



中國人民大學
RENMIN UNIVERSITY OF CHINA

计算传播理论与实务

2019-2020秋季学期

第四讲 网络分析

网络分析概述



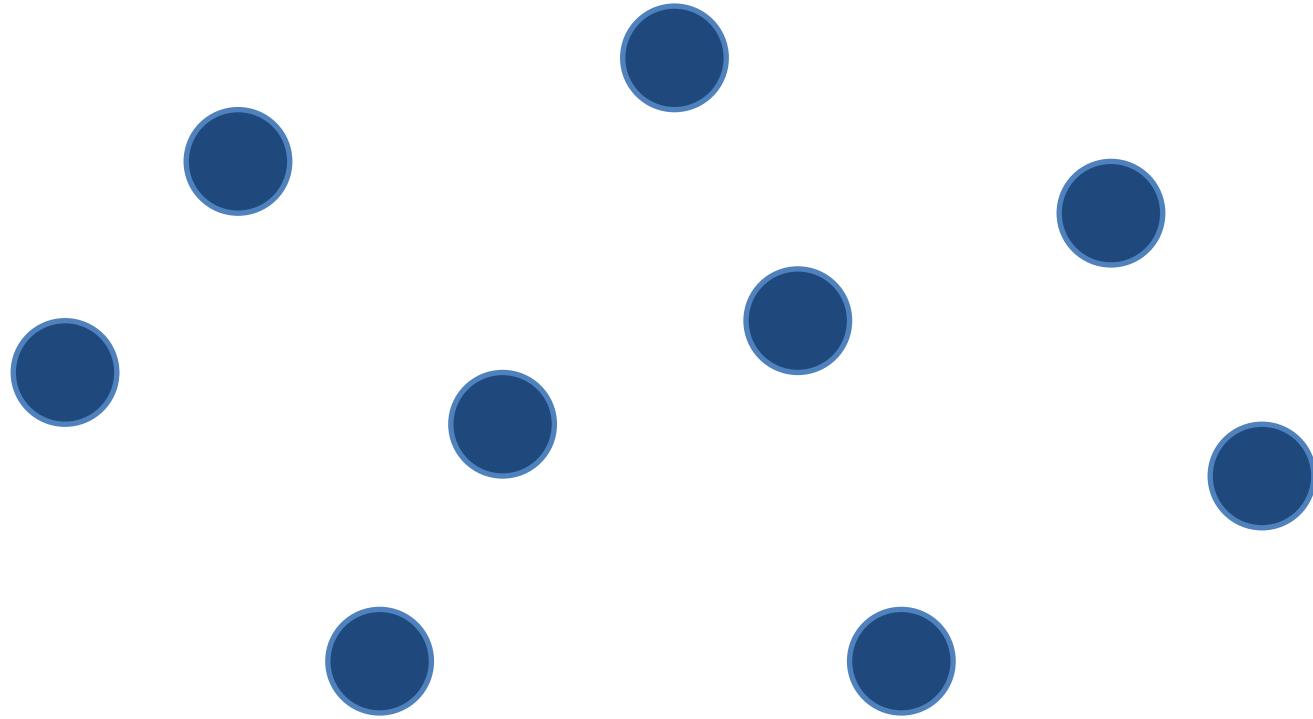
Orion
constellation

授课教师：范举副教授、塔娜讲师
时间：2019年11月18日

复杂系统的背后

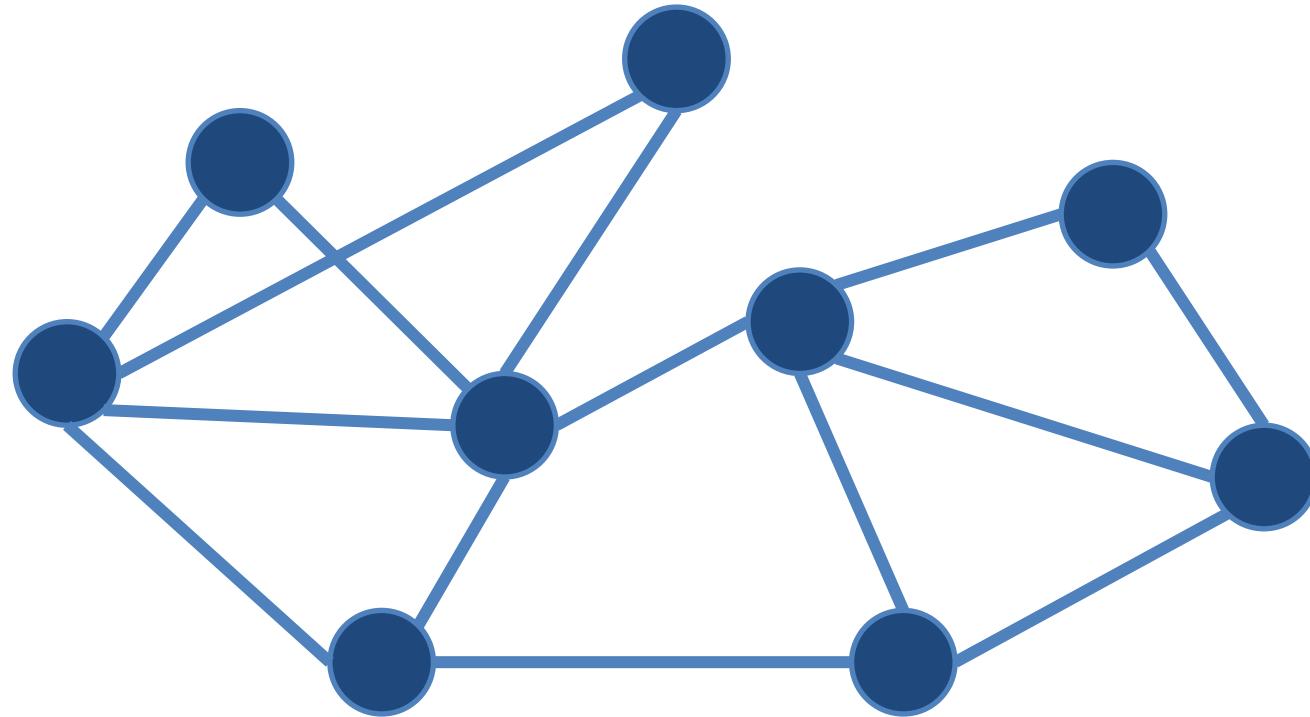
- 我们身边有很多复杂的系统
 - 截止到2019年4月1日，我们的社会包含7,579,185,859个个体，个体与个体之间有着频繁的交流
 - 截止到2018年，中国的手机用户数量突破15亿，这些用户的移动设备之间通过复杂的通信网络彼此连接
 - 截止到2011年10月，全球网站的总量突破5亿，这些网站上汇集了海量的彼此关联的信息与知识
 - 我们的大脑由860亿个彼此相连、传递信息的神经元组成
 - 我们的命体由基因与蛋白质的复杂交互关系构成
 -
- 思考：这些复杂的系统有什么共性？

如何建模上述系统



图模型：表征事物之间的相互关联

如何建模上述系统



图模型：表征事物之间的相互关联

第4.1.1节

为什么用图对数据建模

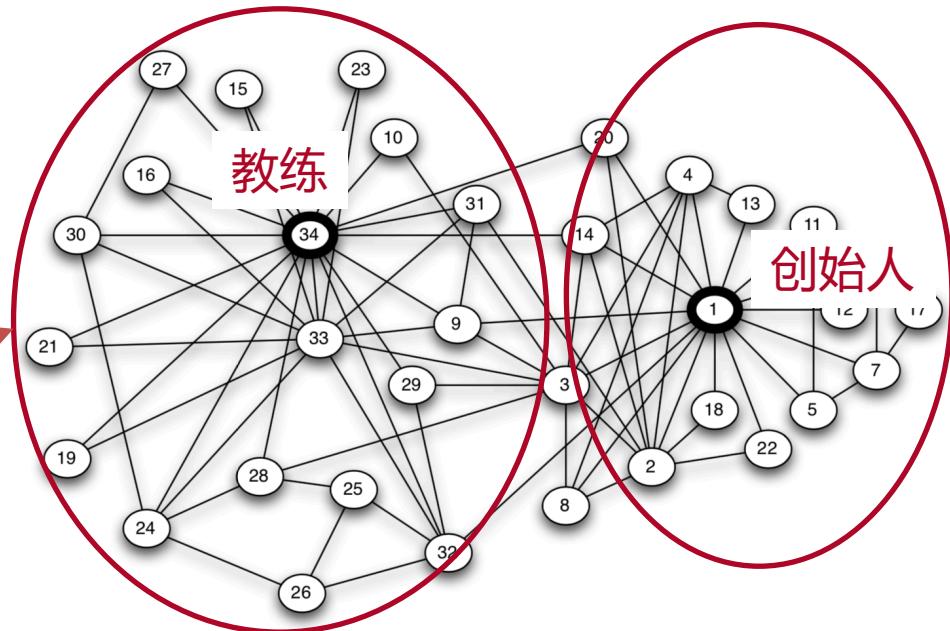
Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角

最终这个俱乐部分裂成两个对立的空手道俱乐部

结构平衡理论

一个空手道俱乐部中34个成员之间朋友关系形成的图。你能发现什么特点？

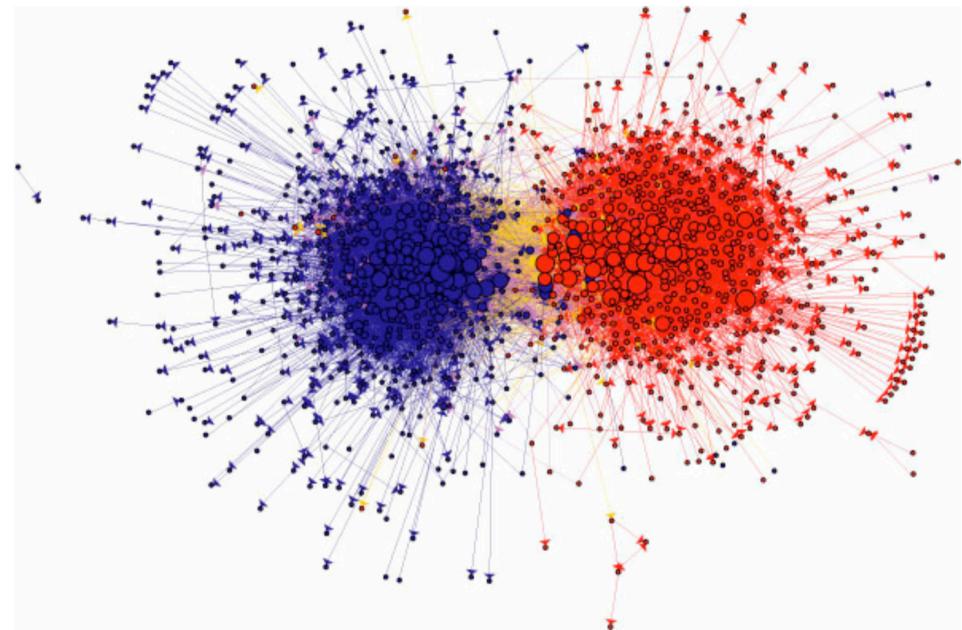


图片来源: [Easley and Kleinberg, Cambridge University Press, 2010]

Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角

2004年美国大选前，政治博客之间的链接网络
你能发现什么特点？



图片来源: [<http://www-personal.umich.edu/ladamic/img/politicalblogs.jpg>]

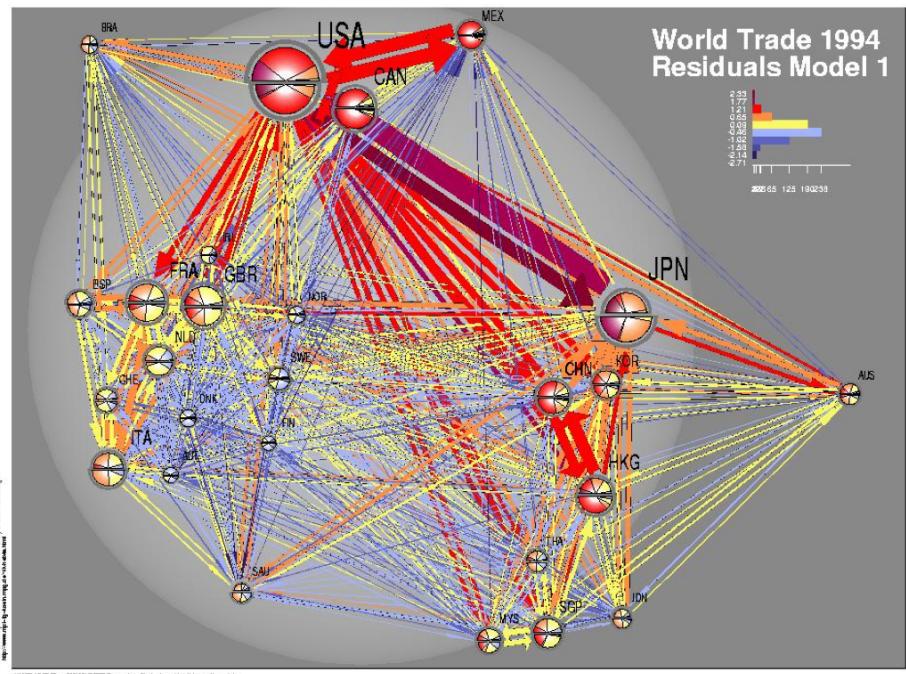
Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角
 - 图提供了一种理解个体行为的分析工具

