Markov Property of Correlated Random Networks and its Application to the Analysis of the Internet Topologies

Shigeo Shioda, Masato Yoshihara
Graduate School of Engineering, Chiba University
1-33, Yayoi, Inage, Chiba 263-8522, Japan
Email: shioda@faculty.chiba-u.jp

Asuka Izumi, Kouhei Tachikawa
Faculty of Engineering, Chiba University
1-33, Yayoi, Inage, Chiba 263-8522, Japan

***Abstract*—We propose a framework for modeling higher-order (beyond two point) degree correlation among nodes, depending on their mutual connectivity. Our focus is on the introduction of the *Markov* property to find maximally unbiased networks under the constraint of a prescribed two-point degree correlation. The topological features of the Markovian networks – networks satisfying the Markov property – are fully characterized solely by the two-point degree correlation. We theoretically investigate the topological characteristics of Markovian networks and derive the analytical formulas for their graph theoretical metrics. We present a comparative analysis of AS- and router-level topologies in terms of whether they are Markovian or not. The results of the analysis show that the studied AS- and router-level topologies are not Markovian. This finding indicates that it is rather difficult to capture the topological characteristics of the Internet either at AS- or at router-level solely by the input of the two-point degree correlation.**

I. Introduction

Network topology has a large impact on the performance of communication protocols and applications [1]. The stability of routing protocols, robustness against network failures, efficiency of file search algorithms, and vulnerability to the DDoS attacks all strongly depend on the topological characteristics of networks. Over the past decade, a considerable number of studies have been made on how to analyze and describe the topological characteristics of networks using graph-theoretical metrics. Various network generators have also been proposed to generate a set of graphs, which reproduce graph-theoretical metrics commonly found in real-world networks [2], [3], [4], [5], [6], [1], [7].

The empirical analysis of real networks (not only in telecommunication filed but also in various scientific fields) reveals the presence of common topological properties. The power-law degree distribution is a widely-known property commonly found in real networks. Most of existing network generators aim at producing graphs with power-law degree distributions or graphs with prescribed degree distributions. The degree distribution does not, however, capture all of topological features, and other metrics beyond the degree distribution are required to more precisely categorize and distinguish various networks observed in real world.

It has been found that empirical networks in the real world show a nontrivial degree correlation between adjacent nodes (two-point degree correlation); for example, large-degree nodes are more likely to be connected to small-degree nodes. Motivated by the observation of the presence of the degree correlation, a number of works have been devoted to construct network models, called correlated network models, to study the origin of correlations in real networks [8], [9], [10], [11], [12]. Mahadevan *et al.* [1], [7] introduced the concept of dK-graphs, which are specified by the joint degree distribution of nodes within a subgraph of size d of the original network. For example, $1K$-graphs reproduce the degree distribution of the original network, and $2K$-graphs reproduce the joint degree distribution of adjacent nodes of the original network. They proposed an algorithm to construct a graph among dK-graphs for $d = 0, 1, 2, 3$, and they claimed that $2K$-graphs reproduce the Internet AS- and router-level topologies with accuracy sufficient for most practical purposes.

The results by Mahadevan *et al.* [1], [7] tell us the importance of considering the two-point degree correlation in the analysis of the network topology. The $2K$-graphs satisfying a prescribed two-point degree correlation, however, still show the structural diversity; for example, they are largely different in their clustering coefficient. Thus, it is natural to seek a principle to find the most *unbiased* (*random*) network among ones satisfying a prescribed two-point degree correlation. To this end, in this work, we introduce a notion of *Markov* property. The topological features of Markovian networks are fully characterized solely by the two-point degree correlation. We propose a framework for modeling the degree correlation of nodes within a subgraph of size three, on which the Markov property is introduced. Our framework allows us to consider the dependence of degree correlation among three nodes on their mutual connectivity (*wedge* or *triangle*). We theoretically investigate the topological characteristics of Markovian networks and derive analytical formulae for some graph theoretical metrics. Based on the analytical results obtained in the work, we numerically investigate whether the Internet AS- and router-level topologies are Markovian or not using the real data.

978-0-9836283-1-6

This paper was peer reviewed by subject matter experts for publication in the Proceedings of Cnet 2011

The manuscript is organized as follow. Section II introduces the notation used in this work and explains some related results. Section III proposes a new framework for modeling the three-point degree correlation as well as the structural detail of a subgraph on which the three-point degree correlation is considered. Section IV defines the Markov property and theoretically derives the topological characteristics of Markovian networks. Section V numerically investigates whether the Internet AS- and router-level topologies are Markovian or not. We conclude the article in Sec. VI.

II. Background

A. Degree distribution

Consider an undirected connected network with N nodes and L links. We treat degree of a randomly chosen node D as a discrete random variable and let $p(k)(= P[D = k])$ denote the probability that the degree of a randomly chosen node is k. If D follows a power-law distribution, that is

$$p(k) \sim ak^{-(\eta)},$$

then the network is called *scale free* and parameter η is called *characteristic exponent*.

B. Edge-based sampling

Here, we consider two sampling methods for nodes in the network. The first one is the *normal* sampling, which chooses any of nodes with equal probability. The other is the *edge-based* sampling, which randomly chooses a link (an edge) in the first step and chooses one of its end nodes in the second step. Let $p_e(k)(= P_e[D = k]$[1]$)$ denote the probability that the edge-based sampling chooses a node with degree k [12]. The edge-based sampling is different with the normal sampling, and there is the following relationship between the degree distributions under the edge-based and normal samplings [12]:

$$p_e(k) = \frac{kp(k)}{E[D]}, \tag{1}$$

where $E[D]$ denotes the average degree of randomly chosen (normally sampled) nodes. The edge-based sampling is more likely to choose nodes with higher degrees, while the normal sampling equally chooses nodes irrespective of their degrees. This fact can easily be understood from the observation that one of links connected to a node with degree k is chosen with probability proportional to $kp(k)$ when any of links in the network are chosen with equal probability.

