含上述两种连接（如右下图所示）
    - 同质性连接：r 网格步以内的节点连接，呈现为“三角形”
    - 弱连接：到网络中其他 k 个节点的链接，呈现为“远程边”
      - 注意：这里的 k 个 节点是随机均匀挑选的

基于W-S模型，节点之间可以通过少量的随机连接实现沟通！



(a) *Nodes arranged in a grid*



(b) *A network built from local structure and random edges*

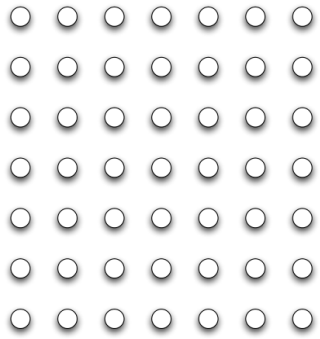
D. J. Watts & S. H. Strogatz, Collective Dynamics of Small-World Networks, Nature, 1998

W-S模型

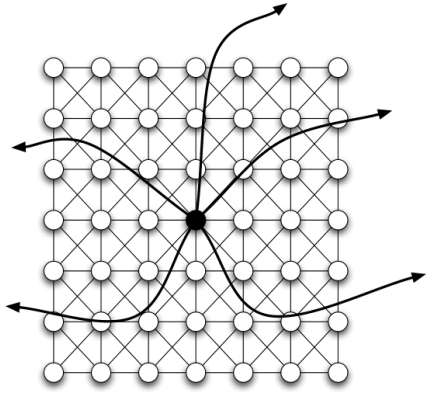
- **从六度分隔到W-S模型**

- 弱连接激发的瓦茨-斯特罗加兹（Watts-Strogatz，W-S）模型
  - W-S模型体现了同质性和弱连接的融合，可视作现实社会网络的合理近似
  - 可以证明，这样的网络中任意两个节点间存在短路径的概率很高
    - 短路径可以近似看做先坐高铁到不同城市，再在城市中通过短途交通到达目的地

然而，W-S模型也存在着局限



(a) *Nodes arranged in a grid*



(b) *A network built from local structure and random edges*

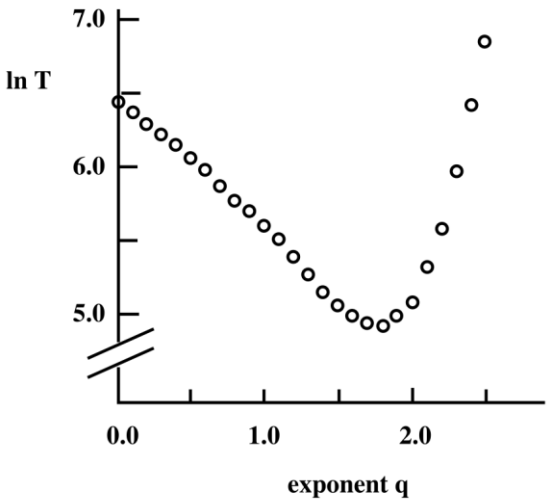
- **从六度分隔到W-S模型**

- 弱连接激发的瓦茨-斯特罗加兹（Watts-Strogatz，W-S）模型
  - W-S模型局限性的由来：路径的随机性与知识的局部性
    - 在W-S模型中，我们只知道每一个节点局部的连接，以及到目标的距离（网格步数）
    - 由于我们并未获得全局的网络信息，因此，我们无法生成确定性的路径
    - 在这种情况下，W-S模型的解决方法是通过随机连接来达到终点

导致的结果是，W-S模型往往无法在最优甚至较优的步数内到达终点

从六度分隔到W-S模型

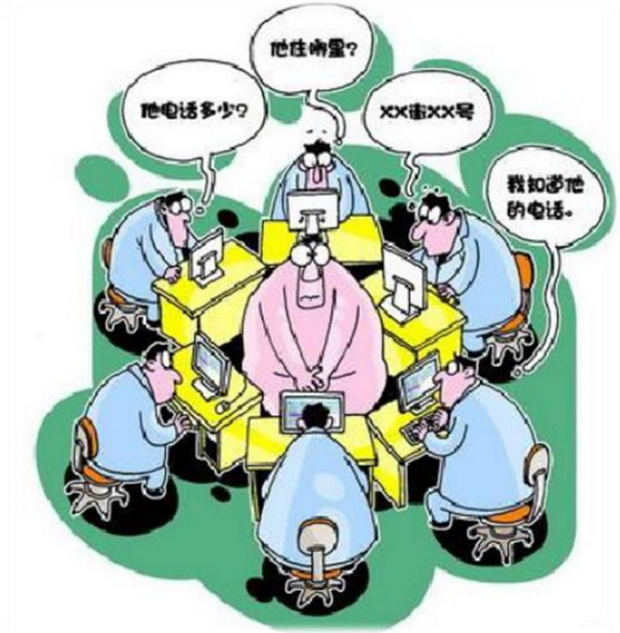
- 如何改进W-S模型来获得更合理的短路径?
  - 2000年，Jon Kleinberg提出W-S-K模型，对W-S的局限性进行了完善
    - 引入参数 q，对于起点 v 和终点 w，以 $d(v,w)^{-q}$ 的概率来生成远程连接
      - 其中，d(v,w)为两个节点之间的距离
      - 显然，W-S模型相当于 q = 0的情况（即所有边等概率，完全随机搜索）
    - 通过这种方式，可以通过调节 q 值来选择合适的路径
      - 大规模实验表明 q = 2 时往往可以获得相对最短的路径



Jon Kleinberg, Navigation in a Small-World Networks, Nature, 2000

- **小世界现象：总结**

- 六度空间理论证实，人们的距离要比想象中要小得多
  - 同时，“有意识的转发”能够有效找到沟通节点的短路径
    - 某种意义上说，这种“有意识的转发”也体现了“群体智能”的思想
  - 一点题外话：1993年，电影《六度分离》上映
    - 豆瓣评分：7.6分
    - 剧中台词：“无论是美国总统还是威尼斯的船夫，只要找到合适的人，我们就能彼此相连”



本章小结

社会网络基础

- 课程背景、问题与挑战
- 社会网络基本概念
- 三元闭包与强弱关系
- 小世界现象与W-S模型