节点大小 : 贸易总额
边的粗细 : 所连接两个国家之间的贸易额

反映了参与贸易的机会与限制

国家之间贸易网络，你能发现什么行为特点？



图片来源:

[<http://www.cmu.edu/joss/content/articles/volume4/KrempelP.html>]

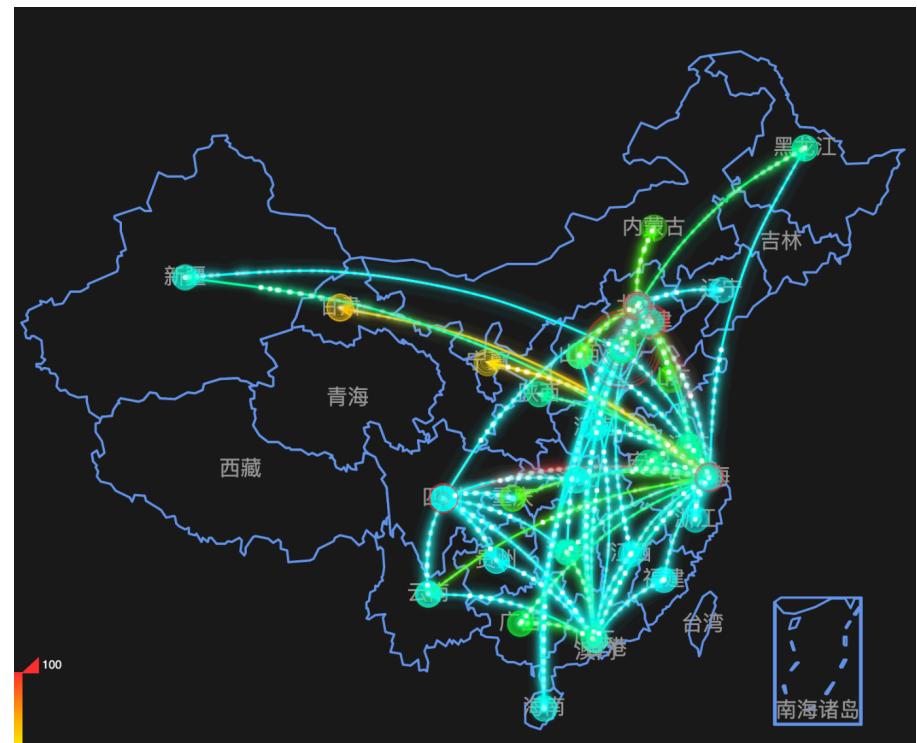
Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角
 - 图提供了一种理解个体行为的分析工具

节点：省、直辖市
边的颜色：颜色越暖表示数量越多

反映了产业迁移演进的规律

中国产业转移演进及发展现状
中国人民大学信息学院研制



图片来源: [<http://playbigdata.ruc.edu.cn/chanyeqianyi/>]

Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角
 - 图提供了一种理解个体行为的分析工具

唱和诗词

醉赠刘二十八使君

【作者】白居易

为我引杯添酒饮，与君把箸击盘歌。
诗称国手徒为尔，命压人头不奈何。
举眼风光长寂寞，满朝官职独蹉跎。
亦知合被才名折，二十三年折太多。

酬乐天扬州初逢席上见赠

【作者】刘禹锡

巴山楚水凄凉地，二十三年弃置身。
怀旧空吟闻笛赋，到乡翻似烂柯人。
沉舟侧畔千帆过，病树前头万木春。
今日听君歌一曲，暂凭杯酒长精神。

图片来源: [http://www.sohu.com/a/248677106_488227]

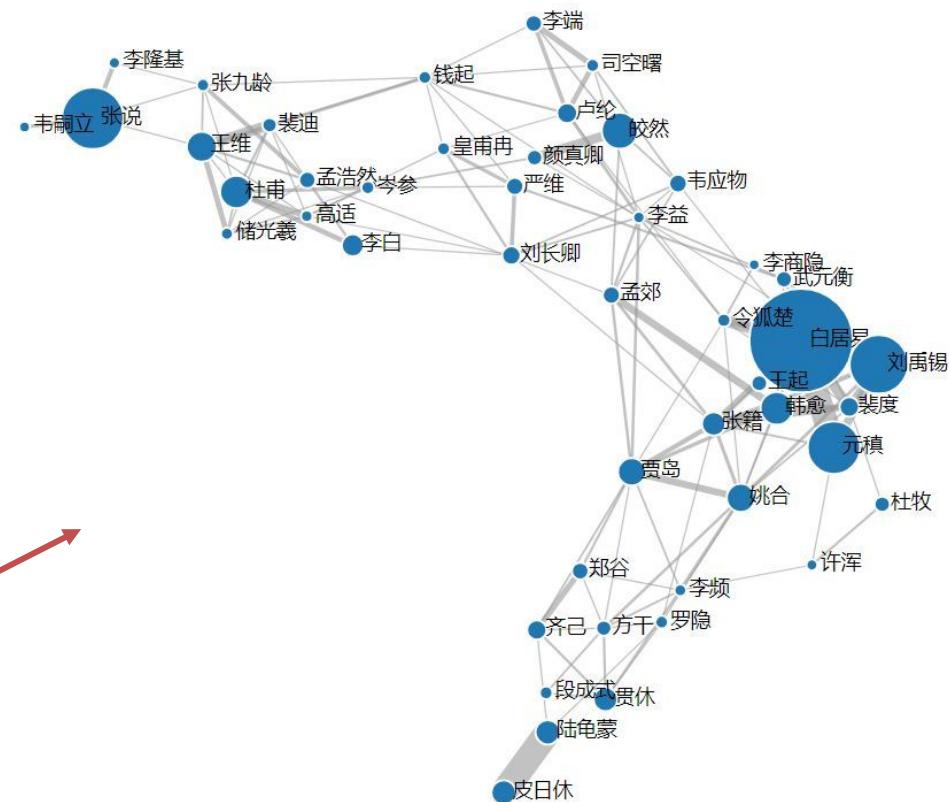
Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角
 - 图提供了一种理解个体行为的分析工具

子曰：诗可以群

元白文学集团
刘白唱和集

唐朝诗人的唱和关系图：节点大小表示和诗数量



图片来源: [http://www.sohu.com/a/248677106_488227]

Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角
 - 图提供了一种理解个体行为的分析工具

唱和诗词

离思五首 · 其四

作者：元稹

曾经沧海难为水，除却巫山不是云。
取次花丛懒回顾，半缘修道半缘君。

醉赠刘二十八使君

【作者】白居易

晨起临风一惆怅，通川湓水断相闻。
不知忆我因何事，昨夜三更梦见君。

酬乐天频梦微之

【作者】元稹

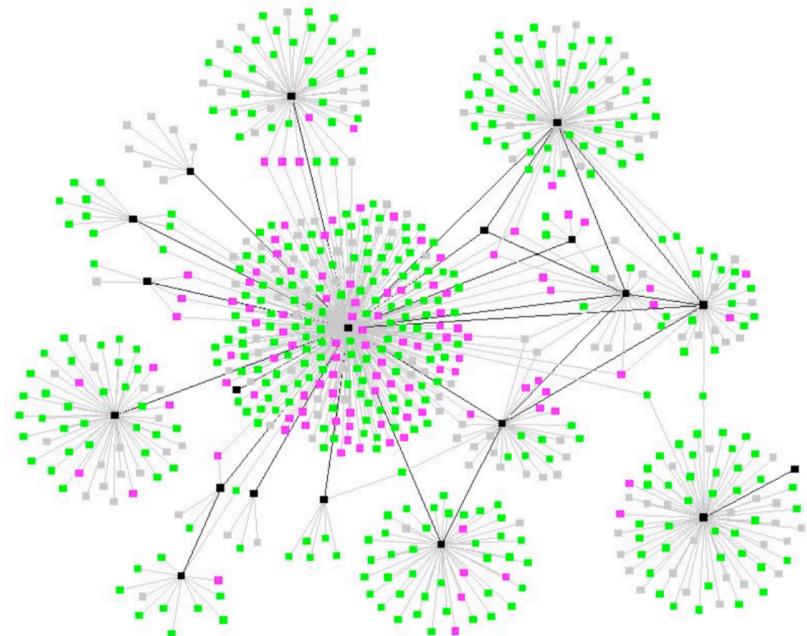
山水万重书断绝，念君怜我梦相闻。
我今因病魂颠倒，唯梦闲人不梦君。

图片来源: [http://www.sohu.com/a/248677106_488227]

Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角
 - 图提供了一种理解个体行为的分析工具
 - 图提供了一种解释信息传播的直观方法

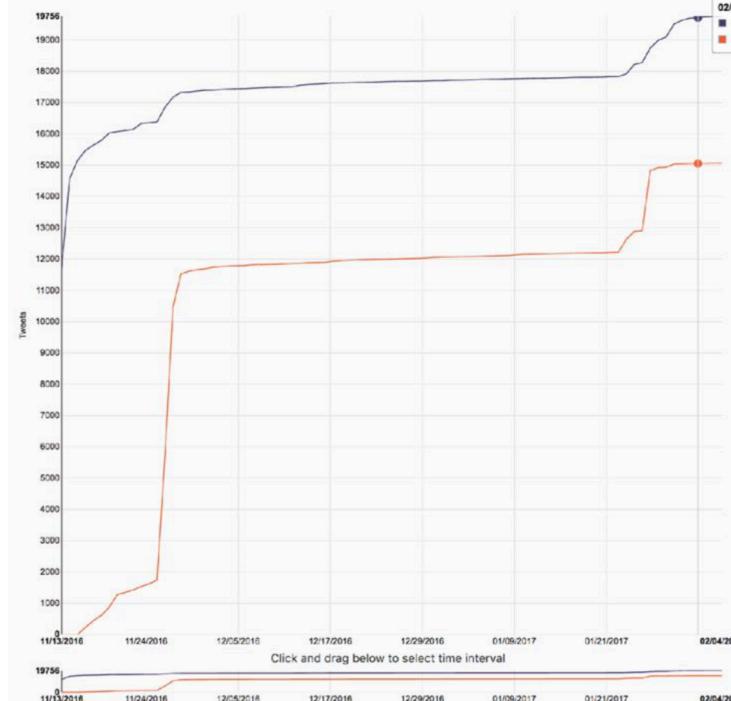
一次肺结核爆发的扩散过程，与信息传播很类似



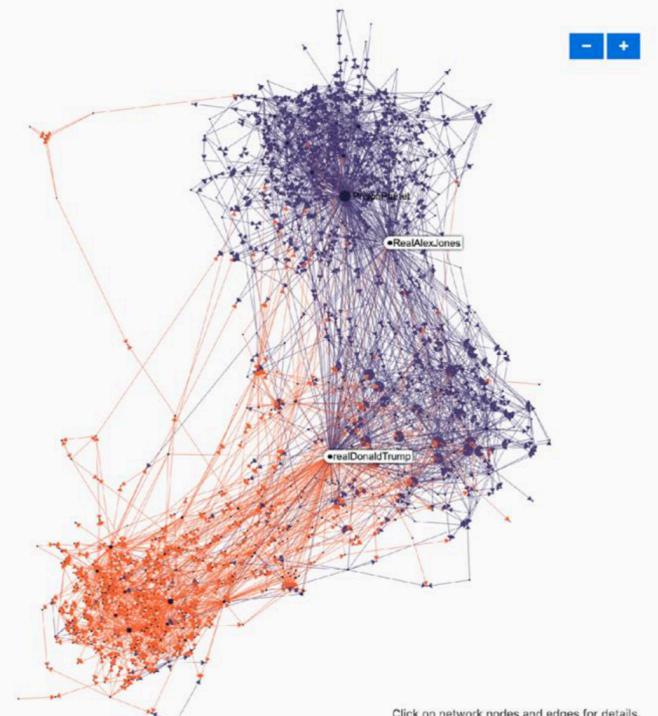
图片来源: [Andre et al., American Journal of Public Health, 2007]

Why Graph?

