Journal of Southeast University (English Edition) Vol. 22, No. 3, pp. 413 – 417 Sept. 2006 ISSN 1003—7985

Structures of semantic networks: how do we learn semantic knowledge

Tang Lu[1,2,4] Zhang Yongguang[1] Fu Xue[1,2,3]

([1] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China)

([2] Graduate School, Chinese Academy of Sciences, Beijing 100080, China)

([3] School of Banking and Public Finance, Jiangxi University of Finance, Nanchang 330013, China)

([4] Bussiness School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Global semantic structures of two large semantic networks, HowNet and WordNet, are analyzed. It is found that they are both complex networks with features of small-world and scale-free, but with special properties. Exponents of power law degree distribution of these two networks are between 1. 0 and 2. 0, different from most scale-free networks which have exponents near 3. 0. Coefficients of degree correlation are lower than 0, similar to biological networks. The BA (Barabasi-Albert) model and other similar models cannot explain their dynamics. Relations between clustering coefficient and node degree obey scaling law, which suggests that there exist self-similar hierarchical structures in networks. The results suggest that structures of semantic networks are influenced by the ways we learn semantic knowledge such as aggregation and metaphor.

Key words: semantic networks; complex networks; small-world; scale-free; hierarchical organization

We are concerned the ontogeny and phylogeny of human knowledge. How did the knowledge systems of a man come into being? How did the semantic structures in natural languages as the knowledge representations of the whole community evolve? In this paper, structures of two large semantic networks, WordNet and HowNet, are analyzed and what dynamic mechanisms caused their structures is discussed.

Semantic networks were proposed by Collins and Quillian. They suggested using networks to represent concepts and their relations[1]. After that, several large scale semantic networks have been constructed. WordNet was proposed by George Miller and developed by him and his colleagues[2]. It uses synonym sets (synsets) to represent lexical concepts. Multiple word forms are connected to a synset if these word forms are synonyms. Synsets are connected to other synsets by semantic relations. WordNet 2. 1 contains more than 14×10^4 word forms and more than 10^5 synsets. There are relations between word forms and synsets or among themselves. HowNet is a bilingual common-sense knowledge base that encodes inter-conceptual relations and inter-attribute relations of concepts[3]. HowNet is an organic knowledge system more than a semantic dictionary.

Received 2006-04-25.

Foundation item: The National Natural Science Foundation of China (No. 60275016).

Biographies: Tang Lu (1977—), male, graduate; Zhang Yongguang (corresponding author), male, professor, yzhang@ iss. ac. cn.

Some research has been done in recent years concerning the network structures of language and semantic networks. The research concerns networks based on conceptual similarity, neighboring words in sentence and association[4-8]. Different from most of these researches, we extract networks from WordNet and HowNet based on their semantic relations. We are only concerned with synsets of WordNet and the relations between synsets in this paper because they represent the meaning structure of English.

1 Some Related Concepts of Complex Networks

The research of complex networks has led to a tremendous amount of interest in the study of complex systems in the real world, including the Internet and the world wide web, the food chain web, social networks and biological networks, etc[9-10]. The most interesting features of complex networks are scale-free[9] and small-world[10].

The statistical quantities characterizing small-world networks are clustering coefficient C and the average length of the shortest path L. The clustering coefficient is the probability that any two nodes are connected to each other, given that they are both connected to a common node. The average length of the shortest path measures the minimal number of links connecting two nodes in the network. Following Watts and Strogatz, for a network with n nodes, we calculate

C by taking the average value of clustering coefficients over all nodes i,

$$C = \frac{\sum_{i=1}^{n} C_i}{n} = \frac{\sum_{i=1}^{n} \frac{2T_i}{k_i(k_i - 1)}}{n}$$

where T_i denotes the number of connections between the neighbors of node i, and $k_i(k_i - 1)/2$ is the number of connections in a fully connected graph with k_i nodes[10]. A path is a sequence of edges that connect one node to another. The path length is the number of edges along the path. Denote the shortest path length between nodes i and j as L_{ij}. The average length of the shortest path L measures the average minimal path connecting two nodes in the network.

