ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理
Chinese Physics B

Evolution of the internet AS-level topology:From nodes and edges to components

Xiao Liu(刘晓), Jinfa Wang(王进法), Wei Jing(景薇), de Menno Jong, Jeroen S Tummers, Hai Zhao(赵海)

Citation: Chin. Phys. B . 2018, 27(12): 120501. **doi:** 10.1088/1674-1056/27/12/120501
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

The Wronskian technique for nonlinear evolution equations

Jian-Jun Cheng(成建军) and Hong-Qing Zhang(张鸿庆)

Chin. Phys. B . 2016, 25(1): 010506. **doi:** 10.1088/1674-1056/25/1/010506

Equivalent formulations of “the equation of life”

Ao Ping

Chin. Phys. B . 2014, 23(7): 070513. **doi:** 10.1088/1674-1056/23/7/070513

Collective composite-rotating consensus of multi-agent systems

Lin Peng, Lu Wan-Ting, Song Yong-Duan

Chin. Phys. B . 2014, 23(4): 040503. **doi:** 10.1088/1674-1056/23/4/040503

Robust finite-time stabilization of unified chaotic complex systems with certain and uncertain parameters

Liu Ping

Chin. Phys. B . 2013, 22(7): 070501. **doi:** 10.1088/1674-1056/22/7/070501

Network evolution driven by dynamics applied to graph coloring

Wu Jian-She, Li Li-Guang, Wang Xiao-Hua, Yu Xin, Jiao Li-Cheng

Chin. Phys. B . 2013, 22(6): 060507. **doi:** 10.1088/1674-1056/22/6/060507

Evolution of the Internet AS-level topology: From nodes and edges to components*

Xiao Liu(刘晓)[1,2], Jinfa Wang(王进法)[2,†], Wei Jing(景薇)[3], Menno de Jong[1], Jeroen S Tummers[1], and Hai Zhao(赵海)[2]

[1] College of Computer Science and Engineering, Northeastern University, Shenyang 110000, China
[2] School of Biosciences, Durham University, Durham, DH1 3LE, UK
[3] School of Information Engineering, Shenyang University, Shenyang 110000, China

(Received 22 June 2018; revised manuscript received 20 September 2018; published online 10 November 2018)

Studying the topology of infrastructure communication networks (*e.g.*, the Internet) has become a means to understand and develop complex systems. Therefore, investigating the evolution of Internet network topology might elucidate disciplines governing the dynamic process of complex systems. It may also contribute to a more intelligent communication network framework based on its autonomous behavior. In this paper, the Internet Autonomous Systems (ASes) topology from 1998 to 2013 was studied by deconstructing and analysing topological entities on three different scales (*i.e.*, nodes, edges and 3 network components: single-edge component M1, binary component M2 and triangle component M3). The results indicate that: a) 95% of the Internet edges are internal edges (as opposed to external and boundary edges); b) the Internet network consists mainly of internal components, particularly M2 internal components; c) in most cases, a node initially connects with multiple nodes to form an M2 component to take part in the network; d) the Internet network evolves to lower entropy. Furthermore, we find that, as a complex system, the evolution of the Internet exhibits a behavioral series, which is similar to the biological phenomena concerned with the study on metabolism and replication. To the best of our knowledge, this is the first study of the evolution of the Internet network through analysis of dynamic features of its nodes, edges and components, and therefore our study represents an innovative approach to the subject.

Keywords: complex system, Internet AS-level topology, evolution, network component.

PACS: 05.10.–a, 05.90.+m, 64.60.aq, 89.20.Hh **DOI:** 10.1088/1674-1056/27/12/120501

1. Introduction

As a complex dynamic system, the Internet evolves with uncertainties over time. It is undergoing many structural changes, including the addition of new facilities and the removal of old or faulty ones. Using complex network theory, several aspects of the dynamic properties of the Internet have been studied, from the inter-dependent relationship of network devices to the economic ecosystem of the Internet.[1,2] When referring to the Internet, which is a typical example of a complex system, the behaviors among the whole system should ideally be described in a well-defined, broad approach. A classic case of using a simple formula to describe complicated processes is the logistic equation that is used to describe population dynamics, which possesses a rich structure that includes chaotic orbits and fractal non-divergent domains. Simple systems are more likely to generate more complicated behavior, in particular interacting particle systems.

Expansion and improvement of the Internet has not stopped since the day it was invented, and so it is fair to assume that changes of nodes and edges drive the network's evolution. All of the activities related to change and evolution are simply to satisfy our needs, from commercial needs to entertainment. In recent decades, complex network theory has become an important approach to investigate most types of complex systems.[3,4] There has been a major research effort made to depict the interaction or inter-dependency among network entities of the network topologies (*e.g.*, node, component). For example, Shen-Orr *et al.*[5] introduced the motif structure (a network component) into complex network studies and found an easily interpretable view of the entire known transcriptional network of the organism. Faloutsos *et al.*[1] worked on the power-law of degree distribution of the Internet topology and corrected the prevailing view of the Internet, according to which the Internet was seen as a random network. The Internet has become an enormous complex system[6] whose topology, function, dynamics and so on are too complicated to be described. A few studies have approached the evolution of the Internet from a different perspective. The statements regarding the changes that the Internet underwent vary from referring to the size of the Internet, the phase of the Internet, to the Internet ASes relationships. In a study of the evolution of the Internet economic ecosystem,[7] the characteristics and interactions of the application and network providers have been investigated, which show how it led to a market equilibrium of the ecosystem. However, what drives this equilibrium state of the Internet remains unclear. Some studies have observed and

*Project supported by the National Natural Science Foundation of China (Grant No. 61671142).
†Corresponding author. E-mail: jinfa.wang@mervin.me
© 2018 Chinese Physical Society and IOP Publishing Ltd

distinguished the birth and death of the nodes and edges.[8,9] The question remains how the network edges drive the Internet's evolution.

The basic mechanisms of biological evolutionary phenomena are applied directly or indirectly to come up with novel designs or solve problems that are difficult to solve otherwise in network science and engineering fields.[10] Imagine a scenario in which some living organisms leave their original group and choose to join other group gradually, while some small groups tend to be involved in the bigger ones, to share certain types of resources (*e.g.*, information, gene diversity, habitats, energy, and so on), in which each individual is trying to maximize their benefits and become robust. These series of behaviors ensure that individuals survive continuously despite limited space or resources. In contrast, those individuals who have not sought for better opportunities or supplies are more likely to die off over time. Thus, when it comes to the Internet system, the question as to how individuals, or some small groups, participate in the Internet's evolution has become a popular topic. We regard the Internet as an instance of a complex system that exhibits simple behavior due to a compendium of innate architectural features that span a range of scientific disciplines. From this perspective, network scientists believe that large-scale studies of the Internet are valuable.[11] To study these phenomena and reveal the behaviors of individuals (*i.e.*, nodes, edges) and groups (*e.g.*, components) of the Internet, the analysis based on a novel perspective is applied in this paper. We analyzed the evolution dynamic of the Internet by focusing on both the birth and the death of nodes, edges and some certain network components of the Internet. Some large scale-properties from the resulting network are observed. The contributions of this work are addressed as follows:

(i) We propose a novel method to observe and analyse the Internet's evolution, which considers nodes, edges and three specific functional network components of the Internet.

