第三章：计算社会学理论——图论

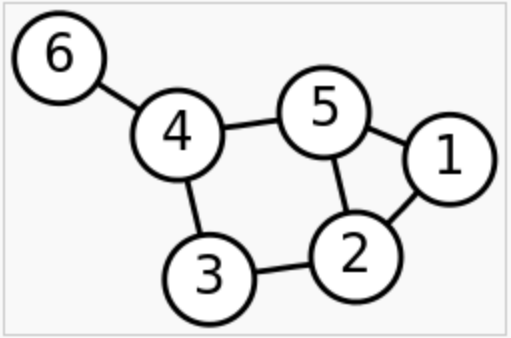
Graph Theory

主　　讲：郭斌 副教授

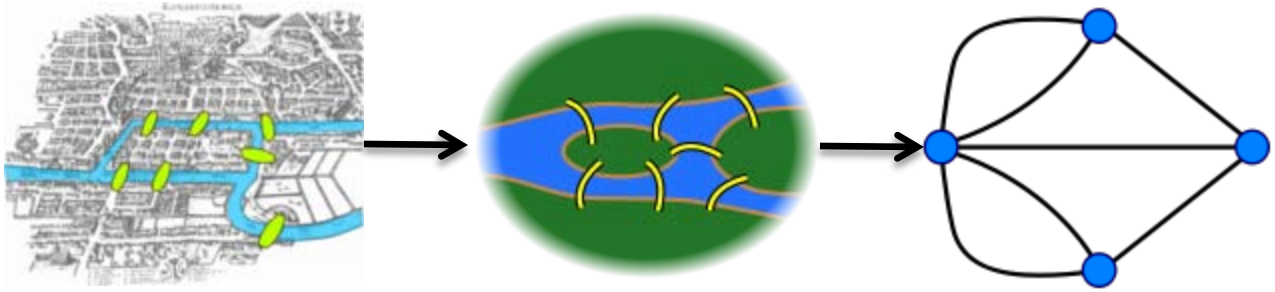
单　　位：西北工业大学计算机学院　陕西省嵌入式系统技术重点实验室
电　　话：18729229010
办 公 室：计算机学院512B房间
电子邮箱：guobin.keio@gmail.com
个人主页：http://www.ayu.ics.keio.ac.jp/~bingo/
课程主页：http://www.ayu.ics.keio.ac.jp/~bingo/course.html

图论——计算机基础理论之一

- 离散数学核心内容
  - ✓与光滑变化的实数不同，离散数学的研究对象——例如集合、图和数理逻辑——不是光滑变化的，而是拥有不连续的值。
  - ✓因此离散数学不包含微积分和分析等"连续数学"的内容。
- 图论以图为研究对象，研究顶点和边组成的图形的数学理论和方法。



像这样的图是离散数学的研究对象之一，它们拥有有趣的数学性质，可以作为现实世界用来解决问题的模型，而且还在计算机算法开发中有着举足轻重的作用。



莱昂哈德·欧拉1735年“**柯尼斯堡七桥问题**”：在所有桥都只能走一遍的前提下，如何才能把这个地方所有的桥都走遍？

本章学习要点

- 人群与网络
- 网络与图
- 路径、连通图、连通分量
- 广度优先搜索

无处不在的网络-自然界网络

- 在自然界，网络无处不在
  - ✓ C. S. Elton(1927)，在生态系统中，物种之间 “吃”与“被吃”也是一个网络，食物链是自然生态系统稳定的重要机制
  - ✓一些自然资源之间的关系，如水的转化关系，也是一个网络
- 网络是自然世界发生演变的机制之一

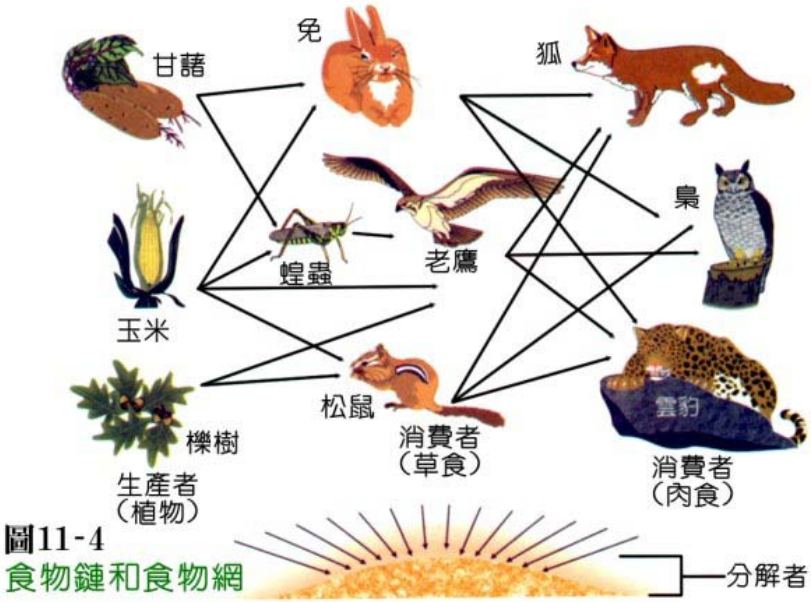
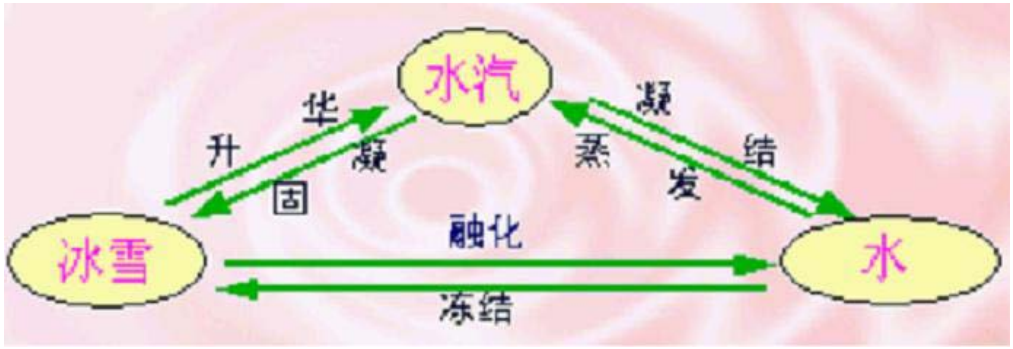
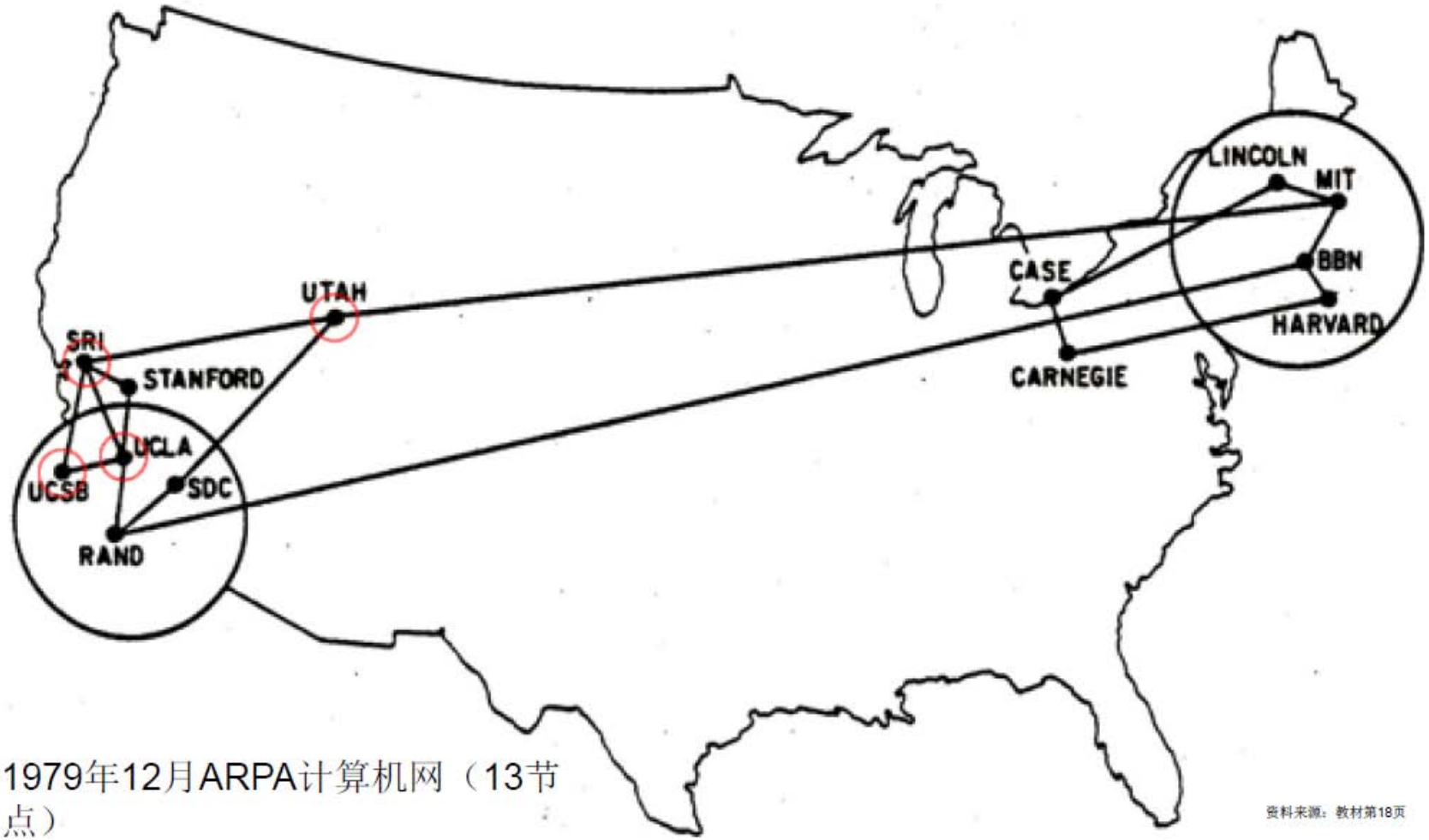


