Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 367 (2006) 613–618
PHYSICA A
www.elsevier.com/locate/physa

Topological properties of integer networks

Tao Zhou[a], Bing-Hong Wang[a,*], P.M. Hui[b], K.P. Chan[b]

[a]Department of Modern Physics and Nonlinear Science Center, University of Science and Technology of China, Hefei Anhui 230026, PR China
[b]Department of Physics, The Chinese University of Hong Kong, New Territories, Shatin, Hong Kong, PR China

Received 26 May 2005; received in revised form 30 October 2005
Available online 9 December 2005

Abstract

Inspired by Pythagoras's belief that numbers represent the reality, we study the topological properties of networks of composite numbers, in which the vertices represent the numbers and two vertices are connected if and only if there exists a divisibility relation between them. The network has a fairly large clustering coefficient $C \approx 0.34$, which is insensitive to the size of the network. The average distance between two nodes is shown to have an upper bound that is independent of the size of the network, in contrast to the behavior in small-world and ultra-small-world networks. The out-degree distribution is shown to follow a power-law behavior of the form k^{-2}. In addition, these networks possess hierarchical structure as $C(k) \sim k^{-1}$ in accord with the observations of many real-life networks.

Keywords: Complex networks; Integer networks; Upper bound of average distance

Many social, biological, and communication systems form complex networks, with vertices representing individuals or organizations and edges representing the interactions between them [1–3]. Examples are numerous, including the Internet, the World Wide Web, social networks of acquaintance and other relationship between individuals, metabolic networks, food webs, etc. [4–10]. Empirical studies on real-life networks reveal some common characteristics different from random networks and regular networks. Among these features, the most noticeable characteristics are the small-world effect and scale-free property [11–13]. In this paper, inspired by Pythagoras' belief that numbers represent absolute reality, we study the topological properties of the most *natural* network in Pythagoras' sense, namely the network consisting of integers.

The distance d between two vertices in a network is defined as the number of edges along the shortest path connecting them. The average distance L of the network is the mean distance between two vertices, averaged over all pairs of vertices. The average distance is one of the most important properties in measuring the efficiency of communication networks. In a store-forward computer network, for example, the most useful measurement characterizing the performance is the transmission delay (or time delay) in sending a message through the network from the source to the destination. The time delay is approximately proportional to the number of edges that a message must pass through. Therefore, the average distance plays a significant role in

*Corresponding author. Fax: +86 551 3603574.
E-mail address: bhwang@ustc.edu.cn (B.-H. Wang).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2005.11.011

measuring the time delay. Other important topological properties include the clustering coefficient and degree distribution. The degree of a vertex x, denoted by $k(x)$, is the number of the edges that are attached to the vertex. Through the $k(x)$ edges, there are $k(x)$ vertices that are correlated with or connected to the vertex x. These neighboring vertices form a set of neighbors $A(x)$ of x. The clustering coefficient $C(x)$ of the vertex x is the ratio of the number of existing edges among all the vertices in the set $A(x)$ to the total possible number of edges connecting all vertices in $A(x)$. The clustering coefficient C of the entire network is the average of $C(x)$ over all x. Empirical studies indicate that many real-life networks have much smaller average distances (with $L\sim \ln S$, where S is the number of vertices in the network) than completely regular networks and have much larger clustering coefficients than the completely random networks. Studies on real-life networks also show a power-law degree distribution of the form $p(k)\sim k^{-\gamma}$, where $p(k)$ is the probability density function for the degrees and γ is an exponent. This distribution falls off slower than an exponential, allowing for a few vertices with large degrees. Networks with power-law degree distribution are referred to as scale-free networks.

Here, we investigate a network in which the vertices represent a set of positive integers. Two vertices x and y are linked by an edge if and only if x is divisible by y or y is divisible by x. By definition, the topological structure of the network is determined, once the set of vertices under consideration is known. For example, Fig. 1 shows the topological structure of a network with the set of vertices representing the integers $\{4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30\}$.

Formally [14], a network is represented by a graph $G(V, E)$, where V is the set of vertices and E is the set of edges. A graph G is said to be connected if any two vertices of G are connected, i.e., one can go from one vertex to another through the edges in the network. If a graph G is not connected, it consists of several disjoint components. The example in Fig. 1 has five components: $V1 = \{4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 24, 27, 28, 30\}$, $V2 = \{21\}$, $V3 = \{22\}$, $V4 = \{25\}$ and $V5 = \{26\}$. Although real-life networks are not always connected, most previous studies have focused on connected graphs. In the analysis of the network of integers, we only keep the largest connected component if the given set of integers gives a disconnected graph. With this choice, the sets of vertices $\{4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 27, 28, 30\}$ and $\{4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 24, 27, 28, 30\}$, for example, will generate the same graph consisting of 15 vertices and 19 edges.

