文章编号：1003-0077（2011）00-0000-00

语言网络研究的数学模型*

——从复杂网络、社会网络到语言网络

赵怿怡[1]，刘海涛[2]

（1.厦门大学，福建省 厦门市 361005；2. 浙江大学，浙江省 杭州市 310058）

摘要：复杂网络技术的发展为大数据时代的语言研究提供了新的视角。网络方法应用到语言研究的重要目的是探索语言网络的结构特征规律和功能演化规律。本文综述了以图论为基础的复杂网络发展及社会网络、语言网络的主要数学模型，试图从复杂网络共性特征——小世界、无标度特征中进一步剥离出语言网络的个性特征，为语言符号多层级网络结构、功能研究提供参考。

关键词：语言网络 网络技术 网络演化 图论 复杂网络特征

中图分类号：TP391 **文献标识码：**A

Mathematical Modeling in Language Networks Research

-- From complex networks to social networks and language networks

Zhao Yiyi[1], Liu Haitao[2]

(1. Xiamen University, Xiamen, Fujian Province 361005, China;

2. Zhejiang University, Hangzhou, Zhejiang Province 310058, China)

Abstract: Networks technology provides a new perspective for linguistics in the age of big data. Network method applied in language networks is to explore the structure of the law and the evolution of language network functions. This article reviews the development of complex network based on Graph Theory and the primary mathematical modeling of social networks, language networks, aiming to strip personality traits of language networks out from the characteristics of complex networks, and giving more references for multi-level language networks studies.

Key words: Language Networks; Network technology; Network evolution; Complex network characteristics; Graph Theory

1 引言

复杂网络技术的发展为语言研究提供了新的视角和手段。“把语言视为网络”具备语言学、认知科学、心理学的理论依据[1]。目前可见语言网络的研究涉及语言符号的字单元、词单元[2]、句法[3-6]、语义[7][8]等多层级符号系统，网络构建与研究的目的除了探索各层级符号对应语言网络之间的差异，还包括探索各类语言网络构建的理据性与网络结构共性[9][10]，但鲜有关于复杂网络、社会网络、语言网络重要规律的综述。语言网络研究科学化的主要目的是发现事物的发展规律，以模型的形式重复验证与预测事物的发展[11]，以此为目标本文综述了迄今从复杂网络到社会网络、语言网络领域的主要数学模型，尝试为语言网络提供普适价值提供参考。

2 网络初步：图论

进入到语言网络研究的操作阶段，图论是打开语言复杂网络研究之门的第一把钥匙。[12-14]网络是节点的集合，所以定义$[X]^k$表示元素为 k 的集合 X。一个简单的无向图 G 表示为 $G=(V, E)$；V 表示图 G 节点集合，E 表示边的集合，$E\subseteq|[V]^2|$；定义 $G = (X, Y)$为图 G，则有 $V(G) = X$，$E(G) = Y$；若有边 $e_2=\{v, w\}\in E$，则表示边 e_2 以节点 v, w 为顶点，同时 v, w

* **收稿日期：** **定稿日期：**

基金项目：国家社会科学基金重大项目——现代汉语计量语言学研究（NO.11&ZD188）；国家社会科学基金青年项目——基于同一文本的句法网络语义网络关系研究（NO.14CYY046）;本成果得到厦门大学哲社科繁荣计划、两岸关系和平发展中心资助。

互为相邻节点（*adjacent neighbors*），如果两条边 e_1, e_2 有共享公共节点，也可以说两条边互为相邻边。*E(v)*是以 *v* 为顶点的边的集合。*N(v)*是节点 *v* 的邻节点集合。 以上是图 2-1 所示无向图 *G* 的组成元素的基本定义。

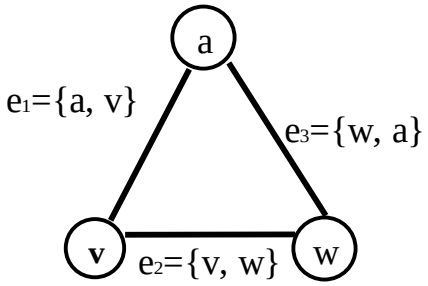

图 2-1 无向图示例，*G = (V,E)，V={a, v,w}, E={ e1, e2, e3}, G=(X, Y), X=3, Y=3, E(a)={ e1, e3}, N (a)={ u, w }, d(G)=2*

