

**T.C.  
BAHÇEŞEHİR ÜNİVERSİTESİ**

**SİNYAL YOLAKLARININ YENİDEN  
OLUŞTURULMASI İÇİN SEZGİSEL  
ALGORİTMA İLE GRAF TEORİK YAKLAŞIM**

**Yüksek Lisans Tezi**

**İREM KARADUMAN**

**İSTANBUL, 2012**

**T.C.  
BAHÇEŞEHİR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ  
MATEMATİK ANA BİLİM DALI  
UYGULAMALI MATEMATİK**

**SİNYAL YOLAKLARININ YENİDEN  
OLUŞTURULMASI İÇİN SEZGİSEL  
ALGORİTMA İLE GRAF TEORİK YAKLAŞIM**

**Yüksek Lisans Tezi**

**İREM KARADUMAN**

**Tez Danışmanı: Yrd. Doç. Dr. Süreyya ÖZÖĞÜR AKYÜZ**

**İSTANBUL, 2012**

**T.C.**  
**BAHÇEŞEHİR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**  
**MATEMATİK ANA BİLİM DALI**  
**UYGULAMALI MATEMATİK**

Tezin Adı : Sinyal Yolaklarının Yeniden Oluşturulması için Genetik  
Algoritma ile Graf Teorik Yaklaşım

Öğrencinin Adı Soyadı : İrem KARADUMAN

Tez Savunma Tarihi : 24 / 04 / 2012

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğu Fen Bilimleri Enstitüsü tarafından onaylanmıştır.

Doç. Dr. F. Tunç BOZBURA  
Enstitü Müdürü  
İmza

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğunu onaylarım.

Prof. Dr. Nuri KURUOĞLU  
Program Koordinatörü  
İmza

Bu Tez tarafımızca okunmuş, nitelik ve içerik açısından bir Yüksek Lisans tezi olarak yeterli görülmüş ve kabul edilmiştir.

Jüri Üyeleri

İmzalar

Tez Danışmanı  
Yrd. Doç. Dr. Süreyya ÖZÖĞÜR AKYÜZ

Üye  
Doç. Dr. Atabey KAYGUN

Üye  
Yrd. Doç. Dr. Devrim ÜNAY

-----  
-----  
-----

## ÖNSÖZ

Çalışmalarım sırasında yardımlarını, bilgisini ve anlayışını esirgemeyen Yard. Doç. Dr. Süreyya ÖZÖĞÜR AKYÜZ'e, görüşlerini ve bilgilerini paylaşan, umutsuz anlarda yüreklendiren Doç. Dr. Atabey KAYGUN'a, katkılarından dolayı Öğr. Gör. Burcu BAKIR GÜNGÖR'e, manevi desteklerini her zaman hissettiğim Gülnur ŞAHİN SEICHAN OGLOU, Zeliha KILIÇ, Ebru DEMİRBAŞ, Ayşen TEZCAN, Özge KARADUMAN ve aileme teşekkürlerimi sunarım.

## ÖZET

### SİNYAL YOLAKLARININ OLUŞTURULMASI İÇİN GENETİK ALGORİTMA İLE GRAF TEORİK YAKLAŞIM

İrem KARADUMAN

Matematik Ana Bilim Dalı

Uygulamalı Matematik

Tez Danışmanı: Yrd. Doç. Dr. Süreyya ÖZÖĞÜR AKYÜZ

Nisan 2012, 31 sayfa

Çok hücreli organizmaların yaşamlarını devam ettirebilmeleri ve gelişebilmeleri için hücrelerin birbirleriyle iletişim kurmaları gerekmektedir. Hücre sinyali hücreler arası iletişimi sağlayan bir çeşit kimyasal tepkimedir. Sinyal yolları ise bu kimyasal tepkimeleri ifade eder. RNAi sistemi virüslü hücrenin RNA'larının vücutta tanımlayıp yok edilmesini sağlar. Sinyal yolları RNAi verileri kullanılarak yapılandırılabilir. Boolean RNAi söndürme verileri kullanılarak protein-protein etkileşimini ifade eden sinyal yollarının topolojik yapısının oluşturulması çözüm uzayı oldukça büyük olan ters bir problemdir (inverse problem). Bu çalışmada sezgisel algoritma olan Genetik Algoritma (GA) kullanılarak tekli Boolean RNAi söndürme deneyleriyle uyumlu olan ve çözüm uzayını küçülten topolojiler bulunmuştur.

**Anahtar Kelimeler:** Sinyal Yolları, RNAi verileri, Genetik Algoritma, Graf Teori

## ABSTRACT

### RECONSTRUCTION OF SIGNALING PATHWAYS BY NOVEL GRAPH THEORETICAL APPROACH

İrem KARADUMAN

Department of Mathematics  
Applied Mathematics

Thesis Supervisor: Assist. Prof. Süreyya ÖZÖĞÜR AKYÜZ

April 2012, 31 pages

Intercellular communication is required for cells to keep on their lives and to be discovered. Cell signaling is a list of chemical reactions that are used for intercellular communication. Signaling pathways denote these chemical reactions. RNAi system enables us to identify and remove RNAs of viral cells. Signaling pathways can be constructed by using RNAi data. Constructing the topologies of signaling pathways which denote protein-protein interaction by using Boolean RNAi knockdown data has a large solution space and is an inverse problem. In this study, topologies compatible with single Boolean RNAi knockdown experiments which reduce the dimension of solution space are found using one of the heuristic algorithms called Genetic Algorithm.

**Keywords:** Signaling Pathways, RNAi data, Genetic Algorithm, Graph Theory

## İÇİNDEKİLER

TABLolar.....	vii
KISALTMALAR.....	viii
SEMBOLLER.....	ix
ŞEKİLLER.....	x
1. GİRİŞ.....	1
2. BİYOLOJİK ÇALIŞMALAR.....	3
2.1 SİNYAL YOLAKLAR.....	3
2.2 RNAi DENEYLERİ.....	4
3. MATEMATİKSEL ÇALIŞMALAR.....	6
3.1 GENETİK ALGORİTMA (GA).....	6
3.1.1 Genetik Algoritmanın Çalışma Prensipleri.....	7
3.1.2 Kromozomların Kodlanması ve İlk Topluluğun Oluşturulması.....	7
3.1.3 Uygunluk Değerinin Hesaplanması.....	9
3.1.4 Genetik Operatörler.....	9
3.2 GRAF TEORİ.....	12
4. SİNYAL YOLAKLARININ OLUŞTURULMASI İÇİN SEZGİSEL ALGORİTMA İLE GRAF TEORİK YAKLAŞIM.....	17
4.1 PROBLEMİN TANIMI.....	17
4.2 GRAF TEORİK YAKLAŞIM.....	18
4.2.1 Eklemler Boyunca Kabul Edilebilir Yönlü Grafların Ayrıştırılması.....	18
4.3 SİNYAL YOLAKLARININ OLUŞTURULMASI İÇİN SEZGİSEL ALGORİTMA İLE GRAF TEORİK YAKLAŞIM.....	21
5. TARTIŞMA ve SONUÇ.....	25
6. GELECEK ÇALIŞMALAR.....	27
KAYNAKÇA .....	28

## TABLÖLAR

Tablo 1.1: Oluřabilecek m¼mk¼n topoloji sayıları.....	2
Tablo 3.1: İki kromozomun arprazlanması.....	10
Tablo 3.2: Mutasyon yöntemleri.....	11
Tablo 5.1: 100 iterasyon için graf topoloji sayıları.....	25
Tablo 5.2: 50 İterasyon için graf topoloji sayıları.....	26



## KISALTMALAR

GA : Genetik Algoritma  
RNAi : RNA İnterferansı

## SEMBOLLER

Graf	:	$G$
Düğüm kümesi	:	$V$
Ayrıtlar kümesi	:	$E$
Kaynak noktası	:	$s$
Hedef noktası	:	$t$

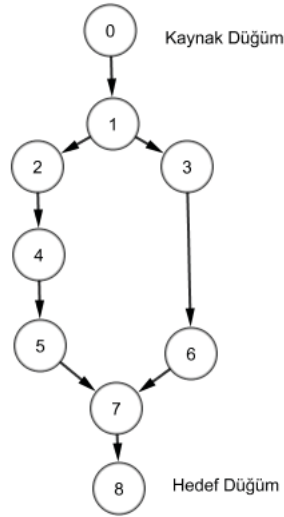
## ŞEKİLLER

Şekil 1.1 : Örnek.....	1
Şekil 2.1 : Sinyal yolları.....	3
Şekil 2.2 : RNAi deneyleri .....	4
Şekil 3.1 : GA akış şeması.....	8
Şekil 3.2 : Yönsüz graf .....	12
Şekil 3.3 : Yönlü graf .....	13
Şekil 3.4 : Çember graf çeşitleri .....	14
Şekil 3.5 : Bitişiklik matrisleri .....	15
Şekil 4.1: Problem.....	17
Şekil 4.2: Grafların parçalanması.....	21
Şekil 5.1: Algoritma akış şeması.....	25

# 1. GİRİŞ

Çok hücreli organizmaların yaşamlarını devam ettirebilmeleri ve hücrelerin gelişimi için birbirleriyle iletişim kurmaları gerekmektedir. Hücre sinyali hücreler arası iletişimi sağlayan bir çeşit kimyasal tepkimedir. Hücre içinde meydana gelen bu tepkimeler diğer kimyasal tepkimeleri etkilemektedir. Kimyasal tepkimelerin oluşumunda meydana gelen herhangi bir aksaklık canlıda hayati fonksiyonların bozulmasına neden olur.

