Empirical analysis of Internet telephone network: From user ID to phone

Qi Xuan,[a)] Fang Du, and Tie-Jun Wu
Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

(Received 11 November 2008; accepted 19 March 2009; published online 20 April 2009)

In order to study the interaction between different communication networks, in this paper, personal computer (PC)-to-phone log data in the year of 2007 are collected from UUCALL database and described as an ID-to-phone bipartite network (ItPBN). The ItPBN contains one giant component (GC) and a large number of satellitic components (SCs), both of which are carefully analyzed. It is found that the ItPBN has power-law incoming/outgoing degree distributions as well as a power-law clustering function (by proposing a new definition of clustering coefficient) indicating a hierarchical and modular structure of the ItPBN. Furthermore, the fact that most of the weak links always surrounding those ID nodes of large degree in the GC suggests that weak links may be more important to keep the structure of the GC than those strong ones. More interestingly, it is also revealed that there is strong correlation between many statistical properties of different SCs and their size, these extra information may be very useful in modeling the ItPBN in the future. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3116163]

In this paper, about 3.65×10^7 PC-to-phone call records in the year of 2007 are collected from UUCALL (the most famous Internet telephone software in China) database, and an ID-to-phone bipartite network (ItPBN) is derived. Studying the structure of such network is of much importance because it can help us better understand the interaction between different communication networks, which, however, is rarely mentioned by many other works in this area. Particularly, the ItPBN contains one giant component (GC) and a large number of small components named as satellitic components (SCs). It is found that such network shows similar characteristics as many other published real-world complex networks, i.e., scale-free, a power-law clustering function indicating a hierarchical and modular structure of the ItPBN, as well as a relatively small average shortest path length of the GC. Furthermore, the fact that most of the weak links always surrounding those ID nodes of large degree in the GC suggests that weak links may be more important to keep the structure of the GC than those strong ones. Additionally, more interesting phenomena are revealed through analyzing those SCs in the ItPBN, e.g., it seems that once the density of a SC reflected by the average weighted degree exceeds the density of the GC, this SC will be magnetized into the GC with a quite high probability. Such phenomena may be also existed in many other real-world complex networks although they have not been revealed yet.

I. INTRODUCTION

Uncovering the structures and evolving rules of social communication networks is of much importance in today's sociology although there is practical difficulty of mapping out interactions among a large number of individuals.[1–5] Fortunately, nowadays more and more human interactions are recorded with the help of advanced communication technologies, i.e., telephone,[6,7] electronic mail,[8,9] blog,[10,11] and so on, which offers unprecedented opportunities to reveal and explore the large-scale characteristics of social communication networks. Through analyzing those recorded data, the topologies of communication networks themselves can be easily figured, e.g., Onnela *et al.*[6] gave a mobile call network through analyzing 18 weeks of all mobile phone call records among ≈20% of a country's entire population, Guimerà *et al.*[9] provided an electronic mail network through analyzing electronic mail records among 1669 users in the University at Rovira i Virgili (URV) in Tarragona, Spain, while Ali-Hasan and Adamic[10] presented three blog networks through analyzing three blog communities, i.e., Kuwait Blogs, Dallas/Fort Worth Blogs, and United Arab Emirates Blogs.

However, most of these researches always focus on just one target communication network and pay little attention to the interaction between different communication networks, which must be inappropriate in such a strongly connected world.[12] Traditionally, one can communicate with others through PC-to-PC (e.g., MSN network) or phone-to-phone connections (e.g., telephone network), while the internet telephone, as a new type of communication method developed in 1996, allows people make calls through PC-to-phone connections.[13] In this paper, we will mainly focus on such intercommunication records from PC to phone just because this type of data has not been studied before and the structure of the resultant bipartite network may provide extra information to understand the interaction between different communication networks.

Particularly, we collect about 3.65×10^7 PC-to-phone call records in the year of 2007 from UUCALL (the most famous internet telephone software in China[14]) database and derive an ID-to-phone bipartite network (ItPBN) which contains one GC and a large number of small components

[a)]Author to whom the correspondence should be addressed. Electronic mail: crestxq@hotmail.com.

named as SCs in this paper. It is found that the ItPBN has power-law incoming/outgoing degree distributions as well as a power-law clustering function (by proposing a new definition of clustering coefficient) indicating a hierarchical and modular structure of the ItPBN. At the same time, it is also revealed that most of the weak links always surrounds those ID nodes of large degree in the GC, which may partly confirm the standpoint of the recent work of Onnela *et al.* that weak links may be more important than strong ones on sustaining the structure of the network. Additionally, we also carefully analyze the characteristics of those SCs which are always neglected in many other works of the same area. Such analysis may help us reveal more interesting phenomena and better understand the hierarchical and modular structure of the GC in the ItPBN.

The rest of the paper is organized as follows. In Sec. II, the background of the internet telephone is introduced and PC-to-phone log data are collected from UUCALL database and described as the ItPBN. In Sec. III, several classical characteristics of the GC in the ItPBN are analyzed and a new definition of clustering coefficient for bipartite networks is proposed. Then in Sec. IV, various statistical properties of the SCs in the ItPBN are carefully studied and some interesting phenomena are revealed. Finally, the paper is concluded in Sec. V.

II. BACKGROUND AND DATA DESCRIPTION

Convenient large-scale communication dates as far back as Alexander Bell and his invention of the telephone with the notion that one person can talk to another person far away using some kind of device which in 1876 is called telephone, whereas in 1996 it can be found on the internet, named as internet telephone, such as SKYPE, UUCALL, and so on.[14] Internet telephone is a new type of communication method based on voice-over-internet protocol (VoIP). Using such technique, one can make telephone calls using a broadband internet connection instead of a regular telephone line, thereby having phone service over the internet delivered through his internet connection, instead of from his local phone company.[15]

VoIP has evolved gradually since it came into existence as a result of work done by a few hobbyists in Israel in 1995. In the first stage, only PC-to-PC communication was in vogue, and in such situation, internet telephone network and the traditional telephone network are totally independent communication networks, i.e., there is no intercommunication between them. Then in 1998, PC-to-phone service is offered, and phone-to-phone service by using a computer to establish the connection is soon followed.[14] In this stage, internet telephone began to quickly penetrate into the traditional telephone network due to its cheapness and convenience.

Here, we have little interest in studying the potential structures of the internet/traditional telephone networks themselves, but mainly focus on the intercommunication records (forming a bipartite network whose nodes are divided into two sets, and only the connection between two nodes from different sets is allowed) between them, with the reason that the structure of such bipartite network may provide extra information to understand the interaction between different communication networks. Particularly, the bipartite PC-to-phone network has two kinds of nodes, i.e., ID nodes (user IDs) and PN nodes (phone numbers). Each link from a user-ID node to a phone node represents that there were call records between them. The sketch map of such bipartite network is shown in Fig. 1.

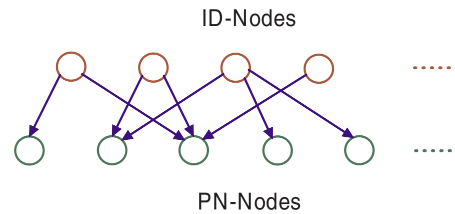


FIG. 1. (Color online) The sketch map of the bipartite network of VoIP's PC-to-phone service. There are two types of nodes, i.e., ID nodes and PN nodes. Each link and its weight represent there being call records from a user ID to a phone number and the accumulative duration of calls between them, respectively.

