ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE

EVOLVING BOOLEAN GRAPHS TO MODEL THE TOPOLOGICAL AND DYNAMICAL BEHAVIOR OF BIOLOGICAL REGULATORY NETWORKS AND THEIR METANETWORKS

M.Sc. THESIS

Burçin DANACI

Department of Phyiscs Engineering

Phyiscs Engineering Programme

MAY 2014

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE

EVOLVING BOOLEAN GRAPHS TO MODEL THE TOPOLOGICAL AND DYNAMICAL BEHAVIOR OF BIOLOGICAL REGULATORY NETWORKS AND THEIR METANETWORKS

M.Sc. THESIS

Burçin DANACI
(509111128)

Department of Phyiscs Engineering

Phyiscs Engineering Programme

Thesis Advisor: Prof. Dr. Ayşe ERZAN

MAY 2014

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

BOOLCU AĞLARIN EVRİMLEŞTİRİLMESİ İLE BİYOLOJİK REGULASYON AĞLARININ VE META-AĞLARININ TOPOLOJİK VE DİNAMİK ÖZELLİKLERİNİN MODELLENMESİ

YÜKSEK LİSANS TEZİ

Burçin DANACI
(509111128)

Fizik Mühendisliği

Fizik Mühendisliği

Tez Danışmanı: Prof. Dr. Ayşe ERZAN

MAYIS 2014

Burçin DANACI, a M.Sc. student of ITU Graduate School of ScienceEngineering and Technology 509111128 successfully defended the thesis entitled **"EVOLVING BOOLEAN GRAPHS TO MODEL THE TOPOLOGICAL AND DYNAMICAL BEHAVIOR OF BIOLOGICAL REGULATORY NETWORKS AND THEIR METANETWORKS"**, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Ayşe ERZAN** Istanbul Technical University	.....
Jury Members :	**Doç. Dr. Cem SERVANTİE** Istanbul Technical University	.....
	Prof. Dr. Muhittin MUNGAN Boğaziçi University	.....
		.....

Date of Submission : **23 June 2014**
Date of Defense : **26 May 2014**

FOREWORD

I would like to thank my advisor Prof. Dr. Ayşe Erzan for her efforts and guidance. I also acknowledge my college Mehmet Ali Anıl for use of his program and for useful discussions.

Finally, I would like to thank my family and friends for their endless support.

May 2014

Burçin DANACI
Physics Engineer

TABLE OF CONTENTS

Page

ABBREVIATIONS

GRN	**:** Gene Regulatory Network
TGRN	**:** Transcriptional Gene Regulatory Network
SP	**:** Significance Profile
GC	**:** Giant Component

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF SYMBOLS

$\mathbf{A}$	:	Adjacency matrix of a Boolean graph
p_0	:	Initial edge density
τ_j	:	State of node j
σ_j	:	Random key of node j
k_j	:	Degree of node j
a	:	Mean attractor length
I	:	Shannon information content
z_μ	:	z-score of motif μ
S_μ	:	Significance profile of motif μ
$\mathscr{O}(\mathbf{S}^{(\alpha)}, \mathbf{S}^{(\beta)}$	:	Overlap between set α and set β
Γ	:	Genotype
Φ	:	Phenotype

EVOLVING BOOLEAN GRAPHS TO MODEL THE TOPOLOGICAL AND DYNAMICAL BEHAVIOR OF BIOLOGICAL REGULATORY NETWORKS AND THEIR METANETWORKS

SUMMARY

Our study aims to capture the topological and dynamical features of gene regulatory networks by evolving Boolean graph populations subject to a fitness function favoring point attractors and two-cycles. According to our model, the set of the dynamical attractors a network has constitutes its "phenotype", whereas its adjacency matrix is considered as its "genotype". The relation between the dynamics (phenotype) and the wiring (genotype) is studied by comparing different features of independently evolved populations.

The distinguishing features of the graphs with short attractor lengths were determined by using a randomized control group. Randomization was carried out by rewiring the evolved graphs by preserving the total number of edges. The topological features of different networks were examined by comparing degree distributions and motif frequencies. For a quantitative comparison information content is used. The information content based on both the topological features such as degree distributions and motif frequencies and the dynamical features such as mean attractor lengths were found to be smaller for evolved graphs compared to those of randomized graphs.

The motif frequencies of the evolved graphs showed that feed back loops are suppressed, whereas the number of feed-forward loops and loopless motifs are relatively high. Significance profiles were used to compare motif frequencies of the evolved and biological networks. The biological networks we have studied are the core computational graphs of gene regulatory networks of *E. coli, B. subtilis* and *S. cerevisae*, extracted by using k-core algorithm. The significance profiles of the evolved networks were found to be similar to those of the biological networks.

Although the significance profiles of different populations of evolved Boolean graphs were similar to each other, in general, their other topological and dynamical features can be very different. As the slow relaxation of the mean attractor lengths to stasis indicates, the fitness landscape is a rugged one and different populations span different areas in genotype and phenotype spaces. To overcome this problem, these populations can be considered as metanetworks whose nodes are Boolean graphs. The structures of these metanetworks are strongly correlated to the vital features of the regulatory networks, such as evolvability and mutational robustness.

The Boolean graphs are connected in genotype space by an edge if they are one mutation away from each other. Coevolved networks are closer to each other in genotype space compared to the randomized networks, in the sense that the average mutational distance between the evolved networks is smaller than those of the randomized networks. The degree distributions of the metanetworks formed in genotype space generally fit Poisson distributions with the same mean value for

independently evolved populations. Most of these metanetworks have connected components containing more than half of the networks within the population. The randomized counterparts of the evolved networks, however, do not have any neighbors in the genotype space, since the minimum mutational distance between the randomized networks is larger than one.

The Boolean graphs form a metanetwork in phenotype space, where two Boolean graphs are connected by an edge, if they share at least one attractor. The edges are weighted by the sizes of the basins of attraction of the shared attractors and different threshold values are used to eliminate weak connections. The metanetworks of the evolved populations are highly connected and the sizes of their largest connected components decrease slower with the increasing threshold compared to the metanetworks of the randomized populations, pointing the robustness of the Boolean graphs under mutations.

BOOLCU AĞLARIN EVRİMLEŞTİRİLMESİ İLE BİYOLOJİK REGULASYON AĞLARININ VE META-AĞLARININ TOPOLOJİK VE DİNAMİK ÖZELLİKLERİNİN MODELLENMESİ

ÖZET

Ağlar, biyolojik sistemler de dahil olmak üzere çok çeşitli yapıların modellenmesinde kullanılmaktadır. Burada sistemi oluşturan elemanlar arasındaki etkileşimler sistemin zaman içindeki davranışını belirler. Her bir elemanın zamanda tanımlı bir durumu vardır. Bu durum ağın çeşidine bağlı olarak bir vektörle veya skaler bir alanla temsil edilebilir. Sistemin durumu ise bütün nodların durumlarının fonksiyonu olarak tanımlanır. Elemanların durumları, aralarındaki etkileşimlere bağlı olarak zaman içinde değişir ve sistemin davranışını belirler. Bu elemanların her biri ağın düğümlerini, aralarındaki bağlantılar ise düğümler arasındaki kenarları oluştururlar. Bu bağlantıların tümü ağın topolojisini, sistemin zaman içinde değişen davranışı ise dinamiğini belirler. Topoloji komşuluk matrisiyle tanımlanır. Komşuluk matrisinin her bir elemanı iki düğüm arasında bir kenar olup olmadığını ifade eder. Ağırlıklandırılmamış bir ağda bu matrisin elemanları 1 (kenar varsa) veya 0 (kenar yoksa) değerlerini alırlar. Ağırlıklandırılmış bir ağda ise bu matrisin elemanları kenarların ağırlıklarıyla orantılı değerler alırlar.

Hücre içi biyolojik fonksiyonları kontrol eden genetik regülasyon ağları Boolcu ağlar tarafından modellenebilmektedir. İlk kez Kauffman tarafından ileri sürülen bu modele göre ağdaki her bir düğüm bir gene karşılık gelir. Genler sadece iki durumda olabilirler. Aktif durumdayken protein kodlarlar. Pasif durumda olduklarındaysa kodlama işlemi gerçekleşmez. Bir genin kodladığı protein, başka bir geni aktifleştirebilir veya pasifize edebilir. Aktive etmesi durumunda bu etkileşim positif, pasifize etmesi durumunda ise negatif etkileşim olarak tanımlanır. Genler sonlu sayıda ise durumları sadece iki değer alabildiğinden sistemin faz uzayı sonlu sayıda durum vektörü içerir. Sistem herhangi bir başlangıç noktasından hareket ederek zaman içinde değişmeyen bir duruma ulaşırsa, bu duruma sabit nokta denir. Ancak sistemin durumu etkileşimlerden dolayı zaman içinde sabit kalmayabilir, fakat faz uzayı sonlu olduğundan sistem bir süre sonra daha önce bulunduğu bir duruma geri dönecektir ve o andan itibaren sabit sayıda durum arasında periyodik olarak salınım yapacaktır. Sistemin herhangi bir başlangıç noktasından başlayarak, yeterli zaman geçtikten sonra ulaştığı sabit noktaya veya tekrarlanan durum kümesine çekici denir. Genetik regülasyon ağlarında farklı çekicilerin farklı hücre tiplerine veya çevredeki değişikliklere verilen farklı tepkilere karşılık geldiği düşünülebilir. Kullandığımız modele göre bir ağın sahip olduğu dinamik çekiciler kümesi, bu ağın "fenotipi"ni, komşuluk matrisi ise "genotipi"ni tanımlamaktadır.

Ağların dinamik özellikleri topolojik özellikleri tarafından belirlenmektedir. Fakat topolojiyle dinamik arasındaki ilişkiyi veren bir formül bulunmamaktadır. Bu çalışmanın amacı, belirli bir dinamik özelliğe sahip olacak şekilde evrimleştirilmiş ağ popülasyonlarının topolojik özelliklerini kıyaslayarak hangi topolojik özelliklerin istenen dinamik özelliğe yol açtığını incelemektir. Burada aranan dinamik özellik

ağların kısa çekicilere sahip olmasıdır. Bunun sebebi hücre başkalaşımı gibi biyolojik fonksiyonlardan sorumlu genetik regülasyon ağlarının kısa çekicilere sahip olmaları gerektiğinin bilinmesidir.

Bu çalışmada, Boolcu ağları evrimleştirmek için sabit nokta çekicilerine veya ikili döngülere sahip ağları (çizgeleri) seçen bir genetik algoritma kullanılmıştır. Birbirinden bağımsız olarak evrimleştirilmiş popülasyonların çeşitli özellikleri karşılaştırılarak Boolcu ağların dinamiği (fenotipleri) ile komşuluk matrisleri (genotipleri) arasındaki ilişki incelenmiştir. Yapılan simülasyonlar sonucu bir popülasyondaki ağların ortalama çekici uzunluklarının simülasyonun ilk evresinde zamanla azaldığı ve bu azalışın $t^{-\gamma}$ şeklinde bir kuvvet yasasına uyduğu gözlemlenmiştir, burada t genetik algoritmanın zaman adımıdır. İlk 100-150 adımdan sonra ise ortalama çekici uzunluğu 2'den küçük bir değerin etrafında küçük salınımlar yapmaya devam etmiştir.

Kısa çekicilere sahip çizgelerin ayırdedici özellikleri bu çizgelerin randomize (rastgele hale getirme) edilmesiyle oluşturulmuş kontrol grubu kullanılarak saptanmıştır. Randomizasyon işlemi evrimleştirilmiş çizgelerin toplam kenar sayılarını koruyarak rastgele bir şekilde yeniden üretilmesiyle gerçekleştirilmiştir. Farklı çizgelerin topolojik özellikleri, derece dağılımları ve motif frekansları karşılaştırılarak incelenmiştir. Niceliksel bir karşılaştırma için enformasyon içeriği kullanılmıştır. Enformasyon içeriği hem motif frekansları ve derece dağılımları gibi topolojik özelliklere dayanarak hem de ortalama çekici uzunluğu gibi dinamik bir özelliğe dayanarak hesaplanmıştır ve her iki durumda da evrimleştirilmiş çizgelerin enformasyon içeriği randomize çizgelere göre daha düşüktür.

Evrimleştirilmiş çizgelerin motif frekansları bazı motiflerin hemen hemen bütün çizgelerde baskılandığını, bazılarının ise yüksek oranda ortaya çıktığını göstermektedir. Evrimleştirilmiş çizgelerde yüksek kenar sayısına sahip veya geri beslemeli motifler düşük oranda görünürken ileri beslemeli motiflerin frekansları oldukça yüksektir. Evrimleştirilmiş çizgelerle biyolojik ağların motif frekanslarını karşılaştırmak için "anlamlılık profili" kullanılmıştır. Karşılaştırma için k-çekirdek algoritması kullanılarak *E. coli, B. subtilis* ve *S. cerevisae*nin genetik regülasyon ağları en çok kenara sahip düğümlerinden oluşan alt çizgelere ayrıştırılmıştır. Biyolojik ağların bu alt çizgeleriyle evrimleştirilmiş Boolcu ağların anlamlılık profillerinin birbirlerine benzer olduğu saptanmıştır.

Evrimleştirilmiş Boolcu ağlardan oluşan farklı popülasyonların anlamlılık profilleri genel olarak birbirine benzese de, diğer topolojik ve dinamik özellikleri birbirinden oldukça farklı olabilmektedir. Ortalama çekici uzunluğunun stasis durumuna yavaş bir şekilde yaklaşmasının da gösterdiği gibi "uyumluluk yüzeyi" engebeli bir yüzeydir ve farklı popülasyonlar genotip ve fenotip uzayının farklı alanlarını taramaktadırlar. Bu sorunu aşmak için bu popülasyonlar düğümlerinde Boolcu çizgeler bulunan meta-ağlar olarak düşünülebilirler. Bu meta-ağların yapıları regülasyon ağlarının evrilebilirlik ve mutasyonlara karşı dayanıklılık gibi hayati önem taşıyan özellikleriyle yakından ilişkilidir.

Genotip uzayındaki meta-ağlar için yaptığımız tanıma göre Boolcu ağlar birbirlerine bir mutasyon adımı uzaklıktaysa bu ağlar bir kenarla bağlı kabul edilirler. Birlikte evrilmiş çizgeler randomize çizgelere kıyasla, genotip uzayında birbirlerine daha yakınlar; başka bir deyişle evrimleştirilmiş çizgeler için ortalama mutasyon uzunluğu randomize çizgelerde oluğundan daha küçüktür. Birbirinden bağımsız olarak evrimleştirilmiş popülasyonların genotip uzayında meydana getirdiği meta-ağların

derece dağılımları aynı ortalama değere sahip Poisson dağılımına uymaktadır. Çoğu meta-ağın popülasyondaki çizgelerin yarısından fazlasını içeren bağlı bir bileşeni bulunmaktadır. Randomize popülasyonlarda çizgeler arası minimum mutasyon uzunluğu birden büyük olduğundan bu poülasyonlardaki çizgelerin komşuları bulunmamaktadır.

Boolcu ağlar fenotip uzayında da bir meta-ağ oluşturabilirler. Burada yaptığımız tanıma göre iki çizgenin en az bir tane ortak çekicisi varsa bu çizgeler bağlı kabul edilmektedirler. Bu kenarlar çizgelerin ortak çekicilerinin çekici havzalarının büyüklükleriyle orantılı olarak ağırlıklandırılmıştır. Çizgelerin dayanıklılığını test etmek için değişen eşik değerlerine göre eşik değerinden düşük ağırlığa sahip kenarlar elenerek meta-ağların en büyük bağlı bileşenleri karşılaştırılmıştır. Buna göre evrimleştirilmiş popülasyonların meta-ağlarının en büyük bağlı bileşenleri eşik değeri artırıldığında randomize meta-ağların en büyük bağlı bileşenlerine göre daha yavaş küçülmektedir. Bu da evrimleştirilmiş çizgelerin, fenotiplerinde fazla değişiklik olmaksızın genotip uzayında geniş alanları tarayabildikleri anlamına gelmektedir.

1. INTRODUCTION

A network is a dynamical system of elements interacting with each other. Each element or node of a network has a state defined in time. The states can be represented by vectors or scalar fields depending on the network's type. The interactions between the nodes are represented by edges which can be directed or undirected depending on the nature of interactions. The overall state of the system is a function of the states of the individual nodes, defined in time and space. In the synchronous description the states of all of the nodes are simultaneously updated at discrete time steps, whereas in the asynchronous description, the nodes are sequentially updated. Asynchronous update is more realistic in the sense that the nodes can be updated at any moment in continous time. However this approach mendates prior knowledge on the sequence of the updates or introduces stochasticity by randomly choosing the nodes to update [1]. Therefore, for a more general and deterministic model, synchronous update of nodes is used in this study.

The dynamcis of a network is determined by its topology [2–7]. Therefore, the topological features of different types of networks were examined over the last years. Some of the distinguishing features of networks are degree distribution, clustering coefficient, average path length and motif structures. For instance, random graphs with sufficiently large number of nodes and a fixed probability of connections between nodes have Poisson distribution of degrees, whereas scale-free networks have power-law distribution of degrees [8].

One of the important applications of dynamical networks is in biology. The idea that gene regulatory networks (GRN) can be modeled by Boolean graphs was presented by Kauffman [9, 10]. The viability of biologically useful networks is usually associated with having short attractor lengths [5]. For instance, in the case of cell differentiation, once the GRN reaches an attractor state, or cell type, no substantial change in the state occurs later on.