Hücre içinde meydana gelen kimyasal tepkimeler Sinyal Yolakları (*signaling pathway*) ile ifade edilir. Her bir yolaktaki kimyasal bileşik, enzimler tarafından değişime uğrar. Hücrelerde pek çok yolak bulunur. Bu yollarda pek çok bileşik ve enzim yer aldığı için karmaşık ağlar oluşturabilirler.



Şekil 1.1a Graf topolojisi örneği

RNAi verisi	
1	Aktif değil
2,3,4,5,6	Aktif
7	Aktif değil

Şekil 1.1b RNAi verisi

## Şekil 1.1: Örnek

Gen ekspresyonu, DNA dizisi olan genlerin, fonksiyonel protein yapılarına dönüşme süreci için kullanılan bir terimdir. Basitçe, bu durum genlerin aktif olup olmadıkları olarak da tanımlanabilir. Canlı hücreler içinde yer alan ve hangi genin aktif olacağını belirleyen ve kontrol eden sisteme RNA interferansı (RNAi) adı verilir. RNAi verisi Şekil 1.1b'de gösterildiği gibi, genlerin iptali (*knockdown*) sonucu gen ekspresyonunun durduğunu gösteren veriler topluluğudur. Sinyal yollarını RNAi verilerini kullanarak

yapılandırabiliriz. Boolean RNAi söndürme verileri kullanarak sinyal yollarının oluşturulması çözüm uzayı oldukça büyük olan ters bir problemdir (*inverse problem*).

Düğüm sayısı  $N$  olan bir graf için, düğümler arasında olası  $n \times n$  yönlü ayrıt vardır. Bu da olası bütün graf topolojilerinin sayısını  $2^{n \times n}$  yapar. Yani  $n = 5$  düğümlü bir graf için  $2^{5 \times 5} = 33.554.432$  yönlü graf oluşturulabilir. Bu sayı Tablo 1.1’de gösterildiği gibi  $n$  büyüdükçe çok hızlı ve üstel olarak arttığı için bütün olasılıkları tek tek test etmek imkansızdır.

**Tablo 1.1: Oluşabilecek mümkün topoloji sayıları**

Düğüm Sayısı ( $n$ )	Ayrıt Sayısı ( $n \times n$ )	Topoloji Sayısı ( $2^{n \times n}$ )
$n = 2$	4	16
$n = 3$	9	512
$n = 4$	16	65.536
$n = 5$	25	33.554.432
$n = 6$	36	68.719.476.736
$n = 7$	49	562.949.953.421.312
$n = 8$	64	18.446.744.073.709.551.616

Bu tezde, çözüm uzayı çok büyük olan problemi çözebilmek için matematiksel olarak graf teorik bir yaklaşım kullanılmıştır. Her bir geni düğüm (*vertex*), düğümler arası ayrıtları (*edges*) genler arasındaki etkileşim, RNAi verisinden alınan bilgiyi ise bağlantılılığı bozan düğümler olarak düşünülürse sinyal yolları başka bir ifade ile graf topolojileri Şekil 1.1a’daki gibi oluşturulabilir. Şekil 1.1a’ya göre 1 nolu genin iptali sonucu düğüm 8’e ulaşabileceğimiz 0-1-3-5-7-8 nolu yolak vardır, dolayısıyla 0-1-3-5-7-8 yolağı Şekil 1.1b’deki RNAi verisi ile uyumludur.

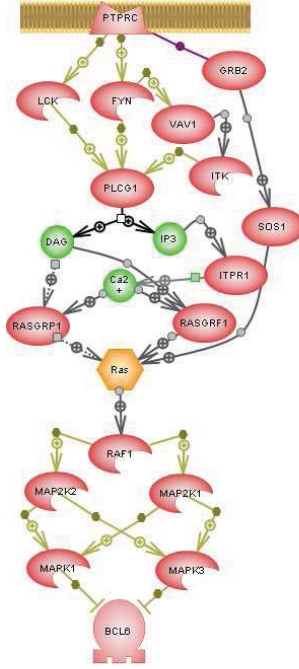
Bu tezde 1-bağlantılı alt graflar Genetik Algoritma ile oluşturularak bütün graf topolojisi elde edilmiş problemin karmaşıklığı (*complexity*) fark edilir şekilde azaltılmıştır.

## 2. BİYOLOJİK ÇALIŞMALAR

Bu kısımda sinyal yollarının biyolojik anlamı ve RNAi verileri hakkında bilgi verilecektir.

### 2.1 SİNYAL YOLAKLARI

Çok hücreli organizmalarda yaşamın devamı ve gelişimi için hücrelerin birbirleriyle iletişim kurmaları gerekmektedir (Pathway Studio 2012). Hücre sinyali hücreler arası iletişimi sağlayan bir çeşit kimyasal tepkimedir. Sinyal yolları ise bu kimyasal tepkimeleri ifade eden Şekil 2.1’de gösterildiği gibi karmaşık ağlar oluştururlar. Hücre içinde enzimlerin yürüttüğü biyokimyasal reaksiyonlarla gerçekleşirler.



Şekil 2.1: Sinyal yolları (Pathway Studio 2012).

Çok hücreli organizmalarda organizmanın bir bütün olarak çalışmasını sağlamak için bireysel hücrelerin davranışlarını kontrol eden bir çok sinyal iletim süreci gerekmektedir.

Sinyal yollarında bozukluk olduğunda pek çok hücrenin işleyişinde aksaklıklar meydana gelir. Pek çok hastalığın temeli sinyal yollarındaki bozukluklara dayanır.

## 2.2 RNAi DENEYLERİ

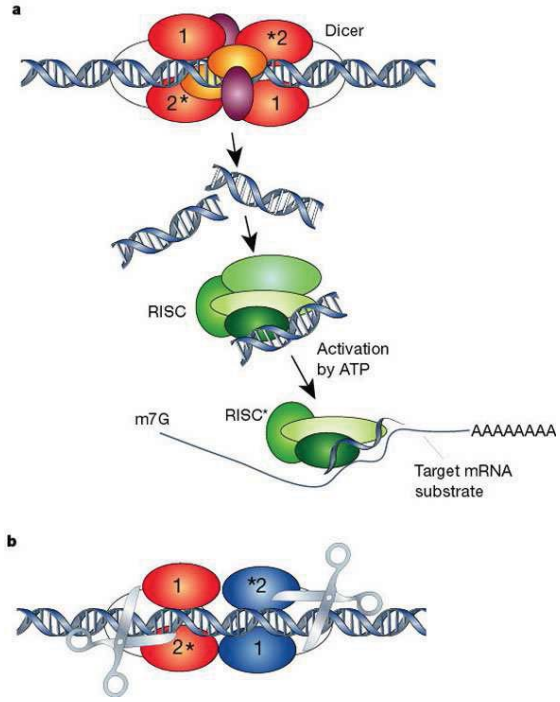
RNA interferansı (RNAi) canlı hücreler içinde yer alan ve hangi genin aktif olacağını belirleyen ve kontrol eden sistemin adıdır. İnsan genomunun ve birçok organizmanın dizi analizlerinin tespit edilmesinde, fonksiyonu henüz bilinmeyen çok sayıda genin açığa çıkarılması ve gen susturma mekanizması için etkili metotların bulunması gen fonksiyonu alanında dikkat çekici bir artış göstermiştir (Kurreck 2003).

