Statistical analysis of airport network of China

W. Li[1,2,*] and X. Cai[1]
[1]*Institute of Particle Physics, Hua-Zhong Normal University, Wuhan 430079, China*
[2]*Max-Planck Institute for Mathematics in the Sciences, Inselstrasse 22-26, D-04103 Leipzig, Germany*
(Received 13 August 2003; published 26 April 2004)

Through the study of airport network of China (ANC), composed of 128 airports (nodes) and 1165 flights (edges), we show the topological structure of ANC conveys two characteristics of small worlds, a short average path length (2.067) and a high degree of clustering (0.733). The cumulative degree distributions of both directed and undirected ANC obey two-regime power laws with different exponents, i.e., the so-called double Pareto law. In-degrees and out-degrees of each airport have positive correlations, whereas the undirected degrees of adjacent airports have significant linear anticorrelations. It is demonstrated both weekly and daily cumulative distributions of flight weights (frequencies) of ANC have power-law tails. Besides, the weight of any given flight is proportional to the degrees of both airports at the two ends of that flight. It is also shown the diameter of each subcluster (consisting of an airport and all those airports to which it is linked) is inversely proportional to its density of connectivity. Efficiency of ANC and of its subclusters is measured through a simple definition. In terms of that, the efficiency of ANC's subclusters increases as the density of connectivity does. ANC is found to have an efficiency of 0.484.

DOI: 10.1103/PhysRevE.69.046106 PACS number(s): 89.75.Da, 89.40.Dd, 89.75.−k

The ER model [1] of random graphs, introduced by Erdös and Rényi, has attracted much attention from both mathematicians and physicists [2–6], and henceforth shaped our understanding of networks for decades. The growing interest in whether randomness dominates real-world networks, however, was eventually prompted by recent advances in two main streams of topics. One part of this work was related to "small worlds," originally described as "six degrees of separation" [7] which manifests that humans are connected through a short, limited chain of acquaintances. The concept was successfully employed by Watts and Strogatz [8,9] in exploring the dynamics of a great variety of networks between order and randomness, e.g., the actor and actress networks [10], the chemical reaction networks [11], the rumor spreading networks [9], the food webs [12], and the electronic circuits [13], etc. Another parallel achievement was made by the research team of Barabási [14–17], which led to the finding of a class of networks with scale-free degree distributions, for example, Internet [14], the networks of coauthorship in natural sciences [18], the web of sexual contacts [19], and the graph of human language [20], etc.

Composed of a number of airports and flights, air networks are simply normal examples of transportation systems among ubiquitous networks in nature. Nevertheless, they appear extraordinary and unique due to the following features: (a) quite limited system sizes, from a few hundred to a few thousand at most; (b) relatively stationary structures with respect to both time and space; (c) bidirectional, weighted links (flights) with slightly fluctuating frequency.

This paper will present investigations of airport network in China (ANC). We demonstrate that on one hand ANC embodies part features of small worlds and of scale-free networks. On the other hand, however, ANC exhibits more features belonging to its topological uniqueness. The whole text is organized as follows. Section I presents the results on degree distributions and degree correlations of ANC. Section II gives the flight weight distributions and the weight-degree correlation of ANC. Section III analyzes the clustering coefficients of ANC. In Sec. IV we calculate the diameter of ANC and discuss the efficiency of ANC by proposing a simple definition for it. Conclusions and discussions are given in the last part, Sec. V.

I. DEGREE DISTRIBUTIONS AND DEGREE CORRELATIONS

ANC consists of $N=128$ [21] airports (nodes) and 1165 flights (edges) that connect most major cities in China. The topology of ANC can be symbolized by a $128\times128\times7$ connectivity matrix C whose entry C_{ijt} is 1 if there is a link pointing from node i to node j at the tth day of a week (herein and after $t=1$, 2, 3, 4, 5, 6, and 7 specifies the seven days within a week, starting from Monday, respectively) and 0 otherwise, and a $128\times128\times7$ weight matrix W [22] whose element is defined as

$$W_{ijt}=\frac{n_{ijt}}{\sum_t\sum_{\{i,j\}}n_{ijt}}, \tag{1}$$

where n_{ijt} is the number of flights $i\to j$ at the tth day. W_{ijt} satisfies the normalization condition, i.e., $\Sigma_t\,\Sigma_{\{i,j\}}\,W_{ijt}=1$. Normally, $C_{ijt}=C_{jit}$ and $W_{ijt}=W_{jit}$ only hold for undirected ANC.

