The Third International Symposium on Optimization and Systems Biology (OSB'09)
Zhangjiajie, China, September 20–22, 2009

Markov Model of Biological Networks

Cai-Feng Du[1] Fu-Zhou Gong[2]

[1]College of Mathematics and Computational Science, China University of Petroleum, Dongying Shandong 257061
[2]Institute of Applied Mathematics, Academy of Mathematics and Systems Science, CAS, Beijing 100080

Abstract A Markov model was proposed to examine the evolution of biological networks. We consider of random elimination processes of node degree as well as the preferential choice mechanisms. For the model,it is proved that there is a stationary power law degree distribution with exponents less than 2,consist with current data on biological networks.

Keywords biological network,degree distribution,random elimination,duplication.

1 Introduction

Networks of interactions are fundamental to all biological system. The interactions are all parts of complex biological networks,such as the PPI network formed by the interactions among proteins. There is considerable interest in biological networks with extensive data acquired by new technological advances[1, 2, 3]. Current research focus on the different topology between the biological networks and other complex networks[4]. Aiello etc.[5]have surveyed the difference. Biological networks often have exponents that are between 1 and 2. The non-biological networks,on the other hand,have exponents that commonly range from 2 to 4.

while it is difficult to draw a strong conclusion from the limited observations to date, it does raise the question as to whether biological networks evolve differently. Some research [6, 7]show that information in the genome will lost partly when the time past.

In the present paper we propose a probabilistic model with random elimination of node degree. Define the degree distribution $P(k)$ as the proportion of vertices with degree k. We say the degree distribution is stationary if $P(k,t)$ converge in probability to $P(k)$. We prove our network model has a stationary power law degree distribution with exponents $-3+\frac{A}{m}$ where A denote the random elimination of node degree. The degree distribution is independent of the initial condition.

2 The Markov model

All the graphs we consider here is simple and undirected without multiple edges and loops.

Define $(G_t)_{t\geq 0}$ as follows: G_m^o is a m-completed graph with vertices v_{-m+1}, v_{-m+2}, $\cdots$, v_0. Given G_{t1},form G_t by adding the vertex v_t with m edges:
(i)(**Elimination**)Form $\widetilde{G}_{t-1}$ by random elimination of node degree of G_{t-1} as following

Denote that degree sequence of G_{t-1} as $d_{v_{-m+1}}, d_{v_{-m+2}}, \cdots, d_{v_{t-1}}$, then the degree sequence of $\widetilde{G}_{t-1}$ is $d_{v_{-m+1}} - A, d_{v_{-m+2}} - A, \cdots, d_{v_{t-1}} - A$
(ii)(**Preference**)the first edge of v_t choose an old vertex v_i of $\widetilde{G}_{t-1}$ with the preferential probability

$$P^0(i)(t) = \frac{k(i,t-1) - A}{\sum_{j=-m+1}^{t-1}(k(j,t-1) - A)},$$

where $v_i \in \{v_{-m+1}, v_{-m+2}, \cdots, v_0, v_1 \cdots, v_{t-1}\}$, $k(i,t-1)$ is the degree of v_i in G_{t-1}.
(iii)(**Duplication**)each of the other $m-1$ edges choose a neighbors of v_i in $\widetilde{G}_{t-1}$ uniformly,that is ,if j is one neighbor of i in $\widetilde{G}_{t-1}$,then,

$$P(j \mid i)(t) = \frac{m-1}{k(i,t-1) - A}.$$

where$P(j \mid i)(t)$denote the probability of a neighbor of v_i in $\widetilde{G}_{t-1}$ chosen by one of the other $m-1$edges of v_t when v_i was chosen by the first edge of v_t.
(iv)From the above three step,we get a network $\widetilde{G_t}$. Denote that degree sequence of $\widetilde{G_t}$ as $d_{v_{-m+1}}, d_{v_{-m+2}}, \cdots, d_{v_t}$, then the degree sequence ofG_t is $d_{v_{-m+1}} + A, d_{v_{-m+2}} + A, \cdots, d_{v_{t-1}} + A$.We can get G_t from the sequence.