$$L = \frac{\sum_{i,j}^{i \neq j} L_{ij}}{n(n-1)}$$

Regular networks have high clustering coefficients and large average lengths of the shortest paths, as opposed to random networks which have low clustering coefficients and small average lengths of the shortest paths. Between these two extremes somewhere, the clustering coefficient is almost as high as that of a regular network while the average length of the shortest path is almost as small as that of a random network with the same number of nodes and edges. This type of networks is called as "small-world" for it is similar to the small-world phenomenon. The average length of the shortest path of small-world networks increases slowly with the total number of its nodes: $\bar{L} \sim \ln(N)$.

The study of scale-free networks concerns behavior in the probability distribution of degree, the possible number of links at a random chosen node in the networks. Unlike the Poisson degree distribution for random networks, in a scale-free network, the distribution of degree follows a power law, $P(k) \propto k^{-\gamma}$, where k is the degree of nodes and $P(k)$ is the probability that the degree of an arbitrary node equals k. In such a network most nodes have only a few connections and a few nodes have very large number of neighbors[9].

Barabasi and Albert demonstrated that the power law distribution could be caused by two basic factors: growth and preferential attachment (BA model). Growth means the number of nodes keeps increasing and the preferential attachment means, as the new nodes appear, that they tend to connect to the more connected nodes. The probability for a new node to be connected to an existing node is proportional to the degree of the existing node. A growing network obeying preferential attachment will have an exponent γ typically near 3.0[9].

It has been discovered recently that aggregation and regeneration of nodes can also lead to the power law distribution of degree[11-12]. Kim and his cooperators proposed a network model in which nodes can merge with one of their neighbors and new nodes be added to the network to maintain the number of nodes[11]. Another model proposed by Alava and Dorogovtsev allows for the aggregation of nodes which are selected at random[12]. Those mechanisms give us new insights into how scale-free networks emerge.

Different from BA model networks, some real scale-free networks have hierarchical structures. A model with a network duplication mechanism can cause such a structure[13]. It displays a hierarchical and coarse-grained similarity. This intrinsic hierarchy can be characterized in a quantitative manner. The clustering coefficient of a node with k links follows the scaling law $C(k) \sim k^{-1}$. This type of structure can give an explanation of the feature of the small-world in many scale-free networks.

Degree correlation coefficient r can distinguish assortative and disassortative networks. In assortative networks, nodes with many connections tend to be connected to other nodes with many connections. It is found that social networks are often assortative while biological networks are often disassortative[14]. r can be measured by

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2}(j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2}(j_i + k_i) \right]^2}$$

where j_i, k_i are the degrees of the vertices at the ends of the i-th edge, with $i = 1, 2, \ldots, M$.

2 Complex Networks Properties of Semantic Networks

We analyze seven kinds of properties of the network: sparsity, diameter, shortest path length, clustering coefficient, degree distribution, relation between clustering coefficient and degree, and degree correlation. Some results are shown in Tab. 1, where N is the number of nodes; C_N is the connectedness, the following data are restricted to the largest components; $\bar{k}$ is the average degree of nodes; γ is the power law exponent for degree distribution; α is the scaling law exponent for $C(k)$; D is the diameter of networks; L is the average shortest path length; $\ln(N)$ is the average shortest path length of small-world networks with N nodes; C is the clustering coefficient; C_r is the clustering coefficient of random graphs with the same size and density; r is the degree correlation coefficient. Clustering coefficients of semantic networks are compared with random networks. These random networks have the same size and connection density as corresponding semantic networks.