We employ $k_{in}^w(i)$ and $k_{ou}^w(i)$ to denote the in-degree and out-degree of a given node i in the directed ANC during a whole week time, and $k_{un}^w(i)$ to represent the undirected degree of the undirected ANC in the same week. Hence, we have

*Electronic address: liwei@mis.mpg.de

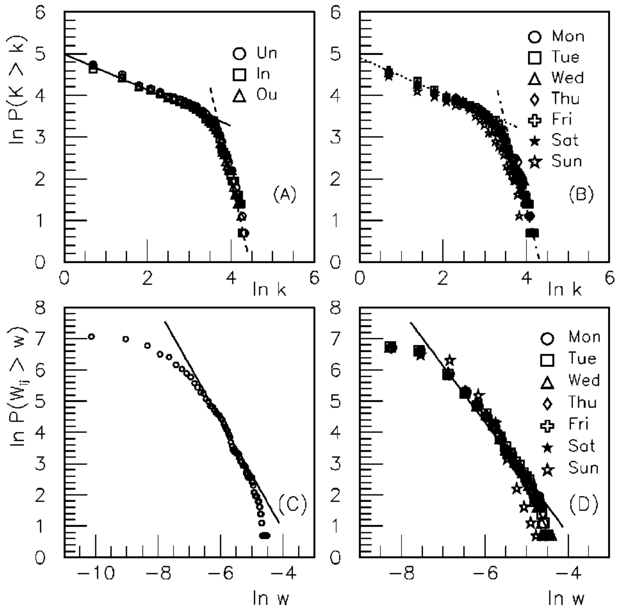

FIG. 1. Cumulative degree distributions of ANC for undirected degree, in-degree, and out-degree of (a) a whole week and (b) each day from Monday to Sunday. Cumulative weight distributions of ANC for (c) a whole week and (d) each day from Monday to Sunday.

$$k_{in}^{w}(i)=\sum_{j}^{j\neq i}\eta\left(\sum_{t=1}^{7}C_{ijt}-1\right),\qquad(2)$$

$$k_{ou}^{w}(i)=\sum_{j}^{j\neq i}\eta\left(\sum_{t=1}^{7}C_{jit}-1\right),\qquad(3)$$

and

$$k_{un}^{w}(i)=\sum_{j}^{j\neq i}\eta\left(\sum_{t=1}^{7}[C_{ijt}+C_{jit}]-1\right),\qquad(4)$$

where $\eta(x)$ is a unit step function, which takes 1 for $x\geqslant 0$ and 0 otherwise.

First we consider the three distributions of $k_{in}^{w}(i)$, $k_{ou}^{w}(i)$, and $k_{un}^{w}(i)$, respectively. Here the cumulative distribution, widely used in economies and well known as the Pareto law [23], is adopted to reduce the statistical errors arising from the limited system size. The cumulative form,

$$P(k_{in}^{w}(i)>k)[P(k_{ou}^{w}(i)>k)\text{ or }P(k_{un}^{w}(i)>k)],$$

gives the probability that a given airport i has an in-degree (out-degree or undirected degree) larger than k. Figure 1(a) presents behaviors of the three distributions. It is amazing to find that all three distributions follow nearly a same two-regime power law with two different exponents, known as *double Pareto law* [24], with a turning point at degree value $k_c\simeq 26$, which can be well prescribed by the following expression:

$$P(K>k)\sim\begin{cases}k^{-\gamma_1}, & \text{for } k\leqslant k_c\\ k^{-\gamma_2}, & \text{for } k>k_c,\end{cases}\qquad(5)$$

