Available online at www.sciencedirect.com
ScienceDirect
Physica A 387 (2008) 2611–2619
PHYSICA A
www.elsevier.com/locate/physa

# Symmetry-based structure entropy of complex networks

Yang-Hua Xiao[a,*], Wen-Tao Wu[a], Hui Wang[b], Momiao Xiong[c,d], Wei Wang[a]

[a] *Department of Computing and Information Technology, Fudan University, ShangHai 200433, PR China*
[b] *Business school, University of Shanghai for Science and Technology, Shanghai 200433, PR China*
[c] *Theoretical Systems Biology Lab, School of Life Science, Fudan University, Shanghai 200093, PR China*
[d] *Human Genetics Center, University of Texas Health Science Center at Houston, Houston TX 77225, USA*

Received 30 September 2007; received in revised form 20 December 2007
Available online 9 January 2008

**Abstract**

Precisely quantifying the heterogeneity or disorder of network systems is important and desired in studies of behaviors and functions of network systems. Although various degree-based entropies have been available to measure the heterogeneity of real networks, heterogeneity implicated in the structures of networks can not be precisely quantified yet. Hence, we propose a new structure entropy based on automorphism partition. Analysis of extreme cases shows that entropy based on automorphism partition can quantify the structural heterogeneity of networks more precisely than degree-based entropies. We also summarized symmetry and heterogeneity statistics of many real networks, finding that real networks are more heterogeneous in the view of automorphism partition than what have been depicted under the measurement of degree-based entropies; and that structural heterogeneity is strongly negatively correlated to symmetry of real networks.


*Keywords:* Structure entropy; Symmetry; Complex network

## 1. Introduction

In recent years, great efforts have been dedicated to the researches on complex networks, due to the fact that many complex systems can be modeled as networks consisting of components as well as relations among these components [1–3]. Previous studies have primarily focused on finding various statistical properties of real networks, especially degree-based statistics, such as degree distribution [4–6], degree correlation [7–9], degree-based structure entropies [10,11]. Studies of many significant properties of networks, such as heterogeneity [4], assortative mixing [12, 13] and self-similarity [14,15], are based on these statistics.

Degree delivers to us the most important information about the number of interconnections of each individual component in the network. However, degree only offers us a view of complex networks in a shallow level, for the reason that vertex partition[1] based on degree is usually coarser than many finer vertex partitions in many networks,

* Corresponding author.
*E-mail address:* shawyanghua@gmail.com (Y.-H. Xiao).

[1] A *partition* of the set $V$ is a set of disjoint non-empty subsets of $V$ whose union is $V$. Elements of a partition are also called its *cells*. A *trivial cell* is the cell with cardinality one. If every cell of a partition is trivial, then the partition is a *discrete partition*; while if the partition only has one cell, the partition is a *unit partition*. For two partitions on set $V$, $P$ and $Q$, if every cell of $P$ is a subset of some cell of $Q$, we say that $P$ is finer than $Q$, or $Q$ is coarser than $P$.

0378-4371/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.01.027

e.g., automorphism partition. In other words, in some networks, vertices with the same degree would be further differentiated from each other, thus forming a finer partition. Consequently, a fascinating problem arises, what will complex network look like if automorphism partition is employed instead of degree partition? Since degree-based statistics underlie many existing studies in complex networks, we believe that studying complex network from the viewpoint of symmetry will lead us to deeper understanding about complex networks.

There is increasing recognition that measuring heterogeneity of complex networks is important in studies of behaviors and functions of complex networks. Works in Ref. [16] have been devoted to directly identifying the regularity or homogeneity of networks to obtain better characterization of complex networks. It has been shown in Ref. [5] that heterogeneity of degree is directly related to the robust-yet-fragile property of scale-free networks, i.e., robustness against random failures of vertices but vulnerability to target attacks. Furthermore, it has been found in Ref. [17] that networks with homogeneous distribution of connectivity are more synchronizable than heterogeneous ones, even though the average network distance is larger.

However, existing heterogeneity measures [10,11,18] of complex networks are all based on degree. Specifically, entropy in Ref. [10] is based on remaining degree [12,13] distribution and entropy in Ref. [11] is based on degree distribution. In fact, degree-based measures of heterogeneity are only the precise quantification of degree heterogeneity of networks, not that of actual heterogeneity in structures of networks. In many cases, degree heterogeneity of networks is only the approximation of structural heterogeneity of networks. For example, as shown in Example 1, in some networks, vertices with the same degree still can be differentiated from each other through measurements on certain structural properties of individual vertex, such as the number of triangles passing through a vertex, the number of shortest paths passing through a vertex (also known as betweenness [19]). Therefore, heterogeneity measured by degree partition cannot precisely describe the structural heterogeneity for any network.

