PHYSICAL REVIEW E **76**, 1 (2007)

Empirical study on clique-degree distribution of networks

Wei-Ke Xiao,[1,2] Jie Ren,[3,2,4] Feng Qi,[2] Zhi-Wei Song,[2] Meng-Xiao Zhu,[2] Hong-Feng Yang,[2] Hui-Yu Jin,[2] Bing-Hong Wang,[4] and Tao Zhou[3,4,*]

[1]*Center for Astrophysics, University of Science and Technology of China, Hefei 230026, China*
[2]*Research Group of Complex Systems, University of Science and Technology of China, Hefei 230026, China*
[3]*Department of Physics, University of Fribourg, Chemin du Muse 3, CH-1700 Fribourg, Switzerland*
[4]*Department of Modern Physics and Nonlinear Science Center, University of Science and Technology of China, Hefei 230026, China*

(Received 21 January 2006; revised manuscript received 29 July 2007)

The community structure and motif-modular-network hierarchy are of great importance for understanding the relationship between structures and functions. We investigate the distribution of clique degrees, which are an extension of degree and can be used to measure the density of cliques in networks. Empirical studies indicate the extensive existence of power-law clique-degree distributions in various real networks, and the power-law exponent decreases with an increase of clique size.

DOI: XXXX PACS number(s): 89.75.Hc, 05.40.−a, 64.60.Ak, 84.35.+i

The discovery of small-world effects [1] and scale-free properties [2] triggered an upsurge in the study of the structures and functions of real-life networks [3–7]. Previous empirical studies have demonstrated that most real-life networks are small world [8]; that is to say, they have a very small average distance like completely random networks and a large clustering coefficient like regular networks. Another important characteristic in real-life networks is the power-law degree distribution—that is, $p(k)\propto k^{-\gamma}$, where k is the degree and $p(k)$ is the probability density function for the degree distribution. Recently, empirical studies reveal that many real-life networks, especially biological networks, are densely made up of some functional motifs [9–11]. The distributing pattern of these motifs can reflect the overall structural properties and thus can be used to classify networks [12]. In addition, the networks' functions are highly affected by these motifs [13]. A simple measure can be obtained by comparing the density of motifs between real networks and completely random ones [12]; however, this method is too rough and thus still under debate now [14,15]. In this paper, we investigate the distribution of *clique degrees*, which are an extension of degree and can be used to measure the density of cliques in networks.

The word *clique* in network science equals the term *complete subgraph* in graph theory [16]; that is to say, the m order clique (m-clique for short) means a fully connected network with m nodes and $m(m-1)/2$ edges. Define the m-clique degree of a node i as the number of different m-cliques containing i, denoted by $k_i^{(m)}$. Clearly, a 2-clique is an edge and $k_i^{(2)}$ equals the degree k_i; thus, the concept of clique degree can be considered as an extension of degree (see Fig. 1). We have calculated the clique degree from order 2 to 5 for some representative networks. Figures 2–8 show the clique-degree distributions of seven representative networks in logarithmic binning plots [17,18]; these are the Internet at the *autonomous systems* (AS) level [19], the Internet at the routers level [20], the metabolic network of *P.aeruginosa* [21], the World-Wide-Web [22], the collaboration network of mathematicians [23], the protein-protein interaction networks of yeast [24], and the BBS friendship networks at the University of Science and Technology of China (USTC) [25]. The slopes shown in those figures are obtained by using a maximum-likelihood estimation [26]. Table I summarizes the basic topological properties of those networks.

Although the backgrounds of those networks are completely different, they all display power-law clique-degree distributions. We have checked many examples (not shown here) and observed similar power-law clique-degree distributions. However, not all the networks can display higher-order power-law clique-degree distributions. Actually, only the relatively large networks could have a power-law clique-degree distribution with order higher than 2. For example, Ref. [21] reports 43 different metabolic networks, but most of them are very small ($N<1000$), in which the cliques with order higher than 3 are exiguous. Only the five networks with most nodes display relatively obvious power-law clique-degree distributions, and the case of *P.aeruginosa* is shown in Fig. 4. Note that, even for small-size networks, the high-order clique is abundant for some densely connected networks such as technological collaboration networks [27] and food webs [28]. However, since the average degree of the majority of metabolic networks is less than 10, the high-order cliques could not be expected with network size $N<1000$. Furthermore, all empirical data show that the power-law exponent will decrease with an increase of clique order. This may be a universal property and can reveal some un-

