World Wide Web (2014) 17:229–254
DOI 10.1007/s11280-012-0171-7

Discovering small-world in association link networks for association learning

Shunxiang Zhang · Xiangfeng Luo · Junyu Xuan · Xue Chen · Weimin Xu

Received: 31 January 2012 / Revised: 16 May 2012
Accepted: 20 May 2012 / Published online: 12 June 2012

Abstract Association Link Network (ALN) is a kind of Semantic Link Network built by mining the association relations among multimedia Web resources for effectively supporting Web intelligent application such as Web-based learning, and semantic search. This paper explores the Small-World properties of ALN to provide theoretical support for association learning (i.e., a simple idea of "learning from Web resources"). First, a filtering algorithm of ALN is proposed to generate the filtered status of ALN, aiming to observe the Small-World properties of ALN at given network size and filtering parameter. Comparison of the Small-World properties between ALN and random graph shows that ALN reveals prominent Small-World characteristic. Then, we investigate the evolution of Small-World properties over time at several incremental network sizes. The *average path length* of ALN scales with the network size, while *clustering coefficient* of ALN is independent of the network size. And we find that ALN has smaller *average path length* and higher *clustering coefficient* than WWW at the same network size and network average degree. After that, based on the Small-World characteristic of ALN, we present an Association Learning Model (ALM), which can efficiently provide association learning of Web resources in breadth or depth for learners.

Keywords semantic network · association link network · small-world · association learning

S. Zhang · X. Luo (✉) · J. Xuan · X. Chen · W. Xu
School of Computer Engineering and Science & High Performance Computing Center,
Shanghai University, Shanghai 200072, China
e-mail: luoxf@shu.edu.cn

S. Zhang
e-mail: sxzhang@shu.edu.cn

J. Xuan
e-mail: xuanjunyu@shu.edu.cn

X. Chen
e-mail: xuechen@shu.edu.cn

W. Xu
e-mail: wmxu@staff.shu.edu.cn

S. Zhang
School of Computer Science and Engineering, Anhui University of Science & Technology,
Huainan 232001, China

1 Introduction

World Wide Web is the most important media platform including text, audio, image, animation, video, and interactivity content forms, which provides not only the worldwide information sharing platform but also plentiful knowledge. However, the hyperlink in WWW has no semantics, which can support free browsing rather than Web intelligent application such as Web-based learning. To solve this problem, many semantic graphs (e.g., similar relation of Web documents [14], media semantic view [17], Semantic Link Network [32], Association Link Network [19]) have been developed to build the semantic relations among Web resources during the last decades. It can set up critical "semantic bridge" between loose Web resources and Web intelligent application. Semantic graphs can represent or deduce semantic relationship and the interaction between its members. The technique of semantic graph has been gradually applied to the related fields such as personalized hierarchical organization of Web documents [14], multimedia data management [17, 22], Web services [6, 24], Web-based learning [9, 16, 32], etc.

Association Link Network (ALN [19]) is a new type of semantic graph, whose nodes can be multimedia Web resources and links are association relations among them. Association relation dislikes similar semantic relation among Web documents in [14], the specific semantic relation (e.g., *a kind of*) among multimedia data [17, 22], Web services [6, 24] and metadata or classes [9, 16, 32]. It is a general relation and is more easily mined by machine automatically, which can be used to manage, retrieve, navigate association documents and efficiently provide the foundation for Web intelligent application such as Web-based learning and semantic search. In addition, nodes of ALN are represented by Element Fuzzy Cognitive Map (E-FCM) [18] whose semantics acquire from the Web resources. Therefore, ALN can be viewed as a complementary method of multimedia technology for supporting Web-based learning. It can semantically organize loose domain[1] Web resources, aiming to bridge the gap between loose data and Web intelligent application such as Web-based learning, and semantic search.

In this paper, we explore whether ALN has Small-World [28] characteristic to efficiently support association learning (a simple idea of "learning from Web resources"). In our previous work [30], we only focus on whether ALN has Small-World characteristic. This paper not only further deeply analyzes the evolution trend of Small-World characteristic of ALN, but also proposes an Association Learning Model based on Small-World characteristic of ALN to assistant/guide association learning for learners. The contributions of this paper are as the follows:

- At a given network size, ALN exhibits prominent Small-World characteristic, i.e., a short *average path length* and a high *clustering coefficient*. In exploring the Small-World characteristic of ALN, a filtering algorithm of ALN is proposed to generate the filtered status of ALN at a given network size and specific filtering parameter.
- The evolution trend of Small-World characteristic of ALN is presented. The *average path length* of ALN scales with the increasing of network size, while *clustering coefficient* of ALN is independent of the network size. In contrast to WWW with the same size, ALN has smaller *average path length* and higher *clustering degree*.
- Based on the Small-World characteristic of ALN, an Association Learning Model (ALM) is proposed to provide association learning of Web resources in depth or breadth for learners. This model can efficiently provide not only the advances of a given

[1] In this paper, a domain can be a classified dataset such as the health news, or documents databases of specialized subject such as computer science, biological science and so on.

associated problem, but also some basic knowledge of a given problem (e.g., the treatment method, possible risk or aftermath about ‘children allergy’).

The rest of this paper is organized as follows: Section 2 presents related work of Small-World and multimedia technologies for Web-based learning. Section 3 introduces essential definitions and experimental setting for studying the Small-World of ALN. Section 4 proposes a filtering algorithm of ALN and gives the analysis of Small-World of ALN. Section 5 explores the evolution of Small-World properties of ALN. Section 6 gives Association Learning Model of Web Resources based on Small-World of ALN. Conclusions are given in Section 7.

2 Related works

In this section, we review the Small-World in some other types of networks and the multimedia technologies for Web-based learning.

2.1 Small-world model and random graph

Traditionally, the study of Small-World model initially evolves from random graph to Small-World. In 1998, Watts and Strogatz published the first Small-World network model [28] which smoothly interpolates a single parameter between a random graph and a lattice. Their model demonstrated that with the addition of only a small number of long-range links, a regular graph, in which the diameter is proportional to the size of the network, can be transformed into a “small world”. Small-World network means the co-occurrence of a small *average path length* and a high *clustering coefficient*.

Therefore, the comparison analysis between real network and random graph can be viewed as a benchmark for empirical studies of Small-World. The *average path length* of random graph (not an exact formula) scales with the number of nodes and average degree [2, 11, 23, 28]. And the *clustering coefficient* of a random graph is equal to the probability p that two randomly selected nodes are connected [2, 28]. Some empirical examples of Small-World networks reveal that *average path length* of real networks (e.g., film actors, power grid) is almost as small as random graph, yet the *clustering coefficient* of real network much larger than random graph [28]. After that, many studies show that networks both natural and man-made world such as the metabolic network, neural network [23], WWW, co-authors network and actors’ network [2] also reveal Small-World characteristic.

2.2 Small-world of the web

Adamic [1] presents that websites are highly clustered yet the path length between them is small. Two randomly chosen documents on the Web are on average 19 clicks away from each other. The logarithmic dependence of <d> (*average path length*) on N (i.e., network size) is important to the future potential probability of the WWW [3]. Andrei [4] shows that, if a directed path exists, its *average path length* will be about 16. Likewise, if an undirected path exists (i.e., links can be followed forwards or backwards), its *average path length* will be about 6. A general conclusion is that World Wide Web is Small-World network.

In addition, some researchers have studied the Small-World properties of Semantic Web (SW) schemas expressed in either RDF/S or OWL. Kim et al. [13] have discovered that the meta-data and some universally accepted semantics constitute a shared ontology, which can be used to bridge local ontology, much as highly connected people who belonged to many cliques (small-worlds). Hoser et al. [12] analyzed the graph structure of OWL ontology, focusing on measures the diameter and the density of the graphs, as well as on different notions of centrality (i.e., degree, betweenness, and eigenvector centrality). They showed the usefulness of these measures for the identification of important concepts, the concepts clusters, and the core conceptual backbone of ontology. In addition, the average distance of the shortest paths, connecting classes whose label is similar to the search terms, is employed.

Besides, existing studies indicate that Small-World property has been widely used in related fields. For example, in the design or interpretation of routes of network topology, the signal integrity of communication networks, the propagation of diseases, searching [5, 7], information browsing and information retrieve [31], *average path length* and *clustering coefficient* are natural and important network statistical property.

2.3 Multimedia technologies and methods for Web-based learning

In multimedia technologies for Web-based learning, Li et al. [15] discuss the emerging technologies and applications on interactive entertainments, including multimedia information standards and processing, networks and architecture, and user interfaces. Stamou et al. [26] examine the problems of semantically annotating multimedia and describe the integration of multimedia metadata with the Semantic Web to enable effective multimedia search and retrieval. Song et al. [25] propose that using category keywords automatically annotate the images. Li et al. [17] have proposed MediaView as an extended object-oriented view mechanism to bridge the "semantic gap" between conventional databases and semantics-intensive multimedia applications. All these multimedia technologies have made great progress, which can be used integrate multimedia Web resources for supporting association learning.