In this study, we aim to model the topological and dynamical features of GRNs by evolving Boolean graph populations subject to a fitness function favoring point attractors and two-cycles. The statistical features of artificially evolved networks having short attractor lengths are studied and compared to those of GRNs. The evolved networks are directed graphs where genes are represented by vertices and the interactions are modeled by Boolean “keys" assigned to vertices [2], rather than random Boolean functions. The networks were evolved according to a genetic algorithm [11] favoring short attractor lengths. Significance profiles of motif frequencies [4] of the networks were calculated and compared to those of GRNs.

2. EVOLVING A MODEL POPULATION OF BOOLEAN GRAPHS WITH SHORT ATTRACTORS

In this chapter, the model [2] used for evolving Boolean graphs is explained. According to this model, random Boolean graphs [12–14] are generated and a matrix of Boolean keys is assigned to each of them. These graphs are evolved according to a genetic algorithm favoring short attractor lengths.

2.1 Boolean Graphs with Random Keys

According to our model, each graph consists of N nodes which can take on the values of 1 or 0. If there is a dicrected edge between two nodes, the corresponding element in the adjacency matrix **A** is 1, and 0 otherwise. Random directed graphs are generated with an expected edge density p_0. Therefore, elements of **A** takes on values 1 with a probability p_0 and 0 with a probability $1-p_0$. The graphs are updated synchronously.

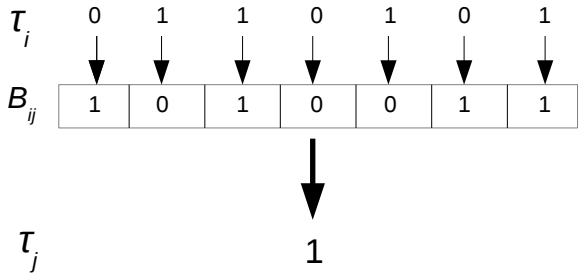

Figure 2.1: Boolean keys and inputs coming from neighbours for an arbitrary vertex.

The dynamics of the graphs are determined by Boolean functions. The Boolean functions are constructed by using random “key"s. A random key is a vector $\sigma_j = (\sigma_{1j}, \ldots, \sigma_{ij}, \ldots, \sigma_{Nj})$ assigned to each node. We generate these Boolean keys randomly in the beginning of the simulations and keep them fixed throughout the process. The XOR function is used to process the random keys. If the input $\tau_i = 0, 1$ from node i complements the corresponding element of the key, the XOR function

returns "true". We compute the updated state of the j'th node as,

$$\tau_j(t+1) \equiv \begin{cases} 1 & \text{for } \sum_i^N A_{ij}(\tau_i(t) \text{ XOR } \sigma_{ij}) > \frac{1}{2}k_j \\ 0 & \text{otherwise} \end{cases}, \quad \textbf{(2.1)}$$

where $k_j = \sum_i^N A_{ij}$ is degree of node j. The new state of the node is equal to the majority of the logical operators (see Fig 2.1).

2.2 Implementation of Genetic Algorithm

In our model, dynamics of the Boolean graphs were defined by the lengths of their attractors. In the case of gene regulatory networks, each attractor corresponds to a phenotype. Although oscillations in the production rates of the proteins are crucial for a cell's life cycle, the core regulatory parts of the graphs, which control most of the genes' activities, were assumed to have either fixed point attractors or limit cycles with short periods. Therefore, a genetic algorithm is used to obtain populations of Boolean graphs with short attractor lengths.

The genetic algorihm [11] utilizes a fitness function $f(a)$ depending on the average attractor length a of the graph. The average is taken over the whole phase space. So all possible initial conditions are taken into account. The fitness function selects graphs with attractor lengths $a \leq 2$ with a given probability. Selected graphs are copied and then mutated by rewiring the edges. This rewiring procedure preserves the in- and out-degrees of each node.

Our simulations run for graphs with $N = 7$, for two different sets of populations with initial edge densities $p_0 = 0.2$ and $p_0 = 0.5$. 16 populations with 10^3 graphs each were generated, in each case.

The steps of the algorithm are described below:

1) Generate a set of randomly wired Boolean graphs, each with the same randomly generated Boolean keys as described above.

2) Choose the graphs according to the fitness function $f(a) = P$ for $a \leq 2$, 0 otherwise. Clone the selected graphs. For rapid convergence the value of P was taken as 1/2.

3) Mutate the clones by rewiring their edges while preserving the in- and out-degrees of each node. This is done by randomly selecting two pairs of connected nodes. The terminals of the edges are switched.

4) Kill an equal number of randomly selected individuals.

5) Go back to step 2.

3. SIMULATIONS

The details of the simulations are covered in this chapter. The randomization process is explained in Section 3.1. The changes in the dynamical and topological features of the Boolean graphs during the course of evolution are discussed in Section 3.2. In Section 3.3, multistationarity, which is related to having multiple attractors, of evolved graphs is discussed. Section 3.4 covers the effects of increasing connectivity on multistationarity, for the random graphs. The module Kreveik is used to for computation [15].

3.1 Randomization

The evolved populations of graphs are compared with a randomized control group, to determine the distinguishing features of the graphs with short attractor lengths. Randomization process was carried out by rewiring the edges while keeping the total number of edges fixed. Our graphs are relatively small, and the resulting graphs have smaller edge density compared to the initial edge densities of the populations, in general. Therefore, the phase space of possible graphs has a limited size and when the graphs are rewired by preserving the in- and out-degrees, a graph that is already present in the evolved population is obtained most of the time. Due to this non-randomizability problem, we relaxed the constraint by preserving the total number of edges, instead of in- and out- degrees, seperately.

The statistical features of both evolved and randomized graphs are examined by comparing their distributions. Shannon information content defined as [16]

$$I = -\sum_{\mu} p(\mu)\log(p(\mu)) \tag{3.1}$$

is used to make a quantitative comparison.

3.2 Evolution of the Populations

The mean attractor lengths change very rapidly during the early course of evolution. After 150-200 time steps the mean attractor lengths stabilize between the values 1 and 2 as seen from Fig. 3.1.

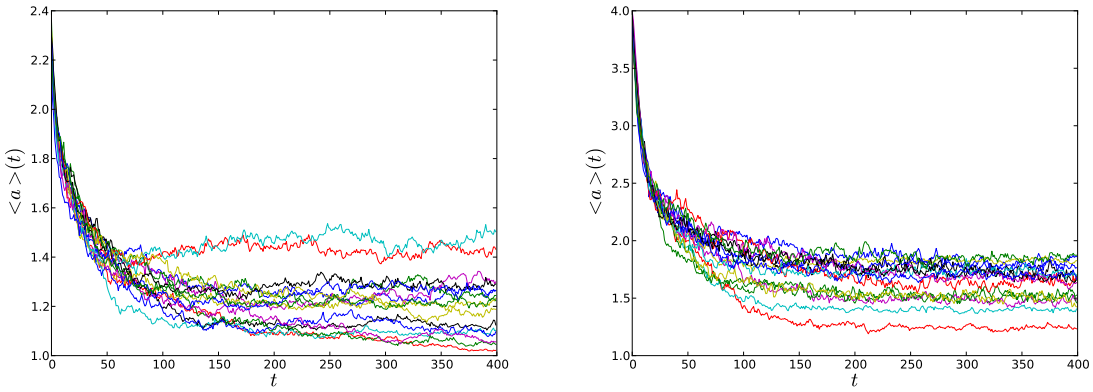

Figure 3.1: Change in the mean attractor lengths with respect to number of iterations of the genetic algorithm, for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

After a transient time, the decrease in the mean attractor lengths with increasing time fits a power law decay,

$$\langle a \rangle \sim t^{-\gamma}, \tag{3.2}$$

over the interval $7 < t < 150$ (see Fig. 3.2). Here $\langle a \rangle$ is the mean attractor length averaged over the whole set, and t is the number of iteration steps. The exponent values are shown in Table 3.1.

Table 3.1: Parameters for the power law decay, over the time interval $7 < t < 150$. The mean attractor lengths $\langle a \rangle_S$, averaged over a hundred time steps within the stasis region, $300 < t < 400$. The rms error of the linear fit are calculated and shown in the error bars for γ. $\langle a \rangle_F$ is the average attractor length at $t = 400$ and $\langle a \rangle_r$ is the average attractor length of the randomized graphs. The average values are calculated over 16 sets. Mean attractor lengths for independent sets are given in Appendix A.

p_0	γ	$\langle a \rangle_S$	$\langle a \rangle_F$	$\langle a \rangle_r$
0.2	0.15 ± 0.03	1.21 ± 0.12	1.21	2.01
0.5	0.20 ± 0.05	1.63 ± 0.17	1.61	3.54

The mean length of the attractors $\langle a \rangle_S$, averaged over a hundred steps, $300 \leq t \leq 400$, and the average of the mean attractor lengths taken at the final time step $t = 400$ $\langle a \rangle_F$, are shown in Table 3.1. The average values are compared with the average attractor lengths of the randomized sets, $\langle a \rangle_r$. (For the values of individual sets see Appendix A.) Table 3.1 shows that average of the mean attractor lengths of the evolved graphs are smaller (almost half as much) than the same average over the randomized graphs.

It can be observed that the average attractor lengths drop rapidly to 2 or below, from Fig. 3.1, implying that most of the attractors in the end are either fixed points or period 2 cycles. As the distributions of average attractor lengths in Fig. 3.3 show, for the evolved sets with $p_0 = 0.2$, the probability of a graph having an attractor with length smaller than 2 is about 0.8, whereas for randomized sets this probability is about 0.45. For the evolved sets with $p_0 = 0.5$, the probability of a graph having an attractor with length smaller than 2 is about 0.55, whereas for randomized sets this probability is about 0.15. These results show that the probability of having smaller attractor lengths is higher for the evolved sets with $p_0 = 0.2$, compared to those of $p_0 = 0.5$. However, in both cases, the evolved graphs have a much narrower distribution compared to their randomized versions. As seen from Fig. 3.4, information content of the evolved sets are substantially lower than the information content of the randomized sets for all the populations.

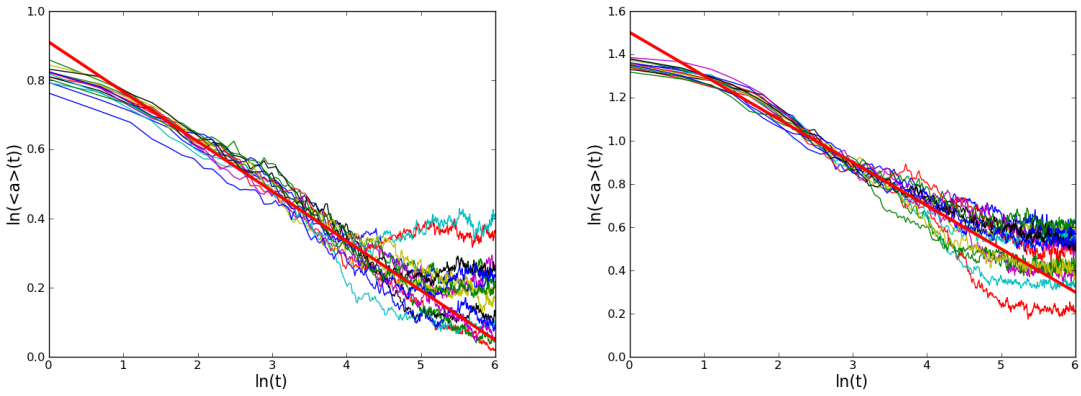

Figure 3.2: Change in the mean attractor lengths with respect to number of iterations of the genetic algorithm, for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel), plotted on a log-log scale. Red thick lines represent lines of best fit found by least square method.

Here, we suggest a measure for *selectivity*, by comparing the information content of the evolved and randomized distributions. If we define the selectivity as the difference

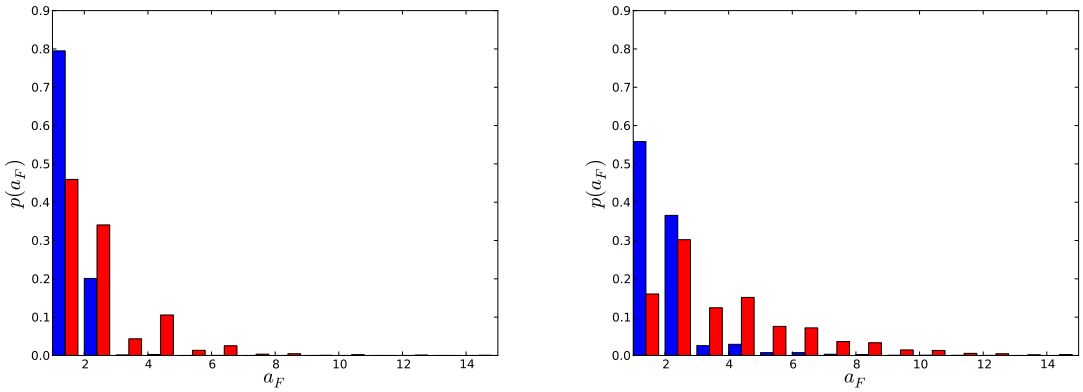

Figure 3.3: Distributions of mean attractor lengths within populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel). Blue and red bars represent evolved and randomized sets, respectively.

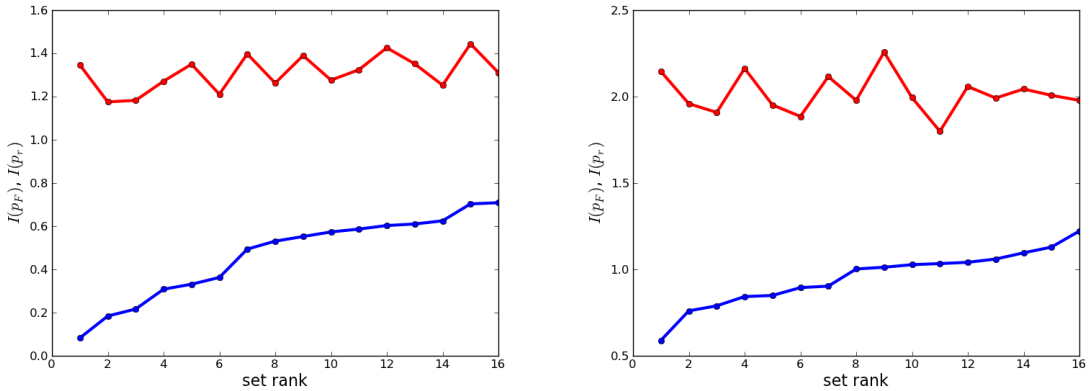

Figure 3.4: Information content $I(p_F)$ and $I(p_r)$ based on the distributions of the attractor lengths within the populations within the evolved (represented by blue dots) and randomized (represented by red dots), respectively. Populations with $p_0 = 0.2$ are shown on the left panel and populations with $p_0 = 0.5$ are shown on the right panel. Sets are ranked according to the mean attractor lengths of the evolved populations.

between the information content I_F and I_r of the evolved and randomized distributions, respectively, and normalizing by I_r, we obtain

$$s \equiv \frac{I_r - I_F}{I_r} \quad . \tag{3.3}$$

For populations with $p_0 = 0.2$, $0.46 < s < 0.94$ is found with average selectivity $\langle s \rangle = 0.65$. For populations with $p_0 = 0.5$, $0.38 < s < 0.73$ is found with $\langle s \rangle = 0.53$. This difference between the sets can be understood in terms of the small phase space of the graphs with lower edge densities.

The mean degrees of the graphs decrease with the number of iterations, in some of the populations (Fig. 3.5). However, this observation does not hold for all of the populations. In the populations with $p_0 = 0.2$, where initial mean degrees are around

1.6, the mean degrees of the resulting graphs change between 1.0 and 1.6 (left panel of Fig. 3.5). In the populations with $p_0 = 0.5$, however, the mean degrees of the graphs fluctuate in a relatively larger range, between 2 and 4, one of them even exceeds their initial mean degree 3.5. Although mean attractor lengths converge very fast, within the first 100 steps, mean degrees of some of the populations continue to fluctuate. This shows that the relationship between the mean degree and the attractor lengths is not so straightforward and one must look further into the topological features of the graph.

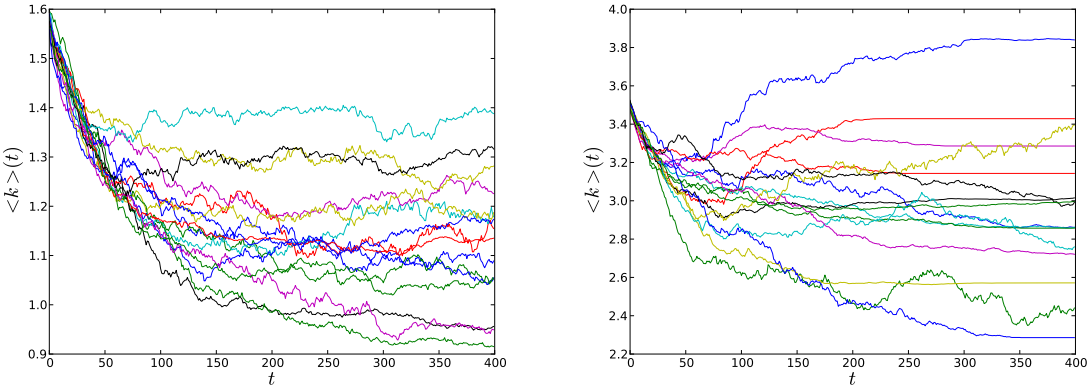

Figure 3.5: Change in the mean degrees with respect to t for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

In Fig. 3.6, we see that, for randomized graphs, there is a correlation between degrees and attractor lengths, average length of attractors increases with increasing degrees, in general. However, for evolved graphs, there is no correlation between degrees and attractor lengths. We see from Fig. 3.5 and Fig. 3.6 that, populations can follow very different evolutionary paths, especially when their initial mean degrees are higher, and there is no direct relationship between changes in attractor lengths and mean degrees.