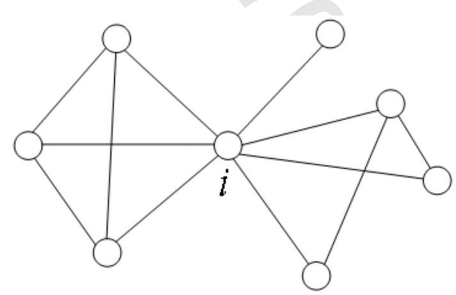
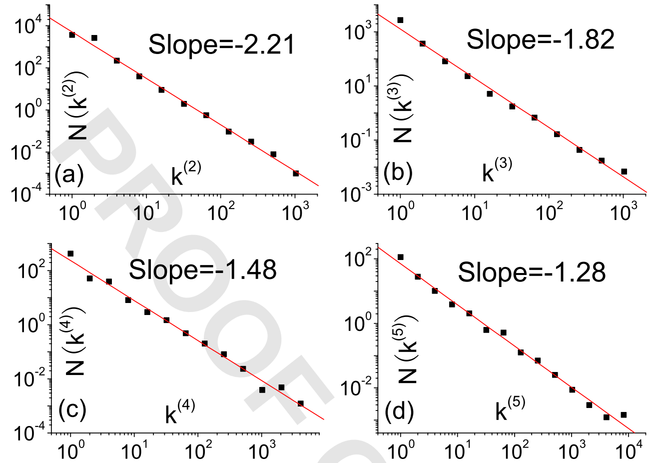


FIG. 1. Illustration of the clique degree of node i. $k_i^{(2)}=7$, $k_i^{(3)}=5$, $k_i^{(4)}=1$, and $k_i^{(5)}=0$.

*zhutou@ustc.edu

PROOF COPY [LN10539EJ] 055709PRE



AQ: #2

FIG. 2. (Color online) Clique-degree distributions of the Internet at the AS the level from order 2 to 5, where $k^{(m)}$ denotes the m-clique degree and $N(k^{(m)})$ is the number of nodes with m-clique degree $k^{(m)}$. In each panel, the marked slope of the red line is obtained by using maximum likelihood estimation [26].

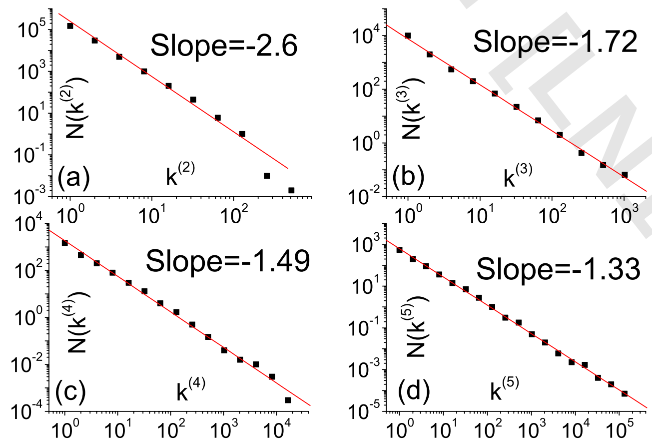


FIG. 3. (Color online) Clique-degree distributions of the Internet at the routers level.

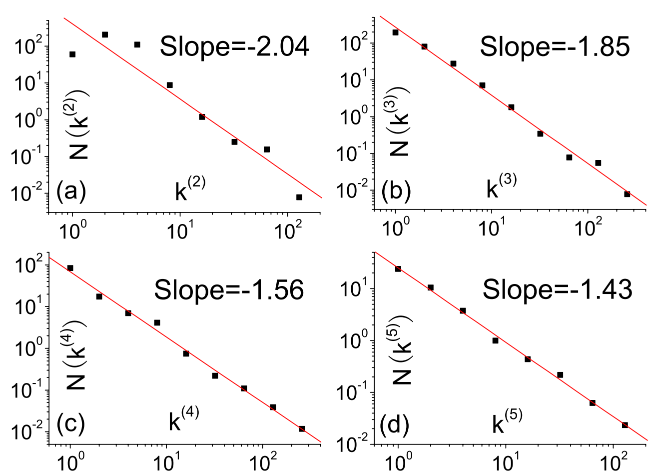


FIG. 4. (Color online) Clique-degree distributions of the metabolic network of *P.aeruginosa*.