In implementation methods for supporting Web-based learning, Ng et al. study the performance on multi-server Distributed Virtual Environment (DVE) to effectively manage the large number of 3D objects and large concurrent users [24]. From the view of maintaining interactivity, they well solve the problem of possible significant server workload in a DVE system. Because the learning contents composed of various multimedia resources induce long latency to display on handheld devices such as smart phones, Chang et al. [8] propose an adaptive caching strategy to solve this problem. Because the multimedia learning materials are massively increasing, Wang et al. [27] utilize the P2P network to manage and retrieve the reusable learning materials.

To the best of our knowledge, the Small-World of Semantic Web with association semantics has not been studied in existing work. At the same time, it is a meaningful task how to manage multimedia Web resources material to realize the simple idea "learning from Web resources". This paper not only deeply studies the Small-World property of ALN, but also discusses how to utilize the Small-World of ALN for supporting association learning of domain multimedia Web resources, which will enrich the learning content of e-learning.

3 Problem formulations

3.1 The benchmark for empirical studies of small-world

The most popular manifestation of Small-World is the "six degrees of separation" which is uncovered by the social psychologist Stanley Milgram [21]. The theoretical study of Small-World may be traced to Small-World model [28] which initially evolves from random graph to Small-World by tuning the 'rewired' probability smoothly interpolates between a random graph and a lattice.

The comparison analysis with random graph can be viewed as a benchmark for empirical studies of Small-World of real network (e.g., ALN, film actors, media views). It mainly includes the following two aspects.

1) The *average path length* L of real network is slight greater than *average path length* L_{random} of random graph at the same network average degree and network size. That is, $L \geq L_{random}$. They all scale with the network size and average degree [2, 11, 23, 28].
2) The *clustering coefficient* C of real network is much larger than *clustering coefficient* C_{random} of random graph at the same network average degree and network size. That is, $C >> C_{random}$, the *clustering coefficient* of a random graph is equal to the probability p that two randomly selected nodes are connected [2, 28].

3.2 Definitions

3.2.1 Dynamic threshold

Dynamic threshold is the smallest detectable value which is used to filter the 'weak' association links related to a given node. If the weight of an association link is smaller than this threshold, it will be filtered; otherwise it will be untouched/retained. Let $\{v_j | j = 1...m\}$ denotes the set of adjacent nodes of node v_k; and w_{kj} denotes the weight of the association link$<v_k, v_j>$from node v_k to node v_j. Dynamic threshold [20] can be defined as,

$$\begin{aligned} & th(k,\alpha) = max(S_k) - \alpha * [max(S_k) - min(S_k)] \\ & S_k = \{w_{k1}, w_{k2}, \cdots, w_{kj}, w_{1k}, \cdots w_{jk}\} \end{aligned} \tag{1}$$

where S_k denotes the set of weights of association links related to the node v_k; $max(S_k)$ and $min(S_k)$ denote the maximal weight and minimal weight in S_k; α is a tunable filtering parameter which falls into the interval [0,1].

From this definition, we can find that:

1) The value of dynamic threshold *th* depends on each node itself, because it depends on the maximal weight $max(S_k)$ and minimal weight $min(S_k)$ of the node itself. When the filtering parameter is given, the dynamic threshold of a node is definite.
2) The dynamic threshold *th* falls into the closed interval [0,1]. Because the weight of association links in ALN are all globally normalized into the interval [0,1]. At the same time, filtering parameter α is indicated as a value in [0,1]. Therefore, we have $th \in [0,1]$.

Applying dynamic threshold to filter 'weak' links of ALN has two advantages:

1) It can decrease the complexity of ALN, which facilitates the analysis and understanding of statistical properties and performance of ALN;
2) It can efficiently keep microscopic (i.e., local) features and semantics of ALN to provide the foundation for association learning of Web resources.

3.2.2 Network status of ALN

ALN is a new type of semantic graph which consists of some nodes and association links among them. Some 'weak' association links may be filtered at a given filtering conditions, which lead to new network status of ALN. In our work, there are two status of ALN, namely initial status and the filtered status of ALN.

1) *Initial status of ALN*. If ALN is built by the generating method of original ALN in Luo's method [19] at given *support* and *confidence*, and no any association links are filtered, then we say that the ALN stays at initial status. Let G denote the initial status of ALN, which is defined as follows.

$$
\begin{array}{l}
G=\{V,E\}, V=\{v_i|i=1,2,\cdots,n\} \\
E=\left\{<v_i,v_j>|v_i,v_j\in V\right\}
\end{array}
\tag{2)}
$$

where V represents a set containing n nodes; E represents the set of association links; and $<v_i,v_j>$ indicates an association link from node v_i to v_j.

2) *The filtered status of ALN*. ALN can be simplified (i.e., some 'weak' links can be deleted) by adjusting the filtering parameter of dynamic threshold. We say that the simplified ALN stays at the filtered status. Let G_α denotes the filtered status of ALN, which is defined as,

$$
G_\alpha=\{V_\alpha,E_\alpha\}, V_\alpha=V, E_\alpha\subset E \tag{3}
$$

where V_α is the set of nodes at the filtered status; E_α represents the set of association links at the filtered status, which is the non-empty proper subset of E.

The filtered status of ALN has favorite Small-World properties, which can keep enough association links for semantic searching and association learning.

3.3 Parameter and data set of experiments

1) The parameter of experiments

To fully explore whether ALN has Small-World property or not, we do experiments from two independent dimensions.

- *The filtering parameter*. At a given network size, the filtering parameter can be adjusted to observe whether ALN satisfies Small-World via the changing tendency analysis of *average path length* and *clustering coefficient*.
- *Network size of ALN*. Fix the filtering parameter of dynamic threshold unchanged, we calculate the *average path length* and *clustering coefficient* of ALN at several incremental network sizes to investigate the evolving trend of Small-World properties of ALN.

2) The data set of experiments

To achieve the experimental data set for exploring the Small-World of ALN, we downloaded the Web resources of three domains news including *health*, *internet* and *environment* from www.reuters.com during Jan. 2007 to Dec. 2011. We confine the search process within the Reuters to guarantee the achieved Web pages belonging to the same domain. Based on Luo's method [18], we automatically generate EFCM (i.e., the representation of a node of ALN) for each Web page and built ALN for the above three domains.

4 The analysis of small-world of ALN

The analysis of Small-World characteristic focuses on two independent structural properties, namely the *average path length* and the *clustering coefficient* [28]. In this section, we first present a filtering algorithm to filter the ALN with a given network size. Then we deeply investigate the *average path length* and *clustering coefficient* of ALN on the filtered status by tuning the filtering parameter. Besides, we also analyzing the trends of Small-World properties along with the changing of filtering parameter and compare them with random graph.

4.1 Filtering algorithm of ALN

To deeply study the Small-World of ALN, we design an algorithm to filter the initial status of ALN by tuning the filtering parameter of dynamic threshold. The filtering process is described as algorithm 1, which converts ALN from the initial status into the filtered status for facilitating investigation of Small-World characteristic.

Algorithm 1: Filtering Algorithm of ALN

Input: $G=(V,E)$ and filtering parameter α

Output: $G^{\alpha}=(V,E^{\alpha})$

1: For (i=1;i<=n; i++)
2: $max(S_i) \leftarrow max(w_{ij})$, $min(S_i) \leftarrow min(w_{ij})$
3: $tag(S_i) \leftarrow$ the j in $max(w_{ij})$
4: end for
5: For (i=1;i<=n; i++)
6: $th = max(S_i) - \alpha * [max(S_i) - min(S_i)]$
7: If ($w_{ij} < th$ and $tag(S_i) \neq j$) then $w_{ij} \leftarrow 0$ // the link is filtered
8: end for

The filtering algorithm of ALN mainly includes two operations.

1) The first operation (i.e., step 1 to 4) is used to add an un-deleted tag for the association link $<v_i, v_j>$. If the weight w_{ij} of this link is maximal related to node v_j, then the un-deleted tag is added. The time complexity of this operation is $O(n^2)$.
2) The second operation (i.e., step 5 to 8) is used to filter the 'weak' association link related to each node v_i. Here, 'weak' association link means that an association link is smaller than dynamic threshold. The time complexity of this operation is also $O(n^2)$.

In the above two operations, the first operation ensures that ALN keeps up the good connectivity. That is, any node at initial status of ALN can not be changed into an isolated node at the filtered status of ALN. For example, the weight w_{ij} is smaller than the dynamic threshold of node v_i, but it is the maximal weight related to node v_j. Then the association link $<v_i,v_j>$ should not be filtered. The second operation not only decreases the complexity of ALN, but also keeps up the local features of ALN, which facilitates the observation, analysis of the structural properties of ALN.

