Journal of Computer Science & Systems Biology - Open Access
Research Article JCSB/Vol.1/ 2008

Symmetry of Metabolic Network

Hua Dong [1, 2], Yanghua Xiao [3], Wei Wang[3], Li Jin[1, 4], Momiao Xiong[1, 2 *]

[1]Laboratory of Theoretical Systems Biology and Center for Evolutionary Biology, State Key Laboratory of Genetic Engineering, School of Life Science, Fudan University, Shanghai 200433, China

[2]Human Genetics Center, University of Texas School of Public Health, Houston, TX 77030, USA

[3] Department of Computing and Information Technology, Fudan University, Shanghai 200433, China

[4]Chinese Academy of Science-Max-Planck-Gesellschaft Partner Institute for Computational Biology, Shanghai Institutes for Biological Science, CAS, Shanghai, 200433, China

*Corresponding author: Momiao Xiong, Human Genetics Center, University of Texas School of Public Health, Houston, TX 77030, USA, Tel: 713-500-9894; Fax: 713-500-0900; E-mail: momiao.xiong@uth.tmc.edu

Present addresses of authors Hua Dong, Yanghua Xiao, Wei Wang, Li Jin: 220 Handan Rd., Shanghai, 200433P.R.China, Momiao Xiong: 1200 Hermann Pressler Dr., Houston, TX, 77030

Received December 14, 2008; Accepted December 24, 2008; Published December 26, 2008

Citation: Hua D, Yanghua X, Wei W, Li J, Momiao X (2008) Symmetry of Metabolic Network. J Comput Sci Syst Biol 1: 001-020. doi:10.4172/jcsb.1000001

Abstract

Previous studies of properties of metabolic works have mainly focused on the statistic properties of networks, including the small world, and power-law distribution of node degree, and building block of network motifs. Symmetry in the metabolic networks has not been systematically investigated. In this report, symmetry in directed graph was introduced and an algorithm to calculate symmetry in directed and disconnected graphs was developed. We calculated several indices to measure the degree of symmetry and compared them with random networks. We showed that metabolic networks in KEGG and BioCyc databases are generally symmetric and in particular locally symmetric. We found that symmetry in metabolic networks is distinctly higher than that in random networks. We obtained all the orbits in networks which are defined as structurally equivalent nodes and found that compound pairs in the same orbit show much more similarity in chemical structures and function than random compound pairs in network, which suggests that symmetry in the metabolic network can generate the functional redundancy, increase the robustness and play an important role in network structure, function and evolution.

Keyword: Symmetry; Metabolic network; Structural equivalence; Orbit

Abbreviations

NCS: Number of Connected Subgraphs

MCS: proportion of nodes in the Maximum Connected Subgraph

ECS: Entropy based on Connected Subgraph

NECS: Normalized Entropy based on Connected Subgraph

Introduction

Metabolic networks are composed of consecutive chemical reactions to produce energy and various molecules. They are represented as directed hyper-graphs, with compounds and(or) enzymes as nodes and the reactions activated by the enzymes as hyper-arcs. How to characterize the structure of metabolic networks and how to link their structure

with function are important in gaining deep understanding of metabolic networks. In the last decade, we have witnessed the great progress in theories and models of complex networks, which provide new powerful tools for study of metabolic networks. Previous research in complex networks have primarily focused on finding the statistical properties of various networks, such as small world properties (Jeong et al., 2000; Ma and Zeng, 2003; Wagner and Fell, 2001; Watts and Strogatz, 1998); power-law distribution of node degree (Mahadevan and Palsson, 2005; Samal et al., 2006); building block of network motifs (Eom et al., 2006; Milo et al., 2002) and hierarchical structure of the network topology (Guimera and Nunes, 2005; Ravasz et al., 2002). A lot of research exploiting the theory or model of complex networks has been dedicated towards metabolic networks. Jeong et al., (2000); Ma and Zeng, (2003) calculated the average path length of the metabolic networks and concluded that metabolic network belongs to small-world network. The small-world characteristic implies that information and energy can be rapidly transferred to the whole network and the cell can response quickly to perturbation of environments. Jeong et al., (2000) also calculated the degree distribution and concluded that metabolic network follows the power-law distribution with exponential index $\gamma \approx 2.2$. However, the small-world property is still disputing (Arita, 2004). Milo et al., (2002) introduced the concept of 'network motifs' and developed algorithms for their identification. Eom et al., (2006) applied the concept of network motifs to metabolic network and identified the network motifs and the statistically significant subgraph patterns as well. Ravasz et al., (2002) (Clauset et al., 2008; Guimera and Nunes, 2005) proposed the hierarchical-modular model according to the characteristics of metabolic network. They calculated the average clustering coefficient for 43 different organisms as a function of the number of distinct substrates present in their metabolism. They found that, for all 43 organisms, the average clustering coefficient is about an order of magnitude larger than expected for a scale-free network of similar size.

However, symmetry, a universal property of real networks, has been rarely studied for metabolic network. Symmetry characterizes the invariance under possible transformations and implies conservation laws of nature (Hatcher, 2002; MacArthur, 2008; MacArthur and Anderson, 2006). Symmetry provides redundancy and robustness against perturbation of environments. There is increasing recognition that the universal evolution is caused by symmetry break, generating diversity and increasing complexity and energy (Mainzer, 2005; Quack, 2003). Symmetry break is often followed by addition of functional modules that usually show local symmetry, increasing network robustness (Felder et al., 2001). Until recently, after the concept of symmetry based on the automorphism has been utilized to explore real networks, quantitative methods for investigation network symmetry have been developed. MacArthur, (2008); MacArthur and Anderson, (2006) first found that large real networks are quite symmetric and such symmetry in real networks can be attributed to the local symmetry which is hidden in local substructures. Xiao et al., (2008b); Xiao et al., (2008c) then proposed a principle referred to as "similar linkage pattern", which means that nodes with similar properties such as degree tend to have similar linkage targets, to explain the emergence of symmetry. In (Mahadevan and Palsson, (2005), symmetry is exploited to characterize the structural heterogeneity of complex networks. Symmetry in real networks has been further used to characterize the simplicity hidden in networks and consequently has been utilized to simplify the network while reserving many key properties of original networks, such as complexity and communication (Xiao et al., 2008a).

To date, symmetry in metabolic networks and the relations between the structural symmetry and function of the network have rarely been investigated. It is still unknown whether symmetry exits in metabolic networks. If it does, the existence of such symmetry also begs a biological explanation. Purposes of this report are (1) to examine the symmetry of the metabolic networks, (2) to measure the abundance of the symmetry in metabolic networks, (3) to obtain the orbits (structurally equivalent nodes) through restricted network quotient and (4) to explore biological implication of the symmetry of the metabolic networks. To accomplish this, we first reconstruct metabolic networks for 705 organisms in KEGG and 373 in BioCyc databases. Then, we study the influence of connectivity of the reconstructed networks on the symmetry of metabolic networks. Previous works about symmetry in the general networks have focused on undirected graphs. However, the metabolic networks are usually handled as directed graphs. Hence, it's necessary to explore symmetry in directed graphs and develope algorithms to find symmetry in the directed and disconnected networks. Based on these results, we then systematically investigate the properties of symmetry in metabolic networks, including the degree of symmetry, restricted network quotient. To explore functional implications of the structural symmetry of metabolic networks, we test the chemical structure similarity of the symmetric compounds in metabolic networks of 705 organisms which allow us to reveal the relationships between the network symmetry and its function.