In this paper, PC-to-phone log data containing about 3.65×10^7 call records in the year of 2007 are collected from UUCALL database. It should be noted that these data are only used to study the interaction between the internet telephone network and the traditional telephone network and far from sufficient to capture the overall social communication network.[6] The resultant ID-to-phone bipartite network (ItPBN) contains about $N_{\text{in}}=1.21\times10^7$ PN nodes, $N_{\text{out}}=2.55\times10^6$ ID nodes, and $E=1.75\times10^7$ links, furthermore, each link has a weight representing the accumulative duration of calls from an ID node to a PN node. These nodes and links make up of one GC (69.2% nodes and 77.6% links) and 7.69×10^5 satellitic SCs (30.8% nodes and 22.4% links) which will be carefully studied in the following two sections.

III. GC ANALYSIS

In this section, we consider the GC of the ID-to-phone bipartite network. The local structures of the GC in the ItPBN around a randomly chosen PN node and a randomly chosen ID node are depicted in Fig. 2 left and right, respectively. Each network in Fig. 2 has 1000 nodes following a broad-first searching algorithm around the selected node. The figure partly shows that the GC in the ItPBN consists of small local clusters, typically grouped around several large degree nodes. Some basic statistical properties characterizing the GC are summarized in Table I.

Each PN node j has an incoming degree $k_{\text{in}}(j)$ denoting the number of different ID nodes around it, and similarly, each ID node i has an outgoing degree $k_{\text{out}}(i)$ denoting the number of different PN nodes around it. Furthermore, if the accumulative duration of calls from ID node i to PN node j is denoted as w_{ij} (minutes) and the total number of links is represented by E, then the average weight $\langle W\rangle$, the weighted incoming degree $s_{\text{in}}(j)$ of PN node j, the average weighted incoming degree $\langle s_{\text{in}}\rangle$, the weighted outgoing degree $s_{\text{out}}(i)$ of ID node i, and the average weighted outgoing degree $\langle s_{\text{out}}\rangle$ can be calculated by Eqs. (1)–(5), respectively,

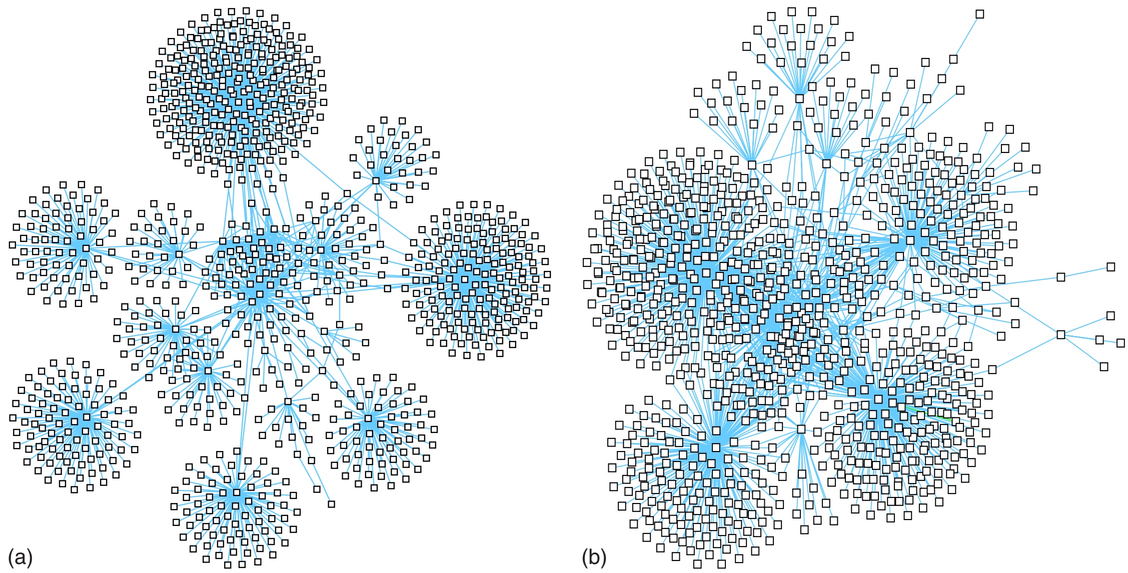


FIG. 2. (Color online) The local structures of the GC in the ItPBN around a randomly chosen PN node (left) and a randomly chosen ID node (right). Each network has 1000 nodes following a broad-first searching algorithm around the selected node.

$$\langle W\rangle = \frac{1}{E}\sum_{i=1}^{N_{\text{out}}}\sum_{j=1}^{N_{\text{in}}} w_{ij}, \quad (1)$$

$$s_{\text{in}}(j) = \sum_{i=1}^{N_{\text{out}}} w_{ij}, \quad (2)$$

$$\langle s_{\text{in}}\rangle = \frac{1}{N_{\text{in}}}\sum_{i=1}^{N_{\text{out}}}\sum_{j=1}^{N_{\text{in}}} w_{ij}, \quad (3)$$

$$s_{\text{out}}(i) = \sum_{j=1}^{N_{\text{in}}} w_{ij}, \quad (4)$$

$$\langle s_{\text{out}}\rangle = \frac{1}{N_{\text{out}}}\sum_{i=1}^{N_{\text{out}}}\sum_{j=1}^{N_{\text{in}}} w_{ij}. \quad (5)$$

It should be noted that all of these parameters will have the same form but with different values for different focused networks, e.g., $N_{\text{in}}=1.21\times 10^7$ for the overall ItPBN, whereas $N_{\text{in}}=8.82\times 10^6$ for the GC in the ItPBN. Additionally, we will keep this manner all over the paper.

In order to overcome the disadvantage of the traditional clustering coefficient[1] that it will lose its signification on bipartite networks due to their lack of triangles, Lind *et al.*[16] provided a new clustering coefficient to measure the density of squares around a node. It should be noted that, in many cases, for the convenience of directly showing the relations among a particular set of nodes or applying the usual measures in complex network theory, bipartite networks are always compressed by one-mode projection.[17–20] However, compression usually means the loss of information,[20] which is especially inappropriate in this paper where the interaction between different sets of nodes is mainly focused. As a result, we would like to define a more natural clustering coefficient for a bipartite network to measure the density of links between the neighbors of a node as usual just through expanding the concept of "neighbors" in the bipartite network.

Specially, in the ItPBN, there are two sets of nodes, i.e., ID nodes and PN nodes, hence each focused node i could be considered to have two types of neighbors, ID neighbors and PN neighbors. If the focused node i is an ID node, then its PN neighbors are those PN nodes directly connected to it and its ID neighbors are those other ID nodes directly connected to its PN neighbors, while if the focused node i is a PN node, then its ID neighbors are those ID nodes directly connected to it and its PN neighbors are those other PN nodes directly connected to its ID neighbors, as is shown in Fig. 3. Then the generalized clustering coefficient of node i in the ItPBN can be calculated by Eq. (6),

TABLE I. Several basic characteristics of the GC in the ItPBN. N_{in} is the number of PN nodes, N_{out} is the number of ID nodes, $\langle W\rangle$ is the average weight, $\langle k_{\text{in}}\rangle$ is the average incoming degree of PN nodes, $\langle k_{\text{out}}\rangle$ is the average outgoing degree of ID nodes, $\langle S_{\text{in}}\rangle$ is the average weighted incoming degree of PN nodes, $\langle S_{\text{out}}\rangle$ is the average weighted outgoing degree of ID nodes, while $\langle C\rangle$ is the average clustering coefficient and $\langle L\rangle$ is the average shortest path length of the corresponding undirected and unweighted bipartite network.