圖11-4
食物鏈和食物網



无处不在的网络-计算机网络



1979年12月ARPA计算机网（13节点）

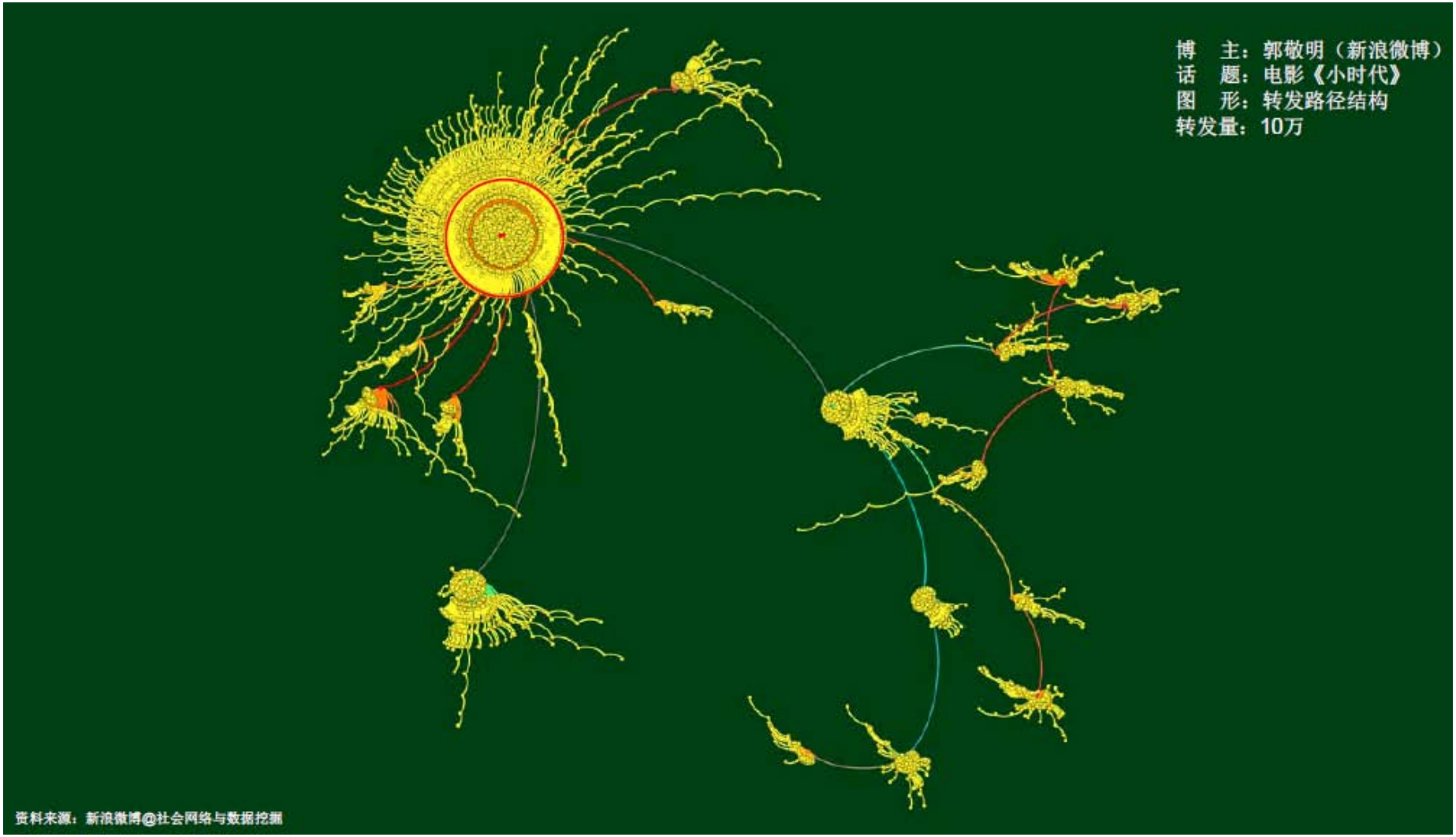
资料来源：教材第18页

最初的互联网，13个节点及节点节的连接关系

无处不在的网络-地铁路线图

无处不在的网络-Facebook联系网络

无处不在的网络-Twitter信息传播



ICT技术发展对网络的影响

•催化了各种网络的发展

✓规模变大

✓范围变广

✓物理特性

✓新型网络涌现

• 微信vs. 微博？

形形色色的网络

交通运输网，邮政网，电话通信网，计算机网，互联网，万维网

- 社会关系网，产品供销网，金融借贷网
- 智能电网，无线网，传感网，物联网
- 神经网，生物代谢网，食物链（网）
- 攻守同盟网，恐怖主义网络
- 人人网，新浪微博网，QQ，团购网，…