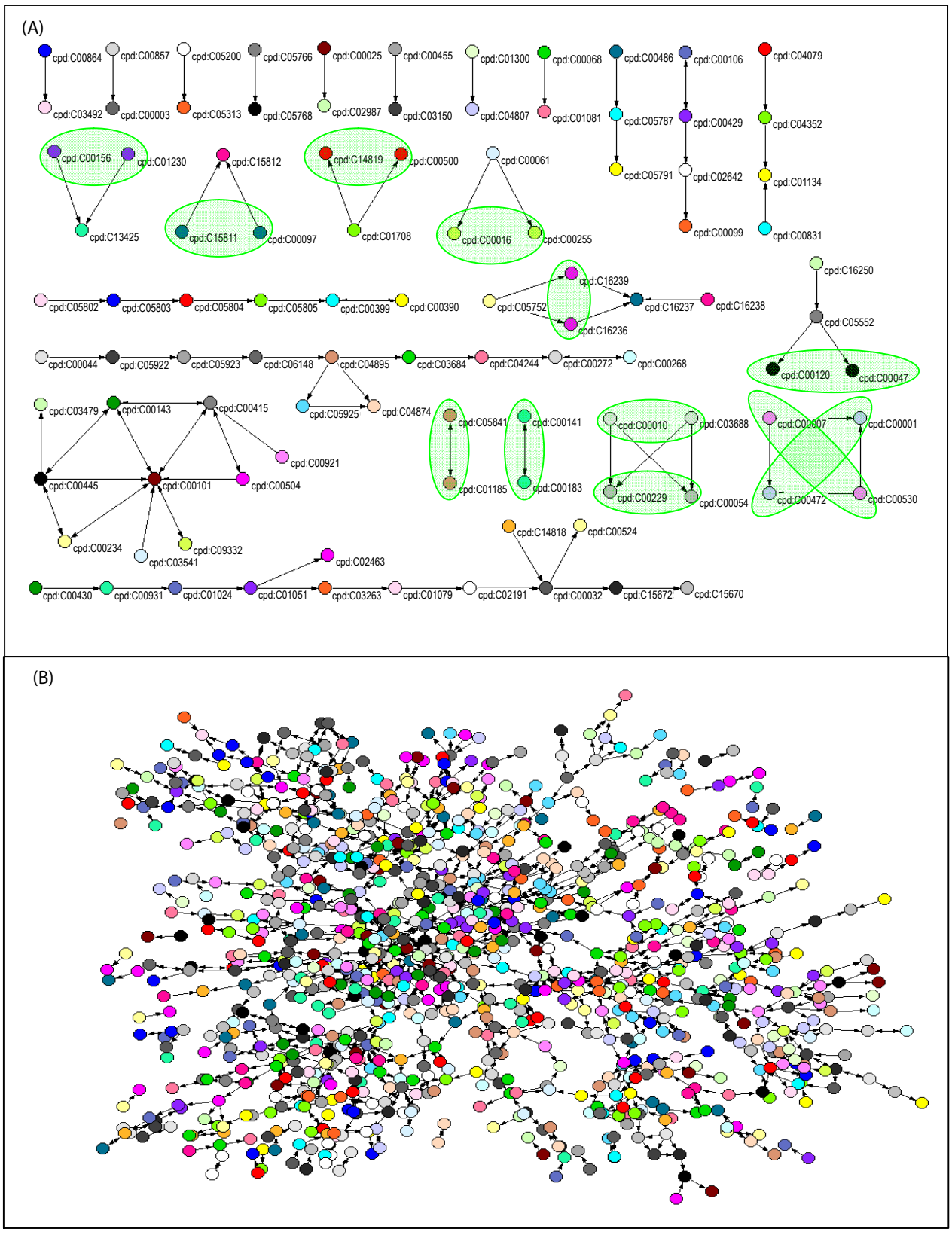


Figure 1: KEGG Metabolic network reconstruction of Drosophila melanogaster

(A) Metabolism of Cofactors and Vitamins module for Drosophila melanogaster (fruit fly) Nodes: 100, Edges:96. The nodes represented compound ID in KEGG database. The nodes in different orbit are marked with different colours and non-trivial orbits are marked in green ellipse. We can see that there are 25 connected subgraphs and 12 orbits in the module. (B) All 100 metabolic pathways of Drosophila melanogaster were integrated into a metabolic network (Nodes: 1050, Edges:1340). The nodes in different orbit are marked with different colours. The compound ID is not shown in the figure.

Results and Discussion

Reconstruction of Metabolic Network

A metabolic network is represented as a directed graph $G(N,E)$ with node set N representing compounds and the edge set E representing the chemical reactions which the compounds participate in. The direction of each edge implies the direction of the chemical reaction. We downloaded the metabolic network data from two major metabolic network databases: KEGG (Kanehisa and Goto, 2000) and BioCyc.

KEGG is a collection of simplified metabolic networks which are manually drawn pathway maps representing knowledge on reactions, while currency metabolites such as $H_2O$and ATP are not included. Xml files of all metabolic pathways for 705 organisms were downloaded from KEGG FTP. We reconstructed the metabolic networks according to the reactions data extracted from the xml files and visualized the network by Pajek (Batagelj and Mrvar, 2003). According to KEGG metabolic functions classification, we integrated single pathways into functional modules. Finally we integrated 11 functional modules into a whole metabolic network. See Figure 1 (a) (b) for Drosophila melanogaster's Cofactors and Vitamins Metabolism module and the integrated metabolic network.

The BioCyc collection of Pathway/Genome Databases (DBs) provides electronic reference sources on the pathways and genomes of different organisms. We collected metabolic pathways of 373 organisms, extracted the reaction data of each organism and integrated the reactions to a network for each organism. Direction of reactions is not given in BioCyc database but currency metabolites are included in. So we finally got 373 undirected graphs including currency metabolites from BioCyc. We first analyze symmetry of KEGG networks and replicate our experiment for BioCyc networks to validate whether our conclusions are ubiquitous in metabolic networks. See additional files 1 for organisms and pathways in KEGG and BioCyc.

The Connectivity of Metabolic Networks

For the integrated networks of 705 organisms in KEGG, the number of the connected subgraphs (NCS) varies from 1 to 271. Only 5 networks (0.7%) contain less than 10 connected subgraphs. The maximum connected subgraph (MCS) contains 7.8%-100% nodes of the whole network. The proportion of MCS, which is defined as the ratio of the number of the nodes in MCS over the total number of nodes in the network, in 99.6% and 56% of the constructed metabolic networks is less than 80% and 60%, respectively. So we can conclude that the connectivity of metabolic is quite low. We introduced normalized entropy based on the connected subgraph (NECS) to measure the degrees of the connectivity of the network (See Materials and Methods). The larger NECS, the less the network connected. NECS value of the metabolic networks ranges from 0 to 0.778064 with the mean value 0.410917.

For the integrated networks of 373 organisms in BioCyc, the number of the connected subgraphs (NCS) varies from 2 to 76, which is distinctly less than that in KEGG. The maximum connected subgraph (MCS) contains 92.1%-99.3% nodes of the whole network, definitely larger than that in KEGG. NECS value of the metabolic networks ranges from 0.007 to 0.075 with the mean value 0.0347128. As the currency metabolites were included in BioCyc , the connectivity of metabolic network is significantly increased.

To gain deeper understanding of the connectivity in metabolic networks, we compared it with the random networks. For each real graph, we generate 100 randomized networks by rewiring the local edges (Maslov, 2004). Then we compute the MCS, NCS and NECS for every network and summarize the mean and variance over the 100 randomized networks. From the error bar in Figure 2, we can see clearly that: most of the NCS in real metabolic network is larger than that in random network (89.8%); most of the MCS in real metabolic network is lower than that in random network (96.9%); NECS in real network is obviously larger than the corresponding random network (96.9%). We also compared the connectivity of BioCyc network with their random networks using the same method. In spite of relatively larger connectivity compared with that in KEGG, the connectivity in BioCyc metabolic network is still smaller than that in random networks. The results are shown in Supplementary Figure 1. The results of NCS, MCS and NECS value for 705 real networks and their corresponding randomized networks for both KEGG and BioCyc networks were shown in additional data file 2. The results above imply that the connectivity in metabolic network is rather small. Consequently connectivity has to be taken into account when exploring the network reduction of metabolic networks.

Symmetry in Metabolic Networks

Given a metabolic network $G(N,E)$, a one-to-one mapping, or bijective mapping, from N onto itself is called a permutation on N. Two nodes are adjacent if there is an arc from one node to the other node. Among all the permutations in $S(N)$, where $S(N)$ is the set of permutations acting

