

汉语依存句法网络的复杂网络性质

刘知远 郑亚斌 孙茂松

智能系统与技术国家重点实验室, 清华信息科学与技术国家实验室,
清华大学计算机科学与技术系, 北京, 100084



摘要: 世界上有几千种人类语言, 它们都有句法规则把词组合成句子。句子中相互关系的词语的排列规则称为句法。语言的句法结构可以通过网络的形式来表示。本文基于大规模语料库, 建立了汉语依存句法网络, 并从复杂网络的角度对该网络进行了系统的实验考察。实验结果表明汉语依存句法网络具有复杂网络的两个基本性质: 小世界效应和无标度特性, 并在其他方面也体现了复杂网络的重要性质。汉语的这些句法上的统计特性, 与捷克语、德语和罗马尼亚语等极为相似, 说明虽然不同语言有着极为不同的句法规则, 但它们具有类似的统计特性, 这种共性对人类语言本质的研究具有一定的启发意义。

关键字: 依存句法 汉语 复杂网络 小世界 无标度 同现

中图分类号: N94, H1

Complex Network Properties of Chinese Syntactic Dependency Network

LIU Zhi-yuan, ZHENG Ya-bin, SUN Mao-song

State Key Laboratory on Intelligent Technology and Systems

Tsinghua National Laboratory for Information Science and Technology

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Abstract: There are thousands of languages on the earth, and they all share some grammar rules to build words into sentences for comprehension. In this paper, we build Chinese syntactic dependency network based on a large corpus and adopt complex network as the tool to analyze the language network. The network shows two important features: the small world effect and the scale-free property. The statistical properties are similar to Czech, German and Romanian which indicate that there are underlying common characteristics among different human languages in despite of their different grammar rules. The common characteristics might make sense for the study of evolution and essence of human languages.

Keywords: dependency grammar, Chinese, complex network, small world, scale free, co-occurrence

1. 简介

最近对复杂网络的研究取得了引人注目的成果[1-5]。研究发现, 各领域中有不同拓扑结构的复杂系统, 如 Internet[6, 7]、万维网[8, 9]、生物学中的食物网[10, 11]以及社会网络[12-14], 都表现出类似的统计规律: 这些网络中节点间的距离都在一个非常小的数值徘徊, 表现出小世界效应; 同时网络节点的连接度呈幂律分布。人类语言也是典型的复杂系统[15], 它在词法、句法、语义等层次上都表现出高度的复杂性。国外已有大量的工作在英语等语言网络上进行[16, 17],

收稿日期: 2008-01-10

基金项目: 国家自然科学基金项目(60573187, 60621062 和 60520130299)。

作者简介: 刘知远(1984 -), 男, 山东泰安人, 清华大学计算机科学与技术系博士研究生。主要研究方向为自然语言处理、复杂网络、数据挖掘等。

包括词同现网络 [18], 词汇联想网络[19, 20], 辞典网络如基于 Roget's Thesaurus 的[2, 20, 21]、基于 Merriam-Webster's Thesaurus 的[3]和基于 WordNet 的[20, 22], 等等。其中对句法网络的研究是语言网络研究的重要组成部分, 吸引了众多国外学者对其深入研究[23-27]。然而据作者所知, 在汉语上句法网络上还没有相应的研究工作。汉语作为世界最重要的语言之一, 它的统计规律和性质, 将具有非常重要参考意义和学术价值, 因此本文将在句法层次考察汉语网络的统计性质。

以往的研究表明人类语言遵循许多统计规律, 其中最为著名的应该是词频上的齐夫定律[28]。但是这种规律没有体现语言的句法性质。而句法是人类语言的关键[29], 因此在句法层次寻找和分析人类语言的统计性质是一项有意义的工作。

句法是把词按一定顺序组合成短语或句子的规则的统称, 这些规则精确地定义了句子中词与词的关系。如果把每个词看作一个节点, 用链接来表示它们在句中的句法关系, 这样就形成句法网络。由于约有 87% 的句法联系发生在距离小于等于 2 的词之间, 所以对词同现网络的研究[18]可以在一定程度上揭示语言的句法统计性质。但这与精确定义的句法关系还有一定差距, 并对远距离的句法关系束手无策。因此对句法网络的研究是非常必要的工作。本文将以汉语的句法网络为研究对象, 揭示其在小世界效应、无标度性、层次性、居间中心性和匹配度等方面表现出来的明显特性。