(ii) By adopting the proposed method on Internet AS-level topology data from January 1998 to August 2013, we observe a type of link adjustment, including inter-edges growth and inter-components reconstruction, which is the key factor of the AS-level Internet evolution.

(iii) Fundamental biological phenomena, such as metabolism and replication, are observed in our study when the analysis method is applied to deconstruct and construct the Internet's network topologies.

The rest of this paper is organized as follows. Section 2 reviews the literature on the complex system and evolution of the Internet. Section 3 introduces the dataset used in this paper and explains the topology extraction method on the Internet AS-level topology. Section 4 proposes a novel method of the Internet evolution analysis, which considers time-series of emerging and vanishing nodes, edges and network components. Based on these methods, Section 5 analyzes the nodes and edges growth, component reconstruction and network fluctuation. Discussions and conclusions are given in Section 6 and Section 7, respectively.

2. Related work

Because the Internet is continuously evolving in a dynamic way due to human activities, understanding how the Internet (both IP-level and AS-level) organizes, evolves and manages its network topology, dynamics, behaviors, function and so on are main tasks for network scientists, this will help to gain more insight into the Internet and enhance the network's infrastructure, and will also contribute some ideas on commercial decisions making.

In 1998, Faloutsos *et al.*[1] was the first to prove that the Internet was a scale-free network topology. By considering the Internet as a complex system, Park *et al.*[6] studied aspects of self-similar traffic, power-law connectivity, WLAN PHY lay dynamics and non-cooperative network games. Ma *et al.*[7] proposed a macroscopic network-aware model which demonstrated the interactions and also the characteristics of the application and network providers. Their study observed the origin of the Internet's evolution and the factors affecting this. Most Internet evolution studies are based on either the physical network layer or the logical topology layer. The studies of the physical network have mainly focused on the Internet's economic ecosystem, especially the commercial relationships between ASes. For over a decade, researchers analyzed the AS-level Internet growth related to AS function and business relationship types.[12,13] Meanwhile, other studies were based on the logical topology layer applied the complex network theory. By analyzing the Internet's evolution, some studies have verified that under the preferential attachment condition, the Internet topology was a scale-free network.[14,15]

While studies have succeeded in better understanding the evolution of the Internet, as of yet, no studies on the systematic behavior of the Internet have been done. Phases have changed in the IPv4 and IPv6 AS-level Internet evolution over a long period of time.[16] Ai *et al.*[17] observed sudden changes in the mean degree, the mean path length and other metrics of IPv6 in time series. Both of these studies only briefly worked on the fluctuation and mutation of the Internet, considering the Internet as a complex network. In 2006, Moore *et al.*[18] proposed models of the time evolution of networks, which were more practical, by adding and deleting nodes. This led to the realization that the essence of network evolution is the changes of nodes and edges in topology. Oliveira *et al.*[19] formulated the topology liveness problem and contrasted it with the completeness problem and developed an empirical model to characterize the trend of AS topology evolution. In a different network model, the edges between the new added node and the

old node are allowed to change when new nodes are added to the network. The positive-feedback preference (PFP) model simulated the growth of the nodes and internal links, using the nonlinear preferential attachment mechanism.[8] Based on PFP, an evolving network model was proposed, with a description of birth and death of the nodes and edges to characterize the evolution of the AS-level Internet.[9]

It is known that nodes need to build connections with other nodes. The more connections that a node can get, the more robust it will be. Hence, some researchers have studied fundamental function-structures analyzing the subgraphs or components of a complex network.[20–22] Maslov[23] identified hierarchical features of the Internet topology to utilize topological pattern units. Topological patterns were used to detect network features and to identify a hierarchical feature of Internet topology. Kiremire *et al.* adopted a network component-based analysis approach to compare the schemes' performances of the Internet topologies.[24]

3. Data acquisition and processing

In 1969, there were only four experimental nodes that were built as an early Internet. By 1983, this number had increased to 300 connected computers. Nowadays, the number of connected Internet devices is tremendous. The Internet is becoming a huge and complicated system that consists of tens of thousands of AS and billions of IP addresses.

The data applied in this study was obtained from the Border Gateway Protocol (BGP) router Table data, which was captured by three projects: Route–Views, RIPE NCC, and PCH. First, we downloaded the raw BGP probing data from the BGP router Table which contains the data from January 1998 to August 2013. Each day's data are provided by all the probing monitors. Second, we merged the whole month's daily data which is provided by all monitors into one snapshot of the Internet to represent the monthly snapshot. By merging data from all of the monitors, we could minimize data incompleteness, this is likely to be caused by equipment failures, hardware or software updating errors and untimely information aggregation. Third, by extracting the AS_PATH records and filtering the located data[23] from merged snapshot dataset files, 184 topologies in time series were obtained.

The AS_PATH field will identify the autonomous systems from the routing information, which is carried by the BGP messages. The components are extracted from AS_PATH: AS-CONFED-SEQUENCEs, AS-CONFED-SETs or AS-SETs. In summary, it is necessary to pass the data through this filter to get the neat and tidy Internet topology. More precisely, to improve validity, the following three steps are carried out to extract the AS_PATH field:

i) Compress AS_PATH paths. For example, the path "a-b-c-c-c-d" will be compressed to "a-b-c-d". Once a BGP router sends an advertisement, it must add its own AS number (ASN) to the far left of the AS_PATH path. The BGP protocol allows self-ASN to be added onto the AS_PATH repeatedly to increase the length of the AS_PATH, and it will affect the routing decision of their neighbors' routes.

ii) Filter the AS_PATH paths of the present loop. Generally, some of the BGP routing rules need to be manually set by an administrator. Thus, some errors could be produced, which will generate self-loops.

iii) Filter these AS_PATH paths, which are contained in AS-CONFED-SEQUENCE, AS-CONFED-SET, and AS-SET. The AS-SET path is disordered, and the AS in the AS-CONFED-SEQUENCE and AS-CONFED-SET belongs to local federations. Therefore, we need to filter out these three types of data and only adopt the AS-SEQUENCE data. Additionally, filter the AS_PATH paths, which contain any private AS. The RFC6996[25] has specified that AS64512∼AS65534 are private ASN. If an AS_PATH contains a private AS, then it apparently suggests that the routing information is wrong.

These pre-processing steps were used to decrease the noise and improve the validity of the data.

4. Methods

The Jaccard similarity theory, known as intersection over union and the Jaccard similarity coefficient, is a statistic that is used to compare the similarity and diversity of sample objects.[26,27] Inspired by Jaccard's theory, to study the object's evolutionary behaviors on a time-series, we focus on the changes that the system undergoes during its evolution and the dynamic evolution mechanism of the Internet. This is different from the classic methods of studying the evolution of the Internet, which have mainly investigated the phenomena and behaviors of the Internet. Based on set theory and Jaccard's similarity theory, we designed a method to investigate the behavior of the Internet, by partially regarding the evolutionary process of the Internet as biological evolutionary process.