4.2 Experimental steps for exploring small-world properties of ALN

Based on news resources of three domains including *health*, *internet* and *environment* downloaded from www.reuters.com during Jan. 2007 to Sep. 2008, we do some experiments to investigate Small-World properties, i.e., *average path length* and *clustering coefficient*. The main experimental steps are as follows.

1) Building the initial status of *ALN*. At a given network size (listed in Table 1), we respectively build ALNs for each of the three domains by the building method of the original ALN [19].
2) Generating the filtered status of ALN. Assign a value to the filtering parameter α and call algorithm 1 to generate the filtered status of ALN.
3) Computing *average path length and clustering coefficient.*

    (1) Dijkstra's algorithm [10] is employed to compute the shortest distance between any pair of nodes of the filtered status of ALN. Further, we achieve the *average path length* of ALN $L(\alpha)$ at a given filtering parameter α.
    (2) The *clustering coefficient* of each node at the filtered status of ALN is computed, further, the network's *clustering coefficient* $C(\alpha)$ achieved, which is the average of *clustering coefficient* of all nodes.

Tuning the filtering parameter α in the interval [0.1,1] (using 0.1 as increment) and executing the above three steps, we can get the *average path length* and *clustering coefficient* at different filtering parameter. The executed results can be denoted as $\{L(\alpha)|\alpha = 0.1, 0.2, \cdots, 1\}$ and $\{C(\alpha)|\alpha = 0.1, 0.2, \cdots, 1\}$.

4.3 The analysis of small-world properties

In this section, we investigate and discuss the *average path length* and *clustering coefficient* of ALN at a given network size. Besides, we compare them between ALN and random graph at the same network size and network average degree.

4.3.1 The average path length of ALN at a given network size

According to the experimental steps for exploring Small-World properties of ALN, we get the *average path length* of ALN of three domains at a given network size and different filtering parameter. Besides, we execute regression analysis to observe the changing trend of *average path length*. The experimental results are listed in Table 2. Herein, R^2 represents the correlation coefficient of regression analysis.

The experimental results and regression analysis of *average path length* of ALN at different filtering parameter are plotted as Figure 1.

According to Figure 1, we summarize two important results:

(1) The *average path length* of ALN is very small, it is about 2 at the initial status of ALN (i.e., $\alpha = 1$), and it gets close to 3 at the filtered status of ALN (i.e., $\alpha = 0.1$). These results indicate that the average value of semantic association steps for any pair of Web pages in the same domain is very small. For example, Web page 'A' discusses mutation about a kind of disease, Web page 'B' may tell you a medicine can efficiently prevent or control the aggravation of this mutation, and Web page 'C' may tell you what producer are yielding this medicine. We can from Web page 'A' to Web page 'C' by only two steps of semantic association.

Table 1 The network size of the three domains.

Health	Internet	Environment	Time span
8168	4158	6690	(01/2007–09/2008)

Table 2 The *average path length* of three domains at a given network size and filtering parameter.

		$L(\alpha)$-health news	$L(\alpha)$-internet news	$L(\alpha)$-environment news
α	0.1	2.62574	2.82746	2.89739
	0.2	2.51834	2.60956	2.71600
	0.3	2.43343	2.53357	2.56257
	0.4	2.38222	2.41845	2.50920
	0.5	2.35103	2.36804	2.42870
	0.6	2.32201	2.28023	2.38436
	0.7	2.28923	2.23074	2.35770
	0.8	2.27896	2.18526	2.34336
	0.9	2.25709	2.16556	2.32659
	1.0	2.23385	2.12302	2.32618
R^2		0.9978	0.9946	0.9812

(2) The *logarithmic* trend of *average path length.* The *average path length* become larger with the emerging of smaller filtering parameter. The reason is that more and more association links are pruned along with α becoming smaller. Further, we find that $L(\alpha)$ and α follow logarithmic function by regression analysis. The correlation coefficients of regression analysis are 0.9978, 0.9946 and 0.9812, respectively. The high correlation coefficient (>0.9) indicates that the regression analysis has a high trustiness.

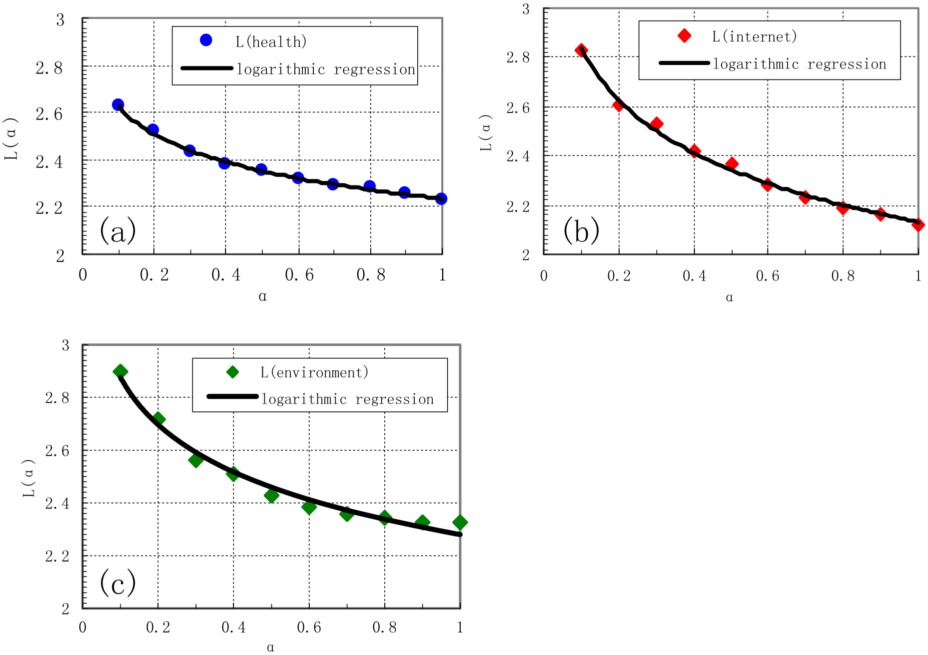

Figure 1 The *average path length* of *ALNs* built for the three domains (i.e., health news, internet news and environment news) by adjusting the filtering parameter α.

Compared with the initial status of ALN, the filtered status of ALN has larger average steps of semantic association for any pair of nodes, but obviously the associated precision among nodes will be enhanced. This result will be used in association learning to satisfy learning requirement of learners with different knowledge background.

4.3.2 The clustering coefficient of ALN at a given network size

Clustering coefficient can measure that network nodes tend to cluster with each other. According to the experimental steps discussed in Section 4.2, we calculate the *clustering coefficient* of ALN built for the three domains at a given network size and filtering parameter. Also, we execute regression analysis according to the achieved *clustering coefficient* to observe its changing trend. The results of experiments and regression are plotted as Figure 2. The correlation coefficients of regression analysis are listed in the bottom row of Table 3.

From this figure, we can get the following conclusions:

- ALN has *high clustering characteristic*. With the changing of filtering parameter from 0.1 to 1, the *clustering coefficient*s of ALN built for three domains are changed alone with them, about from 0.5 to 0.7 in health news, from 0.5 to 0.6 in internet news, and from 0.48 to 0.53 in environment news. Obviously, bigger *clustering coefficient* indicates the higher centrality of the topics of a domain.
- The trend of *clustering coefficient* follows power function, *clustering coefficient*s become smaller with the emerging of smaller filtering parameter. The reason is that some

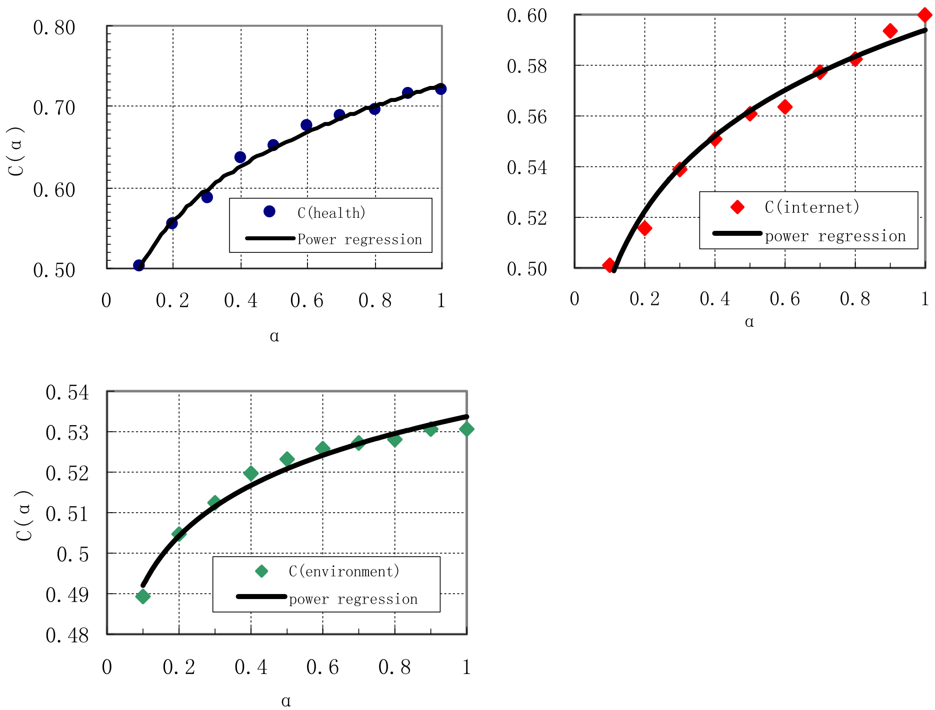

Figure 2 The *clustering coefficient* of ALNs built for the three domains (i.e., health news, internet news and environment news) by adjusting the filtering parameter α.

