

Partially ordered sets in complex networks

Qi Xuan¹, Fang Du and Tie-Jun Wu

Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, People's Republic of China

E-mail: crestxq@hotmail.com

Received 24 November 2009, in final form 5 March 2010

Published 15 April 2010

Online at stacks.iop.org/JPhysA/43/185001

Abstract

In this paper, a partial-order relation is defined among vertices of a network to describe which vertex is more important than another on its contribution to the connectivity of the network. A maximum linearly ordered subset of vertices is defined as a chain and the chains sharing the same end-vertex are grouped as a family. Through combining the same vertices appearing in different chains, a directed chain graph is obtained. Based on these definitions, a series of new network measurements, such as chain length distribution, family diversity distribution, as well as the centrality of families, are proposed. By studying the partially ordered sets in three kinds of real-world networks, many interesting results are revealed. For instance, the similar approximately power-law chain length distribution may be attributed to a chain-based positive feedback mechanism, i.e. new vertices prefer to participate in longer chains, which can be inferred by combining the notable preferential attachment rule with a well-ordered recommendation manner. Moreover, the relatively large average incoming degree of the chain graphs may indicate an efficient substitution mechanism in these networks. Most of the partially ordered set-based properties cannot be explained by the current well-known scale-free network models; therefore, we are required to propose more appropriate network models in the future.

PACS numbers: 89.75.Hc, 89.75.Fb, 02.10.Ab

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Over the past decade, characterizing and modeling complex networks attract many attentions from various areas [1–8]. Interestingly, it is found that lots of real-world complex networks

¹ Author to whom any correspondence should be addressed.

share several common statistical properties [9], such as small-world [1], scale-free [2], power-law clustering function [3], self-similarity [10–12], symmetry [13, 14], etc. In order to explain these properties, a large number of models have been proposed [1–3, 15–18]. For instance, Watts and Strogatz (WS) [1] proposed a simple small-world network model by introducing randomness into a regular network through a rewiring process, and the first scale-free network is proposed by Barabási and Albert (BA) [2] through introducing the preferential attachment (PA) rule into the network growing process. Most of these traditional network properties are based on the single-vertex or pair-vertices measurements, e.g. the degree and the clustering coefficient of a vertex or the symmetry and the shortest path length between a pair of vertices, etc. These types of properties are far from sufficient when the interaction between vertices is especially emphasized in complex networks.

Recently, the basic structural motifs recurring frequently in complex networks were carefully studied [19, 20], and it seems that the rank distributions of these motifs can be used to classify networks [20], e.g. real-world networks always present significantly larger numbers of these motifs compared with randomized networks. This finding is of much interest because it suggests that motifs rather than vertices may be the basic cells to perform the higher level functions of the systems. However, search of large motifs, if any, seems a little blindfold and sometimes is very time-consuming because there must be a mass of different connected motifs of large size. Therefore, in large-scale social networks, researchers are more likely to directly reveal the modules defined by groups (communities) of vertices within which connections are denser than among them [9, 21, 22]. It is widely believed [9] that vertices in the same community are inclined to share common properties and dynamics. However, the ambiguous definition [21] of community produces an inevitable result that every union of communities is also a community. In such a situation, a hierarchy among the communities has to be always assumed *a priori* to overcome such limitation [9, 23].

In this paper, we would like to provide an optional method to rank and classify vertices more precisely based on their contributions to the connectivity of the target network. Firstly, we will introduce several definitions about relations in discrete mathematics [24]. There are many kinds of relationships in the world. A network itself defines a relation among vertices, i.e. two vertices are related if they are connected. And in many cases we are more interested in order relations which could tell us when an element is ‘smaller than’ or ‘preceding’ another. A partially ordered set then could be defined as a set together with a relation describing that one of the elements must precede the other for certain pairs of elements in the set. It should be noted that, different from a linear order, in a partially ordered set some pairs of elements may not be related to each other. A familiar real-world example of a partially ordered set is a collection of people ordered by genealogical descendance [25]. Some pairs of people bear the ancestor–descendant relationship, but other pairs do not.

Generally, in a network, vertices sharing several same neighbors could be compared with each other, i.e. a vertex could be considered to precede another one if each of its neighbors is also the neighbor of its posterior. In fact, a vertex can be completely replaced by its posterior on its contribution to the connectivity of the network. The vertex set together with such relation forms a partially ordered set in the network. A maximum linearly ordered subset of vertices then is defined as a chain and the chains sharing the same end-vertex are grouped as a family. Measurements of chains and families may provide ultra information to understand the intrinsic mechanisms of real-world complex networks and thus are carefully studied in this paper. The results are interesting and challenging because most of the common quantitative and qualitative partially ordered set-based properties revealed in several real-world networks, such as the similar approximately power-law chain length distribution, cannot be explained by current well-known network models [2, 17, 18]. Therefore,

more appropriate network models should be provided in the future to explain these new properties.

The rest of the paper is organized as follows. In the next section, partially ordered set and other related concepts, e.g. chain and family, are defined in a network and a brief algorithm is provided to obtain them, at the same time, the tested data sets are also introduced here. Measurements of chains and families in several real-world and artificial complex networks are carefully studied in sections 3 and 4, respectively. The work is summarized in section 5.