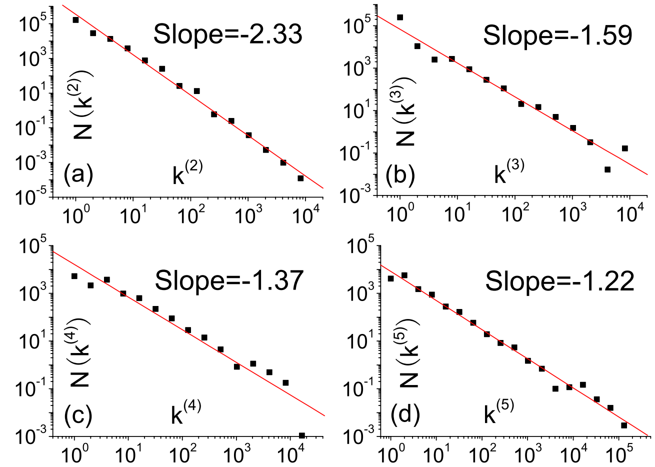


FIG. 5. (Color online) Clique-degree distributions of the World-Wide-Web.

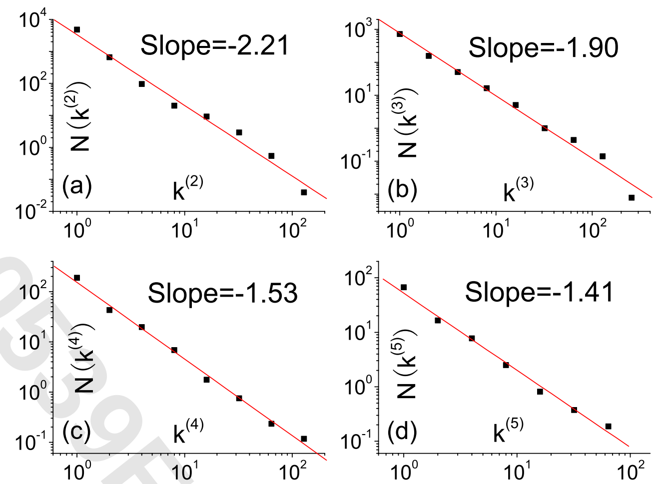


FIG. 6. (Color online) Clique-degree distributions of the collaboration network of mathematicians.

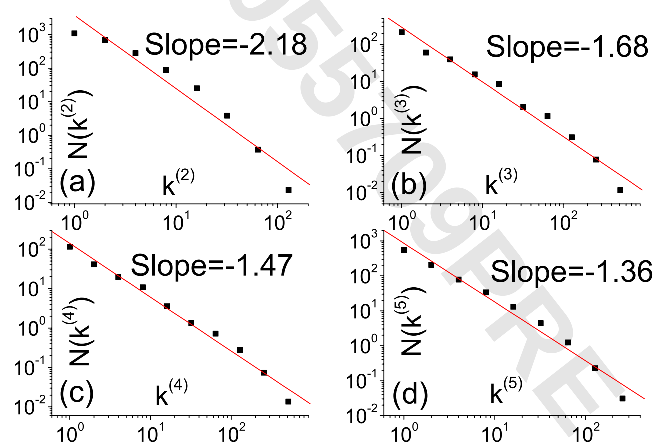


FIG. 7. (Color online) Clique-degree distributions of the protein-protein interaction networks of yeast.

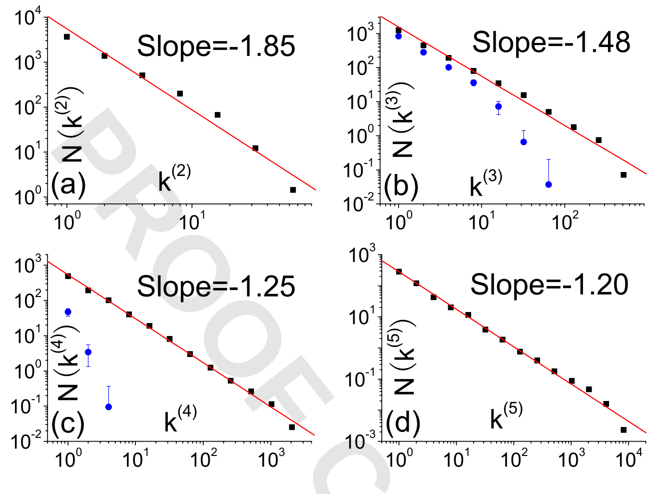


FIG. 8. (Color online) Clique-degree distributions of the BBS friendship networks at the University of Science and Technology of China. The blue points with error bars denote the case of a randomized network.

