Structural Decomposition Model for the Evolution of AS-Level Internet Topologies

BO JIAO[1] AND WENSHENG ZHANG [ID]2

[1]School of Mathematics and Big Data, Foshan University, Foshan 528000, China
[2]Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Corresponding authors: Wensheng Zhang (zhangwenshengaa@outlook.com) and Bo Jiao (jiaoboleetc@outlook.com)

ABSTRACT Modeling Internet graphs at the autonomous-system (AS) level is helpful for recognizing and predicting the development trend of evolving Internet topology from a macro perspective. In contrast to the global statistical models such as the power-law distribution of node degrees, the structural decomposition models can more effectively represent the local connection. In this paper, we propose a structure-based model. Starting with the classification of links among the AS nodes, the proposed model partitions the core and periphery of Internet graphs into 16 atomic-level solid and dotted components. Additionally, the model captures the stable evolving features of these components based on the UCLA dataset that continuously explore Internet graphs over a long historic period from 2001 to 2015. Finally, according to the structure-based model, we design a new Internet-topology generator. Compared with the recently proposed generators, the advantages of our generator are as follows: (1) it accurately captures the structure decomposition property studied in this work, (2) it performs best on three statistical properties of the distance, assortativity coefficient, and maximum degree, and (3) it exhibits the best comprehensive performance in terms of runtime and multiple graph properties.

INDEX TERMS Internet topology, structural model, power-law distribution, graph decomposition, complex network, scale-free network.

I. INTRODUCTION

The autonomous-system (AS)-level Internet topology has evolved over time, and its size (i.e., the number of AS nodes) has rapidly grown from 10,000 in 2001 to approximately 50,000 in 2015 [1]. In addition, many potential Internet structures, e.g., named data networking [2], [3], information centric networking [4], and location-based networking [5], may make the Internet evolve to a new era. Thus, predicting the evolution trend of the topology is critical. Moreover, topological models are commonly used tools to test network technologies, such as routing protocol [6]. In contrast to the large-scale realistic networks, these models can generate small-scale graphs with adjustable characteristics, which greatly reduce the cost and provide a variety of topological scenarios for testing. Therefore, accurately modeling and recognizing the evolution of Internet topology is important.

The development of Internet-topology models has gone through three stages, namely, random, power-law and structure-based graphs. Random graphs are generated by randomly connecting the edges of node pairs independent of one another, such as the Waxman model [7]. Although the random-graph models are easy to use, they are not capable of better capturing the statistical features of the Internet [8]. One of the important metrics used in topology analysis is the distribution of the node degrees. Faloutsos *et al.* [9] demonstrated that this property is well described by the power-laws in the Internet topology. In addition, more studies have focused on power-law models [10], such as the Inet-3.0 model. However, the inherent biases of the traceroute sampling and collection of border gateway protocol (BGP) data from limited vantage points have led researchers to investigate the true existence of power-laws in the topology,

---

The associate editor coordinating the review of this manuscript and approving it for publication was Zehua Guo [ID].

although the degree distribution is widely accepted to be heavy-tailed [11].

In contrast to the power-law models, which focus on global statistical properties, the structure-based models pay more attention to the local structures, which represent the unique characteristics of the Internet. AS is a set of routers within a single administration domain, and the Internet is built on two domain categories, i.e., transit and stub [12], where the transit AS usually carries traffic among other domains and a stub AS that is connected to the end hosts relies on at least one transit AS for connectivity to the rest of the Internet. According to the physical interpretation, transit-stub models [12], [13] have been widely used to represent the hierarchical structure of the Internet. However, fewer types of AS categories usually lead to poor performance of the generators that use these models in terms of statistical properties, such as the power-law of the node degrees [13]. In addition, more studies investigated the local structures [14]–[23]. Zhou *et al.* [16] found that the rich-club connectivity phenomenon, namely, the Internet core, has a high connection density. Carmi *et al.* [17] partitioned the Internet using k-shell decomposition. Çetinkaya *et al.* [18] analyzed the Internet backbone. Jia *et al.* [19] studied the structure of the evolving IPv6 Internet. Liu *et al.* [20] partitioned the Internet into single-edge, binary, and triangular components. Accongiagioco *et al.* [21] found that the Internet core can be broken down into two layers. Lei *et al.* [22] studied the structure entropy. Zu *et al.* [23] proposed a city-level IP geolocation algorithm that decomposed the Internet through the geographical location of nodes. However, the structural models that capture the evolution features of the Internet topology need to be further studied.

Recently, we have partitioned the Internet transit and stub AS nodes into seven categories by analyzing the physical mean of the normalized Laplacian spectral properties [24]–[29]. However, these works only provided a static node classification and have not yet structurally distinguished the core from the periphery. In this paper, the novelty is twofold. First, we establish an Internet core-periphery structural model composed of 16 atomic-level solid and dotted components and find many uniform distribution features that exist in the decomposed components based on the UCLA dataset that spans 15 years [1]. In addition, we observe that most of the node and edge properties of these components remain constant except for the top five highest degrees of transit AS nodes in the process of historical evolution from 2001 to 2015. Second, we propose an Internet-topology generator based on our structural model and numerically verifies that the generator demonstrates the best comprehensive performance in terms of runtime and multiple graph properties compared with the recently proposed generators.

The remainder of this paper is organized as follows. Section II describes the background and related work. Section III presents our Internet core–periphery structural model. Section IV presents the investigation of the evolutionary stability of the different components of our model using the UCLA dataset. Section V proposes a new generator,

namely, SICPS (Simulates Internet graphs using the Core-Periphery Structure). Section VI shows the results obtained by SICPS. Finally, Section VII concludes this paper.

II. BACKGROUND AND RELATED WORK
The Internet topology is usually studied at two levels: router and AS levels [30], [31]. The former adopts routers as nodes, whereas the latter, which is the focus of our research, maps a set of routers to an AS node and describes the BGP connections among these AS nodes.

A. GLOBAL STATISTICAL PROPERTIES
The AS-level Internet topology represents simple and undirected graph $G = (V, E)$, where V and E are the AS node and edge sets, respectively. Because the topology inherits the non-trivial property of complex networks, it is usually described using some statistical properties such as degree, distance, and clustering [30]–[36]. A commonly used degree property is degree distribution $P(k)$, which is the probability that a randomly selected node has k degrees. However, the Internet topology is usually represented by another degree property, namely, the complementary cumulative distribution function (CCDF) degree, which is defined as $F(k) = \sum_{d>k} P(d)$, because this degree property is closer to the power-law [10].

The average neighbor connectivity is a degree correlation property [32], which is defined as

$$K(k) = \sum_{k'=1}^{k_{max}} k' \left(\overline{k} P(k, k') \right) \big/ k P(k), \qquad (1)$$

where $P(k, k')$ denotes the joint degree distribution, namely, the probability that a randomly selected edge connects k- and k'-degree nodes, and $k_{\max}$ and $\overline{k}$ are the maximum and average degrees, respectively.

In addition, the assortativity coefficient [33] shows the statistic of the node interconnectivity, which is defined as

$$r = \frac{\|E\|^{-1} \sum_{i=1}^{\|E\|} j_i k_i - \left[\|E\|^{-1} \sum_{i=1}^{\|E\|} \frac{1}{2} (j_i + k_i) \right]^2}{\|E\|^{-1} \sum_{i=1}^{\|E\|} \frac{1}{2} (j_i^2 + k_i^2) - \left[\|E\|^{-1} \sum_{i=1}^{\|E\|} \frac{1}{2} (j_i + k_i) \right]^2},$$
$$(2)$$

where j_i and k_i are the degrees of the nodes at the ends of the ith edge and $\|\cdot\|$ denotes the cardinality of a set.

This paper chooses spectral property $R = \sum_{i=1}^{\|V\|} (1 - \lambda_i)^4$ to describe the distance, where λ_i ($i = 1, 2, \cdots, \|V\|$) are all the eigenvalues of the normalized Laplacian matrix of graph G [34] because our recent work has demonstrated that the spectral property R is a good indicator of the average path length [26]. However, the former can be calculated faster using the circle enumeration method than the latter [28].

Moreover, clustering is usually described by clustering coefficient $C(k) = 2m(k) \big/ (k(k-1))$, where $m(k)$ is the average number of links between the neighbors of k-degree nodes. Specifically, $C(k)$ can be summarized as average clustering coefficient $C = \sum_k C(k) P(k)$ [33].

Whereas the aforementioned properties represent the global statistical features of the topology, they neglect the local structure of the Internet, which is an essential factor that distinguishes the Internet from other complex networks.

B. INTERNET TOPOLOGY MODELS

1) INET-3.0 MODEL

Inet-3.0 is a classical power-law model of the Internet topology [10]. First, it generates n nodes and uses two power-laws, namely, the degree-rank and degree CCDF, to respectively determine the degrees of the top three and the other nodes. Second, it uses a preference connection rule and the fraction of degree one to construct a tree that includes n nodes. Finally, it fills all free degrees of the n nodes and outputs a graph.

2) ORBIS MODEL

ORBIS uses dK to characterize an Internet graph [14], where dK describes the correlations among the degrees of nodes in subgraphs with d nodes. In particular, $1K$ and $2K$ reproduce the degree and joint degree distributions, respectively. In addition, ORBIS uses two random-graph-construction algorithms to simulate the $1K$ and $2K$ Internet graphs. Mahadevan *et al.* [14] found that the $2K$ random graphs could capture more properties of the Internet topologies except clustering and indicated that the clustering could be reproduced by $3K$ random graphs. However, the design of $3K$ generators is still a problem that needs to be resolved [14].

3) S-BITE MODEL

S-BITE is a model based on the core structure [21]. The model partitions the Internet topology into two distinct blocks: the core, where the nodes are tightly interconnected, and the periphery. Moreover, it separates the core into two layers: the centrum, whose density curve remains almost constant during the five-year evolution of Internet from 2007 to 2011, and layer-1, whose degree distribution in the vertical networks, which is composed of the centrum nodes plus all the layer-1 nodes and their connections to the centrum, is approximately uniform. The model simulates the peripheral nodes using three global statistical parameters: p, P_1, and P_2. Specifically, p is a probability that a peripheral node connects to two interconnected core nodes. P_1 and P_2 describe two statistical curves that are respectively related to the edge-number and preference distributions of connecting to the core for each newly-added peripheral node. In other words, the local structures of the peripheral nodes are not considered. We note that preference distribution P_2 is time-dependent [21], i.e., the distribution is not constant during the Internet evolution from 2007 to 2011.

4) SINETL MODEL

According to Faloutsos' transit-stub model [9], [37], as shown in Fig. 1(a), we add more multi-homed links to construct a static-node-classification model [25], as shown in Fig. 1(b), in which the red dotted lines represent the
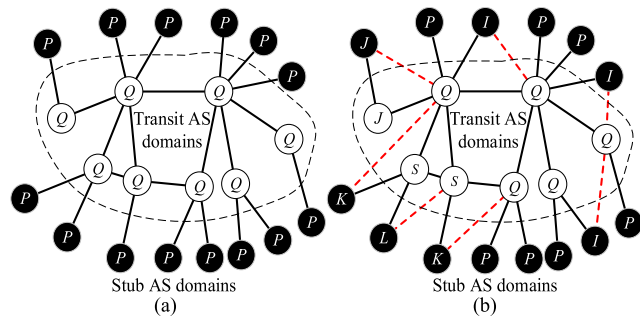

FIGURE 1. Two transit-stub models of the Internet topology. (*a*) Faloutsos's star-based model [9], [37]: deleting all the links between the transit AS nodes results in a non-connected graph that is constituted by the union of star subgraphs. (*b*) Our static node classification model [24], [25]: adding more multi-homed links for better fault-tolerance.

multi-homed links. Fig. 1(a) shows that each stub AS node is directly connected to a transit AS node through only one link in the early Internet topology, which is called a single-homed network. However, in terms of better fault tolerance, an increasing number of stub AS nodes tend to be connected to transit AS nodes through more links, which is called a multi-homed network. Fig. 1(b) shows that more dotted (i.e., multi-homed) edges result in the transit and stub AS nodes to be partitioned into seven categories, namely, Q, S, P, I, J, K, and L. The seven node categories are defined as follows [25]:

$$\begin{cases} P = \{v \in V \,|\, d_G(v) = 1\} \\ Q = \{v \in V \,|\, \exists w, (v,w) \in E, w \in P\} \\ K = \{v \in V_I \,|\, d_{G_I}(v) = 1 \wedge \forall (v,w) \in E_I, d_{G_I}(w) > 1\} \\ S = \{v \in V_I \,|\, \exists w, (v,w) \in E_I, w \in K\} \\ I = \{v \in V_I \,|\, d_{G_I}(v) = 0\} \\ J = \{v \in V_I \,|\, d_{G_I}(v) = 1 \wedge \forall (v,w) \in E_I, d_{G_I}(w) = 1\} \\ L = \{v \in V_I \,|\, d_{G_I}(v) \geq 2 \wedge \forall (v,w) \in E_I, w \in S\}, \end{cases}$$
$$(3)$$

where $G_I = (V_I, E_I)$ is a subgraph of G that is induced by node set $V/(P \bigcup Q)$ and $d_G(v)$ and $d_{G_I}(v)$ are the degrees of node v in graph G and subgraph G_I, respectively.

According to Eq. (3), all one-degree nodes in Internet graph G are marked as P, and all nodes that are connected to P are marked as Q. If all the P and Q nodes are removed from G, the following would result in remaining graph G_I, in which all one-degree nodes with neighbors whose degree is not less than two are marked as K, all nodes that are connected to K are marked as S, all one-degree nodes whose neighbors have a degree of one are marked as J, all zero-degree nodes are marked as I, and all d-degree nodes ($d \geq 2$) that can only be connected to S are marked as L. In addition, we consider all nodes in Internet graph G that have not been marked as noise.

Using the UCLA dataset [1] that contains AS graphs from 2001 to 2015, we find that the average percentage of noise nodes is 3.7% and the average percentage of the L nodes is 1.6%. Thus, the six nodes marked as Q, S, P, I, J, and K

TABLE 1. Cognitive perspectives of different models

Model	Cognitive perspective
Inet-3.0	Two power-law degree distributions, namely degree-rank and degree CCDF. [10]
ORBIS	$2K$ property that can be viewed as the joint degree distribution of a graph. [14]
S-BITE	Core decomposition structure, and the degree distribution of peripheral nodes. [21]
SInetL	Static node classification structure, and the normalized Laplacian spectral features of a graph. [29]
SICPS	Core-periphery decomposition structure that will be proposed in the next section.

FIGURE 2. Connection relationships among the six node categories *Q*, *S*, *P*, *I*, *J* and *K* of Internet graph *G*, where the symbol X_Y denotes the set that consists of all the *X* nodes connected to *Y* nodes. Note that the solid lines represent different types of edges that really exist in the graph, and the unidirectional and bidirectional dotted lines respectively establish the injective and bijection relationships between two node sets that belong to the same category. Specific-ally, *A* and *B* are the same node if the dotted line relation maps *A* to *B*.

can be used to accurately model the static-node-classification of the Internet graph.

Recently, we have designed a model SInetL to sample a given Internet graph using the static-node-classification feature [29]. The model extracts a subgraph using a sampling method of the six types of marked nodes. However, the model neglects the evolving correlation of the Internet graphs at different snapshots. The input of the sampling model is a complete graph that already contains all the topological information, whereas the input of the evolving simulation models only contains some topological features, which are more conducive to human cognition for a nontrivial topological structure. The sampling model is oriented to a static graph and is not suitable for structural cognition and trend prediction.

5) COMPARISON OF THE AFOREMENTIONED MODELS

The AS-level Internet topology is nontrivial [9]. Thus, researchers must recognize the topology from different perspectives. In general, a model is built from a cognitive perspective. Table I lists the cognitive perspectives of the aforementioned models. However, researchers usually do not care about the advantages and disadvantages of the different perspectives but pay more attention on whether the model properties from these perspectives exist in the explored graphs [10], [14], [21]. Thus, in Sections III and IV, we show that we use a series of explored AS graphs in the UCLA dataset to mine the properties of our model.

In the existing literature [10], [14], [21], [29], comparison of the models was mainly realized through comparison of the AS graphs that were simulated by generators because the purpose of the generators was to simulate AS graphs that contained properties of these models.