C. Two-point degree correlation

Let $p_e(k,l)$ denote the probability that a randomly chosen link of the network has nodes with degrees k and l at its ends. The two-point degree correlation is fully expressed in terms of $p_e(k,l)$. If the degrees of the end nodes of any given link are statistically independent, we have

$$p_e(k,l) = p_e(k)p_e(l). \tag{2}$$

[1] $P_e[\cdot]$ and $E_e[\cdot]$ respectively denote the probability measure and the expectation under the edge-based sampling.

If (2) holds, we say that networks are uncorrelated [12]; if $p_e(k,l) < p_e(k)p_e(l)$ ($p_e(k,l) > p_e(k)p_e(l)$), networks are called positively (negatively) correlated. The assortativity introduced by Newman is a metric representing the extent of the two-point degree correlation [13]. In terms of $p_e(k,l)$, the assortativity is expressed as

$$r^{(1)} \overset{\text{def}}{=} \frac{\sum_{k,l} kl(p_e(k,l) - p_e(k)p_e(l))}{E_e[D^2] - (E_e[D])^2},$$

where $E_e[D]$ and $E_e[D^2]$ respectively denote the average degree and the average square degree of edge-based sampled nodes [13]. As apparent from the above expression, the assortativity is the Pearson correlation coefficient of the degrees of two end nodes of a link. It follows from (1) that

$$E_e[D] = \frac{\sum k^2 p(k)}{E[D]} = E[D^2]/E[D],$$
$$E_e[D^2] = \frac{\sum k^3 p(k)}{E[D]} = E[D^3]/E[D],$$

and thus we have another expression for the assortativity.

$$r^{(1)} = \frac{(E[D])^2(\sum_{k,l} klp_e(k,l)) - (E[D^2])^2}{E[D]E[D^3] - (E[D^2])^2}.$$

The assortativity falls into the range $[-1, 1]$ and for uncorrelated networks it is equal to zero.

Remark 1. *Within the authors' knowledge, the notion of edge-based sampling has never been introduced in existing literature. This notion allows us to have a very concise description of the two-point degree correlation. In [12], $p_e(k)$ is called the distribution over edge ends.*

III. Modeling Three-Point Degree Correlation

A. Framework

Consider a subgraph made of a node (node A) and its randomly chosen two neighbors (nodes B and C) (Fig. 1). (Node A has at least two neighbors.) If nodes B and C are not directly connected, the subgraph is called *wedge*, a chain of three nodes connected by two links; otherwise, it is a *triangle*. The degree correlation of nodes A, B and C defines the three-point degree correlation. Let $p_e(k,m;l)$ be the conditional probability that the degree of a edge-based-sampled node is l, and the degrees of its two neighbors randomly chosen are respectively k and m, given that $l \geq 2$. We also let $q(k,m;l)$ denote the probability that the subgraph forms a triangle.

For example, $p(k)$, $p_e(k)$, $p_e(k,l)$, $p_e(k,m;l)$ and $q(k,m;l)$ of a network depicted in Fig. 2 are given below

$$p(1) = 1/4, \quad p(2) = 1/2, \quad p(3) = 1/4,$$
$$p_e(1) = 1/8, \quad p_e(2) = 1/2, \quad p_e(3) = 3/8,$$
$$p_e(1,3) = p_e(3,1) = 1/8,$$
$$p_e(2,3) = p_e(3,2) = p_e(2,2) = 1/4,$$
$$p_e(1,2;3) = p_e(2,1;3) = p_e(2,2;3) = 1/7,$$
$$p_e(2,3;2) = p_e(3,2;2) = 2/7,$$
$$q(1,2;3) = q(2,1;3) = 0,$$
$$q(2,2;3) = q(2,3;2) = q(3,2;2) = 1.$$

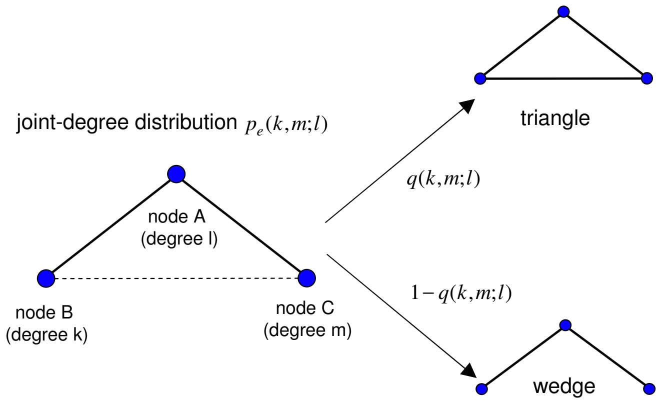

Fig. 1. Subgraph of size three.

Note that $p_e(k,m;l)$ is meaningful only if $p_e(1) < 1$; otherwise any connected-subgraphs with size three do not exist with probability one.

