We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Obtaining Knowledge of Genes' Behavior in Genetic Regulatory System by Utilizing Multiagent Collaborative Computation

Adang Suwandi Ahmad[1] and Arwin Datumaya Wahyudi Sumari[2]
[1]School of Electrical Engineering and Informatics, Institut Teknologi Bandung,
Achmad Bakrie Building, 2nd Floor, Jl. Ganeca 10, Bandung, 40132,
[2]Department of Electronics, Indonesian Air Force Academy,
Jl. Laksda Adisutjipto, Yogyakarta, 55002
Indonesia

1. Introduction

All living organisms are built from millions of cells that work together in a very systematic manner. The trait of each organism is determined by the cells. The function of a living cell is performed by a sequence of well-coordinated activities by a large number of genes. The gene itself is a sequence of Deoxyribonucleic Acid (DNA). The products of a cell are made by the proteins synthesized in accordance with the instructions contained in the DNA (Willett, 2006) since it encodes all the information required for the development and functioning of an organism (Emmert-Streib & Dehmer, 2008).

The expression of a gene is a biological process which a DNA sequence is translated to become a protein. The protein is an important molecule for determining the structure, mortality, metabolism, signaling, reproduction, etc. of a cell. Some genes influence how other gene or genes are expressed. Modifying genes may change the affect of other genes. The expression of a gene or genes is regulated by a mechanism called Genetic Regulatory System (GRS). The GRS' task is to control whether genes are active or inhibit.

Of ways for understanding and analyzing the behavior of GRS is by using microarray data. Microarray is sets of miniaturized reaction areas that may also be used to test the binding of DNA fragment (Reece, 2004). and it exploits the preferential binding of complementary nucleic acid sequences to simultaneously measure expression levels of thousands of genes (Dill, Liu, & Grodzinski, 2009). By having knowledge regarding the influence of a gene or genes to the others in microarray data, we can make estimations of the genes behavior in GRS in order to obtain better products in the future. According to (Ewens & Grant, 2005), there are three basic questions that can be answered by using microarray data:

a. What genes are expressed in a given sample?
b. Which genes are differentially expressed between different samples?
c. How can one find different classes, or clusters, of genes which are expressed in a correlated fashion across a set of samples? How can one find different classes of samples based on their gene expression behavior?

In this paper we address the utilization of Multiagent Collaborative Computation (MCC) to model the genes' interactions in GRS to address questions (a) and (b). Simply, each gene will be represented by an agent that performs internal activities and performs communications with other genes as occurred in biological genes. On each interaction time, we can have knowledge regarding individual gene expressiveness and all genes behavior. The process in obtaining this knowledge will be carried out by Knowledge-Growing System (KGS) based on Observation Multi-time Arwin-Adang-Aciek-Sembiring (OMA3S) information-inferencing fusion and the results will be compared to the MCC results. The comparison has also been done to ensure the accuracy of the KGS-based MCC results. In our research, we use data of yeast cycle database taken from [6] as the case study. The rest of the paper is structured as follows. The concept of MCC and a review of the state-of-the-art of GRS research will be delivered in Section 2. In Section 3, we present the concept of KGS as well as OMA3S method that is used as its knowledge-growing mechanism. The utilization of MCC to GRS as well as the knowledge production by KGS will be delivered in Section 4. Finally, the paper is concluded in Section 5 with some concluding remarks.

2. State-of-the-art of genetic regulatory system research and the concept of multiagent collaborative computation

2.1 State-of-the-art on genetic regulatory system research

The research, studies, and experiments on GRS have been done widely all over the world. These efforts are aimed to one objective, namely how to obtain the most suitable model in order to present the mechanism occurs in GRS. In these endeavors, there are two approaches that have been carried out as studied comprehensively by (Ahmad & Sumari, 2009a) as follows.

In the first approach the GRS mechanism is directly modeled by utilizing available methods such as Dynamic Bayesian Network (DBN) based on knowledge growing, weight matrices, petri nets, artificial gnome model, Neural Network (NN) or its combination with fuzzy technique such as Evolving Connectionist System (ECOS), piece-wise linear, Boolean, and power graph analysis.

The second one is to model the GRS by inferring its structure utilizing techniques such as best-fit extension problem, linear programming combined with supervised learning framework, DBN, evolutionary computation, graph and Linearized Additive Model (LAM), steady-state gene expression, Dynamic Differential BN (DDBN), combination of DBN with Reversible Jump Markov chain Monte Carlo (RJMCC) or with Greedy search algorithm and Markov Chain Monte Carlo (MCMC), Bayesian and MCMC, iterative algorithm based on epistemic approach of conjecture and refutation, NN, and combination of clustering techniques with NNs.

At the time of the writing of this paper, we have not found literatures that study methods in obtaining knowledge regarding the behavior of genes in GRS in order to estimate their expression in the future especially that utilizes multiagent approach. This is the primary matter that will be discussed in this paper. Our hypothesis is by accurately estimating the GRS behavior in the future we can carry out anticipation actions in order to prevent the negative expression of genes that can cause negative effects to living organisms.

2.2 Multiagent collaborative computation

MCC is a new paradigm in MultiAgent System (MAS). It is the emulation of how human beings work in collaborative manner in solving problems or finding solutions of given

problems or in achieving a common or joint goal or objective, by sharing resources they have for the successfulness of their work. Working in collaborative manner does not need leader. What it needs is a divider agent for dividing the work or task to be solved or given into subworks or subtasks, and distributes them to some collaborative agents, and a fusion agent that is tasked to fuse the results from collaborative agents to obtain a comprehensive result or joint goal. By doing it, the subworks can be processed in parallel manner to achieve the subgoals, and therefore it can save the processing time. Before we go deeper into MCC, we start this section with the reintroduction of the concept of agent and the concept of collaborative computation as well as its merit.

a.   The Concept of Agent (Ahmad & Sumari, 2008)

There have been many literatures that study agent, but in this section we view an agent from a comprehensive perspective. In real life, agent is defined as a thing that causes a significant effect on a situation. In order to give this effect, agent must have capabilities. "Capability" in this circumstance is the ability to manage when the tasks will be carried out, knows where to move, knows how to do the tasks, knows the success level of the tasks being carried out, and the consequences of the tasks being done. The essential thing that enables the agent in performing its activities is the brain, a place where the information processing is carried out. This is the most grandeur that is not possessed by other living things.

In accomplishing the assigned tasks, the agent always senses its surrounding environment to get as much as information that can affect its activities. The gathered information is processed in its brain, combining it with the existing information, making inferencing on the fused information, performing information-inferencing fusion, and making the best decision for actions to be done in anticipating the environment dynamics. The relation of an agent with its surrounding environment as well as its information processing mechanism is simply depicted in Fig. 1.
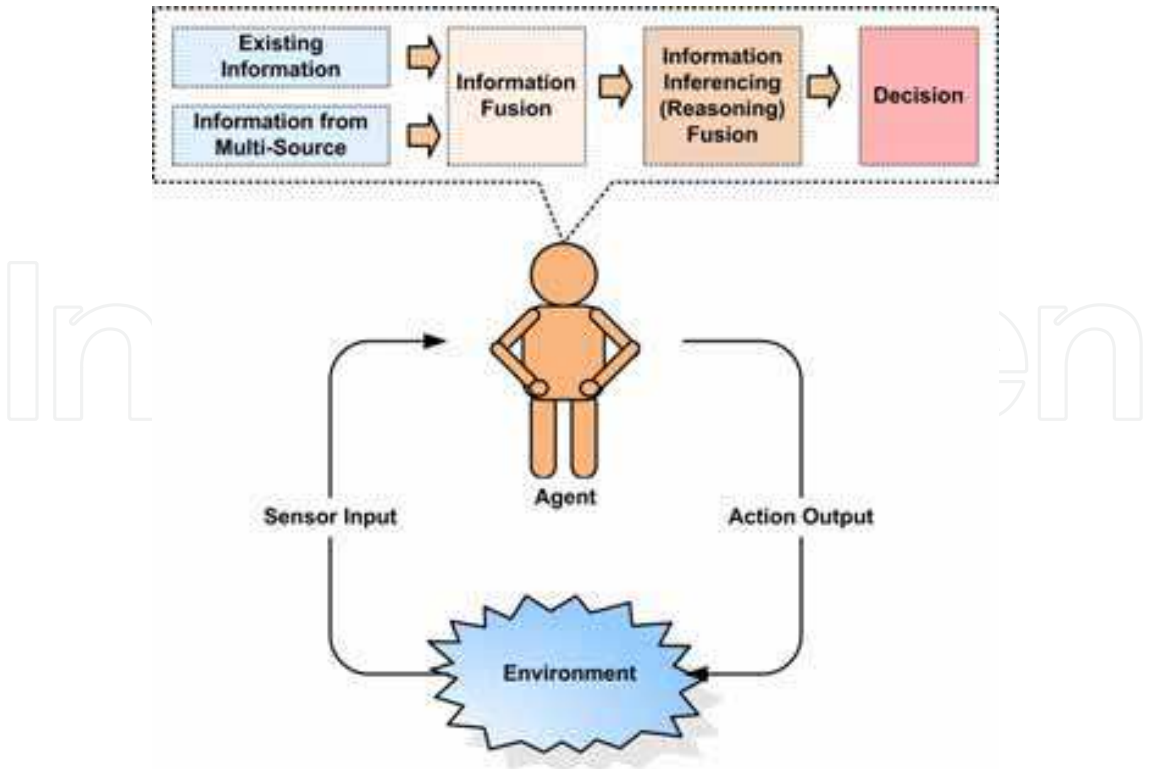

Fig. 1. The concept of agent along with its information-inferencing fusion capability

In many literatures, agent is stated must have an intelligent characteristic. However, because there is no uniform definition of what agent is, one common consensus is taken that the autonomy or we call this as self-governing, as its the essential characteristic. Self-governing means the agent has capability to instruct itself to accomplish the tasks and do self-evaluation to value the success rate of the assigned tasks accomplishment for future enhancement.

b.   The Concept of Collaborative Computation

Collaboration is taken from the Latin word "collaborare" meaning "to work together", while "collaborative" is an adjective form of "collaboration". So, collaboration is defined as to work with another or others on a joint project or in deeper definition is a process in which
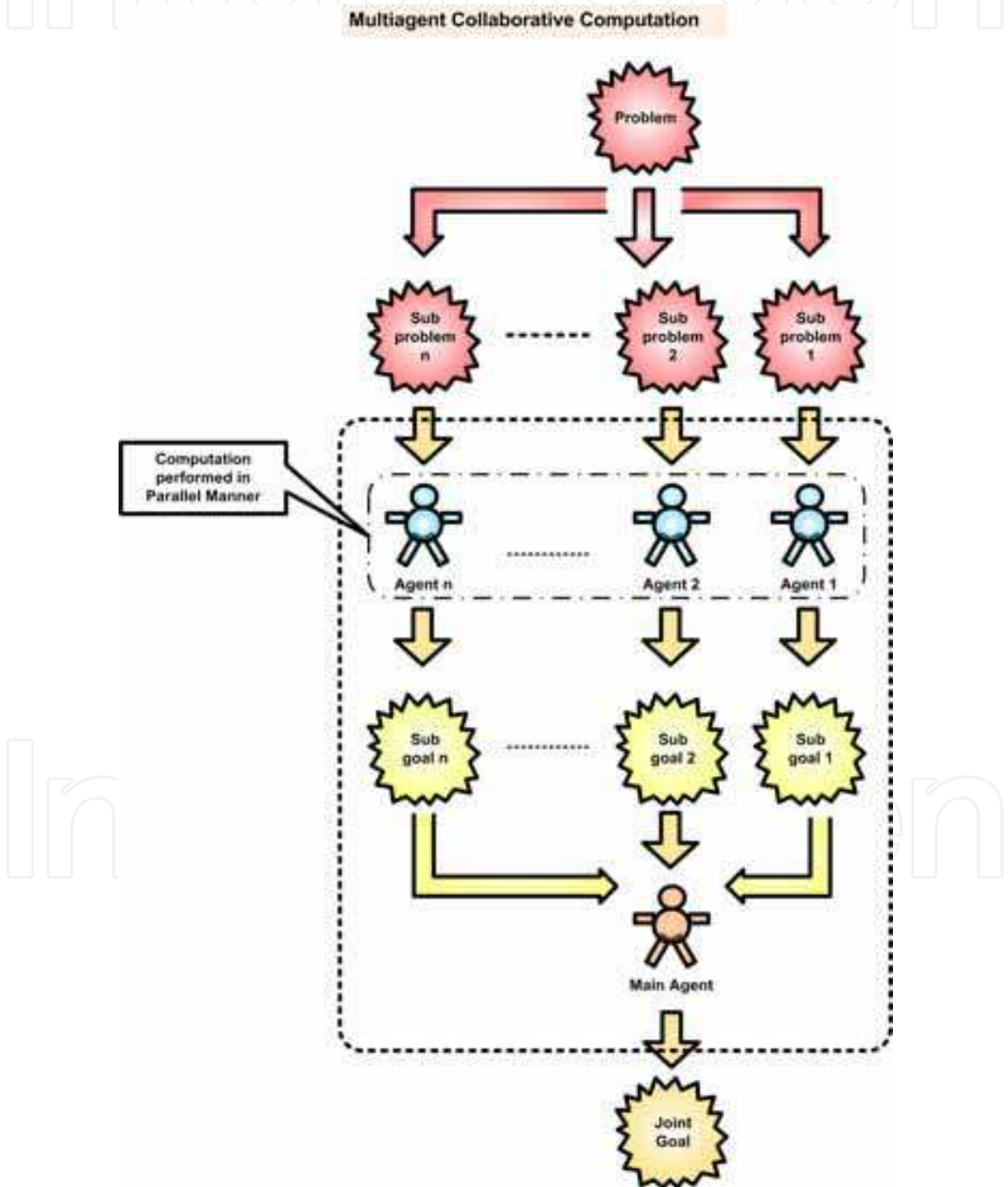

Fig. 2. The concept of collaborative computation in a group of agents (Sumari & Ahmad, 2009)

entities share information, resources and responsibilities to jointly plan, implement, and evaluate a program of activities to achieve a common goal (Camarinha-Matos & Afsarmanesh, 2008).

