

Emergence of symmetry in complex networks

Yanghua Xiao,¹ Momiao Xiong,^{2,3} Wei Wang,¹ and Hui Wang⁴

¹*Department of Computing and Information Technology, Fudan University, Shanghai 200433, People's Republic of China*

²*Theoretical Systems Biology Laboratory, School of Life Science, Fudan University, Shanghai 200433, People's Republic of China*

³*Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas 77225, USA*

⁴*Business School, University of Shanghai for Science and Technology, Shanghai 200093, People's Republic of China*

(Received 21 September 2007; revised manuscript received 7 December 2007; published 11 June 2008)

Many real networks have been found to have a rich degree of symmetry, which is a universal structural property of complex networks, yet has been rarely studied so far. One of the fascinating problems related to symmetry is exploration of the origin of symmetry in real networks. For this purpose, we summarized the statistics of local symmetric motifs that contribute to local symmetry of networks. Analysis of these statistics shows that the symmetry of complex networks is a consequence of similar linkage pattern, which means that vertices with similar degrees tend to share common neighbors. An improved version of the Barabási-Albert model integrating similar linkage pattern successfully reproduces the symmetry of real networks, indicating that similar linkage pattern is the underlying ingredient that is responsible for the emergence of symmetry in complex networks.

DOI: [10.1103/PhysRevE.77.066108](https://doi.org/10.1103/PhysRevE.77.066108)

PACS number(s): 89.75.Fb, 05.40.-a, 02.20.-a

I. INTRODUCTION

In the past decade, we have witnessed great progress in research on complex networks [1–13]. Previous studies have primarily focused on finding various statistical properties of real networks, such as small-world properties [3,5,8], power-law degree distribution [4], network motifs [9], assortative mixing [11], self-similarity [6], community structure [10,12], and hierarchical structure [13]. Based on these properties, many network models, such as the Barabási-Albert (BA) [4] and Watts-Strogatz models [8], have been proposed to help to predict the evolution of a network. However, one property of network structure, *symmetry*, has been rarely studied [14,15].

In general, symmetry is defined as invariance under a group of transformations [16]. The term “symmetry” has been widely used to describe the harmony, beauty, and unity in nature for a long time in an implicit way [17]. In modern physics, symmetry properties have been attributed to physical laws and physical phenomena [18]. In our studies, we will focus on the symmetry in network structures. Similar to the concept of other symmetries in physics, symmetry of network structure also characterizes invariance under certain transformations. However, the detailed definition of symmetry in a network structure is different from that of other symmetries in physics in the meaning of “invariance” and “transformation.” Specifically, symmetry in network structures characterizes the invariance of adjacency of vertices under permutations on the vertices, which implies that invariance of the symmetry in a network structure is the adjacency relation among the vertices and the transformation is the permutations on the vertices.

Contrary to the traditional belief that almost all graphs are asymmetric [19–21],¹ various real complex networks have been shown to have a rich degree of symmetry [22,23]. As a

ubiquitous phenomenon, the existence of symmetry in real networks strongly begs an explanation, since existing ingredients, such as *continuous growth* and *preferential attachment* [4] dominating the construction of network structures, are not dedicated to interpreting the origination of symmetry in real networks.

Recently, symmetry in real networks has attracted some research interest. The work in [23] utilizes symmetry information to characterize the structural heterogeneity of real networks. Reference [22] has verified that the network growth model [24] can produce treelike symmetry, and found that, given that the network is growing as a tree, preferential attachment increases network symmetry. However, many real networks whose structures are far away from trees are also symmetric [22,23]. Hence, symmetry in a majority of real networks has not been reasonably explained yet.

To explore the origin of symmetry in real networks, we summarize various statistics of the local symmetric motifs contributing to the symmetry of real networks. Through the analysis of these statistics, we show that similar linkage pattern,² which indicates that vertices having similar properties, for example degree, tend to have similar neighbors, is a ubiquitous law dominating the construction of structures of a variety of real networks. For example, in a friendship network, it is widely believed that persons with similar properties such as educational background, interest, age, would probably have common friends. Based upon these concepts, we propose an improved version of the BA model integrating the ingredient of similar linkage pattern. The proposed model successfully reproduces the symmetry in real networks, im-

¹This could be understood asymptotically that the proportion of graphs on n vertices that are asymmetric goes to 1 as n tends to ∞ .

²There may be confusion between similar linkage pattern and assortative mixing. Both these two concepts focus on the behavior of those vertices having similar properties. However, assortative mixing emphasizes the interlinkage between these vertices, while similar linkage pattern concerns only neighbor-sharing of these vertices; whether these vertices are interlinked is not significant in similar linkage pattern.

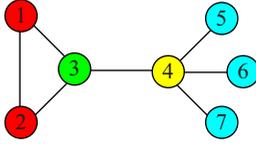


FIG. 1. (Color online) Illustration of a symmetric graph. Vertices with the same color belong to the same orbit.

plying that similar linkage pattern is responsible for the formation of symmetry in real networks.

II. SYMMETRY IN NETWORKS

First, we will give a brief introduction to symmetry in complex networks. The theory of graph symmetry has been well developed in the context of *algebraic graph theory* [20,21,25,26], where the basic idea is to explore the interplay between graph theory and algebra, including matrix algebra and group theory [26].

A. Preliminaries

A graph or 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. If $(v_1,v_2)\in E$, v_1 and v_2 are adjacent. A one-to-one mapping, or bijective mapping, from a vertex set V onto itself is called a *permutation* acting on V . Since a permutation is a mapping, we can define *product* or *composition* operation on permutations. Let $f,g\in S(V)$, where $S(V)$ is the set of all permutations acting on V ; then the product or composition $h=f\circ g$ (or simply fg) is the mapping $h:V\rightarrow V$ for which $x^h=(x^f)^g$. Among all the permutations in $S(V)$, some permutations can preserve the adjacency of the vertices, and these permutations are called *automorphisms* acting on the vertex set. More formally, an automorphism of a graph G is a permutation g of the vertex set of G with the property that, for any vertices u and v , v^g is adjacent to u^g if and only if v is adjacent to u . We denote the set of automorphisms of graph G as $\text{Aut}(G)$, and we have $\text{Aut}(G)=\{g:g\in S(V)\text{ and }E^g=E\}$, where $E^g=\{(u^g,v^g):(u,v)\in E\}$. The set of automorphisms under the product of permutations forms a group [20]. In general, a network is considered as asymmetric if its underlying graph contains only an identity permutation; otherwise, the network is symmetric.

Next, we will introduce the concepts of automorphism partition and orbit. Given the automorphism group acting on vertex set V , we can get a partition $\mathcal{P}=\{V_1,V_2,\dots,V_k\}$ such that x is equivalent to y if and only if $\exists g\in \text{Aut}(G)$, such that $x^g=y$. This partition is called *automorphism partition*, and each cell of the partition is called an *orbit* of $\text{Aut}(G)$. An orbit is *trivial* if it contains only a single vertex; otherwise, the orbit is *nontrivial*.

