Construction of Scale-Free Networks with Partial Information

Jianyang Zeng, Wen-Jing Hsu*, and Suiping Zhou

Center for Advanced Information Systems, Nanyang Technological University, Singapore 639798.
Email:zengjy@gmail.com, hsu@ntu.edu.sg, asspzhou@ntu.edu.sg.

Abstract. It has recently been observed that the node degrees of many real-world large-scale networks, such as the Internet and the Web, follow a power law distributions. Since the classical random graph models are inadequate for explaining this phenomenon, alternative models have been proposed. However, most of the existing models unrealistically assume that each new joining node knows about all the existing nodes in the network. We relax this assumption and propose a model in which each new joining node uniformly and randomly chooses a sample set of existing nodes, and then connects to some nodes in the sample set according to the Preferential Attachment rule. We show that the power law of degree distribution still holds true even if each new joining node knows only a logarithmic number of existing nodes. Compared with the existing models, our construction of scale-free networks based on partial information seems to better approximate the evolution of certain complex networks arising in the real world. Our results may also be applied to the constructions of large-scale distributed systems such as peer-to-peer networks, where the global information is generally unavailable for any individual node in the network.

1 Introduction

It has been reported empirically by several researchers, e.g. in [2, 11, 17], that both the Internet and the Web have a *scale-free* (i.e. size-independent) property: The proportion P_k of nodes with degree k follows a power law distribution: $P_k \sim k^{-r}$, where $r \leq 3$. The well known models of random graphs introduced by Erdős and Rényi [21] do not yield the power law distribution, and thus require modifications for modelling and analyzing these large-scale networks. Barabási and Albert [4] proposed the first scale-free network model referred to as the BA model based on the preferential attachment (PA) rule. The BA model is a dynamic stochastic process, i.e., a new node is added into the network at each time step, and the probability that an existing *old* node gets a link from the new node is proportional to its degree. Kumar et al. [19] independently presented a

* Corresponding author.

so-called *copying model*, motivated by the fact that a new web node is often generated by copying an old one and changing some of its old links. Aiello, Chung, and Lu [1] introduced a scale-free model with a prescribed degree sequence, where the probability of two nodes being connected is proportional to the product of their expected degrees. Fabrikant, Koutsoupias, and Papadimitriou (FKP) [16] considered the underlying geometric metric and proposed a "Highly Optimized Tolerance" (HOT) framework to model the Internet network, i.e., the connection of the new node is constructed according to an adjustable neighborhood consideration. The more details of degree distribution of the FKP model and its extended model are studied in [5] and [6] respectively. A mathematical survey on some scale-free models can be found in [7].

Many variations of the preferential attachment rule originated from the BA model have been developed [9, 10, 14, 15, 12, 13]. Bollobás and Riordan refined the BA model and make it more precise and theoretically analyzable [9, 10]. They found that although the BA model is a dynamic stochastic process, it can be regarded as a "LCD (Linearized Chord Diagrams)" [9], which is a static problem and is easier to solve. A variation on the BA model, in which each node has an "initial attractiveness" was introduced independently by two different groups, Dorogovtsev et al. [14] and Drinea et al. [15]. A precise and rigorous analysis of this model was given by Buckley and Osthus [12]. Cooper and Frieze [13] presented a general model which considers more parameters, such as a variable number of edges or nodes generated at each time step, a mixture of uniform and PA selection, etc.

Although these scale-free models are consistent with the power law observation of real-world large-scale networks, unfortunately, most of existing models assume that a newly added node knows all existing nodes, such as their node degrees [4], or the Euclidean distance [16], or the hop distance to the network center [6], etc. This is an unrealistic assumption with real-world large-scale networks such as the Internet or Web, as it requires a node to access and process the extremely large amount of the global information. We will, instead, assume that each node has access to only a small subset of logarithmic size of all the existing nodes.

