Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Neural Networks 17 (2004) 975–988
Neural Networks
www.elsevier.com/locate/neunet

Modeling of growing networks with directional attachment and communities

Masahiro Kimura*, Kazumi Saito[1], Naonori Ueda[2]

NTT Communication Science Laboratories, 2-4 Hikaridai, Seika-cho, Kyoto 619-0237, Japan

Received 9 December 2002; revised 21 January 2004; accepted 21 January 2004

Abstract

In this paper, we propose a new network growth model and its learning algorithm to more precisely model such a real-world growing network as the Web. Unlike the conventional models, we have incorporated directional attachment and community structure for this purpose. We show that the proposed model exhibits a degree distribution with a power-law tail, which is an important characteristic of many large-scale real-world networks including the Web. Using real Web data, we experimentally show that predictive ability can be improved by incorporating directional attachment and community structure. Also, using synthetic data, we experimentally show that predictive ability can definitely be improved by incorporating community structure.

Keywords: Network growth models; Scale-free models; Communities; Learning; Prediction; Web dynamics

1. Introduction

The World-Wide Web provides a vast repository of information and continues to grow as an important new medium of communication. From the scientific and technological points of view, investigating the Web has become an important and challenging research issue (Albert & Barabási, 2002; Chakrabarti et al., 1999; Dorogovtsev & Mendes, 2002; Kleinberg & Lawrence, 2001; Strogatz, 2001). The Web data set is huge, combines many kinds of features, and changes over time. When mining and modeling this rich collection of data, various problems for learning can be posed. Namely, the Web can provide an interesting new genre of learning problems and also establish a new research field of complex intelligent systems. Therefore, investigating the Web is drawing the attention of the research community involved in Neural Networks as well (Pal, Talwar, & Mitra, 2002).

The pages and hyperlinks of the Web can be viewed as nodes (vertices) and links (edges) of a network (directed graph). This network (graph) structure is useful from various points of view. For example, several algorithms for identifying 'authoritative' or 'influential' pages from hyperlink structures have been proposed to improve Web search engines (Brin & Page, 1998; Kleinberg, 1998; Ng, Zheng, & Jordan, 2001). Also, Web hyperlink structures can be viewed as *social networks*. Thus, from the viewpoint of the social science, it is important to analyze these social networks and understand the ecology of the Web, since these investigations can elucidate, for example, communications among members of a group and economic transactions among corporations (Wasserman & Faust, 1994). On the other hand, the Web constantly grows through the addition of new pages and hyperlinks created by users with their particular interests, and hence it has become a growing network. Recently, considerable attention has been devoted to exploring real-world complex networks, and modeling the growth processes of those networks is becoming one of the most important research issues (Albert & Barabási, 2002; Dorogovtsev & Mendes, 2002; Strogatz, 2001). In this paper, we propose a new network growth model and its learning algorithm to more precisely model such a real-world growing network as the Web. Unlike the conventional models, we have incorporated *directional attachment* and *community structure* for this purpose.

* Corresponding author. Tel.: +81-774-93-5131; fax: +81-774-93-5155.
E-mail addresses: kimura@cslab.kecl.ntt.co.jp (M. Kimura); saito@cslab.kecl.ntt.co.jp (K. Saito); ueda@cslab.kecl.ntt.co.jp (N. Ueda).
[1] Tel.: +81-774-93-5137; fax: +81-774-93-5155.
[2] Tel.: +81-774-93-5130; fax: +81-774-93-5155.

0893-6080/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2004.01.005

In Section 2, we briefly review the previous work on growing network models and give an overview of the basic idea of the proposed model with a description of the related work. In Section 3, we discuss probabilistic models of growing networks and explain the conventional models in detail. In Section 4, we describe the proposed model in detail, and formally show that the proposed model also exhibits a *degree distribution* with a power-law tail, which is one of the most notable characteristics of the graph structures observed in many real-world growing networks including the Web. This result is also numerically confirmed in one of the experiments in Section 6. In Section 5, we give a learning algorithm for the proposed model. In Section 6, we experimentally show that both directional attachment and community structure are quite effective for modeling such a growing network as the Web.

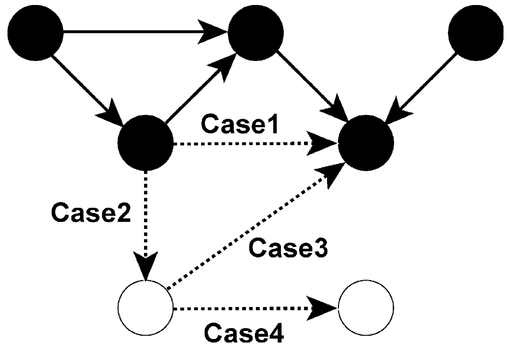

Fig. 1. Example of the four cases of new link creation. Filled circles, unfilled circles, solid arrows, and dashed arrows, respectively, indicate old nodes, new nodes, old links, and new links in a growing network. Cases 1, 2, 3, and 4, respectively, indicate the cases where a new link is created from an old node to an old node, from an old node to a new node, from a new node to an old node, and from a new node to a new node.

2. Overview

2.1. Previous work

A fundamental characteristic of any network is the degree distribution $F(d)$, which represents the fraction of the number of nodes that have d links in the network. Empirical results show that for many large-scale real-world networks including the Web, the degree distributions do not follow Poisson distributions, which the classical random graph theory of Erdös and Rényi expects, but possess power-law tails (Albert & Barabási, 2002; Barabási & Albert, 1999; Broder et al., 2000; Dorogovtsev & Mendes, 2002).[3] These facts suggest that the growing network model for the Web must at least satisfy the following conditions.

- Its growth process must not be completely random but obey certain self-organization principles.
- After it has sufficiently grown, the degree distribution of the resulting network must have a power-law tail.

Barabási and Albert (1999) discovered a growing network model satisfying these conditions. The principal ingredient of their model is a mechanism of *preferential attachment*, in which the probability that an existing node gains a new link is proportional to the number of links it currently has. Some variants of the Barabási–Albert (BA) model have been presented (see, e.g. Albert & Barabási, 2002; Dorogovtsev & Mendes, 2002). In particular, by introducing *mixtures of preferential and uniform attachment*, Pennock, Flake, Lawrence, Glover, and Giles (2002) more accurately accounted for the degree distributions of the networks of category-specific Web pages, the Web as a whole, movie actor collaborations, the western United States electrical power grid, and scholarly citations than the BA model. Since a system with a power-law is known to have a scale-free nature, these growing network models are generally referred to as *scale-free models*.

2.2. The basic idea of the proposed model

With the aim of constructing a more precise model of such a real-world growing network as the Web, we propose a new network growth model that incorporates

- directional attachment
- community structure

into an existing scale-free model.

When a new link is created, the following four cases can happen (cf. Fig. 1). It is attached from an old node to an old node (case 1), from an old node to a new node (case 2), from a new node to an old node (case 3), or from a new node to a new node (case 4). Each growing network has its own bias

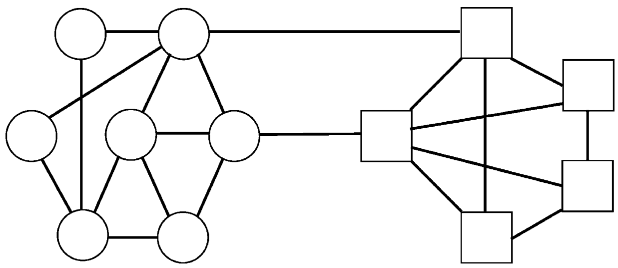

Fig. 2. Example of a network with community structure. In this network, there are two communities whose nodes are, respectively, indicated by circles and squares. Solid lines indicate the undirected links of the network. This example illustrates that communities are subsets of nodes within which the undirected links are dense, but between which the undirected links are less dense.

[3] Other structural characteristics of the Web graph are also reported (see, e.g. Albert & Barabási, 2002; Dorogovtsev & Mendes, 2002). For example, the distribution of the number of pages per site exhibits a power-law (Huberman & Adamic, 1999), and the average shortest-path length is relatively small despite its huge size (Albert, Jeong, & Barabási, 1999; Broder et al., 2000). However, in this paper, we focus on the characteristic degree distribution since it is widely understood, and has a simple and explicit expression.