2. Partially ordered set in a network

2.1. Definitions

A network is denoted by $G = G(V, E)$, where V is the set of vertices and $E \subseteq V \times V$ is the set of edges. Vertices v_i and v_j are adjacent if $(v_i, v_j) \in E$. A subset of vertices $U_i \subseteq V$ is named as the neighbor set of v_i if each vertex in U_i , while none of the vertices in $V \setminus U_i$, is adjacent to v_i . By these notations, a relation R , denoted by the symbol ' \preceq ', on the vertex set V can be defined as

$$v_i \preceq v_j \Leftrightarrow U_i \subseteq U_j. \tag{1}$$

Such a relation \preceq on the vertex set V is *reflexive* and *transitive* because equations (2) and (3) must be always satisfied:

$$v_i \preceq v_i, \tag{2}$$

$$v_i \preceq v_j, v_j \preceq v_k \Rightarrow v_i \preceq v_k. \tag{3}$$

Furthermore, if we consider that v_i and v_j are equal or, more specific, symmetric, denoted by $v_i = v_j$, when $U_i \subseteq U_j$ and $U_j \subseteq U_i$ are both satisfied, then the relation \preceq on the vertex set V is also *antisymmetric* represented by

$$v_i \preceq v_j, v_j \preceq v_i \Rightarrow v_i = v_j. \tag{4}$$

It should be noted that here $v_i = v_j$ just means that the two possible different vertices are equal on their neighbor sets, i.e. $U_i = U_j$. Thus, the relation \preceq can be called a *partial order* on the vertex set V , and the vertex set V together with the partial order \preceq is called a *partially ordered set* [24]. It is worthy of note that the partial order defined between two vertices completely depends on the relationship between their neighbor sets; as a result, it is somewhat related to the degree, i.e. the number of neighbors, of vertices, i.e. denoting the degree of v_i by k_i , equation (5) must be satisfied:

$$v_i \preceq v_j \Rightarrow k_i \leq k_j. \tag{5}$$

A *chain* in such a partially ordered set is denoted by a maximum linearly ordered subset $\theta_i = \{v_{i_1}, v_{i_2}, \dots, v_{i_l}\} \subseteq V$ satisfying equation (6):

$$v_{i_1} \preceq v_{i_2} \preceq \dots \preceq v_{i_l}, \tag{6}$$

at the same time, for each vertex v_j in the set $V \setminus \theta_i$, the vertex subset $\theta_i \cup v_j$ cannot be linearly ordered. For instance, the network in figure 1(a) has two chains $v_a \preceq v_b \preceq v_c \preceq v_d$ and $v_i \preceq v_j \preceq v_k \preceq v_l$. It should be noted that both the subsets $\{v_a, v_b, v_c\}$ and $\{v_a, v_b, v_c, v_d, v_i\}$ in the figure are not chains because the former is not a 'maximum' linearly ordered subset and the latter cannot be linearly ordered at all.

Naturally, each chain has an end-vertex, e.g. v_{i_l} in equation (6); the end-vertex set of a network is denoted by V_E containing all the end-vertices of the chains in the network,

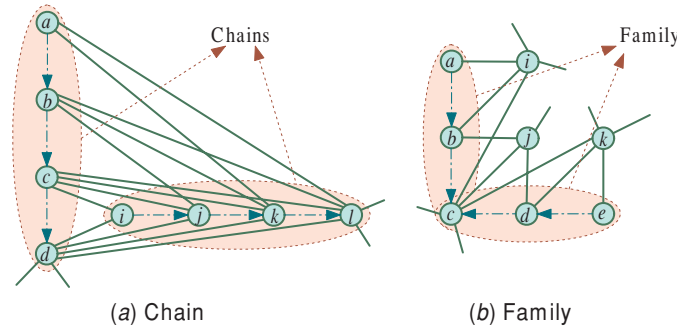


Figure 1. (a) Following the definition of the partial relation \preceq , the network has two chains $v_a \preceq v_b \preceq v_c \preceq v_d$ and $v_i \preceq v_j \preceq v_k \preceq v_l$. It should be noted that both the subsets $\{v_a, v_b, v_c\}$ and $\{v_a, v_b, v_c, v_d, v_i\}$ are not chains here just because the former one is not a ‘maximum’ linearly ordered subset and the latter one cannot be linearly ordered at all. (b) Two chains $v_a \preceq v_b \preceq v_c$ and $v_e \preceq v_d \preceq v_c$ sharing the same end-vertex v_c are grouped as a family $\{v_a, v_b, v_e, v_d, v_c\}$.

e.g. $V_E = \{v_d, v_l\}$ for the network shown in figure 1(a). Furthermore, different chains in a network may share the same end-vertex, and the chains $\theta_{i_j}, j = 1, 2, \dots, \psi_i$, sharing the same end-vertex v_i are grouped as a *family* ϕ_i in the network denoted by

$$\phi_i = \bigcup_{j=1}^{\psi_i} \theta_{i_j}. \tag{7}$$

So there will be totally $|V_E|$ families, and each family corresponds to an end-vertex in the network. For instance, in figure 1(b), the chains $v_a \preceq v_b \preceq v_c$ and $v_e \preceq v_d \preceq v_c$ sharing the same end-vertex v_c can be grouped as a family $\{v_a, v_b, v_e, v_d, v_c\}$.

The main contribution of a vertex in a network is determined by its hinge role in connecting other vertices, which is especially remarkable for those end-vertices in the network. In fact, through the above definitions, we can get the following theorem:

Theorem 1. *Removal of any vertex in a chain except the end-vertex will not influence the average shortest path length of the remaining subnetwork, that is, the end-vertex of a chain can completely replace the other vertices in the same chain on their contributions to the connectivity of the network.*

Theorem 1 can be easily proven with the fact that, considering a chain denoted by equation (6), $U_{i_j} \subseteq U_{i_l}$ must be satisfied for each $j \leq l$, which suggests that each shortest path passing by the vertex v_{i_j} can change its route to pass through the vertex v_{i_l} without increasing its length. It should be noted that, if all the neighbors of v_i are also the neighbors of v_j , i.e. $v_i \preceq v_j$, the betweenness centrality [9] of v_i , denoted by C_i , must not be larger than that of v_j , as is represented by

$$v_i \preceq v_j \Rightarrow C_i \leq C_j. \tag{8}$$

Therefore, an end-vertex must have the largest betweenness centrality in its family.