2. 依存句法网络

本文根据依存句法的定义构造句法网络, 称为依存句法网络。依存句法是法国语言学家 Lucien Tesnière 提出的[30]。他认为句子的主要动词是该句的中心, 支配着其他成分, 而它本身不受任何其他成分支配。后来, Robinson 提出了依存句法的四大公理[31]: (1) 一个句子中只有一个独立成分; (2) 其他成分直接依存于某一成分; (3) 任何一个成分都不能依存于两个或以上的成分; (4) 如果 A 成分能直接依存于 B 成分, 而 C 成分在句子中位于 A 或 B 之间的话, 那么 C 或者直接依存于 A, 或者直接依存于 B, 或者直接依存于 A 和 B 之间的某一成分。

依存句法描述了句子中词与词之间的句法关系, 这种句法关系绝大部分是有向的, 我们假定这个方向由修饰语指向中心词。如“我爱天安门”这一个简单句中, “爱”是句子的中心词, 而“我”和“天安门”是“爱”的修饰语并与之相连。如图 1(A)所示, 链接由修饰语“我”和“天安门”指向中心词“爱”。

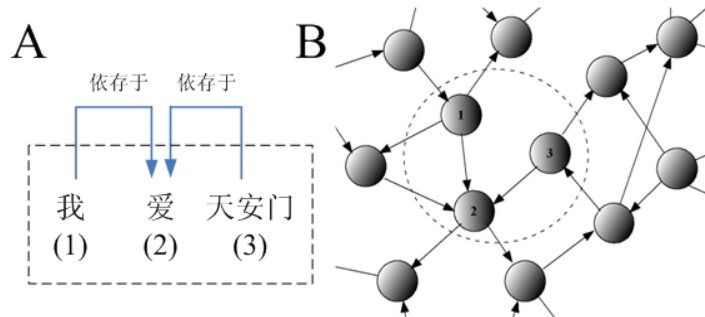


图 1: (A) 句子“我爱天安门”的依存句法结构, 其中动词“爱”是句子中心词, 主语“我”和宾语“天安门”作为修饰语都依存于“爱”。(B) 依存句法网络中, 这三个词对应节点的链接情况。

由于依存句法关系的有向性, 这种关系可以很自然地用有向网络表示。依存句法网络中的节点构成词集合 $V = \{s_i\} (i = 1, \dots, n)$; 网络中的链接集合为 $E = \{a_{ij}\} (i, j = 1, \dots, n)$ 表示, 即如果词 s_i 是词 s_j 的修饰语, 那么 $a_{ij} = 1$, 否则 $a_{ij} = 0$ 。一个句子的依存句法结构可以看作是整个句法关系的子集。如图 1(B)所示, 虚线内的部分就是“我爱天安门”这句话在句法网络中的子图。

在句法网络上，可以计算各种参数来考察其统计特性，如小世界效应[32]，无标度性[33]等。本文以清华大学周强的 100 万词句法标注树库[34]作为语料库，根据周明、黄昌宁在 1994 年提出的汉语依存句法规则[35]构造汉语依存句法网络，对其进行复杂网络特性方面的研究。

3. 网络性质

在依存句法网络上，可以通过以下几个参数考察其复杂网络的统计特性[25]。

1) 小世界效应。有两个关键变量刻画复杂网络的小世界效应。第一个变量是网络的平均路径长度，用 d 表示。定义 $d_{\min}(i, j)$ 是网络中词 s_i 和 s_j 之间的最短路径，那么词 s_i 与其他所有词的平均最短路径为：

$$d(i) = \frac{1}{n} \sum_{j=1}^n d_{\min}(i, j) \quad (1)$$

这样，网络的平均路径长度 d 为：

$$d = \frac{1}{n} \sum_{i=1}^n d(i) \quad (2)$$

其中 n 为网络节点数目。另一个变量是网络的聚合系数，物理意义是网络中某词的任意两个邻居节点互为邻居节点的概率，用 C 表示：

$$C = \frac{1}{n} \sum_{i=1}^n C_i \quad (3)$$

其中 C_i 表示词 s_i 的聚合系数，定义为词 s_i 的邻居节点之间的边数与可能的总边数之比[25]：

$$C_i = \frac{2}{k_i(k_i - 1)} \sum_{j=1}^n a_{ij} \left(\sum_{l=j+1}^n a_{il} a_{jl} \right) \quad (4)$$