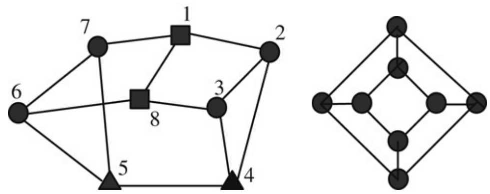
Fortunately, automorphism partitions of networks can naturally partition the vertex set into structurally equivalent cells, thus offering us an ideal alternative to measure the heterogeneity of network structure. An automorphism partition on vertex set of a graph is constructed according to the equivalent relation on vertex set, called as *structural equivalence*,[2] which is defined in the way that two vertices are structurally equivalent to each other if and only if there exists an automorphism transforming one to another, where an automorphism is a permutation on vertex set preserving the adjacency of the whole vertex set. One of the desired properties of automorphism partition is that vertices in each equivalent class will have the same value under many structural measurements on individual vertices,[3] including degree [41], the number of cycles the vertex is incident with [41], the length of the longest path starting from the vertex [41], clustering coefficient of the vertex [27], and so on. In other words, two vertices will not belong to the same cell of automorphism partition provided that these two vertices have different values under any above-mentioned measures on vertices. Hence, automorphism partition offers us a strong power to differentiate vertices from each other from the structure perspective or under certain structural measurements on vertices. Consequently, automorphism partition is the best choice to evaluate structure heterogeneity.

Automorphism partition is one of the core concepts related to the *symmetry in the structure of networks*, which has been widely studied in *algebraic graph theory* [20–23], where the basic idea is to explore the connection between graph theory and algebra, including matrix algebra and group theory [23]. In the community of complex networks, symmetry in real networks has been rarely studied. However, recently, great efforts have been devoted to exploring symmetry in real networks. One of the most important findings is that most real networks are richly symmetric [26, 27], which is surprising since 'almost all' large networks are asymmetric [20,25]. Universal existence of symmetry in real networks strongly begs an explanation, and works in Ref. [28] have found that 'similar linkage pattern', the fact that vertices with the same degree tend to share similar neighbors, is ubiquitous in real networks and responsible for the emergence of symmetry in real networks.

**Example 1.** As shown in Fig. 1, *'cuneane'* and $Q_3$ are all regular graphs with degree 3. Hence, structure of *'cuneane'* and $Q_3$ can be considered as completely homogeneous if degree-based entropy measures are utilized. However,

[2] Note that in Ref. [22], this kind of equivalent relation is also called as 'similar'. In Ref. [42], two vertices are called 'structurally equivalent' to each other, if and only if these two vertices have the same neighbors, which is different from the the meaning of structural equivalence relation defined here.

[3] Note that these measures on vertices are also known as *vertex invariant* [41], or can be characterized as *orbit conserved*, which means that if two vertices are structurally equivalent to each other, they will have the same value under any one of these measures.




(a) Cuneane. (b) $Q_3$.

Fig. 1. Illustration of two 3-regular graphs. Figure (a) shows an abstract structure of molecule known as *'cuneane'*, which is not vertex transitive. Vertices with the same shape belong to the same cell of the automorphism partition. Figure (b) shows an example of 3-cube [20] graph, denoted as $Q_3$, which is vertex transitive.

intuitively, we can see that homogeneity in *'cuneane'* is different from that of $Q_3$. All the vertices in $Q_3$ are equivalent from the structure perspective, thus forming a *unit partition*, and we cannot further partition the vertex set through any other structural measurement on individual vertices. However, in *'cuneane'*, we can easily find that vertices 1 and 8 play a different role from that of vertices 4 and 5 or the remaining vertices, for the reason that 1 and 8 are the only ones not involved in the two triangles, 4 and 5 are the only ones connected by an edge between two different triangles. Therefore, for *'cuneane'*, we can construct a vertex partition $\mathcal{P} = \{\{1, 8\}, \{4, 5\}, \{2, 3, 6, 7\}\}$, which is finer than degree partition. Furthermore, we can validate that partition $\mathcal{P}$ is just the automorphism partition of *'cuneane'*.

## 2. Symmetry-based structure entropy

A graph or network is denoted as $G = G(V, E)$, where $V$ is the set of vertices and $E \subseteq V \times V$ is the set of edges. If $(v_1, v_2) \in E$, then we say that $v_1$ and $v_2$ are adjacent. An automorphism acting on the vertex set can be viewed as a permutation of the vertices in the graph preserving the adjacency of the vertices. The set of automorphisms under the composition of permutations forms a group [20]. In general, a network is *asymmetric* if its automorphism group is the identity group, which only contains an identity permutation; otherwise, the network is *symmetric*. A graph $G = G(V, E)$ is *vertex transitive* (or just transitive) if its automorphism group acts transitively on $V$, which means that for any two distinct vertices of $V$, there is an automorphism mapping one to the other.

