Structure of a Well-Known Modularity-Inducing Problem Domain*

Author One
Institution
Omitted, Omitted
abc@def

Author Two
Institution
Omitted, Omitted
def@def

Author Three
Institution
Omitted, Omitted
ghi@def

ABSTRACT

Understanding biological organisms better can assist in solving complex engineering problems by applying their desirable characteristics and structures. A long-standing biological question is how organisms can quickly adapt themselves to new environments that are constantly changing, which is called evolvability. A key aspect to understand evolvability is to know the origin of modularity. Although various theories have been proposed to explain the conditions under which modularity arises, there is no consensus. In the computational biology, one prevalent theory argues that gene specialisation drives modularity. Our experiments indicated that there existed an inconsistency between this theory and observations in biology, regarding the dominant status of modular structures on evolvability. Subsequent experiments also indicated that networks with high fitness could be converted into modular structures by removing inter-module connections while their performance improved. Furthermore, a fluctuant landscape can also promote modularity.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability;

KEYWORDS

ACM proceedings, LaTeX, text tagging

ACM Reference Format:
Author One, Author Two, and Author Three. 2018. Structure of a Well-Known Modularity-Inducing Problem Domain. In *Proceedings of the Genetic and Evolutionary Computation Conference 2018 (GECCO '18).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Adaptability is an essential problem for evolutionary algorithms to solve. In other words, what can we do to facilitate engineered robots to evolve in order to adapt themselves to the constantly changing environments, just like biological organisms? Studies have indicated that the lack of modularity is one of the reasons that account for the incapability of artificial biological systems for adapting into and scaling up to higher complexity [7]. For example, artificial neural networks are assumed to be densely connected, whereas human brains exhibit modular components taking different responsibilities, such as hippocampus for dealing with novel situations and amygdala for emotional controls. As such, it is important to understand the conditions under which modularity spontaneously emerged in biology. Afterwards, engineers may leverage these conditions to design modular systems that can solve more complex problems and are able to autonomously adapt themselves to new working environments. One good example of the module-based engineering design is the "high cohesion, low coupling" principle in software engineering [6]. This understanding in modularity led to a software engineering booming in the current and last century [6].

Specifically, modularity is defined as the divisibility of structures or functions into sub-units that perform autonomously with each other [11]. In other words, a module is a group of elements whose associations occur preferentially within the group [4]. Furthermore, Many biological activities and structures can be modeled in the form of networks, such as animal brains, signaling pathways, etc. [3]. A network is modular if it can be partitioned into highly connected components, and between these components, there are only sparse connections [3]. Therefore, elements within a module will demonstrate the tendency of undertaking coherent functions independently from other elements outside of it [4] [8]. In biology, such modules exhibit ubiquity [11]. Specifically, they appear at various levels of biological organizations [4]. Modularity can promote the evolvability of organisms, where evolvability is defined as the capability of rapidly adapting to novel environments [10]. Two reasons can justify this statement. Firstly, a modular network may allow changes in a module without disturbing other modules; Secondly, modular structures can be reutilized and combined in different ways in order to perform new functions [4].

Despite the fact that modularity has gained research interests for decades [15], there is no consensus on its origin and evolutionary direction in biology [14]. Among various scenarios to explain the condition under which modularity emerges, two stand out, because their proposed conditions are commonly encountered in nature [15]. These two scenarios include modularity-varying evolutionary goals [7] and specializations in gene activity patterns [4]. Specifically, the former states that modular changes in environments may impose an impetus in the emergence of modularity [7]. That is, organisms that live in the environment whose sub-components are repeatedly and constantly altered demonstrate higher-level modularity than those living in the stable environment. This explanation is plausible due to the ubiquity of fluctuations in the environment [4]. However, despite the fact that environments are continuously changing, it is unclear to what extent they vary modularly [4].

*Produces the permission block, and copyright information

Espinosa-Soto and Wagner studied the conditions under which gene regulatory networks started exhibiting modular structures [4]. They concluded that modularity could arise as a by-product of gene specializations when gene regulatory networks acquire the ability to regulate towards multiple different patterns. Specifically, the distinct sub-components in the regulatory network to regulate sharing and different gene activity patterns will hamper each other's performance. Thus, modular networks that favor fewer connections between modules of the network will break the pleiotropic effect of regulating sharing and distinct gene activity patterns. Moreover, additional gene activity patterns can further improve the modularity. Their work is persuasive since the phenomena that gene regulatory networks acquire new gene activity patterns is ubiquitous in evolution. To be more specific, the same collection of genes exhibits different activity patterns at different phases of development or different locations in organisms [4]. Their theory can also act as an alternative explanation of why modular-varying environments result in modularity since organisms need to express different gene patterns for different environments [7] [4]. However, the experiments of Espinosa-Soto and Wagner lacked the crossover phase in their evolutionary simulations. Biologically, crossover is necessary.