N_{in}	N_{out}	$\langle W\rangle$	$\langle k_{\text{in}}\rangle$	$\langle k_{\text{out}}\rangle$	$\langle s_{\text{in}}\rangle$	$\langle s_{\text{out}}\rangle$	$\langle C\rangle$	$\langle L\rangle$
8.82×10^6	1.34×10^6	9.86	1.54	9.65	15.14	95.45	0.83	8.50

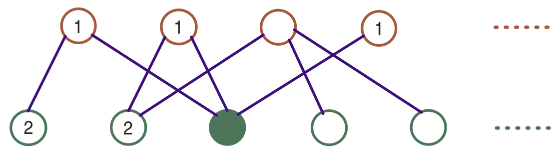


FIG. 3. (Color online) In this bipartite network, the filled node represents the target PN node, the nodes marked by "1" and the nodes marked by "2" denote the ID neighbors and PN neighbors of the target PN node, respectively.

$$C_i = \begin{cases} \dfrac{e_{12}(i)}{n_1(i) \times n_2(i)}, & n_1(i) \times n_2(i) \neq 0 \\ 0, & n_1(i) \times n_2(i) = 0, \end{cases} \tag{6}$$

where $n_1(i)$ is the number of ID neighbors of node i, $n_2(i)$ is the number of PN neighbors of node i, and $e_{12}(i)$ is the number of links between different types of neighbors. The clustering coefficient of a node defined by Eq. (6) can reflect the connection density of its two types of neighbors and is naturally normalized to [0,1] due to there being no links between the nodes of the same set in the ItPBN. How to generalize the clustering coefficient to weighted complex networks is more a scaling argument than a demonstration,[21–24] and the weighted clustering coefficient for weighted bipartite networks has even not been studied yet. So, as we believe, it is better to take a detailed discussion in another work, but not here due to the limitation of the paper length. The average clustering coefficient $\langle C \rangle$ can be calculated by Eq. (7),

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i, \tag{7}$$

where N represents the total number of nodes in the network. Table I shows that the GC in the ItPBN has much large average clustering coefficient and relatively small average shortest path length.

The GC in the ItPBN has very clear power-law incoming and outgoing degree distributions, i.e., $P(k_{\text{in}}) \sim k_{\text{in}}^{-\gamma_{\text{in}}}$ and $P(k_{\text{out}}) \sim k_{\text{out}}^{-\gamma_{\text{out}}}$, as is shown in Figs. 4(a) and 4(b), with the exponents $\gamma_{\text{in}}=2.9$ and $\gamma_{\text{out}}=2.1$, respectively. Besides, Fig. 4(c) shows the weighted incoming degree distribution which also has scale-free property, i.e., $P(s_{\text{in}}) \sim s_{\text{in}}^{-\gamma_{\text{in}}^s}$, with the exponent $\gamma_{\text{in}}^s=1.9$. The weighted outgoing degree distribution, however, behaves totally different from the above three distributions, as is shown in Fig. 4(d). In this figure, as we can see, most users make calls with the total duration close to 30 min and the critical point (30 min) partitions the ID nodes into two classes: the ID nodes with weighted outgoing degree smaller than the critical point almost present uniform weighted outgoing degree distribution, while the ID nodes with weighted outgoing degree larger than the critical point

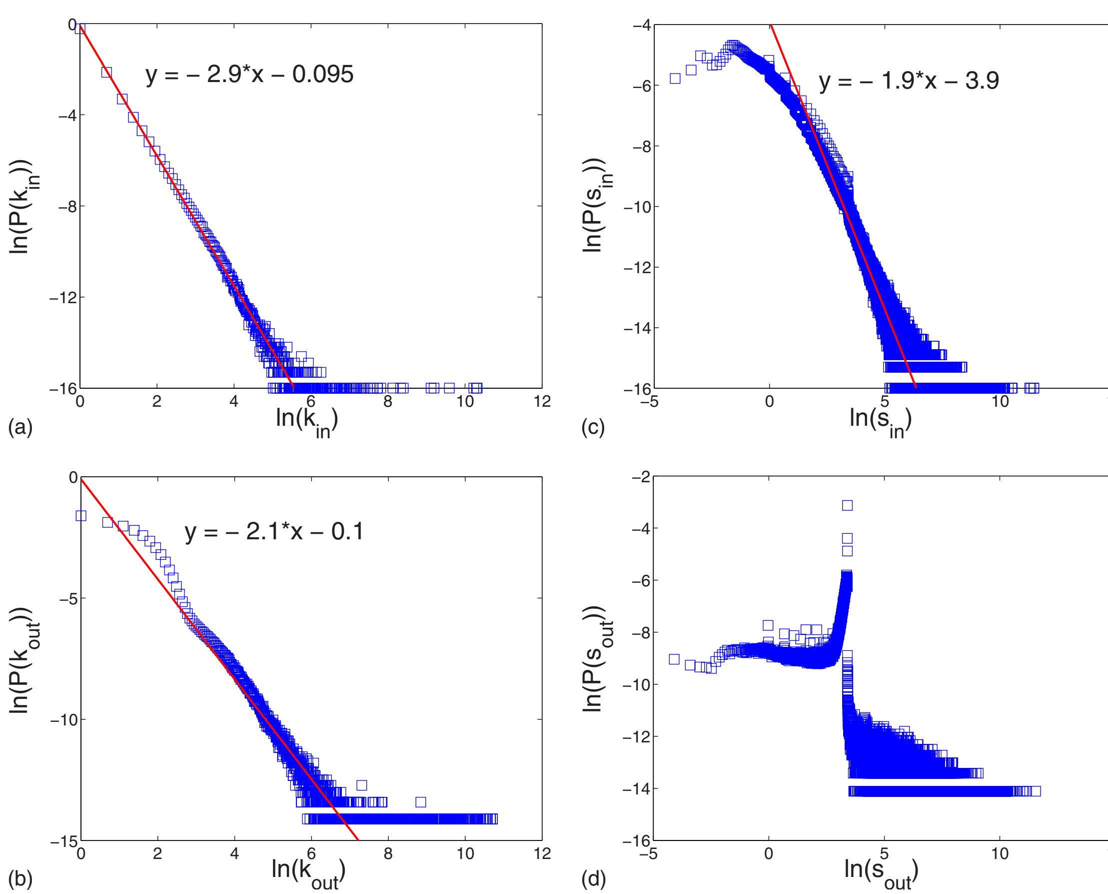


FIG. 4. (Color online) (a) The power-law incoming degree distribution of the GC, i.e., $P(k_{\text{in}}) \sim k_{\text{in}}^{-\gamma_{\text{in}}}$, with the exponent $\gamma_{\text{in}}=2.9$. (b) The power-law outgoing degree distribution of the GC, i.e., $P(k_{\text{out}}) \sim k_{\text{out}}^{-\gamma_{\text{out}}}$, with the exponent $\gamma_{\text{out}}=2.1$. (c) The power-law weighted incoming degree distribution of the GC, i.e., $P(s_{\text{in}}) \sim s_{\text{in}}^{-\gamma_{\text{in}}^s}$, with the exponent $\gamma_{\text{in}}^s=1.9$. (d) The weighted outgoing degree distribution of the GC with a critical point equal to 30 min call duration.