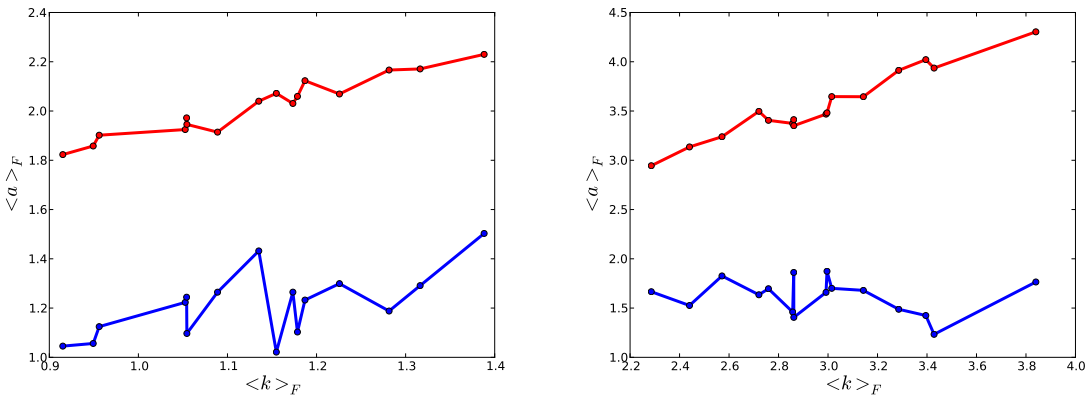

Figure 3.6: Mean attractor lengths versus mean degrees for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel). Blue and red dots represent evolved and randomized sets, respectively.

3.3 Multistationarity of the Set of Evolved Graphs with Short Attractors

A system having more than one dynamical attractor is called multistationary. Thomas and Kaufman have proposed the topological features which lead to multistability and periodic cycles for the systems described by differential equations [6] and for the discrete systems with the asynchronous description [7]. The comparison between the asynchronous and synchronous descriptions show that their findings do not always hold for the synchronous case.

Thomas and Kaufman characterizes the graphs according to the presence of feedback loops and the type of interactions. As they have shown, a feedback loop comprising an even number of negative (repressive) interactions leads to multistability, whereas a feedback loop comprising an odd number of negative interactions leads to oscillatory behavior, in the asynchronous description. Two simple example graphs given below, each consisting of 3 nodes, illustrate the differences between the synchronous and asynchronous descriptions.

The first graph considered by Thomas and Kaufman is a feedback loop with two negative interactions [7], as shown in Fig. 3.7. The sequences of states starting from different initial states and the attractors they reach, for synchronous and asynchronous updates are shown in Table 3.2. The point attractors are shown in brackets.

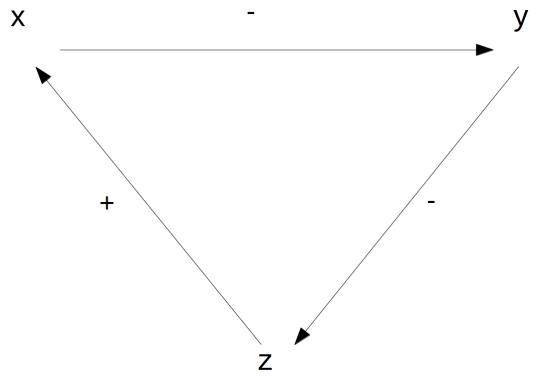

Figure 3.7: A feedback loop with two negative interactions.

In the asynchronous description, the state sequence is determined by the time delays. Here, it is assumed that $t_x < t_y < t_z$, where t_x, t_y and t_z are time delays in the activations of x, y and z, respectively.

Table 3.2: State sequences of a feedback loop with two negative interactions.

synchronous	asynchronous
$x\ y\ z$	$x\ y\ z$
0 0 0	1 0 0
0 1 1	0 0 0
1 1 0	0 1 0
0 0 0	[0 1 0]
1 1 1	0 1 1
1 0 0	1 1 1
0 0 1	1 0 1
1 1 1	[1 0 1]
0 1 0	1 1 0
0 1 0	[0 1 0]
1 0 1	0 0 1
1 0 1	[1 0 1]

As seen from Table 3.2, in the asynchronous description, there are two attractors, [010] and [101], and both of them are stable fixed points, whereas in the synchronous description, there are two attractors with period 3. The states 010 and 101 are fixed points in the synchronous descripition as well, however, they are unstable. The state sequences under the synchronous update show that oscillatory behavior can arise from feedback loops with even number of negative interactions as well, contrary to the case under the asynchronous update.

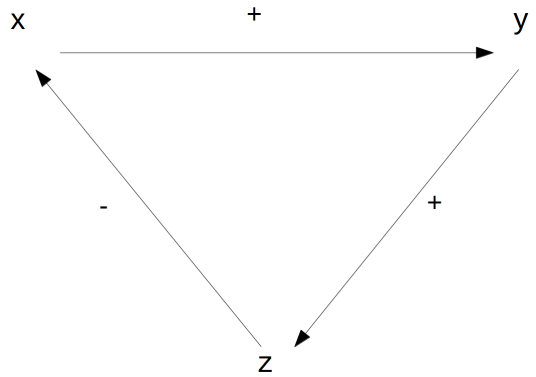

Figure 3.8: A feedback loop with one negative interaction.

The second example given by Thomas and Kaufman is a feedback loop with one negative interaction [7], as shown in Fig. 3.8. The sequence of states are shown in Table 3.3. Both the synchronous and asynchronous updates yield an attractor with period 6 and with the same sequence of states. However, in the asynchronous description there is only one attractor since the two other states 010 and 101 separately fall into the attractor. On the other hand, in the synchronous description, there are two

Table 3.3: State sequences of a feedback loop with one negative interaction.

synchronous	asynchronous
$x\ y\ z$	$x\ y\ z$
0 0 0	0 0 0
1 0 0	1 0 0
1 1 0	1 1 0
1 1 1	1 1 1
0 1 1	0 1 1
0 0 1	0 0 1
0 1 0	0 1 0
1 0 1	1 1 0
0 1 0	. . .
	1 0 1
	0 0 1
	. . .

attractors, since states 010 and 101 form a cycle independent from the attractor with period 6.

These two examples show that the topological features determining the dynamics of the system in the synchronous description are not as straightforward as they are in the asyncronous description, even for simple graphs with 3 nodes.

3.3.1 Number of attractors of the evolved and randomized populations

The distributions of the number of attractors of the evolved and randomized sets are shown in Figure 3.9. Since our fitness function depends only on the attractor length, there were no constraints on the attractor counts. Therefore, the distributions of the number of attractors of evolved graphs are similar to those of randomized graphs. However, the distributions of the populations with $p_0 = 0.2$ are completely different from those of the populations with $p_0 = 0.5$. One difference is the mean number of attractors. For the populations with $p_0 = 0.2$, the mean numbers of attractors of evolved and randomized graphs are around 7, wehereas for the populations with $p_0 = 0.5$, the mean numbers of attractors of evolved and randomized graphs are around 2. We see that p_0 is an important parameter, strongly affecting the mean number of attractors. This implies a correlation between the number of attractors and the mean degrees of the graphs. As seen from Figure 3.10, such a correlation exists for randomized graphs, the number of attractors decreases with the increasing mean degree. However, for the

evolved graphs, there is no correlation between the number of attractors and the mean degree.

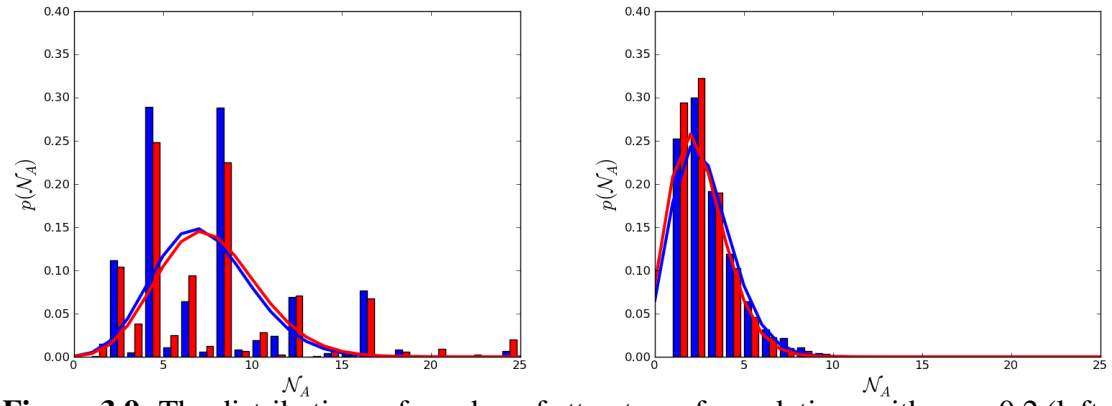

Figure 3.9: The distributions of number of attractors of populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel). Blue and red bars represent evolved and randomized populations, respectively. The blue and red curves represent Poisson distributions with the same mean as evolved and randomized sets, respectively.

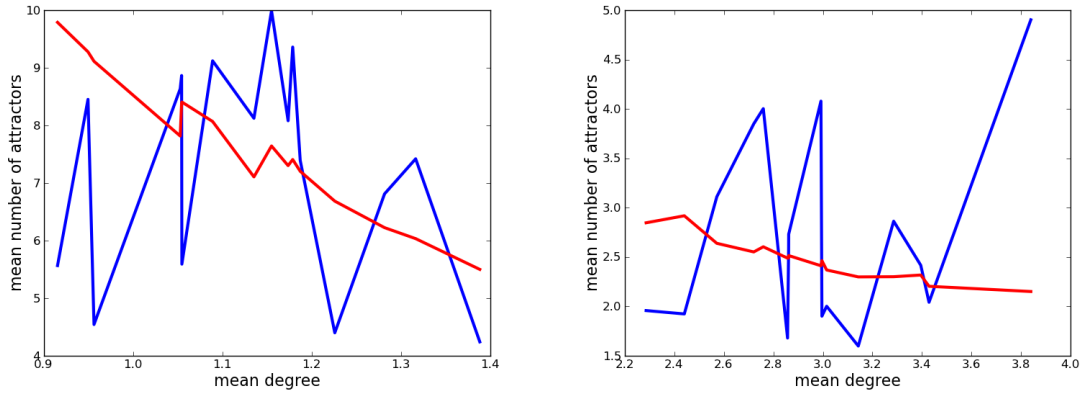

Figure 3.10: Mean number of attractors versus mean degrees for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel). Blue and red lines represent evolved and randomized sets, respectively.

Another important feature about the dynamics of a graph is the sizes of basins of attraction. Attractors with small basins of attractions do not have a significant effect on the overall score of the graph which is attractor length averaged over all the initial conditions. The distributions of the basin sizes are shown in Figure 3.11. For the populations with $p_0 = 0.2$, the distributions of basin sizes of evolved and randomized graphs are generally similar to each other. For the basin sizes larger than 3, both evolved and randomized graphs have the highest probabilities of having attractors with basin sizes powers of 2. The probability of an evolved graph having an attractor with basin size 3 or smaller is very low. The peaks of the evolved distribution are higher

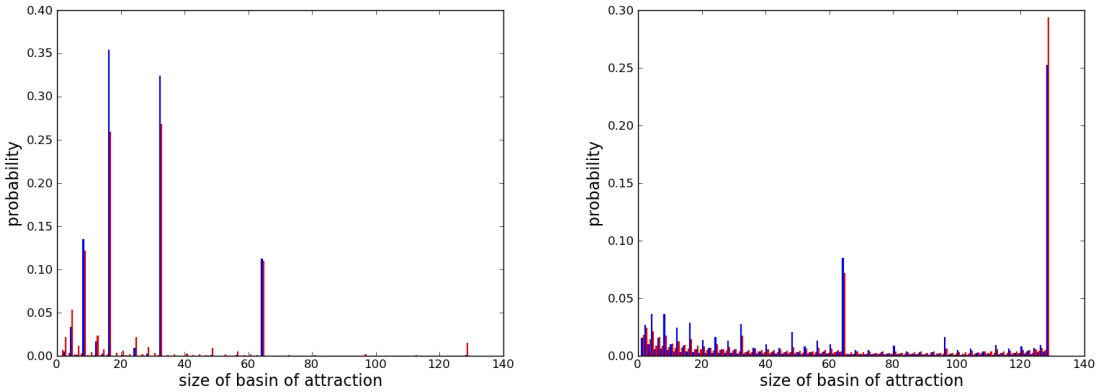

Figure 3.11: Distribution of sizes of attraction basins for $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel) (blue and red bars represent evolved and randomized sets, respectively).

than those of the randomized distribution below the basin size of 40. Both the evolved and randomized distributions have probabilities smaller than 0.05 for the basin sizes larger than 64. In the populations with $p_0 = 0.5$, the distributions of sizes of basins of attractions are more homogenous compared to those in the populations with $p_0 = 0.2$. In the populations with $p_0 = 0.5$, the probability of having a basin size of 128 is the highest (0.3) for the randomized graphs and the probability of having a basin with the same size is about half as much (0.15) for the evolved graphs. These results are consistent with the distributions of number of attractors shown on the right panel of Figure 3.9, where single attractors correspond to basin of attractions with size 128.

3.4 Multistationarity of Random Graphs with Different Connectivities

From Figure 3.6 and Figure 3.10, we see that, for the randomized graphs, the mean attractor length increases with increasing mean degrees, whereas the mean number of attractors decreases with increasing mean degrees. However, the mean degrees of the randomized sets investigated here are within a limited range. For a more thorough analysis, we have generated a set of 1000 random graphs with the same Boolean masks and $p_0 = 0$ initially, and have increased the connectivity of the graphs by adding an edge to each of them, at every time step, until the graphs are fully connected.

The dependences of the attractor lengths and the number of attractors on connectivity are shown in Figure 3.12. Here, connectivity is associated with the second smallest eigenvalue of the laplacian. This eigenvalue is zero for disconnected graphs and it is equal to the number of nodes for fully connected graphs. As seen from the

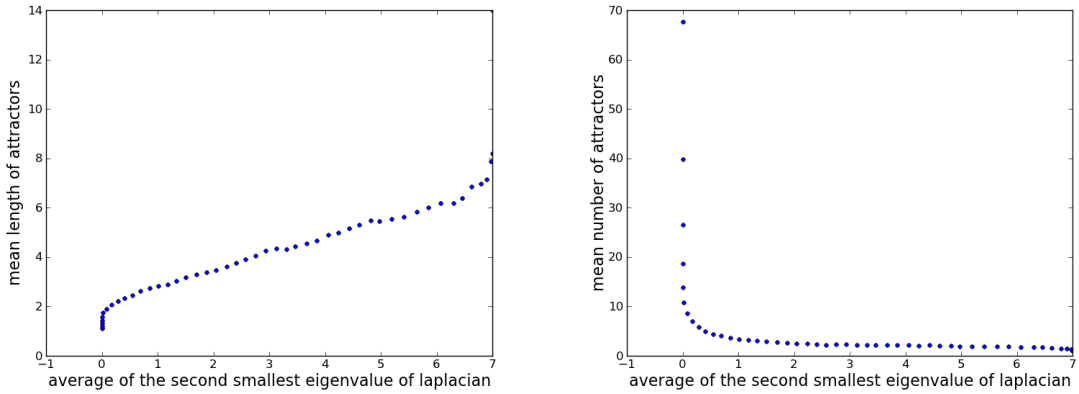

Figure 3.12: The mean length of attractors (left panel) and the mean number of attractors (right panel) versus the average of the second smallest eigenvalue of the laplacian matrices.

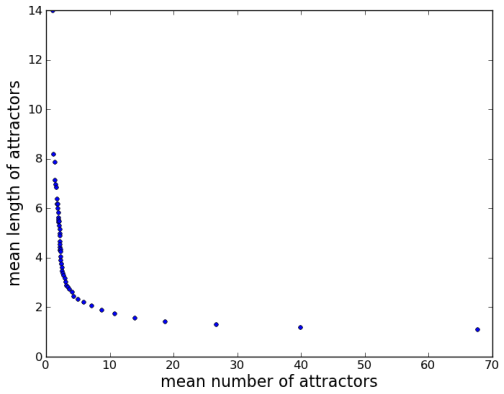

Figure 3.13: The mean lenght of attractors versus the mean number of attractors.

figures, the mean length of attractors increases with increasing connectivity, whereas the mean number of attractors decreases with increasing connectivity. These results are consistent with the findings in Sections 3.2 and 3.3. They also indicate an inverse relation between the mean length of attractors and the mean number of attractors. As seen from Figure 3.13, the mean length of attractors decreases with increasing mean number of attractors.

4. TOPOLOGICAL FEATURES OF EVOLVED NETWORKS

In order to understand the relationship between the graph structure and the dynamical behavior, we examine the degree distributions and motif profiles of the evolved networks as well as their randomized counterparts, in this section.

4.1 Degree Distributions

As mentioned in section 3.2, there is no direct correlation between the mean degrees and the attractor lengths of evolved networks. This suggests that there must be other features which lead to short attractor lengths, beside low degrees. We first consider the degree distributions of graphs. The evolved graphs were randomized preserving their number of edges. Therefore, the resulting randomized graphs have the same mean degrees as the evolved graphs. However, the degree distributions of evolved and randomized populations differ substantially, in some cases. The degree distributions of evolved and randomized graphs are presented in Fig. 4.1, together with a Poisson distribution generated with the same mean value as the mean degree of the evolved and randomized graphs. The mean degrees of the populations with $p_0 = 0.2$ (left panel) and the populations with $p_0 = 0.5$ (right panel) are 2.69 and 6.15, respectively. For populations with $p_0 = 0.2$, the degree distributions of evolved and randomized graphs are similar and they fit the Poisson distribution except for their peak values. For populations with $p_0 = 0.5$, where the mean degree is larger, distributions of evolved and randomized graphs differ. The degree distribution of randomized graphs has a smaller variance (4.06) compared to the Poisson distribution, but it has the same overall shape and it has a peak value around the mean degree 6.15. The degree distribution of evolved graphs has a higher variance (5.00) than the variance of the degree distribution of randomized graphs and has a different shape. The degree distribution of evolved graphs also has a peak value around the mean degree, but it has another peak at a higher degree, degree 8 (Fig. 4.1 right panel).

