Vol 16 No 9, September 2007
1009-1963/2007/16(09)/2600-08
Chinese Physics
© 2007 Chin. Phys. Soc.
and IOP Publishing Ltd

Quasispecies distribution of Eigen model*

Chen Jia(陈 佳), Li Sheng(李 晟)†, and Ma Hong-Ru(马红孺)

Department of Physics, Shanghai Jiaotong University, Shanghai 200240, *China*

(Received 11 December 2006; revised manuscript received 24 January 2007)

We have studied sharp peak landscapes of the Eigen model from a new perspective about how the quasispecies are distributed in the sequence space. To analyse the distribution more carefully, we bring in two tools. One tool is the variance of Hamming distance of the sequences at a given generation. It not only offers us a different avenue for accurately locating the error threshold and illustrates how the configuration of the distribution varies with copying fidelity q in the sequence space, but also divides the copying fidelity into three distinct regimes. The other tool is the similarity network of a certain Hamming distance d_0, by which we can gain a visual and in-depth result about how the sequences are distributed. We find that there are several local similarity optima around the centre (global similarity optimum) in the distribution of the sequences reproduced near the threshold. Furthermore, it is interesting that the distribution of clustering coefficient $C(k)$ follows lognormal distribution and the curve of clustering coefficient C of the network versus d_0 appears to be linear near the threshold.

Keywords: Eigen model, sharp peak landscapes, similarity network, clustering coefficient
PACC: 0540, 0565, 8710, 6120J

1. Introduction

The Eigen model, which first considered that the evolution can be guided other than by chance, is the first mathematical model that makes Darwin's idea of mutation and selection able to work in a simple and seemingly 'lifeless' system of chemical reactants.[1,2] Two main points of the Eigen model are about the quasispecies which is defined as a stationary distribution of genetically close sequences, centred around one or several master sequences, and the existence of an error threshold above which all information is lost because of accumulating erroneous mutations. The formation of quasispecies implies that the target of selection in quasispecies theory is not a single master sequence but a localized distribution in the sequence space. After Leuthäusser mapped Eigen model on a well-known system in statistical mechanics, the two dimensional Ising system,[3] the study of evolution has become easier and clearer.[4−11]

Traditionally, people use the concentration of sequence[2,12,13] or the mean fitness of population[14−16] to measure the error threshold of evolution. Here, we advance a different method by utilizing the variance of Hamming distance of the sequences at a given generation to locate the error threshold of sharp peak landscape (SPL). This new useful method also gives the in-depth results about how the sequence is distributed in the sequence space.

Today, pairwise sequence comparison is the most widely used application of bioinformatics because high sequence similarity usually implies structural, functional or evolutionary relationship among bio-molecular sequences.[17] And the tool of complex network has been already brought in biological study.[18,19] It offers unforeseen possibilities to uncover the organizing principles that govern the formation and evolution of complex systems. Therefore, we also bring in another tool— similarity network to study the Eigen model.

During the evolution, we show that there are two jump discontinuities in the curve of the average of clustering coefficient $\langle C\rangle$ of the network versus copying fidelity q. This phenomenon implies that the phase transition of the similarity network topology occurs just as we expected. So we prove that it is logical to study evolution by bringing in similarity network. We analyse the topology of the similarity networks established in three different areas of q by two main basic measures, that is, the degree distribution and the clustering coefficient. And then we translate the results into biological language. Thereby, we gain a detailed result about how the sequences are distributed in the sequence space. For comparison, we also present the results of the random sequences created according to the random selection principle (select 3000 sequences

*Project supported by the National Natural Science Foundation of China (Grant Nos 10105007 and 10334020).
†E-mail: lisheng@sjtu.edu.cn

randomly from 2^{32} sequences in the sequence space) in this paper.

This paper is organized as follows. In Section 2, we briefly introduce the Eigen model. In Section 3, we introduce the similarity network and explain how its inherent properties describe the distribution of sequences. In Section 4, we study SPL to put our ideas in practice and demonstrate how the sequences generated in SPL are distributed in the sequence space. Finally we end our paper with concluding remarks in Section 5.

2. The Eigen model and sharp peak landscapes

The Eigen model was established by Eigen in 1971. The original model was aimed at describing self-replicating RNA or DNA molecules with the assumptions as follows:[2]

(i) The population of evolution is infinite. The individual is represented by a sequence of n nucleotides $S_k = \{s_1, s_2, s_3, ..., s_n\}$, where s_i contains four possible values to represent the four different bases in the RNA or the DNA molecules. But for simplicity, people usually take them to be binary variables which can be considered as distinguishing only between purines and pyrimidines. Conventionally, s_i is written in the binary form and S_k is thus a binary string (01000...) or equivalent to an integer b ($0 \leq b < 2^n$).