The clustering coefficient, the probability that two neighbors of a node are also neighbors themselves, can be expressed in terms of $p_e(k,m;l)$ and $q(k,m;l)$. For example, the average clustering coefficient of nodes with degree l, $C(l)$, is expressed as

$$\begin{aligned} C(l) &= \frac{\sum_{k,m} p_e(k,m;l) q(k,m;l)}{\sum_{k,m} p_e(k,m;l)} \\ &= \frac{1-p_e(1)}{p_e(l)} \sum_{k,m} p_e(k,m;l) q(k,m;l), \end{aligned} \tag{3}$$

where we use the fact that $\sum_{k,m} p_e(k,m;l) = p_e(l)/(1-p_e(1))$. (Node A should have at least two neighbors.)

The joint degree distribution of nodes within a wedge $p_\wedge(k,m;l)$ is expressed in terms of $p_e(k,m;l)$ and $q(k,m;l)$ as follows:

$$p_\wedge(k,m;l) = A(1-q(k,m;l))p_e(k,m;l),$$

where

$$\begin{aligned} A &\stackrel{\text{def}}{=} 1/\sum_{k,m,l}(1-q(k,m;l))p_e(k,m;l) \\ &= \frac{1-p_e(1)}{1-p_e(1)-\sum_{l\geq 2} p_e(l)C(l)} \\ &= \frac{1}{1-E_e[C]}, \\ E_e[C] &\stackrel{\text{def}}{=} \frac{1}{1-p_e(1)} \sum_{l\geq 2} p_e(l)C(l). \end{aligned}$$

Note that $E_e[C]$ is the average clustering coefficient, where the average is taken in the edge-based-sampling sense. Using $p_\wedge(k,m;l)$, we define the second-order assortativity:

$$r^{(2)} \stackrel{\text{def}}{=} \frac{\sum_{k_1,k_3} k_1 k_3 \left(\sum_{k_2} p_\wedge(k_1,k_3;k_2) - p_\wedge(k_1)p_\wedge(k_3)\right)}{E_\wedge[D^2]-(E_\wedge[D])^2}, \tag{4}$$

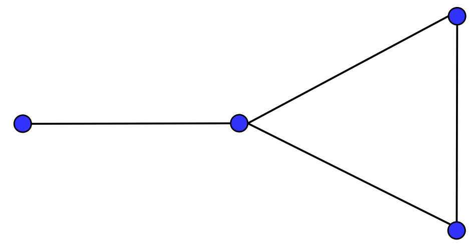

Fig. 2. Exapmle of networks.

where

$$\begin{aligned} p_\wedge(k) &\stackrel{\text{def}}{=} \sum_{l,m} p_\wedge(k,m;l), \\ E_\wedge[D] &\stackrel{\text{def}}{=} \sum_k k p_\wedge(k), \quad E_\wedge[D^2] \stackrel{\text{def}}{=} \sum_k k^2 p_\wedge(k). \end{aligned}$$

Note that $p_\wedge(k)$ is the degree distribution under the *wedge-based* sampling, which randomly chooses a wedge structure and chooses one of its end nodes. The second-order assortativity is the Pearson correlation coefficient of the degrees of two nodes located at a distance of two hops.

The second-order assortativity and the clustering coefficient are independent metrics for the three-point degree correlation.

B. *Uncorrelated network*

Although most of real networks show the existence of degree correlation, the uncorrelated networks are practically important to check the accuracy and the analytical solutions of dynamical processes defined on random networks [14]. The topological features of uncorrelated networks are fully characterized solely by the degree distribution. Uncorrelated networks are maximally random networks under the constraint of two-hop degree correlation [1]. To see this, let $H(D)$ denote the entropy of the degree distribution under the edge-based sampling and $H(D_1,D_2)$ denote the entropy of the joint degree distribution of end nodes of a link. It follows from the definition of the entropy that $H(D_1,D_2) \leq 2H(D)$, but for the uncorrelated network

$$\begin{aligned} &H(D_1,D_2) \\ &= -\sum_k \sum_l p_e(k,l)\log p_e(k,l) \\ &= -\sum_k \sum_l p_e(k)p_e(l)\log p_e(k) - \sum_k \sum_l p_e(k)p_e(l)\log p_e(l) \\ &= -\sum_k p_e(k)\log p_e(k) - \sum_l p_e(l)\log p_e(l) = 2H(D). \end{aligned}$$

That is, the Markov property maximizes the mutual information of joint degree distribution $p_e(k,l)$.

In uncorrelated networks, a neighbor of any nodes is a m-degree node with probability $p_e(m)$. Thus, a k-degree node has $kp_e(m)$ adjacent nodes of degree m in average. Since there are $Np(m)$ m-degree nodes in the network, k-degree and m-degree nodes are connected with the following probability:

$$\frac{kp_e(m)}{Np(m)} = \frac{km}{NE[D]}.$$

This observation yields

$$\begin{aligned} &q(k,m;l) \\ &= q(k,m) = \begin{cases} \frac{(k-1)(m-1)}{NE[D]} & \text{for } (k-1)(m-1) \le NE[D], \\ 1 & \text{otherwise,} \end{cases} \end{aligned} \tag{5}$$

where $q(k,m)$ is the probability that k- and m-degree nodes are connected, and terms $k-1$ or $m-1$ comes from the fact that one of the connections of each node has already been used.

In uncorrelated networks, we have

$$p_e(k,m;l) = p_e(k)p_e(m)p_e(l)/(1-p_e(1)).$$

Thus, by substituting (5) into (3) and assuming $(k-1)(m-1) \le NE[D]$, we obtain

$$\begin{aligned} C(l) &= \frac{\sum_{k,m} p_e(k)p_e(m)p_e(l)\frac{(k-1)(m-1)}{NE[D]}}{p_e(l)} \\ &= \sum_{k,m} \frac{(k-1)(m-1)p_e(k)p_e(m)}{NE[D]} \\ &= \frac{(E[D^2]-E[D])^2}{N(E[D])^3}. \end{aligned}$$

The same expression has been derived in [15], [9].