Let $\mathrm{Aut}(G)$ be the automorphism group acting on the vertex set $V$. Let $\mathcal{P} = \{V_1, V_2, \ldots, V_k\}$ be the automorphism partition of the graph, which is obtained in the way that $x$ is equivalent to $y$ if and only if $\exists g \in \mathrm{Aut}(G)$, s.t. $x^g = y$. Each cell of automorphism partition is called as an orbit of the automorphism group $\mathrm{Aut}(G)$. In other words, an orbit is the set of all vertices which are obtainable from one another by the actions of the permutations in $\mathrm{Aut}(G)$, which can be denoted as $x^{\mathrm{Aut}(G)} := \{x^g : g \in \mathrm{Aut}(G)\}$ for any $x \in V$ [20]. Automorphism partition offers us an in-depth insight into the heterogeneity of networks. Compared with the degree partition, automorphism partition is much finer in most of the networks.

To accurately measure the structural heterogeneity of complex networks, we define an *entropy based on automorphism partition*, abbreviated as EAP, as follows:

$$\mathrm{EAP} = -\sum_{1 \le i \le |\mathcal{P}|} p_i \log p_i, \tag{1}$$

where $\mathcal{P}$ is the automorphism partition of the network, $p_i$ is the probability that a vertex belongs to the cell $V_i$ of $\mathcal{P}$. Note that given a network's automorphism partition $\mathcal{P} = \{V_1, V_2, \ldots, V_k\}$, we can calculate $p_i$ as:

$$p_i = \frac{|V_i|}{\sum_j |V_j|} = \frac{|V_i|}{N}, \tag{2}$$

where $N$ is the number of vertices in a graph.

Obviously, the maximum value of EAP for networks with $N$ vertices, denoted as $\mathrm{EAP}_{\max}$, equal to $\log(N)$, which is obtained when $p_i = \frac{1}{N}$ for each $1 \le i \le |\mathcal{P}|$, i.e., the graph has a *discrete* automorphism partition. The minimum value of EAP for networks with $N$ vertices, denoted as $\mathrm{EAP}_{\min}$, equal to 0, which is obtained when the automorphism partition of the network is a *unit partition*, implying that all the vertices belong to the same cell or all vertices are structurally equivalent. The maximum value of EAP corresponds to the completely structure-heterogeneous networks,

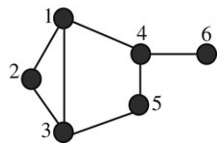



Fig. 2. Illustration of an asymmetric graph. The degree partition $\mathcal{D} = \{\{1, 3, 4\}, \{2, 5\}, \{6\}\}$ is much coarser than automorphism partition, which is a discrete partition in this graph. In the cell {1, 3, 4} of degree partition, all vertices have degree 3, however, vertex 4 is the only one adjacent to a vertex with degree 1, which could distinguish vertex 4 from {1, 3, 4}. Vertex 1 is adjacent to two vertices with degree 3, while vertex 3 is only adjacent to one vertex with degree 3, which could differentiate vertex 1 from vertex 3. Hence, vertices 1, 3, 4 are not structurally equivalent to each other. Vertex 2 and 5 in the cell {2, 5} of degree partition also can be differentiated from each other, because adjacent vertices of vertex 2 and adjacent vertices of vertex 5 are not structurally equivalent, i.e. vertex 1 and vertex 4 are not structurally equivalent.

i.e. asymmetric networks, and the minimum value of EAP corresponds to the completely structure-homogeneous networks, i.e. transitive networks (shown in Fig. 3).

The normalized entropy based on automorphism partition (NEAP) can be defined as:

$$\mathrm{NEAP} = \frac{\mathrm{EAP} - \mathrm{EAP}_{\min}}{\mathrm{EAP}_{\max} - \mathrm{EAP}_{\min}} = \frac{\mathrm{EAP}}{\log N}, \tag{3}$$

where $N$ is the number of vertices in the network.

For comparison, we denote entropy based on remaining degree distribution by ERDD [10] and entropy based on degree distribution by EDD [11]. These two entropies are all defined in the same way as Eq. (1), however, for EDD, $p_i$ is the probability that a vertex has degree $i$; and for ERDD, $p_i$ is the probability that one end of an edge has remaining degree $i$. We also define their corresponding normalized entropies in the form similar to Eq. (3), which are denoted by NERDD and NEDD, respectively. Example 2 illustrates the computation of these three entropies.

**Example 2.** As shown in Fig. 1, since *'cuneane'* is a regular graph, we have EDD = ERDD = NEDD = NERDD = 0. However, the automorphism partition of *'cuneane'* is not a *unit partition*, and we have $p_1 = p_2 = \frac{1}{4}$, $p_3 = \frac{1}{2}$. Thus $\mathrm{EAP} = -\frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{2}\log\frac{1}{2} = \frac{1}{2}\log 8$, $\mathrm{NEAP} = \frac{\frac{1}{2}\log 8}{\log 8} = 0.5$, which is a value larger than 0, thus we could quantify *'cuneane'* as heterogeneous to a certain degree rather than completely homogeneous. Hence, in this case, EAP is more appropriate for quantifying structure-heterogeneity than ERDD and EDD.

