Physica A 388 (2009) 2561–2570

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

English and Chinese languages as weighted complex networks

Long Sheng, Chunguang Li *

Centre for Nonlinear and Complex Systems, School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China

A R T I C L E I N F O

Article history:
Received 12 December 2008
Received in revised form 23 February 2009
Available online 12 March 2009

PACS:
89.75.Hc
89.75.Fb

Keywords:
Weighted complex networks
Written human language
Zipf's law
Power–law distribution

A B S T R A C T

In this paper, we analyze statistical properties of English and Chinese written human language within the framework of weighted complex networks. The two language networks are based on an English novel and a Chinese biography, respectively, and both of the networks are constructed in the same way. By comparing the intensity and density of connections between the two networks, we find that high weight connections in Chinese language networks prevail more than those in English language networks. Furthermore, some of the topological and weighted quantities are compared. The results display some differences in the structural organizations between the two language networks. These observations indicate that the two languages may have different linguistic mechanisms and different combinatorial natures.

1. Introduction

In the past decade, complex network science has received a lot of interest [1–5] since the seminal works of Watts and Strogatz [6] as well as Barabási and Albert [7]. A lot of real world systems have been examined from the viewpoint of complex networks — for example, the Internet and WWW, brain networks, protein-protein interaction networks, disease transmitted networks, and social interactions. These empirical studies establish that the complex network is a powerful tool in the analysis of complex systems by providing useful representations for system elements and their interactions.

The written human language is one of the most important examples of complex systems in nature. Words are the simple elements that combine to form complex structures of this system. If we consider each word as a vertex and their interactions as links between them, then the written human language can be modeled by complex networks. We are able to gather much information about the construction of such a language network: we can know the sequence in which words are linked into a text and we can explore the linguistic rule under which a literature is composed. Moreover, we can analyze similarities and differences about syntax between different languages. Recently, many important properties about language networks have been reported [8–15], such as Zipf's law properties, power law properties, small world properties and scale free properties. But they were all restricted to unweighted networks and did not expand to weighted networks.

Networks are specified not only by their topology but also by the dynamics of information or traffic flow taking place on their structures. In particular, the heterogeneity in the intensity of connections may be very important in the understanding of many network systems. Recently, the intensity of connections has been taken into account in many real network systems. It is found that most of them display the heterogeneous phenomenon of their connections. Examples are the existence of strong and weak ties between individuals in social networks [16–19], the diversity of the predator–prey interactions in food webs [20–22], different capabilities of transmitting electric signals in neural networks [18,23,24], unequal traffic on the

* Corresponding author.
E-mail address: cgli@uestc.edu.cn (C. Li).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2009.02.043

Table 1
Some parameters of the networks. $\langle k\rangle$, $\langle w\rangle$ and $\langle s\rangle$ mean the average degree over all vertices, average weight over all edges, and average strength over all vertices respectively.

	ELN	CLN
Text length	122 142	569 708
Num. of vertices	8811	3294
Num. of edges	51 338	104 023
$\langle k\rangle$	5.8	31.6
$\langle w\rangle$	2.38	5.48
$\langle s\rangle$	13.86	172.95

Internet [25], and the passengers in airline networks [26,27]. These systems can be better described in terms of weighted networks, in which each connection carries a numerical value measuring the strength of the connection. The link strength in a syntactic network may also contain important information about the network structures, so there is a need for research into language networks that goes beyond the purely topological notion.

Motivated by these observations, we perform in this paper the statistical analysis of two weighted networks, which are based on an English novel and a Chinese biography, respectively. Working with the comparison results of the two syntactic networks' properties, we try to find some similarities or differences of organization mechanism between English and Chinese written language.

The rest of this paper is organized as follows. In Section 2 we introduce the construction of the two language networks which are based on two pieces of literature written in English and Chinese, respectively. The structure analysis and results are presented in Section 3. Section 4 is our conclusions and discussions.

2. Construction of language networks

English is the most widely spoken language in the world, while the Chinese language has the largest number of speakers in the world. We construct written language networks based on a novel of English and a biography of Chinese respectively. We consider George Orwell's 1984 [28] as our English language network (ELN). Masucci and Rodgers [10] have studied the topology of this novel by constructing an unweighted network, and found that the second order vertex correlations are an essential component of the network architecture. In order to get more information about the structural organization of the language network, we extend their work to a weighted language network. We use the biography of Mao Zedong (Mao Tse-Tung) [29] as our Chinese language network (CLN). We select this book partially because Mao Zedong is well-known by many people in the world. In our research, both the pieces of literature are treated as directed and weighted graphs.

Words are the basic elements of sentences, the most correlated words in a sentence are usually the closest, and written human language is one of the natural networks closest to the absolute reciprocity [10]. Thus we define the networks as follows: Every different English word/Chinese character in the text is defined as different vertex, an edge exists if two vertices are neighbors, the direction of the edge is pointed from the antecedent to the consequent, and the weight of an edge denotes the frequency of the connection between the two English words/Chinese characters appearing in the text. We consider punctuations as functional words in both the texts, because they have parasitological functions despite their meaningless. Following this construction method, we get two weighted networks, which are denoted by ELN and CLN as above. Given the way in which the networks have been constructed, the degree of a vertex equates the number of different consequent neighbors that a English word/Chinese character has, the weight of an edge denotes the frequency of a connection appearing in the text, and the strength of a vertex (the sum of weight values of all links pointed out from it) equates the frequency of a English word/Chinese character appearing in the text. Without loss of generality, in our study, we consider the out-degree as the degree of a vertex and the out-strength as the strength of a vertex. The in-degree and in-strength condition exhibit the same result.

The length of the English text is 122 142 with 8811 different words, while the length of the Chinese text is 569 708 with only 3294 different characters. The average weight of edges in CLN is almost twice of the one in ELN (see Table 1). Naturally, we would have questions as to why could Chinese language construct a longer text by fewer characters than English? Are the connections of Chinese characters playing a more important role in the Chinese text than that of English words in the English text? By the structure analysis of the weighted language networks of ELN and CLN, we try to find some clues.