Yıllardır çok sayıda gen oluşturma mekanizması ortaya çıkmıştır ve temel olarak genlerin kendisini veya kodladıkları mRNA'ları hedef almaktadır (Gewirtz 2007). RNAi, çift zincirli RNA'nın (dsRNA) hücreye girdiği zaman komplementer (tamamlayıcı) mRNA dizisinin parçalanmasına yol açması ile sonuçlanan transkripsiyon sonrası gen susturma mekanizmasıdır (Atalay 2007). Bu mekanizma doğal bir işlem olup, canlı organizmadaki biyolojik fonksiyonu, virüs kalıtım materyali ve transpozonlar gibi hareketli genetik elementlerin istilasına karşı genomu koruyarak hücrel savunmada rol almaktadır (Agrawal 2003).

RNAi mekanizmasının başlangıç adımı hücredeki uzun dsRNA'ların Dicer enzimi tarafından siRNA yani küçük engelleyici RNA (*Small interfering*) moleküllerine parçalanması işlemidir (bkz Şekil 2a) (Bernstein 2001).

Hücrelerde RNA mekanizmasını uzun dsRNA'lar başlatırken çoğu deneysel çalışmada etki edici moleküller olarak siRNA'lar kullanılmaktadır. İkinci adımda ise, ilk aşamada üretilen çift zincirli siRNA'lar bağlandığı substratla yol olarak RISC-RNA İndükleyici Baskılama Operatörü (*RNA-Induced Silencing Complex*) kompleksi adı verilen bileşene bağlanmaktadır (bkz Şekil 2b) (Filipowicz 2005). Uygun siRNA zincirinin rehberliğinde, komplementer hedef mRNA'yı parçalamaktadır. Kısaca RISC kompleksi mRNA'nın bilgisini taşıdığı genin ekspresyonunu durdurur (bkz Şekil 2b). Diğer bir deyişle bağlandığı mRNA'dan protein üretimine engel olur.

İnsan genomunun ve birçok organizmanın genomunun belirlenmesi, fonksiyonu henüz bilinmeyen çok sayıda geni açığa çıkarmakta ve RNAi diğer tekniklere göre daha hızlı ve



**Şekil 2.2: RNA interferans (Hannon 2002)**

ucuz bir metot olarak fonksiyonel genomik arařtırmalarında çok ideal bir teknik olarak etkin bir şekilde kullanılmaktadır (Downward 2004).

Bu çalışmada sinyal iletim yollarının yapılandırılması için RNAi yöntemiyle herbir genin ekspresyonunu tek tek durdurulması suretiyle elde edilen veriler kullanılmıştır.



### 3. MATEMATİKSEL ÇALIŞMALAR

Bu bölümde sezgisel bir algoritma olan genetik algortima ve graf teori hakkında bilgi verilecektir.

#### 3.1 GENETİK ALGORİTMA (GA)

Algoritmalar, matematiksel ifadeler veya belli kurallar ile tanımlanan problemleri çözmek için kullanılan yöntemlerdir. Algoritmaların tamamı optimum (en iyi) çözümü bulmayı garanti edemezler. Bunun yerine optimuma yakın olan çözümü bulmayı garanti eden algoritmalar sezgisel (*heuristic*) algortimalar olarak adlandırılırlar. GA, Darwin'in "En iyi olan yaşar" (*survival of the fittest*) prensibine dayalı olarak biyolojik olayların gelişim sürecini simüle eder. Evrim ve genetiğin doğal sürecine dayalı bir araştırma tekniği olan GA ilk defa Holland (1975) tarafından önerilmiştir (Goldberg 1989).

Karmaşık, kısıt sayısı fazla, amaç fonksiyonu kurulamayan ve kesin çözüm yöntemi olmayan optimizasyon problemleri GA'nın ilgi alanına girmektedir. GA'lar, çözümü zor optimizasyon problemlerinde, kesin çözüm olmasa bile yaklaşık bir çözümü kısa sürede bulma özelliğine sahiptir (Mitchell 1998).

Geniş bir problem uzayı boyunca rastgele bir araştırmaya imkan sağlayan GA'nın avantajlarından biri, optimize etmeye çalıştıkları problem ile ilgili herhangi bir bilgiye ihtiyaç duymamasıdır.

Problemin parametre değerlerinin ve bunların uygunluğunun değerlendirilmesinde sadece kodlamaya ihtiyaç duyan GA'lar sezgisel bir fonksiyon kullanılarak çalışırlar. Bu araştırma karmaşıklığı azaltır ve diğer araştırma metotlarına göre daha etkin bir yol sağlar. GA'lar birçok problem için optimale yakın çözümlerin bulunmasında başarılı olmuşlardır.

GA'lar rastgele örneklenen çözüm uzayı ile çalışmaya başlamaktadır. Doğal evrimsel süreçte olduğu gibi başlangıçta rastgele seçilebilen veriler üzerinde üreme, çaprazlama ve mutasyon şeklinde üç temel işlem yapılmaktadır. Temel işlemler, belirsiz (deterministik olmayan) işlemlerdir ve sezgisel ya da rastgele olma özelliği taşımaktadırlar.

Bu çalışmada ele alınan algoritma, zor ve karmaşık sayılabilecek bir probleme çözüm aramaktadır. Bu zorlukların ve bunların etkilerinin matematiksel yolla ifadesi her zaman mümkün olmamaktadır. Problemdeki çözümlerin sayısı çok fazladır.

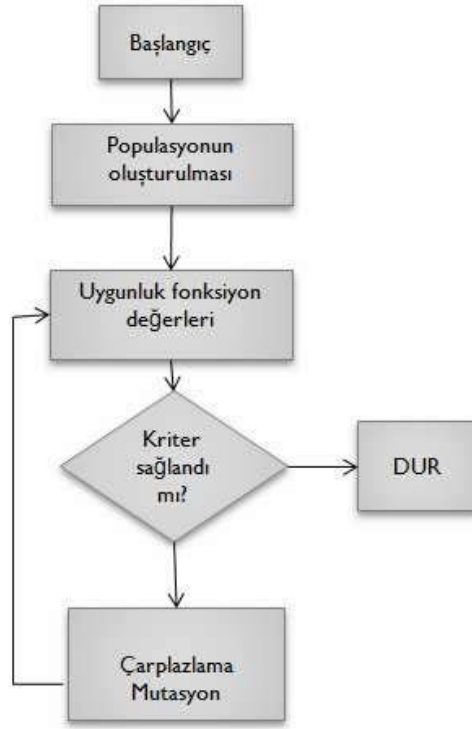
### **3.1.1 Genetik Algoritmanın Çalışma Prensipleri**

1. Arama uzayındaki tüm gerçekleşmesi mümkün çözümler dizisi olarak kodlanır.
2. Genellikle gelişmiş güzel olarak bir çözüm kümesi seçilir ve başlangıç topluluğu olarak kabul edilir.
3. Her bir dizi için uygunluk değeri hesaplanır, bulunan uygunluk değerleri dizilerin çözüm kalitesini gösterir.
4. Bir grup dizi belirli bir olasılık değerine göre gelişmiş güzel olarak seçilip çoğalma işlemi gerçekleştirilir.
5. Yeni bireylerin uygunluk değerleri hesaplanarak, çarpazlama ve mutasyon işlemleri gerçekleştirilir.
6. Önceden belirlenen iterasyon sayısı kadar adım (1-5) tekrarlanır.
7. Uygunluk fonksiyonuna göre en uygun olan dizi seçilir (Engin 2001).

GA'nın genel akış şeması Şekil 3.1'deki gibidir.

### **3.1.2 Kromozomların Kodlanması ve İlk Topluluğun Oluşturulması**

GA, belirli problemlerin mümkün çözümlerini kromozom benzeri veri yapıları şeklinde kodlar. Her çözüm belirli kriterlere göre kodlanmış diziler haline getirilir ve bu diziler kromozom olarak adlandırılır. Dizilerden her biri problemin olası çözümler uzayındaki rastgele bir noktasını simgeler. Kromozomların kodlanması farklı şekillerde olabilir. Kodlamalar arasında en kullanışlı olanlar ikili kodlama, permütasyon kodlaması ve değer kodlamasıdır (Mitchell 1998).



**Şekil 3.1: GA akış şeması**

1. İkili Kodlama: GA'da ilk kullanılan ve en yaygın olan kodlama şeklidir. Her kromozom 1 ve 0'lardan oluşan katarlar halinde ifade edilir.
2. Permütasyon Kodlaması: Daha çok sıralama problemlerinde kullanılan kodlama yöntemidir. Görev, güzergah sıralaması problemlerinde kullanılabilir.
3. Değer Kodlaması: İçerisinde gerçek sayısal değerler barındıran problemlerde kullanılan kodlama şeklidir. Kromozomlar gerçek değerlerle doğrudan kodlanır (Fafali 2003).

