Modeling the Growth of Future Web[1]

Hai Zhuge, Xue Chen and Xiang Li
China Knowledge Grid Research Group (http://kg.ict.ac.cn),
Key Laboratory of Intelligent Information Processing,
Institute of Computing Technology, Chinese Academy of Sciences, China
01062562703, 086
zhuge@ict.ac.cn

ABSTRACT
The future Web can be imagined as a life network consisting of resource nodes and semantic relationship links between them. Any node has a life span from birth – adding it to the network – to death – removing it from the network. Through establishing and investigating two types of models for such a network, we obtain the same scale free distribution of semantic links. Simulations and comparisons validate the rationality of the proposed models.

Categories and Subject Descriptors
H.1.1 [**Information Systems**]: Systems and Information Theory – *general system theory, information theory, value of information.*

General Terms
Theory, Experimentation.

Keywords
Distribution, evolution, power law, Web.

1. INTRODUCTION
The future Web is a vision of a life network consisting of nodes representing versatile resources and edges symbolizing the semantic links or subnets connecting similar semantic resources. Each node has a life span from birth – adding it to the network – to death – removing it from the network. The value of a node's life can be reflected by the number of links it connects. For example, once the number of links connecting a node becomes zero — that means it is not reachable from other nodes, then the node can be regarded as dead and hence it could be removed from the network. Our notion is that a network model in which both creation and deletion operations coexist could better form a competitive environment: the more robust (having much more connections to the environment), the more likely to survive, conversely, the less robust, the more likely to be eliminated.

Some previous models addressed the concept of finite life span for nodes, but they only aim at adding nodes and links to the network, and the nodes are allowed to permanently exist. So they are not suitable for describing the future Web. Amaral suggests limiting the addition of new links on account of nodes' life span or link capacity [1]. His model follows the growth constraint that new links no longer connect to it when a node reaches a certain age or has more than a critical number of links. Another case in point is the decaying model considered by Dorogovtsev and Mendes [2]. At each time step, number of links between sites is removed with equal probability. They have demonstrated that the permanent removing of links to a scale-free network does not break the scaling behavior in a relatively wide range.

2. URN TRANSFER MODEL FOR FUTURE WEB
The urn transfer model can be regarded as a series of urns contain balls having pins attached to them [4]. We assume a countable number of urns, urn_k, k = 0, 1, 2, 3, …, where each ball in urn_k has k pins attached to it. Let $F_k(t)$ be the number of balls in urn_k. p be the expected value of adding a new ball into urn_0, p' be the delete factor and α be non-preferential factor. Then, at each time step one of two kinds of behaviors may occur in this model: (i) add a new ball having no pins attached into urn_0 with the probability:

$$p_{t+1} = 1 - \frac{(1-p)\sum_{k=0}^{t}(k+\alpha)F_k(t)}{t[(1-p)(1-2p')+\alpha p]+\alpha(1-p)}, \quad 0 \le p_{t+1} \le 1 \qquad (1)$$

(ii) add/remove one pin to/from a selected ball, then transfer the ball into the urn containing balls with the same number of pins.

$$p_{remove\_pin} = \frac{p'(1-p)(k+\alpha)F_k(t)}{t[(1-p)(1-2p')+\alpha p]+\alpha(1-p)} \quad \text{and}$$

$$p_{add\_pin} = \frac{(1-p')(1-p)(k+\alpha)F_k(t)}{t[(1-p)(1-2p')+\alpha p]+\alpha(1-p)} \qquad (2)$$

In the boundary case, k=0, one ball is either removed out of the urn_0 with probability p_{remove_pin} or transfer to urn_1 with probability p_{add_pin} after attaching a pin to it. Consequently, for k>0, we have the following expected value of $F_k(t+1)$.

$$E_t(F_k(t+1)) = F_k(t) + \frac{(1-p')(1-p)(k-1+\alpha)}{t[(1-p)(1-2p')+\alpha p]+\alpha(1-p)}F_{k-1}(t) - \frac{(1-p)(k+\alpha)}{t[(1-p)(1-2p')+\alpha p]+\alpha(1-p)}F_k(t)$$

[1] The research work was supported by the National Science Foundation of China.

$$+\frac{p^{'}(1-p)(k+1+\alpha)}{t[(1-p)(1-2p')+\alpha p]+\alpha(1-p)}F_{k+1}(t) \qquad (3)$$

It has been proved $E(F_k(t))/t$ tends to a limit f_k as t tends to infinity [4]. Through some simplification on equation (3), We yield the following approximate resolution:

$$f_k \sim Ck^{-(1+\rho)}(1-p^{'})^k$$

where C is a constant . f_k visually resembles a power law, but it is not the case otherwise.

3. DIRECTED EVOLVING GRAPH FOR FUTURE WEB

The urn transfer model, however, cannot directly reflect the concept of web for not simultaneously considering both incoming and outgoing links. The Web can be described as a directed evolving graph $G=(N, E)$, where N is the node set and E is the link set. Let $F_{i,j}(t)$ be the expected numbers of nodes with i incoming semantic links (i.e. in-degree) and j outgoing semantic links (i.e. out-degree) at time step t. α_{in} and α_{out} are the non-preferential factors for in-degree and out-degree respectively.