4.1. Characterizing related network entities

This paper uses the dynamic network G to model the process of evolution of the Internet over time. The relevant definitions are given below, and a glossary of terms and abbreviations used are given in Table 1.

Dynamic network: a dynamic network G is a set of networks: $G = \{g(i)\}_{i=1}^{T}$, where i is the time index, T is the total number of time slots, $g(i)$ is the sampled network at time i, and $N(i)$ and $E(i)$ are the nodes set and the edges set of $g(i)$. Every node connects with at least one other node in the static network, unless the static network consists of one node only.

Steady node: for a node $n \in N(i)$, if $n \in N(i+1)$, then n is considered as a steady node during $[i, i+1]$. All

steady nodes comprise the steady nodes set: $N_{\text{steady}}(i,i+1) = N(i) \cap N(i+1)$, which are nodes {E, H, G, F, I} in Fig. 1.

Table 1. Glossary and abbreviation.

Symbol	Description
$g(i)$	A network topology snapshot at time tick i
G	All network topology snapshots
N	The number of network nodes
E	The number of network edges
$t(i)$	The time at tick i
N_{steady}	The set of steady nodes
N_{dead}	The set of dead nodes
N_{born}	The set of born nodes
$E_{\text{dead}}^{\text{internal}}$	The set of dead edges in the internal area
$E_{\text{dead}}^{\text{boundary}}$	The set of dead edges in the boundary area
$E_{\text{dead}}^{\text{external}}$	The set of dead edges in the external area
$E_{\text{born}}^{\text{internal}}$	The set of born edges in the internal area
$E_{\text{born}}^{\text{boundary}}$	The set of born edges in the boundary area
$E_{\text{born}}^{\text{external}}$	The set of born edges in the external area
$E_{\text{steady}}^{\text{internal}}$	The set of steady edges
$M1$	A component as shown in Fig. 2(a)
$M2$	A component as shown in Fig. 2(b)
$M3$	A component as shown in Fig. 2(c)

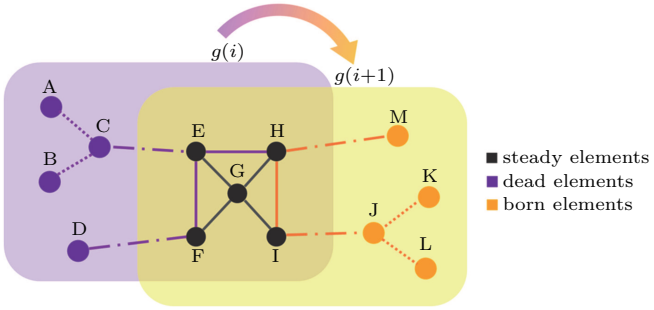


Fig. 1. (color online) Schematic diagram of network evolution from $t(i)$ to $t(i+1)$. The two shades $g(i)$ and $g(i+1)$ refer to the topology snapshot at time i and $i+1$, respectively. The overlap of both shades denotes the area that remains unchanged from $t(i)$ to $t(i+1)$, covering the steady nodes {E, H, G, F, I} and edges. The purple shade on the left contains the vanishing nodes {A, B, C, D} and edges $\{E_{\text{dead}}^{\text{inner}}, E_{\text{dead}}^{\text{boundary}}, E_{\text{dead}}^{\text{outer}}\}$, whereas the yellow shade on the right contains the new born nodes {J, K, L, M} and edges $\{E_{\text{born}}^{\text{internal}}, E_{\text{born}}^{\text{boundary}}, E_{\text{born}}^{\text{external}}\}$.

Dead node: for a node $n \in N(i)$, if n does not belong to $N(i+1)$, then n is considered as a dead node during $[i,i+1]$. All dead nodes comprise the dead nodes set:

$$N_{\text{dead}}(i,i+1) = N(i) - N(i) \cap N(i+1),$$

which are {A, B, C, D} in Fig. 1.

Newborn node: for a node $n \in N(i+1)$, but n does not belong to $N(i)$, then n is considered as a newborn node during $[i,i+1]$. All newborn nodes comprise the newborn nodes set:

$$N_{\text{born}}(i,i+1) = N(i+1) - N(i) \cap N(i+1),$$

which are nodes {J, K, L, M} in Fig. 1.

Dynamic nodes: all dead nodes and newborn nodes are considered as dynamic nodes.

Internal edge, boundary edge, and external edge: for an edge, if its two vertices are both steady nodes, then it is considered as an internal edge, such as {E–G, F–G, G–H, G–I, E–F, E–H, H–I} if its two vertices are both dynamic nodes, then it is considered as an external edge, such as {A–C, B–C, J–K, J–L}; otherwise it is considered as a boundary edge, such as {C–E, D–F, H–M, I–J}.

Dead edge: for an edge, if it belongs to $E(i)$ but does not belong to $E(i+1)$, then it is considered as a dead edge.

Newborn edge: for an edge, if it belongs to $E(i+1)$ but does not belong to $E(i)$, then it is considered as a born edge.

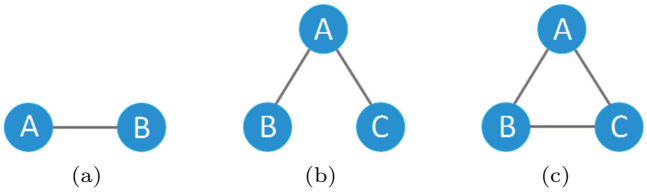


Fig. 2. Graphs of 3 simple components in complex networks. (a) Single-edge component M1; (b) Binary component M2, with three nodes and two edges; (c) The triangle component M3.

In some studies, the new-born nodes and dead nodes of the Internet IP-level network are treated equally, without considering the fact that some elements work together as one unit.[28,29] These cooperating elements contain up to three nodes, which are the most basic and fundamental elements after one singular node, so that each component consists of only two or three nodes and a few edges. A diagram of these three basic components ($M1$, $M2$, and $M3$) is shown in Fig. 2. The individual components can be defined as:

$M1$: the most basic component, which consists of two nodes and one edge. If this edge is a newborn internal edge, this subgraph is considered as an internal-born-$M1$ component, $M1_{\text{born}}^{\text{internal}}$. $M1_{\text{born}}^{\text{boundary}}$, $M1_{\text{born}}^{\text{external}}$, $M1_{\text{dead}}^{\text{internal}}$, $M1_{\text{dead}}^{\text{boundary}}$, $M1_{\text{dead}}^{\text{external}}$ are defined similarly.

$M2$: a network consists of three nodes and two edges. If both edges are new born internal edges, then it is considered as an internal-born-$M2$ component, $M2_{\text{born}}^{\text{internal}}$. $M2_{\text{born}}^{\text{external}}$, $M2_{\text{dead}}^{\text{internal}}$, $M2_{\text{dead}}^{\text{external}}$ are defined similarly. If at least one of the two edges is a new born boundary edge, then it is considered as a boundary-born-$M2$ component $M2_{\text{born}}^{\text{boundary}}$. $M2_{\text{dead}}^{\text{boundary}}$ is defined similarly.