世界充满各种“网络”，掌握网络分析的基本理论和方法可以对自然、社会和组织进行更好理解。

本章学习要点

- 人群与网络
- 网络与图
- 图的基本定义
- 路径、连通图、连通分量
- 广度优先搜索

网络与图

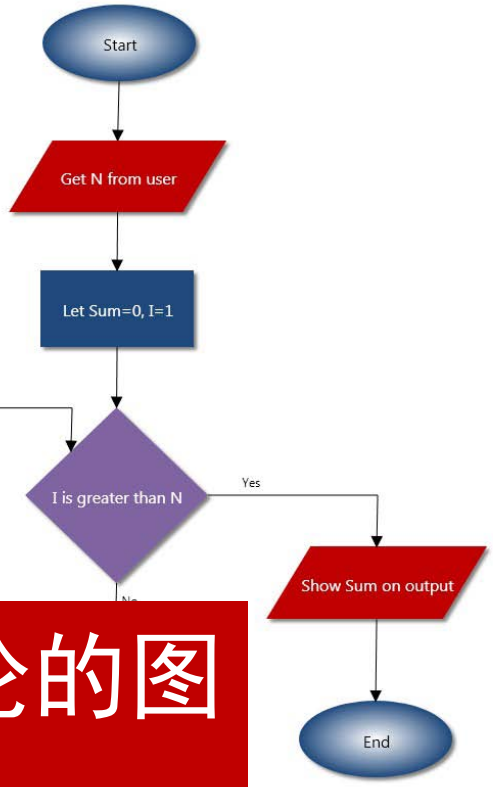
当我们构想网络这个词…

Dictionary: an interconnected system of things or people.

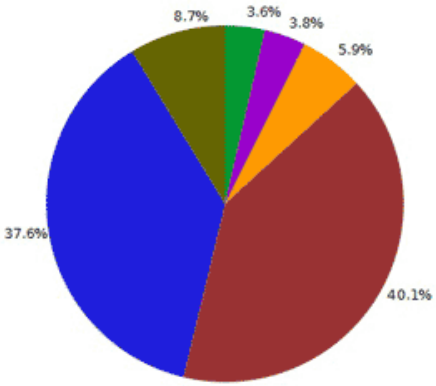
“事物” + “联系”

