IISc-CTS/10/98

Emergence and Growth of Complex Networks in Adaptive Systems

Sanjay Jain[1*] and Sandeep Krishna[2†]
[1] *Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560 012, India*
[2] *Department of Physics, Indian Institute of Science, Bangalore 560 012, India*

Abstract

We consider the population dynamics of a set of species whose network of catalytic interactions is described by a directed graph. The relationship between the attractors of this dynamics and the underlying graph theoretic structures like cycles and autocatalytic sets is discussed. It is shown that when the population dynamics is suitably coupled to a slow dynamics of the graph itself, the network evolves towards increasing complexity driven by autocatalytic sets. Some quantitative measures of network complexity are described.

PACS numbers: 87.10.+e, 05.40.+j, 82.40.Bj, 89.80.+h
Keywords: evolution, complexity, autocatalytic sets, graph theory

Complex networks of interacting components are a characteristic feature of adaptive systems. Examples include prebiotic chemical evolution which produced complex organizations of molecules culminating in a living cell, biological evolution which produces complex ecologies with networks of interdependent species, and economic and social evolution in which webs of interacting agents appear spontaneously. It is of interest to understand how such webs originate and become more complex.

In this paper we discuss a model [1] which describes the network in terms of a directed graph, and treats it as a dynamical variable. The dynamics of the graph is determined by an underlying population dynamics on a fast time scale in which certain structures, autocatalytic sets (ACSs) [2], play an important role. ACSs arise spontaneously in the network and then trigger

* *jain@cts.iisc.ernet.in*
† *sandeep@physics.iisc.ernet.in*

an increase in its complexity. The main purpose of this paper is to describe the mathematical properties of ACSs and explain why they cause an increase of complexity of the network in this model.

Population dynamics on a fixed artificial chemistry

The system is described by a directed graph with s nodes. Associated with each node i is a population $y_i \geq 0$ which evolves in time according to

$$\dot{y}_i = \sum_{j=1}^{s} c_{ij} y_j - \phi y_i, \tag{1}$$

where ϕ is some function of time and $C \equiv (c_{ij})$, $i, j = 1, \ldots, s$ is the adjacency matrix of the graph, i.e., $c_{ij} = 1$ if there is a directed link from j to i and zero otherwise. Links from a node to itself are disallowed; $c_{ii} = 0$.

(1) can be interpreted as a set of rate equations in an artificial chemistry similar to the models studied in [3][4][5][6]. The nodes of the graph correspond to molecular species and a directed link from node j to node i indicates that species j catalyses the production of species i. Let i be produced by a reaction of the type $a + b \xrightarrow{j} i$, i.e., from the ligation of reactants a and b catalysed by j. In the approximation of a well stirred chemical reactor with a constant dilution flux ϕ, the rate of growth of i is given by $\dot{y}_i = k_f(1 + \nu y_j)ab - \phi y_i$, where a and b denote reactant populations, k_f is the rate constant for the spontaneous reaction, and ν is the catalytic efficiency [3]. Assuming that the spontaneous reaction is much slower than the catalysed reaction, and that the concentrations of the reactants are fixed and large, the first term will be proportional to y_j: $\dot{y}_i = k y_j - \phi y_i$, k being a constant. Making the further idealization that all catalytic strengths are equal the rate equations reduce to (1).

We are interested in the attractors of the relative population dynamics which follows from (1), namely

$$\dot{x}_i = \sum_{j=1}^{s} c_{ij} x_j - x_i \sum_{k,j=1}^{s} c_{kj} x_j, \tag{2}$$

where $x_i = y_i / \sum_{j=1}^{s} y_j$. By definition, the relative population vector $\mathbf{x} \equiv (x_1, \ldots, x_s)$ belongs to the simplex $J \equiv \{\mathbf{x} \in \mathbf{R}^s | 0 \leq x_i \leq 1, \ \sum_{i=1}^{s} x_i = 1\}$. J is invariant under (2) since $c_{ij} \geq 0$. Let $\mathbf{X} \equiv (X_1, \ldots, X_s)$ denote an attractive fixed point of (2), $\mathbf{y}^\lambda \equiv (y_1^\lambda, \ldots, y_s^\lambda)$ (or equivalent column vector) a right eigenvector of C with eigenvalue λ, and λ_1 the eigenvalue of C which has the largest real part. Then $\mathbf{x}^\lambda = \mathbf{y}^\lambda \ / \sum_{j=1}^{s} y_j^\lambda$ is a fixed point of (2). If