As in [4, 9], we also allow only one node to be added into the network at each time step. The node uniformly and randomly chooses a sample set of existing nodes, and then connects to some nodes in the sample set according to the PA rule. Our results show that the power law of degree distribution still holds true even if each new joining node knows only a logarithmic number of existing nodes. Compared to the existing models, our construction of scale-free networks based on partial information seems to better approximate the evolution of real-world complex networks.

Our results may also be applied to the constructions of large-scale distributed systems such as peer-to-peer networks, where the global information is generally unavailable for any individual node in the network.

1.1 The Model and Notations

Let m denote an integer constant. Our construction is described as follows.

Step 1**:**
Start with G_1^m*, the graph with only one single node denoted by* v_1 *with* m *self-loops.*
Step t**:**
A new node denoted by v_t *is added to the graph* G_{t-1}^m *to form* G_t^m*. This new node sends* m *edges to the existing nodes in* G_{t-1}^m *according to the following rule: First randomly and independently choose* S_{t-1} *nodes from* G_{t-1}^m *to form the sample set* T_{t-1}*. Then node* v_t *sends* m *edges to the nodes in the sample set* T_{t-1} *according to the preferential attachment rule, i.e., the probability that* v_t *is connected to a node* $u \in T_{t-1}$ *is*

$$Pr[v_t \to u] = \frac{deg_{t-1}(u)}{\sum_{i \in T_{t-1}} deg_{t-1}(i)},$$

where $deg_{t-1}(x)$ *denotes the degree of node* x *in the graph* G_{t-1}^m*.*

We will show that to ensure the power law, it suffices to choose $S_t = \beta \lg t$ [1], where β denotes a constant to be specified later. During the initial steps, it is possible that $S_t \leq t$. In this case, we choose all the existing nodes as the sample set. Such initial choices will not affect our asymptotic results. We assume that all edges in the graph are undirected. The directed variant of our model can be easily obtained by applying a similar method as in [13].

Below are the notations that will be used in our analysis.

$deg_t(x)$: The degree of node x in the graph G_t^m;
T_t: The sample set chosen at the time step t;
S_t: The size of the sample set T_t, i.e., $S_t = |T_t|$;
D_t: The sum of the node degrees of all nodes in the sample set T_t, i.e., $D_t = \sum_{x \in T_t} deg_t(x)$;
$d_t(i)$: The number of nodes with degree i in the graph G_t^m;
$d_{t,s}(i)$: The number of nodes with degree i in the sample set T_t;
$\bar{d}_t(i)$: The expectation of $d_t(i)$;
$\bar{d}_{t,s}(i)$: The expectation of $d_{t,s}(i)$;
M_t: The maximum degree of the graph G_t^m;
Diam(G_t^m): The diameter of the graph G_t^m;
n: The size of the final graph G_n^m.

Throughout the paper, we will identify the node v_x with the integer number x to simplify the notation.

1.2 Our Main Results

Our main results are stated as follows:

[1] The logarithmic symbol log is with the base 2, if not otherwise specified. Also, we remove the ceiling or floor for simplicity throughout the paper.

Result 1 (Node Degree Distribution): *Let m and β denote sufficiently large constants, and let $S_t = \beta \lg t$ denote the size of the sample set in Step t. Let $d_n(i)$ denote the number of nodes with degree i in G_n^m. Then* **whp**

$$d_n(i) = \frac{cn}{i(i+1)(i+2)} + O(n^{\theta}),$$

where θ and c denote constants, and $0 < \theta < 1$.

Since each node has a degree of at least m, it is trivial to calculate $d_t(i)$ for $i \le m$. In the subsequent analysis, when we talk about $d_t(i)$, we assume that $i > m$.

Result 2 (Maximum Degree): *Let M_n denote the maximum degree of the graph G_n^m. Then for any constant $\epsilon \in (0, \frac{1}{2})$,* **whp**

$$C_2 n^{\frac{1}{2}-\epsilon} \le M_n \le C_1 n^{\frac{1}{2}+\epsilon},$$

where C_1 and C_2 denote constants.