3. Network structure analysis

Although ELN and CLN are generated in the same way, there are some different statistical properties of their organizations (see Table 1 and Fig. 1). These differences may imply different structural organizations of the two languages, and we will gather more information under the weighted network analysis.

3.1. English words/Chinese characters distribution

Zipf's law is one of the most important statistical regularities of language networks, which states that the frequency of words decay as a power function of its rank [30]. Researchers show that many languages, such as English, French, Spanish,

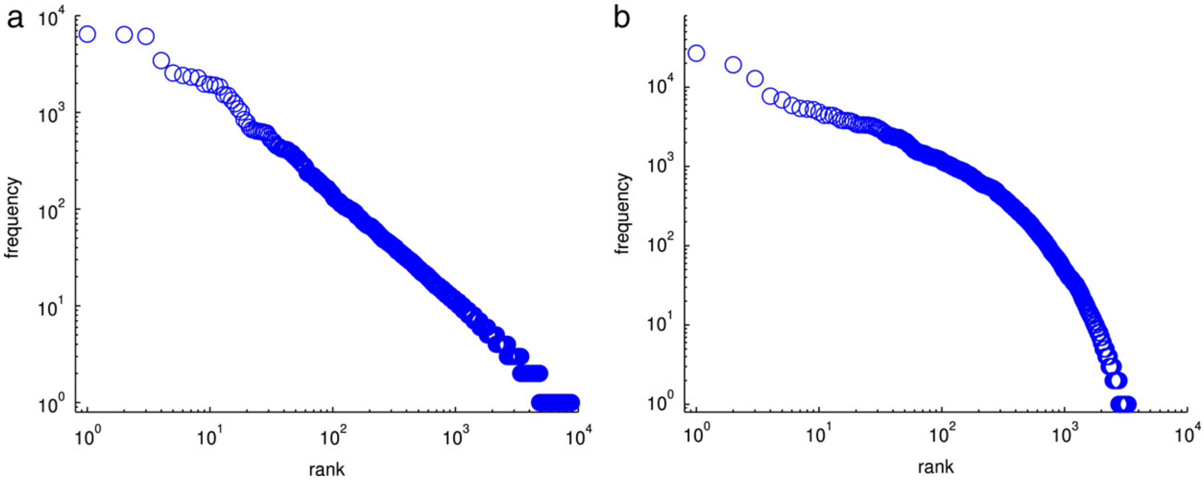

Fig. 1. Measures of Zipf's law. Figure (a) and (b) correspond to the English novel 1984 with a slope of −1.2 and the Chinese biography of Mao Zedong (Mao Tse-Tung) respectively.

have been found to fit Zipf's law [8,30–32]. In our research, the English novel fits Zipf's law with a slope of −1.2, but the Chinese biography doesn't fit the Zipf's law well (see Fig. 1). This result is consistent with other scholars' [12]. In Ref. [12], the authors have found that the words distribution of Shakespeare's corpus conforms to Zipf's law, while the character distribution of the Collection of Chairman Mao departs from Zipf's law. It should be noted that although both Refs. [12,29] are about Mao, the contexts of are quite different.

It is known that Zipf's law is a consequence of the match between structure and dynamics [33]. It does not depend on the syntactic structure of language [32], while it might be due to the growth and preferential selection mechanism of words or characters in a given language [12]. In fact, the difference in Fig. 1 maybe includes the different evolution history of English words and Chinese Characters. After Emperor Qin Shihuang unified the characters, the Chinese language became mature and the number of Chinese characters has grown very slowly. On the other hand, in English, new words are introduced constantly and the number of words grows very fast compared with Chinese characters [12].

Fig. 2 shows typical plots of degree distributions and strength distributions for ELN and CLN — they all follow a power-law distribution. Power-law distribution of vertex degree in Fig. 2(a) and (b) means that the majority of vertices in both the networks have only a few connections to other vertices, whereas some vertices are connected to many other vertices. This behavior of network is called scale-free. ELN and CLN are both scale free networks.

In fact, such scale-free behavior is related to frequency of occurrence of a word [10]. Following the construction of ELN and CLN, strength s corresponds to the frequency of occurrence of a English word/Chinese character. Fig. 2(c) and (d) plot the vertex strength distribution for ELN and CLN respectively, and they also fit the power-law distribution. It is the case that the majority of vertices in the two networks have low strengths, whereas some vertices have high strengths.

The degree and strength distributions of Chinese characters are both flatter than that of English words (see Fig. 2), while the two written language human networks both have scale-free property in text organization. The power-law distribution phenomenon of Fig. 2 can be interpreted as the evidence that the most English words/Chinese characters in the text have a low frequency of occurrence and a small number of different consequent English words/Chinese characters.

Further, in Fig. 3, we report the relationship between the vertices strength and the vertices degree. It is found that the strength of a vertex is almost proportional to its degree in both the two networks, which is also observed in other real-life systems [34]. This correlation demonstrates the phenomenon of "the rich get richer". Corresponding to the two texts, the higher frequency of occurrence an English word/Chinese character has, the more consequent neighbors it has. These English words/Chinese characters with high degrees are usually called hubs. These hubs correspond to some functional words, such as conjunction words, or adverbial words related with time and location. They are important for both the English and Chinese languages, because people can communicate conveniently by connecting hubs with some meaningful words according to syntax rules. In both the texts, these hubs usually have high degree and high strength simultaneously.

3.2. Words interactions

Words interactions play an important role in information expression. Under the way of the construction of ELN and CLN, the weight of edges represents the intension of these interactions. Fig. 4 shows distributions of the weight of edges for ELN and CLN, and they both fit power-law distribution. This means that the majority of edges in both the networks have a small weight, whereas some edges are high weighted. The edges with high weight correspond to some special connections between English words/Chinese characters. Some of these connections are meaningless, and some of these connections can form two-word phrases or parts of multi-word phrases. We define these connections which can form phrases are fixed connections.