2.2. The algorithm

Based on the above definitions, here, we will briefly introduce our method to find all the chains in a network. In the procedure, the chain graph is constructed at first, then the chains

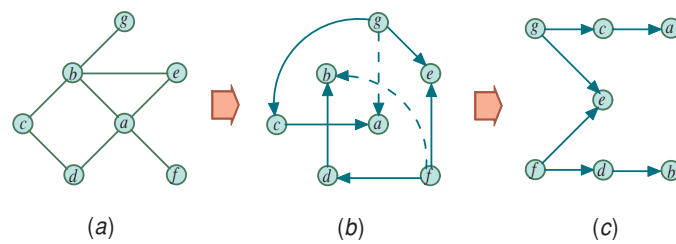


Figure 2. (a) A simple network with seven vertices. (b) The partial-ordered relations among these vertices. Each pair of ordered vertices are connected by a directed edge, and the dotted edges are redundant. (c) The chain graph obtained by dropping those redundant edges.

are extracted by a depth first search algorithm in the chain graph, and the families can be obtained by simply combining these chains. Particularly, the algorithm is summarized by the following three steps.

- (1) *Build chain graph.* For each pair of vertices v_i and v_j sharing at least one common neighbor, determine if they could be ordered, i.e. $v_i \leq v_j$ or $v_j \leq v_i$, by comparing their neighbor sets. Compress those symmetric vertices sharing exact same neighbors, i.e. $v_{k_1} = v_{k_2} = \dots = v_{k_p}$ ($k_1 < k_2 < \dots < k_p$), as one vertex v_{k_1} . Then a compressed digraph can be derived by considering that each pair of ordered vertices $v_i \leq v_j$ are connected by a directed edge pointing from v_i to v_j . Reveal all the triangle motifs in the compressed digraph, e.g. $v_i \leq v_j$, $v_j \leq v_k$, and $v_i \leq v_k$, and remove all the redundant edges, e.g. $v_i \leq v_k$, simultaneously. Replace each compressed vertex v_{k_1} by a directed chain $v_{k_1} \rightarrow v_{k_2} \rightarrow \dots \rightarrow v_{k_p}$; at the same time, the incoming edges of the compressed vertex v_{k_1} still point to the beginning vertex v_{k_1} in the directed chain, while the outgoing edges of the compressed vertex v_{k_1} are now originated from the last vertex v_{k_p} . Thus, the chain graph is constructed. A simple example is shown in figure 2.
- (2) *Extract chains.* In the chain graph, there are three types of vertices: the original-vertices with no incoming edges, the end-vertices with no outgoing edges and the middle-vertices with at least one incoming edge and one outgoing edge. Then all the chains could be revealed by a depth first search algorithm in the chain graph: begin with the original-vertices, along the outgoing edges, till reaching the end-vertices.
- (3) *Combine to families.* When all the chains are collected, each family can be obtained by simply combining all the chains sharing the same end-vertex.

2.3. Data sets

We mainly study the partially ordered sets in three kinds of real-world networks. The first is the protein–protein interaction networks collected from *BioGrid* database [26], where each vertex represents a protein or a gene, and each edge denotes the interaction between these proteins or genes. We use the giant components of four protein–protein interaction networks, corresponding to four species including *Caenorhabditis elegans* (CAE), *Saccharomyces cerevisiae* (SAC), *Homo sapiens* (HOM) and *Drosophila melanogaster* (DRO), derived by the two-hybrid experimental system. The second is the autonomous system (AS) relationship networks collected from *CAIDA* database [27], where each vertex represents an autonomous system, and each edge denotes a certain relationship between two autonomous systems, such as provider–customer, peer-to-peer, or sibling-to-sibling. We use six AS relationship networks captured at different times from 2004 to 2009, i.e. 2004 (5 Jan.), 2005 (3 Jan.), 2006 (2 Jan.),

Table 1. The basic properties including the number of vertices N , the average degree $\langle k \rangle$ and the average clustering coefficient $\langle C \rangle$, for the three kinds of real-world networks and the artificial networks derived by the three models. All the networks have a power-law degree distribution $P(k)$ and almost all the networks except those derived by the BA model have a power-law clustering function $C(k)$.

Networks	N	$\langle k \rangle$	$\langle C \rangle$
<i>BioGird</i>			
CAE	2575	3.34	2.11×10^{-2}
SAC	3085	5.00	2.08×10^{-1}
HOM	4583	3.96	3.48×10^{-2}
DRO	6853	6.43	3.47×10^{-2}
<i>CAIDA</i>			
040105	16 301	4.04	2.33×10^{-1}
050103	18 740	4.11	2.47×10^{-1}
060102	20 542	3.89	2.24×10^{-1}
070101	24 013	4.11	2.20×10^{-1}
080107	26 960	4.03	2.04×10^{-1}
090105	30 610	4.48	2.37×10^{-1}
<i>Douban</i>			
Friendship	504 912	11.3	9.23×10^{-2}
Model			
BA	4000	4.00	$(9.68 \pm 1.96) \times 10^{-3}$
LW	4000	4.00	$(3.10 \pm 0.08) \times 10^{-1}$
DMS	4000	4.00	$(7.39 \pm 0.02) \times 10^{-1}$

2007 (1 Jan.), 2008 (7 Jan.), 2009 (5 Jan.). The third is the friendship network crawled on 18 March 2009 from *Douban* database [28], where each vertex represents a *Douban* user, and each edge denotes an online friendship between two users.