$M3$: a network consists of three fully connected nodes. If all three edges are new born internal edges, then it is considered as an internal-born-$M3$ component, $M3_{\text{born}}^{\text{internal}}$. $M3_{\text{born}}^{\text{external}}$, $M3_{\text{dead}}^{\text{internal}}$, $M3_{\text{dead}}^{\text{external}}$ are defined similarly. If at least one of the three edges is a new born boundary edge, then it is considered as a boundary-born-$M3$ component $M3_{\text{born}}^{\text{boundary}}$. $M3_{\text{dead}}^{\text{boundary}}$ is defined similarly.

4.2. Related measurement

Researchers have modified a few entropy change formulas in various interdisciplinary applications.[30–33] To observe the system structure state during the evolution of the Internet,

we selected one which depicts topological feature. The structural entropy of complex network was defined as in Eq. (1).[33]

$$H = \frac{-2\sum_{i=1}^{N} \ln I_i - \ln 4(N-1)}{\ln N^2 - \ln 4(N-1)}, \quad (1)$$

where I_i is $k_i/\sum_{i=1}^{N} k_i$, N is the number of network nodes, and k_i denotes the degree of node i in the network. H is a topological feature of complex systems, which measures the disorder or randomness of a complex network, which is in the range [0, 1], the greater the entropy, the higher the disorder.[34,35] In particular, when applied to life, which is an organization, at every scale, the evolution of life is the increase of biological organization. Entropy is often used as a measure of the structural order and hierarchical structure within any scale of living organizations.[36]

5. Behavioural analysis

5.1. Growth of nodes and edges

Communication networks, social networks, power grids, protein–protein interaction networks and neural networks can all be considered as complex networks consisting of a series of nodes and edges, on which signals and information can be transmitted from the source nodes to the target nodes by specific routing rules. The global topology of a complex network can influence the local information flow direction and the nodes' load. Local changes can cause system fluctuation and lead to large scale network crashes due to a cascading failure reaction.

For example, with the development of the Internet, nodes and edges in the Internet have increased significantly. Figures 3(a) and 3(b) show numbers of nodes and edges of the Internet from January 1998 to August 2013, and piecewise regressions are used to reflect the trends of the growths of nodes and edges. In Fig. 3(a), according to the distribution of the number of nodes vs date, in the range of January 1998 to July 2001, as well as in the range of January 2003 to August 2013, an exponential regression model and a linear regression model are used to fit the data, respectively. This regression model is described by (where x is the number of days from January 1998, and y is the number of nodes):

$y = 2411.35 * e^{x/25.34} + 631.75$, where x is in the range of January 1998 to July 2003;

$y = -92.43 + 244.89 * x$, where x is in the range of January 2003 to August 2013.

In Fig. 3(b), according to the distribution of the number of edges vs. date, in the range of January 1998 to July 2001, as well as in the range of July 2001 to August 2013, one exponential regression model and a polynomial regression model are used. This regression model is described by (where x is the number of days from January 1998, and y is the number of edges):

$y = 5099.52 * e^{x/24.06} + 181.44$, where x is in the range of January 1998 to July 2001;

$y = 3.57 * x^2 + 18.93 * x + 22190.24$, where x is in the range of July 2001 to August 2013.

The regression models show the Internet global progress from a very simple topology with a few nodes and edges in 1998 to a fairly complicated topology with an enormous number of nodes and edges in 2013, they briefly show the trend of developments of nodes and edges of the Internet topology. The number of nodes and edges of the Internet have been increasing continuously (Figs. 3(a) and 3(b)), with nodes of the Internet's topology emerging and vanishing simultaneously. The y axises show the measures of the number of nodes and edges, where are 0 to 5×10^4 and 0 to 1.8×10^5, respectively (Figs. 3(a) and 3(b)), which imply the growth of network edges outpaces the growth of network nodes. Given that the network topology is adding and removing nodes constantly, each node change will have a follow-up effect. Like the butterfly effect, a small change at one place in a complex system can have large effects elsewhere. With these simulation models, we make an effort to give a brief estimation, or a prediction, that this trend of the Internet will progress in the near future.

We investigated the process of evolution of the birth and dead activities separately (Fig. 3(c) to Fig. 3(f)). Figure 3(c) gives the numbers of newborn nodes and dead nodes, whereas Figs. 3(d), 3(e), and 3(f) demonstrate the numbers of newborn and dead internal-edges, boundary-edges, and external-edges, respectively. In each time slot, the number of born nodes is always greater than that of dead ones (Fig. 3(c)). The amount of the newborn nodes and dead nodes evolve synchronously.

The **internal-edges** (E^{internal}) mainly take part in the local restructuring to make the Internet not only meet the requirement of system growth but also improve system robustness. The proportion of E^{internal} is about 90%. These edges participate in the Internet evolutionary process whenever necessary. They are formed by the nodes which select better neighbors for themselves by optimizing the connections. The **boundary-edges** (E^{boundary}) ensure that every newborn node can be part of the Internet in time and that dead nodes are removed from the Internet topology effectively. The proportion of these edges is about 9%. Only about 1% of the edges are **external-edges** (E^{external}). They are removed from the network topology together with the related endpoints. This implies that there are interdependencies in between these two newborn nodes or dead nodes, and none are discrete individuals. The internal-edges dominate the dynamical growth of the Internet (Fig. 3) in the sense of the activities of birth or death of nodes and their correlated edges. Here, trends of $E^{\text{internal}}_{\text{dead}}$ and $E^{\text{internal}}_{\text{born}}$ were fitted to exponential regression mod-

els, and it was found that the number of newborn and dead edges can be described by: $y_{\text{born}} = 712.34 * e^{x/79.97} + 253.41$ and $y_{\text{dead}} = 1064.67 * e^{x/103.65} - 390.44$, respectively.