We illustrate the above concepts through an example. Figure 1 shows a graph with seven vertices. Thus the number of all possible permutations acting on vertices of this graph is $7!=5040$. Among all these permutations, some cannot preserve the adjacency of the vertex set. For example, the permutation g that maps vertex 3 and vertex 4 to each other while fixing any other vertices cannot preserve the original

adjacency, since vertices 1 and 2 are not adjacent to vertex 4. However, we also can find some nontrivial automorphisms, e.g., the permutation h that only switches vertex 1 and vertex 2 with each other will preserve the adjacency. Furthermore, utilizing some tools, such as GAP [27], or NAUTY [28], we can find all automorphisms for this graph (a total of 12 automorphisms can be found). Given these automorphisms, we can find all orbits and construct the automorphism partition. For example, since automorphism g transforms vertex 1 into vertex 2, these two vertices will belong to the same orbit. Similarly, all orbits can be found and the resulting automorphism partition is $\mathcal{P}=\{\{1,2\},\{3\},\{4\},\{5,6,7\}\}$ (vertices in each orbit are marked with the same color in Fig. 1).

Since nontrivial automorphisms have been found, the graph in Fig. 1 should be characterized as symmetric. However, merely characterizing a network as symmetric or asymmetric is not sufficient for real applications. If a network is symmetric, measures of the extent to which the network is symmetric will provide us more information about the symmetry in the network. Thus, it is necessary to explore measurements of symmetry in networks.

B. Measures of symmetry in networks

Intuitively, the size of the automorphism group $\alpha_G=|\text{Aut}(G)|$ [29] gives a direct quantification of the abundance of symmetry in a network. In order to compare the symmetry of networks with different sizes, β_G is used in [22], defined as

$$\beta_G=(\alpha_G/N!)^{1/N}, \quad (1)$$

where N is the number of vertices in the network. β_G measures the symmetry relative to the maximal number of possible automorphisms of a network with N vertices.

In our studies, another symmetry measure γ_G is also used, which is based upon the intuitive observation that a network having more nontrivial orbits will be more symmetric. This observation motivates us to define γ_G as the ratio of the number of vertices in all nontrivial orbits to the vertex number of the network. Formally, γ_G could be defined as

$$\gamma_G=\frac{\sum_{1\leq i\leq k,|V_i|>1}|V_i|}{N}, \quad (2)$$

where V_i is the i th orbit in the automorphism partition.

As an example, we perform some preliminary analysis on symmetry in circles with n vertices (usually denoted as C_n). In algebraic graph theory [20,21,25,26], it is well known that the automorphisms of an undirected cycle C_n form a group D_n , called the *dihedral group*, whose size has been proved to be $2n$. Thus, for such a circle C_n , we have $\beta_G=(2n/n!)^{1/n}=[2/(n-1)!]^{1/n}$. It is easy to check that all vertices of a circle belong to the same orbit. Consequently, we have $\gamma_G=100\%$. Hence, it is reasonable to believe that C_n is richly symmetric.

Note that, in the community of complex networks, measuring symmetry in complex networks has also attracted some research interest, such as measurements of degree symmetry [14] and its extensions [15]. All these symmetry measures are vertex oriented, not graph oriented, and are used to

measure the local centrality of vertices. However, the symmetry measures used in this paper are all graph oriented; this allows measurement of the symmetry of the whole network, or the ability for a network to preserve vertex adjacency under possible permutations acting on the vertex set.

C. Introduction to data sets

To study the symmetry of real networks, we use the following well-known real network data sets. The first is the electrical power grid of the western United States [8], in which the vertices represent generators, transformers, and substations, and the edges correspond to the high-voltage transmission lines between them. We refer to this network as the *USPowerGrid*. The second data set is the citation network of the high-energy physics theory (hep-th) community [30]. In the citation network, each vertex represents an article and a directed edge from article *A* to article *B* indicates that *A* cites *B*. We refer to this data set as *arXiv*. The third data set is the internet at the autonomous system level [31]. Each vertex in the network represents an autonomous system, and each vertex has a certain relationship with other vertices, such as consumer, provider, peer, or sibling. We refer to this data set as *InternetAS*. The fourth data set is *BioGrid*, which represents protein-gene interactions of various species. Interaction data can be represented as a network, where each vertex represents a protein or a gene, and each edge represents the interaction between these proteins or genes. We use interaction network of five species including *Saccharomyces cerevisiae* (SAC), *Caenorhabditis elegans* (CAE), *Drosophila melanogaster* (DRO), *Homo sapiens* (HOM), and *Mus musculus* (MUS).

III. SIMILAR LINKAGE PATTERN

In this section we will show that similar linkage pattern is a ubiquitous law that holds across many structures of real networks. For this purpose, we first need to have a deeper insight into the local substructures contributing to the abundance of symmetry of real networks. In the mathematical context, such symmetric substructures of real networks can be represented by *symmetric bicliques*.

A. Symmetric bicliques

Let V_1 and V_2 be two disjoint vertex sets. Then K_{V_1, V_2} is a complete bipartite graph, if vertices in the same subset are not adjacent and every two vertices from different subsets are adjacent. Furthermore, if a complete bipartite K_{V_1, V_2} is a subgraph of $G(V, E)$ and for each $v \in V_1$, $N_K(v) = N_G(v) = V_2$, we call K_{V_1, V_2} a symmetric biclique³ [32] of G , where $N_K(v)$ and $N_G(v)$ are the neighbor sets of vertex v in graphs K and G , respectively. This definition of the symmetric biclique implies that, for each vertex $v \in V_1$, v is adjacent to all vertices

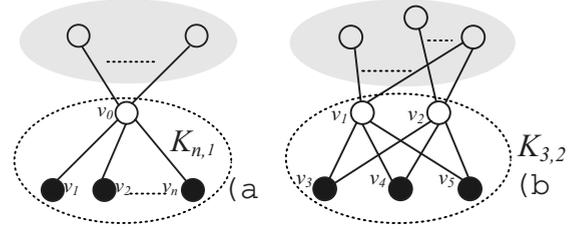


FIG. 2. Illustration of symmetric bicliques. (a) shows an example of $K_{n,1}$, which contributes to the size of the automorphism group of the network with a factor $n!$. (b) shows an example of $K_{3,2}$, which contributes to the size of the automorphism group of the network with a factor $3!$.

in V_2 and only adjacent to vertices in V_2 , whether in subgraph K or supergraph G .

If graph $G(V, E)$ contains a symmetric biclique K_{V_1, V_2} , we can find $n!(n=|V_1|)$ automorphisms of graph G , such that each of these automorphisms only permutes vertices in V_1 with the other vertices in $V-V_1$ fixed. Furthermore, all these automorphisms form a subgroup of the automorphism group of G . Then according to the *Lagrange theorem*⁴ in group theory, the above subgroup will contribute with a factor $n!$ to the size of the whole automorphism group. Thus, the symmetric biclique K_{V_1, V_2} becomes a *local symmetric motif* [32] contributing to the symmetry of the network. Figure 2 illustrates two such bicliques.

If we do not care about what V_1 and V_2 are, we also use $K_{i,j}$ to denote K_{V_1, V_2} , where $|V_1|=i$ and $|V_2|=j$. The set consisting of all $K_{i,j}$ is denoted by $\mathcal{K}_{i,j}$. Note that $K_{1,i}$ does not necessarily contribute to the local symmetry of the network; hence, in the following discussion, only $K_{n,i}$ with $n \geq 2$ will be considered.