In collaborative scheme, a group of entities enhance the capabilities of each other that implies in sharing risks, resources, responsibilities, and reward. Collaboration involves mutual engagement of participants to solve a problem together, which implies mutual trust and thus takes time, effort, and dedication. On the other side, "computation" is defined as a calculation involving numbers or quantities. By employing computational approach, we can manipulate a large database to solve problems at hand very fast. The consequence is faster, more complete, and more accurate results than that can be achieved by conventional approach.

By combining the two definitions previously explained, we define "collaborative computation" as a calculation on quantities done by a group of agent or multiagent that works together to achieve common or joint goals (Ahmad, Sumari, & Zubir, 2009) as depicted in Fig. 2. In an MCC scheme, each agent in the collaborative framework performs its own tasks to achieve its own goals. The goals achieved by each agent actually are partial parts of joint goal, which is the combination of partial goals. In simple word, the system is given a problem along with the goal that has to be achieved. The problem is then divided into several sub-problems that have to be solved by the agents. In achieving the sub-goals, the agents perform computation in parallel. The achieved sub-goals are then combined to become a single comprehensive goal that is called as joint goal (Sumari & Ahmad, 2009).

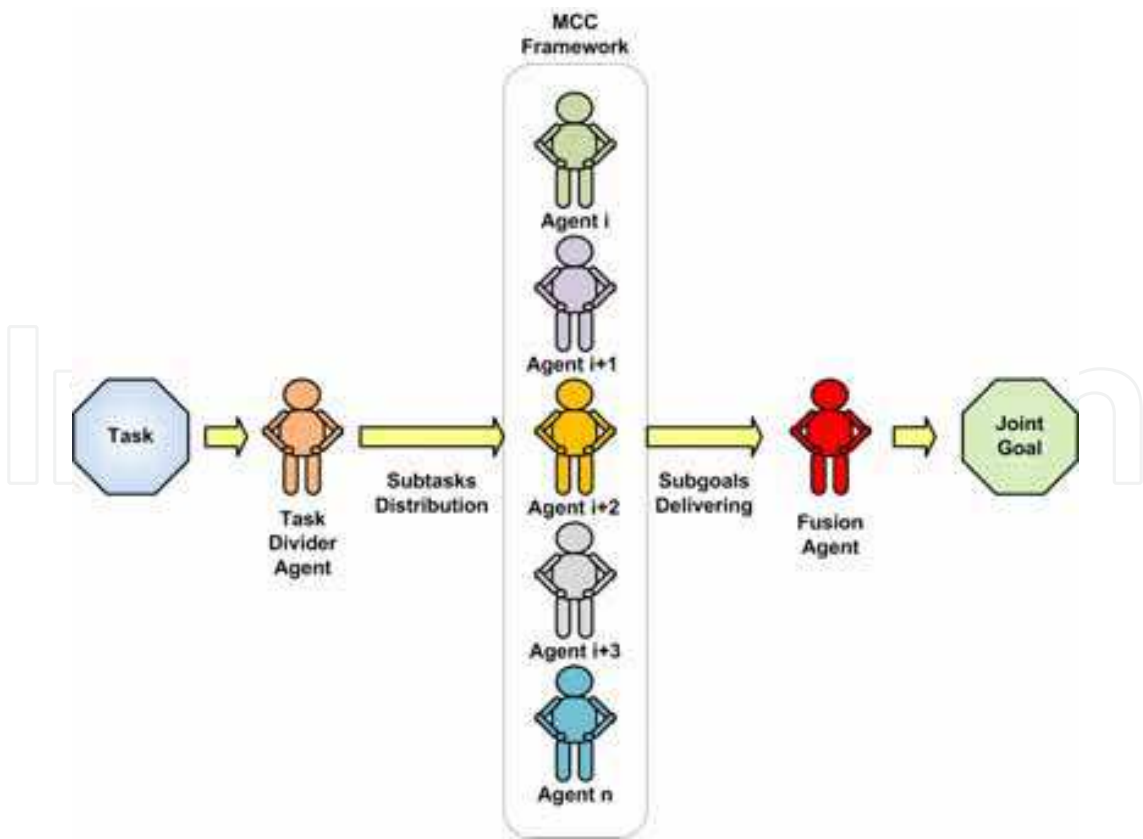c. The Merits of Multiagent Collaborative Computation Approach (Sumari, 2010)

Fig. 3. The concept of MCC (Sumari, 2010)

GRS is a very complex structure that regulates millions of genes in living organism's body. We believe that multiagent approach is suitable to model the genes' interactions in GRS with some strong reasons. MAS is aimed to cope with large-scale, realistic, and complex problems that cannot be handled by single agent in order to find a solution of global problems or to regulate or control complex systems. In MCC framework, the given task will be divided into subtasks according to each agent specific task. By performing this, the information processing can be done in simultaneous manner by all agents so that the processing time can be reduced to minimum.

Assume that we have n agents in MCC framework. Given a task that a single agent can process it in t time. If the task is divided into n subtasks and distributed them to n agents, theoretically, the result of the process can be obtained in t/n time with total $t/n + \tau$ where τ is the time reserved for dividing the task and fusing or combining the subgoals into a joint goal or the ultimate result. Therefore, by utilizing MCC the information-processing time can be reduced from t to $t/n+\tau$, or $(t/n+\tau) <t$. This mechanism is illustrated in Fig. 3.

3. Knowledge-growing system and OMA3S information-inferencing fusion method

3.1 What is knowledge-growing? (Sumari et.al, 2010a)

It is not easy to find any literature that defines or describes what the term "Knowledge-Growing" is. The only research that used this term was for industrial application, that is, examined how the knowledge growing old in human brain and used the analogy of it for building a reconfiguration system for car application. It concentrated on the optimization of knowledge retrieval rather than emulating the way of human brain grows the knowledge over time, and used actuality measurement to measure the knowledge that is very often used to solve an actual problem.

Up to the writing of this paper, there is no research that examines and develops an intelligent method for an intelligent system that emulates the mechanism how the human brain grows the knowledge from time to time. Our approach in developing KGS was started from our observation to an intelligence characteristic displayed by human brain in performing such matter by fusing information perceived by human sensory organs and delivers it to the brain. This approach is discussed in detail in (Ahmad & Sumari, 2008). The original concept of information fusion, even though it also adopted the mechanism occurs in living things' brain, it is just defined how to fuse the information for making estimation regarding a phenomenon, see (Hall & Llinas, 2001) for the detail.

3.2 The concept of KGS

Essentially, KGS is a system that is capable of growing its knowledge along with the accretion of information it receives as the time passes (Ahmad & Sumari, 2009b). The concept of KGS emerges from the observation of the mechanism occurs in human brain when performing information-inferencing fusion to obtain new knowledge. In order to have a more depth understanding on this mechanism, we developed a model of Human Inference System (HIS) as the basis for our model of KGS as depicted in Fig. 4. The general formula to obtain the number of inferencing is presented in (1).

$$\lambda = \left(2^{\delta} - \delta\right) - 1 \tag{1}$$

with λ is the number of inferencing from fused information and δ is the number of information sources or sensors.
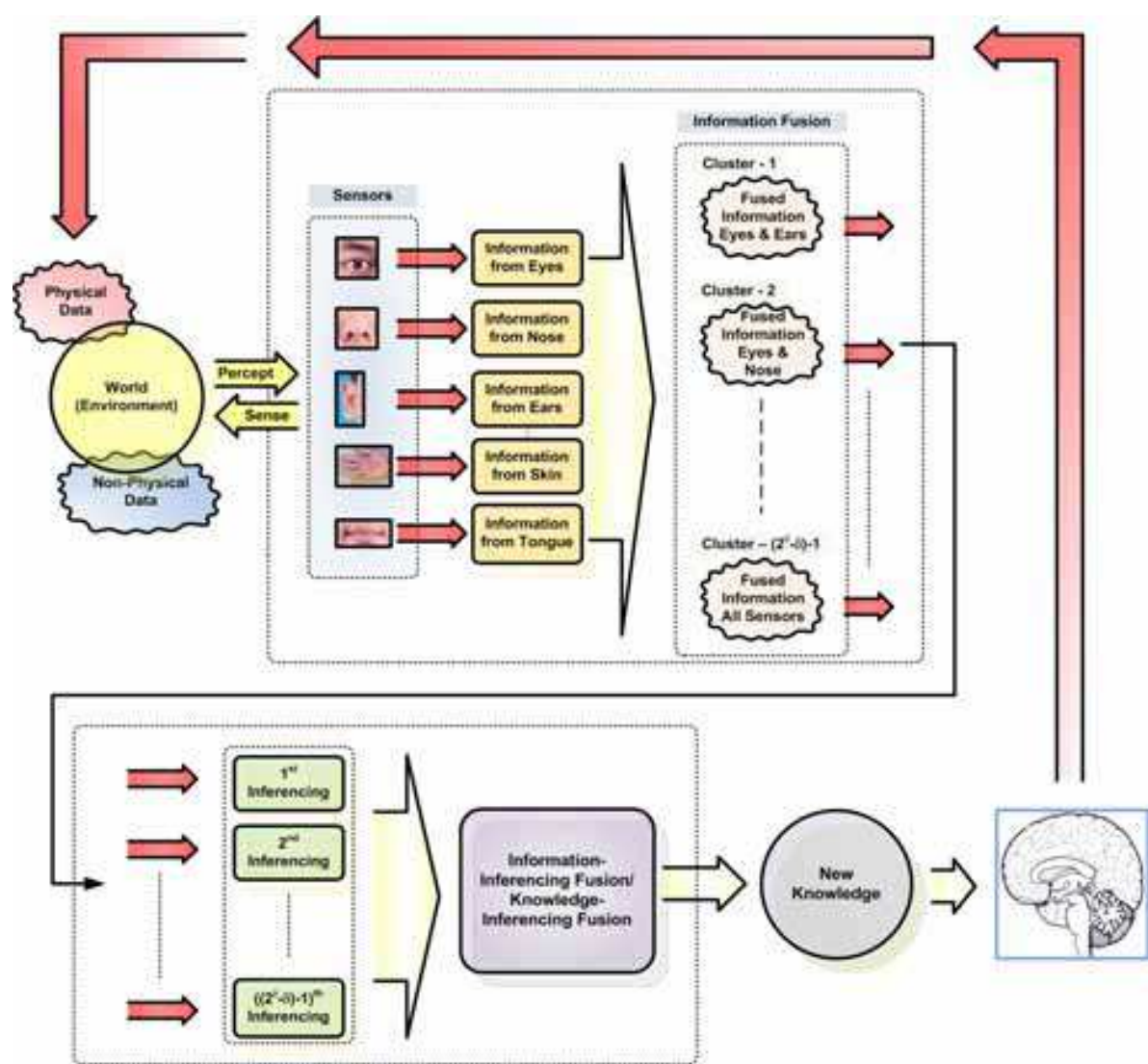

Fig. 4. A simplified illustration of human inference system (Sumari et.al, 2010b)

3.3 Knowledge growing mechanism

As we can see in Fig. 4, the process in obtaining new knowledge from the gathered-and-delivered information of a phenomenon observed by the sensory organs will have to get through a five-step process, i.e. information fusion, information inferencing, information-inferencing fusion, knowledge inferencing, and knowledge-inferencing fusion. Based on the HIS model, we developed the mechanism that will be occurred in our model of KGS as depicted in Fig. 5.