The normalization constant A in the expression of $p_\wedge(k,m;l)$ is equal to $(1-\bar{C})^{-1}$ for uncorrelated networks, where

$$\bar{C} \stackrel{\text{def}}{=} \frac{1}{1-p(1)} \sum_{l \ge 2} p(l)C(l) \left(= \frac{(E[D^2]-E[D])^2}{N(E[D])^3} \right)$$

is the average clustering coefficient in the normal-sampling sense. Thus, the degree distribution under the *wedge-based* sampling, $p_\wedge(k)$, is expressed as

$$\begin{aligned} p_\wedge(k) &= \frac{1}{1-\bar{C}} \sum_{m,l} (1-q(k,m))p_e(k,m;l) \\ &= \frac{1}{1-\bar{C}} \sum_{m} (1-q(k,m))p_e(k)p_e(m) \\ &= \frac{1}{1-\bar{C}} p_e(k) \left(1 - \sum_{m} q(k,m)p_e(m) \right), \end{aligned}$$

which reveals a non-trivial result: in general $p_\wedge(k)$ is not equal to $p_e(k)$ even in uncorrelated networks. We also have a rather surprising result concerning the assortativity coefficients. The (1st-order) assortativity coefficient $r^{(1)}$ is zero, but the 2nd-order assortativity coefficient $r^{(2)}$ is not always equal to zero because the term in the numerator of (4)

$$\begin{aligned} &\sum_{k_1,k_3} k_1k_3 \sum_{k_2} p_\wedge(k_1,k_3;k_2) \\ &= \frac{1}{1-\bar{C}} \sum_{k_1,k_3} k_1k_3(1-q(k_1,k_3))p_e(k_1,k_3) \end{aligned}$$

cannot be factorized into the product of $E_\wedge[D]$ because of the term $q(k_1,k_3)$. The wedges are likely to have small-degree nodes at theirs ends, which generates the degree correlation between end nodes of a wedge even in uncorrelated networks.

The two-point degree correlation is characterized by the average degree of adjacent nodes of a k-degree node, which is formally defined as

$$\overline{D_{nn}}(k) = \sum_l l p_e(l|k),$$

where $p_e(l|k) = p_e(k,l)/p_e(k)$ is the probability that a randomly-selected adjacent node of a degree-k node has degree l. In uncorrelated networks, $\overline{D_{nn}}(k)$ does not depend on k as shown below.

$$\begin{aligned} \overline{D_{nn}}(k) &= \sum_l l p_e(l|k) \\ &= \sum_l l p_e(l) \\ &= E[D^2]/E[D]. \end{aligned}$$

It has been shown [16] that the average distance between two nodes is the order of $\log(N)/\log \tilde{d}$.

It has been an interesting issue to generate uncorrelated networks. The configuration model is the most widely-used algorithm to construct uncorrelated networks from a prescribed degree distribution and the topological properties of networks constructed by the configuration model has been extensively studied [6], [17], [18], [19], [20], [21], [22], [23], [24], [14]. The PLRG (Power-Law Random Graph) [6] is an algorithm of scale-free network generation based on the configuration model.

IV. Markov Property

A. Definition

Definition 1. *If a network meets the following two conditions, then we say that it has the Markov property:*

$$p_e(k,m;l) = \frac{p_e(k,l)}{1-p_e(1)} p_e(m|l) \tag{6}$$

$$q(k,m;l) = q(k,m). \tag{7}$$

To see the implication of condition (6), consider the following conditional probability:

$$p_e(m|k,l) \stackrel{\text{def}}{=} \frac{p_e(k,m;l)}{p_e(k,l)/(1-p_e(1))},$$

which represents the conditional probability that a (randomly selected) neighbor of an edge-based-sampled node has degree m, given that the edge-based-sampled node has degree l, the other end node of the sampled edge has degree k, and l is greater than 1. (It intuitively corresponds to the probability that, in a structure depicted in Fig. 1, the degree of node C is m given that the degrees of nodes A and B are respectively l and k.) It follows from (6) that

$$p_e(m|k,l) = p_e(k,l)p_e(m|l)/p_e(k,l) = p_e(m|l),$$

meaning that the degree of node C is independent of the degree of node B under the condition that the degree of node A is given. In other word, the degree of node C depends on node B only through the degree of node A, which is *so-called* Markov property.

The second condition (7) has the similar meaning with (6): the connectivity between a pair of nodes does not depend on the degree of common neighbor given that the degrees of the pair of nodes are specified. Note that $q(k,m)$ can be analytically expressed in terms of joint degree distribution. To see this, first observe that a node with degree k is connected to $kp_e(m|k)$ nodes with degree m in average. Since there are $Np(m)$ nodes with degree m in the network, we have

$$q(k,m) = \begin{cases} \frac{(k-1)p_e(m|k)}{Np(m)} = \frac{(k-1)(m-1)p_e(k,m)}{NE[D]p_e(k)p_e(m)} & \text{for } \frac{(k-1)(m-1)p_e(k,m)}{NE[D]p_e(k)p_e(m)} \le 1, \\ 1 & \text{otherwise,} \end{cases} \quad (8)$$