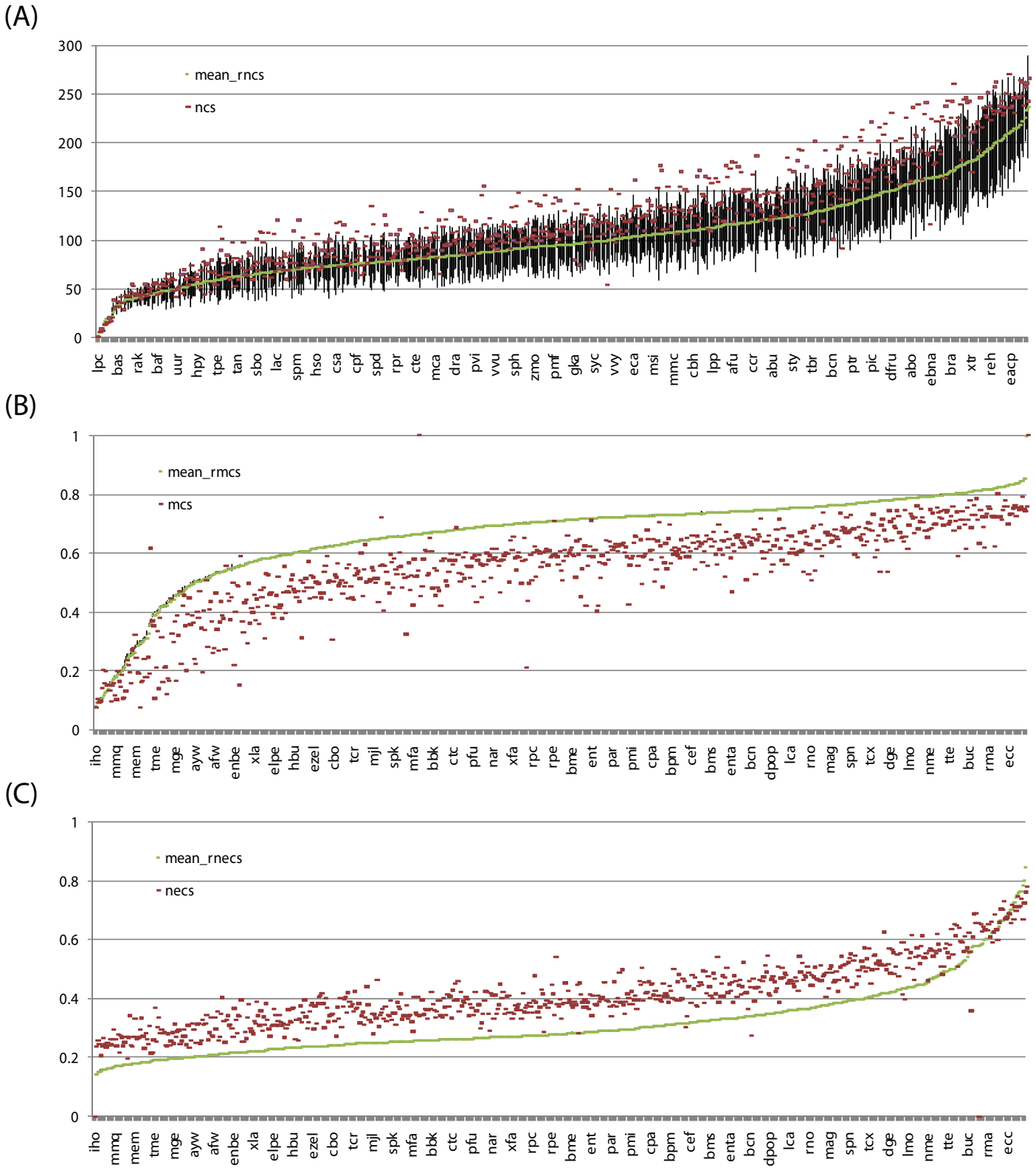
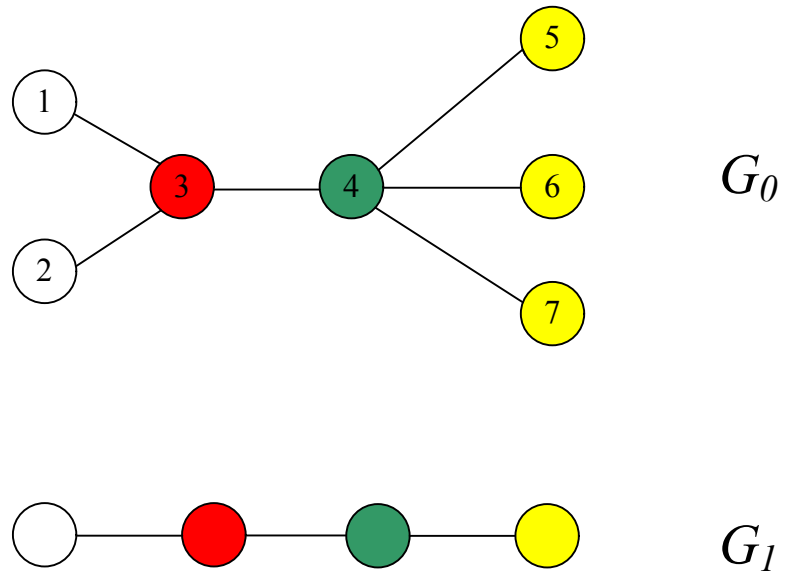
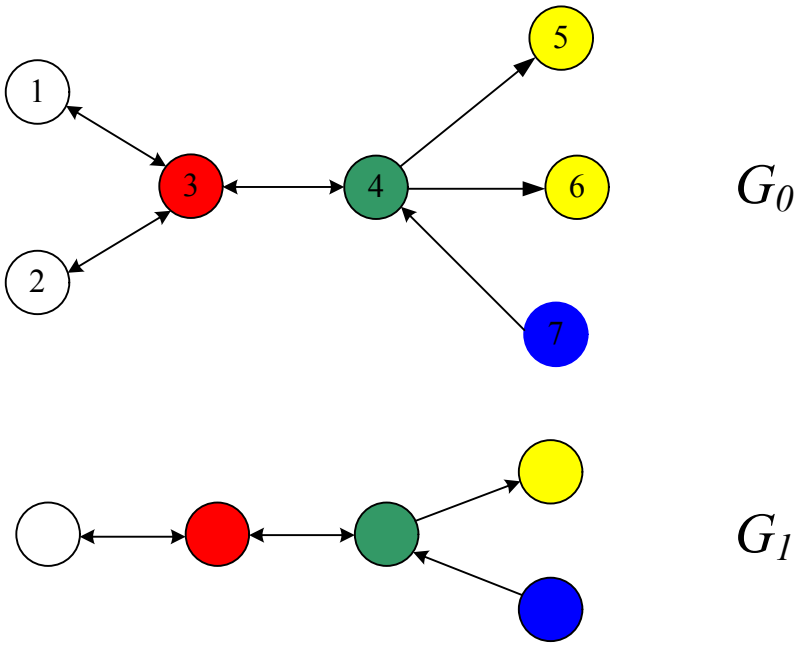


Figure 2: Comparison of the connectivity of metabolic network in KEGG with random network

The random networks are produced by randomly rewiring the local edge in the given real networks. (A) NCS of the real metabolic networks and their random networks. (B) MCS of the real metabolic networks and their random networks. (C) NECS of the real metabolic networks and their random networks. Please note that due to the tiny variances of MCS and NECS of random networks, the error bar looks like the scatter line.



(A)



(B)

Figure 3: The automorphism partition of underlying graph and the directed graph

(A) Underlying graph (B) The directed graph. Different orbits are marked with different colours.

on N, some permutations can preserve the adjacency of the nodes and these permutations are called automorphisms acting on the node set. The set of automorphisms under the product of permutations forms a group: $Aut(G)$ (Tinhofer and Klin, 1999).Two nodes x and y are automorphically equivalent to each other if there is an automorphism that transforms node x to node y. In the context without confusion, such equivalence relation of nodes is also called structural equivalence. A set of structurally equivalent nodes is defined as an orbit of $Aut(G)$. According to such equivalent relation on node set, we can get a partition $P=\{N_1,N_2,\ldots N_m\}$, called as automorphism partition, which is composed of orbit sets $N_1,N_2,\ldots N_m$. An orbit is non-trivial when it contains more than one node, otherwise it's trivial. A network is called symmetric if we can find at least one non-trivial orbit in this network, otherwise the network is asymmetric. The quotient graph of an undirected graph is defined as a reduced graph which has every orbit (structurally equivalent nodes) as its new node and adds an edge to connect two nodes if there is at least one edge from any one node in the orbit to any one node in another orbit. The quotient graph of a directed graph is similar to that of undirected graph except that the direction of the arcs is preserved in the quotient of directed graph.

Consider the undirected graph G_0 in Figure 3(A), the automorphism partition is $P_0=\{N_1,N_2,N_3, N_4\}$, where $N_1=\{1,2\}$, $N_2=\{3\}$, $N_3=\{4\}$, and $N_4=\{5,6,7\}$. There are four orbits which are classified by different colours. The quotient graph of G_0 is shown in G_1. The four orbits of G_0 are reduced to four nodes in G_1 , and as long as there is an edge between any two orbits of G_0, the corresponding nodes in $G_1$are adjacent. For directed graph G_0 in Figure 3(B) , the automorphism partition is $P_0=\{N_1,N_2,N_3,N_4,N_5\}$, where $N_1=\{1,2\}$, $N_2=\{3\}$, $N_3=\{4\}$, and $N_4=\{5,6\}$ and $N_5=\{7\}$. The five orbits of G_0 are reduced to five nodes in G_1 , and as long as there is an arc from one orbit to another orbit in G_0, there is an arc from one corresponding node to another corresponding node in G_1 (shown in Figure 3(B)). Please note that in the directed graph G_0 in Figure 3(B), the edges between orbits N_1 and N_2 and that between N_2 and N_3 are both bidirectional, which determines that the edge between corresponding nodes in G_1 are bidirectional. Because the direction of arc <7, 4> is different from that of <4, 5> and <4, 6>, nodes 5 and 6 belong to the fourth orbit while node 7 belongs to the fifth orbit.

Consider Figure1 (A) where we take the Drosophila melanogaster's Cofactors and Vitamins metabolism module as a real example. There are 100 nodes and 96 arcs in this module. The NCS (number of connected subgraph) is 25, which implies the connectivity of this network is very small. There are 12 non-trivial orbits in the network, each of which is marked in green ellipse.

To measure the degree of the symmetry of metabolic networks, we calculate α :the size of $Aut(G)$, β : the relative degree of network symmetry of 705 metabolic networks[1] . α reflects the absolute symmetry degree of network directly. β is used to measure the symmetry of networks with different sizes. Generally, the larger α and β is, the more symmetric the network is. Among all the tested metabolic networks, 99.3% of them have α larger than 10^{10} and 82.3% of them have α larger than 10^{100}, which implies that most of metabolic networks are far away from asymmetric network. Hence, generally metabolic networks can be characterized as symmetric networks. Statistics of β shows that 98.7% of the metabolic network has β smaller than 0.1 and 83.3% of the metabolic network has β smaller than 0.01, which suggests that relative symmetry degree of metabolic network is quite low compared to the maximal symmetry degree of networks with equivalent number of nodes.

We also summarize φ : the degree of global symmetry for the networks. For some networks, there is some of automorphisms moving all the nodes or most of nodes. The action of such automorphism on node set will have global influence on the structure of the graph. However, for some other networks, all the automorphisms only move a small part of vertices or only action on the local subgraph of the network. Hence, when studying symmetry of networks, it's necessary to characterize the degree of global symmetry or local symmetry of the network. Generally, the larger φ is, the more globally symmetric the network is. Among all the tested metabolic networks, 98.6% of which has φ smaller than 0.1 and 72.8% has φ between 0.05-0.01, which suggest that metabolic network is very local symmetric.