Table 3 The *Clustering coefficient* of ALN at a given network size and different α.

		$C(\alpha)$-health news	$C(\alpha)$-internet news	$C(\alpha)$-environment news
α	0.1	0.50306	0.51963	0.48858
	0.2	0.55517	0.53152	0.49314
	0.3	0.58781	0.55050	0.49985
	0.4	0.63632	0.55692	0.50492
	0.5	0.65142	0.56538	0.50922
	0.6	0.67509	0.57750	0.51367
	0.7	0.68902	0.58296	0.51759
	0.8	0.69604	0.59511	0.52121
	0.9	0.71604	0.59726	0.52529
	1.0	0.72032	0.60763	0.52605
R^2		0.9427	0.9647	0.9566

association links are pruned at the lower filtering parameter. Also we find that $C(\alpha)$ and α roughly follow power function according to the regression simulation. The correlation coefficients of regression analysis are 0.9427, 0.9647 and 0.9566, respectively. The high correlation coefficient (>0.9) indicates regression simulations have a high trustiness.

4.4 Comparison of SW properties between ALN and random graph

First, we do comparison analysis between the *average path length* of ALN and random graph. According to the random graph theory [2, 11], its *average path length* L_{rand} satisfies

$$L_{rand} \sim \ln(N)/\ln(<k>) \tag{4}$$

where N denotes network size and $<k>$ denotes the average degree of random graph.

We calculate *average path length* of random graph at the same network average degree and network size, and compare them with the *average path length* of ALN built for the three domains. We plot the comparison results between ALN and random graph in Figure 3.

According to Figure 3, we get the comparison result of *average path length* between ALN and random graph. The *average path length* of ALN is slight greater than *average path length* of random graph at the same average degree and network size. This result fully satisfies the benchmark of *average path length* for empirical studies of Small-World of real network.

Further, we do comparison between the *clustering coefficient* of ALN and random graph. According to the random graph theory [2, 11], its *clustering coefficient* C_{rand} satisfies

$$C_{rand} = <k>/N \tag{5}$$

where N denotes network size and $<k>$ denotes the average degree of random graph.

We compare the real *clustering coefficient* (results have been listed in Table 3) of ALN of the three domains and the theoretical value of random graph which is calculated by the random graph theory on the condition of the same network average degree and network size. We plot the comparison results between the *clustering coefficient* of three domains (i.e., health news, internet news and environment news) and random graph in Figure 4.

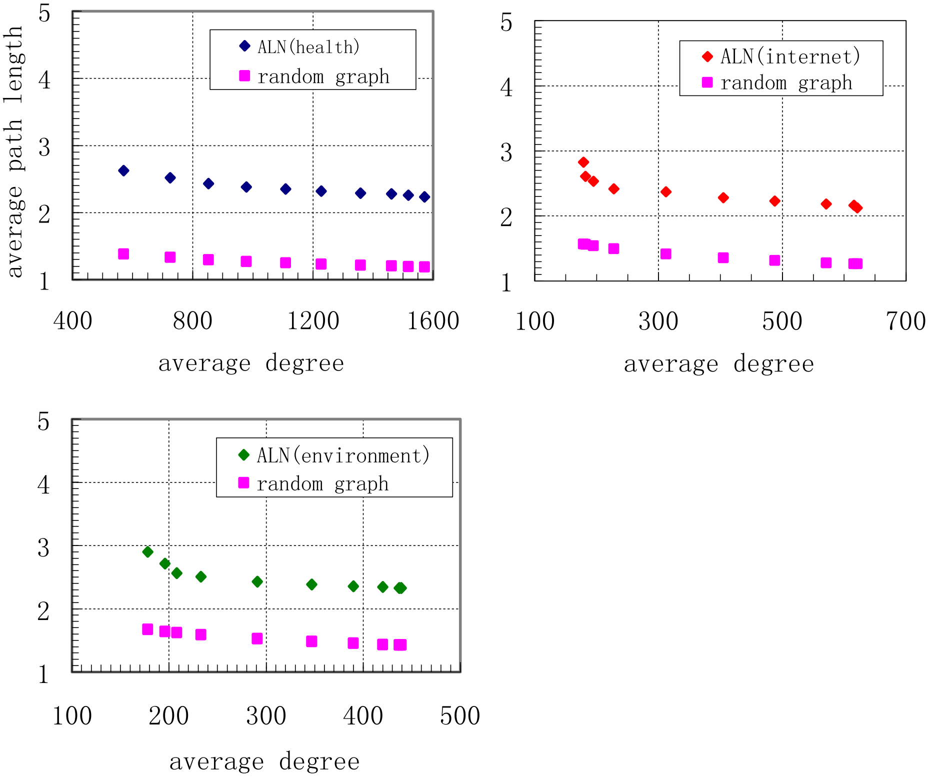

Figure 3 Comparison between the *average path length* of ALNs built for the three domains (i.e., health news, internet news and environment news) and random graph.

From Figure 4, we can get the comparison result of *clustering coefficient* between ALN and random graph. The *clustering coefficient* of ALN is much larger than *clustering coefficient* of random graph at the same network average degree and network size. The greater difference in the value of *clustering coefficient* presents evidence that ALN is prominently different from random network in clustering property.

According to the above analysis, we can conclude that ALN reveals prominent Small-World characteristic, that is, $L \geq L_{random}$, but $C >> C_{random}$. It needs to be pointed out that the Small-World characteristic of ALN has no dependency on a specific domain.

5 The evolution of small-world properties of ALN

In this section, we explore the evolution of Small-World properties of ALN along with the increasing of network size. First, we select three domains including *health*, *internet* and *environment* news and construct seven ALNs according to the given network size for each domain at given time span. The experimental data set are listed in Table 4. For example, *health* domain has 445 web pages during the Jan. 2007. Then we calculate of Small-World properties of ALN to observe their evolution. In the experiments of the evolution, only the network size of ALN is incrementally changed, the filtering parameter of dynamic threshold is fixed to 1 (i.e., ALN stays at initial status, no any association links are filtered).

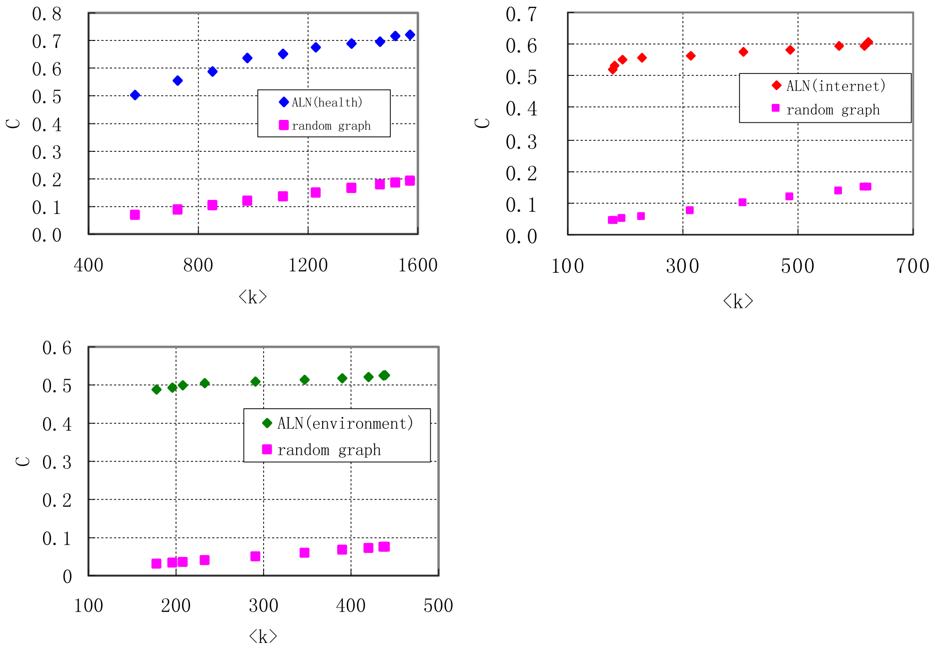

Figure 4 Comparison of *clustering coefficient* between ALN built for the three domains (i.e., health news, internet news and environment news) and random graph.