Different sets of vertices (positive integers) will, in general, lead to graphs of different properties. For instance, the set of integers $\{1, 2, 4, \ldots, 2^n\}$ generates a complete graph, and the set consisting of all the prime

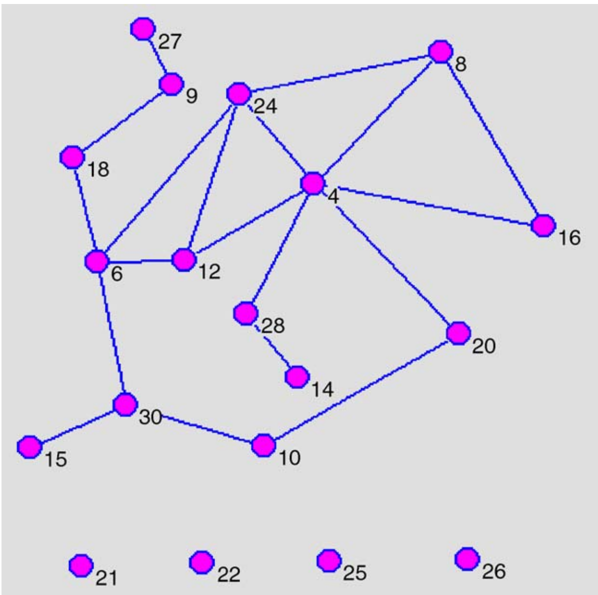

Fig. 1. (Color online) The topological structure of an integer network with the set of vertices representing the integers $V = \{4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30\}$. The corresponding graph consists of 19 vertices and 19 edges, and shows 5 disjoint components. The largest component has 15 vertices and 19 edges, which is denoted by G_{30}.

numbers and 1 leads to an infinite star graph. Here we focus on a special class of networks of integers in which the vertices are *composite numbers*. A composite number is a positive integer greater than 1 and is not prime. We use the symbol G_N to denote the network generated by the set of composite numbers less than or equal to N, i.e., $V_N = \{x|x \in Q, 4 \leqslant x \leqslant N\}$, where Q is the set of all the composite numbers. For example, the largest component of the network in Fig. 1 is the graph G_{30}. Note that the number of vertices in G_N grows with N. We have calculated the clustering coefficients C in the networks G_N up to $N = 20,000$ and found that the clustering coefficients only fluctuate very slightly about the value of 0.34 with N (see inset of Fig. 2).

Fig. 2 also shows the dependence of the average distance L in the networks G_N for N up to 20,000. After a range of N less than 5000 for which L increases from about 1.8 to 2.4, the average distance increases fairly slowly with N for $N > 5000$. Note that the average distance $L \sim \ln N$ in small-world networks [11] and $L \sim \ln \ln N$ in ultrasmall-world networks [15]. Here, we show that the diameter [16] D of G_N must be smaller than a constant M that is independent of network size characterized by N, i.e., $D < M$ in G_N for arbitrary N, where M is a constant. For any vertex $x \in G_N$, we first prove that the distance between x and the smallest integer 4 in G_N satisfies $d(x,4) \leqslant 4$. Since $x \in Q$, it can be written as $x = p_1 p_2 \cdots p_q$, where $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_q$ are prime numbers. If $q \geqslant 4$, then $x/p_1p_2 \in Q$, thus $x/p_1p_2 \in A(x)$. Noting that $4x/p_1p_2 \leqslant x \leqslant N$, there exists a path of length 3 from x to 4 through the vertices x/p_1p_2 and $4x/p_1p_2$. If $q = 3$ and $p_1 > 4$, analogously, a path through the vertices x/p_1 and $4x/p_1$ exists between x to 4. If $q = 3$ and $p_1 < 4$, then x can only be $x =$ 8, 12, 18,