在一个拥有更多节点的网络 *G* 中，节点 v_i 的节点度表示为 $d(v_i) = k_i$，k_i 也反映图中节点的连通性，反映节点邻里规模。节点 v_i 节点度 k_i 也等于节点 v_i 的边集合$|E(v_i)|$，$|E(v_i)|$表示所有以 v_i 为顶点的边数。很明显，在图 2-1 所示无向图 *G* 中，$|E(w)| = |N(w)|$，$|E(v)| = |N(v)|$ ，$|E(a)| = |N(a)|$。这表明 *E* 是不包含多重边的集合。在包含多重边的网络中，多重边可以通过赋予边值来表示，包含多重边的图通常被称为加权图或加权网络。

对于整个网络 *G* 来说，平均节点度（*average node degree*）可以表示为：

$$d(G) = \frac{1}{|V|}\sum_{vi \in V} d_G(vi) = 2\,|E(G)|/|V(G)| \qquad (1)$$

平均节点度反映网络中节点的平均连通性。衡量此问题的标准化参数是网络密度(density)D（0<D<1）。密度为0的网络是一个无边相连、节点孤立的网络，相反，一个节点完全连通的网络密度为1。孤立节点数提供了一个考察网络密度分布的视角。另一个反映密度的相关参数是网络中心度(centralization)。一个星形状拓扑的网络中心度接近1，分散的网络中心度接近0。

路径长度（*path length*）是形成节点间路径的边数。网络中指定两个节点可能有多条路径相连。如图2-1示例，图*G*有节点*v*, *w*，它们可以通过两条路径$L(v, w)=|e_2|=1$，$L(v, w)=|e_{1+}e_3|=2$相连。其中，$L(, w)=|e_2|$为两个节点间的距离，是两个节点最短的路径长度（*shortest path length*），节点*v*和*w*($v \neq w$)的距离表示为（*distance*）$\delta(v, w)=1$。

用*P*表示无向图中所有节点间距离的集合，无向图直径*D(G)*是任意两个节点间最大的最短路径长度，即*P*中最大δ。平均最短路径通常被称为网络的平均路径长度。所有节点间路径长度的均值为网络的平均路径长度（*average path length*）表示为*L*(*G*)：

$$L(G) = \frac{1}{|\mathrm{V(G)}|/2}\sum_{\{v,w\}\in[V]^2}\delta(\mathrm{v,w}) \qquad (2)$$

以语言网络为例来说明，如图2-2所示，G_1是一个由词为节点根据句子“ROOT人体,是,由,数以亿计,的,微小,而,有,生命,的,细胞,构成,的 ROOT,这,些,细胞,构成,各,个,不同,的,组织,器官,保证,了,人体,的,正常,工作”中词的前后邻接的同现关系[1]构成的无向图，节点集合*V*={这，些，各，个，……，细胞，构成}，边集合$E=\{e_{这些}, e_{些各}, e_{各个}, \ldots\ldots e_{细胞}, e_{构成}\}$，$G_1$的值表示为$|G_1|=|V_1|=23$（节点数）,$|E_1|=29$（边数）。$N_{G1}$(些) = {这，细胞}，$d_{G1}$(些) = $|E$(些)| = $|N_{G1}$ (些)| = 2。δ (这，些)=1。直径$D(G_1)$=5。

[1] 同现网络是根据词的上下文同现关系构造的网络，是语言工程领域最常见的构造语言网络方法。

有向图 G_2 相比无向图 G_1 最为明显的变化就是节点度分化为出度、入度。例如，d_{G2}(的) = |E (的)| = |N_{G2} (的)| = k_{in}(的)+ k_{out}(的)=9，k_{in}(的)=4，k_{out} (的)=5。