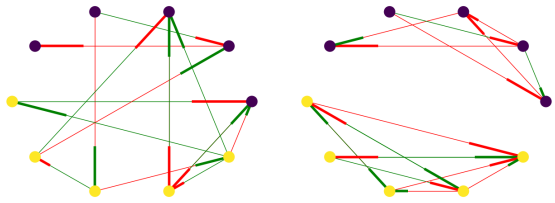
In this paper, we aim to investigate the plausibility of the theory stating that gene specialization drives modularity of organisms [4]. We will first explore whether there exist methods that can expedite the evolutionary process. For example, crossover is assumed to be an effective method to enhance the efficacy of combining useful traits in evolutionary simulations. Therefore, it is beneficial to explore whether there exists a crossover mechanism that can promote modularity. Similarly, the elitism, which is another common mechanism utilized in the artificial evolution, is also worthwhile exploring its contribution to the computational evolution. Furthermore, will different fitness evaluation methods give rise to different modularity levels?

Moreover, experiments in [4] did not demonstrate whether structures with high modularity has gained a dominant status on survivability. In biology, there is no organism that exhibits non-modular structures. As such, one may assume non-modular creatures have been extinct. Therefore, modular individuals are expected to have far better performance than non-modular ones, especially for complicated environments. As such, within the surviving simulated organisms in the gene specialization experiments, we will investigate the dominant status of modularity on survivability by comparing the fitness values of the eminent modular organisms to less modular ones.

Additionally, we also wish to discover what properties of modular structures will obtain in a long-term evolution. That is, towards what direction is the system with high modularity evolving? Although the experiments suggested a significant emergence of modular structures by gene specialization, they only reveal specialization is the origin of modular structures. It did not explain the evolutionary direction of modular systems.

2 BACKGROUND

Figure 1 demonstrates examples of non-modular and modular networks. In the inception phase of this project, we utilized the Louvain heuristics to compute the partition of the network vertices in order



(a) A non-modular example **(b) A modular example**

Figure 1: Non-modular and modular networks

to maximize the modularity of the given graph [2]. We applied the tournament selection scheme with the tournament size being three and the elitism mechanism with ten elites in every generation. As a result of this setting, the partition of the gene regulatory networks by the Louvain heuristics demonstrated a very low modularity score. As Figure 2 indicates where the green line represents the generation to introduce specialization, by simulating the work in [4], we had expected there would be a spike after gene specialization on modularity. Nevertheless, we observed a modularity decrease as a result after gene specialization. In order to understand this puzzling

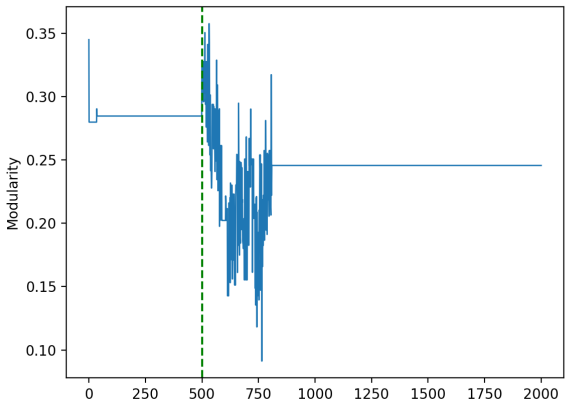


Figure 2: Modularity went worse after gene specialization

phenomenon, we removed the elitism mechanism and changed the tournament to proportional selection scheme. In consequence, we eliminated the deviant phenomenon as Figure 3 indicates. Therefore, we hypothesized that the elitism mechanism or the tournament selection scheme hamper the evolutionary process on evolving out modular structures.

3 METHODS

We utilize genetic algorithms as our evolutionary simulation tools. The gene regulatory network that we used in this paper was originally proposed by Wagner [13] and customized by Espinosa-Soto and Wagner [4] as well as Larson et al. [8].

All simulation code was implemented in Java 1.8.0 and Python 2.7.10. They are all publicly available at https://github.com/xxxxxxxxxx.

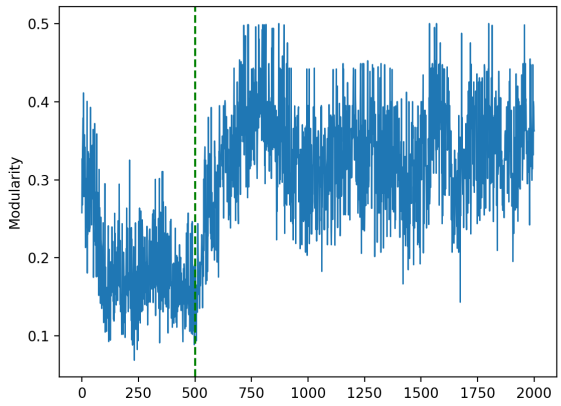


Figure 3: Modularity emerged after gene specialization without elitism

Modularity was evaluated using the NetworkX package with the community API [5]. All the generated data can be downloaded at: https://drive.google.com/file/d/0B9dNEi7lDXnldy1sNmZuTWNXMDg/view?usp=sharing.