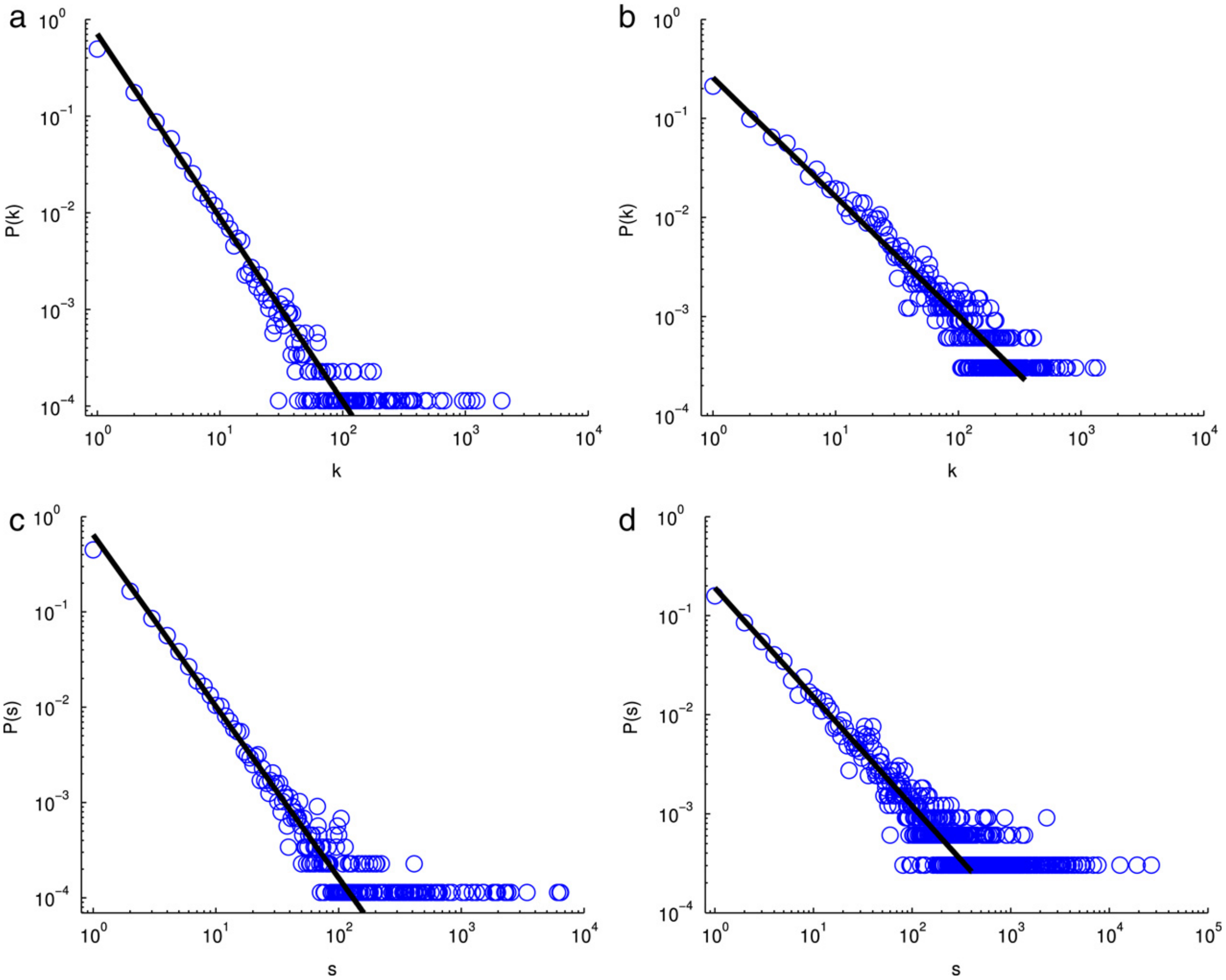

Fig. 2. Power-law distributions of vertices, where k is degree of each vertex, s is strength of each vertex. (a) Degree distribution $P(k)$ of ELN with a slope of -1.9. (b) Degree distribution $P(k)$ of CLN with a slope of -1.2. (c) Strength distribution $P(s)$ of ELN with a slope of -1.8. (d) Strength distribution $P(s)$ of CLN with a slope of -1.1.

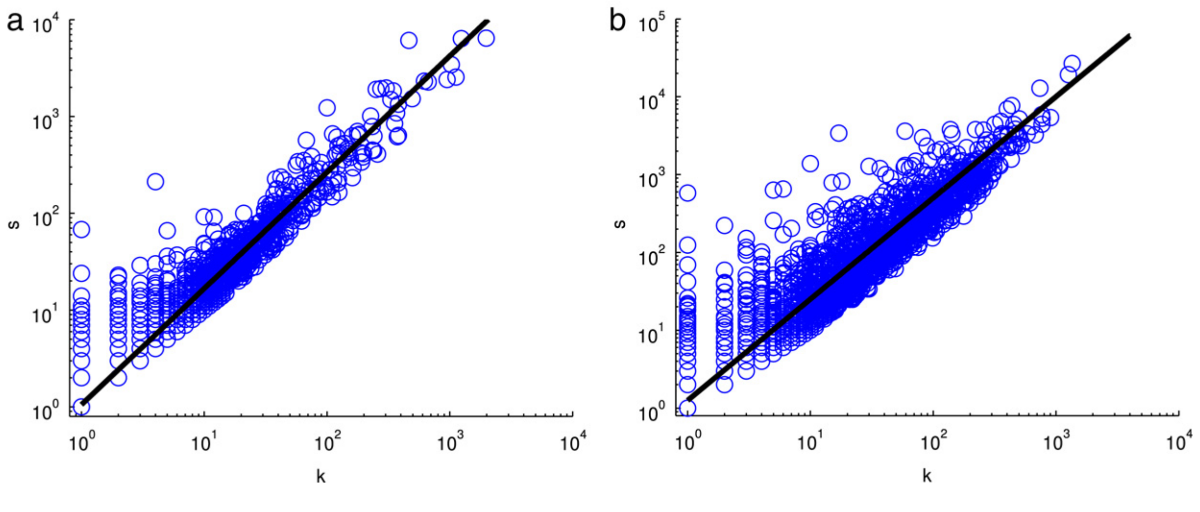

Fig. 3. Vertices strength s as a function of their degree k. (a) For ELN with a slop of 1.2. (b) For CLN with a slope of 1.3.