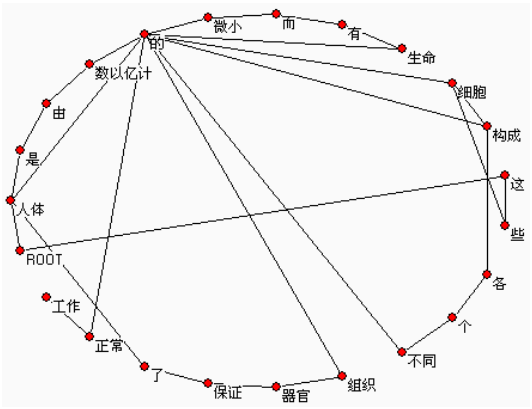

图 2-2（同现网）无向图例 *G1*

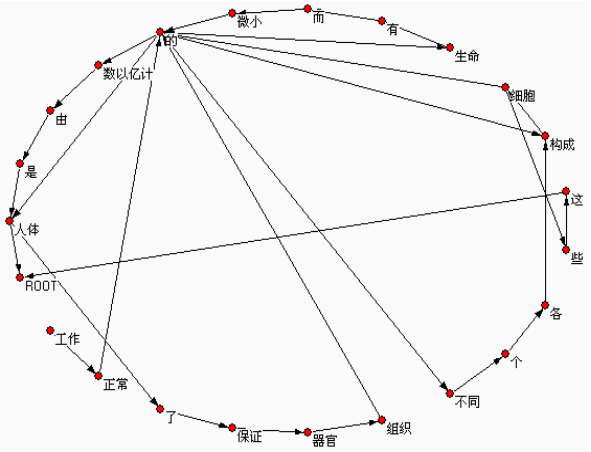

图 2-3（同现网）有向图例 *G2*

表 2-1 G*1* 和 *G2* 的基本参数

	|N|	*|E|*	*L(G)*	*D(G)*	*k(G)*	*density*
G1	23	29	2.85771	5	2.5217	0.1096
G2	23	29	4.74877	12	1.2608	0.0548

通过 PAJEK[2]获得示例网络 G_1 和 G_2 的基本参数，发现相同节点和边构成的无向网络和有向网络在平均路径长度、直径、网络密度、节点度方面存在差异。以同现网络为例的概念解释和参数比较说明，用语言材料不同颗粒的单位构建网络是可行的且有差别的[15]。

网络科学是一门以物理学（图论）为基础的分支学科，但其发展受到社会学的重要影响。网络科学的重要组成复杂网络和复杂科学的发展就是源于社会网络的研究。社会网络注重计量角度考察网络特征[16]。除了观察网络基本参数节点度、平均路径长度、聚集系数之外，网络密度、网络中心度和网络凝聚度（*cohesion*）[17]也是社会网络研究的焦点，社会网络借此来考察更为局部的网络的组成（*components*）和网络中的次集团（*cliques*）现象。

3 小世界网络

欧拉(Euler)开创图论学科促成了网络科学的兴起，网络科学接下来的重要发展始于20世纪中叶由Solomonoff和Rapopor[18]以及Erdös和Rényi[19]引入的随机网络。网络是由节点和连线组成的图，成分简单，但是却能呈现高度的复杂性。20世纪以来的大量研究证明了随机网络并不能描述真实网络的拓扑特性。而自然界的从技术到生物乃至人类社会中的各种开放系统都表现为更为复杂的网络形式。20世纪末统计物理学出现的小世界网络[20]和无标度网络[21]开启了网络科学中对于复杂网络特性的探索。

小世界特征（*Small World*）是区分随机网络和大规模复杂网络的可测特征。该研究开始于Milgram[22]，Milgram最初关注人们和他们熟人间的社会网络关系，即在特定人群中，两个人如何实现联系的最短连接。Wasserman和Faust[23]研究让一个人通过熟人传递的方式把一封信寄给目标人，在这样的社会网络中，网络结合度（和网络信息流动高效性、脆弱性有关的参数）显现出来。在Milgram的模型(*short-cut property*)中，网络中两个随机节点间的最短路径可以被视作小世界网络的指标。但这个单一指标并不能成为社会网络区别于随机网络（也有最短路径特征）的特征。作为补充，Watts和Strogatz [20]提出了两个特征奠定了小世界网络的基础：相比于随机网络，小世界网络有更高的聚集系数；相比于随机网络，小世界网络有近似的最短路径。

为了解释这个问题，Watts和Strogatz[20]引用了两个指标：聚类(*clustering*)和密度。在无

[2] 社会网络分析工具。

向网络中，聚类是节点$v_i \in V$ (G)的聚集度（*cluster value*）$C_{vi}(G)$的均值。更为准确地说，节点的聚类等于节点v_i 的实际边数*adj(vi)*与相应完全图|*NG*(*vi*)|中节点v_i 边数的平均比值：

$$C_{vi}(G) = \frac{\text{adj(vi)}}{\text{dG(v)/2}} = \frac{\text{adj(vi)}}{\text{d(vi)(d(vi)-1)/2}} \in [0,1] \qquad (3)$$