New knowledge in this scheme is called as information with Degree of Certainty (DoC) whcih is introduced by (Sumari et.al, 2009b), that is, a value that determines the certainty of the new knowledge regarding the observed phenomenon is considered accurate. Term "γ" is

the representation of the time needed to obtain the new knowledge. The DoC is obtained by applying A3S (Arwin-Adang-Aciek-Sembiring) method (Ahmad & Sumari, 2008).
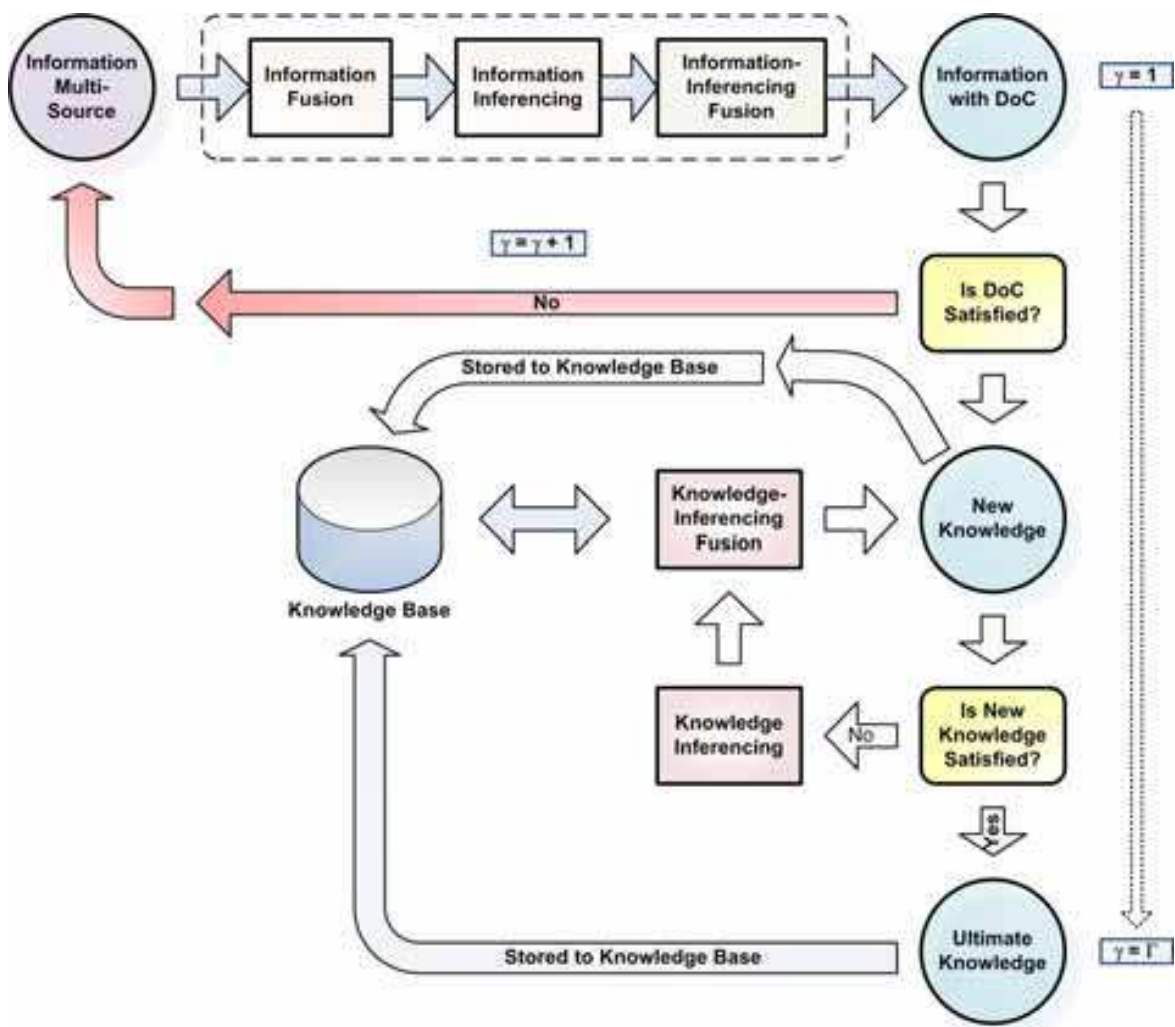

Fig. 5. The mechanism of growing the knowledge in KGS as the representation of Fig. 4 (Sumari et.al, 2009a)

3.4 OMA3S information-inferencing fusion method

To grow the knowledge, the requisites that have to be fulfilled are there has to be a knowledge base and fusion mechanism. The fusion will be applied to the received information with the existing or prior information/ knowledge. The knowledge is the result of the combination between the new information and the existing one which is called as after-processed information or posterior information. The mature technique for this situation is Bayes Inference Method (BIM) (Hall, 1992), a special case of probability theory.

For performing the information-inferencing fusion, we have developed a method called A3S information-inferencing fusion which is aimed to improve BIM for managing multi-hypothesis multi-indication problems (Ahmad & Sumari, 2008). The A3S version which involves the time parameter in the computation is called OMA3S (Sumari et.al, 2009c). The detail how this method works is illustrated in Table I followed by its paired equations given in (2) and (3).

$$P\left(\psi_1^j\right) = \frac{\sum_{i=1}^{\delta} P\left(\upsilon_i^j\right)}{\delta} \qquad (2)$$

$$P(\psi)_{estimate} = \odot\left[P\left(\psi_1^j\right)\right] \qquad (3)$$

with $P\left(\psi_1^j\right)$ is New-Knowledge Probability Distribution (NKPD), $P\left(\upsilon_i^j\right)$ is information j from sensor i, while $i = 1,...,\delta$ is the number of sensor and $j = 1,...,\lambda$ is the number of fused information. The $P\left(\psi_1^j\right)$ and $\psi_1^j \in \Psi$ is the representation of "fused information" of the information delivered from multi-sensor. The word "estimated" means the selected fused information is the most likely inferencing/ knowledge from all available inferencing/ knowledge given information from multi-sensor. $P(\psi)_{estimate}$ is the largest value of $P\left(\psi_1^j\right)$ that is called as DoC and $\odot[...] = \max\limits_{j=1,...,m}[...]$. Term $P\left(\psi_1^j\right)$ defines NKPD at $\gamma = 1$.

Information from -th Sensor	Probability Distribution of Information from i^{th} Sensor and Probability Distribution of jth Fused Information						
	1	2	3	...	j	...	λ
1	$P\left(\upsilon_i^1\right)$	$P\left(\upsilon_i^2\right)$	$P\left(\upsilon_1^3\right)$	...	$P\left(\upsilon_1^j\right)$	...	$P\left(\upsilon_1^\lambda\right)$
...	...	...	...	...	...	...	...
i	0	0	$P\left(\upsilon_i^3\right)$	...	$P\left(\upsilon_i^j\right)$	...	$P\left(\upsilon_i^\lambda\right)$
...	...	...	...	...	...	...	...
δ	0	0	0	...	0	...	$P\left(\upsilon_\delta^\lambda\right)$
ψ (DoC)	$P\left(\psi_1^1\right)$	$P\left(\psi_1^2\right)$	$P\left(\psi_1^3\right)$	...	$P\left(\psi_1^j\right)$	...	$P\left(\psi_1^\lambda\right)$

Table 1. The illustration of the computation mechanism of A3S information-inferencing method

Next observations on the same phenomenon will result in new NKPD at next time series and these NKPD will form NKPD over Time (NKPDT) as presented in (4) and (5).

$$P\left(\theta_j\right) = \frac{\sum_{\gamma=1}^{\Gamma} P\left(\phi_\gamma^j\right)}{\Gamma} \qquad (4)$$

$$P(\theta)_{estimate} = \odot\left[P\left(\theta_j\right)\right] \qquad (5)$$

with $P\left(\theta_j^j\right)$ is NKPDT, $P\left(\psi_t^j\right)$ is information j observed at observation time γ, while $\gamma = 1,...,\Gamma$ is the number of observation time and $j = 1,...,\lambda$ is the number of fused information. The $P\left(\theta_j\right)$ and $\theta_j \in \Theta$ is the representation of "knowledge" of the information delivered from multi-sensor at observation time γ. $P(\theta)_{estimate}$ is the value of DoC of $P\left(\theta_j\right)$ or the ultimate knowledge after a certain observation time.

3.5 Measuring KGS-based MCC's degree of certainty

The KGS-based MCC's certainty of the phenomenon it observes is measured by using DoC in (6)

$$DoC = \left| P(\theta)_{estimate} - \phi_1^j \right| \tag{6}$$

where $j = 1,...,\lambda$ and ϕ_1^j is the knowledge in term of probability value of the j best hypothesis at observation time γ_1.

4. The utilization of KGS-based MCC to GRS

4.1 GRS database

In this research we use five kinds of yeast25 database as presented in Table 2.

Database	Type	Number of Gene	Time Sequence (t)
Yeast25	Alpha	25	18
	Cdc15	25	24
	Cdc28	25	17
	Cho	25	17
	Elu	25	18

Table 2. Yeast25 database for validating KGS-based MCC

As previously mentioned in earlier section, the genes regulated by GRS will be represented by agents that will be working together in MCC paradigm. The experiments and the results delivered in this section are taken from (Sumari, 2010). We use data from (Zubir, 2009) namely yeast25-cho database as presented in Table 2. In this database, there are 25 genes from ACE2 to SIC1 as listed in column 1 with an interaction time interval from t-1 to t-17 or 17t times as shown in line 1. In order to give a view on the challenge in obtaining the knowledge regarding what genes are expressed in a given sample as well their behavior in a certain interaction time, the complexity of the data given in Table 3 is depicted in Fig. 6 and Fig. 7. The expressiveness of individual genes in 17t of interaction time is presented in Fig. 6, while the genes behavior in 17t of interaction time is presented in Fig. 7. The challenge in this case is to obtain knowledge from this phenomenon to answer the questions raised in Section 1.

Gene/Time	t1	t2	t3	t4	t5	t6	t7	t8	t9
ACE2	148	113	144	209	319	360	434	411	462
ASH1	380	256	201	256	320	327	245	366	795
FKH1	52	64	67	167	330	348	227	177	167
MBP1	360	326	321	425	502	484	487	410	401
MCM1	1350	1815	2095	1835	1907	1861	1734	1991	1446
NDD1	210	150	312	409	394	413	322	292	234
STB1	78	123	226	145	54	53	47	11	75
SWI4	87	115	183	104	79	77	71	88	101
SWI5	121	167	272	268	323	540	634	591	606
SWI6	421	520	699	846	869	853	740	721	558
ALG7	97	70	156	168	153	137	164	112	124

CDC20	93	235	325	208	108	168	158	124	145
CDC21	17	59	106	93	57	37	18	18	19
CDC5	86	106	89	154	219	367	549	602	480
CDC6	554	471	967	930	936	876	809	831	864
CLB2	92	158	122	145	259	348	489	554	422
CLB5	562	785	1756	949	659	612	467	563	455
CLN1	149	1045	1344	1305	998	1013	717	571	323
CLN2	445	1620	2485	1303	916	808	738	636	303
CTS1	552	1080	447	258	287	466	435	299	217
EGT2	989	1244	732	786	1083	1284	816	678	1293
FAR1	235	263	160	165	303	468	429	880	903
HTA1	905	810	1490	1594	2037	999	1178	976	843
PCL2	153	510	620	260	224	235	157	130	238
SIC1	352	295	355	308	361	356	294	212	541

Table 3a. Genes' names and their interaction values over time for yeast25-cho from t1 to t9 (continued to Table 3b)

Gene/Time	t10	t11	t12	t13	t14	t15	t16	t17
ACE2	398	226	187	325	433	343	466	344
ASH1	1862	1180	640	544	388	327	517	959
FKH1	91	74	118	172	194	163	144	106
MBP1	277	341	364	507	467	484	473	382
MCM1	2227	1945	1334	1179	1830	1506	1517	1440
NDD1	198	215	319	365	333	246	250	238
STB1	120	140	129	98	66	45	75	44
SWI4	191	172	134	111	74	43	70	83
SWI5	540	463	357	449	937	743	1024	659
SWI6	651	758	641	741	887	885	744	692
ALG7	124	151	170	202	186	172	155	70
CDC20	160	172	166	175	111	113	112	112
CDC21	28	82	90	89	36	19	12	15
CDC5	238	219	202	324	326	435	496	471
CDC6	1149	902	689	859	799	832	707	921
CLB2	236	185	151	218	267	365	350	365
CLB5	1207	1325	643	647	657	509	464	508
CLN1	1152	1870	1388	989	1074	769	614	601
CLN2	791	1288	1624	1035	839	687	868	433
CTS1	1161	1840	1181	830	979	828	507	553
EGT2	4297	3430	3240	1598	1355	1044	2123	1813
FAR1	1794	1437	725	535	570	621	739	1194
HTA1	1934	1815	1998	1103	1888	2346	1708	1580
PCL2	428	659	409	284	276	178	203	211
SIC1	1286	813	595	490	398	400	389	692