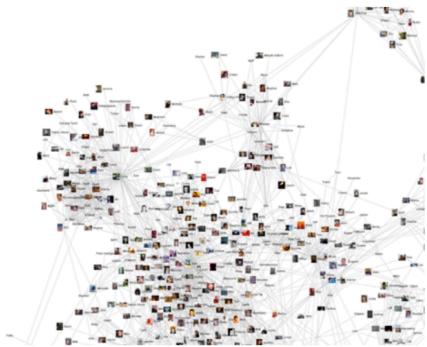
- 为什么使用图模型对数据建模?
 - 图提供了一种观



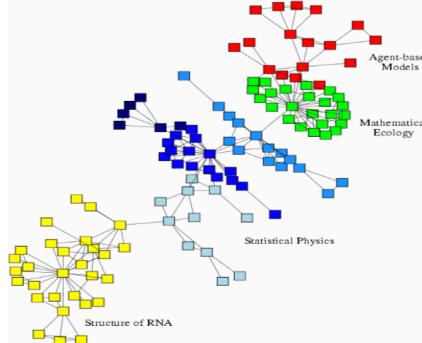
谣言与辟谣信息的传播：
<http://hoaxy.iuni.iu.edu/>



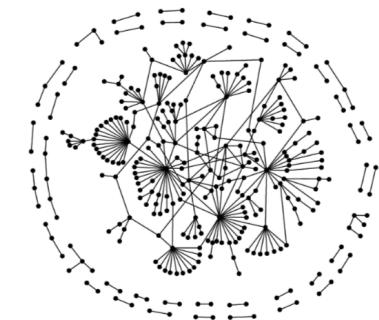
丰富多彩的图数据



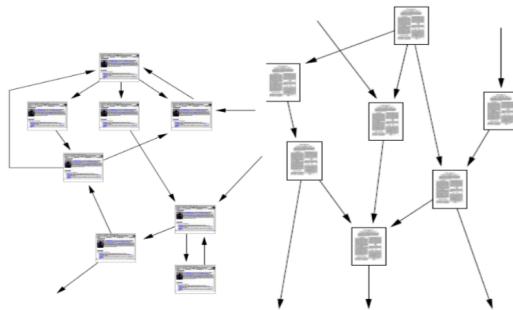
Social networks



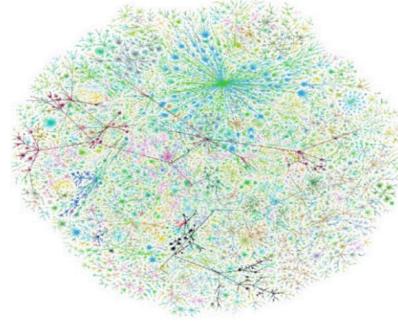
Economic networks



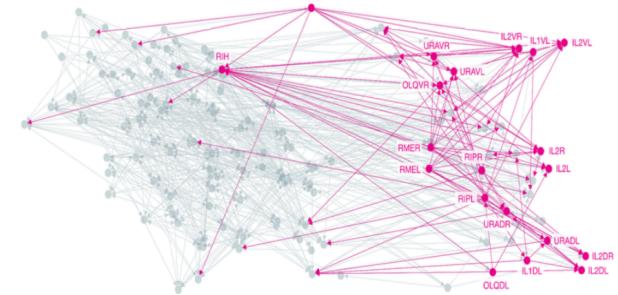
Communication graphs



Information networks:
Web & citations



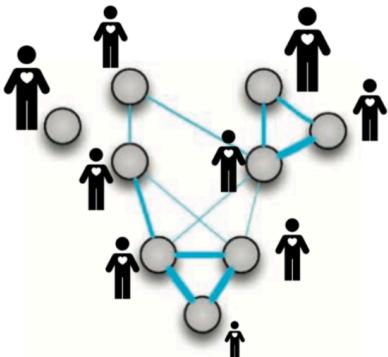
Internet



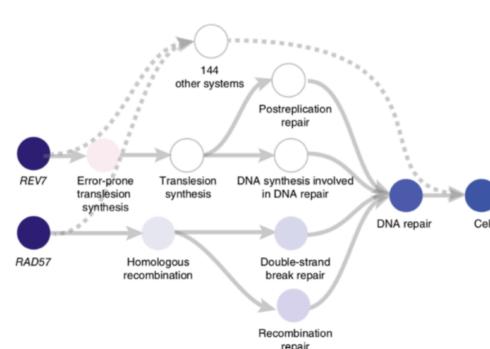
Networks of neurons

图片来源: [Jure Leskovec, CS224W, Stanford University]

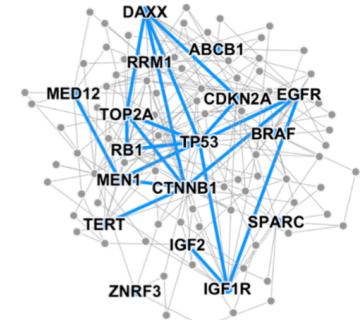
丰富多彩的图数据



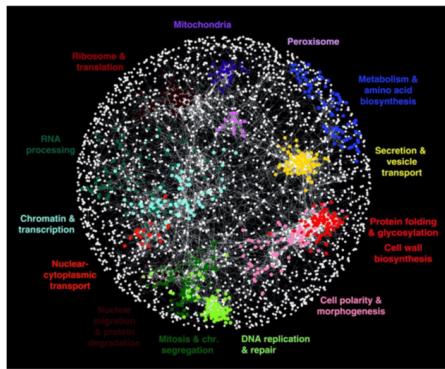
Patient networks



Hierarchies of cell systems



Disease pathways



Genetic interaction networks



Gene co-expression networks

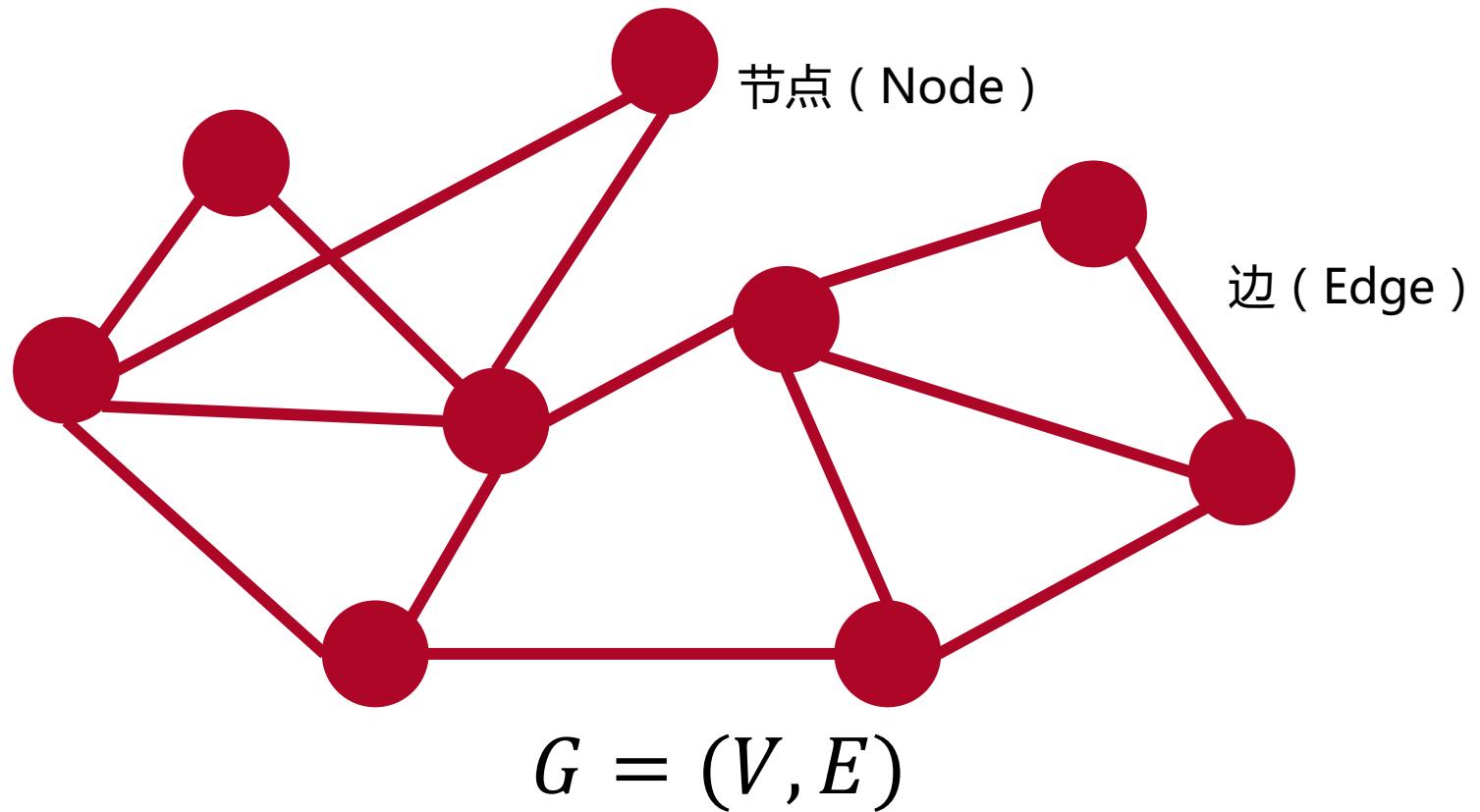


Cell-cell similarity networks

第4.1.2节

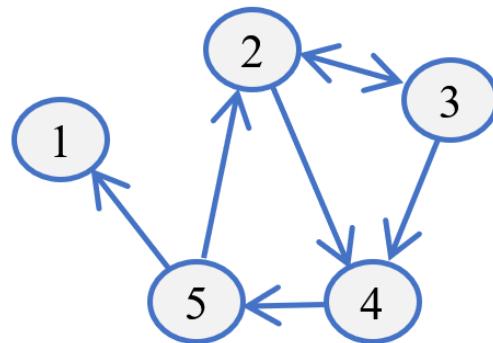
图论的基本概念

图的形式化定义



图的形式化定义

- 图可以使用三种方式进行表示
 - 邻接矩阵(Adjacency Matrix)
 - 边列表(Edge List)
 - 邻接关系列表(Adjacency List)。