where γ_1 and γ_2 are the respective degree exponents of two separate power laws. By means of fitting, exponents pairs (γ_1,γ_2) of the three distributions in Fig. 1(a) are (0.428, 4.161), (0.416, 4.453), and (0.45, 4.535). Using a simple algebra, the original distributions of $k_{in}^{w}(i)$ [$k_{ou}^{w}(i)$ or $k_{un}^{w}(i)$] can be written as

$$P(k)=\frac{\partial P(K>k)}{\partial k}\sim\begin{cases}k^{-(\gamma_1+1)} & \text{for } k\leqslant k_c\\ k^{-(\gamma_2+1)} & \text{for } k>k_c,\end{cases}\qquad(6)$$

where k specifies the three different degrees above. Correspondingly, the mean values of $k_{in}^{w}(i)$, $k_{ou}^{w}(i)$, and $k_{un}^{w}(i)$ are 18.931, 17.156, and 18.203. This conveys that each airport, on average, is connected to around 18 other airports.

The undirected degree of a certain airport i at the tth day of a week is given by

$$k_{un}^{t}=\sum_{j}^{j\neq i}\eta(C_{ijt}+C_{jit}-1).\qquad(7)$$

The cumulative distributions of k_{un}^{t}, with $t=1$, 2, 3, 4, 5, 6, and 7, shown in Fig. 1(b), reflects the daily evolution of the topology of the undirected ANC within a week. It is evident from Fig. 1(b) that the distributions of days from Monday to Saturday nearly coincide with one another, on the same double Pareto law. The distribution of Sunday, however, deviates apparently from the shared curve and itself obeys another law. By checking the original data, one may find out the discrepancy is mainly caused by the fact that some flights are not available on Sundays. Exponents pairs and average undirected degrees of the undirected ANC for each day of one week are listed in Table I. As we can see, the values of γ_1 and γ_2 in the table are in general (except on Sundays)

TABLE I. Comparison of relevant variables of daily undirected ANC (from Monday to Sunday): γ_1 and γ_2 are exponents of two power laws of cumulative degree distributions; $\langle k\rangle$, the average degree; γ, the exponent of flight weight distributions; C, the clustering coefficient of the whole system.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
γ_1	0.582	0.569	0.568	0.603	0.558	0.574	0.463
γ_2	4.398	4.190	4.338	3.949	4.308	4.264	3.992
$\langle k\rangle$	13.570	14.376	13.967	14.017	14.033	14.586	12.264
γ	1.744	1.682	1.729	1.699	1.679	1.747	2.329
C	0.626	0.621	0.614	0.590	0.638	0.620	0.576

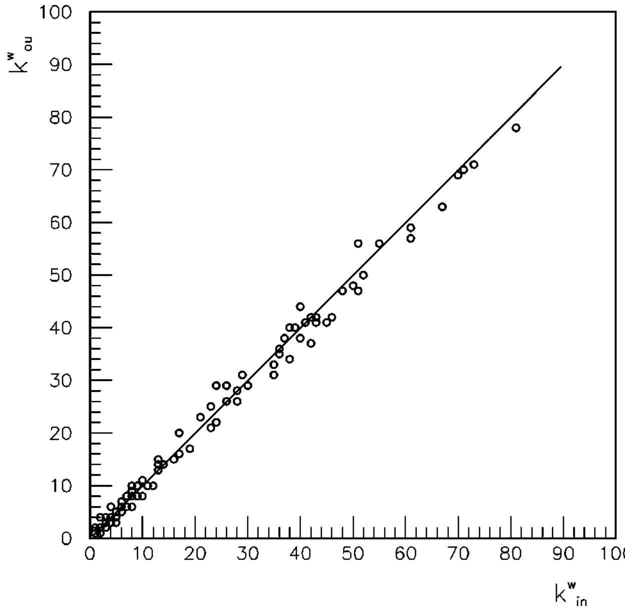

FIG. 2. Correlation between in-degrees and out-degrees of the directed ANC in a whole week.