To validate our conclusion, we replicate our experiment using BioCyc datasets. Among all the tested metabolic networks, 97.6% of them have α larger than 10^{10} and only one network (metaCyc) has α larger than 10^{100}. Statistics of β shows that only 1metabolic network has β smaller than 0.001 and 5 of the metabolic networks has β larger than 0.01, β values of the other networks (98.4%) are all between 0.001 and 0.01. φ of BioCyc data varies from 0.002071 to 0.0434783 with a mean value of 0.008364. Please see additional data file 3 for the results of α ,β and

[1]All symmetry statistics including and ,in this section are summarized from the underlying graphs of the metabolic networks. The underlying graph was derived by removing directions and self loops from the original network.

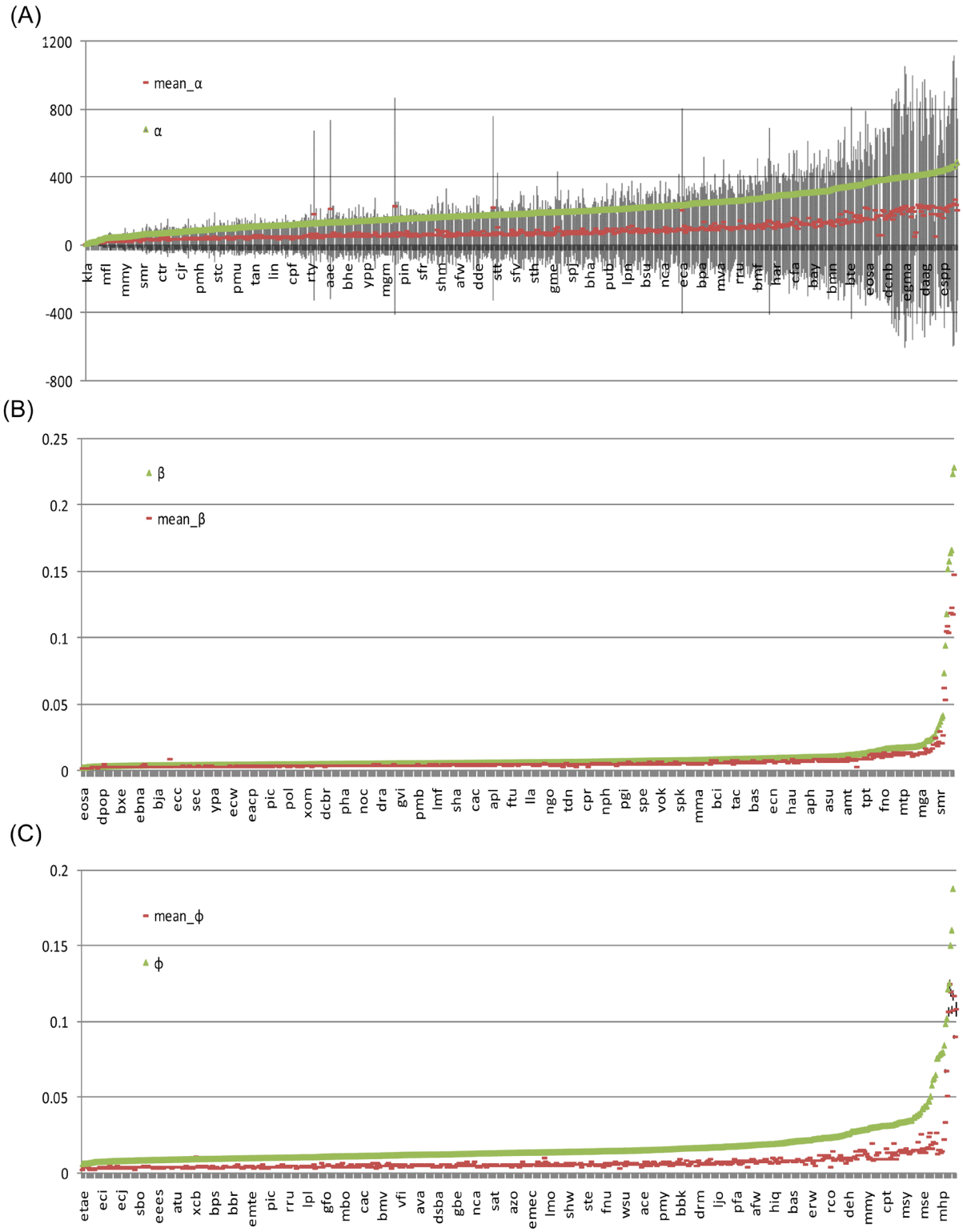


Figure 4: Comparison of the symmetry indices of real metabolic networks with random networks

The random networks are produced by Erdos-Renyi model in Pajek . (A) Error bar of the symmetry index α of real metabolic networks and corresponding randomized networks. (B) Error bar of the symmetry index β of real metabolic networks and their corresponding randomized networks. (C) Error bar of the symmetry index φ of real metabolic networks and their corresponding randomized networks. Please note that the variances of β and φ of randomized networks is so small that the error bar can not be observed obviously.

φ value for metabolic networks in KEGG and BioCyc.

To further investigate the symmetry of metabolic network, we compared three indices α β and φ of metabolic network with corresponding randomized networks. For each real metabolic networks for703 organisms, we generated 100 randomized networks with the same number of nodes and edges as the real network following Erdos-Renyi random graph model(Erdos and Renyi, 1960)[2]. (Two organisms: Debaryomyces hansenii (dha) and Legionella pneumophila Corby (lpc) were not included because they cannot produce Erdos-Renyi random networks due to their small network size). Then we compute α β and φ for every random network and summarize the mean and variance over the 100 randomized networks for each real network. The comparisons of α β and φ between real network and random networks are shown in Figure 4. We can see clearly from Figure 4 that 99.4% of α, 99.4% of β and 99.9% of φ in real metabolic network is larger than that in random network. These results demonstrated that the symmetry in metabolic network is obviously strong than that in random network, suggesting that symmetry is an important and non-ignored feature in metabolic network.

Again, we replicated our study in BioCyc metabolic networks. We found that in spite of relatively less symmetry compared with KEGG, the symmetry in BioCyc network is still substantially larger than that in random networks, suggesting that metabolic network is far away from asymmetric network. The comparisons of α β and φ between real network in BioCyc and random networks are shown in Supplementary Figure 2. Please see additional data file 4 for the results of α β and φ value for randomized networks in both KEGG and BioCyc.

Comparison of the Cconnectivity and Symmetry of Metabolic Networks between KEGG and BioCyc Datasets

In table 1, we compared the range, mean value and vari-

database		KEGG	BioCyc
Number of organisms		705	373
NCS	range	1-271	2-76
	mean	124.048227	16.07238606
	variance	3128.634034	44.65872467
MCS	range	7.8%-100%	92.1%-99.3%
	mean	0.548437869	0.966615067
	variance	0.023976601	0.00011063
NECS	range	0-0.778064	0.007-0.075
	mean	0.410917	0.0347128
	variance	0.012554904	0.000108611
$\log_{10}\alpha$	range	0.301-488.52	0.903-612.608
	mean	199.39	30.69700184
	variance	11423.74545	1099.766327
β	range	0.00196-0.72478	0.000598-0.0581303
	mean	0.01	0.004181307
	variance	0.001709304	1.04908E-05
φ	range	0.00576-0.666667	0.00207-0.0434783
	mean	0.02	0.008363534
	variance	0.001298575	2.01729E-05

Table 1: Comparison of the connectivity and symmetry of metabolic networks between KEGG and BioCyc datasets.

[2] Although it is desired to preserve the degree of vertices when meaningfully randomizing the network, however, it has been shown that symmetry is significantly relying on the degree of vertices(note 5 in xiao 2008c). In other words, when preserving degree of each vertex by exploiting the approach of edge rewiring (Maslov et al., 2004), the randomized networks will have almost the same symmetry properties with the real network, Hence, in this section, we relax the constraint when generating the randomized networks.

ance of three connectivity indices NCS, MCS, NECS and three symmetry indices α β and φ between networks of 705 species in KEGG and networks of 373 species in BioCyc. We can see clearly that for all the six measures, the range and variances in BioCyc datasets is significantly less than that in KEGG datasets. (For the range of $\log_{10}\alpha$ in BioCyc, if we get rid of the largest value 612.608 and the smallest value 0.903; remaining values range from 6.85-96.367). The mean values of NCS and NECS of BioCyc datasets are less than that in KEGG datasets, however, MCS of BioCyc is larger than that of KEGG. All these facts suggest that metabolic network in BioCyc is more connected. Since symmetry is typically greater for lower connectivity and shorter branches networks (MacArthur, 2008), it's naturally to find that symmetry in BioCyc networks is less than that in KEGG networks.

However, for both datasets, we clearly found that symmetry in real networks is obviously larger than that in randomized networks. No matter which metabolic network reconstruction methods were used, we came to the same conclusion that the symmetry of metabolic network, specifically, local symmetry, does exist.