For any symmetric biclique K_{V_1, V_2} in a network, if there does not exist a symmetric biclique $K_{V'_1, V'_2}$ such that $V_2 = V'_2$ and $V_1 \subset V'_1$, we call K_{V_1, V_2} a *maximal* symmetric biclique in the network. For example, in the network shown in Fig. 2(b), we can find three symmetric bicliques $K_{\{v_3, v_4\}, \{v_1, v_2\}}$, $K_{\{v_3\}, \{v_1, v_2\}}$, and $K_{\{v_3, v_4, v_5\}, \{v_1, v_2\}}$. Furthermore, we can easily check that $K_{\{v_3, v_4, v_5\}, \{v_1, v_2\}}$ is the maximal one; it is the union of two other symmetric bicliques.

Recall that two graphs are the same graph if and only if they have the same vertex set and the same edge set; otherwise, these two graphs will be distinct from each other. In the following sections, we will consider only the number of distinct maximal symmetric bicliques for $\mathcal{K}_{n,i}$ with $n \geq 2$, which implies that if there exist two symmetric bicliques K_{V_1, V_2} and $K_{V'_1, V'_2}$ such that $V_2 = V'_2$, we shall merge the two bicliques into a larger biclique $K_{V_1 \cup V'_1, V_2}$; and these three symmetric bicliques will be counted as only one occurrence of symmetric bicliques in this graph.

Based on the above concepts, we can easily find all symmetric bicliques in a network. First, we use $N[V]$ to denote the subgraph consisting of a vertex set V as well as its inci-

³In [32], a more concise characterization of the symmetric biclique in the context of algebraic graph theory is also given, where symmetric bicliques are characterized as complete bipartite subgraphs invariant under the action of $\text{Aut}(G)$.

⁴Lagrange's theorem states that, if H is a subgroup of G and G is a finite group, then $|G|=|H| \cdot [G:H]$, where $[G:H]$ is the number of different cosets.

TABLE I. Symmetric biclique statistics of a variety of real networks. We measure the size of the networks by the number of vertices and edges, denoted by N and M , respectively. For each $i \leq 7$, the statistics of distinct maximal symmetric bicliques contained in $\mathcal{K}_{n,i}$ with $n \geq 2$ is summarized. We use a triple $(S, \text{Min}, \text{Max})$ to show the statistics of $\mathcal{K}_{n,i}$, where S is the number of distinctive maximal $K_{n,i}$ and Min and Max are the minimal and maximal sizes of symmetric bicliques (the size of a symmetric biclique K_{V_1, V_2} is measured by $|V_1|$), respectively. If $K_{n,i}$ does not appear in the network, $S=0$, and Min and Max are not available; this is denoted by a dash. For some larger i , corresponding statistics of $\mathcal{K}_{n,i}$ are also given.

Network	i in $\mathcal{K}_{n,i}$ with $n \geq 2$							Some larger i
	1	2	3	4	5	6	7	
arXiv [30] ^a	(135,2,7)	(42,2,4)	(17,2,3)	(13,2,2)	(11,2,2)	(1,2,2)	(2,2,2)	$i=16, (1,2,2)$
InternetAS ^b	(916,2,343)	(1057,2,285)	(90,2,25)	(9,2,4)	(2,2,2)	(0,-,-)	(0,-,-)	(0,-,-)
BioGrid [33]								
SAC	(51,2,15)	(7,2,5)	(0,-,-)	(0,-,-)	(0,-,-)	(0,-,-)	(0,-,-)	(0,-,-)
MUS	(7,2,44)	(8,2,12)	(4,2,6)	(2,2,2)	(0,-,-)	(1,2,2)	(0,-,-)	(0,-,-)
HOM	(366,2,44)	(53,2,12)	(21,2,6)	(5,2,2)	(2,2,2)	(1,2,2)	(0,-,-)	$i=8, 10, 21, (1,2,2)$
DRO	(418,2,40)	(16,2,11)	(6,2,3)	(6,2,3)	(3,2,2)	(0,-,-)	(3,2,3)	$i=8, 10, 21, (2,2,2), i=15, 27, (1,3,3)$ $i=17, 18, 19, 23, 25 (1,2,2) i=39, (1,6,6)$
CAE	(245,2,47)	(9,2,5)	(1,2,2)	(0,-,-)	(0,-,-)	(0,-,-)	(0,-,-)	(0,-,-)
USPowerGrid [8]	(137,2,9)	(25,2,3)	(0,-,-)	(1,2,2)	(0,-,-)	(0,-,-)	(0,-,-)	(0,-,-)

^aHere, the snapshot at 2006-03 of hep-th (high-energy physics theory) citation graph [30] is used.

^bHere, the snapshot at 2006-07-10 of CAIDA [31] is used.

dent edges. For example, as shown in Fig. 2(b), if $V = \{v_3, v_4, v_5\}$, then $N[V]$ is just the $K_{3,2}$ that has been marked by a dotted ellipse. Please note that, even though v_1 and v_2 are adjacent, the edge (v_1, v_2) should not belong to $N[V]$. The procedure to enumerate all distinct maximal symmetric bicliques is shown as follows.

For each $i \geq 1$, we find the corresponding vertex set consisting of vertices with degree i , denoted by $V(i)$. Then, for each $V(i)$, we partition $V(i)$ into $\{V(i)_1, V(i)_2, \dots, V(i)_k\}$, according to the equivalence of the neighbor set of each vertex in $V(i)$, which means that vertices with the same neighbors are classified as the same class. Then, each $N[V(i)_j]$ will be identified as a symmetric biclique in the form of $K_{n,i}$ with $n=|V(i)_j|$ {Please note that $N[V(i)_j]$ with $|V(i)_j|=1$ will be ignored in the statistics in Table I.}

B. Similar linkage pattern

From the above definition of symmetric bicliques, we can see that in such bicliques vertices in V_1 have the same degree and share the same neighbors, which is just the meaning of the phenomenon that we called similar linkage pattern. Thus, if real networks contain significant numbers of such symmetric bicliques, similar linkage pattern must play an important role in network evolution. Hence, we will summarize the statistics of symmetric bicliques in real networks to explore similar linkage pattern.