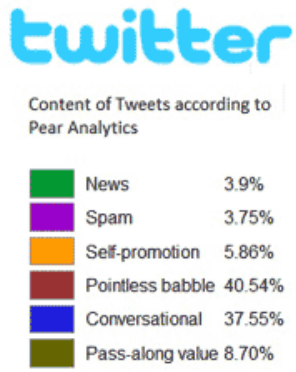
- 计算思维之“抽象”——从网络到图

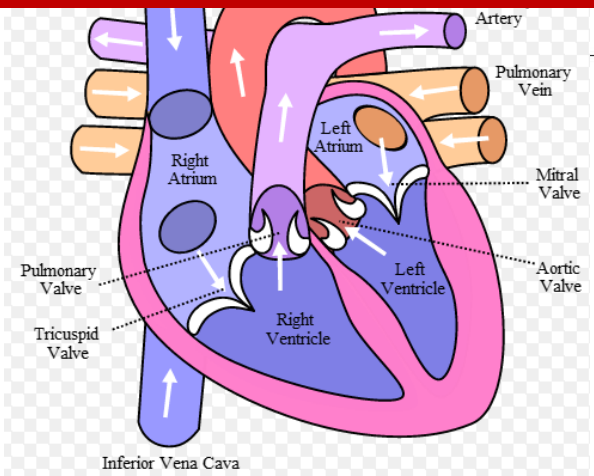
网络与图



但这些都不是我们要讨论的图（Graph）！

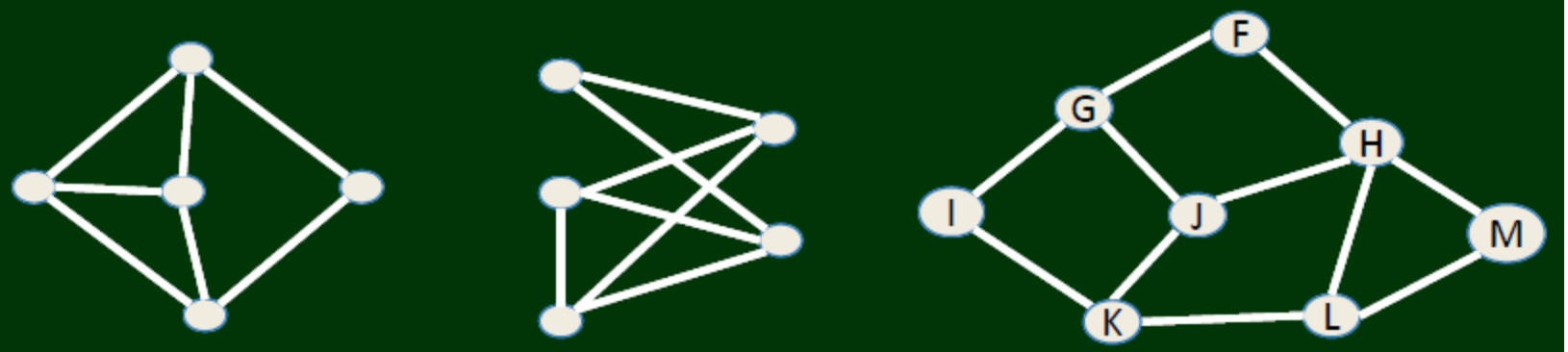






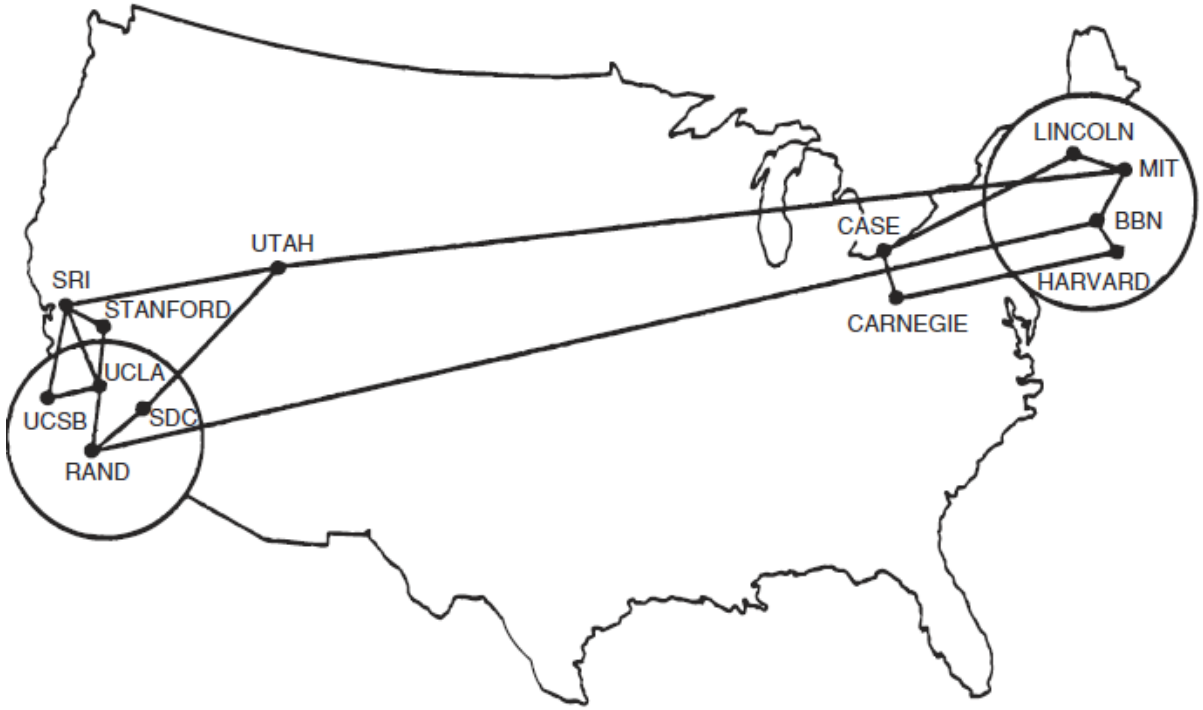
网络与图

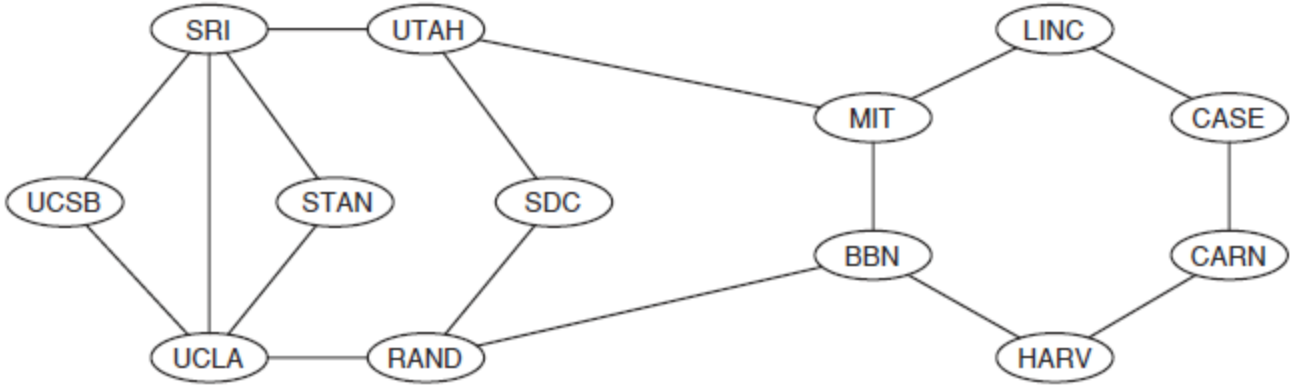
图＝事物＋联系



- 节点（vertex, node, point）
- 边（连接, 链接, 关系, 联系；edge, link, tie）

网络与图


LINCOLN
MIT
BBN
HARVARD
CASE
CARNEGIE
UTAH
SRI
STANFORD
UCLA
UCSB
SDC
RAND

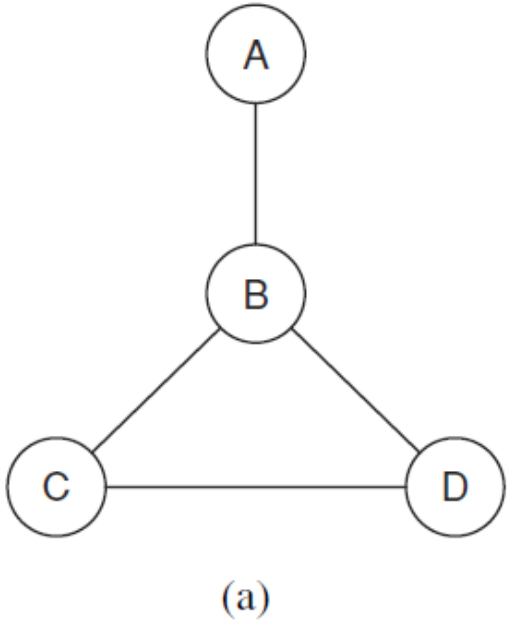

SRI
UTAH
LINC
MIT
CASE
UCSB
STAN
SDC
BBN