III. CORE-PERIPHERY STRUCTURAL MODEL

Eq. (3) shows that Internet graph *G* has many one-degree *P* nodes, and each *Q* node must be connected to at least one *P* node. If the one-degree nodes are called pendants and the nodes connected to the one-degree nodes are called quasi-pendants, subgraph G_I is also composed of pendants (*K* and *J* nodes), quasi-pendants (*S* nodes), and the remaining zero-degree *I* nodes, as expressed in Eq. (3), which exhibits the fractal structure of an Internet graph. We need to note that in subgraph G_I, each one-degree *J* node can only be
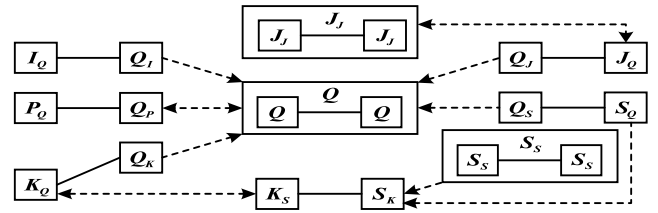
connected to another *J* node. In addition, Eq. (3) shows that in graph *G*, each *I* node must be connected to at least two *Q* nodes and each *K* (or *J*) node must be connected to at least one *Q* node.

According to the aforementioned analysis, we derive the connection relationships among *Q*, *S*, *P*, *I*, *J*, and *K*, as shown in Fig. 2, which classify all the edges of Internet graph *G* to six bipartite-graph sets and three interconnected sets, where a bipartite-graph set means that a traveler can move from one node set to another by one jump. By considering bipartite-graph edge set $I_Q - Q_I$ in Fig. 2 as an example, the subgraph induced by the edge set is a bipartite graph, where I_Q and Q_I are two disjoint node sets, and only edges from I_Q to Q_I are included in the subgraph.

In addition, Fig. 2 shows that the nine subgraphs induced by the distinct edge sets can be merged again using the nine injective or bijection relationships. According to Eq. (3), the structure shown in Fig. 2 can be analyzed as follows.

- First, we neglect the *Q* nodes that are not connected to other *Q* nodes because the percentage of such nodes is extremely small. According to this simplification, at the center in Fig. 2, the *Q* subgraph that is induced by all the edges linking two *Q* nodes includes all the *Q* nodes that are defined in Eq. (3). Except for the *Q* nodes, *S* represents the set of remaining transit AS nodes. However, the probability of connecting two *S* nodes is significantly less than that of connecting two *Q* nodes because the vast majority of the *S* nodes have low degrees. In other words, as shown in the lower right corner in Fig. 2, the S_S subgraph that is induced by all the edges linking two *S* nodes includes only part of the *S* nodes that are defined in Eq. (3).

- Second, according to Eq. (3), we can determine $P = P_Q$, $I = I_Q, J = J_Q = J_J, K = K_S = K_Q, Q = Q_P, Q_I \subseteq Q$, $Q_J \subseteq Q$, $Q_k \subseteq Q$, $S = S_K$, $S_S \subseteq S_k$, $S_Q \subseteq S_k$, and $Q_S \subseteq Q$.

- Finally, we use $X - Y$ to define a bipartite subgraph induced by all the edges that link the X_Y and Y_X nodes, where $X, Y \in \{P, Q, K, S, I, J\}$, and analyze some characteristics of these subgraphs as follows. (1) The degrees of all the P_Q nodes in the $P - Q$ subgraph are

one. (2) The degrees of all the K_S nodes in the $K - S$ subgraph are one. (3) The degrees of all the I_Q nodes in the $I - Q$ subgraph are not less than two. (4) The $J - J$ subgraph has even nodes, and the degree of each J_J node in the subgraph is one.

To extract the core, which is tightly interconnected, from the structure shown in Fig. 2, we predefine low-degree threshold d_{low}^{tra} for the transit AS nodes, high-degree threshold d_{hig}^{cor} for the core, and set $d_{low}^{tra} = 10$ and $d_{hig}^{cor} = 100$, which can capture the more stable characteristics of the evolution of the Internet (a detailed analysis is presented in Section IV).

In addition, we use the following steps to separate the Q subgraph denoted by $Q - Q$, which is induced by all the edges between the Q nodes, into five subcomponents.

Step 1: Derive a low-degree node set

$$Q_{bip}^{B}(l) = \left\{ v \mid v \in Q \land d_Q(v) \leq d_{low}^{tra} \right\},$$

where $d_Q(v)$ is the degree of node v in $Q - Q$. We let $Q_{bip}^{U} = Q / Q_{bip}^{B}(l)$ and obtain first subcomponent Q_{bip}, i.e., a subgraph of $Q - Q$ that is induced by all the edges between the $Q_{bip}^{B}(l)$ and Q_{bip}^{U} nodes.

In addition, we partition Q_{bip}^{U} into low-degree node set

$$Q_{bip}^{U}(l) = \left\{ v \mid v \in Q_{bip}^{U} \land d_{Q_{bip}}(v) \leq d_{low}^{tra} \right\}$$

and high-degree node set $Q_{bip}^{U}(h) = Q_{bip}^{U} / Q_{bip}^{U}(l)$, where $d_{Q_{bip}}(v)$ is the degree of node v in subcomponent Q_{bip}.

Step 2: Let Q_{cor} be a subgraph of $Q - Q$ that is induced by node set Q_{bip}^{U} and decompose Q_{bip}^{U} into

low-degree node set

$$Q_{cor}^{U}(l) = \left\{ v \mid v \in Q_{bip}^{U} \land d_{Q_{cor}}(v) \leq d_{low}^{tra} \right\},$$

high-degree node set

$$Q_{cor}^{U}(h) = \left\{ v \mid v \in Q_{bip}^{U} \land d_{Q_{cor}}(v) \geq d_{hig}^{cor} \right\},$$

and middle-degree node set

$$Q_{cor}^{U}(m) = Q_{bip}^{U} / \left(Q_{cor}^{U}(l) \cup Q_{cor}^{U}(h) \right),$$

where $d_{Q_{cor}}(v)$ is the degree of node v in subgraph Q_{cor}.

Then, we derive second-fourth subcomponents $Q_{cor}^{U}(h) - Q_{cor}^{U}(h)$, $Q_{cor}^{U}(h) - Q_{cor}^{U}(m, l)$, and $Q_{cor}^{U}(m, l) - Q_{cor}^{U}(m, l)$, namely three subgraphs of Q_{cor} induced by different types of edges among node sets $Q_{cor}^{U}(h)$ and $Q_{cor}^{U}(m, l) = Q_{cor}^{U}(m) \cup Q_{cor}^{U}(l)$.

Step 3: Let $Q_{red}^{B}(l)$ be a subset of node set $Q_{bip}^{B}(l)$, where each $Q_{red}^{B}(l)$ node has at least one $Q_{bip}^{B}(l)$ neighbor in $Q - Q$. Then, we can obtain fifth subcomponent

$$Q_{red}^{B}(l) - Q_{red}^{B}(l),$$

which is a subgraph of $Q - Q$ that is induced by all edges between the $Q_{red}^{B}(l)$ and $Q_{red}^{B}(l)$ nodes. ∎

Because node set Q is derived by P, the topological structure associated with the Q and P nodes is shown in Fig. 3, in which the structural partition of the $P - Q$ subgraph is
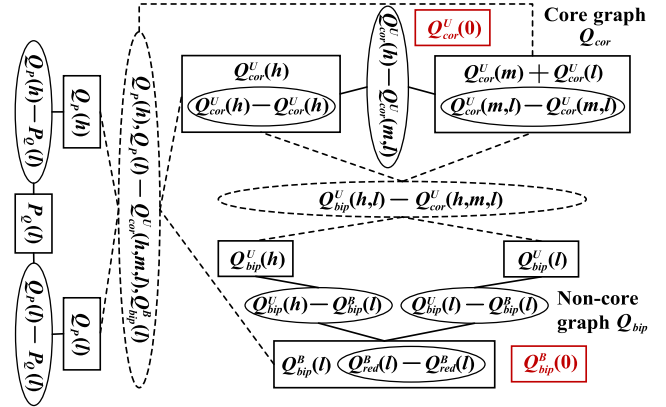

FIGURE 3. Structural partition of the subgraph induced by Q and P nodes, where the solid ellipses represent the distinct types of edges that really exist in Inter-net graph G, the dotted ellipses establish the mapping (injective and bijection) relationships described in Fig. 2, and the rectangles denote node sets. Note that the left solid connected component is the P-Q subgraph, $Q_{bip}^{U}(h, l) = Q_{bip}^{U}(h) \cup Q_{bip}^{U}(l)$ and $Q_{cor}^{U}(h, m, l) = Q_{cor}^{U}(h) \cup Q_{cor}^{U}(m) \cup Q_{cor}^{U}(l)$.

on the left and that of the $Q - Q$ subgraph is on the right. We define $Q_{cor}^{U}(0)$ as the set of zero-degree nodes in Q_{cor} and $Q_{bip}^{B}(0)$ as a set of $Q_{bip}^{B}(l)$ nodes that are not connected to the Q_{bip}^{U} nodes. In Fig. 3, we neglect the $Q_{cor}^{U}(0)$ and $Q_{bip}^{B}(0)$ nodes because the percentage of such nodes is much less than that of the rest of the topology. In the $P - Q$ subgraph in Fig. 3, $P_Q(l) = P_Q$ means that all the P_Q nodes have low degrees. In addition, Q_P can be divided into low-degree node set $Q_P(l) = \left\{ v \mid v \in Q_P \land d_{P-Q}(v) \leq d_{low}^{tra} \right\}$ and high-degree node set $Q_P(h) = Q_P / Q_P(l)$, where $d_{P-Q}(v)$ is the degree of node v in the $P - Q$ subgraph. As well known, Fig. 3 shows the extraction of the core Q_{cor} and other two components, i.e., Q_{bip} and $P - Q$, from the perspective of edge classification. We need to note that the dotted ellipses in Fig. 3 can also be viewed as edge connections.

Fig. 2 shows that the Q subgraph forms the connection center of the other peripheral structures. At the center of Fig. 4, we reduce the Q subgraph into four types of node sets, namely, $Q_{cor}^{U}(h)$, $Q_{cor}^{U}(m)$, $Q_{cor}^{U}(l)$, and $Q_{bip}^{B}(l)$, which include all the Q nodes. Using the UCLA dataset [1], we confirm that all stub AS nodes P, I, J, and K have low degrees because the maximum degree of these nodes did not exceed 30 from 2001 to 2015. Thus, Fig. 4 shows the corresponding node sets as $I_Q(l)$, $J_Q(l)$, $J_J(l)$, $K_Q(l)$, and $K_S(l)$. However, transit AS nodes Q and S have two types of degrees, i.e., low and high degrees. Thus, in each subgraph X–Y shown in Fig. 4 ($X, Y \in \{P, Q, K, S, I, J\}$), which is induced by node sets X_Y, Y_X and all the edges between X_Y and Y_X, we partition Y_X into low-degree node set $Y_X(l) = \left\{ v \mid v \in Y_X \land d_{X-Y}(v) \leq d_{low}^{tra} \right\}$ and high-degree node set $Y_X(h) = Y_X / Y_X(l)$, where $Y \in \{S, Q\}$ and $d_{X-Y}(v)$ is the degree of node v in the X–Y subgraph. We note that threshold d_{low}^{tra} is reset to 30 for $Q_S(l)$ and $S_Q(l)$ in the S–Q subgraph based on the analysis presented in Section IV.F. According to the node-set decomposition,
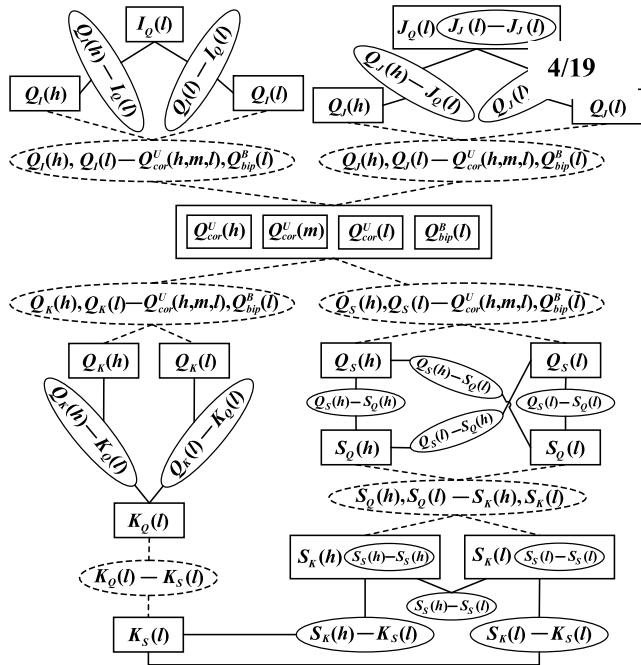
FIGURE 4. Structural partition of Internet graph G except for P nodes, where the solid ellipses represent the distinct types of edges that really exist in the graph, the rectangles denote node sets, and the dotted ellipses U-W establish the map-ping relationships between two node sets U and W that belong to the same cat-egory, for example, $S_Q(h)$, $S_Q(l) - S_K(h)$, $S_K(l)$ means $U = S_Q(h)$, $S_Q(l) = S_Q(h) \cup S_Q(l)$ and $W = S_K(h)$, $S_K(l) = S_K(h) \cup S_K(l)$.

we extend the relationships shown in Fig. 2 to the structural model shown in Fig. 4.

In summary, the combination of Figs. 3 and 4 shows our core-periphery structural model of the Internet topology. Figs. 3 and 4 show eight solidly connected branches in which each branch consists of one solid component and some additional edge sets. Eight dotted-connected branches are also present in which each consists of one dotted component. The sixteen components of our model are listed in Table 2.

In Section IV, we present the similarity between the Accongiagioco's core model [21] and our model. In addition, our model exhibits a peripheral structure using fruitful bipartite-graph connection relationship in detail. Furthermore, we explore the stable characteristics of our model using the UCLA dataset, as presented in Section IV.

IV. EVOLUTIONARY STABILITY ANALYSIS

Our analysis uses the UCLA dataset because it provides public AS graphs that span 15 years from January 2001 to January 2015 [1]. This section first lists some notations associated with our structural model in Table 3 and then presents the analyses of the evolutionary stability of the distinct components of our model, as shown in Figs. 3 and 4 and Table 2.

A. Q_{cor} COMPONENT

First, we analyze the node properties of the Q_{cor} component, which has three node sets, namely, $Q_{cor}^U(h)$, $Q_{cor}^U(m)$, and

TABLE 2. Sixteen solid and dotted components of our structural model.

| Item	Description
1	Solid component Q_{cor} (core subgraph)
2	Solid component Q_{bip} + Edge set $Q_{red}^B(l) - Q_{red}^B(l)$
3	Solid component $P - Q$
4	Solid component $I - Q$
5	Solid component $J - Q$ + Edge set $J_J(l) - J_J(l)$
6	Solid component $K - Q$
7	Solid component $K - S$ + Edge set $S_S(h) - S_S(h)$ + Edge set $S_S(h) - S_S(l)$ + Edge set $S_S(l) - S_S(l)$
8	Solid component $S - Q$ + Edge set $Q_S(h) - S_Q(h)$ + Edge set $Q_S(h) - S_Q(l)$ + Edge set $Q_S(l) - S_Q(h)$ + Edge set $Q_S(l) - S_Q(l)$
9	Dotted component $Q_{bip}^U(h, l) - Q_{cor}^U(h, m, l)$
10	Dotted component $Q_P(h)$, $Q_P(l) - Q_{cor}^U(h, m, l)$, $Q_{bip}^B(l)$
11	Dotted component $Q_I(h)$, $Q_I(l) - Q_{cor}^U(h, m, l)$, $Q_{bip}^B(l)$
12	Dotted component $Q_J(h)$, $Q_J(l) - Q_{cor}^U(h, m, l)$, $Q_{bip}^B(l)$
13	Dotted component $Q_K(h)$, $Q_K(l) - Q_{cor}^U(h, m, l)$, $Q_{bip}^B(l)$
14	Dotted component $Q_S(h)$, $Q_S(l) - Q_{cor}^U(h, m, l)$, $Q_{bip}^B(l)$
15	Dotted component $S_Q(h)$, $S_Q(l) - S_k(h)$, $S_K(l)$
16	Dotted component $K_Q(l) - K_S(l)$
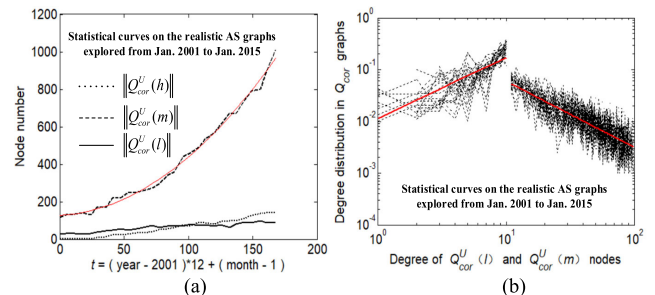
FIGURE 5. Evolving node properties of the Q_{cor} component in the UCLA dataset. (a) Node number *vs.* t (the number of months since January 2001). (b) Degree distribution *vs.* low- and middle-degrees, where $Q_{cor}^U(l)$ is a low-degree node set and $Q_{cor}^U(m)$ is a middle-degree node set.