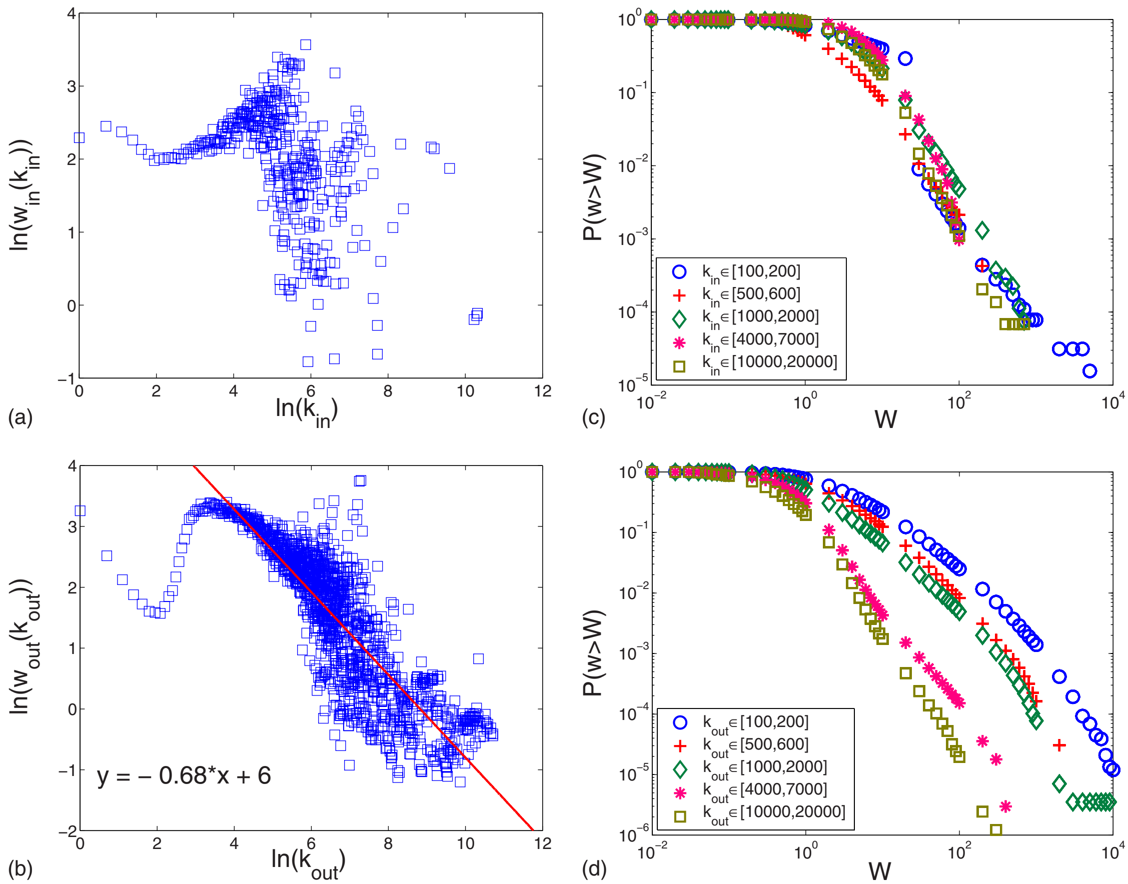


FIG. 5. (Color online) (a) The average $w_{\text{in}}(j)$ as a function $w_{\text{in}}(k_{\text{in}})$ of incoming degree k_{in} for PN nodes. (b) The average $w_{\text{out}}(j)$ as a function $w_{\text{out}}(k_{\text{out}})$ of outgoing degree k_{out} for ID nodes. The function almost presents a power-law property with the exponent equal to 0.68. (c) The link weight cumulative distributions $P(w>W)$ measuring the proportion of links with weight larger than W for five groups of links around the PN nodes with different degree ranges, i.e., $k_{\text{in}}\in[100,200]$, $k_{\text{in}}\in[500,600]$, $k_{\text{in}}\in[1000,2000]$, $k_{\text{in}}\in[4000,7000]$, and $k_{\text{in}}\in[10\,000,20\,000]$. (d) The link weight cumulative distributions $P(w>W)$ measuring the proportion of links with weight larger than W for five groups of links around the ID nodes with different degree ranges, i.e., $k_{\text{out}}\in[100,200]$, $k_{\text{out}}\in[500,600]$, $k_{\text{out}}\in[1000,2000]$, $k_{\text{out}}\in[4000,7000]$, and $k_{\text{out}}\in[10\,000,20\,000]$.

behave a little different, i.e., there is a smaller number of users taking calls with longer total durations. Such phenomenon is mainly caused by the 30 min free experience offered by UUCALL in the year of 2007. This finding reveals that users behave distinctly under different rates of charge. Such result may provide the evidence that the traffic in a real-world network could be controlled to a certain extent by some additional input. For instance, each node in the network could be attached by an award or penalty coefficient which represents what an individual will earn or loss when he passes it. Naturally, the routing strategy of an individual in the network could be influenced by these coefficients, in other words, the traffic in the network then could be controlled through adjusting the values of these additional coefficients.

The average weight of links around PN node j or ID node i is calculated by Eqs. (8) and (9), respectively,

$$w_{\text{in}}(j)=\frac{s_{\text{in}}(j)}{k_{\text{in}}(j)}, \tag{8}$$

$$w_{\text{out}}(i)=\frac{s_{\text{out}}(i)}{k_{\text{out}}(i)}. \tag{9}$$

Then the functions $w_{\text{in}}(k_{\text{in}})$ and $w_{\text{out}}(k_{\text{out}})$, defined as the average of $w_{\text{in}}(j)$ and $w_{\text{out}}(i)$ over the PN nodes with degree k_{in} and ID nodes with degree k_{out}, are shown in Figs. 5(a) and 5(b), respectively. As we can see, the correlation between the degree of PN nodes and their average weight of links are not very clear, as is shown in Fig. 5(a), while Fig. 5(b) shows a distinct trend that the ID nodes with larger degree have relatively lower average weight of links, particularly, the relationship between them almost presents a power-law property with the exponent equal to 0.68.

In order to provide a more credible result, five groups of links around the PN nodes/ID nodes with different degree ranges are selected and their link weight cumulative distributions $P(w>W)$ measuring the proportion of links with weight larger than W are presented in Figs. 5(c) and 5(d), respectively. Figure 5(c) shows that the link weight distributions around PN nodes with different degree ranges are very close, while Fig. 5(d) presents that, statistically, there is indeed a higher proportion of weak links around those ID nodes with larger degree, i.e., the distribution curves are arranged from top to down in the figure as the degree of ID nodes increases, although there is a quite large fitting error mean square error [(MSE)=0.6207] for the power-law tail of the function $w_{\text{out}}(k_{\text{out}})$, as is shown in Fig. 6. As a result, removal of those weak links may result in quickly vanishing of large degree ID nodes, which indicate a structural collapse of the bipartite network. This finding can partly confirm the

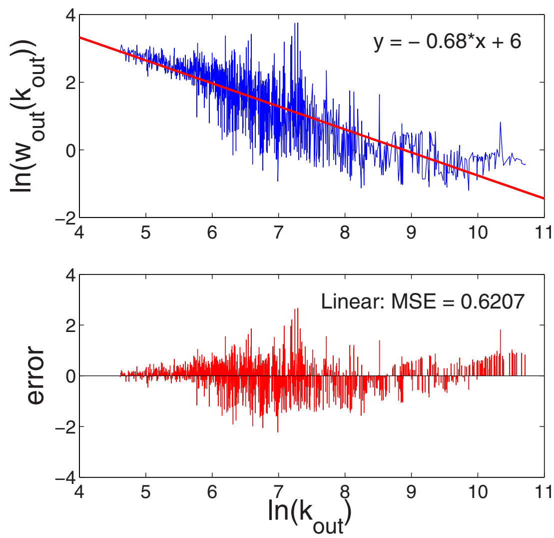


FIG. 6. (Color online) The fitted line (top) of the power-law tail ($k_{out} \geq 100$) shown in Fig. 5(b) satisfying that $\ln(w_{in}(k_{in}))=-0.68\ \ln(k_{in})+6$, as well as its fitting errors (bottom) with the MSE equal to 0.6207 for 1194 different values of k_{out} various from 100 to 44 938.