CARN
UCLA
RAND
HARV

本章学习要点

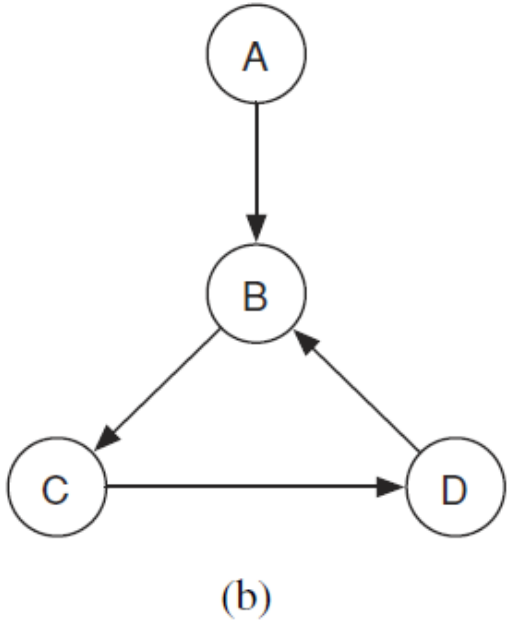
- 人群与网络
- 网络与图
- 路径、连通图、连通分量
- 广度优先搜索

图:节点与边

图是网络结构信息的抽象，表达的是网络中各种事物之间的关系。



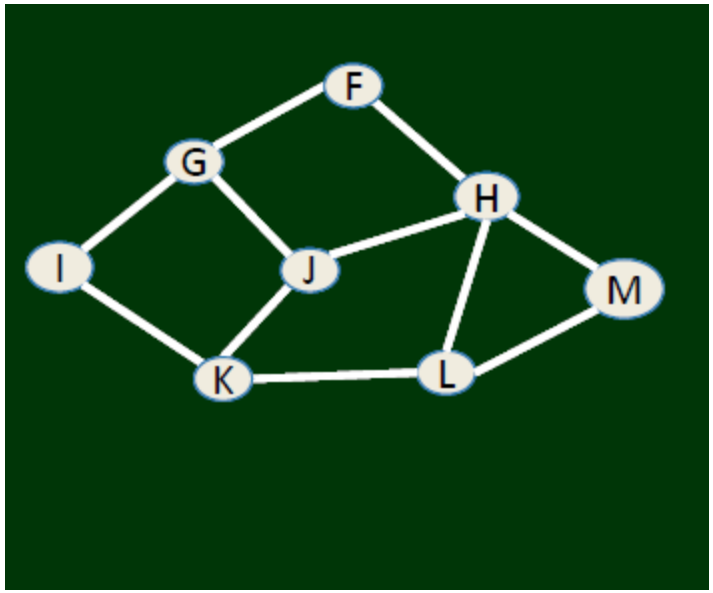
无向图（undirected graph）



有向图（directed graph）

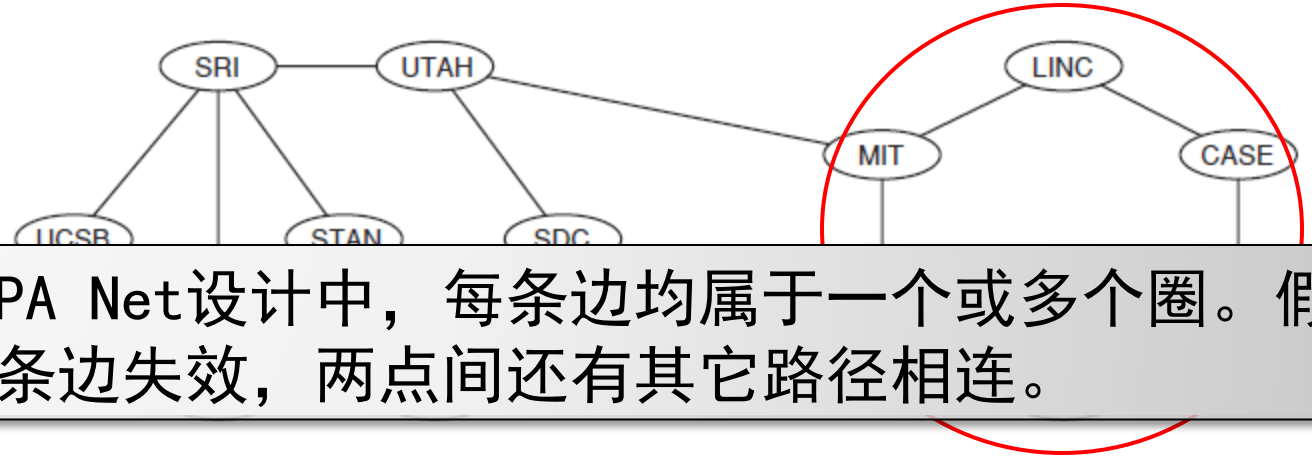
图:路径、最短路径、距离

路径：一个节点序列的集合，序列中任意两相邻节点都有一条边相连；
简单路径：不包含重复节点的路径；
最短路径：两节点间最短的路径；
距离：最短路径的长度为距离



圈（Cycles）

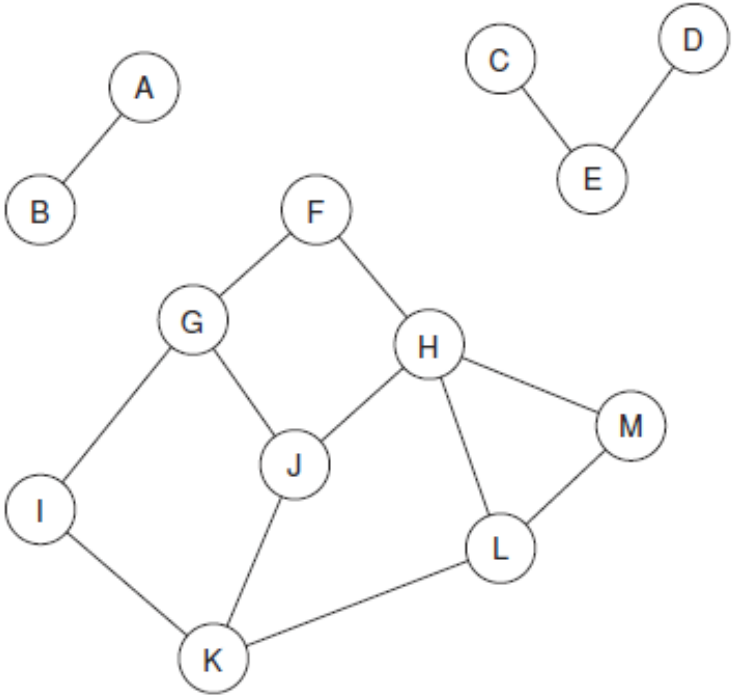
- 有一类很重要的非简单路径称为圈（Cycle）。
- 定义：至少包含三条边，且起点和终点相同，而除此之外所有节点均不重复的路径。
- 最短圈：仅包含三条边的圈。

