



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

第二次作业 文本分析

虚假信息分析

授课教师：范举副教授、塔娜讲师  
时间：2019年11月25日

# 从微信群里的一则消息谈起.....

刘娜09社人

愿你出走半生，归来仍是童年  
愿你出走半生，归来仍然是少年。

刘娜09社人

为了展示人大校友二代才艺，促进家庭合作，活跃群内气氛，增强校友凝聚力，2018年6月2日晚在“人大校友互助平台教育群（微信群）”里举办了一场别开生面的“庆六一”线上联欢会。

6月10日 晚上20:40

吴秀伦

0:55

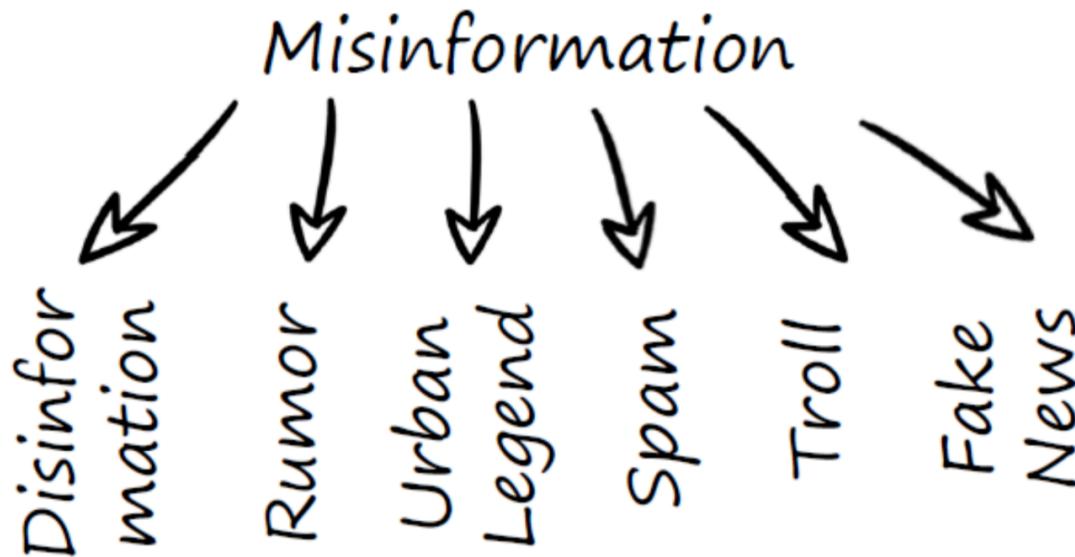
6月10日 晚上20:59

(,,) (:) (+)



# 虚假信息 (Misinformation)

- 在社交媒体上自发传播的错误（假）或不准确（虚）的信息
- 与虚假信息 (Misinformation) 相近的概念



# 思考题：区分两个概念

请连线

• Misinformation

• Disinformation

- 某CEO向股东做报告时说营业额本市本行业Top-3，但他没有说本市同行业企业只有3家，该公司排最后。

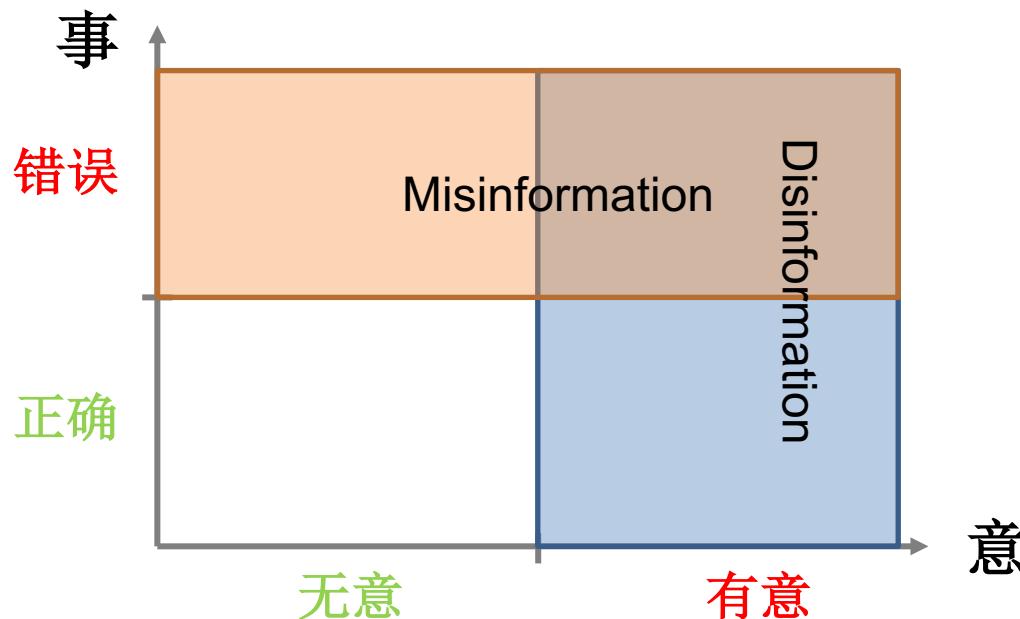
- 某CEO向股东做报告时把营业额说高了，事后证实是财务把数据搞错了，财务和CEO在报告的时候都不知情

春秋之义，原情定过，赦事诛意

——《后汉书·霍谞传》

# 分析：区分两个概念

- 引入两个维度
  - 事：信息所陈述的事实是否正确
  - 意：发布信息的动机是否存心误导



本讲侧重  
**Misinformation**

# 社交媒体上的虚假信息

虚假信息  
散布者

欺诈者

攻击者

.....

虚假信息  
• 谣言  
• 不实新闻

虚假信息  
Misinformation

影响群体  
• 社交媒体用户  
• 微信DAU: 10亿

社交  
媒体



动机

# 虚假信息的危害

- 世界经济论坛十全球危机 (2014年)
  - *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*
- Macquarie词典2016年年度单词