其中 k_i 是 s_i 的度。对于 Erdos-Renyi(ER)随机图网络，设网络的平均度为 \bar{k} ，那么该网络的聚合系数为 $C_{\text{random}} \approx \bar{k}/n$ ，平均路径长度为 $d_{\text{random}} \approx \ln n / \ln \bar{k}$ [36]。当网络的平均路径长度 $d \approx d_{\text{random}}$ [32]

的时候，我们称该网络具有小世界现象。而实际网络与 ER 随机图网络的主要区别在于而前者聚合系数 $C \gg C_{\text{random}}$ [1-3]。

2) 无标度性。网络的度分布 $P(k)$ 是刻画网络统计性质的另一个重要参数。ER 随机图网络的度分布近似为 Poisson 分布。而大部分复杂网络的连接度分布具有幂律形式，即满足：

$$P(k) \sim k^{-\gamma} \quad (5)$$

其中 $P(k)$ 表示度为 k 的节点出现在网络中的概率，这种性质也称为无标度性。网络的无标度性一般按照所有度、出度和入度分别考虑。

3) 层次性。研究表明，许多复杂网络同时存在模块性、局部聚类和无标度性，这些模块会按照等级组织起来。一般通过观察 k 与聚合系数 $C(k)$ 的分布来研究层次性，它表示网络中节点的度 k 与 $C(k)$ 的对应关系。某些网络如演员网、近义词网[37]的 $k-C(k)$ 分布图明显呈幂律分布 [37, 38]：

$$C(k) \sim k^{-\theta} \quad (6)$$

且 $\theta \approx 1$ [37, 38]。这表明度很小的节点具有较高的聚合系数而且属于高度连接的模块；而度很高的节点具有较低的聚合系数，其作用只是把不同的模块连接起来。

4) 居间中心性。节点 v 的居间中心值 $g(v)$ 定义如下：设 $G_v(i, j)$ 表示节点 s_i 和 s_j 之间通过节点 v 的最短路径的条数， $G(i, j)$ 表示节点 s_i 和 s_j 之间所有的最短路径条数，即有 $G(i, j) = \sum_v G_v(i, j)$ 。那么 $g_v(i, j) = G_v(i, j) / G(i, j)$ 可以表示节点 v 在节点 i 和 j 的联系中的重要性。而 $g(v)$ 则是所有节点对 i 和 j 的 $g_v(i, j)$ 之和，表示为[39, 40]：

$$g(v) = \sum_{i \neq j} g_v(i, j) = \sum_{i \neq j} \frac{G_v(i, j)}{G(i, j)} \quad (7)$$

许多实际复杂网络满足：

$$P(g) \sim g^{-\eta} \quad (8)$$

虽然大多数复杂网络的度分布都遵循幂律分布，但是 γ 值却不尽相同。而不同复杂网络的 η 却变化很小[39]。

5) 匹配度。如果网络中高连接度的节点倾向于与高连接度的节点相连，就称该网络具有同配性，如果高连接度的节点倾向于与低连接度的节点相连，就称该网络具有异配性。网络匹配度可以通过计算匹配度系数 Γ 测量[41]：

$$\Gamma = \frac{c \sum_i j_i k_i - \left[c \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{c \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[c \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (9)$$

其中 j_i 和 k_i 是第 i 条边两端点的度数。设 m 是网络中边的条数，则 $c = 1/m$ 。若 $\Gamma > 0$ 则网络是同配的，若 $\Gamma < 0$ 则是异配的。研究表明在实际网络中，Internet、WWW、蛋白质交互网、神经网络和食物网满足 $\Gamma < 0$ ，而各种社会关系网络满足 $\Gamma > 0$ [41, 42]。

4. 实验及分析

本文以清华大学周强的 100 万词句法标注树库[34]作为语料库，构建汉语依存句法网络的基本数据如表格 1 所示。从语料库规模来讲，远大于捷克语(562820 词)、德语(21275 词)和罗马尼亚语(153007 词)。在此基础上构建的汉语依存句法网络更能反映句法网络的本质。

数据表明网络的平均路径长度 $d = 3.8$ ，而聚合系数 $C = 0.13$ 。该网络的 $d \approx d_{random}$ 而 $C \gg C_{random}$ ，说明汉语依存句法网络具有小世界效应。这种效应表明虽然网络的规模巨大，但是人们可以只用很短的路径从一个词到达另一个词。这样的句法网络既能够保证人类交流的速度，又能保证交流中用语的丰富性。图 2 显示句法网络路径长度分布情况，可以看到 $d = 3$ 和 $d = 4$ 的路径占了绝大多数。