(ii) Only point mutations are considered. Then the mutation is an exchange of 0 and 1 on a certain locus. For the sake of mathematical tractability, the point mutation rate $1-q$ (q is the copying fidelity of each position of the sequence) is constant in time and independent of the position of sequence (uniform error model).

Years later, the Eigen model has been developed to be applicable in a much wider area. For instance, the results for finite populations have been gained in both static fitness landscapes[20−22] and dynamic fitness landscapes.[23,24] And the error threshold of finite population $1-q_f$ may be lower than the infinite population's.[22,24] Thus, for simulated simplicity, we apply the Eigen model to finite population in this article.

The equation for describing the reproduction-mutation process is

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{j=0}^{2^n-1} Q_{ij} A_j x_j - x_i \sum_{j=0}^{2^n-1} A_j x_j, \tag{1}$$

where x_i is the concentration of the ith sequence S_i with normalization $\sum_{i=0}^{2^n-1} x_i = 1$. Q_{ij} is the probability of obtaining the sequence S_i as the offspring of the replication of sequence S_j, and is given by

$$Q_{ij} = q^{n-d_{ij}}(1-q)^{d_{ij}}, \tag{2}$$

where d_{ij} is the Hamming distance between S_i and S_j. The set of all sequences together with the Hamming distance forms the sequence space. A_i named fitness is the reproduction rate of S_i. It means that the sequence S_i will reproduce A_i offspring every evolution step. The fitness landscape is then obtained by assigning a numerical value A_i to each point in the sequence space.[25] The notion of 'fitness landscape' is one of the most powerful concepts in evolutionary theory because the evolutionary process can be considered as a random walk in it.

A broad set of different fitness landscapes has been studied recently.[4,13,14,16,24,25] Most of the theoretical work has focused on a particular fitness landscapes, that is, the sharp peak landscapes,[2,4,5,13,26] in which there is a master sequence $S_m = \{s_1, s_2, s_3, ..., s_n\}$ with a larger fitness A_m, while all other sequences have uniform fitness (smaller than the master sequence). In mathematics, it could be expressed as

$$\begin{aligned} A_m[S_m] = A_0 \ \text{ and } \ A_i[S_i] \\ = A_1 < A_0 \ \text{ for any } \ S_i \neq S_m. \end{aligned}$$

Ever since Eigen found phase transition in SPL almost 35 years ago, Eigen's paper has generated substantial interest among the physicists. The error threshold, a minimal replication accuracy necessary to maintain the genetic information in the population, can be viewed as the critical point of a phase transition separating two regimes of the quasispecies evolution. Above the error threshold, the distribution of sequences is centred around the master sequence. Below it, the outcome of a replication event can be considered a random sequence. This fact is well established through a great many works.[2,13,26] The exact formulas for the threshold of ferromagnet and antiferromagnet in the SPL are (see Ref.[2])

$$q_{\text{ferror}} = \left(\frac{1}{A_0}\right)^{1/n}, \tag{3}$$

$$q_{\text{antiferror}} = 1 - \left(\frac{1}{A_1 A_0}\right)^{1/2n}. \tag{4}$$

The two equations here indicate that the threshold of ferromagnet is close to 1 and the antiferromagnetic

threshold is close to 0 for large n. In order to prepare for study of the SPL of the Eigen model, the similarity network will be introduced in the next section.

3. Similarity network

The study of complex networks has been initiated by a desire to understand various real systems.[27−32] The complex networks can describe a wide range of systems in nature and society. The direct-viewing and obvious practical relevance enlighten us to apply this useful tool to the evolutionary study. A network is composed of a set of nodes and edges.[18,19,27,33] In our similarity network, the node i represents the sequence S_i and every node is different from others. If d_{ij} (the Hamming distance between S_i and S_j) is no larger than the given Hamming distance d_0 ($d_0 \neq 0$), we will connect node i to node j. By comparing all the sequences of population, an undirected complex network is then established.[34] It is apparent that the similarity network will become a complete network if d_0 is close to n and there will be little edges in the network if d_0 is close to 0. The similarity network actually reflects the relationships between the sequences of population. It makes the study of the distribution of sequences easier and more vivid.

