[1. Introduction 2](#_Toc87)

[1.1 复杂网络无处不在 2](#_Toc29319)

[1.2 小世界网络 2](#_Toc1946)

[1.3 网络科学的崛起 3](#_Toc17514)

[2. 网络和图 4](#_Toc7445)

[2.1 网络和图 4](#_Toc7219)

[2.2 图论概念 4](#_Toc8234)

[2.3 三元闭包 4](#_Toc31296)

[2.4 社会归属网 5](#_Toc15421)

# 1. Introduction

## 1.1 复杂网络无处不在

互联网的时间变化：

1969 ARPAnet(Advanced Research Projects Agency Network美国高等研究计划署网络)

1986 NSFNET(The National Science Foundation Network 国家科学基金会网络)

1995 INTERNET 互联网

网络分层：IP层，路由层，域名层

【移动通信网络】

人类通讯动态的呼叫模式 (幂律分布 [基于机器人的呼叫者，电信欺诈，电话销售])

(韦博分布 [普通用户])

网络科学：1.构建网络 2.研究细节 3.研究动态 4.关联动态与突发事件

【全球航空公司网络】

【交通网络】(公路，铁路，水路)

从最近的城市开始地面旅游超过48小时的不超过10%的土地。

！！！网络让世界更加互联：互联网，移动通信网，交通网

【悖论】交通博弈

【金融网络】

搜索引擎，级联故障，电网，金融危机

复杂网络在金融上的应用

主要课题：复杂金融网络对经济稳定性的挑战

网络科学家，经济学家和监管机构首次合作讨论

！！！太中心不行，太大不行。关注节点位置而不是大小。

【生物网络】

网络生物学，网络医学，生态网络

人的大脑有100~1000亿个神经元

【专利网络】

【软件系统网络】为大型软件系统提供探索性分析。设计，效率，可重用性，脆弱性，可控性

【语言网络】

【科学合作网络】

【恐怖网络】

【篮球队策略网络】球员是节点，球的轨迹是边

【组织结构】

【如何评估学生的表现】早期发现问题能及时纠正，可以减少辍学人数。 连续的方式

【成功研讨会】社会科学家，计算机科学家，经济学家，物理学家，数学家，讨论成功的规律。过去几年个人效绩的大量数据促使科学家探索个人成功的路径模式。

【社会动员关键时间】 复杂网络的时间演变

【网络的网络】网络相互依赖

【网络科学】什么是连接，网络的出现和演化有共同的基本规律和可再现的机制驱动的

## 1.2 小世界网络

世界上任意两个人之间的平均距离是多少？

【1967小世界实验】信件链 20%到达，距离6.5

【六度分离】美国任意2个人的距离是6

2002实验:4.0 2006msn:6.6 2009renren:5.38 2011facebook:4.74 贝肯数 埃尔德什数

## 1.3 网络科学的崛起

网络地图的出现，网络特性的普遍性

【网络传播】通讯，运输，电网，社会，生物，经济

【社会网络传播】病毒，时尚，行为，谣言，意见，信仰

【理论科学】图论，博弈论，统计物理学，计算机科学

【应用科学】通讯科学，电力工程，生命科学，社会科学

【网络建模】测试数据，分析行为，发现财产，设计性能

【目标】收集足够信息以便：正确的描述，正确的量化，数学的公式，合理的预测，强力的控制

【游戏改变者】对于每种技术，前十年发展，第二个十年市场跟进

【挑战】将复杂问题分解为小问题

【特点】跨学科，数据驱动，定量，计算性

【典型网络科学研究】发现，建模，验证

【数学风格】

【物理风格】

【控制风格】复杂网络可控性，复杂网络可观测性

【计算机科学风格】社区检测风格，链路预测，推荐算法

【实验风格】6100万人的社会影响和政治动员实验

【社会影响偏差】随机实验

【大数据时代】数据资源，数据产业，数据战略，数据科学

【数据量增长快？】1.数据产生方式改变，2.人类生活依赖数据

【大数据】海量数据或巨量数据，其规模巨大到无法通过目前主流的计算机系统在合理时间内获取、存储、管理、处理并提炼以帮助使用者的策略

【特点】Value(价值密度低)，Velocity(快速)，Volume(数据量大)，Vari ety(多样)，

Complexity(复杂度)

# 2. 网络和图

## 2.1 网络和图

图=事物+联系=节点(vertex,node,point)+边(连接,链接,关系,联系;edge,link,tie)

【起源】哥尼斯堡七桥问题：每架桥只经过一次，走过所有桥。

【图论之父】Leonhard Eular(莱昂哈德 欧拉) ：陆地为节点，桥为边

【证明】奇数个边的节点必须为起点或终点，七桥问题有4个奇数边的节点，所以不能遍历。

【贡献】将问题用图表示，能变得更简单易于处理。

图和网络的结构是理解复杂世界的关键

隐藏在图和网络中的属性，会限制或增强我们处理问题的能力

图论是网络的基础

【网络(图)表示现实世界】不同性质的实际系统可以具有相同的网络表示

例如：互联网，好莱坞演员网络，蛋白质作用网络

【区别】网络：通常指实际系统，术语：网络(network)，节点(node)，连接(link)

图论：网络的数学表示，术语：图(graph)，顶点(vertex)，边(edge)

## 2.2 图论概念

【数学表示】G(V,E),V={a,b,...},E包含于{(x,y)|x,y属于V,x不等于y}

【异构】

【图的个数】=2^C(n,2) Polya定理：?