At each time step, one of two kinds of operations for nodes and links may occur in the graph: (i) with probability p, a new unlinked node is introduced. (ii) with probability $(1-p)(1-p')$, a node is chosen. Then, the node is to receive a new incoming semantic link with the probability proportional to $(i+\alpha_{in})F_{i,j}(t)$, or the node is to be deleted a outgoing semantic link with the probability proportional to $(j+\alpha_{out})F_{i,j}(t)$. By employing rate equation approach, we can obtain the joint distribution $F_{i,j}(t)$ [3]:

$$\frac{dF_{i,j}}{dt}=(1-p)\frac{(1-p')(i-1+\alpha_{in})F_{i-1,j}-(i+\alpha_{in})F_{i,j}+p'(i+1+\alpha_{in})F_{i+1,j}}{[(1-p)(1-2p')+\alpha_{in}p]t}$$
$$+(1-p)\frac{(1-p')(j-1+\alpha_{out})F_{i,j-1}-(j+\alpha_{out})F_{i,j}+p'(j+1+\alpha_{out})F_{i,j+1}}{[(1-p)(1-2p')+\alpha_{out}p]t}$$
$$+p\delta_{i,0}\delta_{j,0}$$

$F_{i,j}(t)$ can be resolved straightforwardly into the separated in-degree and out-degree distributions by means of summing i and j respectively: $\sum_j \frac{dF_{i,j}}{dt}=\frac{dI_i}{dt}$ and $\sum_i \frac{dF_{i,j}}{dt}=\frac{dO_j}{dt}$.

Therefore, we found the in- and out-degree distributions evolve in the same manner except the difference of the factor α_{in} and α_{out}, which reflects the fact that the governing rules of this model are symmetric. Hence it is safe to say that both the incoming and outgoing links share the same evolving trend of the degree distribution and the only difference lies in the values of non-preferential factors. By means of deduction, we yield the same approximate resolution as that of urn transfer model

4. EXPERIMENTS AND ANALYSIS

Figure 1 show the two groups of comparisons for the in-degree distribution. The parameters are derived from *the Web graph model* [3]: (a) p=0.125, α =0.75 and p'=0.01and (b) p=0.125, α =0.75 and p'=0.1. The solid lines are results from numerical simulations with 10^6 time steps and points are approximate resolutions. The two curves are basically consistent. It becomes conscious that the smaller for the value of link deleting factor p', the more similar between two curves.

Figure 2 show the comparisons for the out-degree distribution. (a) is for p=0.125, α =3.55 and p'=0.1, while (b) is for p=0.125, α =3.55 and p'=0.2. Through comparison, we obtain the same conclusion as that of in-degree distribution.

5. CONCLUSION

In this paper, we propose two different modeling measures for the future Web and deduced an asymptotic formula to describe the distribution of the semantic links. Considering the simplification we have made, the approximate resolution may be not accurate enough. Then, we have also implemented a simulation with 10^6 time steps to mimic such a dynamic evolution process. Comparisons show the approximate resolution is consistent with our model in general. The proposed model can be used to evaluate future web's experimental data sets [5].

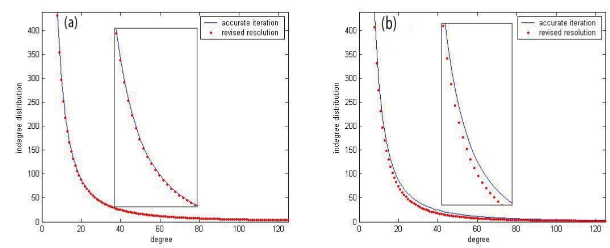

Figure 1. Two groups of comparisons for in-degree distribution between simulations and approximate resolutions.

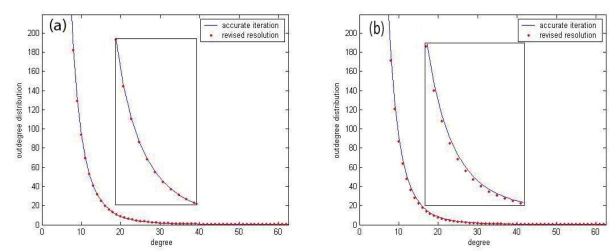

Figure 2. Two groups of comparisons for out-degree distribution between simulations and approximate resolutions.

6. REFERENCES

[1] L.A.N.Amaral, A.Scala et al., Classes of Small-World Networks, Proceedings of the National Academy of Sciences, 97, 2000, pp.11149-11152.

[2] S.N.Dorogovtsev, J.F.F.Mendes, Scaling Behaviour of Developing and Decaying Networks, Europhys. Lett.52, 2000, pp.33-39.

[3] P.L.Krapivsky, S.Redner, A statistical physics perspective on Web growth, Computer Networks 39, 2002, pp.261-276.

[4] M.Levene, T.Fenner, G.Loizou, R.Wheeldon, A stochastic model for the evolution of the Web, Computer Networks, 39, 2002, pp.277-287.

[5] Knowledge Grid Forum, http://www.knowledgegrid.net.