

Research Article

Complex Networks: Statistical Properties, Community Structure, and Evolution

Lei Zhang,¹ Jianxiang Cao,¹ and Jianyu Li²

¹*School of Computer Science, Communication University of China, Beijing 100024, China*

²*Engineering Center of Digital Audio and Video, Communication University of China, Beijing 100024, China*

Correspondence should be addressed to Lei Zhang; zhanglei2012@cuc.edu.cn

Received 26 October 2014; Accepted 31 December 2014

Academic Editor: Yang Tang

Copyright © 2015 Lei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We investigate the function for different networks based on complex network theory. In this paper, we choose five data sets from various areas to study. In the study of Chinese network, scale-free effect and hierarchical structure features are found in this complex system. These results indicate that the discovered features of Chinese character structure reflect the combination nature of Chinese characters. In addition, we study the community structure in Chinese character network. We can find that community structure is always considered as one of the most significant features in complex networks, and it plays an important role in the topology and function of the networks. Furthermore, we cut all the nodes in the different networks from low degree to high degree and then obtain many networks with different scale. According to the study, two interesting results have been obtained. First, the relationship between the node number of the maximum communities and the number of communities in the corresponding networks is studied and it is linear. Second, when the number of nodes in the maximum communities is increasing, the increasing tendency of the number of its edges slows down; we predict the complex networks have sparsity. The study effectively explains the characteristic and community structure evolution on different networks.

1. Introduction

In recent years, complex networks have had a profound effect on many discipline researches [1], such as systems science, statistical physics, social sciences, and biology [2–4]. Networks structure theory may help us to understand the properties, function, and evolution mechanism of complex networks [5]. As we know, Chinese is one of the most widely used languages in the world. Chinese characters play an important role in its well-known civilization. So Chinese character structure analysis based on radicals is a challenging, interesting, and very important problem. It will help to study the characteristics and the evolution of Chinese character and the universal rule of Chinese character combination principles.

In recent years' research, it is found that any language in the word including Chinese character by connecting basic units is based on complex grammar, syntax, and semantics [6]. Recently, with the rapid development of complex networks studies, Chinese complex networks are actively

studied [7–9]. For example, word cooccurrence networks [10] show two important characteristics: small-world effect and scale-free distribution. What is more, the study of Wordnet lexicon [11] demonstrates that Wordnet has global properties common to many self-organized systems, and polysemous links have a profound impact on the organization of the semantic graph which may be crucial for metaphoric thinking, imagery, and generalization. Again, network of free word associations [12] represents a proxy of the way in which our mind stores and organizes all words and related meanings.

Radicals of the Chinese character are treated as the basic units of Chinese. They are monosyllabic, square-shaped, and primitive, having some relationship to iconicity and combination [13]. They combine into Chinese characters based on certain rules.

In this paper, we mainly investigate the statistical properties and community structure of Chinese character network. In addition, we also research the evolution of community

structure in different networks. The paper is organized as follows. Firstly, we focus on the graph features of Chinese character network based on complex network theory. Secondly, we present an analysis aimed at community structure in Chinese character network. Thirdly, we study the community structure evolution of five different networks and the community structure can help us to understand the global structures [14]. Finally, we draw the conclusions in the last section.

2. Data and Network

In order to study the features for different networks, we need to choose data sets from various areas. Table 1 lists the data sets we have used.

(1) *MATLAB Help Document*. The nodes are key terms in MATLAB help document. If two nodes have a hyperlink relationship, they are connected.

(2) *Chinese Characters*. The nodes are radicals in Chinese characters, and the relation of two radicals exists if they can cooccur in a word.

(3) *Yeast*. The nodes are proteins in yeast, and the relation of two proteins exists if they have chemical reaction with each other [15].

(4) *Electronic Collaboration Networks*. The nodes are authors who research general relativity and quantum cosmology. The relation of two authors exists if they cooperate to complete a paper [16].

(5) *Peer-to-Peer Files Sharing Networks*. The nodes are hosts in peer-to-peer files sharing networks topology, and the relation of two hosts exists if they connect to each other [17].