表格 1: 汉语依存句法网络基本数据，作为对照列出捷克语、德语和罗马尼亚语数据[25]。

参数	捷克语	德语	罗马尼亚语	汉语
n	33,336	6,789	5,563	56,232
\bar{k}	13.4	4.6	5.1	16.7
C	0.1	0.02	0.09	0.13
C_{random}	4×10^{-4}	6×10^{-6}	9.2×10^{-4}	3.0×10^{-4}

d	3.5	3.8	3.4	3.8
d_{random}	4	5.7	5.2	3.9
γ	2.29	2.23	2.19	1.90/2.44
γ_{in}	1.99	2.37	2.20	1.99/2.27
γ_{out}	1.98	2.09	2.20	1.98/2.68
η	1.91	2.10	2.10	1.10/1.86
ζ	1.03	1.18	1.06	1.25
Γ	-0.06	-0.18	-0.2	-0.13

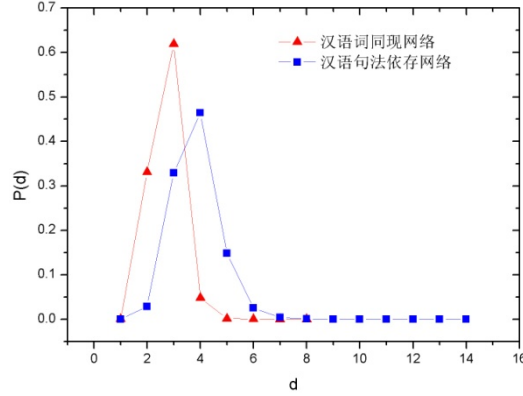


图 2: 汉语依存句法网络的路径长度分布(正方形)。作为比较, 同时画出汉语词同现网络的路径长度分布(三角形)。

图 3、图 4和图 5是汉语依存句法网络的所有度、入度和出度的累积分布曲线。度的累积分布 $P_{\geq}(k)$ 是度不小于 k 的节点的概率, 它与度分布的关系可以表示为:

$$P_{\geq}(k) = \sum_{j \geq k} P(j) \quad (10)$$

当度分布曲线呈幂律分布时, 其累积度分布曲线也呈指数值相差1的幂律分布[2], 根据式(5)可得

$$P_{\geq}(k) = \sum_{j \geq k} j^{-\gamma} \propto k^{-(\gamma-1)} \quad (11)$$

可以看到三个累积度分布曲线都大体呈幂律分布, 显示了汉语依存句法网络的无标度特性。如果对这些曲线进行更为细致的观察, 则会发现其度分布并非一条直线, 而是可以划分为两个斜率不同的线段, 图中箭头所指即为转折点。两段曲线的斜率见表 1。相比词同现网络, 句法网络的两段曲线斜率相差较小[43]。在词同现网络或句法网络上的这种度分布斜率划分为两段的现象, 暗示了人类语言核心词典的存在[18, 43]。核心词典的词汇为该语言的使用者所共用, 其规模不随语言的进化而显著变化, 约为 10^3 量级。核心词典在词同现网络中表现为两个斜率不同的度分布线段。其中属于第二段的词汇度较高, 构成了核心词典; 而第一段则为特定领域所使用的词。

汉语依存句法网络的度-聚合系数 $[k-C(k)]$ 分布图如图 6 所示。该分布呈明显下降趋势, 可以看出网络中既存在很多度虽小却聚合系数很高的节点, 也存在很多度非常高而聚合系数较低的节点。但汉语与捷克语、德语和罗马尼亚语类似, 其依存句法网络的 $k-C(k)$ 分布图并不那么明显遵循 $\theta \approx 1$ 的幂律分布。各语言句法网络的这一相似之处表明它们具有相似的层次结构。

汉语依存句法网络的居间中心性的累计分布曲线如图 7 所示。居间中心性的分布曲线类似于连接度, 因此:

$$P_{\geq}(g) = \sum_{j \geq g} j^{-\eta} \propto g^{-(\eta-1)} \quad (12)$$

图示表明,汉语依存句法网络的 g 累计分布曲线比较明显地分为两段,两段的 η 值如表格 1 所示。这与其他三种语言的分布曲线有明显的不同,可能的原因是汉语网络中缺少更多的关键词(hub word)。而图 8 是度-居间中心性 $[k - g(k)]$ 分布图,可以看到两者之间有较强的正相关关系。