All of these real-world networks have both power-law degree distribution and approximately power-law clustering function which have been well explained by many published models. For instance, The power-law degree distribution can be easily explained by the BA model [2] which could be further improved to explain the relatively high average clustering coefficient and the power-law clustering function through applying the PA rule in a predefined local world of a newly added vertex. Such improved network models are named as local-world (LW) models [15–17]. Besides, Dorogovtsev, Mendes, and Samukhin (DMS) [18] proposed a more straightforward scale-free network model. In the DMS model, a newly added vertex is connected to both ends of a randomly selected edge at every time step. This very simple model can explain both the power-law degree distribution and the power-law clustering function very well. In order to provide comparative results, we also study the partially ordered sets in the artificial networks created by the three models, i.e. the BA model, the LW model proposed by Xuan *et al* [17] and the DMS model. For each model, we create 100 networks with the same number of vertices and edges, and the mean value as well as the standard deviation of the other network properties will be recorded. Several basic properties, including the number of vertices N , the average degree $\langle k \rangle$ and the average clustering coefficient $\langle C \rangle$, of the tested real-world and artificial complex networks are presented in table 1, while the partially ordered set-based properties are recorded in table 2.

Table 2. The partially ordered set-based properties, i.e. the number of chains N_θ , the average chain length $\langle L_\theta \rangle$, the average incoming degree $\langle k_{in} \rangle$ of the chain graph, the number of families N_ϕ , the average family diversity $\langle \psi \rangle$, the average family size $\langle L_\phi \rangle$ and the average family overlapping size $\langle L_\phi^o \rangle$, for the three kinds of real-world networks and the artificial networks derived by the three models.

Networks	N_θ	$\langle L_\theta \rangle$	$\langle k_{in} \rangle$	N_ϕ	$\langle \psi \rangle$	$\langle L_\phi \rangle$	$\langle L_\phi^o \rangle$
<i>BioGrid</i>							
CAE	3 317	10.5	1.55	1044	3.18	24.1	9.39×10^{-1}
SAC	5 013	5.35	1.69	1859	2.70	12.1	3.67×10^{-1}
HOM	7 218	5.00	1.64	2443	2.95	11.0	1.35×10^{-1}
DRO	17 624	3.39	2.66	4444	3.97	10.3	8.56×10^{-2}
<i>CAIDA</i>							
040105	10 916	31.8	2.68	3647	3.60	62.4	8.28×10^{-1}
050103	12 991	30.8	2.80	4140	3.77	64.3	7.94×10^{-1}
060102	13 496	36.2	2.61	4504	3.56	68.0	7.95×10^{-1}
070101	16 591	31.3	2.82	5319	3.78	61.0	5.43×10^{-1}
080107	18 028	33.2	2.73	5926	3.78	61.8	4.98×10^{-1}
090105	22 431	30.4	3.12	6808	4.05	63.9	5.21×10^{-1}
<i>Douban</i>							
Friendship	2646 508	6.30	5.72	316 774	8.40	45.5	6.15×10^{-2}
Model							
BA (mean)	3933.9	1.05	5.46×10^{-2}	3831.6	1.03	1.06	1.26×10^{-5}
BA (std)	8.79	7.77×10^{-3}	7.55×10^{-3}	19.1	5.19×10^{-3}	8.27×10^{-3}	7.01×10^{-6}
LW (mean)	3204.6	1.54	4.60×10^{-1}	2746.6	1.17	1.63	1.88×10^{-4}
LW (std)	21.9	1.50×10^{-2}	1.07×10^{-2}	23.8	8.61×10^{-3}	2.02×10^{-2}	2.23×10^{-5}
DMS (mean)	3214.2	1.88	7.41×10^{-1}	2002.7	1.61	2.41	6.04×10^{-4}
DMS (std)	22.0	1.33×10^{-2}	1.03×10^{-2}	18.0	1.75×10^{-2}	3.19×10^{-2}	3.76×10^{-5}

3. Measurements of chains in complex networks

The vertices in a chain play similar roles in connecting other vertices in the network. In fact, a vertex can be completely replaced by its posterior in the same chain on its contribution to the connectivity of the network. So studying the properties of the chains in complex networks is of much scientific interest. The length of a chain θ_i is denoted by the number of vertices it contains, i.e. $l_\theta(i) = |\theta_i|$. Denoting the number of chains in a network by N_θ , the *average chain length* of the network by $\langle L_\theta \rangle$ and the *chain length distribution* expressing the fraction of chains in the network with length l_θ by $P(l_\theta)$, it is found that the chain distributions of all the three kinds of real-world networks are somewhat consistent with a power-law property, i.e. $P(l_\theta) \sim l_\theta^{-\alpha}$, with different exponents (here only the chains with their length larger than 1, i.e. $l_\theta > 1$, are fitted), as shown in figures 3(a)–(c). All the parameters are determined by fitting the preprocessed (e.g. logarithmized) data adopting the method of multiple linear regression using least squares (the same below). The chain length distributions for three artificial networks are shown in figure 3(d); however, none of the chains in the networks derived by the three models has the length longer than 7. In fact, the average chain lengths of these artificial networks are much shorter than those of the real-world networks.