Result 3 (Network Diameter): *Let $\mathrm{Diam}(G_n^m)$ denote the diameter of the graph G_n^m. If $m \ge C_3 \lg n$ for a sufficiently large constant C_3, then* **whp**

$$\mathrm{Diam}(G_n^m) \le 2 \lg n.$$

Although we only show that $\mathrm{Diam}(G_n^m) = O(\lg n)$ for the case $m = \Omega(\lg n)$, we conjecture that $\mathrm{Diam}(G_n^m) = O(\lg n)$ also holds for $m = \Theta(1)$.

The proof of **Result 1** is relatively more challenging and its proof is given in Section 2. Due to page limitation, the proofs of **Result 2** and **Result 3** are not shown in this conference paper. The reader is referred to [22] for more details.

2 The Degree Distribution

The following is a roadmap for the proof of **Result 1**. Firstly, Section 2.1 gives the expectation of D_t and shows that D_t concentrates around its expectation by applying an extension of the martingale method [20]. Secondly, based on the concentration result of D_t, a recurrence relation of $\bar{d}_t(i)$ is given in Section 2.2. We then inductively show that its solution follows a power law. In Section 2.3, we argue that $d_t(i)$ concentrates around its expectation by applying a similar analysis of the concentration of D_t.

2.1 Concentration of D_t

We first analyze the expectation of D_t.

Since the sample nodes in T_t are selected randomly and independently from the current nodes in G_t^m, we can easily obtain the following lemma.

Lemma 1. $E[d_{t,s}(i)] = \frac{S_t}{t} E[d_t(i)]$.

Based on the above lemma, the expectation of D_t can be easily obtained.

Lemma 2. *$E[D_t] = 2mS_t$.*

Proof:

$$\begin{aligned} E[D_t] &= E[\sum_{i>0} d_{t,s}(i)] = \sum_{i>0} E[d_{t,s}(i)] = \sum_{i>0} \frac{S_t}{t} E[d_t(i)] \qquad \text{(by Lemma 1)} \\ &= \frac{S_t}{t} \sum_{i>0} E[d_t(i)] = \frac{S_t}{t} E[\sum_{i>0} d_t(i)] = \frac{S_t}{t} 2mt = 2mS_t. \qquad \blacksquare \end{aligned}$$

Our analysis of the concentration result is mainly based on the following probabilistic tool, which is an extension of the martingale method [3].

Lemma 3 (Martingale extension [20]). *Let $X = (X_1, \cdots, X_n)$ be a family of random variables with X_k taking value in a set A_k, and let f be a bounded real-valued function defined on $\Pi A_k = A_1 \times A_2 \times \cdots \times A_n$. Let $x_i \in A_i$ for each $i = 1, \cdots, k-1$. For $x \in A_k$, let $g_k(x) = E[f(X)|X_k = x] - E[f(X)]$. Let $r_k = \sup\{|g_k(x) - g_k(y)| : x, y \in A_k\}$, and $R^2(X) = \sum_{k=1}^{n} r_k^2$. Let $\hat{r}^2 = \sup\{R^2(X) \text{for all } X \in \Pi A_k\}$. Then*

$$\Pr[|f(X) - E[f(X)]| \geq c] \leq 2\exp(-2c^2/\hat{r}^2),$$

where $c > 0$

The following lemma gives a lower bound of D_t which is useful for our subsequent analysis. Due to space limitation, its proof is referred to our full version [22].

Lemma 4. *Let m and β denote sufficiently large constants, and let $S_t = \beta \lg t$ as defined earlier and let $n_0 = n^{5/6}$. Then there exists a constant δ where $0 < \delta < 1$, such that*

$$\Pr\left[D_t \geq (1+\delta)mS_t\right] > 1 - \frac{1}{n^2},$$

for all $t > n_0$.