Denote N_s as the set of neighbors of $v_s$in $\widetilde{G}_{t-1}$. Using the totally probability formula , we obtain the marginal probability of vertex v_s is chosen by one of m edges of v_t at time t,

$$\begin{aligned} P_1(s)(t) &= P^0(s)(t) + \sum_{j \in N_s} P^0(j)(t) \frac{m-1}{k(j,t-1) - A} \\ &= \frac{m(k(s,t-1) - A)}{\sum_{j=-m+1}^{t-1}(k(j,t-1) - A)}. \end{aligned} \tag{1}$$

The boundary condition is

$$k(t,t) = m \tag{2}$$

3 Stationary power law degree distribution

All the random variables we consider here are defined on the same probability space $(\Omega, \mathscr{F}, P)$[8].

Let $N(k,t)$ denote the number of vertices with degree k at time t, $k(s,t)$ denote the degree of v_s at time t, then, we have $N(k,t) = \sum_{s=1}^{t} \delta_{[k(s,t)=k]}$,where δ is the indicator function.

Let$\overline{N}(k,t) = \mathrm{E}[N(k,t)]$,we have

Lemma 1. *when $t \to \infty$,we have*

$$\frac{\overline{N}(k,t)}{t} \propto k^{-3+\frac{A}{m}}.$$

Proof. Firstly,we give the recursive equation of $\overline{N}(k,t)$.

Let$P(k(s,t)=k)$ denote the probability of degree of vertex v_s is k at time t, then we have

$$\overline{N}(k,t)=\mathrm{E}[\sum_{s=1}^{t}\ \delta_{[k(s,t)=k]}]=\sum_{s=1}^{t}\mathrm{E}[\delta_{[k(s,t)=k]}]=\sum_{s=1}^{t}P(k(s,t)=k). \tag{3}$$

Let $Y_s(k,t+1)$ denote the number of vertex v_s with degree k at time $t+1$, then

$$N(k,t+1)=\sum_{s=1}^{t}Y_s(k,t+1)+\delta_{km}$$

and

$$Y_s(k,t+1)=\begin{cases}1, & on\{k(s,t)=k-1,k(s,t+1)=k(s,t)\\ & +1\}\bigcup\{k(s,t)=k,k(s,t+1)=k(s,t)\},\\ 0, & else.\end{cases}$$

So,

$$\overline{N}(k,t+1)=\mathrm{E}\sum_{s=1}^{t}Y_s(k,t+1)+\delta_{km}. \tag{4}$$

According(1), we know

$$\begin{aligned}&P(\{k(s,t)=k-1,k(s,t+1)=k(s,t)+1\})\\ =\ &P(\{k(s,t)=k-1\})P(\{k(s,t+1)=k(s,t)+1\\ &\mid k(s,t)=k-1\})\\ =\ &m\frac{k-1-A}{\sum_{j=-m+1}^{t}(k(j,t)-A)}P(\{k(s,t)=k-1\}).\end{aligned}$$

Similarly,

$$\begin{aligned}&P(\{k(s,t)=k,k(s,t+1)=k(s,t)\})\\ =\ &P(\{k(s,t)=k\})[1-P(\{k(s,t+1)=k(s,t)\mid k(s,t)\\ &=k\})]\\ =\ &[1-m\frac{k-A}{\sum_{j=-m+1}^{t}(k(j,t)-A)}]P(\{k(s,t)=k\}).\end{aligned}$$