$Q_{cor}^U(l)$. Fig. 5(a) shows that $\|Q_{cor}^U(h)\|$ and $\|Q_{cor}^U(l)\|$ linearly increases and $\|Q_{cor}^U(m)\|$ quadratically increases with exploration time t, which is defined as the number of months since January 2001. In addition, Fig. 5(b) shows that the degrees of the $Q_{cor}^U(l)$ nodes range from 1 to 10, whereas those of the $Q_{cor}^U(m)$ nodes range from 11 to 99. From the statistical curves of the UCLA graphs shown in Fig. 5(b), we observe that the low degrees of the $Q_{cor}^U(l)$ nodes and the middle degrees of the $Q_{cor}^U(m)$ nodes obey two distinct power-law distributions.

Fig. 5(b) shows that the low and middle degrees have constant distributions and do not change with time, whereas the high degrees tend to increase with time. We sort the degrees of the nodes in $Q_{cor}^U(h)$ by decreasing order and let the ith highest-degree be $d_{cor}^h(i)$. Fig. 6(a) shows that the $d_{cor}^h(i)$ versus t curve approximately obeys a linear relationship when $i \geq 6$, which is defined as

$$d_{cor}^h(i) = m_i \cdot t + b_i, \tag{4}$$

where m_i and b_i are the slope and intercept of the linear relationship, respectively.

TABLE 3. Notations of our structural model.

Notation	Description
$G = (V, E)$	The AS graph with node set V and edge set E.
P, Q, K, S, I, J	The six node sets defined in Eq. (3).
X_Y	The set that consists of all the X nodes attached to Y nodes, where $X, Y \in \{P, Q, K, S, I, J\}$.
$X - Y$	The subgraph that is induced by all the edges whose two ends belong to X_Y and Y_X respectively.
$d_{low}^{tra} = 10$	A low-degree threshold for the transit AS nodes (except for the $S - Q$ subgraph).
$d_{hig}^{cor} = 100$	A high-degree threshold for the core.
$Q - Q$	A subgraph of graph G that is induced by node set Q.
$d_Q(v)$	The degree of node v in the subgraph $Q - Q$.
$Q_{bip}^B(l)$	$= \{v \mid v \in Q \wedge d_Q(v) \leq d_{low}^{tra}\}$.
Q_{bip}^U	$= Q / Q_{bip}^B(l)$.
Q_{bip}	A subgraph of $Q - Q$ that is induced by all the edges between $Q_{bip}^B(l)$ and Q_{bip}^U nodes.
$d_{Q_{bip}}(v)$	The degree of node v in the subgraph Q_{bi} **5/19**
$Q_{bip}^U(l)$	$= \left\{v \mid v \in Q_{bip}^U \wedge d_{Q_{bip}}(v) \leq d_{low}^{tra}\right\}$.
$Q_{bip}^U(h)$	$= Q_{bip}^U / Q_{bip}^U(l)$.
Q_{cor}	A subgraph of $Q - Q$ that is induced by node set Q_{bip}^U.
$d_{Q_{cor}}(v)$	The degree of node v in the core Q_{cor}.
$Q_{cor}^U(l)$	$= \{v \mid v \in Q_{bip}^U \wedge d_{Q_{cor}}(v) \leq d_{low}^{tra}\}$.
$Q_{cor}^U(h)$	$= \{v \mid v \in Q_{bip}^U \wedge d_{Q_{cor}}(v) \geq d_{hig}^{cor}\}$.
$Q_{cor}^U(m)$	$= Q_{bip}^U / (Q_{cor}^U(l) \cup Q_{cor}^U(h))$.
$Q_{red}^B(l)$	$= \{v \mid v \in Q_{bip}^B(l) \wedge \exists w \in Q_{bip}^U(l), (v, w) \in E\}$
$d_{X-Y}(v)$	The degree of node v in the $X - Y$ subgraph.
$Y_X(l)$	$= \{v \mid v \in Y_X \wedge d_{X-Y}(v) \leq d_{low}^{tra}\}$, where $X \in \{P, I, J, K\}$ and $Y \in \{S, Q\}$.
$Y_X(h)$	$= Y_X / Y_X(l)$.
$S_S(h), S_S(l)$	$S_S(h) = S_S \cap S_K(h), S_S(l) = S_S \cap S_K(l)$. Note that $S_K = S, S_S \subseteq S$ and $S_Q \subseteq S$.
$S_Q(l)$	$= \{v \mid v \in S_Q \wedge d_{S-Q}(v) \leq d_{low}^{S-Q}\}$, where $d_{low}^{S-Q} = 30$.
$Q_S(l)$	$= \{v \mid v \in Q_S \wedge d_{S-Q}(v) \leq d_{low}^{S-Q}\}$, where $d_{low}^{S-Q} = 30$.
$S_Q(h), Q_S(h)$	$S_Q(h) = S_Q / S_Q(l), Q_S(h) = Q_S / Q_S(l)$.
$P_Q(l), I_Q(l)$	$P_Q(l) = P, I_Q(l) = I$.
$J_Q(l), K_Q(l)$	$J_Q(l) = J, K_Q(l) = K$.
$J_J(l), K_S(l)$	$J_J(l) = J, K_S(l) = K$.
$\|\cdot\|$	The cardinality of a specified set.

FIGURE 6. Evolving properties of high-degrees in the UCLA dataset. (*a*) Degree of Q_{cor}^U (*h*) nodes in the Q_{cor} component *vs. t* (the number of months since Ja-nuary 2001). (*b*) Degree of Q_{bip}^U (*h*) nodes in the Q_{bip} component *vs. t*. (*g*) Degree of Q_I (*h*) nodes in the $I - Q$ component *vs. t*. In terms of (*a*), (*b*) and (*g*), when $i \geq 6$, the ith highest-degree *vs. t* approximately obey a linear relation that can be represented by a tuple consisting of slope m_i and intercept b_i. Specifically, in (*c*) and (*e*), the parameter i is defined as the rank of Q_{cor}^U (*h*) nodes sorted by the decreasing order of degree in Q_{cor}; in (*d*) and (*f*), the parameter i is defined as the rank of Q_{bip}^U (*h*) nodes sorted by the decreasing order of degree in Q_{bip}; in (*h*), the parameter i is defined as the rank of Q_I (*h*) nodes sor-ted by the decreasing order of degree in $I - Q$.

However, the d_{cor}^h (i) versus t curve approximately exhibits some piecewise linear features when $i \leq 5$, which means that the evolutionary trends of the top five highest-degrees are not constant over the long historical process in 15 years. From the comparison of Figs. 6(*a*), (*b*), and (*g*), we observe that the high-degree feature widely exists in other components of the Internet topology. In addition, Figs. 6(*a*), (*b*) and (*g*) show the following characteristics.

- In Q_{cor} and Q_{bip} that consist of transit AS nodes, the probabilities of the top five highest degree nodes connected to other transit AS nodes rapidly increases because the slope of the lines shown in Figs. 6(*a*) and (*b*) indicates an increasing trend with time.
- In subgraphs $X - Q$, where $X \in \{P, I, J, K\}$, the probabilities of the top five highest degree nodes connected to stub AS nodes do not significantly changed, as shown in Fig. 6(*g*).

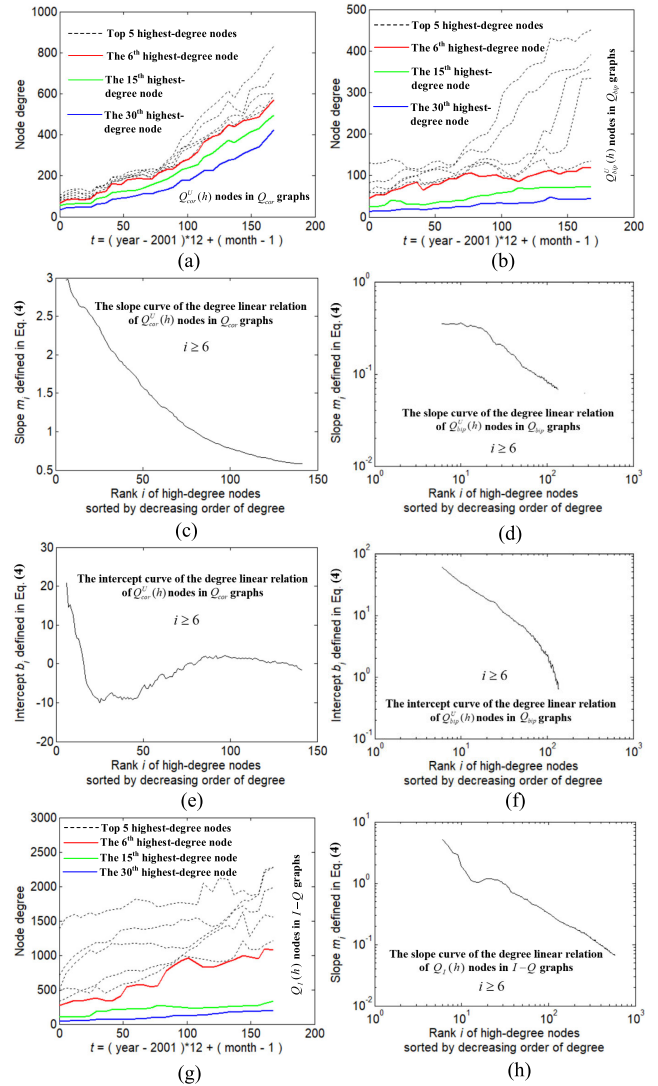In addition, Fig. 6 shows that the high degrees in the peripheral structure that consists of Q_{bip} and $X - Y$ subgraphs

exhibit better power-law relationships in the m_i (b_i) versus i curve when $i \geq 6$.

Next, we analyze the edge property of the Q_{cor} component and point out the similarity between Q_{cor} and the Accongiagioco's core model [21]. In [21], the core nodes are divided into two layers, namely, centrum and layer-1. The core is divided into three networks, namely, centrum network that consists of all the centrum nodes and their mutual connections, vertical network that consists of the centrum network plus all the layer-1 nodes and their connections to
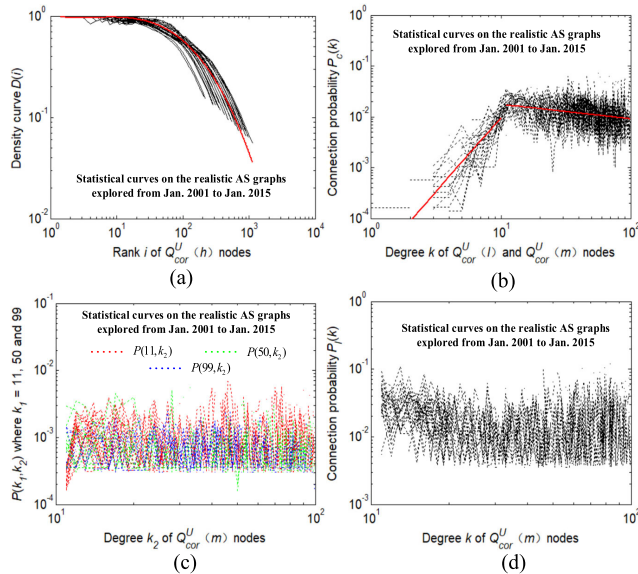
FIGURE 7. Evolving edge properties of the Q_{cor} component in the UCLA dataset. *(a)* Density curve $D(i)$ of the $Q_{cor}^U(h) - Q_{cor}^U(h)$ edge connections *vs.* rank i of $Q_{cor}^U(h)$ nodes. *(b)* Connection probability $P_c(k)$ *vs.* degree k of $Q_{cor}^U(m, l)$ nodes. *(c)* Joint degree distribution $P(k_1, k_2)$ of the $Q_{cor}^U(m) - Q_{cor}^U(m)$ edge connections ($k_1 = 11, 50$ and 99) *vs.* degree k_2 of $Q_{cor}^U(m)$ nodes. *(d)* Connection probability $P_l(k)$ *vs.* degree k of $Q_{cor}^U(m)$ nodes.

the centrum, and horizontal network that consists of only the layer-1 nodes plus their connections to other layer-1 nodes. Fig. 3 shows that our model also decomposes core Q_{cor} into three subgraphs, which are respectively induced by three types of edges in core Q_{cor}, namely, $Q_{cor}^U(h) - Q_{cor}^U(h)$, $Q_{cor}^U(h) - Q_{cor}^U(m, l)$, and $Q_{cor}^U(m, l) - Q_{cor}^U(m, l)$. In [21], the centrum network contains stable density curve $D(i) = 2e/(i(i-1))$ ($i = 1, 2, \cdots, n_c$), where $1, 2, \cdots, n_c$ represents all the nodes in the network sorted according to a certain order and e is the number of edges in the subgraph of the network induced by the first i nodes, $1, 2, \cdots, i$. In the present study, we also use the density curve to present the edge property in the $Q_{cor}^U(h) - Q_{cor}^U(h)$ subgraph and sort all the nodes in the subgraph according to a decreasing order of degree $d_{Q_{cor}}(v)$. Fig. 7(a) shows that the $Q_{cor}^U(h) - Q_{cor}^U(h)$ subgraphs of our model also exhibit a stable density curve on the UCLA dataset that spans 15 years.

Moreover, the degrees of the layer-1 nodes in the vertical network follow a uniform distribution [21]. To capture the feature, we define $P_c(k)$ as the probability that an $Q_{cor}^U(h) - Q_{cor}^U(m, l)$ edge connects a k-degree $Q_{cor}^U(m, l)$ node in core Q_{cor}. Fig. 7(b) shows that the $P_c(k)$ versus k curve also approximately obeys a uniform distribution as k falls in the middle degrees from 11 to 99. Thus, we infer that the $Q_{cor}^U(m)$ node set is associated with layer-1. We note that Fig. 5(a) shows that $\|Q_{cor}^U(m)\| \gg \|Q_{cor}^U(l)\|$, i.e., $Q_{cor}^U(m)$ occupies most of the nodes in core Q_{cor}. In addition, Fig. 7(b) shows that the $P_c(k)$ distribution remains stable in the UCLA dataset over the span of 15 years.

Because the degrees of the $Q_{cor}^U(l)$ nodes are obviously less than those of the $Q_{cor}^U(m)$ nodes in the Q_{cor} component, the $Q_{cor}^U(m) - Q_{cor}^U(m)$ edges represent the principal part of the $Q_{cor}^U(m, l) - Q_{cor}^U(m, l)$ connections. We use the joint degree distribution, which is defined as [32]

$$P(k_1, k_2) = \mu(k_1, k_2) \cdot m(k_1, k_2) / (2m), \quad (5)$$

where $m(k_1, k_2)$ is the number of edges connecting the nodes with k_1 and k_2 degrees, m is the total number of edges, and $\mu(k_1, k_2)$ is one if $k_1 = k_2$ and two otherwise to characterize the $Q_{cor}^U(m) - Q_{cor}^U(m)$ connection feature. Specifically, k_1 and k_2 represent the degrees in core Q_{cor}. Through the UCLA dataset analysis that spans 15 years, we find that $P(k_1, k_2)$ of the $Q_{cor}^U(m) - Q_{cor}^U(m)$ connections obeys a uniform distribution where $k_1, k_2 \in [11, 99]$, which is the range of the middle degrees, as shown in Fig. 7(c). We note that Fig.7(c) includes three extracted distributions $P(11, k_2)$, $P(50, k_2)$, and $P(99, k_2)$. Moreover, we define $P_l(k)$ as the probability that a $Q_{cor}^U(l) - Q_{cor}^U(m)$ edge connects a k-degree $Q_{cor}^U(m)$ node in core Q_{cor}. Fig. 7(d) shows that the $P_l(k)$ versus k curve also approximately obeys a uniform distribution. We note that the $Q_{cor}^U(l) - Q_{cor}^U(l)$ edges are neglected in our model because the number of such edges is very small.

B. Q_{bip} COMPONENT
First, we analyze the node properties of the Q_{bip} component, which contains three node sets, namely, $Q_{bip}^U(h)$, $Q_{bip}^U(l)$, and $Q_{bip}^B(l)$. Fig. 8(a) shows that the cardinalities of these node sets $\|Q_{bip}^U(h)\|$, $\|Q_{bip}^U(l)\|$, and $\|Q_{bip}^B(l)\|$ linearly increases with exploration time t. From the statistical curves shown in Figs. 8(b) and (c), we observe that the degrees of the $Q_{bip}^U(l)$ and $Q_{bip}^B(l)$ nodes have constant distributions. We note that the feature of high degrees associated with $Q_{bip}^U(h)$ has been analyzed, as shown in Figs. 6(b), (d), and (f).