一个简单的有向图

“邻接矩阵”表示

$$\text{ADJ} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

“邻接关系列表”表示

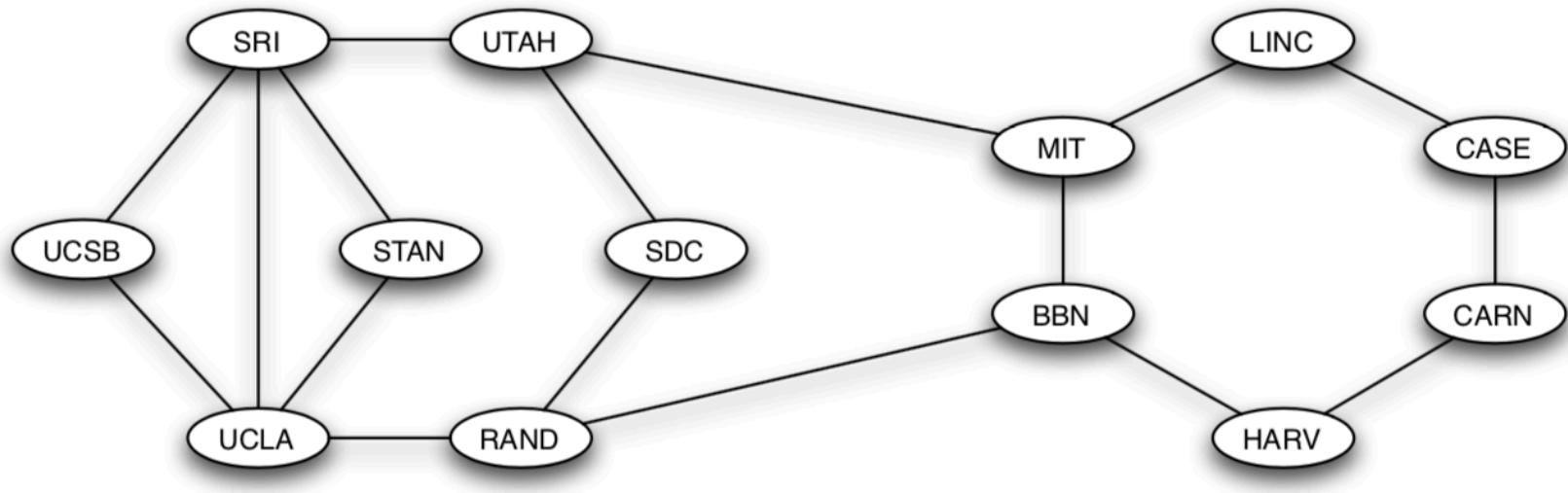
1: 2: 3 4 3: 2 4 4: 5 5: 1 2
--

“边列表”表示

2,3 2,4 3,2 3,4 4,5 5,1 5,2

图的基本定义

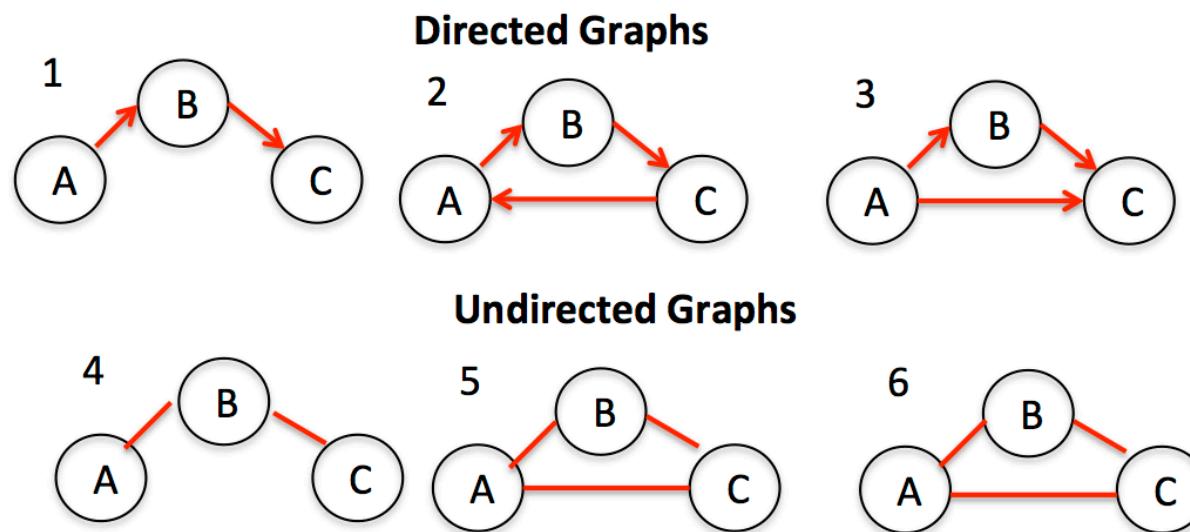
- 练习：请写出下图的邻接矩阵、边列表与邻接关系列表



1970年代的Internet

关系的对称性

- 无向图与有向图 (Undirected Graph vs. Directed Graph)
 - 一条边两端的两个节点是否具有对称关系



本讲主要探讨无向图

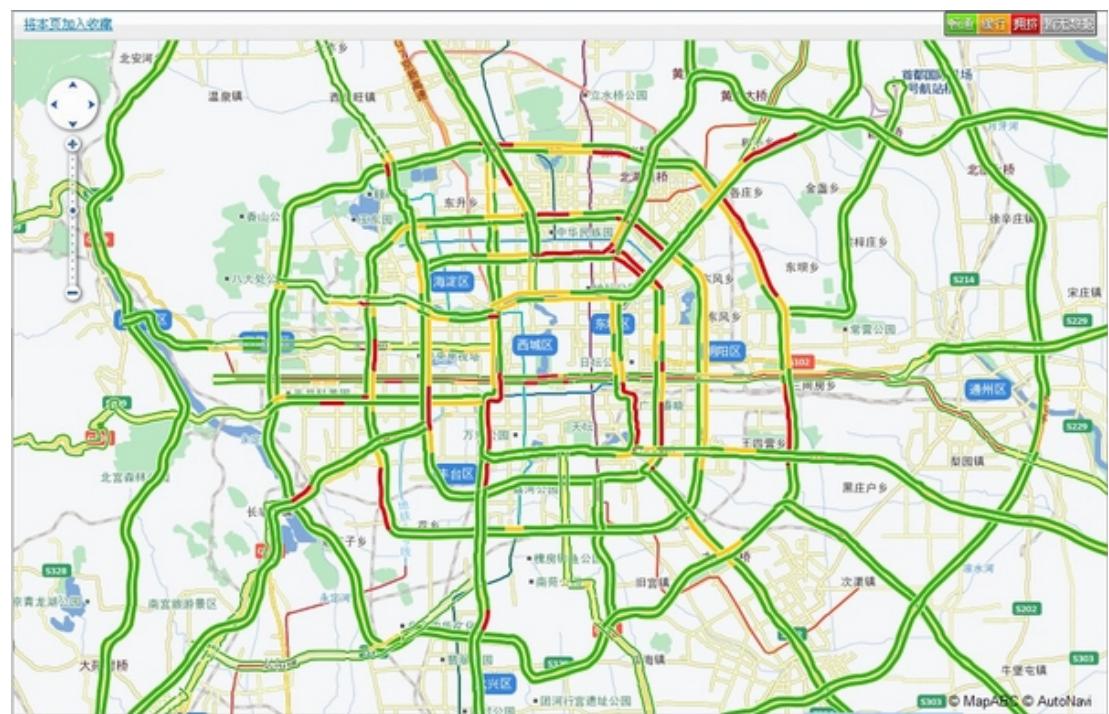
边的权重

- 图上的每一条边 e 关联一个数字 $w(e)$ ，用来表示边的重要性或成本
- 例：道路网中的权重

节点：路口

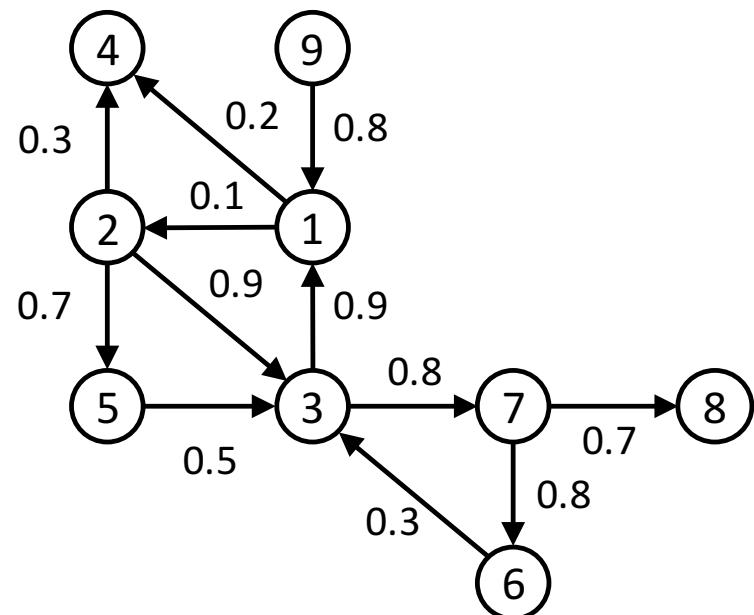
边：道路

权重：拥堵情况



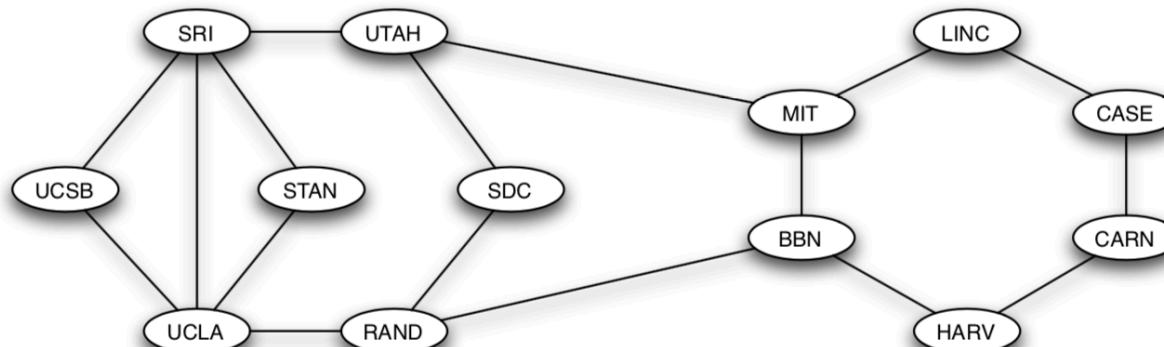
边的权重

- 图上的每一条边 e 关联一个数字 $w(e)$ ，用来表示边的重要性或成本
 - 例：信息传播中的权重
-
- Independent Cascade (IC) 模型
 - 假设用户 u 到 v 之间有一个概率值，称为影响概率
 - 概率值越大， v 相信 u 传播消息的可能性就越大
 - 概率值可以通过 u 和 v 的互动历史学习得到



路径

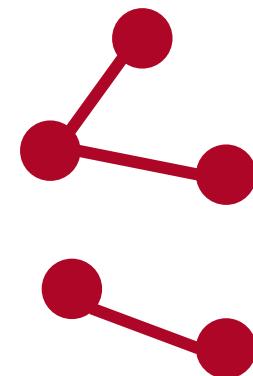
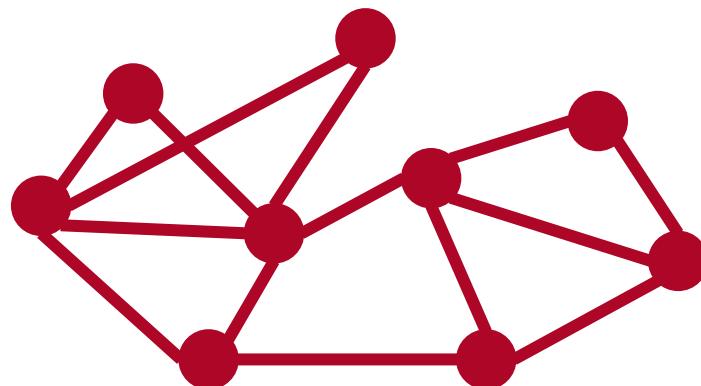
- 路径：图上节点的序列，序列中任意两个相邻的节点都有边相连
 - Path $p = (v_0, v_1, \dots, v_m)$ where any $(v_i, v_{i+1}) \in E$
 - 简单路径：不包含重复节点的路径
- 环：起点与终点相同的路径
 - $p = (v_0, v_1, \dots, v_m, v_0), (v_i, v_{i+1}) \in E$



- 请写出UCSB到MIT的路径
- 该图是否存在环？
- 分析该图存在环的利与弊

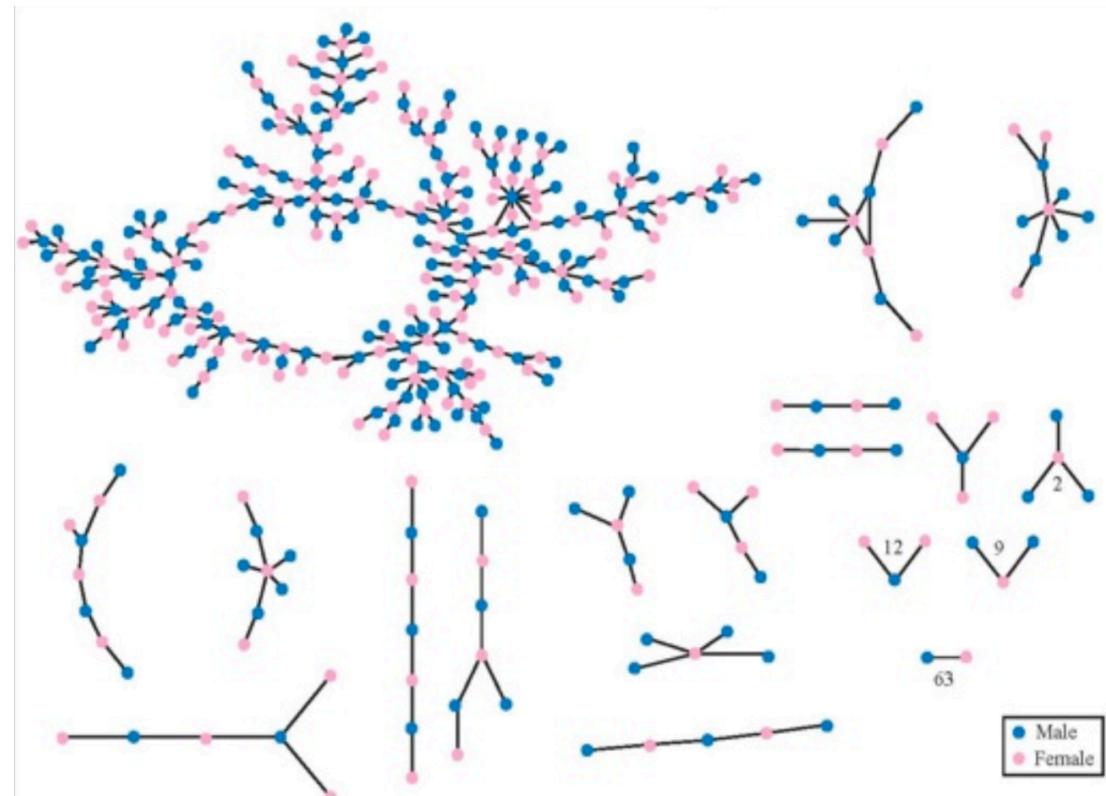
连通图与连通分支

- 如果一个图中任意两点之间都有路径相通，则称此图为**连通图**
- 连通分支/连通分量(Connected Component)
 - 给定图 G ，它的子图 G' 是 G 的连通分支，如果满足以下两个条件：
 - G' 是连通图，即 G' 中任意两个点有路径相连 - **连通性**
 - G' 不是其它任何满足条件I的图的子集 - **独立性**



