

小世界网络浅介

汪秉宏¹ 李平²

(1 中国科学技术大学 230026; 2 南京工程学院 211167)

一、小世界网络的基本定义

复杂网络由许多节点及节点之间相连接的边构成，可以刻画自然和社会中大量的复杂系统。网络中的节点代表系统的构成元素，连边则描述两个元素之间的相互作用。一个网络的规模大小和稀疏稠密性质可由网络平均路径长度及网络群聚性这两个网络结构性质的参数来度量。网络中两个节点之间的连边数称为两节点之间的路径长度，网络的平均路径长度是网络中任意两个节点之间的最短路径长度的平均值。网络的群聚性是对网络中各个节点的群聚系数的平均值（一个节点的群聚系数定义为该节点之所有邻居节点之间的实际连边数目除以所有这些邻居节点之间可能出现的最大连边数）。具有小的网络平均路径长度却有大的网络群聚系数的网络，便称之为小世界网络。例如在与朋友的聊天中，经常会发现你的某个朋友恰好也是你正在聊天的朋友的朋友；而某个你觉得与你隔得很远的人，其实与你很近，因为你正在聊天的朋友与这个“遥远”的人非常地熟悉。描写人与人之间朋友关系的社交网络往往都是小世界网络。

二、小世界网络的历史

1929年，匈牙利著名作家考林蒂（F.Karinty）在其短篇小说《链条》中给出了著名的“六度分隔”假说。小说中的一个人可以通过其认识的网球冠军、网球冠军的球友瑞典国王、瑞典国王又是诺贝尔奖的颁奖者这一路径，以简单明了的方式就与一个诺贝尔奖的获得者连接上了。考林蒂甚至将自己与远在美国的福特汽车公司的一个工人，仅仅通过三个中间人就联系了起来。考林蒂的关于人与人之间最多需要5层关系就能联系起来的推断，可以说是六度分隔假说的

最早表述，也是“小世界网络”概念的雏形。

1967年，美国哈佛大学社会心理学教授米尔格拉姆（S. Milgram）明确提出了“小世界问题”。他通过设计一个信件传递实验，以验证“六度分隔”假说，并使其成为“小世界问题”研究的基本概念。米尔格拉姆以美国堪萨斯州的威奇托市（Wichita）和内布拉斯加州的奥马哈（Omaha）作为实验起点，在两地随机地选择了296位志愿者，要求每位志愿者向指定的目标人——一位居住在马萨诸塞州首府波士顿郊区的股票经纪人传递一封信件。实验结果是其中的64位志愿者平均通过约5.3个中间人转手后，成功地将信件送达到目标人手中。这与38年前匈牙利作家考林蒂的“5个相互认识的人”的假设惊人地吻合。米尔格拉姆由此得出结论：任意两个人都可通过平均6个熟人联系起来，这是“六度分隔”假说的一次成功的实验验证。米尔格拉姆实验展示了大型社会网络的两个重要事实：大型社会网络一定包含了丰富的最短路径；人们只需要通过大型社会网络的局域信息，而不是全局信息，就可以找到这些最短路径。

1990年，米尔格拉姆的有趣实验结果激发了美国百老汇剧作家格雷（J. Guare）的创作灵感，他在电影《六度分隔》中留下了这样一句经典台词：“在这个世界上，任意两个人之间，只隔着6个人。在这星球上的任何两人之间，只有六度分离。”

1998年，“六度分隔”研究取得全新的进展。美国康奈尔大学的博士生瓦茨（D. Watts）和他的导师斯特罗加茨（S. Strogatz）在《自然》上发表了题为《小世界网络的集体动力学》一文，在“六度分隔”假说的基础上，第一次提出了“小世界网络”的概念和模型。他们把“六度分隔”现象归类为可以由两个网络结构参数——较大的网络群聚系数和较短的网络平均距离

来描述的“小世界现象”。他们发现，在各种社会网络、电力网络、神经网络、生物网络中均存在小世界现象。

三、小世界网络模型

早期的网络研究中，人们只讨论了规则网络和随机网络模型。所谓规则网络是一种具有规则连接结构的网络，例如网络中各个节点的连接形式互为相同。规则网络的特点是具有较大的平均路径长度和较小的群聚系数。而随机网络是与规则网络相对应的一种网络，其最经典的模型是由匈牙利数学家厄多斯 (P. Erdos) 和伦伊 (A. Renyi) 于 20 世纪 50 年代提出的，在节点数和节点的连边数确定的情况下，在每对节点之间随机地安排或不安排连边，由此形成的网络简称为 ER 随机网络。随机网络的特点是具有较小的平均路径长度和较大的群聚系数。