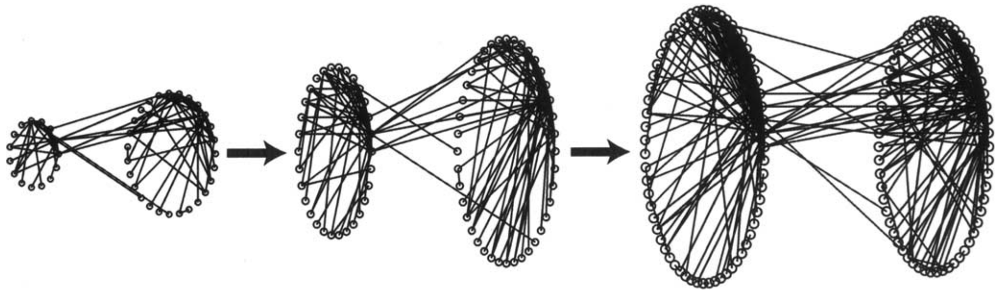

Fig. 3. Example of a growing network with community structure. In this example, a growth process of a network with two communities is displayed. The arrows indicate the time direction. As time goes on, the numbers of nodes and links increase. In particular, the links within each community increase while the links between communities remain sparse. Thus, the network grows as two clusters are formed.

for these four cases. Namely, the probabilities that cases 1, 2, 3 and 4 happen can change depending on the growing network to be modeled. For example, in citation networks, a new link is always created from a new node to an old node. The mechanism that appropriately biases these four cases in a new link creation is referred to as *directional attachment*. To precisely model given growing networks, it is necessary to incorporate the directional attachment proper to them. However, the existing scale-free models do not take into account directional attachment.

A *community* is defined as a collection of nodes in which each member node has more links to nodes within the community than to nodes outside the community (Flake, Lawrence, & Giles, 2000; Girvan & Newman, 2002; Kleinberg & Lawrence, 2001). For example, communities in a citation network might represent related papers on a single topic. Namely, the community structure of a network is defined by the decomposition of the set of its nodes into the clusters that arise from its undirected graph structure (cf. Fig. 2). One characteristic of the Web is the existence of community structure, and the Web grows as various clusters are formed (Chakrabarti et al., 1999; Eckmann & Moses, 2002; Flake et al., 2000; Girvan & Newman, 2002; Kleinberg & Lawrence, 2001; Kumar, Raghavan, Rajagopalan, & Tomkins, 1999). Namely, the following situation can often be observed in its growth process: an increasing number of links are created within each community while the links between communities remain sparse (cf. Fig. 3). Incorporating community structure enables us to model this sort of detailed growth process. However, the existing scale-free models also do not take into account community structure.

In attempts to identify community structure, there have been several investigations using graph-theoretic methods (Eckmann & Moses, 2002; Flake et al., 2000; Girvan & Newman, 2002; Kumar et al., 1999). Also, there have been some investigations using latent variable models such as PHITS (Cohn & Chang, 2000; Cohn & Hofmann, 2001), provided that the definition of community might be slightly changed. However, these investigations dealt with only static networks, that is, the number of nodes and links were not allowed to increase. Therefore, introducing community structure to growing network models may be a promising approach.

3. Scale-free models

Here, we discuss probabilistic models of growing networks. We assume that nodes and links do not disappear in the growth processes.[4]

At an arbitrary time $t \geq 0$, a growing network is represented by an adjacency matrix A_t whose (i,j)-element $A_t(i,j)$ is the number of links from node i to node j, where the size of A_t is regarded as being large enough, and $A_t(i,j) = 0$ if node i or node j does not exist in the network at time t. Let $\mathcal{N}_t$ denote the set of nodes in the growing network at time t and N_t denote the number of elements of $\mathcal{N}_t$. For any $v \in \mathcal{N}_t$, the degree $D_t(v)$ of v at time t is defined by the number of the links attaching to v for the growing network at time t, that is

$$D_t(v) = \sum_{i \in \mathcal{N}_t} \{A_t(v,i) + A_t(i,v)\}.$$

For any $t \geq 1$, we define the matrix ΔA_t of the link increments at time t as follows: if $i,j \in \mathcal{N}_{t-1}$, then the (i,j)-element $\Delta A_t(i,j)$ of ΔA_t is $A_t(i,j) - A_{t-1}(i,j)$, otherwise it is $A_t(i,j)$.

We suppose that the growth process of a network is described as a stochastic process $P(\Delta A_t | A_{t-1})$, $(t \geq 1)$. Our aim is to model the true growth process $P(\Delta A_t | A_{t-1})$, $(t \geq 1)$ based on a time-sequence of observed adjacency matrices. For this purpose, we first construct a network growth model $P(\Delta A_t | A_{t-1}, \theta)$, where θ denotes the set of model parameters. Next, we acquire an optimal model $P(\Delta A_t | A_{t-1}, \hat{\theta})$ for the true growth process by learning the observed data.

In a scale-free model, given an adjacency matrix A_{t-1} at time $t-1$, the probability $P(\Delta A_t | A_{t-1}, \theta)$ that the matrix of link increments at time t is ΔA_t is assumed to be given by

[4] For example, this assumption is always true for citation networks. For the Web, this assumption can also be regarded as almost true if we deal with its short-term changes.

the following multinomial distribution

$$P(\Delta A_t|A_{t-1},\theta) \propto \prod_{u_t,v_t} P([u_t,v_t]|A_{t-1},\theta)^{\Delta A_t(u_t,v_t)}, \quad (1)$$

where $P([u_t,v_t]|A_{t-1})$ indicates the probability that a new link at time t, denoted by $[u_t,v_t]$, from an originating node u_t to a target node v_t is added to the network represented by A_{t-1}.

In the BA model (Barabási & Albert, 1999), at each time t a new node is introduced and it immediately attaches to some of the preexisting nodes. The probability $P(v_t|A_{t-1})$ of choosing node v_t as a target node at time t given A_{t-1} is defined by the fraction of the number of links that node v_t has at time $t-1$, that is

$$P(v_t|A_{t-1}) = \frac{D_{t-1}(v_t)}{\sum_{i\in\mathcal{N}_{t-1}} D_{t-1}(i)}.$$

This is a mechanism of preferential attachment. Then

$$P([u_t,v_t]|A_{t-1}) = P(u_t|A_{t-1})P(v_t|A_{t-1}),$$

and the probability $P(u_t|A_{t-1})$ of choosing node u_t as an originating node at time t given A_{t-1} is 1 iff u_t is the new node. Note that the BA model does not have any control parameters and self-links are allowed.

In Pennock et al.'s (2002) (PFLGG) model, two parameters, α and β ($0 \le \alpha,\ \beta \le 1$), are introduced and self-links are not allowed. At every time-step, a new node and some new links are added to the current network, where both endpoints of a new link are chosen according to a mixture of preferential and uniform attachment. That is, $P([u_t,v_t]|A_{t-1},\alpha,\beta)$ is defined by

$$P([u_t,v_t]|A_{t-1},\alpha,\beta) = P(u_t|A_{t-1},\alpha)P(v_t|u_t,A_{t-1},\beta),$$

and $P(u_t|A_{t-1},\alpha)$ is defined by

$$P(u_t|A_{t-1},\alpha) = \alpha\frac{D_{t-1}(u_t)}{\sum_{i\in\mathcal{N}_{t-1}} D_{t-1}(i)} + (1-\alpha)\frac{1}{N_t},$$

and $P(v_t|u_t,A_{t-1},\beta)$ is defined by

$$P(v_t|u_t,A_{t-1},\beta) = \beta\frac{D_{t-1}(v_t;u_t)}{\sum_{i\in\mathcal{N}_{t-1}\setminus\{u_t\}} D_{t-1}(i;u_t)} + (1-\beta)\frac{1}{N_t-1},$$

where $D_{t-1}(i;u_t)$ denotes the degree of node i for the network obtained by eliminating node u_t from the growing network at time $t-1$, that is

$$D_{t-1}(i;u_t) = \sum_{j\in\mathcal{N}_{t-1}\setminus\{u_t\}} \{A_{t-1}(i,j) + A_{t-1}(j,i)\}.$$

4. The proposed model

We propose a network growth model incorporating two new mechanisms, directional attachment and community structure, into a scale-free model. From the existing scale-free models, we borrow the following two ideas: one is the idea of the BA model that new nodes are always introduced with new links, and the other is the idea of the PFLGG model that new links are generated according to a mixture of preferential and uniform attachment. In the proposed model, self-links are not allowed.[5]

4.1. Directional attachment

We incorporate the directional attachment by introducing a set of control parameters, $\eta = \{\eta_{00}, \eta_{01}, \eta_{10}, \eta_{11}\}$, as shown in Table 1(a), where for a new link creation, η_{00} denotes the probability that both the originating and target nodes are old, η_{01} denotes the probability that the originating node is old and the target node is new, η_{10} denotes the probability that the originating node is new and the target node is old, and η_{11} denotes the probability that both the originating and target nodes are new. That is, we introduce the fully correlated model (Table 1(a)) for directional attachment. To clarify the effectiveness of incorporating directional attachment, we also consider the independent model shown in Table 1(b) for comparison (cf. Section 6.3.2). That is, we consider the following special case. For a new link creation, whether the target node is new or old is independent of whether the originating node is new or old. In Table 1(b), α_0 indicates the probability that a new node is chosen as the originating node of a new link, and β_0 denotes the probability that a new node is chosen as the target.

In comparing the BA and PFLGG models for directional attachment, they can be represented by the 2×2 matrices of probabilities shown in Table 2 at time t. For example, in the BA model, the probability that a new link is attached from a new node to an old node is one, and the other probabilities are zero. Namely, the previous scale-free models have smaller degrees of freedom than the proposed model for directional attachment.