【分类】(有标号图，无标号图)(有向图，无向图)(权重图，无权重图)

【简单图】无多边，无自边

【完全图】边 = n(n-1)/2

【计算机表示】邻接矩阵，关联矩阵

【路径】path,通路：节点序列(相邻节点存在边) 长度：节点数-1 长度：涉及的边条数，

【距离】两个节点之间的最短路径长度

【连通图】任何两个节点之间存在路径

【连通分量】连通子图，不包含在其他连通子图中，孤立的子图

关注路径长度而不是物理距离

【最大连通分量】

【最大分量】包含图的显著部分

【桥】特别性质的边，删除之后两个端点就不再有路(删除会增加连通分量)

【捷径】删除之后两个节点之间的距离至少为3(桥是捷径的特例)

【有向路径】节点序列，相邻节点之间有从前往后的有向边

【强连通分量】任意两个节点之间存在有向路径，不真包含在其他强连通分量中

【二部图】没有长度为奇数的圈：？

从任意节点出发，在广度优先搜索过程中，同一层节点有边，则存在长度为奇数的圈

【遍历】广度优先搜索

【应用】合作图，交流图，信息链接图

## 2.3 三元闭包

【三元闭包】如果两个互不认识的人有一个共同的朋友，则将来成为朋友的可能性提高

【聚集系数】刻画网络结构中的节点属性

节点A的聚集系数

=与A相邻任意两个朋友之间也是朋友的概率

=与A相邻的朋友对的个数/总的对数

【另一种描述】两个互不认识的人的共同朋友数越多，则在未来成为朋友的可能性越大

【验证】电子邮件网络：100对节点，某时刻没边，有5个共同好友，一个月里，20对节点发生通信，80没有，则在一个月里成为朋友的概率为0.2

【共同朋友个数】 = 邻接矩阵相乘

【强关系、弱关系】

【嵌入性】边的属性，一条边两端共同的邻里个数

【结构洞】一个节点，移除该节点会使网络变成多个连通分量的节点

【图划分算法】

分割法：逐步去掉“跨接边”

聚集法：“滚雪球”

近似：准确与效率的平衡

【Girvan-Newman方法】？

【介数】一条边承载的一种“流量” 流量均分到所有最短路径上，

考虑所有对后，累积流量就是介数

【介数计算】从一个节点开始，做宽度优先搜索，确定最短路径条数，确定向其他节点发送流量时，经过每条边的流量。重复以上过程，累计除以2，即得每条边的介数

【总结】边的属性用强弱性、嵌入性表示，点的属性可以用聚集系数、结构洞表示

【强三元闭包】如果一个节点有两个强关系的邻居，则这两个邻居不能什么关系都没有

【捷径】A符合强三元闭包，则与A相连的捷径为弱关系。

【验证】共同朋友越多，关系强度越高

## 2.4 社会归属网

【社会归属网】描述从属关系与社会关系

【社团闭包】选择。由于参与同一件事，原本没联系的人建立联系

【会员闭包】影响。由于朋友参与这件事，另一个人也加入这件事

【同质性】两种特征。节点数n,边数e,不同节点占比p,1-p,两端节点相同的边s

【自动化】n:人物节点个数 m:社交聚点个数 A(i,j):i与j的连接关系，B(i,j):i是否加入j

A(i,\*):A的第i行向量 B(\*,j):B的第j列向量

三元闭包：若A(i,j)=0 且A(i,\*)\*A(j,\*)>=3,则A(i,j)将=1

社团闭包：若A(i,j)=0 且B(i,\*)\*B(j,\*)>=2,则A(i,j)将=1

会员闭包：若B(i,j)=0 且A(i,\*)\*B(\*,j)>=2,则B(i,j)将=1

【谢林模型】起源：芝加哥黑人居住区的改变

同质性动态：越来越多的黑人在某个区域聚集

模型:一个空间居住两类自然特征人群，邻居指紧邻的居住单元（8个）。

t<3就搬家，随机执行

居住隔离为例，模拟了同质性的动态变化

1. 小世界现象

【六度分割】斯坦利·米尔格兰姆(Stanley Milgram)