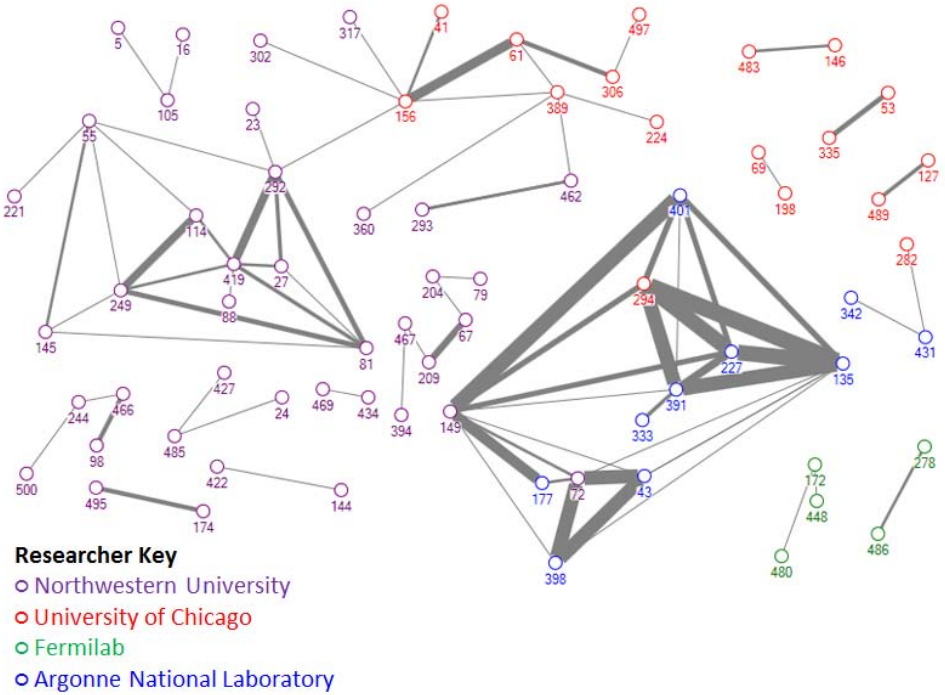


在ARPA Net设计中，每条边均属于一个或多个圈。假设任意一条边失效，两点间还有其它路径相连。

连通性（Connectivity）

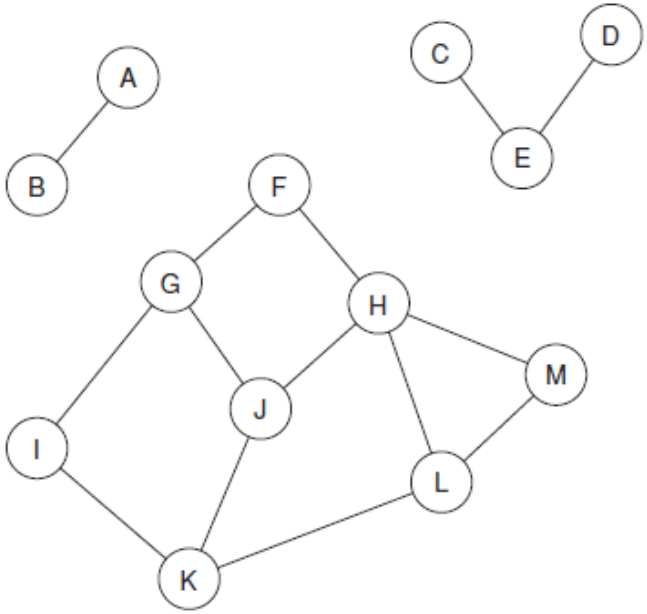
- 问题：是否任意节点均可通过某条路到达其它任一节点。
- 连通图：一个图中任两点间有路相通。
- 例子：通信、交通网等以此为目标。





连通分量（Components）

- 一个非连通图可以由一组连通“部件（Pieces）”表示，每个部件都是一个独立的连通图，且之间无交集。
- 连通分量(Connected Component)：1）子集中任意两节点间均有路径相连，2）该子集不是任何满足条件1）子集的一部分。



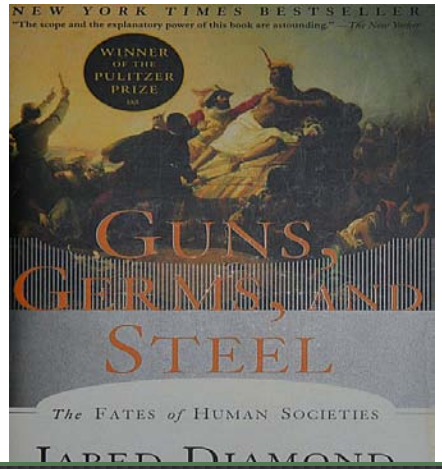
问题：F、G、H、J是否连通分量？

从连通分量中理解网络性质：中心节点

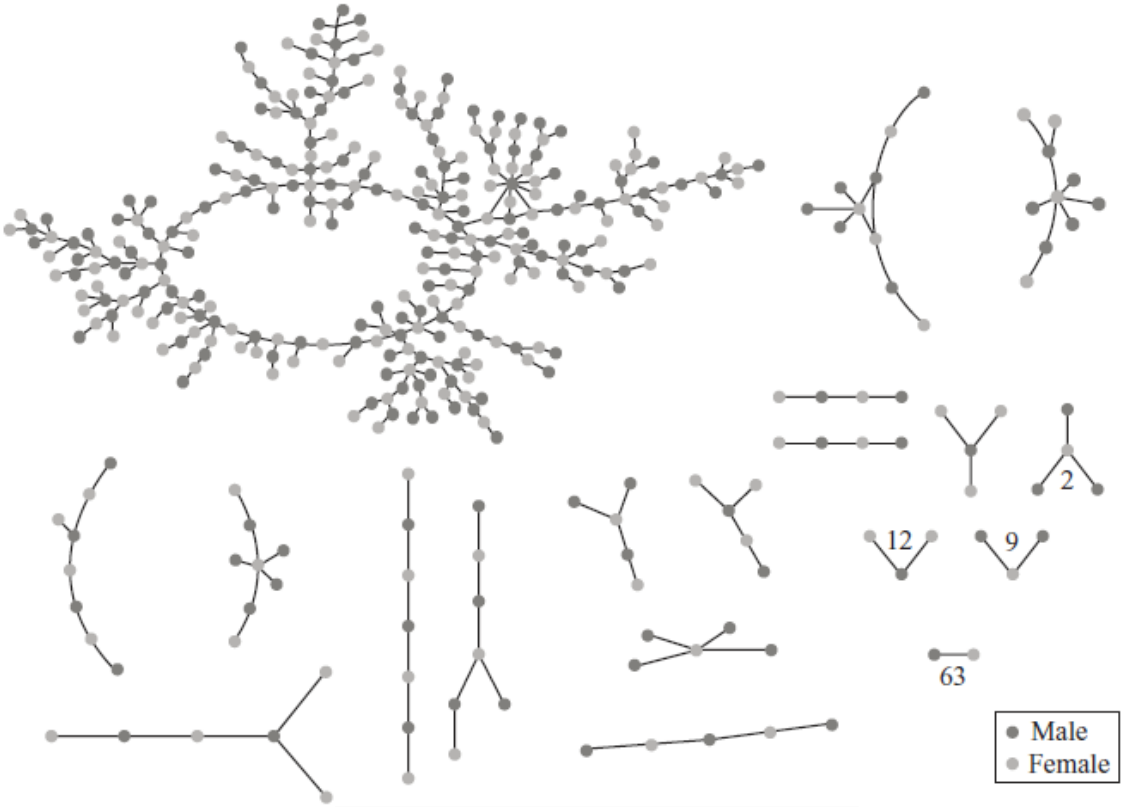
超大连通分量（Giant Component）

- 问题：Is the global social network connected?
- Presumably not；连通性是一个脆弱(brittle)的属性。
- 但在大型、复杂网络中存在包含大部分节点的连通分量：超大连通分量。