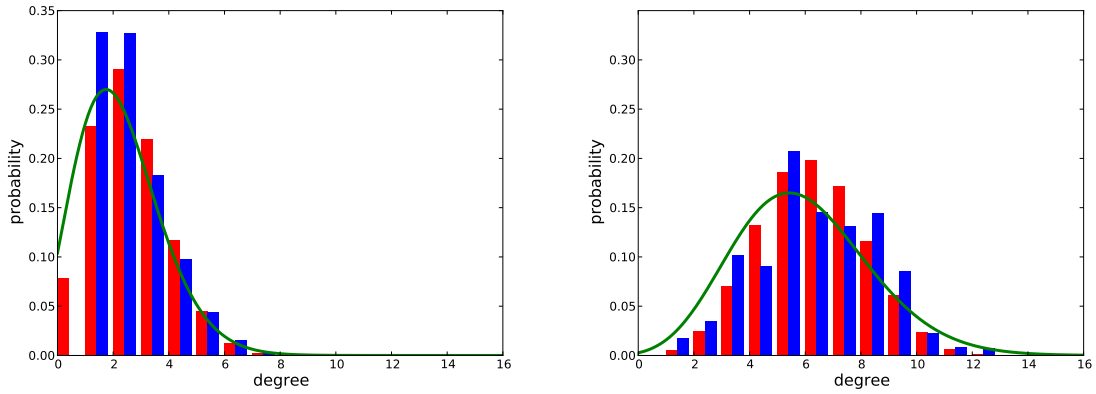

Figure 4.1: Degree distributions of populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel), the green line represents the Poisson distribution with the same mean (see text). Blue and red bars represent evolved and randomized sets, respectively.

Table 4.1: Properties of set A and set B.

Properties	Set A	Set B
p	0.5	0.5
mean score	1.40	1.53
mean degree	2.86	2.85
information content of the degree distribution	1.55	1.00
information content of the in- degree distribution	1.48	1.28
information content of out degree distribution	1.38	1.15
degree of the in-hub	5	5
degree of the out-hub	5	6

Since the shapes and mean values of the degree distributions differ between the sets, when we take an average over the 16 sets, the resulting distribution does not differ significantly from the degree distribution of randomized graphs. When we examine the populations separately, the differences between the degree distributions of the evolved and randomized sets are seen more clearly. Properties of two different sets are shown in Table 4.1. Here, set A and set B have similar mean degrees and their scores are relatively close to each other. However the information contents of the degree distributions of the two sets are substantially different. The degree distributions of evolved and randomized graphs are relatively similar for set A. However, the difference is much more significant for set B (Fig. 4.2).

Although the shapes and mean values of the degree distributions of the evolved sets are different from each other, uneven distribution of the degrees around the mean degree is observed in most of the sets. This might be due to the different distributions of the in-degrees and out-degrees. The out- and in-degree distributions are shown in

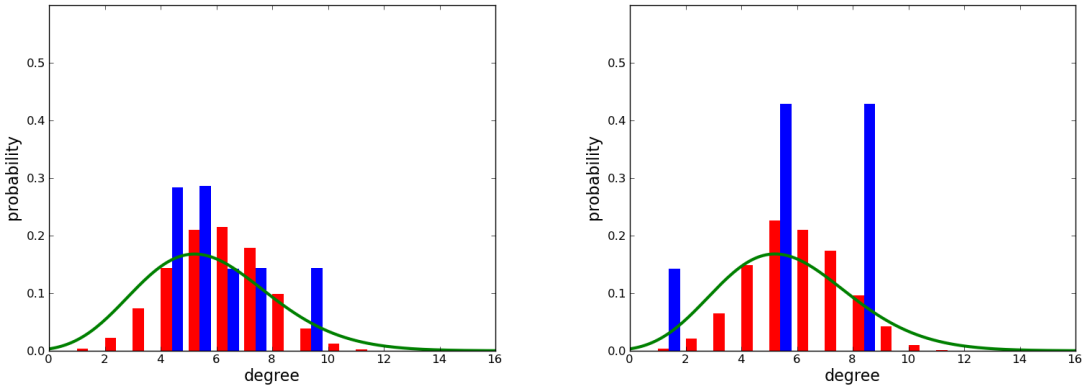

Figure 4.2: Degree distributions of set A (left panel) and set B (right panel), green line represents Poisson distribution with the same mean. Blue and red bars represent evolved and randomized sets, respectively.

Fig. 4.3 and Fig. 4.4, respectively. The figures show that for randomized sets the in-degree and out-degree distributions are exactly the same. For populations with $p_0 = 0.2$ (left panels), the in- and out-degree distributions of evolved graphs are also similar. However, the differences between the in- and out-degree distributions of populations with $p_0 = 0.5$ (right panels) shows that in- and out-degrees are not equally distributed. Both of the distributions have higher probabilities at high degrees, compared to randomized graphs. This indicates the presence of hubs that either regulate many other nodes in the network or are regulated by them.

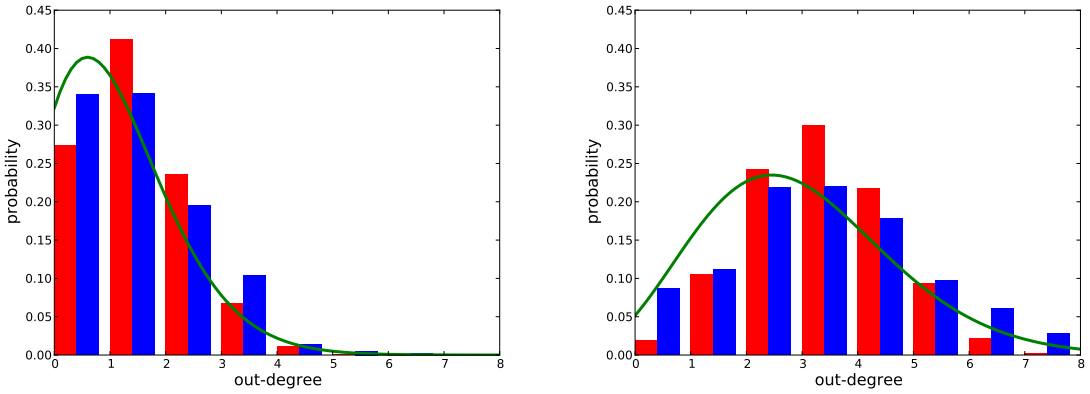

Figure 4.3: Out-degree distributions of populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel), green line represents Poisson distribution with the same mean. Blue and red bars represent evolved and randomized sets, respectively.

The in- and out-degree distributions of set A and set B are shown in Fig. 4.5 and Fig. 4.6, respectively. For both of the sets, the in- and out-degree distributions are different from each other, although they have approximately the same mean value. The out-degree distribution of set A is similar to those of its randomized counterpart.

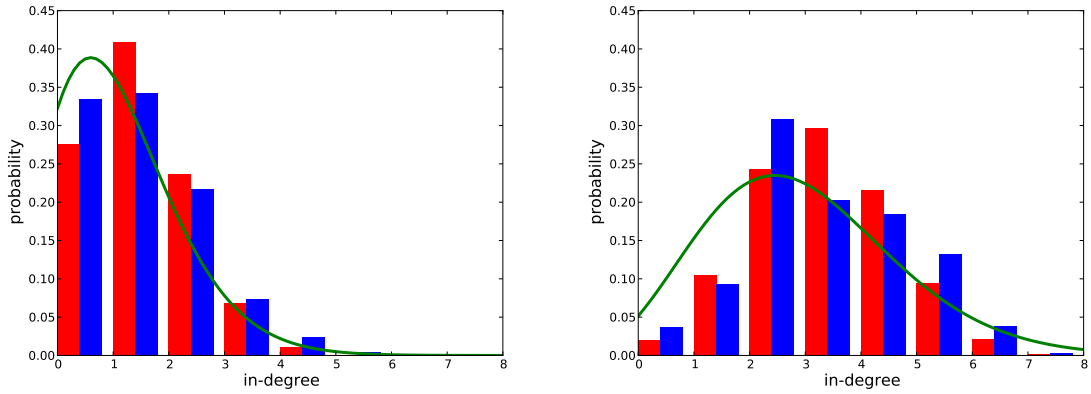

Figure 4.4: In-degree distributions of populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel), green line represents Poisson distribution with the same mean. Blue and red bars represent evolved and randomized sets, respectively.

Both the in- and out-degree distributions of the evolved set the frequencies of the nodes with degree 5 are significantly higher than those of in the randomized set. However, the nodes with out-degree 5 have a higher probability than the nodes with in-degree 5. Both the in-degree and out-degree distributions of set B are very different from those of its randomized counterpart. The in-hubs of set B has degree 5 and the out-hubs has degree 6.

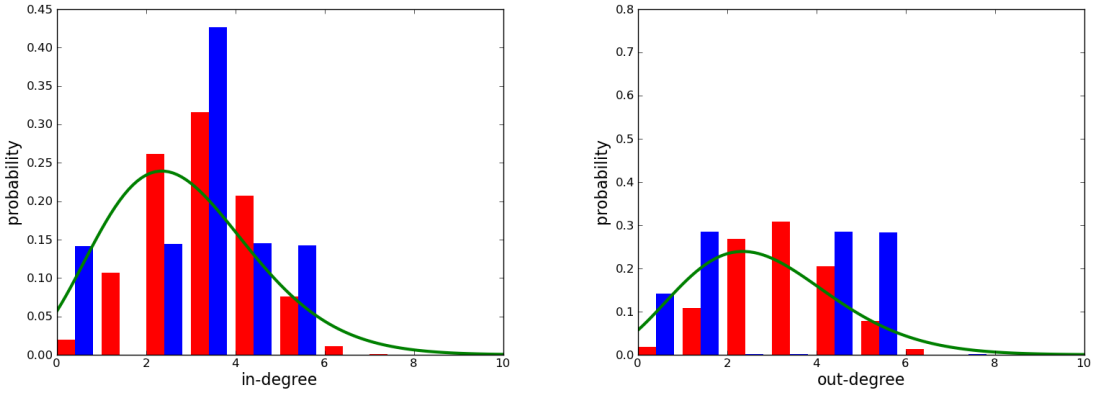

Figure 4.5: In-degree (left panel) and out-degree distributions of set A, green line represents Poisson distribution with the same mean. Blue and red bars represent evolved and randomized sets, respectively.

The outliers shown in Figures 4.5 and 4.6 are not seen clearly in Figure 4.1, since the in- and out-hubs of the networks do not corespond to the same nodes. The hubs are observed better when the graphs have undirected edges. Undirected versions of the graphs are generated by symmetrizing their adjacency matrices. As seen in Fig. 4.7, evolved networks in the populations with $p_0 = 0.2$ and $p_0 = 0.5$ have higher probabilities at high degrees compared to randomized networks. The mean degrees

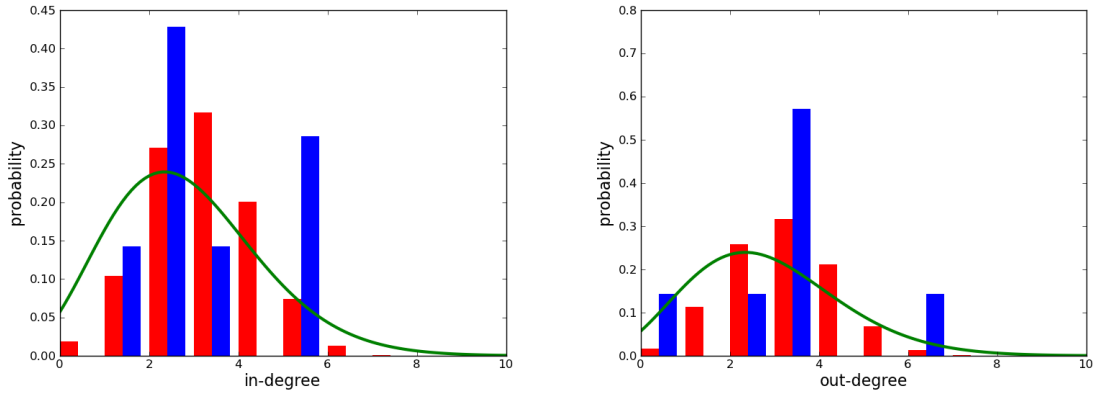

Figure 4.6: In-degree (left panel) and out-degree distributions of set B, green line represents Poisson distribution with the same mean. Blue and red bars represent evolved and randomized sets, respectively.

of randomized graphs are smaller than those of evolved graphs both in populations with $p_0 = 0.2$ and $p_0 = 0.5$. In the populations with $p_0 = 0.2$, the mean degree of the symmetrized randomized graphs is 2.36, whereas the mean degree of the symmetrized evolved graphs is 2.42. The difference between the mean degrees of evolved and randomized graphs is larger for the populations with $p_0 = 0.5$. The mean degree of the symmetrized randomized graphs is 4.55, whereas the mean degree of the symmetrized evolved graphs is 4.77. Since the mean degrees of the evolved and randomized graphs in the original populations are the same, the differences in the mean degrees of symmetrized graphs are due to the mutual interactions. When two edges with opposite directions coincide, they are combined in a single edge during symmetrization. Therefore, the differences between the symmetrized versions of evolved and randomized graphs show that evolved graphs have fewer mutual interactions.

The difference in the degree distributions of separate sets suggests a quantitative approach is needed to evaluate the difference between distributions of evolved and randomized sets. Information content is used for this purpose. Information content of the degree distributions of evolved and randomized populations are shown in Fig. 4.8. For populations with $p_0 = 0.2$, randomization does not make a significant change in the information content of degree distributions. On the other hand, for the populations with $p_0 = 0.5$, information content of degree distributions differ substantially between evolved and randomized graphs. Information content is significantly lower for

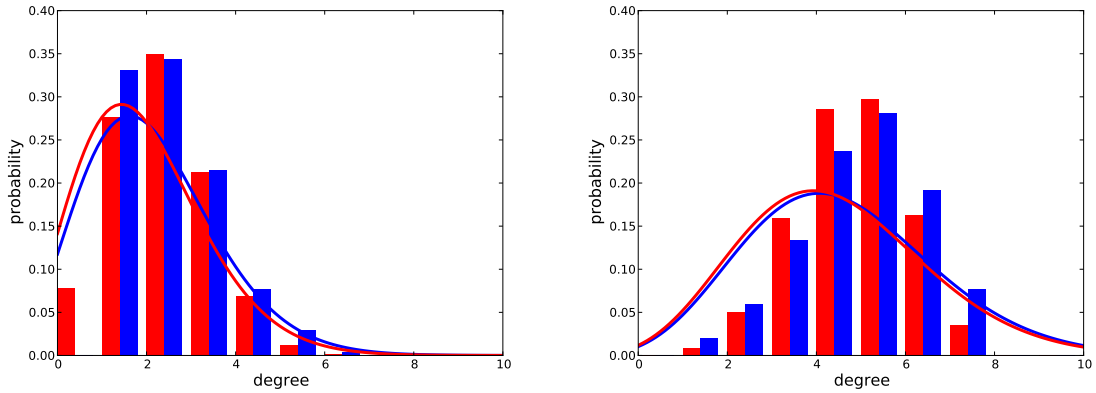

Figure 4.7: Degree distributions of undirected network populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel), green line represents Poisson distribution with the same mean. Blue and red bars represent evolved and randomized sets, respectively.

evolved populations. As the information content of evolved populations increases, the difference between evolved and randomized sets becomes less significant.