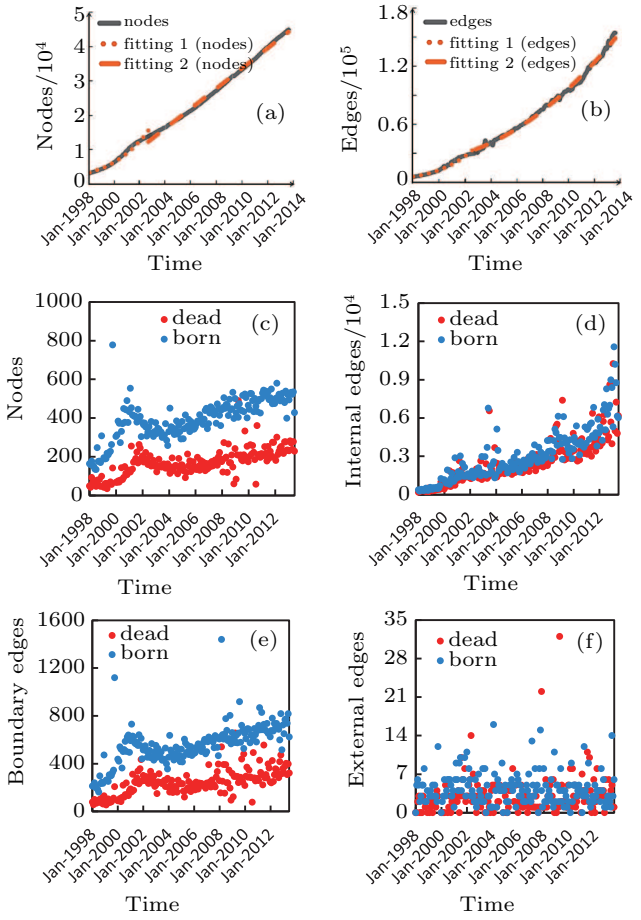


Fig. 3. (color online) Growth of nodes and edges of the AS-level Internet from January (Jan) 1998 to August (Aug) 2013. (a) Increase of the Internet nodes over time. (b) Increase of the Internet edges over time. (c)–(f) Comparisons of the dead (red dots) and born (blue dots) elements of the Internet network topology from January 1998 to August 2013, in which dead elements are represented by red dots and born elements are represented by blue ones. (c) The Internet nodes, N_{dead} and N_{born}; (d) Internal edges of the Internet, $E^{\text{internal}}_{\text{dead}}$ and $E^{\text{internal}}_{\text{born}}$; (e) Boundary edges of the Internet, $E^{\text{boundary}}_{\text{dead}}$ and $E^{\text{boundary}}_{\text{born}}$; (f) External edges of the Internet, $E^{\text{external}}_{\text{dead}}$ and $E^{\text{external}}_{\text{born}}$.

To observe the overall growth of the network edges, an estimate of the mean birth rate is described by:

$$r = \frac{1}{n=1}\sum_{i=1}^{n-1}\frac{|E_{\text{born}}(i,i+1) - E_{\text{dead}}(i,i+1)|}{|E(i)|}, \qquad (2)$$

where $E_{\text{born}}(i,i+1)$ and $E_{\text{dead}}(i,i+1)$ denote the born edges and dead edges accrue in the sampled network in the duration from $t(i)$ to $t(i+1)$, and n represents the number of sampled networks. The mean birth rate of the Internet from 1998 to 2013 is calculated, where $r = 0.96\%$. Furthermore, we infer that the Internet is currently in a process of slow growth. This state of slow growth ensures that the Internet system does not diverge from the system equilibrium. Interestingly, a similar trend was observed for numbers of nodes and boundary edges (Figs. 3(c) and 3(e)), likely because a newborn node has to build connections ($E^{\text{boundary}}_{\text{born}}$) with the existing nodes in the Internet, and dead nodes cut off the connections ($E^{\text{boundary}}_{\text{dead}}$) from the Internet system.

According to these results and the general principle of networking, an Internet evolution scenario is described as a newborn node usually builds connections (boundary edges) with one or more than one existing nodes first to join into the Internet system. Furthermore, some nodes are interconnected by external-edges which are added to the Internet at the same time. They then participate in the local restructuring activities by adjusting local connections. Although the nodes and edges are the basic elements of the complex network, the network components representing the function structures are more relevant to study the complex behaviors of the Internet's evolution. Therefore, in the next section, we will investigate the evolution of network components.

5.2. Component reconstruction

The internal-edges are the main factors governing the evolution of the Internet, while the effect of the two other types of edges is marginal. Therefore, whether the internal-edges will cooperate with other types of edges to co-build a network is of interest, such as by sharing common nodes. In other words, might the internal-edges participate to form any network components, while the boundary edges and external edges form the network components dynamically? How do those network topology components evolve? A router cannot establish the connections with an unlimited number of routers instantly and the old Internet architecture cannot adapt to the modern technology. Therefore, network administrators have to change the old Internet architecture or the local network connections to improve the Internet's efficiency. This is a scenario of network component reconstruction, and the reality becomes even more complicated. This study simplified these complicated scenarios into the following procedure: Any changes within the network, either addition or removal of nodes, edges or components, will break the system's equilibrium. However, the system itself will strive to maintain equilibrium by adding or removing some certain elements. Our work investigated three particular types of network component, $M1$, $M2$, and $M3$; as shown in Fig. 2.

Table 2 demonstrates the evolution of components $M1$, $M2$, and

$M3$

from 1998 to 2013 by showing the numbers of born and dead components of 16 sampled topologies of the Internet. It can be found that the internal components play a major role in the evolution of the Internet. This phenomenon of evolution is more obvious among internal components than among boundary components and external components. From Table 2 it can be seen that the number of the external-components is

lower than the number of the internal-components and the boundary-components. This can be concluded from the fact that the number of the external-edges is lower than the number of both the internal-edges and the boundary-edges (Fig. 3(d)–Fig. 3(f)). Second, according to Table 2, component $M3$ appears to be absent from boundary-components and external-components. Third, the number of newborn $M2$ components is greater than that of born $M1$, while the number of dead components $M2$ is similar to that of dead $M1$s. This indicates that the newborn node prefers to build up connections with at least three nodes and that the dead nodes are the leaf nodes of the network. It also verifies the notion that the number of connections of a node is positively related to the robustness of a node.[28] Referring to internal-components, which are composed of internal edges, component $M2$ is the main contributor to the dynamical structure of internal-components, which accounts for 95% $\sim$ 97% (Table 2), and the number of $M3$ is the lowest contributor (*ca*. 1.04%). These results indicate that component $M2$ is more important than the other components in the process of the Internet's evolution. This result also confirms the observation that the Internet consists mainly of star-like or tree-like structures. In other words, a node prefers to get started by connecting with multi-nodes to form an $M2$ component in order to take part in the network topology. Afterwards, an $M2$ component connects to other $M2$s to form a tree. The network keeps replicating $M2$s and connecting them together continuously. As for $M3$, the particular triangle shape could supply more stability to its own and also to the whole topology. Therefore, the stability of the Internet is positively related to the number of $M3$ components that it contains. The network components' reconstruction is completed by the activities of the edges. Since edge evolution is the process of system optimization, it can be suggested that components reconstruction also participates in the main process of system optimization.

So far in this study, the Internet's activities are analyzed from the node level, the edge level, and the component level. During the evolution of the Internet, the whole topology structure adds new nodes, whilst components are formed to reconstruct the system topology. Table shows that components are added into and removed from the Internet at the same time to keep the system equilibrium. For example, both the birth and death of elements can cause the system to diverge from equilibrium. The system reacts as a self-organizing system[25] to maintain the equilibrium, by adding or removing certain elements. Although it is impossible to distinguish whether birth will precede death, is there a correlation between components' births and deaths? To answer this question, the details of the Internet reconstruction are studied as follows.

Table 2. The evolution activities of network components $M1$, $M2$, and $M3$.