3.1 Model

Cells in an organism display heterogeneity in functionalities and morphologies, while they contain the same set of genes. In other words, cells interpret the same genetic material in different ways so that their behaviors and structures vary. These distinct interpretations are due to the regulation via the activation and repression of genes [13]. In brief, effects of different genes are not mutually independent. A protein that is generated by a gene may activate or repress other genes. A gene regulatory network can be a mathematical directed graph to express these relationships of genes in an organism [13]. Specifically, genes can have two different patterns, namely activation and repression. The term "gene activity pattern" is adopted to represent the activeness status of the entire set of genes. Different gene activity patterns mean the distinct cellular functions and forms [4].

We reconstructed the model that was utilized in the work done by Espinosa-Soto and Wagner, which is a model to represent a gene regulatory network [4]. In this model, a gene regulatory network with N genes will be in the form of an adjacency matrix $A = a_{ji}$, which acts as a genotype of an individual. Each entry a_{ji} is restricted to be either 1, 0 or -1, which represents an activation, absence or repression interaction from gene j to gene i, respectively. The gene activity pattern of this network at time t can be expressed as a Boolean row vector $s_t = [s_t^0, ..., s_t^{N-1}]$. A certain gene i can either be active ($s_t^i = 1$) or inactive ($s_t^i = -1$). The transition of state activity is modeled by the equation below

$$s_{t+\tau} = \sigma[\sum_{j=1}^{N} a_{ji} s_t^j] \quad (1)$$

where $\sigma(x)$ equals 1 if $x > 0$ and is 0 otherwise.

3.2 Fitness

The fitness here evaluates the likelihood that an attractor is obtained when facing perturbations [4]. In other words, Espinosa-Soto and Wagner imposed a bias of robustness on their gene regulatory network models in order to indirectly select modular networks. This is because modular networks can limit perturbations in a module so that the overall structure will not be heavily affected [1]. That is, more modular networks are more robust.

There are two or more stages in their experiments on discovering the conditions under which modularity starts emerging. In the first stage, gene regulatory networks are evolved under selective pressure towards regulating a particular gene activity pattern, while facing some perturbations. The original gene activity pattern before perturbation is called a target. In the second and further stages, networks are evolved under selective pressure to regulate new gene activity patterns, while preserving the ability to regulate the old patterns. In the particular case where there were two gene activity patterns, the first stage lasted for 500 generations and the second took another 1500 generations.

The perturbations of targets are randomly generated in every generation when evaluating the fitness of gene regulatory networks. In Espinosa-Soto and Wagner's experiments, a network would face 500 perturbations comprising different corrupted versions of gene activity patterns. Each gene will have a probability of 0.15 to be perturbed into its opposite activity. A further study was conducted to explore a sufficient number of perturbations in order to shorten the computational time while maintaining a similar eventual improved modularity. It was concluded that 75 or 100 perturbations would lead to the noteworthy emergence of modularity [12]. Therefore, 75 perturbations are undertaken for evaluating the fitness of each gene regulatory network in order to reduce the running time.

Larson et al. applied another approach for evaluating the fitness of networks [8]. They generated a static set of perturbations at the beginning and utilised this same set of corrupted targets whenever network fitness was calculated. This method converts the original stochastic fitness evaluation into a deterministic one. That is, the evolutionary landscape of individuals under this fitness evaluation will remain unchanged in each generation. On contrast, Espinosa-Soto and Wagner's fitness evaluation will lead to the evolutionary landscape to shift every generation.

The fitness value of a gene regulatory network reflects its robustness in recovering from various perturbations. The error function compares an attractor of the network dynamics to the original gene activity pattern. That is, a successful network is able to regulate a corrupted pattern to its initial form. Then, the Hamming Distance G between the attractor and the original pattern was calculated. Previous experiments indicated that it normally took fewer than 20 transitions to reach the attractor [13]. Thus, non-stable attractors are assumed to be those gene regulatory networks that take more than 20 steps to attain the stability, or are cyclically stable. They are treated to have a maximum Hamming distance D_{max}. This is followed by a calculation of the contribution from each perturbation attractor to the fitness, which is defined as a developmental trajectory $\gamma = (1 - D/D_{max})^5$ [4]. Afterwards, this process is repeated to determine 75 γ_i, $1 \leq i \leq 75$. Finally, the fitness of a network is

calculated as

$$f(g) = 1 - e^{-3g} \tag{2}$$

where g represents the arithmetic mean of the sum of all γ_i [4]. As to cases where there are more than one gene activity patterns, the arithmetic mean of $f(g)$ for all the patterns was take. Consequently, a gene regulatory network with a high fitness is able to lead to different attractors matching different targets.