Table 3b. Genes' names and their interaction values over time for yeast25-cho from t10 to t17
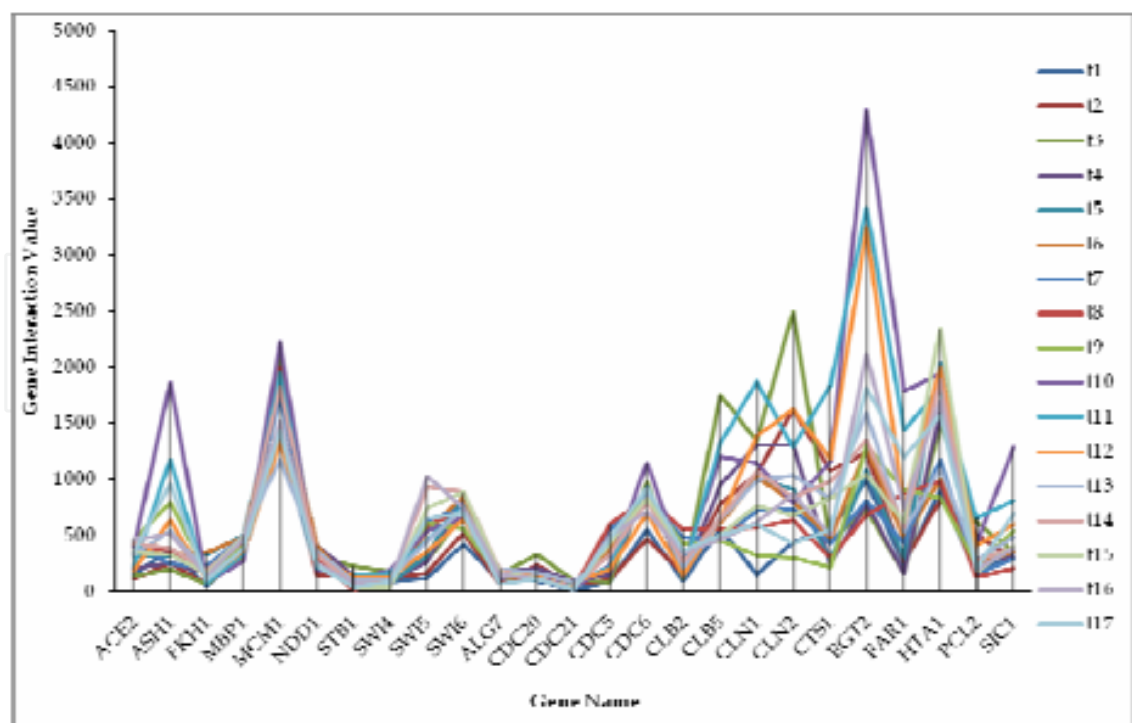
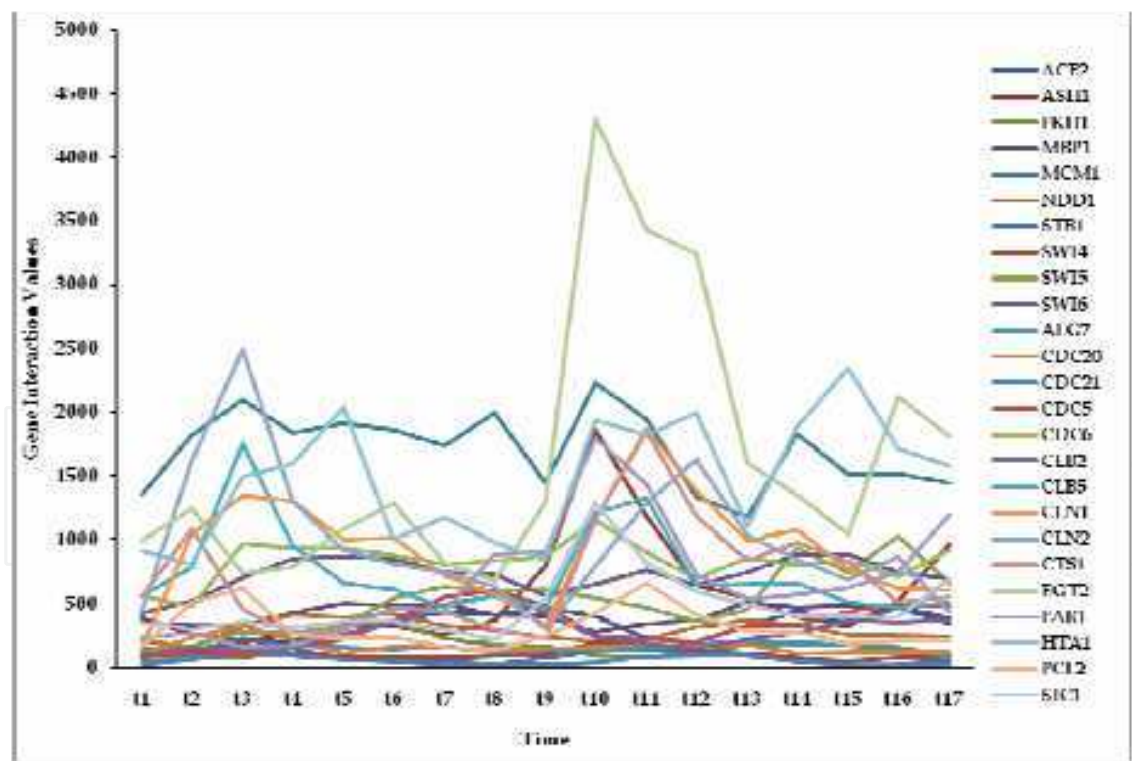Fig. 6. The expressiveness of individual genes in 17t of interaction time (Sumari, 2010)

Fig. 7. The genes behavior in 17t of interaction time (Sumari, 2010)

4.2 The KGS-based MCC's task and objective

In this situation the task of KGS-based MCC is to carry out computation to obtain values of genes behavior in a unity and individual genes expressiveness in an interval of interaction

time. Its objective is to obtain DoCs based on those obtained values to show genes behavior in a comprehensive manner and to find out the most expressive genes, that is, the most probable genes that give big influence to the product of their interaction. Therefore, for this purpose we create a strategy as follows.

- First, represent the genes into agents in multiagent system and construct the MCC framework for it.
- Second, represent the genes' interaction values into two-value state namely active and inhibit, where an active gene will be assigned binary value '1' and binary value '0' if it is inhibited or not active.
- Third, observe the genes' interaction state in time-by-time manner and put arcs on the genes which relate to each other as the representation of the existence of communication amongst agents. In this case only the active ones. Carry out this procedure until the last t in time series. During the interaction time, agents apply OMA3S method to obtain knowledge in collaborative manner. The result will be system's knowledge at certain interaction time t.
- Fourth, obtain the inferencing of the KGS-based MCC behavior as the representation of genes behavior in biological GRS. This inferencing will become the ultimate knowledge obtained by the system.
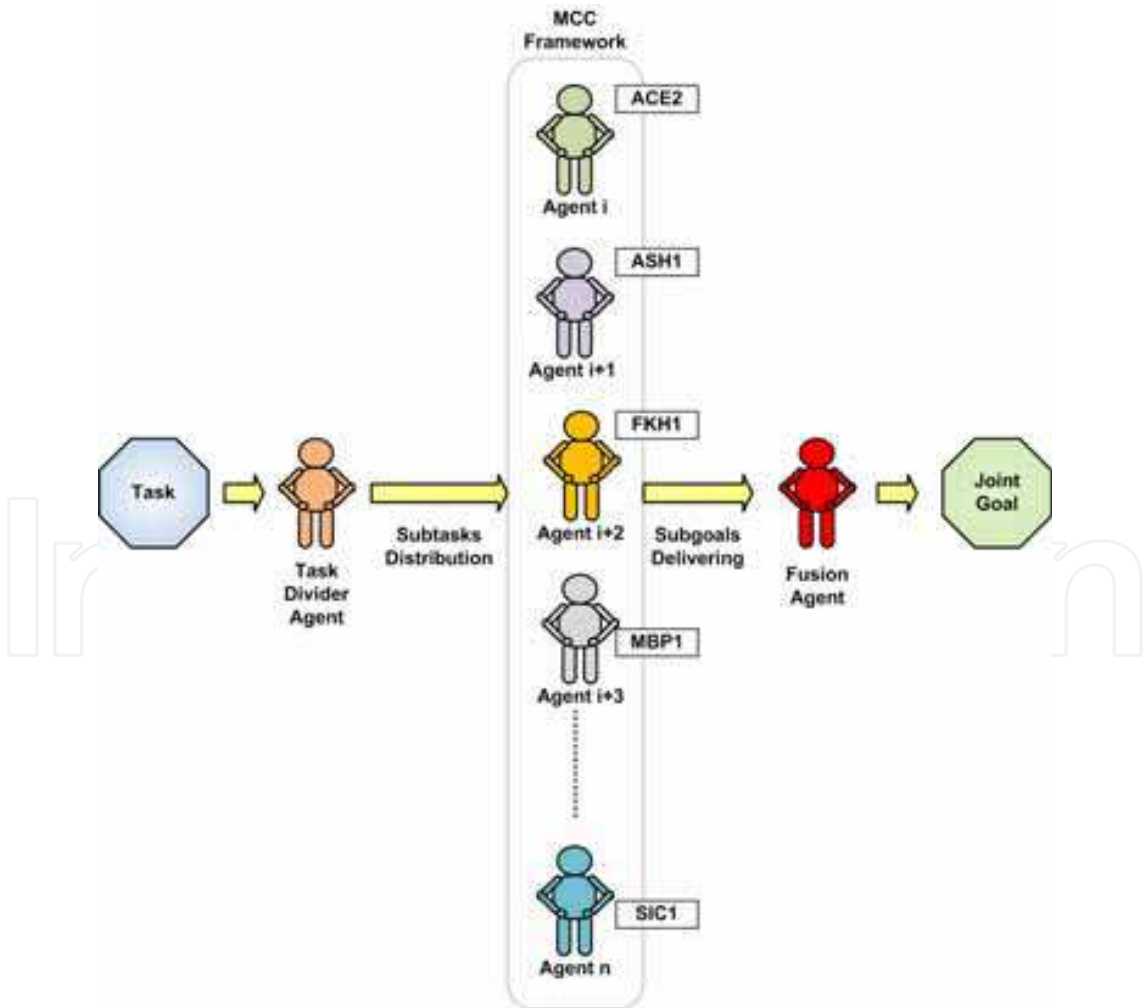
a. The Representation of Genes into MCC Structure

Fig. 8. The representation of genes of yeast25-cho by means agents in MCC structure

We assume that the task divider agent is already done its task and produces Table 2. Refer to Figure 3 in Section 2.2 we can simply convert the genes of yeas25-cho into agents in MCC structure as follows. The first and the last genes in the list namely gene ACE2 and gene SIC1 are represented by agent i and agent n where $i = 1,...,n$ and $n = 25$. The structure of the MCC is depicted in Fig. 8.

b.   The Representation of Genes' Interaction Values into Two-Value State

The simplest means to represent the existence of the interaction among genes is by representing them with value '1' for 'active' interaction and '0' for 'inhibit' interaction as alsi done by (Pasanen and Vihinen, 2005). This is what is called as Boolean representation of genes interaction in GRS. We realize that each gene has its own highest and lowest value that cannot be treated identically one to another. In order to minimize the chances of error in collaborative computation process, we use the threshold-value equation as presented in (7).

$$\phi_\tau^i = \begin{cases} 1, & \textbf{if } \psi_\tau^i < \dfrac{\sum_{i=1}^{\lambda} \psi_\tau^i}{\lambda} \\ 0, & \textbf{if } \psi_\tau^i \geq \dfrac{\sum_{i=1}^{\lambda} \psi_\tau^i}{\lambda} \end{cases} \qquad (7)$$

with $\phi_\tau^i \in \Pi$ is a binary-sequence which represents a set of interaction value at observation time t, ψ_τ^i is the interaction value of agent j at observation time t, and τ is the number of interaction sequence t when the observation is carried out. The parameter t in this case is dimensionless. The application of (7) to the data given in Table 3 is presented in Table 4.