Inferring Functional Equivalence from Structural Equivalence

We calculated the automorphism group (See material and methods for its definition) and obtained the orbits (structurally equivalent nodes) considering the connectivity and direction constraints for the reconstructed metabolic networks of all 705 organisms in KEGG. Since nodes in the same orbit are structurally equivalent to each other, it motivates us to further explore whether structurally equivalent nodes are functional equivalent. To accomplish this, we first generated two datasets which consist of similarity scores (See Materials and Methods for definition of similarity score). One dataset is referred to as orbit dataset. For each orbit in the metabolic network we calculated similarity scores for all pairs of compounds in the orbit and averaged them as the similarity score of the orbit. All similarity scores of the orbits in the networks of 705 organisms in KEGG were collected to form the orbit dataset where the replicated orbits were just calculated once. Another dataset which is referred to as random dataset was generated by collections of similarity scores of all pairs of compounds in the metabolic modules of all 705 organisms in KEGG. We used *t* statistics and

Modules	Orbits #	Compound pairs #	P_t^1	P_t^2	P_r
Glycolysis	55	406	3.04E-13	1.5814E-10	2.3055E-12
TCA cycle	40	171	0.009	0.0192	0.0497
Amino_Acid	557	131328	2.205E-81	8.0344E-32	2.9589E-38
Carbohydrate	618	92665	4.6399E-110	1.5533E-50	1.475E-65
Cofactors_Vitamins	211	37128	2.7279E-92	1.3082E-33	2.5883E-47
Energy	142	5050	0	0	0
Glycan_Biosynthesis	24	2278	4.4982E-26	0	8.2242E-12
Lipid	301	81406	3.6647E-171	1.1755E-63	3.395E-90
Nucleotide	193	9453	4.6127E-66	2.1246E-40	2.5256E-49
Other_Amino_Acids	102	10440	0	0.0015	0.0131
Polyketides	9	780	0	0.0027	0.0007
Secondary_Compounds	76	75078	1.5539E-78	4.5943E-17	9.5074E-24
Xenobiotics	134	73920	8.0514E-59	7.1642E-31	1.7688E-43

Table 2: P-values for testing significance of the similarity of the compounds in the orbits.

See additional files 5 for the module description. Orbits # denotes the number of pairs of compounds in the orbit, compound pairs # denotes the number of pairs of compounds in the random dataset, P_t^1 is the *P*-values of right tail *t*-test with equal variance; P_t^2 is the *P*-values of right tail t-test with unequal variance, P_r is the *P*-values of Wilcoxon two-sided rank sum test.

[3]In this section, all results are obtained from KEGG data, since we can only construct undirected graph from BioCyc database, which make the orbits we got is less biologically meaningful for metabolic networks than in directed graphs in KEGG.

Glycolysis / Gluconeogenesis		TCA cycle	
orbits	**similarity**	**orbits**	**similarity**
C00668 C01172	1	C00042 C00122	1
C00084 C00469	1	C00311 C05379	1
C00221 C00267	1	C00036 C00149	1
C00111 C00118	1	C00311 C00417	0.923077
C00031 C00267	1	C00158 C00417	0.923077
C00022 C00186	1	C00158 C00311 C00417	0.901099
C00668 C01172 C05345	0.921569	C00122 C00149	0.888889
C00111 C00197	0.909091	C00036 C00042	0.888889
C00074 C00631	0.909091	C00036 C00122	0.888889
C01172 C05345	0.882353	C00042 C00149	0.888889
C00103 C01172	0.882353	C00068 C05381	0.787879
C00103 C00668	0.882353	C00074 C00149	0.727273
C00668 C05345	0.882353	C00036 C00074	0.727273
C06187 C06188	0.88	C00149 C00158	0.692308
C00197 C00631	0.833333	C00022 C00036	0.666667
C00236 C06189	0.8125	C00022 C00149	0.666667
C05345 C05378	0.8	C00022 C00074 C00149	0.664647
C00197 C06189	0.785714	C00074 C00122	0.636364
C00033 C00084	0.75	C00042 C00074	0.636364
C00033 C00469	0.75	C00022 C00074	0.6
C00221 C01172	0.75	C00022 C00122	0.555556
C00111 C00118 C00668 C01172	0.75	C00033 C00122	0.5
C00103 C00221	0.75	C00033 C00036	0.444444
C00118 C00631	0.75	C00011 C 00026 C00311	0.433333
C00103 C00267	0.75	C00036 C00158 C00566	0.333625
C00221 C00668	0.75	C00011 C00026	0.3
C00197 C00236	0.733333	C00022 C00024 C00033	0.273504
C00197 C01159	0.733333	C00011 C00311	0.230769
C00631 C01159	0.733333	C00158 C00566	0.177419
C00074 C00197 C01159	0.716667	C00010 C00158	0.173077
C01451 C06186	0.695652	C00024 C00074	0.173077
C00031 C00267 C06187	0.681159	C00024 C00036	0.132075
C00111 C00236	0.666667	C00024 C00149	0.132075
C00111 C00118 C05378	0.666667	C00036 C00566	0.131148
C00074 C00111	0.666667	C00042 C00091	0.125
C00074 C00118	0.666667	C00091 C00122	0.125
C00031 C00267 C06188	0.666667	C00091 C00149	0.122807
C00118 C00236	0.666667	C00036 C00091	0.122807
C00031 C00267 C06187 C06188	0.653913	C00022 C00024	0.096154
C00103 C05378	0.636364	C00024 C00033	0.057692

Table 3: Orbits and their similarity in Glycolysis/Gluconeogenesis and TCA cycle pathways
Orbits were sorted in the descending order of similarity score. See additional files 4 for the orbit similarity and random compound similarity of another 11 modules.

wilcoxon rank sum statistics(Pagano and Gauvreau, 2000) to test whether there were significant differences in the similarity scores between orbit dataset and random dataset[3].

The results were shown in Table 2, from which we can see that in all pathway/modules, the compounds in the orbit showed significant evidence of similarity in compound chemical structures, which implied that the structurally equivalent nodes in metabolic networks are similar in their chemical structure. The compounds with similar chemical structure will have similar functions and play similar roles in biochemical reactions (Gutteridge et al., 2007). In Table 3, we listed the values of similarity of the compounds in all the orbits in Glycolysis/Gluconeogenesis and TCA cycle pathways. In Glycolysis/Gluconeogenesis pathway, 92.7% of orbits' similarity is larger than 0.5, while in TCA cycle pathways 55% of orbits' similarity is larger than 0.5. To gain further understanding of the nature of similarity among the compounds in the orbit, we presented the results of Glycolysis/Gluconeogenesis pathway and TCA cycle pathway.

(1) The orbit [C00668, C01172] in Glycolysis/Gluconeogenesis pathway.

It included C00668 (alpha-D-Glucose 6-phosphate) and C01172 (beta-D-Glucose 6-phosphate). Their chemical structures were shown in Table 3. C01172 is an isomer of C0066, so their chemical structures are identical. We examined the pathways which two compounds C00668 and C01172 participate in and the enzymes catalyzing the biochemical reactions of these two compounds. We found that they shared most of the pathways and enzymes (some enzymes are the same; some enzymes are in the same category, like 3.2.1.86 and 3.2.1.26). Interested readers please see Table 4 for details.

(2) The orbit [C00158, C00311, C00417] in TCA cycle pathway.

Their names, chemical structures, pathways which they participate in and enzymes they were catalyzed by were shown in Table 5. The compound C00311 (Isocitrate) is the isomer of the compound C00158 (Citrate). C00417 (cis-Aconitate) is the product of dehydrolysis from C00158 and C00311. In TCA cycle, the three reactions among these three compounds are catalyzed by the same enzyme EC4.2.1.3:

R01325: Citrate<=> cis-Aconitate + H_2O

R01900: Isocitrate <=> cis-Aconitate + H_2O

R01324: Citrate <=> Isocitrate

From the Table 5 we can see clearly that three compounds C00158, C00311 and C00417 participated in the same pathways and were catalyzed by same enzymes in most cases.

Since metabolic function is mainly determined by chemical structure (Gutteridge et al., 2007), our work showed that structural equivalent nodes in the metabolic networks were more likely to have the same or similar function.

Recently, symmetry in general complex networks has attracted certain research interest. All previous works (MacArthur, 2008; Xiao et al., 2008a; Xiao et al., 2008b; Xiao et al., 2008c) about symmetry in the networks have focused on undirected graphs. To explore symmetry in the metabolic networks, in this report we first introduced the concept of symmetry and developed algorithms to search symmetry in the directed networks. Then we further systematically investigated the symmetry properties of metabolic network, including the degree of symmetry, restricted network quotient. We observed much higher symmetry in metabolic networks which are reconstructed from KEGG and BioCyc datasets than that in random networks. Our preliminary results showed that metabolic networks are generally symmetric and in particular locally symmetric. To explore functional implications of the structural symmetry of metabolic networks, we tested significance of the chemical structure similarity of the compounds in the same orbit of the network. We found that compounds that are structurally equivalent to each other tend to have high similarity in chemical structures and that the structurally equivalent compounds often take part in the activities of the same pathway and are catalyzed by same enzymes. This may suggest that the symmetry in the metabolic network can generate the functional redundancy, and increase the robustness and the ability against attack of external disturbances.

Symmetry may arise from duplications in evolution of metabolic networks. In this report, we have focused on the symmetry of the metabolic networks. Due to the strong correlation between symmetry and duplication (a universe mechanism in biological networks) (Bhan et al., 2002; Chung et al., 2003; Teichmann and Babu, 2004), symmetry is expected to be ubiquitous in a variety of other biological networks and to play an important role in the evolution of biological networks.

Despite increasing interests in exploring the role of symmetry in evolution of networks, mechanism of evolution of symmetry has not been well investigated. It is worth studying the mechanism of generating symmetry of the networks and the role of structurally equivalent compounds in cell and molecular functions in the future.