3.3 Evolutionary Simulations

Espinosa-Soto and Wagner imposed a bias towards low-density gene regulatory networks in mutation [4]. A node in the network has a probability $\mu = 0.05$ to mutate every generation, and it either can lose or gain an interaction. The probability for a node to lose an interaction can be calculated as

$$p(u) = \frac{4r_u}{4r_u + N - r_u} \tag{3}$$

where N is the number of gene nodes in a gene regulatory network, and r_u equals to the number of regulators of gene u [4]. That is, the number of genes that exert effects on gene u. In contrast, the probability for a gene u to obtain an interaction is defined to be $1 - p(u)$. That is, it can keep the sparseness of the network, which computational biology research suggests is necessary for the emergence of modularity.

Espinosa-Soto and Wagner did not apply a crossover mechanism in their simulation [4]. In the reconstructed model by Larson et al., they limited crossover to nine possible partition locations of a 10-node network, corresponding to nine possible rows for splitting the adjacency matrix of a network horizontally [8]. We call this horizontal crossover. When two matrices A_1 and A_2 are selected for crossover at index i, matrices of their children will be produced as

$$C_1[0 : i-1, :] = A_1[0 : i-1, :]$$
$$C_1[i : 9, :] = A_2[i : 9, :]$$
$$C_2[0 : i-1, :] = A_2[0 : i-1, :]$$
$$C_2[i : 9, :] = A_1[i : 9, :]$$

However, this horizontal crossover may not only make the parental networks exchange modular clusters, but also exchange some interactions between the two modules. This may corrupt modularity. In contrast, we use a crossover mechanism that swaps interactions between modules in a gene regulatory network with connections between modules in another network. We refer this as diagonal crossover. Compared with the crossover mechanism of Larson et al., this approach, as Figure 4 illustrates, will better preserve the community structure (Wilcoxon signed-rank test; $p < 0.0372$).

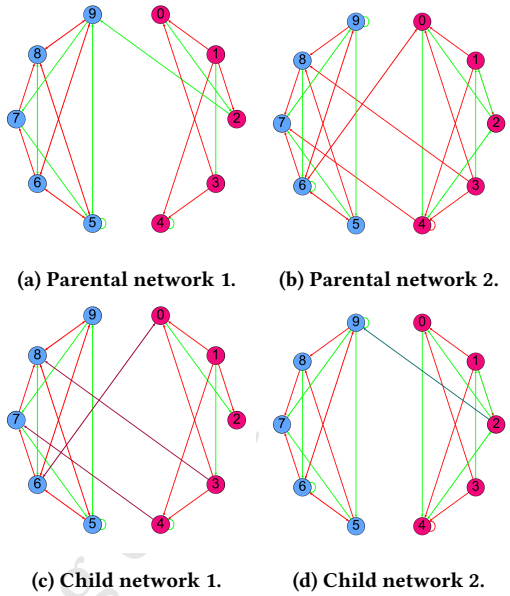
3.4 Modularity Metric

We adopted the Q scoring system to quantify modularity in a network based on the algorithm proposed by Newman [9]. Briefly, this approach is defined as the difference between the ratio of the number of edges in the network connecting nodes within a module over the number of all the edges, and the same quantity when assigning the nodes into the same modules yet edges are assumed to be randomly connected in the network [7]. Formally, Q is calculated as

$$Q = \sum_{i}^{K} [\frac{l_i}{L} - (\frac{d_i}{2L})^2] \tag{4}$$

where i represents one of the K potential modules within a network, L is the total number of connections in a network, l_i stands for the number of interactions in the module i, and d_i is the sum of degrees of all the nodes in module i [4]. In other words, Q considers the two ratios of both intra-module connection density and inter-module connection density [9]. A network that is considered to be good on modularity must consist of as many within-module edges and as few inter-module edges as possible. However, it will result in $Q = 0$ if all the nodes are partitioned into the same module.

The value Q will sit in the range of $[-\frac{1}{2}, 1)$. Nodes in the gene regulatory network are partitioned into different groups according to their regulating gene activity patterns.



(a) Parental network 1. (b) Parental network 2.

(c) Child network 1. (d) Child network 2.

Figure 4: Demonstration of diagonal crossover

4 EXPERIMENTS

Gene activity patterns and the essential parameters of our evolutionary simulations are provided in the form of Tables 1 and 2 in order to facilitate repeatability of these experiments. Unless specified, all the experiments are conducted using the original stochastic fitness evaluation proposed by Espinosa-Soto and Wagner in [4]. The detailed explanations of these parameters are given in Table 3. Overall, only the elite number and the tournament size will be specified in each experiment, since only those may vary in different experiments. All the other parameters are specified in Tables 1 and 2 are consistent in the experiments.