连通图与连通分支

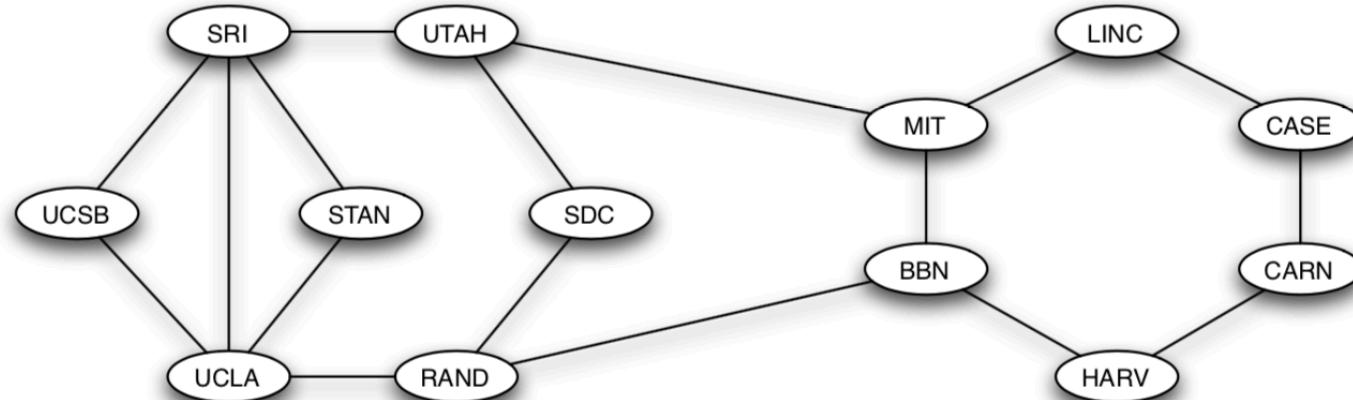
- 例：美国高中生恋爱关系图（边代表二人在18个月内恋爱过）



图片来源: [Bearman et al., American Journal of Sociology, 2004]

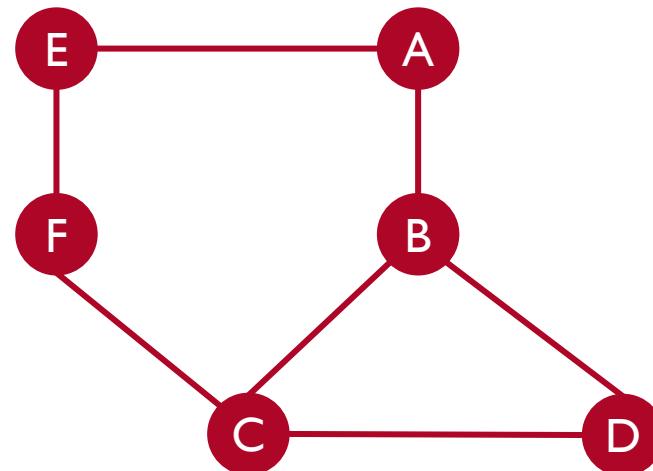
最短路径与距离

- 定义某条路径 p 的长度为它所包含边的个数
 - 例：路径MIT, BBN, RAND, UCLA的长度为3
- 定义图上两点的距离为它们之间最短路径的长度
 - 例：LINC与SRI之间的距离为3



最短路径与距离

- 思考题：如果节点Z在节点X和Y所有的最短路径上，则称Z为X和Y的关键节点（Pivot node），其中Z与X和Y均不重合
 - 请构造一个图：每个节点均为至少一对节点的关键节点
 - 请构造一个图：每个节点均为至少两对节点的关键节点
 - 请构造一个图：该图中至少包含四个节点，并存在一个节点X，它是图中所有节点对的关键节点（不包含X）

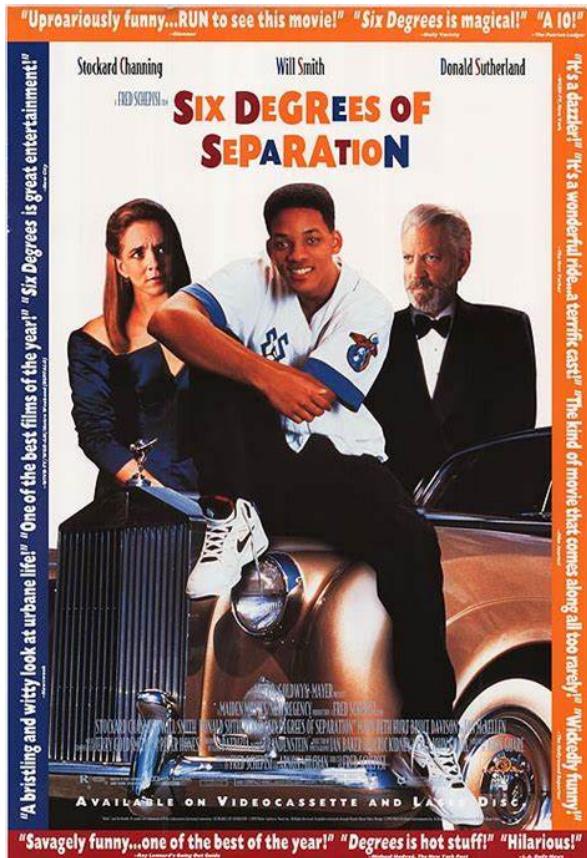


第4.1.3节

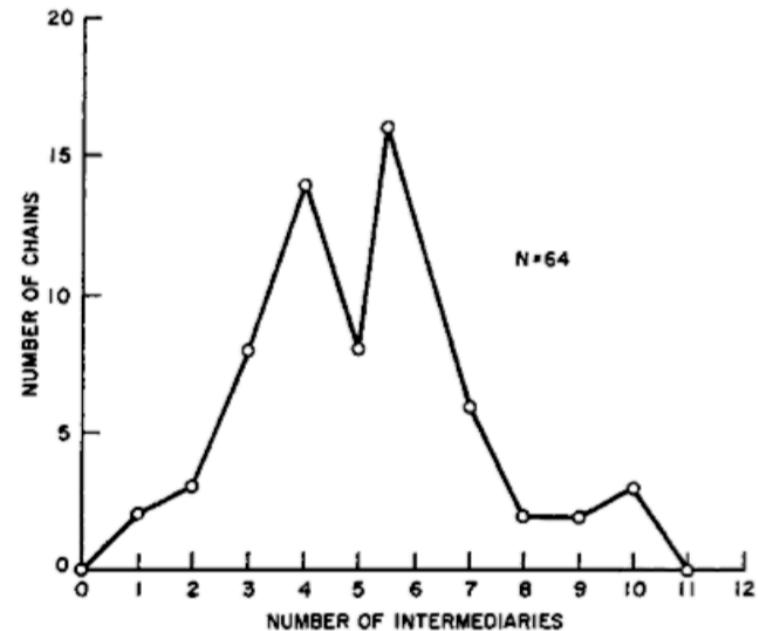
使用图能解决什么问题

小世界 Small World

- 六度分隔理论 (Six Degrees of Separation)



“在这个世界上，任意两个人之间，只隔着六个人”



小世界 Small World

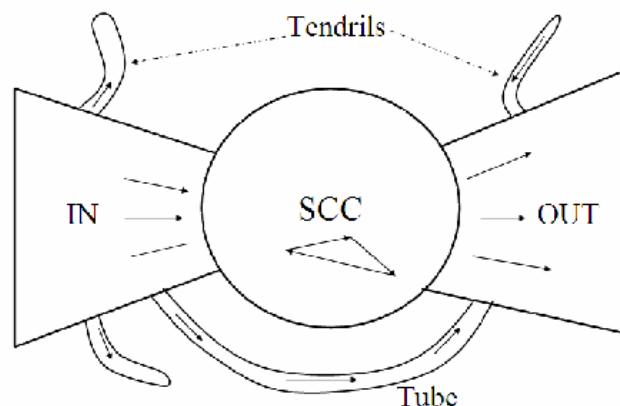
- 四度分隔！



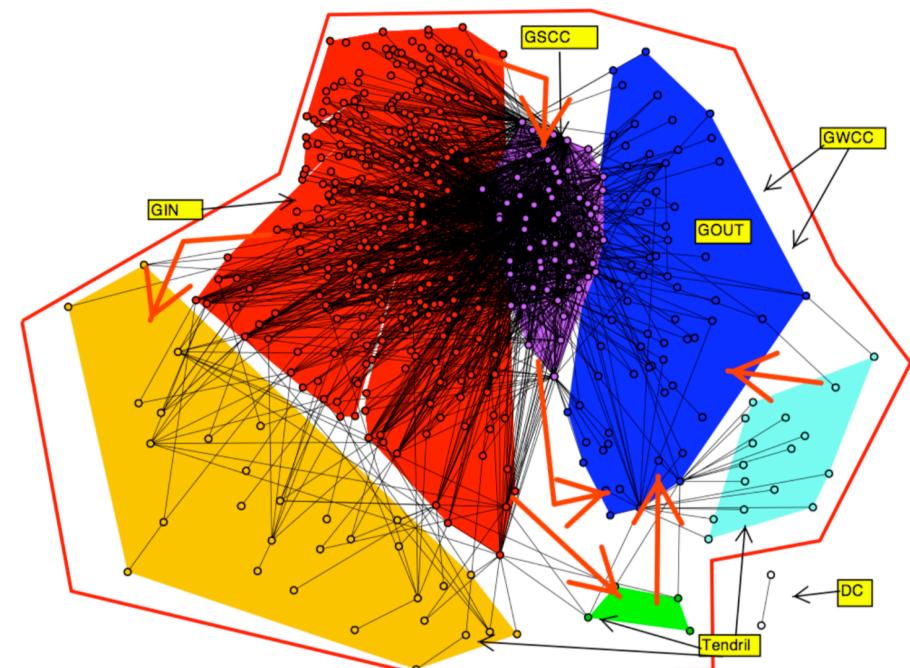
4-degrees of separation [Backstrom-Boldi-Rosa-Ugander-Vigna, 2011]

结构通常决定了地位

金融机构之间的借贷网络，你能发现什么结构特点？



领结结构
The bowtie Structure



图片来源: [Bech and Atalay, Technical Report, 2008]

图数据分析Python第三方类库

NetworkX

Stable (notes)

2.2 — September 2018

[download](#) | [doc](#) | [pdf](#)

Latest (notes)

2.3 development

[github](#) | [doc](#) | [pdf](#)

Archive

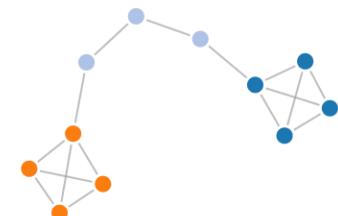
Contact

[Mailing list](#)

[Issue tracker](#)

Software for complex networks

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.



Features

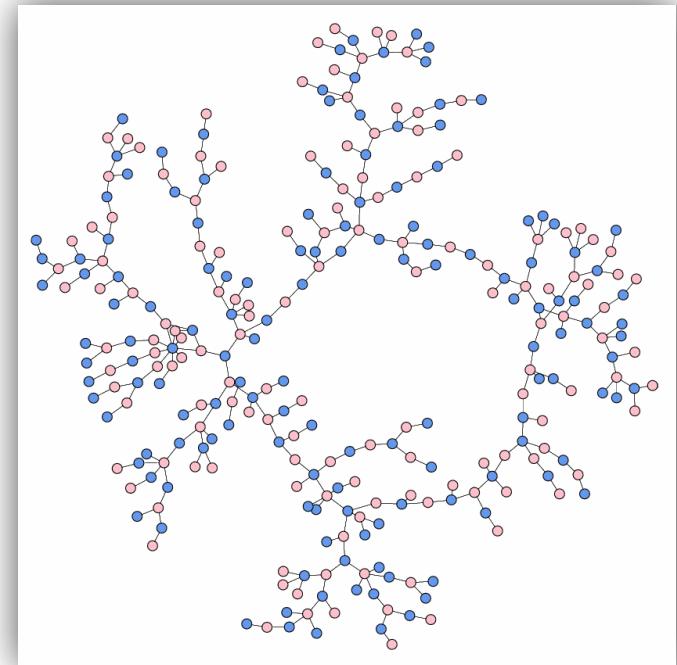
- Data structures for graphs, digraphs, and multigraphs
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g., text, images, XML records)

节点中心度 (Node Centrality) 分析

- 在网络中，不同节点的“地位”是不平等的
 - 例子：美国高中生恋爱关系图
 - 边表示18个月内谈过恋爱



- 思考：
 - 你觉得哪些节点更重要？
 - 你怎么解释这种重要性？



图片来源: [Bearman et al., American Journal of Sociology, 2004]