In the following we attempt to uncover the features of different networks based on graph theory. Let us consider the undirected graph, $G = (W, E)$, where $W = \{w_i\}$ ($i = 1, 2, \dots, N_L$) is the set of nodes and $E = \{\{w_i, w_j\}\}$ is the set of connections. Here, $\zeta = \{w_i, w_j\}$ indicates that there is an edge between w_i and w_j .

3. Scale-Free and Hierarchical on Chinese Character Network

In this section, we will study the characteristics of the Chinese character network. First, we research the degree distribution $P(k)$. The degree of a given radical is the number of edges that connect the given radical with other radicals. Degree distribution is defined as the existence probability of nodes with degree. Degree distribution means the word-formation ability of radicals which is reflected in Chinese character network.

Figure 1 shows that degree distribution of Chinese character network follows power-law distribution $P(k) \sim k^{-r}$. The exponent r is 2.07. A network that exhibits power-law degree distribution is called a scale-free network. Scale-free network indicates that the majority of the nodes have a small amount

TABLE 1: Five various data sets and all nodes in each data set and the existing edges.

Source data	Node	Number of nodes	Number of edges
MATLAB help document	Key terms	2385	10091
Chinese characters	Radicals	1174	4000
Yeast	Proteins	2361	7182
Electronic collaboration networks	Authors	5242	28980
Peer-to-peer files sharing networks	Hosts	6301	20777

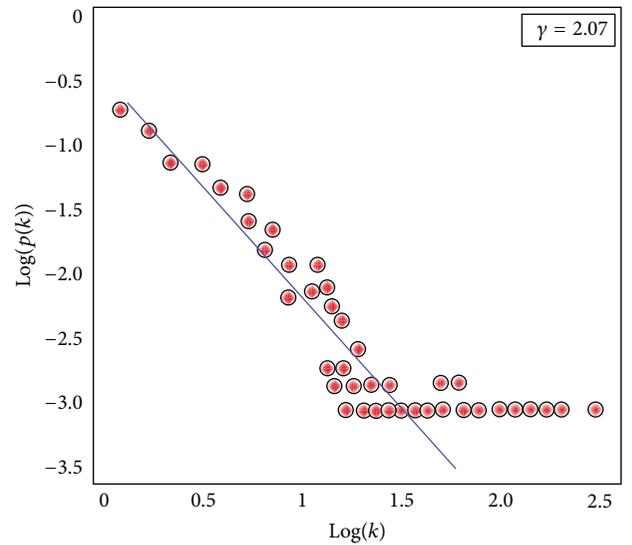


FIGURE 1: $\text{Log}(k)-\text{log}(p(k))$ degree distribution of Chinese character network.

of links, but a few nodes, called hubs, can link to most of the nodes in the network. For example, in Chinese character network, the characters “口” (mouth) and “木” (wood) are highly connected nodes because they are familiar to us and they can form a great deal of Chinese characters.

Another feature is clustering coefficient [18]. Clustering coefficient means the probability that two neighbors of a node are also neighbors to each other (nodes w_i and w_j are neighbors if there is a link between w_i and w_j). For a node with neighbors, the local clustering coefficient C_i is defined as the ratio between the number of links among the k_i neighbors and the maximum possible number of links among these neighbors. This can be expressed as follows:

$$C_i = \frac{2|S_i|}{k_i(k_i - 1)}, \quad (1)$$

where S_i is the number of existing links between the k_i neighbors.

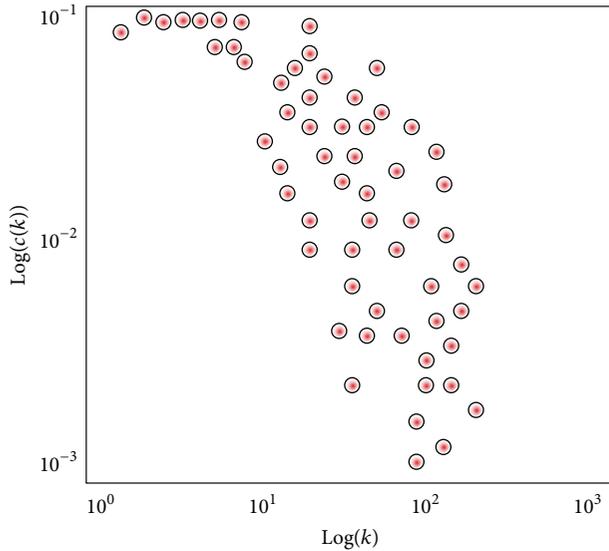


FIGURE 2: $\text{Log}(k)$ - $\text{log}(c(k))$ clustering distribution of Chinese character network.