Long chains in real-world networks may be attributed to the similar quasi-star structure in these networks, where many vertices are only connected to just a very few central vertices

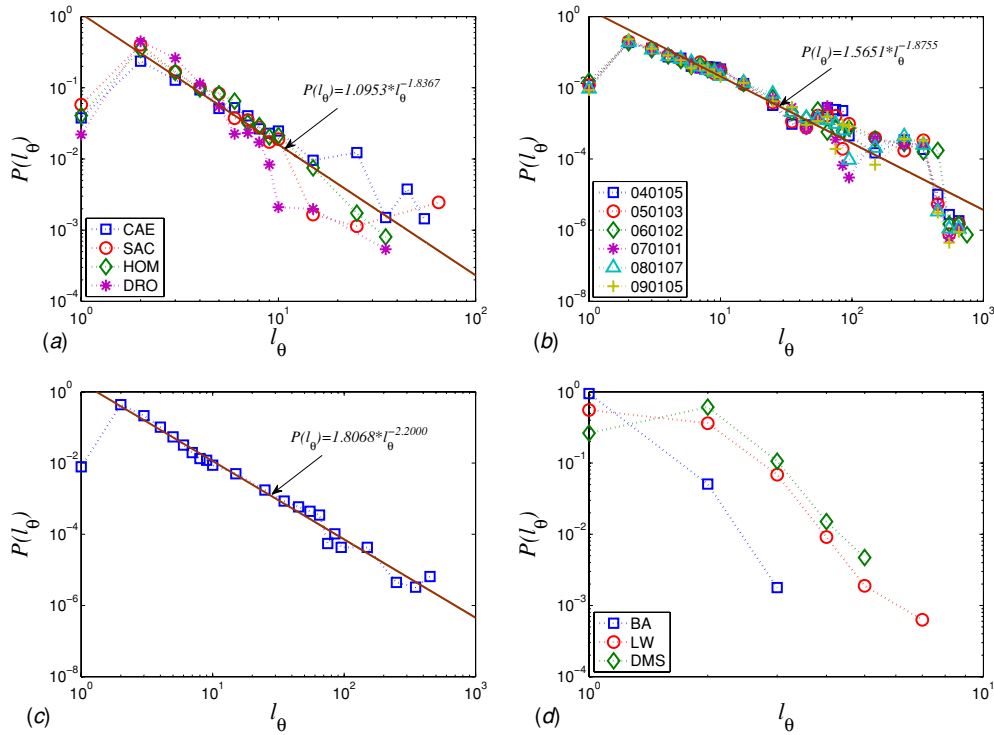


Figure 3. The chain length distributions for (a) the four protein–protein interaction networks; (b) the six AS relationship networks; (c) the *Douban* friendship network and (d) the three artificial networks. All of the distributions for the three kinds of real-world networks seem to be consistent with a power-law property, i.e. $P(l_\theta) \sim l_\theta^{-\alpha}$, with different exponents, while those for the artificial networks do not have such property or, at least, have too narrow chain length distributions. Here only the chains with their length larger than 1, i.e. $l_\theta > 1$, are fitted. It should be noted that the data in the figures (the same for figure 6) experience the same smoothing process: denoting $a_q = p \times 10^q$ and $b_q = (p + 1) \times 10^q$ where $p = 1, 2, \dots, 9$ and $q \geq 1$ is a natural number, the average probability that the chains with length $l_\theta \in (a_q, b_q]$ appear in the network then is calculated by $P(l_\theta^q) = \frac{1}{N_\theta \times 10^q} \sum_{l_\theta \in (a_q, b_q]} N_{|\theta|=l_\theta}$, where l_θ^q is simply set to the middle value of the interval $(a_q + b_q)/2$ and $N_{|\theta|=l_\theta}$ is the number of chains with length l_θ .

(hubs) and thus always have the same neighbor sets. In fact, Carmi *et al* also revealed that the proportion of this type of vertices is close to 30% in AS relationship network, and such vertices always leave or form small clusters and can reach the rest of the network only through the nucleus [29]. Figures 4(a) and (b) present two chains extracted in the AS relationship network and *Douban* friendship network, respectively. For social networks, the power-law chain length could further be explained by considering that there must be a relatively fixed order in which one recommends his friends or other famous people to the newly added individual, and the freshman then is more likely to be associated with those recommendees of higher rank; in most cases, such recommendees are also the friends of his precursors sharing the same recommender, as shown in figure 5. Based on the traditional PA rule, a freshman is more likely to be associated with a person of higher prestige and thus, under the mechanism shown in figure 5, may have more precursors; in other words, here the PA rule also means a chain-based positive feedback mechanism, i.e. a newly added vertex may join in a longer chain with a higher probability. That is, the PA rule with a well-ordered recommendation

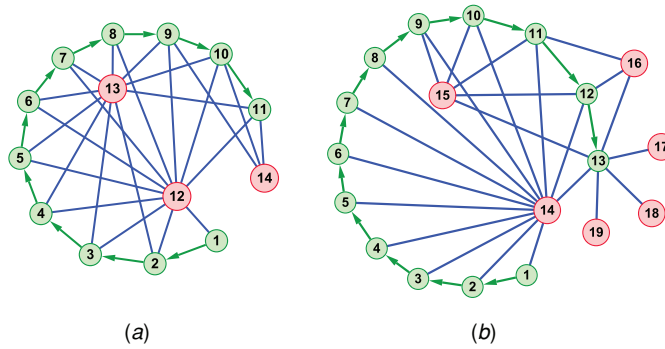


Figure 4. Chains extracted in real-world networks: (a) AS relationship network and (b) *Douban* friendship network. The first chain associated with directed lines consists of 11 vertices marked by 1, 2, . . . , 11, and these vertices have totally three neighbors marked by 12, 13, 14 respectively, while the second chain consists of 13 vertices which are connected to totally six neighbors.