Materials and Methods

Metabolic Network Reconstruction

At present, two major metabolic reconstruction methods are usually used. One method is introduced in Ma and Zeng(Ma and Zeng, 2003), where "currency" metabolites like H_2O, ATP, ADP are not included as nodes in network. This simplified metabolic network is biochemically meaningful in calculating path length. KEGG PATHWAY database uses the simplified metabolic network reconstruction method. The xml files of totally 152 metabolic pathways for every organism (705 organisms in total) were downloaded from KEGG FTP(Release date: Dec. 18, 2007). Reaction data were read from the xml files and represented as directed graph. The direction of each link implies the direction from an input compound (reactant) to an output compound (product) (See Figure 1(a)). Single pathways are combined to modules according to their metabolic functions, such as Carbohydrate Metabolism, Metabolism of Cofactors and Vitamins. There are 11 modules and each module contains 2-23 pathways. See additional data file 1. We also integrated all the 152 metabolic pathways into a metabolic network for every organism. Another metabolic network reconstruction method includes currency metabolites as nodes in network (Jeong et al., 2000), which makes metabolic network more connected. The metabolic network reconstruction in BioCyc database is a representative of this method. Totally 373 available Pathway/Genome Databases was downloaded (Release date: Oct. 15, 2008. Each database in the BioCyc collection describes the genome and metabolic pathways of a single organism, including another independent database AraCyc which has not been combined into BioCyc yet). We processed the tabular flat files of reaction data and combined the reactions into an integrated network for each organism. Direction information was not given in reaction files of BioCyc. So the integrated networks are unidirectional networks.

Connectivity of Metabolic Networks

Since many enzymes have not been found in some organisms yet, the reactions (edges in network) which are catalyzed by these enzymes are absent in the current network. Hence, we expect that the connectivity of metabolic network is quite low compared to corresponding randomized synthetic networks. To verify such conjecture, we use the number of connected subgraphs (NCS) and ratio of size of maximum connected subgraph to the whole network size (MCS) to measure the connectivity of metabolic network. We calculate the number of connected subgraph (NCS) in every single network. Apparently, the larger the NCS is, the lower the connectivity of metabolic network is; the larger MCS is, the more connected the network is. Furthermore, based on these concepts, we can define a new index: entropy based on connected subgraph (ECS) to measure the

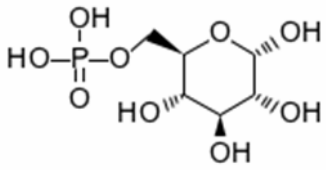
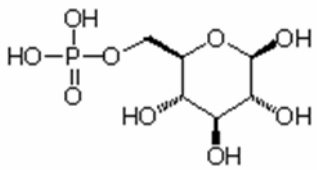
	C00668	C01172
Name	alpha-D-Glucose 6-phosphate	beta-D-Glucose 6-phosphate
Chemical structure	 C00668	 C01172
Pathway	map00010 Glycolysis / Gluconeogenesis map00030 Pentose phosphate pathway map00052 Galactose metabolism map00500 Starch and sucrose metabolism	map00010 Glycolysis / Gluconeogenesis map00030 Pentose phosphate pathway map00500 Starch and sucrose metabolism
Enzyme	2.4.1.15 2.7.1.1 2.7.1.2 2.7.1.63 2.7.1.69 3.1.3.9 3.2.1.26 3.1.3.9 3.2.1.26 5.3.1.9 5.1.3.15 5.4.2.2 5.4.2.5	1.1.1.49 2.7.1.1 2.7.1.2 2.7.1.63 3.2.1.86 5.3.1.9 5.1.3.15 5.4.2.6

Table 4: Compounds in Orbits [C00668, C01172].

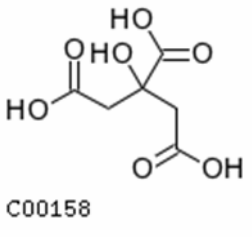
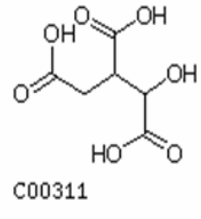
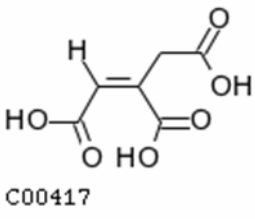
	C00158	C00311	C00417
Name	Citrate	Isocitrate	cis-Aconitate
Chemical structure	 C00158	 C00311	 C00417
Pathway	map00020 Citrate cycle (TCA cycle) map00251 Glutamate metabolism map00252 Alanine and aspartate map00630 Glyoxylate and dicarboxylate map00720 Reductive carboxylate cycle	map00020 Citrate cycle (TCA cycle) map00630 Glyoxylate and dicarboxylate map00720 Reductive carboxylate cycle	map00020 Citrate cycle (TCA cycle) map00630 Glyoxylate and dicarboxylate map00660 C5-Branched dibasic acid metabolism map00720 Reductive carboxylate cycle
Enzyme	2.3.3.1 2.3.3.3 2.3.3.8 2.8.3.10 3.4.13.20 4.1.3.6 4.2.1.3 4.2.1.4 6.2.1.18 6.3.2.27	1.1.1.41 1.1.1.42 1.1.1.286 2.3.1.126 4.1.3.1 4.2.1.3	4.1.1.6 4.2.1.3 4.2.1.4 5.3.3.7

Table 5: Compounds in Orbits [C00158, C00311, C00417].

average connectivity of a metabolic network:

$$ECS = -\sum_{0 \le i \le |C|} p_i \log p_i$$

, where C is the set of connected subgraphs of the network and p_i is the probability that a node belongs to a connected subgraph C_j. Given all connected subgraphs $C=\{C_1, C_2, \ldots, C_k\}$, we can calculate p_i as:

$$p_i = \frac{|C_i|}{\sum_j |C_j|} = \frac{|C_i|}{N}$$

, where N is the number of nodes in a graph. Clearly, for networks with N nodes, $ECS_{max} = \log N$ when $p_i = 1/N$ for each $0 \le i \le |C|$ (in this case, every node in the network is isolated from each other); $ECS_{min} = 0$, when the network is a connected graph and consequently $p=1$. Thus, we can define normalized entropy based on ECS_{max} and ECS_{min} :

$$NECS = -\frac{ECS - ECS_{\min}}{ECS_{\max} - ECS_{\min}} = \frac{ECS}{\log(N)}$$

In general, networks that contain a large connected subgraph tend to have a relatively small value under the measurement of NECS. Whereas metabolic networks are expected to be less connected and consequently the value of NECS is expected to be relatively large. The connectivity statistics including NCS, MCS and NECS in this section are summarized from the underlying graphs of the metabolic networks, where the direction and self loops were removed

from the original network.

Assessing Symmetry of Complex Networks

The degree of the symmetry of a graph *G* usually could be quantified by the following formula:

$$\alpha = |Aut(G)|,$$

which is the size of the automorphism group of graph *G*. Generally, α is very large and we usually use $\log_{10}|Aut(G)|$.

In order to compare the symmetry of networks with different sizes, symmetry measure β is often used, which is defined as:

$$\beta = (\alpha / N!)^{1/N},$$

where *N* is the node number of network, β measures the symmetry relative to maximal possible automorphism group of a graph with *N* nodes.

In general, for empirical networks, when network grows its symmetry is often destroyed. As evolution proceeds we rarely find global symmetry in the network, which means we can rarely find automorphisms that transforms most of nodes. In fact, many real networks have been shown to be locally symmetric (MacArthur, 2008) , which means that we can only find automorphisms which transform only small part of nodes in the network. Here, we use φ to quantify the degree to which graph *G* is globally symmetric(Xiao Y et al., "Efficiently Indexing Shortest Paths by Exploiting Symmetry in Graphs". In Proceedings of the 12th International Conference on Extending Database Technology (EDBT'09), March 23-26, 2009.):

$$\varphi = \max\{|\mathrm{supp}(g)| : g \in ID(G)\} / N,$$

where $supp(g)=\{v_i: v_i^g \neq v_i\}$ and $ID(G)$ is the set of indecomposable automorphisms of graph *G*. An indecomposable automorphism of $Aut(G)$ is a non-identity automorphism in Aut(G) that can not be decomposed into the product of two automorphisms g_1 and g_2 such that $g_1 \neq e$ and $g_2 \neq e$ and $supp(g_1) \cap supp(g_2) = \varnothing$, where *e* is the identity permutation that transform each vertex to itself.

To compute the above measures, the well known *nauty* program(McKay, 1981), which is one of the most efficient graph isomorphism algorithms available, is used to calculate the size and structure of automorphism groups.

Symmetry in Metabolic Networks

Symmetry in Directed Graph

In most of the previous studies of complex networks, networks are usually pre-processed as their underlying graphs: where weights, directions and self-loops are omitted. However, in the studies of metabolic network, the direction can't be omitted since many reactions are irreversible and the direction determines the reaction rates and the product output. Hence, when exploring symmetry in metabolic networks, we need to investigate symmetry in directed networks first.

In general, a directed network is a pair (*N*, *E*) with *N* representing the node set and *E* representing the set of ordered pairs of *N*. The related concepts of symmetry in directed graph is completely the same as that in undirected graphs. The fact that we need to highlight when exploring symmetry in directed networks is that any automorphism in a directed network need to preserve the oriented relation instead of un-oriented relation in undirected graph.