Let N_i denote the set of neighbors of node i when it first enters the network at step i. Then N_i is a tuple of nodes $\boldsymbol{x} = (x_1, \cdots, x_m) \in \{1, \cdots, i-1\}^m$. Let $g_{\tau,t}(\boldsymbol{x}) = E[D_t|N_1, \cdots, N_{\tau-1}, N_\tau = \boldsymbol{x}]$, where the sequence of $N_1, \cdots, N_{\tau-1}$ is fixed and $\boldsymbol{x} \in \{1, \cdots, \tau-1\}^m$, $1 \leq \tau \leq t$. Let $r_{\tau,t} = \sup\left\{|g_{\tau,t}(\boldsymbol{x}) - g_{\tau,t}(\boldsymbol{y})| : \boldsymbol{x}, \boldsymbol{y} \in \{1, \cdots, \tau-1\}^m\right\}$. In order to bound $r_{\tau,t}$ and apply the martingale extension in Lemma 3 to analyze the concentration of D_t, we introduce the following node-edge marking rule:

As in [10, 9], we regard one edge as two "half-ward" directed edges. At a time step τ, we initially mark the nodes in the set N_τ as "τ-influenced nodes", and mark all the half edges connected to N_τ as "τ-influenced half edges". During the next time step $\tau + 1$, if an outgoing half edge sent by the new node $\tau + 1$ is connected to a τ-influenced node, it is also marked as a τ-influenced edge.

Applying the same rule for the constructions of the sequence of graphs $G^m_{\tau+2}, \cdots, G^m_t$, then the value of $E[D_t|N_1, \cdots, N_{\tau-1}, N_\tau = \boldsymbol{x}] - E[D_t|N_1, \cdots, N_{\tau-1}, N_\tau = \boldsymbol{y}]$ is upper bounded by the number of all τ-influenced edges that are attached to the sample set T_t. Let Δ_σ denote the expected number of τ-influenced edges in the graph G^m_σ, then we have $r_{\tau,t} \leq \frac{S_t}{t}\Delta_t$.

Based on the above observations, we have the following lemma.

Lemma 5. *Define $r_{\tau,t}$ and Δ_τ as above. Let m and β denote sufficiently large constants, and let $S_t = \beta \lg t$. Let δ denote a constant such that $0 < \delta < 1$. Then,*

$$r_{\tau,t} \leq \begin{cases} \frac{S_t}{t}\Delta_\tau \frac{n_0}{\tau}\left(\frac{t}{n_0}\right)^{\frac{1}{1+\delta}}, & \text{when } 1 \leq \tau \leq n_0; \\ \frac{S_t}{t}\Delta_\tau \left(\frac{t}{\tau}\right)^{\frac{1}{1+\delta}}, & \text{when } n_0 < \tau \leq t. \end{cases}$$

where $n_0 = n^{5/6}$.

Proof: Recall that Δ_σ is the expected number of τ-influenced edges in the graph G^m_σ. Let Y_σ denote the expected number of new τ-influenced edges generated from G^m_σ during step $\sigma+1$. Then by linearity of expectation, we have $\Delta_{\sigma+1} = \Delta_\sigma + Y_\sigma$.

Let e^σ_i denote the number of nodes with i τ-influenced edges among G^m_σ, then $\Delta_\sigma = \sum_{i>0} i \cdot e^\sigma_i$. So

$$Y_\sigma \leq \sum_{i>0} \frac{m \cdot e^\sigma_i \cdot \frac{S_\sigma}{\sigma} \cdot i}{D_\sigma} = \frac{mS_\sigma}{\sigma D_\sigma} \sum_{i>0} e^\sigma_i \cdot i = \frac{mS_\sigma \Delta_\sigma}{\sigma D_\sigma}.$$

We bound Y_σ according to two different cases.