Tab. 1 Summary statistics for meaning structure of semantic networks

Semantic network	N	$C_N/\%$	$\bar{k}$	γ	α	D	L	$\ln(N)$	C	C_r	r
WordNet	117 948	96. 8	3. 1	1. 65	0. 99	21	8. 17	11. 65	0. 045	0. 000 05	−0. 07
HowNet	55 559	100	5. 9	1. 14	0. 88	10	3. 90	10. 93	0. 238	0. 000 15	−0. 13

A node has only 5. 9 neighbors on average in HowNet. In WordNet this data is 3. 1. Despite their sparsity, all of them have a very large component, while the largest connected component of the random network only consists of almost half of all the nodes. Those features suggests that human beings can efficiently (in a sense use only a few connections) organize their knowledge in an integrated system.

Contrasted with the large component, network exhibits very low diameter and short average path length. The diameter of HowNet is 10 and the average shortest path length is 3. 9. That means that there only need be an average of four links to associate two nodes in HowNet. The corresponding statistics of WordNet are larger due to larger node numbers and lower connection numbers but they are still lower than ln(N). The clustering coefficient of HowNet is 0. 238 and WordNet is 0. 045. All are distinctly larger than the corresponding random networks with the same size and density. We can conclude that they are all small-world networks.

The degree distributions are measured by grouping all values of k into several bins of consecutive values based on logarithmic coordinates and computing the mean value of k for each bin. Different from most scale-free networks which have exponents near 3. 0, the exponents of the best fitting power law of HowNet and WordNet are in the range of 1. 0 and 2. 0 (see Fig. 1). This property shows that it is a very special type of scale-free network. It also suggests that it is not enough to use only preferential attachment mechanism or something similar to it to simulate the development of semantic knowledge because such a mechanism leads to an exponent near 3. 0[7,9]. The degree correlation coefficients of these networks are all lower than 0, which cannot be explained by the preferential attachment mechanism. The degree correlation of the BA model nearly equals 0[14].

We also check the relationships between clustering coefficients and degrees. Fig. 2 shows the average clustering coefficients of nodes with corresponding degrees. They are also measured by grouping all values of k into bins of consecutive values and computing the mean value of clustering coefficients for each bin. All of them display a scaling law with exponents near 1. 0. These evidences strongly suggest that human beings organize their knowledge in hierarchically organized structures with self-similar properties. It also provides an explanation of the small-world feature of these networks[13].

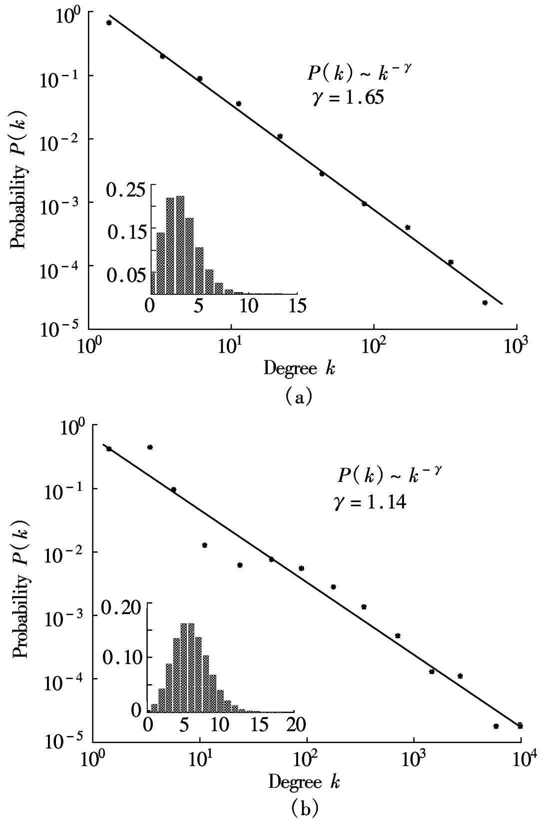

Fig. 1 Degree distributions with the best fitting power law distributions. (a) WordNet synsets; (b) HowNet (The insets depict the degree distributions for equivalent random networks.)