Clustering spectrum $C(k)$ is defined as an average clustering coefficient of nodes with degree k . As Figure 2 shows, clustering coefficient decreases linearly with the degree. This implies that the small nodes are part of highly cohesive, densely interlinked clusters, while the hubs are not, as their neighbors have a small chance of linking to each other. The power-law clustering spectrum $C(k) \sim k^{-a}$ shows that the network has a hierarchical feature [19].

The hierarchical feature of Chinese character network is consistent with hierarchical network models in lexical networks [20]. As shown in Figure 3, the most common and important characters radicals, such as “木” (wood), “氵” (water), “目” (eye), “丨” (single side), and “辶” (cursive head), should be stored in higher level so that people can learn Chinese characters conveniently and efficiently.

4. Community Structure in Chinese Character Network

Chinese characters have many hundreds of thousands of words, most of which are created by combining just with a few thousand radicals. Although tens of thousands of Chinese characters have been created in history, the Chinese characters in common use are about thousands. According to the characters structure, there are about hundreds of traditional radicals, which are mainly used to index and look up characters in the dictionary.

In this section, we will study community structure of the Chinese character network. Community structure [21–26] refers to a high density of links between nodes of the same group and a comparatively low density of links between nodes of different groups. Community structure analysis has a wide range of application in biology, physics, computer graphics, and sociology [27, 28]. For example, in social groups, people with the same hobbies or beliefs always appeal to each other.

In molecular response network, we could distinguish roles or features of molecular from aggregated functional module nodes.

In order to detect community in Chinese character network, we calculate the degree of each node in the network, and then we cut all the nodes in the network from low degree to high degree and obtain many networks with different scale.

In this paper, we only analyze the nodes with high degree. As Figures 4, 5, and 6 show, the most common and important radicals, such as “钅” (metal), “木” (wood), “氵” (water), “火” (fire), and “土” (earth), should be stored in higher level so that people can learn Chinese characters conveniently and efficiently. Furthermore, we found an interesting phenomenon where “钅” (metal), “木” (wood), “氵” (water), “火” (fire), and “土” (earth) correspond to the Yin-Yang and the five elements. It shows the relationship between Chinese traditional culture and Chinese characters. And it also reveals that there is a natural internal relationship between Yin-Yang and the five elements of Chinese philosophy and Chinese characters [29]. The research of radicals of Chinese characters could help us to understand the Chinese culture better.

5. Evolution of Community Structure in Five Different Networks

In this section, we try to explore the relationship between the nodes number of the maximum communities and the number of communities in the corresponding networks to investigate the evolution characteristic of community structure, as shown in Figure 7.

Figure 7 displays the relationship, from which we can observe that the number of communities is gradually declining as the number of nodes is increasing. However, the trends of rate are diverse and we can find that different trends can reflect the structure of different complex networks. For example, the linear correlation of MATLAB help document is steep, which reveals they are connected tensely inside and connected sparsely outside. Meanwhile, the linear correlation of yeast is relatively smooth, which means they are connected sparsely inside and connected tensely outside. Depending on the linear correlation between the numbers of communities in whole network, we are able to study the characteristics of community structure of complex network better.