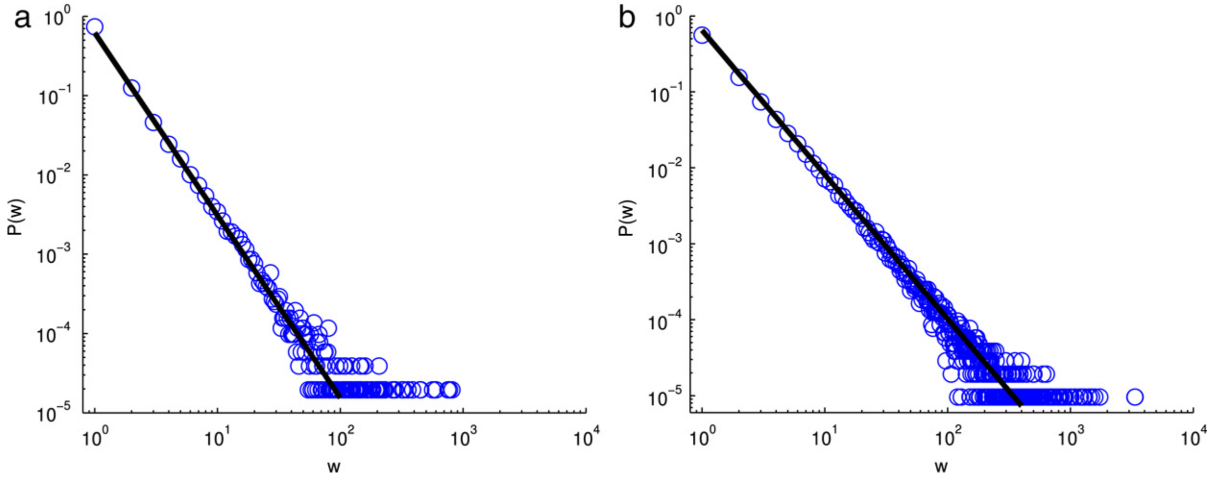

Fig. 4. Edge weight distributions. (a) For ELN with a slop of −2.3. (b) For CLN with a slop of −1.9.

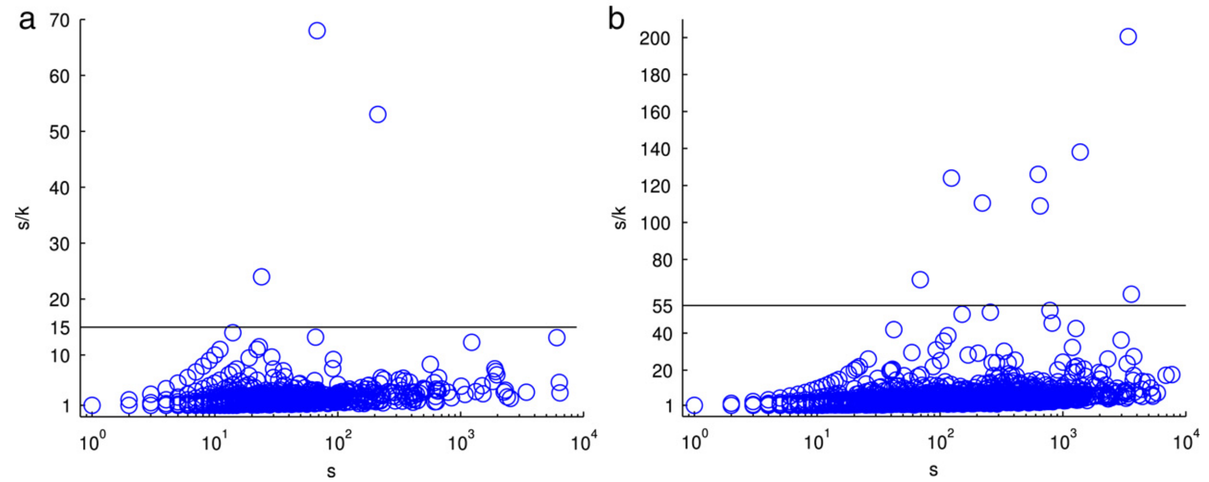

Fig. 5. s/k vs. vertex strength. (a) There are only 4 vertices have a value of s/k above 15 in ELN. (b) There are 8 vertices have a value of s/k above 55 in CLN.

In a text, if words form phrases, there should be fixed connections between them. In other words, words arranging fixed in pairs or queues in a text may form phrases. Under the way of the construction of ELN and CLN, the word strengths increase with the increasing of the word appearance frequency, and their degree increase with the increasing of the number of their different consequent neighbors. If the antecedent word has a high appearance frequency and a small number of different consequent neighbors in the text, there would be a fixed connection between the two words. It is that if the antecedent word have a high strength and a low degree, that the word and its consequents are most likely to form phrases. So if a vertex has a high value of s/k, the vertex and its consequent vertex trend to form a two-word phrase or a part of multi-word phrase (see Figs. 5 and 6). This is unrelated with the text length.

In Fig. 5, we plot the value of s/k of vertices vs. vertex strength. The figures show that the vertices with highest strength don't have the highest value of s/k in both the networks, because these vertices may also have high degrees. This is consistent with the phrase forming conditions. Specially, in both the texts, the name of the protagonists is one of the most appearance phrases (see Fig. 6).

Comparing Fig. 5(a) with (b), the number of vertices in CLN, which have high values of s/k, is more than that in ELN. The proportion of vertices, whose s/k values are above 10, is 0.2% in ELN and is 4.71% in CLN. So in the Chinese text, there should be more phrases than in the English text. There are many characters in Chinese have multiple-meaning, and forming phrases may help to express the correct meaning of the word and make it easy to comprehend.