Generally speaking, the topology of a network can be quantified by two main basic quantities. One quantity is the degree distribution, $P(k)$, giving the probability that a selected node has exactly k links. The degree distribution allows us to distinguish between different classes of networks. It can offer us the approximate configuration about how sequences are distributed in the sequence space. For example, it helps us know whether the distribution is homogeneous or not in the sequence space. The other one is the clustering coefficient, a measure of the cliquishness of a typical neighbourhood, defined as $C_i = 2n_i/k_i(k_i-1)$, where n_i is the number of links among the k_i nearest neighbours of node i. People usually use the function $C(k)$ and clustering coefficient C of the network to measure the network's structure. $C(k)$ is defined as the average clustering coefficient of all nodes with k links. C is the mean clustering coefficient over all nodes in the network ($C = \langle C_i \rangle$), characterizing the overall tendency of nodes to form clusters. So the second quantity can picture the compactibility and the local property of the distribution of sequences. Therefore, we use similarity network to illustrate in detail how the sequences distribute in the sequence space and what the relationship is among the sequences of population by calculating the degree distribution and clustering coefficient in the coming section.

4. Quasispecies distribution of sharp peak landscapes

In this section, we will study the simple model, SPL, to put our ideas in practice. The error threshold of SPL will be located by utilizing the variance of Hamming distance of the sequences at a given generation. Besides, the distribution of sequences reproduced in this case will be characterized by analysing the topology of the similarity network.

In computer simulation aspect, we use a population of 3000 sequences with $n = 32$, $A_0 = 10$, $A_1 = 1$. The total number of population, the length and the fitness of sequence are invariant in the process of simulation. The initial condition satisfies that the homogeneous population is made up of 3000 zeros at $t = 0$. Because the sequence S_i will reproduce A_i offspring every evolution step, we need to choose 3000 sequences with the probability $n_i/\sum_i n_i$ from the offspring after one replication step in order to keep the total number of sequences fixed, where n_i is the number of sequence S_i and $\sum_i n_i$ is the total number of offspring.

Now we begin to realize our ideas. Firstly, we calculate the variance of Hamming distance of the sequences at a given generation to locate the error threshold and characterize how the configuration of the distribution varies with q in the sequence space, on the whole.

Figure 1 shows the variance of Hamming distance $\langle \mathrm{var}(d_{ij}) \rangle$ as a function of q for SPL where $\langle \cdots \rangle$ represents the average over 1000 evolution steps from $t = 9000$ to $t = 10000$ for the given q. It has been proved that it is long enough for the evolution in the SPL to balance at $t = 9000$ for the q chosen in Fig.1. $\mathrm{var}(d_{ij}) = \langle d_{ij}^2 \rangle_{t=t_0} - \langle d_{ij} \rangle_{t=t_0}^2$ where $\langle \cdots \rangle_{t=t_0}$ represents statistical average over different kinds of the sequences at a given generation $t = t_0$. The peaks of the curve in Fig.1 are located at $q = 0.9355$ and $q = 0.0336$, well consistent with the theoretical values of threshold for the corresponding parameters, $q_{\mathrm{ferror}} = 0.9306$, $q_{\mathrm{antiferror}} = 0.0353$. The deviation of the threshold between theory and simulation may be caused by the finite size effect. The jump disconti-

nuities are located at $q = 0.934$ and $q = 0.0344$.

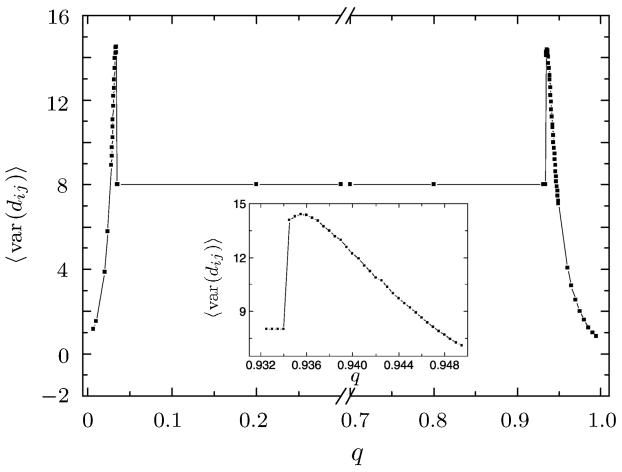


Fig.1. The average of the variance of Hamming distance of sequences $\langle \mathrm{var}(d_{ij}) \rangle$ versus copying fidelity q in SPL. $\langle \cdots \rangle$ represents the average over 1000 evolution steps from $t = 9000$ to $t = 10000$ for the given q. The peaks of the curve are at $q = 0.9355$ and $q = 0.0336$.