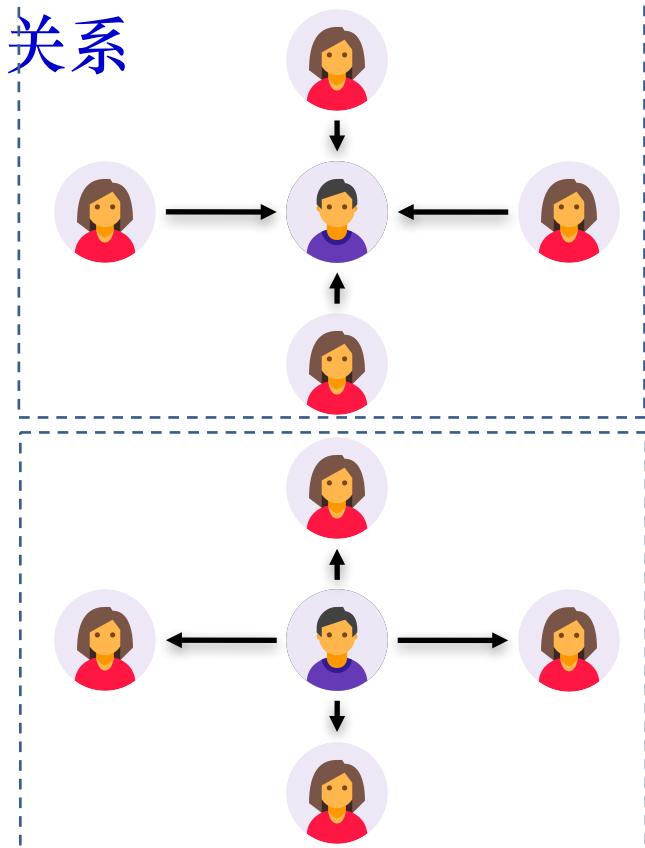
节点中心度 (Node Centrality) 分析

- 在网络中，不同节点的“地位”是不平等的
 - 例子：美国高中生恋爱关系图
 - 如果定义有向边：“追求”关系

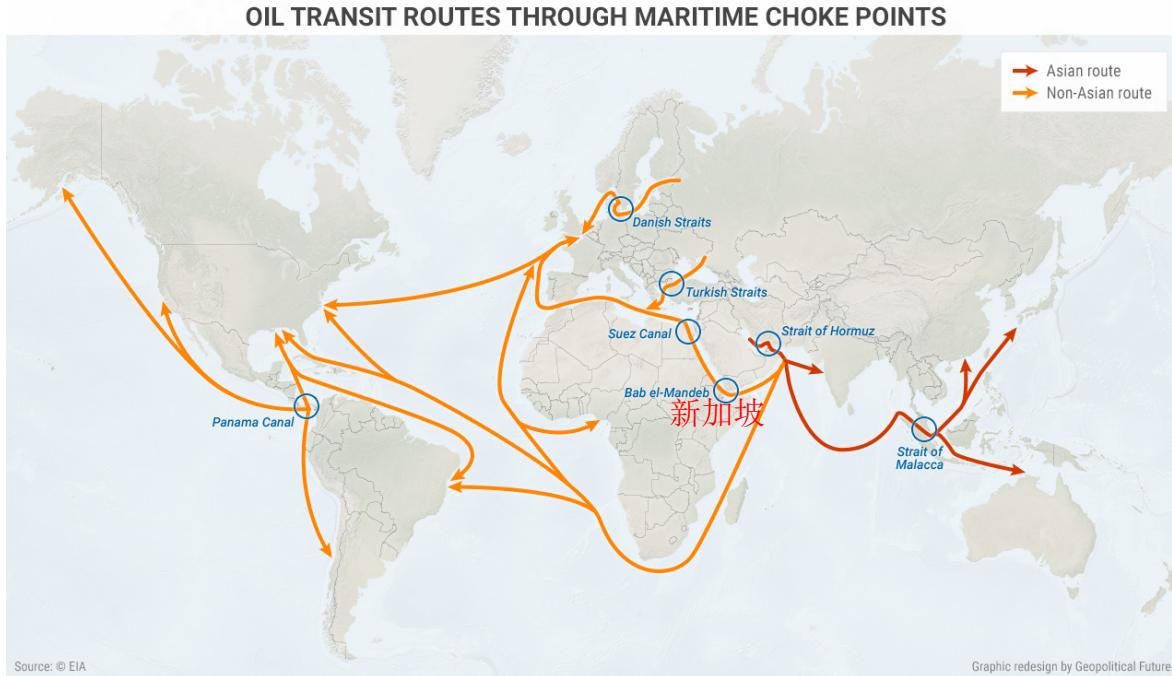


有向图

- 思考：
 - 右边两图中男生的重要性一样吗？
 - 你怎么解释这种重要性？



思考：新加坡为什么成为全球贸易港？



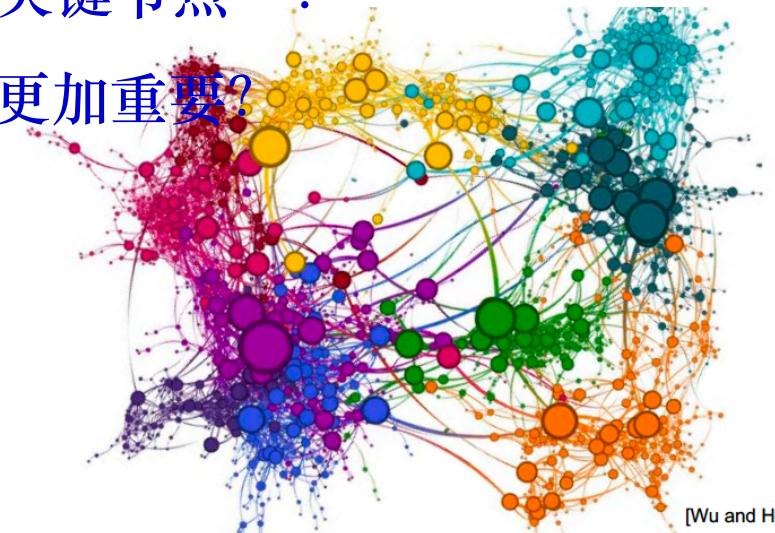
波斯湾石油运输航路图

The Persian Gulf is a leading oil-producing region, accounting for 30% of global supply. Meanwhile, **East Asia** is a major oil-consuming region and accounts for **85%** of the Persian Gulf's exports, according to the Energy Information Administration (EIA)

地缘政治
与
节点中心度

节点中心度 (Node Centrality)

- 节点中心度
 - 给定一个图，哪些节点更重要或更有影响力？
- 为什么要研究节点中心度 (Node Centrality) ?
 - 在社交网络中，每个人的影响力如何？
 - 在道路网络中，有哪些“关键节点”？
 - 在Web网络中，哪些网页更加重要？
 - 你能想到其它吗？



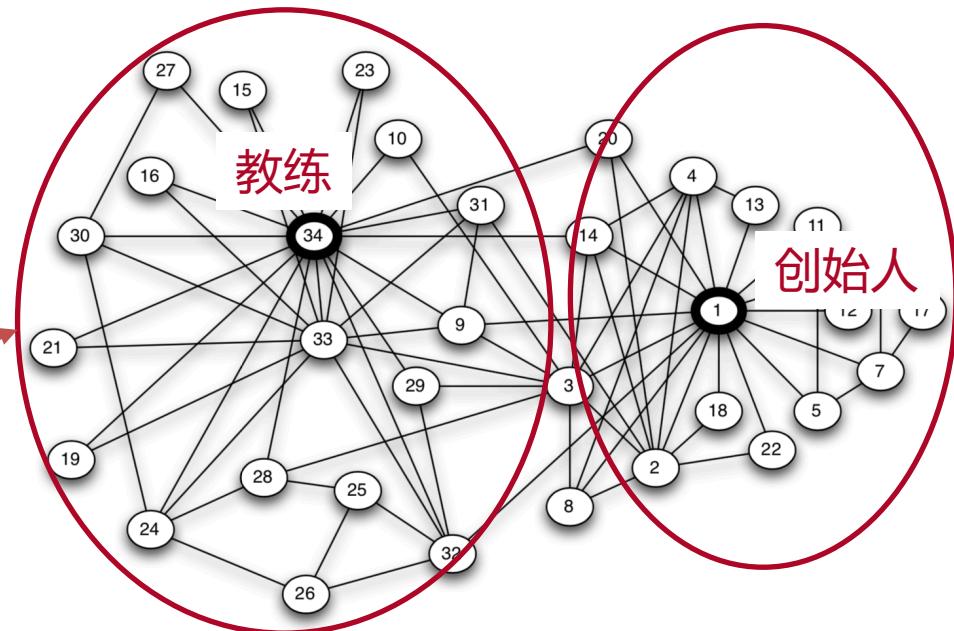
Why Graph?

- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据结构特征的视角

最终这个俱乐部分裂成两个对立的空手道俱乐部

结构平衡理论

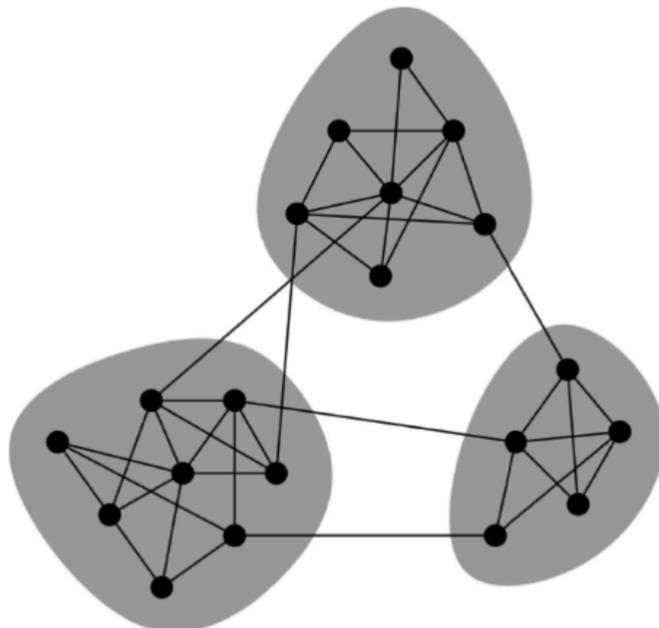
一个空手道俱乐部中34个成员之间朋友关系形成的图。你能发现什么特点?



图片来源: [Easley and Kleinberg, Cambridge University Press, 2010]

真实网络中为什么能形成社区？

- 在分析数据之前，首先要对数据的性质有足够的认识
- 我们认为真实网络（如社交网络）长这样



果真如此吗？如果真是的话，为什么会是这样？

Mark Granovetter的研究

- 上世纪60年代末，Mark Granovetter在做他博士论文研究
 - 研究题目：人们是如何找到新的工作的
 - 发现1：人们通常是通过人际关系获取了新工作的信息
 - 发现2：获取新工作的人际关系通常是“点头之交”（casual acquaintances）而并非“亲密好友”（close friends）
- 这个发现很让人惊讶
 - 一般认为，亲密好友对你的帮助应该大于点头之交的熟人。

为什么？

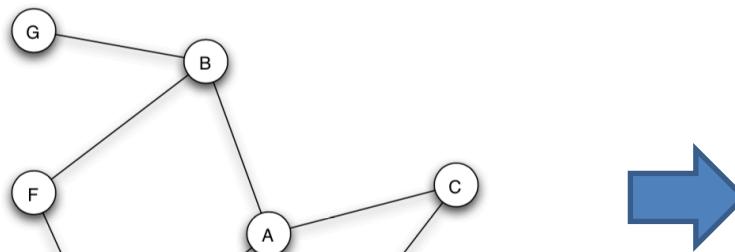


三元闭包 (Triadic Closure)

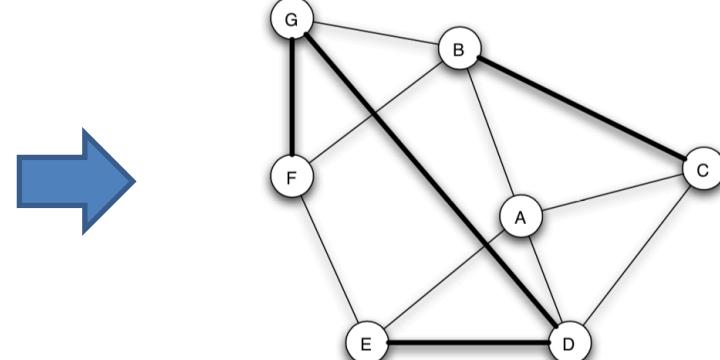
- 三元闭包 (Triadic Closure)
 - 如果两个人在网络中有共同的好友，他们成为好友的几率也会提升
- 几点原因：如果B和C都有共同的好友A，那么
 - B更有可能遇到C——因为他们都与A有交集
 - B和C更有可能互相信任——因为他们有共同的好友
 - A更有可能介绍B和C认识

三元闭包 (Triadic Closure)