5.1 Average path length of ALN with the increasing of network size

According to experimental data set of three domains at given time span listed in Table 4, we first construct seven ALNs for each domain by the building method of the original ALN [19]. Then we calculate the *average path length* of these ALNs and compare them with the *average path length* of random graph, WWW at the same network size and network average degree. The comparison results are as follows.

1) Comparison with random graph

According to the random graph theory [2], the *average path length* of random graph L_{rand} satisfies $L_{rand} \sim \ln(N)/\ln(<k>)$. Therefore the product of L_{rand} and $\ln(<k>)$ is close to $\ln(N)$.

Table 4 The experimental data set of ALN for three domains at given time span.

Health news	Internet news	Environment news	Time span
445	213	384	(01/2007–01/2007)
1352	591	836	(01/2007-03/2007)
2816	1165	2223	(01/2007–06/2007)
4055	1821	3396	(01/2007–09/2007)
7185	4004	6295	(01/2007–06/2008)
10739	4706	8500	(01/2007–04/2009)
14627	6357	10365	(01/2007–04/2010)

Thus, we can regard the value of L_{rand} ln(<k>) as a linear function of ln(N) for random graph, which gives a straight line of slope 1.

In Figure 5, we plot a similar product for ALNs built for the three domains, $L(\alpha = 1)$ ln(<k>), as a function of the network size, comparing it with random graph. Each type symbol indicates the real data of ALN belonging to an identical domain, and seven data points of every type symbol are corresponding to the seven ALNs at different network sizes.

From Figure 5, we can see that the evolution of *average path length* of ALN is similar to the theoretical prediction of *average path length* of random graph. That is, the *average path length* of ALN scales with the network size in same way as *average path length* of random graph. In addition, the products $L(\alpha = 1)$ ln(<k>) of ALNs for the three domains are slightly higher than the products of random graph at the same network size and network average degree. This result is accord with the comparison of *average path length* at different filtering parameter and fixed network size (discussed in Section 4.4).

2) Comparison with WWW

Further, we compare the *average path length* of ALNs with it of WWW at the same network size and network average degree. According to our experimental results, we plot the *average path length* of ALNs with the increasing of the network size (discrete points in Figure 6), compare it with WWW (bold line, which is plotted by the experiential formula given by Albert et al. [3]).

Figure 6 shows that the *average path length* of ALN is far smaller than it of WWW at the same network size and network average degree. We know that the Web pages in WWW contain Web resources of all domains, while the Web resources in ALN come from a given domain. Higher centrality of ALN in content leads to the bigger average degree than it of WWW. Accordingly, ALN owns smaller *average path length* than WWW.

5.2 Clustering coefficient of ALN at different network sizes

As we mentioned in Section 4.3, ALN exhibits high clustering characteristic at a given network size. To analyze the evolution of *clustering coefficient* along with the increasing of network size, we calculate *clustering coefficient*s of the seven ALNs at different network sizes (listed in Table 4). The *clustering coefficient* of ALN along with the increasing of network size for the three domains is plotted as Figure 7.

According to the experimental results of the three domains, we summarize two important conclusions:

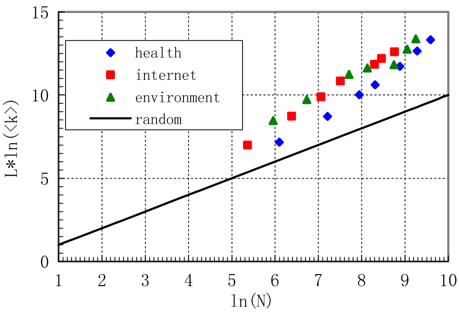

Figure 5 The evolution trend of *average path length* of ALN with the increasing of network size compared with the prediction of random graph theory (*bold line*).

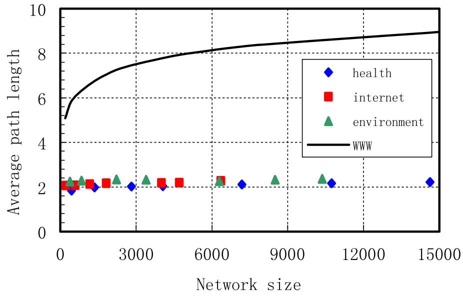

Figure 6 the *average path length* of ALN with the increasing of the network size compared with the experiential value [3] of WWW (*bold line*).

1) The *clustering coefficient* of ALN is independent of the network sizes. With increasing of the network size, *clustering coefficient* of same domain has small change. The plot in Figure 7 convincingly indicates that the *clustering coefficient* of ALN is independent of network size.
2) The *clustering coefficient* of ALN is related to a specific domain. From Figure 7, we find that $C_h > C_i > C_e$, that is, the descending order of *clustering coefficient* of the three domains are health news, internet news and environment news. It also means the topics of health domain have higher clustering degree.

By the way, *clustering coefficient*s of ALN for the three domains are all greater than it of WWW (about 0.29 [2]). The reason is that all the Web resources in ALN come from a given domain, while WWW contains multi-domains Web resources.

6 Association learning model

In this section, we present the Association Learning Model (ALM) of Web resources based on Small-World characteristic of ALN to satisfy the requirements of association learning both in breadth and depth.

6.1 Why we need association learning?

Association learning means that learners can learn/master some associated problems or the basic knowledge about a given problem from Web resources. Compared with the traditional e-learning, association learning from Web resources has two advantages.

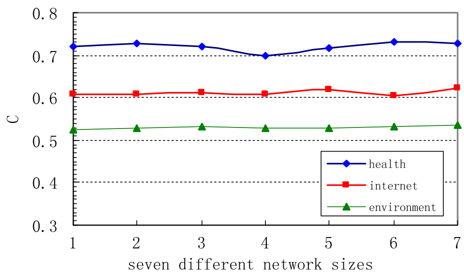

Figure 7 The evolution trend of *clustering coefficient*s of ALN for health news with the increasing of network size compared with the prediction of random graph theory.

(1) *Freshness*. We know that many questions in Web are still difficult to discover suitable answers. Association learning from Web resources can provide the study advances of a special associated problem. For example, in Table 5, the webpage “*Quick brain scan could screen for autism*” tells you that brain scan could be used to test for autism in future, helping doctors diagnose the complex condition more cheaply and accurately. Webpage “*Study finds first evidence that ADHD is genetic*” tells you that British scientists have found the first direct evidence attention deficit/hyperactivity disorder (ADHD) is a genetic disorder.

(2) *Abundance*. We know that Web resources provide plentiful knowledge besides some basic knowledge of a given problem. For example, webpage “*Honey eases nighttime cough*” tells you a simple treatment method for ‘*cough*’. Webpage “*Peanut allergy linked to worse asthma in kids*” tells you possible risk or aftermath of ‘*allergy*’. More examples are listed in Table 5.

6.2 What is association learning both in depth and in breadth?

We know that the learning requirements of users/learners are different in breadth and depth for the same domain because of their different knowledge backgrounds. For example, some people may pay attention to a series of associated problems (listed in Table 5) with children health such as “autism”, “Polio”, “ADHD” and so on. But the

Table 5 Examples of closely associated problems with children health.

Associated problems	Associated sub-problems	The instances of learning resources on the Web
 **autism**	*What cause autism?;* *possible risks or harm;* *the treatment of autism;* *study advances in autism;*	*Genetic study offers answers to autism families* *Autism gene linked to childhood language disorder* *Doctors group calls for universal autism screening* *Quick brain scan could screen for autism*
 **Polio**	*The aftermath of polio;* *polio campaign all over the world;* *The study advances in polio;* *The problem of polio vaccine*	*Polio victims may suffer psychiatric ills later* *Global polio fight to get needed funds- Bill Gates* *Double polio vaccine holds promise for ending virus* *Volcanic ash delays West African polio vaccination*
 ADHD	*Different types of ADHD;* *Could ADHD be Genetic?* *possible risks or harm;* *treatment method;*	*Types: Inattentive, Hyperactive, and Combined Types* *Study finds first evidence that ADHD is genetic* *Brain development slower in children with ADHD* *Group urges heart test before kids get ADHD drugs*
 **Allergy**	*Plants that causes Allergy;* *Food Allergy* *Drug Allergy* *Related role of allergy drugs;*	*Causing Grasses, Weeds, Trees, Flowers and Plants* *Peanut allergy linked to worse asthma in kids* *Drug Allergy Symptoms, Diagnosis, &Treatment* *Allergy drugs may fight diabetes, obesity*
 **Cough**	*What cause cough;* *How to stop cough;* *Active role of allergy drugs;* *Related report;*	*Chronic cough- Iron deficiency could be to blame* *Honey eases nighttime cough* *Cough medicine ingredient may treat prostate cancer* *Few people cover coughs and sneezes*

other people only care of the associated information direct to their interesting problem such as "peanut allergy", "milk allergy" or "egg allergy", which are related to the "Food Allergy" problem.