λ_1 is nondegenerate, $\mathbf{X} = \mathbf{x}^{\lambda_1}$ is the unique asymptotically stable attractor of (2) [1]. If λ_1 is degenerate the attractor configuration is still a fixed point; now $\mathbf{X}$ is a linear superposition of the $\mathbf{x}^{\lambda_1}$ that depends on the initial condition $\mathbf{x}(0)$. The Perron-Frobenius theorem for non-negative matrices ensures that λ_1 is real and ≥ 0. Table 1 shows a few simple graphs and their corresponding λ_1 and attractors. Chains and trees have $\lambda_1 = 0$. Simple cycles of any size have $\lambda_1 = 1$ while more complicated structures such as the double loop or the eight structure have higher λ_1.

Autocatalytic sets and spectral properties of directed graphs

An autocatalytic set (ACS) is a subset of the species that contains the catalysts for all its members [2]. In the present context, one can define an ACS as a subgraph, each of whose nodes has at least one incoming link from a node belonging to the same subgraph. The simplest ACS is a 2-cycle. Every cycle is an ACS but the converse is not true.

(i) *If a graph has no cycle then* $\lambda_1 = 0$. If C is the adjacency matrix of a graph then $(C^n)_{ij}$ counts the number of distinct paths of length n from node j to node i. If λ_i are the eigenvalues of C then λ_i^n are the eigenvalues of C^n. If a graph has no cycle then let the length of the longest path between any two nodes of the graph be denoted r. Clearly $C^m = 0$ for $m > r$. Therefore all eigenvalues of C^m are zero. Hence, all eigenvalues of C are zero which implies $\lambda_1 = 0$.

(ii) *If a graph has a cycle then* $\lambda_1 \geq 1$. If a graph has a cycle then there is some vertex i which has at least one path to itself of length n, i.e. $(C^n)_{ii} \geq 1$, for infinitely many values of n. Since the sum of the diagonal entries of a matrix equals the sum of the eigenvalues of the matrix, $\sum_{i=1}^{s}(C^n)_{ii}$ is equal to the sum of the n^{th} powers of the eigenvalues of C. Thus the sum of the n^{th} powers of the eigenvalues is at least 1, for infinitely many values of n. Therefore, there exists an eigenvalue with modulus ≥ 1. By the Perron-Frobenius theorem, λ_1 is the eigenvalue with the largest modulus, hence $\lambda_1 \geq 1$ [7].

An ACS is not in general an irreducible subgraph, because there need not be a path from every node of an ACS to another. Thus the Perron-Frobenius theorem for irreducible non-negative matrices [8] does not apply to ACSs. Nevertheless, ACSs share some important properties with cycles (which are irreducible):

(iii) *An ACS must contain a cycle.* Let C be the adjacency matrix of a graph which is itself an ACS. Then by definition, every row of C has at least one non-zero entry. Construct C' by removing all non-zero entries in each

row of C except one which can be chosen arbitrarily. Thus C' has exactly one non-zero entry in each row. Clearly the column vector $\mathbf{x} = (1, 1, \ldots, 1)$ is a right eigenvector of C' with eigenvalue 1. This implies that the graph for which C' is the adjacency matrix contains a cycle. Since the construction of C' from C involved only removal of some links, it follows that the original graph must also contain a cycle.

(iv) *If a graph has no ACS then* $\lambda_1 = 0$. This follows from (i) and (iii).

(v) *If a graph has an ACS then* $\lambda_1 \geq 1$. This follows from (ii) and (iii).

