

T.C.
GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ

AĞYAPI ÇIKARIMI TEKNİKLERİNİN
METABOLOM VERİLERİNE UYGULANARAK
HÜCRENİN BİYOLOJİK AMACININ
İNCELENMESİ

Melik ÖKSÜZ
YÜKSEK LİSANS TEZİ
KİMYA MÜHENDİSLİĞİ
ANABİLİM DALI

GEBZE

2012

T.C.
GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ

AĞYAPI ÇIKARIMI TEKNİKLERİNİN
METABOLOM VERİLERİNE UYGULANARAK
HÜCRENİN BİYOLOJİK AMACININ
İNCELENMESİ

Melik ÖKSÜZ
YÜKSEK LİSANS TEZİ
KİMYA MÜHENDİSLİĞİ
ANABİLİM DALI

TEZ DANIŞMANI
Doç. Dr. Hasan SADIKOĞLU
Yrd. Doç. Dr. Tunahan ÇAKIR

GEBZE

2012



YÜKSEK LİSANS TEZİ JÜRİ ONAY SAYFASI

G.Y.T.E. Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 15.06.2012 tarih ve 2012/22... sayılı kararıyla oluşturulan jüri tarafından 07.09.2012 tarihinde tez savunma sınavı yapılan Melik ÖKSÜZ'ün tez çalışması Kimya Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Doç. Dr. Hasan SADIKOĞLU

ÜYE

(2. TEZ DANIŞMANI) : Yrd. Doç. Dr. Tunahan ÇAKIR

ÜYE

: Doç. Dr. Mustafa NADAR

ÜYE

: Doç. Dr. Fatih Erdoğan SEVİLGİN

ÜYE

: Yrd. Doç. Dr. Mahmut ŞEKER

ONAY

G.Y.T.E. Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../20... tarih ve/..... sayılı kararı.

İMZA/MÜHÜR

ÖZET

TEZİN BAŞLIĞI: AĞYAPI ÇIKARIMI TEKNİKLERİNİN METABOLOM VERİLERİNE UYGULANARAK HÜCRENİN BİYOLOJİK AMACININ İNCELENMESİ

YAZARIN ADI: MELİK ÖKSÜZ

Yapılan çalışmada amaç, yukarıdan aşağı hesaplamalı sistem biyolojisi yöntemleri kullanılarak metabolik ağyapıya şekil veren biyolojik amacın belirlenmesidir. Bu amaç için, *Escherichia coli* ve *Saccharomyces cerevisiae* olmak üzere iki farklı mikroorganizma kullanıldı. Farklı biyolojik değişkenlikler için literatürden bu iki mikroorganizmaya ait dinamik modeller kullanılarak *in silico* deneysel veriler üretildi. *In silico* üretilen deneysel verilere yönlü ve yönsüz ağyapı çıkarımı için üç farklı yaklaşım uygulandı: istatistik-bazlı, optimizasyon-bazlı ve nedensellik bazlı. Gerçekleştirilen çalışmalarda, biyolojik amaç olarak minimum enzim üretimi (biyolojik ağyapıların seyrek aralıklı yapıda olması) incelenmiş ve çalışmalar bu doğrultuda gerçekleştirilmiştir. Yapılan çalışmaların bir sonucu olarak, Graphical Gaussian Model'in (GGM) yönsüz metabolik ağyapı çıkarımında en başarılı yöntem olduğu bulunmuştur. Seyrek aralıklı bir ağyapı üreten GGM metodu hücrenin bir amacının da minimum enzim üretimi olduğunu kanıtlamıştır. Yönlü ağyapı çıkarımı için iyi sonuçlar elde edilmemiştir. Ancak, ortaya çıkan problemlere odaklanılmış ve bu problemlerin üstesinden gelmek amacıyla yeni bir yöntem önerilmiştir. Genetik algoritma tabanlı önerilen yöntem ile diğer yöntemlere göre daha iyi sonuçlar elde edilmesine rağmen genel olarak yönlü ağyapı çıkarımı için iyi sonuçlar elde edilemedi. Bu tez çalışması metabolik ağyapı çıkarımının kapsamlı bir analizini temsil etmektedir. Metabolom verilerine dayalı ağyapı çıkarımında yönsüz yöntemlerin daha başarılı olduğu, bu çalışma sayesinde gösterilmiştir.

SUMMARY

THESIS TITLE: INVESTIGATION OF THE BIOLOGICAL OBJECTIVE OF THE CELL BY APPLYING NETWORK INFERENCE TECHNIQUES TO METABOLOME DATA

AUTHOR: MELİK ÖKSÜZ

The aim of this study is the determination of biological objectives of the cell which gives shape to metabolic networks by using ‘top-down’ computational systems biology methods. For this purpose, two different microorganisms including *Escherichia coli* and *Saccharomyces cerevisiae* were used. *In silico* experimental data were generated for different biological variability by simulating dynamic models from literature belonging to these two microorganisms. Three different approaches were employed to infer directed and undirected metabolic networks from data: statistics-based, optimization-based, and causality-based. The tendency of cell metabolism for minimal enzyme production as a biological objective was investigated in detail. As a result, Graphical Gaussian Model (GGM), a statistics-based network inference method, was found to be the most successful method for undirected metabolic network inference. GGM method generated a sparse network, which indicated that minimum enzyme production is a biological objective of cells. Results obtained for directed network inference were not promising. However, focusing on the reasons behind, a new method was proposed in order to improve directed network inference methods. The proposed method based on genetic algorithm gave better results than other methods. This thesis study presents a comprehensive analysis of metabolic network inference. It was shown through this thesis work that best methods for network inference based on metabolome data are undirected methods.

TEŞEKKÜRLER

Tez çalışmam boyunca benden ilgisini, bilgisini ve hoşgörüsünü eksik etmeyen değerli hocam Yrd. Doç. Dr. Tunahan ÇAKIR'a ve bu tez çalışmasında bana çalışma fırsatı veren değerli hocam Doç Dr. Hasan SADIKOĞLU'na tüm samimiyetimle teşekkür ederim.

Tezin nihai şeklini almasında katkıda bulunan jürideki tüm saygıdeğer hocalarıma ayrı ayrı şükranlarımı sunarım. Ayrıca bu çalışmayı 110M464 nolu projeyle maddi olarak destekleyen TÜBİTAK'a da teşekkürlerimi bir borç bilirim.

SİMGELER VE KISALTMALAR

k	Hız Sabitleri
R	Korelasyon Katsayısı
GGM	Graphical Gaussian Model
p-value	Tahmin Değeri
u-test	Ranksum Test
Γ	Kovaryans Matrisi
f	Spesifik Ağırlık
D	Dalgalanma Matrisi
DAG	Yönlü Döngüsüz Ağyapı
r	Tepkime Hızı
GES	Eşitlik Arama Algoritması
SDE	Stokastik Diferansiyel Denklem
J	Jacobian Matrisi
j	Jacobian Vektörü
TPR	Doğru Pozitif Oran
FPR	Yanlış Pozitif Oranı
PC	Pearson Korelasyonu
PPC1	Birinci Dereceden Kısmi Korelasyon
PPC2	İkinci Dereceden Kısmi Korelasyon
CCD	Döngülü Nedensel Keşif Algoritması
GES	Açgözlü Eşitlik Arama Algoritması
E	Enzim Konsantrasyonu

İÇİNDEKİLER DİZİNİ

	<u>Sayfa</u>
ÖZET	iv
SUMMARY	v
TEŞEKKÜRLER	vi
SİMGELER VE KISALTMALAR	vii
İÇİNDEKİLER DİZİNİ	viii
ŞEKİLLER DİZİNİ	xii
TABLolar DİZİNİ	xiv
1 GİRİŞ	1
2 KURAMSAL TEMELLER	3
2.1 Deneysel Sistem Biyolojisi	3
2.1.1 Sistem Biyolojisinde <i>In vivo</i> Deneysel Veri Üretimi	4
2.1.2 Sistem Biyolojisinde <i>In silico</i> Deneysel Veri Üretimi	5
2.1.3 Sistem Biyolojisi Verilerinde Biyolojik Değişkenlik	6
2.2 Hesaplamalı Sistem Biyolojisi	9
2.2.1 Aşağıdan Yukarı Yaklaşım	9
2.2.2 Yukarıdan Aşağıya Yaklaşım	10
2.3 Metabolik Ağyapıların Yapısı	11
2.4 İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımı	12
2.4.1 Temel Kavramlar	12
2.4.2 Korelasyon	13
2.4.3 Kısmi Korelasyon Katsayısı	14
2.4.4 Grafiksel Gaus Model (GGM)	15
2.4.5 Bazı İstatistik Terimlerinin Açıklanması	16
2.4.6 Anlam Testleri	17
2.4.6.1 Student t-test	17
2.4.6.2 Ranksum test (u-test)	17
2.4.6.3 Rastgele Sıralama Testi (‘Randomization Test’)	18
2.4.7 Korelasyonun anlamlandırılması	18
2.5 Optimizasyona Dayalı Yönlü Ağyapı Çıkarımı	18
2.5.1 Temel Kavramlar	19

2.5.2 Doğrusal Programlama	21
2.5.3 Karesel Programlama	21
2.5.4 Karma Tamsayılı Doğrusal Programlama	22
2.5.5 Genetik Algoritma	22
2.5.6 Lyapunov Denklemi	27
2.6 Nedenselliğe Dayalı Ağyapı Çıkarımı	30
2.6.1 Temel Kavramlar	30
2.6.2 PC Algoritması	34
2.6.3 Döngülü Nedensel Keşif ('Cyclic Causal Discovery', CCD) Algoritması	34
2.6.4 Ağgözlü Eşitlik Arama ('Greedy Equivalent Search', GES) Algoritması	35
3 METOT	36
3.1 <i>In silico</i> Metabolom Verilerinin Üretilmesi	37
3.1.1 Kullanılan Dinamik Modeller	37
3.1.2 Enzimatik değişkenliğin modellenmesi ile Metabolom Verilerinin Üretilmesi	42
3.1.3 İçsel Değişkenliğin Modellenmesi ile Metabolom Verilerinin Üretilmesi	43
3.1.4 Hem İçsel Hem de Çevresel Değişkenliğin Birlikte Modellenmesi ile Metabolom Verilerinin Üretilmesi	45
3.2 İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımı Tekniklerinin <i>In Silico</i> Metabolom Verilerine Uygulanması	46
3.2.1 Pearson Korelasyon Yönteminin <i>In silico</i> Metabolom Verilerine Uygulanması	47
3.2.2 Birinci Dereceden Kısmi Korelasyon Yönteminin <i>In silico</i> Metabolom Verilerine Uygulanması	47
3.2.3 İkinci Dereceden Kısmi Korelasyon Yönteminin <i>In silico</i> Metabolom Verilerine Uygulanması	48
3.2.4 (n-2)'nci dereceden Kısmi Korelasyon Yönteminin <i>In Silico</i> Metabolom Verilerine Uygulanması	49
3.3 Optimizasyona Dayalı Yönlü Ağyapı Çıkarımı Tekniklerinin <i>In Silico</i> Metabolom Verilerine Uygulanması	49
3.3.1 Karesel Programlama Kullanılarak Yönlü Ağyapı Çıkarımı	51

3.3.2 Doğrusal Programlama Kullanılarak Yönlü Ağyapı Çıkarımı	51
3.3.3 Karma Tamsayılı Doğrusal Programlama Kullanılarak Yönlü Ağyapı Çıkarımı	52
3.3.4 Genetik Algoritma Tabanlı Önerilen Yöntem	54
3.4 Nedenselliğe Dayalı Yönlü Ağyapı Çıkarımı Tekniklerinin <i>In Silico</i> Metabolom Verilerine Uygulanması	58
3.5 Elde Edilen Sonuçların Değerlendirilmesi	58
4 <i>Saccharomyces cerevisiae</i> MİKROORGANİZMASI İÇİN YAPILAN ÇALIŞMALAR	60
4.1 İstatistiksel Ağyapı Çıkarımı Tekniklerinin <i>Saccharomyces cerevisiae</i> Mikroorganizmasına ait <i>In Silico</i> Metabolom Verilerine Uygulanması	60
4.2 Optimizasyona Dayalı Ağyapı Çıkarımı Tekniklerinin <i>Saccharomyces cerevisiae</i> Mikroorganizmasına ait <i>In Silico</i> Metabolom Verilerine Uygulanması	65
4.2.1 Karesel Programlama ile Yönlü Ağyapı Çıkarımı	66
4.2.2 Doğrusal Programlama ile Yönlü Ağyapı Çıkarımı	66
4.2.3 Karesel ve Doğrusal Programlama Sonuçlarının İyileştirilmesi	67
4.2.4 Karma Tamsayılı Programlama ile Yönlü Ağyapı Çıkarımı	69
4.2.5 Önerilen Genetik Algoritma Tabanlı ile Yönlü Ağyapı Çıkarımı	70
4.3 Nedenselliğe Dayalı Ağyapı Çıkarımı Tekniklerinin <i>Saccharomyces cerevisiae</i> Mikroorganizmasına ait <i>In Silico</i> Metabolom Verilerine Uygulanması	73
4.3.1 PC Algoritmasının Uygulanması	74
4.3.2 CCD Algoritmasının Uygulanması	74
4.3.3 GES Algoritmasının Uygulanması	75
5 <i>Escherichia coli</i> MİKROORGANİZMASI İÇİN YAPILAN ÇALIŞMALAR	76
5.1 İstatistiksel Ağyapı Çıkarımı Tekniklerinin <i>Escherichia coli</i> Mikroorganizmasına ait <i>In Silico</i> Metabolom Verilerine Uygulanması	76
5.2 Optimizasyona Dayalı Ağyapı Çıkarımı Tekniklerinin <i>Escherichia coli</i> Mikroorganizmasına ait <i>In Silico</i> Metabolom Verilerine Uygulanması ile Yönlü Ağyapı Çıkarımı	79
5.2.1 Karesel Programlama ile Yönlü Ağyapı Çıkarımı	79
5.2.2 Doğrusal Programlama ile Yönlü Ağyapı Çıkarımı	80

5.2.3	Karesel ve Doğrusal Programlama Sonuçlarının İyileştirilmesi	80
5.2.4	Önerilen Genetik Algoritma Tabanlı ile Yönlü Ağyapı Çıkarımı	81
5.3	Nedenselliğe Dayalı Ağyapı Çıkarımı Tekniklerinin <i>Escherichia coli</i> Mikroorganizmasına ait <i>In Silico</i> Metabolom Verilere Uygulanması	83
5.3.1	PC Algoritmasının Uygulanması	84
5.3.2	CCD Algoritmasının Uygulanması	84
5.3.3	GES Algoritmasının Uygulanması	85
6	TARTIŞMA VE YORUM	86
6.1	İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımı Sonuçlarının Değerlendirilmesi	87
6.2	Optimizasyona Dayalı Yönlü Ağyapı Çıkarımı Sonuçlarının Değerlendirilmesi	91
6.3	Nedenselliğe Dayalı Yönlü Ağyapı Çıkarımı Sonuçlarının Değerlendirilmesi	94
7	SONUÇ	96
8	ÖNERİLER	97
	KAYNAKLAR	98
	ÖZGEÇMİŞ	101
	EKLER	
	EK 1: <i>S. cerevisiae</i> MİKROORGANİZMASINA AİT DİNAMİK MODEL	
	EK 2: <i>E. coli</i> MİKROORGANİZMASINA AİT DİNAMİK MODEL	
	EK 3: RASTGELE SIRALAMA TEST FONKSİYONU	
	EK 4: <i>M.CLEAN</i> MODELİ İÇİN ÖNERİLEN YÖNTEM İLE YAPILAN HESAPLAMALAR	
	EK 5: <i>S.cerevisiae</i> MİKROORGANİZMASI İÇİN ÖNERİLEN YÖNTEM İLE YAPILAN HESAPLAMALAR	
	EK 6: <i>E.coli</i> MİKROORGANİZMASI İÇİN ÖNERİLEN YÖNTEM İLE YAPILAN HESAPLAMALAR	

ŞEKİLLER DİZİNİ

<u>Sekil</u>		<u>Sayfa</u>
2.1	Sistem Biyolojisi Yaklaşımının Şematik Gösterimi.	3
2.2	Metabolizma ve DNA Arasındaki Hiyerarşik İlişkinin Omik Ölçümleriyle Bağlantısı.	4
2.3	<i>In silico</i> Deneysel Veri Üretiminin Şematik Olarak Gösterimi.	6
2.4	Enzimatik Değişkenliğin Gösterilmesi.	7
2.5	İçsel Değişkenliğin Gösterilmesi.	8
2.6	Çevresel Değişkenliğin Gösterilmesi.	8
2.7	Hesaplamalı Sistem Biyolojisi Yaklaşımının Şematik Olarak Gösterimi.	9
2.8	Yukarıdan Aşağıya Sistem Biyolojisi Yaklaşımında Yönlü ve Yönsüz Ağyapı Çıkarımının Şematik Olarak Gösterilmesi.	10
2.9	Yoğun ve Seyrek Aralıklı Ağyapı Örneklerinin Şematik Gösterimi.	11
2.10	Normal ve Normal Dışı Dağılım Örnekleri.	13
2.11	Birinci Dereceden Kısmi Korelasyonun şematik olarak gösterilmesi.	15
2.12	Bir sistemin optimizasyon için tanımlanması örneğinin şematik gösterimi.	20
2.13	Farklı x değerleri için bir $f(x)$ fonksiyonun değerlerinin grafiksel olarak gösterimi.	23
2.14	Genetik Algoritma Basamakları.	23
2.15	Rastgele İkili Sayı Üretilerek Oluşturulan İlk Popülasyona Bir Örnek.	24
2.16	Tek Nokta Çapraz Değişime Örnek. Popülasyonda yer alan iki kromozom arasında rastgele bir nokta belirlenerek çapraz değişim yapılması.	25
2.17	Çift Nokta Çapraz Değişime Örnek; Popülasyonda yer alan iki kromozom arasında rastgele iki nokta belirlenerek çapraz değişim yapılması.	25
2.18	Genetik Algoritmanın Akış Şeması.	26
2.19	Örnek Bir Jacobian Matristen Yönlü Ağyapı Çıkarımının Şematik Gösterimi.	29

<u>Sekil</u>		<u>Sayfa</u>
2.20	A,B ve C deęiřkenlerinden oluřan bir ynsz aęyapı rneęinin alabileceęi ynl aęyapı ihtimalleri.	30
2.21	Nedensellięe Dayalı Ynl Aęyapı ıkarımının Genel Bir Algoritması.	31
2.22	Meek'in Oryantasyon Kuralı 1'in řematik Olarak Gsterimi.	32
2.23	Meek'in Oryantasyon Kuralı 2'nin řematik Olarak Gsterimi.	32
2.24	Meek'in Oryantasyon Kuralı 3'n řematik Olarak Gsterimi.	33
2.25	Meek'in Oryantasyon Kuralı 4'n řematik Olarak Gsterimi.	33
3.1	Yapılan Tez alıřmasında İzlenen Yolun řematik Olarak Gsterimi.	36
3.2	Teusink ve ark. tarafından geliřtirilen <i>Saccharomyces cerevisiae</i> mikroorganizmasına ait metabolik aęyapı.	38
3.3	Chassagnole ve ark. tarafından geliřtirilen <i>Escherichia coli</i> mikroorganizmasına ait metabolik aęyapı.	40
3.4	Enzimatik Deęiřkenlięin Simulasyon rneęi.	43
3.5	İsel Deęiřkenlięin Simulasyon rneęi.	44
3.6	İsel-evresel Deęiřkenlięin Simulasyon rneęi.	46
3.7	İstatistięe Dayalı Ynsz Aęyapı ıkarımında Kullanılan Yntemler.	47
3.8	Optimizasyona Dayalı Ynl Aęyapı ıkarımında Kullanılan Yntemler.	50
3.9	Nedensellięe Dayalı Ynl Aęyapı ıkarımında Kullanılan Yntemler.	58
6.1	<i>S.cerevisiae</i> ve <i>E.coli</i> mikroorganizmaları sonularına ait ROC eęrisi.	87
6.2	Faklı veri sayıları iin <i>S.cerevisiae</i> sonularına ait eęriler.	90

TABLolar DİZİNİ

<u>Tablo</u>	<u>Sayfa</u>
3.1 Kullanılan <i>Saccharomyces cerevisiae</i> mikroorganizmasına ait modelde yer alan 13 metabolitin isimleri ve kararlı hal konsantrasyonunun milimolar (mM) cinsinden değerleri.	39
3.2 Kullanılan <i>Escherichia coli</i> mikroorganizmasına ait modelde yer alan 18 metabolitin isimleri ve kararlı hal konsantrasyonlarının milimolar (mM) cinsinden değerleri.	41
4.1 <i>S.cerevisiae</i> mikroorganizmasına ait Enzimatik Değişkenliğin Modellenmesi ile Üretilen <i>in Silico</i> Metabolom Verilerine Uygulanan İstatistiksel Ağyapı Çıkarımı Yöntemlerinin İstatistiksel Olarak Değerlendirme Sonuçları.	61
4.2 <i>S.cerevisiae</i> mikroorganizmasına ait içsel değişkenliğin modellenmesi ile üretilen <i>in silico</i> metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı yöntemlerinin istatistiksel olarak değerlendirme sonuçları.	62
4.3 <i>S.cerevisiae</i> mikroorganizmasına ait hem içsel hem de çevresel değişkenliğin modellenmesi ile üretilen <i>in silico</i> metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı sonuçlarının istatistiksel olarak değerlendirme sonuçları.	62
4.4 <i>S.cerevisiae</i> mikroorganizmasına ait üç farklı veri tipi için GGM sonuçlarını iyileştirmek için yapılan denemelerin sonuçları.	65
4.5 <i>S.cerevisiae</i> mikroorganizması için Karesel Programlama ile Ağyapı Çıkarımı Sonuçları.	66
4.6 <i>S.cerevisiae</i> mikroorganizması için doğrusal programlama ile ağyapı çıkarımı sonuçları.	67
4.7 <i>S.cerevisiae</i> mikroorganizması için doğrusal ve karesel programlama ile elde edilen ağyapı çıkarımı sonuçlarını iyileştirmek için yapılan denemelerin sonuçları.	69