The Markovian network would be maximally random networks under the constraint of two-point degree correlation in the sense that it maximizes the entropy of the joint-degree distribution of degrees of nodes within a subgraph of size three. To see this, let $H(D_2, D_3; D_1)$ be the entropy of the joint degree distribution of a size-three subgraph depicted in Fig.1 (D_1, D_2, and D_3 respectively denote the degrees of nodes A, B and C). It follows from the definition of the entropy that $H(D_2, D_3; D_1) = H(D_1, D_2) + H(D_3|D_1, D_2) \le H(D_1, D_2) + H(D_3|D_2)$, but if (7) holds

$$\begin{aligned} &H(D_2, D_3; D_1) \\ &= -\sum_k \sum_l \sum_m p_e(k,m;l) \log p_e(k,m;l) \\ &= -\sum_k \sum_l \sum_m \frac{p_e(k,l)}{1-p_e(1)} p_e(m|l) \log\left(\frac{p_e(k,l)}{1-p_e(1)} p_e(m|l)\right) \\ &= -\sum_k \sum_l \sum_m \frac{p_e(k,l)}{1-p_e(1)} p_e(m|l) \log \frac{p_e(k,l)}{1-p_e(1)} \\ &\quad -\sum_k \sum_l \sum_m \frac{p_e(k,l)}{1-p_e(1)} p_e(m|l) \log p_e(m|l) \\ &= -\sum_l \sum_k \frac{p_e(k,l)}{1-p_e(1)} \log \frac{p_e(k,l)}{1-p_e(1)} \\ &\quad -\sum_l \sum_m \frac{p_e(l)}{1-p_e(1)} p_e(m|l) \log p_e(m|l) \\ &= H(D_1, D_2) + H(D_3|D_2), \end{aligned}$$

That is, in the Markovian network, the mutual information of joint degree distribution $p_e(k,m;l)$ is maximized.

Conditions (6) and (7) of the Markov property yields the following analytical expression for the degree distribution of a wedge structure $p_\wedge(k,m;l)$ in terms of the joint degree distribution of adjacent nodes $p_e(k,l)$:

$$\begin{aligned} &p_\wedge(k,m;l) \\ &= A\left(1 - \frac{(k-1)(m-1)p_e(k,m)}{NE[D]p_e(k)p_e(m)}\right) p_e(k,l)p_e(m|l)/(1-p_e(1)) \\ &= \frac{p_e(k,l)p_e(m|l)}{1-p_e(1)-\sum_{l\ge 2} C(l)p_e(l)} \left(1 - \frac{(k-1)(m-1)p_e(k,m)}{NE[D]p_e(k)p_e(m)}\right). \end{aligned} \quad (9)$$

From the above expression, we obtain the degree distribution under the wedge-based sampling $p_\wedge(k)$ and the average degree of a wedge-based sampled node $E_\wedge[D]$, which are used in the evaluation of the second-order assortativity coefficient. These values can be used as benchmarks to investigate whether a given network is Markovian or not.

Remark 2. *Boguna et al. [25] proposed a Markovian network model for describing higher-order degree correlation, but their model does not consider the dependence of the degree correlation among nodes on their mutual connectivity. Thus, it is difficult for their model to evaluate the joint degree distributions of nodes forming a wedge (or triangle).*

B. Metrics of three-point-degree correlation of Markovian network

The Markov property allows us to evaluate the metrics of the three-point degree correlation (e.g. the clustering and the second-order assortativity coefficients) based on the joint degree distribution of adjacent nodes $p_e(k,l)$. For example, substituting (8) into (3) yields

$$C(l) = \frac{1}{NE[D]} \sum_k \sum_m \frac{(k-1)(m-1)p_e(l,k)p_e(l,m)p_e(k,m)}{p_e(k)p_e(m)(p_e(l))^2}. \quad (10)$$

The second-order assortativity coefficient can also be evaluated by substituting (9) into (4) although we cannot have simpler expressions than (4). The results concerning the clustering coefficient and the second-order assortativity can be used as benchmarks of the Markov property.

C. Generation of Markovian networks

It would be an interesting issue to find the algorithm for generating a network satisfying the Markov property. One possible approach is to perform $2K-$*preserving* rewiring a sufficient number of times to a graph satisfying a prescribed two-point degree correlation [1]. We have proposed an algorithm to construct a graph satisfying a prescribed two-point degree correlation [26]. Our algorithm generates a connected graph that does not have any self-loops and any multiple links between nodes. Thus, we can use the output of our algorithm as an input for the $2K-$*preserving* rewiring. Through numerical experiments, we have found that the above mentioned procedure yields a network that has the clustering and the 2nd-assortativity coefficients very close to the expectations by the Markov property.

V. Are the Internet topologies are Markovian?

We have numerically investigated whether the AS-level and router-level topologies have the Markov property.

A. Network data

We used eight measured AS-level and ten router-level topologies. The AS-level topologies are snapshots obtained from BGP routing tables collected at BGP beacon of RIPE NCC RIS (Route Information Service) project [27]. The BGP data used in the analysis were collected on the September 3rd of every year from 1999 to 2007. The router level topologies

TABLE I
AS-LEVEL TOPOLOGY

Year	# of links	# of nodes	cluster. coef.	assort. coef.
1999	7825	5817	0.209792	-0.174473
2000	16814	8594	0.431251	-0.184715
2001	22360	11816	0.431538	-0.187681
2002	25385	13739	0.398226	-0.196199
2003	29801	15871	0.334195	-0.195170
2004	34185	18100	0.342195	-0.195262
2005	37811	20534	0.343488	-0.195812
2006	43357	23149	0.298065	-0.189331