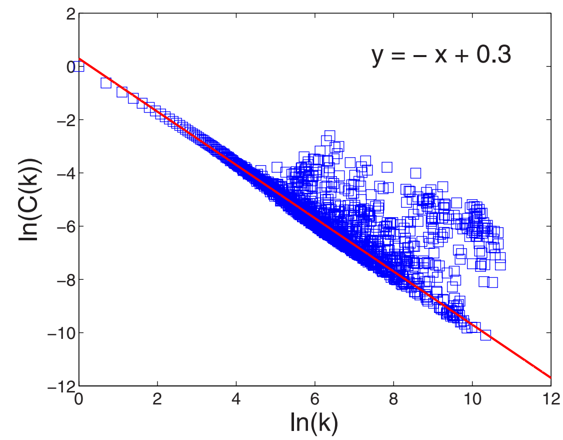


FIG. 7. (Color online) The power-law clustering function of the GC in the ItPBN, i.e., $C(k) \sim k^{-1}$. Such nontrivial scaling of clustering may be an indicator of hierarchical and modular structure of the GC.

standpoint of the recent work of Onnela *et al.*, i.e., weak links may be more important than strong ones on sustaining the structure of the network.

Interestingly, it is observed that, in Figs. 4(b) and 5(b), $w_{out}(k_{out})$ and $P(k_{out})$ both fluctuate slightly when k_{out} is relatively small. Such phenomena may be also due to the 30 min free experience offered by UUCALL in the year of 2007. In fact, denoting the functions $s_{in}(k_{in})=k_{in}\times w_{in}(k_{in})$ and $s_{out}(k_{out})=k_{out}\times w_{out}(k_{out})$ as the average of $s_{in}(j)$ and $s_{out}(i)$ over the PN nodes with degree k_{in} and ID nodes with degree k_{out}, respectively, it is found that $s_{out}(k_{out})$ is always close to 30 min when $k_{out}\leq 8$, i.e., most of ID nodes with degree $k_{out}\leq 8$ have accumulative durations of calls close to 30 min, thereby, $w_{out}(k_{out})$ drops as k_{out} increases from 1 to 8. However, it seems that 30 min may be not enough for a user ID to call more phone numbers, consequently, $s_{out}(k_{out})$ climbs up dramatically as k_{out} further increases, i.e., a user ID with larger k_{out} could be a consumer (a user ID making calls with an accumulative duration longer than 30 min) with a higher probability, which further leads to an increasing $w_{out}(k_{out})$ when $k_{out}>8$. As a result, $w_{out}(k_{out})$ has a local minimum at about $k_{out}=8$, as is shown in Fig. 5(b).

According to the clustering coefficient defined by Eq. (6), the clustering function can be intuitively defined as Eq. (10),[25]

$$C(k)=\frac{1}{N_k}\sum_{i=1}^{N}C_i(k), \tag{10}$$

where N is the total number of nodes, N_k is the number of nodes with degree k in the focused network, and $C_i(k)$ can be calculated by Eq. (11),

$$C_i(k)=\begin{cases}C_i, & k_i=k\\ 0, & k_i\neq k,\end{cases} \tag{11}$$

where k_i denotes the degree of node i. As is shown in Fig. 7, the clustering function of the GC in the ItPBN presents a power-law property, i.e., $C(k)\sim k^{-1}$, suggesting that ItPBN may possess a hierarchical and modular structure.[25–27]

IV. SCs ANALYSIS

Besides the GC containing most of the nodes, in a real-world network, there are always a large number of independent SCs, each of which only possess very small number of nodes.[6,9,17,28] Analyzing these SCs may provide extra information to study the evolution of the network and help us better understand the modular structure of the GC in the network. In fact, as the network grows, new SCs will be continuously created and most of which will finally join in the GC forming modules. However, it its real a pity that nowadays most of complex network researchers still pay little attention to these SCs. Like many other real-world networks, there are also a great number (about 7.69×10^5) of various SCs in the ItPBN which will be carefully studied in this section.

The structures of four independent SCs in the ItPBN with 100, 120, 167, and 224 nodes, respectively, are shown in Fig. 8. It is found that the fragmentized network, composed by all the SCs in the ItPBN, has similar structural properties as the GC of the ItPBN, i.e., power-law incoming/outgoing degree distributions with larger exponents $\gamma_{in}=3.6$ and $\gamma_{out}=4.0$, as is shown in Figs. 9(a) and 9(b), respectively, power-law weighted incoming degree distribution with similar exponent $\gamma_{in}^{s}=2.0$, as is shown in Fig. 9(c), similar weighted outgoing degree distribution with the same critical point (30 min), as is shown in Fig. 9(d), and similar power-law clustering function $C(k)\sim k^{-1}$, as is shown in Fig. 10.

Besides, it will be extra interesting to study the relationship between statistical properties of SCs and their size. First of all, the size distribution of SCs in the ItPBN can be defined by Eq. (12),

$$P(S)=\frac{N_{SC}(S)}{N_{SC}}, \tag{12}$$

where N_{SC} denotes the number of total SCs in the ItPBN and $N_{SC}(S)$ represents the number of SCs whose size is equal to S. Figure 11 shows the power-law SC size distribution of the

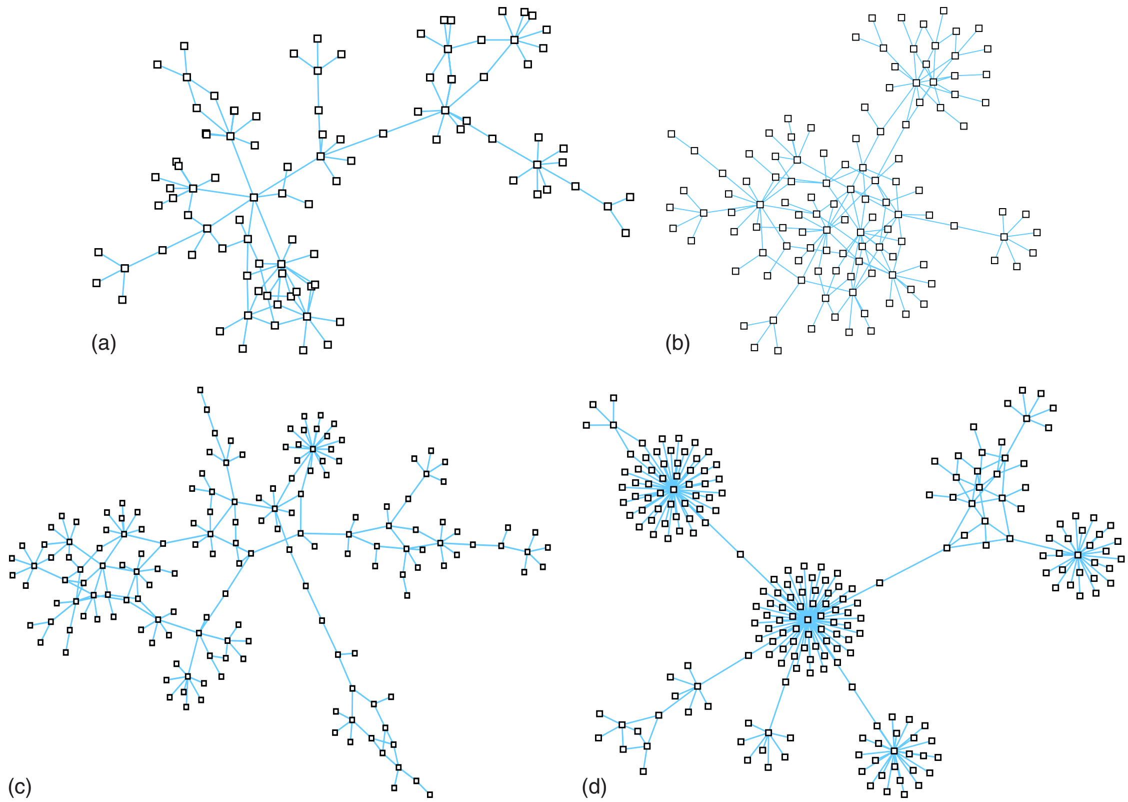


FIG. 8. (Color online) The structures of four independent SCs in the ItPBN with 100 nodes (top left), 120 nodes (top right), 167 nodes (bottom left), and 224 nodes (bottom right), respectively.