Tablo**Sayfa**

4.8	<i>S.cerevisiae</i> mikroorganizması için genetik algoritma tabanlı karma tamsayılı doğrusal programalama sonuçları.	70
4.9	<i>S.cerevisiae</i> mikroorganizması için uygunluk değeri kriteri olarak hata değerinin kullanıldığı genetik algoritma tabanlı önerilen yöntemin <i>in silico</i> metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi.	71
4.10	<i>S.cerevisiae</i> mikroorganizması için uygunluk değeri kriteri olarak sıfır sayısı ve hata değerinden oluşan ortak fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin <i>in silico</i> metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi.	71
4.11	<i>S.cerevisiae</i> mikroorganizması için uygunluk değeri kriteri olarak hem sıfır sayısı hem de hata değerinden oluşan ortak fonksiyonun kullanıldığı ve ek olarak GGM sayılarının dikkate alındığı genetik algoritma tabanlı önerilen yöntemin <i>in silico</i> metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi.	72
4.12	<i>S.cerevisiae</i> mikroorganizması için uygunluk değeri kriteri olarak hem GGM sonucunu hem de hata değerini içeren bir ortak fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin <i>in silico</i> metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi.	73
4.13	<i>S.cerevisiae</i> mikroorganizmasına ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen <i>in silico</i> verilere PC algoritmasının uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi.	74
4.14	<i>S.cerevisiae</i> mikroorganizmasına ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen <i>in silico</i> verilere CCD algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi.	75

Tablo**Sayfa**

4.15	<i>S.cerevisiae</i> mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen <i>in silico</i> verilere GES algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi.	75
5.1	<i>E.coli</i> mikroorganizmasına ait enzimatik değişkenliğin modellenmesi ile üretilen <i>in silico</i> metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı yöntemlerinin istatistiksel olarak değerlendirme sonuçları.	77
5.2	<i>E.coli</i> mikroorganizmasına ait içsel değişkenliğin modellenmesi ile üretileni <i>in silico</i> metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı yöntemlerinin istatistiksel olarak değerlendirme sonuçları.	77
5.3	<i>E.coli</i> mikroorganizmasına ait hem içsel hem de çevresel değişkenliğin modellenmesi ile üretilen <i>in silico</i> metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı sonuçlarının istatistiksel olarak değerlendirme sonuçları.	78
5.4	<i>E.coli</i> mikroorganizmasına ait üç farklı veri tipi için GGM sonuçlarını iyileştirmek için yapılan denemelerin sonuçları.	78
5.5	<i>E.coli</i> mikroorganizması için karesel programlama ile ağyapı çıkarımı sonuçları.	79
5.6	<i>E.coli</i> mikroorganizması için doğrusal programlama ile ağyapı çıkarımı sonuçları.	80
5.7	<i>E.coli</i> mikroorganizması için doğrusal programlama ve karesel programlama ile yapılan ağyapı çıkarımı sonuçlarını iyileştirmek için yapılan denemelerin sonuçları.	81
5.8	<i>E.coli</i> mikroorganizması için uygunluk değeri kriteri olarak hem sıfır sayısı hem de hata değerinden oluşan ortak fonksiyonun kullanıldığı ve ek olarak GGM sayılarının dikkate alındığı genetik algoritma tabanlı önerilen yöntemin <i>in silico</i> metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi.	82

Tablo**Sayfa**

5.9	<i>E.coli</i> mikroorganizması için uygunluk değeri kriteri olarak hem GGM sonucunu hem de hata değerini içeren ortak bir fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin <i>in silico</i> metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi.	83
5.10	<i>E.coli</i> mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen <i>in silico</i> verilere PC algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi.	84
5.11	<i>E.coli</i> mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen <i>in silico</i> verilere CCD algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi.	84
5.12	<i>E.coli</i> mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen <i>in silico</i> verilere GES algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi.	85
6.1	<i>E.coli</i> ve <i>S.cerevisiae</i> mikroorganizmaları için üç farklı biyolojik değişkenliğe ait ROC eğrilerinin AUC değerleri.	88

1 GİRİŞ

Sistem biyolojisi, genlerin, proteinlerin ve biyokimyasal reaksiyonların oluşturduğu ağyapısal sistemin bütünleşik olarak incelenmesini amaçlayan çok disiplinli bir bilim dalıdır [Klipp et al., 2005]. Bu yeni bilim dalının odaklandığı noktalardan bir tanesi hücre içindeki sürece veren ağyapıların farklı koşullarda nasıl ayarlandığı sorusudur. Sistem biyolojisi bu hedef doğrultusunda hem deneysel hem de hesaplamalı teknikleri bütünleşik olarak kullanır.

Hızlı-tarama teknikleri yoluyla hücre içindeki protein, mRNA, akı ve metabolitlerin ölçümleri yapılarak elde edilen ve genellikle ‘omics data’ olarak adlandırılan veriler hesaplamalı sistem biyolojisinin deneysel temelini oluşturur. Bu amaç doğrultusunda, hesaplamalı sistem biyolojisi yukarıdan-aşağı (‘top-down’) ve aşağıdan yukarı (‘bottom-up’) olmak üzere iki temel yaklaşım kullanır. Aşağıdan Yukarı Yaklaşım, literatürde mevcut genel hücre ağyapısı bilgisini kullanır ve farklı çevresel koşullar için bu ağyapıdaki aktif yolları hesaplamalı yöntemlerle tespit eder. Yukarıdan Aşağıya Yaklaşım ise; hiçbir ağyapı bilgisi kullanmadan, istatistik, nedensellik ve optimizasyon teknikleri ile sadece deneysel verileri analiz ederek ağyapının yapısını ortaya çıkarmaya çalışır.

Ağyapı Çıkarımı (“Network Inference”) olarak da anılan yukarıdan aşağı hesaplamalı sistem biyolojisi yaklaşımının deneysel gen ekspresyonu verilerine uygulanması ile ilgili literatürde çeşitli örnekler bulunmaktadır [Bansal ve ark., 2007; Soranzo ve ark., 2007]. Fakat, Ağyapı Çıkarımı yaklaşımının metabolik ağyapıyı öngörmek için metabolit verilerine uygulanmış örnekleri, gen ekspresyonu/gen-kontrol ağyapıları çalışmalarına nazaran oldukça azdır [Nemenman ve ark., 2007; Çakır ve ark., 2009]. Bu nedenle, yapılan tez çalışmasında metabolik ağyapı çıkarımı üzerine odaklanılmış ve farklı mikroorganizmalar için metabolom verileri kullanılarak bu mikroorganizmaların metabolik ağyapılarının öngörülmesi hedeflenmiştir.

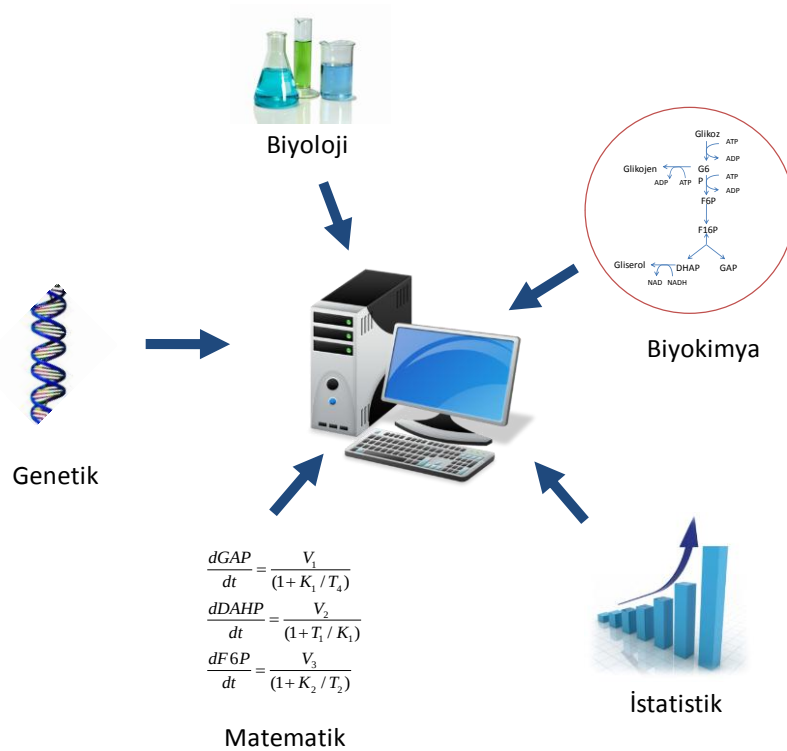
E.coli ve *S.cerevisiae* mikroorganizmaları literatürde yer alan çalışmalarda en yaygın olarak kullanılan iki mikroorganizmadır. Bu nedenle yapılan çalışmada,

E.coli ve *S.cerevisiae* mikroorganizmaları kullanılmış ve bu mikroorganizmaların aktif olarak kullandıkları metabolik ağyapılarının ortaya çıkarılması amaçlanmıştır. Bu doğrultuda her iki mikroorganizmaya ait dinamik modellerden yararlanılarak bilgisayar ortamında (*in silico*) metabolom verisi üretilmiş ve farklı biyolojik değişkenlikler için gerekli düzenlemeler yapılarak üretilen bu verilere istatistik, bayez ağlar ve optimizasyon yöntemleri uygulanarak mikroorganizmaların kullandığı metabolik ağyapılar ortaya çıkarılmaya çalışılmıştır.

Yapılan tez çalışmasında ağyapı çıkarımı yaklaşımında kullanılan istatistik, optimizasyon ve nedensellik metotları ve *in silico* deneysel veri üretme süreci hakkında ayrıntılı bilgi tezin ikinci bölümünde yer almaktadır. Yapılan çalışmalar için izlenen yol tezin üçüncü bölümünde yer almaktadır. Tezin dördüncü bölümünde *Saccharomyces cerevisiae* mikroorganizması için yapılan ağyapı çıkarımı hesaplamalarına, tezin beşinci bölümünde ise *Escherichia coli* mikroorganizması için yapılan ağyapı çıkarımı hesaplamalarına ayrıntılı olarak verilmiştir. Altıncı bölüm, bulguların tartışma ve yorumlanmasına ayrılmıştır. Son bölümde ise tez çalışmasının ana sonuçları ve öneriler verilmektedir.

2 KURAMSAL TEMELLER

Sistem biyolojisi, genlerin, proteinlerin ve biyokimyasal reaksiyonların oluşturduğu ağyapısal sistemin bütünleşik olarak incelenmesini amaçlayan çok disiplinli bir bilim dalıdır [Klipp et al., 2005]. Bu amaçla sistem biyolojisi Şekil 2.1’de görüldüğü gibi matematik, istatistik, biyokimya, biyoloji, genetik gibi birçok disiplinden yararlanır. Sistem biyolojisi hücre içindeki süreçlere şekil veren ağyapıları temel alarak birçok farklı alanda hücreyi yakından inceleme imkanı sunmaktadır. Bu doğrultuda sistem biyolojisi, deneysel ve hesaplamalı yaklaşımları bir bütün olarak kullanır.



Şekil 2.1 Sistem Biyolojisi Yaklaşımının Şematik Gösterimi

2.1 Deneysel Sistem Biyolojisi

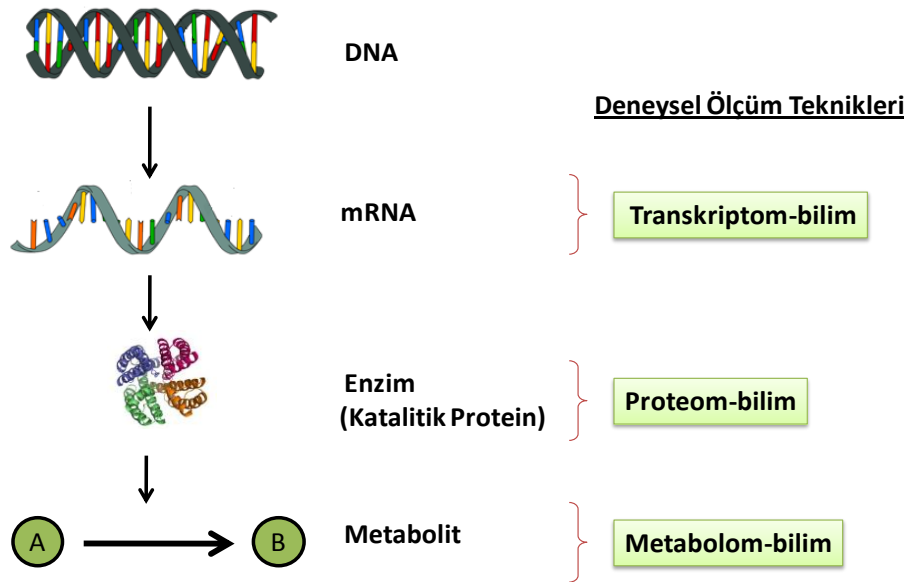
Sistem biyolojisi ile, elde edilen bilgiler ve mevcut veriler kullanılarak hücresel ağyapılar modellenir ve bu modeller üzerinden daha kolay bir şekilde hücrenin

çalışma prensiplerinin incelenmesine imkan sağlanır. Bu modelleme çalışmalarında deneysel veriler önemli rol oynamaktadır.

In silico, *in vitro* ve *in vivo* olmak üzere üç çeşit deneysel veri tipi vardır. Bunlardan *in vitro* deneysel veriler, laboratuvar ortamında canlı kullanılmadan deneysel olarak yapılan çalışmaları kapsamaktadır. *In vivo* deneysel veriler ise canlı organizma kullanılarak yapılan deneysel çalışmaları kapsamaktadır. *In silico* deneysel veri tipi ise, mikroorganizmalara ait mevcut bilgiyi içeren dinamik modeller yardımıyla hücre içerisindeki süreci modelleyerek bilgisayar yoluyla hesaplanan verileri içermektedir.

2.1.1 Sistem Biyolojisinde *In vivo* Deneysel Veri Üretimi

Şekil 2.2’de başlıca omik veri teknikleri ve hücre içinde DNA’dan başlayarak metabolik bir reaksiyonun oluşumuna kadar geçen süreç şematik olarak gösterilmiştir.



Şekil 2.2 Metabolizma ve DNA Arasındaki Hiyerarşik İlişkinin Omik Ölçümleriyle Bağlantısı

Hücresinin içerisinde, üzerinde kalıtsal bilgiyi taşıyan temel yapı taşı DNA bulunmaktadır. Bu yapı üzerinde bulunan yüzlerce genin belli bir koşulda aktif olup

olmadığını gösteren aktiflik derecesi o genden kodlanan mRNA seviyelerinin ölçülmesi ile tespit edilebilir. Hücre içi aktif mRNA seviyelerini tek deneyde aynı anda ölçmeye yarayan deneysel yöntem transkriptom-bilim ('transcriptomics') denir. Bu deneyler sonucunda elde edilen mRNA seviyesi verilerinin tümüne transkriptom adı verilmektedir.

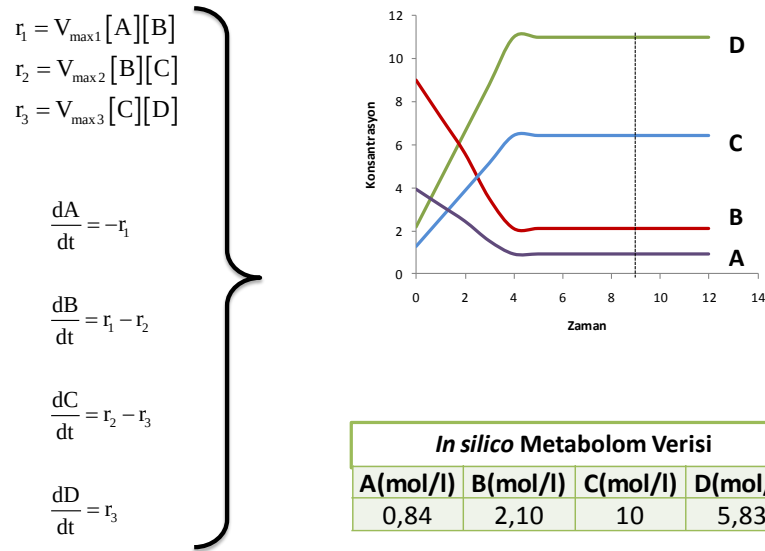
mRNA'lardaki kod bilgileri kullanılarak proteinler sentezlenir. Proteinler; aminoasitlerden oluşan polimerik yapıdır. Toplam 21 aminoasit vardır ve proteinler bunların farklı kombinasyonlarda bir araya gelmesiyle oluşur. Proteinlerden katalitik aktivitesi olanlar (enzimler) hücre içindeki metabolitler arasındaki reaksiyonları gerçekleştirir. Hücre içi aktif proteinleri tek deneyde aynı anda ölçmeye yarayan deneysel yöntem proteom-bilim ('proteomics'), bu deneysel ölçümler sonucu elde edilen protein verilerinin tümüne de proteom denir.

Metabolit, glikozun ve besinlerin hücre içine alınarak parçalanması veya birleşmesiyle elde edilen kimyasal maddelere denir [Demain, 1980]. Hücre içi metabolitleri kalitatif ve kantitatif olarak ölçmeye yarayan deneysel yöntem metabolom-bilim ('metabolomics') ve bu yöntem sonucu elde edilen verilerin tümüne de metabolom adı verilir. Transkriptom-bilim, Proteom-bilim ve Metabolom-bilim gibi tekniklerle üretilen ve sistem biyolojisi çalışmalarında kullanılan verilere genel olarak omik veriler denir. Sistem biyolojisinin deneysel kısmı bu omik verilere dayanmaktadır.

2.1.2 Sistem Biyolojisinde *In silico* Deneysel Veri Üretimi

Sistem biyolojisinde yapılan çalışmaların birçoğu bir mikroorganizmanın temel ağyapı bilgisini içeren ve bir mikroorganizmayı bilgisayar üzerinde inceleme olanağı sağlayan dinamik modellere dayanmaktadır. Sistem biyolojisinin hızla gelişmesi ile literatürde mikroorganizmalara ait dinamik modellerin sayısı günden güne artmaktadır. Bu doğrultuda artık, literatürde pek çok mikroorganizmaya ait dinamik modeller mevcuttur [Chassagnole et al., 2002; Teusink et al., 2000; Shinto et al., 2007]. Daha da öte, günümüzde Biomodels (<http://www.ebi.ac.uk/biomodels-main/>) ve JJJ (<http://www.jjj.bio.vu.nl/>) gibi dinamik modellere özgü veritabanları vardır. Bu veritabanlarında yer alan dinamik modeller üzerinden gerçekleştirilen *in silico*

veri üretimi özellikle yöntem geliştirme ve test etmede fayda sağlamaktadır. *In silico* veri üretimi sayesinde çok hızlı bir şekilde farklı koşullara ait veri üretilirerek deney tasarımı ile ilgili bilgi edinilebilir. Bu nedenle *in vivo* deneysel veri üretiminin maliyet ve zaman açısından dezavantajları düşünüldüğünde, yapılan araştırmaların bir kısmında *in silico* veri üretiminin kullanılması şaşırtıcı değildir.



Şekil 2.3 *In silico* Deneysel Veri Üretiminin Şematik Olarak Gösterimi

In silico olarak deneysel veri üretmek için öncelikle incelenecek mikroorganizmaya ait literatürden uygun dinamik modeller seçilir. Seçilen model çalıştırılarak bilgisayar yardımıyla deneysel metabolom verisi üretilir (Şekil 2.3).

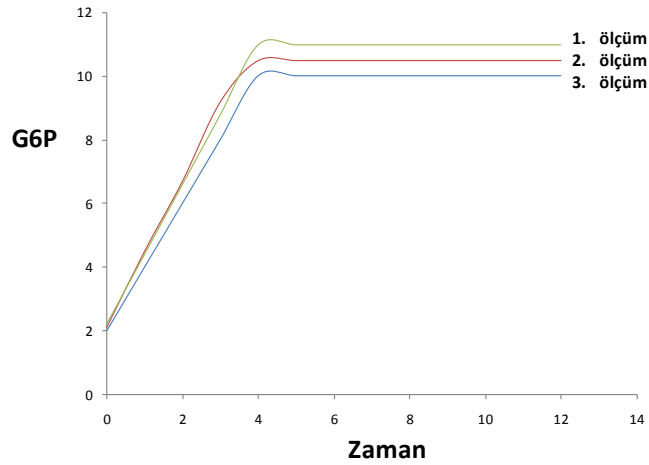
Modeldeki parametreler değiştirilerek kolaylıkla farklı koşullara ait *in silico* veriler elde edilebilir. Aynı mikroorganizmanın farklı popülasyonları arasında biyolojik değişkenlikten kaynaklanan küçük farklılıklar da bu yaklaşım yoluyla istenilen verilere yansıtılabilir.

2.1.3 Sistem Biyolojisi Verilerinde Biyolojik Değişkenlik

Yapılan deneysel çalışmaların güvenilirliği açısından omik ölçümler birkaç kez tekrar edilmelidir. Bu amaçla yürütülen çalışmalarda, aynı genetik geçmişe dayalı ve aynı şartlarda yürütülen omik deneyleri her zaman aynı sonuçları vermemektedir. [Çakır ve ark., 2009]. Çünkü organizmaların hiçbirisi birbirinin aynısı değildir ve

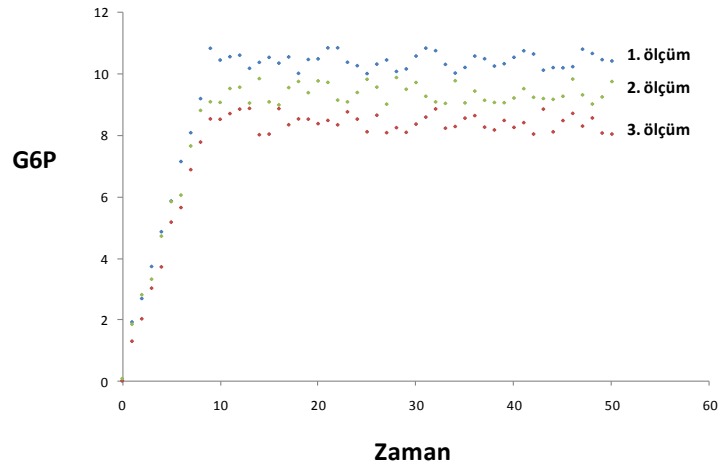
aralarında küçük farklılıklar olması normaldir. İşte bu farklılıklar genel olarak biyolojik değişkenlik olarak adlandırılır ve bu farklılıklar, enzimatik değişkenlikten, çevresel değişkenlikten veya iç değişkenlikten kaynaklanabilir. Hücreler arası ve hücre içinde varolan bu değişkenlikler alt bölümlerde açıklanmıştır.