Take the ferromagnetic phase ($q > 0.5$) as an example to discuss the sharp peak landscapes. In ferromagnetic phase, we normally expect q is close to unity. The curve of the ferromagnetic phase in Fig.1 is apparently made up of three portions, that is, the area above the threshold, near the threshold and below the threshold. In the area far below the threshold, the curve of $\langle \mathrm{var}(d_{ij}) \rangle$ varying with q approximates to a horizontal line and its value is equal to $n/4$. It indicates that the configuration of the distribution far below the threshold is independent of q. In the area near the threshold, there is a peak and its amplitude will become large with the increase in n. The existence of the peak reflects that the scattered band of the distribution is wide and how complicated the distribution of the sequences generated in this landscapes is. In the area far above the threshold, $\langle \mathrm{var}(d_{ij}) \rangle$ is close to zero. It says that the distribution of sequences will converge in the sequence space if q approaches to unity.

Secondly, it is proved that it is logical to study the Eigen model by bringing in similarity network.

Figure 2 shows $\langle C \rangle$ as a function of q for SPL, where $\langle \cdots \rangle$ represents the average over 1000 evolution steps from $t = 9000$ to $t = 10000$ of the similarity network with $d_0 = 8$ for a given q. In Fig.2, there are two jump discontinuities at $q = 0.934$ and $q = 0.0344$, the same as those in Fig.1. We also have analysed the other similarity networks with different Hamming distance d_0. Most of them give the same result when d_0 is not very large. This phenomenon confirms that the phase transition of the similarity network topology indeed occurs during the evolution just as we expected. So it is logical to study evolution by bringing in similarity network.

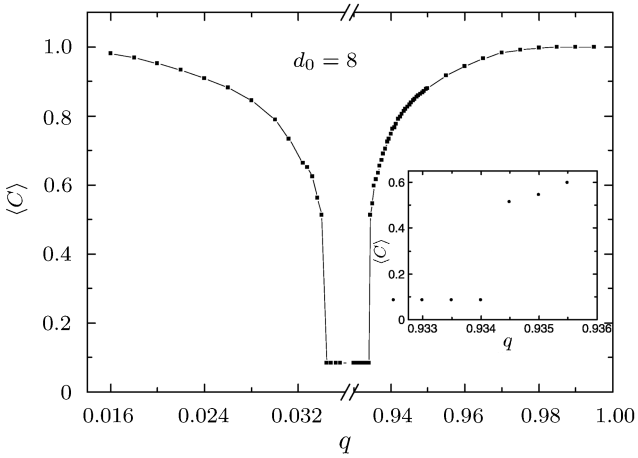


Fig.2. The average of clustering coefficient $\langle C \rangle$ of the similarity network with $d_0 = 8$ $\langle C \rangle$ versus copying fidelity q in SPL. $\langle \cdots \rangle$ represents the average over 1000 evolution steps from $t = 9000$ to $t = 10000$ of the similarity network with $d_0 = 8$ for a given q. The jump discontinuities are at $q = 0.934$ and $q = 0.0344$, the same as those in Fig.1.

Thirdly, the distributions of the sequences generated in the three areas are measured in depth by establishing similarity network. Furthermore, the results of random sequences are presented for further analysis.

Figures 3 and 4 show respectively the degree distribution and the function $C(k)$ of similarity networks for $q = 0.9355$ (at the threshold) and $q = 0.6$ (far below the threshold). Figures 6(a) and 6(b) show C as a function of d_0 for the similarity networks established by the sequences generated at the threshold and by the random sequences, respectively.

We now turn to analyse the topology of the similarity network to interpret the result of SPL. Near the threshold, $\langle C \rangle$ is large (see Fig.2). It shows that the similarity network of this case is compact, namely, the outcomes of the evolution in this case are genetically close to each other. In Fig.3, there is a relatively small number of nodes which possess a large number of links (large degree) but not large clustering coefficient ($C(k)$ is not large). They are known as hubs, which play an important role of bridging the many small communities of clusters into a single, integrated network. The presence of hubs suggests that the master sequence, the centre in the distribution,

exists assuredly in the quasispecies. Moreover, the degree distributions are oscillating in Figs.3(a) and 3(b) and the oscillation amplitude will become small if the ratio of A_0 to A_1 increases. It is more interesting that the curve of $C(k)$ follows lognormal distribution (see Figs.3(c) and 3(d)) and the curve of C versus d_0 appears linear (see Fig.6(a)).