In Table I, we count the number of distinct maximal symmetric bicliques for $\mathcal{K}_{n,i}$ with $n \geq 2$ and record the corresponding maximal and minimal sizes of the bicliques in $\mathcal{K}_{n,i}$. We can see that similar linkage pattern is a universal phenomenon in the process of structure construction of many real networks, including social, biological, and technological networks. For instance, for $\mathcal{K}_{n,1}$ in InternetAS, there are totally 916 symmetric bicliques, among which there exist some

larger symmetric motifs, e.g., the maximal motif has 343 vertices in V_1 . For all the networks we tested, simple symmetric motifs such as $K_{n,1}$ and $K_{n,2}$ can be frequently observed. Moreover, for some networks, e.g., the BioGrid network DRO, complex symmetric motifs, i.e., $K_{n,i}$ with larger i , frequently occur. As shown in Fig. 3, among those simple symmetric bicliques with $i=1, 2$, the biclique size distributions are right skewed with a long tail for larger sizes, which

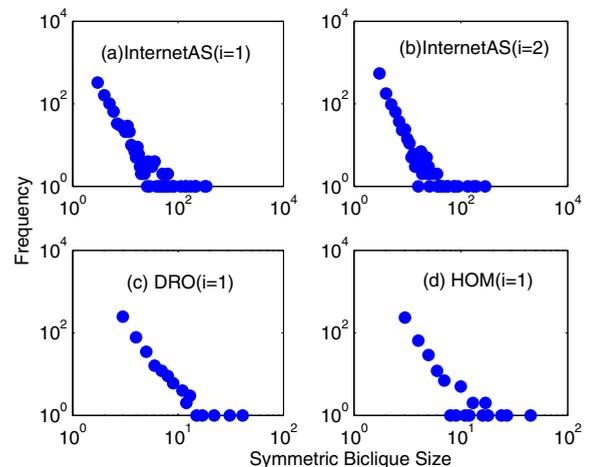


FIG. 3. (Color online) Size distribution of symmetric bicliques for real networks. The horizontal axis for each panel is the size of symmetric bicliques and the vertical axis is the occurrence frequency of the symmetric bicliques with the corresponding size. (a) and (b) show the biclique size distribution of the Internet at the autonomous level. Here, the snapshot at 2006-07-10 of CAIDA [31] is used for $\mathcal{K}_{n,1}$ and $\mathcal{K}_{n,2}$, respectively. (c) shows the biclique size distribution of $\mathcal{K}_{n,1}$ of *D. melanogaster* [33]. (d) shows the biclique size distribution of $\mathcal{K}_{n,1}$ of *H. Sapiens* [33].

TABLE II. Symmetry biclique statistics for ER random networks. For each network tested in Table I, we generate 100 ER random networks with the same size as the real networks using PAJEK [36]. We use two parameters, the vertex number N and the average degree z , to ensure that the simulated ER random networks have the same size as the corresponding real networks. Similar to the results in Table I, for each i , we summarized the number of occurrences of $K_{n,i}$ in randomized networks, which is represented by the statistics $N_{\text{rand}} \pm \text{SD}$ with N_{rand} denoting the average number of occurrences of $K_{n,i}$ over all randomized networks and SD representing the standard deviation of the number of occurrences of $K_{n,i}$ over all randomized networks. For each i , we also summarized the minimal and maximal sizes of $K_{n,i}$ over all 100 randomized networks, respectively.

Network	N	z	$N_{\text{rand}} \pm \text{SD}$ for $K_{n,1}$ with $n \geq 2$
arXiv	27770	25.37	$(0 \pm 0, -, -)$
InternetAS	22442	4.06	$(53.66 \pm 7.87, 2, 4)$
BioGrid			
SAC	5437	26.86	$(0 \pm 0, -, -)$
MUS	218	3.65	$(0.88 \pm 0.94, 2, 3)$
HOM	7522	5.32	$(2.45 \pm 1.49, 2, 2)$
DRO	7528	6.69	$(0.25 \pm 0.54, 2, 2)$
CAE	2780	3.13	$(24.27 \pm 5.33, 2, 4)$
USPowerGrid	4941	1.49	$(222.86 \pm 14.70, 2, 5)$

implies that a certain number of larger symmetric bicliques do exist in these real networks.

Following the general scheme for detecting statistically significant topological patterns [34], we compare the statistics of symmetric bicliques to that summarized from an ensemble of Erdős-Rényi (ER) [35] randomized networks which preserve certain low-level topology properties (here, we just preserve the overall number of vertices and the average degree of real networks).⁵ We find that similar linkage pattern does not tend to happen in corresponding ER random networks. As shown in Table II, randomized networks having the same size as the corresponding real networks have fewer symmetric motifs in the form of $K_{n,1}$, and the complexity of the motifs in randomized networks is much lower than that of the corresponding real networks. We also summarized the statistics for $K_{n,i}$ with $i \geq 2$, which is omitted in the table. When $i=2$, for all eight real networks tested, the corresponding occurrence probability of $K_{n,i}$ in randomized networks, which is defined as the fraction of the number of randomized networks in which $K_{n,i}$ will occur divided by the number of overall sampled randomized networks (here 100 samples are used), is less than 6%; and when $i \geq 3$, the occurrence probability of $K_{n,i}$ is 0. Thus, large or complex symmetric bicliques have less chance to appear in randomized networks.

⁵Although it is desired to preserve the degree of an individual vertex in a meaningful randomization process [41], to explore the statistical significance of the occurrence of symmetric bicliques, we need to relax the constraint of degree preservation due to the fact that statistics of symmetric bicliques strongly rely on the degree distribution of the networks.

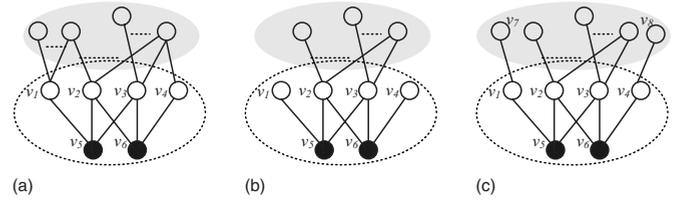


FIG. 4. Illustration of nonexact similar linkage pattern. All vertices in V_1 , i.e., dark vertices, have the same degree. However, the neighbors of these vertices are not exactly the same. Motifs consisting of V_1 as well as their incident edges do not necessarily contribute to the local symmetry of networks. In (a), the subgraph marked by a dotted ellipse will not lead to any automorphism, while the subgraph marked by a dotted ellipse in (b) will result in an automorphism $p = (v_1, v_4)(v_5, v_6)$ and the subgraph marked by a dotted ellipse in (c) also contributes an automorphism $p = (v_1, v_4)(v_5, v_6)(v_7, v_8)$ to the symmetry of the graph.

The frequent occurrence of complex $K_{n,i}$ in real networks and the infrequent occurrence of complex $K_{n,i}$ in random networks strongly suggest that the occurrence of symmetric bicliques in real networks is statistically significant (it is just in this sense that these symmetric bicliques are referred to as symmetric motifs), which implies that there exist some laws dominating the structure construction process of real networks.

Consider the dynamic process of network growth. We assume that at some time a new vertex v joins a symmetric biclique K_{V_1, V_2} of a network, and that the new substructure $K_{V_1 \cup \{v\}, V_2}$ is still a symmetric biclique; then v will attach to all vertices in V_2 . Thus, it is reasonable to believe that the newly added vertex v will be attached to the existing vertices under the principle of preferential linkage to those vertices to which other vertices in V_1 attach. Since vertices in V_1 have the same degree, it is likely that vertices having the same degree will have the same neighbors.