3.3. Architecture analysis

Complex networks display an architecture imposed by the structural organization of many real systems, that is not fully characterized by the distribution $P(k)$ and $P(s)$. Indeed, the structural organization of many complex networks is

ELN words	Word strength	Word degree	s/k value	Two/multi-word phase (phase frequency)
O	212	4	53	**O '** Brien (207)
sort	68	1	68	a **sort of** (68)
don	68	1	68	**don '** t (68)
didn	24	1	24	**didn '** t (24)
CLN characters	**Character frequency**	**Character degree**	**s/k value**	**Two/multi-word phase (phase frequency)**
毛	3608	59	61.2	毛泽东 (3353)
泽	3409	17	200.53	毛泽东 (3361)
第	1381	10	138.1	第一 (497); 第二 (403)... 第十 (68)
介	653	6	108.83	蒋介石 (595)
阶	630	5	126	阶级 (441); 阶段 (175)
什	221	2	110.5	什么 (207)
巩	124	1	124	巩固 (124)
努	69	1	69	努力 (69)

Fig. 6. English words/Chinese characters whose s/k value is above 15 (55) in ELN/CLN.

mathematically encoded in the various correlations existing among the properties of different vertices. For this reason, a set of topological and weighted quantities of ELN and CLN will be studied in order to uncover the networks' architecture.

A first and widely used quantity of complex network is given by the clustering of vertices. The clustering of a vertex i is defined as

$$c_i = \frac{1}{k_i(k_i-1)} \sum_{j,h \in N} a_{ij} a_{ih} a_{jh}, \tag{1}$$

where k_i is the degree of vertex i, a_{ij} is 1 if there is a connection between vertex i and vertex j, otherwise a_{ij} is 0, N is the total number of vertices in a network. c_i measures the local cohesiveness of the network in the neighborhood of the vertex i. More information can be gathered through the inspection of the average clustering coefficient $C(k)$ restricted to classes of vertices with degree k

$$C(k) = \frac{1}{NP(k)} \sum_{i/k_i=k} c_i. \tag{2}$$

In many networks, the degree-dependent clustering coefficient $C(k)$ is a decreasing function of k, which shows that low-degree vertices generically belong to well interconnected groups, while high-degree sites are linked to many vertices that may belong to different groups which are not directly connected [35,36].

The clustering coefficient as defined in Eq. (1) does not take into account the fact that, in a weighted network, some neighbors are more important than the others. In Ref. [26], Barrat et al. have defined the weighted clustering coefficient of a given vertex i as

$$c_i^w = \frac{1}{s_i(k_i-1)} \sum_{j,h \in N} \frac{(w_{ij}+w_{ih})}{2} a_{ij} a_{ih} a_{jh}, \tag{3}$$

where s_i is the strength of vertex i, w_{ij} is the weight of the edge between vertex i and vertex j. The definition in Eq. (3) reduces to the topological vertex clustering coefficient of Eq. (1) when all the weights are equal. In the general cases, the weighted clustering coefficient considers both the number of closed triangles in the neighborhood of vertex i and their total relative weight with respect to the vertex strength.

The average over vertices of a given degree k defines the quantity $C^w(k)$. The comparison between $C(k)$ and $C^w(k)$ provides global information on the correlation between weights and topology. In the case of a large randomized network (lack of correlations) it is to see that $C^w(k) = C(k)$. In real weighted networks, however, we can face two opposite cases. If $C^w(k) > C(k)$, it indicates that edges with large weights have a tendency to form triples, while the opposite case $C^w(k) < C(k)$ indicates a lower relevance of the triangles [37].

Our measures of $C(k)$, $C^w(k)$ and their comparisons for ELN and CLN are shown in Fig. 7. There are two differences between ELN and CLN. In Fig. 7(a), $C(k)$ and $C^w(k)$ of ELN both decrease with the increasing of degrees. In Fig. 7(b), $C(k)$ and $C^w(k)$ of CLN have a fluctuation around 0.1 for low values of degree k under 200 approximately. Which means the majority of low degree vertices display a stronger hierarchical behavior in ELN than that in CLN. The other difference between ELN and CLN in Fig. 7 is the comparison results of $C(k)$ and $C^w(k)$. $C^w(k) > C(k)$ persists in almost the whole spectrum of the vertex degree in ELN, but in CLN, $C^w(k) < C(k)$ appears under the value of the vertex degree of 30 approximately. This

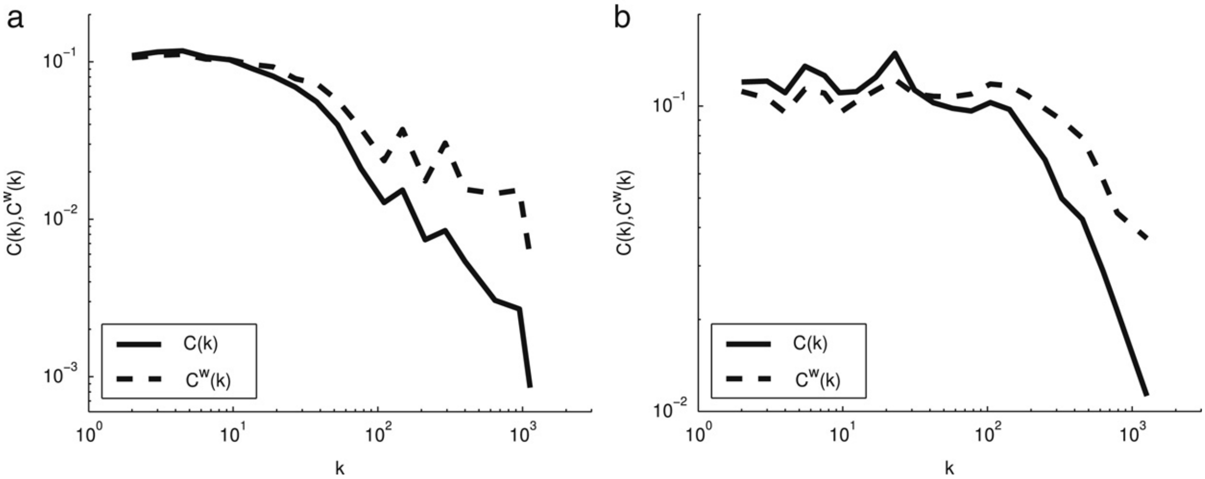

Fig. 7. The comparisons between average clustering coefficients $C(k)$ and weighted average clustering coefficients $C^w(k)$. (a) The comparison result of ELN. (b) The comparison result of CLN.

means that the clustering of low degree vertices has a minor effect in the organization of the CLN than that in ELN, because the largest part of the interactions is occurring on edges not belonging to interconnected triplets in CLN.