ItPBN, i.e., $P(S) \sim S^{-\theta}$, with the exponent $\theta=3.4$.

The average incoming degree of PN nodes $\langle k_{in} \rangle$, the average outgoing degree of ID nodes $\langle k_{out} \rangle$, and the average degree $\langle k \rangle$ of SCs all rise as the SC size S increases in the ItPBN, as is shown in Figs. 12(a)–12(c), respectively, which indicates that, generally, the frequency of communication between different types of nodes in a SC will be enhanced as the SC grows. In detail, Fig. 12(a) shows a power-law relationship between $\langle k_{in} \rangle$ and S, i.e., $\langle k_{in} \rangle \sim S^{\delta}$, with the exponent $\delta=0.17$, whereas Figs. 12(b) and 12(c) show that the relationships that $\langle k_{out} \rangle$ and $\langle k \rangle$ versus S follow the same pattern, i.e., $\langle k_{out} \rangle \sim \eta_1 \ln \ln S$ and $\langle k \rangle \sim \eta_2 \ln \ln S$, with different parameters $\eta_1=5.2$ and $\eta_2=0.73$. Due to the lack of undirected links in the ItPBN, in each SC, the number of outgoing links and the number of incoming links are both equal to the number of total links, as is presented in Eq. (13),

$$2 \times N_{in} \times \langle k_{in} \rangle = 2 \times N_{out} \times \langle k_{out} \rangle = S \times \langle k \rangle. \tag{13}$$

Then it can be easily inferred that the number of PN nodes N_{in} increases as a function of S following $N_{in} \sim (\eta_2/2)S^{1-\delta} \ln \ln S$, which is much slower than that of ID nodes N_{out} following $N_{out} \sim (\eta_2/2\eta_1)S$. Such phenomenon reveals the fact that ID nodes are the determinant force in the expand of the ItPBN.

The average weighted degree of a SC with its size equal to S, as is defined by Eq. (14),

$$\langle s \rangle = \frac{2}{S} \sum_{i=1}^{N_{out}} \sum_{j=1}^{N_{in}} w_{ij}, \tag{14}$$

can reflect some "density" of the SC. The relationship between such density and the SC size S is plotted in Fig. 13, where we can see that the densities of SCs reflected by their average weighted degree $\langle s \rangle$ slightly rise as the SC size increases, and there is an upper bound limiting such trend. Interestingly, this upper bound is very close to the average weighted degree $\langle s \rangle=25.73$ of the GC, which indicates that, in the ItPBN, once the density of the SC reflected by $\langle s \rangle$ exceeds the density of the GC, this SC will be magnetized into the GC with a quite high probability.

The average clustering coefficient $\langle C \rangle$ as a power-law function of the SC size S, i.e., $\langle C \rangle \sim S^{-\rho}$ with the exponent $\rho=0.03$ when $S>8$, is shown in Fig. 14, which suggests that the average clustering coefficient of a SC will slightly decrease even when there is an increasing trend of its average degree as the SC grows in the ItPBN. Such phenomenon is a little different from many existed findings and models confirming that the average clustering coefficient of a network with modular and hierarchical structure will not change much when its average degree keeps a constant as the network grows. Such difference may be attributed to the new definition of clustering coefficient in this paper.

Both of the relationships that the average shortest path length $\langle L \rangle$ and the diameter D versus the SC size S present

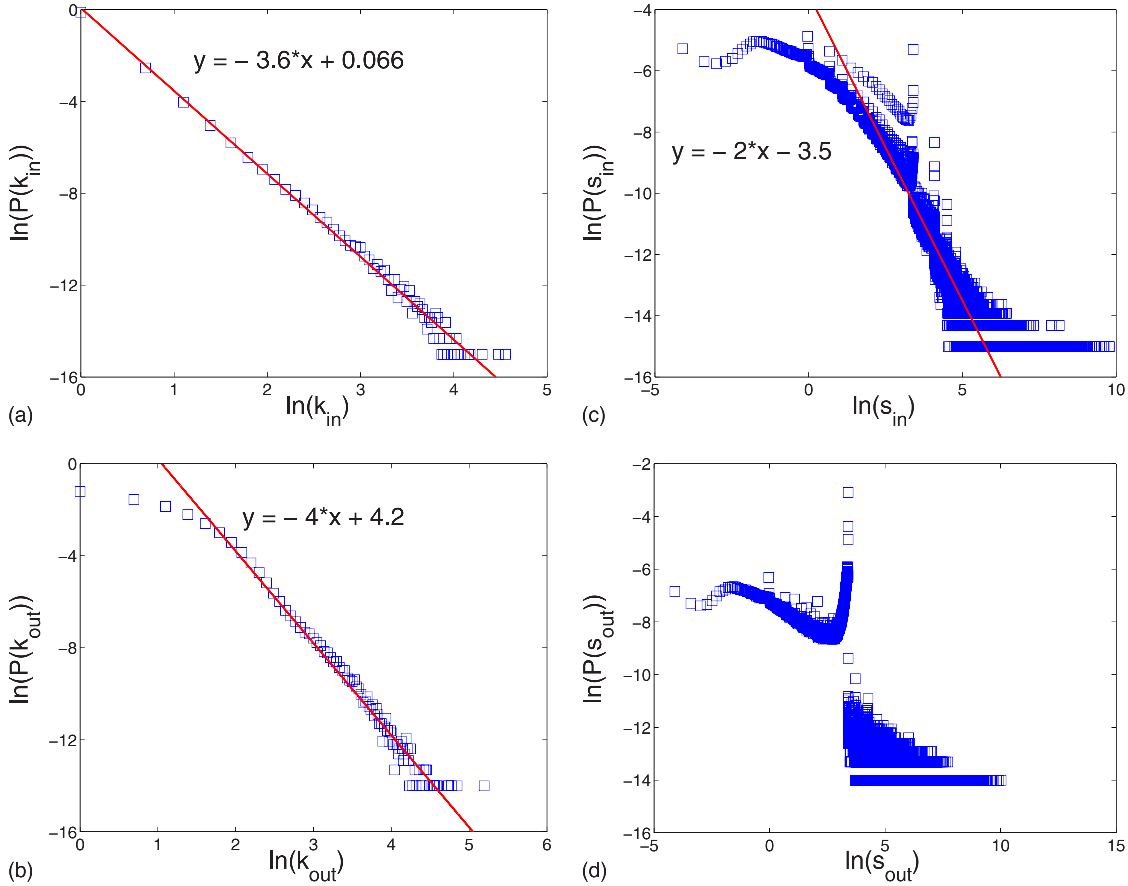


FIG. 9. (Color online) Four degree distributions of the fragmentized network composed by all the SCs in the ItPBN. (a) The power-law incoming degree distribution, i.e., $P(k_{\mathrm{in}}) \sim k_{\mathrm{in}}^{-\gamma_{\mathrm{in}}}$, with the exponent $\gamma_{\mathrm{in}}=3.6$. (b) The power-law outgoing degree distribution, i.e., $P(k_{\mathrm{out}}) \sim k_{\mathrm{out}}^{-\gamma_{\mathrm{out}}}$, with the exponent $\gamma_{\mathrm{out}}=4.0$. (c) The power-law weighted incoming degree distribution, i.e., $P(s_{\mathrm{in}}) \sim s_{\mathrm{in}}^{-\gamma_{\mathrm{in}}^{s}}$, with the exponent $\gamma_{\mathrm{in}}^{s}=2.0$. (d) The weighted outgoing degree distribution with the same critical point equal to 30 min call duration.