那么整个图 *G* 的聚集度 $C_{WS}(G)$可以定义为：

$$C_{\text{WS}}(G) = \frac{1}{n}\sum_{i=1}^{n} \text{Cvi(G)} \in [0,1] \qquad (4)$$

所以C_{WS}描述一个网络中节点相互连接的程度。聚类和社会网络的传递性相关。C_{WS}的缺陷是不能很好地在多重边的图中操作。原因是：两个节点间如果有多条边，边数只能计算一次。因此，Bollobás和Riordan [23]提出了聚集系数*(cluster coefficient)*$C_{BR}(G)$，用来表示网络中三角关系数量*(number of triangles)*和相邻边数量*(number of pairs of adjacent edges)*的比值：

$$C_{BR}(G) = \frac{3 \times \text{number of triangles of G}}{\text{number of pairs of adjacent edges of G}} \in [0,1] \qquad (5)$$

高聚集系数C_{BR} (*G*)和聚集度C_{WS} (*G*)一样，表示图*G*的联结是可传递的，如果某节点*u*∈ *V(G)*和节点*a*, *w*∈ *V(G)*相连接，那么节点*a*, *w*也可能是相连的。在一个好友网络中，高聚集系数意味着一个人甲的朋友乙的朋友丙也可能是甲的朋友。很明显，聚集系数的概念并没有与聚集度分析混淆，但在某种程度上是相似测量。

Watts和Strogatz[19]考察的核心是规则网络具有高聚集度，随机网络有低聚集度，聚集度的分布相反于平均路径长度，聚集系数越大网络的平均路径长度越小。Bollobás和Riordan[24]指出尽管平均路径长度小于或等于网络直径，但是平均路径长度比起直径并非小很多。因此平均路径长度可以作为小世界的测度之一。

从*L*(*G*)和C_{WS}(*G*)的角度，Watts和Strogatz [20]细化了小世界的概念（此后小世界被称作*WS model*）。小世界网络表现出类似规则网络较高的聚集度和类似随机网络较小平均路径长度：

$$CX(Gregular) \sim CX(Gsw) \gg CX(Grandom)$$
$$L(Gregular) \gg L(Gsw) \sim L(Grandom) \qquad X \in \{WS, BR\} \qquad (6)$$

L(*G*)表示网络的平均路径长度，*L*(*G*)显示了“全局网络特征”，它聚合了网络所有成对节点的相关性。相比之下，C_{WS} (*G*)表示“局部网络特性”。按照这个标准，语言网络呈现出小世界的特性。研究表明，汉语词同现网络与英语词同现网络一样，平均最短路径远小于网络规模而聚集系数非常高，具有明显的小世界效应[4]。汉语句法、语义网络和*ER*随机网络[3]的平均路径长度和直径大致相当，但句法网络的聚集系数要远远大于ER随机网络，汉语句法、语义网具有小世界特征[8][25]。

但小世界的模型也存在缺陷，在于：*L*(*G*)和C_{WS} (*G*)关注相应输入值的某一时刻（静态）的分布，丢失了对这个分布更为细节的描述。而无标度模型可以用来弥补这个缺陷。

4 无标度网络

小世界模型描述网络静态特征，而小世界网络动态增长特征被描述为择优模型（*preferential attachment model*）[21]，后来也被称为BA模型（*Barabási-Albert model, BA model*）。Barabási和Albert观察到：复杂网络节点连通根据无标度规律分布，这些网络中的节点连接到网路中其他节点的最短路径和局部聚类具有共同特性。更确切地说，Barabási和Albert确认了许多社会网络中节点连接方式区别于随机网络中节点的连接方式，即每个节点的连接数

[3] Erdös 和 Rényi(1961)引入的随机网络模型。

符合幂律分布。节点连接的概率$P(k)$（随机选定节点与其他k个节点相互作用的概率）近似于：

$$P(k) \sim k^{-r}, \quad \gamma [1.5, 3.5] \text{ [26] [27]} \qquad (7)$$

如果一个无向图节点度分布服从幂律分布，则表示这个网络的连通性是无标度的。很多社会现象（*social-semiotic phenomena*）[28]服从*Zipf*定律[29]，比如语言单位的频率分布，它们也因此被称作无标度网络（*scale-free networks*）[21]。无标度意味着没有代表其他节点的典型节点[30] [31]。