The maximum values of ERDD and EDD are both $\log(N)$, however, the maximum values of two entropies correspond to two different kinds of networks, respectively. For EDD, the maximum entropy value corresponds to the completely degree-heterogeneous networks, i.e., networks with $N$ vertices partitioned into $N$ non-empty cells. For ERDD, the maximum entropy value corresponds to the completely remaining-degree-heterogeneous networks, i.e., networks with remaining degree equally distributed.

Completely degree-heterogeneous networks are the most heterogeneous cases under entropy measure of EDD. However, as shown in Fig. 2, completely structure-heterogeneous networks are not necessarily completely degree-heterogeneous, note that the inverse statement necessarily holds true. As long as a network is asymmetric, i.e., the automorphism group contains no non-trivial permutations, the network structure will be completely heterogeneous. Hence, extremely heterogeneous cases should be extended to asymmetric networks to precisely evaluate structural heterogeneity of networks (shown in Fig. 3).

The minimum values of ERDD and EDD both equal to 0, both correspond to regular networks, which are the most homogeneous networks under these two entropy measures. However, as shown in Fig. 1, regular networks can be subdivided into transitive and non-transitive ones, and only transitive networks are the extremely structure-homogeneous networks. Hence, extremely homogeneous cases should be limited to transitive networks if more precise evaluation of structural heterogeneity is desired (shown in Fig. 3).

According to the above facts, the relations between degree-based entropies and symmetry-based entropy can be stated as following statements:

1. $\mathrm{EDD}(G) = \mathrm{EDD}_{\max} \Rightarrow \mathrm{EAP}(G) = \mathrm{EAP}_{\max}$, however, it does not necessarily hold true vice versa;
2. $\mathrm{EAP}(G) = \mathrm{EAP}_{\min} \Rightarrow \mathrm{EDD}(G) = \mathrm{EDD}_{\min}$, however, it does not necessarily hold true vice versa;
3. $\mathrm{EDD}(G) = \mathrm{EDD}_{\min} \Leftrightarrow \mathrm{ERDD}(G) = \mathrm{ERDD}_{\min}$;

where EDD($G$), ERDD($G$) and EAP($G$) represent the EDD, ERDD and EAP of network $G$, respectively.

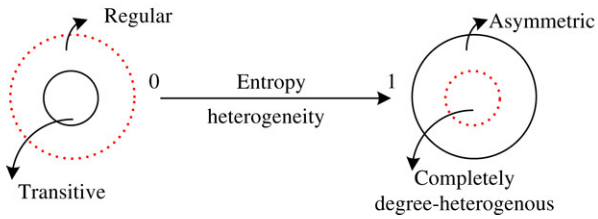
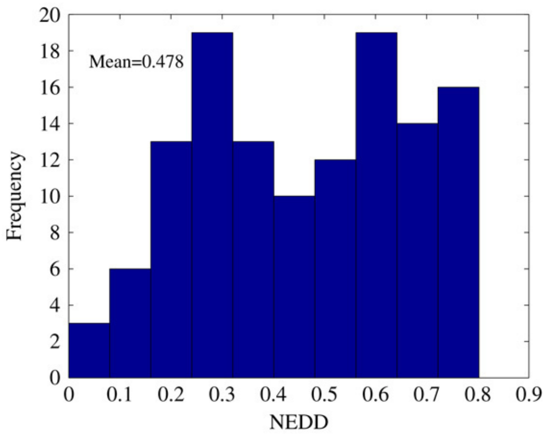
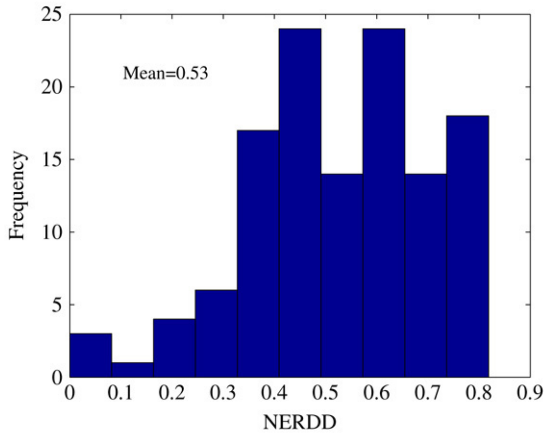


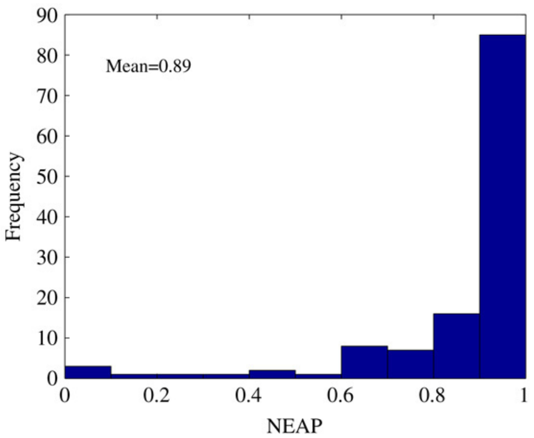

Fig. 3. Illustration of extreme cases under measurements of different entropies. The dotted circles represent the two extreme cases of NEDD. The solid circles represent the two extreme cases of NEAP. The embedding relation between circles expresses the containment relation between corresponding network sets. Note that, the minimum cases of NERDD also can be represented by the left dotted circle, while the maximum cases of NERDD should lie in the middle range of [0, 1] in terms of the measurement of NEDD or NEAP.