The Wilcoxon Signed-Rank Test was used to statistically determine the validity of the experimental conclusions. Each experiment contains 40 independent trials. The evaluation metrics include both the eventual fitness values and final modularity Q scores in the last generation.

Table 1: Table to test captions and labels

Gene Activity Pattern	Generation to Add a New Pattern
1, -1, 1, -1, 1, -1, 1, -1, 1, -1	0
1, -1, 1, -1, 1, 1, -1, 1, -1, 1	500

Table 2: Parameters of the evolutionary simulation

Edge Size	Perturbation Number	Perturbation Rate
20	75	0.15
Mutation Rate	Population Size	Tournament Size
0.05	100	Proportional
Reproduction Rate	Maximum Generation	Elite Number
0.9	2000	0 or 10

Table 3: Explanations of simulation parameters

Gene Activity Patterns	the patterns that are perturbed, and towards which gene regulatory networks evolve.
Generations to add a new pattern	the generations to add new gene activity patterns towards which networks evolve.
Edge Size	the initial number of edges in the original gene regulatory networks.
Perturbation Number	the number of corrupted versions of each gene activity pattern.
Perturbation Rate	the expectation of the number of corrupted genes in a pattern.
Mutation Rate	the probability of a gene node to gain or lose an interaction in a network.
Population Size	the number of individuals in the population in every generation.
Tournament Size	the size of the tournament selection; where tournament selection is used, the size of the tournament; where proportional sections is used, it is annotated as "proportional".
Reproduction Rate	the proportion of children reproduced over the entire population. Any vacancy will be filled by the tournament scheme selecting individuals from the previous generation.
Maximum generation	the generation when the simulation will terminate after reaching it.

4.1 Diagonal Crossover Mechanism Promotes Modularity

We simulated 40 independent evolutions for the development with no crossover and with each of the two crossover mechanisms, namely horizontal crossover and diagonal crossover, respectively. None of these simulations applied elitism. Overall, the diagonal crossover mechanism performed better than no crossover and the horizontal crossover, regarding both regulatory performance and modularity emergence, as Tables 4 and 5 indicate.

The Boolean model that we have utilised to simulate biological networks was originally proposed by Wagner in his study on "epigenetic stability" [13]. His work indicated that random recombination made no difference for the evolution of stability, which may be due to the freeness of random recombination on choosing locations to undertake crossover. This can corrupt the modular structures in biological networks.

Table 4: Results for diagonal crossover driving modularity

	Diagonal	Horizontal	No Crossover
Fitness	0.9492	0.9444	0.9476
Q Score	0.3278	0.2901	0.1919

Table 5: Statistical significant results for diagonal crossover driving modularity

	Fitness P	Q Score P
No < Horizontal	0.2415	9.2918e-7
Horizontal < Diagonal	0.0002	0.0372

Conversely, our experimental results suggested that proper recombination methods can contribute to the evolvability of organisms. The diagonal crossover proposed in this report is able to preserve underlying network modules. Although the crossover mechanism utilised by Larson et al. did not preserve community structures as well as diagonal crossover, its partitioning is still based on a network-like structure. This can be the reason why both of these two crossover mechanisms could help in obtaining modularity, with diagonal crossover better than horizontal crossover. Meanwhile, different combinations of parental traits can increase the diversity of the population so that the evolution can be more exploratory.

4.2 Greed Hampers Modularity

4.2.1 Elitism Hampers Modularity.

We simulated 40 evolutionary trials with 10 elites and without any elites. That was 80 trials in total. The experimental results indicate that elitism will hamper both the networks' regulatory capabilities and modularity emergence, as shown in Table 6 and 7.

Table 6: Results for elitism hampering the modularity

	Without Elites	With 10 Elites
Fitness	0.9492	0.9472
Q Score	0.3278	0.2745

Table 7: Statistical significant results for elitism hampering the modularity

	Fitness P	Q Score P
With 10 Elites < Without Elites	0.0019	0.0044

4.2.2 Proportional Exceeds Tournament Selection on Generating Modularity.

Similar to the elitism scheme in the evolutionnary simulation, tournament selection impose stronger selecting pressure than proportional towards invididuals for generating offspring. This is because the former only considers the relative order of individual fitness values.

We simulated evolutions with tournament size being 3 and 10, also proportional selection. The detailed results are demonstrated in

Tables 8 and 9. In summary, when the tournamennt size increases, i.e., the simulation gets more greedy, both the fitness values and modularity Q scores will decrease.