Next, we analyze the edge properties of the Q_{bip} component. Specifically, $Q_{bip}^U(h) - Q_{bip}^B(l)$, $Q_{bip}^U(l) - Q_{bip}^B(l)$, and $Q_{red}^B(l) - Q_{red}^B(l)$ represent the corresponding edge connections in the component.

Because the $Q_{bip}^B(l)$ degree distribution is constant in the UCLA dataset over the span of 15 years, we adopt $P_b(k)$, which is defined as the probability that a $Q_{bip}^U(h) - Q_{bip}^B(l)$ edge connects a k-degree $Q_{bip}^B(l)$ node in Q_{bip}, to study the $Q_{bip}^U(h) - Q_{bip}^B(l)$ connection feature. Fig. 9(a) shows that the $P_b(k)$ versus k curve obeys a constant distribution in the UCLA dataset that spans 15 years.

In addition, because both the $Q_{bip}^U(l)$ and $Q_{bip}^B(l)$ degree distributions are constant in the UCLA dataset over the span of 15 years, we use joint degree distribution $P(k_1, k_2)$, which is defined in Eq. (5), to analyze the $Q_{bip}^U(l) - Q_{bip}^B(l)$ connection feature, where k_1 and k_2 are the degrees of the $Q_{bip}^U(l)$ and $Q_{bip}^B(l)$ nodes in Q_{bip}, respectively. According to the UCLA dataset, we observe that $P(k_1, k_2)$ of the $Q_{bip}^U(l) - Q_{bip}^B(l)$ connections for each given k_2 approximately follows
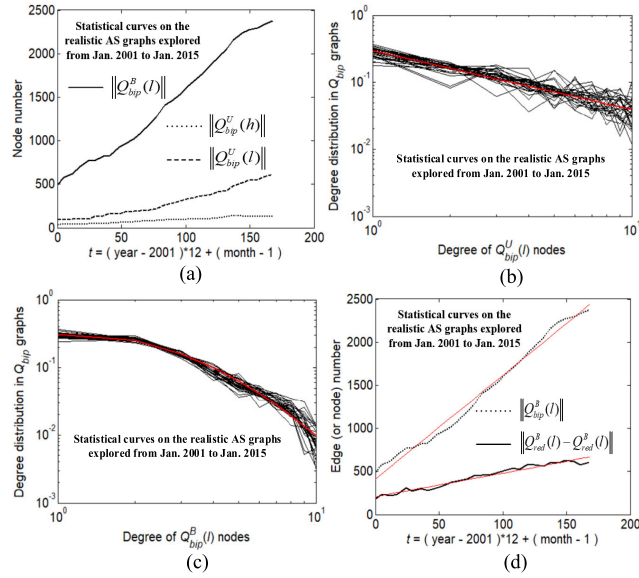
FIGURE 8. Evolving node properties of the Q_{bip} component in the UCLA dataset. (*a*) Node number *vs. t* (the number of months since January 2001). (*b*) Degree distribution of $Q_{bip}^U(l)$ nodes. (*c*) Degree distribution of $Q_{bip}^B(l)$ nodes. (*d*) Edge (or node) number *vs.t* (the number of months since January 2001), where $\left\|Q_{bip}^B(l)\right\|$ denotes the number of $Q_{bip}^B(l)$ nodes and $\left\|Q_{red}^B(l) - Q_{red}^B(l)\right\|$ denotes the number of edges connecting two $Q_{bip}^B(l)$ nodes in the component.
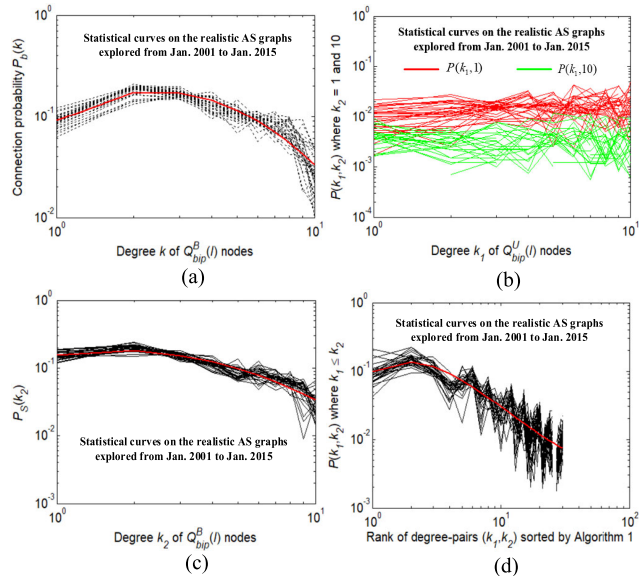

FIGURE 9. Evolving edge properties of the Q_{bip} component in the UCLA dataset. (*a*) Probability $P_b(k)$ *vs.* degree k of $Q_{bip}^B(l)$ nodes. (*b*) Joint degree distribution $P(k_1, k_2)$ of the $Q_{bip}^U(l) - Q_{bip}^B(l)$ connections where $k_2 = 1$ and 10 *vs.* degree k_1 of $Q_{bip}^U(l)$ nodes. (*c*) Probability $P_S(k_2)$ *vs.* degree k_2 of $Q_{bip}^B(l)$ nodes. (*d*) Joint degree distribution $P(k_1, k_2)$ of the $Q_{red}^B(l) - Q_{red}^B(l)$ connections *vs.* rank of degree-pairs (k_1, k_2) sorted by Algorithm 1.

a uniform distribution for $k_1 \in [1, 10]$, i.e., the range of the degrees of $Q_{bip}^U(l)$ nodes, as shown in Fig. 9(*b*).

For given k_2, we derive the following probability

$$P_S(k_2) = \sum_{k_1=1}^{d_{low}^{tra}} P(k_1, k_2), \quad (6)$$

and learn that the distribution of the $P_S(k_2)$ versus k_2 relationship is constant in the span of 15 years, as shown in Fig. 9(*c*). According to the aforementioned analysis, we can determine that

$$P(k_1, k_2) \approx P_S(k_2)\big/d_{low}^{tra}, \quad (7)$$

where $d_{low}^{tra} = 10$ is the maximum degree of the $Q_{bip}^U(l)$ nodes in the Q_{bip} component.

According to the list in Table 3, the $Q_{red}^B(l) - Q_{red}^B(l)$ edges are not included in Q_{bip}, but each $Q_{red}^B(l) - Q_{red}^B(l)$ edge connects two $Q_{bip}^B(l)$ nodes (with degrees k_1 and k_2) in Q_{bip}, i.e., the edge corresponds to degree pair (k_1, k_2), where $k_1, k_2 \in \left[1, d_{low}^{tra}\right]$ and $k_1 \leq k_2$. We sort the degree pairs associated with the $Q_{red}^B(l) - Q_{red}^B(l)$ connections using Algorithm 1.

Algorithm 1 Degree-Pair Sorting

1: **Input:** Upper bound U_b of the degrees.
2: **Output:** Degree-pair list *DPL* sorted by a certain order.
3: Initialize *len* $\leftarrow$ 2 and *DPL* $\leftarrow \emptyset$.
4: **While** *len* $\leq 2U_b$ **do**
5: Initialize $k_1 \leftarrow 1$ and $k_2 \leftarrow len - k_1$.
6: **While** $k_1 \leq k_2 \wedge k_1 \leq U_b \wedge k_2 \leq U_b$ **do**
7:    Add degree-pair (k_1, k_2) to the end of the degree-pair list *DPL*.
8:    Update $k_1 \leftarrow k_1 + 1$ and $k_2 \leftarrow len - k_1$.
9: **End while**
10: Update *len* $\leftarrow len + 1$.
11: **End while**

In Algorithm 1, line 1 inputs the maximum value of the degrees that may occur in the degree pairs. If the maximum value is U_b, line 2 outputs a list that contains all possible degree pairs (k_1, k_2), which satisfies $k_1 \leq k_2 \leq U_b$. Moreover, lines 3–10 confirm that degree pair (k_1, k_2) with a smaller value of $len = k_1 + k_2$ is located in front of the list.

We use joint degree distribution $P(k_1, k_2)$ defined in Eq. (5) to show the $Q_{red}^B(l) - Q_{red}^B(l)$ connections and use Algorithm 1 to sort all possible degree pairs (k_1, k_2). Fig. 9(*d*) shows that $P(k_1, k_2)$ of the $Q_{red}^B(l) - Q_{red}^B(l)$ connections also remains stable in the UCLA dataset over the span of 15 years. Furthermore, Fig. 8(*d*) shows that both $\left\|Q_{bip}^B(l)\right\|$ and $\left\|Q_{red}^B(l) - Q_{red}^B(l)\right\|$ linearly increases with exploration time t. Thus, at any t, we can establish the following relationship:

$$\left\|Q_{red}^B(l) - Q_{red}^B(l)\right\| = 0.23\left(\left\|Q_{bip}^B(l)\right\| - 413\right) + 207, \quad (8)$$

Eq. (8) shows that $\left\|Q_{red}^B(l) - Q_{red}^B(l)\right\|$ is very small because the slope of the linear expression of the edge number relative to the node number is approximately 0.23.

C. X − Y COMPONENT

This section illustrates that the $X - Y$ component is one of the five instances, namely, $P - Q$, $I - Q$, $J - Q$, $K - Q$ and $K - S$, which are listed in Table 2. Specifically, X denotes P, I, J, or K, and Y denotes Q or S. These instances share two common features. One is that X and Y are stub and transit AS node sets, respectively, and the other is that $X_Y (l) = X$, i.e., all the X nodes have low degrees and are included in the $X - Y$ component. Y_X is divided into high-degree node set $Y_X (h)$ and low-degree node set $Y_X (l)$, where the corresponding definitions are listed in Table 3. Because the five instances have similar properties, this section presents their combined analyses. Specifically, we take the $I - Q$ subgraph as an example to analyze the $X - Y$ component because the subgraph contains the largest number of edges in these instances.

First, we analyze the node properties of the $I - Q$ component, which contains three node sets: $Q_I (h)$, $Q_I (l)$, and $I_Q (l)$. Figs. 10(a)–(c) show that $\|Q_I (h)\|$, $\|Q_I (l)\|$, and $\|I_Q (l)\|$ linearly increases with t and that the degrees of $Q_I (l)$ and $I_Q (l)$ nodes follow constant power-law distributions. We note that the property of high-degrees of $Q_I (h)$ has been analyzed, as shown in Figs. 6(g) and (h).

Next, we analyze the edge properties of the $I - Q$ component, which contains edge sets $Q_I (h) - I_Q (l)$ and $Q_I (l) - I_Q (l)$. For the $Q_I (h) - I_Q (l)$ connections, Fig. 10(d) shows that the $P_i (k)$ versus k curve obeys a constant power-law distribution over a span of 15 years, where $P_i (k)$ is defined as the probability that a $Q_I (h) - I_Q (l)$ edge connects a k-degree $I_Q (l)$ node. In addition, Fig. 10(e) shows that for each k_2, joint degree distribution $P (k_1, k_2)$ of the $Q_I (l) - I_Q (l)$ connections is approximately uniform for $k_1 \in [1, 10]$, where k_1 and k_2 are the degrees of $Q_I (l)$ and $I_Q (l)$ nodes in the $I - Q$ component, respectively. Furthermore, Fig. 10(f) shows that the curve of the probability $P_S (k_2) = \sum_{k_1=1}^{d_{low}^{tra}} P (k_1, k_2)$ versus the degree k_2 of $I_Q (l)$ nodes obeys a constant power-law distribution, where $d_{low}^{tra} = 10$. In the summary shown in Figs. 10(e) and (f), we can deduce that $P (k_1, k_2) \approx P_S (k_2) / d_{low}^{tra}$ when $P_S (k_2)$ is known.

We omit the repetitive analyses of the $P - Q$, $J - Q$, $K - Q$ and $K - S$ components because they are similar to the $I - Q$ component. We note that in the $P - Q$ and $K - S$ components, the degrees of all the $P_Q (l)$ and $K_S (l)$ nodes are one; namely, the analysis shown in Figs. 10(c)–(f) can be neglected for the two components. According to the comparison presented in Sections IV.B and IV.C, we can find plenty of similarities between the Q_{bip} and $X - Y$ components. However, an obvious difference exists between the two components, namely, more distribution characteristics in the $X - Y$ component obey the power-law.

D. J_J (l) − J_J (l) EDGE SET

Please note that the definitions of all notations presented in this section are listed in Table 3. Fig. 4 shows that $J_J (l) - J_J (l)$ connects two $J_Q (l)$ nodes in the $J - Q$ com-
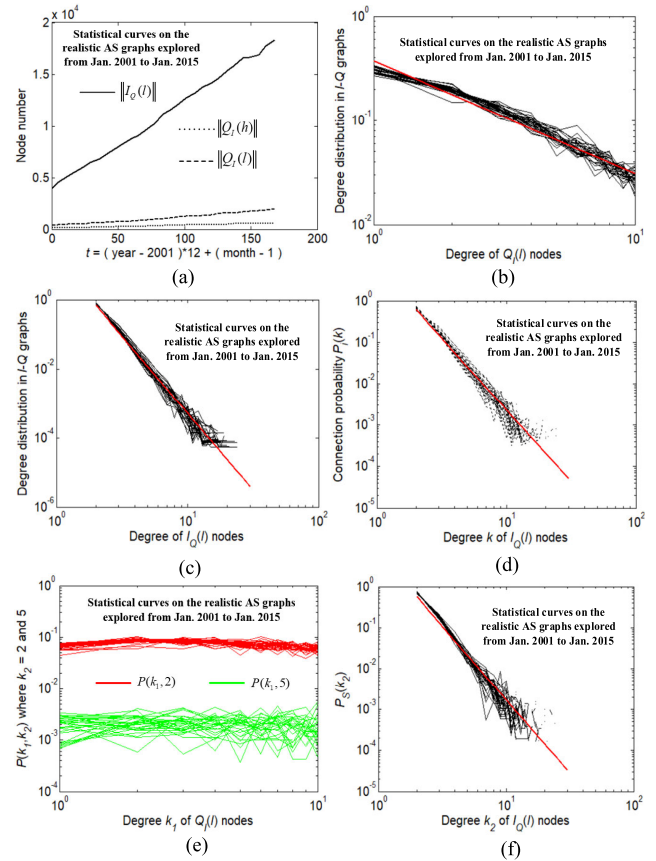

FIGURE 10. Evolving node and edge properties of the $I - Q$ component in the UC-LA dataset. (*a*) Node number *vs.* *t* (the number of months since January 2001). (*b*) Degree distribution of $Q_I (l)$ nodes. (*c*) Degree distribution of $I_Q (l)$ nodes. Note that the degrees of $I_Q (l)$ nodes are not less than 2 and not more than 30. (*d*) Probability $P_i (k)$ *vs.* degree k of $I_Q (l)$ nodes. (*e*) Joint degree distribution $P (k_1, k_2)$ of the $Q_I (l) - I_Q (l)$ connections (where $k_2 = 2$ and 5) *vs.* degree k_1 of $Q_I (l)$ nodes. (*f*) $P_S (k_2)$ *vs.* degree k_2 of $I_Q (l)$ nodes, where $P_S (k_2)$ is defin-ed as $\sum_{k_1} P (k_1, k_2)$ and $P (k_1, k_2)$ is the joint degree distribution in (e).

ponent. Section III confirms that $J_J (l) = J_Q (l) = J$ and $d_{J−J} (v) = 1$ for each node v in the $J - J$ subgraph, which mean that each $J_Q (l)$ node is connected to only one $J_J (l) - J_J (l)$ edge, i.e., edge number $\|J_J (l) - J_J (l)\|$, is strictly equal to half of the number of J nodes, and node number $\|J\|$ is even.

To study the $J_J (l) - J_J (l)$ connection feature, we sort all possible pairs of degrees $d_{J−Q} (v)$ where $v \in J_Q (l)$ using Algorithm 1 whose design is presented in Section IV.B and create degree pair $\left(d_{J−Q} (v_1), d_{J−Q} (v_2) \right)$ for each $J_J (l) - J_J (l)$ edge that connects nodes v_1 and v_2, where $v_1, v_2 \in J_Q (l)$ and $d_{J−Q} (v_1) \leq d_{J−Q} (v_2)$. We note that $1 \leq d_{J−Q} (v) < 30$ for each node $v \in J_Q (l)$. Fig. 11(a) shows that the joint degree distribution $P (k_1, k_2)$ versus the degree-pair rank of the $J_J (l) − J_J (l)$ connections is constant in the UCLA dataset that spans 15 years and is similar to the power-law.