C. Nonexact similar linkage pattern

However, as shown in Fig. 4, in real networks, it is possible to find bicliques in which vertices having the same degree tend to share only some of their neighbors rather than exactly the same neighbors. Furthermore, we can check that these local motifs exhibiting nonexact similar linkage behavior also have the chance to contribute to the symmetry of the network. These local substructures are generalizations of symmetric bicliques, such that the structural constraint of the clique is relaxed from being completely bipartite to satisfying only the requirement that all the vertex of V_1 have the same degree in the clique. Hence, this kind of *generalized symmetric biclique*⁶ can be formally defined as a bipartite K_{V_1, V_2} such that for each $v \in V_1$, $\deg(v) = d$, where $\deg(v)$ is the degree of vertex v and d is a constant. Thus, d becomes

⁶Note that the concept of a dense overlapping regulon (DOR) [42] is also bipartite; however, the DOR is not necessarily a generalized symmetric biclique, i.e., it does not necessarily satisfy the condition that the degrees of vertices in one bipartite system are a characteristic constant.

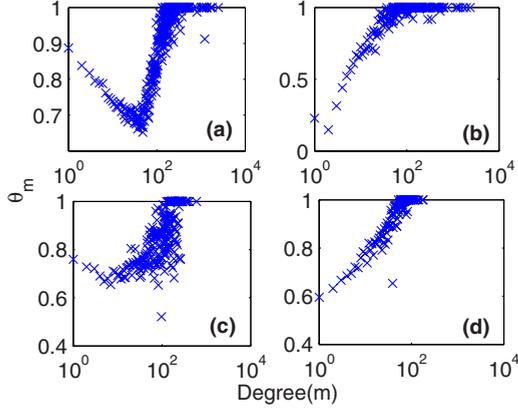


FIG. 5. (Color online) (a), (b), (c), and (d) show θ_m distribution of arXiv, InternetAS, Biogrid SAC, and Biogrid DRO, respectively.

a characteristic of the generalized symmetric biclique. We often denote it as K_{V_1, V_2}^d . In some cases when V_2 is not significant, $K_{V_1}^d$ or $K_{|V_1|}^d$ is often used.

In a network, if nonexact similar linkage pattern does make sense, then vertices with the same degree will tend to share some of their neighbors. Thus, we need to measure to what extent these vertices share the same neighbors. Let $V(m) = \{v : v \in V \text{ and } \deg(v) = m\}$ be all vertices with degree m ; then the neighbors of these vertices can be denoted as $V'(m) = \{v' : (v, v') \in E \text{ and } v \in V(m)\}$. Then we can define θ_m as the ratio of the actual number of distinct neighbors of $V(m)$ to the maximal probable number of distinct neighbors of $V(m)$, as shown in Eq. (3). Note that the maximal set of neighbors of $V(m)$ can be obtained when any pair of vertices in $V(m)$ share no neighbors. θ_m can be used to quantify the overlap ratio of neighbors of vertices in $V(m)$:

$$\theta_m = \frac{|V'(m)|}{m|V(m)|}. \quad (3)$$

From the above equation, we have an immediate consequence $0 < \theta_m \leq 1$. If $|V_m|$ is given, we have $\frac{1}{|V(m)|} \leq \theta_m \leq 1$. Note that the lower θ_m is, the more common neighbors vertices in V_1 tend to share. As shown in Fig. 5, for small values of the degree, all tested networks tend to have relative small θ_m , which strongly suggests that for these real networks, in the process of network growth, vertices with the same small degree tend to share common neighbors.

Please note that θ_m is a *global measure* on the whole network, which is a function of the degree m and indicates to what extent vertices with degree m tend to share common neighbors. In fact, θ_m also can be considered as the quantification of neighbor overlap of the whole network. However, we must note that a variety of local measures are available to measure neighbor overlap in the network, such as structural similarity [37,38] measures and topological overlap [13]. Structural similarity measures, such as the Jaccard index [39] and cosine similarity [40], are used to measure the similarity of vertices solely based on structural information of networks, while topological overlap is used to quantify to what extent two vertices tend to belong to the same module. In spite of the differences, all the above measures including θ_m

are based on the same principle that the number of common neighbors is a significant indication of the similarity between two vertices.

IV. MODEL FOR SYMMETRY

It has been shown that a variety of real networks have power-law degree distribution [1,3,7], which can be attributed to two basic ingredients: (1) growth and (2) preferential attachment [4]. In the BA model, new vertices will be continuously added to the existing networks, and at each time step, a new vertex is preferentially attached to m existing highly connected vertices. We define the number m , i.e., the number of neighbors that a vertex v is linked to at the time when the vertex was added to the network, as the *initial degree*⁷ of the vertex. In the BA and other improved models, whether m is fixed or not is not significant for the resulting degree distribution (choosing m randomly will not change the exponent of the degree distribution [4]); hence usually the initial degree is considered as a constant. However, in our model, which considers symmetry, m should not be treated as a constant, which is significant for the reproduction of symmetry in real networks.

Although many network generation models are available, ingredients producing symmetry have not been considered in the BA model and other network generation models. To reproduce symmetric networks with power-law degree distribution, we propose a network model incorporating similar linkage pattern into the BA model's two ingredients. For this purpose, the BA model's two principles are modified as follows.

(1) Newly added vertices are linked to the existing vertices not only under the principle of preferential attachment, but also that of similar linkage pattern. The latter principle implies that newly added vertices with initial degree m tend to link to the targets to which existing vertices with degree m in the network are linked.

(2) The initial degree m of newly added vertices follows a certain distribution instead of being a constant value. In the BA and other existing models, the initial degree is constant; however, in the following sections, we will show that in some real networks, m follows a certain distribution.

A. Preferential attachment with similar linkage pattern

The probability, denoted by $\Pi(v_i)$, that a new vertex with initial degree m will be connected to vertex v_i relies not only on the degree k_i of vertex v_i but also on whether v_i belongs to $V'_t(m)$ [we use $V_t(m)$ to denote the set of vertices with degree m in the network at time t and $V'_t(m)$ to denote the set consisting of neighbors of $V_t(m)$]. To incorporate the ingredient of similar linkage pattern into the basic BA model, we need to increase $\Pi(v_i)$ for those v_i belonging to $V'_t(m)$. In addition, we need to define the parameter α to control the relative significance of similar linkage pattern in the formation of network structure. Note that, for a given m , $V'_t(m)$ is

⁷The initial degree defined here can be considered as one kind of "fitness" or "hidden variable" of vertices in the network [43,44].

not necessarily nonempty. Hence the probability $\Pi(v_i)$ is defined in two cases: when $V'_t(m)=\emptyset$, then $\Pi(v_i)$ should be defined as

$$\Pi(v_i) = \frac{k_i}{\sum_j k_j}, \quad (4)$$

where k_i is the degree of vertex v_i ; when $V'_t(m) \neq \emptyset$, $\Pi(v_i)$ should be defined as

$$\Pi(v_i) = \begin{cases} \alpha \frac{k_i}{\sum_j k_j} + (1-\alpha) \frac{1}{|V'_t(m)|} & \text{if } v_i \in V'_t(m), \\ \alpha \frac{k_i}{\sum_j k_j} & \text{if } v_i \notin V'_t(m), \end{cases} \quad (5)$$

where $\alpha \in (0, 1]$.

At some time step t , we may have $V'_t(m)=\emptyset$. Consequently, in this case, attachment of vertices will be reduced to purely preferential attachment in terms of the value of the degree. It will happen frequently in the initial stages of network growth in our model, because the abundance of the degree is limited in the initial stage, due to the relative small size of the seed network. For example, if the seed network contains only isolated vertices, then only degree 0 will be found; if the seed network is a regular network, e.g., a complete network, we could also find only one degree in the network.