Gene/Time	t1	t2	t3	t4	t5	t6	t7	t8	t9
ACE2	1	1	1	1	0	0	0	0	0
ASH1	1	1	1	1	1	1	1	1	0
FKH1	1	1	1	0	0	0	0	0	0
MBP1	1	1	1	0	0	0	0	1	1
MCM1	1	0	0	0	0	0	0	0	1
NDD1	1	1	0	0	0	0	0	0	1
STB1	1	0	0	0	1	1	1	1	1
SWI4	1	0	0	1	1	1	1	1	1
SWI5	1	1	1	1	1	0	0	0	0
SWI6	1	1	1	0	0	0	0	0	1
ALG7	1	1	0	0	0	1	0	1	1
CDC20	1	0	0	0	1	0	0	1	1
CDC21	1	0	0	0	0	1	1	1	1
CDC5	1	1	1	1	1	0	0	0	0
CDC6	1	1	0	0	0	0	1	0	0
CLB2	1	1	1	1	1	0	0	0	0

CLB5	1	0	0	0	1	1	1	1	1
CLN1	1	0	0	0	0	0	1	1	1
CLN2	1	0	0	0	1	1	1	1	1
CTS1	1	0	1	1	1	1	1	1	1
EGT2	1	1	1	1	1	1	1	1	1
FAR1	1	1	1	1	1	1	1	0	0
HTA1	1	1	0	0	0	1	1	1	1
PCL2	1	0	0	1	1	1	1	1	1
SIC1	1	1	1	1	1	1	1	1	0

Table 4a. The conversion results of genes interaction values from Table 2a (continued to Table 4b)

Gene/Time	t10	t11	t12	t13	t14	t15	t16	t17
ACE2	0	1	1	0	0	0	0	0
ASH1	0	0	0	1	1	1	1	0
FKH1	1	1	1	0	0	0	1	1
MBP1	1	1	1	0	0	0	0	1
MCM1	0	0	1	1	0	1	1	1
NDD1	1	1	0	0	0	1	1	1
STB1	0	0	0	0	1	1	1	1
SWI4	0	0	0	0	1	1	1	1
SWI5	0	1	1	1	0	0	0	0
SWI6	1	0	1	0	0	0	0	1
ALG7	1	0	0	0	0	0	0	1
CDC20	0	0	0	0	1	1	1	1
CDC21	1	0	0	0	1	1	1	1
CDC5	1	1	1	0	0	0	0	0
CDC6	0	0	1	0	1	0	1	0
CLB2	1	1	1	1	1	0	0	0
CLB5	0	0	1	1	1	1	1	1
CLN1	0	0	0	0	0	1	1	1
CLN2	1	0	0	0	1	1	1	1
CTS1	0	0	0	0	0	0	1	1
EGT2	0	0	0	1	1	1	0	0
FAR1	0	0	0	1	1	1	0	0
HTA1	0	0	0	1	0	0	0	0
PCL2	0	0	0	1	1	1	1	1
SIC1	0	0	0	0	1	1	1	0

Table 4b. The conversion results of genes interaction values from Table 2b

c.   The Observation to Agents Interaction and Obtaining Knowledge from It

Some influence graphs will be presented to show the interaction amongst KGS-based MMC agents in GRS, and knowledge that is acquired by all agents collaboratively at t_1, t_6 and after t_6, at t_{17} and after t_{17}. The expressiveness of single agents and their behaviour at a certain time and at a certain time interval such as from t_1 to t_6 are depicted in Fig. 9 to Fig. 15, while for all 17t of interaction time is presented in Fig. 16.
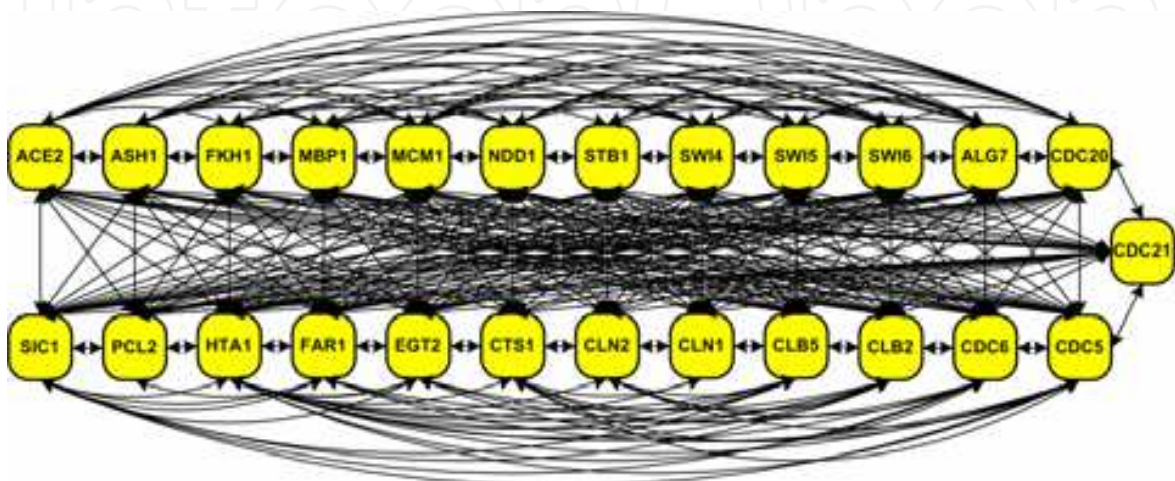
Fig. 9. Influence graph of KGS-based MCC behaviour at t_1. Yellow colour represents the active agents during the interaction at that time
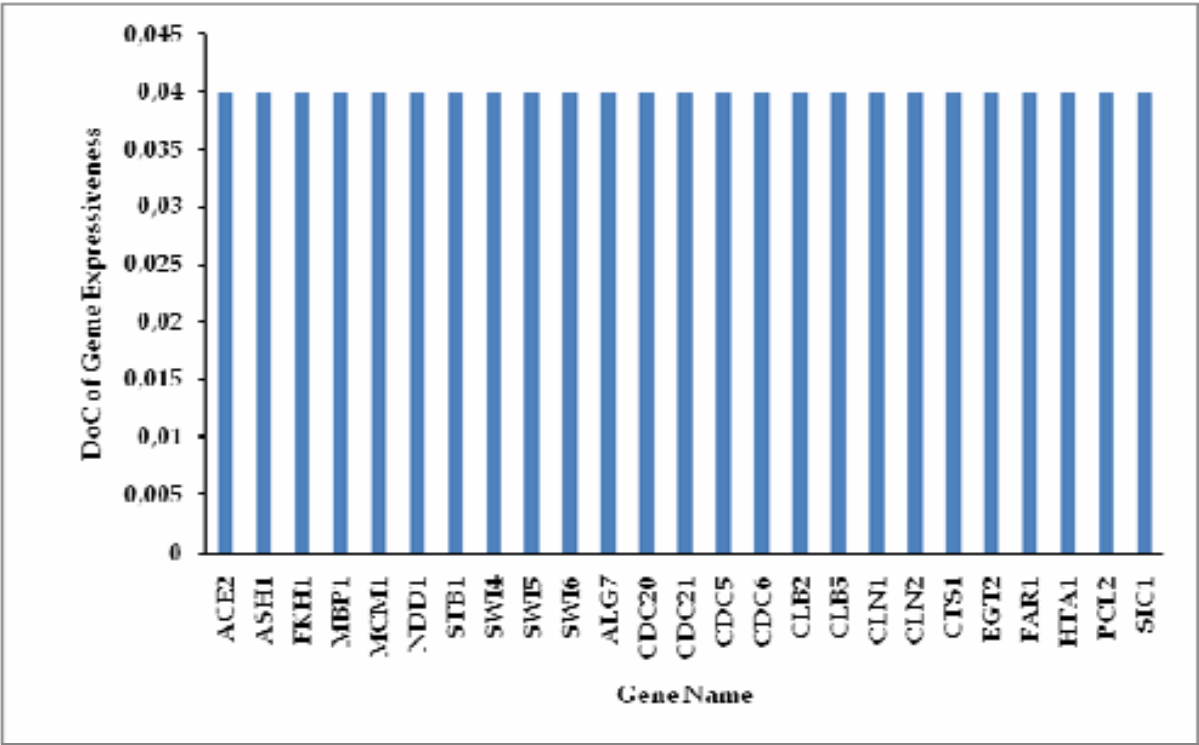
Fig. 10. Knowledge acquired by agents at t_1.
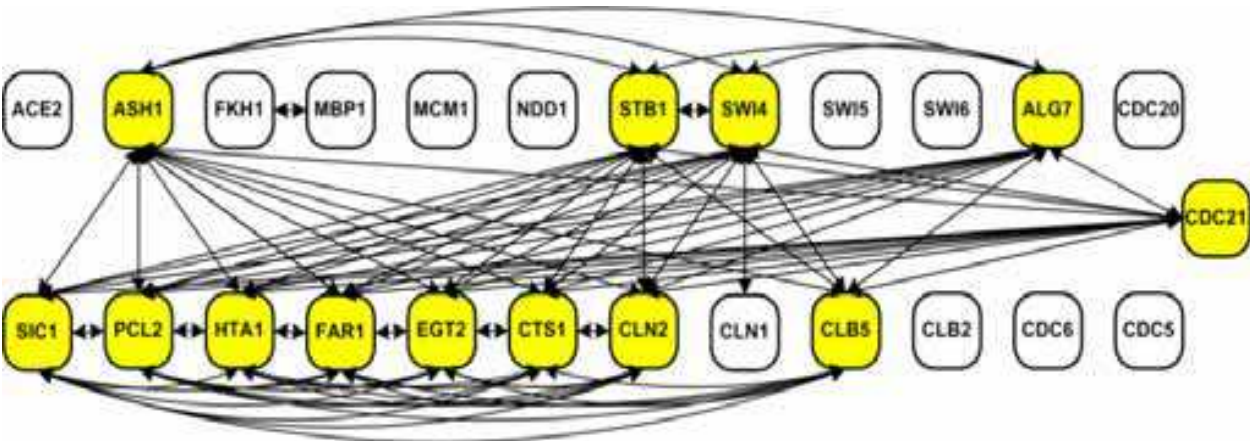
Fig. 11. Influence graph of KGS-based MCC behaviour at t_6
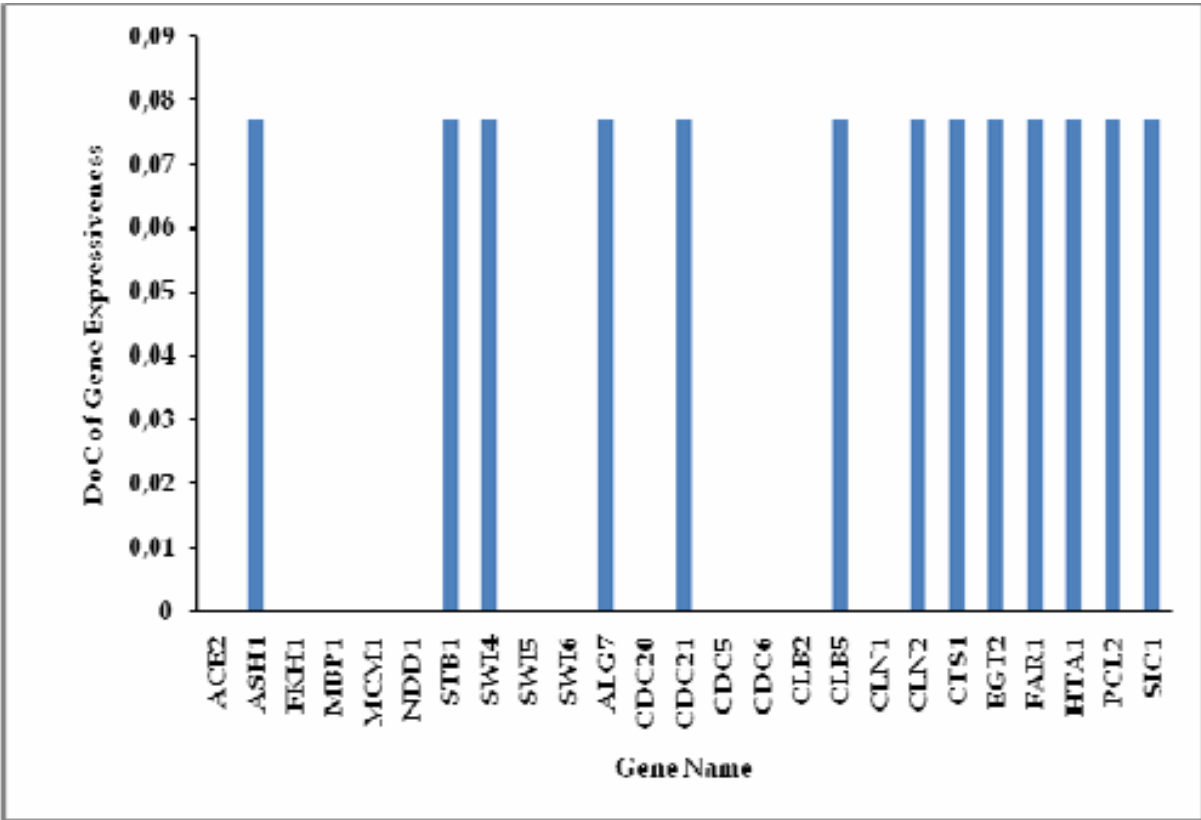
Fig. 12. Knowledge acquired by agents at t_6
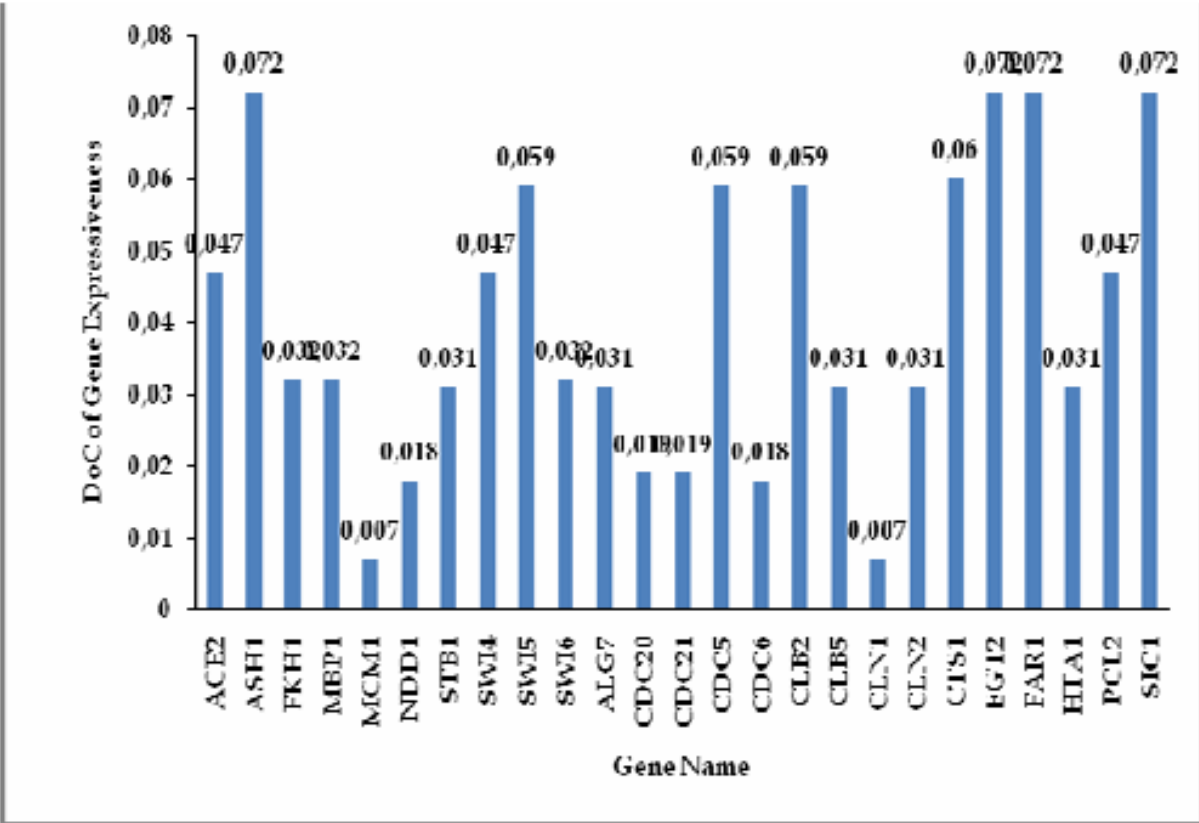
Fig. 13. Knowledge acquired by agents after 6t of interaction time
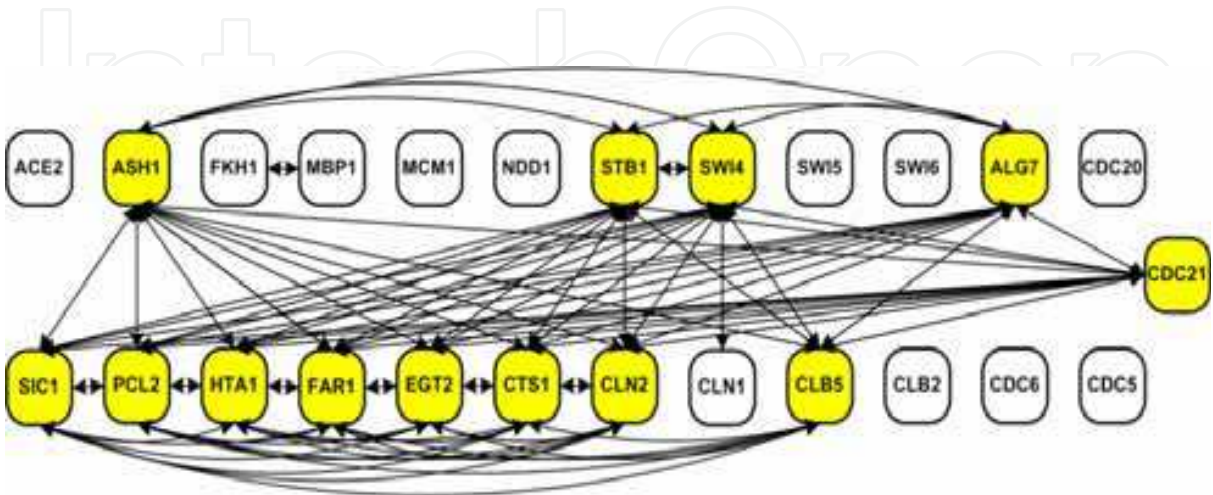
Fig. 14. Influence graph of KGS-based MCC behaviour at t_{17}

