ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 15-20

Modeling of Growing Networks with Directional Attachment and Communities

Masahiro KIMURA, Kazumi SAITO, Naonori UEDA

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

Abstract. With the aim of acquiring a more precise probabilistic model for the future graph structure of such a real-world growing network as the Web, we propose a new network growth model and its learning algorithm. Unlike the conventional models, we have incorporated directional attachment and community structure for this purpose. We formally show that the proposed model also exhibits a degree distribution with a power-law tail. Using the real data of Web pages on the topic "mp3", we experimentally show that the proposed method can more precisely predict the probability of a new link creation in the future.

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 is becoming an important and challenging research issue [8, 1]. Also, applying the theory of adaptive computation such as neural computation is expected to be effective for mining and modeling this rich collection of data [9]. In this paper, we address the problem of learning Web dynamics.

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, for example, in improving Web search engines [7] and understanding the ecology of the Web. Since the Web is constantly growing through the addition of new pages and hyperlinks created by users with their particular interests, modeling the growth process of this network is an important task [8, 1].

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 [2, 1]. Thus, 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 [2] discovered a growing network model satisfying these conditions. The principal ingredient of their model is a mechanism called *preferential attachment* ("rich get richer" mechanism). Some variants of the

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 15-20

Barabási-Albert (BA) model have been presented [1]. In particular, by introducing *mixtures of preferential and uniform attachment*, Pennock *et al.* [10] more accurately accounted for the degree distributions in the Web than the BA model. Since a system with power-law is known to have a scale-free nature, these growing network models are generally referred to as *scale-free models*.

Another characteristic of the Web is the existence of *community structure*, and the Web grows as various clusters are formed [8, 5]. Here, 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 [8, 5]. However, the existing scale-free models do not take into account community structure. On the other hand, there have been several investigations using graph-theoretic methods [5] and latent variable models such as PHITS [3] to identify community structure. However, these investigations dealt with only static networks, where the number of nodes and links were not allowed to increase. In our previous work [6], we proposed a growing network model that incorporates community structure into an existing scale-free model. Also, using synthetic data, we experimentally demonstrated that predictive ability can definitely be improved by incorporating community structure. However, verifying the model with real Web data remained an important task.

When a new link is created, the following four cases can happen. It is attached from a (new / old) node to a (new / old) node. Each growing network has its own bias for these four cases. 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.

To more precisely model such a real-world growing network as the Web, we propose a new network growth model and its learning algorithm. Unlike the conventional models, we incorporate *directional attachment* and *community structure*. We show that the proposed model also exhibits a degree distribution with a power-law tail. Using real Web data, we experimentally show that both directional attachment and community structure are effective for modeling such a growing network as the Web.

2 Proposed Model

Let us describe our proposed model. We assume that nodes and links do not disappear in the growth processes.

2.1 Preliminaries

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. Let $\mathcal{N}_t$ denote the set of nodes in the growing network at time t. 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 $A_t(i,j)$.

In accordance with the exsiting scale-free models, we suppose that the growth process of a network is described as a stochastic process $P(\Delta A_t \mid A_{t-1}, \theta)$, $(t \geq 1)$, where θ denotes the set of model parameters. Also, the probability

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 15-20

$P(\Delta A_t \,|\, A_{t-1}, \theta)$ is assumed to be given by the multinomial distribution

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

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}.

2.2 Directional attachment

We incorporate the directional attachment by introducing a set of control parameters, $\eta = \{\eta_{00}, \eta_{01}, \eta_{10}, \eta_{11}\}$, $(0 \le \eta_{\varepsilon\rho} \le 1,\ \sum_{\varepsilon,\rho} \eta_{\varepsilon\rho} = 1)$, as shown in Table 1. That is, we introduce the fully correlated model for directional attachment. To clarify the effectiveness of incorporating directional attachment, we also consider the independent model shown in Table 1 for comparison, where α_0 (β_0) indicates the probability that a new node is chosen as the originating (target) node. Note that the existing scale-free models can be regarded as independent models for directional attachment.

Table 1: Directional attachment in the proposed model

	Fully correlated model		*Independent model*	
	old node	new node	old node	new node
old node	η_{00}	η_{01}	$(1-\alpha_0)(1-\beta_0)$	$(1-\alpha_0)\beta_0$
new node	η_{10}	η_{11}	$\alpha_0(1-\beta_0)$	$\alpha_0\beta_0$

2.3 New link creation process

We assume that there exist K communities, and each node belongs to only one community without changing the community during the studied period. Let $\{z^1, \cdots, z^K\}$ be the set of community-labels. Suppose that m_t new links are added at each time t. Given the adjacency marix A_{t-1} of the network at time $t-1$, a new link $[u_t, v_t]$ at time t is generated as follows:

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 $\gamma^{k\ell}$ as the community to which the target node v_t belongs. Finally, link $[u_t, v_t]$ is created with probability $P([u_t, v_t] \,|\, [z^k, z^\ell], A_{t-1}, \theta)$. Namely, the probability $P([u_t, v_t] \,|\, A_{t-1}, \theta)$ is defined by

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

After the communities $[z^k, z^\ell]$ 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, $\eta^{k\ell} = \{\eta^{k\ell}_{00}, \eta^{k\ell}_{01}, \eta^{k\ell}_{10}, \eta^{k\ell}_{11}\}$, as shown in Table 1. 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 $P([u_t, v_t] \,|\, [z^k, z^\ell], A_{t-1}, \theta)$ is defined by $P([u_t, v_t] \,|\, [z^k, z^\ell], A_{t-1}, \theta) = \eta^{k\ell}_{11}$. 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

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 15-20

created with probability 1. Namely, the probability $P([u_t, v_t] \,|\, [z^k, z^\ell], A_{t-1}, \theta)$ is defined by

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

where $\mathcal{N}_{t-1}^k$ denotes the set of nodes that belong to community z^k in the growing network at time $t-1$, N_{t-1}^k denotes the number of elements of $\mathcal{N}_{t-1}^k$, and $D_{t-1}^k(i) = \sum_{j \in \mathcal{N}_{t-1}^k} \{A_{t-1}(i,j) + A_{t-1}(j,i)\}$ for any $i \in \mathcal{N}_{t-1}^k$.