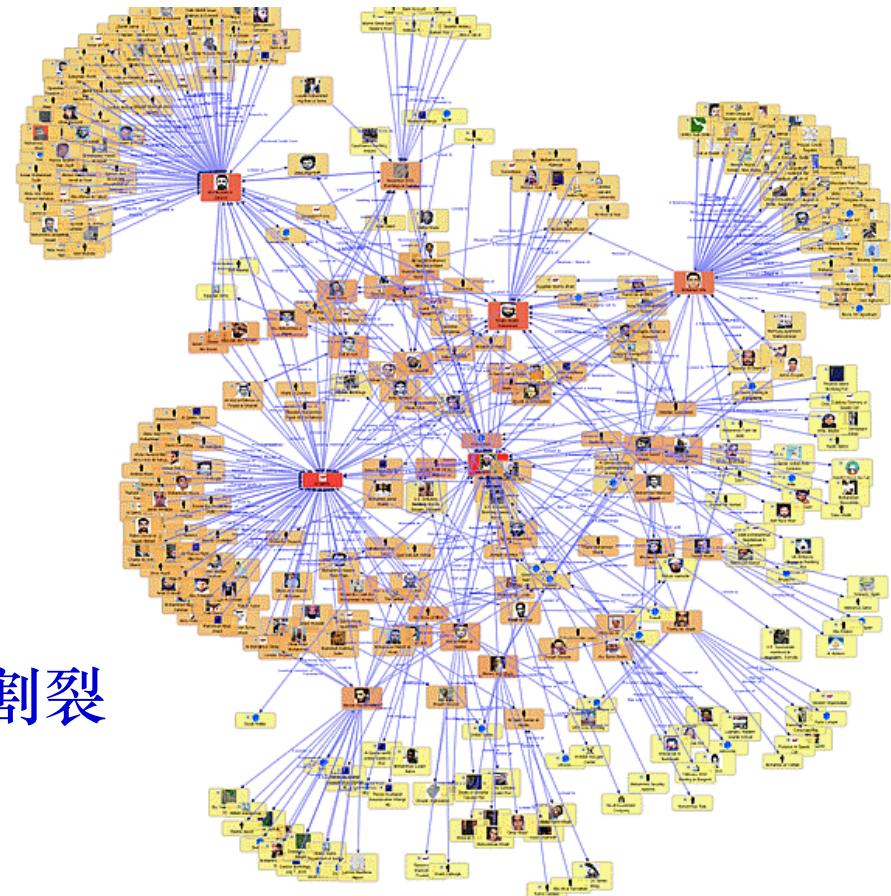
# 虚假信息的危害

- 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

# 社交媒体虚假信息应对的挑战

- 规模广
  - 大规模蔓延
- 发展快
  - 快速发展变形
- 验证难
  - 难以验证真伪
- 危害大
  - 回音室、社群割裂

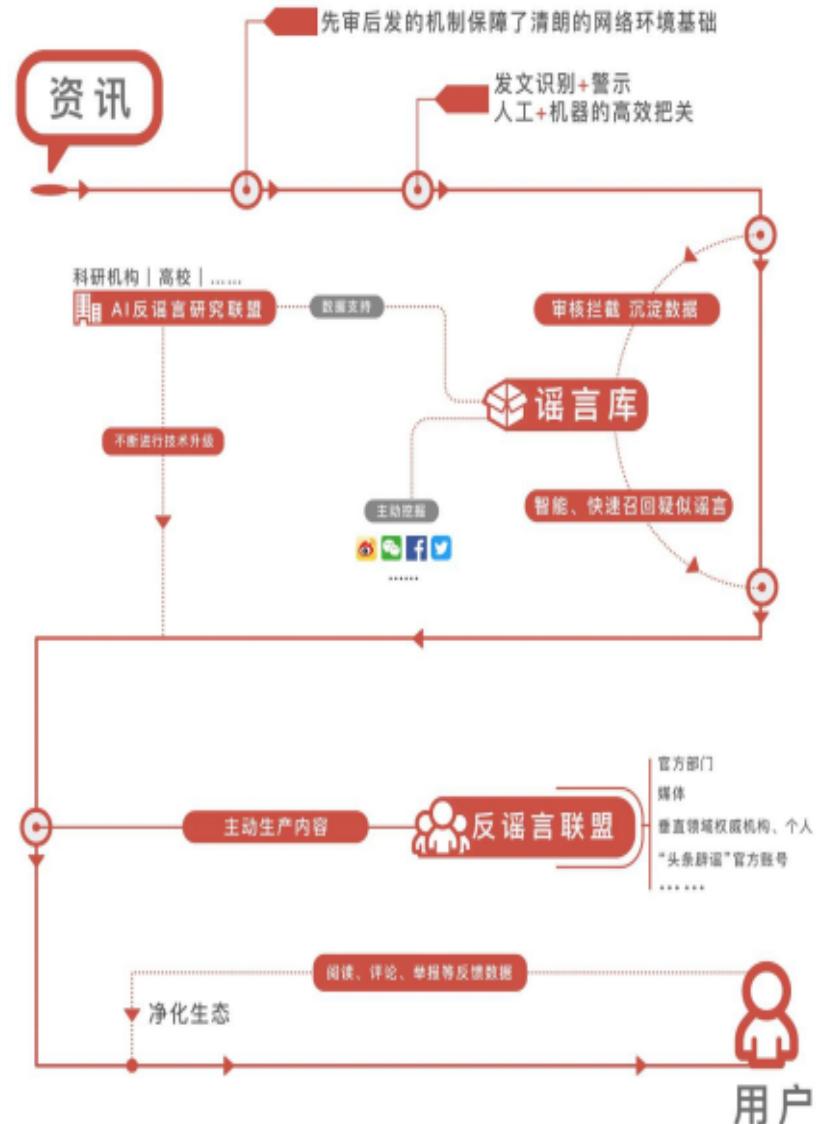


# 今日头条反谣言路线图

- 2018年3月16日今日头  
条（以下简称“头条”）发布反谣言技术  
路线图

- 人民日报、中国日报报道
- 建立系统机制、全链条传播  
监控，及时干预
- 精准辟谣，锁定易感人群
- 悟空间答、微头条、西瓜视  
频……全产品形态鼓励、促  
进辟谣内容生产传播

- 其中所谓的“谣言”很  
大一部分是指虚假信息



# 今日头条在反虚假信息方面的困境

- 今日头条在“谣言拦截”方面已做了不少工作
  - 已构建30多万的谣言（文档）库，涵盖3万多个谣言并配备30余人的标注人员负责审核
  - 基于文本相似性进行谣言的匹配与拦截
- 然而，却面临着以下困境
  - 基于文本相似性的谣言匹配对层出不穷的谣言变种识别拦截效果有限，且不具备“可解释性”——头条做了“幕后英雄”，用户、民众、政府对其反谣言工作的感知不够
  - 突发类的谣言经常出现“多次反转”、相关方“你方唱罢我登场”的现象，简单的文档库不能处理这种“结构性”和“动态性”
  - 专业知识类谣言（如健康类）很能基于匹配识别拦截

# 第二次作业说明

## 实验说明--Lab2

我们生活在一个富媒体的时代，众多繁杂的信息以各种各样的形式涌入到我们日常生活的学习工作之中，可是却同时带来了一个重要的问题——**信息失真**。

它不仅发生在你主动获取信息的过程中——比如说阅读新闻的时候。有可能你赖以获取信息的来源是不真实的，这一类失真的主要特点是**假**；它也可能发生在你被动获取信息的过程中——比如垃圾短信。它让你的信息量变得低效、偏离你获取信息的标的，而这一类失真的主要特点是**杂**。

下面我们将会以**虚假新闻数据集**和**垃圾短信数据集**实现两个数据分析的实例，来看看我们所学到的知识技能能够如何帮助我们解决现实中面临的种种问题。



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

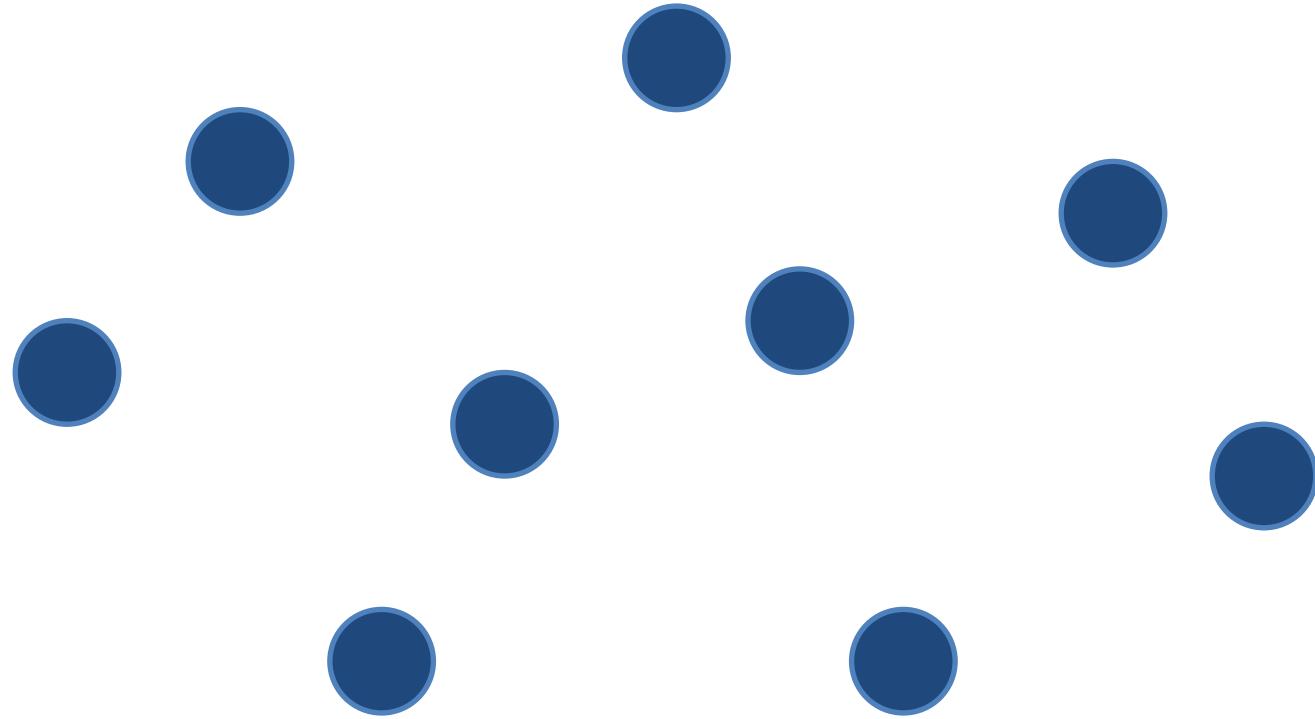
## 第四讲 网络分析

## 节点中心度分析

授课教师：范举副教授、塔娜讲师

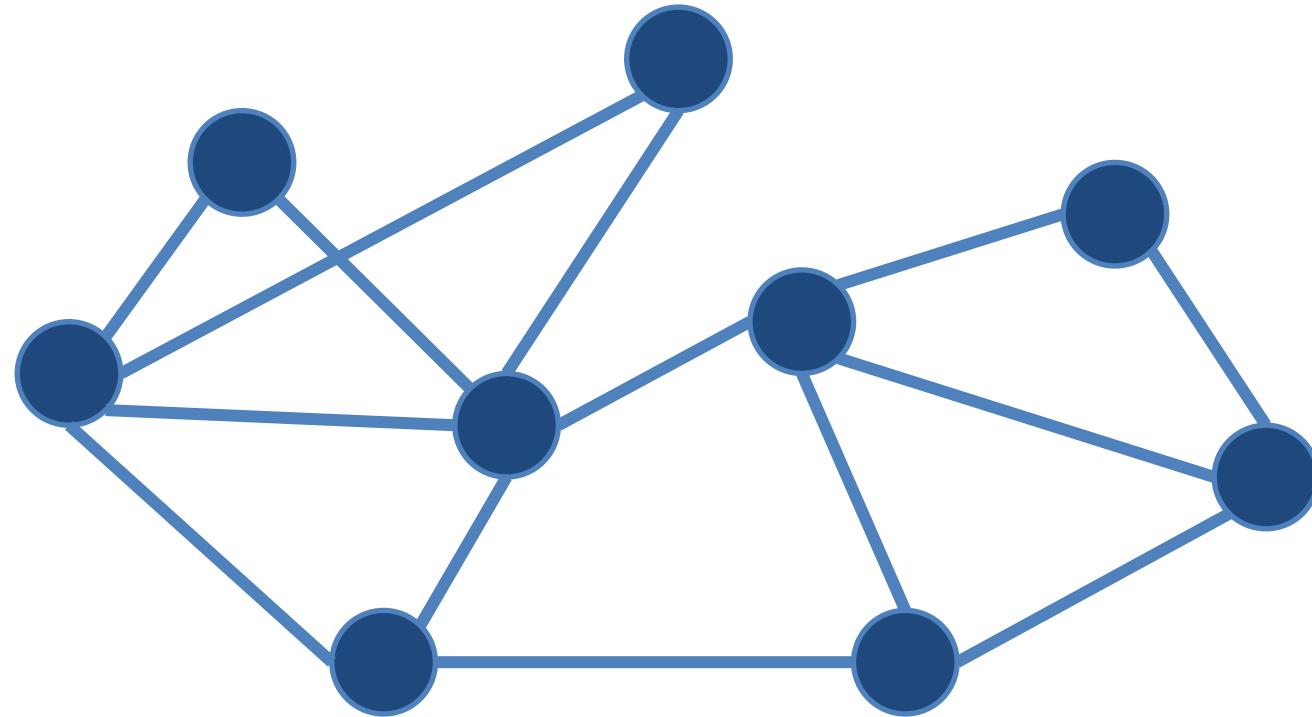
时间：2019年11月25日

# 复习：什么是图



图模型：表征事物之间的相互关联

# 复习：什么是图



图模型：表征事物之间的相互关联

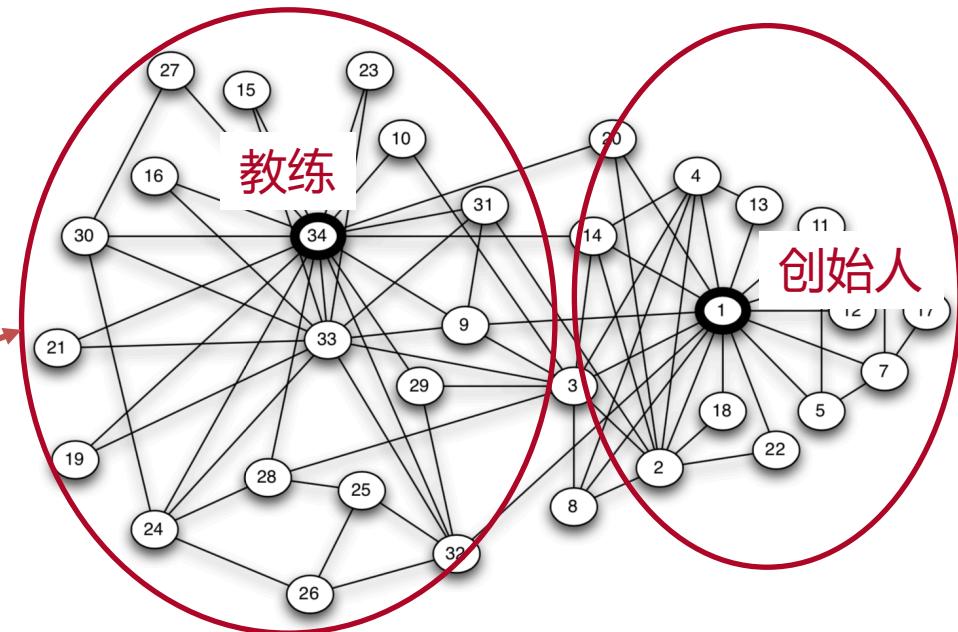
# 复习：为什么用图建模

- 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]