It's trivial to show that if *g* is an automorphism of a directed graph *G*, *g* will also be an automorphism of its underlying graph *G'*. However the inverse does not necessarily hold true. Hence, if *G'* is the underlying graph of graph *G*, we have $Aut(G) \subseteq Aut(G')$, which implies that the degree of symmetry of *G* is smaller than that of *G'*. Consequently, the automorphism partition of the directed graph is finer[4] than that of its underlying graph.

Restricted Network Quotient

Recall that nodes of a symmetric network can be partitioned into disjoint equivalent classes which are called orbits of the graph according to the automorphic equivalence relation on nodes. Nodes in the same orbit are structurally equivalent and cannot be distinguished from each other (Tinhofer and Klin, 1999) by usual measurement on nodes, such as degree, clustering coefficient. Therefore they can be glued together to create a coarse reduced network, known as the quotient. In Figure 3, G_1 are quotient networks of G_0. Since metabolic networks possess a non-trivial automorphism partition they carry a significant amount of redundant information in which more than one node plays

[4] Let *P* and *Q* be two partitions on the same set *X*, we say *P* is finer than *Q* if any cell in *P* is a subset of some cell in *Q*.

the same structural role.

In most of the previous researches about complex network, only the automorphism group of the largest connected subgraph is exploited. In above sections, we have shown that metabolic networks are more disconnected than other empirical networks. Hence it is necessary to explore the symmetry of metabolic networks with all disconnected subgraphs taken into account.

However, preserving all disconnected subgraphs will pose a challenge to the calculation of quotient of metabolic structure. Please note that when calculating quotient of a graph consisting of two isomorphic disconnected subgraphs, these two subgraphs will be merged into one subgraph under the action of the automorphism group of the graph. Hence, calculation of network quotient should take into account the connectivity constraint so that the isomorphic isolated modules will not be merged into one reduced subgraph in the quotient.

Specifically, assume that graph G contains pair-wise isomorphic and disconnected subgraphs G_1, G_2, …G_m. Let $H(G)$ be the set consisting of all those automorphisms that swap nodes between pair-wise and disconnected subgraphs, i.e.

$$H(G)=\{g: x^g=y \text{ and } g \in Aut(G) \text{ and } x \in V(G_i) \text{ and } y \in V(G_j) \text{ and } i \neq j \ \}.$$

Then we can calculate the *restricted quotient* of graph G under the action of $R(G)=Aut(G)-H(G)$.

It's easy to show that in the restricted quotient all disconnected subgraphs in the original graph will not be merged and consequently the number of disconnected subgraphs will be preserved. We obtained the orbits in network through the restricted network quotient.

Given a graph G_0 consisting of two isomorphic subgraphs, as shown in Figure5(A). Obviously, the automorphism group of G_0 can be decomposed into the wreath product(Rotman, 1999) of G_{in} and G_{out}, where G_{in} is the triangle, shown in Figure 5(B), and G_{out} is the abstracted outer graph, as shown in Figure 5(C). Note that in $Aut(G)$, there are some automorphisms, such as g=(1,4)(2,5)(3,6) that swap nodes in different isolated subgraphs(e.g. 1 and 4 are transformed to each other in spite of that these two vertices are in different isolated subgraphs). Under the action of such automorphisms, we finally can obtain one single orbit {1,2,3,4,5,6}for G_0. Thus the quotient of G_0 is just a single node (shown in Figure 5(D)), which contradicts to the fact that two isolated triangle structures often interpreted as two isolated functional modularity for biological networks. However based on the concept of restricted network quotient, the network G_0 can be reduced to a quotient network consisting of two isolated nodes (shown in Figure 5(E)).

In the computation of symmetry of directed networks, we find that nauty program can not ensure the correctness for directed graphs. Hence, we first use nauty to get the automorphism partition of the underlying graph of the network and then refine the automorphism partition. Considering the direction and connectivity of metabolic networks, the algorithm to get the orbits is shown in Algorithm 1. Although theoretically we can not ensure that the resulting partition is equivalent to the automorphism partition under the restricted automorphism group, the resulting partition is practically close to the desired partition and is practically useful in the exploration of functional equivalence of nodes in the same orbits:

Algorithm1: *getOrbits (G)*

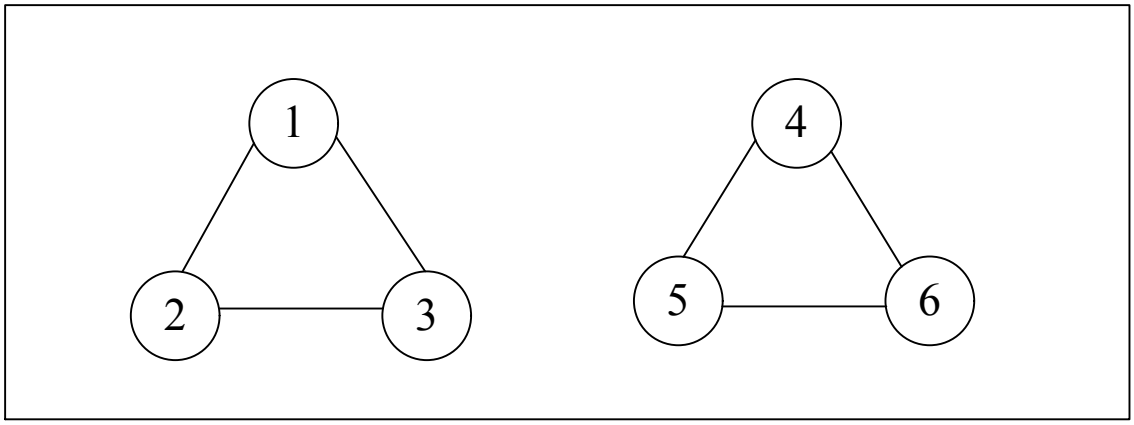
Input: a metabolic network G
Output: new orbit partition P'
{
1. $P'=\varnothing$, $G=\{ G_1, G_2, \ldots G_m \}$ *//get Connected Subgraphs of G ;*
2. $R(G)=Aut(G)-H(G)$ // get restricted automorphisms

3. $P=\{V_1, V_2, \ldots V_k\}$*// obtain partition according to R(G);*
4. for *each* $V_i \in P$ *such that* $|V_i|>1$
5. for *each* $v \in V_i$
6. $L(v)=(|N^+(v)|,|N^-(v)|)$; *//compute the in-degree and out-degree of v*
7. $Order(V_i)$; *//Sort* V_i *according to lexicographic order of L(v)*
8. $\{V_{i1}, V_{i2}, \ldots, V_{ik}\}$ =*Subdivide*(*Vi*);
9. $P'= P' \cup \{V_{i1}, V_{i2}, \ldots, V_{ik}\}$
10. return *P'*

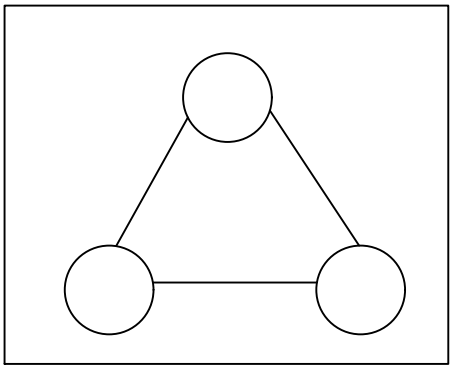
}

Analyze Orbit Similarity

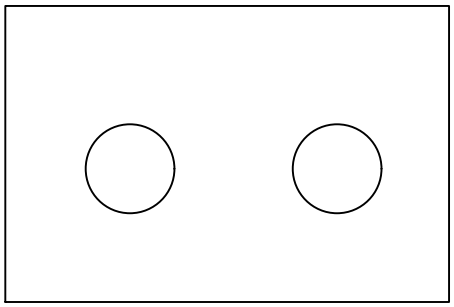
In the above section, we have known that nodes in the same orbit group are structurally equivalent. It is well known that structural equivalence implies functional equivalence. Whether the structurally equivalent nodes in the network are more similar in function has still not been validated. In metabolic networks, nodes are compounds with specific structure which determines its function in reactions. If two compounds are structurally similar to each other, they will



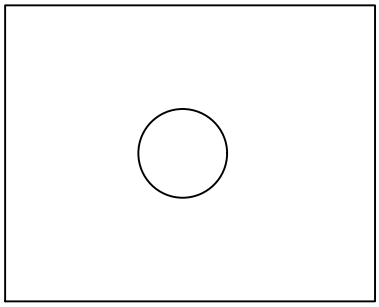
(A) G_0



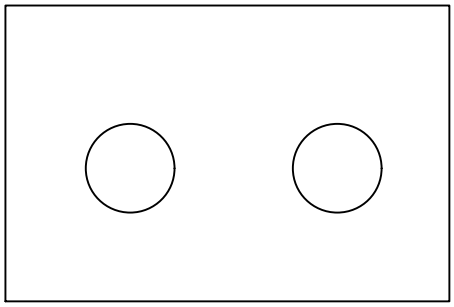
(B) G_{in}



(C) G_{out}



(D) G_1



(E) G_1'

Figure 5: A graph consisting of two isomorphic isolated subgraphs

(A) Underlying graph, (B) G_{in}, (C) G_{out}, (D) the network quotient of the underlying graph and (E) the restricted network quotient of the underlying graph considering the connectivity.

function similarly. Alex Gutteridge et.al have found that chemical structure of small molecular compounds often determines compound suitability for use in regulation and how groups of similar compounds can regulate sets of enzymes (Gutteridge et al., 2007). Hence, it is reasonable to believe that whether the structurally equivalent nodes in the network are functionally equivalent can be inferred from the similarity between their chemical structures.