一般性质：当一个图网络中存在超大连通分量，则它是唯一的。



500年前，欧洲->美洲大陆



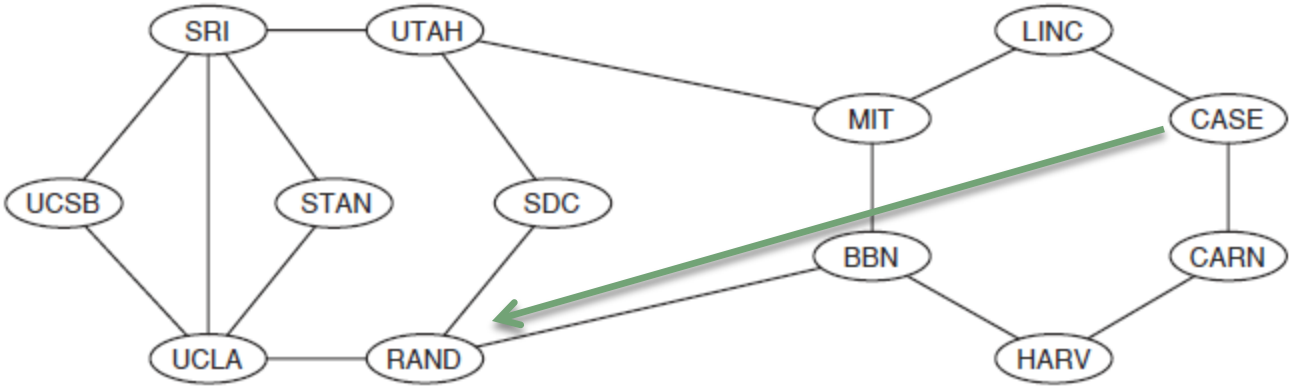
美国高中恋爱关系图

本章学习要点

- 人群与网络
- 网络与图
- 路径、连通图、连通分量
- 距离与先宽搜索

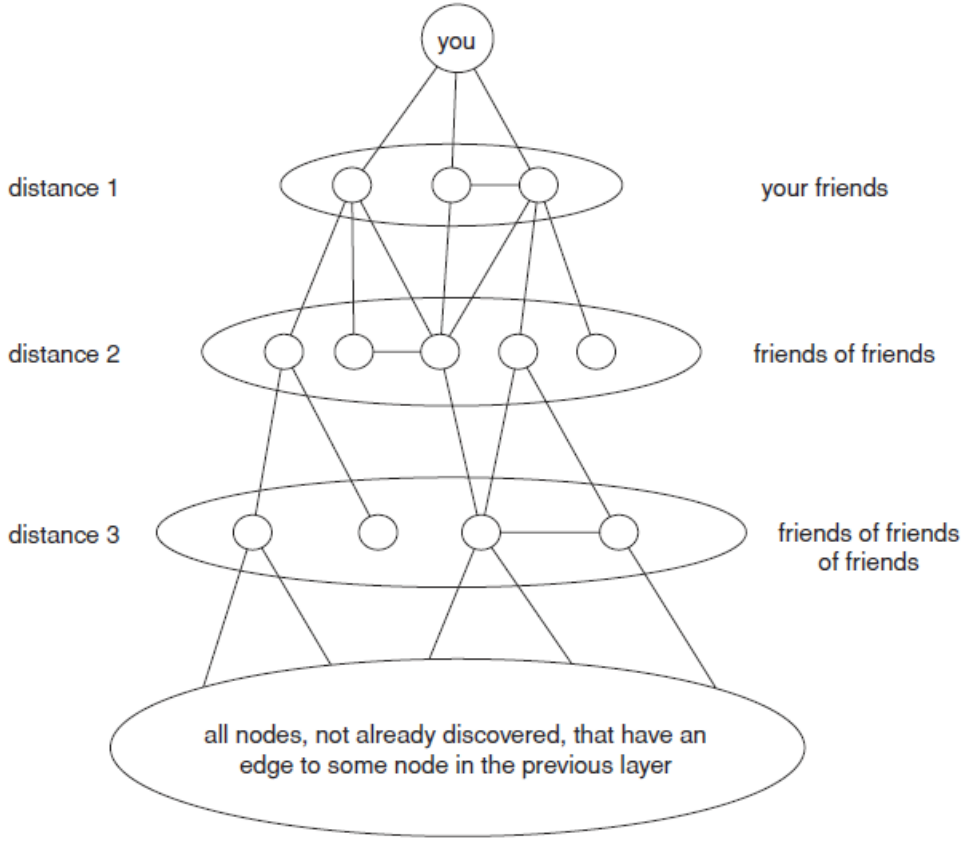
距离

- 两节点间是否有路径是问题一方面；路径的长度则是问题另一方面。
  - 例如：交通运输、互联网通信、疾病传播中中转次数，或“跳数”。
- 长度：一条路径长度为其包含的边数；
- 距离：图中两节点间最短路径长度。



先宽搜索（广度优先搜索）

- 问题：复杂图中如何计算两点距离？
- 先宽搜索（breadth-first search）：Motivated by Global Social Network.