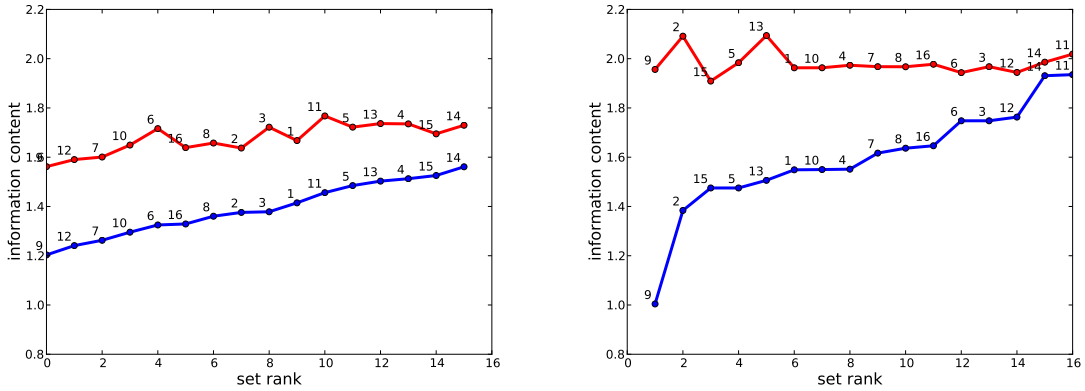

Figure 4.8: Information content of the degree distributions for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

4.2 Motif Frequencies

The topological features of different networks can be examined in more detail by looking for common motif structures. Motifs are subgraphs of networks, consisting of three or more connected nodes [3]. Our model graphs are small, therefore, we considered only 3-motifs. Motif frequencies of evolved and randomized populations are shown in Fig. 4.11. Due to the large number of motifs, it is hard to capture the effect of a particular motif structure on the dynamics of networks. Therefore, self interactions were eliminated to reduce the number of motifs. As a result, there are 13 motifs left to consider as shown in Fig. 4.12. The frequencies of the 13 motifs are shown in

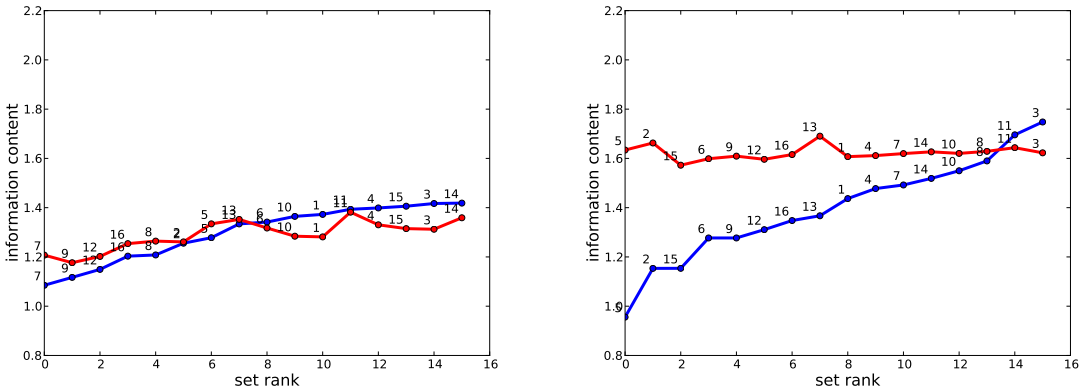

Figure 4.9: Information content of the in-degree distributions for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

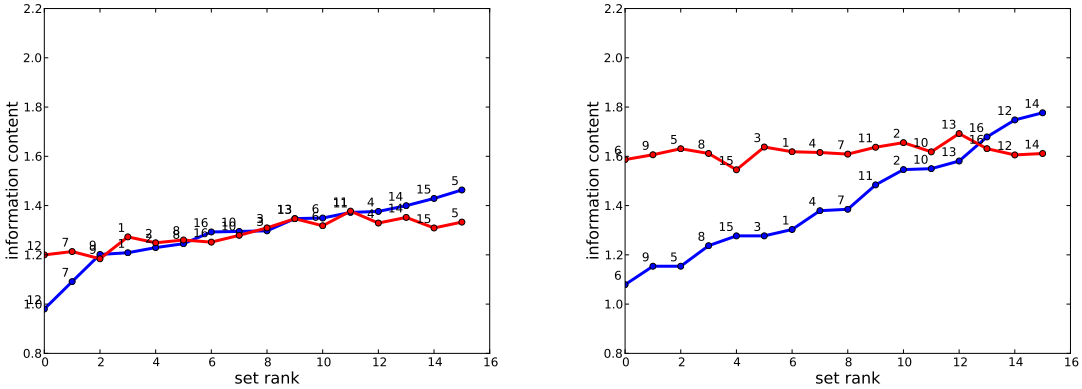

Figure 4.10: Information content of the out-degree distributions for populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

Fig. 4.13. Since low edge density do not permit formation of loops and bidirectional edges, it has a limitation on motif diversity. Therefore, populations with lower number of edges exhibit simpler, mostly loopless motifs (Fig. 4.13 left panel). Apart from the first two motifs, only the frequency of motif 7 in evolved populations is higher than the frequency in randomized populations. This difference is much more emphasized in populations on the right panel, where edge density is higher. Frequencies of motifs 9 and 10 in evolved populations are also higher compared to randomized populations. These three motifs involve feed-forward loops. These findings are consistent with those found by Alon et al. [3,4], who have emphasized the contribution of feed-forward loops to stability. As shown in Fig. 4.13, the frequency of motif 8 is significantly lower in evolved populations than those of in randomized populations. Motif 8 is a feedback loop. As Thomas and Kaufman showed, feedback loops have negative effect on stability yielding periodic cycles depending on the nature of interactions between

the nodes, a feedback loop consisting of odd number of repressive interactions yields oscillations between different states [7].

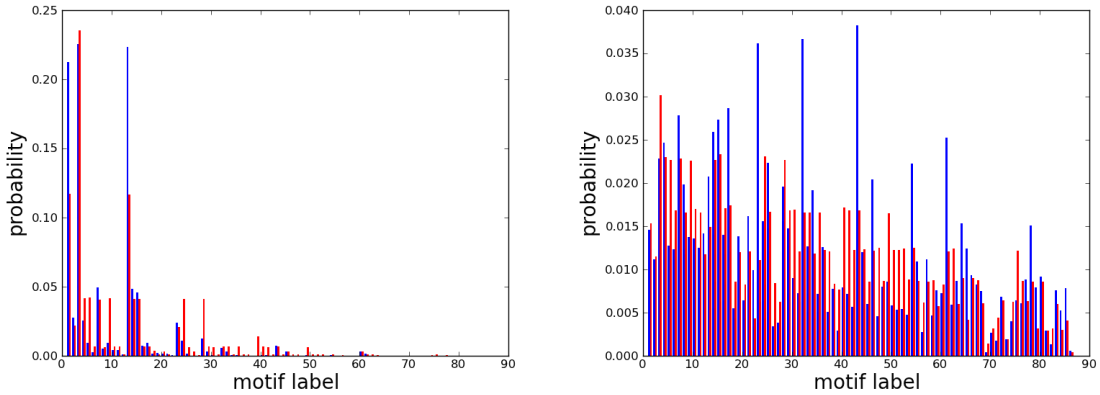

Figure 4.11: The motif frequencies averaged over 16 populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

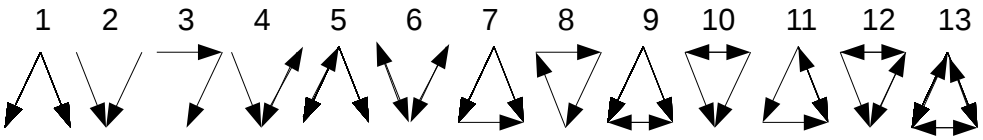

Figure 4.12: Three-motifs without self-interactions.

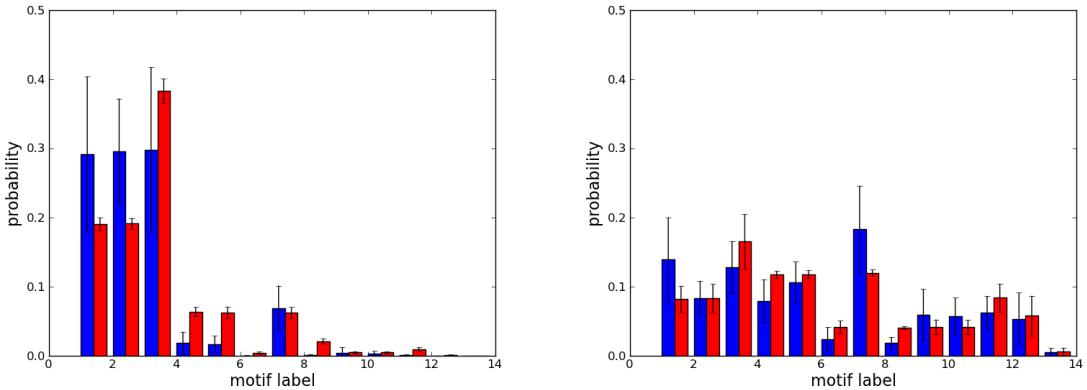

Figure 4.13: Frequencies of 13 motifs averaged over 16 populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

Information content of motif frequencies (Fig. 4.14) also shows populations with low number of edges (left panel) exhibit limited number motifs, since they have a lower information content, and randomization yields to similar motif structures. Whereas information contents of evolved populations with higher edge density are systemitically lower than those of randomized populations.

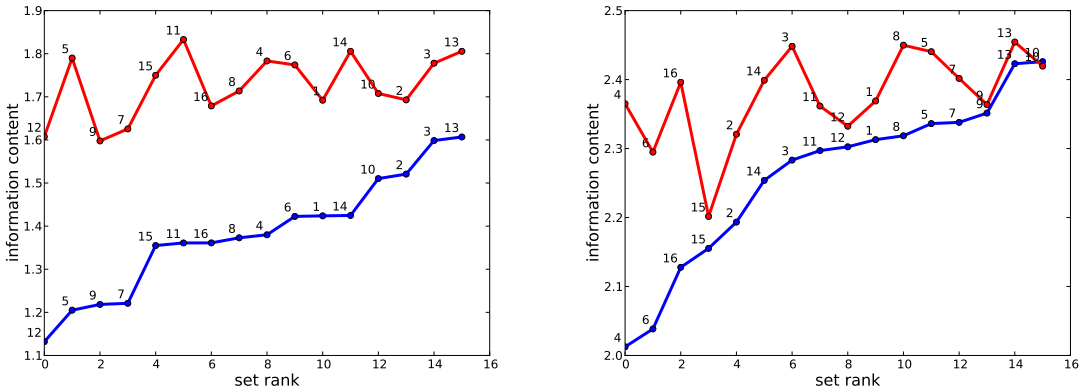

Figure 4.14: Information contents based on motif frequency distributions of populations with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel).

4.3 k-core Decomposition of Empirical Networks

The biological networks considered in this study were transcriptional gene regulatory networks (TGRN) of *E. coli, B. subtilis* and *S. cerevisiae* [17]. k-core decompotion method [18, 19] is used to scale-down the size of these networks, since they were too large compared to our model graphs. The sizes and total number of edges of empirical networks and their core graphs are given in Table 4.2. The plots of the empirical networks and their core-graphs are shown in Figures 4.15 and 4.16; the plots are obtained using Large Networks Visualization tool (LaNet-vi) [26].

Table 4.2: The sizes and total number of edges of empirical networks and their core graphs. Here, N is the number of nodes and E is the number of edges.

	Original Network		Core Graph	
	N	E	N	E
E. coli	1607	4141	14	54
B. subtilis	922	1397	36	116
S. cerevisiae	4441	12900	125	1129

4.4 Significance Profiles

We computed z-scores and Significance Profiles (SP) based on motif frequencies for the comparison of evolved and empirical graphs. The z-score of motif μ is defined as [3],

$$z_\mu = \frac{\langle N_{\text{real}}(\mu)\rangle - \langle N_{\text{rand}}(\mu)\rangle}{\sigma[N_{\text{rand}}(\mu)]}. \qquad \textbf{(4.1)}$$

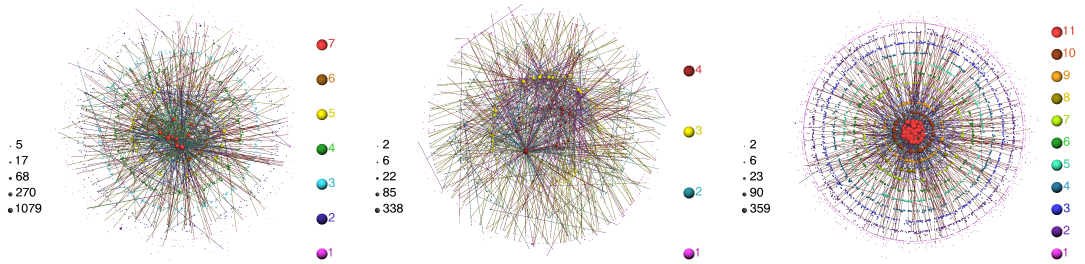

Figure 4.15: Transcriptional gene regulatory networks of *E. coli* (left panel), *B. subtilis* (middle panel) and *S. cerevisiae* (right panel). The columns on the left side side of the figures show how the sizes of the nodes are scaled with their degrees, and the columns on the right of the figures show k-core numbers of the nodes and their corresponding colors.

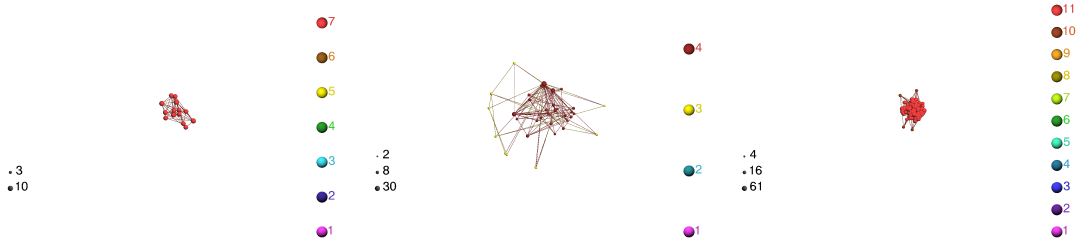

Figure 4.16: Core graphs of *E. coli* (left panel), *B. subtilis* (middle panel) and *S. cerevisiae* (right panel) extracted using k-core decomposition method. The columns on the left side side of the figures show how the sizes of the nodes are scaled with their degrees, and the columns on the right of the figures show k-core numbers of the nodes and their corresponding colors.

Here, $\langle N_{\text{real}}(\mu)\rangle$ and $\langle N_{\text{rand}}(\mu)\rangle$ are motif frequencies (averaged over 10^3 graphs) of evolved and randomized graphs, respectively; $\sigma[N_{\text{rand}}(\mu)]$ is the standard deviation (for 10^3 randomized graphs) of $N_{\text{rand}}(\mu)$. The significance profile of motif μ is defined as [4],

$$S_\mu = z_\mu \left(\sum_\mu z_\mu^2 \right)^{-1/2} . \tag{4.2}$$

The SPs of evolved and empirical networks are compared in Fig. 4.17. Here, we observe that the SPs of evolved graphs are similar to each other, especially at certain motifs such as motifs 1, 2, 7, 9 and 10. Notice that motifs 1 and 2 are loopless motifs and motifs 7,9 and 10 involve feed-forward loops. The SPs of the cores graphs of the empirical networks (*E. coli, B. subtilis* and *S. cerevisiae*) also exhibit the same properties. The motif frequencies of the complete gene regulatory networks, however, differ from those of the evolved graphs.

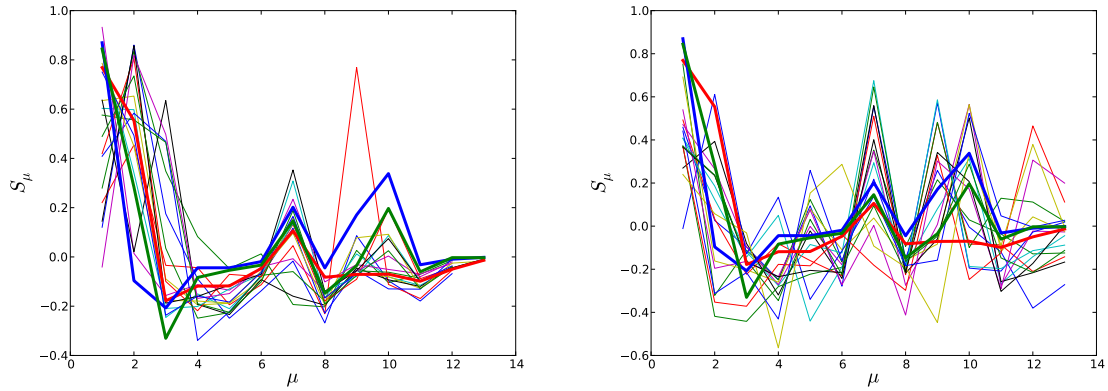

Figure 4.17: Significance profiles for $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel) (red, blue and green bold lines represent *E. coli, B. subtilis* and *S. cerevisiae*, respectively).

We defined a scalar product to make a quantitative comparison between the siginificance profiles of different populations, as well as the TGRNs,

$$\mathscr{O}(\mathbf{S}^{(\alpha)},\mathbf{S}^{(\beta)}) = \sum_{\mu} S_{\mu}^{(\alpha)} S_{\mu}^{(\beta)} \quad , \tag{4.3}$$

where α and β denote set labels. For a given set, the average overlap with all the other sets is,

$$\bar{\mathscr{O}} = \frac{1}{15} \sum_{\alpha \neq \beta} \mathscr{O}(\mathbf{S}^{(\alpha)},\mathbf{S}^{(\beta)}) \quad . \tag{4.4}$$

The results for the evolved sets are shown in Fig.4.18. The average overlaps of the individual sets, as well as their overlaps with biological TGRNs are given in Appendix A. For comparison, 16 sets of 10^3 graphs are randomly generated with edge density $p = 0$. Their overlap matrice is shown in Fig. C.2. The overlaps between biological networks are shown in Table 4.3.

As seen from Fig. 4.18 the overlaps between the sets with $p_0 = 0.2$ (left panel) are higher than those of the sets with $p_0 = 0.5$ (right panel), in general. For the populations with $p_0 = 0.2$ the average overlap changes between 0.46 and 0.80 (see Table A.1), whereas for the populations with $p_0 = 0.5$ it changes between 0.14 and 0.61 (see Table A.2). Although some of the sets may have low average overlap, for the populations with $p_0 = 0.2$, two clusters of sets (one with five sets and the other with three sets) with large mutual overlaps can be seen from the left panel of Fig. 4.18. For the populations with $p_0 = 0.5$, except for three sets all the other sets have average overlaps higher than 0.4 (Table A.2), whereas the mutual overlaps of random sets with $p = 0.5$ are around zero, in general.