Equation (5) has only one parameter α to control the relative significance of purely preferential attachment and similar linkage pattern. It is clear that the larger α is, the less will be the impact of similar linkage pattern on the network. When $\alpha=1$, the model is reduced to purely preferential attachment according to the vertex degree.

B. Initial degree following a certain distribution

In the BA and other improved models, all vertices except for the seed vertices have the same initial degree. However, for some networks, especially social and technological networks, where historical data about the initial degree of real networks are available, we can show that the initial degree of these real networks may be far away from a fixed value or a value independent of degree. For example, Fig. 6 shows the distribution of initial degree of a citation network constructed from the arXiv data set. From this figure, we can see that the frequency of the initial degree of vertices follows a right-skewed distribution instead of being a fixed value.

Assume that we grow the network in a way following the principle of preferential attachment with similar linkage pattern. If the initial degree is constant, then each time a new vertex is added to the network, a fixed number (m) of edges will be introduced into the network. Thus, the local symmetric motifs will concentrate on those subgraphs with structure closer to $K_{V_1}^m$. If m is very much larger than 1, it is contradictory to the observation in Table I that the larger n is, the

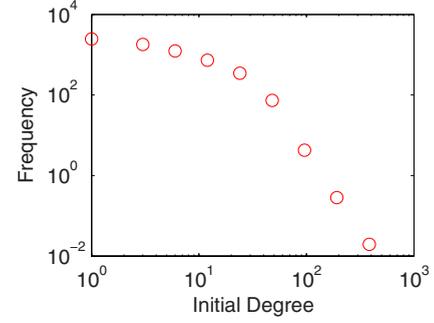


FIG. 6. (Color online) Distribution of initial degree of arXiv data set.

less frequently $K_{V_1}^n$ tends to occur in real networks.

Therefore, it is necessary to extend the initial degree from a fixed value to a certain distribution. From this perspective, the BA model can be considered as a special case of our model, where the initial degree distribution is specified as a constant value.

C. Network model based on similar linkage pattern

The algorithm for network growth incorporating the ingredient of similar linkage pattern is the following.

(1) *Growth*. Starting from a small number (n_0) of isolated vertices, at every time step, we add a new vertex with m edges that link the new vertex to m different vertices already present in the system, where m follows a distribution $F(m)$ with $m \leq \bar{m}$ (\bar{m} is the upper bound of the initial degree, called the maximal initial degree.).

(2) *Preferential attachment with similar linkage pattern*. The probability Π that a new vertex will be connected to vertex v_i is defined by Eqs. (4) and (5).

The above improved model based on similar linkage pattern just needs three input parameters ($n_0, F(m), \alpha$). For notational convenience, the model is denoted by $\text{SLP}(n_0, F(m), \alpha)$, where SLP indicates similar linkage pattern.

As shown in Fig. 7(a), with α varying from 1 to 0.1, i.e., with similar linkage pattern becoming more significant in the network construction, the size of the automorphism group of networks continuously increases through several hundreds of orders of magnitude. The insets (I) and (II) of Fig. 7(a) show that two other symmetry indices β_G and γ_G also increase with the decrease of α . Such facts also can be observed from Figs. 7(b)–7(d). Hence, it is reasonable to believe that similar linkage pattern is responsible for the emergence of symmetry in the networks.

With increase of the size of the SLP network, the symmetry indices have different trends. As shown in Fig. 7(b), the automorphism group size grows exponentially with growth of the network. In contrast, β_G decreases in a power-law way with the growth of the network, as shown in Fig. 7(c). γ_G reaches a steady state when $N \rightarrow \infty$, as shown in Fig. 7(d).

If we remove the ingredient of similar linkage pattern, we can show that the ingredient of preferential attachment with initial degree following a distribution will not necessarily reproduce the symmetry of networks. For this purpose, we

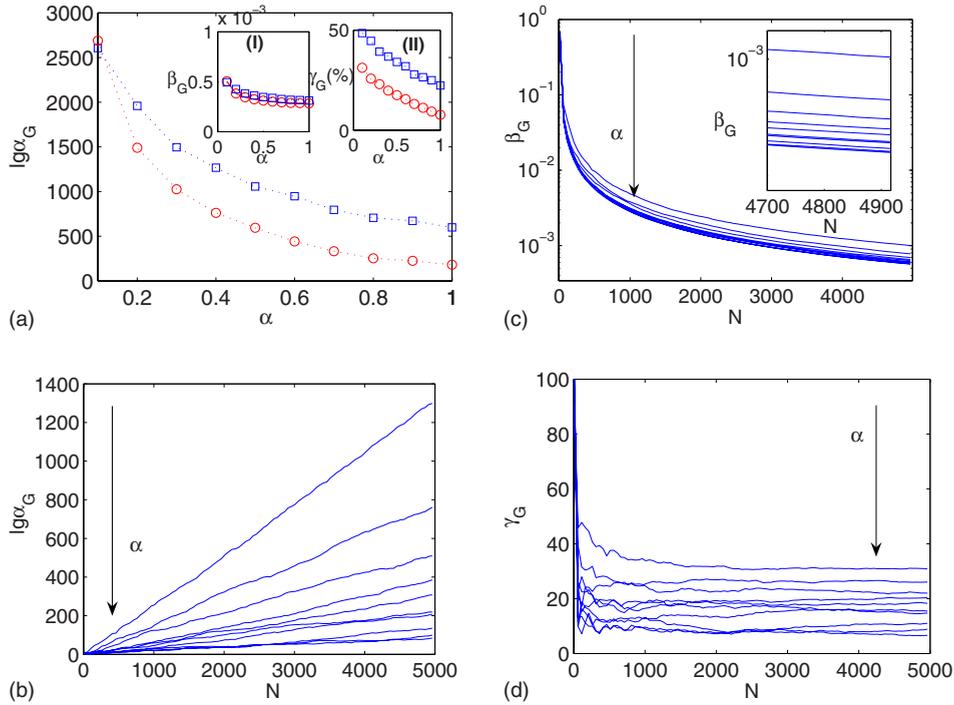


FIG. 7. (Color online) Effect of α and size on symmetry of networks generated by the SLP model. (a) shows a simulation of the SLP model with α varying from 0.1 to 1 in increments of 0.1. The horizontal axis of the figure as well as the two insets (I) and (II) is α ; the vertical axis of the figure is $\log_{10}\alpha_G$, and the vertical axes of the inset figures (I) and (II) are β_G and γ_G , respectively. In the simulations, we use $n_0=10$, $\bar{m}=10$, $t=10\,000$ and employ two kinds of initial-degree distributions. Blue squares (\square) show an exponential distribution $F(m)=a\gamma^{-m}$ with $\gamma=3$; red circles (\circ) show a power-law distribution $F(m)=am^\gamma$ with $\gamma=-1$. (b), (c), and (d) show the growth of symmetry indices including $\log_{10}\alpha_G$, β_G , and $\gamma_G(\%)$ of networks generated by the SLP model. In the simulations shown in (b)–(d), a power-law initial-degree distribution is employed with $\bar{m}=10$ and $\gamma=-1$; we fix n_0 as 10 and vary α from 0.1 to 1 in increments of 0.1 (the arrow shows the direction of increasing α). We vary t from 0 to 5000 and capture snapshots of the network every 50 units of time; thus we could get 100 samples of networks with linearly increasing sizes. Clearly, for all α , the growth of automorphism group size α_G shows an obvious exponential trend, the decrease of β_G shows a power-law trend, and $\lim_{N\rightarrow\infty} \gamma_G=\sigma(\alpha)$. When α varies from 0.1 to 1, all three symmetry indices decrease. The inset of (c) is an amplified local plot, which clearly shows that β_G decreases with the growth of α .