Comparisons between $C(k)$ and $C^w(k)$ show a structural difference of ELN and CLN, and this might be an explanation of why there are more phrases in the Chinese text than in the English text. In CLN, the highest weight edges between the vertices with low degree does not belong to interconnected triplets, which means these vertices may have large values of s/k. So there are more phrases formed by characters with few consequents in the Chinese text, than in the English text.

Another important source of information lies in the correlations of the degree of neighboring vertices [38,39]. Since the whole conditional distribution $P(k'|k)$, a given site with degree k connecting to another site of degree k', is often difficult to interpret, the average nearest-neighbor degree has been proposed to measure these correlations [38]

$$k_{nn,i} = \frac{1}{k}\sum_{i,j=1}^{N} a_{ij}k_j. \tag{4}$$

Once averaged over classes of vertices with degree k, the average nearest-neighbor degree can be expressed as

$$k_{nn}(k) = \sum_{k'} k'P(k'|k). \tag{5}$$

The average nearest-neighbor degree provides a probe on the degree correlation function. If degrees of neighboring vertices are uncorrelated, $k_{nn}(k)$ is independent of k, and $P(k'|k)$ is only a function of k'. When $k_{nn}(k)$ increases with the increasing of k, indicates that large degree vertices are preferentially connected with other large degree vertices. In this case the network is defined as assortative. In contrast, the network is said to be disassortative if $k_{nn}(k)$ decreases with the increasing of k [40].

Analogously, the weighted average nearest-neighbor degree can be defined as [26]

$$k^w_{nn}(k) = \frac{1}{s_i}\sum_{i,j=1}^{N} w_{ij}k_j. \tag{6}$$

Such a quantity allows to characterize the weighted assortative/disassortative properties considering the actual interactions among the system's elements, because the behavior of $k^w_{nn}(k)$ measures the effective affinity to connect with high or low degree neighbors according to the magnitude of the actual interactions. Indeed, $k^w_{nn}(k) > k_{nn}(k)$ when the edges with the larger weights are pointing to the neighbors with larger degree, and $k^w_{nn}(k) < k_{nn}(k)$ in the opposite case [37].

In Fig. 8, both ELN and CLN exhibit disassortative according to the decreasing curves of $k_{nn}(k)$, which implies the existence of a nontrivial correlation property for both ELN and CLN.

In order to confirm this disassortative phenomenon, we also compute the amount of assortative mixing r according to Ref. [40] as

$$r = \frac{M^{-1}\sum_i j_ik_i - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2}, \tag{7}$$

where M denotes the number of edges in the network, j_i, k_i are degrees of the vertices at the ends of the edge i, with $i = 1, 2, \ldots, M$. The coefficient is in the range $[-1, 1]$. If the network is uncorrelated, the correlation coefficient equals 0. Disassortative networks have $r < 0$, while assortative graphs have a value of $r > 0$.

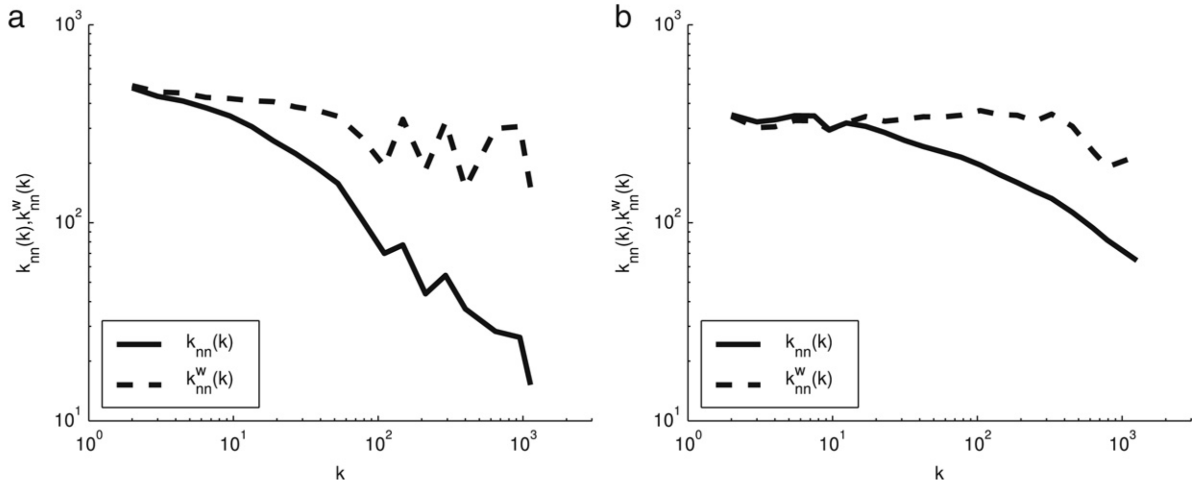

Fig. 8. The comparison between average nearest-neighbor degree $k_{nn}(k)$ and weighted average nearest-neighbor degree $k^w_{nn}(k)$. (a) The comparison result of ELN. (b) The comparison result of CLN.

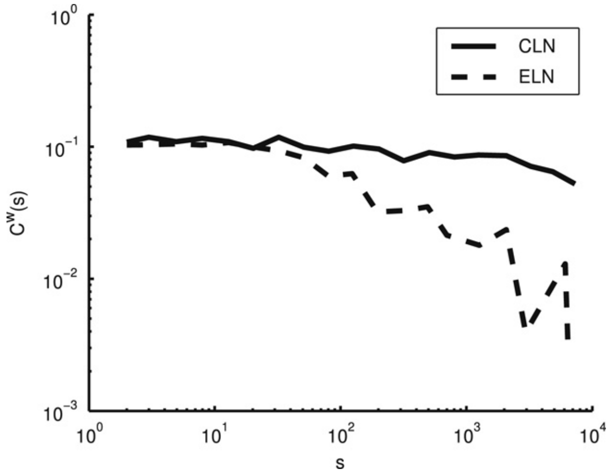

Fig. 9. The comparison of weighted average clustering coefficient C^w between ELN and CLN restricted to classes of vertices strength s.