Sample period	Birth									Death								
	Internal			Boundary			External			Internal			Boundary			External		
	$M1$	$M2$	$M3$	$M1$	$M2$	$M3$	$M1$	$M2$	$M3$	$M1$	$M2$	$M3$	$M1$	$M2$	$M3$	$M1$	$M2$	$M3$
01 Feb. 98-01 Mar. 98	268	1240	0	202	1272	0	10	0	0	253	860	0	80	98	0	0	0	0
01 Feb. 99-01 Mar. 99	423	1980	6	251	1636	0	6	0	0	343	1238	0	57	112	0	0	0	0
01 Feb. 00-01 Mar. 00	733	34952	306	350	3658	0	4	0	0	561	15286	156	117	168	0	0	0	0
01 Feb. 01-01 Mar. 01	1170	64744	846	597	19766	0	12	0	0	921	14122	132	166	642	0	4	0	0
01 Feb. 02-01 Mar. 02	1198	20794	108	440	6260	0	10	0	0	1038	18172	108	229	1546	0	6	0	0
01 Feb. 03-01 Mar. 03	1286	36086	240	379	3748	0	6	0	0	1127	76520	900	180	470	0	2	0	0
01 Feb. 04-01 Mar. 04	1826	531222	29328	559	5652	0	23	20	0	1088	53330	222	170	476	0	0	0	0
01 Feb. 05-01 Mar. 05	1380	59250	330	396	1902	0	4	0	0	1363	36546	330	255	1008	0	0	0	0
01 Feb. 06-01 Mar. 06	1447	113216	1458	464	2298	0	6	0	0	1336	32308	192	211	468	0	4	0	0
01 Feb. 07-01 Mar. 07	1836	94338	990	504	4570	0	7	22	0	1614	22274	102	257	412	0	2	0	0
01 Feb. 08-01 Mar. 08	2226	123778	1794	649	2338	0	4	0	0	1632	94996	1362	183	354	0	2	0	0
01 Feb. 09-01 Mar. 09	2248	303998	2742	616	2628	0	8	2	0	1752	226644	3462	158	252	0	2	0	0
01 Feb. 10-01 Mar. 10	2023	510972	9090	631	2572	0	0	0	0	1897	352470	3828	339	1238	0	13	6	0
01 Feb. 11-01 Mar. 11	2297	585482	6456	764	2732	0	3	6	0	2200	88306	2346	331	482	0	6	2	0
01 Feb. 12-01 Mar. 12	2582	340956	4266	746	1978	0	6	0	0	2323	215892	672	295	608	0	2	2	0
01 Feb. 13-01 Mar. 13	2731	392744	3492	742	1930	0	6	0	0	2425	2755530	59280	391	912	0	7	20	0

To address this problem, an analogy is put forward to depict the process of evolution from a single node. One newborn node is like an infant growing up and making friends with other nodes over time, encountering some anomalies and being restrained by the surroundings. In this process, the system allows its participants to involve in the topology by using materials that already exist. A few of these new nodes may disappear after a short time but most of them will continuously participate in the "life-sustaining" process (which can also be called metabolism) within the network's topology. In this process, some of them even become important or fundamental members (nodes) of the network. Inspired by this analogy, we observed and analyzed the evolutionary process of node AS7713 (AS7713 represents the autonomous system number of a single administrative entity or domain whose information is registered in the RIR). Once AS7713 was added into the Internet in January 1998, its degree grew exponentially after October 2004 (Fig. 4(b)). This indicates that a new born node

adapted to the system environment for some time, and then quickly grew to become an important node (according to the case of AS7713). Figure 4(a) is a visualization of a local network including AS7713 and its neighbors in a snapshot network of the on August 2013.

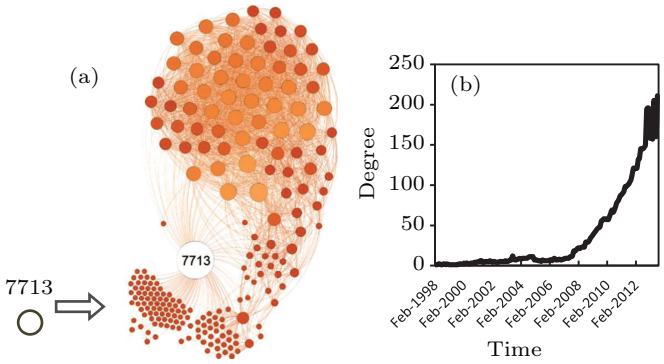


Fig. 4. (color online) A case of the evolution of node AS7713. (a) AS7713 emerges into the Internet in August 1998 and starts to build connections with its neighbors. The visualization shows a snapshot of the local network topology of AS7713 and its neighbors (August 2013). The colour and size of the nodes vary by node's degree, the bigger the node is, the lighter the colour is, and the higher degree it has, (b) the degree of AS7713 evolves since a newborn node to an important hub of the Internet.

In Subsection 5.1, we argued that the internal-edges are fundamental elements determining both the Internet network topology and its process of evolution, whereas the boundary-edges are used to receive the born elements or remove the dead elements. AS7713 was added into the Internet by building up boundary edges with those existing nodes first. By continuously establishing connections with its neighbors, it eventually became a hub in the Internet: a node with a greatly above average number of links. In addition, to describe AS7713's growth, the associated component of the evolutionary process is also analyzed to demonstrate the relationship between the born component and dead component.

Figure 5 contains three cases of reconstruction of internal-edge in the local subnetwork of AS7713 from 2010 to 2012, where the edges are categorized into three types: steady edges, born edges and dead edges. In Fig. 5(a), $M2$ component ({AS6939–AS38202, AS38202–AS4844} {AS6939–AS38202, AS38202–AS4826}, and {AS4844–AS38202, AS38202–AS4826}) are removed from the network, whilst $M3$ component ({AS55658–AS4844, AS4844–AS4826, AS4826–AS55658}) is added. This allows AS7713 access to AS4844 and AS4826 via AS55658, despite the disappearance of the connections of AS38202–AS4844 and AS38202–AS4826. According to Fig. 5(b), the local subnetwork turns from component ({AS23947–AS9304, AS23947–AS45896}) to component ({AS23947–AS55658, AS23947–AS18351}). However, in Fig. 5(c), the AS4844–AS2411 is removed while one $M1$ and one $M2$ are introduced into the network. The first case shows that the changes improve the network robustness because of the stability of $M3$. The second case shows the changes of the connection preference. The third case shows that the changes enhance the network connectivity by applying more internal edges.

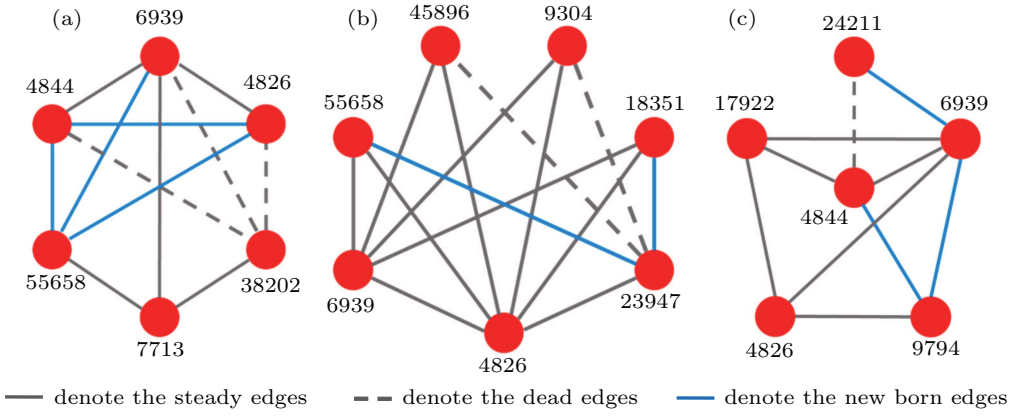


Fig. 5. (color online) Three Illustrations of network component reconstruction.