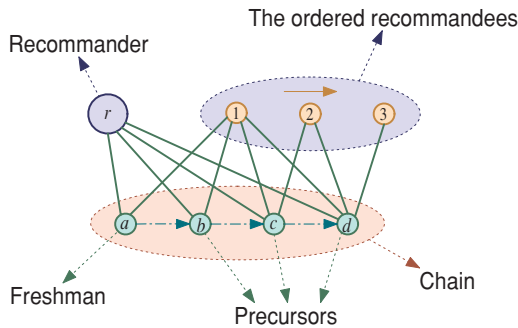


Figure 5. The well-ordered mechanism of social networks. v_r , as a person of high status in a social circle, always recommends his close friends or other people of high prestige, i.e. v_1, v_2 and v_3 , in a relatively fixed order to his adherents, i.e. v_a, v_b, v_c and v_d . As a result, the freshman v_a is more likely to be associated with the recommendees of higher rank, e.g., v_1 , which, in most cases, are also the friends of his precursors sharing the same recommender, and thus long chains will be formed as time goes on.

manner can explain the power-law chain length distribution revealed in the *Douban* friendship network.

Besides, a vertex in a network may appear in various chains, i.e. a vertex, like a multi-functional block, may have the ability to substitute different vertices in different chains when they break down. The average number of chains that a vertex participates in can be easily calculated by $N_\theta \langle L_\theta \rangle / N$, where the number of vertices N , the number of chains N_θ and the average chain length $\langle L_\theta \rangle$ are all recorded in table 2. Here, we would like to measure the multi-functional ability of a vertex by its incoming degree $k_{in}(i)$ and outgoing degree $k_{out}(i)$ in the corresponding chain graph. Denoting the average incoming degree of the chain graph by $\langle k_{in} \rangle$ and the average outgoing degree by $\langle k_{out} \rangle$, the equation $\langle k_{in} \rangle = \langle k_{out} \rangle$ must be satisfied with the reason that each directed edge connects two different vertices in the chain graph. It should be noted that a vertex v_i is isolated in the chain graph, i.e. $k_{in}(i) = k_{out}(i) = 0$, if and only if it belongs to a chain only containing itself. In fact, the chain graph of each artificial network presented in this paper has a large number of isolated vertices (the chains with length equal to 1); as a result, the chain graphs of the artificial networks have extra small average

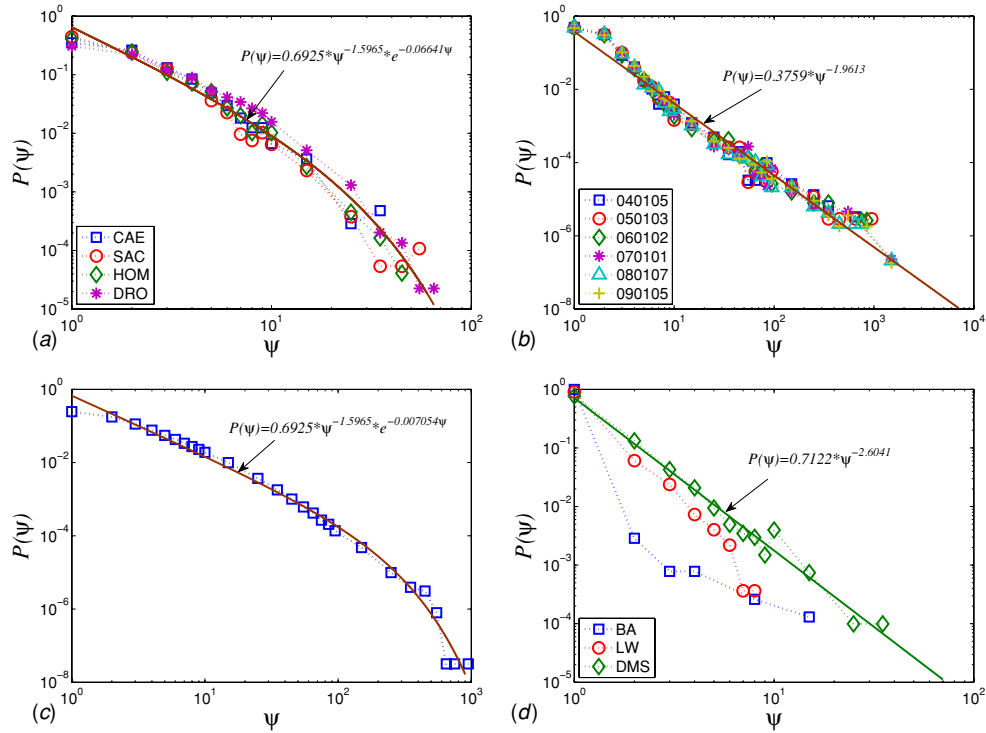


Figure 6. The family diversity distributions. The distributions for the AS relationship networks possess a near power-law property, i.e. $P(\psi) \sim \psi^{-2.0}$, which can be explained by the DMS model to a certain extent, while those for the protein–protein interaction networks and the friendship network present a similar exponential decayed power-law property, i.e. $P(\psi) \sim \psi^{-\gamma} e^{-\lambda\psi}$, with different parameters.

incoming degree $\langle k_{in} \rangle$. Intuitively, in social networks, although a newly added individual is inclined to be connected to the recommendees in regular sequence at the beginning, he/she might also become connected to the individuals in other communities who are not recommended by the initial recommender; therefore, chains in different communities may intersect each other at many different vertices, and thus the chain graphs of social networks may always have much large average incoming degree $\langle k_{in} \rangle$. For example, in this paper, the chain graph of the *Douban* friendship network has the largest $\langle k_{in} \rangle$ in all the tested networks.