Fig. 15. Knowledge acquired by agents at t_{17}
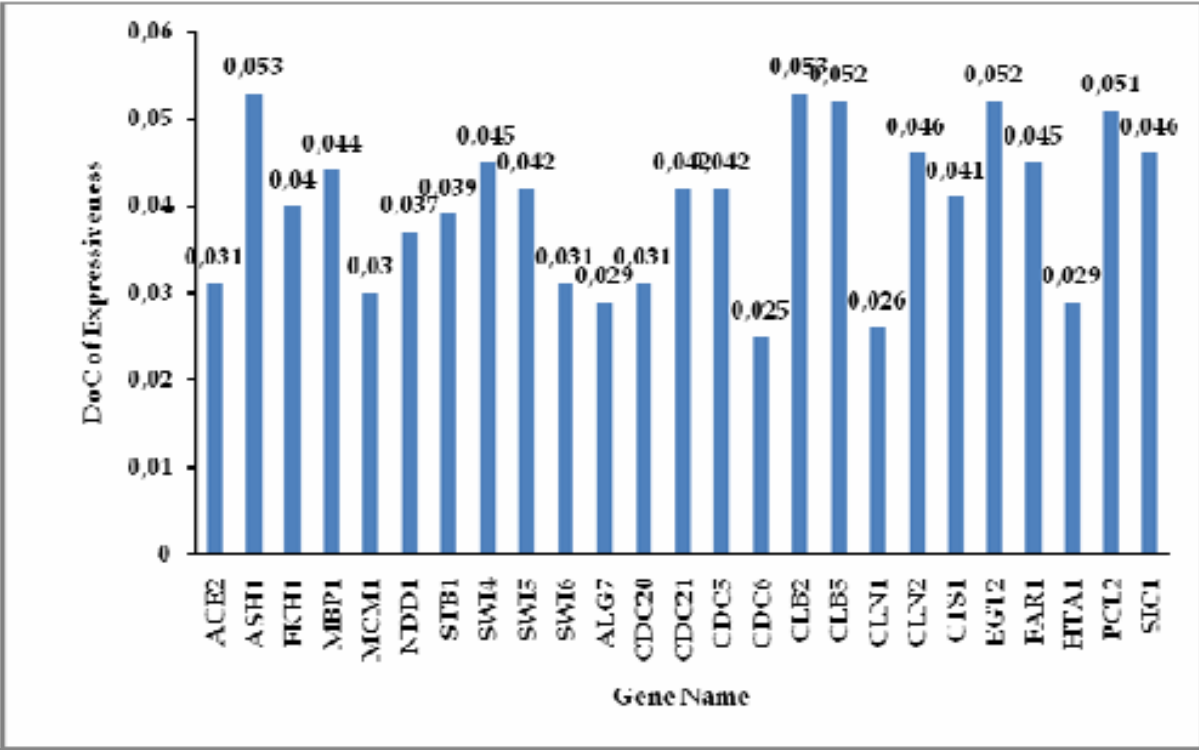
Fig. 16. Knowledge acquired by agents after 17t of interaction time

d.   Knowledge Obtained by KGS-based MCC

The computation results in form of the values of DoCs that are produced by KGS-based MCC's fusion agent which are compared with the number of communication paths formed

during agents interaction is presented in Fig. 17 and Fig. 18. These DoCs become the ultimate knowledge of the system.
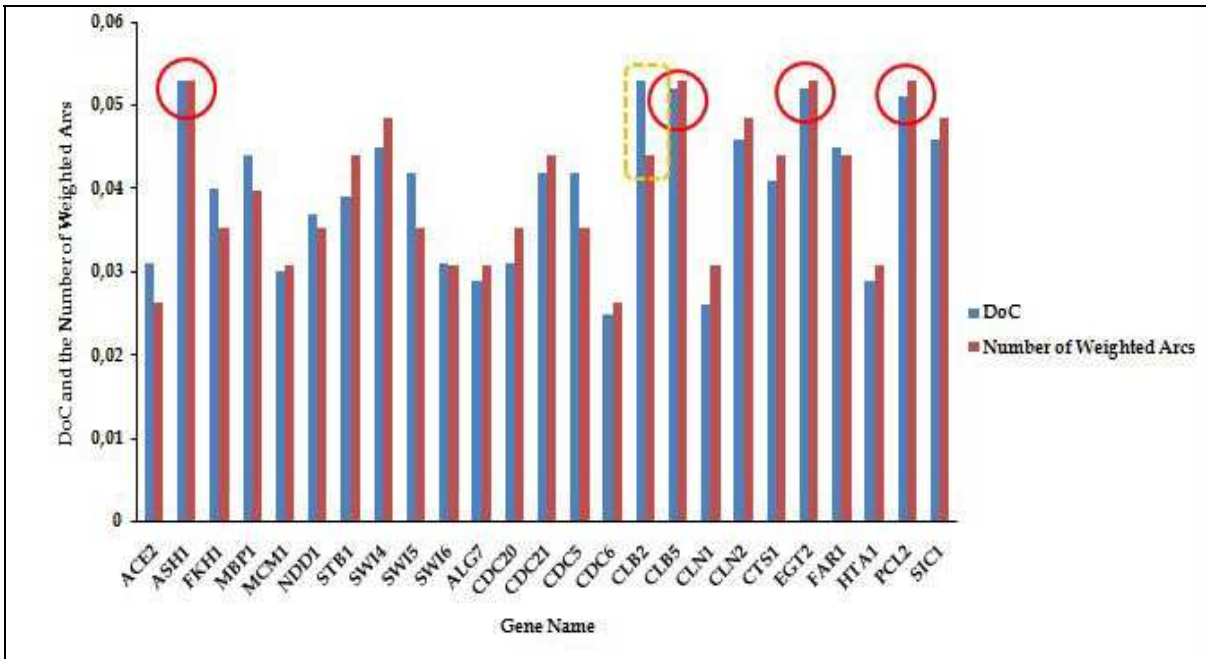
Fig. 17a. The ultimate knowledge acquired by the system regarding individual agents expressiveness after 17t of interaction time. The DoCs of the system's knowledge are compared with the communication paths formed amongst agents
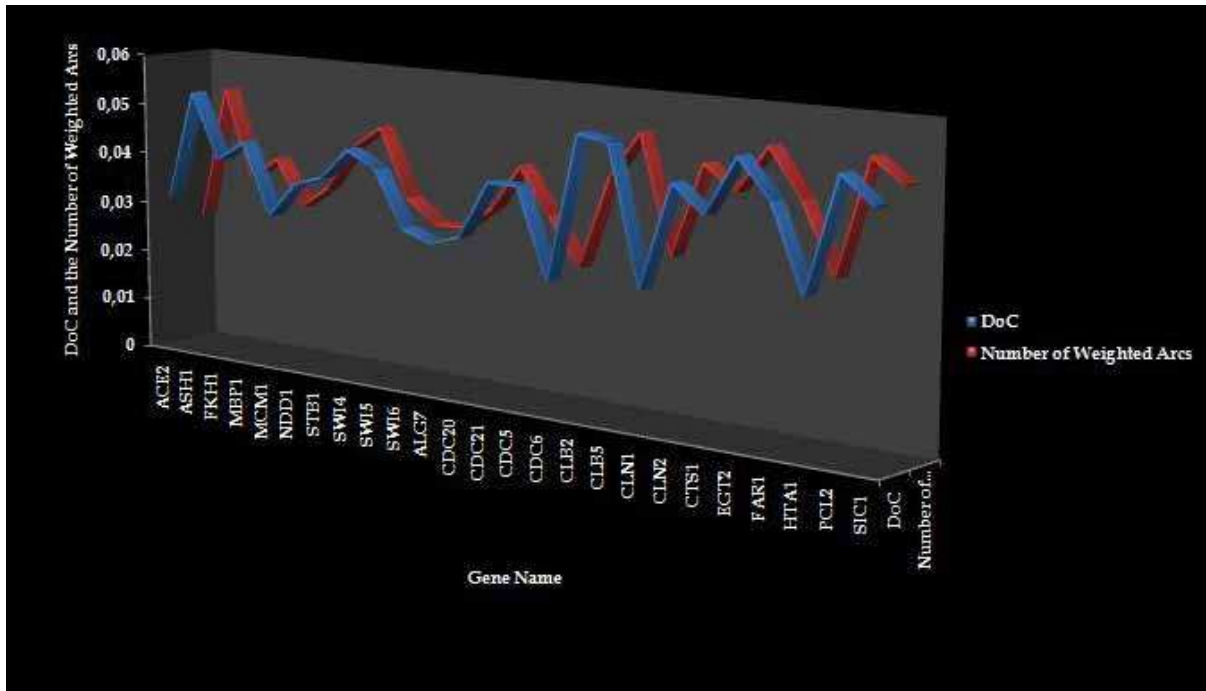
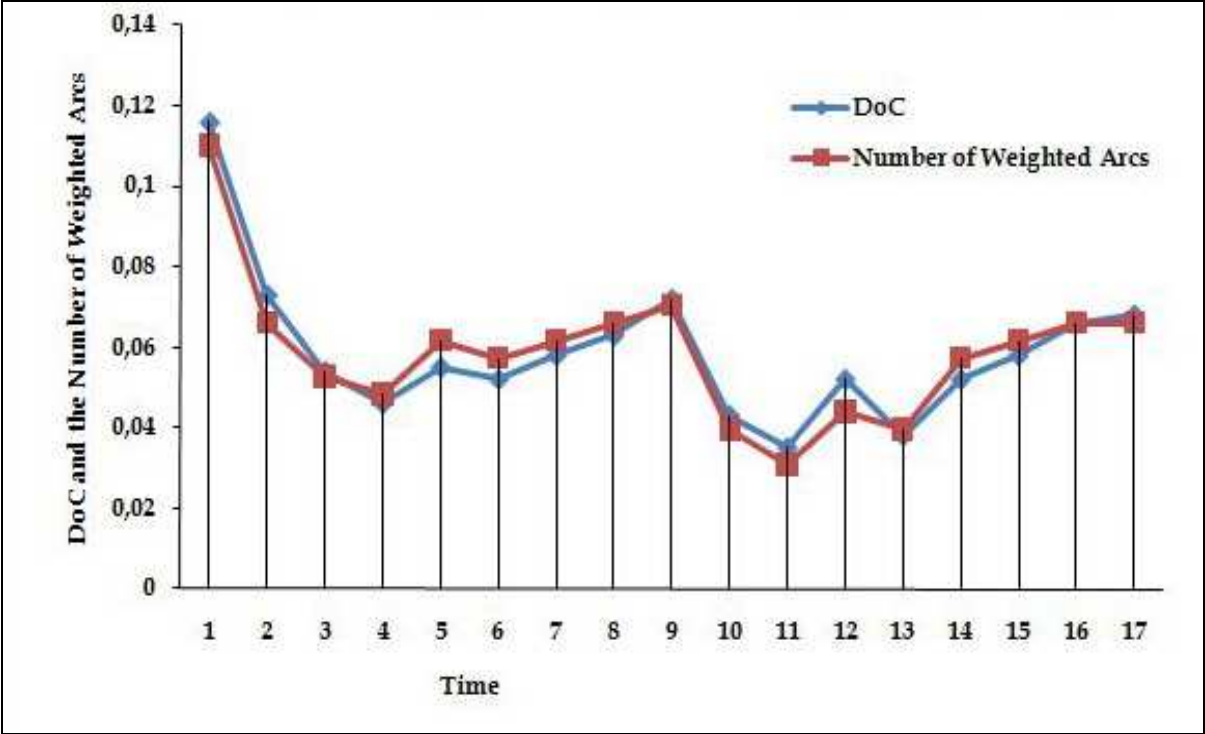Fig. 17b. 3D line graphic representation of system's knowledge regarding individual agents expressiveness after 17t of interaction time