The reason why an ACS is a useful concept in the present context is the following property, not true of cycles in general:

(vi) *If a graph has* $\lambda_1 \geq 1$*, then the subgraph corresponding to the set of nodes* i *for which* $x_i^{\lambda_1} > 0$ *is an ACS.* Renumber the nodes of the graph so that $x_i^{\lambda_1} > 0$ only for $i = 1, \ldots, k$. Let C be the adjacency matrix of this graph. Since $\mathbf{x}^{\lambda_1}$ is an eigenvector of the matrix C we have $\sum_{j=1}^{s} c_{ij} x_j^{\lambda_1} = \lambda_1 x_i^{\lambda_1} \Rightarrow \sum_{j=1}^{k} c_{ij} x_j^{\lambda_1} = \lambda_1 x_i^{\lambda_1}$. Since $x_i^{\lambda_1} > 0$ only for $i = 1, \ldots, k$ it follows that for each $i \in \{1, \ldots, k\}$ there exists a j such that $c_{ij} > 0$. Hence the $k \times k$ submatrix $C' \equiv (c_{ij})$, $i, j = 1, \ldots, k$ has at least one non-zero entry in each row. Thus each node of the subgraph corresponding to this submatrix has an incoming link from one of the other nodes in the subgraph. Hence the subgraph is an ACS. We call this subgraph the 'dominant ACS' of the graph.

(vii) Consider a graph with no cycles and let there be a chain of r links in this graph whose successive nodes are labelled $i = 1, 2, \ldots, r+1$. The node 1 (to which there is no incoming link) has a constant population y_1 since the r.h.s of (1) vanishes for $i = 1$ when $\phi = 0$. For the node 2, we get $\dot{y_2} = y_1$, hence $y_2(t) = y_2(0) + y_1 t \sim t$ for large t. Similarly it can be seen that y_k grows as t^{k-1}. In general, it is clear that for a graph with no cycles, $y_i \sim t^r$ for large t when $\phi = 0$, where r is the length of the longest path terminating at node i. Since the dynamics (2) does not depend upon the choice of ϕ, this proves that *for a graph with no cycles* $X_i = 0$ *for all* i *except the nodes at which the longest paths in the graph terminate.* Similarly if a 2-cycle feeds into another 2-cycle as shown in Table 1, $X_i = 0$ for the nodes in the first 2-cycle. These are examples of the more general situation where $X_i = 0$ for nodes of a subgraph which has another subgraph 'downstream' from it with an equal or larger λ_1.

Evolution of the network

On a short timescale the graph remains fixed and the x_i evolve according to (2). On a longer timescale at discretely spaced intervals (labelled by

$n = 1, 2, \ldots$) the graph itself changes by the elimination of an existing species and the creation of a new one. At the initial time ($n = 0$) the graph is random: $c_{ij} = 1$ (for $i \neq j$) with probability p (p is called the 'catalytic probability') and zero with probability $1 - p$. $m \equiv p(s-1)$ is referred to as the average connectivity. The graph at $n+1$ is obtained from that at n by the following procedure: Determine the 'set of least fit nodes' at n, namely, those that have the smallest X_i for the graph at n. Pick a node at random from this set, and reassign links between it and all other nodes randomly with the same probability p. The idea is that for a fixed graph the relative population stabilizes at $\mathbf{X}$. One of the least fit species (X_i is taken to be a measure of fitness) mutates or is eliminated. This corresponds to an 'extremal dynamics' as in the model of Bak and Sneppen [9]. The graph update corresponds to the appearance of the mutant or replacement, which has random connections (but with the same average connectivity m as in the initial graph) with other species.

Define $s_1(n)$ as the number of species i for which $X_i \neq 0$ at the n^{th} time step. For a graph with $\lambda_1 \geq 1$, s_1 is the number of nodes in the dominant ACS. Whenever $s_1 < s$ it can be shown that the set of least fit nodes, which is the set of nodes with $X_i = 0$, is unique and independent of the initial condition even if λ_1 is degenerate. When $s_1 = s$, λ_1 has turned out to be nondegenerate in the runs displayed. Hence there is no ambiguity in the update procedure arising from initial conditions on $\mathbf{x}$.