4.2. Community structure

Suppose that the set of community-labels is

$$\{z^1, \ldots, z^K\},$$

that is, there exist K communities, and each node belongs to only one community without changing the community during the studied period. We also assume that each community has its own growth dynamics, and in the growth process, the links within a community are more frequently generated than those between different communities.

[5] From the viewpoint of designing Web search engines, it is significant to know the authoritative Web pages concerning a topic based on the link structure (Brin & Page, 1998; Chakrabarti et al., 1999; Kleinberg, 1998). In this respect, self-links should be neglected.

Table 1
Directional attachment in proposed model

	Old node	New node
(a) *Fully correlated model*		
Old node	η_{00}	η_{01}
New node	η_{10}	η_{11}
(b) *Independent model*		
Old node	$(1-\alpha_0)(1-\beta_0)$	$(1-\alpha_0)\beta_0$
New node	$\alpha_0(1-\beta_0)$	$\alpha_0\beta_0$

We incorporate communities into a scale-free model as latent variables.

Here, we introduce some notations. Consider a growing network. We denote by A_t the adjacency matrix of the growing network at time t. Let $\mathcal{N}_t^k$ denote the set of nodes that belong to community z^k in the growing network at time t and N_t^k denote the number of elements of $\mathcal{N}_t^k$. Note that $\mathcal{N}_t^k$ and N_t^k depend on A_t. Let $D_t^k(v)$ denote the degree of node v for community z^k in the growing network at time t; i.e.

$$D_t^k(v) = \sum_{i \in \mathcal{N}_t^k} \{A_t(v,i) + A_t(i,v)\}. \quad (2)$$

Also, for any $u, v \in \mathcal{N}_t$ (the set of nodes in the growing network at time t), let $D_t^k(v;u)$ denote the degree of node v for community z^k in the network obtained by eliminating node u from the growing network at time t; i.e.

$$D_t^k(v;u) = \sum_{i \in \mathcal{N}_t^k \setminus \{u\}} \{A_t(v,i) + A_t(i,v)\}. \quad (3)$$

4.3. A network growth model

The proposed model is characterized by a stochastic process $P(\Delta A_t | A_{t-1}, \theta)$, which has the same form as Eq. (1). In particular, m_t new links are added at each time t. Given the adjacency matrix A_{t-1} of the network at time $t-1$, a new link $[u_t, v_t]$ at time t is generated in the following way.

Table 2
Directional attachment in conventional models

	Old node	New node
(a) *BA model*		
Old node	0	0
New node	1	0
(b) *PFLGG model*		
Old node	$\left(1-\frac{1-\alpha}{N_t}\right) \cdot \left(1-\frac{1-\beta}{N_t-1}\right)$	$\left(1-\frac{1-\alpha}{N_t}\right) \cdot \frac{1-\beta}{N_t-1}$
New node	$\frac{1-\alpha}{N_t}$	0

First, z^k is chosen with probability ξ^k as the community to which the originating node u_t of the new link belongs. Next, z^l is chosen with probability γ^{kl} as the community to which the target node v_t belongs. Finally, link $[u_t, v_t]$ is created with probability $P([u_t,v_t]|[z^k,z^l], A_{t-1}, \theta)$. Namely, taking into account the community structure, the probability $P([u_t,v_t]|A_{t-1},\theta)$ that link $[u_t,v_t]$ is newly added given A_{t-1} is defined by

$$P([u_t,v_t]|A_{t-1},\theta) = \sum_{k,l=1}^{K} \xi^k \gamma^{kl} P([u_t,v_t]|[z^k,z^l],A_{t-1},\theta). \quad (4)$$

Note here that the proposed model is different from the aspect model (PHITS) (Cohn & Chang, 2000; Hofmann, 1999) as a latent variable model. In particular, our model becomes a generative model, and a community-label is directly assigned to each individual node.

After the communities $[z^k, z^l]$ of u_t and v_t are given, whether u_t and v_t are new or old is decided according to the directional attachment

$$\eta^{kl} = \{\eta_{00}^{kl}, \eta_{01}^{kl}, \eta_{10}^{kl}, \eta_{11}^{kl}\}$$

as shown in Table 1(a). For example, with probability η_{11}^{kl}, both u_t and v_t are new nodes.

If both u_t and v_t are new nodes, a new node of community z^k and a new node of community z^l are created with probability 1. Namely, the probability $P([u_t,v_t]|[z^k,z^l], A_{t-1},\theta)$ that link $[u_t,v_t]$ is newly added given $[z^k,z^l]$ and A_{t-1} is defined by

$$P([u_t,v_t]|[z^k,z^l],A_{t-1},\theta) = \eta_{11}^{kl}. \quad (5)$$

If u_t is an old node and v_t is a new node, the probability of choosing u_t is defined by a mixture of preferential and uniform attachment within community z^k, and a new node of community z^l is created with probability 1. Namely, the probability $P([u_t,v_t]|[z^k,z^l], A_{t-1},\theta)$ is defined by

$$P([u_t,v_t]|[z^k,z^l],A_{t-1},\theta) = \eta_{01}^{kl} \left\{ \alpha^k \frac{D_{t-1}^k(u_t)}{\sum_{i \in \mathcal{N}_{t-1}^k} D_{t-1}^k(i)} + (1-\alpha^k)\frac{1}{N_{t-1}^k} \right\}, \quad (6)$$

(cf. Eq. (2)), where $0 \le \alpha^k \le 1$.

Also, if u_t is a new node and v_t is an old node, a new node of community z^k is created with probability 1, and the probability of choosing v_t is defined by a mixture of preferential and uniform attachment within community z^l. Namely, the probability $P([u_t,v_t]|[z^k,z^l], A_{t-1},\theta)$ is defined by

$$P([u_t,v_t]|[z^k,z^l],A_{t-1},\theta) = \eta_{10}^{kl} \left\{ \beta^l \frac{D_{t-1}^l(v_t)}{\sum_{i \in \mathcal{N}_{t-1}^l} D_{t-1}^l(i)} + (1-\beta^l)\frac{1}{N_{t-1}^l} \right\}, \quad (7)$$

(cf. Eq. (2)), where $0 \le \beta^l \le 1$.

If both u_t and v_t are old nodes, the probability of choosing u_t is defined by a mixture of preferential and uniform attachment within community z^k, and the probability of choosing v_t is defined by a mixture of preferential and uniform attachment within community z^l excepting u_t. Namely, the probability $P([u_t, v_t]|[z^k, z^l], A_{t-1}, \theta)$ is defined by

$$P([u_t, v_t]|[z^k, z^l], A_{t-1}, \theta) = \eta_{00}^{kl}\left\{\alpha^k \frac{D_{t-1}^k(u_t)}{\sum_{i \in \mathscr{N}_{t-1}^k} D_{t-1}^k(i)} + (1-\alpha^k)\frac{1}{N_{t-1}^k}\right\} \cdot \left\{\beta^l \frac{D_{t-1}^l(v_t; u_t)}{\sum_{i \in \mathscr{N}_{t-1}^l \setminus \{u_t\}} D_{t-1}^l(i; u_t)} + (1-\beta^l)\frac{1}{N_{t-1}^l - \delta_{k,l}}\right\}, \tag{8}$$

(cf. Eqs. (2) and (3)), where $0 \le \alpha^k, \beta^l \le 1$, and $\delta_{k,l}$ is Kronecker's delta.

Hence, a generative model of growing networks has been constructed. Here, the set θ of model parameters becomes

$$\theta = \{\alpha^k, \beta^l, \eta_{\varepsilon\rho}^{kl}, \xi^k, \gamma^{kl};\ \ k, l = 1, \ldots, K,\ \ \varepsilon, \rho = 0, 1\},$$

where $0 \le \alpha^k,\ \beta^l,\ \eta_{\varepsilon\rho}^{kl},\ \xi^k,\ \gamma^{kl} \le 1$, and $\sum_{\varepsilon,\rho} \eta_{\varepsilon\rho}^{kl} = \sum_k \xi^k = \sum_l \gamma^{kl} = 1$. We assume that m_t is given.

Although we used the mechanism of preferential attachment in the manner of undirected graphs, we can use a more detailed preferential attachment mechanism by taking into account a directed graph nature and decomposing the degree of a node into its in-degree and its out-degree (Krapivsky, Rodgers, & Redner, 2001). Then, we can obtain a more flexible model of growing networks. Note that we can also develop similar analyses and learning algorithms for this extended model.

4.4. Degree distributions

We investigate the degree distributions of the growing networks generated by the proposed model.

As Barabási and Albert (1999) pointed out, the mechanisms of growth and preferential attachment can be considered as the origin to yield a degree distribution with a power-law tail. These mechanisms are also incorporated in the proposed model. Thus, it can be considered that our model also exhibits a degree distribution with a power-law tail.

There are various approaches to formally prove the power-law dependence of the degree distribution of a growing network (Albert & Barabási, 2002; Dorogovtsev & Mendes, 2002). We can give a formal proof of the following proposition using the master equation approach with the continuous approximation of Dorogovtsev and Mendes (2001) (see Appendix A).