The calculations of r are −0.23786 and −0.21409 for ELN and CLN respectively. So it again indicates that both ELN and CLN are disassortative mixing in terms of degree. This disassortative phenomenon in both networks can be intuitively explained in that high degree vertices, which are key words of a theme or which have a function of syntax, are not often related together in both the English and Chinese texts.

The comparisons of $k_{nn}(k)$ and $k^w_{nn}(k)$ in Fig. 8 also uncover a difference between ELN and CLN. In Fig. 8(a), $k^w_{nn}(k) > k_{nn}(k)$ persist in the whole spectrum of k. In Fig. 8(b), $k^w_{nn}(k) < k_{nn}(k)$ appears under the degree value of 10 approximately. This implies that edges with large weights appear between both low degree and high degree vertices in CLN, and these edges only appear between large degree vertices in ELN. Large weight edges between vertices means strong interactions between the English words/Chinese characters, so this may be another reason for the large number of phrases in the Chinese text.

We also take a comparison of the weighted average clustering coefficient $C^w(s)$ and the average nearest-neighbor degree $k^w_{nn}(s)$ between ELN and CLN in Figs. 9 and 10 respectively. In both the figures, the parameters of CLN are flatter than that of ELN. This fact reveals that the structure of ELN varies more with the changes of vertex strength than that of CLN. The effect can be interpreted as the evidence that the English text displays a larger heterogeneity in the intensity of the words interactions than the Chinese text. In Fig. 11, we plot weights of edges between first 200 English words/Chinese characters ranked by their strength. Comparing the figures in Fig. 11 will show the difference more clearly. In Fig. 11, the pixels represent the edges between vertices — the darker the pixel is, the higher weight the edge of the corresponding pair of vertices has. In Fig. 11(a), high weight edges are mostly connected between hubs or connected between hubs and low strength vertices, and they are seldom connected between low strength vertices. In Fig. 11(b), weights are more dispersive on the network structure. The hubs may play a more important role in the structural organization in the English text, while connections between Chinese characters have larger intensity and density.

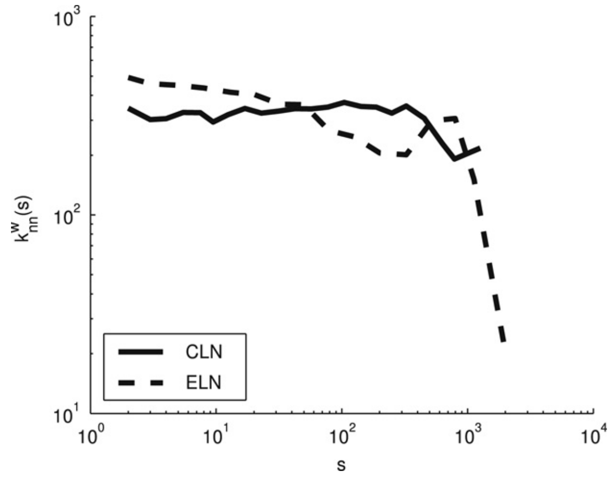

Fig. 10. The comparison of weighted average nearest-neighbor degree k_{nn}^{w} between ELN and CLN restricted to classes of vertices strength s.

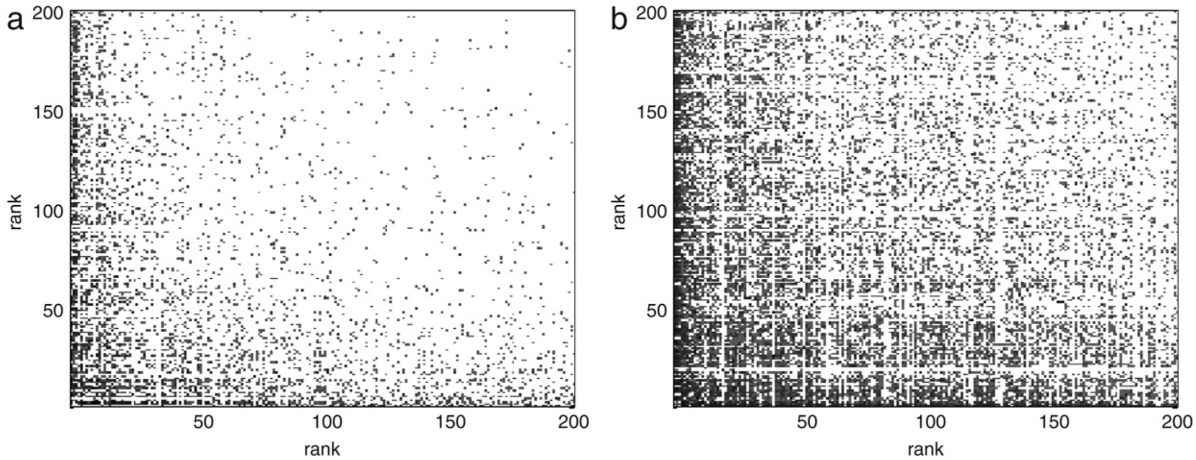

Fig. 11. Weights of edges between first 200 English words/Chinese characters ranked by their strength. The darker the pixel is, the higher weight the edge of the corresponding pair of vertices has. (a) The edge weights of ELN. (b) The edge weights of CLN.

4. Conclusions and discussions

In this work, we analyzed in detail the structure of two written human languages through weighted network representations of Orwell's 1984 and the Biography of Mao Zedong (Mao Tse-Tung). We found some similarities and differences between the two important languages.

Under the way of the construction of the two language networks, they displays tremendous differences. The length of the Chinese text is almost fivefold that of the length of the English one, while the number of different characters in the Chinese text is less than half of the number of different words in the English one.

Then, through the statistical analysis of the English words and the Chinese Characters' distributions $P(k)$ and $P(s)$, we found they both have scale-free properties in the text organization. While distributions $P(k)$ and $P(s)$ of the two networks have different exponents.

Next, the interconnection analysis between English words and between Chinese characters implies that there are more phrases in the Chinese text than in the English text. The large number of phrases may help to distinguish different meanings of Chinese characters, because many Chinese characters have more than one meaning.