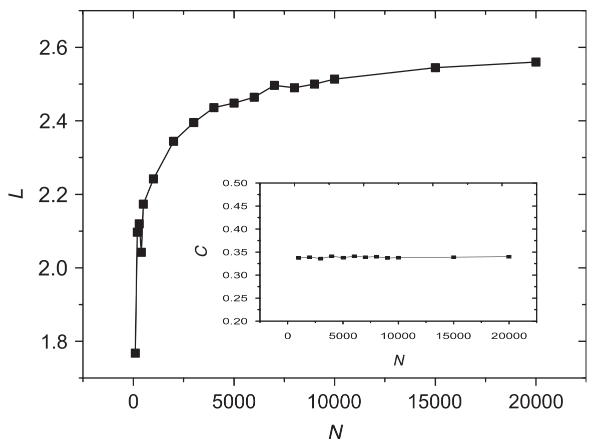

Fig. 2. The average distance L in the network G_N for network size up to $N = 20,000$. For $N > 5000$, L increases only slowly with N. Analytically, there exists an upper bound for L that is independent of N. The inset shows the clustering coefficient C, which fluctuates only slightly about the value 0.34 as N increases.

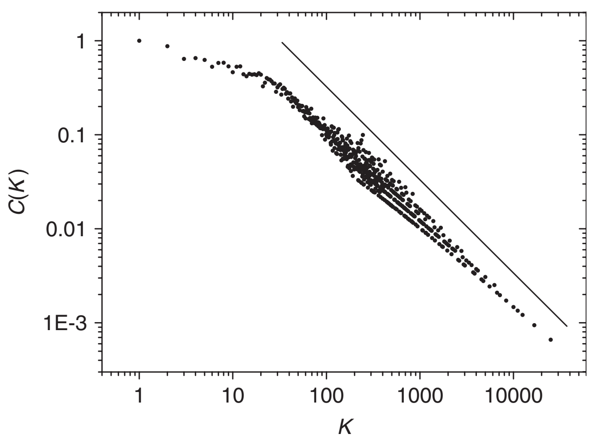

Fig. 3. The scaling of $C(k)$ with k for integer network $G_{100,000}$. The solid line in the log–log plot has slope -1 for comparison.

and 27, for which 8 and 12 are connected directly to 4. For $N > 107$, there exists a path from x to 4 through $4x$. Therefore, for any vertex x with at least 3 prime factors, $d(x,4) \leqslant 3$. Finally, for $q = 2$, since $A(x) \neq \emptyset$, $y = 2x$ must be in $A(x)$. Then y has at least 3 prime factors: p_1, p_2 and 2, leading to $d(y,4) \leqslant 3$ and thus $d(x,4) \leqslant 4$. Therefore, for any two vertices $a, b \in G_N$, we have $d(a,b) \leqslant d(a,4) + d(b,4) \leqslant 8$. This implies $D \leqslant 8$. Since the average distance $L < D$, the networks G_N have a constant upper bound for L. The networks G_N are, thus, distinguished from the other networks by having a large clustering coefficient C (≈ 0.34) and a constant upper bound to the average distance L.