So,

$$\begin{aligned}&\mathrm{E}[\sum_{s=1}^{t}Y_s(k,t+1]\\ =\ &\sum_{s=1}^{t}\mathrm{E}[Y_s(k,t+1)]\\ =\ &\sum_{s=1}^{t}P(\{k(s,t)=k-1,k(s,t+1)=k(s,t)+1\}\bigcup\\ &\{k(s,t)=k,k(s,t+1)=k(s,t)\})\end{aligned}$$

$$
\begin{aligned}
&= \sum_{s=1}^{t}[P(\{k(s,t)=k-1,k(s,t+1)=k\})+P(\{k(s,t)\\
&\quad =k,k(s,t+1)=k\})]\\
&= \sum_{s=1}^{t}(m\frac{k-1-A}{\sum_{j=-m+1}^{t}(k(j,t)-A)}P(\{k(s,t)=k-1\})\\
&\quad +[1-m\frac{k-A}{\sum_{j=-m+1}^{t}(k(j,t)-A)}]P(\{k(s,t)=k\})).
\end{aligned}
$$

Then with(3)(4) ,we obtain

$$
\begin{aligned}
\overline{N}(k,t+1) &= m\frac{k-1-A}{\sum_{j=-m+1}^{t}(k(j,t)-A)}\overline{N}(k-1,t)\\
&\quad +(1-m\frac{k-A}{\sum_{j=-m+1}^{t}(k(j,t)-A)})\overline{N}(k,t)\\
&\quad +\delta_{km}, \qquad (5)
\end{aligned}
$$

where

$$
\delta_{km}=\begin{cases}1 & k=m;\\ 0 & k\neq m.\end{cases}
$$

Secondly,using recursion and the method in [9],we can give the solution of difference equation(5)which is a linear difference equation with rational coefficients,

$$
\overline{N}(k,t)=\gamma(m)k^{-3+\frac{A}{m}}t+\beta(k)t^{-1/2}.
$$

where$\gamma(m),\beta(k)$ is constant independent of t.

So,when $t\to\infty$,we get

$$
\frac{\overline{N}(k,t)}{t}\propto k^{-3+\frac{A}{m}}.
$$

Now, we complete the proof of Lemma 1 . □

Next,we will prove our model has a stationary degree distribution.

To the end, let

$$
\mathscr{F}_t=\sigma(G_0,G_1,\cdots,G_t),
$$

and

$$
M_k(t)=\mathrm{E}[N(k,T)|\mathscr{F}_t](t=1,2,\cdots,T),
$$

then, $M_k(t)$ is martingale, and

$$
M_k(0)=\mathrm{E}[N(k,T)],M_k(T)=N(k,T).
$$

From our early work[8], we know that $\{G_t\}_{t\geq 0}$ is Markov Chain, so we have $M_k(t)=\mathrm{E}[N(k,T)|\mathscr{F}_t]=\mathrm{E}[N(k,T)|G_t](t=1,2,\cdots,T)$.

The following lemma was proved by Professor Z.M.Ma[10]. It is different from the proof in[11, 12].

Lemma 2. *For any k,* $| M_k(t) - M_k(t+1) | \leq 2m$.

Lemma 3 (Azuma inequality)**.** *If* $M(t)$ *is martingale with* $| M(t) - M(t+1) | \leq d$*, then*

$$P(M(T) - M(0) \geq r) \leq \exp(-\frac{r^2}{2d^2T}).$$

Theorem 4. *Let* $P(k) = \gamma(m)k^{-3+\frac{A}{m}}$*,Our model has the stationary degree distribution* $P(k)$*,that is*

$$\frac{N(k,T)}{T} \xrightarrow{P} P(k).$$

Proof: Since

$$| \frac{N(k,T)}{T} - P(k) | \leq | \frac{N(k,T)}{T} - \frac{\overline{N}(k,T)}{T} | + | \frac{\overline{N}(k,T)}{T} - P(k) |,$$

and according lemma1, we have

$$\frac{\overline{N}(k,T)}{T} \to P(k),$$

that is $\forall \varepsilon, \exists T_0, s.t. T \geq T_0$

$$| \frac{\overline{N}(k,T)}{T} - P(k) | < \frac{\varepsilon}{2}.$$

So,we get

$$[| \frac{N(k,T)}{T} - P(k) | \geq \varepsilon] \subseteq [| \frac{N(k,T)}{T} - \frac{\overline{N}(k,T)}{T} | \geq \frac{\varepsilon}{2}].$$