Table 8: Results for diagonal crossover driving modularity

	Tournament Size 3	Tournament Size 10	Proportional
Fitness	0.9432	0.9215	0.9461
Q Score	0.3511	0.2675	0.3223

Table 9: Results for diagonal crossover driving modularity

	Fitness P	Q Score P
Tournament Size 10 < Size 3	0.0052	0.0017
Tournament Size 10 < Proportional	0.0002	0.0091
Tournament Size 3 < Proportional	0.9031	0.2589

4.3 Stochastic Fitness Evaluation Excels Deterministic Fitness Evaluation

We simulated 40 independent evolutions each for both stochastic fitness evaluation and deterministic fitness evaluation. That is a total number of 80 experiments. The results showed that stochastic fitness evaluation outcompleted deterministic fitness evaluation on both survivability and modularity, as Table 10 and 11 indicate.

Table 10: Results for elitism hampering the modularity

	Stochastic	Deterministic
Fitness	0.9461	0.9322
Q Score	0.3223	0.1644

Table 11: Statistical significant results for elitism hampering the modularity

	Fitness P	Q Score P
Deterministic < Stochastic	0.0026	2.4369e-5

5 RESULTS

This is where we present the detailed results. We need fitness and modularity results. We also need the comparison between the optimum and the fittests high-modularity solutions. Then we need the results of deleting non-modular links from optimal solutions, and comparing resulting fitnesses, and showing the fitnesses of intervening paths. Finally, it may be desirable to check the result of evaluating the fitness of a good solution under one sampling method from one generation with the fitness of that solution under some other fitness function (i.e. how much do the fitnesses of an individual vary from generation to generation under Espinosa-Soto evaluation? How different are the Larson fitnesses? Are the differences larger or smaller for modular or non-modular solutions?

6 ANALYSIS

This may end up merged with the methods section. Detailed settings for the experiments, including full evolutionary tableaux.

7 DISCUSSION

7.1 Modular systems did not gain dominance on survivability

Greedy methodologies, including elitism and tournament selection scheme, impede the emergence of modularity under our evolutionary simulations. This implies that individuals who performed optimally in the early stage might not be optimal on modularity. In other words, the most competitive elites in each generation did not have the most modular gene regulatory networks.

Overall, these phenomena suggest that the modularity emergence condition, namely gene specialization promotes modular networks, may not be plausible to explain biological modularity. They indicated that modules in the simulated gene regulatory networks did not gain dominance in determining the survivability of individuals. However, biologically, modular networks are dominant and ubiquitous [11]. In order to further investigate the plausibility of this theory, namely specialization driving modularity, we obtained the most optimal gene regulatory network among networks that were the most modular. Conversely, we also collected the network that was the least modular among those that had the greatest fitness value. These networks were collected from the generated results of simulations in Section 4.1, using the dignonal crossover.

Biologically, I expected the fitness value of the latter would be lower than the fitness of the former. Nevertheless, the situation was converse. That is, some less modular networks were more robust than more modular ones, as Table 12 indicates. This is not consistent with what has been observed in biology. Initially, I hypothesized

Table 12: Modularity dominance analysis results for generated data of Section X.X

Generation Range	Modularity	Fitness
(500, 2000)	0.5000	0.9482
	0.1736	0.9502

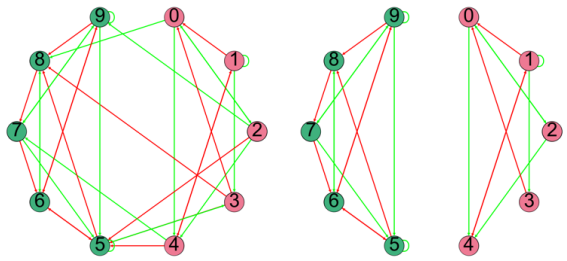
that the inconsistency was due to the targeted gene activity patterns being over-simple. That is, the number of genes in a pattern was not sufficient or the number of patterns was not enough. A modular network may give great performance on complex tasks, but worse than non-modular ones for simple tasks. Thus, I conducted a complicated evolutionary simulation consisting 7 patterns, each of which comprised 15 gene nodes. This evolution lasted for 35,000 generations and during the generation interval of (26000, 35000), it evolve towards all seven gene activity patterns. We conducted the modularity dominance analysis again and the results are in Table 13. Overall, the complex of gene activity patterns could not resolve the issue of non-dominance for modular networks on survivability.

Table 13: Modularity dominance analysis results for generated data of Section X.X

Generation Range	Modularity	Fitness
(26000, 35000)	0.5506	0.9100
	0.4151	0.9419

7.2 Inter-Module Connections Can Hamper Network Fitness

Fitness values of gene regulatory networks were measured after removing interconnections between modules in order to understand the functionality of inter-module interactions. The results indicated that among 40 networks which had the highest fitness values and relatively low modularity Q scores in their corresponding evolutionary simulations, 24 of them demonstrated higher fitness after manually converting them into modular structures by deleting inter-module edges. That is, there existed non-modular networks that exhibited better fitness performance after removing all the inter-module connections. For example, the right network in Figure 5 was the consequence of removing inter-module connections of the network in the left. The fitness value of the latter was 0.9502 after it had removed 6% connections of the former, whose fitness was 0.9472. Further statistical investigations will be conducted in the future. Originally, we suspected that this deviance was due to



(a) Before removal **(b) After removal**

Figure 5: Demonstration of inter-module connection removal

the fact that these modified solutions had a lower density than was expected from the evolutionary operations (sec 2.4), and thus may have been excluded from the search space. Nevertheless, further investigation revealed that the average number of edges for those networks that increased fitness values after trimming their inter-module connections was approximately 30. That is, it was not due to the bias on the sparseness that caused this anomaly.