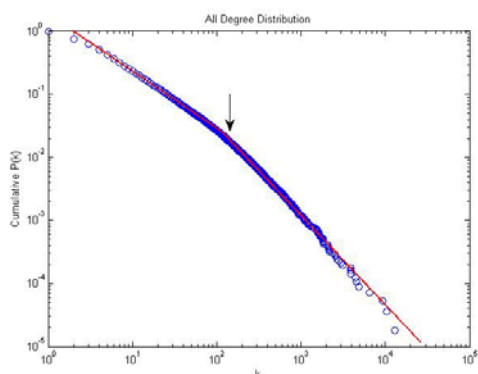


图 3: 所有度的累积分布曲线。

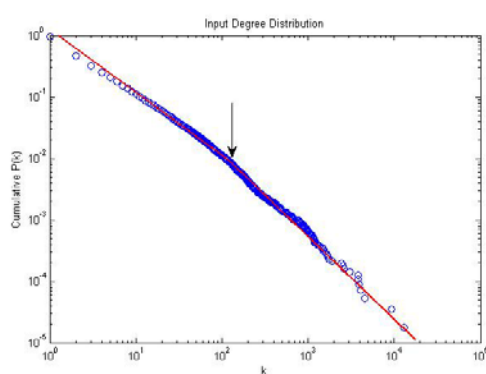


图 4: 入度的累积分布曲线。

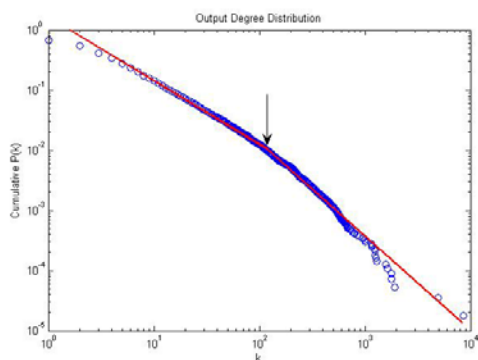


图 5: 出度的累积分布曲线。

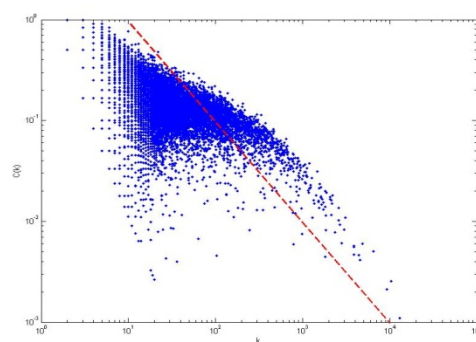


图 6: 度-聚合系数分布图。其中红色虚线的斜率为-1。

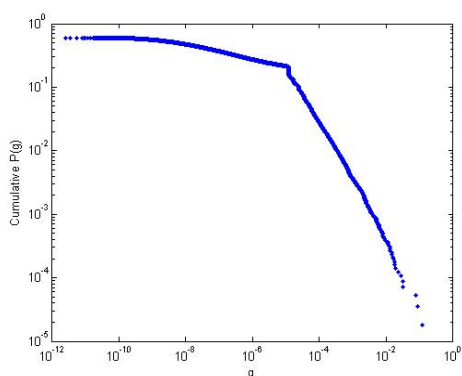


图 7: 居间中心性的累积分布曲线。

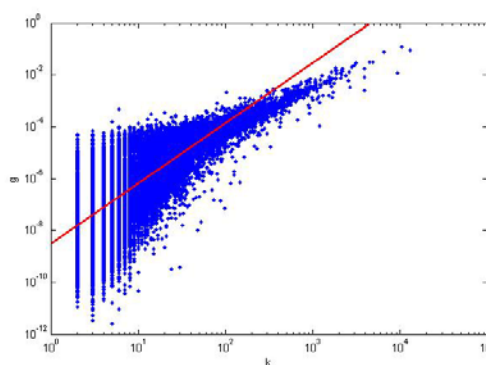


图 8: 度-居间中心性值分布图。

图 9 是汉语依存句法网络中节点(词)的连接度及其在语料库中的词频的分布图,计算表明该分布呈明显的幂律分布:

$$f \sim k^\zeta \quad (13)$$

其中 $\zeta \approx 1$ (见表格 1)。汉语依存句法网络的 ζ 略高于其他三种语言,说明汉语网络更加稀疏,这是语料库的规模较大却并不足够大造成的。其中语料库规模变大使得网络节点增大,而不够大则