Proposition 1. *Suppose that the values of all parameters of the proposed model are not zero, and $m_t = m_0$ for any $t > 1$. We arbitrarily fix an initial network and consider the growing network generated by the proposed model from the initial network. Let $F_t(d)$ be the degree distribution of the growing network at time t. Then, there exists a positive constant ν such that*

$$F_t(d) \propto d^{-\nu} \text{ as } t \to \infty,\ \ d \to \infty.$$

Hence, the proposed model generically exhibits a degree distribution with a power-law tail after it has sufficiently grown.

5. Learning for growing networks

Let $\{A_0, A_1, \ldots, A_T\}$ be the observed time-sequence of adjacency matrices of a growing network. Our task is to estimate the set θ of model parameters from these data. Although the latent variable corresponding to node i should be probabilistically estimated as $P(z^k|i \in \mathscr{N}_t, A_t)$, for simplicity, we deterministically estimate it by the following hard-clustering procedure.

5.1. Clustering

First, we perform clustering for the network of adjacency matrix A_T (that is, the last observed network) to assign the community-label to each node of the network.[6] Then, $P(z^k|i \in \mathscr{N}_T, A_T)$ is 1 for only one k and 0 for the others, and we consider $P(z^k|i \in \mathscr{N}_{t-1}, A_{t-1}) = P(z^k|i \in \mathscr{N}_T, A_T)$ for $t = 1, \ldots, T$. Several methods (e.g. Flake et al., 2000; Girvan & Newman, 2002) can be used for this clustering. In our experiments, we made the network undirected, added self-links to the network, and used the K-means clustering algorithm based on the Kullback–Leibler divergence (see Appendix B).

5.2. Parameter estimation

Next, we estimate θ from the observed data based on the clustering result. Let $\{([u_t^\lambda, v_t^\lambda]; m_t^\lambda);\ \lambda = 1, \ldots, r_t\}$ be the set of links added newly at time $t \ge 1$, where $([u_t^\lambda, v_t^\lambda]; m_t^\lambda)$ means that link $[u_t^\lambda, v_t^\lambda]$ is added m_t^λ times, that is, $\Delta A_t(u_t^\lambda, v_t^\lambda) = m_t^\lambda$. Note that we know the communities of each u_t^λ and each v_t^λ by the clustering result. We also know whether or not each u_t^λ and each v_t^λ were new nodes from the observed data $\{A_t; t = 0, 1, \ldots, T\}$.

First, let us estimate the parameters $\{\xi^k\}$, $\{\gamma^{kl}\}$, and $\{\eta_{\varepsilon\rho}^{kl}\}$. Consider all the links added for the observed period.

[6] First, community structure is defined by the cluster structure of the network. It is also assumed that on the growth process, links are more frequently generated within each community than those between different communities. Therefore, we can consider that by clustering the most recent network, it is fairly possible to determine the community-label of each node of the network.

Let m denote the total number of these links, that is, $m = \sum_{t=1}^{T} \sum_{\lambda=1}^{r_t} m_t^\lambda$. ξ^k is estimated as n_1^k/m, where n_1^k is the number of these links such that the originating node has label z^k. γ^{kl} is estimated as n^{kl}/n_1^k, where n^{kl} is the number of these links such that the originating node has label z^k and the target node has label z^l. Similarly, we can also empirically estimate $\{\eta_{\varepsilon\rho}^{kl}\}$.

It suffices to estimate the parameters $\varphi = \{\alpha, \beta\}$, where $\alpha = \{\alpha^k\}$ and $\beta = \{\beta^l\}$. We perform the maximal likelihood estimation. In this case, the log-likelihood function $\mathscr{L}(\varphi)$ is of the form

$$\mathscr{L}(\varphi) = \sum_{t=1}^{T} \log P(\Delta A_t | A_{t-1}, \varphi)$$

$$= \sum_{t=1}^{T} \sum_{\lambda=1}^{r_t} m_t^\lambda \log P([u_t^\lambda, v_t^\lambda] | A_{t-1}, \varphi) + \text{const.},$$

which is explicitly calculated by Eqs. (4)–(8).

We can efficiently estimate the optimal parameter values by using an iterative algorithm based on the EM algorithm. Let $\bar{\varphi}$ be the current estimate of φ. Then the update formula of φ can be obtained by maximizing the Q-function with respect to φ, which is defined by

$$Q(\varphi|\bar{\varphi}) = M \sum_{t=1}^{T} \sum_{\lambda=1}^{r_t} \sum_{k,l=1}^{K} m_t^\lambda q_t^{\lambda kl}(\bar{\varphi}) \log P([u_t^\lambda, v_t^\lambda], [z^k, z^l] | A_{t-1}, \varphi),$$

where

$$q_t^{\lambda kl}(\bar{\varphi}) = P([z^k, z^l] | [u_t^\lambda, v_t^\lambda], A_{t-1}, \bar{\varphi})$$

$$= \frac{\xi^k \gamma^{kl} P([u_t^\lambda, v_t^\lambda] | [z^k, z^l], A_{t-1}, \bar{\varphi})}{\sum_{k,l} \xi^k \gamma^{kl} P([u_t^\lambda, v_t^\lambda] | [z^k, z^l], A_{t-1}, \bar{\varphi})}$$

which is explicitly calculated by Eqs. (5)–(8).

We decompose as

$$Q(\varphi|\bar{\varphi}) = Q_1(\alpha|\bar{\varphi}) + Q_2(\beta|\bar{\varphi}) + \text{const.}$$

Our task is to calculate α and β, which maximize $Q_1(\alpha|\bar{\varphi})$ and $Q_2(\beta|\bar{\varphi})$, respectively. By Eqs. (5)–(8)

$$Q_1(\alpha|\bar{\varphi}) = \sum_{t=1}^{T} \sum_{\lambda=1}^{r_t} \sum_{k,l=1}^{K} m_t^\lambda q_t^{\lambda kl}(\bar{\varphi}) \chi_{t-1}(u_t^\lambda) \log(\alpha^k \pi_{t-1}^k(u_t^\lambda) + (1-\alpha^k)/N_{t-1}^k),$$

where $\chi_{t-1}(u_t^\lambda)$ is 1 if $u_t^\lambda \in \mathscr{N}_{t-1}$ (that is, u_t^λ is an old node) and 0 otherwise, and

$$\pi_{t-1}^k(u_t^\lambda) = \frac{D_{t-1}^k(u_t^\lambda)}{\sum_{i \in \mathscr{N}_{t-1}^k} D_{t-1}^k(i)}.$$

$Q_2(\beta|\bar{\varphi})$ can also be calculated in a similar way.

Next, we consider maximizing $Q_1(\alpha|\bar{\varphi})$ with respect to α. This is also performed by using an iterative algorithm based on the EM algorithm. Let $\tilde{\alpha}$ be the current estimate of α. We define $Q_1'(\alpha|\tilde{\alpha}, \bar{\varphi})$ by

$$Q_1'(\alpha|\tilde{\alpha}, \bar{\varphi}) = \sum_{t=1}^{T} \sum_{\lambda=1}^{r_t} \sum_{k=1}^{K} y_t^{\lambda k}(\bar{\varphi}) \cdot \frac{\tilde{\alpha}^k \log(\alpha^k) \pi_{t-1}^k(u_t^\lambda) + (1-\tilde{\alpha}^k) \log(1-\alpha^k)/N_{t-1}^k}{\tilde{\alpha}^k \pi_{t-1}^k(u_t^\lambda) + (1-\tilde{\alpha}^k)/N_{t-1}^k},$$

where $y_t^{\lambda k}(\bar{\varphi}) = m_t^\lambda \chi_{t-1}(u_t^\lambda) \sum_{l=1}^{K} q_t^{\lambda kl}(\bar{\varphi})$. Since by Jensen's inequality

$$Q_1(\alpha|\bar{\varphi}) - Q_1(\tilde{\alpha}|\bar{\varphi}) \geq Q_1'(\alpha|\tilde{\alpha}, \bar{\varphi}) - Q_1'(\tilde{\alpha}|\tilde{\alpha}, \bar{\varphi}),$$

the update formula of α can be obtained by maximizing $Q_1'(\alpha|\tilde{\alpha}, \bar{\varphi})$ with respect to α. Hence, it is obtained by

$$\alpha^k = \sum_{t=1}^{T} \sum_{\lambda=1}^{r_t} \frac{y_t^{\lambda k}(\bar{\varphi})}{\sum_{t'=1}^{T} \sum_{\lambda'=1}^{r_{t'}} y_{t'}^{\lambda' k}(\bar{\varphi})} \cdot \frac{\tilde{\alpha}^k \pi_{t-1}^k(u_t^\lambda)}{\tilde{\alpha}^k \pi_{t-1}^k(u_t^\lambda) + (1-\tilde{\alpha}^k)/N_{t-1}^k}.$$

In a similar way, we can obtain the update formula of β for maximizing $Q_2(\beta|\bar{\varphi})$. Hence, we have obtained an algorithm to calculate the set φ of unknown parameters by the maximal likelihood estimation.