(a) NEDD distribution.




(b) NERDD distribution.




(c) NEAP distribution.

Fig. 4. Distribution of values of three entropy measures, NEDD, NERDD and NEAP for 125 real networks.

## 3. Analysis

In this section, we first show that in the view of symmetry, or automorphism partition, most of the real networks should be characterized as more heterogeneous than what have been shown in the view of degree partition. To show this, we calculated NEDD, NERDD and NEAP for 125 real networks. As shown in Fig. 4, NEDD values of real networks tend to lie in the range [0.2, 0.8] (overall 87.2% networks lie in this range) and the mean value is 0.47; NERDD values of real networks tend to lie in the range [0.4, 0.8] (overall 75.2% real networks lie in this range) and

Table 1
Statistics of some real networks and theoretic networks

| Network | $N$ | $M$ | $z$ | $\lg \alpha_G$ | $\beta_G$ | $\gamma_G$ (%) | NEDD | NERDD | NEAP |
|---|---|---|---|---|---|---|---|---|---|
| Technique network | | | | | | | | | |
| USPowerGrid [31] | 4941 | 6594 | 2.67 | 152.71 | $5.90 \times 10^{-4}$ | 16.7 | 0.20 | 0.25 | 0.98 |
| InternetAS[b] | 22442 | 45550 | 4.06 | 11346 | $3.8784 \times 10^{-4}$ | 76.1 | 0.16 | 0.39 | 0.84 |
| Social network | | | | | | | | | |
| arXiv[c] | 27770 | 352285 | 25.37 | 333.26 | $1.01 \times 10^{-4}$ | 3.51 | 0.41 | 0.51 | 0.99 |
| USAir97 [34] | 332 | 2126 | 12.81 | 24.41 | $9.59 \times 10^{-3}$ | 26.20 | 0.539 | 0.68 | 0.95 |
| PairsP [35] | 10617 | 63782 | 12.02 | 632.80 | $2.90 \times 10^{-4}$ | 24.32 | 0.32 | 0.47 | 0.97 |
| foldoc [36] | 13356 | 91471 | 13.6974 | 17 | $2 \times 10^{-4}$ | 0.80 | 0.32 | 0.39 | 1 |
| Erdos02 [37] | 6927 | 11850 | 3.42 | 4222.5 | $1.6 \times 10^{-3}$ | 73.75 | 0.15 | 0.44 | 0.77 |
| Biological network | | | | | | | | | |
| BioGrid-SAC [38] | 5437 | 73054 | 13.43 | 57.79 | $5.12 \times 10^{-4}$ | 3.2739 | 0.48 | 0.61 | 1.00 |
| BioGrid-MUS [38] | 218 | 400 | 3.65 | 126.93 | $4.69 \times 10^{-2}$ | 77.98 | 0.28 | 0.47 | 0.64 |
| BioGrid-HOM [38] | 7522 | 20029 | 5.32 | 935.09 | $4.81 \times 10^{-4}$ | 24.47 | 0.28 | 0.43 | 0.94 |
| BioGrid-DRO [38] | 7528 | 25196 | 6.69 | 624.32 | $4.27 \times 10^{-4}$ | 21.36 | 0.30 | 0.45 | 0.96 |
| BioGrid-CAE [38] | 2780 | 4350 | 3.13 | 829.69 | $1.94 \times 10^{-3}$ | 51.08 | 0.21 | 0.411 | 0.85 |
| ppi [39] | 1870 | 2203 | 4.7123 | 518.6 | $2.7 \times 10^{-3}$ | 53.32 | 0.21 | 0.34 | 0.82 |
| Theoretic networks[d] | | | | | | | | | |
| Star Graph[e] | 2000 | 1999 | 1.99 | 5732.2 | 0.9962 | 99.95 | $5.65 \times 10^{-4}$ | 0.09 | $5.65 \times 10^{-4}$ |
| BA(1) | 2010 | 2000 | 1.99 | 282.09 | $1.90 \times 10^{-3}$ | 56.37 | 0.17 | 0.30 | 0.91 |
| BA(2) | 2010 | 4000 | 3.98 | 0.60 | $1.40 \times 10^{-3}$ | 0.2 | 0.24 | 0.35 | 1 |
| BA(3) | 2010 | 6000 | 5.97 | 0 | $1.35 \times 10^{-3}$ | 0 | 0.28 | 0.39 | 1 |
| BA(4) | 2010 | 8000 | 7.96 | 0 | $1.35 \times 10^{-3}$ | 0 | 0.31 | 0.43 | 1 |
| ER(1) | 2000 | 2081 | 2.08 | 507.97 | $2.4 \times 10^{-3}$ | 34 | 0.225 | 0.228 | 0.89 |
| ER(2) | 2000 | 4002 | 4 | 51.33 | $1.4 \times 10^{-3}$ | 2.65 | 0.276 | 0.274 | 0.99 |
| ER(3) | 2000 | 5923 | 5.90 | 2.07 | $1.36 \times 10^{-3}$ | 0.25 | 0.30 | 0.30 | 1 |
| ER(4) | 2000 | 8137 | 8.14 | 0 | $1.36 \times 10^{-3}$ | 0 | 0.32 | 0.32 | 1 |