4. Measurements of families in complex networks

Denoting the diversity of a family ϕ_i by ψ_i representing the number of chains it contains, the number N_ϕ of families and the *average family diversity* of the tested networks are also recorded in table 2, where we can find that, compared with the three artificial networks, the real-world networks have much higher average family diversity. Generally, a family with high diversity always has a strong end-vertex which, as an important junction, connects many different parts as a whole in the network. The *family diversity distribution* is denoted by $P(\psi)$ representing the fraction of families in the network with diversity equal to ψ . It is found that the protein–protein networks and the friendship network present a similar exponential and power-law family diversity distribution, i.e. $P(\psi) \sim \psi^{-\gamma} e^{-\lambda\psi}$, with different parameters, as shown in figures 6(a) and (c), while the AS relationship networks seem to present a near

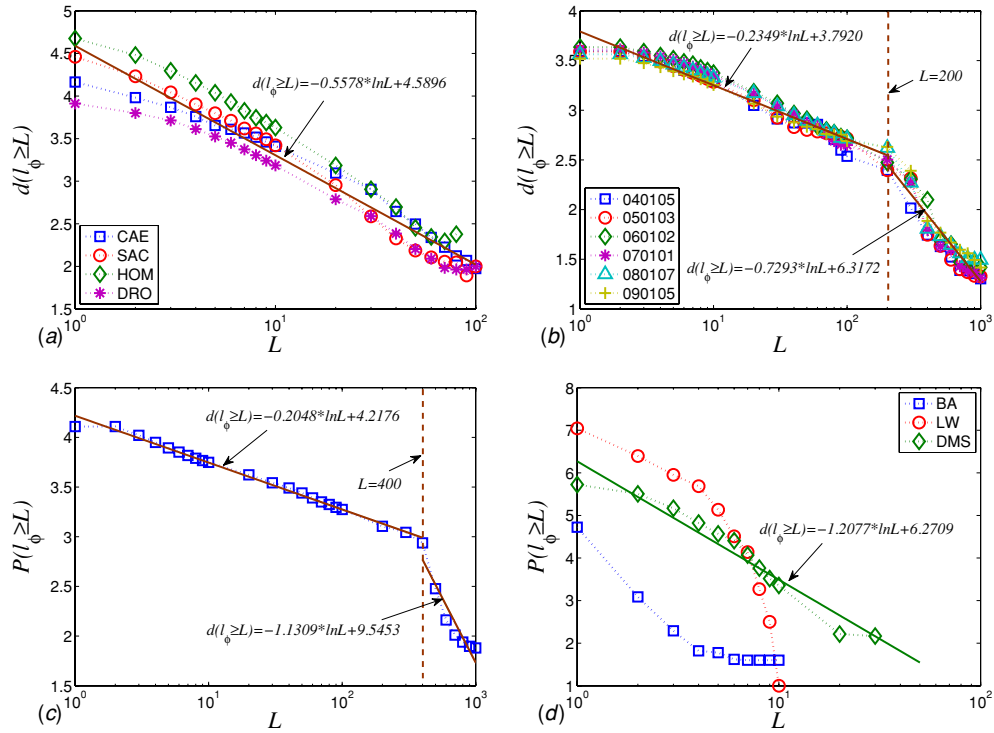


Figure 7. The average distance $d(l_\phi \geq L)$ between families with their size larger than L as functions of size L . Generally, The similar logarithmically decreasing trend of $d(l_\phi \geq L)$ suggests that most of the large families are always centralized around unique centers in these real-world networks and families continuously shrink as they depart from the centers, which can be somewhat explained by the DMS model. Moreover, the centralization seems to be abruptly enhanced when $L > 200$ for the AS relationship networks and $L > 400$ for the friendship network, which may indicate a similar two-level structure of these networks.

power-law family diversity distribution, i.e. $P(\psi) \sim \psi^{-2.0}$, as shown in figure 6(b). By comparison, the diversity distributions of the artificial networks are shown in figure 6(d). The heavy tails of the diversity distributions of real-world networks can be well explained by the Internet model of Carmi *et al*. In the model, many low-degree vertices are connected to the rest of the network only through a few hubs which are densely connected with each other and form a so-called nucleus of the network. So it is probable that a group of low-degree vertices is connected to one or two hubs in the nucleus and thus belongs to the same chain. Since the nucleus is highly connected, there may be another vertex in the nucleus which is connected to these hubs as well and thus serves as the end-vertex of this chain. Certainly, this vertex can be also connected to other hubs, i.e. it is also the end-vertex of many other chains, and thus leading to particularly high-diversity families.

The size of a family ϕ_i is denoted by $l_\phi(i)$ representing the number of vertices in it. And due to the significant roles of the end-vertices in the respective families, the distance between the two families ϕ_i and ϕ_j could be denoted by d_{ij} representing the shortest path length between their end-vertices in the subnetwork $G_E = G(V_E, E_E)$. Denoting $d(l_\phi \geq L)$ as the average distance between families with their size larger than L , the relationships between $d(l_\phi \geq L)$ and L for the real-world networks and the artificial networks are depicted in figures 7(a)–(d) respectively. The similar logarithmically decreasing trend of $d(l_\phi \geq L)$

suggests that most of the large families are always centralized around unique centers in these real-world networks and families continuously shrink as they depart from the centers, which can be somewhat explained by the DMS model. Moreover, the centralization seems to be abruptly enhanced when $L > 200$ for the AS relationship networks and $L > 400$ for the friendship network, which may also be attributed to the densely connected nuclei in these networks. Two families are considered overlapped if they share common vertices and the *family overlapping size* $l_{\phi}^o(i, j)$ between the pairwise families ϕ_i and ϕ_j , representing the correlation between them, is defined by the number of common vertices in them. Also with the reason that there are a large number of chains/families only containing one vertex in the artificial networks, they have very small *average family overlapping size* $\langle L_{\phi}^o \rangle$. Similarly, the long chains as well as the quasi-star structure should be responsible for the relatively large $\langle L_{\phi}^o \rangle$ of the real-world networks, because in these networks, there might be different vertices in the nucleus serving as the end-vertices of a same group of low-degree vertices, which leads to different but highly overlapping chains and therefore highly overlapping families.