E. S − S SUBGRAPH

Fig. 4 shows that the S_S node set is decomposed into $S_S (h) = S_S \cap S_K (h)$ and $S_S (l) = S_S \cap S_K (l)$, where $S_K (h)$ and

$S_K(l)$ are high- and low-degree S_K node sets in the $K - S$ component, respectively. Thus, the $S - S$ subgraph consists of three types of edge sets, namely, $S_S(h) - S_S(h)$, $S_S(h) - S_S(l)$, and $S_S(l) - S_S(l)$.

According to the analysis presented in Section III, $S_S \subseteq S_K = S$. Thus, degree pair $(d_{K-S}(v_1), d_{K-S}(v_2))$ can be used to characterize each $S_S(l) - S_S(l)$ edge that connects two $S_K(l)$ nodes v_1 and v_2, where $d_{K-S}(v_1) \leq d_{K-S}(v_2)$. We sort all the degree pairs using Algorithm 1, and Fig. 11(b) shows that the joint degree distribution $P(k_1, k_2)$ versus degree-pair rank of the $S_S(l) - S_S(l)$ connections is constant in the UCLA dataset over the span of 15 years and is similar to the power-law.

Moreover, to study the $S_S(h) - S_S(l)$ connections, we define $P_{sl}(k)$ as the probability that an $S_S(h) - S_S(l)$ edge connects a k-degree $S_S(l)$ node, where k is the degree of the $S_S(l)$ node in the $K - S$ component. Fig. 11(c) shows that the $P_{sl}(k)$ versus k curve obeys a constant power-law distribution in the UCLA dataset that spans 15 years.

Finally, we analyze the connection feature associated with the $S_S(h)$ nodes. Fig. 11(d) shows that at any exploration time t, we can establish the following relationships:

$$\begin{cases} \|S_S(l) - S_S(l)\| = 1.6314 (\|S_K(l)\| - 576) + 103 \\ \|S_S(h) - S_S(l)\| = 0.8328 (\|S_K(l)\| - 576) - 153 \\ \|S_K(h)\| = 0.0239 (\|S_K(l)\| - 576) + 9 \\ \|S_S(h) - S_S(h)\| = 1.0211 (\|S_K(h)\| - 9) - 5. \end{cases}$$
(9)

Eq. (9) indicates that the $S_S(h) - S_S(h)$ edges, where $S_S(h) = S_S \cap S_K(h)$, are very sparse compared with node number $\|S_K(h)\|$. In addition, the difference between the degrees of the two $S_K(h)$ nodes in the $K - S$ component is not obvious because $S_K(h)$ is far away from the core. Thus, we assume that the $S_S(h) - S_S(h)$ edges uniformly connect two $S_K(h)$ nodes at random, and the probability that an $S_S(h) - S_S(l)$ edge connects a k-degree $S_K(h)$ node is approximately uniform.

F. S − Q COMPONENT
The $S - Q$ component shown in Fig. 4 has two transit AS node sets S_Q and Q_S, which can be divided into $S_Q(h)$, $S_Q(l)$, $Q_S(h)$, and $Q_S(l)$. The division method is listed in Table 3. In addition, the connections in the component are composed of $Q_S(h) - S_Q(h)$, $Q_S(h) - S_Q(l)$, $Q_S(l) - S_Q(h)$, and $Q_S(l) - S_Q(l)$. According to Section IV.C, if all the $S_Q(h)$ nodes are removed from the $S - Q$ component, $S - Q$ can be viewed as an instance of the $X - Y$ component, as defined in Section IV.C. Thus, we reset low-degree threshold of the $S - Q$ component $d_{low}^{S-Q} = 30$ to reduce the number of $S_Q(h)$ nodes. Fig. 12(a) shows that the node and edge numbers in the $S - Q$ component linearly increase except for $\|Q_S(h) - S_Q(h)\|$ and $\|Q_S(l) - S_Q(h)\|$, which quadratically increase. In addition, Fig. 12(a) shows that at any exploration time t, we can establish the following relationships:
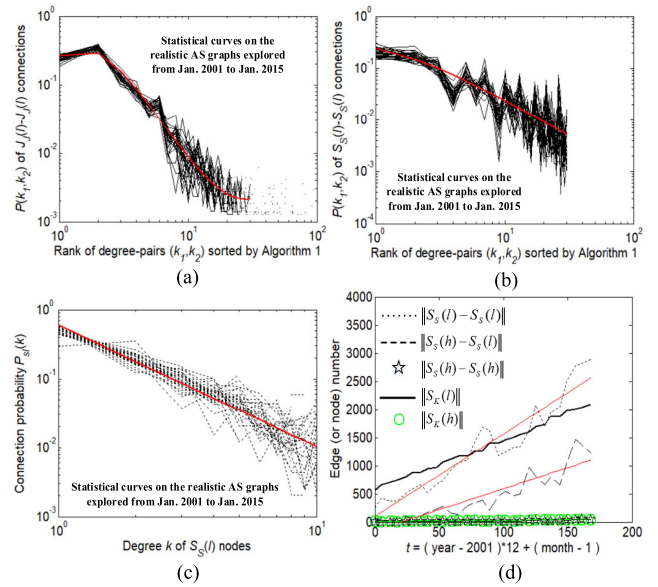

FIGURE 11. Evolving edge properties of the $J_J(l) - J_J(l)$ connections and the S-S subgraph in the UCLA dataset. (a) Joint degree distribution $P(k_1, k_2)$ of the $J_J(l) - J_J(l)$ connections vs. rank of degree-pairs (k_1, k_2) sorted by Algorithm 1, where k_1 and k_2 are the degrees of J nodes in the $J − Q$ component. (b) Joint degree distribution $P(k_1, k_2)$ of the $S_S(l) - S_S(l)$ connections vs. rank of deg-ree-pairs (k_1, k_2) sorted by Algorithm 1, where k_1 and k_2 are the degrees of S nodes in the $K − S$ component. (c) Probability $P_{sl}(k)$ vs. degree k of $S_S(l)$ no-des, where k is the degree of S nodes in the $K − S$ component. (d) Edge (node) number vs. t (the number of months since January 2001), where $\|S_K(h)\|$ is the number of $S_K(h)$ nodes, $\|S_K(l)\|$ is the number of $S_K(l)$ nodes, $\|S_S(l) - S_S(l)\|$ is the number of edges connecting two $S_S(l)$ nodes, $\|S_S(h) - S_S(l)\|$ is the number of edges connecting $S_S(h)$ and $S_S(l)$ nodes, and $\|S_S(h) - S_S(h)\|$ is the number of edges connecting two $S_S(h)$ nodes.

$$\begin{cases} \|Q_S(h) - S_Q(l)\| = 39.58 (\|Q_S(h)\| - 8.62) + 490 \\ \|Q_S(l) - S_Q(l)\| = 1.14 (\|Q_S(l)\| - 254.5) + 1315 \\ \|S_Q(h)\| = 1.497 (\|Q_S(h)\| - 8.622) - 23 \\ \|S_Q(l)\| = 0.679 (\|Q_S(l)\| - 254.49) + 533. \end{cases}$$
(10)

Despite the use of a larger low-degree threshold, Figs. 12(b) and (c) show that the degree distributions of the $Q_S(l)$ and $S_Q(l)$ nodes in the $S - Q$ component remain constant and tend toward the power-law. Owing to the similarity between the $S - Q$ and the $X - Y$ (see Section IV.C) components, such as the high-degree properties shown in Fig. 12(d), the properties of the edges that connect the low-degree nodes, and that of the edges that connect the high- and low-degree nodes, this section only presents the analyses of the difference, i.e., the property of the $Q_S(h) - S_Q(h)$ edges that connects the high- and high-degree nodes.

In Algorithm 2, we introduce the given n nodes into c-ordered categories. Specifically, line 3 ensures that the nodes with higher degrees are introduced in the categories at the front of the list. When $n < c$, lines 4 and 5 separately introduce one node in the first n category and let the last $c - n$

Algorithm 2 Node Classification and Sorting

1: **Input:** High-degree nodes $1, 2, \cdots, n$ and their degrees
$\quad d_1, d_2, \cdots, d_n$, category number c.

2: **Output:** Node category list $L_1, L_2, \cdots, L_c$ sorted by a
$\quad$ certain order.

3: Initialize $L_j \leftarrow \emptyset$ for $j = 1, 2, \cdots, c$, and sort the
$\quad$ high-degree nodes by decreasing order of their degrees.
$\quad$ Without loss of generality, we assume $d_1 \geq d_2 \geq \cdots$
$\quad \geq d_n$.

4: **If** $n < c$ **do**

5: $\quad$ Update $L_j \leftarrow \{j\}$ for $j = 1, 2, \cdots, n$.

6: **Elseif** $n \geq c$ **do**

7: $\quad$ Derive $r = n/c$, and initialize $L_1 \leftarrow \{1, 2, \cdots, round\,(r)\}$,
$\quad$ where $round\,(r)$ rounds r to the nearest integer.

8: $\quad$ **For** $j = 2 : 1 : c$ **do**

9: $\quad\quad$ Update $L_j \leftarrow \{max\,(L_{j-1}) + 1, max\,(L_{j-1})$
$\quad\quad +2, \cdots, round\,(j \times r)\}$,
$\quad\quad$ where $max\,(X)$ returns the largest element in X.

10: $\quad$ **End for**

11: **End if**

categories be empty sets. Otherwise, lines 6–10 introduce the n nodes in the c categories and ensure that the cardinality of each category is approximately equal.

Using Algorithm 2, we classify node sets $Q_S\,(h)$ and $S_Q\,(h)$ into c-sorted categories $L_1^{Q_S}, L_2^{Q_S}, \cdots, L_c^{Q_S}$ and $L_1^{S_Q}, L_2^{S_Q}, \cdots, L_c^{S_Q}$, respectively. In addition, we represent the $Q_S\,(h) - S_Q\,(h)$ connections using joint rank distribution $P\,(r_1, r_2)$, which is defined as follows:

$$P\,(r_1, r_2) = m\,(r_1, r_2)\big/m, \qquad (11)$$

where m denotes the total number of $Q_S\,(h) - S_Q\,(h)$ edges, $m\,(r_1, r_2)$ is the number of edges that connects a node in $L_{r_1}^{Q_S}$ and another node in $L_{r_2}^{S_Q}$, and r_1 and r_2 are the ranks of $L_{r_1}^{Q_S}$ and $L_{r_2}^{S_Q}$, respectively, in the sorted node categories.

Fig. 12(e) shows that for any given r_2, joint rank distribution $P\,(r_1, r_2)$ tends to be stable and can be characterized using a quintic fitting curve. In addition, Fig. 12(f) shows that the $S_Q\,(h)$ nodes with higher r_2 tend to be connected to more high-degree $Q_S\,(h)$ nodes with lower r_1.

G. (DOTTED) NODE MAPPING COMPONENT

Figs. 3 and 4 show that except for the solid-edge sets, our structural model contains dotted-edge sets, which establish injective or bijection relationships between two node sets U and W. Specifically, $u \in U$ and $w \in W$ are in the same node if the dotted edge maps u to w. Thus, we can derive that $\|U\| = e \leq \|W\|$, where e is the number of dotted edges. Figs. 3 and 4 show eight node-mapping components associated with the dotted edges, which implement the merging of the solid components, as presented in the analysis in Sections IV.A–IV.F. This section presents the use of a joint rank distribution to represent the $U - W$ connections.
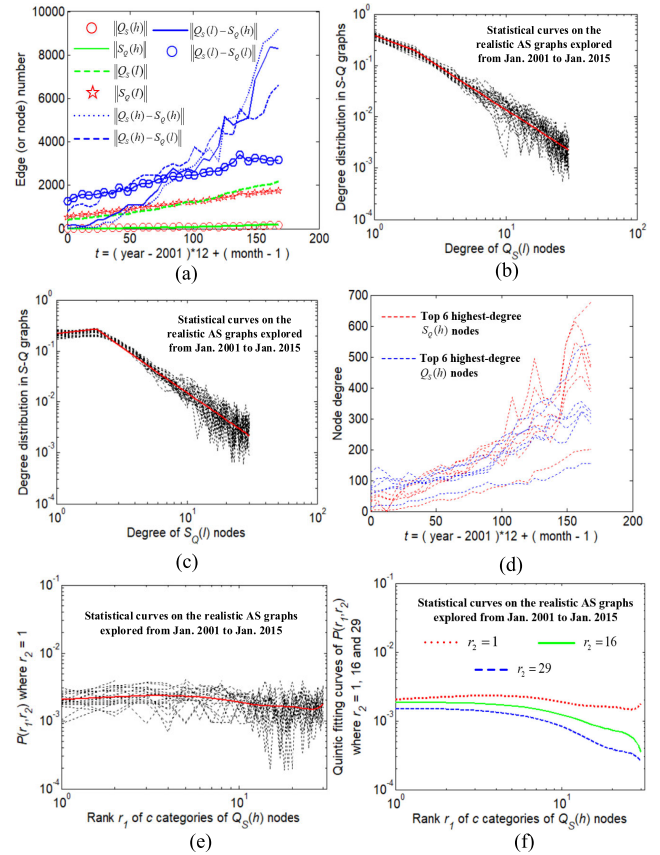

FIGURE 12. Evolving node and edge properties of the $S - Q$ component in the U-CLA dataset. *(a)* Edge (or node) number *vs. t* (the number of months since Ja-nuary 2001), where $\|Q_S\,(h)\|$, $\|Q_S\,(l)\|$, $\|S_Q\,(h)\|$ and $\|S_Q\,(l)\|$ correspond to the number of nodes, and $\|Q_S\,(h) - S_Q\,(h)\|$, $\|Q_S\,(h) - S_Q\,(l)\|$, $\|Q_S\,(l) - S_Q\,(h)\|$ and $\|Q_S\,(l) - S_Q\,(l)\|$ correspond to the number of edges. *(b)* Degree distribution of $Q_S\,(l)$ nodes. *(c)* Degree distribution of $S_Q\,(l)$ nodes. *(d)* The ith highest-degrees of $S_Q\,(h)$ and $Q_S\,(h)$ nodes *vs. t* with $1 \leq i \leq 6$, which approximately obey linear relations that can be represented by tuples consisting of slope m_i and intercept b_i. *(e)* Joint rank distribution $P\,(r_1, r_2)$ of the $Q_S\,(h) - S_Q\,(h)$ co-nnections with $r_2 = 1$ *vs.* rank r_1 of c categories of $Q_S\,(h)$ nodes where $c = 30$. *(f)* Three quintic fitting curves of the joint rank distribution $P\,(r_1, r_2)$ of the $Q_S\,(h) - S_Q\,(h)$ connections with $r_2 = 1, 16$ and 29.

For each dotted (node mapping) component $U - W$, we let $G\,(U)$ and $G\,(W)$ be two solid components that include U and W, respectively, and define $d_{G(U)}\,(u)$ and $d_{G(W)}\,(w)$ as the degrees of node $u \in U$ in $G\,(U)$ and node $w \in W$ in $G\,(W)$, respectively. Then, we can decompose U and W into high-degree and non-high-degree node sets using $d_{G(U)}\,(u)$ and $d_{G(W)}\,(w)$, respectively, as listed in Table 4.

In Algorithm 3, we introduce all the nodes in set U, which belong to dotted component $U - W$, into $c + d$ ordered categories, where c is an input number and d is the degree threshold, as listed in Table 4. Specifically, line 3 decomposes U into high-degree node set U_H and non-high-degree node set U_N based on Table 4. Line 4 calls Algorithm 2 to divide all nodes U_H into top c-ordered categories. In addition, for any $k \in \{1, 2, \cdots, d\}$, lines 5–10 introduce all nodes u in U_N with $d_{G(U)}\,(u) = k$ into the $(c + d - k + 1)^{th}$ category,

TABLE 4. Node classifications of different $U - W$ dotted components in Figs. 3 and 4 (except for $K_S(l)$)

U W	$G(U)$ $G(W)$	High-degree node set	Non-high-degree node set	Degree threshold
$Q_{cor}^U(h, m, l)$	Q_{cor}	$Q_{cor}^U(h)$	$Q_{cor}^U(m, l)$	$d_{hig}^{cor} - 1$
$Q_{bip}^U(h, l)$	Q_{bip}	$Q_{bip}^U(h)$	$Q_{bip}^U(l)$	d_{low}^{tra}
$Q_{cor}^U(h, m, l), Q_{bip}^B(l)$	$Q - Q$	Q_{bip}^U	$Q_{bip}^B(l)$	d_{low}^{tra}
$Q_P(h), Q_P(l)$	$P - Q$	$Q_P(h)$	$Q_P(l)$	d_{low}^{tra}
$Q_I(h), Q_I(l)$	$I - Q$	$Q_I(h)$	$Q_I(l)$	d_{low}^{tra}
$Q_J(h), Q_J(l)$	$J - Q$	$Q_J(h)$	$Q_J(l)$	d_{low}^{tra}
$S_Q(h), S_Q(l)$	$S - Q$	$S_Q(h)$	$S_Q(l)$	d_{low}^{S-Q}
$S_K(h), S_K(l)$	$K - S$	$S_K(h)$	$S_K(l)$	d_{low}^{tra}
$K_Q(l)$	$K - Q$	$\emptyset$	$K_Q(l)$	$d = 30$

$$Q_{cor}^U(m, l) = Q_{cor}^U(m) \cup Q_{cor}^U(l),$$
$$Q_{bip}^B(l) = \{v | v \in Q \wedge d_Q(v) \le d_{low}^{tra}\}, Q_{bip}^U = Q/Q_{bip}^B(l),$$
$$d_{hig}^{cor} - 1 = 99, d_{low}^{tra} = 10, d_{low}^{S-Q} = 30.$$

where $d_{G(U)}(u)$ is the degree of node u in solid component $G(U)$, as listed in Table 4.