### 3.1.3 Uygunluk Deęerinin Hesaplanması

Bir kuşak oluşturulduktan sonraki ilk adım, topluluktaki her üyenin uygunluk deęerini hesaplama adımıdır. Çözümü aranan her problem için bir uygunluk fonksiyonu mevcuttur. Verilen belirli bir kromozom için uygunluk fonksiyonu, o kromozomun temsil ettięi çözümün kullanımıyla orantılı olan sayısal bir uygunluk deęeri verir. Bu bilgi her kuşakta daha uygun çözümlerin seçiminde yol gösterir (Yeniay 2001).

### 3.1.4 Genetik Operatörler

#### Çoęalma Operatörü

Çoęalma operatöründe diziler, amaç fonksiyonuna göre kopyalanır ve iyi kalıtsal özellikleri gelecek kuşaęa daha iyi aktaracak bireyler seçilir (Fıęlalı 2002). Uygunluk deęerinin hesaplanma adımıdan sonra mevcut kuşaktan yeni bir topluluk yaratılmasıdır. Bu yöntemin amacı, ortalama uygunluęun üzerindeki deęerlere çoęalma fırsatı tanımaktır. Bir dizinin kopyalanma şansı, uygunluk fonksiyonuyla hesaplanan dizinin uygunluk deęerine baęlıdır (Jang 1997).

#### Çarpazlama Operatörü

Mevcut gen havuzunun potansiyelini araştırmak ve bir önceki kuşaktan daha iyi nitelikler taşıyan yeni kromozmlar yaratmak için çarpazlama operatörü kullanılmaktadır (Jang 1997). Kısaca, çarpazlama ile iyi uygunluk deęerine sahip iki bireyin iyi özelliklerini birleştirek daha iyi sonuçlar elde etmektir. Fakat hangi özelliklerin iyi performans sağladığına yönelik bir fikre sahip olunamadığı için özelliklerin deęişimi şeklinde birleşim rastgele olarak gerçekleştirilir. Bu şekilde rastgele olarak yapılan birleşimler ile iyi sonuçlar alınması beklenir. Çarpazlama sonucunda iyi olmayan bireyler de oluşabilir ancak onlar uygunluk deęerleri düşük olacağından elenecektir (Whitley 1994).

Programlama sürecinde farklı çarpazlama operatörleri kullanılabilir. Bu operatörlerden hangisinin seçileceğine ise problemin çözümü öncesi yapılacak deęerlendirmelere bakarak karar verilir.

### Çarpazlama Operatörleri:

1. Tek Nokta Çarpazlama: Kromozlar üzerinde tek bir nokta seçilir ve o nokta başlangıç alınarak ebeveyn kromozlar çarpazlanır.
2. İki Nokta Çarpazlama: Tek noktadan farklı iki referans nokta seçilmesidir.
3. Çok Nokta Çarpazlama: İki noktanın gelişmiş halidir. Bir çok başlangıç noktası seçilir ve başlangıç noktalarına göre ebeveynlerin kromozları çarpazlanır.
4. Üniorm Çarpazlama: Rastgele olarak çarpazlama maskesi oluşturulur. Birinci ve ikinci kromozoma karşılık gelen genin koplanmasıyla oluşturulur. Çarpazlama maskesinde 1 genin birinci kromozomundan, 0 ise genin ikinci kromozomdan kopyalanacağı anlamına gelmektedir (Fafali 2003).

Çarpazlama operatörlerine örnek Tablo 3.1’de verilmiştir.

**Tablo 3.1: İki kromozomun çarpazlanması**

	<b>Tek nokta</b>	<b>İki Nokta</b>	<b>Çok Nokta</b>	<b>Üniorm</b>
Kromozom	110100101	110100101	110100101	110100101
Kromozom	011010011	011010011	011010011	011010011
Kromozom Maskesi	...	...	...	101010101
Yeni Kromozom	110100011	110010101	111100111	110000111

## Mutasyon Operatörü

GA'da çeşitlendirmeye gitmede mutasyondan faydalanılır. Sürekli yeni nesil üretimi sonucunda belirli bir süre sonra nesildeki kromozomlar birbirlerini tekrarlama konumuna gelebilir, bunu sonucunda farklı kromozom üretimi durabilir veya azalabilir (Goldberg, 1989). Bu sebepten dolayı kromozomların çeşitliliğini arttırmak için kromozomlardan bazıları mutasyona uğratılır.

Oluşan yeni çözümlerin önceki çözümü kopyalamasını önlemek ve sonuca daha hızlı ulaşmak için yapılır. Mutasyon operatörü olarak, ele alınan problemin yapısına göre en uygun olan mutasyon yöntemlerinden biri seçilir (Braysy 2001).

Mutasyon Operatörleri:

1. Ters Çevirme Mutasyonu: Mutasyon için seçilen kromozomdan iki gen gelişi güzel seçilir, bu iki kısımdaki alt diziler ters çevrilir.
2. Ekleme Mutasyonu: Kromozomdan rastgele bir kısım seçilir, gelişi güzel bir yere yerleştirilir.
3. Yer Değişikliği Mutasyonu: Rastgele bir alt dizi seçilir ve rastgele bir yere yerleştirilir.
4. Karşılıklı Değişim Mutasyonu: Rastgele seçilen iki genin yerleri değiştirilir (Fafali 2003).

Mutasyon operatörlerine örnek Tablo 3.2'de verilmiştir.

**Tablo 3.2: Mutasyon yöntemleri**

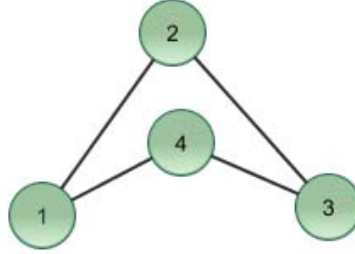
	<b>Ters Çevirme</b>	<b>Ekleme</b>	<b>Yer Değişikliği</b>	<b>Karşılıklı Değişim</b>
Kromozom	0111000101	0111000101	0111000101	0111000101
Yeni Kromozom	0100110101	0101000101	0101011100	0011010101

## 3.2 GRAF TEORİ

**Tanım 3.1.** *Düğüm olarak adlandırılan noktalar ve her biri bu noktaları veya sadece noktanın kendisini birleştiren ve kenar olarak adlandırılan çizgiler topluluğuna **Graf** denir (West 2001).*

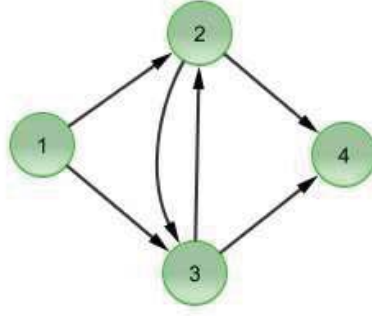
**Tanım 3.2.** *Eğer bir grafta bulunan kenarlar yön bildirmiyorsa **yönsüz (undirected)** graf denir. Bu durumda iki düğüm arasında bulunan kenar, her iki yönde de hareket edilebileceğini ifade eder.*

$V = \{1, 2, 3, 4\}$  düğüm kümesi,  $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$  kenar kümesi olmak üzere oluşacak yönsüz graf Şekil 3.2'deki gibidir.



**Şekil 3.2: Yönsüz graf**

**Tanım 3.3.**  $V$  bir küme ve  $\Delta(V) = \{(v, v) | v \in V\}$  olmak üzere  $E \subseteq (V \times V) \setminus \Delta(V)$  olsun.  $(V, E)$  çifti yönlü graf olarak adlandırılır.  $V$ 'nin elemanları düğüm noktaları,  $E$ 'nin elemanları da ayrıt olarak adlandırılır.  $V = \{1, 2, 3, 4\}$  düğüm kümesi,  $E = \{(1, 2), (1, 3), (2, 3), (3, 2), (2, 4), (3, 4)\}$  kenar kümesi olmak üzere oluşacak yönlü graf Şekil 3.3'deki gibidir.