节点度的幂律分布可以反映节点度的“等级”分布（等级由节点连通性降序决定），也可以反映节点度的“大小”分布（从2度的节点到网络中最高度的节点的数量排序），还可以描述有向图节点出入度的分布。度分布的幂律说明大部分节点是不连接的，仅有少数具有高连接性的中枢节点（*hubs*）[32]。这些中心节点主要任务是提供结合能力，它们把多数节点整合到网络中[33]。因此，对于固定数量的连接来说，幂指数越小，曲线的斜面越窄，存在高连接的中枢节点的概率越高。相比之下，如果一定度的节点数量随着度增长呈指数衰退，高连接节点可能会逐渐靠不住或消失。

为了构建一个能够解释幂律涌现（*emergence*）的模型，Barabási和Albert不再考察节点数，而是统一考察概率。Barabási和Albert的基本思想是：无标度分布的结果是网络增长择优行为导致的。动态网络中的节点集合通过连接到高连接的节点实现增长。这种“择优”行为也被称作马太效应（*Matthew effect*）[34]，它表示已有节点通过连接新节点实现“富有”[32]。文献引用就是一个“富有”的例子，新文献往往趋向于连接高频引用的文献。用公式表示为：假设有概率$P(k_v)$，它表示新节点将要连接到连通性为k_v的节点v上的概率，则存在k_v的函数如下所示，w表示已经连接到网络的节点。

$$P(k_v) = \frac{k_v}{\sum_w k_w} \qquad (8)$$

在一些试验中，Barabási 和 Albert[21]展示了一些根据此模型演变的网络发展为“标度不变”，其中的节点度分布符合幂律分布（幂指数通常为 2.9 ± 0.1）。需要注意的是，按照无标度模型产生的网络并不一定是小世界模型。

尽管无标度模型克服了小世界模型静态表述，但无标度模型也忽略了网络动态增长的其他因素。比如，网络可能通过节点新增和消亡的一定比例实现增长，或者是有些高度节点不一定直接连接到网络新节点。但是无论如何，无标度网络模型促进了对于网络及特征进一步的研究，它从纯粹随机网络中更精确的分离了复杂网络特征。在语言网络研究中，刘海涛[5][6][8][24]对汉语依存句法网络、语义网络的无标度特性进行了测定，结果显示它们的节点度分布均服从幂律分布，幂指数在 2. 18-2.439 之间。汉语句法、语义、同现词网络均符合复杂网络小世界和无标度特征。小世界、无标度模型的出现成为复杂网络研究的里程碑。但是我们不难想象，以语言网络为代表的各类复杂网络仍可能包含了更多具有特殊性的拓扑结构特点和演化规律值得更深入的研究。而网络的相关性匹配和社团结构特征的发现可以称为社会网络特殊性研究的最好例证。

5 相关性匹配

在演化网络中，由于网络下一步的演化依赖于当前每一个节点的度，因此新、旧节点度之间存在相关性。Newman [35][36]提出一个模型，其基本假设就是：两个节点连接的概率依赖于这两个节点的连通性*(connectivity)*，连通性即节点度。这个模型用来统计社会网络中节点倾向于和有相似特征的节点发生连接的程度，这种网络演化的趋势叫做节点的正相关性匹配（*assortative mixing*）。根据Newman和Park[37]的研究，这个标准可以区分都同属于小世界

模型中的社会网络（如人工网络）和非社会网络(如生物、技术网络)。社会事件节点相互连接多为正相关连接；技术网络（比如因特网）节点相互连接多为负相关匹配（*disassortative mixing*）。Newman始创相关系数（*correlation coefficient*）来测量无向网的节点的连接情况，如公式(9)：

$$r(G)=\frac{\frac{1}{m}\sum_i j_i k_i-[\frac{1}{m}\sum_i \frac{1}{2}(j_i+k_i)]^2}{\frac{1}{m}\sum_i \frac{1}{2}(j_i^2+k_i^2)-[\frac{1}{m}\sum_i \frac{1}{2}(j_i+k_i)]^2}\in[-1,1] \qquad (9)$$

*i*表示以节点*j*、*k*为顶点的边，j_i 和k_i 表示节点*j*、*k*的节点度，$m=|E|$，$G=(V, E)$。正相关连接发生条件是$r(G)>>0$, 相反$r(G)<<0$的情况为负相关匹配。刘海涛利用相关系数对语义网和句法网节点连接情况进行测量，结果表明汉语的句法、语义网和大多数生物网络、技术网络一样均为负相关的网络。但其更有益的发现在于：语义网相关系数显示出弱于句法网的特点。据此，刘海涛认为句法网络中虚词的存在和句法连接增强了语言网络的相关性，而语义网因为缺少虚词导致其相关性差是可以被合理解释的[8]。