1967年 发信件 只能发给熟人，继续转发 结果1/3的信件经过6次转发到目标

90年代 贝肯数 133万名演员计算出平均“贝肯数”为2.981，最大为8

埃尔德什数 数学论文作者与埃尔德什的合作距离 40万名数学家平均4.65 最大为13

2003年 6万个用户转发，中间路径5-7步

网页之间直径为18.59次

【现象】网络中两节点包含丰富的短路径，任意两节点存在短路径概论很高

短视搜索能有效找到短路径。（短视搜索：每一步只能看到邻居节点）

对于“十分稀疏”的社会网络，不是必然

完全随机的网络没有这样的性质

【形成】同质性(共同关系，邻里关系，同学，同事，共同兴趣，对应大量“三角形”)

弱关系(偶然原因认识“远程”朋友)

【Watts-Strogatz模型】1.没有确定结构2.高度有序与随机连接3.改变连接随机性p,可以生成

【特点】1.低平均路径长度l：网络中任意两个节点之间平均最短路径

1. 大聚集系数c: 网络中每个节点的平均连接

定点v有kv个邻居，最多存在kv(kv-1)/2个边

Cv表示实际存在的边的分数

c为所有节点v的Cv平均值

【网络特性】完全有序的网络：高L,高C

随机连接的网络：低L,低C

小世界网络：低L,相对高C

【应用】电影演员，电网，线虫：L大于等于随机L，C远大于随机C

【定义】维基百科：小世界网络是一种数学图形,大多数节点不是邻居,但可通过少量跳到达

百度百科：小世界网络模型是一类具有较短的平均路径长度又具有较高的聚集系数 的网络的总称

【概念】同质性连接：某节点到相距r步以内的节点的连接

弱关系连接：每个节点形成到其他k个随机均匀选择的节点连接

证明：这样的网络，任意两点之间存在短路径的概率很高

短视搜索路径太长，

弱连接太“随机”，与同质相似性的连接没有关系

【短视搜索】一种有目标的基于局部信息的搜索

特点： 每个节点有特征，任意两个节点特征不同

每个节点知道自己和邻居的节点特征

搜索过程是信息传递的过程，节点将信息传给离目标节点距离较近的邻居

【回顾】信件过程，实际是有效的，W-S模型效果不好

【Watts-Strogatz-Kleinberg模型】引入衡量远程弱连接跨越距离的“尺度”

节点在r个网格步内与其他节点相互连接

节点的k个随机边以到该节点的距离衰减的方式产生，由聚集系数q控制

【参数】r:同质性连接的丰富程序

K:远程连接的个数

q:控制远程连接的概率随距离递减的程度

【最佳状态】q=2时，WSK模型能达到最佳工作状态(q小，随即边远;q大，随即边近)

q=2解释？？？ q=1/d^2

【应用】地理上均匀分布的网络，两人成为朋友的概率与空间距离成反比

排名：w在v眼里的排名r(w) = 网络中比w离v近的节点的个数

地理上朋友数量随距离平方递减1/d^2，节点的朋友数量随排名递减1/r

【焦点】任何类型的社团，职业，邻居，共享信息，围绕某种活动的社会组织

【核心-外围结构】社会地位较高的人连接在密集连接的核心，地位较低的分散在网络外围

向核心搜索，连接结构越来越丰富，向外围搜索，连接结构越来越贫瘠

【结论】实验现象-理论解释-测量验证

实验现象-理论模型+完善+提炼-实际测量、验证与推广

【例子】微博转发，经济学期刊的引用

1. 有向图

【定义】具有方向性的图

【出度】

【入度】

【有向路径】

【强连通分量】任意两个节点之间存在两个方向上的有向路径，不被包含在更大的节点集合

【领结】链入，链出，管道，游离

万维网包含一个超大强连通分量SCC以及其他部分

A：正向集，B：反向集

SCC:A交B IN:B-A OUT:A-B 其他部分

【搜索引擎】对用户提交的查询，从海量网页集合中找到最可能满足用户需求的结果

利用链接关系蕴含的信息

【传统信息检索】基于词语之间的相关性，查询目标包含查询词

【HITS算法】计算权威值(被很多网页指向)和中枢值(指向很多网页)

【收敛】每轮结束做归一化，相继两次迭代值不变，存在“均衡”

【PageRank】节点重要性的测度

【描述】n个节点的网络，初始值为1/n。每个节点将当前值通过链接均分给所指向的节点

【基本精神】基于信息网络的结构，每个节点不断将自己的重要性分给邻居

【自私节点】不断吸收别人的价值，不向外分享

【同比缩减】每次更新后，将所有节点PageRank值乘一个比例因子

【统一补偿】每个节点的PageRank值加上(1-s)/n

【随机游走】一个人从一篇网页开始，随机选择一个链接浏览到下一页，经过k步随机游走到达X的概率，可以证明：到达X的概论等于运行PageRank算法k步得到的值