6. Experimental evaluation

For simplicity, we focus on the case of undirected graphs in our experiments. Thus, we have $\alpha^k = \beta^k$, $\eta_{10}^{kl} = \eta_{01}^{kl}$, $(k, l = 1, \ldots, K)$.

6.1. Evaluation of degree distributions

We numerically confirm that the proposed model exhibits a degree distribution with a power-law tail.

Fig. 4 shows the degree distributions for three models and the network of the Web pages used in our next experiment (cf. Section 6.3): (a) is for the BA model, where five links are added at each time step, and (b) and (c) are for our models with $m_t = 5$, $\eta_{00}^{kl} = 81/100$, $\eta_{10}^{kl} = 9/100$, $\eta_{11}^{kl} = 1/100$, $\alpha_1^k = 4/5$, and $\alpha_2^k = 1/5$, where (b) is the case of a single community ($K = 1$), which can essentially be regarded as the PFLGG model, and (c) is the case of two communities ($K = 2$) with $\xi^1 = 1/5$, $\xi^2 = 4/5$ and $\gamma^{11} = \gamma^{22} = 4/5$, $\gamma^{12} = \gamma^{21} = 1/5$. (d) is for the real data of the Web. Fig. 4 numerically confirms that our model with two communities also exhibits a degree distribution with a power-law tail, as do the BA model, the model with a single community, and the network of the Web pages. This result supports Proposition 1.

6.2. Performance measure

Consider a growing network, and its observed data $\{A_0, A_1, \ldots, A_T\}$. Let $\hat{\theta}_K$ denote the set of the parameter

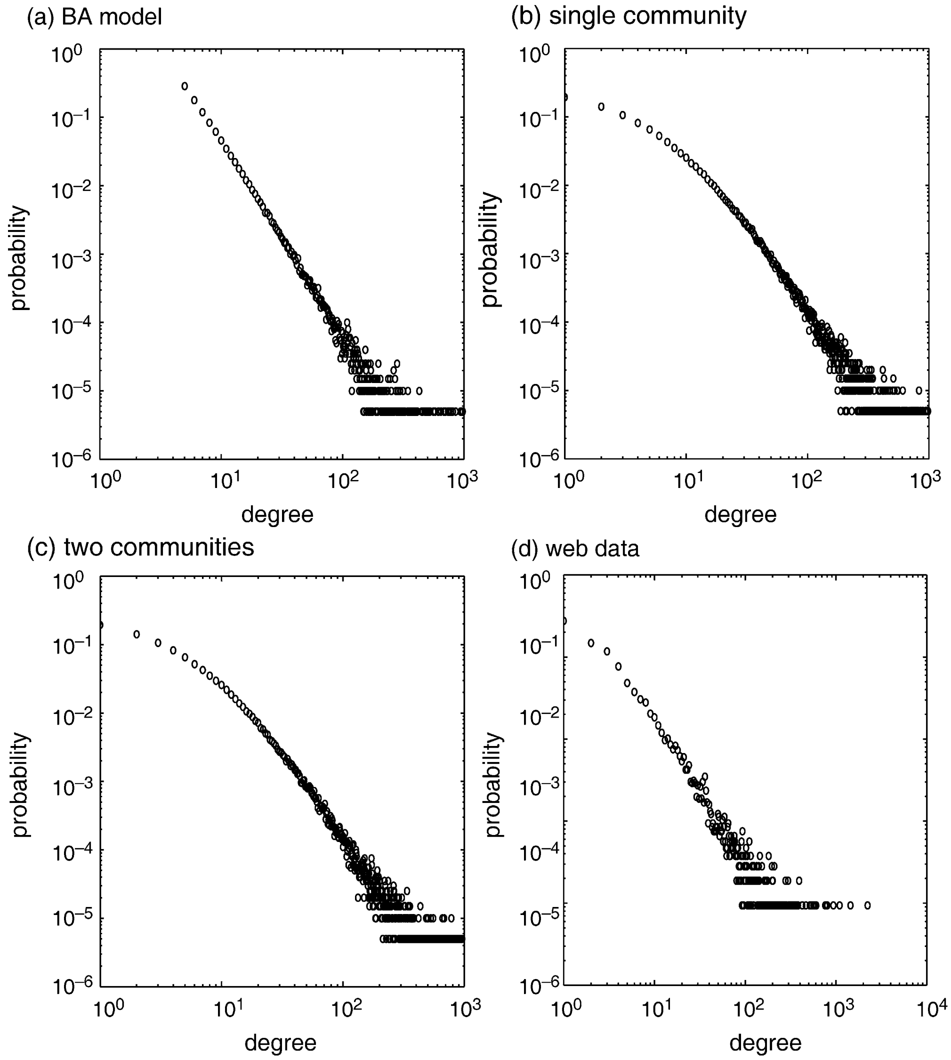

Fig. 4. Degree distributions.

values learned by the proposed model with K communities from this data. We evaluate the ability of the learned model with the prediction performance for dynamic probability matrix defined below.

We define the $(N_T+2)\times(N_T+2)$ matrix $\hat{\Gamma}_K = [\hat{\Gamma}_K(i,j)]$ by

$$\hat{\Gamma}_K(i,j) = P([i,j]|A_T, \hat{\theta}_K), \qquad (i,j = 1,\ldots,N_T+2).$$

$\hat{\Gamma}_K$ is called the *dynamic probability matrix* of the learned model at time T. Note that it represents the probability distribution for a new link creation given A_T. Let $\Gamma = [\Gamma(i,j)]$ denote the dynamic probability matrix of the actual process at time T. We evaluate the prediction performance of the learned model for dynamic probability matrix by the Kullback–Leibler divergence

$$I(\Gamma;\hat{\Gamma}_K) = \sum_{i,j} \Gamma(i,j)\log(\Gamma(i,j)/\hat{\Gamma}_K(i,j)).$$

6.3. Evaluation for real Web data

We evaluate the performance of the proposed model using a real-world growing network, that is, a growing network of Web pages concerning a broad topic.

6.3.1. Real Web data

Based on the method of Kleinberg (1998), we construct the network $G_t(\sigma)$ of the Web pages concerning a topic σ, at time t in the following way. At each time τ, we first collect the 200 highest-ranked pages for the query σ by using a text-based Web search engine. Next, we collect all of the pages linked from these pages, and up to 50 the pages that link these pages. Let $S_\tau(\sigma)$ denote the set of Web pages collected in this way. We define $G_t(\sigma)$ by the network induced on the Web pages in $\cup_{\tau=0}^{t} S_\tau(\sigma)$ at time t.

We consider the real-world growing network $G_t(\sigma)$, $(t \geq 0)$. In the experiment, 'mp3' was used as topic σ, and the time-interval was 1 month. Also, the observed time-sequence of the adjacency matrices were $\{A_0, A_1, A_2\}$. We used $\{A_0, A_1\}$ as the training data, that is, $T = 1$.

6.3.2. Performance evaluation

For the data of this real-world growing network, we investigated the effectiveness of incorporating directional attachment and community structure.

Let Model-0 be our independent model for directional attachment (cf. Section 4.1) and Model-1 be the proposed model, that is, the fully correlated model for directional attachment. Fig. 5 displays $I(\Gamma; \hat{\Gamma}_K)$ with respect to K for Model-0 and Model-1. Here, Γ was empirically calculated from A_1 and A_2.

Fig. 5 first shows that Model-1 could much more accurately predict the actual dynamic probability matrix than Model-0. This result implies that the prediction performance can be improved by incorporating directional attachment.

Fig. 5 also shows that although the prediction performance could be raised by increasing the number K of latent variables, an optimal number (11 in this case) of latent variables could exist. In particular, the proposed model incorporating community structure could more accurately predict the actual dynamic probability matrix than the model not incorporating it ($K = 1$). These results imply that the prediction performance can be improved by incorporating community structure.

Furthermore, it is important to identify the community structure of the network of Web pages on a particular topic from the viewpoint of information retrieval. We note that our method identified the community structure for such a network from time-sequence data without community information. Namely, in this experiment, the inferred community structure for $K = 11$ is regarded as an optimal community structure for this Web network. Incidentally, for the suggested communities, we in particular found the group of Japanese sites, the group of Korean sites and the group of Chinese sites.

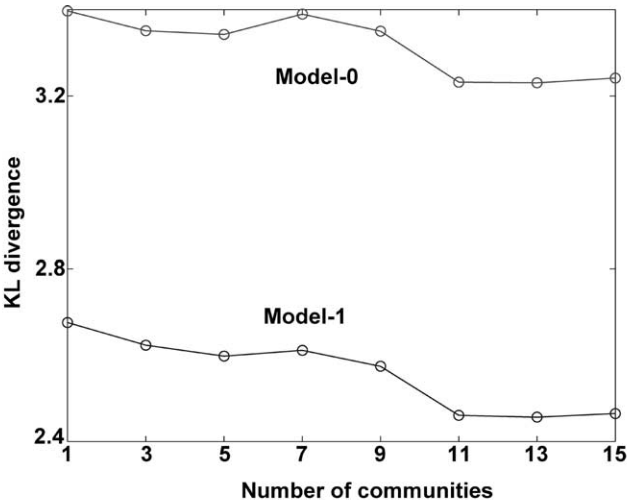

Fig. 5. Prediction performance of learned models.