Case 1 $n_0 < \sigma < t$: Let $\mathcal{E}_\sigma$ denote the event that $D_\sigma \geq (1+\delta)mS_\sigma$. From Lemma 4, we have $\Pr[\mathcal{E}_\sigma] \geq 1 - \frac{1}{n^2}$. Let $\mathcal{E} = \bigcap_{\sigma=\tau}^t \mathcal{E}_\sigma$, then the event $\mathcal{E}$ occurs with probability at least $1 - \frac{1}{n}$. Thus we can assume that $D_\sigma \geq (1+\delta)mS_\sigma$ for all $n_0 < \sigma \leq t$ in the following. Thus,

$$Y_\sigma \leq \frac{mS_\sigma \Delta_\sigma}{\sigma D_\sigma} \leq \frac{\Delta_\sigma}{(1+\delta)\sigma}$$

Case 2 $\tau < \sigma \leq n_0$: It is obvious that $D_\sigma \geq mS_\sigma$. So we have

$$Y_\sigma \leq \frac{mS_\sigma \Delta_\sigma}{\sigma D_\sigma} \leq \frac{\Delta_\sigma}{\sigma}$$

Combining the above two cases, when $1 \leq \tau \leq n_0$, we have

$$\Delta_t \leq \Delta_\tau \prod_{\sigma=\tau+1}^{n_0} \frac{\sigma+1}{\sigma} \prod_{\sigma=n_0+1}^{t} \left(1 + \frac{1}{(1+\delta)\sigma}\right) \leq \Delta_\tau \frac{n_0}{\tau}\left(\frac{t}{n_0}\right)^{\frac{1}{1+\delta}}$$

By using the fact that $1 + ax \leq (1 + x^a)$ for $x > -1$ and $a \geq 1$, we have $1 + \frac{1}{(1+\delta)\sigma} \leq (1 + \frac{1}{\sigma})^{1/(1+\delta)}$. Thus, we have

$$\Delta_t \leq \Delta_\tau \prod_{\sigma=\tau+1}^{n_0} \frac{\sigma+1}{\sigma} \prod_{\sigma=n_0+1}^{t} \left(1 + \frac{1}{\sigma}\right)^{\frac{1}{1+\delta}} \leq \Delta_\tau \frac{n_0}{\tau}\left(\frac{t}{n_0}\right)^{\frac{1}{1+\delta}}$$

When $n_0 < \tau < t$, only case 1 applies, so we have

$$\Delta_t \le \Delta_\tau \prod_{\sigma=\tau+1}^{t} \left(1 + \frac{1}{(1+\delta)\sigma}\right) \le \Delta_\tau \prod_{\sigma=\tau+1}^{t} \left(1 + \frac{1}{\sigma}\right)^{\frac{1}{1+\delta}} \le \Delta_\tau \left(\frac{t}{\tau}\right)^{\frac{1}{1+\delta}}$$

Since $r_{\tau,t} \le \frac{S_t}{t}\Delta_t$, we have

$$r_{\tau,t} \le \begin{cases} \frac{S_t}{t}\Delta_\tau \frac{n_0}{\tau}\left(\frac{t}{n_0}\right)^{\frac{1}{1+\delta}}, & \text{when } 1 \le \tau \le n_0; \\ \frac{S_t}{t}\Delta_\tau \left(\frac{t}{\tau}\right)^{\frac{1}{1+\delta}}, & \text{when } n_0 < \tau \le t. \end{cases}$$ ∎

Theorem 1. *Let m and β denote sufficiently large constants, and let $S_t = \beta \lg t$ as before. Let $n_1 = n^{\frac{11}{12}}$. Then there exists a constant $0 < \varphi < 1$ such that*

$$\Pr[|D_t - 2mS_t| \ge mS_t t^{\varphi-1}] \le \frac{1}{n^2},$$

for all $t > n_1$.