5.3. Network fluctuation

The equilibrium of the Internet can be disturbed by either the growth of nodes and edges, or reconstruction of the network components. As a complex system, changes of network nodes, edges, and components are the essential activities of the Internet. We wondered if those essential behaviors will change the basic characteristics of the Internet system. Therefore, we attempted to investigate it from the view of global Internet topology. The Internet is a typical complex network with scale-free characteristics,[8] where the degree distribution follows the power law, $p(k) \sim k^{\gamma}$ (k is the degree of the nodes in the network. For scale-free network, the γ value is typically in the range [2, 3], although it may occasionally lie outside these bounds, in a number of real world networks it has values between 1.2, 2.9). Power-law values γ of degree distributions of all the sampled networks are in the range [1.78, 2.25] and the standard deviation is 0.13. The degree distribution of the AS-level Internet network appears to follow a power law. The scale-free features of the network are relatively stable, which

refers to the growth process and attachment preference have not been changing much over time.

Figure 6 shows the evolution of the structural entropy of the Internet (H) from 1998 to 2013. According to the trend of the structural entropy over time, we can divide this period into two stages: structural entropy decreasing linearly from 0.205 to 0.175 and undulating within the range of [0.1725, 0.180]. From January 1998 to March 2003, the value of the structural entropy mainly decreased linearly along with some fluctuations, which mainly happened between November 2000 to July 2001. The result shows that the Internet tends towards lower structural entropy over time, and with the evolution of the Internet, the topology is becoming more organized, meaning the Internet becomes less random. This follows the general principle of evolution (*i.e.*, evolution in the extended sense is the increase of biological organization, and an irreversible process occurring in time), resulting in an increase of diversity and level of organization in its products.[37,38]

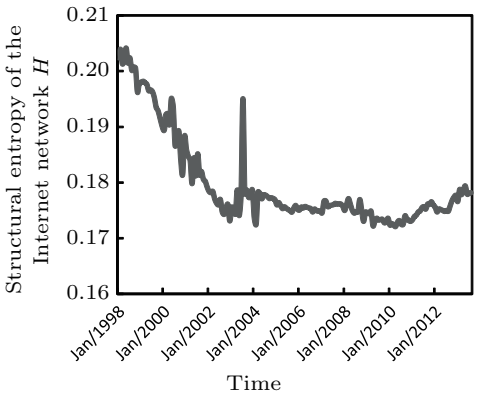


Fig. 6. The evolution of the structural entropy of the AS-level Internet from 1998 to 2013. From January 1998 to March 2003, the structural entropy of the Internet declined linearly. Afterwards, it mainly fluctuated in the range of [0.1725, 0.180].

From April 2003 to August 2013, the value of H of the Internet remained within the range of [0.1725, 0.180]. The tendency of the fluctuation is slow and smooth. It is worth noting that from 2000 to 2003, the Internet experienced a historic speculative bubble when the stock markets in industrialized countries saw their equity value rise rapidly due to growth in the Internet sectors and related fields.[39] Before 2003, the Internet Service Providers (ISPs) expanded the Internet infrastructure tremendously to meet the massive needs of Internet companies. After the Internet bubble, the Internet's resources were in excess because a large number of Internet companies went bankrupt. To decrease the operating costs and to improve the efficiency of network transmission, some ISP companies chose to merge with each other, while other companies optimized their self-network topology to provide better network services. No matter which strategy was chosen, the companies could not change the fact that the Internet topology had made a response to the surrounding environment. The Internet structural entropy decreased linearly from January 1998 to March 2003 (Fig. 6). This procedure is also the process of a rational development for commercial Internet. The Internet's status then entered a fluctuating phase, which implies that the AS-level Internet architecture and also the level of the randomness of the network topology did not change much after the development (from January 1998 to March 2003).

6. Discussion

The birth and death of network nodes, edges and component structures are an external expression of the Internet's metabolism. Our network evolution analysis considers the birth and death of network entities during the network growth process. The aims of this paper were to detect the Internet's evolutionary process from nodes-level, edges-level, and components-level, and precisely distinguish them into internal, boundary, and external units. Table 2 shows that nearly 95% of components reconstruct based on their existing original connections in the network structures. The remaining 5% of components evolve by adding or removing nodes to and from the original topology. This indicates that the Internet metabolism can be divided into two categories: metabolism by adding and removing nodes in and from the topology, and metabolism by reconstructing edges in the topology, but otherwise remaining all the nodes. The former mainly gets the external environment engaged with the Internet system, and the latter aims to optimize the original network structure. Metabolism of nodes mainly occurs on the border of the Internet. In other words, the newborn and dead nodes usually have small degrees. This implies that a new node prefers to join the Internet system from the surface rather than from the deep core of the Internet. In this process, most of the new nodes enter the network, and then some of them will participate in network reconstruction and become important nodes of the network topology. At the same time, some nodes disappear from the Internet because they lose all the connections whilst competing with others or they just simply leave temporarily. Here, the behavior of replicating component structure is termed "network component reconstruction". $M2$ can build star or tree topologies while $M3$ could improve the system stability according to the rectangle principle. Reconstruction does not change the Internet global topology but restructures the local path of the information transmission. The power-law degree distribution verifies that component reconstruction is a behavioral series to make sure the Internet system functions properly in the process of the Internet's evolution. Inevitably, being affected by the external environment, the Internet happens to fluctuate correspondingly. Our analysis shows that the Internet's entropy decreases in the process of evolution. It takes some time to reach an organized state and then its structure condition starts to fluctuate. By tracking and investigating those specific el-

ements, the results lead us to recognize the motions of every element of the network topology.

The facts that we observed in evolutionary process of the Internet can be summed up as follows: first, as a complex system, the Internet presents some complicated characteristics, such as self-organization, scale-free and synergetic. Inspired by the behavioral dynamics of biological groups, the process of the Internet evolution was found similar to some phenomena of biological evolution, such as metabolism and replication. For example, the growth of nodes and edges is comparable to biological metabolic activities. In biology, metabolism is a term that is used to describe all of the chemical reactions involved in maintaining the living state of the cells and organisms. It is usually divided into two categories: catabolism (the breakdown of large molecules to smaller molecules in order to obtain energy) and anabolism (a set of constructive metabolic processes, aimed to synthesize complex molecules). The node and edge entities can be seen as the metabolites during the growth process of Internet. We also suggest that the reconstruction of network components (the self-replicative process of the Internet system) is comparable to replication. Since the resources that it uses to construct the topology are replicas of itself, the new topology is built up by rearranging and connecting them in new position or location. As an analogy, viruses can replicate but only by taking advantage of cell reproduction through a process of infection. Harmful prion proteins can replicate by converting normal proteins into rogue forms. Similarly, computer viruses use an approach in that they reproduce using the hardware and software already present on computers.