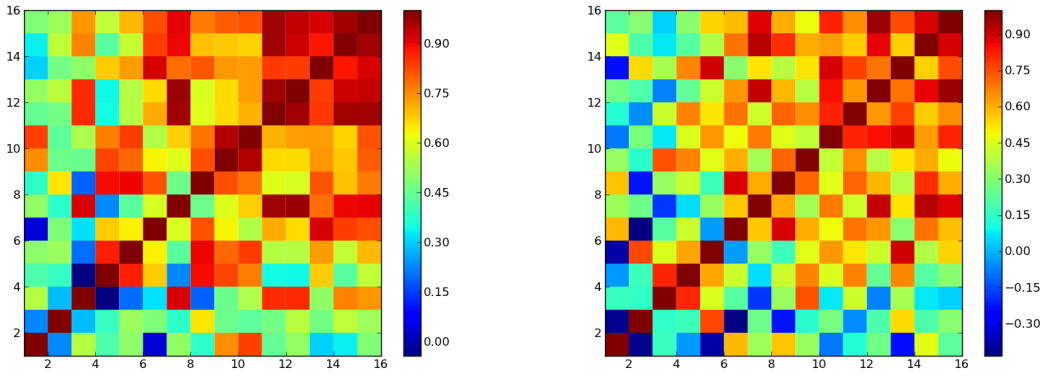

Figure 4.18: The overlap between the significance profiles of evolved sets with $p_0 = 0.2$ (left panel) and $p_0 = 0.5$ (right panel). The sets are sorted with respect to their average overlaps, so blocks of networks having the largest overlaps are displayed along the diagonal. The rank of the sets are shown on the horizontal and vertical axis. The side bar contains the color code.

Table 4.3: The overlap between the SPs, $\mathscr{O}(\mathbf{S},\mathbf{S}')$, of biological TGRNs.

	E. coli	*S. cerevisiae*	*B. subtilis*
E. coli	1.00	0.91	0.67
S. cerevisiae	0.91	1.00	0.88
B. subtilis	0.67	0.88	1.00

In Table 4.3, it is seen that the minimum overlap between the biological networks is 0.67, whereas the maximum overlap is 0.91. According to our measure of similarity, the core graphs of *E. coli* and *S. cerevisiae* have the highest resemblance, whereas those of *E. coli* and *B. subtilis* are the least similar. Populations with $p_0 = 0.2$ and $p_0 = 0.5$ perform differently, in terms of their overlaps with the biological networks. The overlap the populations with $p_0 = 0.2$ and *E. coli* changes between 0.40 and 0.99, whereas their overlaps with *S. cerevisiae* can be as low as 0.08, and with *B. subtilis* -0.20 (Table A.1). On the contrary, the populations with $p_0 = 0.5$ resemble *B. subtilis* and *S. cerevisiae* more (Table A.2). Their overlap with *B. subtilis* changes between 0.04 and 0.86 with a mean value 0.57; and their overlap with *S. cerevisiae* changes between 0.22 and 0.78 with a mean value 0.55, whereas their overlap with *E. coli* changes between 0.18 and 0.62 with a mean value 0.45.

5. NEUTRAL NETWORKS

In the previous chapters we have discussed the topological and dynamical features of different sets of graphs evolved according to the same fitness function. In this chapter we will examine the evolution of the populations of Boolean graphs over genotype and phenotype spaces. The adjacency matrices of the Boolean graphs form the genotype space, whereas the attractors of the Boolean graphs form the phenotype space. The differences between the properties of independently evolved sets show that these sets of graphs are spanning different areas both in the genotype and phenotype spaces. In fact, it is shown that two networks having the same phenotype can have genotypes as different from each other as if they were chosen at random [20]. This property allows one to explore a broad area in genotype space without experiencing a substantial change in the fitness. This is called mutational robustness, an essential future for the evolution of living organisms. Therefore, it is useful to look at the structures of these genotype and phenotype spaces.

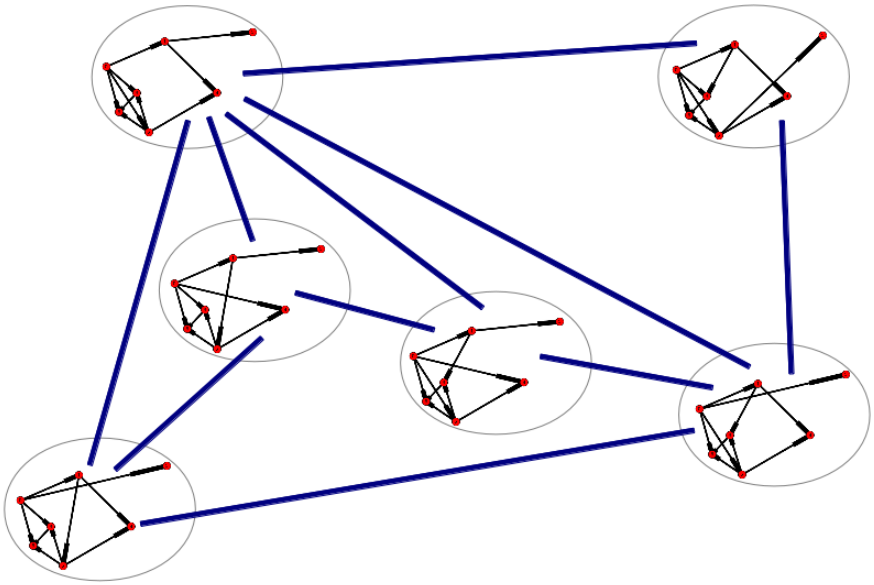

Figure 5.1: An illustration of a network in genotype space.

5.1 Neutral Networks Formed in Genotype Space

5.1.1 Distances between Boolean graphs in genotype space

A set of graphs having the same phenotype, or equal fitness, can be considered as a network itself, where each vertex corresponds to a Boolean graph and it is connected to another if it is one mutation away from it [21]. These networks are called neutral networks. In our simulations, a mutation involves rewiring the edges between two pairs of nodes. This operation changes 4 elements in the adjacency matrix. If we define mutational distance d_{IJ} between two networks with adjacency matrices A^I and A^J as

$$d_{IJ} \equiv \frac{\sum_{k,l}^{N} |A^I_{kl} - A^J_{kl}|}{4}, \tag{5.1}$$

any two neighbours in a neutral network are seperated by a mutational distance 1. The adjacency matrix of the neutral network in genotype space is

$$B_{ij} \equiv \begin{cases} 1, & \text{if } d_{IJ} \leq 1 \\ 0, & \text{otherwise} \end{cases}. \tag{5.2}$$

An illustration of a network in genotype space is shown in Fig. 5.1.

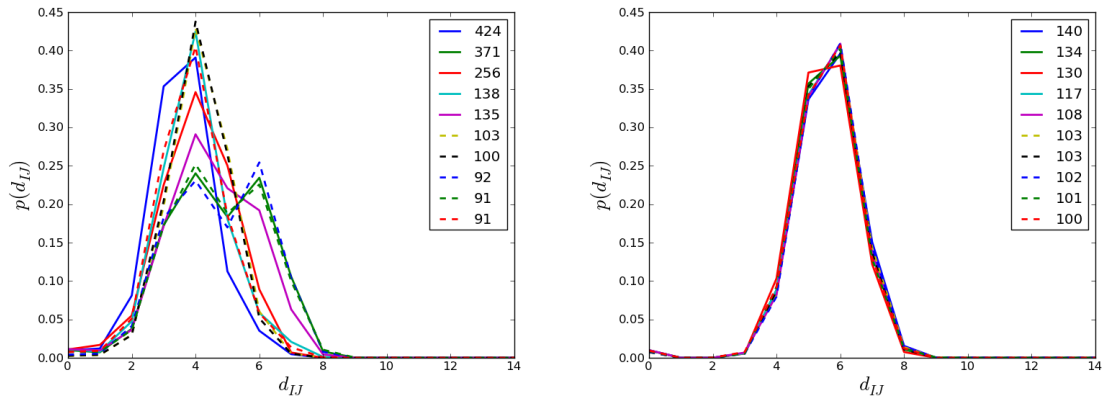

Figure 5.2: Distribution of pairwise distances between networks having a particular attractor in set 1 with $p = 0.5$ for the evolved (left panel) and randomized (right panel) counterparts. Different curvatures correspond to network sets with different attractors. Only 10 largest sets are shown. Sizes of the network sets sharing the same attractors are given in the legend. Note that a network can have more than one attractor, therefore can be found in more than one set.

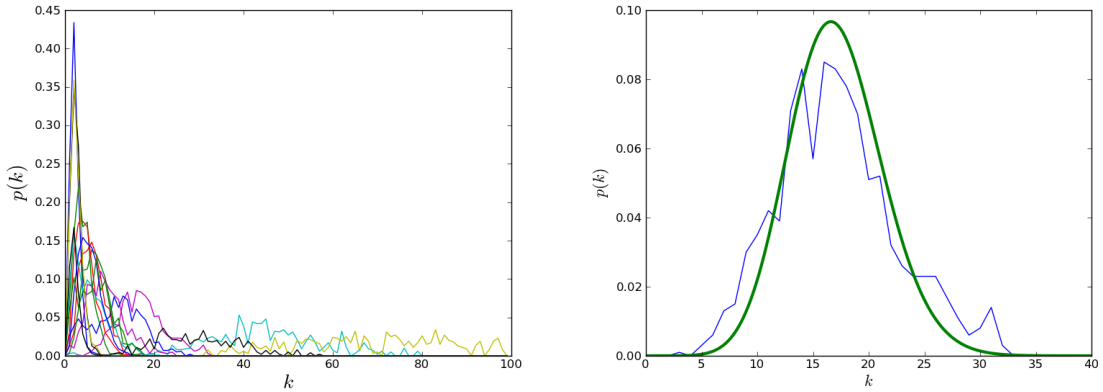

Figure 5.3: Degree distributions of neutral networks formed in genotype space with $p_0 = 0.5$ (left panel) and the degree distribution of the neutral network of set 5 with $p_0 = 0.5$ (blue line), and Poisson distribution with the same mean value (green line) (right panel).

5.1.2 Boolean graphs sharing the same attractors

The distributions of the distances between pairs of Boolean graphs having a particular attractor in set 1 with $p_0 = 0.5$ are shown in Fig. 5.2. Comparison between the sizes of the clusters of graphs sharing the same attractors in evolved and randomized populations reveals that most of the Boolean graphs in the evolved population are found in the 3 largest clusters, whereas Boolean graphs in the randomized population are more homogenously distributed among the 10 largest clusters. However the sizes of the clusters do not affect the distribution of the distances between graphs.

The minimum distance between different graphs is higher than 2 for the randomized populations. This means that randomized graphs do not form neutral networks according to our definition.

The distance takes its maximum value when all the elements of the adjacency matrices of the two networks are different from each other, i.e. $|A^I_{kl} - A^J_{kl}| = 1$ for all k and l. So the maximum possible distance between two Boolean graphs is

$$d_{\max} \equiv \frac{\sum_{k,l}^{N} 1}{4} = 12.25. \qquad \textbf{(5.3)}$$

For the randomized population, the mean distances between the graphs in all the neutral networks is around 6, which is about half of the maximum possible distance 12.25, also the distribution of the distances is symmetric around the mean value and same for all the network clusters. For the evolved population, distributions of distances

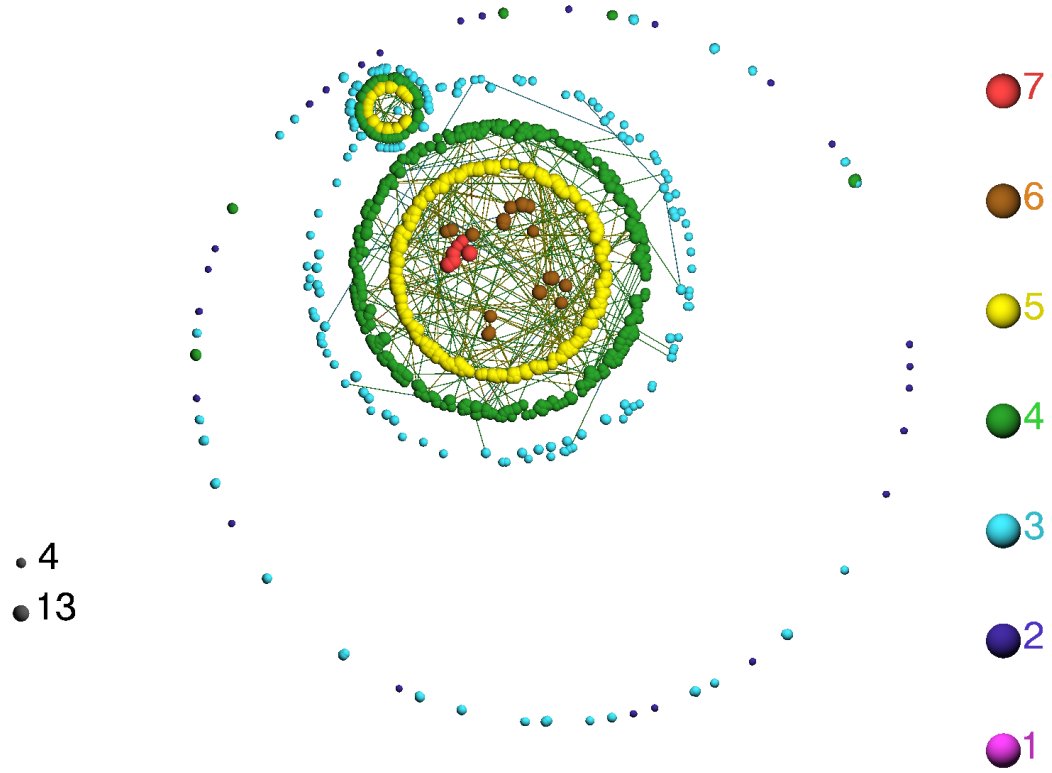

Figure 5.4: Neutral network formed by an evolved population in genotype space.

differ between the clusters, however, most of them have the highest probability around $d_{IJ} = 4$.

5.1.3 Topologies of neutral networks

Since randomized populations do not form neutral networks in genotype space, we will only consider the topologies of the neutral networks formed by evolved populations.

The degree distributions of neutral networks formed in genotype space of the populations with $p_0 = 0.5$ are shown on the left panel of Fig. 5.3. The neutral networks have different mean degrees, however, their degree distributions fit Poisson distributions. This is shown for one of the sets on the right panel of Fig. 5.3. This means that the degree distributions of the neutral networks are similar to those of the random networks. However, when we took a neutral network and compared it with a random network with the same size and number of edges, we found that their structures are not the same (Fig. 5.4 and Fig. 5.5).

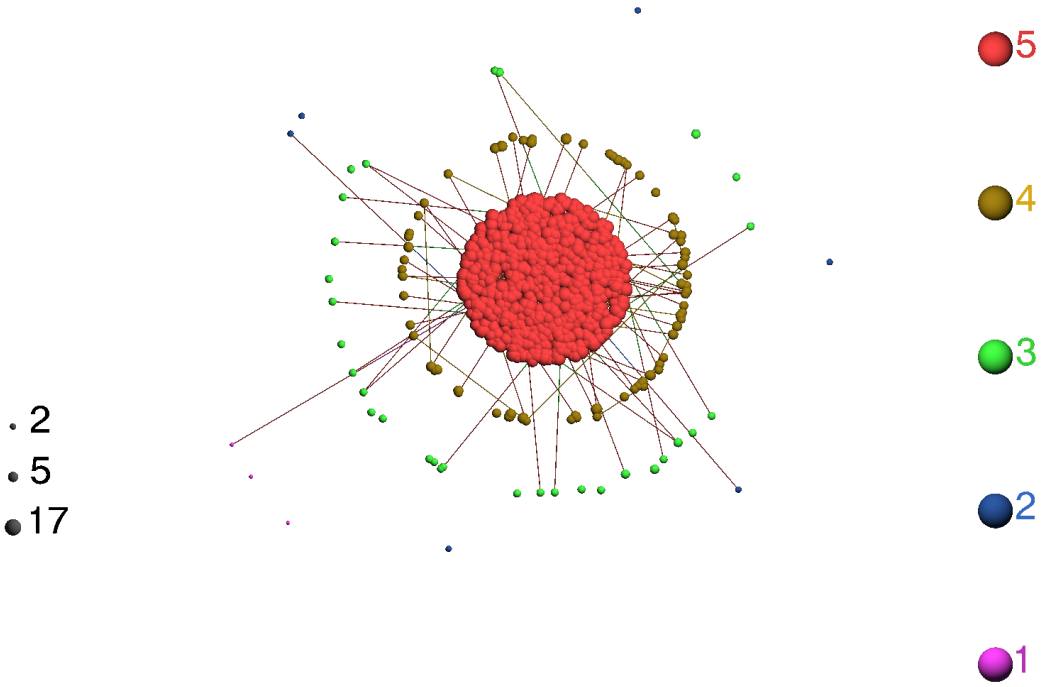

Figure 5.5: A random network with the same size and number of edges as the neutral network in Fig. 5.4.

5.2 Metanetworks Formed in Phenotype Space

A metanetwork is formed by networks, each node of a metanetwork corresponds to network. The neutral networks discussed in Section 5.1 are metanetworks formed in genotype space. In this section, we will define metanetworks formed in phenotype space and compare them to the neutral networks in genotype space.

Different Boolean graphs may share one or more attractors. These graphs form a weighted metanetwork in phenotype space based on the number of attractors they share and the sizes of the basins of attraction of these attractors. We define weight of the edge between the Boolean graph I and J as

$$W_{IJ} \equiv \frac{\Sigma_{\alpha}^{2^N} C_{\alpha}^{I} C_{\alpha}^{J}}{2^{2N}}, \tag{5.4}$$

where C_{α}^{I} is the size of αth basin of attraction of Ith Boolean graph. An illustration of a meta-network formed in phenotype space is shown in Fig. 5.6.