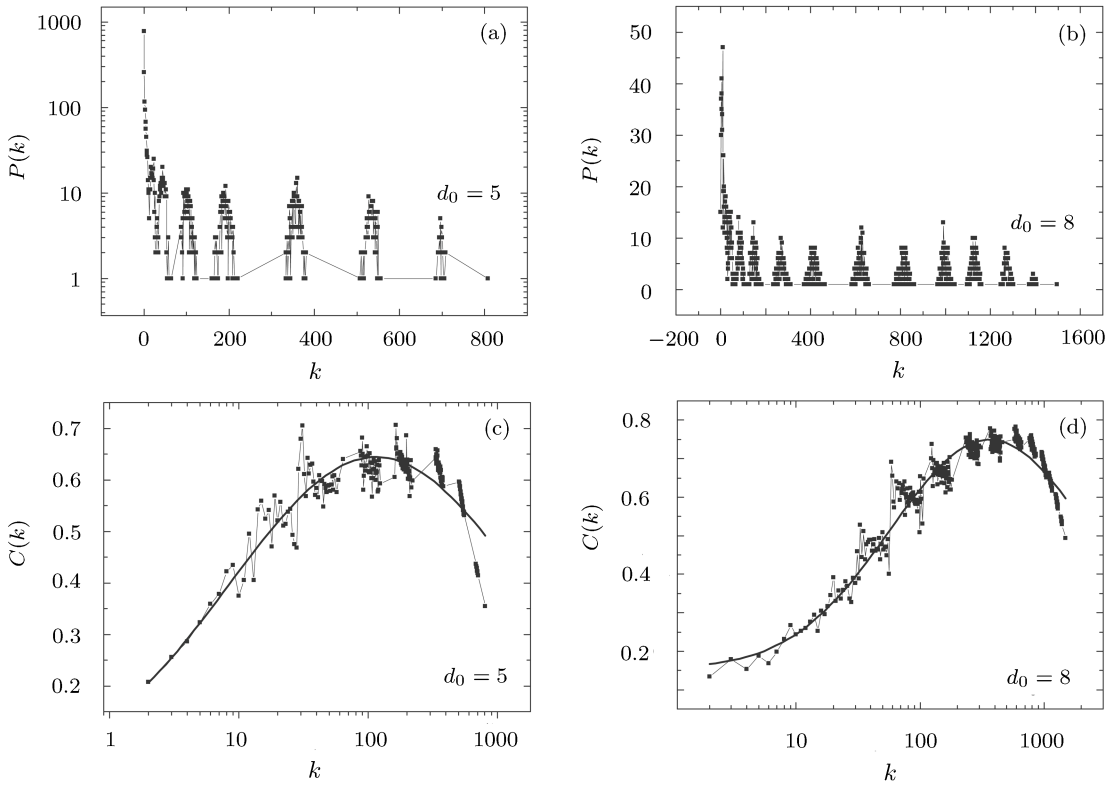


Fig.3. (a), (b) The degree distribution of the similarity network constructed in SPL with $q = 0.9355$ at $t = 10000$. (c), (d) The clustering coefficient $C(k)$ versus k of the same similarity network. The data of degree k have been binned logarithmically. The curves of $C(k)$ versus k follow lognormal distribution.

Far below the threshold, $\langle C\rangle$ is small (see Fig.2) and the curve of $\langle C\rangle$ versus q approximates to a horizontal line just as the curve of $\langle \mathrm{var}(d_{ij})\rangle$ versus q does. The degree distribution of the similarity network with $d_0 = 10$ follows the Gauss distribution in Fig.4(a) and the curve of $C(k)$ of the same similarity network is close to a horizontal line in Fig.4(b). For a deeper discussion, we introduce the random sequences created according to the random selection principle (3000 sequences are selected randomly from 2^{32} sequences in the sequence space). Figure 5 shows the simulations of random sequences. For random sequences, the value of $\langle \mathrm{var}(d_{ij})\rangle$ could be obtained as follows. We consider the special case: since the random sequences are distributed evenly in the sequence space, it is assumed that there are 2^n sorts of sequences in the population (n is the length of sequence). So, we have

$$\mathrm{var}(d_{ij}) = \langle d_{ij}^2\rangle - \langle d_{ij}\rangle^2 = \frac{2^n(2^n-n-1)}{(2^n-1)^2}\frac{n}{4}. \quad (5)$$