自“小世界”问题在社会学研究中被提出后，一直没有相应的数学模型。直到 20 世纪末，一批数学、物理学和计算机专家才创造性地为这个社会概念构建了数学模型，并引用网络平均路径长度和网络群聚系数这两个结构参数来刻画网络系统的小世界性质。小世界网络模型的典型代表是瓦茨-斯特罗加茨 (WS) 小世界网络模型与纽曼-瓦茨 (NW) 小世界网络模型。

瓦茨-斯特罗加茨 (WS) 小世界网络模型 第一个小世界网络模型是由瓦茨和斯特罗加茨在 1998 年《自然》上发表的《小世界网络的集体动力学》一文中提出的。他们发现，规则网络之群聚性较高，但网络之平均距离也大；而随机网络之平均距离较短，其群聚性也低。真实世界的网络既非完全规则，也非完全随机，而是介于这两者之间。因此，他们以一个规则网络为基础，然后以概率 p 对规则网络中的每条边进行随机重连（剔除节点的自我连接和重边连接），这些随机重连的边在规则网络中既引入了随机性，又引入了长程连接（或称为捷径）。这些少量的、随机的长程连接大大地减小了网络节点之间的平均距离，却又保持了原有网络较高的群聚性，使网络具有了小世界特性，其演化过程如图 1 所示。

纽曼-瓦茨 (NW) 小世界网络模型 纽曼 (M. Newman) 和瓦茨对 WS 小世界网络模型作了改进，

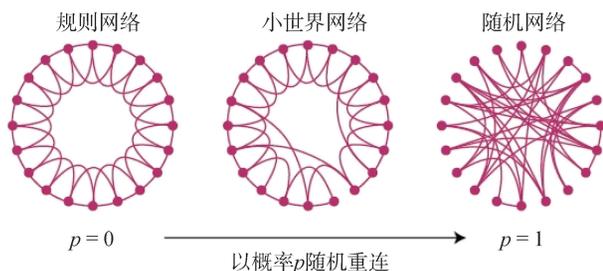


图 1 WS 小世界网络模型

在不改变原规则网络中各节点之间的连边情况下，任选二个节点，由概率 p 决定是否在这 2 个节点之间增加一条边，如图 2 中在 a、b 两节点之间就增加了一条随机连边。这样的操作同样在规则网络中既引入了随机性，又引入了长程连接，从而使网络具有了小世界特性，其演化过程如图 2 所示。

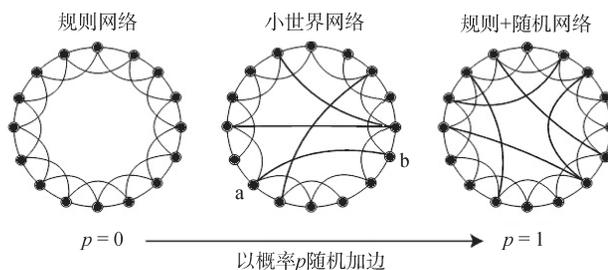


图 2 NW 小世界网络模型

四、小世界网络的特征参数 L 、 C 、 Q

人们引入了平均路径长度 L 和平均群聚系数 C 来定量分析网络的小世界特性。

网络的平均路径长度 L ：如果 L_{ij} 表示网络中第 i 节点与第 j 节点之间的最短路径长度。则 L_{ij} 对于所有节点对 (i, j) 的平均值 L 称之为网络的平均路径长度（也称为特征路径长度或平均最短路径长度）。理论分析、数值模拟和大量的实证研究表明，小世界网络的平均路径长度是网络规模（网络的节点总数） N 的对数增长函数：

$$L \propto \log N$$

网络的群聚系数 C ：对于一般的无向网络，网络中第 i 节点的群聚系数 $C(i)$ 定义为

$$C(i) = \frac{2e(i)}{k(i)(k(i)-1)},$$

其中 $k(i)$ 是第 i 节点之邻居节点数目, 亦称度。 $e(i)$ 则是 $k(i)$ 个邻居节点之间的实际连接边数。 k 个邻居节点之间的最大连接边数是

$$\binom{k}{2} = \frac{k(k-1)}{2}。$$

网络的平均群聚系数 C 定义为 $C(i)$ 对于网络中所有节点 i 之平均值。群聚系数是用来描述网络中的节点之间群聚成团程度的参数, 表征了一个节点的邻居接点之间相互连接的程度。显然, 群聚系数是一个介于 0 与 1 之间的数, 越接近 1, 表示这个节点的所有邻居节点间越有“抱团”的趋势, 它描述了网络节点间局域连接的密集程度。