Summarized statistics about basic information of networks[a] include the number of the vertices $N$, the number of the edges $M$, the average degree $z$. The key measures quantifying symmetry of networks are also summarized, including the automorphism group size of networks $\alpha_G$ [24] (here, we use $\lg \alpha_G$); the ratio of $\alpha_G$ to the maximal automorphism group size of graphs with $N$ vertices, defined as $\beta_G = (\alpha_G/N!)^{1/N}$ [26,27]; the ratio of the number of vertices in the non-trivial orbits to $N$, defined as $\gamma_G = \frac{\sum_{1\le i\le k, |V_i|>1} |V_i|}{N}$ [28]. We also generate four Barabási–Albert (BA) [4] networks with $m$ (the number of vertices that a new node attach to) varying from 1 to 4 in increment of 1. Also we generate four Erdös–Rényi (ER) [29] networks with average degree approximately as one of {2, 4, 6, 8}, using `Pajek` [30]. Here, we use `nauty` [40], one of the most efficient graph isomorphism algorithms, to calculate the size and structure of various automorphism groups.

[a] All the networks are preprocessed as an undirected, unweighted graphs without any self-loops and multi-edges. Note that the whole network instead of the largest connected component is used.

[b] Here, the snapshot at 2006-07-10 of CAIDA [32] is used.

[c] Here, the snapshot at 2003-04 of HEP-TH (high-energy physics theory) citation network [33] is used.

[d] Certain degree of symmetry can be observed from simulated BA and ER networks with small average degrees, which can be attributed to the following two reasons: (1) when average degree is small, the structure is close to tree, consequently producing tree-like symmetry [26,28]; (2) we do not use the largest connected components but instead the whole networks, which probably contain isolated vertices or components when average degrees are small, and these isolated vertices or components will contribute greatly to symmetries in networks.

[e] A star graph is a bipartite in the form of $K_{1,n}$.

the mean value is 0.53; while NEAP of real networks primarily lies in the range [0.8, 1] (overall 80.8% real networks lie in this range) and the mean value is 0.89, close to 1. In addition, for almost all the tested real networks, the value of NEAP is larger than that of NEDD and NERDD, which is shown in Fig. 5. Hence, from these observations, we can see that real networks are very heterogeneous in the view of automorphism partition, and real networks will have a large probability (larger than 80% in our samples) to be quantified with a NEAP value larger than 0.8.

We also need to note that some real networks characterized as very homogeneous in the view of degree partition have been quantified as very heterogeneous in the view of symmetry. As shown in Table 1, for almost all the real networks, the corresponding values of the degree-based entropies are less than 0.5, except for the NERDD of arXiv, USAir97, BioGrid-SAC, and NEDD of USAir97; while for all the networks, the corresponding values of NEAP are

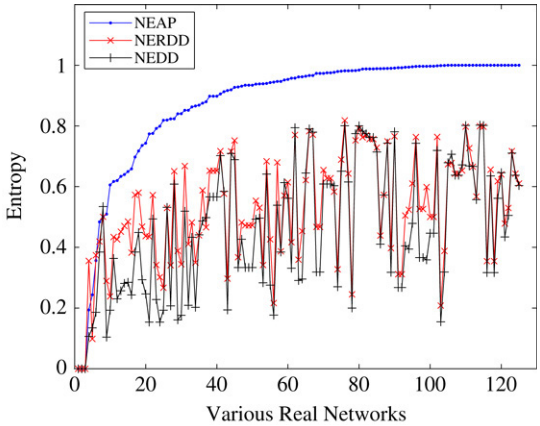



Fig. 5. Comparison of three entropy measures. The horizontal axis represents various networks in the ascending order of corresponding NEAP values.

larger than 0.6 and most of them are larger than 0.8. If the median value of range [0, 1] is taken as the critical value indicating whether a network is heterogeneous, then under the measure of degree-based entropies real networks tend to be quantified as homogeneous; however, under the measure of symmetry-based entropy real networks tend to be quantified as very heterogeneous.

Since heterogeneity based on automorphism partition describes the structure heterogeneity more accurately than degree-based heterogeneity, it is reasonable to believe that most of the real networks are very heterogeneous in their structures.

In Table 1, statistics of some theoretic networks are also summarized. We can see that NEAP of star graph is close to 0, indicating that star graph is very homogeneous. Indeed, in a star graph, all vertices except for the central vertex lie in the same cell of automorphism partition, thus it is natural that star graph is very homogeneous.