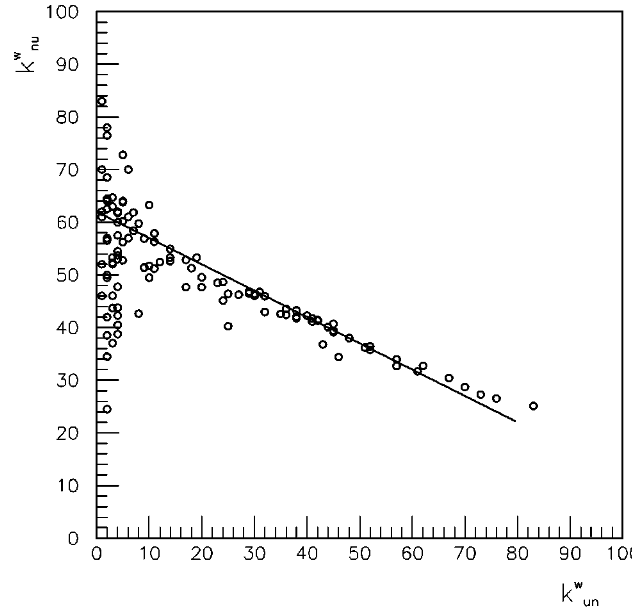

FIG. 3. Correlation between degrees of adjacent airports of the undirected ANC in a whole week.

slightly larger than the counterparts of undirected ANC during a whole week. The average degrees of each day, around 14 (12 on Sundays), are much smaller than 18, the counterpart of a week. This is understandable because many flights are only available on certain days of a week.

We also check an important feature of ANC, the degree correlations. First we come to the correlation between in-degrees and out-degrees, simply called in-out degree correlation. This is quite natural for airport networks because each airport should generally maintain the balance of its traffic flow. Normally, for each airport, the higher its in-degree, the higher its out-degree. We plot $k_{in}^{w}(i)$ versus $k_{ou}^{w}(i)$ $(i=1,2,\ldots,128)$ in Fig. 2. The following expression can be obviously obtained by fitting the data:

$$k_{in}^{w}(i) \simeq k_{ou}^{w}(i). \tag{8}$$

Evidently, the in-out degree correlation is very strong.

Another possible correlation exists between the degrees of adjacent airports, named degree-degree correlation. The degree-degree correlation tells that the degrees are not independent and correlate with those of their neighbors. It can be demonstrated by calculating the mean degree of the neighbors of a given airport as a function of the degree of that airport. Figure 3 presents our analysis of degree-degree correlation in the undirected ANC. As shown, the degrees of adjacent airports have significant anticorrelations, based on which the ANC appear to be disassortative [25]. But the anticorrelation found in ANC is almost linear, different than that found in Ref. [26], which is a power law with exponent of about −0.5.

II. FLIGHT WEIGHT DISTRIBUTIONS AND WEIGHT-DEGREE CORRELATION

An important feature of ANC is that some flights are more frequent than others. The weight or the frequency of a certain flight, henceforth, states the extent to which it is busy. The weight of flight $i \rightarrow j$ in a whole week is given by

$$W_{ij}^{w} = \sum_{t=1}^{7} W_{ijt}. \tag{9}$$

The cumulative distribution of W_{ij}^{w}, $P(W_{ij}^{w} > W)$, gives the probability that a flight has a weight larger than w. Shown in Fig. 1(c), $P(W_{ij}^{w} > W)$ has a power-law tail,

$$P(W_{ij}^{w} > W) \sim W^{-\gamma}, \tag{10}$$

where γ=1.65. Through a simple algebra, one may obtain $P(W) \sim (W)^{-2.65}$. Such a power-law tail indicates that the probability of finding a very busy flight is nonzero, and significant instead. The daily cumulative distributions of W_{ijt} within a week is given in Fig. 1(d). Among the seven distributions, those from Monday to Saturday obey the same power law, while Sunday data reveal a steeper power law that extends a narrower region on the x coordinate. The exponents of flight weight distributions of each day are also presented in Table I and are slightly larger than 1.65.