7. Conclusion and future work

With the continuous growth of the Internet's network, the number of nodes increases while the evolution of the edges enhances the function of the Internet and increases the complexity of its topology. This study shows that the evolution of network edges, especially that of internal edges, is the dominant force that drives the evolution of the Internet. The Internet network mainly consists of internal components, particularly $M2$ internal components. Newborn nodes primarily build multiple connections with existing nodes, while most of the dead nodes fall off from the edge of the network. It is noteworthy that during this evolutionary process, the Internet increased in size and became more efficient, and so on, but did not change the attachment preference and has remained in a stable state (in terms of the hierarchical structure of the Internet architecture) since 2003. The evolution of the Internet is governed by a principle according to which the network topology evolves towards increasing complexity.

By deconstructing the network topology into components, we make a preliminary study on the Internet's evolution considering the reconstruction of the three simplest functional components, namely: $M1$, $M2$, and $M3$. However, further work is needed to gain a better understanding of how more complicated components, such as those consisting of four, five or even six nodes, participate in the evolution of the Internet. Apart from this study on metabolism and replication of the Internet, an observation of the mutation and selection processes within the Internet's evolution would be of interest. In addition, this study will make viewing some of the large scale-properties of the Internet in light of general biological laws acceptable. Although different in certain aspects, the Internet has many similarities with biological intelligence. Consequently, this paper makes an effort to recognize the life phenomena in this artificial complex system.

Moreover, it would be of interest to model metabolism or to use the underlying principles in comparable evolutionary processes to make a comparison between the Internet and biology. This would enable us to predict the outcomes and future use of the Internet. We are currently taking a further observation of AS7713's (a hub) development and interactive activities, and we look forward to seeing some new results which would be of interest.

Acknowledgments

Xiao Liu would like to thank A Rus Hoelzel, Thomas Keggin and Mario Erandi Bonillas Monge for their helpful discussions and suggestions.

References

[1] Faloutsos M, Faloutsos P and Faloutsos C 1999 *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, August 30–September 3, 1999, Cambridge, Massachusetts, USA, p. 251
[2] Alderson D, Li L, Willinger W and Doyle J C 2005 *IEEE/ACM Transactions on networking* **13** 6
[3] Barabási A L 2009 *Science* **325** 5939
[4] Barabási A L 2013 *Phil. Trans. R. Soc. A* **371** 1987
[5] Shen-Orr S S, Milo R, Mangan S and Alon U 2002 *Nat. Genetics* **31** 1
[6] Park, Kihong and Williger W 2005 *The Internet As a Large-Scale Complex System.* (Oxford: Oxford University Press) p. 3:1
[7] Ma R T B, Lui J C S and Misra V 2015 *IEEE/ACM Transactions on Networking* **23** 1
[8] Zhou S and Mondragón R J 2005 *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, March 29–April 1, 2005, Hong Kong, China, pp. 163–167
[9] Guo H, Bu Y J, Lan J L and Liu L K 2011 *J. Syst. Eng.* **26** 2
[10] Eiben A E and Smith J E 2003 *Introduction to evolutionary computing* (Heidelberg: Springer)
[11] Pastor-Satorras R and Vespignani A 2007 *Evolution and structure of the Internet: A statistical physics approach* (Cambridge: Cambridge University Press) pp. 226–228
[12] Dhamdhere A and Dovrolis C 2008 *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, October 2008, Vouliagmeni, Greece, pp. 183–196
[13] Dhamdhere A and Dovrolis C 2011 *IEEE/ACM Trans. Netw.* **19** 5
[14] Pastor-Satorras R, Vázquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 25
[15] Vázquez A, Pastor-Satorras R and Vespignani A 2002 *Phys. Rev. E* **65** 6
[16] Zhang G Q, Quoitin B and Zhou S 2011 *Comput. Commun.* **34** 5
[17] Ai J, Zhao H, Carley K M, Su Z and Li H 2013 *Chin. Phys. B* **22** 078902

[18] Moore C, Ghoshal G and Newman M E J 2006 *Phys. Rev. E* **74** 3
[19] Oliveira R V, Zhang B and Zhang L 2007 *ACM SIGCOMM Comput. Commun. Rev.* **37** 4
[20] Kuramochi M and Karypis G 2004 *IEEE Transactions on Knowledge Data Engineering* **16** 9
[21] Estrada E and Rodriguez-Velazquez J A 2005 *Phys. Rev. E* **71** 5
[22] Li F, Wei L, Zhao H and Hu F 2013 *Discret. Dyn. Nat. Soc.* **2013** 5
[23] Maslov S, Sneppen K and Zaliznyak A 2004 *Physica A: Stat. Mech. Its Appl.* **333** 1
[24] Kiremire A R, Brust M R and Phoha V V 2014 *Comput. Netw.* **72** 14
[25] Mitchell J "Autonomous System (AS) Reservation for Private Use [EB/OL]" http://tools.ietf.org/html/rfc6996
[26] Liu Z, Xiao Q, Zhan Q, Gu C and Yang H 2017 *Physica A: Stat. Mech. Its Appl.* **478** 837
[27] Leskovec J, Rajaraman A and Ullman J D 2014 *Mining of massive datasets* (Cambridge: Cambridge University Press) pp. 68–70
[28] Lv C C, Si S B, Duan D L and Zhan R J 2017 *Physica A: Stat. Mech. Appl.* **471** 837
[29] Heylighen F 2008 *Encyclopedia of library and information sciences* **3** 1215
[30] Wang J B, Wang L and Li X 2016 *IEEE Trans. Cybern.* **46** 2782
[31] Wang L, Li X, Zhang Y Q, Zhang Y and Zhang K 2011 *PloS ONE* **6** e21197
[32] Demetrius L and Manke T 2005 *Physica A: Stat. Mech. Its Appl.* **346** 682
[33] Yan Y, Zhang S, Tang J and Wang X 2017 *Physica A: Stat. Mech. Its Appl.* **477** 149
[34] McGraw-Hill 2004 *Concise encyclopedia of chemistry* (New York: McGraw-Hill Professional)
[35] Sethna J 2006 *Statistical mechanics: entropy, order parameters, and complexity* (Oxford: Oxford University Press) p. 78
[36] Ulanowicz R E 2012 *Growth and development: ecosystems phenomenology* (Berlin: Springer Sci. & Business Media) p. 21
[37] Huxley J 1955 "Evolution and Genetics" in *What Is Science*? ed. Newman J R (New York, Simon and Schuster) p. 278
[38] Mayr E 1970 *Populations Species and Evolution* (Harvard: Harvard University Press) p. 106
[39] Ofek E, Eli and Richardson M 2003 *The Journal of Finance* **58** 3