**Şekil 3.3: Yönlü graf**

**Tanım 3.4.**  $G = (V, E)$  grafının *karşıt grafi*

$$E^{op} = \{(v, v') \in V \times V \mid (v', v) \in E\} \quad (3.1)$$

olmak üzere  $G^{op} = (V, E^{op})$  çiftidir.  $U \subseteq V$  olsun.  $U$  üzerinde  $G$ 'nin bir **Tam Altgrafi (full subgraph)**

$$G(U) := (U, (U \times U) \cap E) \quad (3.2)$$

grafıdır.

**Tanım 3.5.** *Ayrı düğüm çiftlerinin tümü bir kenar ile bağlı olan graflara **Tam (Complete) Graf** denir ve  $n$  düğüm sayısı olmak üzere  $K_n$  ile gösterilir.*

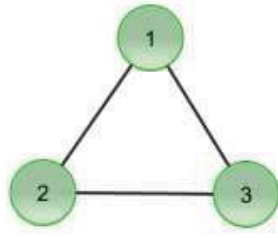
**Tanım 3.6.**  $G = (V, E)$  ve  $G' = (V', E')$  iki yönlü graf olsunlar. Eğer  $f : V \rightarrow V'$  biçiminde bir bijektif fonksiyon (birebir ve örten bir fonksiyon) var ve  $f, f \times f : E \rightarrow E'$  şeklinde bir bijektif fonksiyona genişletiliyor ise bu graflar **izomorfik (isomorphic) graf** olarak adlandırılır.

**Tanım 3.7.**  $G = (V, E)$  yönlü graf (directed graph) olsun.  $(v, w) \in E$  özelliğini sağlayan hiç bir  $v' \in V$  elemanı yok ve  $(v, w) \in E$  olacak şekilde bir  $w \in V$  var ise  $v \in V$  düğüm noktası **kaynak (source) nokta** olarak adlandırılır. Eğer  $v, G^{op}$ 'da bir kaynak ise  $v \in V$  düğüm noktası **hedef (sink) nokta** olarak adlandırılır.

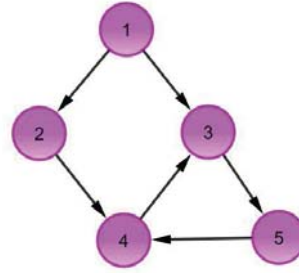


**Tanım 3.8.**  $G = (V, E)$  yönlü graf (directed graph) olsun. Her  $i = 0, \dots, n - 1$  için,  $(v_i, v_{i+1}) \in E$  ise düğüm noktalarının dizisi olan  $(v_0, \dots, v_n)$  **yol (path)** olarak adlandırılır. Eğer  $v_0 = v_n$  ise  $(v_0, \dots, v_n)$  yolu **döngü (cycle)** olarak ifade edilir.

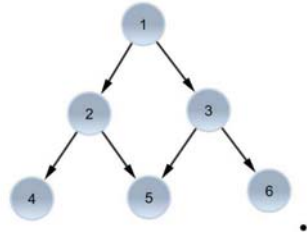
**Tanım 3.9.** Düğüm sayısı  $n$  en az üç olacak şekilde tasarlanan,  $n$  adet düğüm ve bu düğümlerin çiftlerinden oluşan kenarlardan meydana gelen graflara **Çember (Cycles) Graf** (bkz Şekil 3.4a), yönlü bir graf çember içeriyorsa **Yönlü Çember Graf** (bkz Şekil 3.4b), yönlü bir graf çember içermiyorsa **Yönlü Çembersiz Graf (Directed Acyclic Graph)** (bkz Şekil 3.4c) denir.



Şekil 3.4a Çember Graf



Şekil 3.4b Yönlü Çember Graf



Şekil 3.4c Yönlü Çembersiz Graf

### Şekil 3.4: Çember Graf Çeşitleri

**Tanım 3.10.** Düğüm kümesi, tüm kenarları  $V_1$ 'in bir düğümünü  $V_2$ 'nin bir düğümüne bağladığı  $\{V_1, V_2\}$  şeklinde bir parçalanmaya sahip olan graflara **iki parçalı graf (Bipartite Graph)** denilmektedir.

**Tanım 3.11.**  $V_1$ 'in tüm düğümlerini  $V_2$ 'nin tüm düğümlerine tek bir kenar ile bağlayan iki parçalı graflara, **tam iki parçalı graf (Complete Bipartite Graph)** denir.

Tam iki parçalı grafların özellikleri aşağıdaki gibidir:

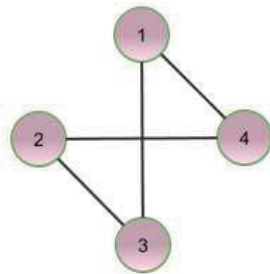
1. Bir tam iki parçalı graf  $K_{m,n}$  şeklinde gösterilir ve ilgili grafın düğümlerinin kümesi m ve n elemanlı iki alt kümeye ayrılır.
2. Bir kenarı birbirine bağlayan iki düğümü de farklı alt kümelerin elemanı olmak zorundadırlar.

**Tanım 3.12.**  $G = (V, E)$  bir graf olsun. Graf üzerindeki köşelerin sayısı  $n$ ,  $A$  matrisi  $n \times n$ 'lik bir matris,  $V(A) = \{n_1, n_2, n_3, \dots, n_n\}$  düğüm noktaları olsun. Eğer  $n_1$  köşesinden  $n_2$  köşesine bir kenar var ise, matriste  $n_1$ 'inci satır ile  $n_2$ 'inci sütunun olduğu eleman 1, aksi takdirde 0 olacak şekilde oluşturulan matrise **Bitişiklik Matrisi (Adjacency Matrix)** (bknz Şekil 3.5) denir.

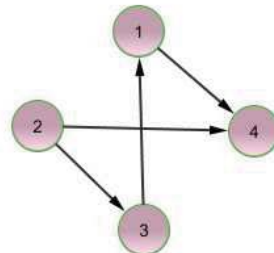
Bitişiklik matrisinin matematiksel ifadesi ise

$$\delta_{ij} = \begin{cases} 1, & (i, j) \notin E, \\ 0, & (i, j) \in E \end{cases}$$

şeklindedir.



Şekil 3.5a A



Şekil 3.5b B

**Şekil 3.5: Bitişiklik matrisleri (Adjacency Matrix)**

Şekil 3.5'deki graflara ait Bitişiklik Matrisleri aşağıdaki gibidir.

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

**Tanım 3.13.**  $G(V, E)$  grafindaki  $v$  düğüm noktası

$$G \setminus v := G(V \setminus \{v\}) \quad (3.3)$$

*bağlantısız ise eklem noktası (articulation point) olarak adlandırılır.*

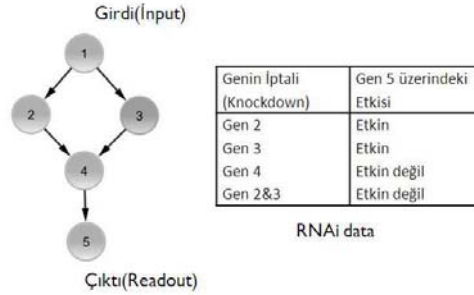
**Tanım 3.14.**  $V \setminus \{s, t\}$ 'nin büyüklüğü  $k$  olan her  $V'$  alt kümesi için,  $V \setminus V'$  üzerindeki  $G(V \setminus V')$  tam alt grafi hedef noktayı kaynak noktaya bağlayan bir yola sahip ise bu durumda  $G = (V, E; s, t)$  kabul edilebilir yönlü grafi  **$k$ -bağlantılı ( $k$ -connected)** yönlü graf olarak adlandırılır.

## 4. SİNYAL YOLAKLARININ OLUŞTURULMASI İÇİN SEZGİSEL ALGORİTMA İLE GRAF TEORİK YAKLAŞIM

Bu bölümde problemin tanımı, sinyal yollarını oluştururken graf teorik olarak nasıl yaklaşıldığı ve graf topolojileri oluştururken genetik algoritmanın nasıl kullanıldığına dair bilgi verilecektir.

### 4.1 PROBLEMİN TANIMI

Bu çalışmadaki problem Şekil 4.1’de örnek olarak verilen RNAi verisinden sinyal yollarını oluşturma problemidir (Kaderali 2009). Bu problem graf teorik yaklaşımla çözülmüştür. Oluşturulacak sinyal yolları graf topolojisi olarak tasarlanmıştır. Graf topolojileri oluşturulurken ise verilen RNAi verisinden faydalanılmıştır. RNAi verisinden alınan gen iptali sonucu etkin olarak işaretlenen genler bağlantılılığı bozan düğümler olarak düşünülerek graf topolojileri oluşturulmuştur.