generate SLP networks with α set as 1 to eliminate the ingredient of the similar linkage pattern. We tune the value of the average degree⁸ $\langle k \rangle$ and the exponent γ of the power-law initial-degree distribution, to observe the growth trend of the symmetry indices. The result is shown in Fig. 8, where we can see that only for small average degree $\langle k \rangle$ and large $|\gamma|$ can obvious symmetry be observed. In any other case, the symmetry in the corresponding networks is limited and thus can be ignored.

As shown in Fig. 8, when the average degree $\langle k \rangle$ is small (close to 1), the network has a higher degree of symmetry. Note that those networks with $\langle k \rangle$ closer to 1 tend to have a tree structure, and it is desirable for the tree to have a higher degree of symmetry. This result conforms to the result reported in [22] that BA random trees and uniform random trees have high degrees of symmetry.

As shown in Figs. 8(a)–8(c), for small $|\gamma|$, when $\langle k \rangle$ increases, the symmetry of the network rapidly decays to a constant level. The symmetries at such constant levels are

⁸In our study, for convenience of notation, the average degree $\langle k \rangle$ is defined as M/N , where N and M are the numbers of vertices and edges, respectively. Obviously, $\langle k \rangle$ is one-half of the actual average degree.

determined by the exponents of the power-law initial-degree distribution, which can be validated by the observation that the steeper the double-logarithmic initial-degree distribution is, the higher is the symmetry level.

We can also see that when $\gamma=0$, i.e., the slope of the double-logarithmic distribution plot is zero, the symmetry indices of networks rapidly decay to zero or values close to 0 as $\langle k \rangle$ increases. However, with the slope becoming steep, the symmetry of the network rapidly decays to an approximately constant value far larger than 0. Thus, for a steep log-log initial-degree distribution, an obvious degree of symmetry would be observed.

Such observed symmetries can be naturally interpreted. Note that steeper initial-degree distribution will result in a higher probability of smaller initial degrees m , especially $m=1$. As a result, more treelike symmetry will be found in the structure of the network. The correlation between skewness of the initial-degree distribution and abundance of treelike symmetries can be observed from Fig. 8(d), where the number of $K_{n,1}$ increases with the growth of $|\gamma|$. Furthermore, the data in Table III shows that the complexity and the size of $K_{n,1}$ also increase with the growth of skewness of the double-logarithmic initial-degree distribution.

Thus, it is rational to conclude that solely preferential attachment with the initial degree following a distribution

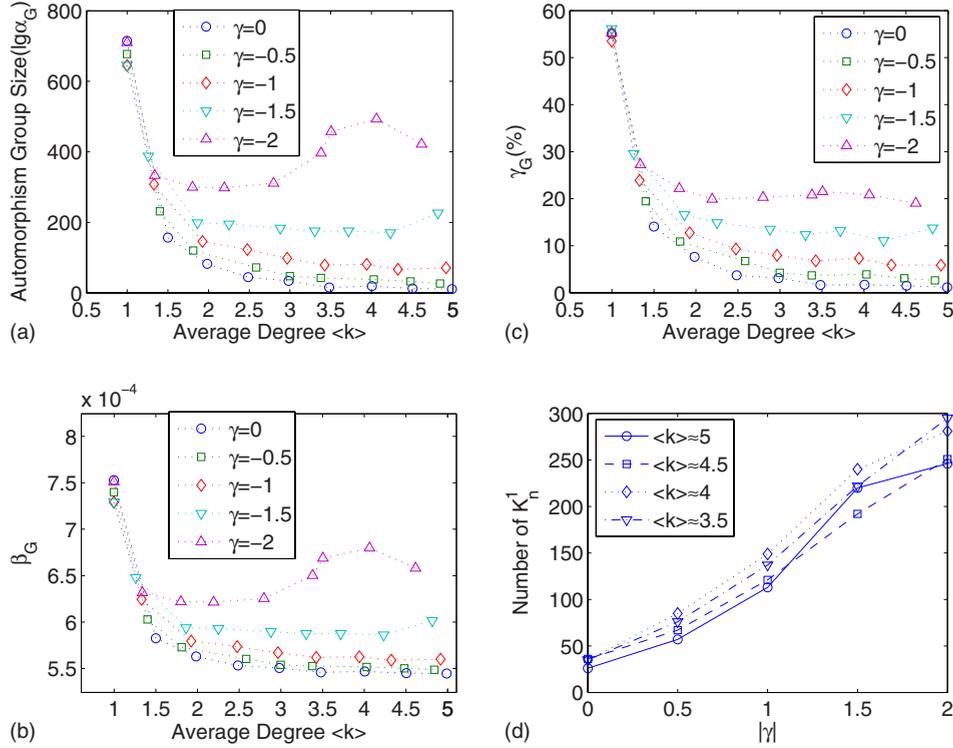


FIG. 8. (Color online) Effect of average degree $\langle k \rangle$ on symmetry of networks generated by preferential attachment with initial degree following a power-law distribution $F(m) = am^\gamma$. In this simulation, we use $\alpha = 1$ to eliminate the influence of similar linkage pattern. Other parameters are set as $n_0 = 10$, $t = 5000$, and $\gamma = 0, -0.5, -1, -1.5, -2$. For each γ , we vary $\langle k \rangle$ from 1 to 5 in increments of 0.5 through increasing \bar{m} . Given a power-law initial-degree distribution $F(m) = am^\gamma$ and upper bound \bar{m} , we can calculate $\langle k \rangle = \frac{\sum_{1 \leq m \leq \bar{m}} m F(m)}{\sum_{1 \leq m \leq \bar{m}} F(m)} = \frac{\sum_{1 \leq m \leq \bar{m}} m a m^\gamma}{\sum_{1 \leq m \leq \bar{m}} a m^\gamma} = \frac{\sum_{1 \leq m \leq \bar{m}} m^{1+\gamma}}{\sum_{1 \leq m \leq \bar{m}} m^\gamma}$. Then, for each γ , for each $\langle k_i \rangle$, we increase \bar{m} from 1 step by step, calculate the $\langle k_{\bar{m}} \rangle$ by the above equation, and, if the value is in the range of $\langle k_i \rangle \pm 0.25$, we let $\langle k_i \rangle \approx \langle k_{\bar{m}} \rangle$. (a)–(c) show the trend of automorphism group size $\log_{10} \alpha_G$, β_G , and γ_G (%) with the growth of the average degree of the network, respectively. It is clear that symmetry of the network will rapidly (superlinearly) decrease to a constant level for less steep initial-degree distribution. (d) shows the relation between the number of local structures $K_{n,1}$ and the slope ($|\gamma|$) of the power-law initial-degree distribution for average degree as one of $\{5, 4.5, 4, 3.5\}$. Parameters in (d) are the same as in (a)–(c). Clearly, with the increase of $|\gamma|$, more $K_{n,1}$ will occur as substructures of the network.

will not necessarily reproduce the symmetry of networks. Only in those cases where small initial degrees have a higher probability of occurrence, especially $m=1$, will produce a treelike symmetry of networks. To reproduce the higher probability of smaller initial degrees, we need to decrease the maximal initial degree or increase the slope of the initial-degree distribution.