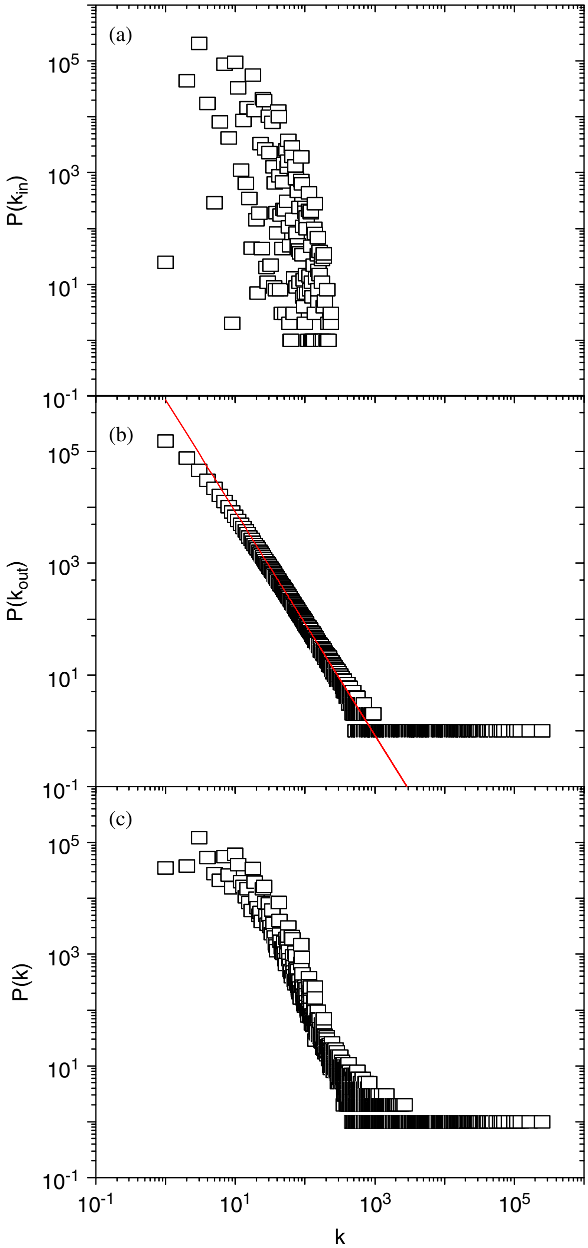

Fig. 4. (Color online) The (a) in-degree and (b) out-degree distributions of the network $G_{1000000}$. The y-axis gives the total number of nodes having degree k. The line in (b) shows that the out-degree distribution follows a power-law of the form $\sim k^{-2}$. The total degree (with in and out degrees counted together) distribution is shown in (c).

The clustering coefficient of a vertex will also depend on the degree of the vertex. Fig. 3 shows the results for $C(k)$, where $C(k)$ is the average clustering coefficient of all the nodes with degree k. We added a solid line with slope -1 in the plot for comparison. The results indicate that $C(k)$ approximately follows a power-law with an exponent ≈ -1. We note that the relationship $C(k){\sim}k^{-1}$ has been observed in many real-life networks, while it is in contrast to the behavior of $C(k)$ in the scale-free growing networks with preferential attachments in which $C(k)$ has been claimed to be k-independent [17].

Fig. 4 shows the degree distributions of the network $G_{1000000}$. The degree of a vertex x is the sum of the in-degree and out-degree: $k(x) = k_{in}(x) + k_{out}(x)$; where the in-degree of x is defined as the number of elements in the set $A(x)$ that are factors of x, and the out-degree of x is the number of elements in $A(x)$ that are divisible by x. Fig. 4(a) and (b) show the in-degree and out-degree distributions. Note that the range of in-degrees is smaller than that of out-degrees in the network G_N. The in-degree distribution does not show any power-law behavior. For the out-degrees, $k_{out}(x) = \lfloor N/x \rfloor - 1$. This leads to $P(k_{out}){\sim}k^{-2}$, for a range of k larger than unity. This behavior is shown by the straight line in (b). The behavior for very large k is limited by the size of the network characterized by N. Counting the in-degree and out-degree together, Fig. 4(c) shows the total degree distribution.

In summary, we proposed and studied a kind of networks that can be regarded as the most natural, namely networks consisting of sets of composite numbers as vertices. Such networks exhibit interesting features that are different from many previously studied real-life networks. Most noticeably, these networks show a fairly large clustering coefficient $C \approx 0.34$, hierarchical structure $C(k){\sim}k^{-1}$, and a power-law out-degree distribution. More interesting, the average distance between two vertices is shown to have an upper bound that is independent of the network size, suggesting that they can be considered as a new class of networks wherein the infectious diseases may spread much more easily and quickly than traditional small-world networks especially when the network size is sufficient large.

Due to their special topological properties, the present networks have significant potential in the designs of communication systems. Firstly, most previous designs of the topology of distribution processing systems, local area networks, data memory allocation and data alignment in single instruction multiple data processors require deterministic structures [18]. Secondly, most widely used networks for the designs of communication systems have infinite average distance when the network size gets infinite $L \to \infty (N \to \infty)$, which may lead to great time delay if the system size becomes huge [19]. These examples include hypercubes [20], crossed cubes [21], double loops [22,23], and so on. While the upper bound of diameter in the present networks ensures an efficient communication. The last problem, also the most important one, is about the routing strategy. The previous studies have proved that one can scheme out highly efficient routing algorithm if the global topological information is known and the dynamical computing is feasible [24,25]. However, the global topological information requires too much memory for a single router, thus the global routing strategy, although much more efficient than local routing strategy [26–28], is usually impractical for large-scale communication networks. Fortunately, in the present networks, the global information is available while no memory is required [29]. This property is completely different from most known ones, providing us the feasibility to work out an efficient searching and routing algorithm for very large-scale networks.