When n is large enough, Eq.(5) approximates to $n/4$. In Fig.1, the simulation results of $\langle \mathrm{var}(d_{ij})\rangle$ of the sequences created in SPL with $q = 0.6$ is equal to $n/4$ (consistent with the random sequences). In Fig.5(a), the degree distribution of the similarity network with $d_0 = 10$ follows the Gauss distribution, the same as that in Fig.4(a). In Fig.5(b), the curve of $C(k)$ of the same similarity network approximates to a horizontal line, the same as that in Fig.4(b). All results indicate that both of the two similarity networks possess the random network topology. Hence the outcomes of the evolution far below the threshold in SPL are equivalent to the random sequences distributing evenly in the sequence space.

Far above the threshold, the similarity network is very compact which is approximate to a complete network as $\langle C\rangle$ is close to 1 no matter what the value

of d_0 is ($d_0 \neq 0$) in this case (see Fig.2), implying that the sequences are highly similar to each other. The degree distribution and the function $C(k)$ also give the same result.

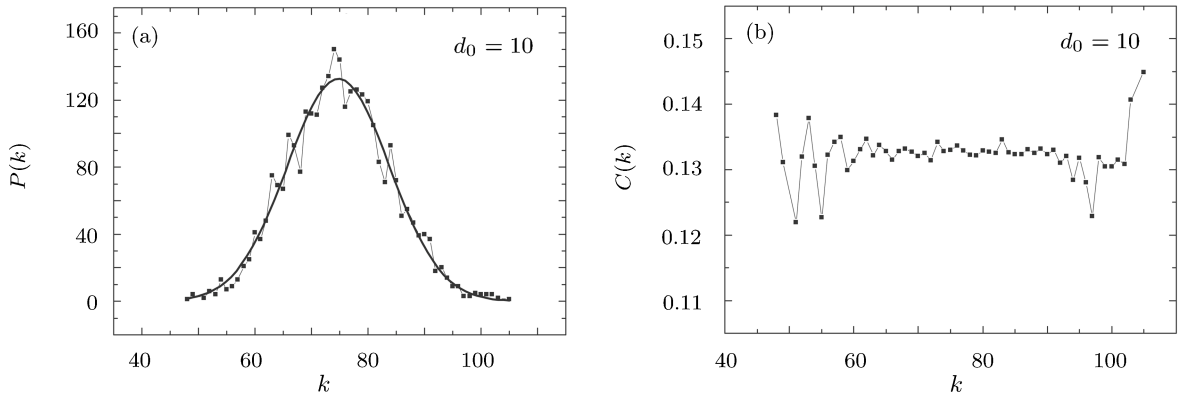


Fig.4. (a) The degree distribution of the similarity network with $d_0 = 10$ constructed in SPL with $q = 0.6$ at $t = 10000$. (b) The clustering coefficient $C(k)$ versus degree k of the same similarity network. The degree distribution follows a Gauss distribution when d_0 is large than 7. All the curves of $C(k)$ approximate to a horizontal line. And when d_0 is less than or equal to 4, the value of $C(k)$ is equal to 0.

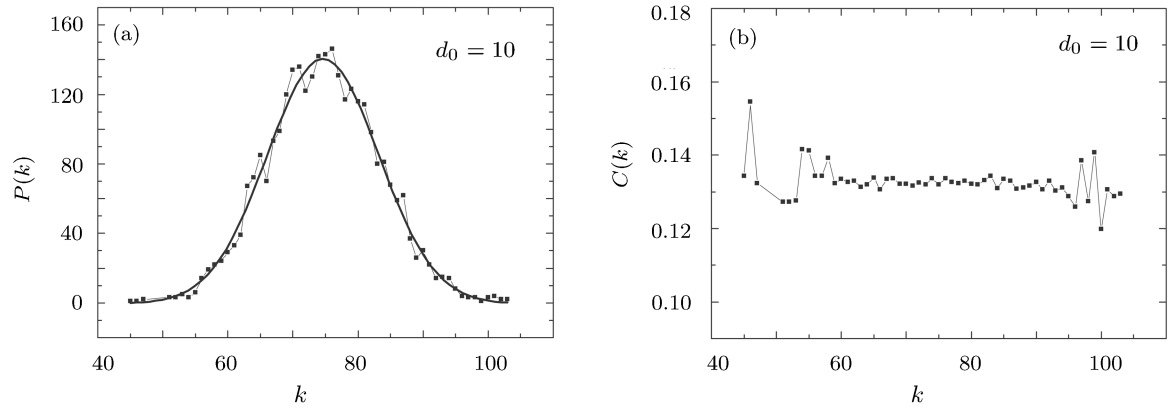


Fig.5. (a) The degree distribution of the similarity network with $d_0 = 10$ constructed by random sequences. (b) The clustering coefficient $C(k)$ versus degree k of the same similarity network. The degree distribution follows a Gauss distribution when d_0 is large than 7. All the curves of $C(k)$ approximate to a horizontal line. And when d_0 is less than or equal to 4, the value of $C(k)$ is equal to 0.