使得网络中的链接较人类语言的实际句法网络而言不够充分。根据齐夫定律[28]:

$$P(f) \sim f^{-\beta} \quad (14)$$

一般其中 $\beta \approx 2$ 。而 $\zeta \approx 1$ ，则 $P(k) \sim k^{-2}$ ，即 $\gamma \approx 2$ ，符合表格 1 中数据。

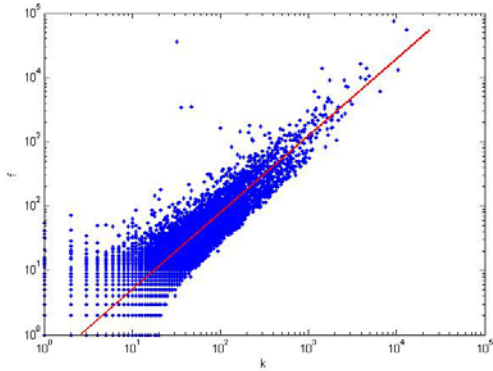


图 9: 度-词频分布图。

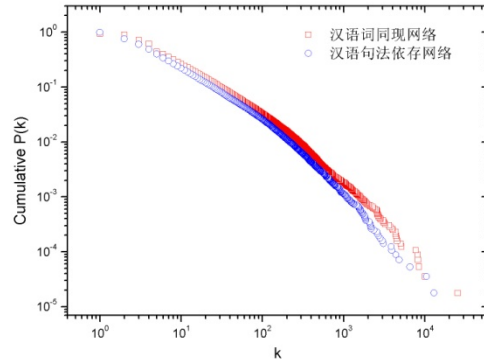


图 10: 汉语词同现网络与句法网络度分布示意图。

度-词频分布图表明了度与词频的统计关系，即词频越高，其度也越高。一般而言，这些高连接度和高词频的节点都是功能词，如“的”，“了”，“在”，“和”等。计算可得汉依存句法网络与其他三种语言类似，其匹配度系数 $\Gamma < 0$ ，说明该网络是异配的，即网络中高连接度的节点倾向于与低连接度的节点相连。实际上，根据句法规则，功能词之间一般不构成句法依存关系，在句法网络中就表现为互相之间没有链接。而在 Roget 词典网络中 $\Gamma = 0.157$ [2, 20, 21]，说明该网络是同配的。

5. 句法网络与词同现网络的比较

根据相同的语料库，我们构造了汉语词同现网络。词同现网络的构造算法十分简单[18, 43]: 语料库所包含的每一个词型 (word type)，对应着词同现网络中的一个节点。如果在一个句子中，两个词之间在 n 阶 Markov 链的条件下存在同现关系，则认为网络中相应的两个节点之间存在一个连接。对语料库中的所有句子进行上述处理，便可构造出词同现网络。一般而言 n 取 2。

首先在度分布上，通过图 10 可以看到汉语句法网络与词同现网络都大致呈幂律分布，并在尾部曲线斜率变大。虽然两者在边数上相差近 1/5，但在度分布上差别并不明显，只是在连接度较高的部分有比较大的偏离。表格 2 可以看到一些主要参数的对比。可以看到，在平均度 \bar{k} 和聚合系数 C 上两者差别较大。由于平均度 \bar{k} 与网络边数有直接联系，两者差别较大并不难理解。对于聚合系数，句法网络相比词同现网络小很多，主要原因是，句法网络中，词是按照句法依存关系建立链接的，与某个词连接的其他词大部分在词性和句法功能上相似(例如都属于动词或名词等)，这些词较少倾向于互相连接，从而导致网络聚合系数较小。例如“文明”分别与“热爱”和“提倡”存在连接，但“热爱”和“提倡”之间存在句法依存关系的概率非常小。图 2 对两种路径长度分布进行了比较，两曲线形状大致相同。如图 11 是“文明”节点在同现网络和依存句法网络中邻居节点的间连接情况的比较，可以明显看出，同现网络的邻居节点间的连接紧密程度高于依存句法网络的邻居节点的紧密程度。

从本质上讲，词同现网络是句法网络的一种近似，可以揭示出人类语言的性质。通过比较，可以看到词同现网络与句法网络存在一定的差别，究竟这些差别是依存句法的什么特性造成的，是一个值得定量分析的问题。

表格 2: 汉语词同现网络与句法网络参数比较。

参数	汉语句法网络	汉语同现网络
n	56,326	56,326
E	447,519	589,654
\bar{k}	15.9	20.9
C	0.1	0.6
C_{random}	2.8×10^{-4}	3.7×10^{-4}
d	3.8	2.7
d_{random}	3.9	3.6
Γ	-0.12	-0.11