在复杂网络科学中, 将既有小的平均路径长度又有大的平均群聚系数特征的网络称为小世界网络。在 WS 小世界网络模型中, 当调节节点间连边的随机重连概率 p 从 0 逐渐增大到 1 的过程中, 可以观察到初始的规则网络将历经“规则网络—小世界网络—随机网络”三个阶段。图 3 给出了 WS 小世界网络模型中归一化的平均路径长度 $L(p)/L(0)$ 和平均群聚系数 $C(p)/C(0)$ 随重连概率 p 增长的变化关系。

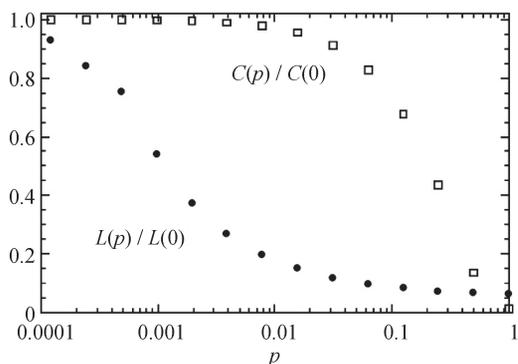


图 3 WS 小世界模型中归一化平均路径长度和平均群聚系数随重连概率 p 的变化关系

小世界商 Q : 与复杂网络科学描述小世界网络的方式不同, 社会学中描述网络是否呈现小世界特性的特征参数是所谓的小世界商 Q , 它是将实际网络的平均群聚系数 C 与平均路径长度 L 的比值 C/L , 去除以对应的具有相同节点数和边数的随机网络的平均群聚系数 C_0 与平均路径长度 L_0 的比值 C_0/L_0 , 这样的

一个商值

$$Q = (C/L) \div (C_0/L_0)$$

大于 1, 则表明该网络具有小世界特性。高值 Q 越大, 网络的小世界特性越显著。

五、小世界网络的实例

互联网的节点是各个路由器, 连边则是连接各个路由器的光纤。在 1995~1999 年对于互联网网站及路由器层次都进行了计算, 发现互联网的平均路径长度是 $L=4.0$, 这与 ER 随机图的 $L=10$ 相比明显为小。互联网的群聚系数则是 $C=0.18\sim 0.3$, 这与 ER 随机图的 $C=0.001$ 相比明显偏大, 是一个小世界网络。

科学引文网络的节点是发表于科学刊物上的各篇论文, 连边则是互相之间的引用, 科学引文网络也是典型的小世界网络。可以从科学论文数据库构建各种科学合作网络。科学合作网络的节点是论文的各个作者, 连边则是两个不同作者在同一篇文章中作为共同作者出现。众所周知的常见的科学论文数据库有洛斯阿拉莫斯电子预印文库 (Los Alamos e-Print Archives preprints (1992~)), 生物医学研究论文库 (Medline: biomedical research articles (1961~)), 斯坦福公共信息检索系统: 高能物理论文库 (Stanford Public Information Retrieval System(SPIRES): high-energy physics articles(1974~)), 网络计算机科学技术文献图书馆 (Network Computer Science Technical Reference Library (NCSTRL): computer science articles (10 years records))。分别对 1 万到 2 百万个节点 (论文), 在相应年限里进行计算, 均符合小世界网络特征。

图 4 所示的语言网络也是小世界网络。每一个单词是一个节点, 两个单词相连接出现在一个句子中即为有连边。人脑可以记忆的单词大约有 $10^4\sim 10^5$ 个。据计算, 两个单词之间的平均距离是 $d=2\sim 3$ (Romaine, 1992), 即两个单词平均只要通过 2~3 个中间单词的相连即可出现在一个句子中。因此, 语言网络是典型的小世界网络。