We also conjecture if there is a certain kind of relation between the weight of a given flight and the degrees of the two airports at both ends of that flight. We simply call it weight-degree correlation. Without losing the generality, we propose the following ansatz for the possible existence of such correlation:

$$W_{ij}^{w} \sim [k_{un}^{w}(i) k_{un}^{w}(j)]^{1/2}. \tag{11}$$

This scaling ansatz has been well supported by the real data, shown in Fig. 4.

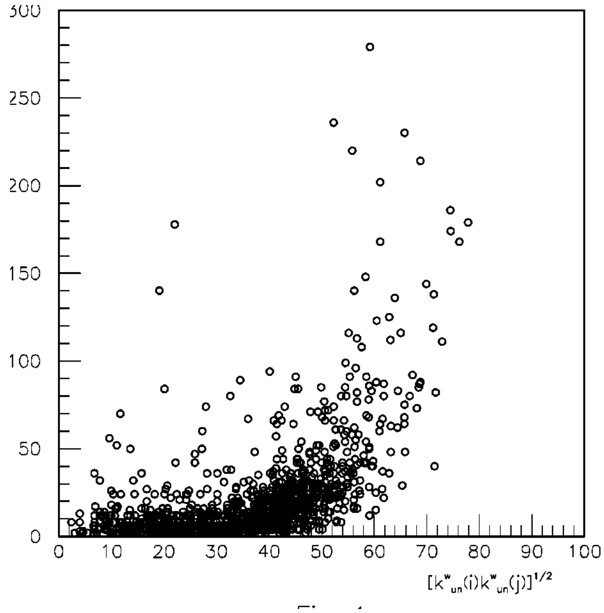

FIG. 4. Weight-degree correlation of the undirected ANC in a whole week.

III. CLUSTERING COEFFICIENT

The neighborhood Γ_v of a given airport v is a graph which includes all nodes which have flights with v. The clustering coefficient [9] $C(\Gamma_v)$ of neighborhood Γ_v of airport v characterizes the extent to which airports in Γ_v are connected to every other. In precise words,

$$C(\Gamma_v) = \frac{E(\Gamma_v)}{C_m^2}, \tag{12}$$

where $E(\Gamma_v)$ is the number of real connections in Γ_v consisting of m airports, and C_m^2 is the total number of all possible connections in Γ_v. The average clustering coefficient of the entire air network is defined as

$$C = \frac{1}{N}\sum_{\Gamma_v} C(\Gamma_v), \tag{13}$$

where N is the number of airports of the whole network. By calculation, C of the entire undirected ANC for a whole week is 0.733, in stark contrast with the low density of connectivity, $\langle k\rangle/N$, 0.143. C of the daily undirected ANC given in Table I centralizes 0.600, the value for Sunday being slightly lower.

IV. DIAMETER AND EFFICIENCY

For a connected network, the diameter D can have the following definition:

$$D = \frac{1}{N(N-1)/2}\sum_{(i,j)} d_{min}(i,j), \tag{14}$$

where $d_{min}(i,j)$ represents the shortest-path length between nodes i and j. In an air network, the diameter D indicates the average number of transfers a passenger needs to take between the start and the end. For ANC, D is around 2.067. Specifically, $d_{min}(i,j)$ in ANC only takes three distinctive values, 1, 2, and 3, with percentages of 0.143, 0.646, and 0.211, respectively. This implies most trips will need one intermediate transfer or two before the final destinations, only a small percent can be reached directly.

The high clustering and the small diameter inevitably indicate the small-world property of ANC. For comparison, random graphs of the same average degree, $\langle k\rangle$, and the same number of nodes, N, with ANC are investigated. It is readily learned that the average clustering coefficient of random graphs, 0.143, is much smaller than 0.733, the weekly average clustering coefficient. The diameter of such random graphs scales as $\ln N/\ln\langle k\rangle$, which is 1.672, less than the counterpart of ANC.