尽管相关系数从复杂网络中区分了社会网络，但它仍不能解释复杂网络节点相关匹配的涌现（*emergence of mixing*）。因为所有系数仅仅停留于图指数的表示，复杂网络更高层的结构次序被忽视。为了弥补对网络结构层次忽视，Newman和Park[37]又提出一个观点，即社团结构（*community structure*）。

6 社团结构和模体

社团结构源于社会网络中成员相互影响的概率依赖于社团（比如家庭，联盟等）和前后关系（*contexts*）。这个关系通常是分享性的[32]。共享社团或前后成员关系建立了相互影响的概率。这意味着，一个行动者（*agents*）进入一个社会网络并不一定具有与网络高连接成员接触的互动机会，这和无标度网络的节点增加方式刚好相反。所以社团构建模型并不适合来考量网络连接上的无标度层级限制。但是 Newman [35]利用社团结构模型来研究从属网络（*affiliation networks*）。从属网络的最佳实例是科学家合作网，其中同一个社团或前后关系被定义为合作者。从属网络是双向图建模，节点行动者（*actor*）是连接到社团中的行动元。双向模型转换为不可分图（*unipartite graph*），图中节点表示至少被连接到一个社团的行动者。不可分图被输入来计算聚集度和平均路径。Newman[35]的讨论核心是相比于随机图（*Erdös-Rényi model*），这种从属网络中聚类是更高级的，原因是社团成员数量越多，网络中会存在更多的三角关系。相互作用的节点 *a*、*w* 也和同社团节点 *v* 相连。社团结构的另一个发现是节点的正相关连接可以出现在具有社团结构的网络中也会出现在没有社团结构的社会网络中。因此网络的社团结构可以代替节点连接相关系数成为更精确的判断网络类型的标准。

与社团结构相似的另一个反映网络局部形式的概念是再生子网络（*recurrent sub-networks*）[27][38]。再生子网络的研究发现图 *G* 的子图 *G'*相比于相同边数和节点的随机网络，能够表现出超预期的特征。这类子图被称为模体（*Motif*）。不同网络的模体反映网络的局部连接模式，复杂网络模体表示的子网络数量明显高于随机形成的网络。特定的几个模体聚集在一起可以形成大的模体簇，这有助于理解网络的增长机制[39]。模体可以很好地区分生物网络、技术网络和信息网络。Ravasz et al.[33]展示了一个包含模体结构的无标度分布模型，发现该类模型有内在的等级结构，节点围绕高聚集度的节点构建网络，而越来越多的节点逐渐减小聚集度形成外围的连接。所以此类模型表现出明显的网络层级性（*hierarchical networks*），可用于区分无标度网络中的层级网络和非层级网络。Ravasz et al.观察到该层级网络模型的节点度 *k* 和聚集系数 *C* 的函数 *C(k)*随着节点度 *k* 幂律衰减，表示如下：

$$C(k) \sim k^{-\theta} \qquad (10)$$

这一模型把复杂网络模体测量简化为节点度与聚集系数的幂律测定。符合该模型的层级网络更具中心模块性，作为应用于语言网络层级性测定的一个模块化的模型，Ferrer i Cancho et al.[3]测得句法网络的$\theta \approx 1$，因为句法网络来源于层级结构的句法树所以也具有明显的层级性。刘海涛[8]对汉语语义网络的 *C(k)*测定显示其不服从幂律分布。由此可见，就汉语句法网与语义网的比较，汉语虚词在语言网络的节点负相关匹配和网络层级性中都扮演重要的角色。

7 以时间为变量的网络演化

目前除了BA模型关注网络增长外，几乎所有的网络特征都集中反映一定时间点上的网络静态图。BA模型源于“假设有一个节点集合，它随时间演化，表现为不断有一定节点度的一定数量的节点连接到该集合中”的推导[24]。尽管这个随时间增长的优先连接模型可以表示为当前网络的度分布指标。但仍有脱离了最初的网络随时间演化的实证研究的嫌疑。Leskovec et al.[40]通过实证研究把网络随时间变化成为网络的“稠化和收缩”（*densification and shrinking*）。他们首先观察到复杂网络，以文献引用网为例，随时间变化越来越密集，这意味着节点的平均度在随时间增长。Leskovec et al.得到了此类网络以进程指数$1 < \alpha < 2$正相关于时间的幂律分布。其中，*e*(*t*)是时间为*t*时边的数量，*n*(*t*)表示时间*t*的节点数量。