5. Summary

In this paper, we have defined a partially ordered relation between pairwise vertices by comparing their neighbor sets, and then proposed a series of new measurements based on predefined chains and families, such as chain length distribution and family diversity distribution, in order to evaluate different network models. Unfortunately, most of these partially ordered set-based properties revealed in real-world networks cannot be explained by those well-known network models, such as the Barabási (BA) model, local-world (LW) model and Dorogovtsev, Mendes, and Samukhin (DMS) model, although, by comparing, the LW model and the DMS model behave a little better than the BA model. Interestingly, it seems that long chains, high diversity, as well as strong overlap can be explained to a certain extent by a quasi-star structure with a densely connected nucleus which has already been revealed in the autonomous system relationship networks by Carmi *et al* adopting a k -shell decomposing algorithm. So we believe that such new measurements can help us better understand the structure of real-world networks and further provide more appropriate models for them.

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 discussion about the ideas presented in this paper. This work has been supported by China Postdoctoral Science Foundation (grant no 20080441256).

References

- [1] Watts D J and Strogatz S H 1998 Collective dynamics of 'small-world' networks *Nature* **393** 440
- [2] Barabási A-L and Albert R 1999 Emergence of scaling in Random networks *Science* **286** 509
- [3] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabási A-L 2002 Hierarchical organization of modularity in metabolic networks *Science* **297** 1551
- [4] Barabási A-L and Oltvai Z N 2004 Network biology: Understanding the cell's functional organization *Nature* **5** 101
- [5] Yook S-H, Jeong H and Barabási A-L 2002 Modeling the Internet's large-scale topology *Proc. Natl Acad. Sci.* **99** 13382
- [6] Guimerà R, Mossa S, Turtschi A and Amaral L A N 2005 The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles *Proc. Natl Acad. Sci.* **102** 7794

- [7] Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J and Barabási A-L 2007 Structure and tie strengths in mobile communication networks *Proc. Natl Acad. Sci.* **104** 7332
- [8] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D-U 2006 Complex networks: structure and dynamics *Phys. Rep.* **424** 175
- [9] Costa L D F, Rodrigues F A, Travieso G and Boas P R V 2007 Characterization of complex networks: a survey of measurements *Adv. Phys.* **56** 167
- [10] Song C, Havlin S and Makse H A 2005 Self-similarity of complex networks *Nature* **433** 392
- [11] Arenas A, Danon L, Guilera A D, Gleiser P M and Guimerà R 2004 Community analysis in social networks *Eur. Phys. J. B* **38** 373
- [12] Goh K I, Salvi G, Kahng B and Kim D 2006 Skeleton and fractal scaling in complex networks *Phys. Rev. Lett.* **96** 018701
- [13] Xiao Y, Xiong M, Wang W and Wang H 2008 Emergence of symmetry in complex networks *Phys. Rev. E* **77** 066108
- [14] Xiao Y-H, Wua W-T, Wang H, Xiong M and Wang W 2008 Symmetry-based structure entropy of complex networks *Physica A* **387** 2611
- [15] Mossa S, Barthélémy M, Stanley H E and Amaral L A N 2002 Truncation of power law behavior in 'scale-free' network models due to information filtering *Phys. Rev. Lett.* **88** 138701
- [16] Li X and Chen G 2003 A local-world evolving network model *Physica A* **328** 274
- [17] Xuan Q, Li Y and Wu T-J 2007 A local-world network model based on inter-node correlation degree *Physica A* **378** 561
- [18] Dorogovtsev S N, Mendes J F F and Samukhin A N 2001 Size-dependent degree distribution of a scale-free growing network *Phys. Rev. E* **63** 062101
- [19] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 Network motifs: Simple building blocks of complex networks *Science* **298** 824
- [20] Xu X, Zhang J and Small M 2008 Superfamily phenomena and motifs of networks induced from time series *Proc. Natl Acad. Sci.* **105** 19601
- [21] Palla G, Derényi I, Farkas I and Vicsek T 2005 Uncovering the overlapping community structure of complex networks in nature and society *Nature* **435** 814
- [22] Girvan M and Newman M E J 2002 Community structure in social and biological networks *Proc. Natl Acad. Sci.* **99** 7821
- [23] Xuan Q, Li Y and Wu T-J 2006 Growth model for complex networks with hierarchical and modular structures *Phys. Rev. E* **73** 036105
- [24] Susanna S E 2004 *Discrete Mathematics with Applications* 3rd edn (Belmont, CA: Thomson Learning)
- [25] http://en.wikipedia.org/wiki/Partially_ordered_set
- [26] Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A and Tyers M 2006 BioGRID: a general repository for interaction datasets *Nucleic Acids Res.* **34** D535
- [27] The CAIDA AS Relationships Dataset, <http://www.caida.org/data/active/as-relationships/>
- [28] <http://www.douban.com/>
- [29] Carmi S, Havlin S, Kirkpatrick S, Shavitt Y and Shir E 2007 A model of Internet topology using k-shell decomposition *Proc. Natl Acad. Sci.* **104** 11150