V. CONCLUSION

In summary, by studying the statistics of certain local symmetric motifs including (generalized) symmetric bi-

cliques in many real networks, we found that similar linkage pattern plays an important role in the origin of symmetry of networks. To incorporate this ingredient into the BA model, we modified it in two respects: (1) we extended the initial degree from a constant value to a distribution; and (2) we increased the link probability of the target vertices. Simulations showed that similar linkage pattern was responsible for the emergence of symmetry of networks, while preferential attachment with the initial degree following a distribution would only reproduce treelike symmetry in some cases.

TABLE III. Statistics of $K_{n,1}$ in some SLP networks with power-law initial-degree distribution. All the parameters are set the same as in Fig. 8(d). In this table, we record the number and the minimal and maximal sizes of $K_{n,1}$ with γ as one of $\{0, -0.5, -1, -1.5, -2\}$ and $\langle k \rangle$ as one of $\{3.5, 4, 4.5, 5\}$.

$\langle k \rangle$	γ				
	0	-0.5	-1	-1.5	-2
3.5	(35,2,4)	(76,2,11)	(137,2,10)	(222,2,14)	(295,2,26)
4	(35,2,5)	(85,2,5)	(149,2,6)	(240,2,9)	(281,2,41)
4.5	(36,2,4)	(67,2,10)	(121,2,9)	(192,2,18)	(251,2,47)
5	(26,2,4)	(57,2,5)	(113,2,6)	(220,2,16)	(246,2,40)

The fact that the occurrence of symmetric bicliques is statistically significant in real networks strongly suggests that the behavior of individual vertices is not randomized but determined by a certain organization principle, which we referred to as similar linkage pattern. A significant consequence of similar linkage pattern is the richness of symmetry in the whole real network. Symmetry, as a universal property in real networks [22,32], provides a new perspective for exploring the static and dynamic properties of complex systems. From this viewpoint, we have proposed a model of network growth which is able to reproduce the symmetry in real networks. This will make the network modeling of real systems more accurate.

An interesting area for future studies, but lacking in this study, is to perform a comprehensive theoretic analysis of the properties of this network model, which allows us to gain deeper insight into the network-generating model reproducing symmetry.

ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China under Grants No. 60303008, No. 60673133, and No. 60703093 and the National Grand Fundamental Research 973 Program of China under Grant No. 2005CB321905 and by Shanghai Leading Academic Discipline Project Under Project No. B114.

-
- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 [2] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **401**, 130 (1999).
 [3] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149 (2000).
 [4] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
 [5] J. Kleinberg, *Nature (London)* **406**, 845 (2000).
 [6] C.-M. Song, S. Havlin, and H. A. Makse, *Nature (London)* **433**, 392 (2005).
 [7] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
 [8] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).
 [9] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
 [10] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
 [11] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
 [12] M. E. J. Newman, *Eur. Phys. J. B* **38**, 321 (2004).
 [13] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, *Science* **297**, 1551 (2002).
 [14] P. Holme, *Phys. Rev. E* **74**, 036107 (2006).
 [15] P. Holme, *J. Korean Phys. Soc.* **50**, 330 (2007).
 [16] H. Weyl, *Symmetry* (Princeton University Press, Princeton, NJ, 1952).
 [17] K. Mainzer, <http://plato.stanford.edu/entries/symmetry-breaking/>
 [18] Elena Casterllani, *On the Meaning of Symmetry Breaking. Symmetries in Physics: Philosophical Reflections* (Cambridge University Press, Cambridge, U.K., 2003).
 [19] B. Bollobás, *Random Graphs*, 2nd ed., Cambridge Studies in Advanced Mathematics Vol. 73 (Cambridge University Press, Cambridge, U.K., 2001).
 [20] C. Godsil and G. Royle, *Algebraic Graph Theory*, Graduate Texts in Mathematics Vol. 207 (Springer, Berlin, 2001).
 [21] J. Lauri and R. Scapellato, *Topics in Graph Automorphisms and Reconstruction*, London Mathematical Society Student Texts Vol. 54 (Cambridge University Press, Cambridge, U.K., 2003).
 [22] B. D. MacArthur and J. W. Anderson, e-print arXiv:cond-mat/0609274, (2006).
 [23] Y.-H. Xiao, W.-T. Wu, H. Wang, M. Xiong, and W. Wang, *Physica A* **387**, 2611 (2008).
 [24] D. M. Pennock, G. W. Flake, S. Lawrence, E. J. Glover, and C. L. Giles, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5207 (2002).
 [25] N. Biggs, *Algebraic Graph Theory* (Cambridge University Press, London, 1974).
 [26] L. W. Beineke, R. J. Wilson, and P. J. Cameron, *Topics in Algebraic Graph Theory, Encyclopedia of Mathematics and its Applications* (Cambridge University Press, Cambridge, U.K., 2004), Vol. 102, pp. 250–266.
 [27] The GAP group, <http://www.gap-system.org>
 [28] B. D. McKay, *Congr. Numer.* **30**, 45 (1981).
 [29] B. Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics Vol. 184 (Springer, New York, 1998).
 [30] J. Gehrke, P. Ginsparg, and J. M. Kleinberg, *ACM SIGKDD Explorations* **5**, 149 (2003).
 [31] The CAIDA group, <http://www.caida.org/data/active/as-relationships/>
 [32] B. D. MacArthur, R. J. Sánchez-García, and J. W. Anderson, e-print arXiv:0705.3215.
 [33] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, *Nucleic Acids Res.* **34**, D535 (2006).
 [34] S. Maslov, K. Sneppen, and A. Zaliznyak, *Physica A* **333**, 529 (2004).
 [35] P. Erdős and A. Rényi, *Publ. Math. (Debrecen)* **6**, 290 (1959).
 [36] V. Batagelj and A. Mrvar, *Connections* **21**, 47 (1998).
 [37] D. S. Goldberg and F. P. Roth, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4372 (2003).
 [38] E. A. Leicht, P. Holme, and M. E. J. Newman, *Phys. Rev. E* **73**, 026120 (2006).
 [39] P. Jaccard, *Bull. Soc. Vaudoise Sci. Naturelles* **37**, 547 (1901).
 [40] G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer* (Addison-Wesley, Reading, MA, 1989).
 [41] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
 [42] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, *Nat. Genet.* **31**, 64 (2002).
 [43] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
 [44] B. Söderberg, *Phys. Rev. E* **66**, 066121 (2002).