Algorithm 3 Node Classification and Sorting in Dotted Components

1: **Input:** Node set U, solid component $G(U)$, node degree $d_{G(U)}(u)$ for $\forall u \in U$ and category number c.
2: **Output:** Node category list L sorted by a certain order.
3: Decompose U into a high-degree node set U_H and a non-high-degree node set U_N based on Table 4.
4: Derive the node category list $L_1, L_2, \cdots, L_c$ of U_H using Algorithm 2.
5: Let d be the degree threshold shown in Table 4 that is the maximum degree of the non-high-degree nodes. Initialize $k \leftarrow d$ and $i \leftarrow 1$.
6: **While** $k \ge 1$ **do**
7: Derive a node set $S = \{u | u \in U_N \wedge d_{G(U)}(u) = k\}$, and let the $(c + i)^{th}$ node category L_{c+i} be S.
8: Update $k \leftarrow k - 1$ and $i \leftarrow i + 1$.
9: **End while**
10: Update $L \leftarrow L_1, L_2, \cdots, L_c, L_{c+1}, \cdots, L_{c+d}$.

Thus, using Algorithm 3, we can classify node sets U and W in dotted component $U - W$ into two sorted category lists, namely, $UL_1, UL_2, \cdots, UL_{cu}$ and $WL_1, WL_2, \cdots, WL_{cw}$.

Then, we define the joint rank distribution of the $U - W$ edges as $P(r_1, r_2) = m(r_1, r_2)/m$, where $m(r_1, r_2)$ is the number of dotted edges that maps the UL_{r_2} nodes to the WL_{r_1} nodes and m is the total number of $U - W$ edges.

By considering the $Q_P(h), Q_P(l) - Q_{cor}^U(h, m, l), Q_{bip}^B(l)$ edges as an example, Fig. 13(a) shows that for any given r_2, joint rank distribution $P(r_1, r_2)$ tends to be stable and can be represented using a quintic fitting curve. Fig. 13(b) shows that the $Q_P(h), Q_P(l)$ nodes with higher r_2 tend to be mapped to the $Q_{cor}^U(h, m, l), Q_{bip}^B(l)$ nodes with higher r_1.

We note that $K_S(l)$ in the $K_Q(l) - K_S(l)$ dotted component cannot be classified using Algorithm 3 because all the degrees of the $K_S(l)$ nodes in the $K - S$ solid component are one. To classify $K_S(l)$ into the c categories, we sort
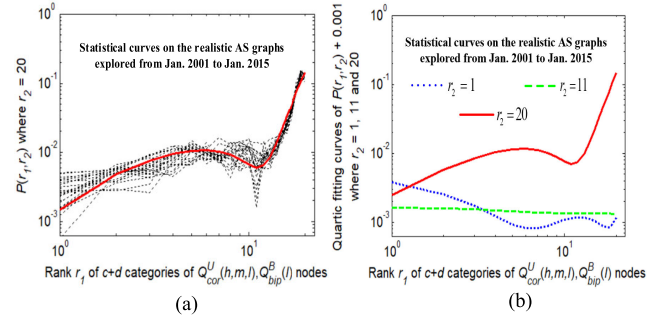

FIGURE 13. Joint rank distribution of the $Q_P(h)$, $Q_P(l) - Q_{cor}^U(h, m, l)$, $Q_{bip}^B(l)$ dotted component. (a) Joint rank distribution $P(r_1, r_2)$ with $r_2 = 20$ vs. rank r_1 of $c + d$ categories of $Q_{cor}^U(h, m, l)$, $Q_{bip}^B(l)$ nodes where $c = 10$ and $d = 10$. (b) Three quartic fitting curves of the joint rank distribution $P(r_1, r_2) + 0.001$ with $r_2 = 1$, 11 and 20, $P(r_1, r_2) + 0.001$ avoids $P(r_1, r_2)$ close to 0.

all the $S_K(h), S_K(l)$ nodes as $S_1, S_2, \cdots, S_n$ according to a decreasing order of $d_{K-S}(v)$, which is the degree of node v in $K - S$, and classify them into c categories:

$$\{S_1\}, \cdots, \{S_{c-1}\}, \{S_c, \cdots, S_n\}.$$

Because each $K_S(l)$ node is connected to only one $S_i (i \in \{1, 2, \cdots, n\})$ node in $K - S$, the S_i node in the c categories can be replaced by all the $K_S(l)$ nodes that are connected to S_i, i.e., $K_S(l)$ is classified into the c sorted node categories using the classification of the $S_K(h), S_K(l)$ nodes.

V. INTERNET-TOPOLOGY GENERATOR SICPS

Our structural model decomposes the Internet topology into many bipartite-graph components, which remain stable in terms of statistical features over a span of 15 years except for the top five highest degrees of transit nodes. However, the unstable factor shows the trend of the Internet, i.e., increasingly more transit nodes tend to be connected to a few highest degree core nodes. This section describes the design of Internet-topology generator SICPS based on our structural model and its evolutionary stability. Specifically, SICPS first generates eight solid components using the stability of the node and the edge properties analyzed in Sections IV.A–IV.F. Then, it realizes the merging of the eight solid components using the dotted-edge properties analyzed in Section IV.G.

A. Q_{cor} COMPONENT GENERATION

According to the analysis presented in Section IV.A, we design Algorithm 4 to generate the Q_{cor} component, which consists of two steps, namely, node and edge generations.

In Algorithm 4, lines 3–6 create nodes and their predefined degrees. We note that the free degree of a node is equal to its predefined degree minus the number of edges that have been connected to the node. Lines 7–9 use the density curve to create $Q_{cor}^U(h) - Q_{cor}^U(h)$ edges. Lines 10–19 use the preference attachment distribution to create $Q_{cor}^U(h) - Q_{cor}^U(m, l)$ edges, and lines 20–29 use the uniform

Algorithm 4 Generation of the Q_{cor} Component

1: **Input:** Exploration time t. Node properties: linear fitting lines of $\left\| Q_{cor}^U (h) \right\|$ and $\left\| Q_{cor}^U (l) \right\|$, quadratic fitting curve of $\left\| Q_{cor}^U (m) \right\|$, degree distribution $P_l (k)$ of $Q_{cor}^U (l)$ nodes where $k \in \{1, 2, \cdots, 10\}$, degree distribution $P_m (k)$ of $Q_{cor}^U (m)$ nodes where $k \in \{11, 12, \cdots, 99\}$, fitting curves $f_1 (t), f_2 (t), \cdots, f_5 (t)$ of top 5 highest-degrees of $Q_{cor}^U (h)$ nodes, fitting curves $S (i)$ and $I (i)$ associated with the slope and intercept of the linear relation of the ith highest-degree of $Q_{cor}^U (h)$ nodes respectively.
Edge properties: density curve $D (i) \left(i = 1, 2, \cdots, \left\| Q_{cor}^U (h) \right\| \right)$, preference attachment distribution $P_c (k)$ defined as the probability that a $Q_{cor}^U (h) - Q_{cor}^U (m, l)$ edge connects a k-degree $Q_{cor}^U (m, l)$ node.

2: **Output:** The Q_{cor} component.

3: Derive the high-degree node number $h = \left\| Q_{cor}^U (h) \right\|$, the middle-degree node number $m = \left\| Q_{cor}^U (m) \right\|$ and the low-degree node number $l = \left\| Q_{cor}^U (l) \right\|$ at the exploration time t.

4: Generate h high-degree nodes $1, 2, \cdots, h$ and assign the degree d_i to the node i ($i = 1, 2, \cdots, h$), where

$$
d_i = \begin{cases} round \, (f_i (t)), & 1 \leq i \leq 5 \\ round \, (max \, (S (i) \cdot t + I (i), 100)), & 6 \leq i \leq h. \end{cases}
$$
(12)

5: Generate m middle-degree nodes $h + 1, h + 2, \cdots, h + m$ and assign the degree $d_{h+i} = k \in \{11, 12, \cdots, 99\}$ to the node $h + i$ ($i = 1, 2, \cdots, m$) using the distribution $P_m (k)$. Assume $d_{h+i} \geq d_{h+i+1}$ for $1 \leq i \leq m - 1$.

6: Generate l low-degree nodes $h + m + 1, h + m + 2, \cdots, h + m + l$ and assign the degree $d_{h+m+i} = k \in \{1, 2, \cdots, 10\}$ to the node $h + m + i$ ($i = 1, 2, \cdots, l$) using the degree distribution $P_l (k)$. Assume $d_{h+m+i} \geq d_{h+m+i+1}$ where $i = 1, 2, \cdots, l - 1$.

7: **For** $i = 2 : 1 : h$ **do**

8: Derive $x = i (i - 1) D (i) / 2 - (i - 1) (i - 2) D (i - 1) / 2$. Uniformly at random, extract x nodes from $\{1, 2, \cdots, i - 1\}$ that still have at least one free degree, and connect node i to the x nodes.

9: **End for**

10: Define $z (i) = i$ if $1 \leq i \leq 5$ and $z (i) = h - i + 6$ if $6 \leq i \leq h$.

11: **For** $i = 1 : 1 : h$ **do**

12: Let x be the number of free degrees of node $z (i)$, and classify the x free degrees into different sets R_k ($k = 1, 2, \cdots, 99$) according to the preference attachment distribution $P_c (k)$.

13: **For** $k = 1 : 1 : 99$ **do**

14: Derive $y = \| R_k \|$, obtain set Y of all the k-degree nodes in $\{h + 1, h + 2, \cdots, h + m + l\}$ that still have at least one free degree, sort the nodes in Y by decreasing order of $fd (v)$ defined as the number of free degrees of node v, extract the first

15: $a_k = min \, (y, \| Y \|)$ nodes in Y, and connect node $z (i)$ to the a_k nodes.

15: **End for**

16: **End for**

17: **For** $i = 1 : 1 : h$ **do**

18: Let x be the number of free degrees of node i and Y be the subset of $\{h + 1, h + 2, \cdots, h + m + l\}$ that still have at least one free degree and have not been attached to node i. Then, extract the top $a_k = min \, (x, \| Y \|)$ nodes in Y with highest $fd (v)$ values and connect node i to the a_k nodes.

19: **End for**

20: **For** $i = 1 : 1 : l$ **do**

21: Let x be the number of free degrees of node $h + m + i$.

22: **While** $x \geq 1$ **do**

23: Uniformly at random select a $k \in \{11, 12, \cdots, 99\}$. Extract the k-degree node subset Y in $\{h + 1, h + 2, \cdots, h + m\}$ that still have at least one free degree, connect node $h + m + i$ to a randomly-selected node in Y that have not been connected by node $h + m + i$ and update $x \leftarrow x - 1$.

24: **End while**

25: **End for**

26: Let x be the total number of free degrees of all the nodes $h + 1, h + 2, \cdots, h + m$, and initialize $t_c \leftarrow 0$.

27: **While** $x \geq 2 \wedge t_c < 500$ **do**

28: Uniformly at random select $k_1, k_2 \in \{11, 12, \cdots, 99\}$. Extract a k_1-degree node v_1 and a k_2-degree node v_2 in $\{h + 1, h + 2, \cdots, h + m\}$ where v_1 and v_2 are not connected and have maximum $min \, (fd (v_1), fd (v_2))$. If the two nodes can be extracted, connect the two nodes and update $x \leftarrow x - 2$, otherwise update $t_c \leftarrow t_c + 1$.

29: **End while**

---

distributions to create $Q_{cor}^U (m, l) - Q_{cor}^U (m, l)$ edges. We note that line 10 defines another sorting method of high-degree nodes because the experimental results show that this method can minimize the number of free degrees while maintaining the $P_c (k)$ distribution and avoiding multiple edges. Algorithm 4 avoids multiple edges by prejudging whether the two nodes are connected. Because all the nodes and their degrees are predefined, we can apply $1 \times e$ array L and two $1 \times n$ pointers P_1 and P_2 to describe all the edges in the component, where e is the sum of all the degrees and n is the total number of nodes. We initialize $P_1 (i) = P_2 (i) = \sum_{j=1}^{i-1} d_j + 1$ for arbitrary node $i \in \{1, 2, \cdots, n\}$, where d_i is the degree of node i, and update $P_2 (i) \leftarrow P_2 (i) + 1$, $L (P_2 (i)) = j$, $P_2 (j) \leftarrow P_2 (j) + 1$, and $L (P_2 (j)) = i$ for each newly added edge connecting nodes i and j. Then, set $L ([P_1 (i) : 1 : P_2 (i)])$ always stores all the nodes that have been connected to arbitrary node i. We note that in line 4 of Algorithm 4, the symbol $round \, (x)$ rounds x to the nearest integer.

Algorithm 5 Generation of the $X - Y$ Component

1: **Input:** Exploration time t. Node properties: linear fitting lines of $\|Y_X(h)\|$, $\|Y_X(l)\|$ and $\|X_Y(l)\|$, degree distribution $P_Y(k)$ of $Y_X(l)$ nodes where $k \in \{1, 2, \cdots, 10\}$, degree distribution $P_X(k)$ of $X_Y(l)$ nodes where $k \in \{1, 2, \cdots, 30\}$, fitting curves $f_1(t), f_2(t), \cdots, f_5(t)$ of top 5 highest-degrees of $Y_X(h)$ nodes, fitting curves $S(i)$ and $I(i)$ associated with the slope and intercept of the linear relation of the ith highest-degree of $Y_X(h)$ nodes respectively. Edge properties: two preference attachment distributions, $P_c(k)$ defined as the probability that a $Y_X(h) - X_Y(l)$ edge connects a k-degree $X_Y(l)$ node, $P_S(k_2)$ defined as the probability that a $Y_X(l) - X_Y(l)$ edge connects a k_2-degree $X_Y(l)$ node. Note that $P_X(k) = 0$ when $k = 11, 12, \cdots, 30$ for the Q_{bip} component since the degree of $Q_{bip}^B(l)$ nodes in the component is not more than $d_{low}^{tra} = 10$. Specifically, 30 is the upper bound of the degrees of all $X_Y(l)$ nodes.
2: **Output:** The $X - Y$ component.
3: Derive the three node numbers $h = \|Y_X(h)\|$, $y = \|Y_X(l)\|$ and $x = \|X_Y(l)\|$ at the exploration time t.
4: Generate h high-degree nodes $1, 2, \cdots, h$ and assign the degree d_i to the node i ($i = 1, 2, \cdots, h$), where

$$d_i = \begin{cases} round\,(f_i(t)), & 1 \le i \le 5 \\ round\,(max\,(S(i) \cdot t + I(i), 11)), & 6 \le i \le h. \end{cases}$$
(13)

where $round(x)$ rounds x to the nearest integer.
Note that max and min return the maximum and minimum values respectively.
5: Generate y low-degree nodes $h + 1, h + 2, \cdots, h + y$ and assign the degree $d_{h+i} = k \in \{1, 2, \cdots, 10\}$ to the node $h + i$ ($i = 1, 2, \cdots, y$) using the degree distribution $P_Y(k)$. Assume $d_{h+i} \ge d_{h+i+1}$ where $i = 1, 2, \cdots, y - 1$.
6: Generate x low-degree nodes $h + y + 1, h + y + 2, \cdots, h + y + x$ and assign the degree $d_{h+y+i} = k \in \{1, 2, \cdots, 30\}$ to the node $h + y + i$ ($i = 1, 2, \cdots, x$) using the degree distribution $P_X(k)$. Assume $d_{h+y+i} \ge d_{h+y+i+1}$ where $i = 1, 2, \cdots, x - 1$.
7: Derive the joint degree distribution $P(k_1, k_2) = P_S(k_2) / 10$, which is the probability that a $Y_X(l) - X_Y(l)$ edge connects a k_1-degree $Y_X(l)$ node and a k_2-degree $X_Y(l)$ node.
8: Derive e_l which is the total number of $Y_X(l) - X_Y(l)$ edges, namely the sum of $d_{h+1}, d_{h+2}, \cdots, d_{h+y}$.
9: **For** $k_1 = 1 : 1 : 10$ **do**
10:   **For** $k_2 = 1 : 1 : 30$ **do**
11:     Derive $z = round\,(e_l \times P(k_1, k_2))$, and initialize $t_c \leftarrow 0$.
12:     **While** $z \ge 1 \wedge t_c = 0$ **do**
13:       Extract a k_1-degree node v_1 in $\{h + 1, h + 2, \cdots, h + y\}$ and a k_2-degree node v_2 in $\{h + y + 1, h + y + 2, \cdots, h + y + x\}$ where v_1 and v_2 are not connected and have maximum $min\,(fd\,(v_1), fd\,(v_2))$. If the two nodes can be extracted, connect the two nodes and update