In order to further comprehend this phenomenon on why our evolutionary simulations could not find a path to the trimed networks, we recorded the fitness value of removing one inter-module edge in turn, until deleting all of them. We plotted graphs as Figure 6, where x-axis represents the number of inter-module edges that have been discarded, y-axis represents the corresponding fitness values. Interestingly, most of our collected plots demonstrated a steady increasing trend for fitness vs deleting edge numbers, whereas genetic algorithms could not find these paths.

7.3 Fluctuant landscapes are essential for generating modularity

The stochastic fitness evaluation used by Espinosa-Soto and Wagner [4] demonstrated much higher fitness and modularity Q score than Larson et al.'s deterministic fitness evaluation. Therefore, we

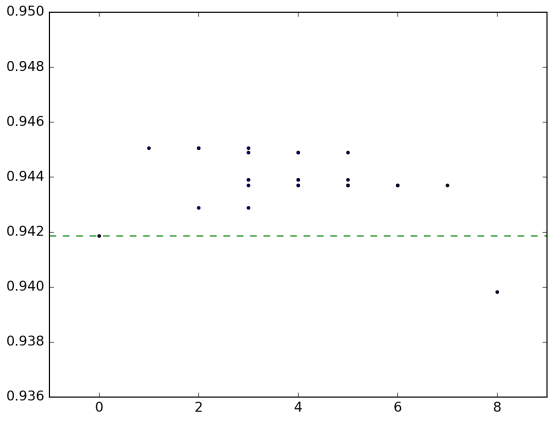


Figure 6: A removing inter-module connection path

hypothesized that a fluctuannt landscapes for individuals during the evolution might be necessary to develop high modularity. In order to verify this hypothesis, we collected the gene regulatory networks of the last generation, and mutated each network 9 times to generate their mutated neighbors. That is, each network would have 10 neighors, given including itself. Afterwards, we measured the fitness values of these neighors with the original target perturbations in the evolution and picked up thier maximum. In this fashion, we would have 40 maximum fitness values for both stochastic and deterministic fitness evaluation. Additionally, we also did the same process for the modularity Q score. Formally, a maximum value for a network N is collected with the formula

$$max(function(mutatedNeighbors(N))) \quad (5)$$

where $function$ can either be $fitness$ or $modularity$. Subsequently, our statistical test indicated that fitness of stochastic neighbors did not demonstrate advantages, whereas their modularity Q scores were much higher than deterministic neighbors, as Tables 14 and 15 indicate. In general, in order to evolve out high modularity, a combination of gene specialization and a constantly changing environments will be desirable, instead of applying gene specialization alone. Moreover, previously the statistics test revealed that the

Table 14: Results for comparsing stochastic and deterministic neighbor fitness

	Stochastic	Deterministic
Fitness	0.9410	0.9323
Q Score	0.3374	0.1851

Table 15: Statistical significant results for comparsing stochastic and deterministic neighbor fitness

	Fitness P	Q Score P
Deterministic < Stochastic	0.7223	2.6879e-5

stochastic approach would lead to a higher fitness value, whereas

this advantage disappeared when evaluating the fitness of mutated neighbors. Further investigation suggested that a deterministic, or static landscape may result in the searching getting stuck at the local optima. This is because for our 40 networks generated by deterministic fitness evaluation, the maximum fitness values for a network's neighbors were all from itself. That is, the neighbors of a network evolving in a static landscape always performed worse than themselves. Formally, for a network N,

$$max(fitness(mutatedNeighbors(N))) = fitness(N) \quad (6)$$

Furthermore, there existed a lot of networks produced by deterministic fitness evaluation whose fitness values were much lower (apprximately 0.88), compared to the rest of networks as well as those generated by stochastic fitness evaluation (approximately 0.93). We hypothesized that these low-performing networks are the cause on why statistically, fitness values generated by deterministic evaluation were lower than stochastic evaluation. Additionally, there may exist some correlation between getting stuck at local optima and modularity evolution.

7.4 More modular networks require fewer connections

As previous results suggested, interactions between modules sometimes do not contribute to and even hamper the regulation activity of networks. That is, a network can gain a better performance by removing those inter-module connections, which indicates that modular networks require fewer connections in total. In order to justify this hypothesis, we collected both of the most and the least modular network among those fittest individuals from each evolutionary simulation in Section 4.1, using the dignonal crossover. That is, given two networks that have the same fitness value, we would like to discover whether the more modular one needs fewer connections. Our statistical test verified this hypothesis to be correct, as Table 16 indicates.