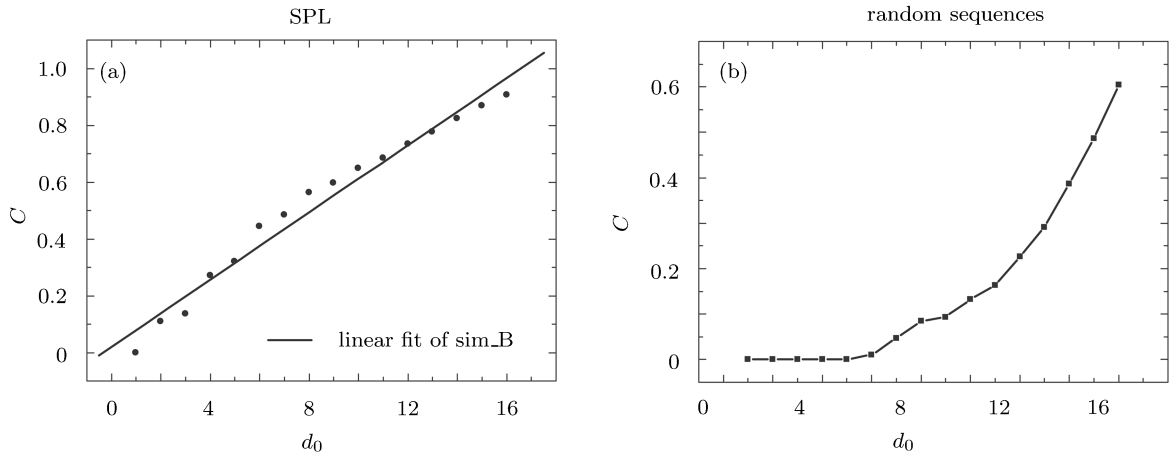


Fig.6. (a) Clustering coefficient C of the similarity network for $q = 0.9355$ at $t = 10000$ in SPL versus Hamming distance d_0. (b) Clustering coefficient C of the similarity network for random sequences versus Hamming distance d_0.

To sum up, far below the threshold, the values of both var(d_{ij}) and $\langle C\rangle$ are independent of q. The distribution is uniform in the sequence space. Near the threshold, the sequences have a large variance of Hamming distance var(d_{ij}) (see Fig.1) and a large clustering coefficient C of the network (see Fig.2), implying that the scattered band of the distribution is wide and the cliques have formed in the distribution of sequences. So, we can conclude that there are several local similarity optima around the centre (global similarity optimum) in the sequence space. Far above the threshold, the sequences have a small variance of Hamming distance var(d_{ij}) (see Fig.1) but a large clustering coefficient C of the network (see Fig.2). So the range of the distribution will reduce to one point in the sequence space if q approaches to unity. The optimum in this paper means the optimum of similarity, that is, there are several clusters around the hub in the network of sequence space. Finally, it is interesting that the curve of $C(k)$ follows lognormal distribution (see Figs.3(c) and 3(d)) and the curve of C versus d_0 appears to be linear (see Fig.6(a)) near the threshold. By the way, the results of antiferromagnet are the same as that of the ferromagnetic phase.

5. Conclusions

We have advanced two ways to study SPL from a different perspective. One way is the curve of $\langle \mathrm{var}(d_{ij})\rangle$ versus q. It not only offers us a different away for accurately locating the error threshold, but also shows how the configuration of the distribution varies with q in the sequence space. The other one is the similarity network, offering a new method to describe the evolution. We have shown that there is a phase transition in the curve of $\langle C\rangle$ versus q for SPL, which proves that it is logical to apply the similarity network to evolutionary study. Then, we further investigate the distribution of sequences by measuring the topology of similarity network. And the results are as follows. Far below the threshold, the values of both var(d_{ij}) and $\langle C\rangle$ are independent of q. The distribution is uniform in the sequence space. Near the threshold, the sequences have a large variance of Hamming distance var(d_{ij}) and a large clustering coefficient C of the network, implying that the scattered band of the distribution is wide and the cliques have formed in the distribution of sequences. So, we can conclude that there are several local similarity optima around the centre (global similarity optimum) in the sequence space. Far above the threshold, the sequences have a small variance of Hamming distance var(d_{ij}) but a large clustering coefficient C. So the range of the distribution will reduce to one point in the sequence space if q approaches to unity. Finally, it has been found that $C(k)$ follows a lognormal distribution and the curve of C versus d_0 is seen as a straight line near the threshold.