Enzimatik Değişkenlik: Enzimatik değişkenlik, iki ayrı mikroorganizma popülasyonu arasında reaksiyon hız sabitleri (k) veya enzim konsantrasyonları arasında ufak da olsa farklılıklar olabileceği gerçeğine dayanmaktadır (Şekil 2.4). Yani Şekil 2.4'te de görüldüğü gibi bir metabolitin konsantrasyon ölçümleri için deneyler tamamen aynı koşullarda yapılsa da, hiçbir zaman aynı sonuçlar elde edilemeyebilir. Ancak değerler birbirine çok yakın seviyelerde ölçülmektedir.



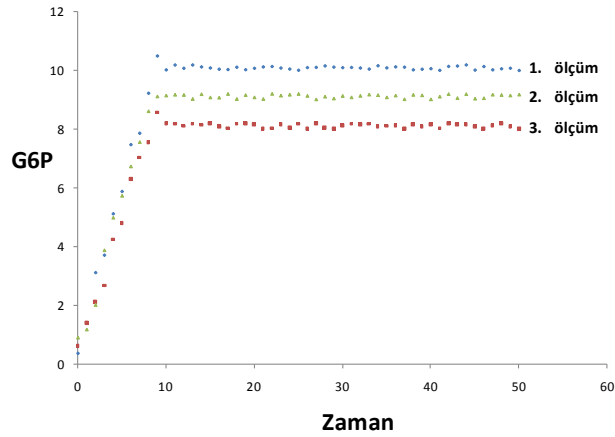
Şekil 2.4 Enzimatik Değişkenliğin Gösterilmesi. Bir metabolitin ayrı popülasyona ait üç farklı örnek üzerinde yapılan ölçüm değerlerinde enzimatik değişkenlikten dolayı küçük farklılıkların mevcut olduğu grafik üzerinde gösterilmiştir.

İçsel Değişkenlik: Hücre içerisindeki metabolit konsantrasyonlarında zamanla dalgalanmalar olabilir. Bu dalgalanmalar, deneysel hatalardan değil hücre içindeki süreçten dolayıdır. Hücre içi normal olan bu dalgalanmalar zamana bağlı toplanan verilerin düzgün bir şekilde birbirini izlememesine neden olur (Şekil 2.5). Bu dalgalanmalar doğal olabilir veya sıcaklık, pH gibi değişkenlerdeki küçük değişimlerden kaynaklanmaktadır. Zaten hücre gibi yüzlerce enzim, gen ve metabolitin çalıştığı bir sistemde zamanla ufak dalgalanmalar olması normaldir.



Şekil 2.5 İçsel Değişkenliğin Gösterilmesi. Bir mikroorganizmaya ait bir metabolitte içsel değişkenlikten dolayı meydana gelen zaman içerisindeki dalgalanmalar grafik üzerinde gösterilmiştir.

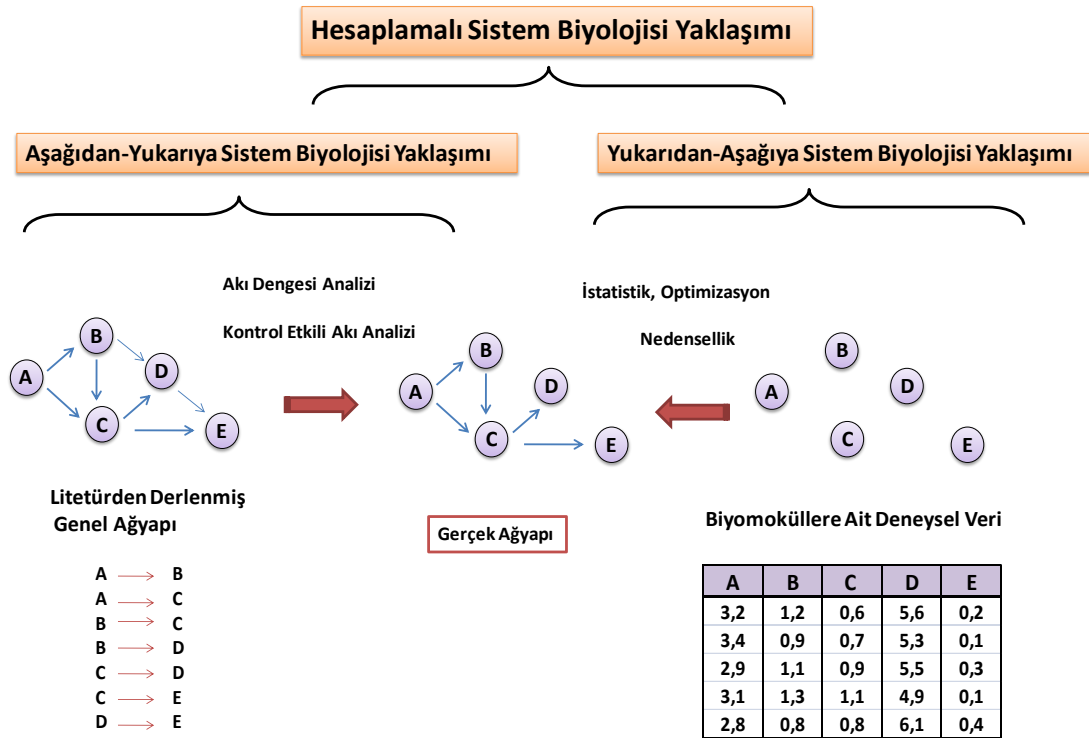
Çevresel Değişkenlik: Hücre dışı substrat/besin konsantrasyonlarındaki küçük değişimlerden dolayı dalgalanmalar olabilir ve bu değişimler hücre içine taşınarak içerideki metabolitlerin seviyelerinde değişkenliğe neden olabilir. Bu nedenle zamana bağlı toplanan veriler çok düzgün bir biçimde birbirini takip etmeyebilir.



Şekil 2.6 Çevresel Değişkenliğin Gösterilmesi. Bir mikroorganizma içerisindeki bir metabolit konsantrasyonunda çevresel değişkenlikten dolayı meydana gelen zaman içerisindeki dalgalanmalar grafik üzerinde gösterilmiştir.

2.2 Hesaplamalı Sistem Biyolojisi

Hesaplamalı sistem biyolojisi, düzenleyici ve metabolik ağyapılar gibi hücre içerisindeki sürece şekil veren ağyapılara odaklanarak bir bütün olarak hücreyi matematiksel yöntemler ışığında inceleme imkanı tanıyan bir sistem biyolojisi yaklaşımıdır. Hesaplamalı sistem biyolojisi, hesaplamalı teknikleri kullanarak omik verilerin analizi sonucunda sistem düzeyinde hücreyi anlamayı amaçlar [Kriete and Eils, 2005]. Bu amaç doğrultusunda hesaplamalı sistem biyolojisi Şekil 2.7’de görüldüğü gibi yukarıdan-aşağı (‘top-down’) ve aşağıdan yukarı (‘bottom-up’) hesaplamalı sistem biyolojisi yaklaşımı olmak üzere iki temel yaklaşımı kullanır. Bu iki yaklaşım sonucunda da çıkış noktaları farklı olmasına rağmen hücrenin aktif olarak kullandığı reaksiyonların yer aldığı ağyapılar elde edilir.



Şekil 2.7 Hesaplamalı Sistem Biyolojisi Yaklaşımının Şematik Olarak Gösterimi

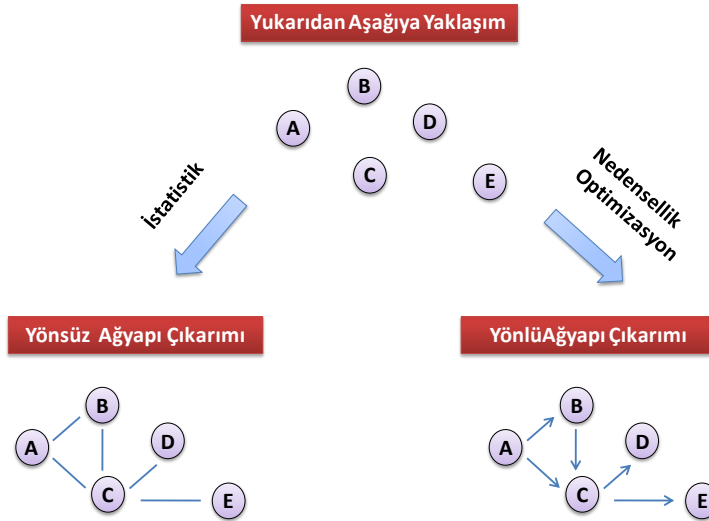
2.2.1 Aşağıdan Yukarı Yaklaşım

Aşağıdan-yukarı yaklaşım, Şekil 2.7’de de görüldüğü gibi mevcut ağyapı bilgisini kullanarak farklı koşullarda hücrenin ağyapıdaki hangi reaksiyonları aktif

olarak kullandığını ortaya çıkaran bir yaklaşımdır. Bu amaç doğrultusunda en yaygın kullanılan yöntem Akı Dengesi Analizidir. Akı Dengesi Analizi, optimizasyon kullanarak metabolik ağyapıları oluşturan reaksiyonların hızlarını ve farklı koşullarda metabolitlerin üretilme kapasitesini hesaplamaya olanak sağlayan bir yöntemdir [Kauffman et al., 2003].

2.2.2 Yukarıdan Aşağıya Yaklaşım

Yukarıdan aşağıya yaklaşım Şekil 2.7’de de görüldüğü gibi aşağıdan yukarı yaklaşımın tam tersi bir yol izleyerek hiçbir ağyapı bilgisi kullanmadan sadece deneysel verileri analiz ederek ağyapının yapısını ortaya çıkarmaya çalışan bir yaklaşımdır. Yukarıdan aşağıya yaklaşım ile literatürde ağyapı bilgisi bulunmayan bir mikroorganizmanın ağyapı bilgisi açığa çıkarılabilir. Ağyapı çıkarımı için kullanılan formülasyonlar istatistikten optimizasyona, nedenselliğe kadar farklılıklar gösterir [Markowitz and Spang, 2007]. Bu yaklaşım yönlü ve yönsüz ağyapı çıkarımı olarak iki kısma ayrılır (Şekil 2.8).



Şekil 2.8 Yukarıdan Aşağıya Sistem Biyolojisi Yaklaşımında Yönlü ve Yönsüz Ağyapı Çıkarımının Şematik Olarak Gösterilmesi [Çakır, 2010]

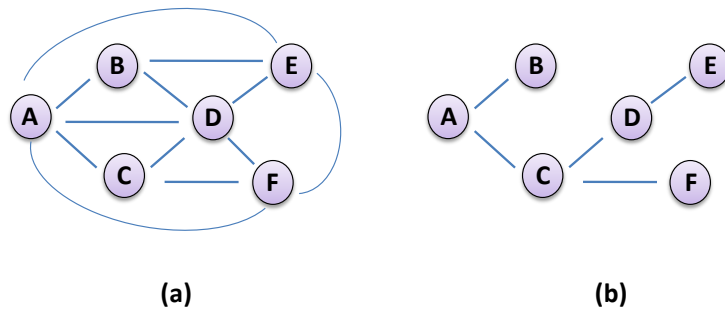
Yönsüz ağyapı çıkarımında istatistiksel yöntemler kullanılarak ağyapının yapısı yönsüz olarak ortaya konur. Bu amaçla korelasyon katsayısı hesaplamaları kullanılır. Biyomoleküllere ait deneysel ölçümler arasında farklı derecelerde

korelasyon katsayıları hesaplamaları yapılarak biyomoleküller arasındaki ilişkiler hakkında öngörude bulunulur. Bu yöntemde iki biyomolekül arasında ilişki olup olmadığı konusunda çıkarımda bulunulurken, bu ilişkinin yönü hakkında herhangi bir bilgi elde edilememektedir.

Yönsüz ağyapı çıkarımı ile ilişkilerin yönü ile ilgili bilgi elde edilemediğinden dolayı yönlü ağyapı teknikleri zamanla geliştirilmiştir. Yönlü ağyapı çıkarımı ile ilişkinin hangi yönde olduğu yani hangi biyomolekülün diğerine dönüştüğü/diğerini etkilediği hakkında bilgi sahibi olunabilmektedir. Bu amaçla nedensellik ilkesi dahilinde veri madenciliği teknikleri ile deneysel ölçümler analiz edilerek yönlü ağyapı hakkında öngörude bulunulur [Husmeier, 2003]. Ayrıca nedensellik dışında hücrenin amacının minimum-etkileşim olduğu kabulünü direk olarak kullanarak, optimizasyon (doğrusal programlama vs.) yöntemleri ile yönlü ağyapıyı öngörmeyi amaçlayan teknikler de mevcuttur. [Thomas ve ark., 2004; Wang ve ark., 2006].

2.3 Metabolik Ağyapıların Yapısı

Genel olarak hücrelerin metabolik ağyapılarının yoğun aralıklı değil seyrek aralıklı ('sparse') oldukları zaten bilinmektedir (Şekil 2.9). Bunun biyolojik olarak anlamı, hücrenin amacının biyomolekülleri arasındaki etkileşim sayısını minimum düzeyde tutmak olduğu yani reaksiyonları gerçekleştirmek için üreteceği enzim miktarını minimum düzeyde tutmak isteyeceği varsayımdır [Yeung et al., 2002]. İstatistik, optimizasyon ve nedenselliğe dayalı ağyapı çıkarımında genellikle bu bilgi temel alınarak çalışmalar gerçekleştirilir.



Şekil 2.9 Yoğun ve Seyrek Aralıklı Ağyapı Örneklerinin Şematik Gösterimi **a.** Yoğun Aralıklı Ağyapı **b.** Seyrek Aralıklı Ağyapı

2.4 İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımı

Benzerlik testlerine dayalı olarak yönsüz ağyapı çıkarımı yapan yöntemler, istatistiğin temel kavramlarını yoğun olarak kullanır.

2.4.1 Temel Kavramlar

İstatistiksel metotların temeli olarak, veriler hakkında sayısal olarak bilgi veren; ortalama, standart sapma ve varyans kavramları gösterilebilir.

Ortalama: Bir veri dizisini tek bir nokta ile temsil etmek için verilerin orta noktası değerinin hesabıdır. Eşitlik 2.1’de yer alan formül ile hesaplanmaktadır.

$$\text{Ortalama}(\bar{X}) = \frac{\sum_{i=1}^n X_i}{n} \quad (2.1)$$

Varyans: Dizideki verilerin ortalamadan uzaklıklarının karelerinin ortalamasıdır. Kısaca istatistiksel anlamda veri dağılımının bir ölçüsüdür. Eşitlik 2.2’deki formül ile hesaplanmaktadır.

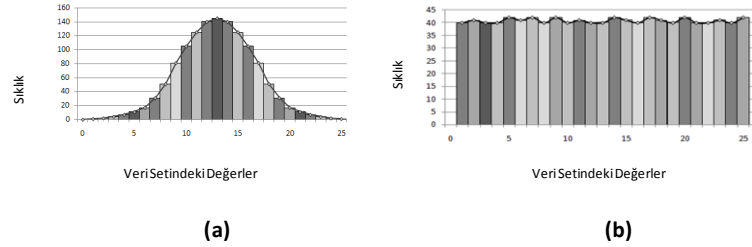
$$\sigma^2(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (2.2)$$

Standart Sapma: Dizideki verilerin ortalamadan ne kadar saptığının bir ölçüsüdür. Varyans da standart sapma da istatistiksel veriler için dağılımın bir ölçüsüdür. Eşitlik 2.3’te yer alan formül ile standart sapma hesaplanabilmektedir.

$$\sigma(X) = \sqrt{\sigma^2} \quad (2.3)$$

Dağılım türü: Verilerin grafiksel olarak dağılımı gösterilerek türü belirlenir. Birçok dağılım türü vardır. Bu dağılım türlerinden en önemlisi istatistiksel anlam testlerinin genelinin temel aldığı normal dağılımdır. Normal bir dağılım gösteren veri setindeki değerlerin çoğu ortalamanın etrafında toplanmıştır (Şekil 2.10.a). Normal dışı bir

dağılım gösteren veri setine örnek olarak tekdüze dağılım verilebilir. Tekdüze dağılımda veriler genellikle ortalama etrafında değil her yere eşit olarak dağılmıştır (Şekil 2.10.b). Bu tür dağılımlar tekdüze dağılım olarak adlandırılır. Ayrıca normal dışı dağılım türüne örnek olarak binom dağılımı, poisson dağılımı da verilebilir.



Şekil 2.10 Normal ve Normal Dışı Dağılım Örnekleri. **a.**Normal Dağılım Gösteren Bir Veri Seti Örneği (<http://msdn.microsoft.com/en-us/library/bb924368.aspx>) **b.** Normal Dışı Dağılım Gösteren Bir Veri Seti Örneği; Tekdüze Dağılım (<http://msdn.microsoft.com/en us/library/bb924368.aspx>)

2.4.2 Korelasyon

Korelasyon, iki veri arasında benzerliğin olup olmadığını analiz etmek için kullanılan istatistiksel bir yöntemdir. Korelasyon hesaplanmadan önce kovaryans terimini açıklamalıyız. Kovaryans; iki değişkenin birlikte değişiminin ölçüsüdür. Kovaryans, Eşitlik 2.4'teki formülle hesaplanmaktadır.

$$\Gamma = Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (2.4)$$

İki değişken arasındaki korelasyon hesabı (R) ise iki değişkenin kovaryansının bu değişkenlerin standart sapmalarının çarpımlarına bölünmesi ile yapılır:

$$R = \frac{Cov(X, Y)}{S(X)S(Y)} \quad (2.5)$$

Burada R değeri X ve Y arasındaki lineer ilişkiyi tanımlamaktadır ve korelasyon katsayısı olarak adlandırılmaktadır. Korelasyon katsayısı -1 ile +1

arasında değer alır. Korelasyon değeri -1 veya +1 değerine yaklaştıkça ilişki o oranda güçlüdür. Bu şekildeki korelasyon yaklaşımına doğrusal korelasyon veya Pearson korelasyonu denir.

2.4.3 Kısmi Korelasyon Katsayısı

Çok değişkenli sistemlerde diğer değişkenlerin iki değişken üzerinde dolaylı etkileri vardır. Bu yüzden ikiden çok değişkenli sistemlerde seçilen iki değişken arasında doğrusal bir ilişki olup olmadığına bakılacaksa Pearson korelasyonu gerçek ilişkiyi doğru bir şekilde yansıtamayacaktır. Çünkü Pearson korelasyonu, dolaylı ve direk ilişkiler arasında bir ayırım yapamaz. Bu nedenle ikiden daha fazla değişken içeren sistemlerde iki değişken arasındaki ilişkilerin gücünü doğru bir şekilde ortaya koymak için kısmi korelasyon katsayısı kullanılır.

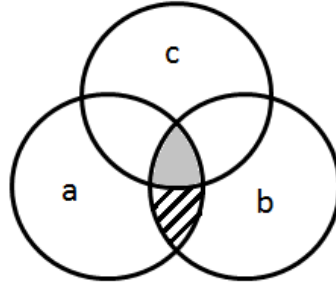
Kısmi korelasyon katsayısındaki amaç; ilişkilerine bakılan iki değişken arasındaki korelasyondan, farklı bir veya daha fazla değişkenin dolaylı etkisini çıkararak daha güvenilir sonuçlara ulaşmaktır.

Örneğin a , b ve c değişkenlerinin olduğu bir sistemde a ve b değişkenlerinin arasındaki ilişkiye bakılacak olsun. Bu ilişki Pearson korelasyonu ile hesaplanırsa; a ile b değişkenlerinin arasındaki direk etkileşim yanında eğer varsa c üzerinden dolaylı etkileşimleri de sonucu etkileyecektir.

Kısmi korelasyon katsayısı ise, a ile b değişkenlerinin arasındaki korelasyondan c 'nin a ve b değişkenleri üzerindeki dolaylı etkisinin çıkarılması ile hesaplanır:

$$R_{ab,c} = \frac{R_{ab} - R_{ac}R_{bc}}{\sqrt{(1 - R_{ac}^2)(1 - R_{bc}^2)}} \quad (2.6)$$

Bu şekilde a ile b 'nin korelasyonundan c 'nin dolaylı etkisi çıkarılır (Şekil 2.11). Kısmi korelasyon katsayısı ile bulunan sonuç, c 'nin a ile b arasındaki korelasyona etkisini dikkate almayan Pearson korelasyon katsayısına göre daha güvenilirdir.



Şekil 2.11 Birinci Dereceden Kısmi Korelasyonun şematik olarak gösterilmesi. Taralı olan kısım; a ile b'nin etkileşiminden c'nin dolaylı etkisinin çıkarılması sonucu elde edilen korelasyonun görsel olarak ifadesidir.

Kısmi korelasyonun derecesi kaç tane değişkenin dolaylı etkisinin çıkarılması ile ilgilidir. Eşitlik 2.6 ile tanımlanan kısmi korelasyon, birinci derecedendir. Eğer iki değişken arasındaki korelasyonu başka iki değişkenin dolaylı etkisini çıkararak hesaplıyorsak ikinci dereceden kısmi korelasyon olarak adlandırılır ve Eşitlik 2.7'deki formüller ile hesaplanır.

$$R_{ab,cd} = \frac{R_{ab,c} - R_{ad,c}R_{bd,c}}{\sqrt{(1 - R_{ad,c}^2)(1 - R_{bd,c}^2)}} \quad (2.7)$$

Ayrıca daha yüksek dereceden katsayıları hesaplamak için benzer eşitlikler mevcuttur [Fuente ve ark., 2004]. Kısmi korelasyon hesabı n terimli bir sistemde $(n-2)$ 'nci dereceye kadar hesaplanabilir. Yani bu n değişkenli bir sistemde iki değişken arasındaki korelasyon hesaplanırken diğer tüm değişkenlerin etkilerinin çıkarılması anlamına gelmektedir. $(n-2)$ 'nci dereceden korelasyonun kolay bir şekilde hesabı için Dempster tarafından önerilen Grafiksel Gaus Metodu ('Graphical Gaussian Model, GGM') olarak adlandırılan bir metot vardır [Demster, 1972].