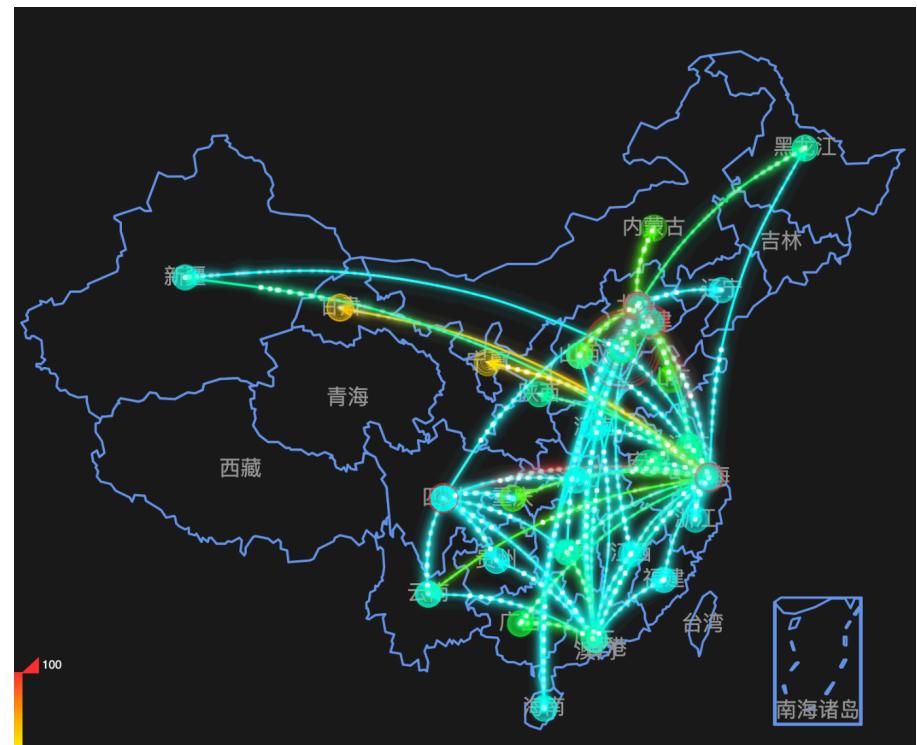
# 复习：为什么用图建模

- 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/>]

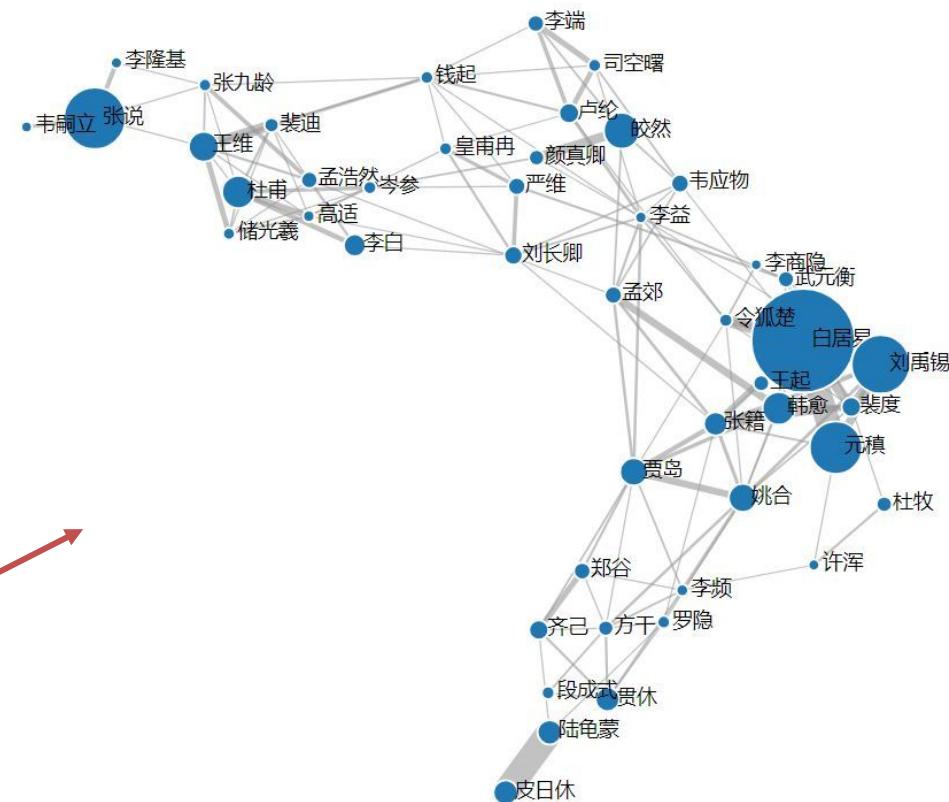
# 复习：为什么用图建模

- 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

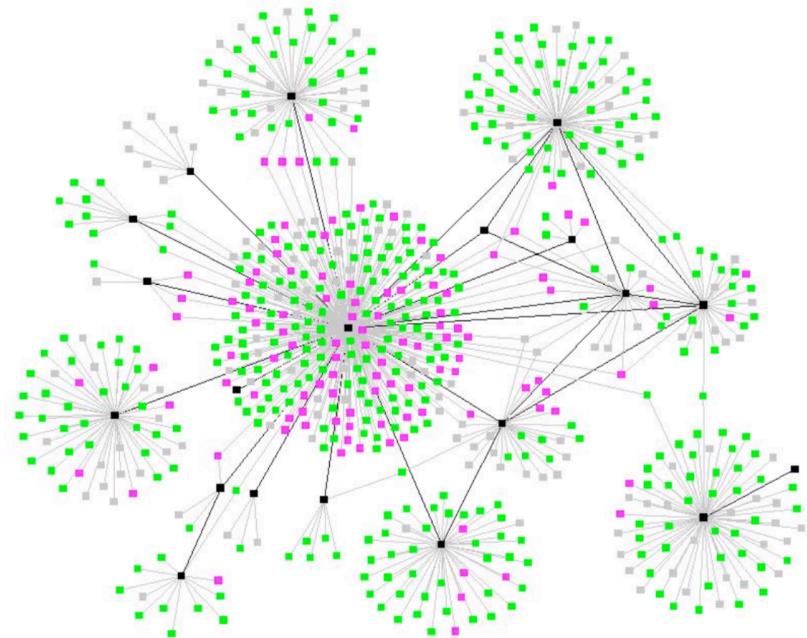


图片来源: [http://www.sohu.com/a/248677106\_488227]