By lemma3 ,let $M(T) = M_k(T) = N(k,T), M(0) = M_k(0) = \overline{N}(k,T)$,we get $\forall \varepsilon > 0$,

$$P([| \frac{N(k,T)}{T} - \frac{\overline{N}(k,T)}{T} | \geq \frac{\varepsilon}{2}]) \leq \exp(-\frac{(\frac{\varepsilon}{2})^2 T}{2(2m)^2}).$$

So, for any $\varepsilon > 0$, we have
$P([| \frac{N(k,T)}{T} - P(k) | \geq \varepsilon]) \leq P([| \frac{N(k,T)}{T} - \frac{\overline{N}(k,T)}{T} | \geq \frac{\varepsilon}{2}]) \leq \exp(-\frac{\varepsilon^2 T}{32m^2})$,
that is,

$$\frac{N(k,T)}{T} \xrightarrow{P} P(k).$$

We complete the proof.

Now,we can draw the conclusion that our model has stationary power law degree distribution with exponents $-3 + \frac{A}{m}$.

Also we give numerical simulations of our model to check the analytical result.(see figures)

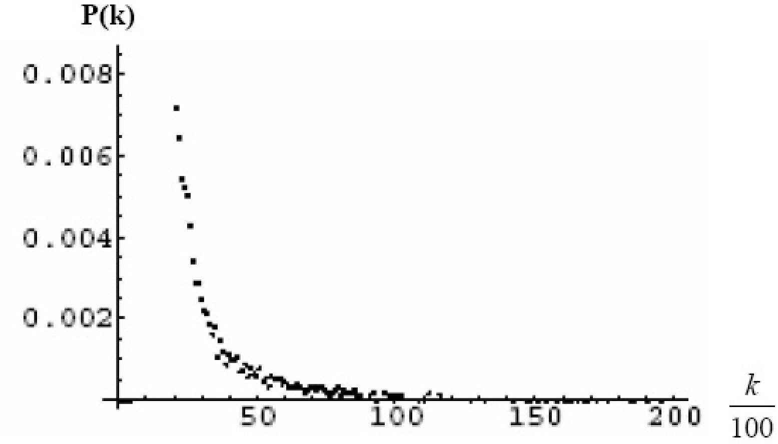

Fig1:degree distribution of model with $m0 = 6, m = 6$ when t=50000

4 Conclusion

We have proposed a precise model and we give the sufficient condition rather than a simple mechanism to power law degree distribution. By random elimination mechanics, our model gets power law degree distribution with exponents $-3+\frac{A}{m}$ which can describe many real biological networks. Yet,there are still many further works to be done. For our models itself,we consider only the properties of degree distribution,many other important properties have not been considered such as diameter of network and spectrum properties.

There are many other mechanism which can not covered by our sufficient condition but they can obtain the scale free properties. Chung etc.[4]gave a duplication model which can produce power law degree distribution with exponents less than 2. In their model, preferential choice was not taken into count. Why can different marginal distribution lead to similar power law degree distribution and what are the relations between them? These questions lead us to further research.

References

[1] E. Ravasz and A.-L. Barabasi,Hierarchical organization in complex networks. Physical ReviewE,67:026112,2003.

[2] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabasi,Hierarchical organization of modularity in metabolic networks. Science,297:1551-1555,2002.

[3] Zhihua Zhang, Changning Liu, Geir Skogerb, Xiaopeng Zhu, Hongchao Lu, Lan Chen, Baochen Shi, Yong Zhang Jie Wang,Tao Wu and Runsheng Chen,Dynamic changes in subgraph preference profiles of crucial transcription factors.PLOS COMPUT BIO,2(5):383-391,2006.