The above two type of learning instance can be illustrated as Figure 8. If learners who want to know a series of associated problems, the learning type is called "association learning in breadth". If learners who only focus on a given detail associated sub-problem, e.g., food allergy, we call the learning type as "association learning in depth". To implement association learning of Web resources in breadth and depth, the main challenge is how to manage the Web learning resources based on ALN. The results of Small-World characteristic of ALN can be applied to association learning of Web resources.

6.3 Two key factors in association learning

According to the above two type of association learning of Web resources, i.e., association learning in depth and breadth, there are two key factors related to learner's learning requirements to be considered here.

1) *The associated steps*. The number of associated steps between two arbitrary associated problems should be very small. Otherwise the learner may lose the interests and patience during the process of association learning of Web resources.
2) *The associated range*. The associated range should be tunable according to the weight of association link in ALN. For example, if some learners prefer to browse more information, then the range of association learning should be large as possible, i.e., association learning in breath. But the other learners only focus on a special problem, the range of association learning should be accurate as possible, i.e., association learning in depth.

Further, to effectively control the above two factors in association learning of Web resources, we list the correlation between these two factors and the two properties of Small-World of ALN in Table 6.

From Table 6, the first factor, i.e., the associated steps in association learning, is positive correlation with the *average path length* of ALN. Larger *average path length* of ALN may lead to larger associated steps in association learning. Therefore, the network of association learning should own small *average path length*. The second factor, i.e. associated range, is negative correlation with the *clustering coefficient* of ALN. If learners prefer association learning in breadth, we should keep the higher

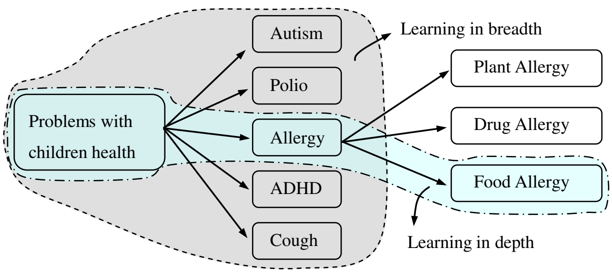

Figure 8 Web resources learning both in breadth and in depth about children health.

Table 6 The correlations between learning factors and properties of Small-World (SW) of ALN.

Factors in association learning	Related properties of SW	Correlation type
associated steps	Average path length	Positive correlation
associated range	Clustering Coefficient	Negative correlation

clustering coefficient for ALN. But if learners prefer association learning in depth, we should keep the smaller *clustering coefficient* for ALN.

Actually, from the analysis in Section 4, filtering the 'weak' association links in ALN can enlarge the *average path length* of ALN and decrease the *clustering coefficient* of ALN. Accordingly, we can control the associated steps and associated range in association learning of Web resources by filtering the 'weak' association links in ALN. Next, we present the basic idea of Association Learning Model (ALM) which incorporates the Small-World results of ALN.

6.4 Framework of association learning model

Building the framework of Association Learning Model of Web Resources mainly includes the following two steps.

1. Built initial status of ALN for the loose Web resources

   (1) Construct the unit of a Web resource. A Web resource unit, as shown in Figure 9 (a), can contain different types of files such as text, audio, image, video clip and so on. We only assign each file an identifier. It dislikes the object re-ranking mechanism for ubiquitous learning environment presented by Yen [29]. It can easily keep the consistency between a Web resource and its files. We represent each Web resource according to unified form as,

$$R = \{R_{id}, id_{text}, id_{image}, id_{audio}, id_{video}\} \tag{6}$$

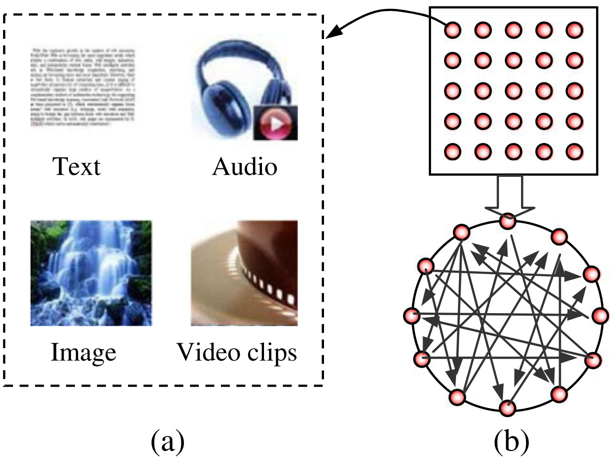

Figure 9 (**a**) A Web resource unit containing text, audio, image, and video clip; (**b**) organize the loose Web resource by initial Associated Link Network.

Where R denotes the unit of a Web resource; R_{id}, id_{text}, id_{image}, $id_{audio}$and id_{video} are respectively the identifiers of a Web resource and its four type files.

(2) Built initial status of ALN for organizing the loose Web resource, as shown in Figure 9(b). According to the text description of each Web resource, we first generate semantic representation, i.e., E-FCM [19]. Then, we generate the association links among E-FCMs to form the ALN of semantic representation of Web resources. Note that, the building process should be carried out at a given *support* and *confidence* to ensure that the built ALN has small *average path length*.

2. Association learning model

Based on the initial status of ALN and the Small-World results of ALN, we construct association learning model, as shown in Figure 10, aim to satisfy different learning requirements of learners. We divide the whole model into five stages according to the learning process.

a) *Select*: When a user/learner wants to deeply understand a given topic, he/she will first search his own background knowledge. The associated knowledge would be activated by his/her consideration. This consideration may lead to two results. One result is, at most time, he/she doesn't know what problems are associated with the given topic. He/she will want to widely browse Web resources, i.e., he/she can select association learning in breadth. The other result is that he/she well understands the given topic, only but want to know a special associated sub-problem such as 'food allergy'. At this condition, he/she can select association learning in depth.
b) *Query*: In query stage, a user/learner first searches his/her memory and organizes relevant pieces of information to make a query decision. In this type of Web resources learning, the representation of query is an important task. The simple style of query can be represented by the keywords set. For example, a user/learner can input 'children health' or 'drug topic' and so on. He/she also can input a sub-problem 'children allergy', 'food allergy', 'statin drug', 'drug price' and so on. Learning model will carry out matching between query and topic. If matching is successful, the model will go to next stage to search database of ALN, otherwise, the model will end the query.

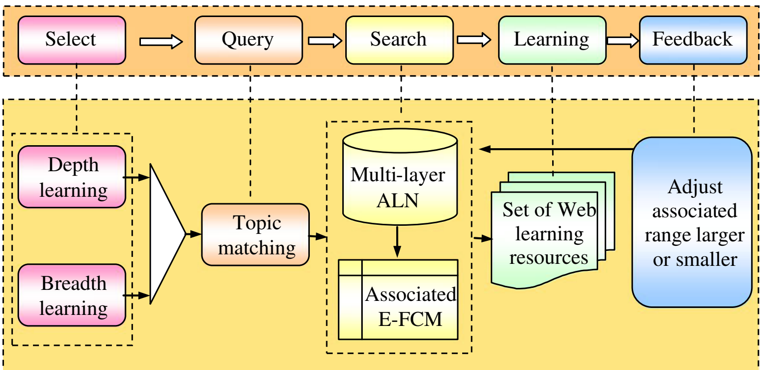

Figure 10 Association learning model based on small-world of ALN.

c) *Search*: In search stage, the association links of ALN will be divided into multi-layer according to the results of Small-World of ALN. All association links are stored in database using the adjacent table. If a user/learner has selected learning in breadth, the model will search all Web resources which have an association with the given query. If a user/learner has selected learning in depth, the model will search the Web resources according to the layer of association links with larger weight. It should be stressed that the searching results are semantic representation of Web resources, i.e., E-FCMs.

d) *Learning*: In this stage, the set of Web learning resources will be provided for learner to carry out association learning in breadth or depth. In our ALN, a Web learning resource and its corresponding E-FCM have the same identifier. So, the model can easily present the set of Web learning resources according to the searching results of E-FCMs.

e) *Feedback*: After a user/learner has browsed the Web resources provided by the model at the first time, he/she needs to evaluate whether these Web resources in content are enough to satisfy his/her knowledge requirement. If the requirement is not satisfying, the model will turn to the search stage, adjust associated range of Web resources larger or smaller for re-searching Web resources.

Note that, in the stage c), we need to classify the association links of ALN into multi-layer. Algorithm 1, i.e., the filtering algorithm of ALN, can provide the set of the filtered status of ALN E^{α}, α=0.1,0.2,· 1. Obviously, there is $E^{0.1} \subset E^{0.2} \subset \cdots E^{1}$. Herein, $E^{0.1}$ contains association links with the larger weight than $E^{0.2}$, which facilitates the association learning of Web resources in depth. While the association links of initial status of ALN E^{1} provide more Web resources at large for learner's association learning in breadth.