Şekil 4.1: RNAi verisi ve oluşabilecek graf topolojisine örnek (Kaderali 2009).

Şekil 4.1’de verilen RNAi verisinden kimyasal tepkimede gen 4’ün iptal edilmesi ile tepkimenin durduğu, gen 2 ya da 3’ün iptal edilmesi ile kimyasal tepkimenin durmadığı, gen 2 ve gen 3’ün birlikte iptal edilmesi ile kimyasal tepkimenin durduğu belirtilmiştir.

Düğüm sayısı  $N$  olan bir graf için, düğümler arasında olası  $n \times n$  yönlü ayrıt vardır. Bu da olası bütün graf topolojilerinin sayısını  $2^{n \times n}$  yapar. Yani  $n = 5$  düğümlü bir graf için  $2^{5 \times 5} = 33.554.432$  yönlü graf oluşturulabilir. Bu sayı  $n$  büyüdükçe çok hızlı ve

üstel olarak arttığı için bütün olasılıkları tek tek test imkansızdır. Bu tezde, Bölüm 3’de tanıtılan GA ve alt graflar yardımı ile üstel artan karmaşıklık (*complexity*) azaltılmıştır.

## 4.2 GRAF TEORİK YAKLAŞIM

Deneysel veriye uyan (RNAi verisi) ve gerekli minimum bilgiyi içeren bütün grafların sayılabilmesi için zayıfça kabul edilebilir (*weakly admissibile*) graflar Tanım 4.1’deki gibi ifade edilir.

**Tanım 4.1.**  $G = (V, E)$  yönlü grafi aşağıdaki koşulları sağladığı takdirde zayıfça kabul edilebilir (*weakly admissibile*) olarak adlandırılır.

1.  $G$ ’de hiç bir döngü yoktur.
2. Her düğüm noktası hedef nokta ve kaynağa bir yol boyunca bağlıdır.
3.  $(a, b)$  ayrıtı  $G$  kaldırılırsa, oluşan yeni grafta ya kaynak nokta  $s$ ’den  $a$ ’ya bir yol yoktur ya da  $b$ ’den hedef nokta  $t$ ’ye hiç bir yol yoktur.

Bir  $G$  grafi zayıf kabul edilebilir olup, bir tek kaynak (*source*) ve bir tek hedef (*sink*) noktaya sahip ise kabul edilebilir olarak adlandırılır.  $V$  düğüm noktaları kümesindeki elemanların sayısı, kabul edilebilir yönlü grafin büyüklüğünü verir.

### 4.2.1 Eklemler Boyunca Kabul Edilebilir Yönlü Grafların Ayırıştırılması

Bu bölümde  $G = (V, E; s, t), a_1, \dots, a_k$  olacak şekilde  $k$  tane eklem noktası (*articulation point*) ve büyüklüğü  $n$  olan kabul edilebilir yönlü graf olduğu varsayılmıştır.

**Önerme 4.1.** Kaynak nokta  $s$ ’yi hedef nokta  $t$ ’ye bağlayan  $(s, v_1, \dots, v_m, t)$  biçimdeki bütün yollar eklem noktalarını içerir. Dahası,  $s$ ’yi  $t$ ’ye bağlayan her yol eklem noktasını aynı sırada içerir.

**İspat:**  $\alpha = (s, v_1, \dots, v_m, t)$ ,  $s$ ’den  $t$ ’ye bağlı bir yol olsun. Her bir  $i = 1, 2, \dots, m$  için  $v_i \neq a_j$  olacak şekilde  $a_j$  eklem noktası (*articulation point*) var olsun. Eklem noktası  $a_j$ ,  $G$  kabul edilebilir yönlü grafindan kaldırmak yolu etkilemeyecektir. Bu durum

$a_j$ 'nin bir eklem noktası olması gerçeğiyle çelişmektedir. Bütün eklem noktaları  $\alpha$ 'dan oluşmaktadır. Şimdi, farzedelim ki  $a$  ve  $a'$  olacak şekilde iki eklem noktası ile,  $\alpha = (s, v_1, \dots, v_m, t)$  ve  $\beta = (s, u_1, \dots, u_m, t)$  gibi iki yolun olduğu varsayılan  $(a, v_i, \dots, v_j, a')$  de,  $\alpha$  içinde  $a'$ 'nü  $a$ 'ya bağlayan ve  $(a', u_l, \dots, u_s, a)$ 'de  $\beta$  içinde  $a'$ 'yi  $a'$ 'ne bağlayan alt yollar olsunlar. Bu durumda  $G$  grafında, bir döngü elde edilmiş oldu ancak bu bir çelişkidir. Çünkü  $G$  kabul edilebilir yönlü grafında hiçbir çember olmayacak şekilde alınmıştır. Bu durumda  $s$ 'yi  $t$ 'ye bağlayan her yolda eklemeler özel bir sırada görünür.

**Tanım 4.2.**  $(a_1, \dots, a_k)$  eklem noktalarının doğal sıralarında görüldüğü sıralı bir dizi olsun. İlk olarak

$$U_0 = \{v \in V | s \text{ 'yi } v \text{ 'ye bağlayan eklem noktası olmayan bütün yollar}\}$$

tanımlayalım.

$1 \leq m \leq k$  için  $U_m \subseteq V$  olacak şekilde aşağıdaki biçimde tanımlansın:

$$U_m = \{v | s \text{ 'yi } v \text{ 'ye bağlayan } a_1, \dots, a_m \text{ eklem noktalarını içeren bütün yollar}\}$$

,

$U_0 \subseteq U_1 \subseteq \dots \subseteq U_k$  olduğu açıktır. Benzer şekilde  $1 \leq m \leq k - 1$  için

$$W_m = \{v | s \text{ 'yi } t \text{ 'ye bağlayan } a_{m+1}, \dots, a_k \text{ 'ya kadar olan eklem noktalarını içeren tüm yollar}\}$$

biçiminde tanımlarsak bu durumda

$$W_k = \{v | v \text{ 'yi } t \text{ 'ye bağlayan hiçbir eklem noktası içermeyen tüm yollar}\} \text{ 'yi tanımlarız.}$$

O halde  $W_0 \supset W_1 \supset \dots \supset W_k$  geçerli olur.

**Lemma 4.1.** *diyelim ki  $V_i = U_i \cap W_i, i = 0, \dots, k$  tanımlansın. Her  $i \neq j$  için  $V_i$  ve  $V_j$  kümelerinin kesişimi aşağıdaki gibi hesaplanır:*

$$V_i \cap V_j = \begin{cases} a_i + 1 & , \text{ eğer } j = i + 1, \\ 0 & , \text{ eğer } |i - j| > 1. \end{cases}$$

**İspat:** Genellemeyi kaybetmeksizin  $i < j$  olduğunu varsayalım. O zaman kesişim şu şekildedir:

$$V_i \cap V_j = U_i \cap W_i \cap U_j \cap W_j = U_j \cap W_i.$$

Çünkü,  $U_j \subset U_i$  ve  $W_i \subset W_j$ 'dir. Diyelim ki  $v \in V_i \cap V_j = U_j \cap W_i$  olsun. O halde  $v \in W_i$  ve sırasıyla  $a_{i+1}, \dots, a_j, \dots, a_k$  olan bütün eklem noktalarını,  $v$ 'yi  $t$ 'ye bağlayan bütün yolları ve sırasıyla  $a_1, \dots, a_j$ 'ye kadar olan eklem noktalarını ve  $s$ 'yi  $v$ 'ye bağlayan bütün yolları içerir. Fakat o zaman  $a_j$ 'den  $v$ 'ye bir yol olduğundan ve  $v$ 'den  $a_j$ 'ye diğer bir yol olduğundan bir döngü elde ederiz ki  $j = a + 1$  ve  $v = a_j$  olamaz ise bu bir çelişkidir.