# 复习：为什么用图建模

- 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

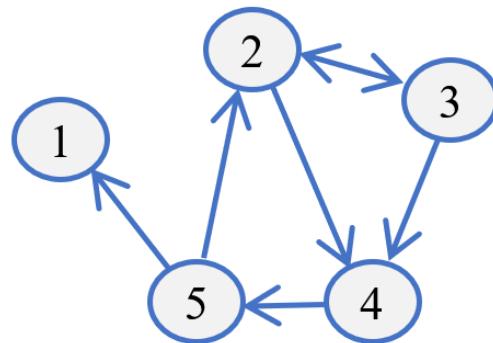
一次肺结核爆发的扩散过程，与信息传播很类似



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

# 复习：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

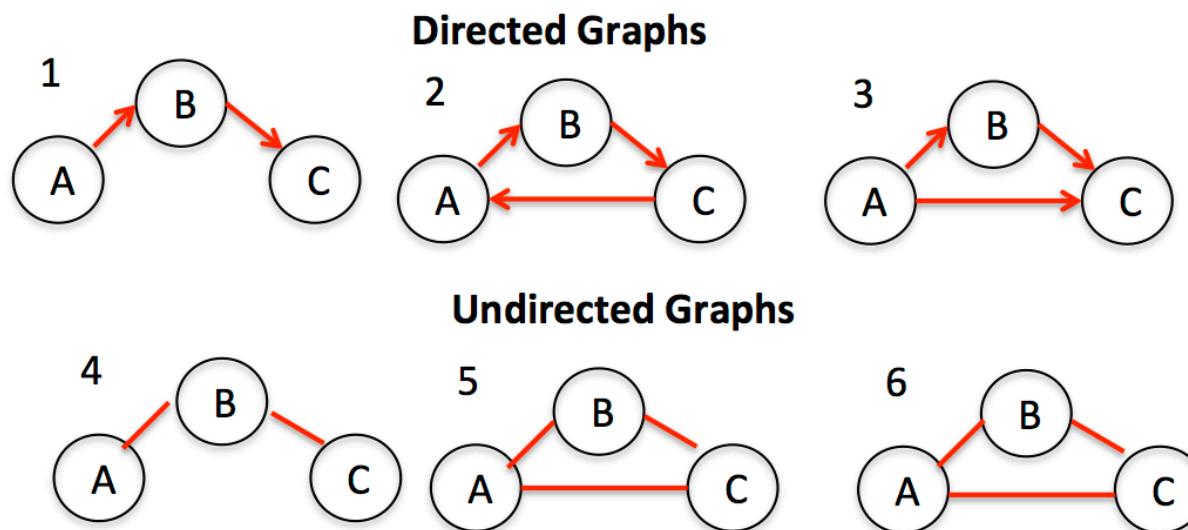
1: 2: 3 4
3: 2 4
4: 5
5: 1 2

“边列表”表示

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

# 复习：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

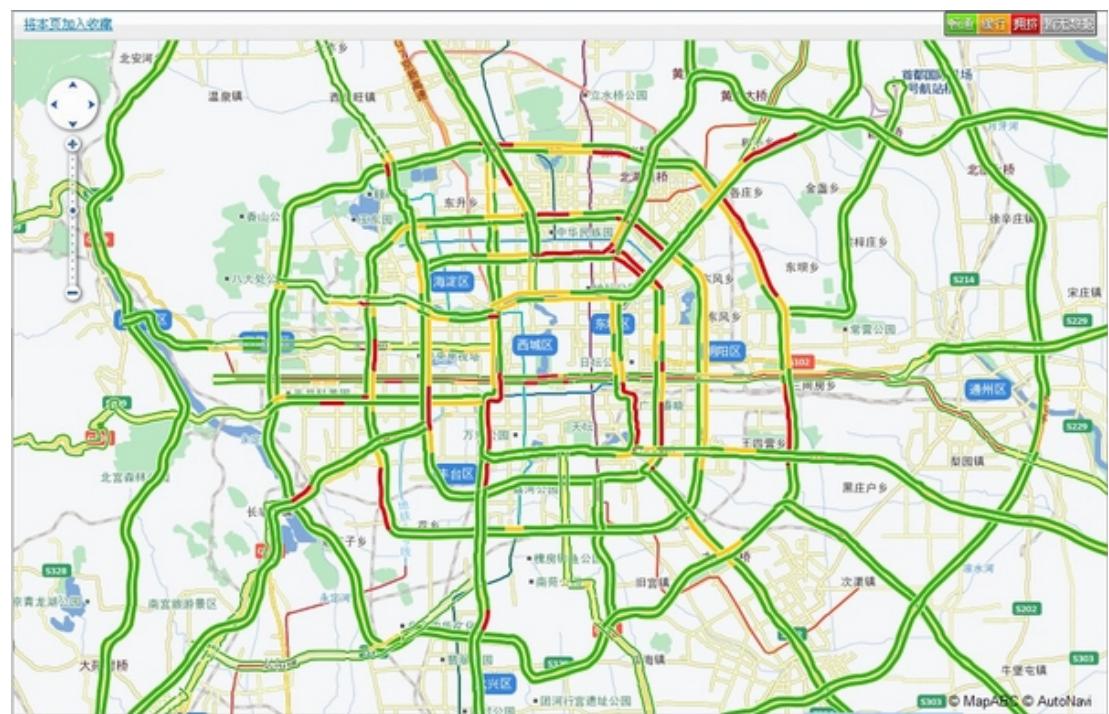
# 复习：边的权重

- 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

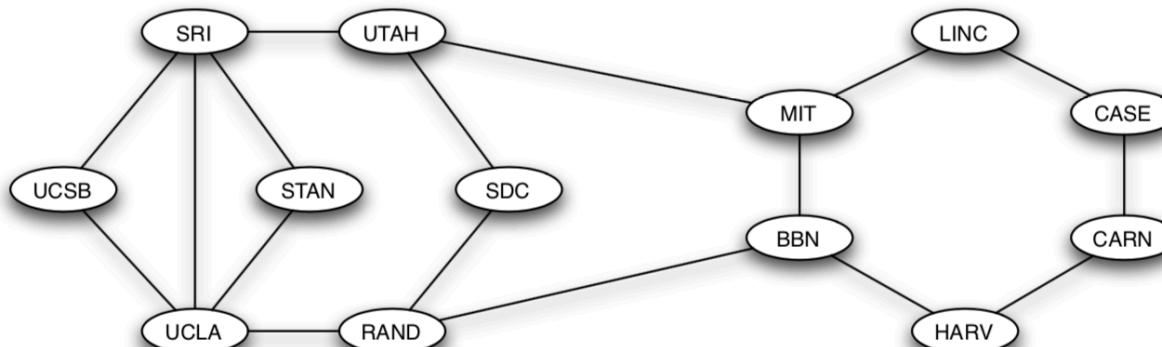
边：道路

权重：拥堵情况



# 复习：路径

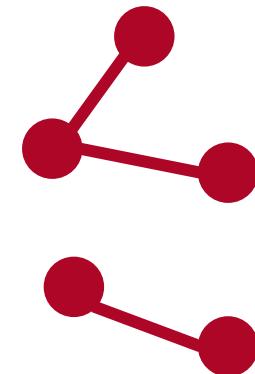
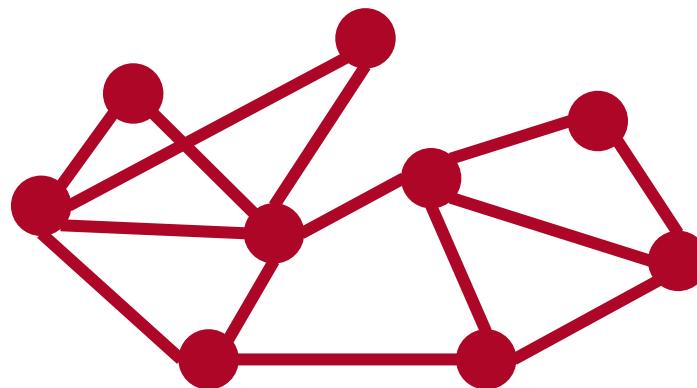
- 路径：图上节点的序列，序列中任意两个相邻的节点都有边相连
  - Path  $p = (v_0, v_1, \dots, v_m)$  where any  $(v_i, v_{i+1}) \in E$
  - 简单路径：不包含重复节点的路径
- 环：起点与终点相同的路径
  - $p = (v_0, v_1, \dots, v_m, v_0), (v_i, v_{i+1}) \in E$



- 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

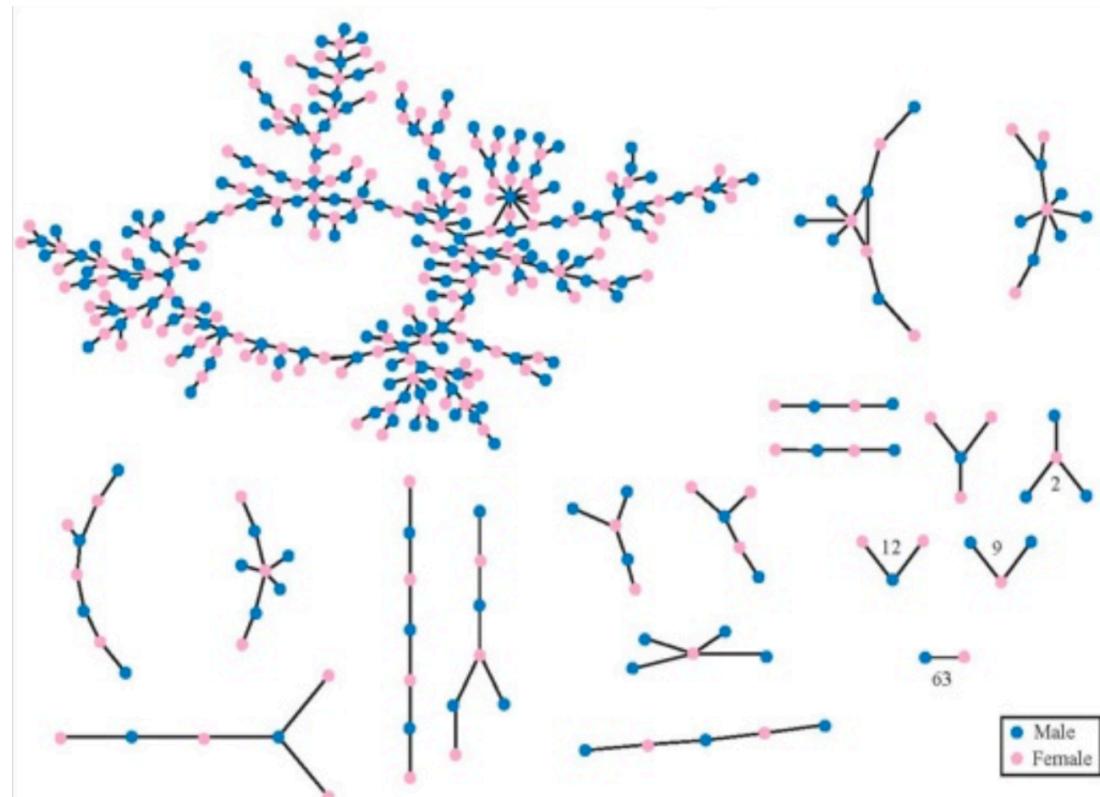
# 复习：连通图与连通分支

- 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**



# 复习：连通图与连通分支

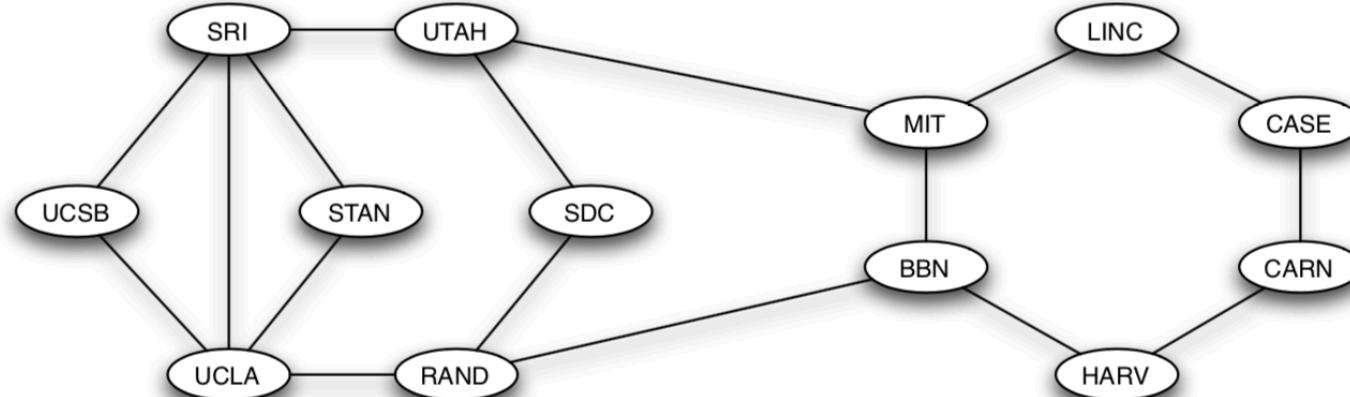
- 例：美国高中生恋爱关系图（边代表二人在18个月内恋爱过）



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

# 复习：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

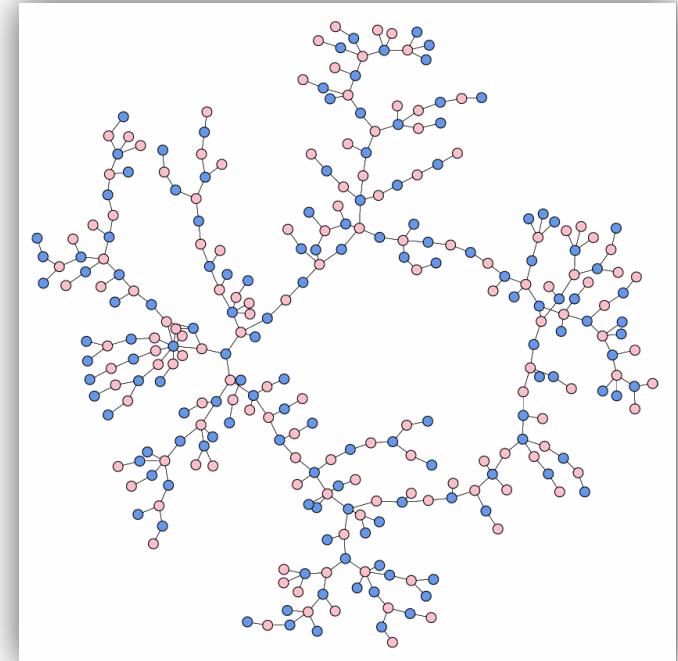


# 节点中心度 (Node Centrality) 分析

- 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？



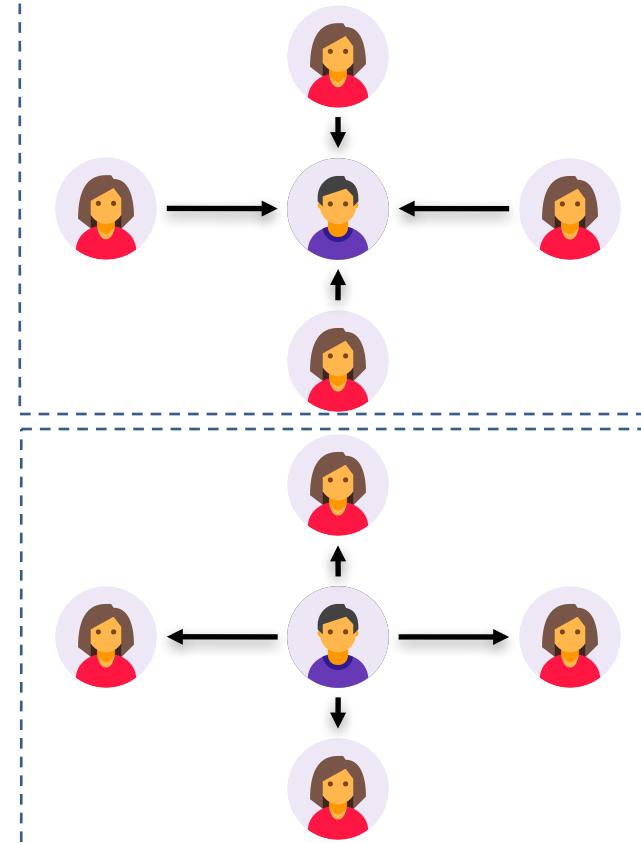
图片来源: [Bearman et al., American Journal of Sociology, 2004]

# 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
  - 定义有向边：“追求”关系

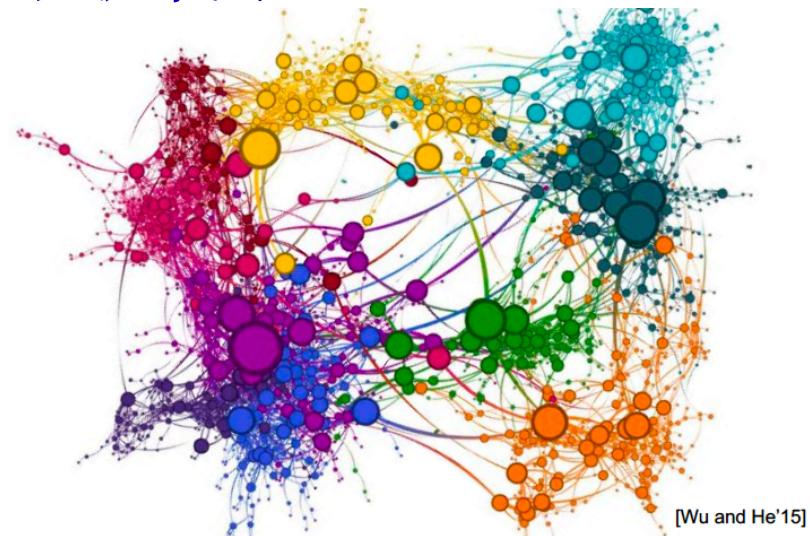


- 思考：
  - 右边两图中男生的重要性一样吗？
  - 你怎么解释这种重要性？



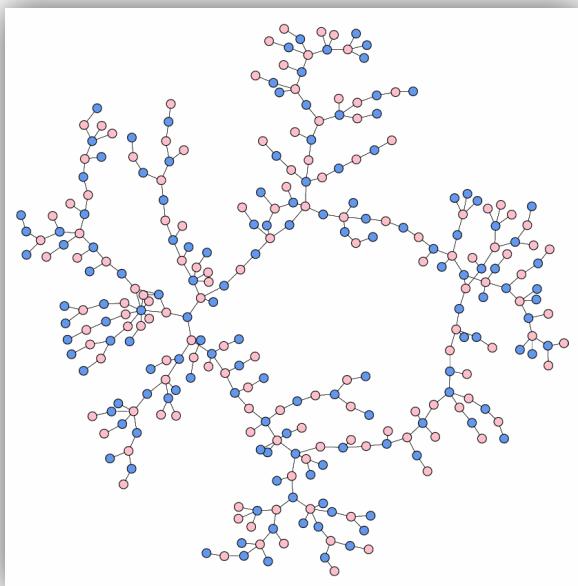
# 节点中心度 (Node Centrality)

- 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”？
  - 哪些网页更加重要？
  - 你能想到其它吗？

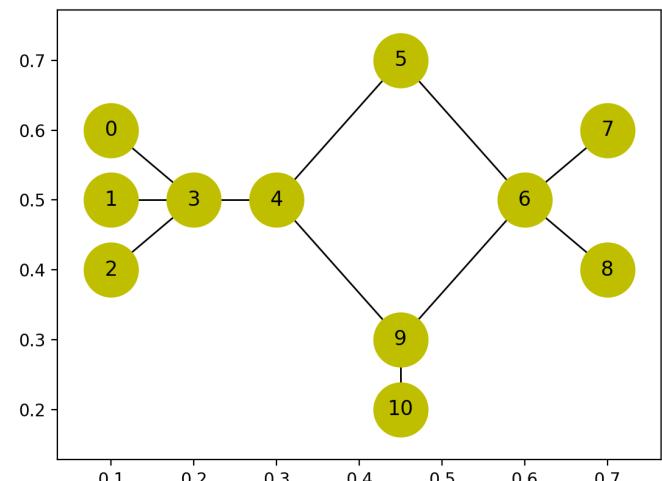


# 我们考虑一个小例子

- 演示示例
  - 简化版的美国高中生恋爱关系图



简化  
→



# 分析工具： NetworkX

## NetworkX

Stable (notes)

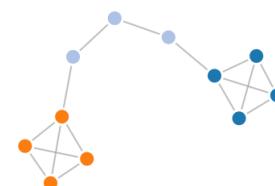
2.2 — September 2018  
[download](#) | [doc](#) | [pdf](#)

Latest (notes)

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

Software for complex networks

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



```
$ pip install networkx
```

```
>>> import networkx as nx
>>> G = nx.Graph()
>>> G.add_edge('A', 'B', weight=4)
>>> G.add_edge('B', 'D', weight=2)
>>> G.add_edge('A', 'C', weight=3)
>>> G.add_edge('C', 'D', weight=4)
```

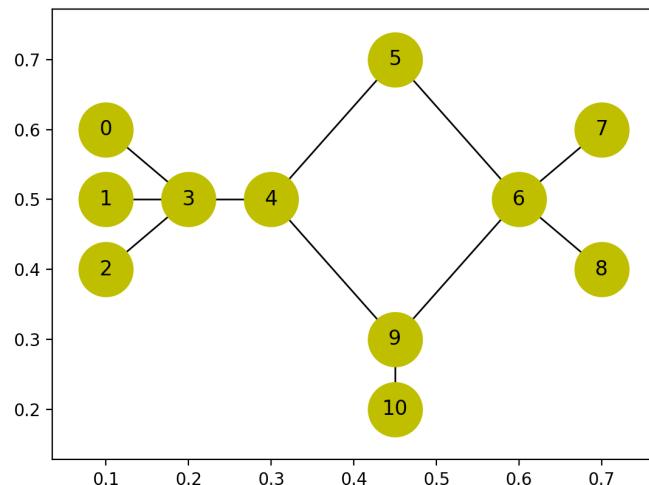
辅助工具：Python plotting lib

```
$ pip install matplotlib
```

```
>>> import matplotlib.pyplot as plt
>>> import matplotlib.colors as mcolors
```

# 分析准备

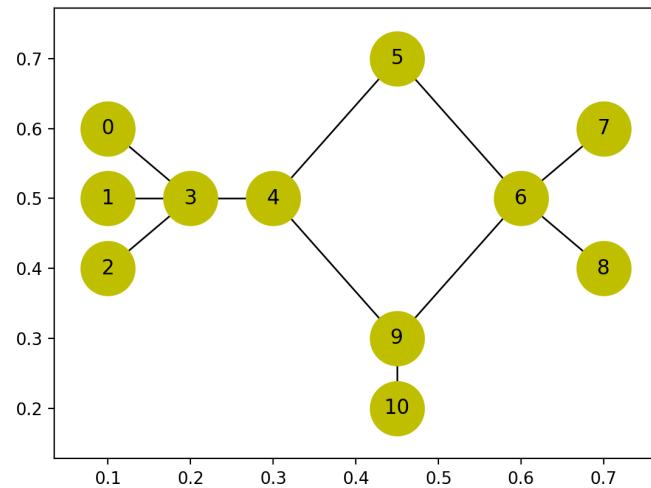
- 步骤I：生成简化版的美国高中生恋爱关系图
  - 请写出下图的邻接矩阵表示



$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

# 分析准备

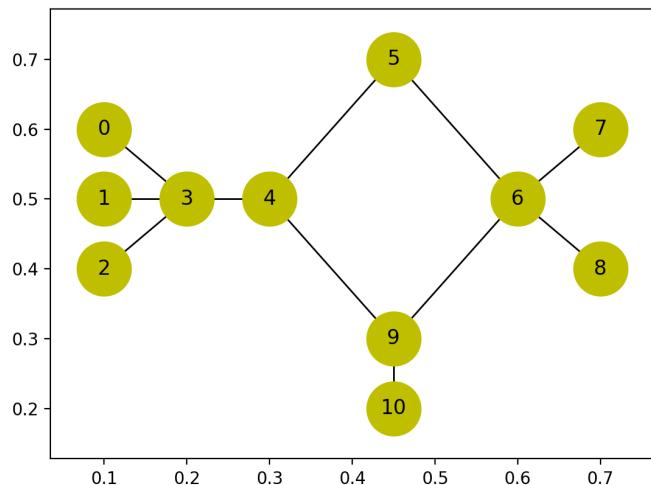
- 步骤I：生成简化版的美国高中生恋爱关系图
  - 请写出下图的表列表表示



(0,3)
(1,3)
(2,3)
(3,4)
(4,5)
(4,9)
(5,6)
(6,7)
(6,8)
(6,9)
(9,10)

# 分析准备

- 步骤1：生成简化版的美国高中生恋爱关系图
  - 使用边列表的方式初始化图



```
import networkx as nx
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors

students = set(range(11))
G = nx.Graph()
G.name = "Simple Dating Graph"
G.add_nodes_from(students)
dating_rel = [
    (0,3), (1,3), (2,3), (3,4),
    (4,5), (4,9), (5,6), (6,7),
    (6,8), (6,9), (9,10)
]
G.add_edges_from(dating_rel)
```

# NetworkX函数讲解

- 创建图结构

```
>>> G=nx.Graph() # 无向图  
>>> G=nx.DiGraph() # 有向图  
>>> G=nx.MultiGraph() # 无向多重图  
>>> G=nx.MultiDiGraph() # 有向多重图
```

- 添加节点

```
>>> G.add_node(1) # 添加一个节点  
>>> G.add_nodes_from([2,3]) # 添加多个节点
```

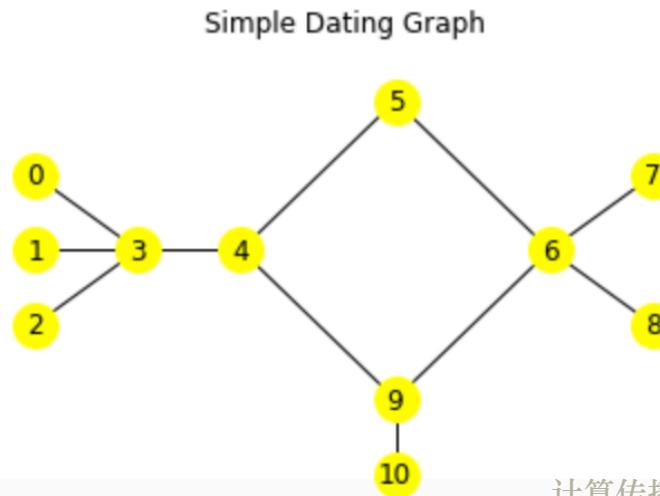
- 添加边

```
>>> G.add_edge(1,2) # 添加一条边  
>>> G.add_edges_from([(1,2),(1,3)]) # 添加多条边
```

# 分析准备

- 步骤2：可视化简化版美国高中生恋爱关系图

```
pos = {0: [0.1, 0.6], 1: [0.1, 0.5], 2: [0.1, 0.4],
       3: [0.2, 0.5], 4: [0.3, 0.5], 5: [0.45, 0.7],
       6: [0.6, 0.5], 7: [0.7, 0.6], 8: [0.7, 0.4],
       9: [0.45, 0.3], 10: [0.45, 0.2]}
nodes = nx.draw_networkx_nodes(G, pos, node_size=400,
                               node_color='yellow')
nx.draw_networkx_labels(G, pos)
nx.draw_networkx_edges(G, pos)
plt.title(G.name)
plt.axis('off')
plt.show()
```



## 第4.2.1节

# 基于几何图形的度量方法

# 基于几何图形的度量方法

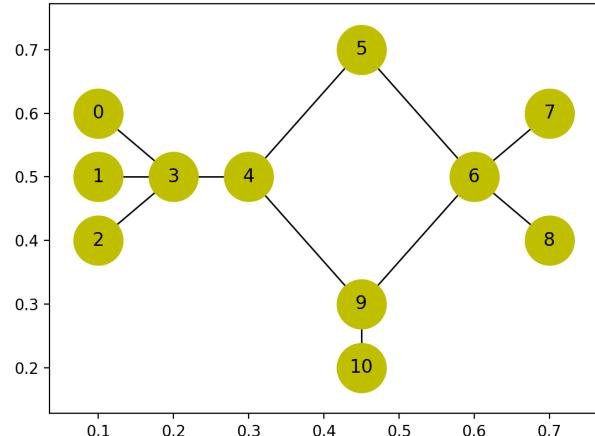
- 基本思想
  - 节点 $v$ 的Centrality是该节点到其它节点的距离的函数
- (In-)Degree Centrality
  - 节点 $v$ 的Centrality取决于它的度（如果是有向图，则为入度），即到节点 $v$ 距离为1的节点的个数

$$C_D(v) = \frac{\deg(v)}{n - 1}$$

为什么除以 $n-1$ ？

- 请计算右图中节点的Degree Centrality

{0: 0.1, 1: 0.1, 2: 0.1, 3: 0.4, 4: 0.3, 5: 0.2, 6: 0.4, 7: 0.1, 8: 0.1, 9: 0.3, 10: 0.1}



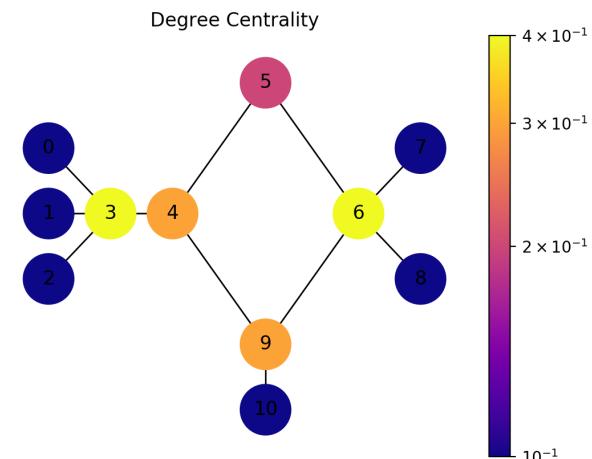
# 基于几何图形的度量方法

- 基本思想
  - 节点 $v$ 的Centrality是该节点到其它节点的距离的函数
- (In-)Degree Centrality
  - 节点 $v$ 的Centrality取决于它的度（如果是有向图，则为入度），即到节点 $v$ 距离为1的节点的个数

$$C_D(v) = \frac{\deg(v)}{n - 1}$$

- 如何解释Degree Centrality?
  - 恋爱网络：情史丰富 😞
  - 微博网络：大V

更直观的可视化——热力图



# 节点度数的分布

- 问题：在大多数真实图（如社交图）中，节点度数呈现什么分布？
- 交互环节
  - 微信扫描并填写调查问卷
  - 填写你的姓名
  - 填写在本课堂中的好朋友姓名，多个姓名以空格分隔

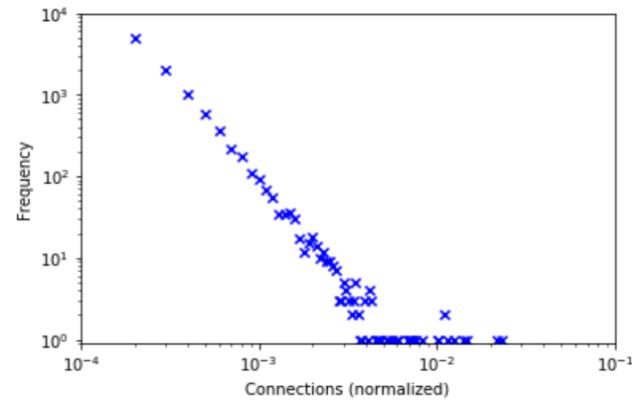


# 基于节点度数的解决思路

- 问题：在大多数真实图（如社交图）中，节点度数呈现什么分布？

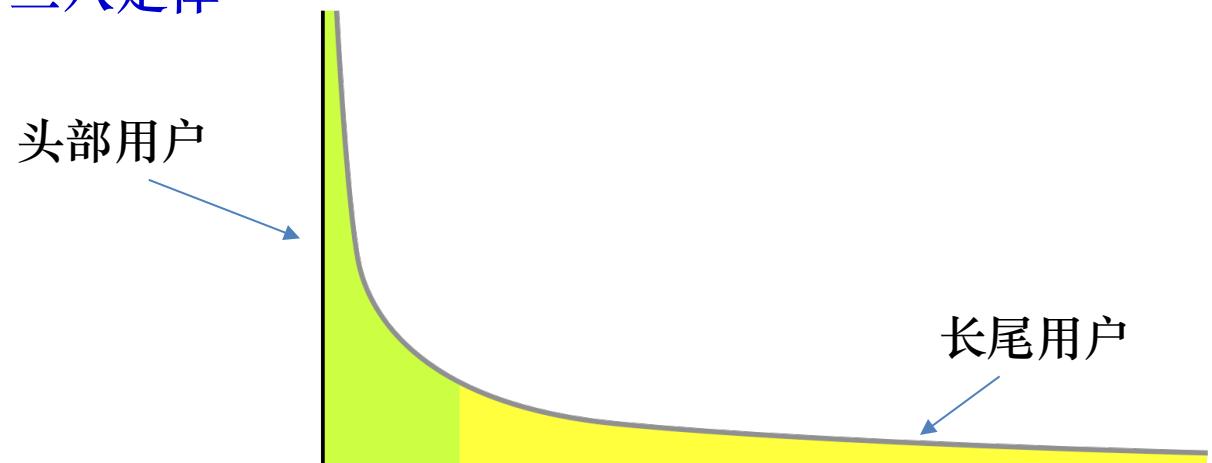
```
def degree_histogram (G):
    deg = nx.degree_centrality(G)
    deg_raw = {k: int(v * (len(G) - 1)) for k, v in
deg.items()}
    deg_hist = dict(Counter(deg_raw.values()))

    plt.xscale('log')
    plt.yscale('log')
    plt.scatter(deg_hist.keys(), deg_hist.values(),
c='b', marker='x')
    plt.xlim((1e0, 1e2))
    plt.ylim((1e0, 1e2))
    plt.xlabel('Degree')
    plt.ylabel('Frequency')
    plt.show()
```



# 基于节点度数的解决思路

- 问题：在大多数真实图（如社交图）中，节点度数呈现什么分布？
  - Power Law幂律分布：度量分布函数 $P(K)$ ，其中 $K$ 为节点度数， $P(K)$ 表示度数为 $K$ 的节点出现的频率
  - Power Law度数分布的直观解释 – 长尾分布：大多数节点的度数都很低，只有少部分节点的度数很高，即对他人能够产生较大影响力
  - 著名的“二八定律”



# 案例分析：公众号的幂律分布

类型	定义	所占百分比	现状概括	重点事项
头部	>100万粉丝	1. 80%	变现	帮别人卖：接广告 自己卖：实物类，虚拟类（知识付费，会员）
腰部	1-100万	19. 60%	折腾	涨粉+变现
尾部	<1万	78. 60%	挣扎	内容+涨粉

- 大号：既得利益者，上有远虑
- 腰部账号：正值当年，尚能饭否
- 尾部账号：生无可恋，回天乏术

来源：<http://www.niaogebiji.com/article-16711-1.html>

# 案例分析

- 如何解释中小学生沉湎游戏
  - 游戏玩得好 → 越容易受欢迎?

# 绘制节点热力图

```
def draw(G, pos, measures, measure_name):
    nodes = nx.draw_networkx_nodes(G, pos, node_size=400,
                                    cmap=plt.cm.plasma,
                                    node_color=list(measures.values()),
                                    nodelist=list(measures.keys()))
    nodes.set_norm(mcolors.SymLogNorm(linthresh=0.01, linscale=1))

    nx.draw_networkx_labels(G, pos)
    nx.draw_networkx_edges(G, pos)

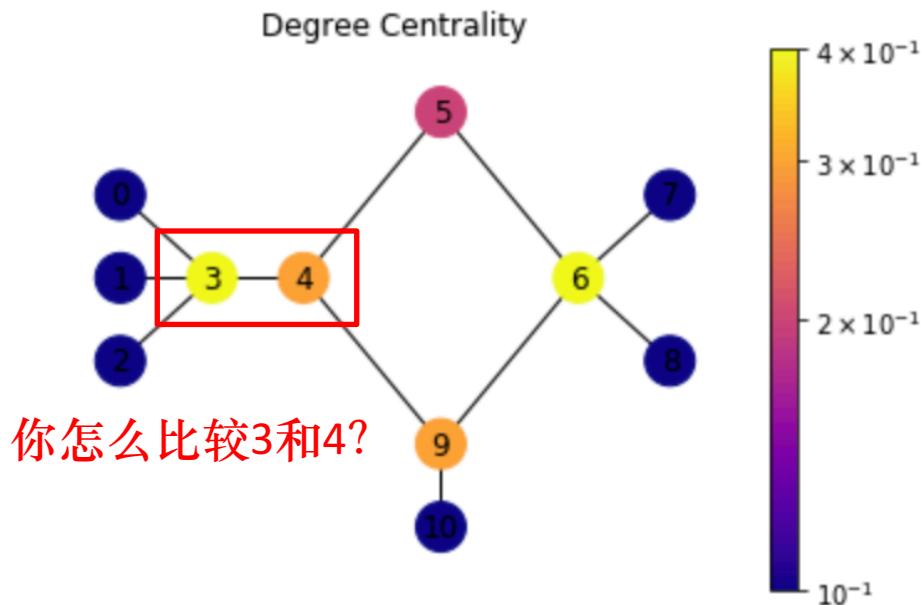
    plt.title(measure_name)
    plt.colorbar(nodes)
    plt.axis('off')
    plt.show()

print ('Degree Centrality')
print(nx.degree_centrality(G))
draw(G, pos, nx.degree_centrality(G), 'Degree Centrality')
```

# 绘制节点热力图

Degree Centrality

```
{0: 0.1, 1: 0.1, 2: 0.1, 3: 0.4, 4: 0.3000000000000004, 5: 0.2, 6: 0.4,  
7: 0.1, 8: 0.1, 9: 0.3000000000000004, 10: 0.1}
```



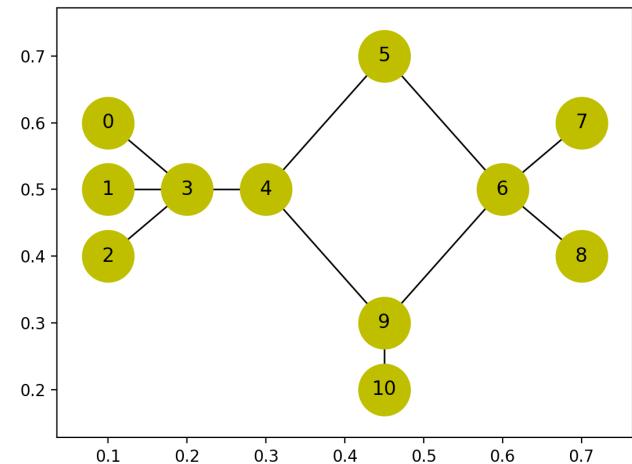
# 基于几何图形的度量方法

- Closeness Centrality

- 节点 $v$ 的Centrality取决于其它节点到它的距离是否接近
- 其它节点到 $v$ 的距离越接近， $v$ 越重要

$$C_C(v) = \frac{n - 1}{\sum_{u \in V - \{v\}} d(u, v)}$$

$d(u, v)$ 为 $u$ 到 $v$ 的最短距离



Closeness  
Centrality {0: 0.32, 1: 0.32, 2: 0.32, 3: 0.45, 4: 0.53, 5:  
0.45, 6: 0.43, 7: 0.31, 8: 0.31, 9: 0.5, 10: 0.34}

# 基于几何图形的度量方法

- Closeness Centrality

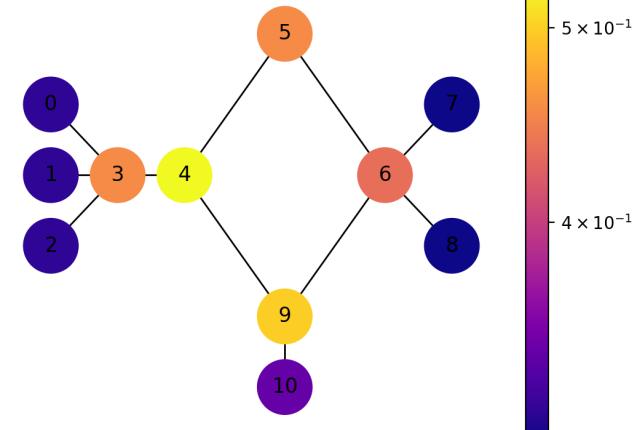
- 节点 $v$ 的Centrality取决于其它节点到它的距离是否接近
- 其它节点到 $v$ 的距离越接近， $v$ 越重要

$$C_C(v) = \frac{n - 1}{\sum_{u \in V - \{v\}} d(u, v)}$$

{0: 0.32, 1: 0.32, 2: 0.32, 3: 0.45, 4: 0.53, 5:  
0.45, 6: 0.43, 7: 0.31, 8: 0.31, 9: 0.5, 10: 0.34}

- 如何解释Closeness Centrality ?
  - 恋爱网络：跟哪个圈子的人都谈过
  - 道路网络：中心地标建筑

Closeness Centrality



## 第4.2.2节

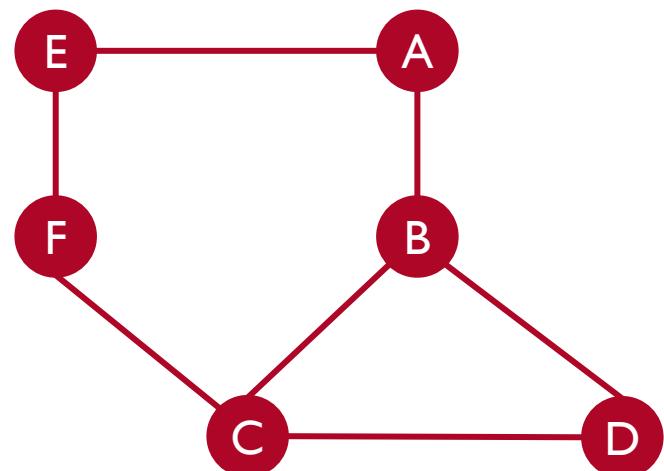
# 基于路径的度量方法

# 思考一道题

- 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）

- 思考：关键节点起到了什么作用？

中介！



# 基于路径的度量方法

- Betweenness Centrality

- 节点 $v$ 的Centrality取决于它是否经常出现在其它节点的最短路径上
- $v$ 出现在其它节点最短路径上次数越多， $v$ 越重要

$$C_C(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

$\sigma(s,t)$ 为 $s$ 到 $t$ 的最短路径个数； $\sigma(s,t|v)$   $s$ 到 $t$ 经过 $v$ 的最短路径个数

Betweenness Centrality

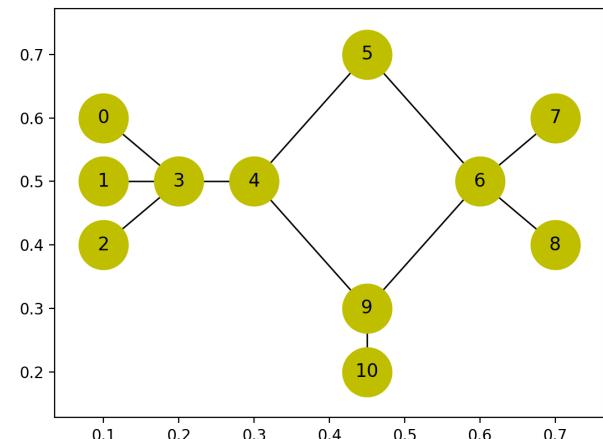
{0: 0.0, 1: 0.0, 2: 0.0, 3: 0.53, 4: 0.56, 5: 0.17, 6: 0.4, 7: 0.0, 8: 0.0, 9: 0.37, 10: 0.0}

Closeness Centrality

{0: 0.32, 1: 0.32, 2: 0.32, 3: 0.45, 4: 0.53, 5: 0.45, 6: 0.43, 7: 0.31, 8: 0.31, 9: 0.5, 10: 0.34}

Degree Centrality

{0: 0.1, 1: 0.1, 2: 0.1, 3: 0.4, 4: 0.3, 5: 0.2, 6: 0.4, 7: 0.1, 8: 0.1, 9: 0.3, 10: 0.1}



# 基于路径的度量方法

- Betweenness Centrality

- 节点 $v$ 的Centrality取决于它是否经常出现在其它节点的最短路径上
- $v$ 出现在其它节点最短路径上次数越多， $v$ 越重要

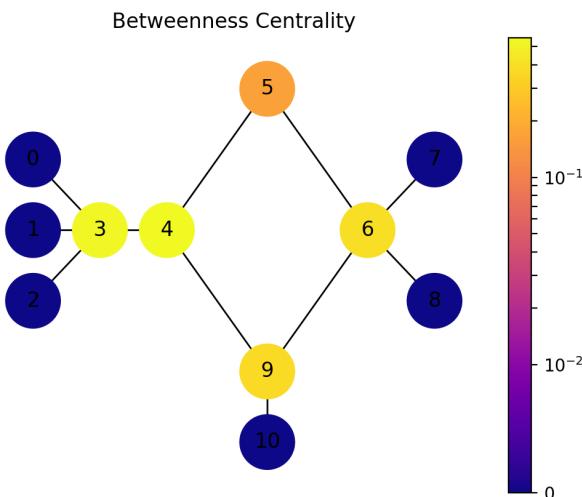
$$C_C(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

$\sigma(s,t)$ 为 $s$ 到 $t$ 的最短路径个数； $\sigma(s,t|v)$   $s$ 到 $t$ 经过 $v$ 的最短路径个数

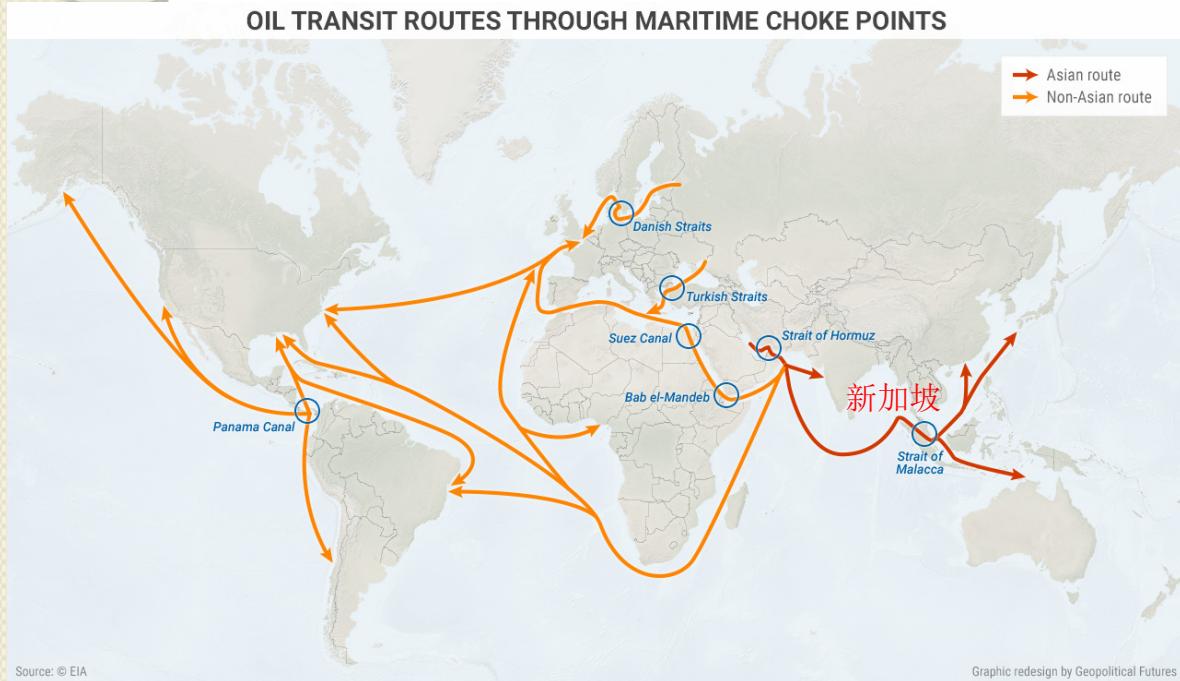
Betweenness Centrality

{0: 0.0, 1: 0.0, 2: 0.0, 3: 0.53, 4: 0.56, 5: 0.17, 6: 0.4, 7: 0.0, 8: 0.0, 9: 0.37, 10: 0.0}

- 如何解释Betweenness Centrality ?
  - 恋爱网络：织就情网 😕
  - 贸易网络：贸易枢纽hubs



# 思考：新加坡为什么成为全球贸易港？

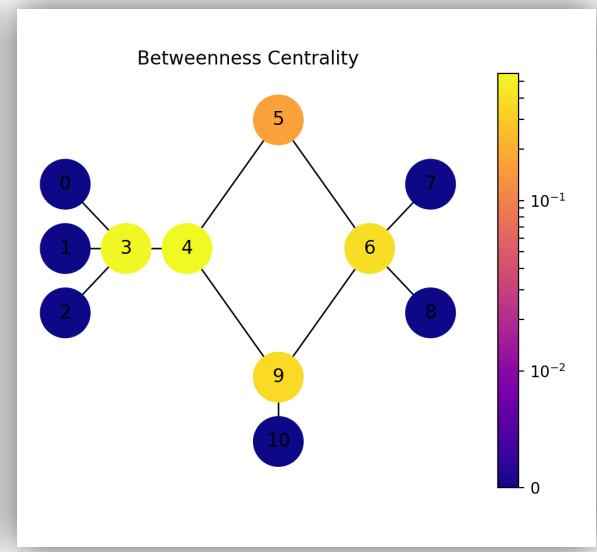
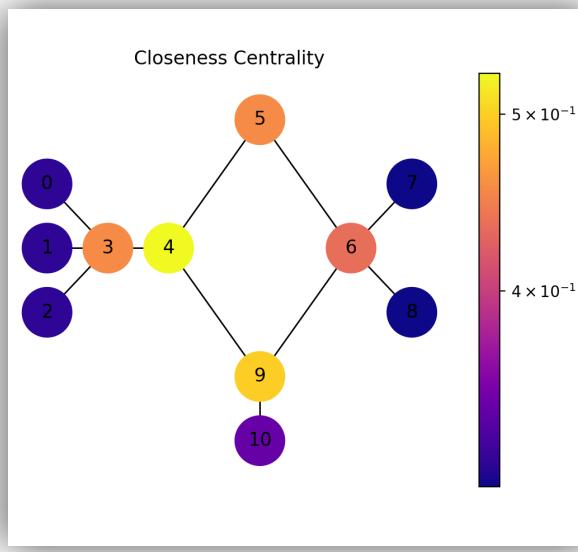
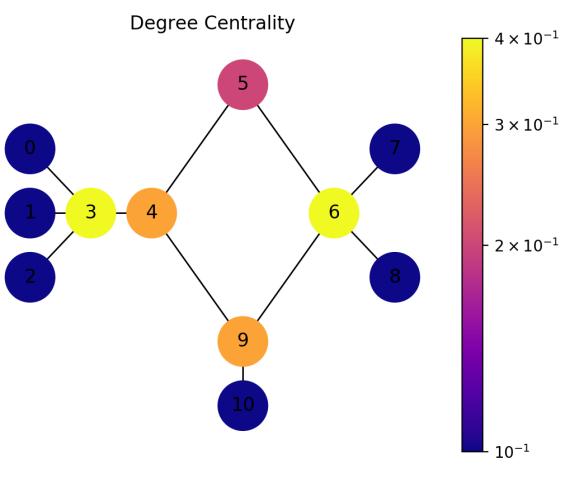


波斯湾石油运输航路图

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  
Betweenness  
Centrality

# 三种Centrality度量的比较



Degree Centrality

Closeness Centrality

Betweenness Centrality

# 三种Centrality度量的比较

- Python代码实现

```
if __name__ == '__main__':
    G, pos = gen.school_dating_graph()

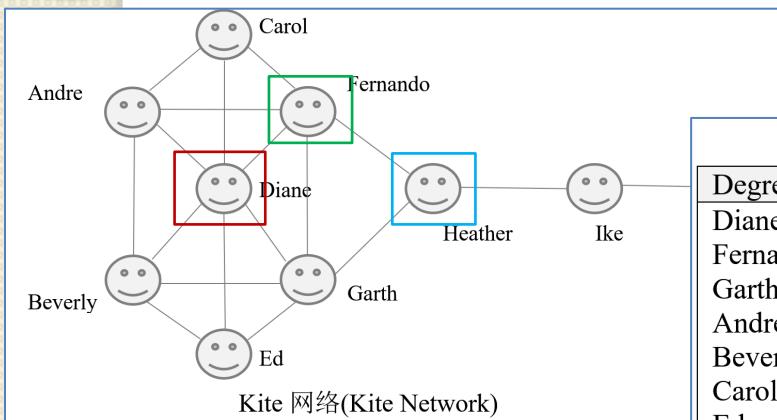
    print ('Degree Centrality')
    print(nx.degree_centrality(G))
    draw(G, pos, nx.degree_centrality(G), 'Degree Centrality')

    print ('Closeness Centrality')
    print (nx.closeness_centrality(G))
    draw(G, pos, nx.closeness_centrality(G), 'Closeness Centrality')

    print ('Betweenness Centrality')
    print (nx.betweenness_centrality(G))
    draw(G, pos, nx.betweenness_centrality(G), 'Betweenness Centrality')
```

# 三种Centrality度量的比较

- 请分别指出Degree, Closeness和Betweenness Centrality最大的节点



网络各个节点的中心性					
Degree		Closeness		Betweenness	
Diane	0.667	Fernando	0.643	Heather	0.389
Fernando	0.556	Garth	0.643	Fernando	0.231
Garth	0.556	Diane	0.600	Garth	0.231
Andre	0.444	Heather	0.600	Ike	0.222
Beverly	0.444	Andre	0.529	Diane	0.102
Carol	0.333	Beverly	0.529	Andre	0.023
Ed	0.333	Carol	0.500	Beverly	0.023
Heather	0.333	Ed	0.500	Carol	0.000
Ike	0.222	Ike	0.429	Ed	0.000
Jane	0.111	Jane	0.310	Jane	0.000

- 扩展学习：Eigenvector Centrality（与PageRank算法很相关）  
<https://www.geeksforgeeks.org/eigenvector-centrality-centrality-measure/>