We explore the relationship between the nodes number of the maximum communities and the number of edges in the corresponding maximum communities. Figure 8 displays the relationship of five various data sets. From Figure 8, we can observe that the number of edges is increasing rapidly while the number of nodes is increasing at the beginning. However, when nodes increase to a certain extent, we can observe that the number of edges, in the corresponding communities, is increasing tend to smooth while the number of nodes is increasing. From Figure 8, we can also know that it is different in five data sets; some data sets change to steep, such as electronic collaboration networks, and other data sets change to smooth, such as peer-to-peer files sharing networks. When nodes increase to a certain extent, the number of edges is not increasing but achieves a smooth state. Based on the

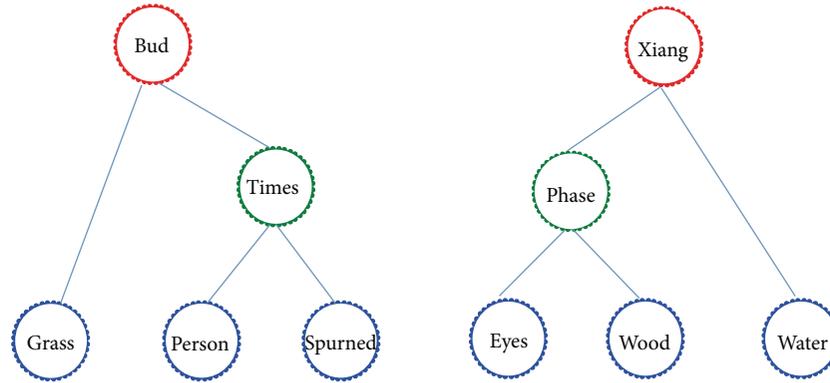


FIGURE 3: Hierarchical structure of Chinese character network. The meanings of the nodes are listed in Table 2.

TABLE 2: Meaning of Chinese characters which are nodes of Figure 3 is listed.

Chinese characters	亻	音	艹	倍	蓓
Chinese pronunciation	dān rén páng	pǒu	cǎo zītóu	bèi	bèi
English meaning	Person	Spurned	Grass	Times	Bud
Chinese characters	氵	木	目	相	湘
Chinese pronunciation	sān diǎn shuǐ	mùzǐpáng	mùzǐpáng	xiāng	xiāng
English meaning	Water	Wood	Eyes	Phase	Xiang

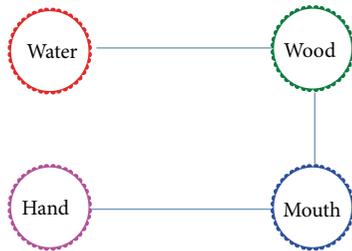


FIGURE 4: The degree of each node is more than 100 in the community structure. The meanings of the nodes are listed in Table 3.

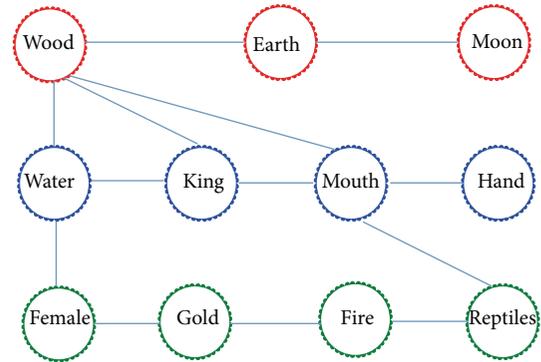


FIGURE 6: The degree of each node is more than 63 in the community structure. The meanings of the nodes are listed in Table 5.

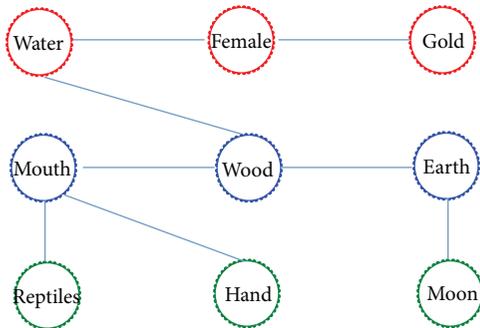


FIGURE 5: The degree of each node is more than 83 in the community structure. The meanings of the nodes are listed in Table 4.

above analysis, we could conclude that complex networks have sparsity feature.

6. Conclusion

In this paper, we investigate the function of five different networks based on complex network theory.