2.4.4 Grafiksel Gaus Model (GGM)

Bu yöntemle $(n-2)$ 'nci dereceden korelasyonu Eşitlik (2.6) ve (2.7)'ye benzer karmaşık formüllerle hesaplamaya gerek kalmaz. Pearson korelasyonu hesabı ile elde edilen korelasyon katsayıları matrisine uygulanan birkaç işlemten sonra $(n-2)$ 'nci dereceden korelasyon matrisi elde edilebilir. Bu işlemler sırasıyla şu şekildedir:

- a) Pearson korelasyonu matrisine P adını verelim. Öncelikle P matrisinin tersi alınır:

$$W = P^{-1} \quad (2.8)$$

- b) Daha sonra elde edilen W matrisinde herbir eleman, ilgili indise karşılık gelen matrisin köşegenlerinin çarpımının kareköküne bölünür ve -1 ile çarpılır:

$$\Omega_{i,j} = \frac{-W_{i,j}}{\sqrt{W_{i,i}W_{j,j}}} \quad (2.9)$$

Bu işlemler sonucunda oluşturulan Ω matrisi (n-2)'nci kısmi korelasyon matrisidir.

2.4.5 Bazı İstatistik Terimlerinin Açıklanması

Mikroorganizmalara ait deneysel verilere korelasyon hesaplamaları uygulanarak hücrenin metabolik ağyapısı hakkında öngöründe bulunmaktadır. Bu amaçla farklı derecelerden kısmi korelasyon katsayıları kullanılabilir. Hesaplanacak korelasyon yöntemi seçildikten sonra elde edilen korelasyon katsayılarından tahmin değerleri (probability value, 'p-value') hesaplanır ve hesaplanan tahmin değerleri de anlamlandırılarak bir sonuca varılır. Bu bağlamda yapılan çalışmalarda sıklıkla kullanılan yokluk hipotezi, hata değeri, tahmin değeri terimlerini açıklamak gerekir.

Yokluk Hipotezi ('null hypothesis') : İki durum arasındaki ilişki için kurulan hipotezdir ve şu şekildedir; 'iki biyomolekül (veya koşul) arasında anlamlı bir ilişki yoktur'. Anlam testlerinde ve benzerlik ölçülerinde, kurulan bu hipotezin doğruluğuna veya yanlışlığına bakılır.

Hata Değeri ('cut-off'): Literatürlerde α değeri olarak da geçer. Genelde yapılan istatistiksel testlerde yanılma payı olmaktadır. Bu yanılma payı hata miktarı ile temsil edilir. Hata miktarı keyfi küçük bir değer seçilir. Hata terimi ne kadar küçük olursa testin güvenilirliği o kadar artar. Genellikle hata terimi 0,01 veya 0,05 seçilir.

Tahmin Değeri ('p-value'): Yokluk hipotezinin doğru olma ihtimalinin matematiksel olarak ifadesidir. Tahmin değeri, incelenen değişkenlere uygulanan anlam testleri sonucunda elde edilir. Tahmin değeri hata değerinden küçükse hipotez yanlıştır. Örneğin; hata değerini 0,05 olarak belirlemiş olalım; bu durumda iki değişken arasında hesaplanan tahmin değeri 0,02 çıkarsa bu hipotezin yanlı olduğunu gösterir. Çünkü hipotezin doğru olma ihtimali 0,02 olarak bulunmuştur, ve bu değer seçilen yanılma payından (hata değerinden) daha düşük bir ihtimaldir. Yani iki koşul arasında anlamlı bir ilişki vardır. Tahmin değeri hata değerinden büyük ise hipotez doğrudur. Yani iki koşul arasında anlamlı bir ilişki yoktur. Tahmin değeri 0 ile 1 arasında bir değer alır.

2.4.6 Anlam Testleri

Anlam testleri iki koşul arasında anlamlı bir değişiklik olup olmadığını istatistiksel olarak ölçer. Bunu, veriye dayalı olarak bir tahmin değeri hesaplayarak yapar. Anlam testleri parametrik ve parametrik olmayan testler olarak ikiye ayrılırlar. Parametrik testlerin en yaygın olanları normal dağılım gösteren veriler arasındaki anlamlı ilişkinin, parametrik olmayan testler ise normal dağılım göstermeyen veriler arasındaki anlam ilişkisinin çıkarılmasında kullanılır.

2.4.6.1 Student t-test

Parametrik testlerden en yaygın kullanılanı t-testtir. t-test iki koşul arasında karşılaştırma yaparken verilerin ortalamasını ve standart sapmasını kullanır. Bu testten doğru şekilde sonuç almak için verilerin normal dağılım göstermesi şarttır.

2.4.6.2 Ranksum test (u-test)

Parametrik olmayan testlerden biridir. Mann-Whitney test olarak da adlandırılır. Verilerin gerçek değerleri yerine sıralamalarını kullanır. Ortalama, standart sapma gibi dağılım bilgilerine gerek yoktur. Ortalama yerine medyanları karşılaştırır. Medyan; veri dizisinin küçükten büyüğe doğru sıralanması sonucunda bu dizinin ortasında bulunan değerdir. Normal dağılım göstermeyen veriler arasında

daha iyi sonuç verir. Dezavantajı, veri sayısı az olduğunda güvenli sonuç vermemesidir.

2.4.6.3 Rastgele Sıralama Testi ('Randomization Test')

Bu yöntemde de diğer anlam testi metotlarında olduğu gibi iki değişken veya iki koşul arasındaki ilişkiyi ölçmek için tahmin değeri ('p-value') hesaplanır. Rastgele Sıralama testi, verinin dağılım tipi ne olursa olsun tahmin değeri hesaplamada güvenilir bir yöntemdir.

Rastgele sıralama testi ile tahmin değeri hesaplanması şu şekilde olur; x ve y değişkenlerine ait verilerin kendi içerisinde yerleri i (i ; 100 veya 1000 gibi yüksek bir sayı seçilir) kez rastgele değiştirilir. Ve her seferinde x ve y arasındaki korelasyon hesaplanır. Başlangıçta, gerçek x ve y arasındaki korelasyon hesaplanır ve i tane rastgele korelasyon değeri ile karşılaştırılır. i tane korelasyon değerinden kaçının başlangıçtaki korelasyon değerinden büyük olduğuna bakılır. Büyük olan değerlerin sayısı i sayısına bölünerek x ve y arasındaki korelasyonun tahmin değeri bulunur. Bu bir anlamda ilgili korelasyon değerinin rastgele elde edilebilme ihtimalidir. Bu yöntemin en önemli özelliği daha önce de bahsedildiği gibi dağılım türünden etkilenmemesi, dağılım türü ne olursa olsun iyi sonuç vermesidir.

2.4.7 Korelasyonun anlamlandırılması

Anlam testlerinde hesaplanan tahmin değeri hata değerinden küçük ise yokluk hipotezi yanlıştır ve iki değişken arasında anlamlı bir ilişki vardır. Büyük ise yokluk hipotezi doğrudur ve anlamlı ilişki yoktur. Bu ortaya çıkan ilişkinin kuvvetini de korelasyon katsayısı açığa çıkarır. Korelasyon kısmında anlatıldığı gibi korelasyon katsayısı -1 veya +1 değerine yaklaştıkça doğrusal ilişki o oranda güçlüdür.

2.5 Optimizasyona Dayalı Yönlü Ağyapı Çıkarımı

Bu çalışmada Lyapunov Deklemine (Bölüm 2.4.6) dayalı olarak yönlü ağyapı çıkarımı yapan yöntemler kullanılmıştır. Bu denklem tanımlanmamış

(‘underdetermined’) bir yapıda olduğundan, çözümü optimizasyon yöntemlerinin kullanılmasını gerektirmektedir. Bu amaçla farklı yöntemler kullanılmaktadır. Bu yöntemlere başlıca doğrusal programlama, karesel programlama, karmaşık tamsayılı programlama ve genetik algoritma örnek gösterilebilir. Bu optimizasyon yöntemleri optimizasyonun temel kavramları üzerine kurulu olduğu için öncelikle temel optimizasyon kavramları açıklanmalıdır.

2.5.1 Temel Kavramlar

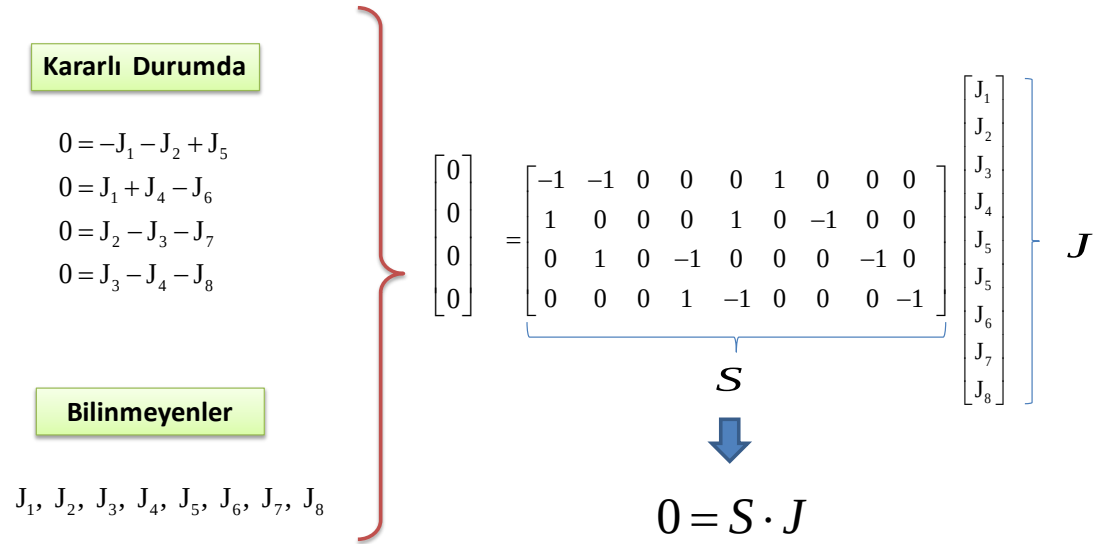
Optimizasyon, belirlenen bir amaç fonksiyonunu belirli kısıtlar altında maksimum veya minimum yapmak için bir problemin veya denklem sisteminin çözülmesidir. Optimizasyonda amaç birden çok çözüm arasından istenen amaca yönelik en iyi çözümü tespit etmektir. Bu nedenle optimizasyon ile çözümü yapılacak bir problemin veya sistemin birden çok çözümü olması gerekmektedir. Bunun tespiti o sistemin serbestlik derecesi hesaplanarak yapılabilir.

Bir sistemin serbestlik derecesi bilinmeyen sayısından bağımsız denklem sayısının farkı alınarak bulunur:

$$\text{Serbestlik Derecesi} = \text{Bilinmeyen Sayısı} - \text{Bağımsız Denklem Sayısı} \quad (2.10)$$

Eğer bir sistemin serbestlik derecesi sıfır ise yani bilinmeyen sayısı bağımsız denklem sayısı birbirine eşit ise bu sistemin tek bir çözümü vardır. Eğer sistemin serbestlik derecesi sıfırdan büyük ise yani bilinmeyen sayısı bağımsız denklem sayısına göre fazla ise bu tür sistemler tanımlanmamış sistem olarak adlandırılır ve bu sistemin birden çok çözümü vardır. Bu tür birden çok çözüme sahip tanımlanmamış bir sistemin çözümünde optimizasyon yöntemlerine ihtiyaç duyulur.

Optimizasyon yöntemleri çözüm için kısıtlar (‘constrains’) ve amaç fonksiyonu (‘objective function’) olmak üzere iki parametre kullanır. Kısıtlar, çözüme ulaşmak için o sistemin özelliklerine göre ölçülmüş değerleri, sınır aralıklarını ve sisteme ait belirli özellikleri kapsayan matematiksel eşitlik ve eşitsizliklerdir. Amaç fonksiyonu ise, optimizasyonun yapılma amacının matematiksel bir ifadesidir.



Şekil 2.12 Bir sistemin optimizasyon için tanımlanması örneğinin şematik gösterimi

Yapılan tanımlamalar Şekil 2.12'deki örnek bir sistem üzerinden açıklanırsa; bu sistemin bilinmeyen sayısı 9'dur. Bağımsız denklem sayısı ise 4'tür. Bu sistemin serbestlik derecesi, bilinmeyen sayısından bağımsız denklem sayısının çıkarılması ile 5 olarak bulunur. Serbestlik derecesi sıfırdan büyüktür. Yani sistem tanımlanmamış bir sistemdir ve çözümü için optimizasyon yöntemlerinin kullanılmasını gerektirmektedir. Bu sistemde amaç fonksiyonu J_2 'nin maksimum değer alması olsun:

$$\text{Amaç Fonksiyonu} = \text{maksimum } (J_2) \quad (2.11)$$

Amaç fonksiyonu için optimizasyon yapılmadan önce var ise kısıtlar tanımlanmalıdır. S matrisi ile bilinmeyenlerin değerlerini içeren J vektörünün çarpımının sıfıra eşit olması bu sistem için bir eşitlik kısıtıdır:

$$S \cdot J = 0 \quad (2.12)$$

Ayrıca sistemde diğer kısıtlara örnek olarak her bir bilinmeyen değerinin belirli bir aralıkta değer almak zorunda olması gösterilebilir. Bu sistem için bir eşitsizlik kısıtıdır:

$$\text{Alt Sınır Değeri} \leq J_i \leq \text{Üst Sınır Değeri} \quad (2.13)$$

Tanımlanmamış sistemin çözümü için bu amaç fonksiyonu ve kısıtların doğrusal olup olmamasına göre bir optimizasyon yöntemi belirlenir. Optimizasyon yöntemleri, bu açıdan doğrusal ve doğrusal olmayan yöntemler olarak ikiye ayrılmıştır. Doğrusal optimizasyon yöntemlerine doğrusal programlama, karma tamsayılı doğrusal programlama örnek verilirken, doğrusal olmayan optimizasyon yöntemlerine karesel programlama örnek gösterilebilir.

2.5.2 Doğrusal Programlama

Sürekli değişkenlerin olduğu doğrusal bir amaç fonksiyonu ve doğrusal kısıtların yer aldığı sistemlerin çözümünde kullanılır. Bu tür sistemlerin çözümünde kullanılan en yaygın metottur. Bir amaç fonksiyonunun en genel matematiksel ifadesi Eşitlik 2.14'te verilmiştir.

$$\text{Amaç Fonksiyonu} = \sum f_i X_i^n \quad (2.14)$$

Burada f amaç fonksiyonu katsayısı vektörüdür ve reel sayı değerleri almaktadır. X değişkenleri kapsayan bir vektördür. n ise üssel ifadedir. Doğrusal programlamada bu değer her zaman 1'dir. Bu sayede amaç fonksiyonu her zaman doğrusaldır.

2.5.3 Karesel Programlama

Amaç fonksiyonu ve kısıtlardan birinin veya ikisinin de doğrusal olmadığı sistemlerin çözümünde doğrusal olmayan optimizasyon yöntemlerine ihtiyaç duyulur. Bu yöntemlerden bir tanesi de karesel programlamadır. Bir doğrusal olmayan programlama türü olan karesel programlamanın özelliği, amaç fonksiyonunda doğrusal terimlere ek olarak karesel terimlerin de yer almasıdır.

Eşitlik 2.14'te yer alan n değeri karesel programlama için 1 veya 2 olabilir. Böylelikle karesel programlama, n ifadesinin 2 değerini alabilmesinden dolayı amaç fonksiyonunda karesel ifadelerin de yer almasına olanak tanımaktadır. Kısıtlar ise bu yöntemde her zaman doğrusaldır.

2.5.4 Karma Tamsayılı Doğrusal Programlama

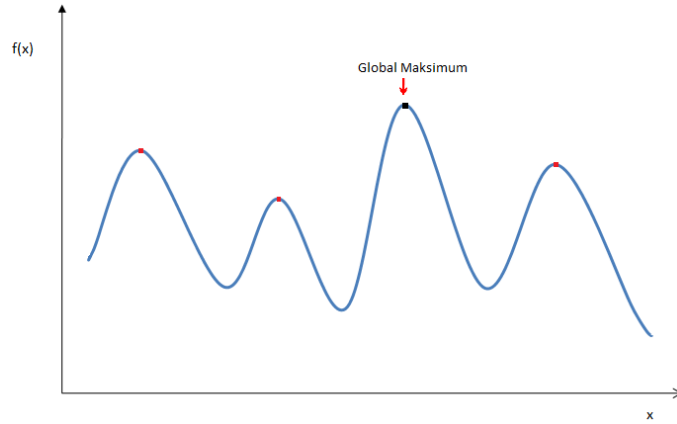
Doğrusal bir amaç fonksiyonu ve doğrusal kısıtların bulunduğu sistemlerin çözümünde değişkenlerin tümünün ve ya bir kısmının tamsayı almak zorunda oldukları durumda kullanılan bir optimizasyon yöntemidir. Doğrusal programlamadaki gibi Eşitlik 2.14'teki n ifadesi karma tamsayılı doğrusal programlama için her zaman 1 değerini almaktadır. Bu anlamda tamsayılı doğrusal programlama doğrusal programlamanın özel bir halidir [Ulucan, 2004].

Doğrusal programlamaya ek olarak, bu yöntem ile optimizasyon yapılırken problemin çözüm vektöründe (X) yer alan değişkenlerin tümünün veya bir kısmının tamsayı değer alacak şekilde hesaplamaya imkan sunmaktadır. Ayrıca; özellikle bilinmeyen sayısının fazla olduğu durumlarda; karışık-tamsayılı programlama problemini genetik algoritma kullanarak daha kısa sürede çözmek mümkündür [Patil et al., 2005].

2.5.5 Genetik Algoritma

Genetik algoritmalar, karmaşık düzenli problemlerin çözümlerini gerçekleştirmek amacıyla, kromozomların yeni diziler üretme esasını temel alan sezgisel bir araştırma ve optimizasyon yöntemidir.

Genetik algoritmaları diğer araştırma ve optimizasyon yöntemlerinden ayıran en önemli özellik bir çözüm seti ile başladıktan sonra geliştirmek için biyolojik evrimi esas alan bir algoritmanın kullanılmasıdır [Cheng et al.,1995]. Genetik algoritma matematiksel fonksiyonların global optimizasyonunu hedefler. Şekil 2.13'deki gibi fonksiyon veren ve maksimum değerinin bulunması amaçlanan bir sistem düşünüldüğünde, genetik algoritma programı kullanarak global maksimum değerini bulabiliriz. Şekil 2.13'deki fonksiyonda görüldüğü gibi bu sistem için 4 farklı bölgesel maksimum değeri vardır. Genetik algoritma programı mantığıyla amaçlanan, tüm sistemi temsil eden ve siyah ile gösterilen global maksimum noktayı bulabilmektir. Bunun için, genetik algoritma tek bir "aday çözüm vektörü" kullanmak yerine, çok sayıda aday çözüm vektörü kullanarak ilerler.



Şekil 2.13 Farklı x değerleri için bir $f(x)$ fonksiyonun değerlerinin grafiksel olarak gösterimi. Burada $f(x)$ fonksiyonun 4 farklı bölgesel maksimum değeri vardır ve bu noktalardan bir tanesi bu fonksiyonun global maksimum değeridir.

Genetik algoritma global optimizasyonu uygularken sırasıyla Şekil 2.14'teki basamakları takip eder.



Şekil 2.14 Genetik Algoritma Basamakları

Çözümlerin Kodlanması: Tüm aday çözüm vektörlerinin aynı boyutlara sahip diziler biçiminde gösterilmesidir.

İlk Popülasyonun Üretimi: Olası çözüm vektörlerinin kodlandığı bir çözüm grubu oluşturulur. Çözüm vektörleri grubu popülasyon olarak, her bir çözüm vektörü de

kromozom olarak adlandırılır. İkili sayı sisteminin kullanıldığı kromozomların gösteriminde, ilk popülasyonu oluşturmak için rastgele sayı üretilebilir (Şekil 2.15).

Kromozom 1 :	<table><tr><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td></tr></table>	1	1	0	1	0	1
1	1	0	1	0	1		
Kromozom 2 :	<table><tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr></table>	0	1	0	0	1	0
0	1	0	0	1	0		
Kromozom 3 :	<table><tr><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td></tr></table>	1	0	1	0	0	1
1	0	1	0	0	1		
Kromozom 4 :	<table><tr><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	1	1	1	0	1	1
1	1	1	0	1	1		

Şekil 2.15 Rastgele İkili Sayı Üretilerek Oluşturulan İlk Popülasyona Bir Örnek. Çözüm kümesi 6 tane elemandan oluşan bir sistemin rastgele olarak üretilen kromozomlarla temsil edilerek ilk popülasyonun oluşturulması

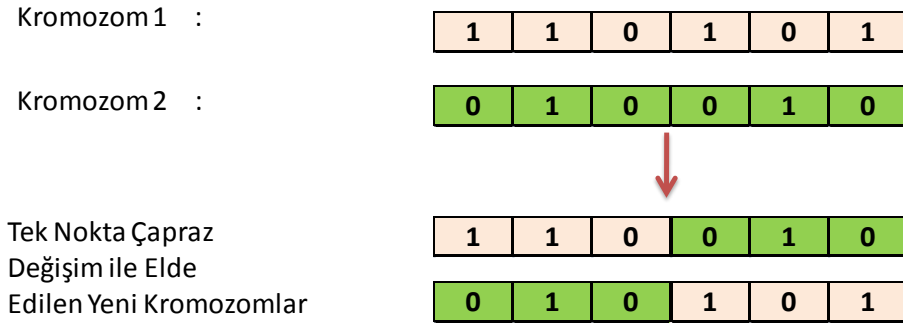
Çözüm vektörleri (kromozomlar) farklı amaçlar doğrultunda kodlanabilir. Örneğin, aranan çözüm vektöründe sıfır değerleri önemli ise, kromozomlardaki sıfır değerleri ilgili vektör elemanını sıfıra eşitlemek şeklinde kullanılarak kromozomlar ve çözümün bağlantısı kurulabilir.

Uygunluk Değerlerinin ('Fitness Value') Hesaplanması: Bu aşamada üretilen popülasyon içindeki bütün bireylere (kromozomlara) denk gelen amaç fonksiyonunun ayrı ayrı değerleri hesaplanır.