From the analysis of Small-World introduced in Section 4, we know that the *average path length* of the filtered status of ALN is still very small. This result indicates that the number of associated steps at filtered status of ALN is small. More importantly the associated range is changed at the same time, which is shown as Figure 11.

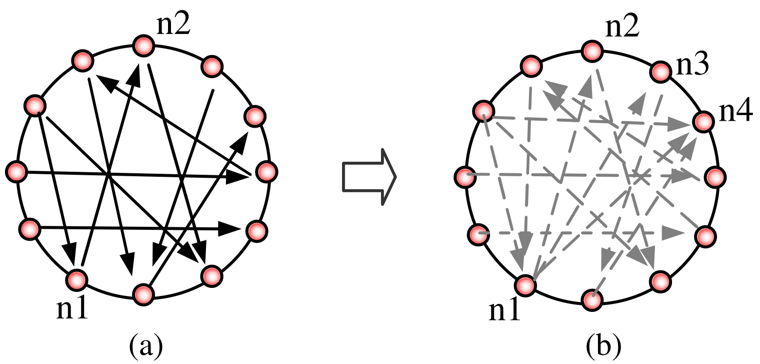

Figure 11 **a** Supporting for learning in depth, at the smaller filtering parameter, only the node 'n2' is reachable from node 'n1'. **b** Supporting for learning in breadth, at the larger filtering parameter, the nodes 'n2, n3, n4' are reachable from node 'n1'.

6.5 Algorithm description of association learning model

According to the framework of Association Learning Model, we present description of two algorithms of Association Learning Model (ALM). One is the layer partition which is described as algorithm2. The other is the learning process which is described as algorithm 3.

Algorithm 2 : Layer partition of ALN

Input: $G = (V, E)$
Output: $G^p = (V, E^{0.1}, E^{0.2}, E^{0.3}, E^{0.4}, E^{0.5}, E^{0.6}, E^{0.7}, E^{0.8}, E^{0.9}, E^1)$

Tag all $links \in E$ using 1 to form E^1
For (i=9;i>=1; i--)
  $\alpha \leftarrow i * 0.1$
  Call algorithm 1 with parameter α
  Tag all $links \in E^{\alpha}$ using α
end for

Algorithm 2 repeats calling algorithm 1 to divide all association links at the initial status of ALN into multi-layers, which satisfy $E^{0.1} \subset E^{0.2} \subset \cdots E^1$. The time complexity of this algorithm is much large, i.e., $O(n^2)$. The layer partition of ALN has completed before the association learning. Therefore this algorithm does not affect the efficiency of association learning.

Algorithm 3: Learning Algorithm of Web resources based on SW

Input: learning type $ltype$, query $query$, and feedback parameter fp
Output: the set of Web learning resources V^s

get $ltype$ of a learner
get $query$ of a learner
For (i=1;i<=n; i++)
  $matching(deacedent \quad or \quad associationrule)$
  $S(topic) \leftarrow matching(associationrules)$
end for
$\alpha \leftarrow 0.1$
while(true)
  if $ltype == 0$ then find E-FCM based on E^{α}
  else find E-FCM based on $E^{1-\alpha}$
  Get fp
  $\alpha \leftarrow \alpha + 0.1$
  If $fp > 0$ then find E-FCM based on E^{α}
  Else If $fp < 0$ find E-FCM based on $E^{1-\alpha}$
    Else exitwhile
Endwhile

In algorithm 3, there are mainly two operations. The first operation implements the topic matching. If a user/learner wants to learn in breadth, then the model executes topic matching which uses query as descendants of association rule. If a user/learner prefers association learning in depth, then the model executes topic matching which uses query as association rule. The second operation automatically selects the layer of association links of ALN to search Web resources for learning both in breadth and in depth. The model can adjust the associated range of Web resources larger or smaller according to the feedback of learner's learning process. For example, if the learner wants to adjust associated range larger, the model may jump from $E^{0.1}$ to $E^{0.2}$ for the searching of Web resources. On the contrary, the model may jump from E^{1} to $E^{0.9}$ to execute the searching of Web resources.

6.6 Learning instance of web resources

Below we give some learning instance of Web resources by the help of Association Learning Model based on ALN. In Table 7, we list three learning instances of Web resources (from http://www.reuters.com/healthnews). The first learning instance belongs to learning in breadth. According to 'children health' topic, ALM model provides its many associated problems such as ADHD, autism, polio, allergy, cough and so on. The high clustering property of Small-world of ALN ensures that ALM can provide many associated problems.

The second learning instance belongs to is learning in depth. According to 'food allergy', ALM model provides accurate associated sub-problems such as milk allergy, peanut allergy, egg allergy and so on. The short average path of Small-world of ALN ensures implementation of learning in depth. The third instance also belongs to learning in depth, which provides suggestion for allergy on holiday. More details see the bottom row of Table 7.

6.7 Evaluation

Here, we demonstrate the advantage of using ALN with Small-world characteristic for association learning. We know that domain Web resources based on hyperlink can not

Table 7 Learning instance of Web resources based on ALM.

Query	The learning instances of Web resources
Advances in children health (Learning in breadth)	Premature birth tied to increased risk of ADHD
	Is ADHD tied to adulthood obesity?
	One in 38 kids in South Korea may have autism
	Brain scans accurate at spotting autism
	Double polio vaccine proves most effective in study
	Childhood pets linked to lower allergy risk
	Fruits and veggies may not lower kids' allergy risk
	How far do your sneeze and cough go? Experts study
Food allergy (Learning in depth)	Peanut allergy linked to worse asthma in kids
	Cake may be the answer to kids' egg allergy
	Most children with milk allergy tolerate warm milk
	Therapy may help some with deadly peanut allergy
Suggestion for allergy on holiday	* Try to keep stress, which can trigger an asthma attack, to a minimum.
	* Bring along your own pillow with an allergen-proof cover, or ask for non-down pillows if you're staying at a hotel or with a relative or friend. Down pillows harbor more dust mites than pillows with synthetic fill.
	* If you're using artificial snow spray to decorate, follow instructions carefully; the spray can be a lung irritant.

provide any help for association learning. However, ALN can effectively assist/guide the association learning process of learners, because it can provide much complete and rich association semantic among domain Web resources. To demonstrate the significance of using ALN for association learning, we design a questionnaire as Table 8.

In this questionnaire, we list five topics of health domain. For each topic, there are one or two questions. We give the questionnaire to 10 post graduates who major in computer science. And ask them write answers according to their own background knowledge. That is, they give answers with no help of ALN. Then we do comparison analysis between their answers and the answers provided by Association learning Model using ALN.

In order to facilitate the comparison analysis, we give the main associated problems of each topic by learning and summarization, which is shown in Table 9. First, we repeatedly learn all health news at http://www.reuters.com during the period from 07/2011 to 12/2011, 2117 Web resources in total. Then, based on our learning process, we summarize all possible associated problems for each topic and seriously consider the reasonability of associated problems. For example, in the 'health risk' topic, there is 'grain' intake related to 'health risk'. The reason is that eating too much grain will make diabetes worse. Also, 'coffee' is related to 'health risk' topic. Because coffee may reduce the risk of developing gallstones, discourage the development of colon cancer, improve cognitive function, and reduce the risk of liver damage.

Using the main associated problems listed in Table 9 as standard answers, we compute the average precision of all respondents (i.e., ten post graduates) and the precision provided by Association Learning Model using ALN with Small-world characteristic on each topic respectively. Based on this statistical method, we plot the experimental results as Figure 12.

From Figure 12, we know that: (1) with no help of Association Learning Model using ALN (i.e., 'no ALN'), the cognition precision degree of all respondents are very low. It is less than 40 % and about 30 % on average for all five topics. In fact, from questionnaire we find that most respondents only know very little associated problems and only a very few of respondents know much associated problems. It is undeniable that the limitation of respondents' background knowledge is one of reasons of low cognition precision degree. (2) With the help of Association Learning Model using ALNwith Small-world characteristic (i.e., 'ALN'), the cognition precision degree of all respondents can be improved. It is close to 90 % on average for all five topics. The experiments verify that Association Learning Model using ALNwith Small-world characteristic can enumerate most associated problems with a given topic according to learners' requirements. Therefore, Association Learning Model

Table 8 A questionnaire about associated problems in health domain.

No.	Topics	Questions
1	Health risk	Can you say some intake risks of food related to diabetes or heart attack?
		Can you say some risks of drug related to diabetes or heart attack?
2	Cancer	Can you say some types of cancer?
		Can you say treatments about breast cancer?
3	Children health	Can you say some diseases about children health?
4	Flu virus	Can you write the main districts where flu virus has happened?
		Can you write the main animals related to flu virus?
5	Woman health	Can you write the main associated problems related to woman health?

Table 9 The main associated problems by learning and summarization.