Firstly, we have presented the results of the analysis performed on Chinese character network. Chinese character network displays scale-free and hierarchical structure features, which are responsible for robustness. A group of Chinese characters tends to share the same radicals in order to communicate efficiently with the least effort. The most interesting phenomenon that antonymous nodes emerge in community structure should be further investigated.

Chinese characters have their special organizing principles; the radicals based on semantics show that they have intimate relations to nature. Chinese culture and our life

TABLE 3: Meaning of Chinese characters which are nodes of Figure 4 is listed.

Chinese characters	扌	口	氵	木
Chinese pronunciation	tíshǒu páng	kǒu zìpáng	sān diǎn shuǐ	mùzìpáng
English meaning	Hand	Mouth	Water	Wood

TABLE 4: Meaning of Chinese characters which are nodes of Figure 5 is listed.

Chinese characters	氵	女	金	口	木
Chinese pronunciation	sān diǎn shuǐ	nǚzìpáng	jīn zìpáng	kǒu zìpáng	mùzìpáng
English meaning	Water	Female	Gold	Mouth	Wood
Chinese characters	土	虫	扌	月	
Chinese pronunciation	títǔpáng	chóng zìpáng	tíshǒu páng	yuèzìpáng	
English meaning	Earth	Reptiles	Hand	Moon	

TABLE 5: Meaning of Chinese characters which are nodes of Figure 6 is listed.

Chinese characters	木	土	月	氵
Chinese pronunciation	mùzìpáng	títǔpáng	yuèzìpáng	sān diǎn shuǐ
English meaning	Wood	Earth	Moon	Water
Chinese characters	王	口	扌	女
Chinese pronunciation	wáng zìpáng	kǒu zìpáng	tíshǒu páng	nǚzìpáng
English meaning	King	Mouth	Hand	Female
Chinese characters	金	火	虫	
Chinese pronunciation	jīn zìpáng	huǒzìpáng	chóng zìpáng	
English meaning	Gold	Fire	Reptiles	

and the combination of Chinese radicals have hierarchical structure.

Secondly, the appearance of Chinese character radical not only enriched human language but also has important effect on word association. It is well known that information in our brain is associative and is retrieved by connecting similar concepts. Our experiment has been brought to the attention of community structure as a valuable tool to understand the basic cognitive mechanisms and information retrieval processes. The structure features of community structure may be related to increasing our memory retention and recall, which is probably necessary for the brain to store information and associate.

Thirdly, we study the evolution of community structure in five different networks. After fitting between the number of nodes in the maximum communities and the number of communities in the corresponding networks, we observed the linear correlation of them. We could extend the linear correlation to other kinds of networks and predict the scale of them. In addition, depend on the study of the relationship between the nodes number of the maximal communities and the number of edges in the corresponding maximal communities, which revealed the sparse feature of the community structure in network and the generality of complex networks.

Although the resulting Chinese character networks display statistical features different from random networks, this does not mean to rule out the random factors. As a complex adaptive system, there exist random factors for combination of Chinese characters. Even in scale-free networks, random

attachment still plays an important role and is a preferential attachment. This problem will be considered in the future.

There is no shared definition of community, which is justified by the nature of the problem itself. What is more, the network in the real world is always dynamic. Most researches on complex networks are focusing on excavating the hidden relations and features in real networks such as social network. Therefore, the improvement of implementing our research on ever-changing dynamic networks will be an innovative and challenging topic in our future work. In addition, more complex and more realistic model should be considered.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank Professor Li for helpful suggestions and comments. This research work was supported by the Engineering Planning Project for Communication University of China (3132014XNG1449 and 3132014XNG1433), the Comprehensive Reform Project of Computer Science and Technology (ZL140103); Guangzhou Research Institute of Communication University of China Common Construction Project, "Sunflower"—the Aging Intelligent Community; the Construction Project of Professional Postgraduate Degree Education Case Library in Communication University of