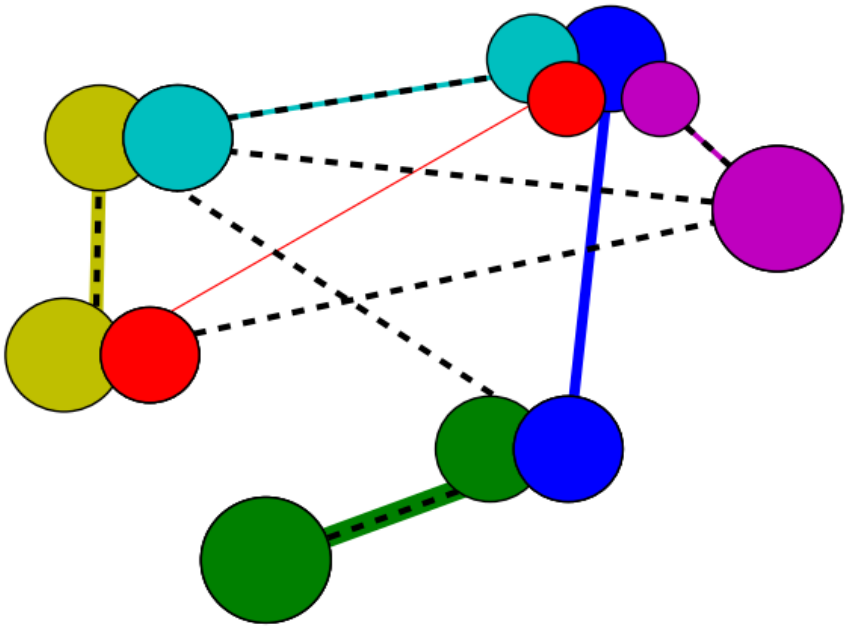

Figure 5.6: An illustration of a meta-network in phenotype space. Circles represent attraction basins of different attractors that a Boolean graph has and their area are proportional to the sizes of the basins of attraction. The colored lines represent edges in phenotype space, whereas black dashed lines represent edges in genotype space. The thickness of the colored edges are proportional to their weights (Eq. 5.4). Different colors correspond to different attractors.

5.2.1 Robustness of the evolved graphs

The frequencies of the attractors the Boolean graphs have are given in Fig. 5.7. The attractors shared by the Boolean graphs in the evolved sets have a much narrower frequency distribution than those of the attractors in the randomized sets. Moreover, there are much fewer attractors shared by Boolean graphs in the evolved sets, than there are in the randomized sets. This causes stronger bonds to form between the Boolean graphs. For the gene regulatory networks, this means that the evolved metanetworks can span a large area in genotype space without significant changes occuring in their phenotypes, which is a sign of robustness.

The robustness of the metanetworks are investigated by eliminating the edges with weights under a threshold value θ. The dependence of the sizes of the giant components $|GC_\phi|$ on the threshold θ for the evolved and randomized metanetworks are shown in Fig. 5.8. Here, we see that the giant components of the randomized metanetworks shrink much faster than those of the evolved metanetworks.

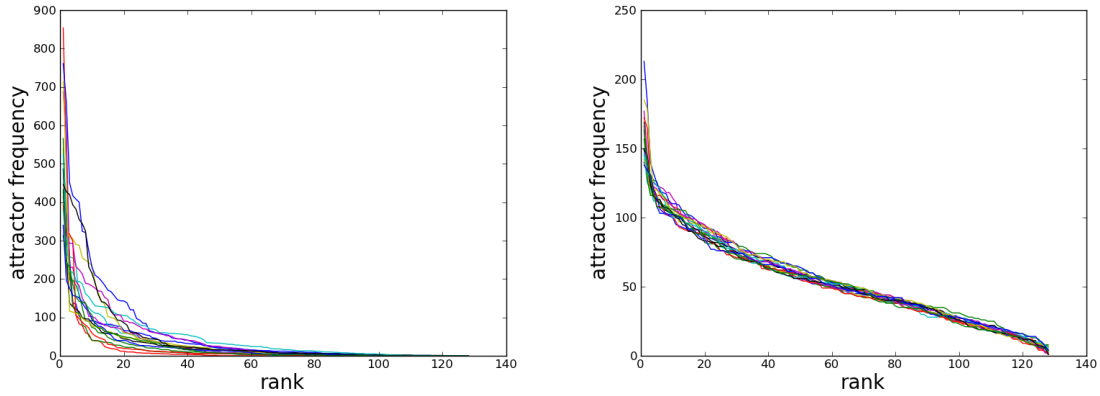

Figure 5.7: Frequencies of attractors in the evolved populations with $p_0 = 0.5$ (left panel) and in their randomized counterparts (right panel).

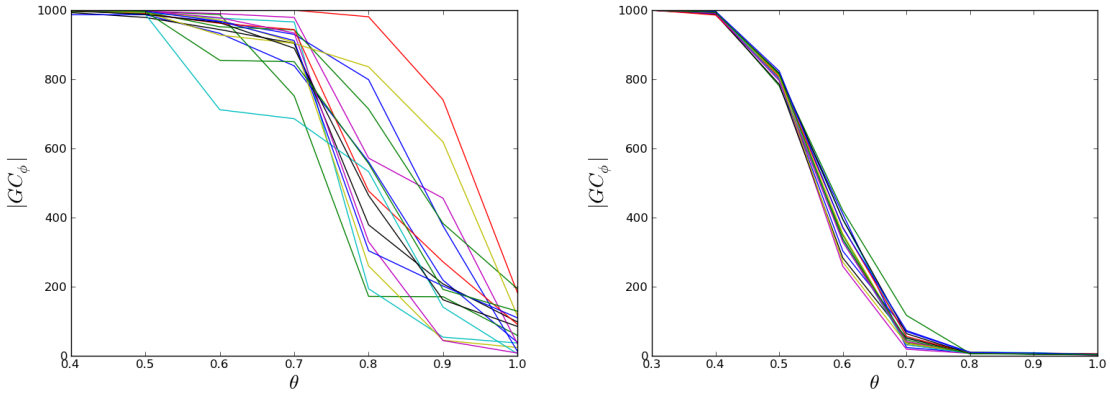

Figure 5.8: Sizes of the giant components of the metanetworks vs. threshold for the evolved populations with $p_0 = 0.5$ (left panel) and their randomized counterparts (right panel).

5.2.2 Topologies of metanetworks in phenotype space

The topologies of the evolved and randomized metanetworks in phenotype space are significantly different from each other, for the thresholds below 0.4. When the threshold is small enough both the evolved metanetworks and the randomized metanetworks are completely connected. However, as the threshold increases the metanetworks start to move away from each other. Fig. 5.9 shows the plots of the metanetworks of evolved and randomized populations with threshold 0.5. There is a huge difference between the mean degrees of the two networks; evolved metanetwork having $\langle k \rangle = 148.79$ and randomized metanetwork having $\langle k \rangle = 1.58$.

The degree distributions of the metanetworks shown in Fig. 5.9 are displayed in Fig. 5.10. As seen from the figures, the degree distributions of the two metanetworks

are completely different. The randomized metanetwork has the highest probability at degree zero, and the probability decrease with increasing degree, whereas the evolved metanetwork has the highest probability at around degree 270, and although it is not regular, probability increases with increasing degree upto degree 150.

5.3 Metanetwork in Phenotype Space Spanned by Genotype Network

In this section, we discuss how much the genotype and the phenotype metanetwoks of the evolved sets coincide with each other. We compare the giant components of the two metanetworks for this. In 3 out of 16 genotype metanetworks, giant components are not formed. Therefore, we only considered 13 evolved sets, where in the genotype metanetwork, giant components are found.

We define the overlap ratio between giant components of genotype and phenotype metanetworks as follows,

$$R \equiv \frac{|GC_{\Phi} \cap GC_{\Gamma}|}{|GC_{\Gamma}|}, \tag{5.5}$$

As seen from Fig. 5.11, in 10 out of 13 sets the the overlap ratios of the metanetworks are approximately 1 above the threshold value 0.7. In 3 sets the ratios drop to 0 above the threshold 0.85. In these 3 sets sizes of the giant components in genotype space are below 100. These results show that in 10 metanetworks, the giant components of the genotype metanetwork span the giant components of the phentoype network compeletely.

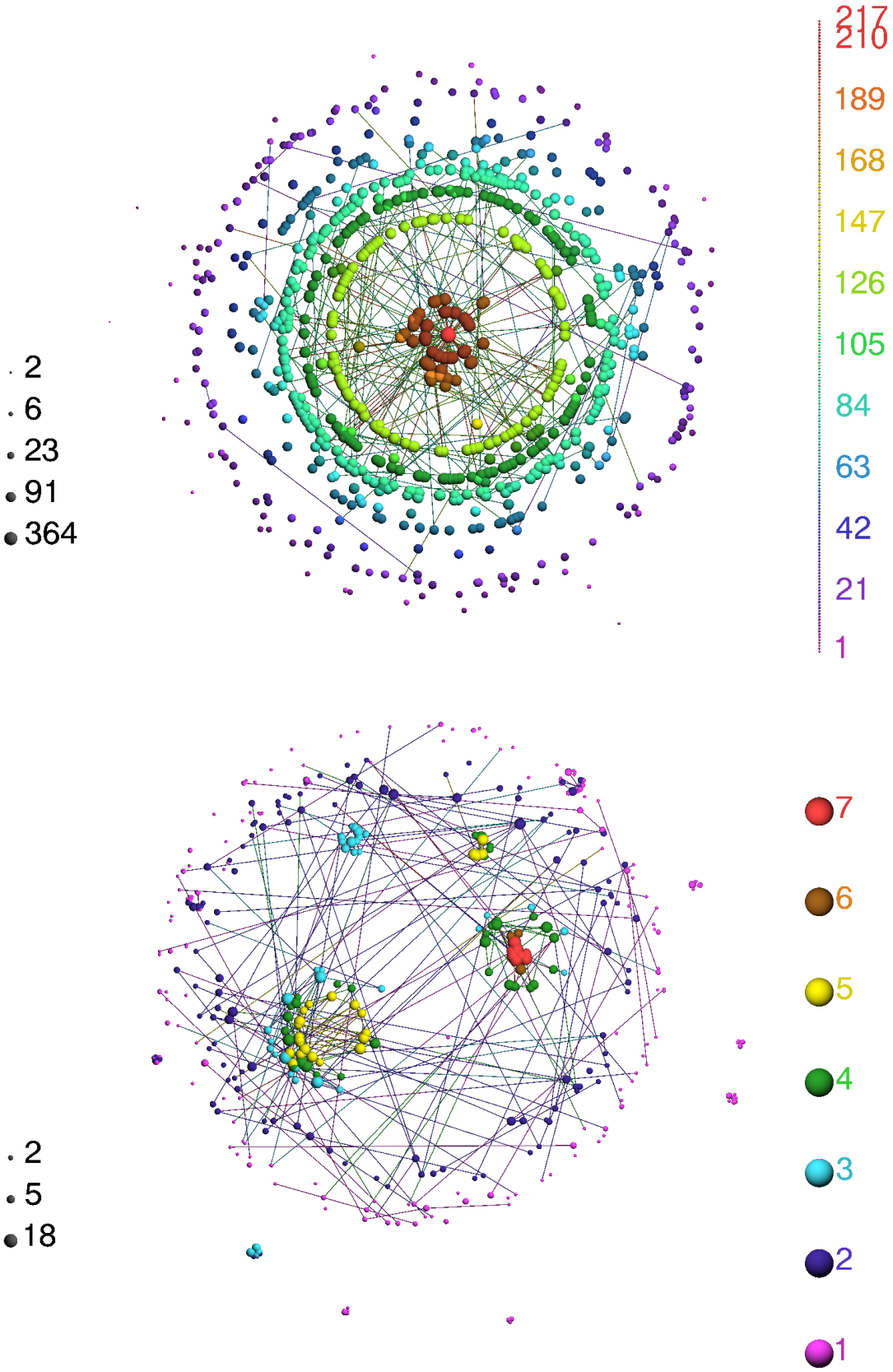

Figure 5.9: Plot of the phenotype metanetwork of an evolved set (upper panel) and its randomized counterpart (lower panel) with threshold 0.5.

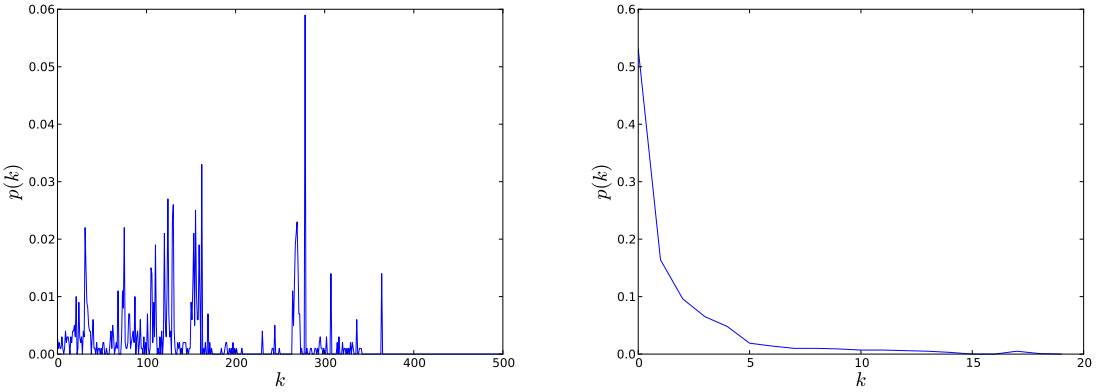

Figure 5.10: Degree distributions of the evolved and randomized metanetworks shown in Fig. 5.9.

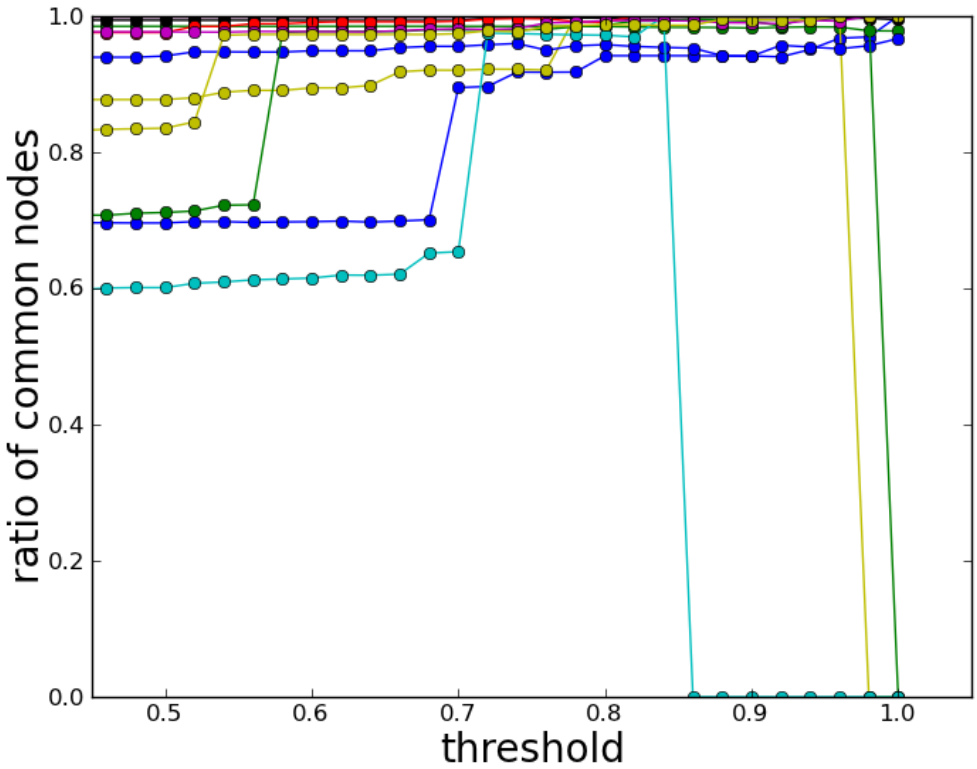

Figure 5.11: Ratio of the common nodes in the giant component of the phenotype network, spanned by the giant component of the genotype metanetwork vs. threshold (Eq. 5.5).

6. DISCUSSION

In this study we have examined the topological and dynamical features of artificially evolved Boolean graphs and compared them to those of the randomized graphs. The Boolean graphs were evolved under a genetic algorithm favoring short attractor lengths. It is found that the mean attractor lengths decay as a power law with time, before the stasis is reached. After the stasis is reached, the comparison between the randomized graphs and the evolved graphs revealed that, the mean attractor lengths of the evolved graphs were shorter than those of the randomized graphs, as expected. However, when the change in the mean degrees during the simulations were examined, the nature of the change in the mean degrees was not the same for all the simulations. This indicates that the fitness landscape is a rugged one, consistent with the slow relaxation with the evolutionary process [22].

Different evolutionary paths followed by independently evolved populations in terms of the mean degrees, suggest that there is no direct correlation between the mean degree and the mean attractor length, for the evolved networks. However, this is not the case for random networks. As one randomly add edges to a set of randomly generated networks, they observe that the average of the mean attractor lengths increase with time. This is consistent with the findings of Aldana et. al., greater edge density leads to longer attractors for netwoks with random Boolean functions [23]. We have also found that the mean number of attractors of the random networks decrease with increasing connectivity.

The topological features examined were degree distributions and motif frequencies. Although the degree distributions of evolved and randomized sets were similar in average, the degree distributions of independent sets differ from those of the randomized sets in general. We have calculated information contents of the distributions to make the comparison. The analysis of the information contents show that the information contents of the degree distributions of the evolved graphs with lower initial edge density are more similar to those of their randomized counterparts,

whereas for the evolved graphs with higher initial edge density information content is significantly lower compared to their randomized counterparts.

The motif frequencies of the evolved graphs were consistent with the findings of Alon et. al. [4] and Thomas et. al. [7]. Feed-forward loops were dominant in evolved graphs, whereas feedback loops were significantly low in frequency. We have calculated significance profiles to compare the evolved graphs to the computational cores of the gene regulatory networks of E. coli, B. subtilus and S. scerevisiae. The comparison yielded that the significance profiles of evolved graphs and gene regulatory networks were similar in nature.