实现过程

(1)和直接朋友距离为1；

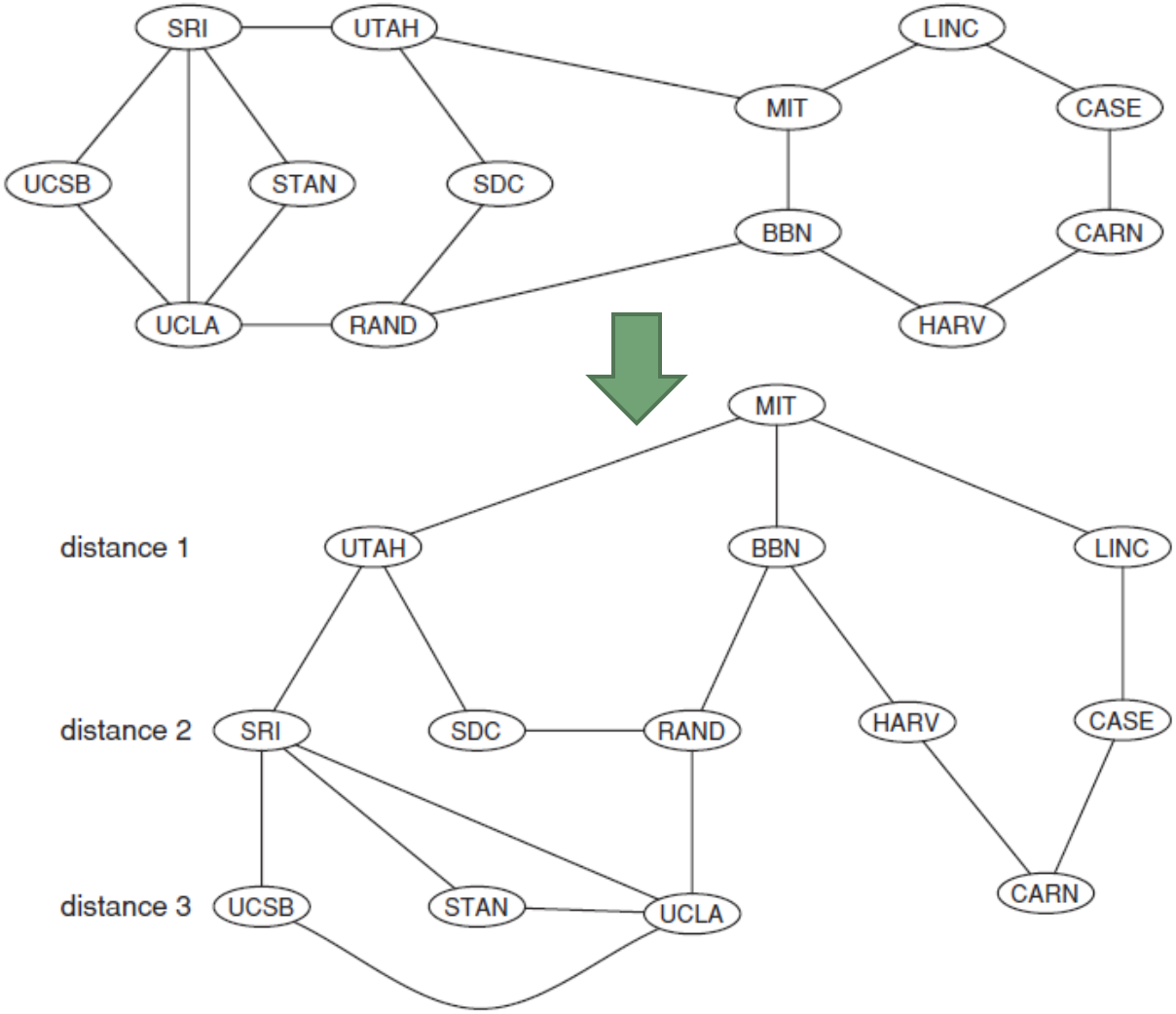
(2)朋友的朋友，距离为2；

(3)再找(2)中所有朋友，要排除1,2中出现的人，距离为3；

(4)以此类推…

先宽搜索（广度优先搜索）

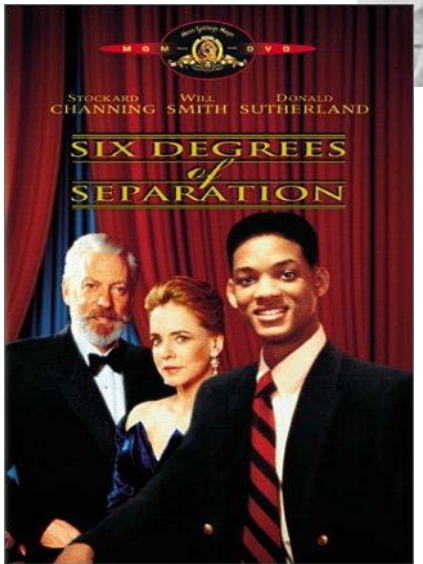
- 问题：MIT到UCSB的距离是多少？



小世界现象

最多只需经过6个人，你可以和奥巴马建立联系！

- 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



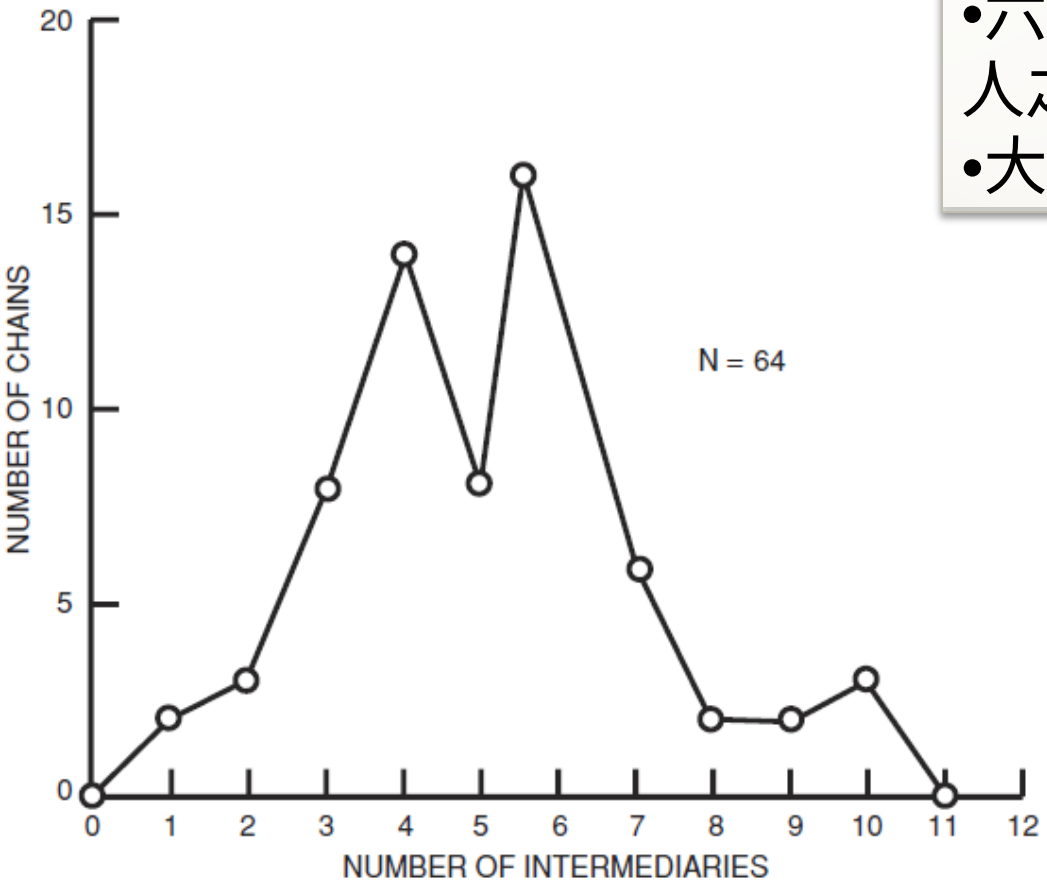
《六度分隔》电影，1993

Stanley Milgram 的实验（1967年）

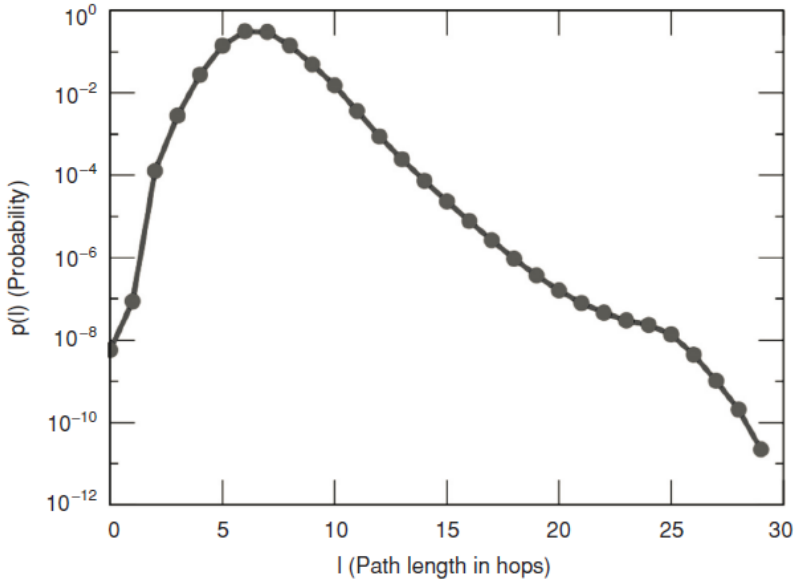
- 几百名“ 初始者 ”，要求每人尝试通过转发让一个指定的人收到一封信；向每个初始者提供了目标收信人的姓名、地址、职业等个人信息；
- 规定：参与者只能将信件转发给能直呼其名的熟人，并请他继续转发。因此，如果一个参与者不认识目标收信人，则他不能直接将信寄给他；
- 要求：参与者力争让这信件能尽早达到目的地；
- 结果：约三分之一的信件经过平均六次转发到达了目标

小世界现象

- 六度分隔：社交网络中，人与人之间相连距离甚短；
- 大型网络中存在丰富的短路径



米尔格拉姆“小世界”实验论文中公布的成功将信送达目标人物的路线长度分布图



微软MSN2.4亿活跃用户距离分布图（平均6.6）

扩展：在网上，从北大如何走到中科大？

北京大学 → 相关链接
→ 研究生院院长联席会
→ 成员单位 → 中科大
研究生院 → 中科大图
书馆 → 中科大

课后思考：

1. 从西工大→本课程主页
2. 西工大→清华/北航
3. 西工大→中国政府网(gov.cn)

新浪微博用户之间的路径

课后作业

- 第二章课后1、2题，本周五交
- 助教：吴文乐（15339202169）
- 地点：学院楼513A

End of Chapter 3.1

Contact: guob@nwpu.edu.cn
http://www.ayu.ics.keio.ac.jp/~bingo/

西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY