



图数据入门、中心度



覃雄派



提纲

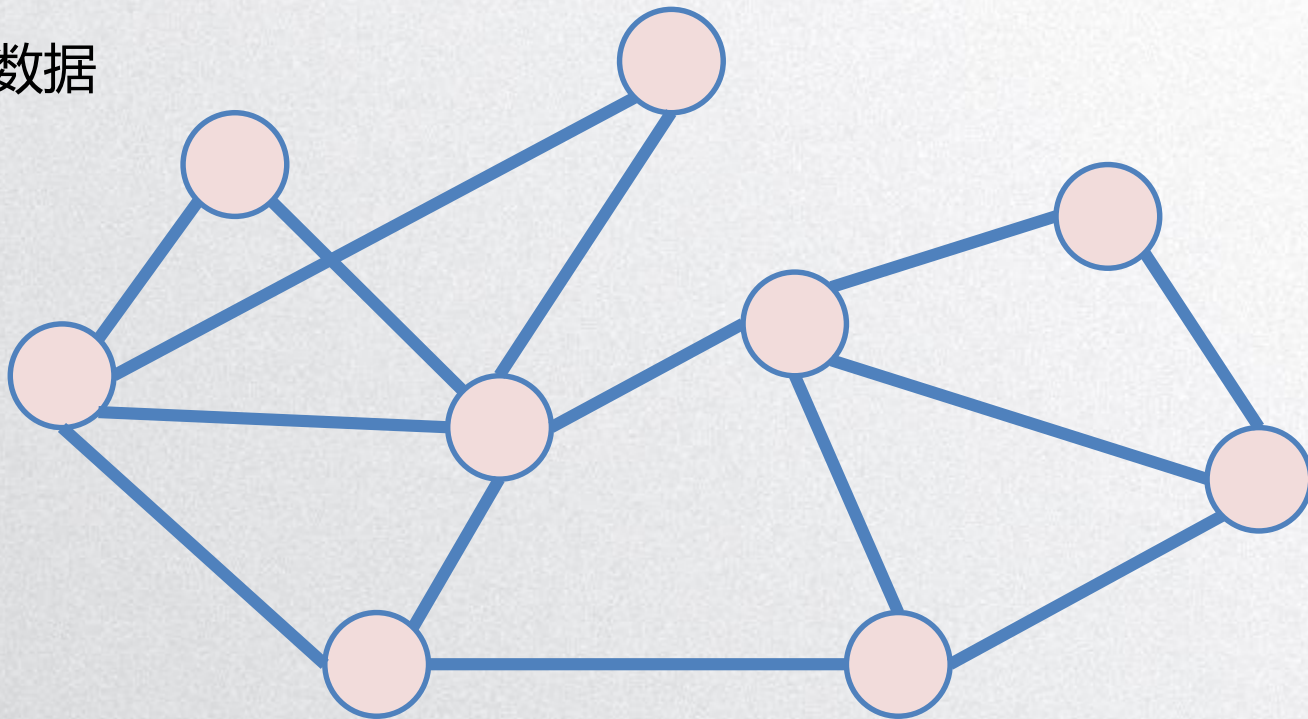
- 图数据处理概述
- 图论的基本概念
- 图数据处理模块的内容
- 中心度(中心性)
- 中心度计算Python实例分析

Orion
constellation

图数据入门、中心度

图数据入门、中心度

- 图数据



图模型：表征事物之间的相互关联

图数据入门、中心度

- 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
 - 我们的生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？



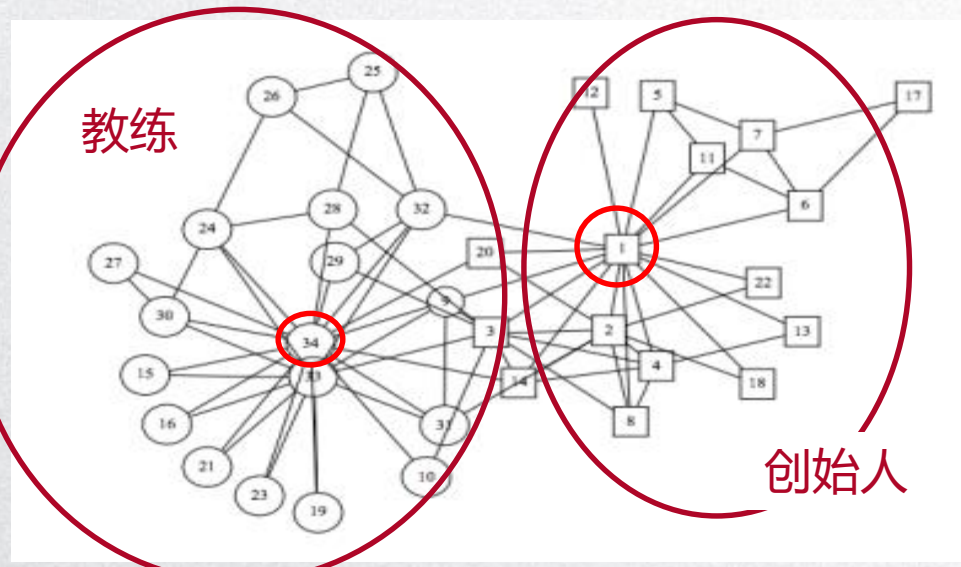
图数据入门、中心度

- 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

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?

最终这个俱乐分裂成两个对立的空手道俱乐部

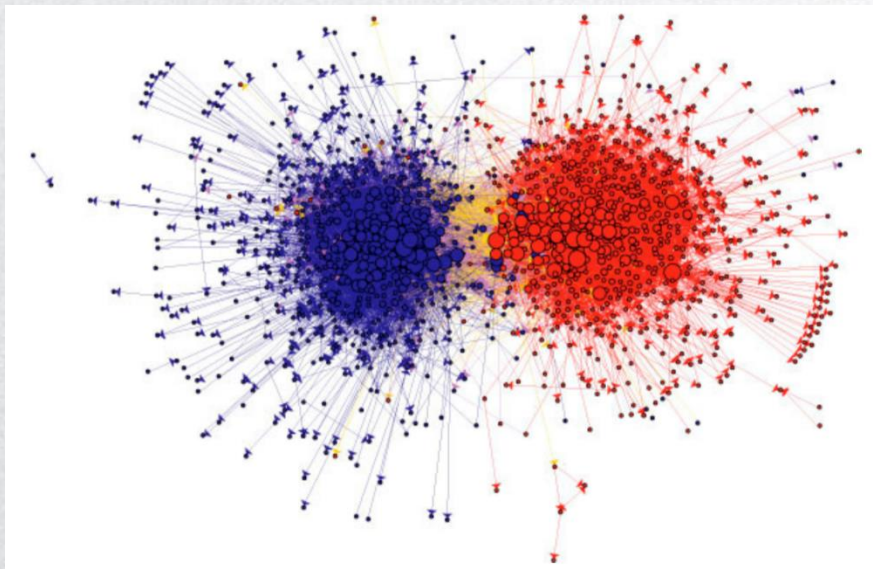
结构平衡理论



图数据入门、中心度

- 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？



图数据入门、中心度

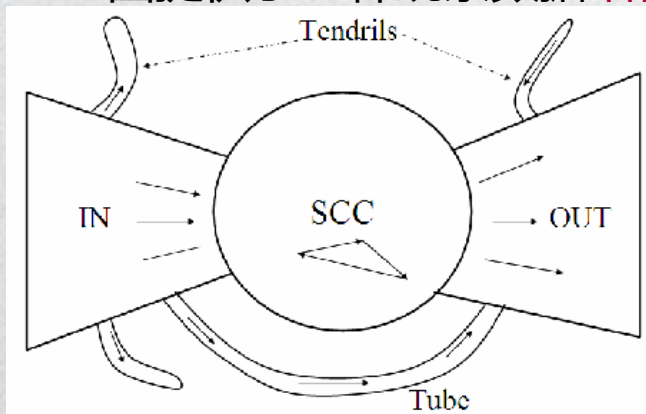
tendrils 英 ['tendrɪlz] 卷须; 卷须状物(如髭发);

n. (攀缘植物的) 卷须; 卷须状物(如髭发);



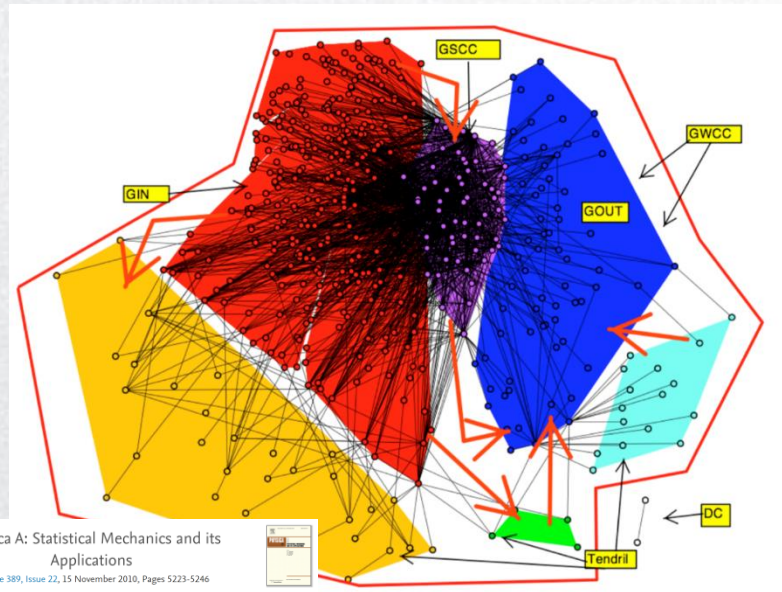
- Why Graph
- 为什么使用图模型对数据建模?
 - 图提供了一种观察数据**结构**特征的视角

金融机构之间的借贷网络,
你能发现什么结构特点?



互联网的领结结构
The bowtie Structure

https://www.researchgate.net/publication/200111010_Mining_the_Inner_Structure_of_the_Web_Graph



Physica A: Statistical Mechanics and its Applications
Volume 389, Issue 22, 15 November 2010, Pages 5223-5246

The topology of the federal funds market

Morten L. Bech^{a,1}, Engin Atalay^b

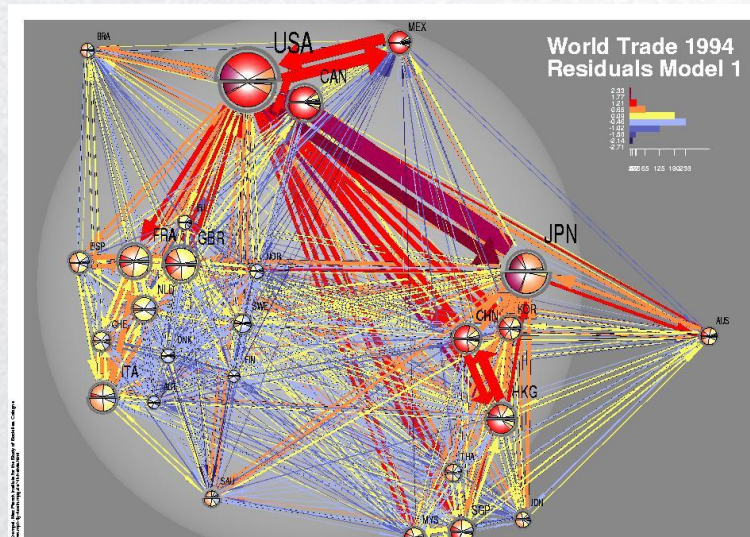
图数据入门、中心度

- 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

国家之间贸易网络，你能发现什么结构特点？

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



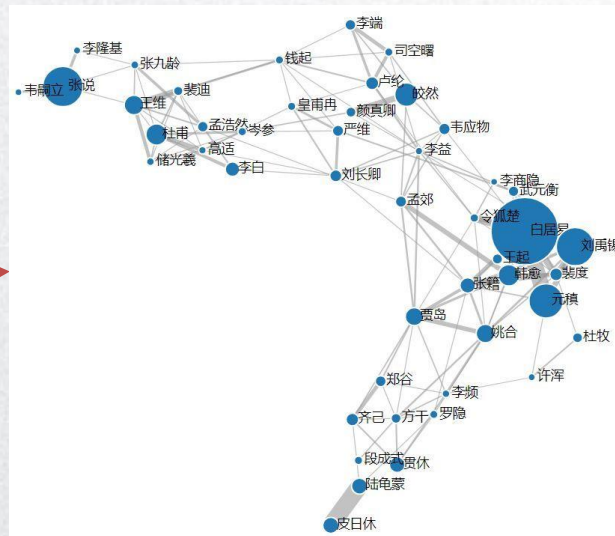
图数据入门、中心度

- 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

唐朝诗人的唱和关系图：节点大小表示和诗数量

子曰：诗可以群

元白文学集团
刘白唱和集



中唐诗人元稹、白居易的并称

https://www.sohu.com/a/248677106_488227

图数据入门、中心度

- 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



[Am J Public Health](#). 2007 March; 97(3): 470–477.

doi: [10.2105/AJPH.2005.071936](https://doi.org/10.2105/AJPH.2005.071936)

PMCID: PMC1805030

PMID: [17018825](https://pubmed.ncbi.nlm.nih.gov/17018825/)

Transmission Network Analysis to Complement Routine Tuberculosis Contact Investigations

McKenzie Andre, MD, [Kashef Ijaz](#), MD, [Jon D. Tillinghast](#), MD, [Valdis E. Krebs](#), MLIR, [Lois A. Diem](#), BS, [Beverly Metchock](#), DrPH, [Theresa Crisp](#), MPH, and [Peter D. McElroy](#), PhD

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

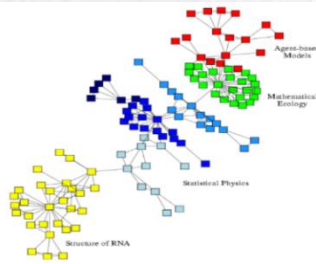


图数据入门、中心度

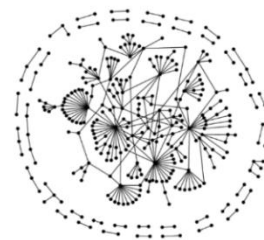
- 丰富多彩的图数据



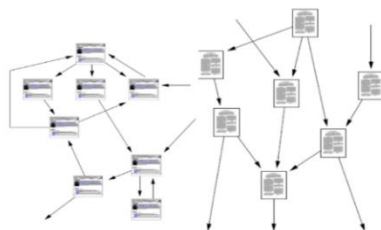
Social networks



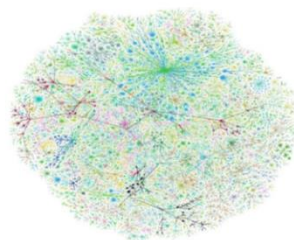
Economic networks



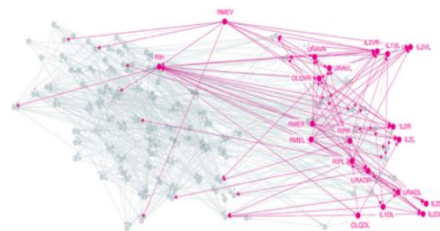
Communication graphs



Information networks:
Web & citations



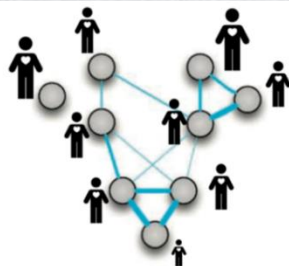
Internet



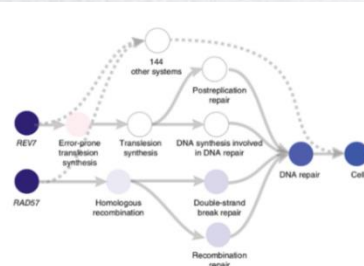
Networks of neurons

图数据入门、中心度

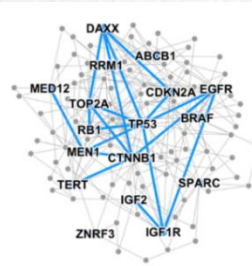
- 丰富多彩的图数据



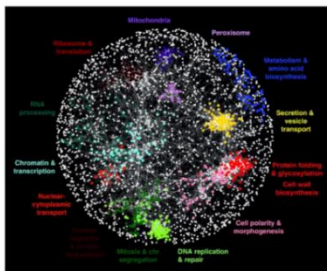
Patient networks



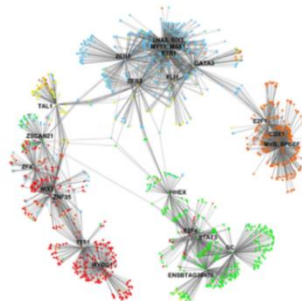
Hierarchies of cell systems



Disease pathways



Genetic interaction networks



Gene co-expression networks



Cell-cell similarity networks

pathways 英 ['pɑ:θweɪz] 美 ['pæ:θweɪz]

n. 小路; 路径; 途径; 行动路线;

图数据入门、中心度



图数据入门、中心度

- 知识图谱

- 知识图谱：语义关联、机器可读的知识表示技术



节点
关系
节点的属性
关系的属性
.....

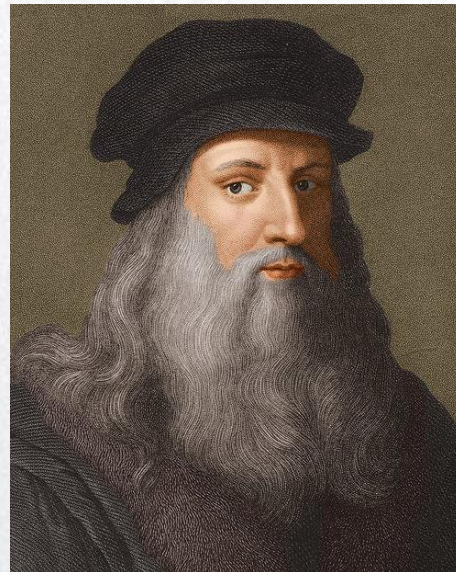
图数据入门、中心度

- 知识图谱



Mona_Lisa

painted-by



Leonardo_da_Vinci

IsA

Painting

IsA

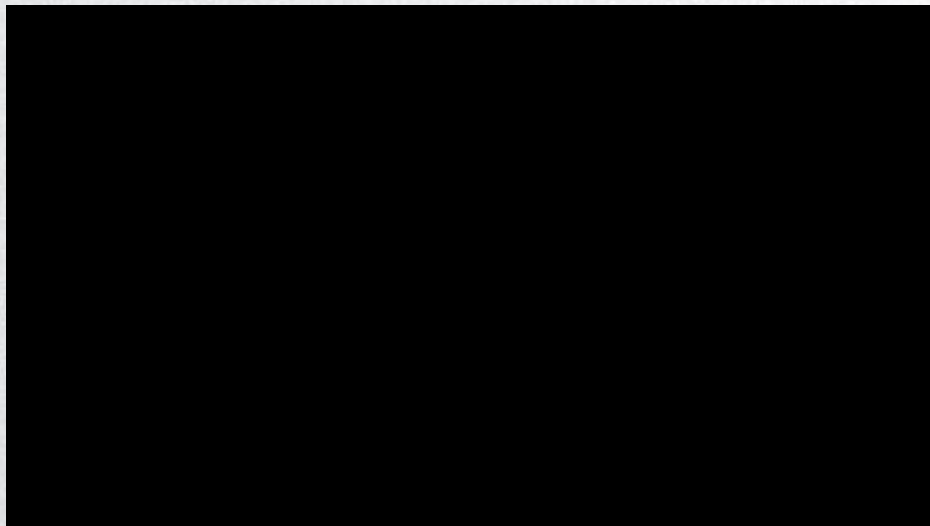
Artist

S P O



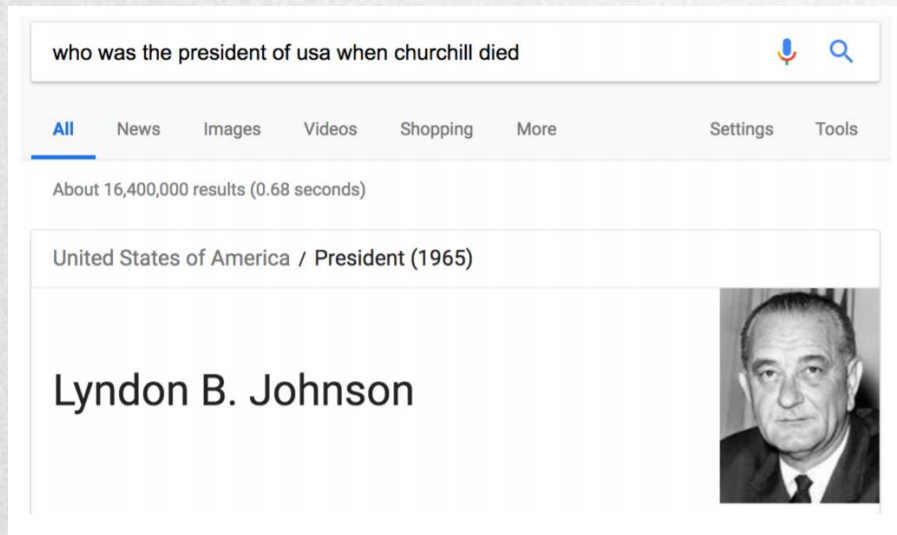
图数据入门、中心度

- 知识图谱的应用
 - 为AI系统提供领域知识



图数据入门、中心度

- 知识图谱的应用
 - 问答系统



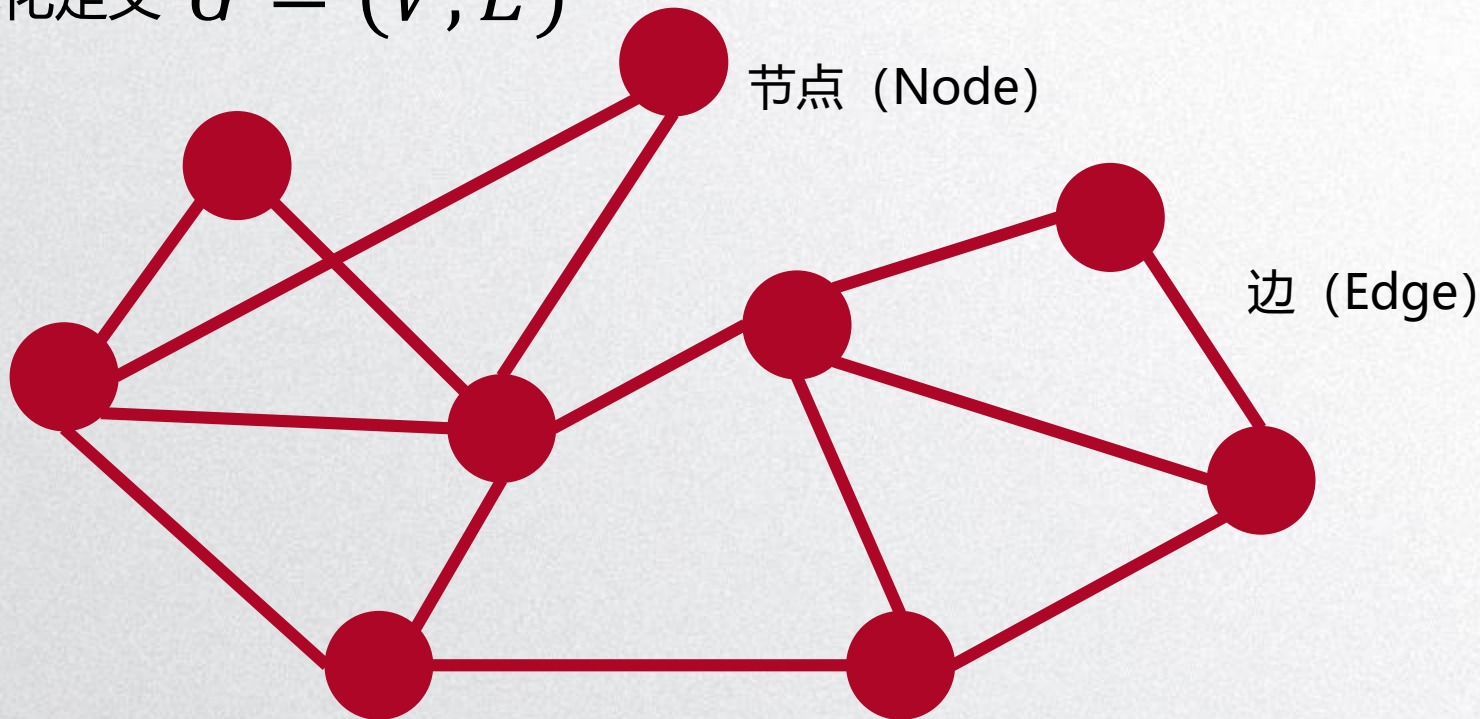
<http://www-bcf.usc.edu/~xiangren/NAACL18-KB-full.pdf>

图数据入门、中心度



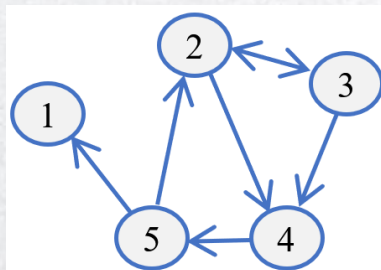
图数据入门、中心度

- 图论的基本概念
- 图的形式化定义 $G = (V, E)$



图数据入门、中心度

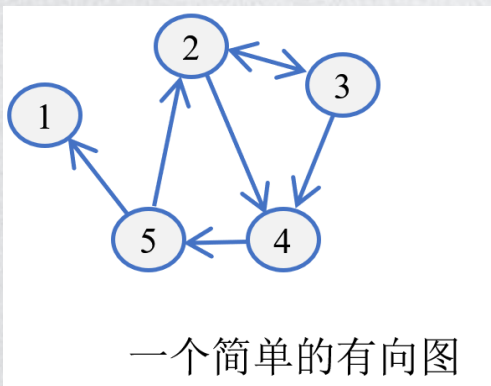
- 图论的基本概念：图的表示
- 可以使用三种方式进行表示(Representation)
 - 分别是邻接矩阵(Adjacency Matrix)
 - 边列表(Edge List)
 - 邻接关系列表(Adjacency List)



一个简单的有向图

图数据入门、中心度

- 图论的基本概念：图的表示
- 可以使用三种方式进行表示(Representation)
 - 分别是邻接矩阵(Adjacency Matrix)
 - 边列表(Edge List)
 - 邻接关系列表(Adjacency List)



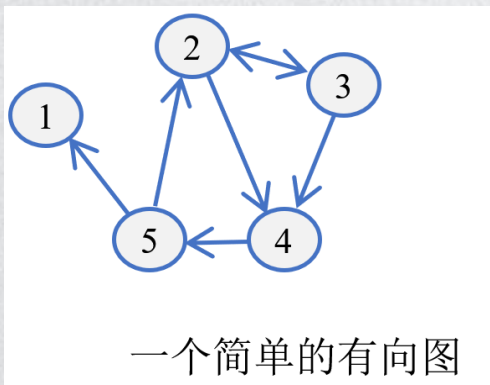
一个简单的有向图

“邻接矩阵”表示

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

图数据入门、中心度

- 图论的基本概念：图的表示
- 可以使用三种方式进行表示(Representation)
 - 分别是邻接矩阵(Adjacency Matrix)
 - 边列表(Edge List)
 - 邻接关系列表(Adjacency List)

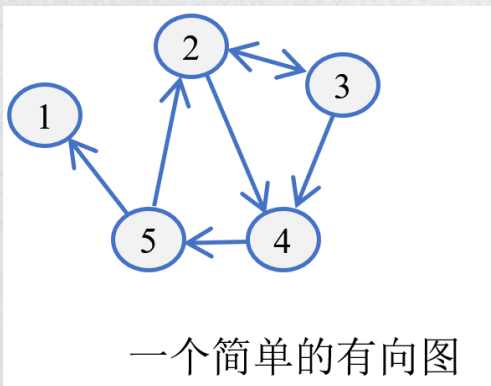


“边列表”表示

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

图数据入门、中心度

- 图论的基本概念：图的表示
- 可以使用三种方式进行表示(Representation)
 - 分别是邻接矩阵(Adjacency Matrix)
 - 边列表(Edge List)
 - 邻接关系列表(Adjacency List)



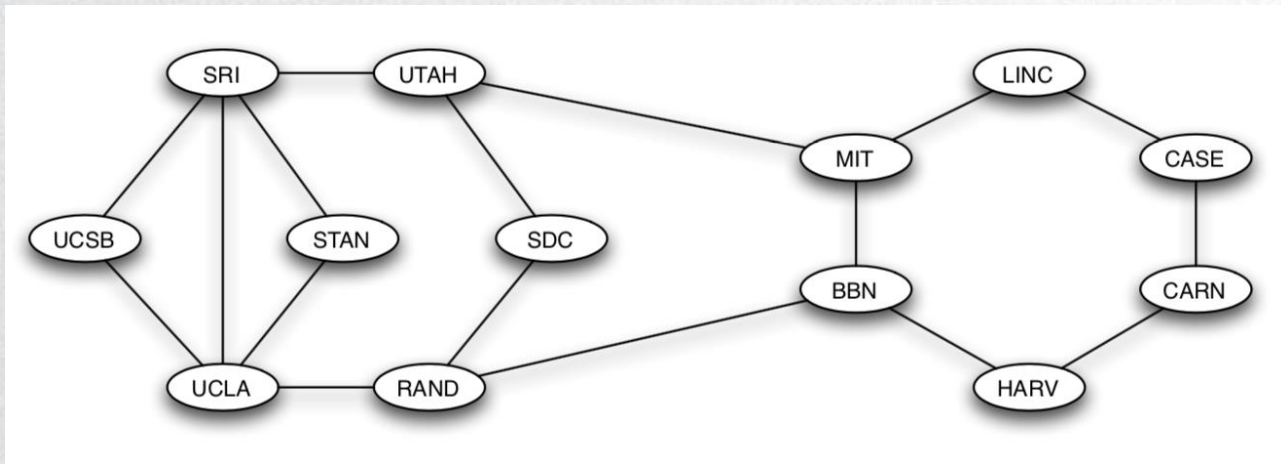
“邻接关系列表”表示

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

图数据入门、中心度

- 图论的基本概念：图的表示

练习：请写出下图的邻接矩阵、边列表与邻接关系列表（写出邻接矩阵的一行）



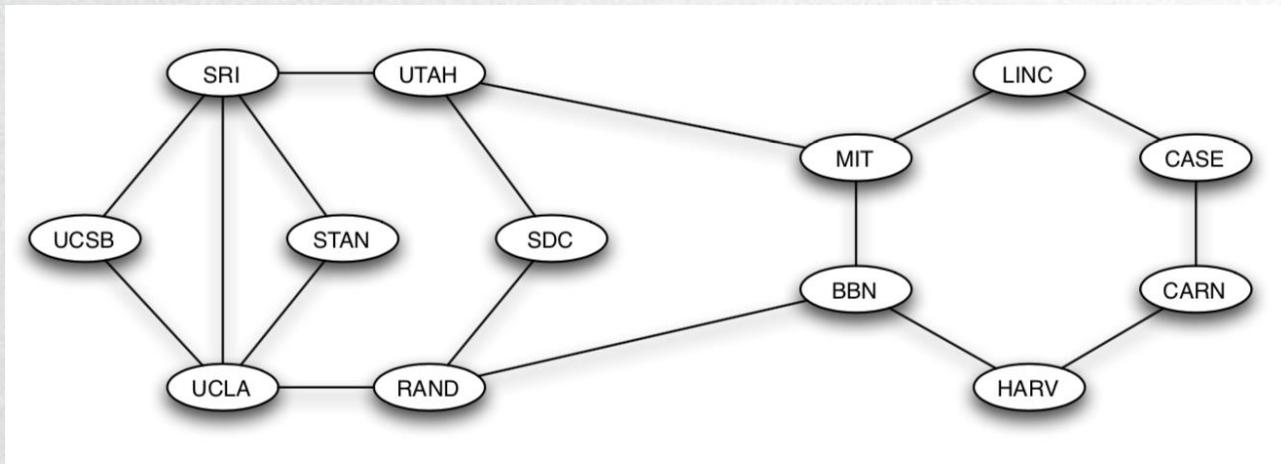
	UCSB	SRI	UCLA	STAN	UTAH	SDC	RAND	MIT	BBN	LINC	HARV	CASE	CARN
SDC													

1970年代的Internet

图数据入门、中心度

- 图论的基本概念：图的表示

练习：请写出下图的邻接矩阵、边列表与邻接关系列表（写出邻接矩阵的一行）

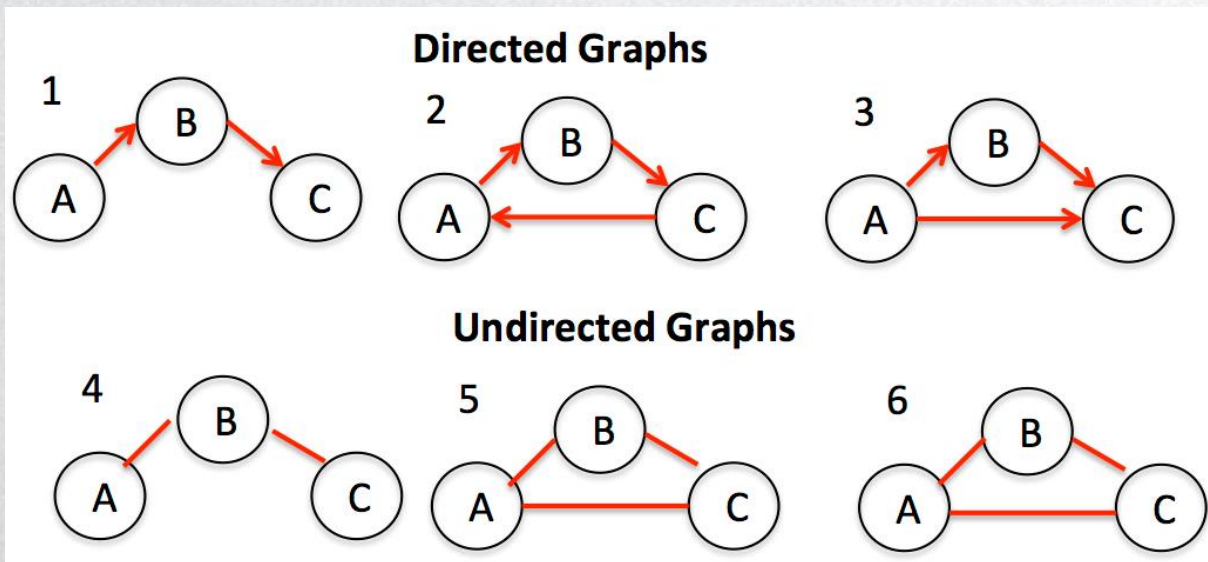


	UCSB	SRI	UCLA	STAN	UTAH	SDC	RAND	MIT	BBN	LINC	HARV	CASE	CARN
SDC	0	0	0	0	1	0	1	0	0	0	0	0	0

1970年代的Internet

图数据入门、中心度

- 图论的基本概念
- 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



微博

微信

在此主要探讨无向图

图数据入门、中心度

- 图论的基本概念：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

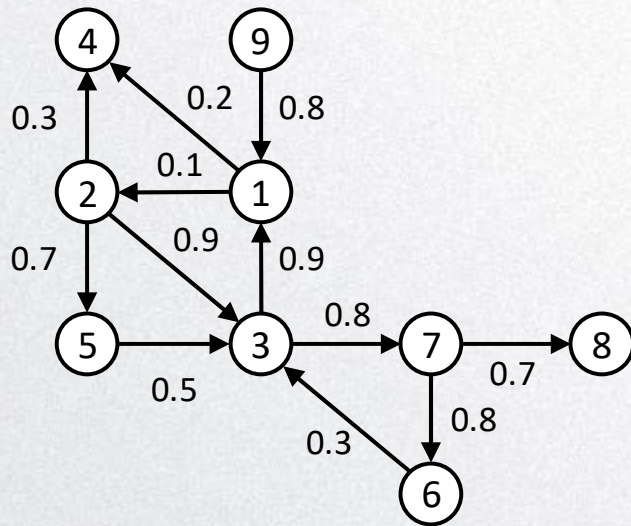
权重：**拥堵**情况



图数据入门、中心度

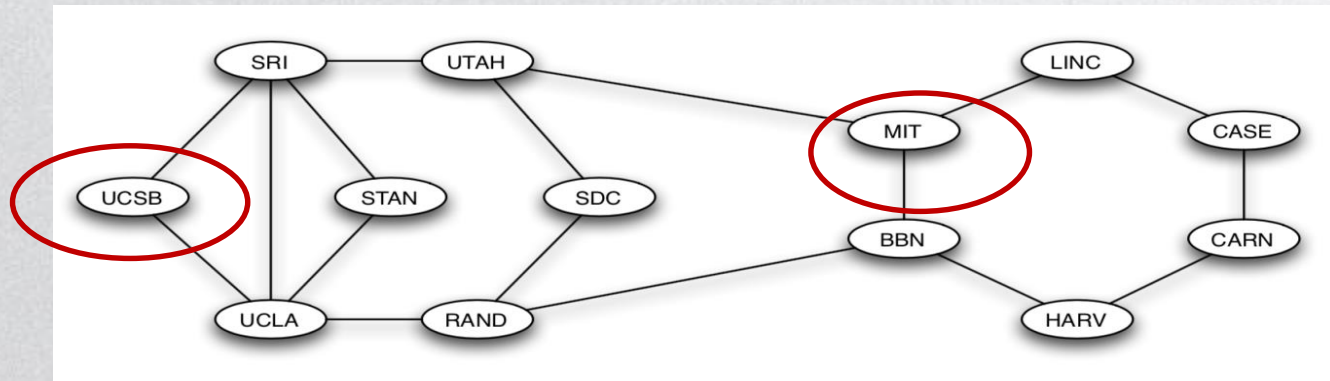
- 图论的基本概念：边的权重
- 图上的每一条边 e 关联一个数字 $w(e)$ ，用来表示边的**重要性或成本**
- 例：信息传播中的权重

- 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



图数据入门、中心度

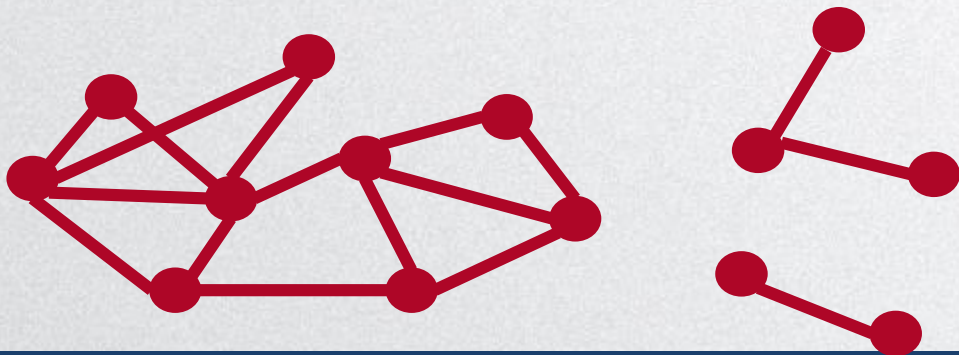
- 图的基本概念：路径
- 路径：图上节点的序列，序列中任意两个相邻的节点都**有边相连**
 - Path $p = (v_0, v_1, \dots, v_m)$ where any $(v_i, v_{i+1}) \in E$
 - 简单路径**：不包含重复节点的路径
 - 环：**起点与终点相同**的路径 $p = (v_0, v_1, \dots, v_m, v_0), (v_i, v_{i+1}) \in E$



- 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

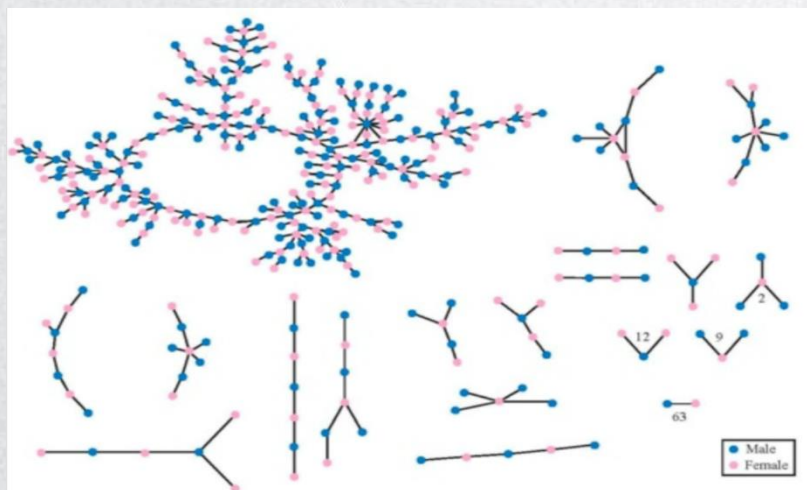
图数据入门、中心度

- 图的基本概念：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1的图的子集 – **独立性**



图数据入门、中心度

- 图的基本概念：连通图与连通分支
- 实例：美国高中生恋爱关系图（边代表二人在18个月内恋爱过）



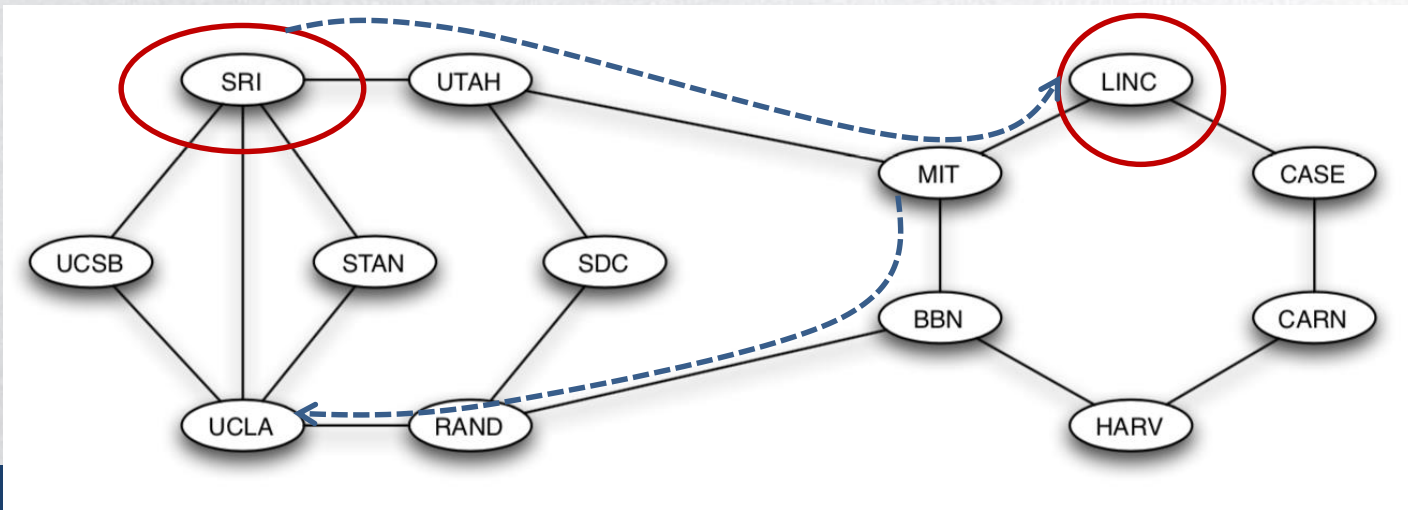
https://www.researchgate.net/publication/245634594_Chains_of_Affection_The_Structure_of_Adolescent_Romantic_and_Sexual_Networks1

图数据入门、中心度

- 图的基本概念：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**
 - **强连通分支** (Strongly Connected Component): 每个顶点和任意其它顶点通过一系列有向连接(Directed Link)到达，也就是任意两个顶点之间是互相可达的；**即对于任意顶点对A和B，既可以从A到达B，也可以从B到达A**
 - **弱连通分支** (Weakly Connected Component): 每个顶点和任意其它顶点，可以从某个方向到达，也就是两个顶点，只需要有一个方向到达即可；即对于任意顶点对A和B，**可以从A到达B，或者可以从B到达A**
 - **大型分支** (Giant Component): 真实网络中，通常存在一个**最大的连通分支**，**它包含了网络中大多数的节点(比如超过80%的节点)**，这样的分支称为大型分支(Giant Component)
 - 在社交网络网站上的用户，一般处在一个超大的连通分支中，真正孤立的顶点或者小的连通分支，所占的比例是很小的
 - 比如，Facebook的数据显示，约99.7%的用户处在一个超大的Giant Component中

图数据入门、中心度

- 图的基本概念：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

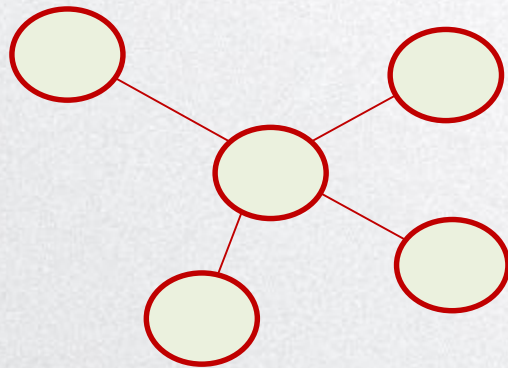
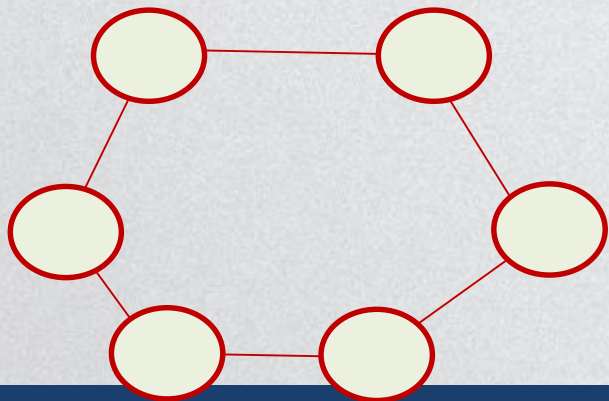


图数据入门、中心度

- 图的基本概念：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)

图数据入门、中心度

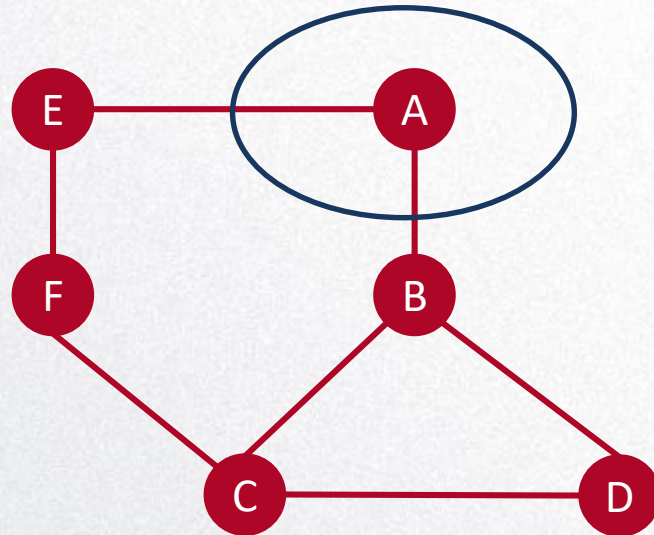
- 图的基本概念：最短路径与距离
- 思考题：如果节点Z在节点X和Y所有的最短路径上，则称Z为X和Y的**关键节点 (Pivot node)**，其中Z与X和Y均不重合
 - 请构造一个图：每个节点均为至少一对节点的关键节点
 - 请构造一个图：每个节点均为至少两对节点的关键节点
 - 请构造一个图：该图中至少包含四个节点，并存在一个节点X，它是图中所有节点对的关键节点（不包含X）



图数据入门、中心度

- 图的基本概念
- 广度优先搜索BFS

A
EB
FCD

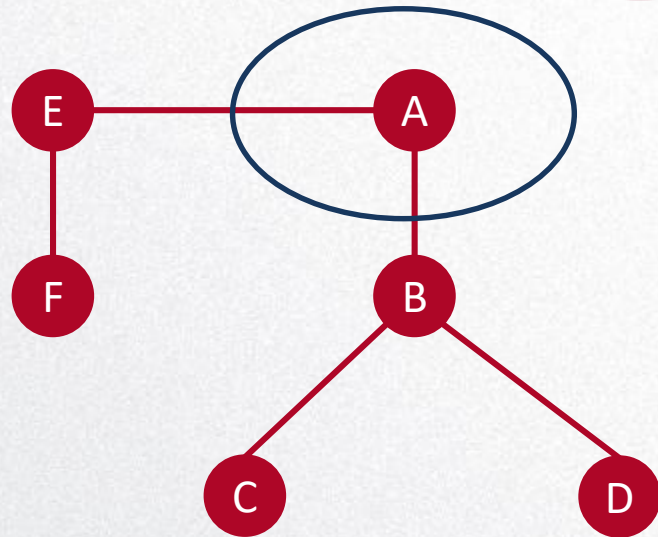


请写出从A开始，做深度优先搜索的过程

图数据入门、中心度

- 图的基本概念
- 深度优先搜索DFS

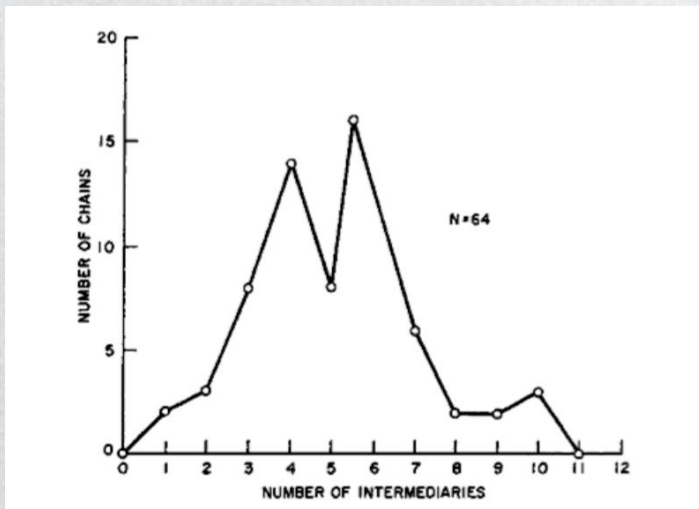
A
E
F
B
C
D



请写出从A开始，做深度优先搜索的过程

图数据入门、中心度

- 图的基本概念：六度分隔现象
- 在这个世界上，任意两个人之间，只隔着六个人
 - 六度分隔



小世界现象（又称小世界效应），也称六度分隔理论（英文：Six Degrees of Separation）

图数据入门、中心度

- 图的基本概念：小世界现象



<https://www.cs.uic.edu/~cornelia/kdsin16/lectures/sna1.pdf>

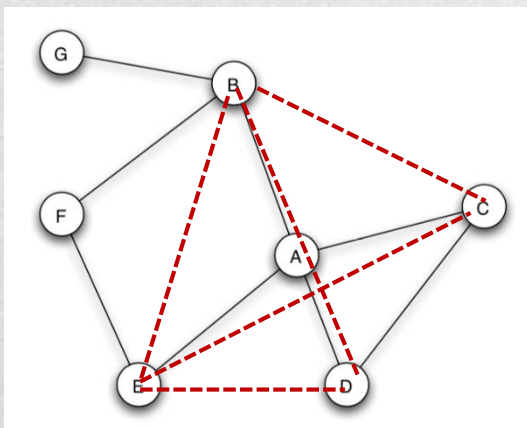


图数据入门、中心度

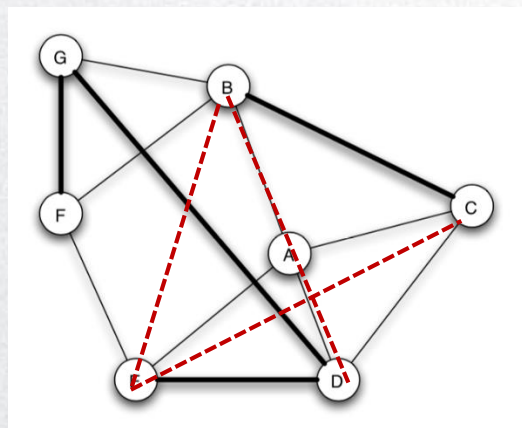
- 图的基本概念：三元闭包 (Triadic Closure)
 - 如果两个人在网络中有共同的好友，他们成为好友的几率也会提升
- 几点原因：如果B和C都有共同的好友A，那么
 - B更有**可能遇到**C——因为他们都与A有交集
 - B和C更有可能**互相信任**——因为他们有共同的好友
 - A更有**可能介绍**B和C认识

图数据入门、中心度

- 图的基本概念：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

图数据入门、中心度





图数据入门、中心度

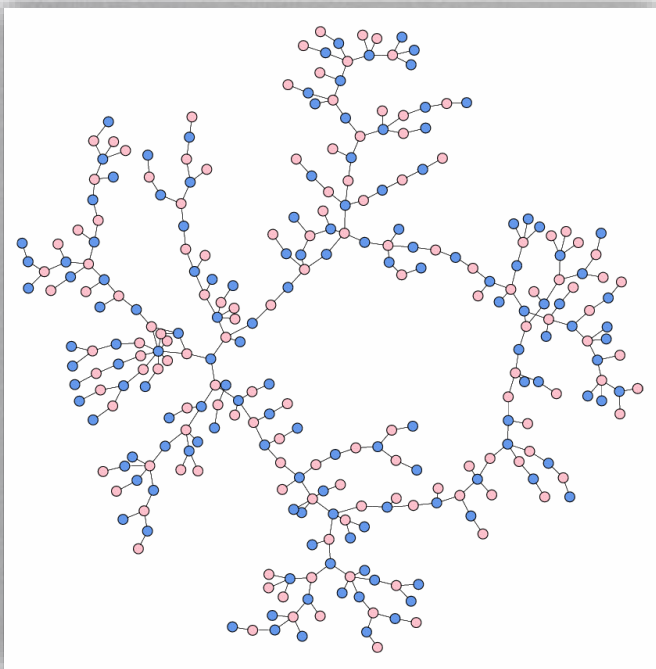
- Graph模块的基本内容
- 基本知识点：
 - **Centrality**: 图里的哪些节点更重要?
 - **Community**: 图是否能够划分为不同的社区
 - **Influence**: 信息如何在图上传播, 如何度量人与人之间的影响力
 - **Query**: 如果利用图回答一些基本的问题
- 基本技能:
 - **Statistical Thinking**: 统计思维
 - **Optimization Thinking**: 优化思维
 - **System Thinking**: 系统思维
- 考核:
 - 作业: 实现PageRank

图数据入门、中心度



图数据入门、中心度

- 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



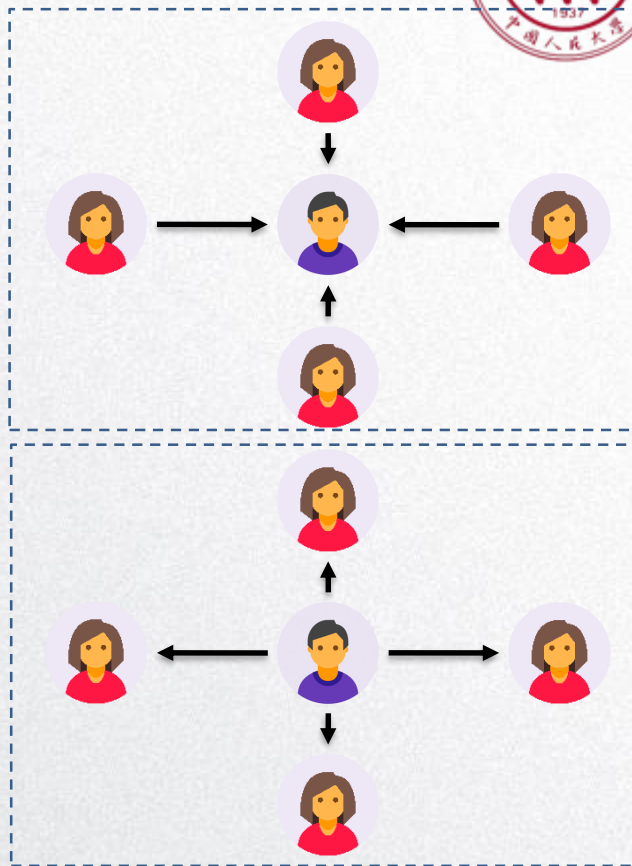
- 思考:
 - 你觉得哪些节点更重要?
 - 你怎么解释这种重要性?

图数据入门、中心度

- 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

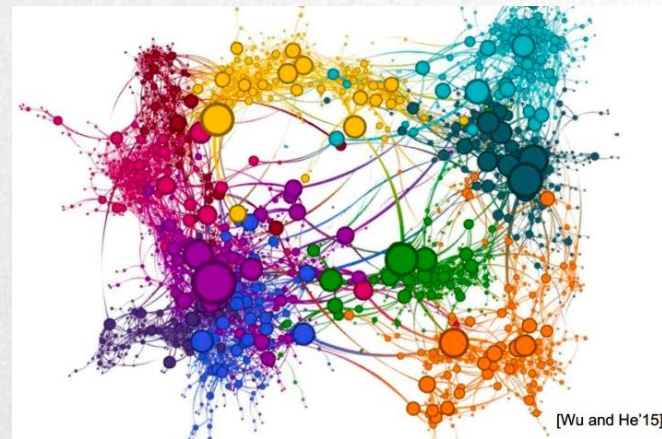


- 思考:
 - 右边两图中男生的重要性一样吗?
 - 你怎么解释这种重要性?



图数据入门、中心度

- 节点中心度 (Node Centrality)
 - 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？



[Wu and He'15]

图数据入门、中心度





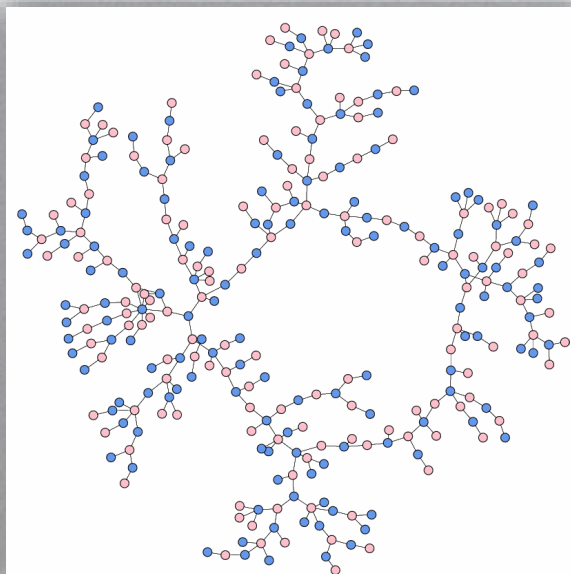
图数据入门、中心度

- 节点中心度 (Node Centrality)
 - 1.基于几何图形的度量方法
 - 2.基于路径的度量方法
 - 3.PageRank算法

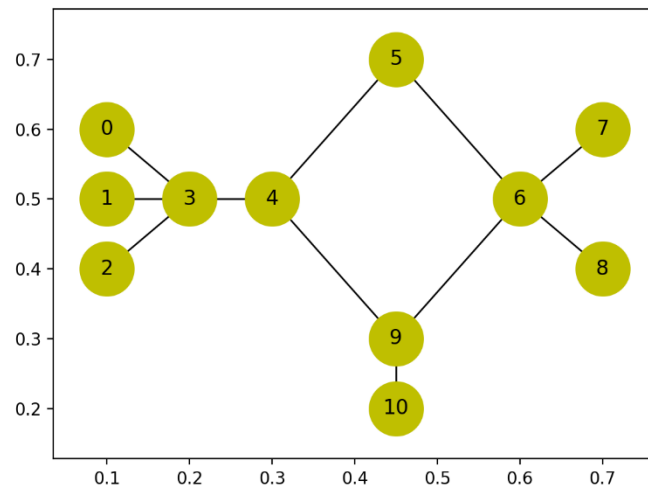
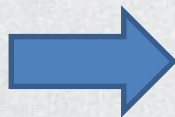
PageRank算法单独讲授
这里讲述前两种方法计算的中心度

图数据入门、中心度

- 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



图数据入门、中心度

- 节点中心度 (Node Centrality)

- 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)

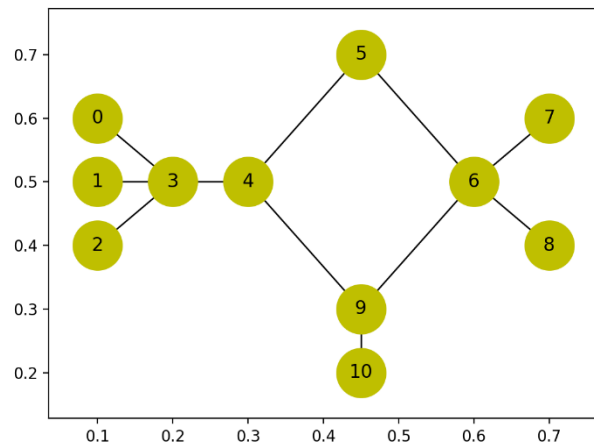
- 即到节点 v 距离为1的节点的个数

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

- 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}

课堂练习, 计算一下节点4的Degree中心度



问题: 公式中为什么除以 $n-1$? 某个节点最多与 $n-1$ 个其它节点有关系

图数据入门、中心度

- 节点中心度 (Node Centrality)
- 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)

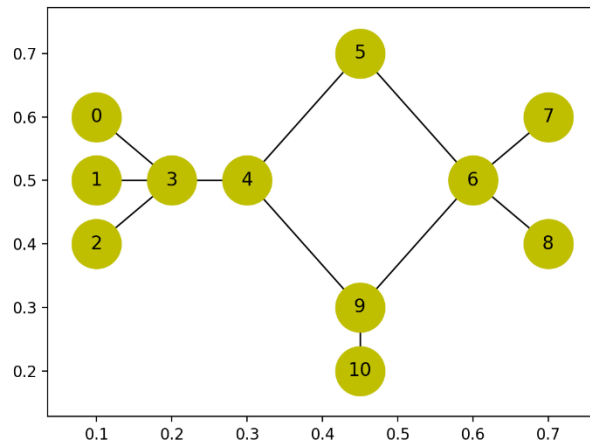
– 即到节点 v 距离为1的节点的个数

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

- 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}

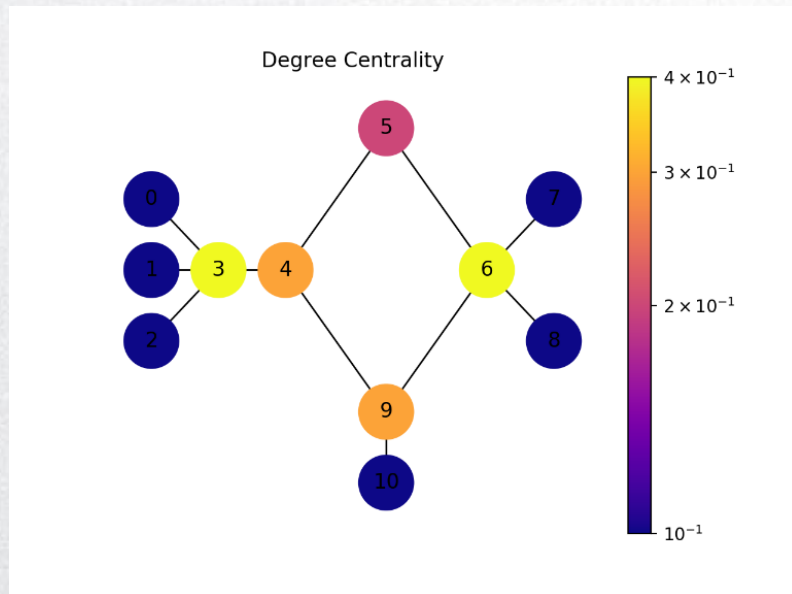
课堂练习, 计算一下节点4的Degree中心度
 $3/(11-1) = 3/10 = 0.3$



图数据入门、中心度

- 节点中心度 (Node Centrality)
- Degree 中心度
- 如何解释Degree Centrality?
 - 恋爱网络: 哥的情史很丰富 😞
 - 微博网络: 大V

{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}



图数据入门、中心度





图数据入门、中心度

- 节点中心度 (Node Centrality)
- Closeness中心度

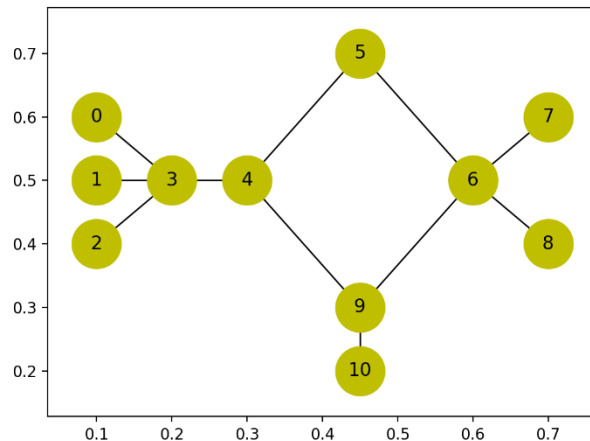
课堂练习，计算一下节点4的Closeness中心度

- 如果一个顶点到其他顶点的距离都比较短，那么它处于网络的中心
- 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}



图数据入门、中心度

- 节点中心度 (Node Centrality)
- Closeness中心度

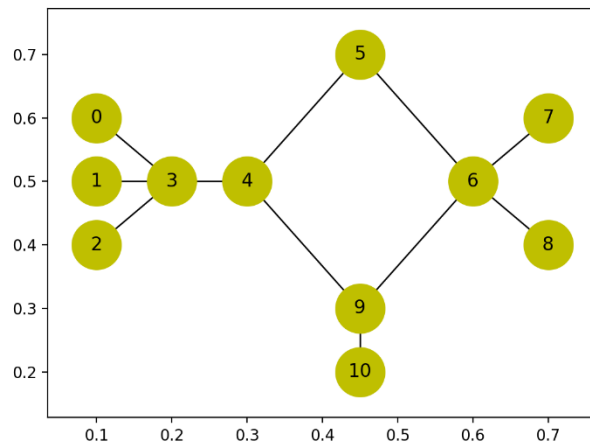
- 如果一个顶点到其他顶点的距离都比较短，那么它处于网络的中心
- 节点 v 的Centrality取决于其它节点到它的距离是否接近
- 其它节点到 v 的距离越接近， v 越重要

课堂练习，计算一下节点4的
Closeness中心度
(11-1)/(2+2+2+1+1+1+2+2+3+3)
=10/19=0.5263

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

$d(u, v)$ 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}



图数据入门、中心度

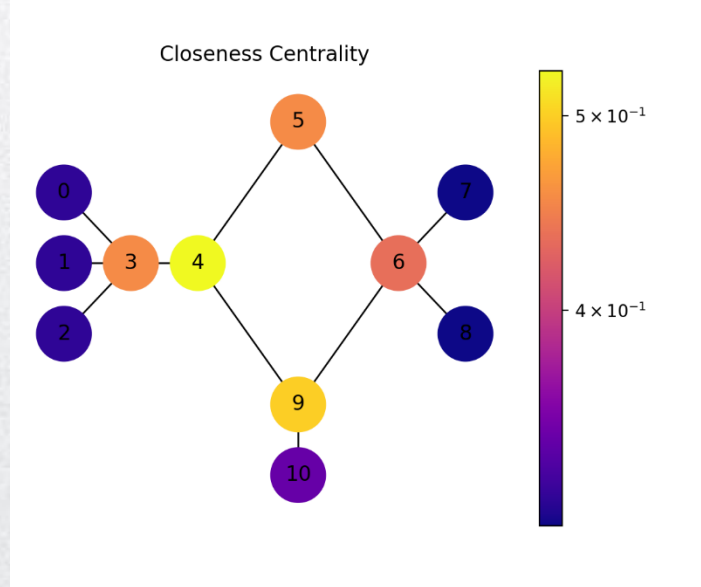
• Closeness中心度

- 如果一个顶点到其他顶点的距离都比较短，那么它处于网络的中心
- 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

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}

• 如何解释Closeness Centrality?

- 恋爱网络：哥不是她的男朋友，就是他女友的前男友 😞
- 道路网络：中心地标建筑



图数据入门、中心度



图数据入门、中心度

- 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

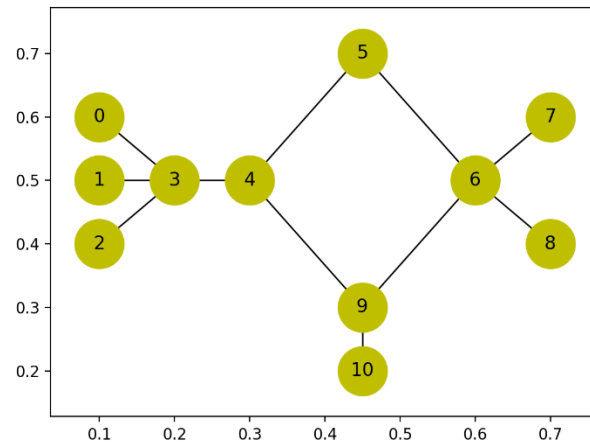
课堂练习, 计算一下节点4的Betweenness中心度

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

$\sigma(s,t)$ 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}



图数据入门、中心度



课堂练习，计算一下节点4的Betweenness中心度

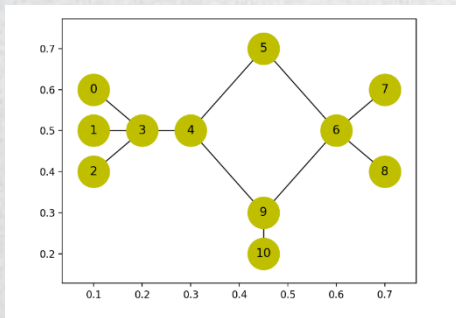
- 计算相当复杂，后续

图数据入门、中心度

• Betweenness Centrality

– 节点 v 的Centrality取决于

在其它节点最短路径



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

$\sigma(s, t)$ 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

某两个点的最短路径

对 betweenness(4) 的贡献

0 1 : [[0, 3, 1]]
0 2 : [[0, 3, 2]]
0 3 : [[0, 3]]
0 4 : [[0, 3, 4]]
0 5 : [[0, 3, 4, 5]]
0 6 : [[0, 3, 4, 5, 6],
0 7 : [[0, 3, 4, 5, 6, 7],
0 8 : [[0, 3, 4, 5, 6, 8],
0 9 : [[0, 3, 4, 9]]
0 10 : [[0, 3, 4, 9, 10]]

[0, 3, 4, 9, 6]]
[0, 3, 4, 9, 6, 7]]
[0, 3, 4, 9, 6, 8]]

$\frac{\sigma_{st}(4)}{\sigma_{st}}$
0
0
0
0
1
1
1
1
1
1

1, 每对节点, 看看端点不为4、路径中间经过4的路径
3, 每对节点, 所有路径里面, 上述路径的比例

1 2 : [[1, 3, 2]]
1 3 : [[1, 3]]
1 4 : [[1, 3, 4]]
1 5 : [[1, 3, 4, 5]]
1 6 : [[1, 3, 4, 5, 6],
1 7 : [[1, 3, 4, 5, 6, 7],
1 8 : [[1, 3, 4, 5, 6, 8],
1 9 : [[1, 3, 4, 9]]
1 10 : [[1, 3, 4, 9, 10]]

[1, 3, 4, 9, 6]]
[1, 3, 4, 9, 6, 7]]
[1, 3, 4, 9, 6, 8]]

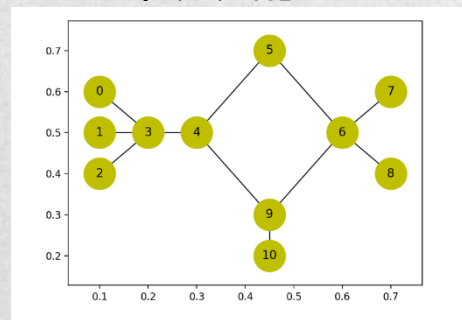
0
0
0
0
1
1
1
1
1
1

图数据入门、中心度

• Betweenness Centrality

– 节点 v 的Centrality取决于

在其它节点的
最短路径上



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

$\sigma(s,t)$ 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

某两个点的最短路径	对betweenness(4)的贡献 $\frac{\sigma_{st}(4)}{\sigma_{st}}$
2 3 : [[2, 3]]	0
2 4 : [[2, 3, 4]]	0
2 5 : [[2, 3, 4, 5]]	1
2 6 : [[2, 3, 4, 5, 6], [2, 3, 4, 9, 6]]	1
2 7 : [[2, 3, 4, 5, 6, 7], [2, 3, 4, 9, 6, 7]]	1
2 8 : [[2, 3, 4, 5, 6, 8], [2, 3, 4, 9, 6, 8]]	1
2 9 : [[2, 3, 4, 9]]	1
2 10 : [[2, 3, 4, 9, 10]]	1
3 4 : [[3, 4]]	0
3 5 : [[3, 4, 5]]	1
3 6 : [[3, 4, 5, 6], [3, 4, 9, 6]]	1
3 7 : [[3, 4, 5, 6, 7], [3, 4, 9, 6, 7]]	1
3 8 : [[3, 4, 5, 6, 8], [3, 4, 9, 6, 8]]	1
3 9 : [[3, 4, 9]]	1
3 10 : [[3, 4, 9, 10]]	1

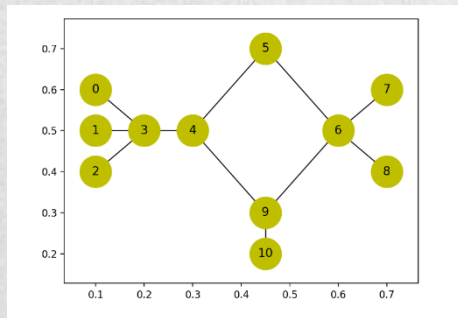
1, 每对节点, 看看
端点不为4、路径中
间经过4的路径
3, 每对节点, 所有
路径里面, 上述路
径的比例

图数据入门、中心度

• Betweenness Centrality

– 节点 v 的Centrality取决于

在其它节点的最短路径上



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

$\sigma(s,t)$ 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

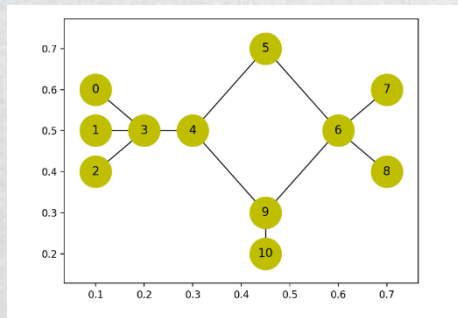
某两个点的最短路径	对betweenness(4)的贡献 $\frac{\sigma_{st}(4)}{\sigma_{st}}$
4 5 : [[4, 5]]	0
4 6 : [[4, 5, 6], [4, 9, 6]]	0
4 7 : [[4, 5, 6, 7], [4, 9, 6, 7]]	0
4 8 : [[4, 5, 6, 8], [4, 9, 6, 8]]	0
4 9 : [[4, 9]]	0
4 10 : [[4, 9, 10]]	0
5 6 : [[5, 6]]	0
5 7 : [[5, 6, 7]]	0
5 8 : [[5, 6, 8]]	0
5 9 : [[5, 4, 9], [5, 6, 9]]	0.5
5 10 : [[5, 4, 9, 10], [5, 6, 9, 10]]	0.5

1, 每对节点, 看看端点不为4、路径中间经过4的路径
3, 每对节点, 所有路径里面, 上述路径的比例

图数据入门、中心度

Betweenness Centrality

— 节点 v 的Centrality取决于



在其它节点的最短路径上

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

$\sigma(s,t)$ 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

某两个点的最短路径

对betweenness(4)的贡献 $\frac{\sigma_{st}(4)}{\sigma_{st}}$

6 7 : [[6, 7]]
6 8 : [[6, 8]]
6 9 : [[6, 9]]
6 10 : [[6, 9, 10]]

7 8 : [[7, 6, 8]]
7 9 : [[7, 6, 9]]
7 10 : [[7, 6, 9, 10]]

8 9 : [[8, 6, 9]]
8 10 : [[8, 6, 9, 10]]

9 10 : [[9, 10]]

0
0
0
0

0
0
0

0
0

0

1, 每对节点, 看看端点不为4、路径中间经过4的路径
3, 每对节点, 所有路径里面, 上述路径的比例

累加起来为25

- 对25进行标准化,
- 除以 $(n-1)(n-2)/2=10*9/2=45$
- $25/45=0.56$

图数据入门、中心度

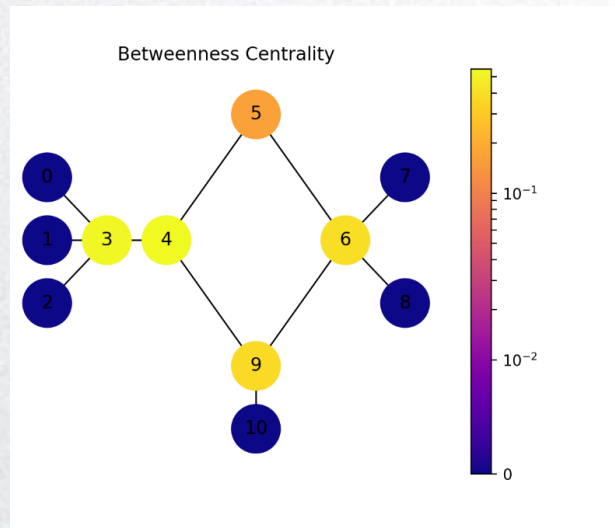


图数据入门、中心度

- 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

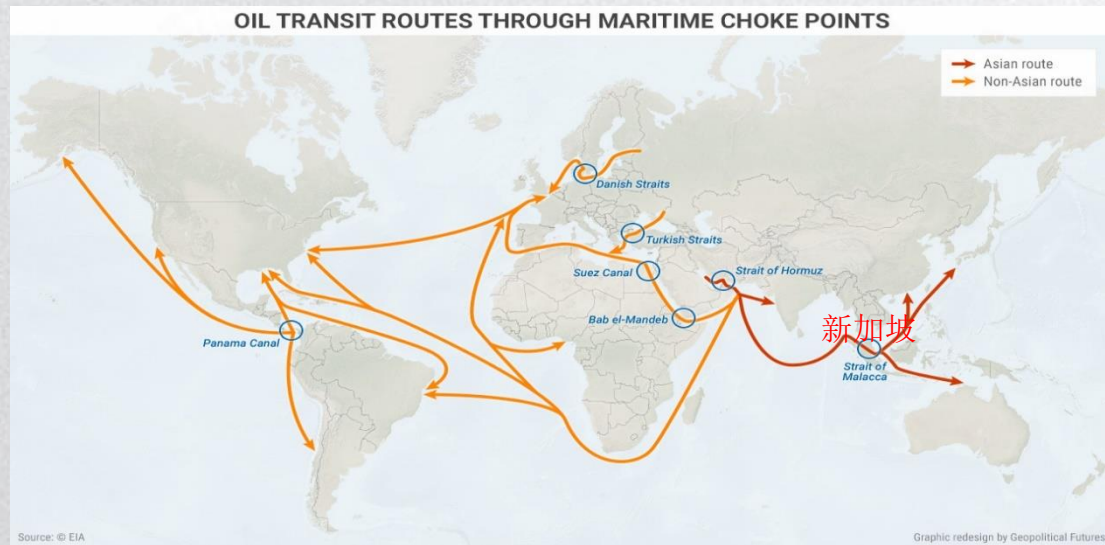
betweenness Centrality {0: 0.0, 1: 0.0, 2: 0.0, **3: 0.53**, **4: 0.56**, 5: 0.17, **6: 0.4**, 7: 0.0, 8: 0.0, 9: 0.37, 10: 0.0}

- 如何解释Betweenness Centrality?
 - **恋爱**网络: 没有哥, 那些妹子们这辈子也不会有什么关联 😞
 - **贸易**网络: 贸易枢纽hubs



图数据入门、中心度

- Betweenness中心度
- 思考：新加坡为什么成为全球贸易港？



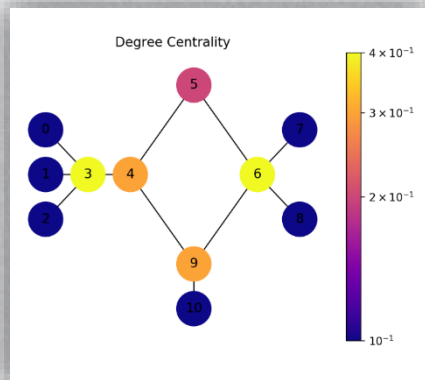
波斯湾石油运输航路图

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)

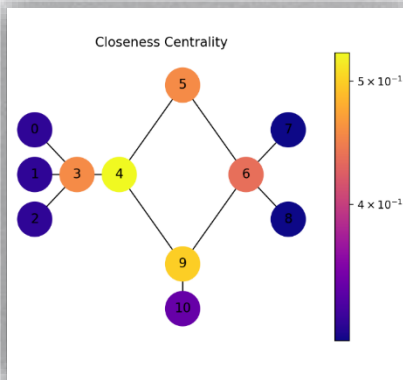
地缘政治与
Betweenness Centrality

图数据入门、中心度

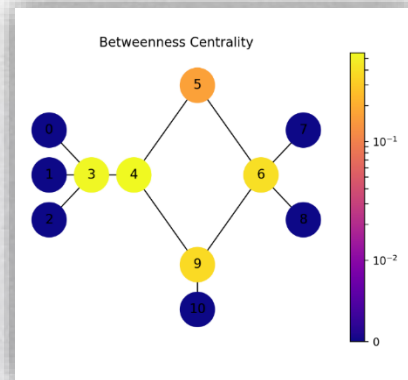
- 三种Centrality度量的比较



Degree Centrality
3/6/4/9



Closeness Centrality
4/9/5/3/6



Betweenness Centrality
4/3/6/9

Degree

{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}

Closeness

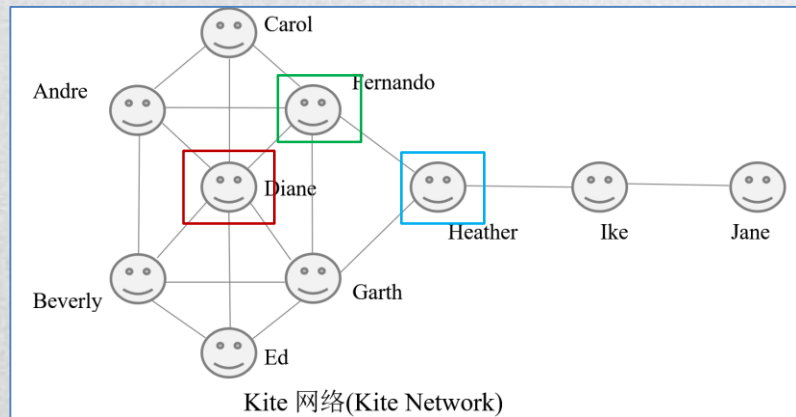
{0: 0.32, 1: 0.32, 2: 0.32, 3: 0.45, 4: 0.53, 5: 0.45, 6: 0.43, 7: 0.31, 8: 0.31, 9: 0.5, 10: 0.34}

Betweenness

{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}

图数据入门、中心度

- 另一个实例
- 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

图数据入门、中心度



图数据入门、中心度

- 中心度计算Python实例分析

```
$ pip install networkx
```

NetworkX

Stable (notes)

2.2 — September 2018

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

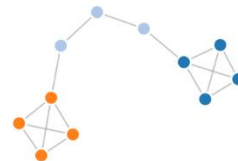
Latest (notes)

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.





图数据入门、中心度

- 中心度计算Python实例分析

名称	类型	大小	修改日期 ^
 01qraph high school love.py	Python File	3 KB	2021/10/25 16:00

<https://aksakalli.github.io/2017/07/17/network-centrality-measures-and-their-visualization.html>



图数据入门、中心度

- 中心度计算Python实例分析
 - 绘制函数

```
1 import networkx as nx
2 import matplotlib.pyplot as plt
3 import matplotlib.colors as mcolors
4
5 def draw(G, pos, measures, measure_name, num):
6
7     plt.figure(num)
8     nodes = nx.draw_networkx_nodes(G, pos, node_size=250, cmap=plt.cm.plasma,
9                                     node_color=list(measures.values()),
10                                    nodelist=measures.keys())
11     nodes.set_norm(mcolors.SymLogNorm(linthresh=0.01, linscale=1, base=10))
12     # Labels = nx.draw_networkx_labels(G, pos)
13     edges = nx.draw_networkx_edges(G, pos)
14
15     plt.title(measure_name)
16     plt.colorbar(nodes)
17     plt.axis('off')
18     plt.show()
```




图数据入门、中心度

- 中心度计算Python实例分析
 - 图的创建函数

```
20 def school_dating_graph():
21     students = set(range(11))
22     G = nx.Graph()
23     G.name = "Simple Dating Graph"
24     G.add_nodes_from(students)
25     dating_rel = [(0,3), (1,3), (2,3), (3,4), (4,5), (4,9),
26                  (5,6), (6,7), (6,8), (6,9), (9,10)]
27     G.add_edges_from(dating_rel)
28     # You may want to try automatic layout
29     #pos = nx.spring_layout();
30     pos = {0: [0.1, 0.6], 1: [0.1, 0.5], 2: [0.1, 0.4], 3: [0.2, 0.5],
31            4: [0.3, 0.5], 5: [0.45, 0.7], 6: [0.6, 0.5], 7: [0.7, 0.6],
32            8: [0.7, 0.4], 9: [0.45, 0.3], 10: [0.45, 0.2]}
33     return G, pos
```

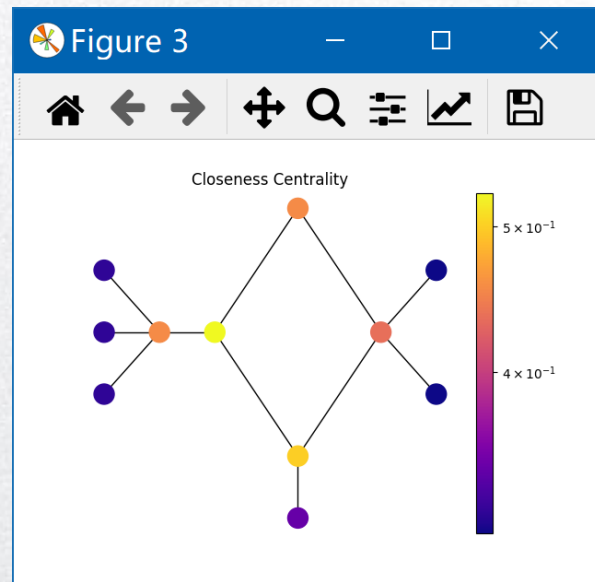
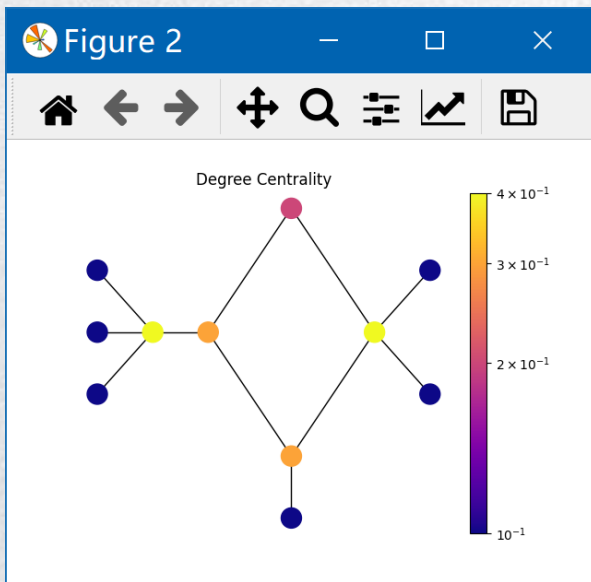
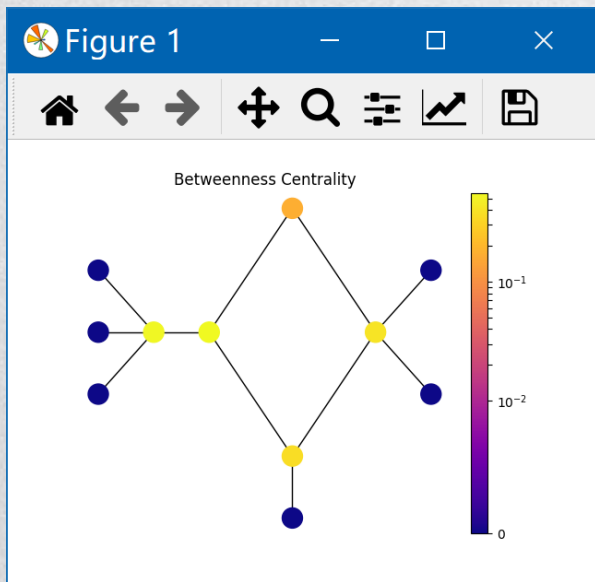
图数据入门、中心度

- 中心度计算Python实例分析
 - 计算并且显示degree,closeness,betweenness中心度

```
35 G,pos = school_dating_graph()
36 draw(G, pos, nx.betweenness_centrality(G), 'Betweenness Centrality',1)
37 print("betweenness_centrality",nx.betweenness_centrality(G))
38
39 draw(G, pos, nx.degree_centrality(G), 'Degree Centrality',2)
40 print("degree_centrality",nx.degree_centrality(G))
41
42 draw(G, pos, nx.closeness_centrality(G), 'Closeness Centrality',3)
43 print("closeness_centrality",nx.closeness_centrality(G))
44
```

图数据入门、中心度

- 中心度计算Python实例分析
 - 计算并且显示degree, closeness, betweenness中心度



图数据入门、中心度

- 中心度计算Python实例分析
 - 计算所有节点间的所有最短路径

```
46 pair_list = [  
47     [0,1],[0,2],[0,3],[0,4],[0,5],[0,6],[0,7],[0,8],[0,9],[0,10],  
48     [1,2],[1,3],[1,4],[1,5],[1,6],[1,7],[1,8],[1,9],[1,10],  
49     [2,3],[2,4],[2,5],[2,6],[2,7],[2,8],[2,9],[2,10],  
50     [3,4],[3,5],[3,6],[3,7],[3,8],[3,9],[3,10],  
51     [4,5],[4,6],[4,7],[4,8],[4,9],[4,10],  
52     [5,6],[5,7],[5,8],[5,9],[5,10],  
53     [6,7],[6,8],[6,9],[6,10],  
54     [7,8],[7,9],[7,10],  
55     [8,9],[8,10],  
56     [9,10]  
57 ]  
58 for pair in pair_list:  
59     first = pair[0]  
60     second = pair[1]  
61     print(first, second, ":", [p for p in nx.all_shortest_paths(G, source=first, target=second)])
```

```
01 : [[0, 3, 1]]  
02 : [[0, 3, 2]]  
03 : [[0, 3]]  
04 : [[0, 3, 4]]  
05 : [[0, 3, 4, 5]]  
06 : [[0, 3, 4, 5, 6], [0, 3, 4, 9, 6]]  
07 : [[0, 3, 4, 5, 6, 7], [0, 3, 4, 9, 6, 7]]  
08 : [[0, 3, 4, 5, 6, 8], [0, 3, 4, 9, 6, 8]]  
09 : [[0, 3, 4, 9]]  
10 : [[0, 3, 4, 9, 10]]  
12 : [[1, 3, 2]]  
13 : [[1, 3]]  
14 : [[1, 3, 4]]  
15 : [[1, 3, 4, 5]]  
16 : [[1, 3, 4, 5, 6], [1, 3, 4, 9, 6]]  
17 : [[1, 3, 4, 5, 6, 7], [1, 3, 4, 9, 6, 7]]  
18 : [[1, 3, 4, 5, 6, 8], [1, 3, 4, 9, 6, 8]]  
19 : [[1, 3, 4, 9]]  
110 : [[1, 3, 4, 9, 10]]  
23 : [[2, 3]]  
24 : [[2, 3, 4]]  
25 : [[2, 3, 4, 5]]  
26 : [[2, 3, 4, 5, 6], [2, 3, 4, 9, 6]]  
27 : [[2, 3, 4, 5, 6, 7], [2, 3, 4, 9, 6, 7]]  
28 : [[2, 3, 4, 5, 6, 8], [2, 3, 4, 9, 6, 8]]  
29 : [[2, 3, 4, 9]]  
210 : [[2, 3, 4, 9, 10]]  
34 : [[3, 4]]  
35 : [[3, 4, 5]]  
36 : [[3, 4, 5, 6], [3, 4, 9, 6]]  
37 : [[3, 4, 5, 6, 7], [3, 4, 9, 6, 7]]  
38 : [[3, 4, 5, 6, 8], [3, 4, 9, 6, 8]]  
39 : [[3, 4, 9]]  
310 : [[3, 4, 9, 10]]  
45 : [[4, 5]]  
46 : [[4, 5, 6], [4, 9, 6]]  
47 : [[4, 5, 6, 7], [4, 9, 6, 7]]  
48 : [[4, 5, 6, 8], [4, 9, 6, 8]]  
49 : [[4, 9]]  
410 : [[4, 9, 10]]  
56 : [[5, 6]]  
57 : [[5, 6, 7]]  
58 : [[5, 6, 8]]  
59 : [[5, 4, 9], [5, 6, 9]]  
510 : [[5, 4, 9, 10], [5, 6, 9, 10]]  
67 : [[6, 7]]  
68 : [[6, 8]]  
69 : [[6, 9]]  
610 : [[6, 9, 10]]  
78 : [[7, 6, 8]]  
79 : [[7, 6, 9]]  
710 : [[7, 6, 9, 10]]  
89 : [[8, 6, 9]]  
810 : [[8, 6, 9, 10]]  
910 : [[9, 10]]
```