TABLE II
ROUTER-LEVEL TOPOLOGY

Network	# of links	# of nodes	cluster. coef.	assort. coef.
AT&T	14261	11745	0.017112	-0.450118
Sprintlink	12816	10180	0.057527	-0.316064
Verio	9450	6252	0.119402	-0.278969
Telstra	4322	3515	0.054640	-0.230357
Level3	6917	1786	0.191861	0.015039
Abovenet	1332	654	0.285643	-0.196377
Tiscail	756	506	0.039277	0.062716
Exodus	893	424	0.292335	-0.210903
Ebone	548	300	0.175320	-0.198499
VSNL	285	226	0.058946	-0.235925

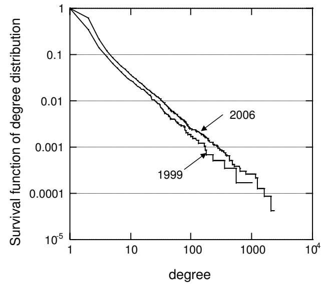

Fig. 3. Degree distribution (AS-level topology)

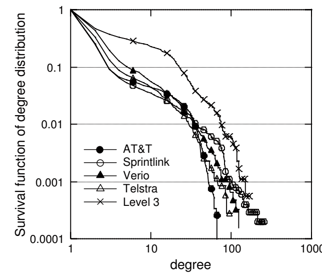

Fig. 4. Degree distribution (router-level topology)

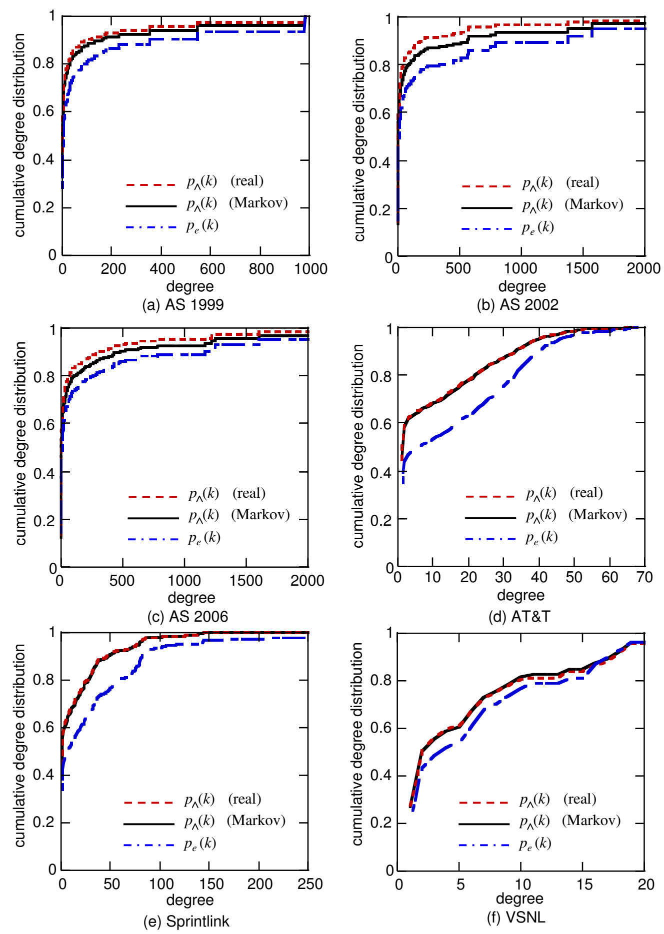

Fig. 5. Degree distribution under wedge-based sampling

were measured by the research group of Washington university using Rocketfuel, traceroute-base tool for topology analysis [28]. We summarize the topological data of the networks in Tables I and II.

The tables indicate that all of the networks are correlated because their assortativity coefficients are far from zero. All of the AS-level topologies and most of the router level topologies have negative assortativity coefficients, but the router-level topologies of Level-3 Communications and Tiscail have positive assortativity coefficients. The AS-level topologies have larger clustering coefficients than the router-level topologies.

Figure 3 shows the log-log plot of survival functions of the degree distributions of the AS-level topologies in 1999 and 2006. Although the AS-level topologies in 1999 and 2006 are very different in terms of the number of nodes and links, they have very similar power-law degree distributions. Figure 4 shows the log-log plot of survival functions of the degree distributions of router-level topologies of AT&T, Sprintlink, Verio, Telstra and Level 3 Communications. As the figures indicate, the degree distributions of router-level topologies do not necessarily exhibit typical power-law distributions.

B. Degree distributions under wedge-based sampling

We first investigate the degree distribution of wedge-based sampled nodes $p_\wedge(k)$ in the following procedure: we evaluate $p_\wedge(k)$ based on the two-hop degree correlation ($\{p_e(k,l)\}_{k,l\in\mathbb{N}}$) of the original topology (one of the AS-level and the router-level topologies) using the Markov property, and compare the resultant distribution with the real one of the original topology. Figure 5 shows the results for six topologies. For reference, in Fig. 5, we also show the cumulative degree distribution under the edge-based sampling. The figure shows that the cumulative degree distribution under the wedge-based sampling is larger than the one under the edge-based sampling. That is, wedges are likely to have small degree nodes at their ends. This fact can be understood from the observation that a three-node subgraph in Fig. 1 is likely to become a wedge when nodes B and C are small degree nodes. The figure also shows that, for AS-level topologies, the Markov property underestimates the cumulative degree distribution under the wedge-based sampling, equivalently implying that the Markov property

TABLE III
Average degree: AS-level topologies.