The differences between independently evolved sets show that fitness landscape is a rugged one. As the trajectories in Fig. 3.1 indicate, there are two stages of evolution: slow relaxation to a target phenotype, and spreading over the regions in genotype space where phenotype is neutral to mutations (see [22]) for a more detailed discussion). Therefore, different sets of evolved graphs span different regions both in genotype and phenotype spaces and the evolution of networks in genotype space [21] can be understood in terms of neutral networks. The structure of the neutral networks is strongly correlated with the robustness and evolvability of the regulatory networks [22].

Ciliberti et. al. have studied robustness and capacity of innovation of gene regulatory networks [20], by taking into account only one arbitrary initial state and the attractors reached from that particular state. However, gene regulatory networks may have many attractors reached from different initial conditions, in principle (see 3.3). Since attractors are shared by different graphs, a weighted meta-network can be formed in phenotype space.

According to Ciliberti et. al., for evolutionary innovations a neutral network should have a large diameter and the mean distance between randomly sampled networks having a particular attractor reached from a prespecified initial state is 80% of the maximum distance [20]. In our study, where we have considered all the attractors the Boolean networks in a population have, the mean distance between the randomized graphs is about half of the maximum distance and the distributions of the distances are exactly the same for all the attractors (5.2 right panel). However, the mean distance

between the evolved graphs is about 30% of the maximum distance, implying that evolved graphs are concentrated in a smaller region of genotype space compared to randomized graphs. Ciliberti et. al. also suggests that robustness and ability to innovate are negatively correlated. Therefore, we have tested the robustness of the Boolean graphs by eliminating the weak edges with weights under a threshold θ from the metanetworks formed in the phenotype space. As seen from Fig. 5.8, both the evolved and randomized metanetworks are connected before removal of edges. Above $\theta = 0.4$, randomized metanetworks start to disintegrate and no edges are left when $\theta = 0.8$. On the other hand, the evolved metanetworks start to disintegrate when $\theta = 0.5$ and the sizes of their giant components decrease more slowly with increasing θ compared to randomized metanetworks, leaving giant components with upto 200 nodes even when $\theta = 1$. This indicates the percolation behavior of evolved networks are similar to those of the real-life networks studied before [24, 25].

Our findings show that, by evolving populations of Boolean graphs under selection for short attractors, it is possible to mimic some of the essential features of real-life networks, such as significance profiles of motif structures and mutational robustness.

REFERENCES

[1] **Gershenson, C.** (2003). Classification of Random Boolean Networks, *Proceedings of the Eighth International Conference on Artificial Life*, ICAL 2003, pp.1–8.

[2] **Anıl, M.A.** (2011). Boolcu ağlarda motif istatistiği için bir model, Diploma thesis, ITÜ.

[3] **Milo, R.**, **Shen-Orr, S.**, **Itzkovitz, S.**, **Kashtan, N.**, **Chklovskii, D. and Alon, U.** (2002). Network Motifs: Simple Building Blocks of Complex Networks, ***298***(5594), 824–827.

[4] **Milo, R.**, **Itzkovitz, S.**, **Kashtan, N.**, **Levitt, R.**, **Shen-Orr, S.**, **Ayzenshtat, I.**, **Sheffer, M. and Alon, U.** (2004). Superfamilies of Evolved and Designed Networks, ***303***(5663), 1538–1542.

[5] **Thomas, R.**, (1981). On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations, Numerical methods in the study of critical phenomena, Springer, pp.180–193.

[6] **Thomas, R. and Kaufman, M.** (2001). Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, ***11***(1), 170–179.

[7] **Thomas, R. and Kaufman, M.** (2001). Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, ***11***(1), 180–195.

[8] **Newman, M.E.**, **Strogatz, S.H. and Watts, D.J.** (2001). Random graphs with arbitrary degree distributions and their applications, *Physical Review E*, ***64***(2), 026118.

[9] **Kauffman, S.A.** (1993). *The origins of order: Self-organization and selection in evolution*, Oxford university press.

[10] **Kauffman, S.A.** (1969). Metabolic stability and epigenesis in randomly constructed genetic nets, *Journal of theoretical biology*, ***22***(3), 437–467.

[11] **Holland, J.H.** (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.*, U Michigan Press.

[12] **Erdös, P. and Rényi, A.** (1959). On random graphs, I, *Publicationes Mathematicae*, ***6***, 290–297.

[13] **Erdös, P. and Rényi, A.** (1960). On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci*, ***5***, 17–61.

[14] **Erdös, P. and Rényi, A.** (1961). On the evolution of random graphs, *Bull. Inst. Internat. Statist*, ***38***(4), 343–347.

[15] **Anıl, M.A.**, `https://github.com/kreveik/Kreveik`, date retrieved 29.10.2013.

[16] **Shannon, C.E.** (2001). A mathematical theory of communication, *ACM SIGMOBILE Mobile Computing and Communications Review*, ***5***(1), 3–55.

[17] **Rodríguez-Caso, C.**, **Corominas-Murtra, B. and Solé, R.V.** (2009). On the basic computational structure of gene regulatory networks, *Molecular BioSystems*, ***5***(12), 1617–1629.

[18] **Bollobás, B.** (1998). *Modern graph theory*, volume184, Springer.

[19] **Batagelj, V. and Zaversnik, M.** (2003). An O (m) algorithm for cores decomposition of networks, *arXiv preprint cs/0310049*.

[20] **Ciliberti, S.**, **Martin, O.C. and Wagner, A.** (2007). Innovation and robustness in complex regulatory gene networks, *Proceedings of the National Academy of Sciences*, ***104***(34), 13591–13596.

[21] **Van Nimwegen, E.**, **Crutchfield, J.P. and Huynen, M.** (1999). Neutral evolution of mutational robustness, *Proceedings of the National Academy of Sciences*, ***96***(17), 9716–9720.

[22] **Wagner, A.** (2013). *Robustness and evolvability in living systems*, Princeton University Press.

[23] **Aldana, M.**, **Coppersmith, S. and Kadanoff, L.P.**, (2003). Boolean dynamics with random couplings, Perspectives and Problems in Nolinear Science, Springer, pp.23–89.

[24] **Albert, R.**, **Jeong, H. and Barabási, A.L.** (2000). Error and attack tolerance of complex networks, *nature*, ***406***(6794), 378–382.

[25] **Dorogovtsev, S.N. and Mendes, J.F.F.** (2002). Evolution of networks, *Advances in Physics*, ***51***(4), 1079–1187.

[26] Large Networks Visualization Tool, `<http://lanet-vi.soic.indiana.edu/index.php>`, date retrieved 10.02.2014.

APPENDICES

APPENDIX A : Set Properties
APPENDIX B : Sets with Exclusively Negative and Exclusively Positive Interactions
APPENDIX C : Significance Profiles of Random Networks

APPENDIX A

Here, we present the numerical data for the evolved and randomized populations with initial edge densities $p_0 = 0.2$ (Table A.1) and $p_0 = 0.5$ (Table A.2). $\langle k \rangle_F$ is the mean degree averaged over all the graphs within the population at the final time step $t = 400$. Mean attractor length of the graph a is averaged over all the initial conditions. $\langle a \rangle$ is the mean attractor length averaged over all the graphs in the set. $\langle a \rangle_F$ is the mean attractor length at $t = 400$, averaged over the graphs in the population. $\langle a \rangle_r$ is the mean attractor length averaged over the randomized graphs. $\langle a \rangle_S$ is the mean attractor length averaged over a hundred steps within the time interval $300 < t < 400$. $\langle \mathscr{N}_{\mathscr{A}} \rangle$ is the mean number of attractors averaged over the graphs in the set. $\langle \mathscr{N}_{\mathscr{A}} \rangle_{\mathrm{r}}$ is the mean number of attractors averaged over the randomized graphs. $\overline{\mathscr{O}}$ is the mean overlap between the significance profiles of a given population and other sets; $\mathscr{O}(\mathbf{S}, \mathbf{X})$ is the overlap between the the significance profile of a given set $\mathbf{S}$ and that of biological network $\mathbf{X}$

Table A.1: Numerical results for $p_0 = 0.2$.

set label	$\langle k \rangle_F$	$\langle a \rangle_S$	$\langle a \rangle_F$	$\langle a \rangle_{\mathrm{r}}$	$\langle \mathscr{N}_{\mathscr{A}} \rangle$	$\langle \mathscr{N}_{\mathscr{A}} \rangle_{\mathrm{r}}$	$\overline{\mathscr{O}}$	$\mathscr{O}(\mathbf{S}, E.coli)$	$\mathscr{O}(\mathbf{S}, B.subt.)$	$\mathscr{O}(\mathbf{S}, S.cerev.)$
1	1.09	1.24	1.26	1.91	9.12	8.06	0.69	0.63	0	0.34
2	1.05	1.26	1.24	1.97	8.86	8.38	0.64	0.62	0.03	0.34
3	1.16	1.04	1.02	2.07	9.99	7.64	0.47	0.46	0.29	0.35
4	1.18	1.09	1.10	2.05	9.36	7.40	0.75	0.95	0.58	0.85
5	1.23	1.28	1.29	2.06	4.39	6.68	0.52	0.83	0.89	0.91
6	1.19	1.21	1.23	2.12	7.39	7.20	0.80	0.97	0.53	0.82
7	0.95	1.13	1.12	1.90	4.54	9.11	0.46	0.47	0.38	0.36
8	1.05	1.10	1.09	1.94	5.59	8.40	0.72	0.68	0.21	0.43
9	0.91	1.05	1.04	1.82	5.56	9.78	0.73	0.75	0.33	0.53
10	1.13	1.42	1.43	2.03	8.12	7.10	0.75	0.86	0.41	0.73
11	1.38	1.45	1.50	2.23	4.24	5.50	0.68	0.96	0.81	0.97
12	0.94	1.07	1.05	1.85	8.45	9.27	0.55	0.40	-0.20	0.08
13	1.28	1.17	1.18	2.16	6.81	6.22	0.74	0.96	0.77	0.94
14	1.31	1.28	1.29	2.17	7.42	6.03	0.64	0.72	0.19	0.54
15	1.17	1.27	1.26	2.03	8.08	7.30	0.74	0.99	0.68	0.92
16	1.05	1.21	1.22	1.92	8.63	7.81	0.80	0.89	0.42	0.72

Table A.2: Numerical results for $p_0 = 0.5$.

set label	$\langle k \rangle_F$	$\langle a \rangle_S$	$\langle a \rangle_F$	$\langle a \rangle_r$	$\langle \mathcal{N}_{\mathcal{A}} \rangle$	$\langle \mathcal{N}_{\mathcal{A}} \rangle_r$	$\mathcal{C}$	$\mathcal{O}(\mathbf{S}, E.coli)$	$\mathcal{O}(\mathbf{S}, B.subt.)$	$\mathcal{O}(\mathbf{S}, S.cerev.)$
1	2.86	1.80	1.86	3.41	2.58	2.49	0.55	0.50	0.60	0.55
2	2.44	1.51	1.52	3.13	1.92	2.92	0.55	0.48	0.84	0.67
3	3.42	1.24	1.23	3.93	2.04	2.20	0.57	0.54	0.66	0.60
4	2.86	1.40	1.40	3.35	2.73	2.51	0.60	0.58	0.65	0.67
5	3.28	1.47	1.48	3.91	2.86	2.30	0.58	0.62	0.70	0.75
6	2.57	1.82	1.82	3.23	3.11	2.64	0.51	0.50	0.86	0.78
7	3.01	1.72	1.70	3.64	2.00	2.37	0.51	0.56	0.49	0.61
8	3.83	1.78	1.76	4.30	4.90	2.15	0.17	0.41	0.04	0.22
9	2.85	1.53	1.45	3.37	1.68	2.49	0.41	0.18	0.55	0.38
10	3.14	1.63	1.67	3.64	1.60	2.30	0.26	0.19	0.40	0.32
11	2.75	1.73	1.69	3.40	4.00	2.60	0.44	0.51	0.50	0.48
12	2.72	1.67	1.63	3.49	3.85	2.55	0.51	0.39	0.67	0.57
13	3.39	1.50	1.42	4.02	2.42	2.32	0.14	0.26	0.25	0.36
14	2.99	1.70	1.65	3.46	4.08	2.41	0.61	0.58	0.62	0.63
15	2.28	1.70	1.66	2.94	1.96	2.85	0.45	0.24	0.67	0.55
16	2.99	1.85	1.87	3.48	1.90	2.46	0.57	0.59	0.59	0.66

APPENDIX B

We would like to show that topology plays the most significance role in determining the lengths of attractors. Two sets of randomly generated graphs with 7 nodes and initial edge density $p_0 = 0.5$ have been evolved. One set of graphs includes exclusively positive interactions, whereas the other includes exclusively positive interactions. The resulting significance profiles are shown in Fig. B.1. As seen from the figures the dominance of positive or negative interactions interactions does not have a significant effect on the overall behavior of the profiles.

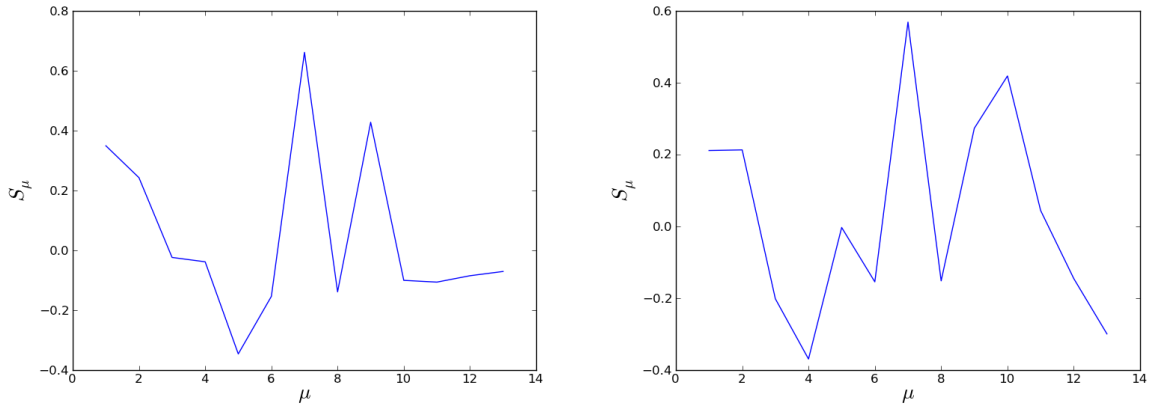

Figure B.1: Significance profiles of two artificially evolved populations with only positive (left panel) and only negative (right panel) interactions and initial connection probability $p_0 = 0.5$.

APPENDIX C

The significance profiles of 16 sets of 10^3 randomly generated graphs are shown in Fig. C.1. Contrary to the evolved sets the significance profiles of the random sets do not display any structure. This is also seen from Fig. C.2 where the overlap matrix between the randomized graphs is shown.

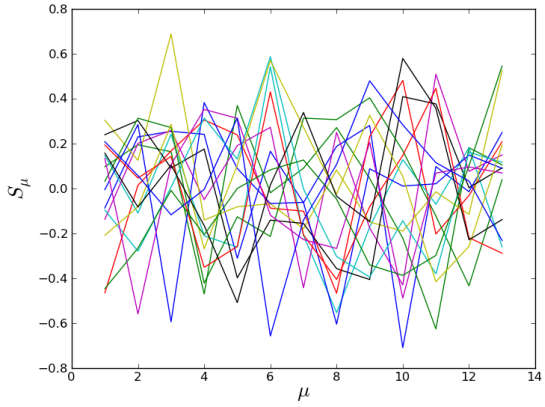

Figure C.1: The significance profiles of 16 randomly generated populations with initial connection probability $\langle p \rangle_0 = 0.5$. No common structure is observed in the profiles.

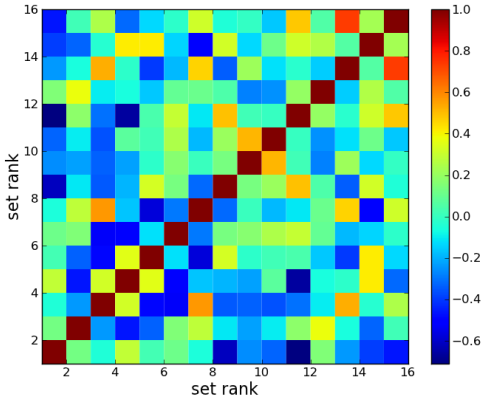

Figure C.2: The overlap between SPs of 16 randomly generated populations with 10^3 graphs and initial edge density $p_0 = 0.5$. The side bar contains the color code.

CURRICULUM VITAE

Name Surname: Burçin Danacı

Place and Date of Birth: Bilecik, 09/01/1984

Adress: Istanbul

E-Mail: danacib@itu.edu.tr

B.Sc.: Istanbul Technical University

M.Sc.: Istanbul Technical University

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- Danacı B., Anıl M. A. and Erzan A., "Motif statistics of articially evolved and biological networks", arXiv:1402.728.

- Danacı B. and Erzan, A., 2014: Model Regulatory Networks Form Metanetworks in Genotype and Phenotype Spaces. *Conference of the Middle European Cooperation in Statistical Pysics*, Coventry, England

- Danacı B., Anıl M. A. and Erzan A., 2013: Motif Statistics Of Artificially Evolved And Biological Networks. *European Conferance on Complex Systems*, Barcelona, Spain

- Danacı B., Anıl M. A. and Erzan A., 2012: Motif Statistics and Dynamics on Boolean Networks. *Third International Workshop on Statistical Mechanics and Dynamical Systems*, Marmaris, Turkey