Proof: Let $R_t^2 = \sum_{\tau=1}^{t}(r_{\tau,t})^2$. From Lemma 5, we have

$$\begin{aligned} R_t^2 = \sum_{\tau=1}^{t}(r_{\tau,t})^2 &\le \sum_{\tau=1}^{n_0}\left(\frac{S_t}{t}\Delta_\tau \frac{n_0}{\tau}\left(\frac{t}{n_0}\right)^{\frac{1}{1+\delta}}\right)^2 + \sum_{\tau=n_0+1}^{t}\left(\frac{S_t}{t}\Delta_\tau\left(\frac{t}{\tau}\right)^{\frac{1}{1+\delta}}\right)^2 \\ &= \left(\frac{S_t}{t}\Delta_\tau t^{\frac{1}{1+\delta}}\right)^2\left(n_0^{2-\frac{2}{1+\delta}}\sum_{\tau=1}^{n_0}\frac{1}{\tau^2} + \sum_{\tau=n_0+1}^{t}\frac{1}{\tau^2}\right) \\ &= O\left(\left(\frac{S_t}{t}\Delta_\tau t^{\frac{1}{1+\delta}} n_0^{1-\frac{1}{1+\delta}}\right)^2\right) \end{aligned}$$

Since node τ affects at most $2m$ degrees at the time step τ, $\Delta_\tau \le 2m$. Hence, $R_t^2 = O\left(\left(\frac{S_t}{t}\Delta_\tau t^{\frac{1}{1+\delta}} n_0^{1-\frac{1}{1+\delta}}\right)^2\right) = O\left(\left(\frac{S_t}{t} m t^{\frac{1}{1+\delta}} n_0^{1-\frac{1}{1+\delta}}\right)^2\right)$.

So $\hat{r}^2 = \sup\{R_t^2\} = O\left(\left(\frac{S_t}{t} m t^{\frac{1}{1+\delta}} n_0^{1-\frac{1}{1+\delta}}\right)^2\right)$. By Lemma 3, we have

$$\Pr\left[|D_t - 2mS_t| \ge \frac{mS_t}{t} t^{\frac{1}{1+\delta}} n_0^{1-\frac{1}{1+\delta}} \lg n\right] = \exp\left(-\Omega(\lg^2 n)\right)$$

Since $n_1 = n^{\frac{11}{12}} > n_0 = n^{\frac{5}{6}}$, there exists a constant $0 < \varphi < 1$ such that

$$\Pr\left[|D_t - 2mS_t| \ge mS_t t^{\varphi-1}\right] \le \frac{1}{n^2},$$

for all $t > n_1$ ∎

2.2 Power Law Distribution of $\bar{d}_t(i)$

Theorem 2. *Let m and β denote sufficiently large constants, and let $S_t = \beta \lg t$. Then there exists a constant $0 < \theta < 1$ such that* **whp**

$$\bar{d}_n(i) = \frac{cn}{i(i+1)(i+2)} + O(n^\theta),$$

for a constant c.

Proof: By construction of our model (cf. Section 1.1), we have the following relation:

$$E[d_{t+1}(i)|G_t^m] = d_t(i) + md_{t,s}(i-1)\frac{i-1}{D_t} - md_{t,s}(i)\frac{i}{D_t}.$$

Taking the expectation on both sides, we have

$$E[d_{t+1}(i)] = E[d_t(i)] + m(i-1)E\big[\frac{d_{t,s}(i-1)}{D_t}\big] - miE\big[\frac{d_{t,s}(i)}{D_t}\big]. \qquad (1)$$

Let $\mathcal{F}$ denote the event that $|D_t - 2mS_t| < mS_t t^{\varphi-1}$, then $\Pr[\neg\mathcal{F}] \le n^{-2}$ for all $t > n_1 = n^{\frac{11}{12}}$ according to Theorem 1. It is obvious that $D_t \ge i \cdot d_{t,s}(i)$, hence

$$\begin{aligned} E\big[\frac{d_{t,s}(i)}{D_t}\big] &= E\big[\frac{d_{t,s}(i)}{D_t}|\mathcal{F}\big]\Pr[\mathcal{F}] + E\big[\frac{d_{t,s}(i)}{D_t}|\neg\mathcal{F}\big]\Pr[\neg\mathcal{F}] \\ &\le E\big[\frac{d_{t,s}(i)}{D_t}|\mathcal{F}\big]\Pr[\mathcal{F}] + \frac{\Pr[\neg\mathcal{F}]}{i}. \end{aligned}$$