**Lemma 4.2.** *Bazı  $i, j$ 'ler için  $u \in V_i$  ve  $v \in V_j$  olsun. O halde  $(u, v)$  bir ayırıt ise  $i = j$  ya da  $j = i + 1$  ve  $v = a_{i+1}$ 'dir.*

**İspat:** Diyelim ki  $i > j$  olsun. O halde  $s$ 'den  $u$ 'ya olan bütün yollar  $a_{j+1}, \dots, a_i, \dots, a_k$ 'ya kadar eklem noktaları içerdiğinden ve  $v$ 'den  $t$ 'ye olan bütün yollar eklem noktası içerdiğinden  $a_i$  üzerinde bir döngü elde ederiz. Bu ise bir çelişkidir. Varsayalım ki  $i < j$  ve  $\alpha = \{s, u_1, \dots, u_l, u\}$ ,  $a_1, \dots, a_i$ 'ye kadar eklem noktalarını içeren  $s$ 'den  $u$ 'ya bir yol,  $\beta = (v, v_1, \dots, v_m, t)$ ,  $a_{j+1}, \dots, a_k$ 'ya kadar eklem noktalarını içeren  $v$ 'yi  $t$ 'ye bağlayan bir yol olsun. Bu durumda

$$(s, u_1, \dots, u_l, u, v, v_1, \dots, v_m, t)$$

yolu  $a_i$  ile  $a_j$  arasındaki eklem noktalarını atlayan  $s$ 'den  $t$ 'ye bir yoldur ve  $j = i + 1$ ,  $v = a_i + 1$  olmaz ise bu bir çelişkidir.

**Önerme 4.2.** Kaynak nokta  $a_0 := s$  ve hedef nokta  $a_{k+1} := t$  olarak tanımlansın. Bu durumda her bir  $i = 0, \dots, k$  için  $G(V_i)$  tam alt grafi  $a_i$ 'nin tek kaynak,  $a_{i+1}$  tek hedef noktası olduğu kabul edilebilir yönlü 1-bağlantılı graftır.

**İspat:**  $w \in V_i$  olacak şekilde bir eleman alalım.  $v$ 'yi  $s$ 'ye bağlayan bütün yollar  $a_i$ 'den geçtiğinden  $v = a_i$  olmaz ise  $v$  elemanı bir kaynak noktası olamaz. Benzer şekilde  $v$ 'yi  $t$ 'ye bağlayan bütün yollar  $v \in V_i$  olduğunda  $a_{i+1}$ 'den geçmek zorundadır. Öyleyse  $v = a_{i+1}$  olmaz ise  $v$  bir hedef noktası olamaz.  $G(V_i)$ ,  $a \in V_i$  eklem noktasına sahip olsun. O halde bu eklem noktası  $G$ 'nin de bir eklem noktası ve Önerme 4.1'den  $a$ , sırasıyla  $a_i$  ile  $a_{i+1}$  arasında bulunmak zorundadır. Bu da bir çelişkidir. Bu durumda  $G(V_i)$  1-bağlantılı kabul edilebilir yönlü graftır.

### 4.3 SİNYAL YOLAKLARININ OLUŞTURULMASI İÇİN SEZGİSEL ALGORİTMA İLE GRAF TEORİK YAKLAŞIM

Bu çalışmada sinyal yolları oluşturulurken Bölüm 4.2'de anlatıldığı üzere bir takım grafsal yaklaşımlarda bulunulmuştur. İstenilen "zayıf kabul edilebilir" (Tanım 4.1) ve minimal graf koşuluna uyacak grafların listesini çıkarmaktır. Eklem noktası (*articulation point*) ve zayıf kabul edilebilir graf koşullarından faydalanarak grafları Şekil 4.2'deki gibi eklem noktalarından parçalayabiliriz. Buradaki her bir graf, 1-bağlantılı (*1-connected*) graftır ve içerisinde eklem noktası bulunmamaktadır. Kısaca ifade edilecek olursa, eğer  $G$  grafi  $n$  noktalı zayıf kabul edilebilir bir graf ise  $s$ 'den  $t$ 'ye giden bütün yollar eklem noktalarını içerir ve  $k + 1$  tane grafa bölünebilir.



**Şekil 4.2: Grafların parçalanması**

Problem, grafları parçaladıktan sonra minimal graflar listesini çıkarmaktan minimal 1-bağlantılı grafların listesini alıp parçalamaları kullanarak grafları yanyana dizmek problemine dönüşmektedir. Bir başka deyişle içinde eklem noktası olmayan grafların listesini çıkarmaya dönüşmektedir.



**1-bağlantılı grafların listesini kullanarak temel problemin çözümü:**

$n = 10$  düğümlü,  $k = 2$  eklem noktalı graf topolojilerini çıkarmak isteyelim. Öncelikle  $12 = n_1 + n_2 + n_3$  olacak şekilde yazılabilecek bütün  $n_1, n_2, n_3$  üçlüleri yazılır. Örnek olarak  $(3, 3, 6)$ ,  $(5, 5, 2)$  üçlüleri verilebilir.  $n_1$  noktalı,  $n_2$  noktalı,  $n_3$  noktalı 1-bağlantılı grafların listesi alınır ve teker teker yan yana eklenerek graf topolojileri oluşturulabilir.

---

**Algorithm 1** Parçalama Algoritması

---

**function** Par( $n+k, k+1$ )**for**  $(n_0, \dots, n_k) \in Par(n+k, k+1)$  ( $n_0 + \dots + n_k = n+k$ )**for**  $(n_0, \dots, n_k) \in G(n_0) \times \dots \times G(n_k)$  $G(n_i) = x_i$  (1-bağlantılı  $n$  noktalı grafların sayısı)Listeye  $x_0 * \dots * x_k$  grafini ekle**end****end**

---

Algoritma 1’de  $n$  düğüm sayısı,  $k$  eklem noktası olacak şekilde  $(n+k)$  sayısının  $(k+1)$ ’li parçalanışlarının sayıları kullanılarak bu sayılara karşılık 1-bağlantılı grafların listesi çıkarılmaktadır. Daha sonrasında bu graflar teker teker birleştirilmektedir. Algoritma 1 büyük parçaları küçük parçalara nasıl parçalayabileceğimizi söylemektedir. Fakat 1-bağlantılı grafların listesinin nasıl oluşturulacağını söylememektedir. Listeleri çıkardıktan sonra  $G(n_0), G(n_1), \dots, G(n_k)$  graflarını oluşturmak için genetik algoritma kullanılmaktadır.

1-bağlantılı graf için GA’ya (Algoritma 2) boş bir liste ile başlanılmakta ve algoritma iterasyon sayısı kadar çalıştırılmaktadır.

İlk başta her iterasyon rastgele bir bitişiklik matrisi oluşturmaktadır. Artan düğümlerin etiketlerini nümerik değerlere diyagonal altı sıfır olacak şekilde ve yönlü çembersiz graf özelliğinden diyagonal sıfır olacak şekilde rastgele bitişiklik matrisi oluşturulur. Bu rastgele matrisi oluştururken  $p = 1 - \frac{1}{1 + \sqrt{\frac{n-2-k}{n-2}}}$  (Ayaz, 2011) olmak üzere  $i$ .inci düğüm ile  $j$ .inci düğüm arasında ayrıt olma olasılığı  $p^{j-i}$  kullanılarak belirlenen bir eşik değeri ile  $p^{j-i}$  eşik değerinin altında ve üstünde olmasına göre 0 ve 1’den oluşan  $n \times n$ ’lik matriste güncellenir (Algoritma 3).

---

**Algorithm 2** 1 bağlantılı graf için GA

---

**List**  $\leftarrow \emptyset$

**for**  $i=1$  to number of generation

$G \leftarrow$  random upper triangular matrix

**Repeat**

*Mutation* $G$

**Until**  $G$  is 1-connected

**if**  $G \notin$  List then add  $G$  to List

**end for**

**return** List

---

Oluşan matrise 1-bağlantılı graf olana kadar mutasyon operatörü uygulanır. 1-bağlantılı altgraflar birleştirilerek istenilen graf topolojisi elde edilir. Jenerasyon sayısı, başka bir deyişle GA birden fazla uygulandığında elde edilen graf topolojilerinin izomorfik olup olmadıklarına karar verilerek çözüm kümesine izomorfik olmayanlar eklenir.

---

**Algorithm 3** Bitişiklik matrisi oluşturma

---

**Let**  $p = 1 - \frac{1}{1 + \sqrt{\frac{n-2-k}{n-2}}}$

**for**  $i=1$  to  $n$

**for**  $j=i+1$  to  $n$

**if**  $p^{j-i} \geq \text{Eşik Değeri}$

**then**  $A(i, j) = 1$

**else**  $A(i, j) = 0$

**end**

**end**

**end**

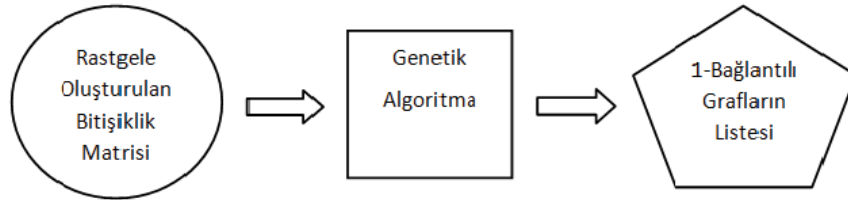
---

## 5. TARTIŞMA VE SONUÇ

RNAi verileri ile sinyal yollarının oluşturulması çözüm uzayı çok büyük olan bir problemdir. Bu tezde graf teori ile graflar matris olarak ifade edilmiştir. 1-bağlantılılığı bozan noktalardan grafları parçalayıp 1-bağlantılı grafların listesi Genetik Algoritma ile oluşturulmuştur. 1-bağlantılı graflar birleştirilerek istenilen sinyal yollarının çözüm uzayı daraltılmıştır.