3 Cognitive Mechanisms

Based on the scale-free network models[9,11-13], there exist three kinds of mechanisms which can produce scale-free networks. But none of them can explain the statistic properties of semantic networks solely. Growth and preferential attachment cannot explain the scale law between clustering coefficient and de-

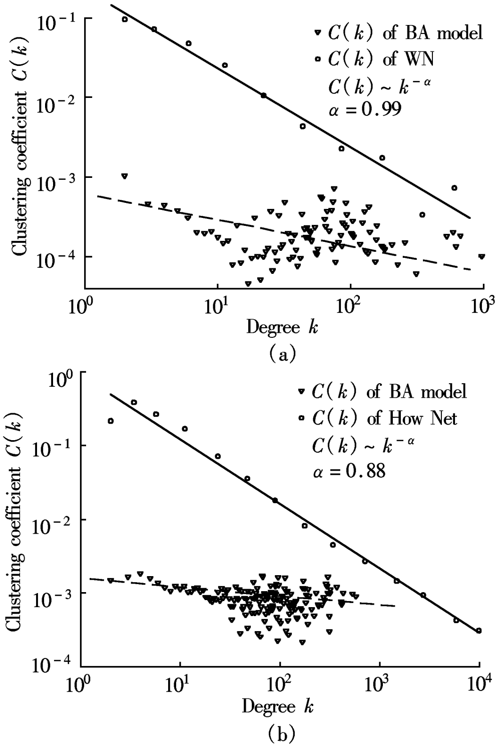

Fig. 2 Statistical $C(k)$ with the best fitting scaling laws. (a) WordNet synsets; (b) HowNet

gree. Aggregation and merging cannot explain the small-world feature. Hierarchical network models cannot explain their disassortative features.

These three mechanisms all exist in semantic knowledge learning processes. Semantic knowledge obviously always grows. There is no evidence of preferential attachment, but we always learn new knowledge based on current knowledge, which means the new knowledge can attach to networks when they have relations to old nodes. When we face new knowledge, we tend not to use our weak domain knowledge but our "hub" nodes, professional domain knowledge, to explain it. For example, an economist tends to use economics core concepts such as market, money, price and commodity to explain social and natural phenomena while a physicist maybe prefers to use physics concepts such as force, movement, dynamics. When the knowledge system of a man is abstracted to nodes and links in networks, it seems as if new nodes prefer to be attached to "hubs".

Aggregation and merging also exist in cognition processes. Cognition processes such as from specific to generic, or induction, can be treated as aggregation. And we use aggregation and merging to form new words in language. The need of aggregation and merging comes from communication and mind operation. Compact and quick communication strengthens our viability. Based on the theory of Consciousness Theater, our capacity of working memory is limited, which means we cannot handle so many elements consciously at the same time[15]. Operations with abstract concepts can help us use fewer elements in our consciousness to handle complex environments and conditions.

One way to obtain hierarchical and self-similar networks is to duplicate part of the networks[13]. We can find that this operation is also a cognition process. A similar operation we frequently use is learning through metaphors[16]. To give an example in cognition science, a computer scientist who knows a little about mind may get some idea on how the mind works if he learned the computer metaphor which says the mind operates like a computer. In this process we can say a duplicate of the knowledge frame of the computer structure is added to his knowledge network as a representation of how the mind works.

None of the mechanisms can explain the structure of semantic networks solely, but all of them exist in semantic cognition processes. We need a model which combines all of these mechanisms.