Let $a = 2mS_t$ and $b = mS_t t^{\varphi-1}$. In the event $\mathcal{F}$, we have $a - b \le D_t \le a + b$. Thus, $\frac{1}{D_t} \le \frac{1}{a-b}$ in this case. So

$$\begin{aligned} & \frac{1}{D_t}\big(\frac{a-b}{a}\big) \le \big(\frac{1}{a-b}\big)\big(\frac{a-b}{a}\big) = \frac{1}{a} \\ \Rightarrow\ & \frac{1}{D_t} \le \frac{1}{a} + \frac{b}{a}\frac{1}{D_t} = \frac{1}{2mS_t} + \frac{t^{\varphi-1}}{2}\frac{1}{D_t} \end{aligned}$$

Thus we have

$$\begin{aligned} E\big[\frac{d_{t,s}(i)}{D_t}\big] &\le E\big[\frac{d_{t,s}(i)}{2mS_t} + \frac{d_{t,s}(i)}{D_t}\frac{t^{\varphi-1}}{2}\big]\Pr[\mathcal{F}] + \frac{\Pr[\neg\mathcal{F}]}{i} \\ &\le \frac{E[d_{t,s}(i)]}{2mS_t} + \frac{t^{\varphi-1}}{2i} + \frac{\Pr[\neg\mathcal{F}]}{i}. \end{aligned}$$

Since $\Pr[\neg\mathcal{F}] \le n^{-2}$ for all $t > n_1 = n^{\frac{11}{12}}$, we have

$$E\big[\frac{d_{t,s}(i)}{D_t}\big] \le \frac{E[d_{t,s}(i)]}{2mS_t} + \frac{t^{\varphi-1}}{i}.$$

According to Lemma 1, $E[d_{t,s}(i)] = \frac{S_t}{t}E[d_t(i)]$. So if $t > n_1 = n^{\frac{11}{12}}$, we have **whp**

$$E\big[\frac{d_{t,s}(i)}{D_t}\big] = \frac{\bar{d}_t(i)}{2mt} + O\big(\frac{t^{\varphi-1}}{i}\big).$$

Similarly, we obtain

$$E\big[\frac{d_{t,s}(i-1)}{D_t}\big] = \frac{\bar{d}_t(i-1)}{2mt} + O\big(\frac{t^{\varphi-1}}{i-1}\big).$$

Thus, Eq. (1) can be converted into

$$\bar{d}_{t+1}(i) = \bar{d}_t(i) + \frac{\bar{d}_t(i-1)}{2t}(i-1) - \frac{\bar{d}_t(i)}{2t}i + O\big(t^{\varphi-1}\big),$$

for all $t > n_1 = n^{\frac{11}{12}}$.

This formula is similar to the recurrence equation in [18]. Thus, by a similar inductive analysis, we can get the following solution for all $1 \leq t \leq n$:

$$\bar{d}_t(i) = \frac{ct}{i(i+1)(i+2)} + O(n^{\theta}),$$

where c denotes a constant, and $\max\{\varphi, \frac{11}{12}\} < \theta < 1$. ■

2.3 Concentration of $d_t(i)$

Theorem 3. *Let m and β denote sufficiently large constants, and let $S_t = \beta \lg t$ for our model. Then there exists a constant $0 < \xi < 1$ such that*

$$\Pr\left[|d_n(i) - \bar{d}_n(i)| \geq n^{\xi}\right] \leq n^{-2}.$$

Proof: The proof can be obtained by applying a similar analysis of the concentration of D_t. ■

3 Concluding Remarks

We have proposed a new scale-free model for large-scale networks where a new joining node connects to some nodes in a small sample set; the connections follow the preferential attachment rule. We show that the power law of degree distribution still holds true. Compared with the existing models, our construction based on partial information can better approximate the evolution of large-scale distributed systems arising in the real world. Our results may also be applied to the constructions of peer-to-peer networks, where the global information is generally unavailable for any individual node in the network.