Yeni Popülasyonun Oluşturulması: Bu aşama kendi içinde çapraz değişim, mutasyon, yeniden üretim ve üstün nitelikli çocukların saklanması olmak üzere 4 bölümde incelenebilir. Genetik algoritma, popülasyondaki kromozomlara (çözüm vektörlerine) denk gelen amaç fonksiyonu değerleri hesaplandıktan sonra, bir sonraki iterasyon için yeni kromozomlar oluşturma esasına göre ilerler.

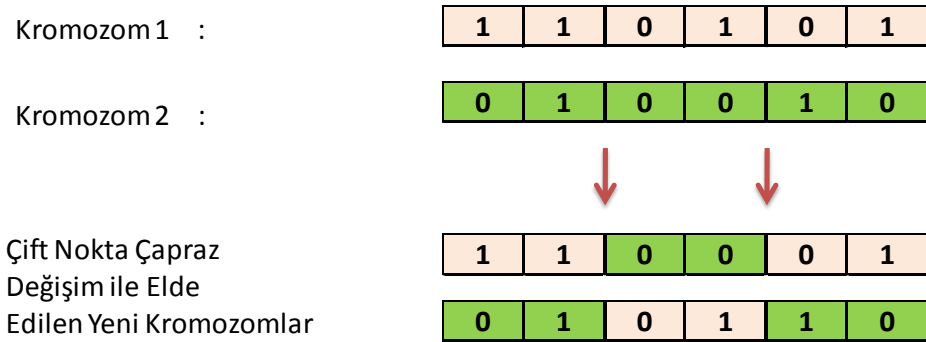
a) Çapraz değişim ('Cross-over') aşaması: Bu aşamada popülasyon içindeki bütün bireylerin uygunluk değerlerini kullanarak, uygunluk değeri daha yüksek bireylere daha fazla şans veren bir mantıkla, popülasyondaki tüm birey sayısının belirli bir sayısı (örneğin; 1/4, 1/5'i) kadar bireye çapraz değişim yapılır. Çapraz

değişim yapılacak gen 1 ve 0 arasında rastgele sayılar üretilerek belirlenir. Tek nokta ve çift nokta olmak üzere iki şekilde çapraz değişim yapılabilir. Adlarından da anlaşılacağı gibi tek nokta çapraz değişimde aşağıdaki gibi rastgele tek bir nokta seçilerek karşılıklı değişim yapılır (Şekil 2.16).



Şekil 2.16 Tek Nokta Çapraz Değişime Örnek. Popülasyonda yer alan iki kromozom arasında rastgele bir nokta belirlenerek çapraz değişim yapılması

Çift noktada ise rastgele iki nokta seçilerek iki kromozom arasında çapraz değişim yapılır (Şekil 2.17).



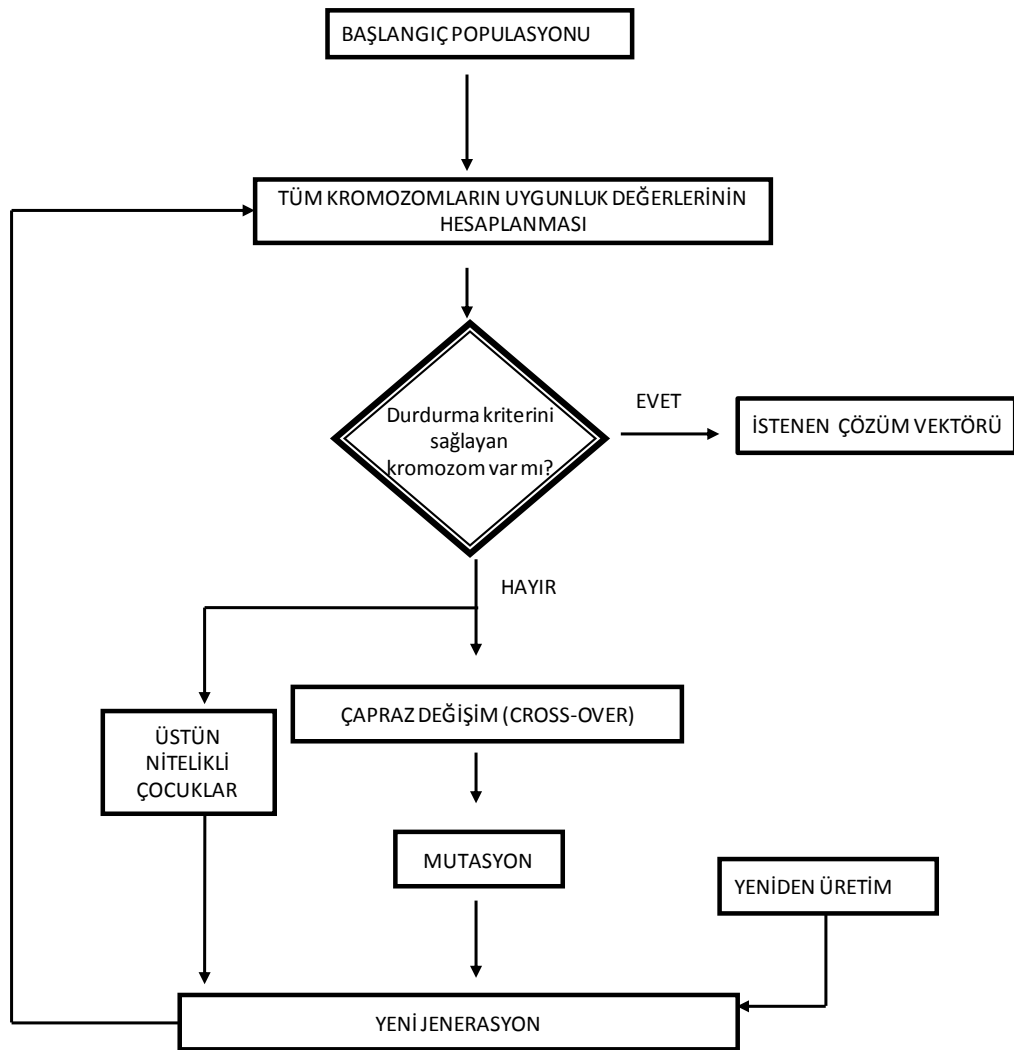
Şekil 2.17 Çift Nokta Çapraz Değişime Örnek; Popülasyonda yer alan iki kromozom arasında rastgele iki nokta belirlenerek çapraz değişim yapılması

b) Mutasyon aşaması: Çapraz değişim yapılmış tüm bireylere belli bir mutasyon oranı belirleyerek mutasyon işlemi uygulanır. Yani; rastgele bir şekilde kromozomdaki 1 değerleri 0'a, 0 değerleri de 1'e dönüştürülür.

c) Yeniden Üretim ('Fresh Blood'): Çapraz değişim ve mutasyon yapılmış bireylere ek olarak yeni kromozomlar üretilerek popülasyona eklenmesidir.

d) Üstün Nitelikli Çocukların ('Elite Children') Saklanması: Tüm popülasyon bireylerinin uygunluk değerleri karşılaştırılarak en iyi uygunluk değerini veren belirli sayıdaki bireyin hiçbir işlem yapılmadan olduğu gibi yeni popülasyona aktarılmasıdır.

Bu şekilde 4 farklı aşamadan elde edilen yeni popülasyon bireyleri bir sonraki kontrol aşamasına aktarılır. Bu işlemler, en iyi bireylerin uygunluk değerlerinin hiç değişmediği duruma kadar devam ettirilir. Tüm bu işlemler sonucunda aranan sonuca ulaşılır (Şekil 2.18).



Şekil 2.18 Genetik Algoritmanın Akış Şeması [Patil et al., 2005]; Başlangıç popülasyonunun üretilmesinden uygun sonucun bulunmasına kadar olan süreç şematik olarak gösterilmiştir.

2.5.6 Lyapunov Denklemi

Metabolik ağyapılar için optimizasyona dayalı yönlü ağyapı çıkarımı çalışmalarında kullanılan yöntemlerden biri Steuer ve arkadaşları tarafından deneysel metabolit verilerinin kovaryansının Jacobian matris ile olan ilişkisinin tanımlandığı bir eşitliktir [Steuer et al., 2003]. Bir metbolitin konsantrasyonundaki zamanla değişimi yaklaşık olarak o metabolitin Jacobian Matris ile çarpımına eşittir:

$$\frac{dX}{dt} \approx JX \quad (2.15)$$

Ayrıca metabolik ağyapıdaki bu metabolitlerin konsantrasyonları hücre içerisindeki süreçten kaynaklanan ξ_i ile Langevin-tipi eşitlik tarafından modellenen dalgalanmalara bağlıdır:

$$\frac{dX_i}{dt} = \sum_j J_{ij} X_j + \sqrt{2D} \xi_i(t) \quad (2.16)$$

Eşitlik 2.16'nın $P(X)$ kararlı çözümü bir çok değişkenli Gauss dağılımı olarak bilinir. Γ kovaryans matrisi elde etmek amacıyla, $P(x)$ için Fokker-Planck eşitliği kullanılarak Eşitlik 2.16 düzenlendi:

$$0 = -\sum_{ij} J_{ij} \frac{\partial}{\partial x_i} x_j P + \sum_{ij} D_{ij} \frac{\partial^2 P}{\partial x_i \partial x_j} \quad (2.17)$$

Burada $D_{ij} = \delta_{ij} D_i$ köşegenlerdir. Burada Γ kovaryans matrisini temsil eden ifadesine odaklanıldı:

$$\langle X_k X_l \rangle = \int x_k x_l P(x) dx \quad (2.18)$$

Eşitlik 2.17 $x_k x_l$ ile çarpılarak integral alınırsa Eşitlik 2.19 elde edilir:

$$0 = \sum_j J_{kj} \langle X_l X_j \rangle + \sum_j J_{lj} \langle X_k X_j \rangle + 2D_{kl} \quad (2.19)$$

Eşitlik 2.19, Γ kovaryans matrisi bakımından tekrar düzenlenerek Eşitlik 2.20 tanımlanabilir:

$$J \Gamma + \Gamma J^T = -2D \quad (2.20)$$

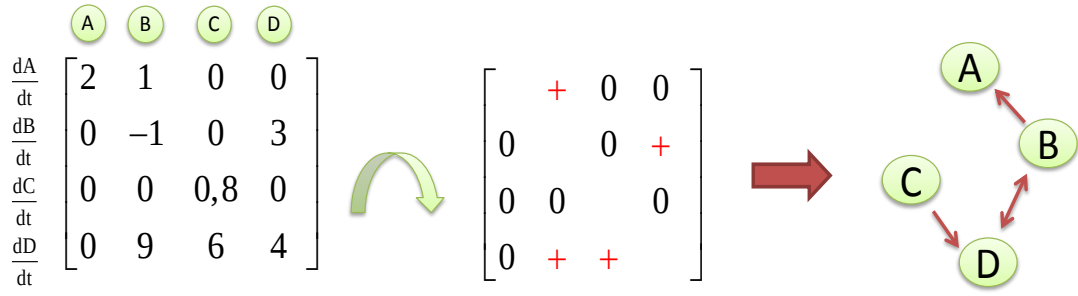
Bu eşitlik genel olarak Lyapunov eşitliği olarak bilinir. Burada D içsel ve çevresel değişkenlikte Bölüm 2.1.3'te bahsedilen zaman içindeki dalgalanma durumu ile bağlantılıdır. Dalgalanmanın sayısal olarak büyüklüğünü gösterir. Bu parametrenin matris şeklinde ifade edilmiş halidir. J , $n \times n$ boyutunda Jacobian matrisi, Γ ise $n \times n$ boyutunda deneysel verilere ait kovaryans matrisidir.

Lyapunov eşitliği, deneysel veriler kullanılarak hesaplanabilecek kovaryans matrisi ile ağyapı bilgisini içinde barındıran Jacobian matrisi arasındaki ilişkiyi vermesi açısından önemlidir. Burada amaç deneysel verilerin kovaryansını ve dalgalanma katsayısını kullanarak ağyapı bilgisini içeren Jacobian matrisi elde etmek, böylece yönlü ağyapıyı ortaya koymaktır.

Jacobian matris dinamik modelle tanımlanmış bir ağyapıdaki metabolit çiftleri arasında ilişki olup olmadığını ve varsa bu ilişkilerin yönünü ve gücünü gösterir. Jacobian matrisin (i,j) ögesi, j metabolitinin seviyesinde son derece küçük bir değişime karşılık i metabolitinin zamanla davranışının değişiminin büyüklüğünü göstermektedir [Çakır ve ark., 2009].

$$J_{i,j} = \frac{\partial \left(\frac{dC_i}{dt} \right)}{\partial C_j} \quad (2.21)$$

Bir modelin Jacobian matrisi, o modelin ağyapı bilgisini içerir. Jacobian matrisin elemanları kendi içinde değerlendirilerek gerçek ağyapı çıkarılabilir. Örneğin, Jacobian matrisinde (3,4)'e denk gelen elemanda sıfırdan farklı bir sayı olması, üçüncü metabolit ile dördüncü metabolit arasında bir ilişki olduğunu, ve bu ilişkinin yönünün üçüncüden dördüncüye olduğunu gösterir. Bu şekilde Jacobian matris bilgisi ile yönlü ağyapı çıkarılabilir (Şekil 2.19).



Şekil 2.19 Örnek Bir Jacobian Matristen Yönlü Ağyapı Çıkarımının Şematik Gösterimi

Eşitlikte yer alan kovaryans matris (Γ) her zaman simetrik bir matristir. Bu nedenle, matris $n \times n$ boyutunda olmasına rağmen simetrik terimlerin yer almasından dolayı içinde sadece $n(n+1)/2$ tane farklı değer içermektedir. Jacobian matris için ise bu durum söz konusu değildir. $n \times n$ sayıda bilinmeyen içeren Jacobian matrisini çözmek için $n(n+1)/2$ sayıda bilinen içeren bir kovaryans matrisi tanımlanmaktadır. Böyle bir sistemin serbestlik derecesi Eşitlik 2.10'da yer alan formül ile hesaplanacak olursa;

$$\text{Serbestlik Derecesi} = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2} \quad (2.22)$$

olarak bulunacaktır. Yani bu sistem için serbestlik derecesi her zaman sıfırdan büyük olacaktır. Sistem tanımlanmamış bir sistemdir. Bu nedenle Lyapunov eşitliğini çözerek Jacobian matrisi elde etmek ancak Bölüm 2.4.2-2.4.5'te bahsedilen optimizasyon yöntemleri ile mümkündür. Eşitlik 2.20, çözüm için daha basit bir formülasyona indirgenerek Eşitlik 2.23'a dönüştürülebilir.

$$A_j = -2d \quad (2.23)$$

Burada j , n metabolit içeren bir sistemde jacobian matrisin $n^2 \times 1$ boyutunda (J) bir vektöre dönüştürülmüş halidir. d , köşegenleri *in silico* metabolom verisi üretmek amacıyla içsel değişkenliği modellemek için kullanılan Eşitlik 3.3'teki spesifik ağırlık (f) değerine eşit, diğer elemanları ise sıfır olan $n \times n$ boyutundaki dalgalanma matrisinin (D) boyutu $n^2 \times 1$ olan vektörel formudur. A ise $n \times n$ boyutlu kovaryans matrisinin (Γ) bilgisini içeren $n^2 \times n^2$ boyutunda bir matristir.

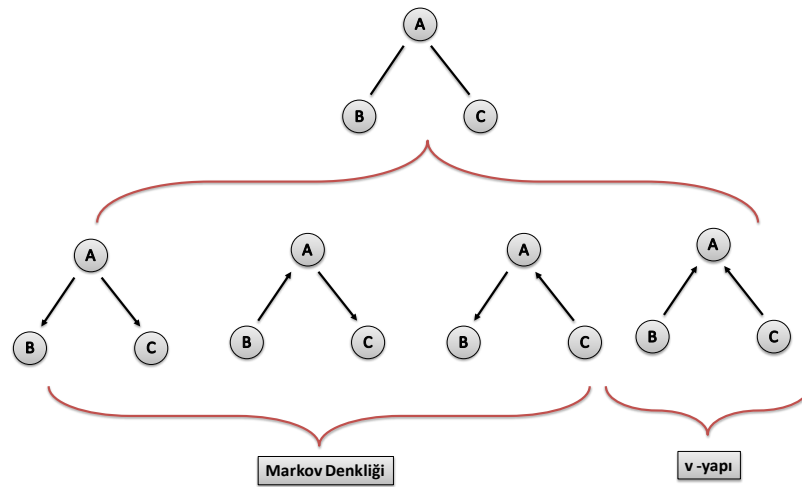
2.6 Nedenselliğe Dayalı Ağyapı Çıkarımı

Nedensellik ilkesi kapsamında ağyapı çıkarımı uygulamaları literatürde mevcuttur [Husmeier, 2003]. Nedenselliğe dayalı ağyapı çıkarımı için literatürde kısıt bazlı ('constrain-based') algoritmalar ve Bayez ağlar olarak da adlandırılan skor bazlı algoritmalar olmak üzere iki temel sınıf algoritma vardır. Bu algoritmalar ağyapıda bulunan ilişkilerin tümünü değil, bir kısmının yönünü bulabilmektedir. Nedenselliğe dayalı yapılan ağyapı çıkarımı yöntemleri nedenselliğin temel kavramlarına dayanmaktadır.

2.6.1 Temel Kavramlar

Nedenselliğe dayalı ağyapı çıkarımı için birçok farklı algoritma bulunmaktadır. Ama bu algoritmaların tümü Pearson ve farklı derecelerden kısmi korelasyon katsayılarının hesaplanarak anlam testlerinde değerlendirildikten sonra elde edilen yönsüz ağyapıyı temel almaktadır. Bu yapı üzerinden aralarında anlamlı ilişki bulunan çiftler arasındaki yön belirlenmeye çalışılır. Bu amaçla öncelikli olarak üçlü yapılar arasında v-yapılar ('v-structures') taranır.

v-yapı: Burada A-B ve C değişkenlerinden oluşan ve A-B ve A-C arası yönsüz bir ilişkinin olduğu bir yönsüz ağyapı düşünölsün:

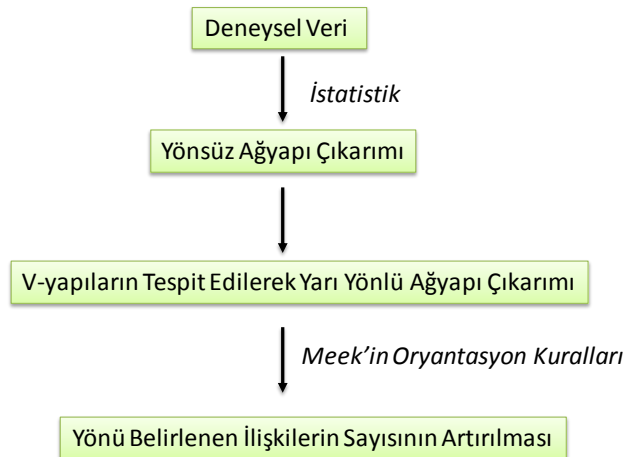


Şekil 2.20 A,B ve C değişkenlerinden oluşan bir yönsüz ağyapı örneğinin alabileceği yönlü ağyapı ihtimalleri

Şekil 2.20’da yer alan ilk üç yönlü ağyapı da B-C arasındaki kısmi korelasyonun (A’nın etkisinin çıkarıldığı) sıfır yani anlamsız olduğu durumlardır. Bu üçü arasında nedensellik yöntemleri ile ayırım yapmak mümkün olmadığından Markov Denkliği (‘Markov Equivalent’) olarak bilinirler. Ama dördüncü durum farklıdır ve bu yapı v-yapı olarak bilinmektedir. Çarpışma nodu (‘Collision node’) olarak da bilinen v-yapılar, aralarında Pearson korelasyonuna göre anlamsız, başka bir şartlandırılmış kısmi korelasyon katsayısına göre anlamlı ilişki olan değişkenlerin oluşturduğu üçlü yapılardır ve bu değişkenler arasında yön bellidir. A, B ve C değişkenlerinden oluşan Şekil 2.20’deki gibi yönsüz bir sistem düşünüldüğünde; B-C çiftleri arasında Pearson korelasyona göre ilişki yoksa ama birinci dereceden kısmi korelasyon katsayısına göre bir ilişki bulunursa bu üçlü yapı bir v-yapıdır ve bu yapının gerçek hali Şekil 2.20’deki gibidir. Buna göre temel alınan yönsüz ağyapıda tespit edilebilen yönler ağyapı üzerinde gösterilerek yönlü ağyapı elde edilmeye çalışılır. Yönsüz ağyapıdaki v-yapılar tespit edilerek belirlenen yönler ağyapıya uygulanarak yarı-yönlü döngüsüz ağyapı elde edilir.

Yönlü Döngüsüz Ağyapı (‘Directed Acyclic Graphs’, DAG): İçerisinde yönlü olarak hiçbir döngüsel yapı olmayan ağyapılardır. Nedenselliğe dayalı algoritmaların çoğu yönlü döngüsüz yapı varsayımına göre geliştirilmiştir.

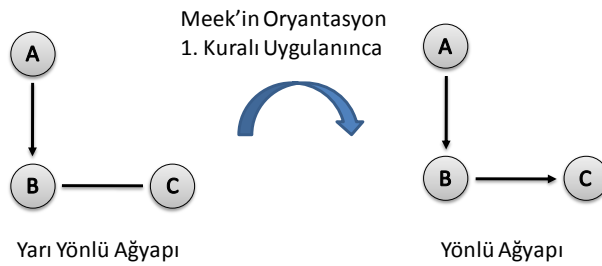
Elde edilen yarı-yönlü döngüsüz ağyapıya Meek’in oryantasyon kuralları uygulanarak yönü belirlenmemiş ilişkilerden bazılarının yönü belirlenebilir (Şekil 2.21).



Şekil 2.21 Nedenselliğe Dayalı Yönlü Ağyapı Çıkarımının Genel Bir Algoritması

Meek'in Oryantasyon Kuralları: Meek tarafından yarı yönlü ağyapılar için yön tespit etmek için bazı kurallar öne sürülmüştür [Meek,1995]. Bu kurallar v-yapılar tespit edildikten sonra yani yarı yönlü ağyapı çıkarıldıktan sonra uygulanır. Meek'in oryantasyon kuralları aşağıda örneklerle açıklanmıştır.