Year	$E[D]$	$E_e[D]$	$E_\wedge[D]$	
			real	Markov
1999	2.690390	110.5357	54.35840	68.45700
2000	3.912963	201.4241	75.08271	124.0863
2001	3.784699	269.1720	88.62603	162.8148
2002	3.695320	289.9202	107.3246	179.2369
2003	3.755403	266.4795	123.4499	178.5335
2004	3.777348	270.8629	121.7359	181.1810
2005	3.682770	270.7002	122.0396	182.1087
2006	3.745907	254.6451	126.3163	178.7363

TABLE IV
Average degree: router-level topologies.

Network	$E[D]$	$E_e[D]$	$E_\wedge[D]$	
			real	Markov
AT&T	2.428438	15.85211	10.15614	10.26589
Sprintlink	2.517878	31.82498	15.16892	15.63006
Verio	3.023033	18.70159	14.66031	15.87358
Telstra	2.459175	14.78968	10.86147	11.28003
Level3	7.745801	36.91615	34.53724	34.91744
Abovenet	4.073394	11.26577	9.111358	10.32635
Tiscail	2.988142	7.003968	6.608177	6.829092
Exodus	4.212264	7.977604	6.705841	6.946289
Ebone	3.653333	7.047445	6.462871	6.632233
VSNL	2.522124	6.712281	6.016624	5.907109

TABLE V
Verification of Markov property: AS-level topologies.

Year	clustering coefficient		2nd assortativity coefficient	
	real	Markov	real	Markov
1999	0.209792	0.112524	0.088542	0.133358
2000	0.431251	0.235564	-0.04367	0.096591
2001	0.431538	0.238117	-0.03945	0.103146
2002	0.398226	0.232803	-0.03756	0.085701
2003	0.334195	0.190260	0.001133	0.052462
2004	0.342195	0.187181	0.009139	0.042524
2005	0.343488	0.178313	0.023968	0.035633
2006	0.298065	0.151402	0.047786	0.027109

TABLE VI
Verification of Markov property: router-level topology.

Network	clustering coefficient		2nd assortativity coefficient	
	real	Markov	real	Markov
AT&T	0.017112	0.003581	0.225254	0.127726
Sprintlink	0.057527	0.007260	0.255543	0.150171
Verio	0.119402	0.008947	0.398492	0.128577
Telstra	0.054640	0.009537	0.008529	0.064626
Level3	0.191861	0.062740	0.328257	0.038356
Abovenet	0.285643	0.037228	0.088277	0.040090
Tiscail	0.039277	0.022826	0.073944	0.010652
Exodus	0.292335	0.020878	0.269113	0.047952
Ebone	0.175320	0.031574	0.137128	0.039227
VSNL	0.058946	0.042803	0.315413	0.111277

overestimates the average degree of a wedge-based sampled node. In router-level topologies, the Markov property yields very accurate estimates of the cumulative degree distribution under the wedge-based sampling.

We evaluate the average degree of a wedge-based sampled node ($E_\wedge[D]$) based on $\{p_e(k,l)\}_{k,l\in\mathbb{N}}$ of the original topology using the Markov property, and compare the result with the real average degree of the original topology. The results of the AS-level topologies are summarized in Table III. In the table, we also show the average degrees of nodes chosen by the normal sampling ($E[D]$) and the ones chosen by the edge-based sampling ($E_e[D]$). The edge-based sampling has a tendency to choose large-degree nodes, so $E_e[D]$ is much larger than $E[D]$. The wedge-based sampling also has a tendency to choose large-degree nodes and thus $E_\wedge[D] > E[D]$, while the wedges are likely to have small degree nodes at their ends and thus $E_\wedge[D] < E_e[D]$. As we expect, the Markov property overestimates the average degree of a wedge-based sample node. That is, wedges in the AS-level topologies have smaller degree nodes at their ends than the expectation by the Markov property.

Table IV summarizes the results of the router-level topologies. The Markov property yields accurate estimates of the average degree of wedge-based-sampled nodes.

In summary, in terms of the degree distribution under wedge-based sampling, the router-level topologies are more Markovian than the AS-level topologies.

C. Clustering coefficient and 2nd-order assortativity

Next, we evaluate the clustering and the 2nd-order assortativity coefficients based on $\{p_e(k,l)\}_{k,l\in\mathbb{N}}$ of the original topology (one of the AS-level and router-level topologies) using the Markov property, and compare them with the actual ones of the original topology. Table V summarizes the results for the AS-level topologues. The Markov property expects smaller clustering coefficients than those of the real AS-level topologies, while it expects larger values for the 2nd-order assortativity coefficient than the real topologies.

Table VI summarizes the results for the router-level topologies. Both the clustering and the 2nd-order assortativity coefficients of the router-level topologies are larger than the expectations by the Markov property. The router-level topologies show structural diversity; Exodus or Ebone is found to be nearly Markovian in the sense that its clustering and the 2nd-order assortativity coefficients are comparatively close to the expectations by the Markov property, but these metrics of Sprintlink and Verio are far from the expectations by the Markov property.

The clustering coefficients of the AS and router-level topologies are plotted in Fig. 6 where the x-axis shows the real data and the y-axis shows the expectation by the Markov property. Most of the real topologies used in the analysis exhibit larger clustering coefficients than the expectation of the Markov property. These results are not surprising because in general the uncorrelated network has small clustering coefficient and the Markovian network is the simplest generalization of the uncorrelated network. In terms of the clustering and the 2nd-order assortativity coefficients, the AS-level topologies are comparatively closer to the Markovian network than the router-level topologies.