4 Discussion

It is clear that those semantic networks as representations of meaning structures of human beings have features of complex networks, small-world and scale-free. These features will make human knowledge structures have advantages that complex networks possess. However, there are noticeable differences between semantic networks and typical scale-free networks. The exponents are a prominent difference. All of the semantic networks have exponents lower than 2.0. And the degree correlation coefficients are lower than 0. These are similar to technological and biological networks. There also exist scaling laws between clustering coefficients and degrees. This provides evidence that semantic knowledge may be organized in a hierarchical and self-similar structure. All these remind us that a single mechanism cannot explain the statistical results. We need a combinatorial model. The next work that needs to be done is to find a model which combines these mechanisms and it is also necessary to clearly explain the structures and dynamics of semantic networks.

References

[1] Collins A M, Loftus E F. A spreading-activation theory of

semantic processing [J]. *Psychological Review*, 1975, **82** (6): 407 - 428.
[2] Fellbaum C. *WordNet*, *an electronic lexical database* [M]. Cambridge: MIT Press, 1998. 1 - 20.
[3] Dong Z D. HowNet2000 [EB/OL]. (2000-10-05) [2005-07-30]. http://www.keenage.com.
[4] Ferrer R, Sole R V. The small world of human language [J]. *Proc Roy Soc London B*, 2001, **268** (1482): 2261 - 2265.
[5] Dorogovtsev S N, Mendes J F F. Language as an evolving word web [J]. *Proc Roy Soc London B*, 2001, **268**(1485): 2603 - 2606.
[6] Sigman M, Cecchi G A. Global organization of the wordnet lexicon [J]. *Proc Natl Acad Sci*, 2002, **99** (3): 1742 - 1747.
[7] Steyvers M, 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] Motter A E, de Moura A P S, Lai Y C, et al. Topology of the conceptual network of language [J]. *Phy Rev E*, 2002, **65**(6): 065102.
[9] Barabasi A L, Albert R. Emergence of scaling in random networks [J]. *Science*, 1999, **286**(5439): 509 - 512.
[10] Watts D J, Strogatz S H. Collective dynamics of small-world networks [J]. *Nature*, 1998, **393** (6684): 440 - 442.
[11] Kim B J, Trusina A, Minnhagen P, et al. Self-organized scale-free networks from merging and regeneration [J]. *Euro Phys J B*, 2005, **43**(3): 369 - 372.
[12] Alava M J, Dorogovtsev S N. Complex networks created by aggregation [J]. *Phys Rev E*, 2005, **71**(3): 036107.
[13] Ravasz E, Barabasi A-L. Hierarchical organization in complex networks [J]. *Phys Rev E*, 2003, **67** (2): 026112.
[14] Newman M E J. Assortative mixing in networks [J]. *Phys Rev Lett*, 2002, **89**(20): 208701.
[15] Baars B J. *In the theater of consciousness—the work-space of the mind* [M]. New York: Oxford University Press, 1997. 39 - 61.
[16] Lakoff G, Johnson M. *Metaphors we live by* [M]. Illinois: University of Chicago Press, 1980. 106 - 155.

语义网络的结构:我们怎样学习语义知识

唐　璐[1,2,4]　张永光[1]　付　雪[1,2,3]

([1] 中国科学院数学与系统科学研究院,北京 100080)
([2] 中国科学院研究生院,北京 100080)
([3] 江西财经大学公共管理学院,南昌 330013)
([4] 上海理工大学管理学院,上海 200093)

摘要:分析了 2 个大型语义网络 HowNet 和 WordNet 的全局意义结构. 发现两者都是具有小世界和无尺度特征的复杂网络,但具有一些独特的属性. 两者连接度分布的幂律指数介于 1.0 和 2.0 之间,而不是像许多常见的无尺度网络一样接近于 3.0. 连接度相关系数都小于 0,与生物性网络相似. BA 模型以及与其相似的一些模型不能对其动力学加以解释. 节点连接度与其聚集度指数之间遵循标度律,表明网络中可能存在自相似的层次结构. 认为人类学习语义知识的几种主要方式如聚合与隐喻等影响了语义网络的这些结构特征.

关键词:语义网络;复杂网络;小世界;无尺度;层次结构
中图分类号:TP182