Fig. 18a. The ultimate knowledge acquired by the system regarding all agents behavior
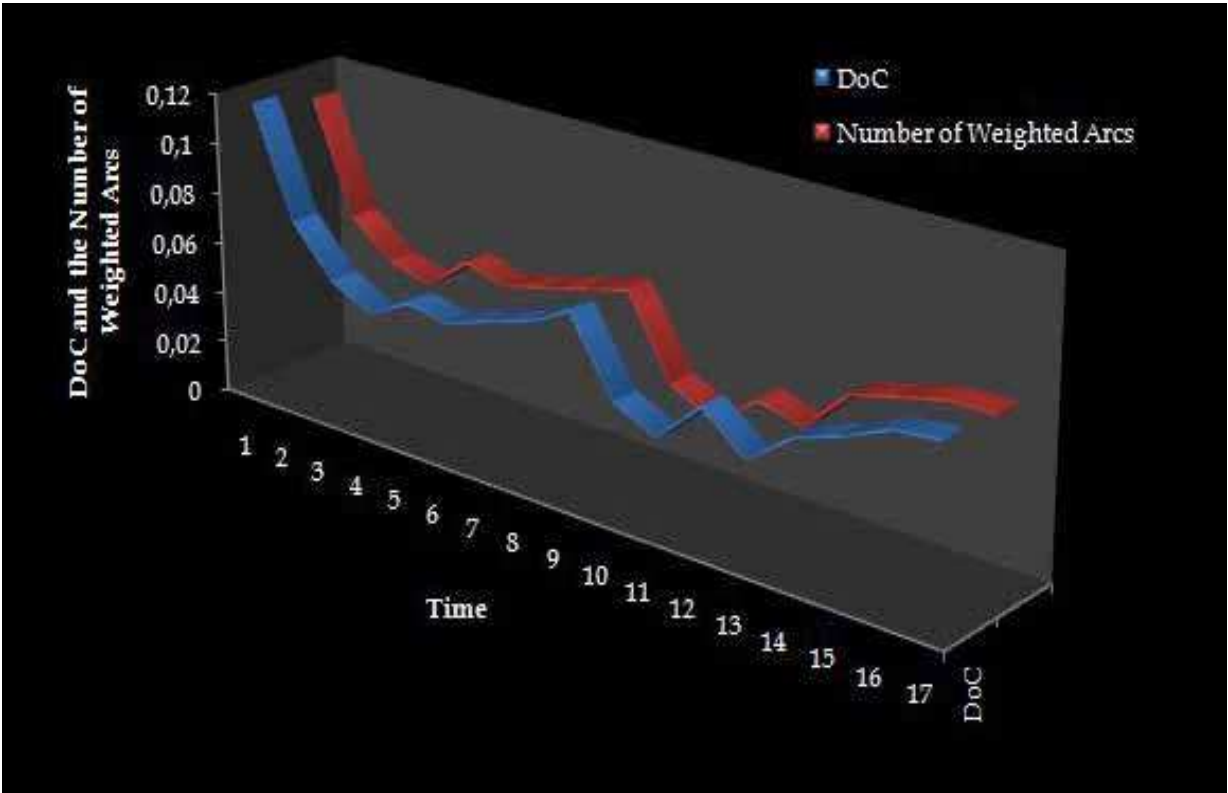during 17t of interaction time

Fig. 18b. 3D line graphic representation of the system's knowledge regarding all agents
behavior during 17t of interaction time

After observing Fig. 17 and Fig. 18, we can have knowledge regarding KGS-based MCC agents' behaviour as follows.

1. Viewed from individual agents expressiveness we can obtain knowledge that gene **ASH1** exhibit a dominant role even though there are genes CLB5, EGT2, and PCL2 that also show their expressiveness with **DoC = 0.053**. The certainty that gene ASH1 is the most expressive gene amongst other expressive genes is shown by the difference between DoC and the number of weighted arcs that is very minimum, that is, **0.01%**. Fig. 17 is the comprehensive inferencing of the information given in Fig. 6.

2. The time where the highest interaction is achieved is at t_1 or $t_{high} = t_1$, when 100% of yeast25-cho agents involve in the interaction with **DoC = 0.116**. On the other hand, the time where the lowest interaction is achieved is at t_{11} or $t_{low} = t_{11}$ with **DoC = 0.035**. Fig. 18 is the comprehensive inferencing of the information given in Fig. 7.

4.2 Estimations based on the knowledge obtained by KGS-based MCC

Based on the knowledge obtained by KGS-based MCC regarding the phenomena displayed by all agents, we can make estimations as follow.

a. The most probable gene that will be playing the important role in the interaction time between t_{18} to t_{24}, with the assumption that the interaction cycle will restart every 17t times, is **gene ASH1**. This phenomenon is in line with the information presented in (http://www.yeastgenome.org/) regarding the role of that gene as an inhibitor in transcription process.

b. Estimation of $t_{high} = t_{18}$ and estimation of $t_{low} = t_{28}$

5. Measuring the accurateness of KGS-based MCC results

Validating the results is of importance matter to ensure that the new paradigm we deliver in this paper is valid. For this purpose we have developed some parameters to compare DoC values with the number of arcs which is the representation of the existence of communication paths amongst agents in the influence graphs depicted in previous section. These parameters are:

a. The number of arcs, ξ and weighted arc, Ξ. Arcs are the arcs that connect the active agents in each interaction time and are observed by the help of influence graphs. Weighted arc is normalized value of arc in each interaction time, t or for each gene, g by the total number of arcs formed in each interaction time, t.

$$\Xi = \frac{\xi[i]}{\sum_{i=1}^{n} \xi[i]} \tag{7}$$

with $i = 1,...,n$ and n is the number of genes or one cell cycle, as example 17t.

b. ΔDKJ, $\%\Delta DK$, and $\%\Delta DKJ\_total$. The absolute value of the difference between DoC and the number of arcs formed at each interaction time, t or for each gene, g. $\%\Delta DK$ is the percentage of ΔDKJ while $\%\Delta DKJ\_total$ is the total number of $\%\Delta DK$.

$$\Delta DKJ[i] = |DoC[i] - \xi[i]| \tag{8}$$

$$\%\Delta DKJ = \Delta DKJ[i] \times 100\% \tag{9}$$

$$\%\Delta DKJ\_total = \sum_{i=1}^{n} \%\Delta DKJ \tag{10}$$

with $i = 1,...,n$ and n is the number of genes or one cell cycle, as example 17t.

c. Behavior match, χ, the total number of match, X, and percentage of the behavior match, $\%X$. Parameter that shows the difference of genes behaviour is viewed from DoC and Ξ with an assumption that the arrangement of genes as given in Table 2 is not changed. Behaviour matching comparison is done by observing the graphic dynamic from gene i^{th} to gene n^{th}.

$$\chi[i] = \begin{cases} 1, & if\, DoC[i] = \Xi[i] \\ \\ 0, & if\, DoC[i] \neq \Xi[i] \end{cases} \tag{11}$$

$$X = \sum_{i=1}^{n} \chi[i]_1 \tag{12}$$

$$\%X = \frac{\sum_{i=1}^{n} \chi[i]_1}{n} \tag{13}$$

with $i = 1,...,n$ and n is the number of genes or one cell cycle, as example 17t. $\chi[i]_1$ is match gene behaviour otherwise it will be $\chi[i]_0$, and $\%X$ is the percentage of behaviour match.

d. Erroneous of behavior match, ε. The percentage of the comparison between the total number of match of gene behavior, X, with the total number of absolute erroneous, $\%\Delta DKJ\_total$.

$$\varepsilon = \frac{\%\Delta DKJ\_total}{X} \tag{13}$$

In this case even though $\%X = 100\%$, ε does not always have value of 0% because of the influence of $\Delta DKJ[i]$ factor.