A practical thing of ANC is related to its transportation efficiency, which tells us how one can travel from one place to another both quickly and economically. Let us first take a look at the efficiency of subclusters of ANC. A subcluster here is composed of a hub v, the central node, and its neighborhood $\Gamma(v)$ consisting of whoever has flights with the hub. The largest subcluster of ANC includes 84 airports, and the smallest one, only 2. In terms of graph theory, the subclusters consist of only two kinds of structure, trees and triangles. The density of connectivity of a subcluster with m nodes and $E(\Gamma_v)$ edges in $\Gamma(v)$ of the hub is

$$\rho_{dc} = \frac{2[E(\Gamma_v) + m]}{m^2 + m}. \tag{15}$$

The diameter of the subcluster, d_{sc}, can be readily derived:

$$D_{sc} = \frac{2[m^2 - E(\Gamma_v)]}{m^2 + m}. \tag{16}$$

The plot of D_{sc} versus ρ_{dc}, for all 128 subclusters of ANC, is presented in Fig. 5(a), which can be well fitted by a straight line. The larger ρ_{dc} is, the more direct connections there exist in the subclusters, and the smaller the diameter will be. In the case of a complete graph, the diameter will be definitely 1.

We simply define the efficiency of subclusters of ANC, E_{sc}, as

$$E_{sc} = \frac{1}{D_{sc}} = \frac{m^2 + m}{2[m^2 - E(\Gamma_v)]}. \tag{17}$$

After a simple calculation, E_{sc} versus ρ_{sc} is presented in Fig. 5(b). It is clearly shown that the higher the density of connectivity, the higher the efficiency of a subcluster. The efficiency is 1 when the subcluster is totally connected. This agrees with our intuition.

Compared with its subclusters, ANC itself displays no more difference in structure. The ANC can be viewed as a cluster with hierarchical structure [27], composed of a center, e.g., Beijing, and whoever has direct connections with the center, and whoever has no direct connections with the center, but with whoever has, and so on. For a connected network, such a cluster can include all nodes in the same system. By analyzing the real data, each node of ANC is

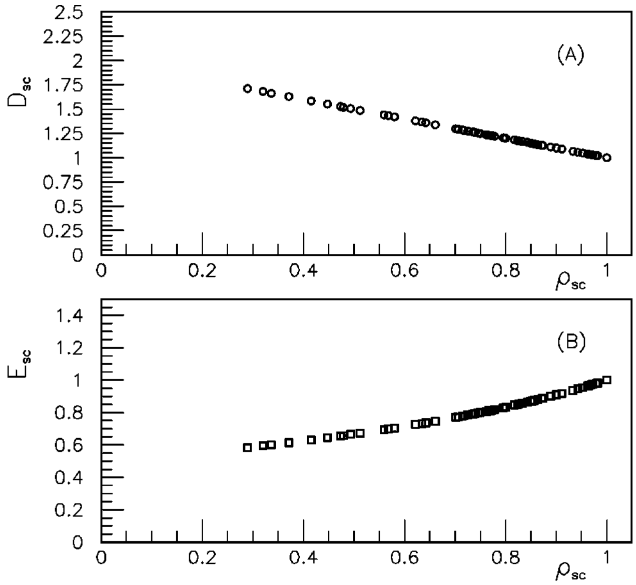

FIG. 5. (a) Diameter and (b) Efficiency vs density of connectivity for subclusters of the undirected ANC.

connected to any other with no more than three steps. In this sense, Eq. (17) also applies to ANC. After some algebra, we find the efficiency of ANC is 0.484.

V. CONCLUSIONS AND DISCUSSIONS

In conclusion, our analysis reveals two characteristic small-world properties of ANC, a short average path length and a high degree of clustering. Another important feature of ANC, the degree distribution, however, is strikingly different from counterparts of both scale-free networks and of random graphs. In ANC there exist strong, positive correlations between in-degrees and out-degrees of each airport, and significant anticorrelations between degrees of adjacent airports. The weekly and daily weight distributions of ANC display power-law behaviors. The existence of weight-degree correlation of ANC shows that there is a dependence of the weight of a certain flight on the degrees of the two airports at both ends of that flight. In particular, we suggest a rough idea to measure the efficiency of ANC and that of its subclusters.