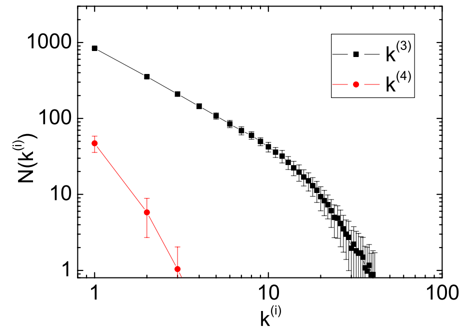


FIG. 9. (Color online) The clique-degree distributions in the randomized network corresponding to the BBS friendship network of USTC. The black squares and red circles represent the clique-degree distributions of order 3 and 4, respectively. All the data points and error bars are obtained from 100 independent realizations.

known underlying mechanism in network evolution.

In order to illuminate that the power-law clique-degree distributions with order higher than 2 could not be considered as a trivial inference of the scale-free property, we compare these distributions between original USTC BBS friendship network and the corresponding randomized network. Here the randomizing process is implemented by using the edge-crossing algorithm [12,29–31], which can keep the degree of each node unchanged. The procedure is as follows: (i) Randomly pick two existing edges $e_1=x_1x_2$ and $e_2=x_3x_4$, such that $x_1\neq x_2\neq x_3\neq x_4$ and there is no edge between x_1 and x_4 as well as x_2 and x_3. (ii) Interchange these two edges; that is, connect x_1 and x_4 as well as x_2 and x_3, and remove the edges e_1 and e_2. (iii) Repeat (i) and (ii) for $10M$ times.

We call the network after this operation the *randomized network*. In Fig. 9, we report the clique-degree distributions in the randomized network. Obviously, the 2-clique degree distribution (not shown) is the same as that in Fig. 8. One can find that the randomized network does not display power-law clique-degree distributions with higher order; in fact, it has very few 4-cliques and none 5-cliques. The direct comparison is shown in Fig. 8.

TABLE I. The basic topological properties of the present seven networks, where N, M, L, and C represent the total number of nodes, the total number of edges, the average distance, and the clustering coefficient, respectively.

Networks/Measures	N	M	L	C
Internet at AS level	10515	21455	3.66151	0.446078
Internet at routers level	228263	320149	9.51448	0.060435
Metabolic network	1006	2957	3.21926	0.216414
World-Wide-Web	325729	1090108	7.17307	0.466293
Collaboration network	6855	11295	4.87556	0.389773
ppi-yeast networks	4873	17186	4.14233	0.122989
Friendship networks	10692	48682	4.48138	0.178442

The discoveries of new topological properties of networks infuse the network science with ozone [1,2,7,9,32–34]. These empirical studies not only reveal new statistical features of networks, but also provide useful criteria in judging the validity of evolution models. (For example, the Barabási-Albert model [2] does not display high-order power-law clique-degree distributions.) The clique degree, which can be considered as an extension of degree, may be useful in measuring the density of motifs; such subunits not only play a role in controlling the dynamic behaviors, but also refer to the basic evolutionary characteristics. More interesting, we find that various real-life networks display power-law clique-degree distributions of decreasing exponent with the clique order. This is an interesting statistical property, which can provide a criterion in the studies of modeling.

It is worthwhile to recall a prior work [13] that reported a similar power-law distribution observed for some cellular networks. They divided all the subgraphs into two types. Moreover, they derived the analytical expression of the power-law exponent δ'_m for m-clique degree distribution as [13] $\delta'_m=1+(\gamma-1)/[m-1-\alpha(m-1)(m-2)/2]$, where α denotes the power-law exponent of clustering-degree correlation $C(k)\sim k^{-\alpha}$. Table II displays the predicted power-law exponents δ'_m, compared with the empirical observation δ_m. For the type-I cases, the predicted results are, to some extent, in accordance with the empirical data. Note that, although the power law is detected for type-II cases, the analytical expression of δ'_m loses its validity in those cases. The qualitative difference in type-II cases and quantitative departure in type-I cases may be attributable to the structural bias (e.g., assortative connecting pattern [32], rich-club phenomenon [35], etc.) since the derivation in Ref. [13] is based on uncorrelated networks. In addition, the predicted accuracy decreases as the increase of clique size m, because the clustering coefficient takes into account only the triangles [36]. Therefore, a more accurate analysis may involve a higher-order clustering coefficient [7]. In other words, Ref. [13] provides a starting point of an in-depth understanding of the network structure at the clique level, while the diversity and complexity of real networks require further explorations on this issue.