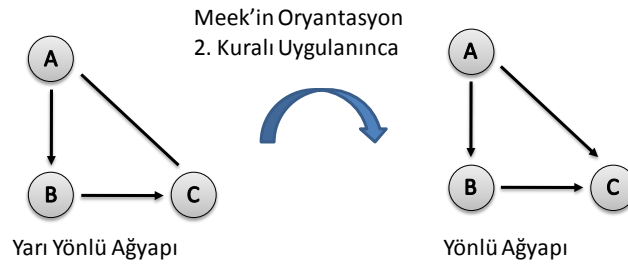
Kural 1: A,B ve C değişkenlerinin olduğu bir sistem düşünülürse; bu üç değişken arasında Şekil 2.22'deki gibi A'dan B'ye yönlü bir ilişki, B-C arasında ise yönsüz olarak bir ilişki bulunuyorsa; B-C arasındaki ilişkinin yönü B'den C'ye doğrudur (Şekil 2.22).



Şekil 2.22 Meek'in Oryantasyon Kuralı 1'in Şematik Olarak Gösterimi

Eğer B-C arasındaki ilişkinin yönü C'den B'ye olsaydı bir v-yapı oluşacaktı ve v-yapılar taranırken zaten ilişkinin yönünün tespit edilmiş olması gerekirdi. Bundan yola çıkarak "B-C arasındaki ilişkinin yönü ancak B'den C'ye doğrudur" şeklinde çıkarım yapılmaktadır.

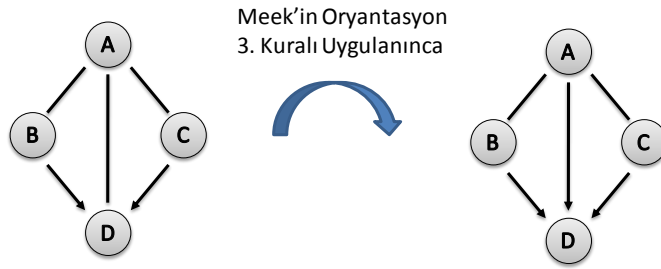
Kural 2: A,B ve C değişkenlerinin olduğu bir sistem düşünülürse; bu üç değişken arasında Şekil 2.23'deki gibi A'dan B'ye ve B'den C'ye yönlü bir ilişki, A-C arasında ise yönsüz olarak bir ilişki bulunuyorsa; A-C arasındaki ilişkinin yönü A'dan C'ye doğrudur (Şekil 2.23).



Şekil 2.23 Meek'in Oryantasyon Kuralı 2'nin Şematik Olarak Gösterimi

Eğer A-C arasındaki ilişki C'den A'ya doğru olsaydı A,B ve C arasında yönlü döngüsüz ağ yapıya (DAG) zıt olan bir döngü oluşacaktı. Bundan dolayı bu tür yapılarda yön A'dan C'ye doğru kabul edilir.

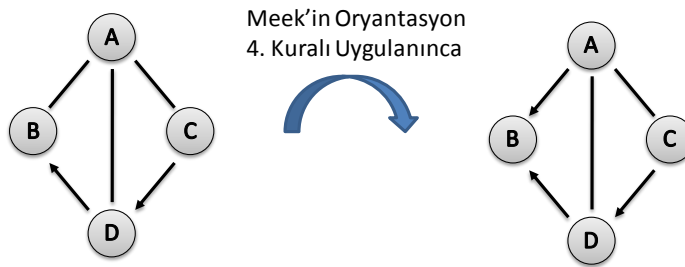
Kural 3: A,B,C ve D değişkenlerinin olduğu bir sistem düşünülürse; bu dört değişken arasında Şekil 2.24'teki gibi B'dan D'ye ve C'den D'ye yönlü bir ilişki, A-C, A-B ve A-D arasında ise yönsüz olarak bir ilişki bulunuyorsa; A-D arasındaki ilişkinin yönü A'dan D'ye doğrudur (Şekil 2.24).



Şekil 2.24 Meek'in Oryantasyon Kuralı 3'ün Şematik Olarak Gösterimi

Eğer A-D arasındaki yön D'den A'ya olsaydı bu yapıya Kural 2 uygulanacaktı ve ortaya iki yeni v-yapı çıkacaktı. Fakat v-yapılar Meek'in kuralları uygulanmadan önce tespit edildiğinden sonradan ortaya yeni bir v-yapının çıkması olası değildir. Bu nedenle yön D'den A'ya değil A'dan D'ye doğrudur.

Kural 4: A,B,C ve D değişkenlerinin olduğu bir sistem düşünülürse; bu dört değişken arasında Şekil 2.25'teki gibi D'den B'ye ve C'den D'ye yönlü bir ilişki, A-C, A-B ve A-D arasında ise yönsüz olarak bir ilişki bulunuyorsa; A-B arasındaki ilişkinin yönü A'dan B'ye doğrudur (Şekil 2.25).



Şekil 2.25 Meek'in Oryantasyon Kuralı 4'ün Şematik Olarak Gösterimi

Eğer yön B'den A'ya doğru olsaydı, Kural 2 bu yapıya uygulanacak ve ortaya ortaya iki yeni v-yapı çıkacaktı. Fakat v-yapılar daha önceden tespit edildiğinden sonradan v-yapının çıkması muhtemel olamaz. Bu nedenle yön A'dan B'ye doğrudur.

Meek'in oryantasyon kuralları bu şekilde yarı yönlü ağyapı üzerinde uygulanarak yönsüz olarak bulunan ilişkilerin yönleri bulunmaya çalışılır.

2.6.2 PC Algoritması

Spirtes ve ark. tarafından geliştirilen en genel kısıt bazlı ('constraint-based') algoritmadır [Spirtes et al., 2000]. Benzerlik değeri olarak Pearson ve kısmi korelasyon katsayısını kullanır. İlk olarak sistemdeki değişkenlerin her biri arasında ilişki var olarak kabul edilir ('Full-connected network') ve bu şekilde ağyapı çıkarımına başlanır. Olası tüm çiftler arasında Pearson korelasyon katsayısı hesaplanır. Hesaplanan korelasyon katsayıları anlam testlerinde değerlendirilerek anlamsız olanlar arasındaki ilişkiler ağyapıdan kaldırılır. Daha sonra yine olası tüm çiftler arasında birinci dereceden veya daha yüksek bir dereceden kısmi korelasyon katsayıları hesaplanır, aynı şekilde anlam testlerinde değerlendirilerek anlamsız olan ilişkiler ağyapıdan kaldırılır. Sonunda yönsüz ağyapı ('undirected graph') elde edilir. Bu ağyapı nedensellik analizi için temel olarak kullanılır. Öncelikle ağyapıdaki v-yapılar ('collision node') kontrol edilir. Bir sonraki aşamada ise Meek'in oryantasyon kuralları [Meek, 1995] uygulanarak yönsüz olan yapılar arasında yön tespit edilmeye çalışılır. Bu şekilde algoritma yönlü olarak ağyapıya ulaşmayı amaçlar.

2.6.3 Döngülü Nedensel Keşif ('Cyclic Causal Discovery', CCD) Algoritması

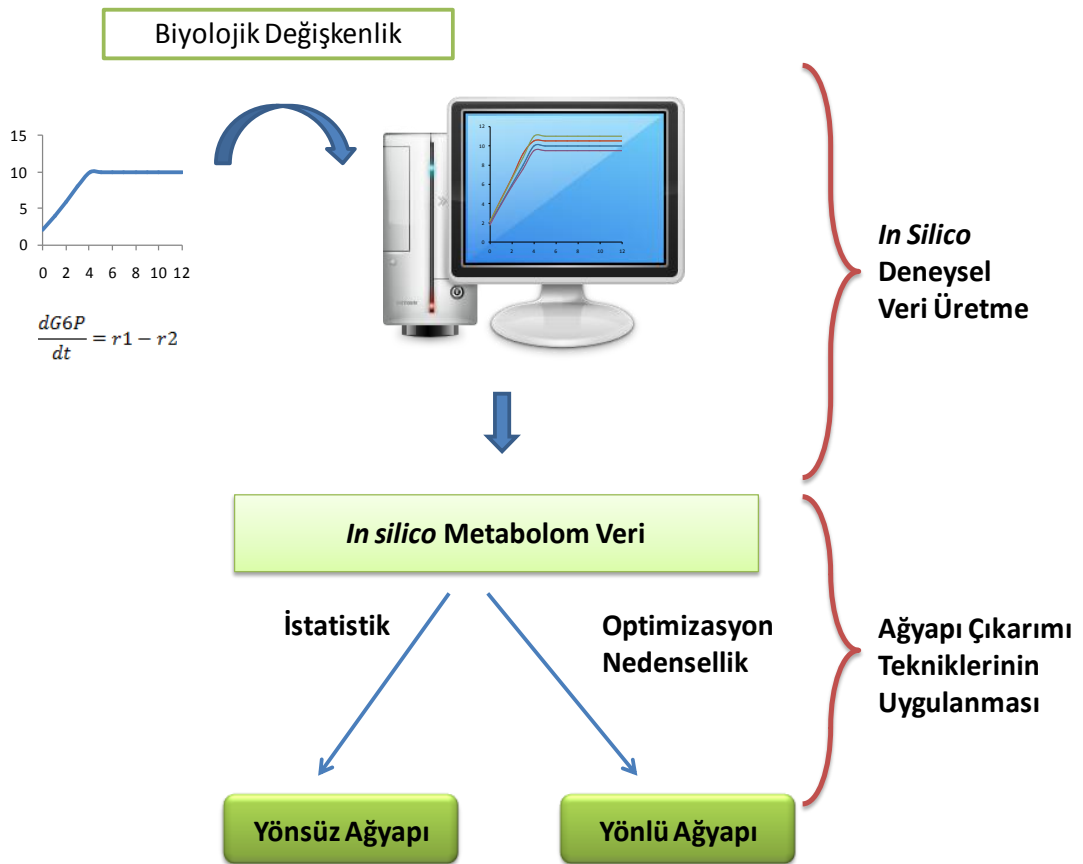
PC algoritma yönlü döngüsüz ağyapı çıkarımı için geliştirilmiştir. Yani, geribesleme sistemleri gibi döngüleri yakalayamamaktadır. Bunun için CCD olarak adlandırılan ve döngüsel ağyapıları açığa çıkarabilen bir algoritma geliştirilmiştir [Ricardson, 1996]. CCD algoritması kısıt bazlı bir algoritmadır.

2.6.4 Açgözlü Eşitlik Arama ('Greedy Equivalent Search', GES) Algoritması

GES algoritması skor-bazlı yeni nesil 'greedy search' algoritmasıdır. Yönlü döngüsüz ağyapılar üzerinden arama yapmak yerine, daha hızlı şekilde sonuca ulaştıran yönlü döngüsüz yapı kalıpları üzerinden arama yapar. Algoritma, ileri ve geri tarama olmak üzere iki tarama basamağından oluşur. İleri taramada, skoru artıran ek kenarları ('edge') arar. Skoru artırabilecek kenar kalmayınca, geri tarama başlar. Geri taramada, kaldırıldığında skoru artıran ilişkiler aranır. Skoru artırabilecek ilişkilerin kaldırılması tamamlanınca program durur [Chickering, 2002].

3 METOT

Yapılan tez çalışmasında, *Escherichia coli* ve *Saccharomyces cerevisiae* mikroorganizmalarına ait metabolom verilerine ağyapı çıkarımı teknikleri uygulanarak hücrenin biyolojik amacının incelenmesi amaçlanmıştır. Bu amaçla yapılan tez çalışması genel olarak deneysel metabolom verilerinin *in silico* olarak üretilmesi ve bu verilere ağyapı çıkarımı tekniklerinin uygulanması olmak üzere iki aşamadan oluşmaktadır (Şekil 3.1).



Şekil 3.1 Yapılan Tez Çalışmasında İzlenen Yolun Şematik Olarak Gösterimi

Öncelikle dinamik modeller üzerinde değişiklikler yapılarak farklı biyolojik değişkenlikler için *in silico* deneysel veriler üretilmiş, daha sonra ise bu verilere yönlü ve yönsüz ağyapı çıkarımı teknikleri uygulanarak sonuçlar elde edilmiştir. Elde edilen sonuçlar değerlendirilerek hücrenin biyolojik amacı belirlenmeye çalışılmıştır. Aynı zamanda, kullanılan ağyapı çıkarımı tekniklerinin analiz kabiliyetleri karşılaştırılarak en etkili yöntem belirlenmeye çalışılmış ve yöntemlerin

birbirine göre avantaj ve dezavantajları ortaya çıkarılmıştır. Ayrıca, bu yöntemlerin haricinde genetik algoritma tabanlı yeni bir yöntem önerilmiş ve bu yöntem ile yönlü ağyapı çıkarımı çalışmaları yapılmıştır. Bu bölüm, yapılan tüm bu çalışmaların yöntemsel ayrıntılarını kapsamaktadır.

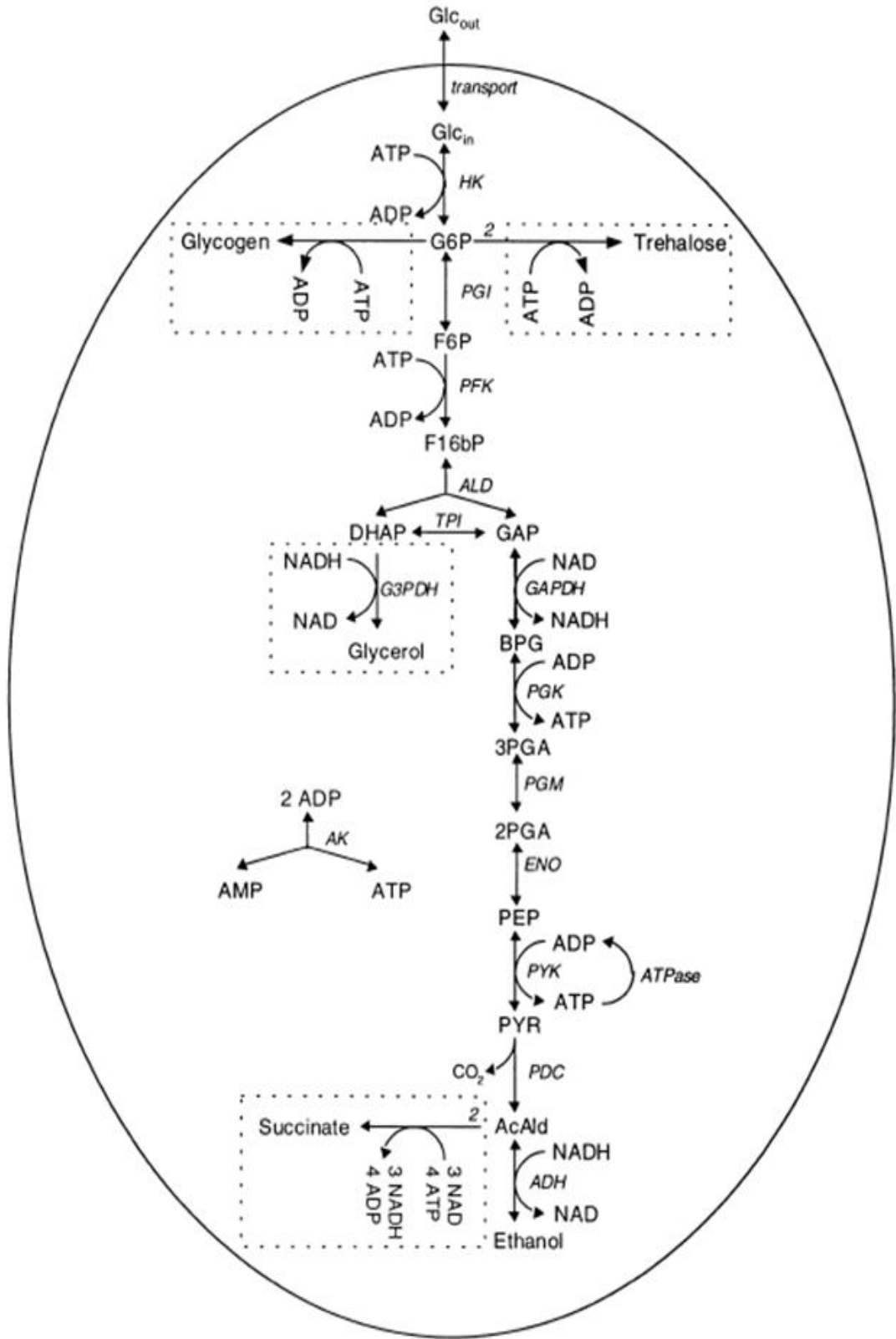
3.1 *In silico* Metabolom Verilerinin Üretilmesi

Yapılan tez çalışmasında, gerek literatürde yeterli sayıda metabolom verisi bulunmamasından dolayı gerekse uygulanan ağyapı çıkarımı tekniklerinin analiz kabiliyetlerini ortaya koymak adına deneysel veriler dinamik modeller kullanılarak *in silico* olarak üretilmiştir. Bu bağlamda, mikroorganizmalara dayalı bilimsel çalışmalarda en çok çalışılan iki organizma olan *Escherichia coli* ve *Saccharomyces cerevisiae* mikroorganizmalarına ait dinamik modeller kullanılmış, Bölüm 2.1.3'te anlatılan farklı biyolojik değişkenlik türleri için *in silico* metabolom verisi üretilmiştir. Çalışmalar aşağıdaki bölümlerde daha detaylı olarak anlatılmıştır.

3.1.1 Kullanılan Dinamik Modeller

Yapılan tez çalışmasında iki farklı tür mikroorganizma ele alınmıştır. Bir ökaryot olan *Saccharomyces cerevisiae* mikroorganizması için Teusink ve ark. tarafından geliştirilen 13 metabolit ve 18 reaksiyondan oluşan Şekil 3.2'deki glikoliz yoluna ait dinamik model kullanılmıştır [Teusink et al., 2000].

Mikroorganizmada etil alkol üretimi ile birlikte süksinat, glikojen ve trihaloz üretilmektedir. Model içerisinde NADH-NAD ve ATP-ADP-AMP dengeleri mevcuttur. Matematiksel bunun anlamı olarak açıklanırsa, NAD ve NADH metabolitlerin konsantrasyonlarının toplamı ve ATP-ADP-AMP metabolitlerinin konsantrasyonlarının toplamı zamana bağlı olarak değişmemekte her zaman sabit kalmaktadır. Biyolojik olarak anlamı ise, mikroorganizmanın bu dengeyi kendiliğinden sağlamasıdır. Örneğin ATP'nin artması durumunda AMP ve ADP metabolitlerinin konsantrasyonları mikroorganizma tarafından azaltılarak dengelenmektedir. Modelde NAD ve NADH metabolitlerinin konsantrasyonlarının toplamı 1,59 mM, ATP-ADP-AMP metabolitlerinin toplamı ise 4,1 mM'dir



Şekil 3.2 Teusink ve ark. tarafından geliştirilen *Saccharomyces cerevisiae* mikroorganizmasına ait metabolik ağyapı

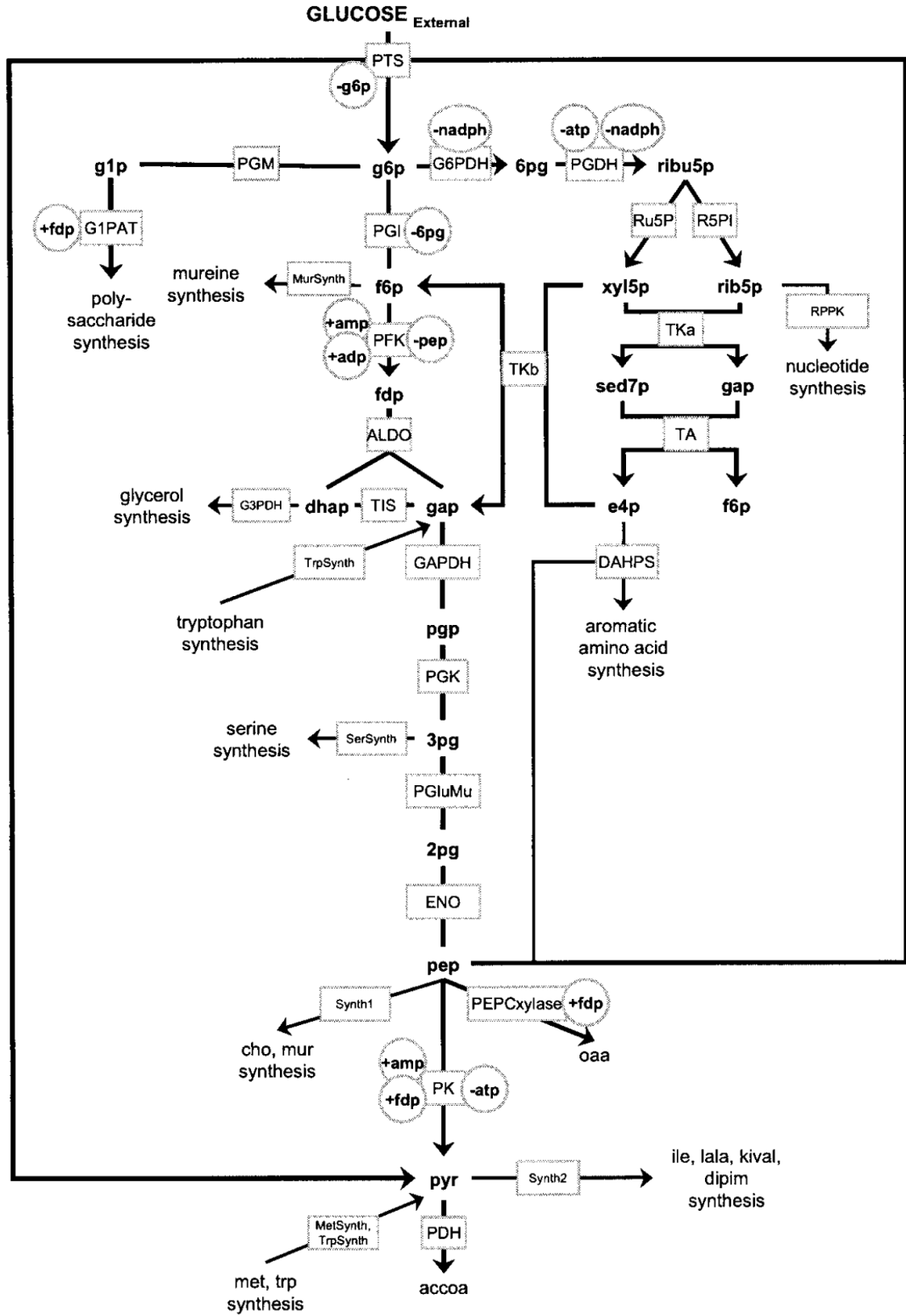
Glikoz yolu içerisinde yer alan 13 metabolit arasında yönsüz olarak toplam 21 ilişki bulunmaktadır. Dinamik modelde yer alan metabolitlerin isimleri ve kararlı hal ('steady state') konsantrasyonları Tablo 3.1'de yer almaktadır.