$$e(t) \sim n(t)^{\alpha} \qquad (11)$$

接着，他们还发现有效直径（*effective diameter*）随时间缩减。有效直径表示为网络中相连节点间距离的累积分布。实际上网络增长过程中可能只具备上述特征之一，但研究经验显示有必要将其分开考察。需要指出的是Leskovec的网络依赖时间模型的演变并非要求网络属于小世界模型的前提，这为重新考虑和进一步发展复杂网络中依赖时间的模型（*time-dependent models*）提供了参考。目前依赖时间的模型可利用于研究网络文件的变化，例如，研究维基网站中文本节点和链接变化。这种网络时间历时演变的考察方法也是复杂文本网络的语料库语言学分析的现实做法。

前面六小节讨论了从复杂网络的静态特征小世界模型到网络动态增长的无标度模型，从网络增长中节点连接的相关性到比节点更高层次的网络模体和社团结构，最后谈到网络演化的时间模型。这一个个模型渐进地限制了不同类复杂网络的节点连通性和网络结构形式。目的都是为了层层剥离出隐藏在系统复杂性背后的形成机制和演化规律。当今复杂性科学的研究也不再满足于把复杂网络简单描述为“一个由较短的平均路径，较高的聚集系数，度分布符合幂律的多节点网络”[7]，而是要发现更有效的适合大规模节点的网络模型[26][41]来预测社会网络、生物网络、技术网络和语言网络的行为，同时也要发现更多的具有特殊性的模型来区别广泛的复杂系统类型。这一目的也将是语言网络研究的任务。语言网络研究是否能像语言计量研究发现齐夫定律一样，从语言网络中探索出普适的规律和模型来辅助复杂网络分析。在网络结构这一共同的基础上，语言网络的分析是否能为计算机模拟大脑语言能力提供更可行的和可靠的依据？这些问题是我们研究的目标也是动力。

参考文献

[1] 赵怿怡，刘海涛.基于网络观的语言研究[J].厦门大学学报(哲学社会科学版), 2014, 226(6):127-136.

[2] Sigman, M. and Cecchi, G.A. Global organization of the Wordnet lexicon[M]. Procs. Natl. Acad. Sci. USA, 2002, 99(3): 1742-1747.

[3] Ferrer i Cancho, R., Solé, R.V., Köhler, R. Patterns in syntactic dependency networks [J]. Physical Review E, 2004, 69: 051915.

[4] 刘知远, 孙茂松. 汉语词同现网络的小世界效应和无标度特性[J]. 中文信息学报, 2007, 21 (6): 52-58.

[5] Liu, H. The complexity of Chinese dependency syntactic networks[J]. Physica A., 2008, 387: 3048-3058.

[6] Liu, H. Statistical Properties of Chinese Semantic Networks[J]. Chinese Science Bulletin. 2009, 54(16): 2781-2785.