TABLE II. The empirical (δ_m) and predicted (δ'_m) power-law exponent of the clique-degree distribution, where γ and α denote the power-law exponents of the degree distribution and clustering-degree correlation. The symbol "/" denotes the cases with $\alpha(m-2)>2$, leading to negative and meaningless δ'_m.

Networks	γ	α	m	δ_m	δ'_m	Type
Internet at AS level	2.21	1.04	3	1.82	2.26	II
			4	1.48	/	II
			5	1.28	/	II
Internet at routers level	2.60	0.16	3	1.72	1.86	I
			4	1.49	1.63	I
			5	1.33	1.53	I
Metabolic network	2.04	0.80	3	1.85	1.87	I
			4	1.56	2.73	II
			5	1.43	/	II
World-Wide-Web	2.33	1.15	3	1.59	2.56	II
			4	1.37	/	II
			5	1.22	/	II
Collaboration network	2.21	0.90	3	1.90	2.10	II
			4	1.53	5.03	II
			5	1.41	/	II
ppi-yeast networks	2.18	0.91	3	1.68	2.08	II
			4	1.47	5.37	II
			5	1.36	/	II
Friendship networks	1.85	0.32	3	1.48	1.51	I
			4	1.25	1.42	I
			5	1.20	1.41	I

We thank Dr. Ming Zhao for useful discussions. This work is supported by the National Natural Science Foundation of China under Grants Nos. 10472116, 70471033, and 10635040.

[1] D. J. Watts *et al.*, Nature (London) **393**, 440 (1998).

[2] A.-L. Barabási *et al.*, Science **286**, 509 (1999).

[3] R. Albert *et al.*, Rev. Mod. Phys. **74**, 47 (2002).

[4] S. N. Dorogovtsev *et al.*, Adv. Phys. **51**, 1079 (2002).

[5] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[6] S. Boccaletti *et al.*, Phys. Rep. **424**, 175 (2006).

[7] L. da F. Costa *et al.*, Adv. Phys. **56**, 167 (2007).

[8] L. A. N. Amaral *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).

[9] R. Milo *et al.*, Science **298**, 824 (2002).

[10] A.-L. Barabási *et al.*, Nat. Rev. Genet. **5**, 101 (2004).

[11] S. Itzkovitz, R. Milo, N. Kashtan, G. Ziv, and U. Alon, Phys. Rev. E **68**, 026127 (2003).

[12] R. Milo *et al.*, Science **303**, 1538 (2004).

[13] A. Vázquez *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 17940 (2004).

[14] Y. Artzy-Randrup *et al.*, Science **305**, 1107c (2004).

[15] R. Milo *et al.*, Science **305**, 1107d (2004).

[16] I. Derényi, G. Palla, and T. Vicsek, Phys. Rev. Lett. **94**, 160202 (2005).

[17] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).

[18] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).

[19] http://www.cosin.org/extra/data/internet/nlanr.html

[20] http://www.isi.edu/scan/mercator/map.html

[21] H. Jeong *et al.*, Nature (London) **407**, 651 (2000).

[22] R. Albert *et al.*, Nature (London) **401**, 130 (1999).

[23] http:/www.oakland.edu/~grossman

[24] http://dip.doe-mbi.ucla.edu/

[25] This network is constructed based on the BBS of USTC, wherein each node represents a BBS accounts and two nodes are neighboring if one appears in the other one's friend list. Only the undirected network is considered.

[26] M. L. Goldstein *et al.*, Eur. Phys. J. B **41**, 255 (2004).

[27] P.-P. Zhang *et al.*, Physica A **360**, 599 (2006).

[28] S. L. Pimm, *Food Webs* (University of Chicago Press, Chicago, 2002).

[29] S. Maslov *et al.*, Science **296**, 910 (2002).

[30] B. J. Kim, Phys. Rev. E **69**, 045101(R) (2004).

[31] M. Zhao *et al.*, Physica A **371**, 773 (2006).

[32] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[33] E. Ravasz and A. L. Barabasi, Phys. Rev. E **67**, 026112 (2003).

[34] C. Song *et al.*, Nature (London) **433**, 392 (2005).

[35] S. Zhou *et al.*, IEEE Commun. Lett. **8**, 180 (2004).

[36] The theory in Ref. [13] is really accurate for δ_3 if belongs to type I; for example, δ_3 in *random Apollonian networks* [37] can be exactly predicted by the analytical result δ'_3.

[37] T. Zhou, G. Yan, and B. H. Wang, Phys. Rev. E **71**, 046141 (2005).