六、小世界网络上的动力学问题

网络的结构对于以这个网络为基础架构的系统的

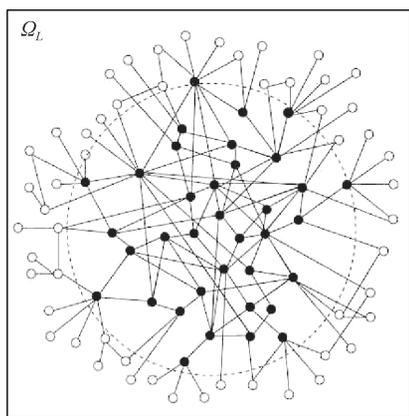


图4 语言网络

集体动力学行为具有重要的影响与关联。小世界网络上的动力学主要讨论其上的导航（搜索）、同步、传播及博弈等问题。

1. 小世界网络上的导航（搜索）问题

网络导航又称为网络搜索，如在研究某一概念（方法、模型）最初究竟是由谁第一个提出的问题，人们可以借助科学引文网络通过对相关论文的搜索来获得相应的答案。小世界网络特有的拓扑结构能不能进行网络导航（搜索）？如果能够进行，那相对随机（或规则）网络究竟是有利于还是不利于导航？这些问题对小世界网络上的动力学都有重要意义。美国康奈尔大学的乔恩·克莱因伯格（J. Kleinberg）研究了均匀网络上有向化的NW小世界模型上的导航问题，如果随机加边的概率正比于两个节点“距离”倒数的 α 幂次，这里的“距离”既可以是纯粹的地理上的距离，也可以是一种隐喻意义上的社会距离，如职业、种族、收入、教育、信仰等方面的差别或不同。当 α 恰好等于网络的维数时（如对一维网络， $\alpha=1$ ，对二维网络， $\alpha=2$ ），应用网络的局域信息，并通过贪婪算法（即每次导航到的下一个节点都应离目标节点更近，以保证最终可以达到目标），可以最有效地实施导航（或搜索）。幂次 α 的特定取值说明网络只有特定的结构才使寻找最短路径成为可能，这也正是米尔格拉姆实验只有在小世界网络中才能成功的原因。显然，在小世界网络上可以有效地实施导航（或搜索）是有条件的，即网络中的节点在“距离”上是均匀分布的，且网络中两节点之间随机添加长程连接（捷径）的

概率是基于两节点间的“距离”。而在实际网络中，节点的“距离”分布往往是不均匀的，且两节点间是否实现长程连接（添加捷径）既有“距离”的因素，也有非“距离”的因素在起作用，如节点在动力学行为上的惯性与情感等。

2. 小世界网络上的同步问题

在自然科学和社会科学中都广泛存在着同步问题。如一场精彩表演结束后，观众给予的热烈掌声，往往都是从最初零乱的鼓掌到最后趋于一致的、有节奏性的鼓掌。把同步问题放在复杂网络科学的视域下进行研究，就是用网络中的诸多节点来代表不同的动力系统（如不同的观众），网络节点间的连边代表不同节点（不同的观众）之间的相互作用，在不同的初始状态下，由于彼此之间的相互作用，使众多不同的动力系统（如不同观众）的状态相互接近，并最终达到（或基本达到）全同状态的过程即所谓的网络同步。显然，在复杂网络的框架下研究同步问题，节点所代表的动力系统既可以是全同的、也可以是不同的；而不同系统间的相互作用既可以是全局的作用、也可以是局域的作用。特别的是人们可以深入的研究网络拓扑结构对复杂动力系统的同步过程与行为的重要影响。

中国学者汪小帆、陈关荣最早研究了具有小世界结构性质的复杂网络系统的同步问题，发现小世界特性可以显著地提高复杂网络系统的同步能力，即小世界网络实现同步的能力远远高于其对应的规则网络。斯特罗加茨认为同步实质上是小世界效应的一种表征，即小世界网络中节点间具有较短的平均距离，使得处于网络不同部分的节点能够共享网络的共同属性而形成同步。小世界网络的基本结构是具有较大的局域群聚特性，同时又具有较短的全局性平均距离。其局域群聚特性使得网络耦合得更加紧密，促进了局域范围中节点间的相互关联程度，使相应网络具有了较强的同步能力；较短的全局性平均距离，使规模越大的网络其同步能力越强。随着重连概率 p 的增长，网络系统会发生从非同步相到同步相的相变。必须注意的是，“节点间的平均距离越小，同步能力越强”这一结论只适用于小世界网络，而对其他类型的复杂网