1-bağlantılı grafların listesini elde etmek için gerekli olan akış şeması Şekil 5.1’de verilmiştir.

Literatürde benzer çalışma Ayaz (2011)’de verilmiştir. Ayaz (2011),  $n$  düğüm sayısı,  $k$  eklem noktası verilen grafin topolojisini GA ile oluşturmuştur. GA kullanırken uygunluk fonksiyonu oluşturulan rastgele bitişiklik matrisinin  $k$  tane eklem noktasına yakınsamasını amaçlamaktadır. Bu tezde önerilen algoritmanın en büyük farkı ve avantajı grafların 1-bağlantılı graflara parçalanarak Ayaz (2011)’de önerilen yöntemden daha hızlı olması ve çözüm uzayını daha da daraltılmasıdır.



**Şekil 5.1: Algoritma akış şeması**

Ayaz algoritması (iterasyon başına 1/3 sn) ve bu tez çalışmadaki algoritma düğüm sayısı 8, eklem noktası 1, 2, 3, 4 olacak şekilde 100 iterasyon çalıştırılarak oluşturulan topoloji sayıları Tablo 5.1’de karşılaştırılmıştır.

Yazdığımız algoritmanın 50 iterasyon için oluşacak topoloji sayıları ve zamanı Tablo 5.2’de verilmiştir.

**Tablo 5.1: 100 iterasyon için graf topoloji sayıları**

	<b>Parçalama olmadan (Ayaz, 2011)</b>	<b>Geliştirilen Algoritma</b>
n=8, k=1	386.118	188 (6sn)
n=8, k=2	10.077	572 (15sn)
n=8, k=3	500	1105 (30sn)
n=8, k=4	45	473 (45sn)

**Tablo 5.2: 50 iterasyon için graf topoloji sayıları**

	<b>Graf sayısı</b>	<b>Çalışma Süresi(Sn.)</b>
n=8, k=1	94	1.873483
n=8, k=2	572	5.440001
n=8, k=3	543	7.500265
n=8, k=4	232	33.973656

Bölüm 4'te anlatıldığı gibi olası bütün grafları ele alıp aralarından uygun olanları seçmek zordur. Kullandığımız yöntem aradığımız çözüm uzayını oldukça daraltıp çözüme daha çabuk ulaşmamızı sağlamaktadır. Öte yandan kullanılan Genetik Algoritma sezgisel bir algortima olduğundan bize tam sonuç vermeyi garanti etmemektedir. Dolayısıyla oluşturulmak istediğimiz 1-bağlantılı graf topolojilerinin tam listesini elde edemeyebiliriz. Ancak graf listesini çok kısa bir sürede verdiği için zamandan kazancımız büyük olmaktadır.

## 6. GELECEK ÇALIŞMALAR

Bölüm 5’de anlatıldığı gibi noktalar arasındaki bağlantılar sadece teorik olarak belirlenmeye çalışılırsa çözüm uzayımız üstel olarak büyüdüğünden bütün çözümleri düşünmek hesapsal olarak oldukça zorlaşır. Bu konuda olası bir çözüm düğümler arası bağlantıların rastgele değil literatürde metin madenciliği (*text mining*) yapılarak belirlenmesi olacaktır. Başka bir gelecek çalışmada da kullanılan GA metodunun geliştirilmesi ya da GA yerine başka metotlar kullanılması olacaktır. Tam sayılı programlama ya da kesin kombinatorik metodların geliştirilmesi gelecek çalışmalara bırakılmıştır.

## KAYNAKÇA

### *Kitaplar*

- Ayaz E.S., 2011. *Reconstructing Signaling Pathways From RNAi Data Using Genetic Algorithms*, Metu, Master Thesis.
- Braysy O., 2001. *Local Search and Variable Neighborhood Search Algorithms for The Vehicle Routing Problem With Time windows*, PHD Thesis.
- Engin, O., 2001. *Akış Tipi Çizelgeleme Problemlerinin Genetik Algoritma ile Çözüm Performansının Arttırılmasında Parametre Optimizasyonu*, Fen Bilimleri Enstitüsü, Doktora Tezi.
- J. G. Michaels and K. H. Rosen, 1991. *Applications of Discrete Mathematics*, McGraw-Hill.
- Goldberg, D.E., 1989. *Genetic Algorithms in search, Optimization and Machine Learning*, Addison Wesley.
- Gross J., Yellen J., 1999. *Graph Theory and Its Applications*.
- Mitchell, M., 1989. *Introduction To Genetic Algorithms*, Bradford Press, London,10-48.
- West, D. B., 2001. *Introduction to Graph Theory*.

### ***Sürekli Yayınlar***

- Aldenderfer, M. S. & Blashfield, R. K., 1984. Cluster Analysis, *Sage University Paper series on Quantitative Applications in the Social Science*, series no 07 – 044.
- Agrawal N., et al., 2003. RNA Interference: Biology, Mechanism, and Applications, *Microbiology and Molecular Biology Reviews*, **67** (4), pp: 657- 685.
- Atalay A., 2007. RNA İnterferans Kursu, s. 25, *Ankara Üniversitesi Biyoteknoloji Enstitüsü Yaşam Bilimleri Kursları Serisi*, Ankara.
- Bernstein E., et al., 2001 Role For A Bidentate Ribonuclease in The Initiation Step of RNA Interference, *Nature*, 409, pp: 363- 366.
- Downward J., 2004 RNAi and Cancer Research, *The Biochemist*, **26** (5), pp: 12- 14.
- Gündoğdu R., Çelik V., 2009 RNA interferans, 2009, *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi* **25** (1-2) ss: 34 – 47.
- Fafali, P., 2003. Patrikakis, C., Michalas, A., Loumos, V., Internet Traffic Engineering: History Monitoring Information Featuring Routing Algorithms, *Automatics and Informatics Conference*, Sofia Bulgaria, pp: 46-51.
- Fiğlai A., 2002 Engin O. Genetik Algoritmalarla Akış Tipi Çizelgelemede Üreme Yöntemi Optimizasyonu, *İTÜ Dergisi*, ss:1-6.
- Filipowicz W., 2005. RNAi: The Nuts andBolts of The RISC Machine, *Cell*, 122, pp: 17- 20.



- Gewirtz A.M., 2007. On Futures Doorstep: RNA Interference and The Pharmacopeia of Tomorrow, *The Journal of Clinical Investigation*, **117** (12), pp: 3612- 3614.
- Jang J. S. R., 1998 Neuro-Fuzzy and Soft Computing: *A Computational Approach to Learning and Machine Intelligence*, Cahpter 7: Derivative-Free Optimization, Prentice-Hall, USA, pp: 173-196.
- Hannon G.J., 2002. RNA Interference, *Nature*, 418, pp: 244- 251.
- Kaderali L., 2009, Dazert E., Zeuge U., Frese M., Bartenschlager R. Reconstructing Signaling Pathways from RNAi Data Using Probabilistic Boolean Threshold Networks. *Bioinformatics*, **25**(17), pp: 2229-2235.
- Kurreck J., 2003. Antisense Technologies Improvement Through Novel Chemical Modifications, *Europea Journal of Biochemistry*, 270, pp:1628-1644.
- Whitley, L. D. 1994, A Genetic Algoritihm Tutorial , *Statistical Computing*. **4**,pp: 65-85.
- Yeniay Ö., 2002. An Overview of Genetic Algorithms, *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, **2(1)**, ss:. 37-49.

### ***Diğer Yayınlar***

Cell Signaling Technology, Signaling Pathways and Tables,

<http://www.cellsignal.com/reference/pathway/index.html>

Pathway Studio, Protein Tyrosine Phosphatase Receptor Signaling Pathway,

<http://www.ariadnegenomics.com/products/pathway-studio/sample-pathways/ptprc-signaling/>