In the previous sections we do not answer why the structure of ANC obeys double Pareto law. Here we come up with a simple idea which can be realized through computer simulation. Suppose one constructs a whole airport network from the very beginning, with only a few airports in major cities, following two simple rules. Under the first rule, preferential attachment [14], a newly established airport tends to connect to the hubs with more flights, which naturally drives the airport network to develop a structure beyond those of random graphs. The second rule manifests the existence of different growth rates of airports between the region of smaller airports and that of larger ones. That is, in the early history of airport network construction, smaller airports have considerable probabilities to be growing to accommodate more flights. Gradually, as most major airports have been established, the smaller airports were unlikely to expand any more. Hence more small-sized airports were established. This limited growth endows the airport network features part of scale-free topology. It may be more appropriate to say that ANC has an intermediate topology between random graphs and scale-free networks.

Another issue should be addressed to the efficiency. The efficiency based on our definition is solely idealistic and only limited to the structure of the network itself. It is more appropriate to call it structural efficiency. In the reality of air transportation, the carriers (airlines) should consider more factors in order to have a higher and reasonable efficiency. That is, one needs to know how an air network can satisfy the passengers' needs on one hand, and ensure the profits of airlines, on the other hand. This should be an interesting topic and worth investigating.

W.L. would like to thank Alexander von Humboldt Stiftung for research funding and Professor Juergen Jost of Max-Planck Institute for Mathematics in the Sciences for hosting. This work was supported in part by National Natural Science Foundation of China Under Grant No. 70271067, a key program of Hua-Zhong Normal University, and a grant from the international program of the Santa Fe Institute of the United States of America.

[1] P. Erdös and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959); Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960); Bull. Inst. Int. Stat. **38**, 343 (1961).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[4] B. Bollobás, *Random Graphs* (Academic, London, 1985).
[5] D. ben Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, 2000).
[6] K. Christensen, R. Donangelo, B. Koiller, and K. Sneppen, Phys. Rev. Lett. **81**, 2380 (1998).
[7] S. Milgram, Psychol. Today **1**, 60 (1967).
[8] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[9] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press, Princeton, NJ, 1999).
[10] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[11] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, Proc. Natl. Acad. Sci. U.S.A. **97**, 1143 (2001).
[12] J. M. Montoya and R. V. Solé, J. Theor. Biol. **214**, 405 (2002); R. J. Williams, N. D. Martinez, E. L. Berlow, J. A. Dunne, and A.-L. Barabsi, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 913 (2002).

[13] R. F. i Cancho, C. Janssen, and R. V. Solé, Phys. Rev. E **64**, 046119 (2001).
[14] A.-L. Barabási and R. Albert, Nature (London) **286**, 509 (1999).
[15] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[16] H. Jeong, B. Tombor, R. Albert, Z. N. Oltval, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[17] S. H. Yook, H. Jeong, A.-L. Barabási, and Y. Tu, Phys. Rev. Lett. **86**, 5835 (2001).
[18] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001);, Phys. Rev. E **64**, 016131 (2001); **64**, 016132 (2001).
[19] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).
[20] R. F. i Cancho and R. V. Solé, Proc. R. Soc. London, Ser. B **268**, 2261 (2001).
[21] The data were taken from a website. According to it, the total number of airports in China is 129, within which Pudong and Hongqiao belong to the same city Shanghai and are regarded as united.
[22] Strictly speaking, the air network should also include a matrix of spatial distance L. In this paper, L is not taken into account.
[23] V. Pareto, *Cours d'Economie Politique*, edited by F. Pichou (University of Lausanne Press, Lausanne, 1897), Vol. 2.
[24] W. J. Reed, Physica A **319**, 469 (2003).
[25] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002); Phys. Rev. E **67**, 026126 (2003).
[26] R. Pastor-Satorras, A. Vaźquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[27] W. Li and X. Cai, Phys. Rev. E **61**, 771 (2001); **61**, 5630 (2001); **62**, 7743 (2001).