络不适用。如对于无标度网络，这一结论恰好相反。这再一次地印证了结构决定动力学这一结论。

3. 小世界网络上的流行病传播问题

传染病在人群中的流行、病毒在计算机网络上的蔓延、谣言在人类社会中的扩散等，都可视为流行病在网络上的传播问题。

美国学者穆尔 (C. Moore) 和纽曼最早对小世界网络上的传播行为进行了较系统的研究。他们研究了 NW 小世界网络上的 SIR 模型 (易感-感染-恢复模型) 的传播行为。他们把小世界网络上的疾病传播问题等价于逾渗问题，如座愈渗问题——删除一定比例的网络中的节点、键愈渗问题——删除一定比例的网络中的连边、或座键混合愈渗问题——删除一定比例的网络中的节点与连边，以此来观察删除网络中一定比例的节点与连边对阻碍疾病在网络中传播的效果。穆尔和纽曼给出了愈渗的临界值与网络长程边 (捷径) 密度的函数关系，发现随着长程边 (捷径) 的出现，流行病在网络上的传播阈值 (即愈渗的临界值) 迅速下降，即使少量的长程边 (捷径) 就可以显著地增加流行病在网络中的传播能力。

阿根廷学者库珀曼 (M. Kuperman) 和艾布拉姆森 (G. Abramson) 研究了 WS 小世界网络上的 SIRS 模型 (易感-感染-恢复-再易感模型) 的传播行为。发现当重连概率 p 很小的时候，疾病可以在网络中长期存在，但患病比率很小且波动不大，可以近似地看作收敛到一个不动点；而当重连概率 p 很大的时候，患病人数会出现周期性的波动。即随着重连概率 p 的逐渐增加，对应于一定的人群结构，网络中被传染的人数从不规则的、小幅度的增加发展到自发的、大范围的振荡状态，其中在 $p=0.1$ 附近传染人数明显增加，显示出小世界效应。

流行病在小世界网络上的传播行为与小世界网络的规模 N 、重连概率 p (或加入长程边的概率) 和 k (网络上单方向连接的边数) 等网络结构参数有关，这些网络结构参数对流行病在小世界网络上的传播规律和趋势有很大影响。改变这些网络结构参数的数值，可以得到流行病在不同情况下的传播情况，显示流行病的发展过程，预测流行病的传播趋势，分析流行的原因和关键因素，寻找对其进行控制和预测的最佳策略。

4. 小世界网络上的博弈问题

具有争斗或竞争性的问题往往可以通过博弈模型来进行描述与分析，这类问题又统称为博弈问题，如生物体进化过程中的优胜劣汰问题，社会关系中的合作与背叛问题等。复杂网络上的博弈问题是指群体博弈者之间的关系构成了一个复杂网络，基于对复杂网络拓扑结构的全新认识，来研究群体博弈者之间的合作与竞争问题。特别关注网络的拓扑结构如何影响合作行为的涌现与演化特性。社会网络都具有明显的小世界效应，因此，在小世界网络上探讨群体博弈行为无论在理论研究和实际应用等方面都很有意义。人们在小世界网络上的博弈研究中发现，博弈群体的演化行为与小世界网络的拓扑结构参数密切相关，其较短的平均路径长度和较高的平均群聚系数特性，使得信息可以在网络中快速传递并在局域群体中较快达成共识，有利于合作行为的涌现，其典型特征就是重连概率与合作的涌现程度有着明显的关联性。

阿根廷学者艾布拉姆森 (G. Abramson) 和库珀曼 (M. Kuperman) 率先研究了小世界网络上的博弈问题，探讨了从规则网络到小世界网络的演化过程中合作行为的变化特点，发现重连概率与合作行为的涌现程度有明显的关联性，网络平均度和重连概率在某些范围内能够促进合作，而在另一些情况下则会抑制合作。人们还基于 NW 小世界网络模型研究了同质性与度异质性的影响，表明适当的异质性能够有效促进网络中合作行为的涌现。

* 本文为《中国大百科全书》小世界网络词条而写。