Similarly, the probability $P([u_t, v_t] \,|\, [z^k, z^\ell], A_{t-1}, \theta)$ is also defined for the case where u_t is a new node and v_t is an old node and the case where both u_t and v_t are old nodes. Hence, a genetative model of growing networks has been constructed. Here, the set θ of model parameters becomes $\theta = \{\alpha^k, \beta^\ell, \eta_{\varepsilon\rho}^{k\ell}, \xi^k, \gamma^{k\ell}\,;\, k, \ell = 1, \cdots, K,\ \varepsilon, \rho = 0, 1\}$, where $0 \le \alpha^k,\ \beta^\ell,\ \eta_{\varepsilon\rho}^{k\ell},\ \xi^k,\ \gamma^{k\ell} \le 1$, and $\sum_{\varepsilon,\rho} \eta_{\varepsilon\rho}^{k\ell} = \sum_k \xi^k = \sum_\ell \gamma^{k\ell} = 1$. We assume that m_t is given.

2.4 Degree distribution

Let us investigate the degree distributions of the growing networks generated by the proposed model. It is possible to give a formal proof of the following proposition using the master equation approach with the continuous approximation [4], but we omit the proof due to the lack of space.

Proposition 1 *Suppose that the values of all parameters of the proposed model are not zero. 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 $F_t(d) \propto d^{-\nu}$ 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.

3 Learning Algorithm

Let $\{A_0, A_1, \cdots, A_T\}$ 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.

We first perform clustering for the network of adjacency matrix A_T (the last observed network) to assign the community-label to each node of the network. Several methods [5] 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.

Next, we estimate θ. Let $\{([u_t^\lambda, v_t^\lambda]; m_t^\lambda); \lambda = 1, \cdots, r_t\}$ be the set of the links added newly at time t, 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, $\Delta A_t(u_t^\lambda, v_t^\lambda) = m_t^\lambda$. We can empirically estimate the parameters $\{\xi^k\}$, $\{\gamma^{k\ell}\}$ and $\{\eta_{\varepsilon\rho}^{k\ell}\}$ from these data based on the clustering result. Now it suffices to estimate the parameters $\varphi = \{\alpha, \beta\}$, where $\alpha = \{\alpha^k\}$ and $\beta = \{\beta^\ell\}$. We perform the maximal likelihood estimation. In this case, the log-likelihood function $\mathcal{L}(\varphi)$ is of the form

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

ESANN'2003 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 23-25 April 2003, d-side publi., ISBN 2-930307-03-X, pp. 15-20

Optimal parameter values can be efficiently estimated by using an iterative alogorithm based on the EM algorithm, but we omit the details due to the lack of space.

4 Experimental Evaluation

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

4.1 Performance measure

Let $\hat{\theta}_K$ denote the set of parameter values of the learned model with K communities. To evaluate the prediction performance of the learned model, we define the *dynamic probability matrix* $\hat{\Gamma}_K$ by $\hat{\Gamma}_K(i,j) = P([i,j] \,|\, A_T, \hat{\theta}_K)$, which represents the probability distribution for a new link creation given A_T. Let Γ denote the dynamic probability matrix of the actual process given A_T. We evaluate the prediction performance of the learned model by the Kullback-Leibler divergence, $I(\Gamma; \hat{\Gamma}_K) = \sum_{i,j} \Gamma(i,j) \log(\Gamma(i,j)/\hat{\Gamma}_K(i,j))$.

4.2 Evaluation for real Web data

We evaluate the proposed model using a growing network of Web pages concerning a broad-topic.

4.2.1 Real Web data

Based on Kleinberg's method [7], 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 one 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, $T = 1$.

4.2.2 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 (see Table 1) and Model-1 be the proposed model, that is, the fully correrated model for directional attachment.

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

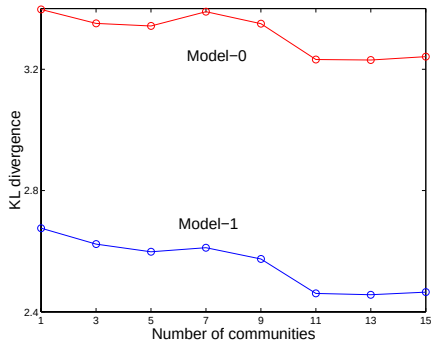

Figure 1: Prediction performance of learned models.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] D. Cohn and H. Chang. Learning to probabilistically identify authoritative documents. In *Proceedings of the 17th International Conference on Machine Learning*, pages 167–174, 2000.

[4] S. N. Dorogovtsev and J. F. F. Mendes. Scaling properties of scale-free evolving networks: Continuous approach. *Physical Review E*, 63:056125-1–056125-19, 2001.

[5] G. W. Flake, S. Lawrence, and C. L. Giles. Efficient identification of Web communities. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 150–160, 2000.

[6] M. Kimura, K. Saito, and N. Ueda. Modeling of growing networks with communities. In *Neural Networks for Signal Processing XII*, pages 189–198. IEEE, 2002.

[7] J. Kleinberg. Authoritative sources in a hyperlinked environment. In *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, pages 668–677, 1998.

[8] J. Kleinberg and S. Lawrence. The structure of the Web. *Science*, 294:1849–1850, 2001.

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

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