      $z \leftarrow z - 1$, otherwise update $t_c \leftarrow 1$. Note that $fd\,(v)$ is defined as the number of free degrees of node v.
14:     **End while**
15:   **End for**
16: **End for**
17: Define $z(i) = i$ if $1 \le i \le 5$ and $z(i) = h - i + 6$ if $6 \le i \le h$.
18: **For** $i = 1 : 1 : h$ **do**
19:   Let z_f be the number of free degrees of node $z(i)$, and classify the z_f free degrees into different sets R_k ($k = 1, 2, \cdots, 30$) according to the preference attachment distribution $P_c(k)$.
20:   **For** $k = 1 : 1 : 30$ **do**
21:     Derive $z_k = \|R_k\|$, obtain set L of all the k-degree nodes in $\{h + y + 1, h + y + 2, \cdots, h + y + x\}$ that still have at least one free degree, sort the nodes in L by decreasing order of $fd\,(v)$ values, extract the first $a_k = min\,(z_k, \|L\|)$ nodes in L, and connect node $z(i)$ to the a_k nodes.
22:   **End for**
23: **End for**
24: **For** $i = 1 : 1 : h$ **do**
25:   Let z_f be the number of free degrees of node i and L be the subset of $\{h + y + 1, h + y + 2, \cdots, h + y + x\}$ that still have at least one free degree and have not been attached to node i. Then, extract the top $a_k = min\,(z_k, \|L\|)$ nodes in L with highest $fd\,(v)$ values and connect node i to the a_k nodes.
26: **End for**

B. $X - Y$ COMPONENT GENERATION

According to the analysis presented in Sections IV.B and IV.C, the Q_{bip} component can be divided into $X - Y$ categories. Hence, Algorithm 5 is designed to generate the Q_{bip} component and other five components, namely, $P - Q$, $I - Q$, $J - Q$, $K - Q$, and $K - S$, as listed in Table 2.

In Algorithm 5, lines 3–6 create nodes and their predefined degrees. Lines 7–16 use joint degree distribution $P(k_1, k_2)$ that is derived by $P_S(k_2)$ to create $Y_X(l) - X_Y(l)$ edges, and lines 17–6 use preference attachment distribution $P_c(k)$ to create $Y_X(h) - X_Y(l)$ edges. We note that lines 24–26 are used to fill the remaining free degrees of the $Y_X(h)$ nodes.

When Algorithm 5 is applied to create the Q_{bip} component, we should map $Q_{bip}^U(h)$, $Q_{bip}^U(l)$, and $Q_{bip}^B(l)$ to $Y_X(h)$, $Y_X(l)$, and $X_Y(l)$, respectively. After the generation of the aforementioned six components, namely, $Q_{bip}, P - Q, I - Q$, $J - Q, K - Q$, and $K - S$. According to Sections IV.B, IV.D, and IV.E, the $Q_{red}^B(l) - Q_{red}^B(l)$ connections should be added to the Q_{bip} component, the $J_J(l) - J_J(l)$ connections should be added to the $J - Q$ component, and the three types of connections, namely, $S_S(h) - S_S(h)$, $S_S(h) - S_S(l)$, and $S_S(l) - S_S(l)$, should be added to the $K - S$ component.

Figs. 9(d) and 11(a) and (b) show that the joint degree distributions of the $Q_{red}^B(l) - Q_{red}^B(l)$, $J_J(l) - J_J(l)$, and

$S_S(l) - S_S(l)$ connections remain stable over a span of 15 years. In addition, Eqs. (8) and (9) indicate that the number of these connections can be calculated using the inputs of Algorithm 5, namely, the number of nodes. Thus, using the constant fitting curves of these distributions and the edge numbers, the three types of connections can be easily added to the corresponding components. Moreover, Fig. 11(c) shows that distribution $P_{sl}(k)$, which is defined as the probability that an $S_S(h) - S_S(l)$ edge connects a k-degree $S_S(l)$ node, also remains stable. According to the analysis presented in Section IV.E, the $S_S(h) - S_S(h)$ edges uniformly connect two $S_K(h)$ nodes at random, and the probability that an $S_S(h) - S_S(l)$ edge connects a k-degree $S_K(h)$ node is approximately uniform. Thus, using the above-mentioned stable distribution and edge numbers, the $S_S(h) - S_S(h)$ and $S_S(h) - S_S(l)$ connections can also be easily added to the $K - S$ component.

C. S − Q COMPONENT GENERATION

The $S - Q$ component is an extension of the $X - Y$ component, and the main difference between them is that the former has an additional type of connection, namely, the edges with two high-degree ends. Hence, the methods for predefining the degrees, those that connect the high- and low-degree nodes and those that connect the low-degree nodes in Algorithm 5 can also be reused to generate the $S - Q$ component, as indicated by lines 3–5 and 15–16 of Algorithm 6. We note that lines 6–14 of Algorithm 6 use the joint rank distribution to create the connection of the $Q_S(h) - S_Q(h)$ edges.

D. DOTTED COMPONENT GENERATION

Sections V.A–V.C present the creation of eight solid components of the Internet topology. To obtain a complete graph, we need to merge these components using the eight dotted connections shown in Figs. 3 and 4, which can be modeled by $U - W$ that maps each node $u \in U$ to only one node $w \in W$. The instances of U and W and solid components $G(U), G(W)$, including U and W, are listed in Table 4. Once the mapping from $u \in U$ to $w \in W$ is completed, $G(U)$ and $G(W)$ can be combined into one graph by merging u and w to a single node. The generation method of the $U - W$ node mapping connections is demonstrated in Algorithm 7.

VI. EXPERIMENTAL RESULTS

This section presents the comparison of the realistic AS graphs that are provided by the UCLA dataset [1], including the results obtained by our SICPS generator and those by four other generators, namely, Inet-3.0 [10], ORBIS [14], S-BITE [21], and SInetL [29], which are introduced in Section II.B. Inet-3.0 is a classical Internet-topology generator that considers both the hierarchical structure and degree power-law properties. ORBIS aims to obtain the $2K$ degree distribution, S-BITE captures the topological core structure, and SInetL extracts a subgraph from a given Internet topology while maintaining the normalized Laplacian spectral properties. We note that the comparison uses the graph properties discussed in Section II.A. First, a comparison of the largest

---

Algorithm 6 Generation of the $S - Q$ Component

1: **Input:** Exploration time t. Node properties: linear fitting lines of $\|Q_S(h)\|, \|Q_S(l)\|, \|S_Q(h)\|$ and $\|S_Q(l)\|$, degree distributions $P_Q(k)$ of $Q_S(l)$ nodes and $P_S(k)$ of $S_Q(l)$ nodes where $k \in \{1, 2, \cdots, 30\}$, fitting curves associated with the slope and intercept of the linear relations of the ith highest-degree of $Q_S(h)$ and $S_Q(h)$ nodes. Edge properties: three preference attachment distributions, $P_{Q,c}(k)$ defined as the probability that a $Q_S(h) - S_Q(l)$ edge connects a k-degree $S_Q(l)$ node, $P_{S,c}(k)$ defined as the probability that a $S_Q(h) - Q_S(l)$ edge connects a k-degree $Q_S(l)$ node, $P_S(k_2)$ defined as the probability that a $Q_S(l) - S_Q(l)$ edge connects a k_2-degree $S_Q(l)$ node. Joint rank distribution $P(r_1, r_2)$ defined in Eq. (11), which is the probability that a $Q_S(h) - S_Q(h)$ edge connects r_1-rank and r_2-rank node categories.
2: **Output:** The $S - Q$ component.
3: Derive node sets $Q_S(h), Q_S(l), S_Q(h), S_Q(l)$ and their degrees using the methods of lines 3-6 in Algorithm 5.
4: Derive the edge numbers $\|Q_S(l) - S_Q(l)\|$ and $\|Q_S(h) - S_Q(h)\|$ using Eq. (10) and the sum of $Q_S(h)$ degrees.
5: According to $e_l = \|Q_S(l) - S_Q(l)\|$ and the input $P_S(k_2)$, connect $Q_S(l) - S_Q(l)$ edges using the methods of lines 7-16 in Algorithm 5. Note that line 9 in Algorithm 5 should be replaced by "9: **For** $k_1 = 1 : 1 : 30$ **do**".
6: Respectively classify $Q_S(h)$ and $S_Q(h)$ into $c = 30$ sorted categories $L^Q(1), L^Q(2), \cdots, L^Q(c)$ and $L^S(1), L^S(2), \cdots, L^S(c)$ using Algorithm 2, and initialize $e_h \leftarrow \|Q_S(h) - S_Q(h)\|$.
7: **For** $r_1 = 1 : 1 : c$ **do**
8: **For** $r_2 = 1 : 1 : c$ **do**
9:   Derive $z = round(e_h \times P(r_1, r_2))$, and initialize $t_c \leftarrow 0$.
10:   **While** $z \geq 1 \wedge t_c = 0$ **do**
11:    Extract a node v_1 in the category $L^Q(r_1)$ and a node v_2 in the category $L^S(r_2)$ where v_1 and v_2 are not connected and have maximum $min(fd(v_1), fd(v_2))$. If the two nodes can be extracted, connect the two nodes and update $z \leftarrow z - 1$, otherwise update $t_c \leftarrow 1$. Note that $fd(v)$ is defined as the number of free degrees of node v.
12:   **End while**
13: **End for**
14: **End for**
15: According to $P_{Q,c}(k)$, connect $Q_S(h) - S_Q(l)$ edges using the methods of lines 17-26 in Algorithm 5.
16: According to $P_{S,c}(k)$, connect $S_Q(h) - Q_S(l)$ edges using the methods of lines 17-26 in Algorithm 5.

AS graph of the UCLA dataset, which was investigated in January 2015, is shown in Fig. 14.

Because the output of SInetL is a series of subgraphs of the largest AS graph, the simulated SInetL graph related

Algorithm 7 Generation of the $U - W$ Node Mapping Connections

1: **Input:** Two node sets U and W where $\|U\| \leq \|W\|$ and $U - W$ is one of the eight dotted connections in Figs. 3 and 4; two solid components $G(U)$ and $G(W)$ that include U and W respectively; category number c; degree threshold d in Table 2; the joint rank distribution $P(r_1, r_2)$, which is the probability that a $U - W$ edge connects two nodes respectively belonging to r_1-rank W category and r_2-rank U category.

2: **Output:** The $U - W$ node mapping connections.

3: Respectively classify U and W into two sorted category lists $UL(1), UL(2), \cdots, UL(c+d)$ and
   $WL(1), WL(2), \cdots, WL(c+d)$ sing Algorithm 3,
   sort all the nodes in U as $u_1, u_2, \cdots, u_m$ and initialize
   $z(w) \leftarrow 0$ for each node $w \in W$.

4: **For** $i = 1 : 1 : m$ **do**

5:   Determine rank r_2 where $u_i \in UL(r_2)$, and initialize
     $b \leftarrow 0$.

6:   **While** $b = 0$ **do**

7:     For the given rank r_2, randomly select a rank
       $r_1 \in \{1, 2, \cdots, c+d\}$ with the rank distribution
       $P(r_1) = P(r_1, r_2) / \sum_{r_1} P(r_1, r_2)$.

8:     **If** $\|WS = \{w \mid w \in WL(r_1) \wedge z(w) = 0\}\| > 0$ **do**

9:       Uniformly at random, select a node $w \in WS$,
         establish a dotted connection between u_i and w,
         namely u_i and w are viewed as the same
         node in the AS-level Internet topology, and
         update $b \leftarrow 1$ and $z(w) \leftarrow 1$.

10:    **End if**

11: **End while**

12: **End for**

FIGURE 14. Graph properties of the real AS graph explored in January 2015 and the corresponding simulated graphs. (*a*) Degree CCDF $F(k)$ *vs.* degree k where $F(k) = \sum_{d>k} P(d)$ and $P(d)$ is the probability that a randomly selected node is d-degree. (*b*) Average neighbour connectivity $K(k)$ *vs.* degree k where $K(k)$ is simply the average neighbour degree of the average k-degree node [32]. (*c*) Clustering coefficient $C(k)$ *vs.* degree k where $C(k) = 2m(k)/(k(k-1))$ and $m(k)$ is the average number of links between the neighbours of k-degree nodes. (*d*) Percentage of total nodes *vs.* node categories P, Q, I, J, K and S defined in Eq. (3) that are the basic node classification of our structure model.

to January 2015 is the largest AS graph. Thus, the results of SInetL are not shown in Fig. 14. Fig. 14(*a*) shows that SICPS and all other generators perform well in terms of the degree power-law property because SICPS adopts the local degree distributions of the different structural components as inputs and the other three generators use three types of global degree distributions as inputs. From the comparison of Fig. 14(*b*)–(*d*), we find that SICPS performs best on all three properties because it partitions the AS graph into atomic-level components, separately generates diverse solid components, and merges them using different dotted connections. The structural decomposition feature enables SICPS to capture not only the global degree properties but also the correlation among the different local components. The average neighbor connectivity represents the correlation of the node degrees, which is not adopted by Inet-3.0 and S-BITE. Because ORBIS uses the $2K$ degree distribution, which is actually the joint degree distribution, it performs well, as shown in Fig. 14(*b*) and (*d*). However, Mahadevan *et al.* [14] indicated that ORBIS does not consider the clustering property.

Fig. 15 shows our comparison of the five generators, namely, Inet-3.0, ORBIS, S-BITE, SInetL, and SICPS, using
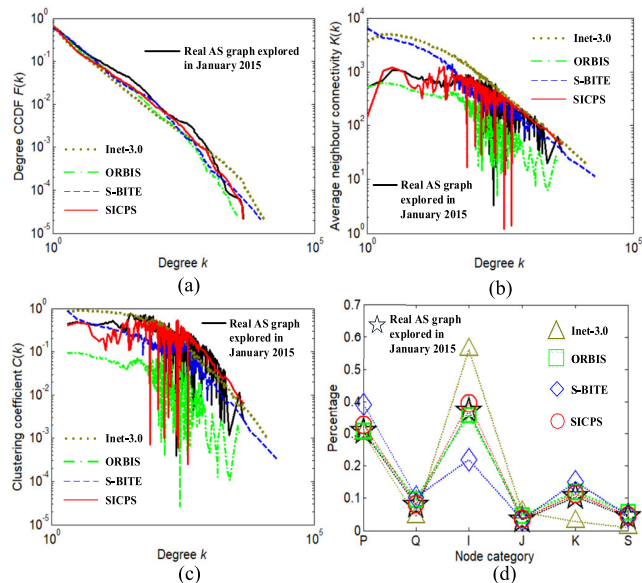
a series of AS graphs in the UCLA dataset investigated from January 2001 to January 2015.