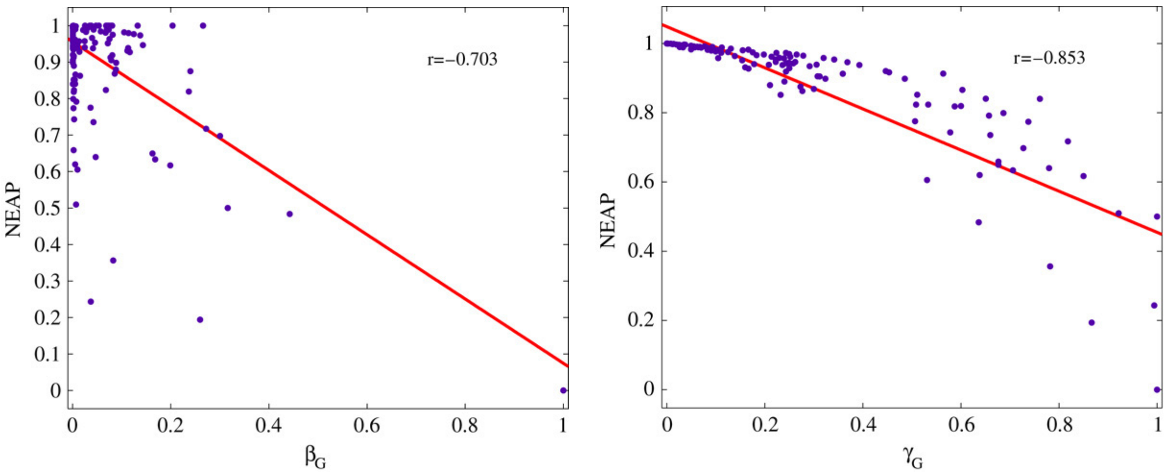
Generally, Barabási–Albert (BA) [4] scale-free networks are considered to be more heterogeneous than Erdös–Rényi (ER) [29] random networks [4], which is based on the intuitive observation that BA scale-free networks are right-skewed in double-log degree distribution while the degrees of ER random networks are exponentially distributed with an obvious scale. However, no quantification or theoretic proof has been provided to verify the above notion, which can be partly attributed to the lack of appropriate measures of heterogeneity of real networks. However, utilizing three entropy measures, we can see that under the measurement of NEDD and NERDD, the difference of heterogeneity between BA networks and ER networks are very small, less than 0.05; and under the measure of NEAP, both of them tend to be quantified as very heterogeneous, the difference is less than 0.02.

Next, we will show that structural heterogeneity is strongly negatively correlated to the symmetry of networks, which means that the less symmetric a network is, the more structure-heterogeneous the network is. As shown in Fig. 6, strong negative relation could be observed from the $\beta_G -$ NEAP and $\gamma_G -$ NEAP correlation curves. In fact, if a network is very symmetric, vertices in the network will have a high probability to be equivalent in the structure, thus the automorphism partition will be close to a *unit partition*, which is extremely homogeneous. Conversely, if the network is close to an asymmetric network, vertices can be easily differentiated from each other from the structural perspective, leading to a nearly discrete automorphism partition. Consequently, the whole network tends to be structure-heterogeneous.

## 4. Conclusion

We have shown that entropies based on degree partition cannot precisely describe the structural heterogeneity of complex networks in many cases due to its inability to differentiate vertices with the same degree. Instead, due to the strength of automorphism partition that can naturally partition vertex set into equivalent cells from the structure perspective, entropy based on automorphism partition can quantify the heterogeneity of networks more accurately.

Networks with extreme heterogeneity and homogeneity under different entropy measures, including two degree-based entropies and symmetry-based entropy, have been analyzed, showing that symmetry-based entropy is more accurate in quantifying the heterogeneity or disorder of a network system than degree-based entropies. We also




(a) Correlation between NEAP and $\beta_G$.

(b) Correlation between NEAP and $\gamma_G$.

Fig. 6. NEAP appears to be negatively correlated to $\beta_G$ and $\gamma_G$, and corresponding correlation coefficients are −0.703 and −0.853, respectively. 125 Real networks and 28 theoretic networks, overall 153 samples, are used.

calculated symmetry and heterogeneity statistics for hundreds of real networks and several theoretic networks, and found that real networks are more heterogeneous in the view of automorphism partition than what have been depicted under the measurement of degree-based entropies. We also found that structural heterogeneity measured by entropy based on automorphism partition is highly negatively correlated to the abundance of symmetry in real networks.

Generally, heterogeneity of networks is strongly correlated to the complexity of a network system, i.e., more heterogeneous, more complex. Thus, we believe that precisely characterizing the heterogeneity of a network can allow us to gain deeper insight into the complexity of systems represented by networks.