- 量化指标: clustering coefficient
 - 节点A的clustering coefficient是A任意两个邻居是好友的概率



clustering coefficient of A = 1/6

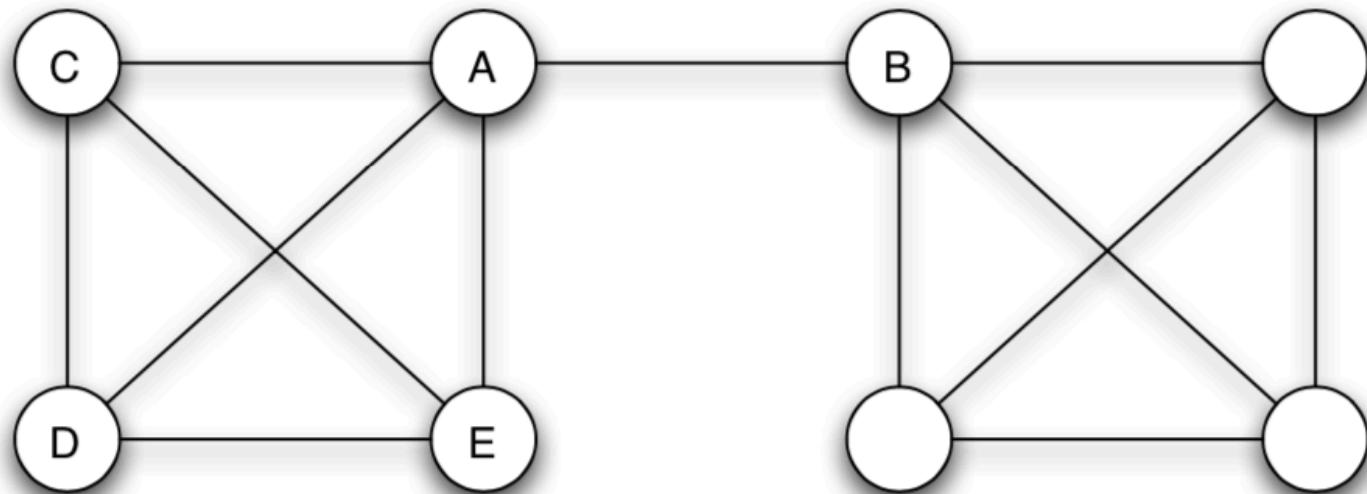


clustering coefficient of A = 1/2

- [Bearman and Moody]青少年女孩调查 : clustering coefficient越低越易自杀

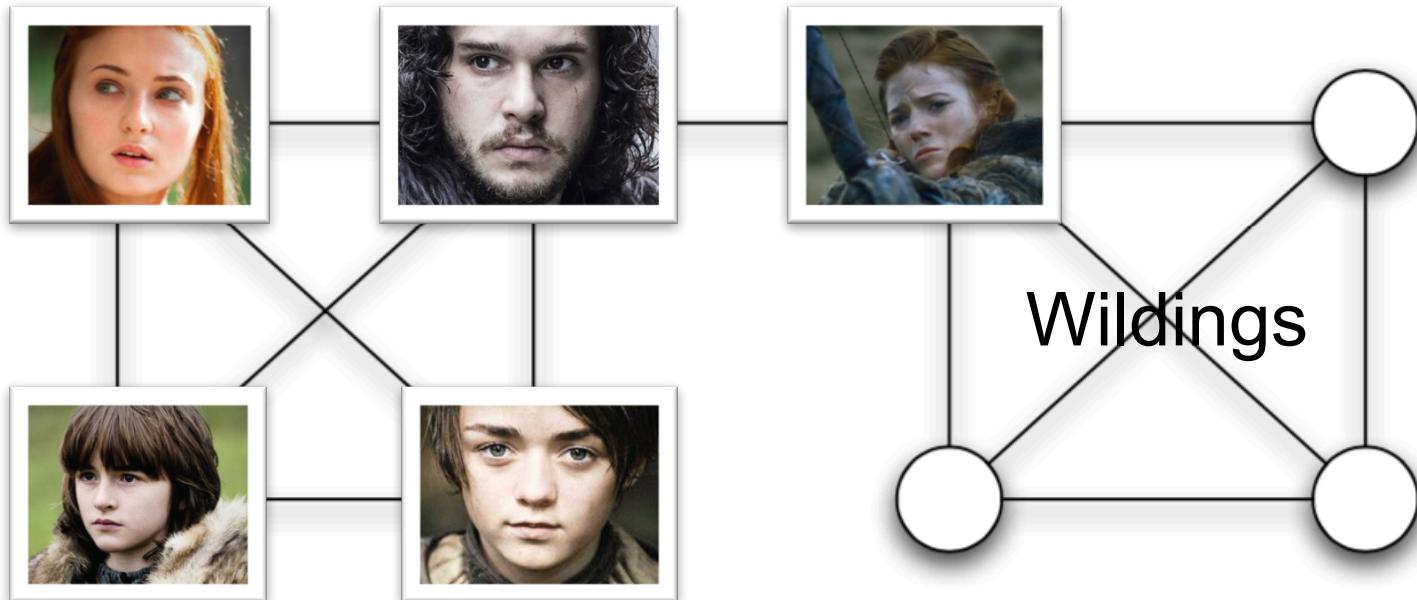
强关系 vs. 弱关系

- 怎么解释下图中A和B之间的关系？



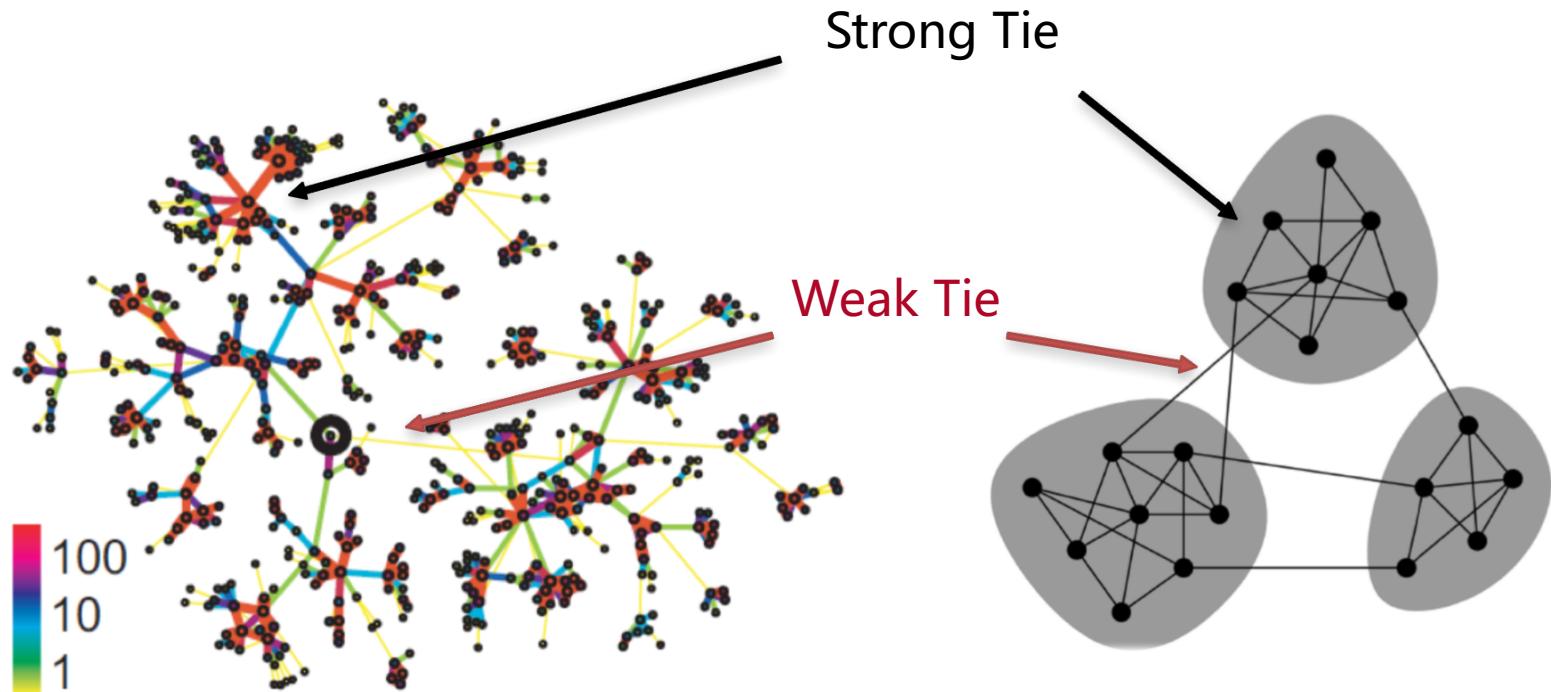
强关系 vs. 弱关系

- 怎么解释下图中A和B之间的关系？



真实数据中的强弱关系

- [Onnela et al. 2007] 测量了电话通信网络
 - 边的强弱表示打电话的次数

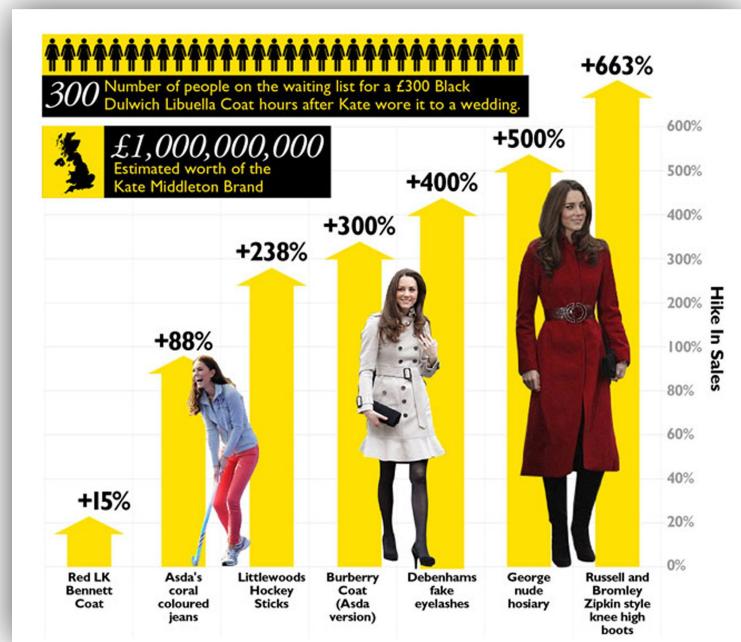


图片来源: [Onnela et al. 2007]

社交影响力 (Social Influence)

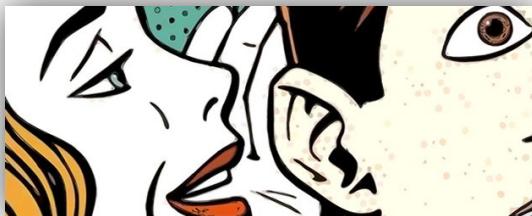
- 社会影响力 Social Influence
 - 凯特王妃效应：穿衣品味影响力拉动百亿时装消费
 - 最佳的广告推广形式来自熟人之间的口耳相传

Recommendations from people known (90%)

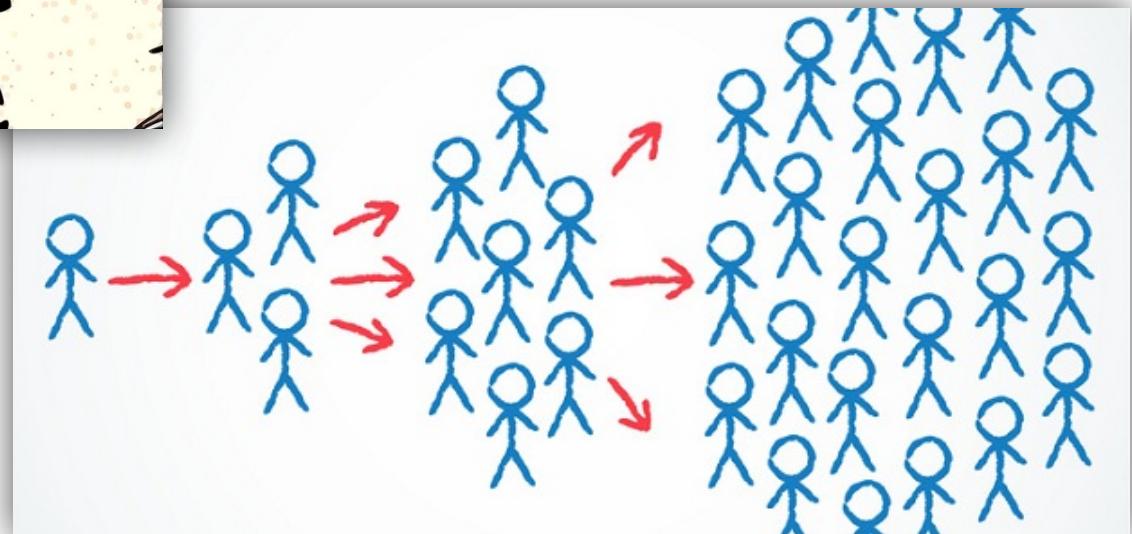


社交影响力 (Social Influence)

- Social influence occurs when a person's emotions, opinions, or behaviors are affected by others