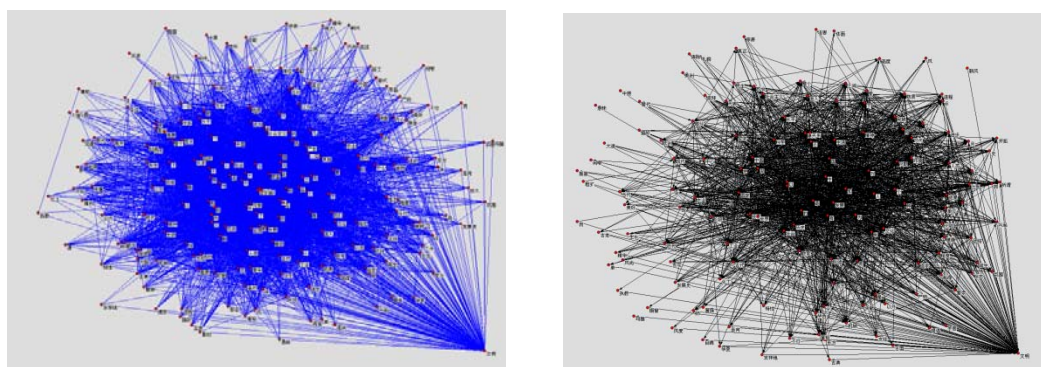


图 11: 节点“文明”在同现网络(左)和依存网络(右)中邻居节点之间的连接情况比较。两幅图的右下角节点是“文明”。

6. 结论

本文展示了汉语依存句法网络的小世界效应、无标度特性以及其他复杂网络性质，与捷克语、德语和罗马尼亚语等做了定量和定性的比较，此外还将汉语词同现网络与句法网络进行了初步比较。本文首次以句法网络的形式研究汉语的句法性质，并与其他语言进行了比较，有利于更全面的把握和分析人类语言的本质。在未来我们可以进行以下工作：(1)词同现网络实际上是句法网络的一种近似，那么如何更深层次挖掘这两者之间的异同，将对于理解人类语言具有重要意义；(2)目前我们还仅在无权的句法网络上进行了研究，而网络连接的权重是非常重要的信息，那么进一步研究带权句法网络将有助于更深入了解语言的性质；(3)目前对于语言网络的研究还停留在整体统计性质上的分析，进一步分析和研究语言网络的局部结构的性质将帮助我们更全面认识语言的本质。

参考文献

- [1] Dorogovtsev S N, Mendes J F F. Evolution of networks[J]. Advances in Physics, 2002. 51(4): 1079-1187.
- [2] Newman M E J. The structure and function of complex networks[J]. Siam Review, 2003. 45(2): 167-256.
- [3] Albert R, Barabási A-L. Statistical mechanics of complex networks[J]. Reviews of Modern Physics, 2002. 74(1): 47-97.
- [4] 郭雷, 许晓鸣. 复杂网络[M]. 上海: 上海科技教育出版社, 2006.
- [5] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用[M]. 北京: 清华大学出版社, 2006.