Figures 1 and 2 show how λ_1 and s_1 evolve for a run with $s = 100$ and $m = 0.25$. For $n < n_1 = 1515$, $\lambda_1 = 0$ and there is no ACS in the graph. For $n < n_1$, s_1 is below 10 most of the time. At $n = n_1$ the graph update forms a 3-cycle which becomes the dominant ACS and λ_1 jumps from zero to one. The nodes outside the dominant ACS by definition have $x_i^{\lambda_1} = 0$ and constitute the set of least fit nodes. Therefore, as long as the dominant ACS at step n (let us denote this subgraph as A(n)) does not include the whole graph, i.e., $s_1 < s$, the mutating node will be outside it. At step $n+1$ the mutant species can either (a) get linked to A(n), or (b) form another ACS with other nodes which were not part of A(n), or (c) be a singleton or part of a non-ACS structure. In all cases $\lambda_1(n+1)$ cannot be less than $\lambda_1(n)$. (For case (a) this depends upon the fact that the mutating node, being outside A(n), cannot destroy any of its links.) Thus, *whenever* $s_1 < s$, λ_1 *is a non-decreasing function of* n. It follows that once an ACS is formed by chance, the autocatalytic property of the graph will be preserved until the dominant ACS engulfs the whole graph.

After the appearance of the ACS at $n = n_1$, whenever a node gets an

incoming link from the dominant ACS it becomes part of the dominant ACS. Most of the time this increases s_1 and the dominant ACS grows until it spans the entire graph at $n = n_2 = 3010$ when $s_1 = s$ for the first time. For $n \in [n_1, n_2]$, s_1 locally averaged in time grows exponentially.

In the given run a 2-cycle disconnected from the existing dominant ACS is created $n = 1807$. Both the cycles coexist and grow together until $n = 2145$ when the graph update creates a chain from the 3-cycle to the 2-cycle. Thus a situation occured in which a subgraph with $\lambda_1 = 1$, comprising the 3-cycle and the nodes being fed by the 3-cycle, was now feeding into the 2-cycle 'downstream' which also had $\lambda_1 = 1$. Hence all the nodes of the first subgraph entered the set of least fit nodes (see remarks under (vii)) and s_1 decreased from 15 to 10. Later at $n = 2240$ the chain joining the cycles was broken by the graph update and both cycles became part of the dominant ACS. At $n = 2607$, s_1 decreases from 58 to 31. This time also a chain formed between the cycles with the 2-cycle now feeding into the 3-cycle. s_1 rises back up to 54 at $n = 2611$ when this chain too was broken by the graph update. There is one more fall in s_1 at $n = 2780$. The graph update results in the 3-cycle being converted to a double loop which has $\lambda_1 \approx 1.17$ which overshadows the ACS whose core is the 2-cycle.

Even in regions where s_1 is increasing the ACS structure could be changing quite a lot. Many complicated processes are possible, such as formation of new disconnected ACSs or the reinforcement of an old overshadowed ACS before it was completely broken up. Thus the actual growth of the ACS is usually a very complicated process with the dominant ACS often undergoing drastic changes in structure caused by purely chance events. We wish to emphasize, however, that notwithstanding the historical particularities of a given run, every history respects the rule that λ_1 is a nondecreasing function of time (unless $s_1 = s$) in this simple model. (In a 'non-extremal' dynamics, one can expect a stochastic version of this monotonicity, as long as selection is still sufficiently strong.) Further, the *ensemble* of runs with the same parameter values m, s has identifiable characteristics like the average time of appearance of an ACS $\tau_a \equiv \langle n_1 \rangle \sim s/m^2 \sim 1/(p^2 s)$, and the exponential growth time scale $\tau_g \sim s/m\ 1/p$ (for sufficiently small m and large s) [1].

At $n = n_2$ the whole graph becomes an ACS and for the first time the graph update will alter a node from the dominant ACS. Then λ_1 can decrease. It eventually settles in a statistical steady state with large fluctuations in the run shown. For very low m, λ_1 can even become zero in a graph update.

In addition to λ_1, s_1 and the total number of links l, another measure

of the complexity of the graph is 'interdependency', denoted $\bar{d}$, which we define as follows: $\bar{d} \equiv (1/s) \sum_{i=1}^{s} d_i$, where d_i is the 'dependency' of the i^{th} node. d_i is the total number of links in all paths that terminate at node i, each link counted only once. Since d_i counts how many links ultimately 'feed into' the node i, it is a measure of how 'dependent' species i is on other species. Thus $\bar{d}$ is a measure of how interdependent the species in the graph are. From Table 1 it is evident that graphs with the same number of links can have different $\bar{d}$.