This work unfolds an alternative perspective in the study of complex network, that is, to explore deterministic mathematical networks instead of search for real data in nature. Mathematical objects, although unintentionally defined to be so at the beginning, very often have close connections to the physical world [30–33]. It is, therefore, worth investigating the secret of other networks of mathematical objects such as those consisting of groups and rings as vertices.

The authors thank Miss Jun Liu and Dr. Jian-Guo Liu for their assistance in preparing the manuscript and Dr. Yee Jiun Yap for discussion. This work has been supported in part by the National Natural Science Foundation of China under Grant nos. 10472116, 10532060, 10547004, 70471033, 70271070 and 70571074, the Specialized Research Fund for the Doctoral Program of Higher Education of China (SRFDP No.20020358009), and the Foundation for Graduate Students of University of Science and Technology of China under Grant No. KD200408. One of us (P.M.H.) acknowledges the support of a Direct Grant for Research at CUHK.

References

[1] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47.
[2] S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. 51 (2002) 1079.
[3] M.E.J. Newman, SIAM Rev. 45 (2003) 167.
[4] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. 87 (2001) 258701.
[5] R. Albert, H. Jeong, A.-L. Barabási, Nature 401 (1999) 130.
[6] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, Y. Åberg, Nature 411 (2001) 907.
[7] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, Nature 407 (2000) 651.
[8] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, E.D. Gilles, Nature 420 (2002) 190.
[9] J. Camacho, R. Guimerà, L.A.N. Amaral, Phys. Rev. Lett. 88 (2002) 228102.
[10] P.-P. Zhang, K. Chen, Y. He, et al., Physica A 360 (2006) 599.
[11] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[12] A.-L. Barabási, R. Albert, Science 286 (1999) 509.
[13] A.-L. Barabási, R. Albert, H. Jeong, Physica A 272 (1999) 173.
[14] B. Bollobás, Modern Graph Theory, Springer, New York, 2002.
[15] A.F. Rozenfeld, R. Cohen, D. Ben-Avraham, S. Havlin, Phys. Rev. Lett. 89 (2002) 218701.
[16] The diameter D of a network $G(V, E)$ is the maximal distance: $D = \max\{d(x, y) \mid x, y \in V\}$, where $d(x, y)$ is the distance between two vertices x and y.
[17] E. Ravasz, A.-L. Barabási, Phys. Rev. E 67 (2003) 026112.
[18] J.-M. Xu, Topological Structure and Analysis of Interconnection Networks, Kluwer Academic Publishers, Dordrecht, 2001.
[19] A.S. Tanenbaum, Computer Networks, Prentice-Hall, Englewood Cliffs, NY, 1996.
[20] K. Efe, IEEE Trans. Comput. 40 (1991) 1312.
[21] P. Kulasinghe, S. Bettayeb, IEEE Trans. Comput. 44 (1995) 923.
[22] G.H. Barnes, R.M. Brown, M. Kato, D.J. Kuck, D.L. Slotnick, R.A. Stokes, IEEE Trans. Comput. 17 (1968) 746.
[23] T. Zhou, M. Zhao, B.-H. Wang, arXiv: cond-mat/0508368.
[24] L. Zhao, Y.-C. Lai, K. Park, N. Ye, Phys. Rev. E 71 (2005) 026125.
[25] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, B.-H. Wang, arXiv: cond-mat/0505366.
[26] B. Tadić, S. Thurner, G.J. Rodgers, Phys. Rev. E 69 (2004) 036102.
[27] J.D. Noh, H. Rieger, Phys. Rev. Lett. 92 (2004) 118701.
[28] C.-Y. Yin, B.-H. Wang, W.-X. Wang, T. Zhou, H.-J. Yang, arXiv: physics/0506204.
[29] For arbitrary pair of nodes, one can immediately know whether the two nodes are connected by checking the divisibility relation between them. Therefore, the global information in available without memory requirement.
[30] G. Corso, Phys. Rev. E 69 (2004) 036106.
[31] J.S. Andrade Jr., H.J. Herrmann, R.F.S. Andrade, L.R. da Silva, Phys. Rev. Lett. 94 (2005) 018702.
[32] T. Zhou, G. Yan, B.-H. Wang, Phys. Rev. E 71 (2005) 046141.
[33] Z.-M. Gu, T. Zhou, B.-H. Wang, et al., arXiv: cond-mat/0505175.