On the next tables we will see how these parameters are used to measure the accurateness of the application of KGS-based MCC to obtain the knowledge regarding the GRS behaviour. Table 5 contains the system's knowledge taken from Fig. 17 while Table 6 contains the system's knowledge taken from Fig. 18.

g^{th}	$DoC[i]$	$\xi[i]$	$\Xi[i]$	$\Delta DKJ[i]$	$\%\Delta DKJ[i]$	$\chi[i]$
ACE2	0.031	6	0.026	0.005	0.46	1
ASH1	0.053	12	0.053	0.000	0.01	1
FKH1	0.04	8	0.035	0.005	0.48	1
MBP1	0.044	9	0.040	0.004	0.44	1
MCM1	0.03	7	0.031	0.001	0.08	1
NDD1	0.037	8	0.035	0.002	0.18	1
STB1	0.039	10	0,044	0,005	0,51	1
SWI4	0.045	11	0.048	0.003	0.35	1
SWI5	0.042	8	0.035	0.007	0.68	1
SWI6	0.031	7	0.031	0.000	0.02	1
ALG7	0.029	7	0.031	0.002	0.18	1
CDC20	0.031	8	0.035	0.004	0.42	1
CDC21	0.042	10	0.044	0.002	0.21	1
CDC5	0.042	8	0.035	0.007	0.68	1
CDC6	0.025	6	0.026	0.001	0.14	1
CLB2	0.053	10	0.044	0.009	0.89	0
CLB5	0.052	12	0.053	0.001	0.09	1
CLN1	0.026	7	0.031	0.005	0.48	1
CLN2	0.046	11	0.048	0.002	0.25	1
CTS1	0.041	10	0.044	0.003	0.31	1
EGT2	0.052	12	0.053	0.001	0.09	1
FAR1	0.045	10	0.044	0.001	0.09	1
HTA1	0.029	7	0.031	0.002	0.18	1
PCL2	0.051	12	0.053	0.002	0.19	1
SIC1	0.046	11	0.048	0.002	0.25	1
25 gen	**1.002**	**1**	**227**	**0.076**	**7.63**	**24**
X						**24**
%X						**96**
ε						**0.32%**

Legend

        The most expressive genes based on the results of $\xi[i]$

        The least expressive genes based on the results of $\xi[i]$

        Genes with $DoC[i] \neq \Xi[i]$ viewed from $\xi[i]$ perspective

Table 5. The comparison of individual genes expressiveness for yeast25-cho genes based on the similarity between DoCs and the number of arcs after 17t of interaction time

t^{th}	$DoC[i]$	$\xi[i]$	$\Xi[i]$	$\Delta DKJ[i]$	$\%\Delta DKJ[i]$	$\chi[i]$
1	0.116	25	0.110	0.006	0.59	1
2	0.073	15	0.066	0.007	0.69	1
3	0.054	12	0.053	0.001	0.11	1
4	0.046	11	0.048	0.002	0.25	1
5	0.055	14	0.062	0.007	0.67	1
6	0.052	13	0.057	0.005	0.53	1
7	0.058	14	0.062	0.004	0.37	1
8	0.063	15	0.066	0.003	0.31	1
9	0.072	16	0.070	0.002	0.15	1
10	0.043	9	0.040	0.003	0.34	1
11	0.035	7	0.031	0.004	0.42	1
12	0.052	10	0.044	0.008	0.79	1
13	0.038	9	0.040	0.002	0.16	1
14	0.052	13	0.057	0.005	0.53	1
15	0.058	14	0.062	0.004	0.37	1
16	0.066	15	0.066	0.000	0.01	1
17	0.068	15	0.066	0.002	0.19	1
17t	**1.001**	**227**	**1**	**0.06**	**6.46**	**17**
X						**17**
%X						**100**
ε						**0.38**

Legend

(yellow)	t_{high}
(green)	t_{low}

Table 6. The behaviour comparison for all yeast25-cho genes based on the similarity
between DoCs and the number of arcs after 17t of interaction time

From the comparison results presented in Table 4 and Table 5, we can conclude as follows.

3. Viewed from the individual genes expressiveness in interaction in 17t times, the accurateness of system's result is **24/25 * 100% = 96%**, with the erroneous of behaviour match as much as **0.32%**.

4. Viewed from the behaviour of all genes in 17t times of interaction time, the accurateness of system's result is **17/17 * 100% = 100%**, with the erroneous of behaviour match as much as **0.38%**.

6. The results of the application of KGS-based MCC to other types of yeast25

With the same manner, we apply KGS-based MCC to the other types of yeast25 and the results are presented in Table 7 for individual gene expressiveness and Table 8 for all genes behaviour.

Type	Accurateness	The Most Expressive Genes	Time Interval
Alpha	88%	MBP1, CDC5, FAR1	18t
Cdc15	72%	ASH1, STB1, CLN2, PCL2	24t
Cdc28	92%	CLB5, EGT2	18t
Cho	96%	ASH1	17t
Elu	92%	SWI6, SIC1	18t
Average	**88%**		

Table 7. The accurateness of the application KGS-based MCC to yeast25 database for obtaining knowledge regarding the most expressive genes

Type	Accurateness	Interaction Time	
		t_{high}	t_{low}
Alpha	100%	t_5 , t_6	t_1
Cdc15	100%	t_9 , t_{22}	t_6
Cdc28	100%	t_5	t_{15}
Cho	100%	t_1	t_{11}
Elu	94.4%	t_6	t_{13}
Average	**98.8%**		

Table 8. The accurateness of the application KGS-based MCC to yeast25 database for obtaining knowledge regarding al genes behaviour

Based on the information delivered in the two tables above, we can conclude the accurateness of KGS-based MCC in obtaining knowledge regarding GRS behaviour as follows.

1. The average accurateness of the knowledge regarding the most expressive genes reaches **88%**.
2. The averaged accurateness of the knowledge regarding the behaviour of all genes reaches **98.8%**.

7. Conclusion

GRS is a unique system within all living things which regulates the "core" of living cells called gene, that is, a sequence of DNA. The regulation performed by GRS to the genes displays a interesting phenomena that if we can obtain knowledge about them we can use it to refine the products of the cells in the future. Harmonious with it, in this paper we have presented in concise manner the process in obtaining knowledge regarding GRS behaviour by applying KGS-based MCC paradigm on several types of yeast25 database. The core of KGS-based MCC paradigm is OMA3S information-inferencing fusion method that is developed primarily for KGS, a new perspective in Artificial Intelligence (Sumari, 2010).

From the strategy as well as the research results we have presented in this paper we can conclude that the representation of GRS by means of KGS-based MCC paradigm is a very good approach and very advantegous for obtaining knowledge regarding the behavior of genes in it. The knowldege includes the most expressive genes and all genes behaviour in a

single cell cycle, that is, a certain interval of interaction time. One example, that is yeast25-cho, has already been given and the accurateness of the knowledge acquired by the system is 96% and 100% for having a knowledge of the most expressive genes and all genes behaviour subsequently. By applying the same technique to other types of yeast25, we found that in average the knowledge accurateness of the knowledge acquired by the system is **88**% and **98.8**% for having a knowledge of the most expressive genes and all genes behaviour subsequently. These results show that KGS-based MCC paradigm is one of appropriate means for obtaining knowledge regarding GRS behaviour.

One of our further works will be applying KGS-based MCC paradigm on a larger number of genes from the same database. We have been conducting a research on this matter and in the same time perfecting our method so it can be applied to diverse problems.

8. References

Ahmad, A.S. & Sumari, A.D.W. (2008). *Multi-Agent Information Inferencing Fusion in Integrated Information System*, Institut Teknologi Bandung Publisher, ISBN 978-979-1344-31-9, Bandung.

Ahmad, A.S. & Sumari, A.D.W. (2009a). The modeling of genes' interactions by means of multiagent collaborative computation, *Proceedings of International Conference on Instrumentation, Communication, and Information Technology and Biomedical Engineering 2009* (ICICI-BME2009), *Plenary Speech Paper*, pp. 6-11, ISBN 978-979-1344-67-8, IEEE CFP0987H-CDR, Institut Teknologi Bandung, November 2009, Institut Teknologi Bandung, Bandung.

A Ahmad, A.S. & Sumari, A.D.W. (2009b), A novel perspective on Artificial Intelligence: information-inferencing fusion for knowledge growing, *The 2nd International Conference on Electrical Engineering and Informatics 2009* (ICEEI2009), *Keynote Speech Paper*, Universiti Kebangsaan Malaysia, August 2009, Universiti Kebangsaan Malaysia, Kuala Lumpur.

Ahmad, A.S.; Sumari, A.D.W. & Zubir, H.Y. (2009) Collaborative computation in bioinformatics for better life, *Proceedings of National Seminar on Information Technologi and Its Applications 2009*, *Keynote Speech Paper*, ISBN 977-208-5234-00-7, Malang State Politechnic, March 2009, Malang State Politechnic, Malang.

Camarinha-Matos, L. M. & Afsarmanesh, H. (eds) (2008). *Collaborative Network: Reference Modeling*, Springer, ISBN 978-0387794259, New York.

Dill, K.; Grodzinski, P. & Liu, R.H. (eds) (2009). *Microarrays: Preparation, Microfluidics, Detection Methods, and Biological Applications*, Springer, ISBN 978-0387727165, New York.

Emmert-Streib, F. & Dehmer, M. (eds) (2008). *Analysis of Microarray Data: A Network-Based Approach*, Wiley-VCH, ISBN 978-3527318223, Weinheim.

Ewens, W. & Grant, G. (2005). *Statisical Methods in Bioinformatics: An Introduction, 2nd Ed*, Springer, ISBN 0-387-40082-6, New York.

Hall, D.L. (1992). *Mathematical Techniques in Multisensor Data Fusion*, Artech House, ISBN 0-890065586, London.

Hall, D.L. & Llinas, J.L. (2001). *Handbook of Multisensor Data Fusion*, CRC Press, ISBN 978-0849323799, New York.

Pasanen, T. dan Vihinen, M. (2005). Gene regulatory networks, In: *DNA Microarray Data Analysis*, Tuimala, J & Laine, M.M. (Ed.), pp. 135-142, CSC-Scientific Computing Ltd., ISBN 952-5520129, Helsinki.

Reece, R.J (2004). *Analysis of Genes and Genomes*, John Wiley & Sons, ISBN 978-0470843802, Chichester.

Sumari, A.D.W. & Ahmad, A.S. (2009). Multiagent collaborative computation paradigm for decision-making support system, *Proceedings of International Industrial Informatics Seminar 2009* (IIIS2009), pp. 17-22, ISSN 2085-4854, Sunan Kalijaga Islamic State University, August 2009, Sunan Kalijaga Islamic State University, Yogyakarta.

Sumari, A.D.W.; Ahmad, A.S.; Wuryandari, I. & Sembiring, J (2009a). A mathematical model of knowledge-growing system: a novel perspective In Artificial Intelligence, *Proceedings* of IndoMS *International Conference on Mathematics and Its Applications 2009* (IICMA2009), pp. 229-240, ISBN 978-6029642605, Gadjah Mada University, October 2009, Gadjah Mada University, Yogyakarta.

Sumari, A.D.W.; Ahmad, A.S.; Wuryandari, I. & Sembiring, J (2009b), Multiagent-based information fusion system in network-centric warfare paradigm, *Journal of Computer Technology "COMPILE"*, Vol.2, No. 1 (January 2009) pp. 39-55, ISSN 1978-4678.

Sumari, A.D.W.; Ahmad, A.S.; Wuryandari, I. & Sembiring, J (2009c), The application of knowledge-growing system to multiagent collaborative computation for inferring the behavior of genes interaction, *International Journal of Computer Science & Network Security* (IJCSNS), Vol. 9, No. 11 (November 2009) pp. 82-92, ISSN 1738-7906.

Sumari, A.D.W. (2010). Knowledge-Growing System: a novel perspective in Artificial Intelligence, *Doctor (Dr.) Dissertation*, Institut Teknologi Bandung, July 2010, Institut Teknologi Bandung, Bandung.

Sumari, A.D.W.; Ahmad, A.S.; Wuryandari, I. & Sembiring, J (2010a). Knowledge sharing in knowledge-growing-based systems, *Proceedings of the First International Conference on Green Computing 2010 and the Second AUN/SEED-Net Regional Conference on ICT* (ICGC-RCICT2010), Gadjah Mada University, pp. 329-333, ISSN 2086-4868, March 2010, Gadjah Mada University, Yogyakarta.

Sumari, A.D.W.; Ahmad, A.S.; Wuryandari, I. & Sembiring, J (2010b), Constructing brain-inspired knowledge-growing system: a review and a design concept, *The Second International Conference on Distributed Framework and Applications 2010* (DFmA2010), Gadjah Mada University, pp. 95-102, ISBN 978-602974704, August 2010, Gadjah Mada University, Yogyakarta.

Zubir, H.Y. (2008). The modeling of genetic regulatory system using knowledge growing basis, *Doctor (Dr.) Dissertation*, Institut Teknologi Bandung, December 2009, Institut Teknologi Bandung, Bandung.

Willett, E. (2006). *Genetic Demystified*, McGraw-Hill, ISBN 978-0071459303, New York.
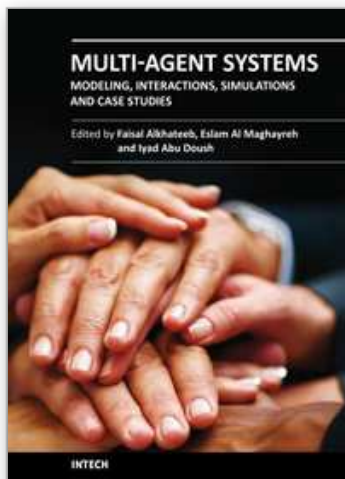
Multi-Agent Systems - Modeling, Interactions, Simulations and Case Studies

Edited by Dr. Faisal Alkhateeb

A multi-agent system (MAS) is a system composed of multiple interacting intelligent agents. Multi-agent systems can be used to solve problems which are difficult or impossible for an individual agent or monolithic system to solve. Agent systems are open and extensible systems that allow for the deployment of autonomous and proactive software components. Multi-agent systems have been brought up and used in several application domains.

INTECH
open science | open minds