Figs. 15(*a*) and (*b*) show that SICPS performs well on both the distance and clustering properties because it can accurately model the local structure and their correlations using the solid and dotted components, respectively. S-BITE accurately captures the core structure of the Internet topology. However, it does not consider the peripheral structure, which accounts for more than 95% of the nodes and is more important for the performance in terms of the statistical characteristics. S-BITE performs well in terms of the clustering coefficient because it uses statistical parameter p, which is the probability that a newly added peripheral node connects to two interconnected core nodes, to control the property [21]. ORBIS performs well in terms of most of the properties but neglects the clustering property [14]. Inet-3.0 captures the degree of the power-law properties, but it does not consider the degree correlation [10]. We note that the degree and rank correlations are critical tools for modeling the solid and dotted components of our structural model. Hence, SICPS also performs best in terms of the assortativity and maximum degree properties, as shown in Figs. 15(*c*) and (*d*). SInetL extracts a series of subgraphs from the given unique AS graph that was explored in January 2015 while maintaining some properties of the AS graph. In other words, it neglects the evolution of the UCLA dataset from 2001 to 2015, as shown in Figs. 15(*a*) and (*e*). Fig. 15(*e*) shows that the average degrees of the graphs simulated by SICPS are less than those
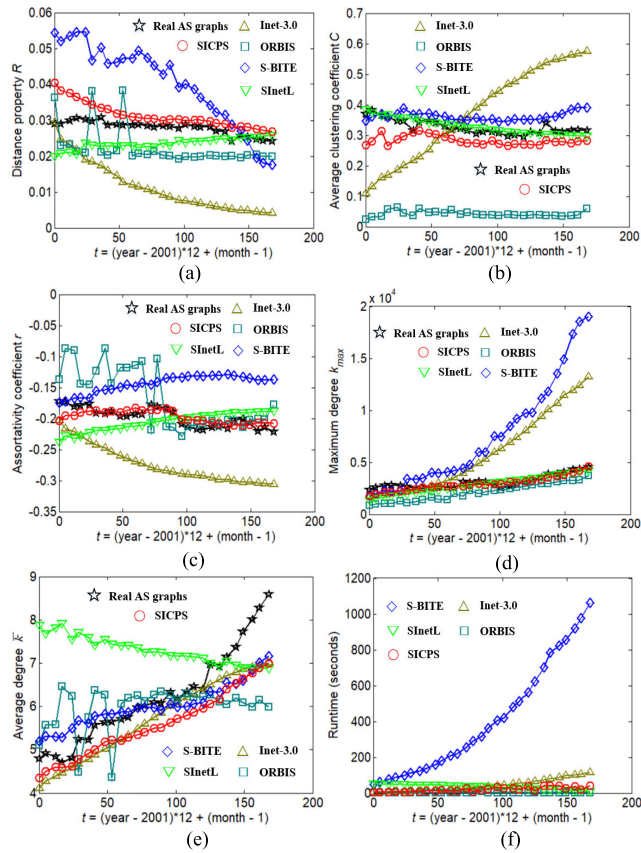
FIGURE 15. Graph evolving properties on the UCLA dataset from January 2001 to January 2015 and the corresponding simulated graphs. (a) Distance property R vs. t (the number of months since January 2001) where $R = \sum_i (1 - \lambda_i)^4 / \|V\|$ and λ_i ($i = 1, 2, \cdots, \|V\|$) are all the eigenvalues of the normalized Laplacian spectrum [34]. (b) Average clustering coefficient C vs. t where C is defined in section II.A. (c) Assortativity coefficient r vs. t where r is defined in section II.A. (d) Maximum degree k_{max} vs. t. (e) Average degree $\bar{k}$ vs. t. (f) Runtime of the five generators vs. t. Note that the number of nodes increases from 10,301 to 49,448 as t grows from 0 to 168.

of the real-world AS graphs because some free degrees presented in Section V have not been filled and the L and noise nodes defined in Eq. (3) are removed to simplify the problem. However, these phenomena also exist in S-BITE, ORBIS, and Inet-3.0.

Finally, we compare the runtimes of the five generators, as shown in Fig. 15(f). SICPS, ORBIS, and Inet-3.0 all generate graphs by filling the free degrees; thus, the time complexity of the three generators is $O(\|E\|)$, where $\|E\|$ is the total number of edges. In [29], the time complexity of SInetL was proven to be $O(\|E\|^2)$. For each newly added peripheral node, S-BITE first randomly selects a node that already exists in the network and then connects the peripheral node to the selected node. Because the peripheral nodes account for more than 95% of the Internet topology, the generators that are based on the free-degree filling realize better time efficiency.

To quantitatively compare the five generators using the results shown in Fig. 15, for each graph property x in explored graph G_e and corresponding simulated graph G_S, we define

TABLE 5. Average deviation degree and average runtime of the five generators

Property x	Inet-3.0	ORBIS	S-BITE	SInetL	SICPS
R	0.597	0.263	0.479	0.138	0.121
C	0.477	0.871	0.122	0.037	0.187
r	0.380	0.209	0.262	0.138	0.046
k_{max}	0.870	0.336	1.299	0.157	0.092
$\bar{k}$	0.081	0.127	0.073	0.266	0.096
Runtime	43.5 s	1.96 s	409 s	33.6 s	20.5 s

R: Distance property; C: Average clustering coefficient; r: Assortativity coefficient; k_{max}: Maximum degree; $\bar{k}$: Average degree. The above runtime of a generator is the average runtime of the generator to simulate all the explored graphs in the UCLA dataset.

the deviation degree as follows:

$$D(x, G_e, G_s) = |x(G_s) - x(G_e)| / |x(G_e)|, \quad (14)$$

where $x(G_s)$ and $x(G_e)$ denote the property values of x in G_S and G_e, respectively, and then define average deviation degree $D(x, y)$ of the UCLA dataset for a series of graphs simulated by certain generator y as follows:

$$D(x, y) = \sum_{G_e \in UCLA} D(x, G_s, G_e) / \|UCLA\|, \quad (15)$$

where $UCLA$ is the set of explored AS graphs included in the UCLA dataset, $\|UCLA\|$ is the cardinality of the set, and y is the generator that simulates graphs G_s for the explored AS graphs $G_e \in UCLA$.

In Table 5, for graph properties $x \in \{R, C, r, k_{max}, \bar{k}\}$ and $y \in \{Inet - 3.0, ORBIS, S - BITE, SInetL, SICPS\}$, we list average deviation degrees $D(x, y)$ and the average runtime of the five generators. The list in Table 5 illustrates that SICPS performs best on properties $x \in R, r, k_{max}$. Further, we observe that $D(\bar{k}, SICPS)$ is closer to $D(\bar{k}, Inet - 3.0)$ and $D(\bar{k}, S - BITE)$. Therefore, from the perspective of multiple factors, SICPS demonstrates a greater advantage.

VII. CONCLUSION AND FUTURE WORK
The Internet is a complex network system. According to different requirements, researchers can model the Internet topologies from various perspectives such as macro, micro, wired, wireless, and information centric networking. From different perspectives, the physical meaning of nodes and edges in the topology may be quite different. In the present study, we focus on the AS-level Internet topology in which the nodes represent ASs and the edges describe the data-communication paths among these nodes. Research on the AS-level topology can satisfy many application requirements of interdomain systems, such as routing optimization.

Although many AS-level Internet-topology models have been studied, the models based on deep structural decomposition still need to be seriously investigated because the exponential growth of the topological scale has become a critical obstacle in the analysis of the current Internet behavior [38, 39]. Structural decomposition is a useful means to reduce the complexity of the problems.

In the present study, we introduce the periphery that accounts for more than 95% of AS nodes into the structural model and decompose the Internet topology into 16 atomic-level solid and dotted components from the viewpoint of local connection and evolutionary stability. In contrast to the global statistical characteristics, our structural decomposition model is helpful for researchers to more precisely distinguish the Internet interdomain topology from other complex networks. In addition, our structural model provides many adjustable local characteristic parameters, which can help engineers generate topological environments in diverse scenes.

According to the UCLA dataset that spans 15 years, we obtain many uniform distribution characteristics that exist in the decomposed components of the Internet topology. In contrast to the power-law distribution, the uniform distribution is simpler. The discovery of these simpler distribution characteristics is more conducive to the recognition of the evolution stability of the Internet topology.

In addition, we find that most of the node and edge properties of these components remain constant except for the top five highest degrees of transit AS nodes. The inconstant property implies that the top five transit AS nodes are attracting more nodes (especially other transit AS nodes) to connect to them. The evolution of the Internet topology cannot be stable at all times. In other words, capturing the inconstant property is also important for the recognition and prediction of the topology. Furthermore, we design topology generator SICPS based on our structural model. The comparison results show that SICPS performs best on both global statistical and local structural properties. Although our generator needs more detailed statistical parameters from the 16 components as inputs, we can view the inputs as an accurate portrait of the Internet topology, which is an important reflection of the accuracy of the structural models.

In future work, we will analyze the relationship between the Internet behavior and the 16 decomposed components and apply our structural model and topological generator to network management and other engineering fields.

REFERENCES

[1] (2020). *IRL: UCLA Internet Research Lab.* Accessed: Feb. 16, 2020. [Online]. Available: http://irl.cs.ucla.edu

[2] Y. Kan, Q. Zheng, J. Yang, and X. Tan, "A cache invalidation strategy based on publish/subscribe for named data networking," *IEEE Access*, vol. 8, pp. 80074–80085, 2020.

[3] M. A. Naeem, M. A. U. Rehman, R. Ullah, and B.-S. Kim, "A comparative performance analysis of popularity-based caching strategies in named data networking," *IEEE Access*, vol. 8, pp. 50057–50077, 2020.

[4] G. Li and Z. Wu, "Information-centric networking cache robustness strategy for system wide information management," *IEEE Access*, vol. 8, pp. 82432–82441, 2020.

[5] H. Sun, H. Yi, F. Zhuo, X. Du, and G. Yang, "Precise fault location in distribution networks based on optimal monitor allocation," *IEEE Trans. Power Del.*, vol. 35, no. 4, pp. 1788–1799, Aug. 2020.

[6] I. Voitalov, R. Aldecoa, L. Wang, and D. Krioukov, "Geohyperbolic routing and addressing schemes," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 47, no. 3, pp. 11–18, Sep. 2017.

[7] B. M. Waxman, "Routing of multipoint connections," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 9, pp. 1617–1622, Dec. 1988.

[8] J. Wang, S. Jia, H. Zhao, J. Xu, and C. Lin, "Internet anomaly detection based on complex network path," *IEICE Trans. Commun.*, vol. E101.B, no. 12, pp. 2397–2408, 2018.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, Oct. 1999.

[10] J. Winick and S. Jamin, "Inet-3.0: Internet topology generator," Univ. Michigan, Ann Arbor, MI, USA, Tech. Rep., 2002.

[11] R. Motamedi, R. Rejaie, and W. Willinger, "A survey of techniques for Internet topology discovery," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 1044–1065, 2nd Quart., 2015.

[12] A. Lodhi, A. Dhamdhere, and C. Dovrolis, "Peering strategy adoption by transit providers in the Internet: A game theoretic approach?" *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 40, no. 2, pp. 38–41, Oct. 2012.

[13] M. Meddeb, A. Dhraief, A. Belghith, T. Monteil, K. Drira, and S. Gannouni, "AFIRM: Adaptive forwarding based link recovery for mobility support in NDN/IoT networks," *Future Gener. Comput. Syst.*, vol. 87, pp. 351–363, Oct. 2018.

[14] P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat, "Orbis: Rescaling degree correlations to generate annotated Internet topologies," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, pp. 325–336, Aug. 2007.

[15] D. L. Alderson, J. C. Doyle, and W. Willinger, "Lessons from 'a first-principles approach to understanding the Internet's router-level topology,'" *ACM SIGCOMM Comput. Commun. Rev.*, vol. 49, no. 5, pp. 96–103, Nov. 2019.

[16] S. Zhou and R. J. Mondragon, "The rich-club phenomenon in the Internet topology," *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 180–182, Mar. 2004.

[17] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, "A model of Internet topology using k-shell decomposition," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 27, pp. 11150–11154, Jul. 2007.

[18] E. K. Çetinkaya and T. A. Shatto, "Eigenvalues for resilience analysis of backbone networks," in *Proc. SIAM Workshop Netw. Sci.*, Pittsburgh, PA, USA, Jul. 2017, pp. 1–10.

[19] S. Jia, M. Luckie, B. Huffaker, A. Elmokashfi, E. Aben, K. Claffy, and A. Dhamdhere, "Tracking the deployment of IPv6: Topology, routing and performance," *Comput. Netw.*, vol. 165, Dec. 2019, Art. no. 106947.

[20] X. Liu, J. Wang, W. Jing, M. de Jong, J. S. Tummers, and H. Zhao, "Evolution of the Internet AS-level topology: From nodes and edges to components," *Chin. Phys. B*, vol. 27, no. 12, Dec. 2018, Art. no. 120501.

[21] G. Accongiagioco, E. Gregori, and L. Lenzini, "S-BITE: A structure-based Internet topology gEnerator," *Comput. Netw.*, vol. 77, pp. 73–89, Feb. 2015.

[22] M. Lei, L. Liu, and D. Wei, "An improved method for measuring the complexity in complex networks based on structure entropy," *IEEE Access*, vol. 7, pp. 159190–159198, 2019.

[23] S. Zu, X. Luo, S. Liu, Y. Liu, and F. Liu, "City-level IP geolocation algorithm based on PoP network topology," *IEEE Access*, vol. 6, pp. 64867–64875, 2018.

[24] J. Bo, L. Zhi-yong, D. Jing, L. Ying-long, H. Cheng-dong, and Z. Ying, "Study on the stability of the topology interactive growth mechanism using graph spectra," *IET Commun.*, vol. 8, no. 16, pp. 2845–2857, Nov. 2014.

[25] B. Jiao and J.-M. Shi, "Graph perturbations and corresponding spectral changes in Internet topologies," *Comput. Commun.*, vol. 76, pp. 77–86, Feb. 2016.

[26] B. Jiao, J. Shi, X. Wu, Y. Nie, C. Huang, J. Du, Y. Zhou, R. Guo, and Y. Tao, "Correlation between weighted spectral distribution and average path length in evolving networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 26, no. 2, Feb. 2016, Art. no. 023110.

[27] B. Jiao and X. Wu, "The 3-cycle weighted spectral distribution in evolving community-based networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 27, no. 3, Mar. 2017, Art. no. 033109.

[28] B. Jiao, Y.-P. Nie, J.-M. Shi, G. Lu, Y. Zhou, and J. Du, "Accurately and quickly calculating the weighted spectral distribution," *Telecommun. Syst.*, vol. 62, no. 1, pp. 231–243, May 2016.

[29] B. Jiao, J. Shi, W. Zhang, and L. Xing, "Graph sampling for Internet topologies using normalized Laplacian spectral features," *Inf. Sci.*, vol. 481, pp. 574–603, May 2019.

[30] G. Sun and S. Bin, "Router-level Internet topology evolution model based on multi-subnet composited complex network model," *J. Internet Technol.*, vol. 18, no. 6, pp. 1275–1283, 2017.

[31] R. B. da Silva and E. S. Mota, "A survey on approaches to reduce BGP interdomain routing convergence delay on the Internet," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2949–2984, 4th Quart., 2017.

[32] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, K. C. Claffy, and A. Vahdat, "The Internet AS-level topology: Three data sources and one definitive metric," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, pp. 17–26, Jan. 2006.

[33] D. Yang, L. Pan, and T. Zhou, "Lower bound of assortativity coefficient in scale-free networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 27, no. 3, Mar. 2017, Art. no. 033113.

[34] D. Fay, H. Haddadi, A. Thomason, A. W. Moore, R. Mortier, A. Jamakovic, S. Uhlig, and M. Rio, "Weighted spectral distribution for Internet topology analysis: Theory and applications," *IEEE/ACM Trans. Netw.*, vol. 18, no. 1, pp. 164–176, Feb. 2010.

[35] E. K. Çetinkaya, M. J. F. Alenazi, A. M. Peck, J. P. Rohrer, and J. P. G. Sterbenz, "Multilevel resilience analysis of transportation and communication networks," *Telecommun. Syst.*, vol. 60, no. 4, pp. 515–537, Mar. 2015.

[36] M. Cinelli, G. Ferraro, and A. Iovanella, "Rich-club ordering and the dyadic effect: Two interrelated phenomena," *Phys. A, Stat. Mech. Appl.*, vol. 490, pp. 808–818, Jan. 2018.

[37] F. Comellas and S. Gago, "A star-based model for the eigenvalue power law of Internet graphs," *Phys. A, Stat. Mech. Appl.*, vol. 351, nos. 2–4, pp. 680–686, Jun. 2005.

[38] K. Claffy and D. Clark, "Workshop on Internet economics (WIE2018) final report," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 49, pp. 25–30, May 2019.

[39] R. Motamedi, B. Yeganeh, B. Chandrasekaran, R. Rejaie, B. M. Maggs, and W. Willinger, "On mapping the interconnections in Today's Internet," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 2056–2070, Oct. 2019.

BO JIAO received the Ph.D. degree in control science and engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2009. From 2010 to 2017, he was a Research Associate/Research Assistant with the Luoyang Electronic Equipment Test Center, China. He is currently a Distinguished Researcher with the School of Mathematics and Big Data, Foshan University, Foshan, China. He has authored more than 30 refereed articles, including those published in *Information Sciences*, *Computer Communications*, and *Chaos*. His current research interests include Internet topology, spectral graph theory, complex networks, and evolving systems.

WENSHENG ZHANG received the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, in 2000. He joined the Institute of Software, CAS, in 2001. He is currently a Professor of machine learning and data mining and the Director of the Department of Research and Development, Institute of Automation, CAS. His research interests include artificial intelligence, knowledge mining, and probability graph model.

• • •