power-law properties, i.e., $\langle L\rangle \sim S^{\mu}$ and $D \sim S^{\nu}$, with the exponents $\mu=0.40$ and $\nu=0.56$, as is shown in Figs. 15(a) and 15(b) respectively. Traditionally, a network is considered to have small-world property if the average shortest path length of the network increases logarithmically or more slowly as the network grows with its average degree keeps a constant. According to this standard, the SCs in the ItPBN are not small-world networks. However, the GC in the ItPBN seems to possess a small-world property, i.e., having a very small average shortest path length $\langle L\rangle=8.50$ for a connected component with a huge number $N_{\mathrm{GC}}=1.02\times 10^{7}$ of nodes as well as a relative small average degree $\langle k\rangle=2.34$. The reason of

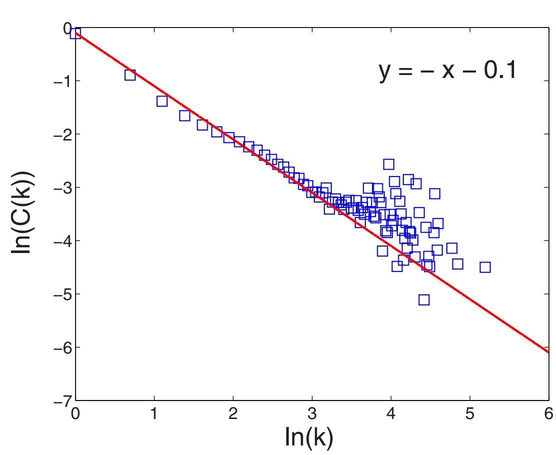


FIG. 10. (Color online) The clustering function of the fragmentized network composed by all the SCs in the ItPBN. As we can see, this clustering function also presents a power-law property, i.e., $C(k) \sim k^{-1}$, which suggests that hierarchical and modular structure may also exist in most of SCs in the ItPBN.

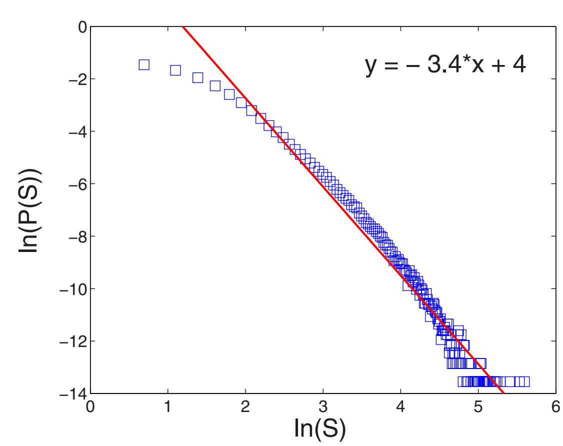


FIG. 11. (Color online) The power-law SC size distribution in the ItPBN, i.e., $P(S) \sim S^{-\theta}$, with the exponent $\theta=3.4$.

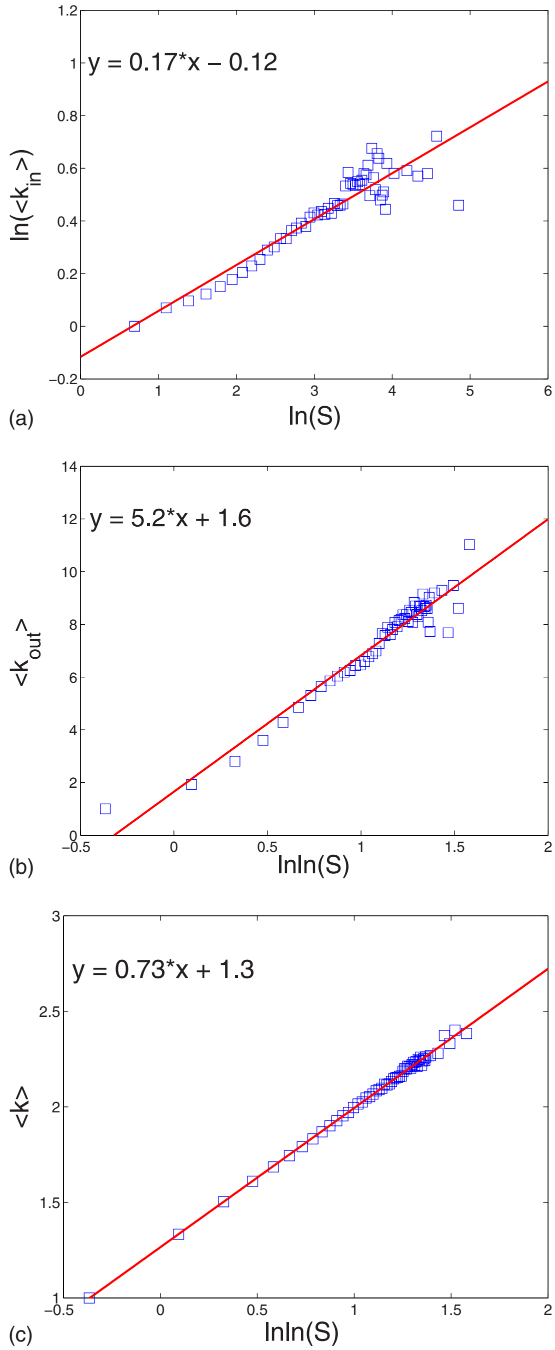


FIG. 12. (Color online) (a) The average incoming degree $\langle k_{\mathrm{in}}\rangle$ as a function of SC size S, the relationship between them follows $\langle k_{\mathrm{in}}\rangle \sim S^{\delta}$ with the exponent $\delta=0.17$. (b) The average outgoing degree $\langle k_{\mathrm{out}}\rangle$ as a function of SC size S, the relationship between them follows $\langle k_{\mathrm{out}}\rangle \sim \eta_1 \ln \ln S$ with the parameter $\eta_1=5.2$. (c) The average degree $\langle k\rangle$ as a function of SC size S, the relationship between them follows $\langle k\rangle \sim \eta_2 \ln \ln S$ with the parameter $\eta_2=0.73$.

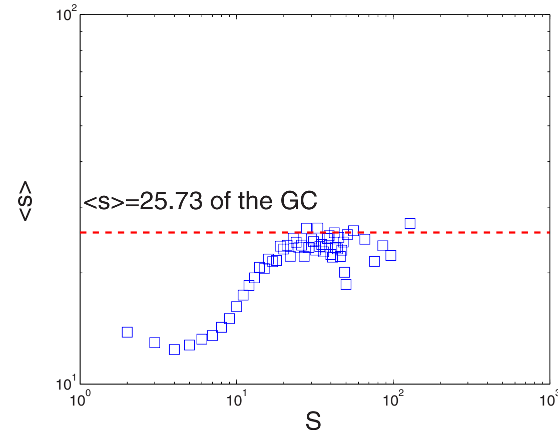


FIG. 13. (Color online) The relationship between the average weighted degree $\langle s\rangle$ and the SC size S. It is shown that the average weighted degree $\langle s\rangle=25.73$ of the GC in the ItPBN can be seemed as an upper bound of the average weighted degree of SCs.

V. CONCLUSION

In order to figure the interaction between different communication networks, 3.65×10^{7} PC-to-phone call records in the year of 2007 derived from UUCALL database have been described as an ItPBN in this paper. ItPBN contains two different types of nodes, i.e., PN nodes and ID nodes, forming one GC and a large number of SCs.