## Acknowledgements

The work was supported by the National Natural Science Foundation of China under Grant No. 60303008, No. 60673133, No. 60703093; the National Grand Fundamental Research 973 Program of China under Grant No. 2005CB321905; Natural Science Foundation for Colleges and Universities of Jiangsu Province No. 05KJB520107.

## References

[1] R. Albert, A.-L. Barabási, Rev. Modern Phys. 74 (2002) 47–97.
[2] M.E.J. Newman, SIAM Rev. 45 (2003) 167–256.
[3] S.H. Strogatz, Nature (London) 410 (2001) 268–276.
[4] A.-L. Barabási, R. Albert, Science 286 (1999) 509–512.
[5] R. Albert, H. Jeong, A.-L. Barabási, Nature (London) 406 (2000) 378.
[6] K.-I. Goh, B. Kahng, D. Kim, Phys. Rev. Lett. 87 (2001) 278701.
[7] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. 87 (2001) 258701.
[8] S. Maslov, K. Sneppen, Science 296 (2002) 910.
[9] J. Berg, M. Lassig, Phys. Rev. Lett. 89 (2002) 228701.
[10] R.V. Solé, S.V. Alverde, Lect. Notes Phys. 650 (2004) 189.
[11] B. Wang, H.W. Tang, C.H. Guo, Z.L. Xiu, Physica A 363 (2005) 591.
[12] M.E.J. Newman, Phys. Rev. Lett. 89 (2002) 208701.
[13] M.E.J. Newman, Phys. Rev. E 67 (2003) 026126.
[14] C.-M. Song, S. Havlin, H.A. Makse, Nature 433 (2005) 392–395.
[15] Z.-Z. Zhang, S.-G. Zhou, T. Zou, Eur. Phys. J. B 56 (2007) 259–271.
[16] L.da F. Costa. arxiv:physics/0702102.
[17] T. Nishikawa, A.E. Motter, Y.C. Lai, F.C. Hoppensteadt, Phys. Rev. Lett. 91 (2003) 014101.
[18] L.da F. Costa, F.A. Rodrigues, G. Travieso, P. Villas Boas, Adv. Phys. 56 (2007) 167–242.
[19] L.C. Freeman, Sociometry 40 (1977) 35–43.
[20] C. Godsil, G. Royle, Algebraic Graph Theory, in: Graduate Texts in Mathematics, vol. 207, Springer, 2001.
[21] N. Biggs, Algebraic Graph Theory, Cambridge University Press, London, 1974.

[22] J. Lauri, R. Scapellato, Topics in Graph Automorphisms and Reconstruction, in: London Mathematical Society Student Texts, vol. 54, Cambridge University Press, Cambridge, 2003.

[23] L.W. Beineke, R.J. Wilson, P.J. Cameron, Topics in Algebraic Graph Theory, in: Encyclopedia of Mathematics and its Applications, vol. 102, Cambridge University Press, Cambridge, 2004.

[24] B. Bollobás, Modern Graph Theory, in: Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998.

[25] B. Bollobás, Random Graphs, second ed., in: Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[26] B.D. MacArthur, J.W. Anderson. arXiv:cond-mat/0609274.

[27] B.D. MacArthur, R.J. Sánchez-García, J.W. Anderson. arXiv:0705.3215.

[28] Y.-H. Xiao, M.-M. Xiong, W. Wang, H. Wang. arXiv:0709.1249, 2007.

[29] P. Erdös, A. Rényi, Publ. Math. (1959) 290–297.

[30] V. Batagelj, A. Mrvar, Connections 21 (1998) 47–57.

[31] D.J. Watts, S.H. Strogatz, Nature (London) 393 (1998) 440–442.

[32] The CAIDA Group, The caida as relationships dataset, 2003–2007. http://www.caida.org/data/active/as-relationships/.

[33] J. Gehrke, P. Ginsparg, J.M. Kleinberg, SIGKDD Explorations 5 (2003) 149–151.

[34] North American Transportation Atlas Data. http://www.bts.gov/.

[35] D.L. Nelson, C.L. McEvoy, T.A. Schreiber, The University of South Florida Word Association, Rhyme and Word Fragment Norms. http://w3.usf.edu/FreeAssociation/.

[36] V. Batagelj, A. Mrvar, M. Zaveršnik, Network analysis of texts, Language Technologies 2002, Ljubljana, 14–15th October 2002, pp. 143–148.

[37] V. Batagelj, A. Mrvar, Social Networks 22 (2000) 173–186.

[38] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, Nucl. Acids Res. 34 (2006) D535–D539.

[39] H. Jeong, S.P. Mason, A.-L. Barabási, Z.N. Oltvai, Nature (London) 41 (2001) 411.

[40] B.D. Mckay, Congr. Number. 30 (1981) 45–87.

[41] G. Tinhofer, M. Klin, Algebraic combinatorics in mathematical chemistry. Methods and algorithms. III, Graph Invariants and Stabilization Methods (Preliminary Version), Technical Report, TUM-M9902, Technische Universitat Munchen, 1999.

[42] P. Holme, M. Huss, J. Roy. Soc. Interface 2 (2005) 327.