At last, we performed architecture analysis of the two networks, finding that different structure behaviors are exhibited. The analysis of topological and weighted quantities implies that the English text displays a larger heterogeneity in the intensity of the words' interactions than the Chinese text. This structure difference might be a factor of the more phrases in the Chinese text than in the English text.

The above results together indicate that the hub-like words in the English text play a more important role of structural organization than in the Chinese text, while the connections between Chinese characters have larger intensity and density than between English words, which means phrases prevail more in Chinese language than in English. This might be a reason why Chinese language could construct a longer text by fewer characters than English.

Acknowledgments

The authors would like to thank the CNCS members for their valuable comments on an early version of this paper. This work was supported by the National Natural Science Foundation of China (NSFC 60871094), the Foundation for the Author of National Excellent Doctoral Dissertation of PR China, the MOE Key Project (Grant No. 107099), and the Youth Foundation of Sichuan Province (Grant No. 07ZQ026-019).

References

[1] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.
[2] S.N. Dorogvtsev, J.F.F. Mendes, Evolution of networks, Adv. Phys. 51 (2002) 1079–1187.
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. 424 (2006) 175–308.
[4] R. Albert, A.L. Barabási, Statistical mechanics of complex networks, Rev. Modern. Phys. 74 (2002) 47–97.
[5] L.D.F. Costa, F.A. Rodrigues, G. Travieso, Analyzing trails in complex networks, Phys. Rev. E 76 (2007) 046106.
[6] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small world' networks, Nature 393 (1998) 440–442.
[7] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[8] I.C.R. Ferrer, R.V. Sole, The small world of human language, Proc. R. Soc. B: Biol. Sci. 268 (2001) 2261–2265.
[9] S.N. Dorogvtsev, J.F.F. Mendes, Language as an evolving word web, Proc. R. Soc. B: Biol. Sci. 268 (2001) 2603–2606.
[10] A.P. Masucci, G.J. Rodgers, Network properties of written human language, Phys. Rev. E 74 (2006) 1539–3755.
[11] H. Liu, The complexity of Chinese syntactic dependency networks, Physica A 387 (2008) 3048–3058.
[12] D.H. Wang, M.H. Li, Z.R. Di, True reason for Zipf's law in language, Physica A 358 (2005) 545–550.
[13] S. Zhou, G.Hu. Zhang, Z. Guan, An empirical study of Chinese language networks, Physica A 387 (2008) 3039–3047.
[14] J.Li. Zhou, Chinese character structure analysis based on complex networks, Physica A 380 (2007) 629–638.
[15] D.R. Amancio, L. Antiqueira, T.A.S. Pardo, L.da.F. Costa, O.N. Oliveira Jr., M.G.V. Nunes, Complex networks analysis of manual and machine translations, Inteernat. J. Modern. Phys. C 19 (2008) 583–598.
[16] M.E.J. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. USA 98 (2001) 404–409.
[17] M. Marchiori, V. Latora, Harmony in the small-world, Physica A 285 (2000) 539–546.
[18] V. Latora, M. Marchiori, Efficient behavior of small-world networks, Phys. Rev. Lett. 87 (2001) 198701.
[19] M. Granovetter, The strength of weak ties, Amer. J. Sociol. 78 (1973) 1360–1380.
[20] G.A. Polis, Ecology: Stability is woven by complex webs, Nature 395 (1998) 744–745.
[21] E.L. Berlow, Strong effects of weak interactions in ecological communities, Nature 398 (1999) 330–334.
[22] A.E. Krause, K.A. Frank, D.M. Mason, R.E. Ulanowicz, W.W. Taylor, Compartment revealed in food-web structure, Nature 426 (2003) 282–285.
[23] V. Latora, M. Marchiori, Economic small-world behavior in weighted networks, Eur. Phys. J. B 32 (2003) 249–263.
[24] O. Sporns, Network analysis: Complexity and brain function, Complexity 8 (2003) 56–60.
[25] R. Pastor-Satorras, A. Vespignani, Evolution and Structure of the Internet: A Statistical Physics Approach, Cambridge Univ. Press, Cambridge, UK, 2004.
[26] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. USA 101 (2004) 3747–3752.
[27] W.Li. Cai, Statistical analysis of airport networks of China, Phys. Rev. E 69 (2004) 046106.
[28] G. Orwell, Nineteen Eighty Four, Middlesex, Penguin, 1978. http://www.online-literature.com/orwell/1984.
[29] C.J. Chong, Biography of Mao Zedong, Central Party Literature, Press, 2004, (in Chinese) http://www.zywxpress.com/Page/page1-1.asp?id=605.
[30] G.K. Zipf, Human behavior and the principle of least effort, Addison-Wesley, Cambridge, MA, 1949.
[31] I. Kanter, D.A. Kessler, Markov processes: Linguistics and Zipf's Law, Phys. Rev. Lett. 74 (1995) 4559–4562.
[32] W. Li, Random texts exhibit Zipf's-law-like word frequency distribution, IEEE Trans. Inform. Theory 38 (1992) 1842–1845.
[33] L. da F.Costa, O. Sporns, L. Antiqueira, M.G.V. Nunes, O.N. Oliveira Jr., Correlations between structure and random walk dynamics in directed complex networks, Appl. Phys. Lett. 91 (2007) 054107.
[34] M. Barthelemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Velocity and hierarchical spread of epidemic outbreaks in scale-free networks, Phys. Rev. Lett. 92 (2004) 178701.
[35] A. Vázquez, R. Pastor-Satorras, A. Vespignani, Large-scale topological and dynamical properties of the Internet, Phys. Rev. E 65 (2002) 066130.
[36] E. Ravasz, A.L. Barabási, Hierarchical organization in complex networks, Phys. Rev. E 67 (2003) 026112.
[37] A. Barrat, M. Barthélemy, A. Vespignani, Modeling the evolution of weighted networks, Phys. Rev. E 70 (2004) 066149.
[38] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the Internet, Phys. Rev. Lett. 87 (2001) 258701.
[39] S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks, Science 296 (2002) 910–913.
[40] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.