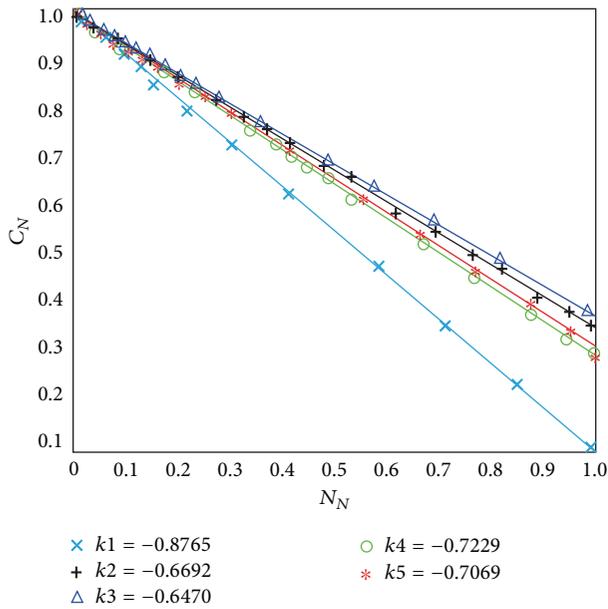


FIGURE 7: Description of the relationship between the nodes number of the maximum communities and the number of communities is linear, but the slopes of the lines from five data sets are different, where k_1 , k_2 , k_3 , k_4 , and k_5 represent the slopes of five linear fittings of MATLAB help document, Chinese characters, yeast, electronic collaboration networks, and peer-to-peer files sharing networks, respectively (N_N : the node number of the maximum communities and C_N : the number of communities in the corresponding networks).

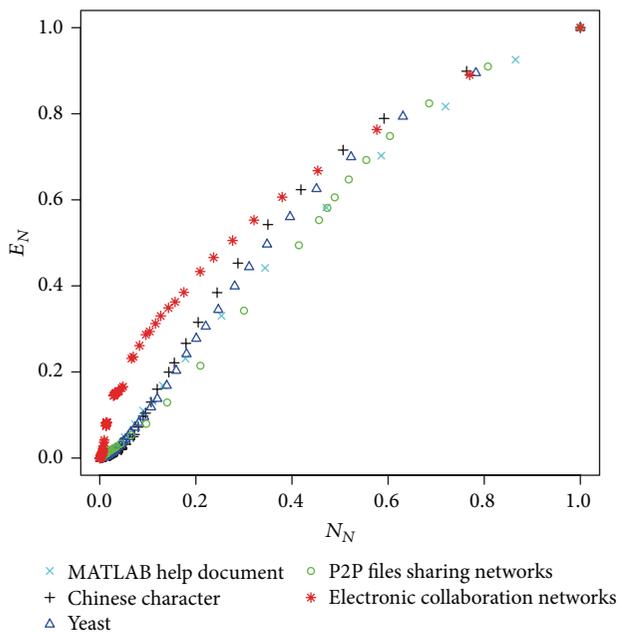


FIGURE 8: Visualized relationship of five different networks (N_N : the nodes number of the maximum communities and E_N : the number of edges in the corresponding maximum communities).

China (ZYKC201448); and the Discipline Construction Project School of Computer Science and Technology, Division of Science and Technology.