The 2nd-order assortativity coefficients of the AS and the

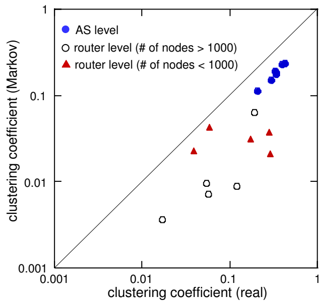

Fig. 6. Verification of Markov property: clustering coefficient.

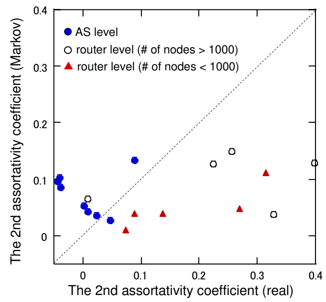

Fig. 7. Verification of Markov property: 2nd-order assortativity.

router-level topologies are plotted in Fig. 7. The AS-level topologies show smaller values while the router-level topologies show larger values than the expectation of the Markov property. Overall, the AS and the router-level topologies are not Markovian with respect to the clustering coefficient and the 2nd-order assortativity.

VI. Conclusion

We propose the Markov property for correlated random networks to find maximally unbiased networks under the constraint of a prescribed two-point degree correlation. The topological characteristics of the Markovian networks are fully characterized solely by the two-point degree correlation. We theoretically investigate the topological characteristics of Markovian networks and derive the analytical formulas for their graph theoretical metrics. A comparative analysis of AS- and router-level topologies shows that the studied topologies are not Markovian. This finding indicates that it is rather difficult to capture the topological characteristics of the Internet either at AS- or at router-level solely by the input of the two-point degree correlation.

There are several possible reasons why the real AS- and router-level topologies are not Markovian. The most promising one is the existence of hidden parameters: each node has some hidden parameters (other than degree), which influences the interconnectivity between nodes. For example, the locations of nodes in the real seem to be typical hidden parameters because they would affect the interconnectivity. Estimating the hidden parameters of nodes based on the difference from the expectation by the Markov property would be an interesting subject.

References

[1] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," in *ACM SIGCOMM*, 2006.
[2] C. Jin, Q. Chen, and S. Jamin, "Inet: Internet topology generator," University of Michigan, Tech. Rep. CSE-TR443-00, 2000.
[3] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.
[4] R. Albert and A.-L. Barabasi, "Topology of evolving networks: local events and universality," *Phys. Rev. Lett.*, vol. 85, pp. 5234–5237, 2000.
[5] T. Bu and D. Towsley, "On distinguishing between Internet power law topology generators," in *IEEE INFOCOM*, 2002.
[6] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in *STOC*, 2000.
[7] P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat, "Orbis: Rescaling degree correlations to generate annotated Internet topologies," in *ACM SIGCOMM*, 2007.
[8] G. Caldarelli *et al.*, "Scale-free networks from varying vertex intrinsic fitness," *Phys. Rev. Lett.*, vol. 89, no. 25, 2002.
[9] M. Boguna and R. Pastror-Satorras, "Class of correlated random networks with hidden variables," *Phys. Rev. E*, vol. 68:036112, 2003.
[10] V. Servedio, M. Boguna, and P. Butta, "Vertex intrinsic fitness: how to produce arbitrary scale-free networks," *Phys. Rev. E*, vol. 70:056126, 2004.
[11] N. Masuda, H. Miwa, and N. Konno, "Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights," *Phys. Rev. E*, vol. 70:036124, 2004.
[12] S. Weber and M. Porto, "Generation of arbitrarily two-point-correlated random networks," *Phys. Rev. E*, vol. 76:046111, 2007.
[13] M. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, no. 20, 2002.
[14] M. Catanzaro, M.Boguna, and R.P.-Satorras, "Generation of uncorrelated random scale-free networks," *Phys. Rev. E*, vol. 71:027103, 2005.
[15] M. Newman, "Random graphs as models of networks," in *Handbook of Graphs and Networks: From the Genome to the Internet*, S. Bornholdt and H. Schuster, Eds. WILEY, 2003.
[16] F. Chung and L. Lu, "Connected components in random graphs with given expected degree sequences," *Annals of Combinatorics*, vol. 6, pp. 124–145, 2002.
[17] E. Bender and E. Canfield, "The asymptotic number of labeled graphs with given degree sequences," *J. Comb. Theory*, vol. A24, pp. 296–307, 1978.
[18] N. Wormald, "Asymptotic connectivity of labelled regular graphs," *J. Combinat. Theory*, vol. B31, pp. 156–167, 1981.
[19] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures and Algorithms*, vol. 6, pp. 161–179, 1995.
[20] ——, "The size of the giant component of a random graph with a given degree sequence," *Combinatorics, Probability and Computing*, vol. 7, pp. 295–305, 1998.
[21] T. Luczak, "Sparse random graphs with a given degree sequence," *Random Graphs*, vol. 2, pp. 165–182, 1992.
[22] P. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of growing random networks," *Phys. Rev. Lett.*, vol. 85, pp. 4629–4632, 2000.
[23] M. Newman, S. Strogatz, and D. Watts, "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E*, vol. 64:026118, 2001.
[24] M. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, pp. 167–256, 2003.
[25] M. Boguna and R.P.-Satorras, "Epidemic spreading in correlated complex networks," *Phys.Rev.E*, vol. 66:047104, 2002.
[26] S. Shioda, "On the existence and reconstruction of a simple graph with a given joint-degree sequence," *IEICE Technical Report, COMP2010-23*, pp. 53–60, 2010, (in japanese).
[27] Http://www.ripe.net.
[28] Http://www.cs.washington.edu/research/networking/rocketfuel.