Generally, like many other published real-world networks, the ItPBN has power-law incoming/outgoing degree distributions as well as a power-law weighted incoming degree distribution. At the same time, it is revealed that the traffic in a real-world network could be significantly influenced by some additional input, e.g., the 30 min free experience offered by UUCALL in the year of 2007 causes a phase transition in the weighted incoming degree distribution of the ItPBN. Through proposing a new definition of clustering coefficient, we also find that the GC of the ItPBN has a much large average clustering coefficient and a power-law clustering function which may indicate a hierarchical and modular structure. The fact that most of the weak links always surrounding those ID nodes of large degree may suggest that such difference may be that there are some independent central nodes in the GC connecting different modules whose pre-existence are those SCs, as a result, the average shortest path length of the GC will not increase too quickly as it grows. In fact, there are indeed many independent service nodes, e.g., QQ (the most popular free communication tool in China) service phone, in the GC of the ItPBN. These service nodes may play very important roles in the evolution of the ItPBN.

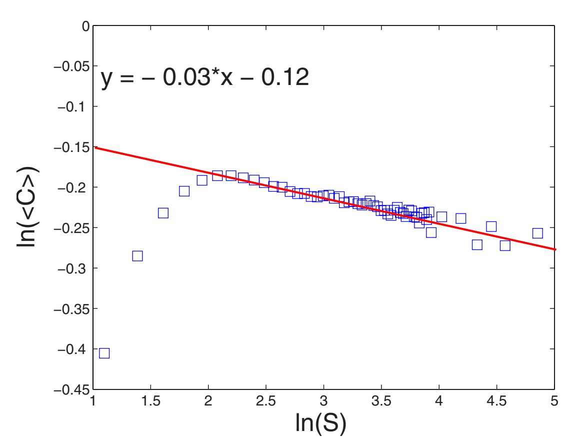
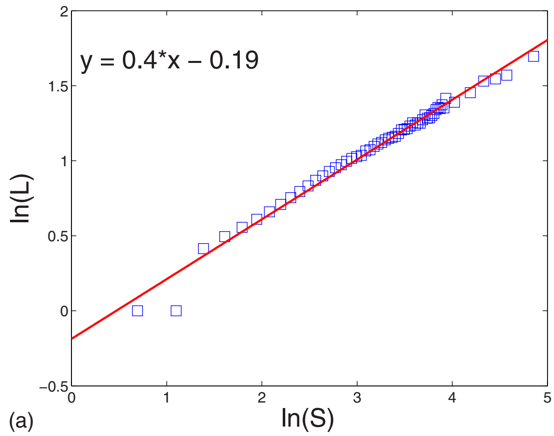


FIG. 14. (Color online) The average clustering coefficient $\langle C\rangle$ as a function of SC size S. The relationship presents a power-law property, i.e., $\langle C\rangle \sim S^{-\rho}$ with the exponent $\rho=0.03$, when $S>8$.



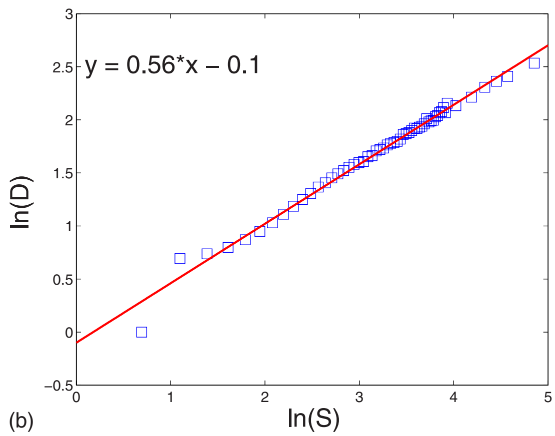


FIG. 15. (Color online) (a) The average shortest path length $\langle L\rangle$ as a function of SC size S. The relationship presents a power-law property, i.e., $\langle L\rangle \sim S^{\mu}$, with the exponent $\mu=0.40$. (b) The diameter D as a function of the SC size S. The relationship also presents a power-law property, i.e., $D\sim S^{\nu}$, with the exponent $\nu=0.56$.

weak links is more important to keep the structure of the GC in the ItPBN than those strong ones.

Besides, in this paper, characteristics of SCs in the ItPBN are also carefully studied, and several interesting results are derived. For examples, the SC size distribution in the ItPBN shows similar power-law property as the module size distribution revealed in several other real-world networks, and such result may provide optional explanation to the modular structure of these real-world networks. It is also found that ID nodes may be the determinant force in the expand of the ItPBN because the proportion of ID nodes is always higher in those SCs of larger size. More interestingly, in the ItPBN, it seems that once the density of a SC reflected by the average weighted degree exceeds the density of the GC, the SC will be magnetized into the GC with a quite high probability. Additionally, the possible small-world property of the GC may be attributed to many independent service nodes in it because the SCs themselves have no small-world property at all, i.e., different modules (in the GC) evolved from the SCs are connected with each other through these service nodes, as a result, the average shortest path length of the GC will not increase too quickly as it grows.

ACKNOWLEDGMENTS

We would like to thank all the members in our research group in the Department of Control Science and Engineering, Zhejiang University at Yuquan Campus, for the valuable discussions about the ideas presented in this paper. This work has been supported by China Postdoctoral Science Foundation (Grant No. 20080441256) and the National 973 Program of China (Grant No. 2002CB312200).

[1]D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2]A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3]S. H. Strogatz, Nature (London) **410**, 268 (2001).
[4]P. S. Dodds, R. Muhamad, and D. J. Watts, Science **301**, 827 (2003).
[5]S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature (London) **429**, 180 (2004).
[6]J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **104**, 7332 (2007).
[7]K. Dasgupta, R. Singh, and B. Viswanathan, EDBT'08, Nantes, France, 2008 (unpublished).
[8]M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. E **66**, 035101(R) (2002).
[9]R. Guimerà, L. Danon, A. D. Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103(R) (2003).
[10]N. F. A. Hasan and L. A. Adamic, ICWSM'07, Colorado, 2007 (unpublished).
[11]T. Furukawa, M. Ishizuka, Y. Matsuo, I. Ohmukai, and K. Uchiyama, *Analyzing Reading Behavior by Blog Mining* (AAAI, Vancouver, Canada, 2007).
[12]M. Kurant and P. Thiran, Phys. Rev. Lett. **96**, 138701 (2006).
[13]http://www.utdallas.edu/bjackson/history.html.
[14]http://www.uucall.com.
[15]http://www.utdallas.edu/bjackson/whatis.html.
[16]P. G. Lind, M. C. González, and H. J. Herrmann, Phys. Rev. E **72**, 056127 (2005).
[17]M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[18]M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).
[19]T. Zhou, B.-H. Wang, Y.-D. Jin, D.-R. He, P.-P. Zhang, Y. He, B.-B. Su, K. Chen, Z.-Z. Zhang, and J.-G. Liu, Int. J. Mod. Phys. C **18**, 297 (2007).
[20]T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang, Phys. Rev. E **76**, 046115 (2007).
[21]P. Grindrod, Phys. Rev. E **66**, 066702 (2002).
[22]A. Barrat and M. Barthélemy, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[23]J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, Phys. Rev. E **71**, 065103(R) (2005).
[24]J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertész, Phys. Rev. E **75**, 027105 (2007).
[25]E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).
[26]D.-H. Kim, G. J. Rodgers, B. Kahng, and D. Kim, Physica A **351**, 671 (2005).
[27]Q. Xuan, Y. Li, and T.-J. Wu, Phys. Rev. E **73**, 036105 (2006).
[28]A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).