[4] Chung,F.,and L.LučňT.Dewey,D.Galas,Duplication models for biological networks.J.computational biology,vol10,No.5,677-687,2003.

[5] Aiello,W., F.Chung,and L.Lu,A random graph model for massive graphs. Proceedings of the 32nd ACM Symposium on the Theory of Computing (ACM, New York):171-180,2000.

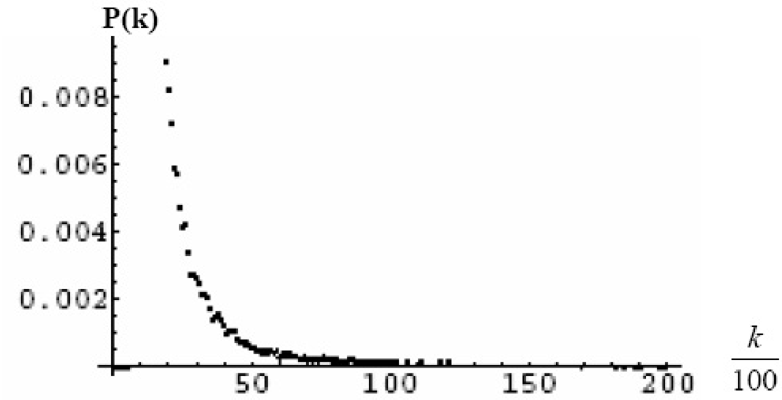

Fig2:degree distribution of model with $m0 = 6, m = 6$ when t=100000

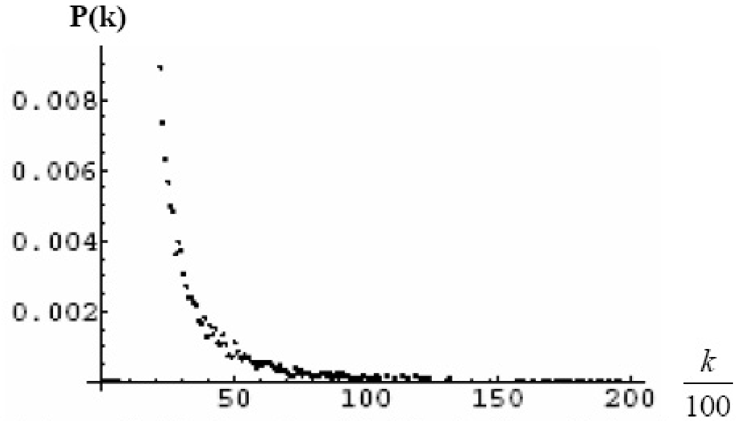

Fig3:degree distribution of model with $m0 = 7, m = 7$ when t=50000

[6] Wagner,A.,Evolution of gene networks by gene duplications:A mathematical model and its implications on genome or organization. Proc.Natl.acad.Sci.USA 91,4387-4391,1994.

[7] Friedman,R.,hughes,A.,Gene duplication and the structure of eukaryotic genomes.Gemome Res.11,373-381,2002.

[8] C.F.Du.Stationary properties of random networks and applications, Academy of Nathematics and systems Science,Chinese Academy of Sciences,Doctoral dissertation,2006.

[9] Colin Cooper, Alan M.Frieze, A general model of web graphs.Random Struct.Algorithms22:33, 311-335,2003.

[10] Z.M.Ma.Markovian random network processes,in Abstracts,The jointed meeting of The chinese Society and Statistics and the Institute of Mathematical Statictics,July 9-12,peking University,Beijng,China,Abstract 2(P.SL2),Special Lectures,P.1,2005.

[11] B. Bollobas, Mathematical results on scale-free graphs,Handbook of Graphs and Networks,1-34,2003.

[12] B. Bollobas, O.M. Riordan, J. Spencer, and G. Tusnady, The degree sequence of a scale-free random graph process.Random Structures and Algorithms 18,279-290,2001.