- [6] Chen Q, Chang H, Govindan R, et al. The origin of power laws in Internet topologies revisited[C]. in Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. 2002. Los Alamitos, CA: IEEE Computer Society.
- [7] Michalis F, Petros F, Christos F. On power-law relationships of the Internet topology[C]. in Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication. 1999. Cambridge, Massachusetts, United States: ACM Press.
- [8] Albert R, Jeong H, Barabási A-L. Diameter of the world-wide web[J]. *Nature*, 1999. 401: 130-131
- [9] Barabási A-L, Albert R, Jeong H. Scale-free characteristics of random networks: the topology of the world-wide web[J]. *Physica A*, 2000. 281(1-4): 69-77.
- [10] Iamnitchi A, Ripeanu M, Foster I, Locating Data in (Small-World?) Peer-to-Peer Scientific Collaborations[A], in Proceedings of the First International Workshop on Peer-to-Peer Systems. Springer-Verlag: Berlin. 2002: 232-241.
- [11] Martinez N D. Constant Connectance in Community Food Webs [J]. *The American Naturalist*, 1992. 139(6): 1208-1218.
- [12] Ebel H, Mielsch L-I, Bornholdt S. Scale-free topology of e-mail networks[J]. *Physical Review E*, 2002. 66(3): 035103.
- [13] Newman M E J, Moore C, Watts D J. Mean-Field Solution of the Small-World Network Model[J]. *Physical Review Letters*, 2000. 84(14): 3201.
- [14] Bearman P S, Moody J, Stovel K. Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks[J]. *American Journal of Sociology*, 2004. 110 (1): 44-91.
- [15] Steels L. Language as a complex adaptive system[C]. in Proceedings of PPSN-VI. 2000. Berlin: Springer-Verlag.
- [16] Mehler A, Large Text Networks as an Object of Corpus Linguistic Studies[A], in *Corpus Linguistics. An International Handbook.*, de Gruyter: Berlin/New York. 2007.
- [17] Sole R V, Murtra B C, Valverde S, et al. Language Networks: their structure, function and evolution[J]. *Trends in Cognitive Sciences*, 2006.
- [18] Cancho R F I, Sole R V. The small world of human language[J]. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 2001. 268(1482): 2261-2265.
- [19] Capocci A, Servedio V D P, Caldarelli G, et al. Detecting communities in large networks[J]. *Physica A*, 2005. 352(2-4): 669-676.
- [20] Steyvers M, Tenenbaum J B. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth[J]. *Cognitive Science: A Multidisciplinary Journal*, 2005. 29(1): 41-78.
- [21] Motter A E, de Moura A P S, Lai Y C, et al. Topology of the conceptual network of language[J]. *Physical Review E*, 2002. 65(6): 065102.
- [22] Sigman M, Cecchi G A. Global organization of the Wordnet lexicon[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. 99(3): 1742-1747.
- [23] Cancho R F I. Euclidean distance between syntactically linked words[J]. *Physical Review E*, 2004. 70(5): 056135.
- [24] Sole R. Language - Syntax for free?[J]. *Nature*, 2005. 434(7031): 289-289.
- [25] Cancho R F I, Sole R V, Kohler R. Patterns in syntactic dependency networks[J]. *Physical Review E*, 2004. 69(5): 051915.
- [26] Cancho R F I, The structure of syntactic dependency networks: insights from recent advances in network theory[A], in *Problems of quantitative linguistics*. 2005: 60-75.

- [27] Cancho R F i, Capocci A, Caldarelli G. Spectral methods cluster words of the same class in a syntactic dependency network[J]. *International Journal of Bifurcation and Chaos*, 2007. 17(7): 2453-2463.
- [28] Zipf G K. *Human Behavior and the Principle of Least Effort*[M]. Cambridge, MA: Addison-Wesley, 1949.
- [29] Hauser M D, Chomsky N, Fitch W T. The faculty of language: What is it, who has it, and how did it evolve?[J]. *Science*, 2002. 298(5598): 1569-1579.
- [30] Tesnière L. *Elements de Syntaxe Structurale*[M]. Klincksieck, Paris: Libraire C., 1959.
- [31] Robinson J. Dependency structures and transformational rules[J]. *Language*, 1970. 46(2): 259-285.
- [32] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998. 393(6684): 440-442
- [33] Barabási A-L, Albert R. Emergence of Scaling in Random Networks [J]. *Science*, 1999. 286(5439): 509-512.
- [34] 周强. 汉语句法树库标注体系[J]. *中文信息学报*, 2004. 18(4): 1-8.
- [35] 周明, 黄昌宁. 面向语料库标注的汉语依存体系的探讨[J]. *中文信息学报*, 1994. 8(3): 35-52.
- [36] Newman M E J. Models of the Small World [J]. *Journal of Statistical Physics*, 2000. 101(3-4): 819-841.
- [37] Ravasz E, Barabási A-L. Hierarchical organization in complex networks[J]. *Physical Review E*, 2003. 67(2): 026112.
- [38] Ravasz E, Somera A L, Mongru D A, et al. Hierarchical Organization of Modularity in Metabolic Networks[J]. *Science*, 2002. 297(5586): 1551-1555.
- [39] Goh K-I, Oh E, Jeong H, et al. Classification of scale-free networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. 99(20): 12583-12588.
- [40] Goh K I, Kahng B, Kim D. Universal Behavior of Load Distribution in Scale-Free Networks[J]. *Physical Review Letters*, 2001. 87(27): 278701.
- [41] Newman M E J. Assortative Mixing in Networks[J]. *Physical Review Letters*, 2002. 89(20): 208701.
- [42] Newman M E J. Mixing patterns in networks[J]. *Physical Review E*, 2003. 67(2): 026126.
- [43] 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.