No.	Topics	Associated problems
1	Health risk	Food intake risks mainly includes: meat, vegetable, soda, grain, calcium Vitamin and coffee etc.
		Drugs intake risks mainly includes: aspirin, statin, avandia etc.
2	Cancer	Types of cancer mainly includes: prostate, breast, oncology, throat etc.
		Treatments about breast cancer: diagnosis, testing, radiation and mammogram etc.
3	Children health	The long-term concerned diseases about children health: autism, ADHD, polio, cough, chickenpox, asthma and allergy etc.
4	Flu virus	The main districts: Egypt, Indonesia, china and Vietnam etc.
		The main animals related to flu virus: bird, chicken, duck, and swine etc.
5	Woman health	The main associated problems related to woman health: pregnancy, menopause, hpv virus, fertility, abortion, fetus, delivery, breast and estrogen etc.

using ALN can assistant/guide the association learning process of learners, make the learning content clearer, and relieve the cognition burden of learners.

7 Conclusions

Association Link Network (ALN) can organize the loose multimedia Web resources with association semantics, aiming to bridge the gap between the loose Web resources and the Web-based learning. This paper has investigated the Small-World properties of ALN to effectively provide a theoretical support for association learning (a simple idea of "learning from Web resources"). Our contributions mainly include three aspects.

1) ALN exhibits prominent Small-World characteristic, a short *average path length* and high *clustering coefficient*. To facilitate deeply exploring the Small-World properties of

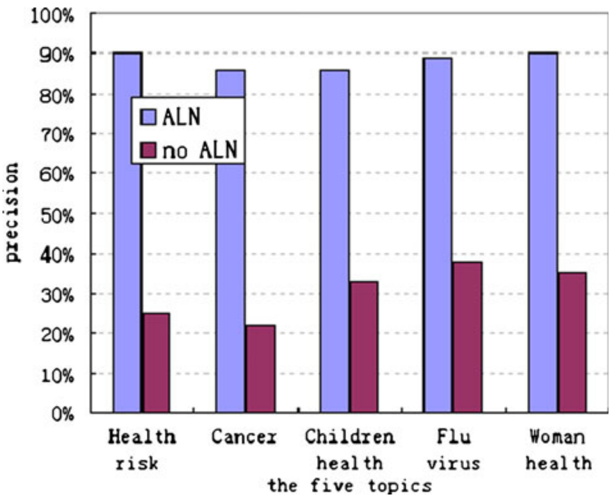

Figure 12 Comparison experiments between using ALN and not using ALN.

ALN, the filtering algorithm of ALN is proposed for generating the filtered status of ALN at given network size and filtering parameter.

2) The evolution of Small-World properties of ALN is presented, the *average path length* of ALN scales with the increasing of network size, while *clustering coefficient* of ALN is independent of the network size. Besides, ALN has smaller *average path length* and higher *clustering coefficient* than WWW at the same network size and average degree.
3) Based on the Small-World characteristic of ALN, an Association Learning Model (ALM) of Web Resources is proposed to satisfy the learners' requirements of association learning both in depth and in breadth. This model can efficiently provide both some basic knowledge and the advances of a special topic.

The Association Learning Model (ALM) is a kind of new multimedia technology, which effectively supports association learning from Web resources. The Small-World characteristic of ALN can be used to control the process of association learning. ALM not only can provide freshness and abundant learning content, but also can relieve the cognition burden of association learning process. Our future works include deducing the mathematical theory of Small-World of ALN and how to further extend the theory of Small-World of ALN to the related Web activities such as intelligent browsing, learning and recommendation of Web knowledge.

Acknowledgments Research work reported in this paper was partly supported by the key basic research program of Shanghai under grant no. 09JC1406200, by the National Science Foundation of China under grant nos. 91024012, 61071110 and 90612010, the National High-tech R&D Program of China under grant no. 2009AA012201 and by the Shanghai Leading Academic Discipline Project under grant no. J50103.

References

1. Adamic, L.-A.: The small world Web. In Lecture Notes in Computer Science. Springer, New York, **1696**, 443–454 (1999)
2. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 48–97 (2002)
3. Albert, R., Jeong, H., Barabási, A.-L.: Diameter of the world-wide Web. Nature **401**, 130–131 (1999)
4. Andrei, B., Ravi, K., Farzin, M.: Graph structure in the Web. Comput. Network **33**, 309–320 (2000)
5. Brin, S., Page, L.: The anatomy of a large scale hyper textual Web search engine. In: Proc. 7th WWW, (1998)
6. Cao, J., Wang, J., Law, K., Zhang, S., Li, M.: An interactive service customization model. J. Inform. Software Tech. **48**(4), 280–296 (2006)
7. Chakrabarti, S., Dom, B., Gibson, D. et al.: Automatic resource compilation by analyzing hyperlink structure and associated text. In: Proc. 7th WWW, (1998)
8. Chang, H.P., Shih, Tk, Li, Q., Wang, C.C., Wang, T.H., Chao, L.R.: An adaptive caching strategy for m-learning based on SCORM sequencing and navigation. World Wide Web **11**(3), 387–406 (2008)
9. Chen, J., Li, Q., Jia, W.: Automatically generating an E-textbook on the web. World Wide Web **8**, 377–394 (2005)
10. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**, 269–271 (1959)
11. Fronczak, A., Fronczak, P., Hołyst, J.A.: Average path length in random networks. Phys. Rev. E **70**, 056110-1–056110-7 (2004)
12. Hoser, B., Hotho, A., Jäschke, R. et al.: Semantic network analysis of ontologies. Third European Semantic Web Conf., (2006)
13. Kim, H.-M., Biehl, M.: Exploiting the small-worlds of the semantic web to connect heterogeneous, local ontologies. Inform. Tech. Manag. **6**, 89–96 (2005)
14. Lau, R.W.H., Si, A., Kam, L.: Personalized hierarchical organization of web documents. J. Internet Technol. **2**(4), 253–261 (2001)

15. Li, F., Lau, R.: Emerging technologies and applications on interactive entertainments. J. Multimed. **6**(2), 107–114 (2011)
16. Li, Q., Lau, R.W.H., Shih, T.K., et al.: Technology supports for distributed and collaborative learning over the internet. ACM Transactions on Internet Technology, **8**(2), 10:1–10:24 (2008)
17. Li, Q., Zhao, J., Zhu, X.: Multimedia data modeling through a semantic view mechanism. World Wide Web **11**(2), 287–311 (2008)
18. Luo, X.F., Fang, N., et al.: Semantic representation of scientific documents for the e-science Knowledge Grid. Concurrency Comput. Pract. Exp. **20**(7), 839–862 (2008)
19. Luo, X.F., Xu, Zh, Yu, J., et al.: Building association link network for semantic link on web resources. IEEE Trans. Autom. Sci. Eng. **8**(3), 482–494 (2011)
20. Luo, X.F., Zhang, J., Liu, F.F., et al.: Merging textual knowledge represented by element fuzzy cognitive maps. J. Softw. **5**(2), 225–234 (2010)
21. Milgram, S.: The small-world problem. Psychol. Today **1**(1), 61–67 (1967)
22. Moxley, E., Tao, M., Manjunath, B.S.: Video annotation through search and graph reinforcement mining. IEEE Trans. Multimed. **12**(3), 184–193 (2010)
23. Newman, M.E.J.: The structure and function of complex networks. SIAM Rev. **45**, 167–256 (2003)
24. Ng, B., Li, F., Lau, R., Si, A., Siu, A.: A performance study on multi-server DVE systems. Inform. Sci. **154**, 85–93 (2003)
25. Song, Y., Wang, W., Zhang, A.: Automatic annotation and retrieval of images. World Wide Web **6**(2), 209–231 (2003)
26. Stamou, G., van Ossenbruggen, J., Pan, J.Z., Schreiber, G., Smith, J.R.: Multimedia annotations on the semantic web. IEEE Multimed. **13**(1), 86–90 (2006)
27. Wang, G.R., Yuan, Y., Sun, Y.J., Xin, J.C., Zhang, Y.: PeerLearning: a content-based e-learning material sharing system based on P2P network. World Wide Web **13**(3), 275–305 (2010)
28. Watts, D.J., Strogatz, S.H.: Collective dynamics of "Small-World" networks. Nature **393**, 440–442 (1998)
29. Yen, N.Y., Shih, T.K., Jin, Q., et al.: Adaptive object re-ranking mechanism for ubiquitous learning environment. J. Multimed. **6**(2), 129–138 (2011)
30. Zhang, S.X., Luo, X.F., Xuan, J.Y., Chen, X., Xu, W.M.: Discovering small-world in association link networks for web-based learning, MTDL'11, 19–24 (2011)
31. Zhang, Zh., Chen, L., Zhou, Sh., et al.: Analytical solution of average path length for Apollonian networks. Phys. Rev. E **77**, 017102-1–017102-4 (2008)
32. Zhuge, H.: Communities and emerging semantics in semantic link network: discovery and learning. IEEE Trans. Knowl. Data Eng. **21**(6), 785–799 (2009)