Table 16: Results for verifying more modular networks require fewer connections

	Most Modular	Least Modular	Most < Least Modular p
Edge Number	24.6	29.925	6.1913e-7

Clune et al. stated that the evolutionary origin of modularity is due to the cost associated with every connection in the network [3]. They demonstrated this by their experiments indicating that there was a significant emergence of modular networks after imposing a penalty on the number of edges in the network [3].That is, modularity arose in order to minimise the connection costs. Specifically, they made simulated organisms evolve towards two objectives, namely to maximise the performance and to minimise the edge costs. However, in reality, biological organisms evolve in a single-objective fashion. That is, they are only selected under the pressure of fitting the living environments. Therefore, the theory stating that modularity comes from minimising connection costs may not be sufficiently plausible.

Our results revealed a converse causality of Clune et al.'s explanation on modularity. To be specific, the connecting costs of modular networks are lower may be because modular networks need fewer edges to support their activities than non-modular ones. It may be also due to this, Clune et al. can recognise and select more modular systems by choosing structures in which there are fewer connections. Nevertheless, containing fewer edges is a property of more modular networks, not their evolutionary origin.

8 CONCLUSIONS

In summary, we found that the diagonal crossover mechanism can promote the emergence of modularity. In contrast, elitism hampers the rise of modular networks, which indicates that early optimal individuals did not demonstrate high-level modularity. Further experiments also indicated that the theory on the origin of modularity resulting from specialization has limitations on explaining the surviving dominance of modular systems in biology. Furthermore, networks that have high fitness values could demonstrate better performance after converting them into modular structures by removing their inter-module edges. This suggests that modular networks initialized by gene specialization may evolve towards structures requiring a fewer number of connections in total. However, evolutionary simulations could not find these more optimal solutions. Moreover, individuals that live in more fluctuant environments can result in more modular network structures. Therefore, fluctuant landscapes can be essential for modularity evolution. In the future, we will aim to understand the reason why genetic algorithms could not find a path to individuals with better fitness and modularity. Additionally, we would also investigate the correlation between fluctuation of landscapes and the level of modularity.

ACKNOWLEDGEMENTS

?Do we need to acknowledge grants here? Assistance from Bongard? etc.

REFERENCES

[1] Alan Aderem. 2005. Systems biology: its practice and challenges. *Cell* 121, 4 (2005), 511–513.
[2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.
[3] Jeff Clune, Jean-Baptiste Mouret, and Hod Lipson. 2013. The evolutionary origins of modularity. In *Proc. R. Soc. B*, Vol. 280. The Royal Society, 20122863.
[4] Carlos Espinosa-Soto and Andreas Wagner. 2010. Specialization can drive the evolution of modularity. *PLoS computational biology* 6, 3 (2010), e1000719.
[5] Aric Hagberg, Pieter Swart, and Daniel S Chult. 2008. *Exploring network structure, dynamics, and function using NetworkX.* Technical Report. Los Alamos National Laboratory (LANL).
[6] Martin Hitz and Behzad Montazeri. 1995. Measuring coupling and cohesion in object-oriented systems. (1995).
[7] Nadav Kashtan and Uri Alon. 2005. Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America* 102, 39 (2005), 13773–13778.
[8] Ari Larson, Anton Bernatskiy, Collin Cappelle, Ken Livingston, Nicholas Livingston, John Long, Jodi Schwarz, Marc Smith, and Josh Bongard. 2016. Recombination Hotspots Promote the Evolvability of Modular Systems. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion.* ACM, 115–116.
[9] Mark EJ Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical review E* 69, 2 (2004), 026113.
[10] Massimo Pigliucci. 2008. Is evolvability evolvable? *Nature Reviews Genetics* 9, 1 (2008), 75–82.
[11] Gerhard Schlosser and Günter P Wagner. 2004. *Modularity in development and evolution.* University of Chicago Press.
[12] Mariko Totten. 2015. Exploring the Evolution of Modularity in Gene Regulatory Networks. (2015). https://scholarworks.uvm.edu/cgi/viewcontent.cgi?referer=

https://www.google.com.au/&httpsredir=1&article=1081&context=hcoltheses
[13] Andreas Wagner. 1996. Does evolutionary plasticity evolve? *Evolution* 50, 3 (1996), 1008–1023.
[14] Günter P Wagner and JG Mezey. 2004. The role of genetic architecture constraints in the origin of variational modularity. *Modularity in development and evolution* (2004), 338–358.
[15] Günter P Wagner, Mihaela Pavlicev, and James M Cheverud. 2007. The road to modularity. *Nature Reviews Genetics* 8, 12 (2007), 921–931.