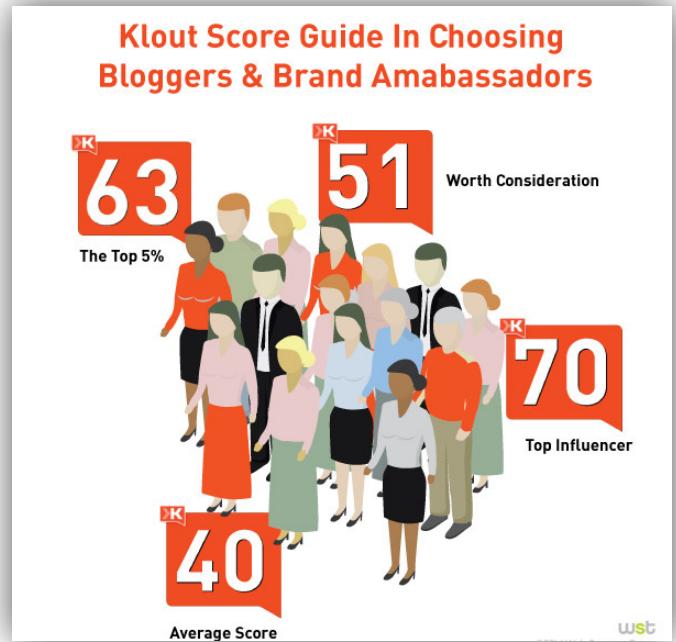
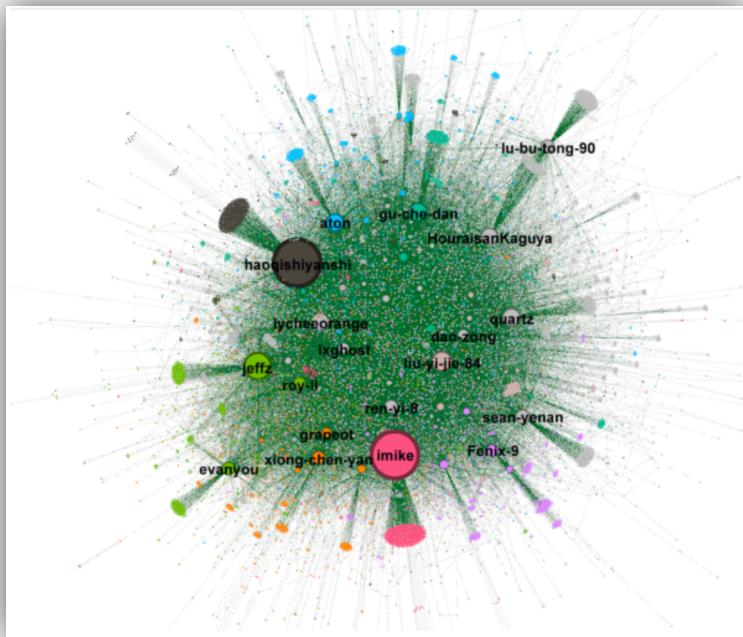


主要应用
病毒营销



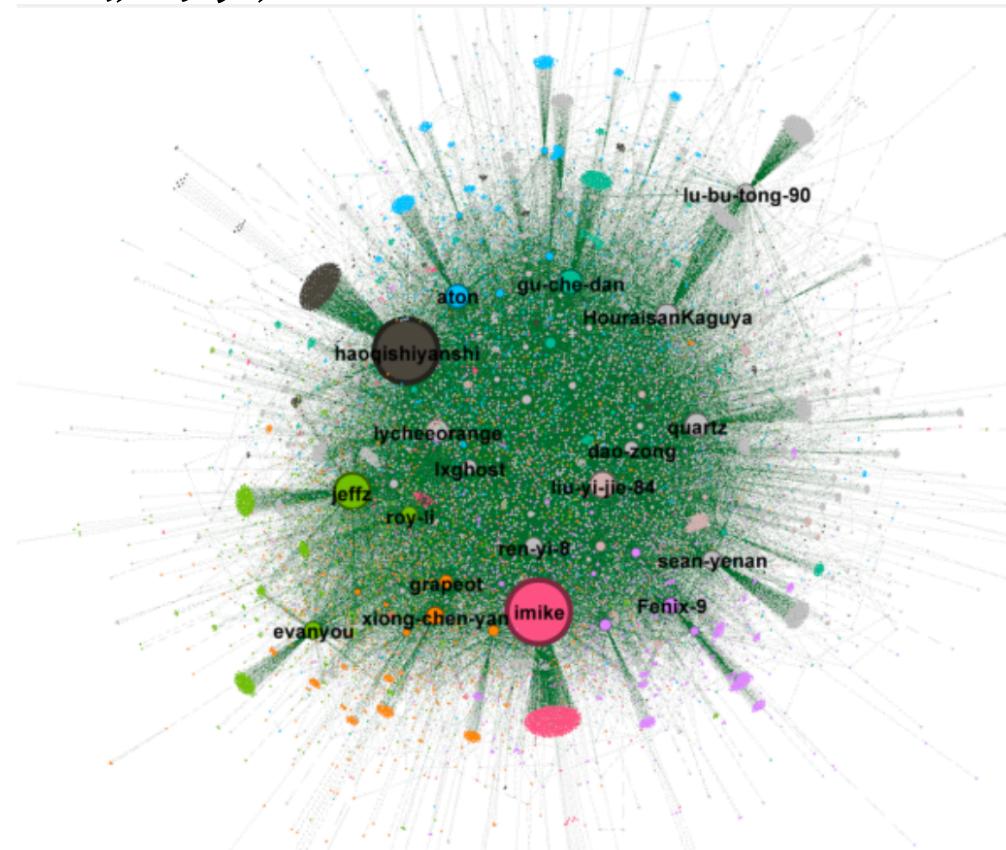
社交影响力 (Social Influence)

- 在线社交网络迅猛发展
- 在线社交影响力：度量用户在社交网络上对
其听众能够影响



信息在社交网络的传播

- 一夜之间从知乎小透明到万赞。。。需要多少个大V扶持呢？



<https://news-at.zhihu.com/story/9305265>

社交影响力 (Social Influence)

- 应用：谣言与虚假信息防控



世界经济论坛十大小危机 (2014)

- 1. Rising societal tensions in the Middle East and North Africa
- 2. Widening income disparities
- 3. Persistent structural unemployment
- ...
- *10. The rapid spread of misinformation online*

信息在社交网络的传播

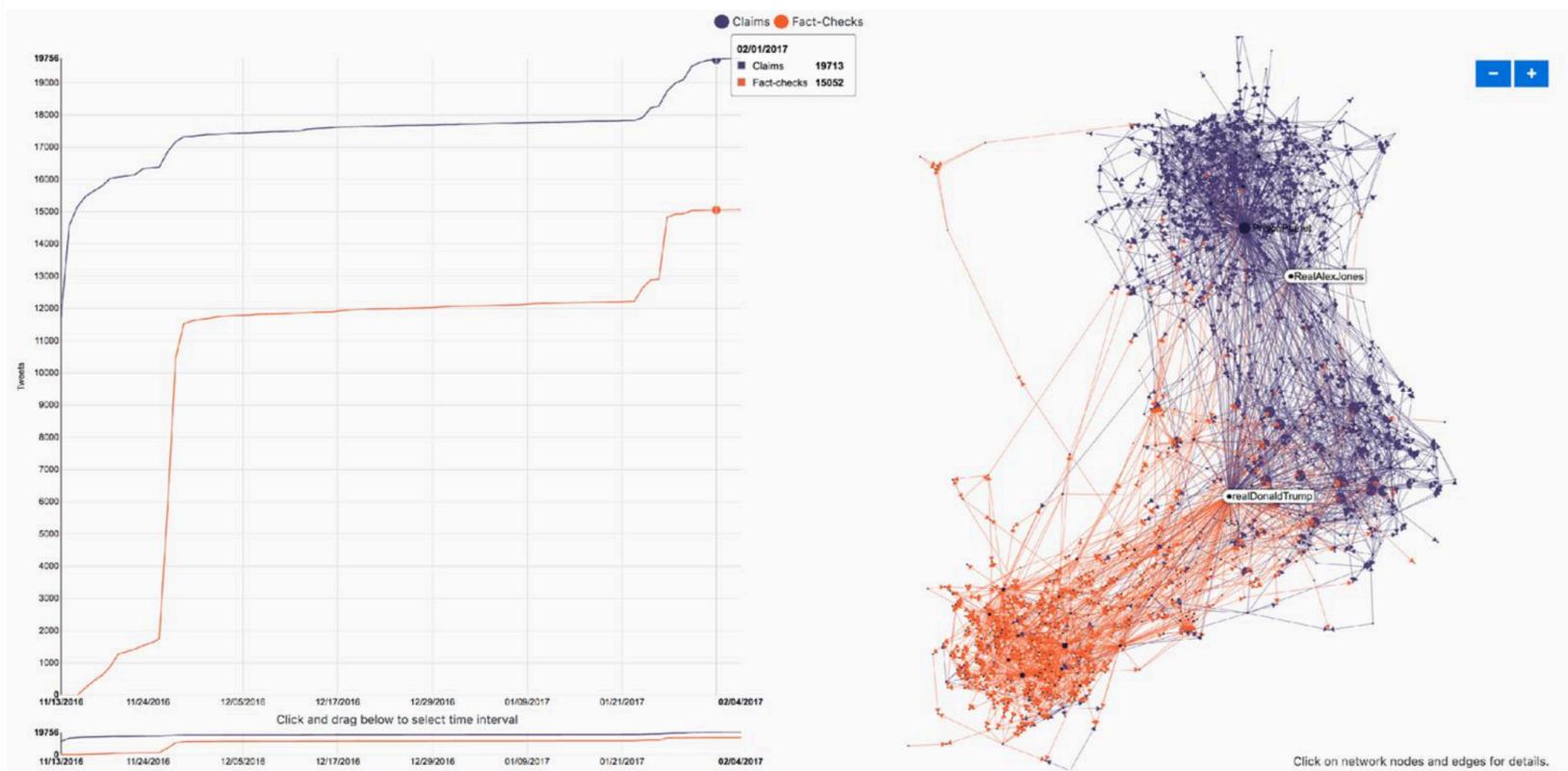


- 2016年底，Facebook上的一篇帖子引爆网络：
- 3 million votes in presidential election cast by illegal aliens
- 多个事实核查机构发布核查信息

问题：虚假/核查信息在社交网络上是如何传播的？

信息在社交网络的传播

- 美国网站：<http://hoaxy.iuni.iu.edu/>



Hoaxy is a tool that **visualizes the spread** of articles online. Articles can be found on Twitter, or in a corpus of claims and related fact checking.

社交影响力 (Social Influence)

- 应用：谣言与虚假信息防控
- 披萨门事件：
 - 维基解密泄露希拉里竞选团队主任John Podesta邮件
 - 社交媒体Reddit用户从泄密邮件中搜罗证据，并发现有关Pizza、Dinner Plan的词汇
 - 单词cheese pizza被联想为恋童child pornography
 - 有人通过邮件定位到一家乒乓彗星披萨店，认定民主党高层参与了娈童案
 - 事件通过PizzaGate标签迅速在社交媒体上滚雪球
 - 一位28岁男子来到乒乓彗星披萨店，枪击服务生未遂

整个过程发生在**不到一个月**的时间内

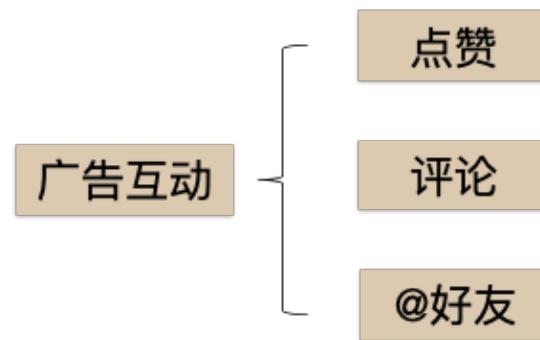
真实应用实例

- 应用场景：微信朋友圈广告
 - 微信朋友圈广告是典型feeds流广告
 - feeds广告就是与内容混排在一起的广告。
 - 最不像广告的广告，长得最像内容的广告。
 - Feeds广告操作性简单，打扰性低，已经成为移动互联网时代主流的广告形式。
- 建立在用户行为记录和大数据
 - 分析基础上，个性化推荐
- 微信朋友圈广告中的社交影响力
 - 微信朋友圈广告具有互动属性



真实应用实例

- 朋友圈广告中的社交影响力
 - 朋友圈广告具有互动属性



- 关键节点先投放，逐层影响其它节点

宝马中国
越是期待已久，悦是如期而至。
查看详情 ↗

悦

22小时前

zbf (张必锋), Lala-Du, 黄识-hanks, 林瑾jomar, nat, aragon, Amy, 孙超, 陈源, Stin张新辉, 李小叉, 爷巍, summer妹子, Tira 微信广州, iNane, 陈红, caden, elise, 东北小黑龙, lex, shawn, 丹Sir, berryxie(谢平进), 涠山Linda, 棱輝, tenfyguo郭懿心, tata, 晓翼, iris, A Du晓峰, 我是MG, 小王子-我要奋斗, 步步, 余传亮, 肉兔, 小仙女, 涠山小颖, 蛋蛋【Enid】-微信测试, 雪松, 乐宇, 肖爽, Shandidi, 李梅

Lala-Du: mark one

Lala-Du: 我开i8比X5更合适, 有木有 😊

zbf (张必锋): 宝马好车

nat: 哈哈 收到

林瑾jomar: 我是目标客户 😊

hudson: 赞赞赞

节点价值

$$quality_score = \alpha \times heart_rate_{ee} + \beta \times comment_rate_{ee} + \gamma \times influence_{ee}$$

点赞概率

评论概率

个人影响力