Tablo 3.1 Kullanılan *Saccharomyces cerevisiae* mikroorganizmasına ait modelde yer alan 13 metabolitin isimleri ve kararlı hal konsantrasyonunun milimolar (mM) cinsinden değerleri.

Metabolit İsmi	Kısaltması	Konsantrasyon (mM)
Glikoz	GLC	0,09874
Glikoz-6-fosfat	G6P	1,03346
Fruktoz-6-fosfat	F6P	0,11285
Fruktoz-1,6-bifosfat	F16P	0,60195
D-gliserinaldehit-3-fosfat	GAP	0,77756
1,3-bifosfoglisarat	BGP	0,00033
3-fosfoglisarat	3PG	0,35652
2-fosfoglisarat	2PG	0,04485
Fosfoenolpirüvat	PEP	0,07363
Pirüvat	PYR	8,52341
Asetaldehit	ACAL	0,17012
Fosfat	P	6,30940
Nikotinamidadenin dinükleotit(ind)	NADH	0,04444

Modele ait metabolitlerin zamanla değişimini içeren adi diferansiyel denklemler EK-1'de verilmiştir.

Yapılan çalışmalarda bir prokaryot olan *Escherichia coli* mikroorganizması için ise Chassagnole ve ark. tarafından geliştirilen 18 metabolit ve 30 reaksiyondan oluşan merkezi karbon metabolizmasına ait dinamik model kullanılmıştır [Chassagnole et al., 2002]. Şekil 3.3'te yer alan merkezi karbon metabolizması içerisinde yer alan 18 metabolit arasında yönsüz olarak 39 ilişki bulunmaktadır. Model içerisinde NADH, NADP, NADPH, ATP, ADP ve AMP metabolitlerinin konsantrasyonları zamana bağlı değildir ve bu metabolitlerin konsantrasyonları sırasıyla 0,1 mM, 0,195 mM, 0,062 mM, 4,27 mM, 0,595 mM ve 0,955 mM'dır. Bu nedenle bu metabolitlerin konsantrasyonlarının zamanla değişimini içeren adi diferansiyel denklem eşitlikleri model içerisinde yer almamaktadır.



Şekil 3.3 Chassagnole ve ark. tarafından geliştirilen *Escherichia coli* mikroorganizmasına ait metabolik ağıyapı

Dinamik modelde yer alan metabolitlerin isimleri ve kararlı hal ('steady state') konsantrasyonları Tablo 3.2'de yer almaktadır.

Tablo 3.2 Kullanılan *Escherichia coli* mikroorganizmasına ait modelde yer alan 18 metabolitin isimleri ve kararlı hal konsantrasyonlarının milimolar (mM) cinsinden değerleri.

Metabolit İsmi	Kısaltması	Konsantrasyon (mM)
Glikoz	GLC	0,05535
Glikoz-6-fosfat	G6P	3,47532
Fruktoz-6-fosfat	F6P	0,59919
Fruktoz-1,6-difosfat	FDP	0,26964
Gliseraldehit-3-fosfat	GAP	0,21701
Dihidroksiasteonfosfat	DHAP	0,16625
1,3-difosfoglisarat	PGP	0,00798
3-fosfoglisarat	3PG	2,12514
2-fosfoglisarat	2PG	0,39791
Fosfoenolpirüvat	PEP	2,66270
Pirüvat	PYR	2,66849
6-fosfoglisarat	6PG	0,80661
Riboloz-5-fosfat	RIBU5P	0,11077
Ksilüloz-5-fosfat	XY15P	0,13770
Sedoheptüloz-7-fosfat	SED7P	0,27598
Riboz-5-fosfat	RIB5P	0,39715
Eritroz-4-fosfat	E4P	0,09766
Glikoz-1-fosfat	G1P	0,65186

Modele ait metabolitlerin zamanla değişimini içeren adi diferansiyel denklemler EK-2'te verilmiştir.

Modeller enzimatik değişkenlik, içsel değişkenlik ve içsel-çevresel değişkenlikler olmak üzere farklı biyolojik değişkenlikler için modellenerek üç tip metabolom verisi üretilmiştir:

1. Enzimatik Değişkenliğin Modellenmesi ile Üretilen Metabolom Verisi
2. İçsel Değişkenliğin Modellenmesi ile Üretilen Metabolom Verisi
3. Hem İçsel Hem de Çevresel Değişkenliğin Birlikte Modellenmesi ile Üretilen Metabolom Verisi

3.1.2 Enzimatik değişkenliğin modellenmesi ile Metabolom Verilerinin Üretilmesi

Enzimatik değişkenlik olarak adlandırdığımız biyolojik değişkenlik türü, mikroorganizma popülasyonları arası hız sabiti (k) veya enzim konsantrasyonunun çok az bir değişiminden kaynaklanmaktadır. Dinamik modellerdeki her bir tepkime hızı (r) ifadesi bu iki etkinin her ikisinden oluşan bir V_{max} parametresine sahiptir:

$$\frac{dC_i}{dt} = \sum r_i \quad ; \quad r = f(C)V_{max,i} \quad ; \quad V_{max,i} = k_i [E] \quad (3.1)$$

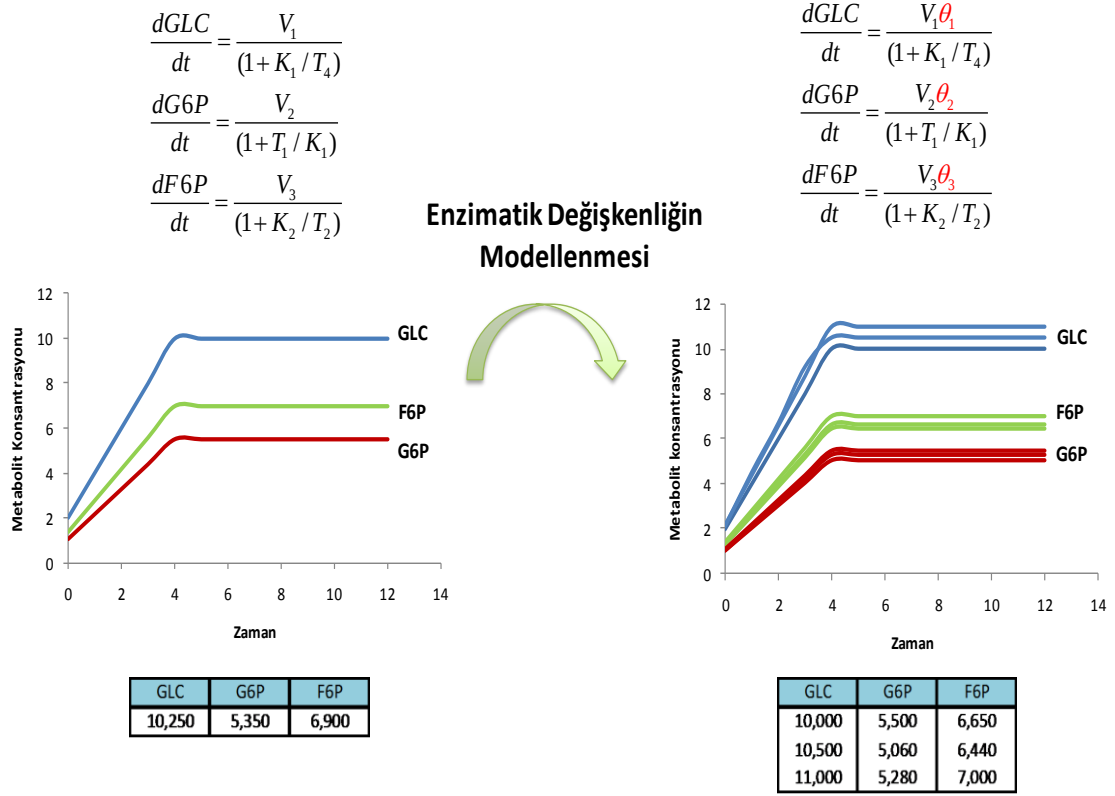
Bu yüzden enzimatik değişkenliği dinamik modeller üzerinde modellemek için $V_{max,i}$ ifadeleri belirli bir standart sapma ve ortalamaya sahip normal dağılım gösteren rastgele sayılardan oluşan bir θ_i ifadesi ile çarpılarak relatiflik sağlandı:

$$V_{max,i} = k_i [E] \theta_i \quad (3.2)$$

Farklı standart sapmalara sahip θ_i ifadeleri için çalışmalar gerçekleştirildi fakat standart sapması 0,05'ten küçük değerler için veriler birbirine çok yaklaştığı ve gerçek deneysel ölçümleri yansıtmadığı görüldü. Standart sapması 0,05'ten daha büyük kullanılan θ_i ifadelerinin ise, verilerde çok büyük sapmalara neden olduğu ve tekrarlanan deneylerde gözlemlenen sapmaların çok üzerinde veri dağılımı verdiği saptandı. Bu nedenle gerçekleştirilen çalışmalar sonucunda her iki mikroorganizma modeli için de deneysel ölçümleri en iyi şekilde yansıtan ortalaması 1, standart sapması 0,05 olan normal dağılım gösteren θ_i ifadesi kullanıldı. Bu ifadenin dinamik modele eklenmesi ile $V_{max,i}$ değerlerinin her yeni simülasyonda belirli oranlarda (~%5) değişim göstermesi sağlanmış oldu. Bu şekilde dinamik modeller ard arda 1000 kez eş zamanlı olarak MATLAB'ın içerisinde bulunan adi diferansiyel denklem çözücülerden ode15s (euler metodu) ile sayısal olarak çözdürülerek kararlı hal metabolit konsantrasyonları toplandı (Şekil 3.2).

Enzimatik değişkenliğin modellenmesi ile elde edilen metabolom verileri Tablo 3.1 ve 3.2'de yer alan kararlı hal metabolit konsantrasyonları etrafında ufak

farklılıklar göstermektedir. Bu şekilde, hem *Escherichia coli* hem de *Saccharomyces cerevisiae* modellerinde yer alan her metabolit için 1000 metabolom verisi üretilmiş oldu.



Şekil 3.4 Enzimatik Değişkenliğin Simulasyon Örneği

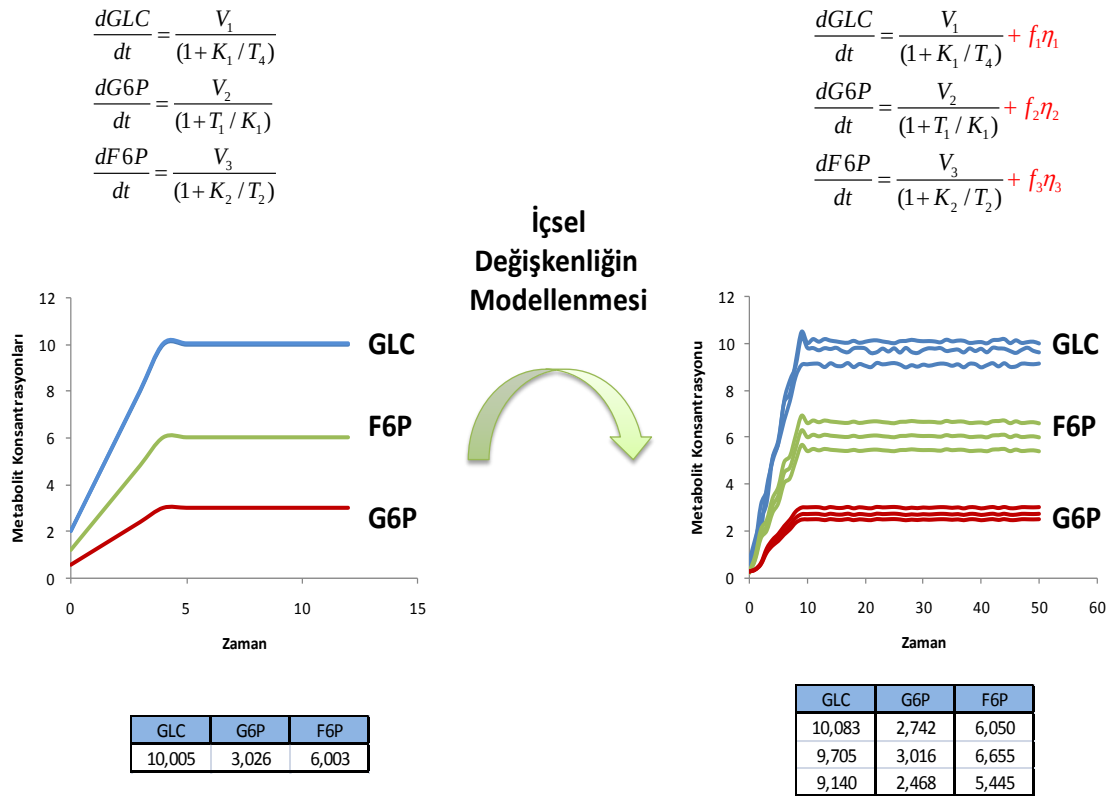
3.1.3 İçsel Değişkenliğin Modellenmesi ile Metabolom Verilerinin Üretilmesi

İçsel değişkenlik, hücre içerisindeki metabolitlerde zamanla dalgalanmalara neden olmaktadır. Bu tür dalgalanmaları matematiksel olarak modellemek, model içerisine stokastik bir terim eklemek ile mümkündür. Bu amaçla, her bir metabolit konsantrasyonunun zamanla değişimini içeren adi diferansiyel denkleme stokastik bir terim ($f_i \times \eta_i$) eklendi:

$$\frac{dC_i}{dt} = \sum r_i + f_i \eta_i ; r = f(C) V_{\max,i} ; V_{\max,i} = k_i [E] \quad (3.3)$$

Böylelikle adi diferansiyel denklem sistemi stokastik diferansiyel denklem takımına dönüştürüldü. Burada η_i , ortalaması 0, standart sapması 1 olan normal dağılım gösteren rastgele bir sayıdır. f ise spesifik ağırlıktır. Bu ağırlık *Escherichia coli* modeli için 0,01, *Saccharomyces cerevisiae* modeli için ise 0,1 olarak kullanılmıştır.

Modeller düzenlenerek oluşturulan stokastik denklem takımları, stokastik denklem takımını çözen ve MATLAB altında çalışan SDE (Stochastic Differential Equation) yazılım paketi ('toolbox') kullanılarak çözdürüldü [Picchini, 2007]. Eş zamanlı olarak SDE yazılım paketi ile 1000 kez üst üste denklem takımı çözdürülerek her seferde kararlı halde metabolom verileri toplandı. Bu sayede, iki mikroorganizma için de her bir metabolite ait 1000 tane metabolom verisi üretilmiş oldu (Şekil 3.3).



Şekil 3.5 İçsel Değişkenliğin Simulasyon Örneği

Üretilen metabolom verileri Tablo 3.1 ve 3.2’de yer alan kararlı hal metabolit konsantrasyonları etrafında ufak farklılıklar göstermektedir.

3.1.4 Hem İçsel Hem de Çevresel Değişkenliğin Birlikte Modellenmesi ile Metabolom Verilerinin Üretilmesi

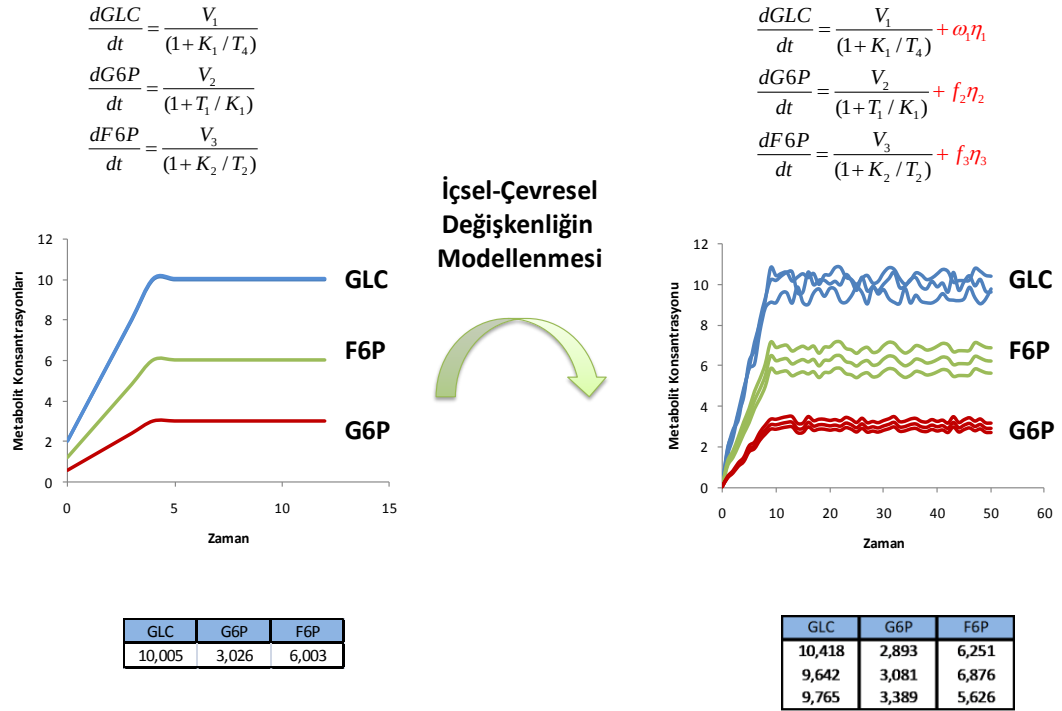
Bölüm 2.1.3'te anlatıldığı gibi, çevresel değişkenlik, hücre dışı substrat/besin konsantrasyonlarındaki küçük değişimlerden dolayı hücre içi metabolit seviyelerinde değişkenliğe neden olmaktadır. Çevresel değişkenlik, içsel değişkenlikle doğrudan bağlantılı olduğundan tek başına çevresel değişkenliği modellemek yerine içsel değişkenlikle birlikte modellemek tercih edilmiştir. Bu amaçla içsel değişkenlik için modelde yapılan düzenlemelere ek olarak çevresel değişkenliği modellemek amacıyla sadece dışarıdan hücre içine taşınan metabolite yani glikoza ait diferansiyel denkleme farklı bir spesifik ağırlık eklendi:

$$i=1 : \quad \frac{dC_i}{dt} = \sum r_i + \omega \eta_i ; \quad r = f(C)V_{\max,i} ; \quad V_{\max,i} = k_i [E] \quad (3.4)$$

Burada ω glikoza ait spesifik ağırlıktır ve bu ağırlık *Escherichia coli* modeli için 0,025, *Saccharomyces cerevisiae* modeli için ise 0,25 olarak kullanılmıştır. Bu ağırlık içsel değişkenlikte kullanılan η_i 'ye göre daha baskın olacak şekilde ayarlanmıştır. Diğer diferansiyel denklemlere ait ifadeler içsel değişkenlik için yapılan düzenlemelerde olduğu gibidir:

$$i \neq 1 : \quad \frac{dC_i}{dt} = \sum r_i + f_i \eta_i ; \quad r = f(C)V_{\max,i} ; \quad V_{\max,i} = k_i [E] \quad (3.5)$$

Yine burada daha önceden bahsedildiği gibi, η_i stokastik terimi ortalaması 0 , standart sapması 1 olan normal dağılım gösteren rastgele bir sayı, f ise spesifik ağırlıktır. Bu ağırlık içsel değişkenlik için kullanıldığı gibi *Escherichia coli* modeli için 0,01, *Saccharomyces cerevisiae* modeli için ise 0,1 olarak kullanılmıştır. Modeller düzenlenerek tekrar oluşturulan denklem takımları, stokastik denklem takımını çözen ve MATLAB altında çalışan SDE yazılım paketi kullanılarak çözdürüldü. Eş zamanlı olarak SDE yazılım paketi ile 1000 kez üst üste denklem takımı çözdürülerek her seferde kararlı halde metabolom verileri toplandı. Bu sayede, iki mikroorganizma için ayrı her bir metabolite ait 1000 tane metabolom verisi üretilmiş olundu (Şekil 3.4).



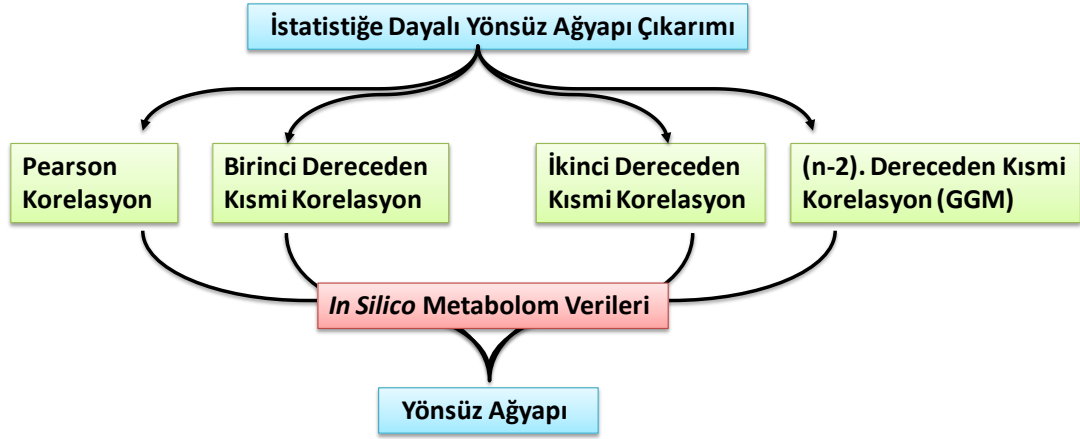
Şekil 3.6 İçsel-Çevresel Değişkenliğin Simulasyon Örneği

Üretilen metabolom verileri Tablo 3.1 ve 3.2’de yer alan kararlı hal metabolit konsantrasyonları etrafında ufak farklılıklar göstermektedir.