The oscillation amplitude will not remarkably change if the total number of population increases. We think this phenomenon may be caused by the existence of local similarity optima. From the viewpoint of statistics, as n increases, the total number of population in the simulation should also greatly increase in order to obtain good results. The computation is difficult for large population. So, we did not choose a large n for simplicity, which would not affect the results.

Our study for the first time applies the similarity network to study the Eigen model. We hope that this work can serve as a motivation for uncovering the mechanism of evolution. Moreover, finding a suitable way to offer a bench mark for evaluating various fitness landscapes to estimate how the real-valued landscapes is on earth will be considered in forthcoming work.

References

[1] Eigen M 1971 *Naturwissenschaften* **58** 465
[2] Eigen M, McCaskill J S and Schuster P 1989 *Adv. Chem. Phys.* **75** 149
[3] Leuthäusser I 1986 *J. Chem. Phys.* **84** 1884
[4] Franz S and Peliti L 1997 *J. Phys.* A **30** 4481
[5] Saakian D and Hu C K 2004 *Phys. Rev.* E **69** 021913
[6] Leuthäusser I 1987 *J. Stat. Phys.* **48** 343
[7] Tarazona P 1992 *Phys. Rev.* E **45** 6038
[8] Baake E, Baake M and Wagner H 1997 *Phys. Rev. Lett.* **78** 559
[9] Wagner H, Baake E and Gerisch T 1998 *J. Stat. Phys.* **92** 1017
[10] Galluccio S 1997 *Phys. Rev.* E **56** 4526
[11] Saakian D B, Muñoz E, Hu C K and Deem M W 2006 *Phys. Rev.* E **73** 041913
[12] Aoki K and Furusawa M 2003 *Phys. Rev.* E **68** 031904
[13] Tarazona P 1992 *Phys. Rev.* A **45** 6038
[14] Nilsson M and Snoad N 2000 *Phys. Rev. Lett.* **84** 191
[15] Tannenbaum E and Shakhnovich E I 2004 *Phys. Rev.* E **70** 021903
[16] Bonhoeffer S and Stadler P F 1993 *J. Theor. Biol.* **164** 359

[17] Liao L and Noble W S 2003 *J. Comput. Biol.* **10** 857
[18] Alm E and Arkin A P 2003 *Curr. Opin. Struct. Biol.* **13** 193
[19] Barabasi A L and Oltvai Z N 2004 *Nat. Rev. Genet.* **5** 101
[20] Alves D and Fontanari J F 1998 *Phys. Rev.* E **57** 7008
[21] Campos P R A and Fontanari J F 1999 *J. Phys.* A **32** L1
[22] Nowak M and Schuster P 1989 *J. Theor. Biol.* **137** 375
[23] Schmitt L, Nehaniv C L and Fujii R H 1998 *Theor. Comp. Sci.* **200** 101
[24] Wilke C O, Ronnewinkel C and Martinetz T 2001 *Phys. Rep.* **349** 395
[25] Weinberger E D and Stadler P F 1993 *J. Theor. Biol.* **163** 255
[26] Eigen M and Schuster P 1977 *Naturwissenschaften* **64** 541
[27] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[28] Hu K and Tang Y 2006 *Chin. Phys.* **15** 2782
[29] Li Y, Liu Y, Shan X M, Ren Y, Jiao J and Qiu B 2005 *Chin. Phys.* **14** 2153
[30] Wu J J, Gao Z Y and Sun H J 2007 *Chin. Phys.* **16** 0047
[31] Pan Z F and Wang X F 2006 *Acta Phys. Sin.* **55** 4058 (in Chinese)
[32] Liu H K and Zhou T 2007 *Acta Phys. Sin.* **56** 0106 (in Chinese)
[33] Ravasz E and Barabási A L 2003 *Phys. Rev.* E **67** 026112
[34] Wei F, Meng M, Li S and Ma H R 2006 *Mol. Phylogenet. Evol.* **38** 1