Figure 3 shows $\bar{d}$ versus n for the same run (the lower curve in the figure corresponds to a 'random run' in which the node to be eliminated is picked randomly from the entire set of s nodes). The qualitative features are similar to l which has been discussed in [1]. A quantitative difference is that from $n = 0$ to the steady state, $\bar{d}$ increases by a factor of approximately 50 which is much more than the increase in l (a factor of about 5, see [1]). The evolution of the system does not seem to lead to maximally dense graphs, $m^* \sim s$, but only $m^* \sim O(1)$ (m^* is the steady state connectivity). Nevertheless the system achieves a high interdependency. A steady state $\bar{d}$ value of about 25 means that each species is being fed with an appreciable fraction of the total number of links, suggestive of long range correlations having developed in the system. The fluctuations of $\bar{d}$ in the steady state are also more pronounced than for l. At each graph update the change in the number of links is $O(1)$ for fixed m but the addition or removal of a few links causes large changes in $\bar{d}$, reflecting its nonlocal character.

Acknowledgement: S. J. acknowledges the affiliation and support of the Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, as well as Associate membership and hospitality of the Abdus Salam International Centre for Theoretical Physics, Trieste.

References

[1] S. Jain and S. Krishna, Autocatalytic sets and the growth of complexity in an evolutionary model, preprint no. IISc-CTS/9/98, *adap-org/9809003*.

[2] S. A. Kauffman, J. Theor. Biol. **119**, 1 (1986); *The Origins of Order* (Oxford University Press, Oxford, 1993).

[3] J. D. Farmer, S. A. Kauffman and N. H. Packard, Physica **D22**, 50 (1986); R. J. Bagley and J. D. Farmer, in *Artificial Life II*, edited by

C.G. Langton et al., (Addison Wesley, Redwood City, 1992), p. 93; R. J. Bagley, J. D. Farmer and W. Fontana, in *Artificial Life II*, p. 141.

[4] W. Fontana, in *Artificial Life II*, p. 159; W. Fontana and L. W. Buss, Bull. Math. Biol. **56**, 1 (1994).

[5] P. F. Stadler, W. Fontana and J. H. Miller, Physica **D63**, 378 (1993).

[6] D. Segré, D. Lancet, O. Kedem, and Y. Pilpel, Origins of Life and Evolution of the Biosphere **28**, 501 (1998); D. Segré and D. Lancet, in *Chemical Evolution: Exobiology, An Abdus Salam Memorial*, edited by J. Chela-Flores and F. Raulin (Kluwer, Dordrecht, The Netherlands, 1998) p. 123.

[7] The above proofs of statements (i) and (ii) were supplied to us by V. S. Borkar and R. Hariharan, respectively.

[8] E. Seneta, *Non-Negative Matrices* (George Allen and Unwin, London, 1973).

[9] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).

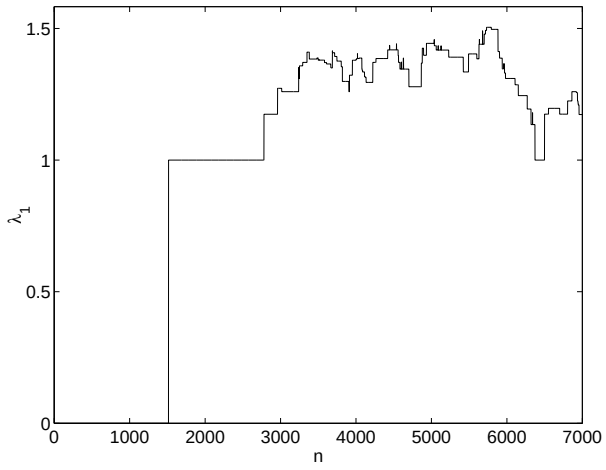


Figure 1: λ_1 versus n for $s = 100, m = 0.25$.