References

- [1] A. Vespignani, "Modelling dynamical process in complex socio-technical systems," *Nature Physics*, vol. 8, no. 1, pp. 32–39, 2012.
- [2] E. A. Leicht and M. E. J. Newman, "Community structure in directed networks," *Physical Review Letters*, vol. 100, no. 11, Article ID 118703, 2008.
- [3] C. Durugbo, W. Hutabarat, A. Tiwari, and J. R. Alcock, "Modelling collaboration using complex networks," *Information Sciences*, vol. 181, no. 15, pp. 3143–3161, 2011.
- [4] P. de Meo, A. Nocera, G. Terracina, and D. Ursino, "Recommendation of similar users, resources and social networks in a social interneting scenario," *Information Sciences*, vol. 181, no. 7, pp. 1285–1305, 2011.
- [5] D.-I. O. Hein, S. W. I. M. Schwind, and W. König, "Scale-free networks," *Wirtschaftsinformatik*, vol. 48, no. 4, pp. 267–275, 2006.
- [6] R. G. Gordon and B. F. Grimes, *Ethnologue: Languages of the World*, SIL International, Dallas, Tex, USA, 2005.
- [7] J. Ke, "Complex networks and human language," <http://arxiv.org/abs/cs/0701135>.
- [8] J. Li and J. Zhou, "Chinese character structure analysis based on complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 380, no. 1-2, pp. 629–638, 2007.
- [9] J. Li, J. Zhou, X. Luo, and Z. Yang, "Chinese lexical networks: the structure, function and formation," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 21, pp. 5254–5263, 2012.
- [10] R. F. I. Cancho and R. V. Solé, "The small world of human language," *Proceedings of the Royal Society B: Biological Sciences*, vol. 268, no. 1482, pp. 2261–2265, 2001.
- [11] M. Sigman and G. A. Cecchi, "Global organization of the Wordnet lexicon," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 3, pp. 1742–1747, 2002.
- [12] P. Gravino, V. D. P. Servedio, A. Barrat, and V. Loreto, "Complex structures and semantics in free word association," *Advances in Complex Systems*, vol. 15, no. 3-4, Article ID 1250054, 2012.
- [13] C. Lu, *Linguistics for Knowledge Engineering*, Tsinghua University Press, 2010.
- [14] K. Takemoto, C. Oosawa, and T. Akutsu, "Structure of n-clique networks embedded in a complex network," *Physica A: Statistical Mechanics and its Applications*, vol. 380, no. 1-2, pp. 665–672, 2007.
- [15] D. Bu, Y. Zhao, L. Cai et al., "Topological structure analysis of the protein-protein interaction network in budding yeast," *Nucleic Acids Research*, vol. 31, no. 9, pp. 2443–2450, 2003.
- [16] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: densification and shrinking diameters," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, article 2, 2007.
- [17] M. Ripeanu, I. Foster, and A. Lamnitchi, "Mapping the gnutella network: properties of large-scale peer-to-peer systems and implications for system design," *IEEE Internet Computing Journal*, vol. 6, no. 1, pp. 1–12, 2002.
- [18] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertész, "Generalizations of the clustering coefficient to weighted complex networks," *Physical Review E*, vol. 75, no. 2, 4 pages, 2007.

- [19] Y. Li, L. Wei, Y. Niu, and J. Yin, "Structural organization and scale-free properties in Chinese Phrase Networks," *Chinese Science Bulletin*, vol. 50, no. 13, pp. 1304–1308, 2005.
- [20] B. Y. Ge, *Modern Chinese Lexicology*, Shandong Prioivence Renmin Press, Jinan, China, 2001.
- [21] G. Caldarelli and A. Vespignani, Eds., *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science*, World Scientific, 2007.
- [22] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.
- [23] M. A. Porter, J.-P. Onnela, and P. J. Mucha, "Communities in networks," *Notices of the American Mathematical Society*, vol. 56, no. 9, pp. 1082–1097, 2009.
- [24] A. Arenas, A. Fernández, and S. Gómez, "Analysis of the structure of complex networks at different resolution levels," *New Journal of Physics*, vol. 10, Article ID 053039, 2008.
- [25] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 1, pp. 36–41, 2007.
- [26] S. Fortunato and C. Castellano, "Community structure in graphs," in *Computational Complexity*, pp. 490–512, Springer, New York, NY, USA, 2012.
- [27] J. Liu and G.-S. Deng, "A collaborative recommendation method based on user network community with weighted spectral analysis," *Journal of Dalian University of Technology*, vol. 50, no. 3, pp. 438–442, 2010.
- [28] R. Cun, D. Xiao, X. Liu, and L. I. Zhi-jie, "A physical community discovery algorithm," *Microelectronics and Computer*, vol. 27, no. 9, pp. 33–36, 2010.
- [29] N. Wu, X. Zhou, and H. Shu, "Sublexical processing in reading Chinese: a development study," *Language and Cognitive Processes*, vol. 14, no. 5-6, pp. 503–524, 1999.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