Many chemical structure comparison methods have been proposed to analyze the compound similarity in the metabolic pathways (Hattori et al., 2003; Nobeli et al., 2003; Raymond et al., 2002; Raymond and Willett, 2002). In most of the algorithms, the chemical structure is treated as a two dimensional (2D) object, which can be presented as a graph consisting of nodes (atoms) and edges (bonds). In this paper, we use the method proposed by Hattori et al., (2003) to compute the similarities between chemical compounds. In this method, the similarities between compounds are measured by the size of maximal common subgraph (MCS) between two graphs representing these two compounds. A normalized similarity score based on Jaccard coefficient (Watson, 1983) is used in this method, which is defined as the ratio of the size of the maximal common substructure to the size of the non-redundant set of all substructures:

$$JC(G_1, G_2) \equiv \frac{|G_1 \cap G_2|}{|G_1 \cup G_2|} = \frac{|MCS(G_1, G_2)|}{|G_1| + |G_2| - |MCS(G_1, G_2)|}$$

, where $|G|$ is defined as the number of nodes of graph G.

After obtaining the similarity of all the compound pairs by Masahiro's algorithm, we compared the compound similarities between nodes in the same orbits to the compound similarity between nodes in the network and test the significance of similarity scores of nodes in the same orbit.

All similarity score of the orbits in the networks of 705 organisms were collected to form the orbit dataset where the replicated orbits were just calculated once. Another dataset which is referred to as random dataset was generated by collections of similarity scores of all pairs of compounds in the metabolic modules of all 705 organisms.

Three statistics were used to test differences in the similarity scores between the orbit dataset and random dataset: right tail *t*-test with equal variance, right tail *t*-test with unequal variance and Wilcoxon two-sided rank sum test (Pagano and Gauvreau, 2000).

We assume that the compounds in the same orbit should have similar chemical structure and function. So we use Hattori's maximal-common-subgraph based algorithm with loose weighting condition to do chemical structure comparison. We summarized the average over similarity scores of all compound pairs in the same orbit:

$$S = \frac{\sum_{1 \le i < j \le n} S_{i,j}}{n(n-1)/2}.$$

Authors' Contributions

Hua Dong analyzed the data and wrote the draft; Yanghua Xiao was responsible for the algorithms and the program. Hua Dong and Yanghua Xiao contribute equally to this work. All authors wrote and approved the final manuscript.

Acknowledgements

This work was partially supported by grant from Shanghai Commission of Science and Technology (04dz14003) and Grant from Hi-Tech Research and Development Program of China(863) (2007AA02Z300). Thanks for Hoicheong Siu's help on network reconstruction from BioCyc dataset.

References

1. Arita M (2004) The metabolic world of Escherichia coli is not small. Proc Natl Acad Sci USA 101: 1543-1547. » CrossRef » Pubmed » Google Scholar
2. Batagelj V, Mrvar A (2003) Pajek - Analysis and Visualization of Large Networks. In Graph Drawing Software, M. Junger, and P. Mutzel (eds), 28. Berlin: Springer. » CrossRef » Pubmed » Google Scholar
3. Bhan A, Galas DJ, Dewey TG (2002) A duplication growth model of gene expression networks. Bioinformatics 18: 1486-1493. » CrossRef » Pubmed » Google Scholar
4. Chung F, Lu L, Dewey TG, Galas DJ (2003) Duplication models for biological networks. J Comput Biol 10: 677-687. » CrossRef » Pubmed » Google Scholar
5. Clauset A, Moore C, Newman ME (2008) Hierarchical structure and the prediction of missing links in networks. Nature 453: 98-101. » CrossRef » Pubmed » Google Scholar
6. Eom YH, Lee S, Jeong H (2006) Exploring local structural organization of metabolic networks using subgraph patterns. J Theor Biol 241: 823-829. » CrossRef » Pubmed » Google Scholar

7. Erdos P, Renyi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5: 45.» Google Scholar

8. Felder G, Garcia BJ, Greene PB, Kofman L, Linde A, et al. (2001) Dynamics of symmetry breaking and tachyonic preheating. Phys Rev Lett 87: 011601.» CrossRef » Pubmed » Google Scholar

9. Guimera R, Nunes ALA (2005) Functional cartography of complex metabolic networks. Nature 433: 895-900. » CrossRef » Pubmed » Google Scholar

10. Gutteridge A, Kanehisa M, Goto S (2007) Regulation of metabolic networks by small molecule metabolites. BMC Bioinformatics 8: 88. » CrossRef » Pubmed » Google Scholar

11. Hatcher A (2002) Algebraic Topology. London: Cambridge University Press.» CrossRef » Google Scholar

12. Hattori M, Okuno Y, Goto S, Kanehisa M (2003) Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. J Am Chem Soc 125: 11853-11865. » CrossRef » Pubmed » Google Scholar

13. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. Nature 407: 651-654. » CrossRef » Pubmed » Google Scholar

14. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30. » CrossRef » Pubmed » Google Scholar

15. Karp PD, Ouzounis CA, Moore KC, Goldovsky L, Kaipa P, et al. (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res 33: 6083-6089. » CrossRef » Pubmed » Google Scholar

16. Ma H, Zeng AP (2003). Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19: 270-277. » CrossRef » Pubmed » Google Scholar

17. MacArthur B, Sanchez GRJ, Anderson JW (2008) Symmetry in complex networks. Discrete Applied Mathematics. » CrossRef » Google Scholar

18. MacArthur BD, Anderson JW (2006) Symmetry and Self-Organization in Complex Systems. arXiv:cond-mat/0609274. » CrossRef » Google Scholar

19. Mahadevan R, Palsson BO (2005) Properties of metabolic networks: structure versus function. Biophys J 88: L07-09. » CrossRef » Pubmed » Google Scholar

20. Mainzer K (ed) (2005) Symmetry And Complexity: The Spirit And Beauty Of Nonlinear Science Singapore: World Scientific Publishing Company. » CrossRef » Google Scholar

21. Maslov S, Sneppen K, Zaliznyak A (2004) Detection of topological patterns in complex networks: correlation profile of the internet Physica A: Statistical Mechanics and its Applications 333: 12. » CrossRef » Google Scholar

22. McKay BD (1981) Practical Graph Isomorphism. Congressus Numerantium 30: 43.» CrossRef » Google Scholar

23. Milo R, Shen OS, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: simple building blocks of complex networks. Science 298: 824-827. » CrossRef » Pubmed » Google Scholar

24. Nobeli I, Ponstingl H, Krissinel EB, Thornton JM (2003) A structure-based anatomy of the E.coli metabolome. J Mol Biol 334: 697-719. » CrossRef » Pubmed » Google Scholar

25. Pagano M, Gauvreau K (2000) Principles of Biostatistics, 2nd Edition Belmont,California: Duxbury Press. » CrossRef » Google Scholar

26. Quack M (2003) Molecular Spectra, Reaction Dynamics, Symmetries and Life CHIMIA International. Journal for Chemistry 57: 14. » CrossRef » Google Scholar

27. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297: 1551-1555. » CrossRef » Pubmed » Google Scholar

28. Raymond JW, Gardiner EJ, Willett P (2002) Heuristics for similarity searching of chemical graphs using a maximum common edge subgraph algorithm. J Chem Inf Comput Sci 42: 305-316. » CrossRef » Pubmed » Google Scholar

29. Raymond JW, Willett P (2002) Maximum common subgraph isomorphism algorithms for the matching of chemical structures. J Comput Aided Mol Des 16: 521-533. » CrossRef » Pubmed » Google Scholar

30. Rotman J (ed) (1999) An Introduction to the Theory of Groups New York: Springer-Verlag. » Google Scholar

31. Samal A, Singh S, Giri V, Krishna S, Raghuram N, et al. (2006) Low degree metabolites explain essential reactions and enhance modularity in biological networks. BMC Bioinformatics 7: 118. » CrossRef » Pubmed » Google Scholar

32. Teichmann SA, Babu MM (2004) Gene regulatory network growth by duplication. Nat Genet 36: 492-496. » CrossRef » Pubmed » Google Scholar

33. Tinhofer G, Klin M (1999) Gragh invariants and Stabilization Methods. In Algebraic combinatorics in mathematical chemistry. Methods and algorithms: Technische Universitat Munchen. » CrossRef » Google Scholar

34. Wagner A, Fell DA (2001) The small world inside large metabolic networks. Proc Biol Sci 268: 1803-1810. » CrossRef » Pubmed » Google Scholar

35. Watson GA (1983) An algorithm for the single facility location problem using the Jaccard metric. SIAM J. Sci Stat Comput 4: 9. » CrossRef » Google Scholar

36. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440-442. » CrossRef » Pubmed » Google Scholar

37. Xiao Y, MacArthur B, Wang H, Xiong M, Wang W (2008a) Network Quotients: Structural Skeletons of Complex Systems. Phys Rev E Stat Nonlin Soft Matter Phys 78: 046102. » CrossRef » Pubmed » Google Scholar

38. Xiao Y, Wu W, Wang H, Xiong M, Wang W (2008b) Symmetry-based structure entropy of complex networks. Physica A: Statistical Mechanics and its Applications 387: 9. » CrossRef » Google Scholar

39. Xiao Y, Xiong M, Wang W, Wang H (2008c) Emergence of symmetry in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 77: 066108. » CrossRef » Pubmed » Google Scholar