We have also experimentally evaluated the distribution of node degree in our model. Due to space limitation, the reader is referred to [22] for more details. Our simulations show that, when $|S_t| \geq 5$, the proportion P_k of nodes with degree k follows a power law distribution: $P_k \backsim k^{-r}$, where $r \approx 3$ denotes as the slope of the log-log curve. Our experimental results indicate that our theoretical result (Result 1) is a little conservative. We conjecture that when $|S_t| = \Omega(1)$, the distribution will obey a power law distribution. The rigorous proof of this conjecture remains open.

Acknowledgement We wish to thank Abraham D. Flaxman for useful discussions over emails during the preparation of this work. We would also like to thank anonymous referees for their numerous remarks.

References

1. W. Aiello, F.R.K. Chung, and L. Lu. A Random Graph Model for Massive Graphs. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, 2000.
2. R. Albert, H. Jeong, and A.-L. Barabási. The Diameter of the World Wide Web. *Nature*, 401(9):130–131, 1999.

3. N. Alon and J.H. Spencer. *The Probabilistic Method*. John Wiley, 2000.
4. A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, 1999.
5. N. Berger, B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Degree Distribution of the FKP Network Model. In *Proceedings of the 13th International Colloquium on Automata, Languages and Programming*, 2003.
6. N. Berger, C. Borgs, J.T. Chayes, R.M. D'Souza, and R.D. Kleinberg. Degree Distribution of Competition-Induced Preferential Attachment Graphs. In *Proceedings of the 14th International Colloquium on Automata, Languages and Programming*, 2004.
7. B. Bollobás and O. Riordan. *Mathematical Results on Scale-Free Random Graphs.* In Handbook of Graphs and Neworks, 2002.
8. B. Bollobás and O. Riordan. Coupling Scale-Free and Classical Random Graphs. *Internet Mathematics*, 1(2):215–225, 2003.
9. B. Bollobás and O. Riordan. The Diameter of a Scale-Free Random Graph. *Combinatorica*, to appear.
10. B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The Degree of Sequence of a Scale-Free Random Graph Process. *Random Structures and Algorithms*, 18:279–290, 2001.
11. A. Broder, R. Kumar, F. Maghoul, P. Raghavan S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph Structure in the Web. In *Proceedings of the 9th International World Wide Web Conference*, 2002.
12. G. Buckley and D. Osthus. Popularity Based Random Graph Models Leading to a Scale-Free Degree Distribution, Submitted. 2001.
13. C. Cooper and A. Frieze. A General Model of Web Graphs. *Random Structures and Algorithms*, 22:311–335, 2003.
14. S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Sturcture of Growing Networks with Preferential Linking. *Physical Review Letters*, 85(21):4633–4636, 2000.
15. E. Drinea and M. Enachescu and M. Mitzenmacher. Variations on Random Graph Models for the Web. Technical report, Department of Computer Science, Harvard University, 2001.
16. A. Fabrikant, E. Koutsoupias, and C. Papadimitriou. Heuristically Optimized Trade-Offs: a New Paradigm for Power Laws in the Internet. In *Proceedings of the 29th International Colloquium on Automata,Languages and Programming*, 2002.
17. M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-Law Relationships of the Internet Topology. In *Proceedings of ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, 1999.
18. A.D. Flaxman, A.M. Frieze, and J. Vera. A Geometric Preferential Attachment Model of Networks. In *Proceedings of the 3rd International Workshop on Algorithms and Models for the Web-Graph*, 2004.
19. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic Models for the Web Graph. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 2000.
20. C. McDiarmid. *Concentration*. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed (eds.), Probabilistic Methods in Algorithmic Discrete Mathematics, Springer, 1998, pp. 195-248.
21. P. Erdős and A. Rényi. On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
22. J. Zeng and W.-J. Hsu and S. Zhou. Construction of Scale-Free Networks with Partial Information. Available at http://www.cais.ntu.edu.sg/∼zjy. 2005.