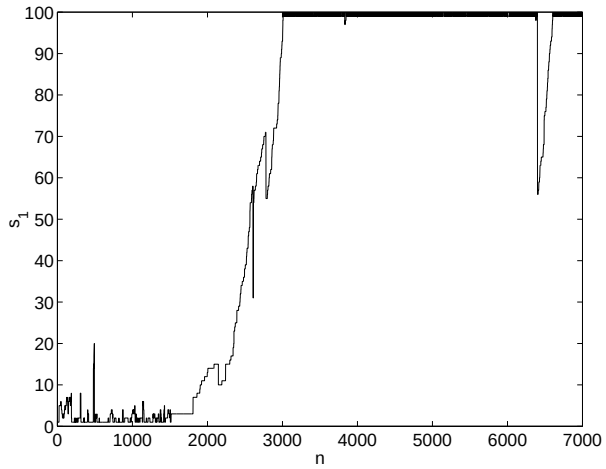


Figure 2: s_1 versus n for $s = 100, m = 0.25$.

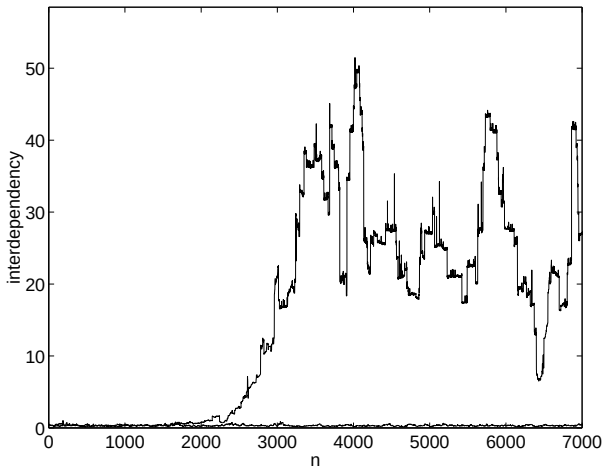
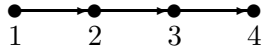
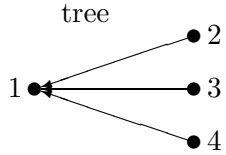
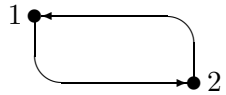
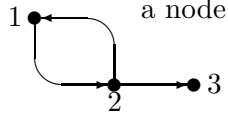
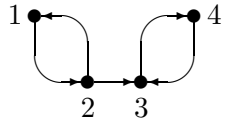
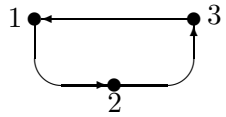
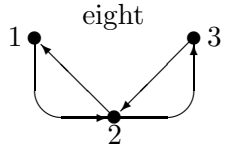
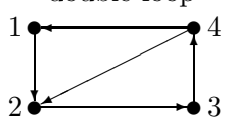


Figure 3: Interdependency versus n for $s = 100, m = 0.25$. The lower curve is for a random run with $s = 100, m = 0.25$.

Table 1: Adjacency matrices, largest eigenvalues , the corresponding eigenvectors and interdependency of some simple graphs

Graph	C	λ_1	$\mathbf{x}^{\lambda_1}$	$\bar{d} = (d_1 + d_2 + \ldots + d_s)/s$
chain 	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	0	$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$	1.5=(0+1+2+3)/4
tree 	$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	0	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	0.75=(3+0+0+0)/4
2-cycle 	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	1	$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$	2=(2+2)/2
2-cycle feeding into a node 	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	1	$\begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$	2.33=(2+2+3)/3
2-cycle feeding into a 2-cycle 	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	1	$\begin{pmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix}$	3.5=(2+2+5+5)/4
3-cycle 	$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	1	$\begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$	3=(3+3+3)/3
eight 	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$\sqrt{2}$	$\begin{pmatrix} 0.293 \\ 0.414 \\ 0.293 \end{pmatrix}$	4=(4+4+4)/3
double loop 	$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	1.22	$\begin{pmatrix} 0.18 \\ 0.33 \\ 0.27 \\ 0.22 \end{pmatrix}$	5=(5+5+5+5)/4