6.4. Evaluation for synthetic data

Using synthetic data of a growing network with two communities, we investigate the effectiveness of incorporating community structure in further detail.

6.4.1. Synthetic data

We consider the following growing network, which has the same form as the proposed model except for having time-dependent $P([z^k, z^l]|t) = P(z^k|t)P(z^l|z^k, t)$.

- Five links are added at each time-step; i.e. $m_t = 5$, $(t \geq 0)$.
- The growing network has two communities; i.e. $K = 2$.
- $\alpha^1 (= \beta^1) = \alpha^2 (= \beta^2) = 8/9$, that is, the intensity of preferential attachment of community z^1 is the same as that of community z^2.
- The growth process is an independent model for directional attachment. Moreover, $\alpha_0^1 (= \beta_0^1) = \alpha_0^2 (= \beta_0^2) = 1/10$, that is, for the probability that a new node is chosen as an end node of a new link, community z^1 and community z^2 are identified.
- $P(z^1|z^1, t) = P(z^2|z^2, t) = 9/10$, $(t \geq 0)$, that is, for the probability that both end nodes of a new link belong to the same community, community z^1 and community z^2 are identified.

$$P(z^1|t) = 1 \Big/ \left\{ 1 + \exp\left(1 - \frac{t}{5000} \right) \right\}, \qquad (t \geq 0),$$

that is, when an end node of a new link is chosen, there is the difference between the probability that it is chosen from community z^1 and the probability that it is chosen from community z^2. In particular, $P(z^1|t)$ is small at time $t = 0$ and gradually increases while $P(z^2|t)$ is large at time $t = 0$ and gradually decreases.

We started the growth process from an initial network with 100 nodes, and observed the networks at $t = T_0 (= 5000)$ and $t = T (= 10{,}000)$. Namely, the observed time-sequence of the adjacency matrices were A_{T_0}, A_T. Note that the population of community z^1 becomes more than that of community z^2 in the future while the former is less than the latter at time T_0 and the two are almost equal at time T. Fig. 6(a) indicates the actual network at time T together with the community structure, and (b) indicates its observable data, where only the first 100 nodes are displayed.

6.4.2. Performance evaluation

For the prediction of this artificial growing network, we compared the model incorporating community structure ($K = 2$) and the model not incorporating it ($K = 1$) within the proposed model. For the model of $K = 1$, we used the true parameter values, and for the model of $K = 2$,

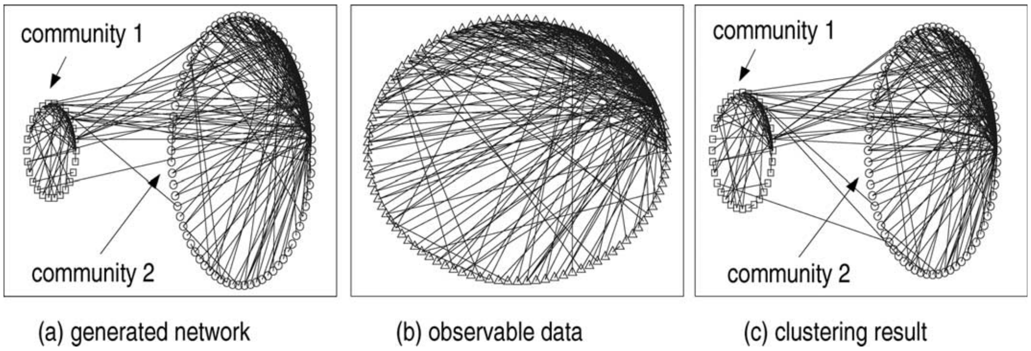

Fig. 6. Clustering performance.

we used the parameter values estimated from A_{T_0}, A_T by our learning method. The experiment was performed for five different initial networks with 100 nodes, which are randomly generated. The results are shown in Table 3.

The second column of Table 3 indicates the clustering accuracy. In particular, Fig. 6(c) visually indicates the clustering result for one of these five trials. These show that our clustering algorithm works well. Namely, at time T, our method could fairly accurately recover the community structure of this growing network from the data without community information.

The third and fourth columns, respectively, indicate the prediction performances of the $K=1$ model and the $K=2$ model. Here, Γ, $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are, respectively, the dynamic probability matrices at time T of the true model, the $K=1$ model and the $K=2$ model. The results show that the $K=2$ model could much more accurately predict the true dynamic probability matrix than the $K=1$ model. This implies that the prediction performance can definitely be improved by incorporating community structure.

6.4.3. Prediction of future degrees

Moreover, we compare the $K=1$ model and the $K=2$ model for the ability to predict the future degrees of existing nodes. Several methods have been proposed to measure the importance of a node in a network from only its graph structure; for example, the HITS algorithm (Kleinberg, 1998), the PageRank algorithm (Brin & Page, 1998), the Randomized HITS algorithm and the Subspace HITS algorithm (Ng et al., 2001). Using the degree of a node in the network is one of the simplest approaches to this task. For these synthetic data, we consider predicting the future degrees of the nodes in the network observed at time T.

Let Δt be a positive integer. For any $v \in \mathcal{N}_T$ (the set of nodes in the network at time T), we consider the degree $D_{T+\Delta t}(v)$ of node v at time $T+\Delta t$. We approximate the true network at time $T+\Delta t$ by the network obtained by adding $M = 5\Delta t$ new links at time $t = T+1$. Moreover, we incorporate the nodes created at time $t \geq T$ into nodes N_T+1 and N_T+2, since we restrict our attention to the nodes in $\mathcal{N}_T$. Let A'_{T+M} denote the adjacency matrix of the network constructed in this manner as an approximation of the true network at time $T+\Delta t$. Therefore, for any $v \in \mathcal{N}_T$, we regard the degree $D_{T+\Delta t}(v)$ as

$$D_{T+\Delta t}(v) = \sum_{j=1}^{N_T+2} A'_{T+M}(v,j). \tag{9}$$

Note here that we deal with the case of undirected graphs. In the same way, for $v = N_T+1, N_T+2$, we define $D_{T+\Delta t}(v)$ by Eq. (9).

Let $\langle D_{T+\Delta t}(v)\rangle$ denote the mean of $D_{T+\Delta t}(v)$. Then, the following relation holds (see Appendix C):

$$\langle D_{T+\Delta}(v)\rangle = \sum_{j=1}^{N_T} A_T(v,j) + \frac{M}{2}\sum_{j=1}^{N_T+2}\{\Gamma(v,j)+\Gamma(j,v)\}. \tag{10}$$

Let $h^{(\Delta t)}(v)$ denote the normalized mean degree of v at time $T+\Delta t$; i.e.

$$h^{(\Delta t)}(v) = \frac{\langle D_{T+\Delta t}(v)\rangle}{\sum_{i=1}^{N_T+2}\langle D_{T+\Delta t}(i)\rangle}.$$

We define $h(v)$ by

$$h(v) = \lim_{\Delta t\to\infty} h^{(\Delta t)}(v).$$

By Eq. (10), it turns out that

$$h(v) = \frac{1}{2}\sum_{j=1}^{N_T+2}\{\Gamma(v,j)+\Gamma(j,v)\}.$$

We call $h(v)$ the *dynamic degree* of v at time T, and $h = [h(1),\ldots,h(N_T+2)]$ the *dynamic degree vector* at time T.

For this artificial growing network, we predict the future degrees of the nodes in the network observed at time T by

Table 3
Effects of incorporating community structure

Trial	Clustering accuracy	$I(\Gamma;\hat{\Gamma}_1)$	$I(\Gamma;\hat{\Gamma}_2)$	$I(h;\hat{h}_1)$	$I(h;\hat{h}_2)$
1	0.910	0.793	0.033	0.195	0.079
2	0.852	0.872	0.137	0.252	0.118
3	0.891	0.869	0.089	0.249	0.092
4	0.912	0.829	0.085	0.221	0.109
5	0.856	0.842	0.078	0.229	0.095

the dynamic degree vector h at time T. Let $\hat{h}_1$ and $\hat{h}_2$ be the dynamic degree vectors of the $K=1$ model and the $K=2$ model at time T, respectively. In Table 3, the fifth and sixth columns, respectively, indicate the Kullback–Leibler divergences $I(h;\hat{h}_1)$ and $I(h;\hat{h}_2)$ for the five trials. The results show that the $K=2$ model could more accurately predict the true dynamic degree vector than the $K=1$ model. This implies that the prediction performance can be improved by incorporating community structure.

7. Concluding remarks

In this paper, we proposed a network growth model incorporating directional attachment and community structure. We showed that the proposed model also exhibits a degree distribution with a power-law tail. We presented a learning algorithm for the proposed model to estimate the actual growth process from the observed data of adjacency matrices. Using the growing network of the Web pages concerning a certain topic, we experimentally showed that the prediction performance for dynamic probability matrix can be improved by incorporating directional attachment and community structure. Moreover, using synthetic data, we experimentally showed that the prediction performance for dynamic probability matrix can definitely be improved by incorporating community structure.