[7] Steyvers, M. and Tenenbaum, J.B. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth[J]. Cognitive Science, 2005, 29(1): 41-78.
[8] 刘海涛. 汉语语义网络的统计特征[J]. 科学通报，2009, 54(14): 2060-2064.
[9] Cong J, Liu H. Approaching human language with complex networks[J]. Physics of Life Reviews，2014(4): 598-618.
[10] Zhao, Y. Three lines to view language network: Comment on “Approaching human language with complex networks” by Cong and Liu[J]. Physics of Life Reviews, 2014(4):637-638.
[11] 赵怿怡，刘海涛.歧义结构理解中依存距离最小化倾向[J].计算机工程与应用，2014,50(6):7-11.
[12] Mehler, A. Large Text Networks as an Object of Corpus Linguistic Studies[A]. In: Lüdeling, A. and Kytö, M. eds. Corpus Linguistics. An International Handbook[M]. de Gruyter: Berlin/New York, 2008: 328-382.
[13] Diestel, R. Graph Theory[M]. Springer, Heidelberg, 2005.
[14] Melnikov, O., Sarvanov, V., Tyshkevich, R. and Yemelichev, V. Exercises in Graph Theory. Kluwer, Dordrecht, 1998.
[15]赵怿怡，刘海涛语言同现网、句法网、语义网的构建与比较[J].中文信息学报，2014,28(5):24-31.
[16] Otte, E., Rousseau, R., 2002, Social Network Analysis: a Powerful Strategy, Also for the Information Sciences[J]. Journal of Information Science, 28, 443-455.
[17] Egghe, L. and Rousseau, R. A measure for the cohesion of weighted networks[J]. Journal of the American Society for Information Science, 2003, 53(3): 193-202.
[18] Solomonoff, R. and Rapoport, A. Connectivity of random nets. Bull. Math. Biophys[M]. 1951, 13: 107.
[19] Erdös, Rényi. On the Evolution of Random Graphs[J]. Bulletin of the Institute of International Statistics, 1961.
[20] Watts, D.J. and Strogatz, S. H. Collective dynamic of “small-world” networks[J]. Nature, 1998, 393: 440－442.
[21] Barabási, A-L., Albert, R. Emergence of scaling in random networks[J]. Science , 1999, 286, 509－12.
[22] Milgram, S. The small-world problem[J]. Psychology Today, 1967, 2:60－67.
[23] Wasserman, S. and Faust, K. Social Network Analysis. Methods and Applications[M]. Cambridge: Cambridge University Press, 1999.
[24] Bollobás, B. and Riordan, O.M. Mathematical results on scale-free random graphs[A]. In Bornholdt, S. and Schuster, H.G., editors. Handbook of Graphs and Networks. From the Genome to the Internet[M]. Wiley-VCH, Berlin, 2003: 1-34.
[25] 刘海涛. 汉语句法网络的复杂性研究[J]. 复杂系统与复杂性科学, 2007, 4(4): 38-44.
[26] Newman, M.E.J. The structure and function of complex networks[J]. SIAM Review, 2003a, 45:167－256.
[27] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N. and Alon, D.C.U. Network motifs: simple building blocks of complex networks[J]. Science, 2002, 298(5594):824－827.
[28] Rapoport, A. Zipf’s law re-visited[A]. In Guiter, H. and Arapov, M. V., editors, Studies on Zipf’s Law[M]. Bochum: Brockmeyer, 1982: 1－28.
[29] Zipf, G.K. Human Behavior and the Principle of Least Effort[A]. An Introduction to Human Ecology[M]. New York: Hafner Publishing Company, 1972.
[30] Barabási, A-L. and Oltvai, Z. N. Network biology: Understanding the cell’s functional organization[J]. Nature Reviews. Genetics, 2004, 5(2): 101－113.
[31] Newman, M.E.J. Power laws, Pareto distributions and Zipf’s law[J]. Contemporary Physics, 2005, 46:323－351.
[32] Watts, D.J. Six Degrees. The Science of a Connected Age[M]. New York/London: W. W. Norton & Company, 2003.
[33] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. and Barabási, A-L. Hierarchical organization of modularity in metabolic networks[J]. Science, 2002, 297:1551－1555.
[34] Simon, H. A. On a class of skew distribution functions[J]. Biometrika, 1955: 42:425-440.
[35] Newman, M.E.J. Assortative mixing in networks[J]. Physical Review Letters, 2002, 89(20): 208701.
[36] Newman, M.E.J. Mixing patterns in networks[J]. Physical Review E, 2003b, 67: 026126.
[37] Newman, M. E J., Park, J. “Why social networks are different from other types of networks,”. Physical Review E, 2003, 68: 036122.
[38] Itzkovitz, S., Milo, R., Kashtan, N., Ziv, G., and Alon, U. Subgraphs in random networks[J]. Physical Review E, 2003, 68: 026127.
[39]Motter, A.E, De, M.A, Lai, Y.C, et al. Topology of the conceptual network of language[J].

Physical Review E. 2002, 65(6): 065102.
[40]Leskovec, J., Kleinberg, J. and Faloutsos, C. Graphs over time: densification laws, shrinking diameters and possible explanations[A]. In KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining[C]. New York. ACM Press, 2005: 177-187.
[41] Bornholdt, S. and Schuster, H.G. Handbook of Graphs and Networks. From the Genome to the Internet[M]. Wiley-VCH, Weinheim, 2003.

作者简介：作者一赵怿怡（1982——），女，博士，副教授、硕士生导师，主要研究领域为应用语言学、依存语法、语言复杂网络。Email: zhaoyiyi@xmu.edu.cn；作者二刘海涛（1962——），通讯作者，男，博士，浙江大学求是特聘教授、博士生导师，主要研究领域为计量语言学、依存语法、语言复杂网络等。 Email: lhtzju@gmail.com。