Our research aims to model the Web dynamics. Extensive verification of the proposed model with various real Web data remains an important task. However, we have already made substantial progress, and we are encouraged by the initial results of our efforts to model the Web dynamics.

Appendix A. Proof of Proposition 1

Here, we give a formal proof of Proposition 1 using the master equation approach with the continuous approximation of Dorogovtsev and Mendes (2001). For simplicity, we assume $K=2$ and $m_0=1$, since the generalization to the case of $K>2$ and $m_0>2$ is straightforward.

Note first that $F_t(d)$ can be regarded as the probability that a node in the growing network at time t has d links. Let $F_t(d_1,d_2)$ be the probability that a node in the growing network at time t has d_1 links to the nodes of community z^1 and d_2 links to the nodes of community z^2. We denote the stationary distributions of $F_t(d)$ and $F_t(d_1,d_2)$ by $F(d)$ and $F(d_1,d_2)$, respectively, that is

$$F(d)=\lim_{t\to\infty}F_t(d),$$

$$F(d_1,d_2)=\lim_{t\to\infty}F_t(d_1,d_2).$$

Then, we have

$$F(d)=\sum_{d_1+d_2=d}F(d_1,d_2). \tag{A1}$$

We consider the growing network at time t. Given a node $v\in\mathcal{N}_t$ (the set of nodes in the growing network at time t), let $F_t(d_1,d_2;v)$ denote the probability that at time t, node v has d_1 links to the nodes of community z^1 and d_2 links to the nodes of community z^2. Then, we have

$$F_t(d_1,d_2)=(1/N_t)\sum_{v\in\mathcal{N}_t}F_t(d_1,d_2;v).$$

Next, given a node $v\in\mathcal{N}_t^k$ (the set of nodes that belong to community z^k in the growing network at time t), $(k=1,2)$, let $F_t^k(d_1,d_2;v)$ denote the probability that at time t, node v has d_1 links to the nodes of community z^1 and d_2 links to the nodes of community z^2. Then, we have

$$F_t^k(d_1,d_2)=(1/N_t^k)\sum_{v\in\mathcal{N}_t^k}F_t^k(d_1,d_2;v),\quad (k=1,2). \tag{A2}$$

We also have

$$F_t(d_1,d_2)=\sum_k(N_t^k/N_t)^2F_t^k(d_1,d_2),$$

since

$$\sum_{v\in\mathcal{N}_t}F_t(d_1,d_2;v)=\sum_k(N_t^k/N_t)\sum_{v\in\mathcal{N}_t^k}F_t^k(d_1,d_2;v).$$

Note that the mean $\langle N_t^1\rangle$ of N_t^1 is given by

$$\langle N_t^1\rangle=\{2\xi^1\gamma^{11}\eta_{11}^{11}+\sum_{\varepsilon=0}^{1}(\xi^1\gamma^{12}\eta_{1\varepsilon}^{12}+\xi^2\gamma^{21}\eta_{\varepsilon 1}^{21})\}t+N_0^1, \tag{A3}$$

and the mean $\langle N_t^2\rangle$ of N_t^2 can also be given by a linear expression of t. Thus, we can consider that there exist positive constants g_1 and g_2 such that

$$\lim_{t\to\infty}(N_t^k/N_t)=\sqrt{g_k},\qquad (k=1,2).$$

Hence

$$F(d_1,d_2)=\sum_k g_kF^k(d_1,d_2), \tag{A4}$$

where

$$F^k(d_1,d_2)=\lim_{t\to\infty}F_t^k(d_1,d_2).$$

We consider calculating $F^1(d_1,d_2)$. When we arbitrarily fix a $v\in\mathcal{N}_t^1$, the master equation for $F_t^1(d_1,d_2;v)$ is

$$\begin{aligned}F_{t+1}^1(d_1,d_2;v)=&f_1(d_1-1,t)F_t^1(d_1-1,d_2;v)\\&+f_2(d_1,t)F_t^1(d_1,d_2-1;v)\\&+f_3(d_1,t)F_t^1(d_1,d_2;v),\end{aligned} \tag{A5}$$

where $f_1(d_1-1,t)$ represents the probability that when $F_t^1(d_1-1,d_2;v)=1$, both end nodes of a new link are chosen from community z^1 and either of them is v; i.e.

$$f_1(d_1-1,t)=\xi^1\gamma^{11}\left\{(\eta_{00}^{11}+\eta_{01}^{11})\left(\frac{\alpha^1(d_1-1)}{L_t^1}+\frac{1-\alpha^1}{N_t^1}\right)+(\eta_{00}^{11}+\eta_{10}^{11})\left(\frac{\beta^1(d_1-1)}{L_t^1}+\frac{1-\beta^1}{N_t^1}\right)\right\},$$

$f_2(d_1,t)$ represents the probability that when $F_t^1(d_1,d_2-1;v)=1$, one of the end nodes of a new link is v and the other is chosen from community z^2; i.e.

$$f_2(d_1,t)=\xi^1\gamma^{12}(\eta_{00}^{12}+\eta_{01}^{12})\left(\frac{\alpha^1 d_1}{L_t^1}+\frac{1-\alpha^1}{N_t^1}\right)+\xi^2\gamma^{21}(\eta_{00}^{21}+\eta_{10}^{21})\left(\frac{\beta^1 d_1}{L_t^1}+\frac{1-\beta^1}{N_t^1}\right),$$

and $f_3(d_1,t)$ represents the probability that when $F_t^1(d_1,d_2;v)=1$, both end nodes of a new link are not v; i.e.

$$f_3(d_1,t)=1-f_1(d_1,t)-f_2(d_1,t).$$

Here, L_t^1 denotes the number of links within community z^1 at time t, that is, $L_t^1=\sum_{v\in\mathcal{N}_t^1}D_t^1(v)$. Since the degree of a new node is always one,

$$\sum_{v\in\mathcal{N}_t^1}F_{t+1}^1(d_1,d_2;v)=\sum_{v\in\mathcal{N}_{t+1}^1}F_{t+1}^1(d_1,d_2;v)-\delta_{d_1+d_2,1}=N_t^1F_{t+1}^1(d_1,d_2)+(N_{t+1}^1-N_t^1)\cdot F_{t+1}^1(d_1,d_2)-\delta_{d_1+d_2,1}$$

by Eq. (A2), where $\delta_{d,d'}$ is Kronecker's delta. Applying $\sum_{v\in\mathcal{N}_t^1}$ to both sides of Eq. (A5) and passing it to the $t\to\infty$ limit, we obtain

$$c_1F^1(d_1,d_2)+(a_{11}d_1+b_{11})F^1(d_1,d_2)-\{a_{11}(d_1-1)+b_{11}\}\cdot F^1(d_1-1,d_2)+(a_{12}d_1+b_{12})\{F^1(d_1,d_2)-F^1(d_1,d_2-1)\}=\delta_{d_1+d_2,1}, \quad \text{(A6)}$$

where

$$c_1=\lim_{t\to\infty}(N_{t+1}^1-N_t^1),$$

$$a_{11}=\xi^1\gamma^{11}\{(\eta_{00}^{11}+\eta_{01}^{11})\alpha^1+(\eta_{00}^{11}+\eta_{10}^{11})\beta^1\}\lim_{t\to\infty}(N_t^1/L_t^1),$$

$$b_{11}=\xi^1\gamma^{11}\{(\eta_{00}^{11}+\eta_{01}^{11})(1-\alpha^1)+(\eta_{00}^{11}+\eta_{10}^{11})(1-\beta^1)\},$$

$$a_{12}=\{\xi^1\gamma^{12}(\eta_{00}^{12}+\eta_{01}^{12})\alpha^1+\xi^2\gamma^{21}(\eta_{00}^{21}+\eta_{10}^{21})\beta^1\}\lim_{t\to\infty}(N_t^1/L_t^1),$$

$$b_{12}=\xi^1\gamma^{12}(\eta_{00}^{12}+\eta_{01}^{12})(1-\alpha^1)+\xi^2\gamma^{21}(\eta_{00}^{21}+\eta_{10}^{21})(1-\beta^1).$$

Here, we used

$$\lim_{t\to\infty}N_t^1\{F_{t+1}^1(d_1,d_2)-F_t^1(d_1,d_2)\}=0.$$

Since the mean $\langle L_t^1\rangle$ of L_t^1 is

$$\langle L_t^1\rangle=\xi^1\gamma^{11}t+L_0^1,$$

we can consider that $\lim_{t\to\infty}(N_{t+1}^1-N_t^1)<\infty$ and $\lim_{t\to\infty}(N_t^1/L_t^1)<\infty$ by Eq. (A3). Thus, c_1, a_{11} and a_{12} are positive constants. Eq. (A6) implies that

$$F^1(d_1,d_2)\propto(d_1+b_{11}/a_{11})^{-(1+\nu_1)} \text{ as } d_1\to\infty$$

for a positive constant ν_1, and

$$F^1(d_1,d_2)\propto(\kappa_1)^{-w_1d_2} \text{ as } d_2\to\infty$$

for positive constants κ_1 and w_1. On the other hand, in the continuous d_1, d_2 limit (Dorogovtsev & Mendes, 2001), Eq. (A6) has the form

$$c_1F^1(d_1,d_2)+\frac{\partial}{\partial d_1}\{(a_{11}d_1+b_{11})F^1(d_1,d_2)\}+(a_{12}d_1+b_{12})\frac{\partial}{\partial d_2}F^1(d_1,d_2)=0.$$

By this continuous approximation, we can obtain $\nu_1=c_1/a_{11}$ (cf. Dorogovtsev & Mendes, 2001). Hence, there exists a positive constant ν_1 such that

$$\sum_{d_1+d_2=d}F^1(d_1,d_2)\propto d^{-\nu_1} \text{ as } d\to\infty. \quad \text{(A7)}$$

We can show the same results for $F^2(d_1,d_2)$; for example, there exists a positive constant ν_2 such that

$$\sum_{d_1+d_2=d}F^2(d_1,d_2)\propto d^{-\nu_2} \text{ as } d\to\infty. \quad \text{(A8)}$$

Eqs. (A1) and (A4) imply that

$$F(d)=\sum_{d=d_1+d_2}(g_1F^1(d_1,d_2)+g_2F^2(d_1,d_2)). \quad \text{(A9)}$$

Hence, from Eqs. (A7)–(A9)

$$F(d)\propto d^{-\nu} \text{ as } d\to\infty,$$

where $\nu=\min\{\nu_1,\nu_2\}$. We have completed the proof of Proposition 1.

B. The clustering method

We specifically describe the clustering method used in our experiments. Basically, we use the K-means algorithm (see, e.g. Bishop, 1995).

First, we identify each $u\in\mathcal{N}_T$ with the point $\mathbf{x}_u=[x_u(1),\ldots,x_u(N_T)]$ of the (N_T-1)-dimensional simplex in the N_T-dimensional Euclidean space, which is defined by

$$x_u(i)\propto\begin{cases}(A_T(u,i)+A_T(i,u))/2, & (i\neq u),\\ \max_{1\le j\le N_T}(A_T(u,j)+A_T(j,u))/2, & (i=u),\end{cases}$$

$$\sum_{i=1}^{N_T}x_u(i)=1.$$

Then, our task is to partition the points $\mathbf{x}_u$, $(u\in\mathcal{N}_T)$ into K disjoint subsets $C^1,\ldots,C^K$, where subset C^k corresponds to community z^k for each k.

The algorithm begins by assigning at random the points $\mathbf{x}_u$, $(u \in \mathcal{N}_T)$ to K sets and then computing the representative vectors $\mathbf{r}^k = [r^k(1), \ldots, r^K(N_T)]$, $(k = 1, \ldots, K)$, in such a way as to minimize the objective function

$$J = \sum_{k=1}^{K} \sum_{\mathbf{x}_u \in C^k} I(\mathbf{x}_u; \mathbf{r}^k)$$

under the condition that each $\mathbf{r}^k$ is a point of the $(N_T - 1)$-dimensional simplex in the N_T-dimensional Euclidean space. Here, $I(\mathbf{x}_u; \mathbf{r}^k)$ is the Kullback–Leibler divergence

$$I(\mathbf{x}_u; \mathbf{r}^k) = \sum_{i=1}^{N_T} x_u(i)(\log x_u(i) - \log r^k(i)).$$

It is easily shown that these representative vectors $\mathbf{r}^1, \ldots, \mathbf{r}^K$ are obtained by

$$r^k(i) = \frac{\sum_{\mathbf{x}_u \in C^k} x_u(i)}{|C^k|}, \qquad (i = 1, \ldots, N_T, \ \ k = 1, \ldots, K),$$

where $|C^k|$ denotes the number of the data points belonging to set C^k. Next, each data point $\mathbf{x}_u$ is reassigned to a new set according to which is the nearest representative vector with respect to the Kullback–Leibler divergence, that is, the set C^{k_*} to which $\mathbf{x}_u$ newly belongs is determined by $k_* = \mathrm{argmin}_{1 \le k \le K} I(\mathbf{x}_u; \mathbf{r}^k)$. Then, the value of the objective function J is recomputed. This procedure is repeated until there is no further change in the grouping of the points $\mathbf{x}_u$, $(u \in \mathcal{N}_T)$. Since this algorithm finds a local optimum solution, in our experiments, we changed the initial assignment of the data points to K sets 10 times, and selected the best result.

C. Proof of relation (10)

It suffices to prove that for $u, v = 1, \ldots, N_T + 2$

$$\langle A'_{T+M}(u,v) \rangle = A_T(u,v) + M\frac{1}{2}\{\Gamma(u,v) + \Gamma(v,u)\}. \qquad \text{(A10)}$$

Note that the left-hand side of Eq. (A10) represents the mean number of links between nodes u and v at the next time-step. We put $\mathcal{N}' = \{1, \ldots, N_T + 2\}$, and

$$S_M = \left\{ s = (s_{ij})_{i,j \in \mathcal{N}'}; \ \ 0 \le s_{ij} \le M, \ \ (i,j \in \mathcal{N}'), \sum_{i,j \in \mathcal{N}'} s_{ij} = M \right\}.$$

Then, it is easily seen that

$$\langle A'_{T+M}(u,v) \rangle - A_T(u,v)$$

$$= \sum_{s \in S_M} \frac{s_{uv} + s_{vu}}{2} \frac{M!}{\prod_{i,j \in \mathcal{N}'} s_{ij}!} \prod_{i,j \in \mathcal{N}'} \Gamma(i,j)^{s_{ij}}$$

$$= \sum_{s \in S_{M-1}} M \frac{\Gamma(u,v) + \Gamma(v,u)}{2} \frac{(M-1)!}{\prod_{i,j \in \mathcal{N}'} s_{ij}!} \prod_{i,j \in \mathcal{N}'} \Gamma(i,j)^{s_{ij}}$$

$$= M\frac{1}{2}\{\Gamma(u,v) + \Gamma(v,u)\}.$$

This completes the proof of relation (10).

References

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*, 47–97.

Albert, R., Jeong, H., & Barabási, A.-L. (1999). Diameter of the World-Wide Web. *Nature*, *401*, 130.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.

Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Brin, S., & Page, L. (1998). The anatomy of a large scale hypertextual Web search engine. *In Proceedings of the Seventh International World Wide Web Conference*, pp. 107–117.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. (2000). Graph structure in the Web. *Computer Networks*, *33*, 309–320.

Chakrabarti, S., Dom, B., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D., & Kleinberg, J. (1999). Mining the Web's link structure. *IEEE Computer*, *32*(8), 60–67.

Cohn, D., & Chang, H. (2000). Learning to probabilistically identify authoritative documents. *In Proceedings of the 17th International Conference on Machine Learning*, pp.167–174.

Cohn, D., & Hofmann, T. (2001). The missing link—a probabilistic model of document content and hypertext connectivity. *In Advances in Neural Information Processing Systems 13*, pp. 430–436.

Dorogovtsev, S. N., & Mendes, J. F. F. (2001). Scaling properties of scale-free evolving networks: continuous approach. *Physical Review E*, *63*, 056125.1–056125.19.

Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, *51*, 1079–1187.

Eckmann, J.-P., & Moses, E. (2002). Curvature of co-links uncovers hidden thematic layers in the World Wide Web. *Proceedings of the National Academy of Sciences of USA, 99*, pp. 5825–5829.

Flake, G. W., Lawrence, S., & Giles, C. L. (2000). Efficient identification of Web communities. *In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 150–160.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of USA, 99*, pp. 7821–7826.

Hofmann, T. (1999). Probabilistic latent semantic analysis. *In Proceedings of the 15th International Conference on Uncertainty in Artificial Intelligence*, pp. 289–296.

Huberman, B. A., & Adamic, L. A. (1999). Growth dynamics of the World-Wide Web. *Nature*, *401*, 131.

Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. *In Proceedings of the Ninth ACM-SIAM Symposium on Discrete Algorithms*, pp. 668–677.

Kleinberg, J., & Lawrence, S. (2001). The structure of the Web. *Science*, *294*, 1849–1850.

Krapivsky, P. L., Rodgers, G. J., & Redner, S. (2001). Degree distributions of growing networks. *Physical Review Letters*, *86*, 5401–5404.

Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). Trawling the Web for emerging cyber-communities. *Computer Networks*, *31*, 1481–1493.

Ng, A. Y., Zheng, A. X., & Jordan, M. I. (2001). Stable algorithms for link analysis. *In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 258–266.

Pal, S. K., Talwar, V., & Mitra, P. (2002). Web mining in soft computing framework: relevance, state of the art and future directions. *IEEE Transactions on Neural Networks*, *13*, 1163–1177.

Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., & Giles, C. L. (2002). Winners don't take all: characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences of USA, 99*, pp. 5207–5211.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, *410*, 268–276.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications.* Cambridge: Cambridge University Press.