3.2 İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımı Tekniklerinin *In Silico* Metabolom Verilerine Uygulanması

Yapılan çalışmada *Escherichia coli* ve *Saccharomyces cerevisiae* için üretilen *in silico* metabolom verilerine istatistik, optimizasyon ve nedensellik bazlı ağyapı çıkarımı teknikleri uygulanarak gerçek ağyapıya ulaşılmaya çalışıldı.

Bölüm 3.1’de anlatıldığı gibi *Escherichia coli* ve *Saccharomyces cerevisiae* için enzimatik değişkenliğin, içsel değişkenliğin ve hem içsel hem de çevresel değişkenliğin modellenmesi ile üretilen 3 tip metabolom verisine farklı korelasyon teknikleri uygulanarak yönsüz olarak gerçek ağyapıya ulaşılmaya çalışılmış ve bu yöntemlerin analiz kabiliyetleri ölçülmüştür. Bu doğrultuda metabolom verilerine Pearson korelasyonu, birinci dereceden, ikinci dereceden ve (n-2)’nci dereceden kısmi korelasyon olmak üzere dört farklı yöntem uygulanmıştır (Şekil 3.7).



Şekil 3.7 İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımında Kullanılan Yöntemler

Escherichia coli ve *Saccharomyces cerevisiae* için yapılan hesaplamalara dair sonuçlar ise Bölüm 4 ve Bölüm 5’te verilmiştir.

3.2.1 Pearson Korelasyon Yönteminin *In silico* Metabolom Verilerine Uygulanması

Pearson korelasyonu ile ilgili detaylı bilgiye Bölüm 2.3.2’de yer verilmişti. Her iki mikroorganizma için *in silico* olarak üretilen metabolom verilerine Pearson korelasyonu uygulanmıştır. Bu amaçla, MATLAB’ın istatistik paketinde yer alan *corr* fonksiyonu kullanılarak korelasyon katsayısı ve tahmin değerleri hesaplamaları gerçekleştirilmiştir. Yapılan hesaplamalar sonucu elde edilen tahmin değerleri (‘p-value’) belirlenen hata değeri (‘cut-off value’) ile değerlendirilerek metabolit çiftleri arasında anlamlı olan ilişkiler tespit edilerek yönsüz ağyapı öngörülme çalışılmıştır. Hata değeri yapılan tüm hesaplamalarda 0,01 olarak kullanılmıştır.

3.2.2 Birinci Dereceden Kısmi Korelasyon Yönteminin *In silico* Metabolom Verilerine Uygulanması

In silico metabolom verilerine birinci dereceden kısmi korelasyon yöntemi uygulanmış ve metabolit çiftleri arasındaki kısmi korelasyon katsayıları ile bunlara ait tahmin değerleri MATLAB istatistik paketinde yer alan *partialcorr* fonksiyonu ile hesaplanmıştır.

n değişkenin bulunduğu bir sistemde iki metabolit arasındaki ilişkiye etki edebilecek $n-2$ tane metabolit vardır. Bu nedenle iki metabolit çifti arasında birinci dereceden kısmi korelasyon katsayısı o çift için her bir metabolit ayrı ayrı şartlandırılarak hesaplanmış ve bunun sonucunda $n-2$ tane kısmi korelasyon katsayısı ile bunlara ait $n-2$ tane tahmin değeri bulunmuştur. Fakat değerlendirmede bir metabolit çifti için tek bir korelasyon katsayısı ve tahmin değeri kullanılacağından, o çiftte ait $n-2$ tane korelasyon değerinden en küçüğü bu çift arasındaki korelasyon katsayısı olarak kabul edildi, böylece en çok etkisi olanın dolaylı etkisi çıkarılmış oldu. Bu şekilde her bir metabolit çifti için korelasyon katsayılarının en küçükleri seçilerek bunlara karşılık gelen tahmin değerleri, hata değerine göre değerlendirilerek aralarında anlamlı ilişki bulunan metabolit çiftleri tespit edilmiş ve yönsüz ağyapı çıkarılmıştır. Hata değeri 0,01 olarak kullanılmıştır.

3.2.3 İkinci Dereceden Kısmi Korelasyon Yönteminin *In silico* Metabolom Verilerine Uygulanması

In silico metabolom verilerine ikinci dereceden kısmi korelasyon yöntemi uygulanmış ve metabolit çiftleri arasındaki kısmi korelasyon katsayıları ile bunlara ait tahmin değerleri birinci dereceden korelasyonda olduğu gibi MATLAB istatistik paketi içerisinde yer alan *partialcorr* fonksiyonu ile hesaplanmıştır.

n değişkenin bulunduğu bir sistemde iki metabolit arasındaki ilişkiye etki edebilecek $n-2$ 'nin 2'li kombinasyonlarının sayısı kadar metabolit çifti vardır. Bu nedenle birinci dereceden kısmi korelasyon yönteminde olduğu gibi her bir metabolit çifti arasında birden fazla kısmi korelasyon katsayısı ve tahmin değeri hesaplanılmıştır. Değerlendirme yapabilmek için bir metabolit çiftine ait $(n-2)(n-3)/2$ tane ikinci dereceden kısmi korelasyon değerinden en küçüğü bu çift arasındaki korelasyon katsayısı olarak kabul edilmiş, böylece en çok etkisi olan çiftin dolaylı etkisi çıkarılmıştır. Bu şekilde her bir metabolit çifti için korelasyon katsayılarının en küçükleri seçilerek bunlara karşılık gelen tahmin değerleri, hata değerine göre değerlendirilerek aralarında anlamlı ilişki bulunan metabolit çiftleri tespit edilmiş ve buna bağlı olarak yönsüz ağyapı çıkarılmıştır. Anlam testi için hata değeri 0,01 olarak kullanılmıştır.

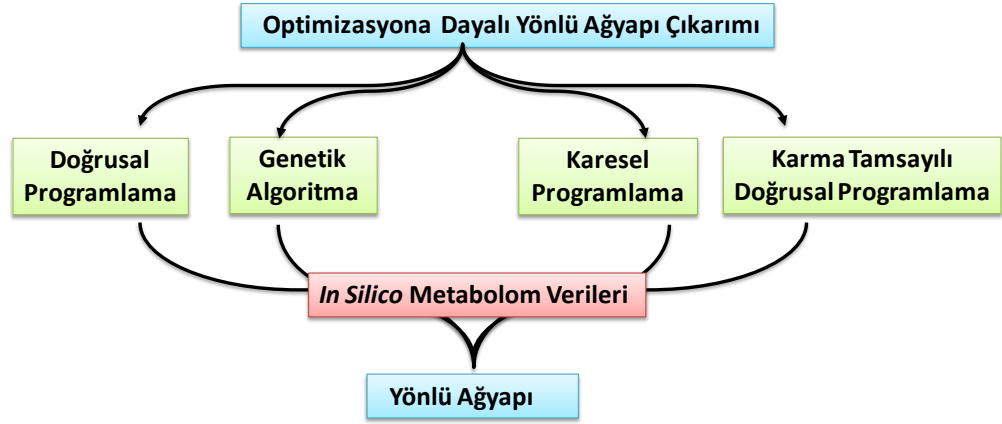
3.2.4 (n-2)'nci dereceden Kısmi Korelasyon Yönteminin *In Silico* Metabolom Verilerine Uygulanması

Modellere ait *in silico* verilere (n-2)'nci dereceden kısmi korelasyon yöntemi uygulanarak metabolit çiftleri arasındaki kısmi korelasyon katsayıları hesaplanmıştır. Korelasyon katsayı hesabı için birinci ve ikinci dereceden kısmi korelasyon yöntemlerinde izlenen yolu takip etmek yerine, daha basit bir şekilde (n-2)'nci dereceden kısmi korelasyon katsayılarını hesaplamaya olanak sağlayan GGM yöntemi kullanılmıştır. Bu amaçla MATLAB'ın içerisinde bulunan *corr* fonksiyonu ile *in silico* verilere ait Pearson korelasyon katsayılar matrisi hesaplanmış ve bu matris temel alınarak Bölüm 2.3.4'te anlatıldığı gibi GGM korelasyon katsayı matrisi elde edilmiştir.

Metabolit çiftleri arasında GGM yöntemi ile hesaplanan kısmi korelasyon katsayılarının tahmin değerleri bilinmemektedir. Bu nedenle tahmin değerleri, rastgele sıralama testi (Bölüm 2.3.6.3) için MATLAB'da yazılan bir fonksiyon yardımı ile hesaplanmıştır (EK 3). Elde edilen tahmin değerleri, hata değeri 0,01 seçilerek değerlendirilmiş ve aralarında anlamlı ilişki bulunan metabolit çiftleri tespit edilerek yönsüz ağyapı çıkarımı yapılmıştır.

3.3 Optimizasyona Dayalı Yönlü Ağyapı Çıkarımı Tekniklerinin *In Silico* Metabolom Verilerine Uygulanması

Yapılan çalışmada *Escherichia coli* ve *Saccharomyces cerevisiae* mikroorganizmalarına ait içsel değişkenlik için üretilen *in silico* metabolom verilerine farklı optimizasyon yöntemleri uygulanarak gerçek ağyapıya yönlü olarak ulaşılmaya çalışılmıştır. Bu doğrultuda dört farklı optimizasyon yöntemi kullanılmış ve elde edilen sonuçlar gerçek ağyapı ile karşılaştırılarak yöntemlerin analiz kabiliyeti ölçülmüştür. Bu amaç doğrultusunda doğrusal programlama, karesel programlama, karma tamsayılı doğrusal programlama ve genetik algoritma yöntemleri kullanılmıştır (Şekil 3.8).



Şekil 3.8 Optimizasyona Dayalı Yönlü Ağyapı Çıkarımında Kullanılan Yöntemler

Bölüm 2.4.6’da anlatıldığı gibi bir modelin Jacobian matrisi, o modelin ağyapı bilgisini içermektedir. Bu, bir mikroorganizmanın Jacobian matris bilgisine sahip olduğu sürece modelin yönlü ağyapısının bilindiği anlamına gelmektedir. Ağyapı çıkarımında temel amaç deneysel verileri kullanarak ağyapıyı elde etmek olduğundan Bölüm 2.4.6’da yer alan deneysel verilerin kovaryansı ile Jacobian matris arasındaki ilişkiyi tanımlayan Eşitlik 2.23 kullanılmıştır (Lyapunov Denklemi).

Eşitlik 2.23 farklı amaç fonksiyonları için optimizasyon yöntemleri ile çözülerek ağyapı bilgisini içeren Jacobian matris hesaplanarak yönlü ağyapı çıkarımı yapılmıştır. Jacobian matriste eğer bir (i,j) elemanının değeri 0 ise i metaboliti ile j metaboliti arasında ilişki yoktur. Eğer bu değer sıfırdan farklı ise; i ile j metabolitleri arasında yönlü bir ilişki vardır ve bu ilişkinin yönü j ’den i ’ye doğrudur (Şekil 2.19).

Bu şekilde, elde edilen Jacobian matris üzerinden yönlü ağyapı çıkarılmaktadır. Lyapunov Denkleminden (Eşitlik 2.8) Jacobian matrisi hesaplamak için 3 farklı amaç fonksiyonu kullanılmıştır:

1. Jacobian Vektörünün Köşegen Olmayan Elemanlarının Karelerinin Toplamının Minimizasyonu (Karesel Programlama)
2. Jacobian Vektörünün Köşegen Olmayan Elemanlarının Mutlak Değerleri Toplamının Minimizasyonu (Doğrusal Programlama)
3. Jacobian Vektörünün Köşegen Olmayan Elemanların Sıfır Olmayan Sayısının Minimizasyonu (Karma Tamsayı Doğrusal Programlama)

Amaç fonksiyonlarında köşegenlerin değerlendirilmemesinin nedeni yön bilgisi içermemesidir. Kullanılan üç amaç fonksiyonunun da biyolojik olarak anlamı; yukarıdan aşağıya yaklaşımla ağyapı çıkarımı analizlerinde, hücrenin amacının reaksiyonları gerçekleştirmek için hücre üreteceği enzim miktarını minimum düzeyde tutmak isteyeceği yaklaşımıdır. Bölüm 2.3’de de anlatıldığı gibi biyolojik ağyapıların seyrek aralıklı (‘sparse’) oldukları zaten bilinmektedir. Bu nedenle kullanılan 3 amaç fonksiyonu da Jacobian matristeki sıfır sayısını maksimum tutmak adına seçilmiştir. Yukarıda belirtilen amaç fonksiyonlarına uygun optimizasyon yöntemleri kullanılarak Lyapunov Denklemleri çözülerek yönlü ağyapı bilgisini içeren Jacobian matris hesaplanmış ve yönlü olarak ağyapı çıkarımı yapılmıştır.

3.3.1 Karesel Programlama Kullanılarak Yönlü Ağyapı Çıkarımı

Burada, amaç fonksiyonu olarak, Jacobian matrisinin köşegen olmayan elemanlarının karelerinin toplamının minimizasyonu kullanılmıştır (Eşitlik 3.6). Amaç fonksiyonunun doğrusal olmaması ve içerisinde karesel ifadelerin olması nedeniyle karesel programlama kullanılmıştır.

$$\min \sum_{i=1}^{n \times n} j_i^2 \quad (3.6)$$

$i \neq 1:(n+1):n \times n$

Her iki mikroorganizma için de Eşitlik 3.6’daki amaç fonksiyonu için IBM Şirketinin geliştirdiği optimizasyon paketi CPLEX’in 2010 Akademik sürümünde yer alan MATLAB programıyla uyumlu *cplexqp* fonksiyonu kullanılarak Jacobian matris hesaplanmış ve elde edilen Jacobian matrise göre yönlü ağyapı çıkarılmıştır.

3.3.2 Doğrusal Programlama Kullanılarak Yönlü Ağyapı Çıkarımı

Karesel programlamada kullanılan amaç fonksiyonunun köşegen olmayan elemanları tek tek sıfıra zorlamak yerine her elemana küçük değerler vermesinden dolayı amaç fonksiyonu olarak Jacobian vektörünün köşegen olmayan elemanlarının mutlak değerleri toplamının minimizasyonu kullanılmıştır. Kullanılan amaç fonksiyonu ve kısıtlar doğrusal olduğundan dolayı Eşitlik 2.23 doğrusal

programlama ile çözülmüştür. Doğrusal programlama için IBM Şirketinin geliştirdiği optimizasyon paketi CPLEX'in 2010 Akademik sürümünde yer alan MATLAB programıyla uyumlu *cpexlp* fonksiyonu kullanılmıştır. Mutlak değer minimizasyonunu doğrusal programlama ile çözebilmek için, j vektörü iki pozitif vektörün (q ve z) farkı olarak tanımlanmış ve amaç fonksiyonu olarak bu iki vektörün toplamının minimizasyonu seçilmiştir:

$$q_i, z_i \geq 0 \quad (3.7)$$

$$j_i = q_i - z_i \quad (3.8)$$

$$\min \sum_{i=1}^{n \times n} q_i + z_i \quad (3.9)$$

$i \neq 1:(n+1):n \times n$

Burada n , sistemdeki metabolit sayısıdır. q ve z olmak üzere iki pozitif vektörün farkı olarak tanımlanan jacobian (j) vektörünün, bu iki vektörün toplamının minimum yapılması ile köşegen olmayan elemanlarının mutlak değerlerinin toplamının minimum olması sağlanmıştır.

3.3.3 Karma Tamsayılı Doğrusal Programlama Kullanılarak Yönlü Ağyapı Çıkarımı

Karesel programlama ve doğrusal programlama ile köşegen olmayan elemanların kendileri değil de kareleri veya mutlak değerleri toplamı minimize edildiğinden, köşegen olmayan elemanlar tek tek sıfır olmaya zorlanmamaktadır. Bu nedenle, amaç fonksiyonu olarak Jacobian vektörünün köşegen olmayan elemanlarının sıfır olmayanlarının sayısının minimizasyonu seçilerek Lyapunov Denklemi karma tamsayılı doğrusal programlama ile çözdürülmüştür. Bu amaçla öncelikli olarak sistemde tamsayılarla işlem yapabilmek adına j vektörüyle aynı boyutta bir y vektörü tanımlanmıştır. Bu y vektörü ikili (elemanları 0 ve 1'lerden oluşan) bir vektördür. Optimizasyon probleminin çözümünde aşağıdaki kısıtlar kullanılmıştır:

$$j \leq y \cdot ub \quad (3.10)$$

$$j \geq y \cdot lb \quad (3.11)$$

$$y \in \{0,1\} \quad (3.12)$$

Burada ub , değişkenlerin alabileceği üst sınır değeridir. Amaç fonksiyonu ise bu y vektörü için tekrar tanımlanarak y vektörünün elemanları toplamının minimizasyonu olarak seçildi:

$$\min \sum_{i=1}^{n \times n} y_i \quad (3.13)$$

$i \neq 1:(n+1):n \times n$

Burada n , daha önceden tanımlandığı gibi sistemdeki metabolit sayısıdır. y ise metabolit çiftleri arasında reaksiyonun aktif olup olmadığını yani aralarında bir ilişki olup olmadığını gösterir. y değeri eğer ilişki varsa 1, ilişki yok ise 0 olacaktır. Bu sayede, Jacobian matrisin köşegenler haricindeki elemanlarının değerlerinin toplamı küçük olmasa bile, sıfırların sayısı maksimumda tutulacaktır. Bu amaçla IBM Şirketinin geliştirdiği optimizasyon paketi CPLEX'in 2010 Akademik sürümünde yer alan MATLAB programıyla uyumlu *cplexmilp* fonksiyonu kullanılarak hesaplamalar iki model için de gerçekleştirilmiş, fakat işlem süresinin çok uzun sürmesi ve bulunan sonuçların uygun olmaması nedeniyle, çözüm bu tür büyük sistemlerde karma tamsayılı doğrusal programlamaya göre daha hızlı çalışan genetik algoritma kullanılarak tekrar edilmiştir. Genetik algoritmanın temelleri Bölüm 2.4.5'te kapsamlı bir şekilde anlatılmıştır. Aşağıda genetik algorithmada kullanılan parametreler ve basamaklar kısaca açıklanmıştır.

1.Çözümlerin kodlanması: Jacobian matrisin boyutu kadar yani $n^2 \times 1$ boyutunda rastgele 0 ve 1'lerden oluşan kromozomlar tanımlanmıştır. Bir kromozom üzerindeki 0 değeri o indekse karşılık gelen metabolit çiftleri arasında ilişkinin olmadığını, 1 ise o metabolit çiftleri arasında bir ilişkinin bulunduğunun göstergesidir. Üretilen kromozomların Eşitlik 2.23'i sağlayıp sağlamadığı öncelikle kontrol edilmiştir. Bunun için bir kromozom üzerindeki sıfırlar kısıt olarak kullanılmış ve Jacobian vektörde o indeksin değeri 0'a eşitlenerek Eşitlik 2.23'in çözülebilir olup olmadığı rastgele bir amaç fonksiyonu ile test edilmiştir. Eğer üretilen kromozom Lyapunov Denklemini sağlamıyorsa popülasyon içerisinde çıkarılmıştır. İlk popülasyon olarak bu kromozomlardan 400 tane üretilerek algorithmaya başlanmıştır.

2.Uygunluk Değerinin Hesaplanması: Uygunluk değeri, amaca uygun olarak maksimum sıfır sayısı olarak seçilmiştir. Her bir kromozomda yer alan sıfır sayısı toplanarak, o kromozomun uygunluk değeri hesaplanmıştır.

3.Çapraz Değişim ve Mutasyon: Çapraz değişim türü olarak iki nokta çapraz değişimi, mutasyon türü olarak ise basit mutasyon kullanıldı. Basit mutasyon için; kromozomun boyutu kadar rastgele sayılar üretilerek mutasyon oranından küçük olanlar tespit edilir. Küçük olanların kromozom üstündeki değeri 1 ise 0, 0 ise 1 yapılır. Yapılan hesaplamada mutasyon oranı 0,4 olarak seçilmiştir.

4.Yeni Popülasyonun Oluşturulması: Her popülasyondaki en iyi uygunluk değerine sahip olan üstün nitelikli çocuklardan ('elit children') 10 tane, çapraz değişim ve mutasyondan 150 tane ve yeniden üretimden ('Fresh Blood') 40 tane kromozom alınarak toplam 200 kromozomdan oluşan yeni bir popülasyon oluşturulmuştur.

Bu parametreler için algoritma yeterli iterasyon sayısına ulaşınca kadar çalıştırılarak hesaplamalar yapılmış ve elde edilen Jacobian matrisleri değerlendirilerek yönlü ağyapı çıkarımı yapılmıştır.

3.3.4 Genetik Algoritma Tabanlı Önerilen Yöntem

Karesel programlama, doğrusal programlama gibi optimizasyon yöntemleri, tanımlanmamış sistemler için değişkenlere istenen amaç fonksiyonu doğrultusunda değerler vererek, sistemi birebir sağlayan tek bir çözüm önerirler. Bu yöntemlerle bulunan çözümler, eşitlik kısıtı olan Lyapunov denklemini ($A_j + 2d = 0$) katı katıya sağlamaktadır. Bu nedenle elde edilen çözümler en fazla sistemlerin serbestlik derecesi kadar sıfır içermektedir. Fakat metabolik ağyapıların yapısı seyrek aralıklıdır yani yapılarında serbestlik derecesinden fazla sayıda sıfır değeri yer almaktadır. Bu problemi aşmak amacıyla yapılan tez çalışmasında Lyapunov denklemini katı katıya sağlayan bir çözüm yerine, sisteme belirli bir esneklik katarak sistem artık-tanımlanmış ('overdetermined') olarak çözülmüştür. Bu amaçla çözüm için $A_j + 2d$ ifadesini sıfıra eşitlemek yerine mutlak değerini minimum tutmak şeklinde yeni bir yöntem önerilmiştir. Böylelikle gerçek çözüme yakın birden çok

çözüm bulunmuştur. Genetik algoritma tabanlı bu yöntemde, sistemi artık-tanımlamak için en küçük kareler yöntemi kullanılmıştır. Önerilen yöntem, genetik algoritmanın basamakları ile birlikte aşağıda detaylı olarak açıklanmıştır.

Amaç Fonksiyonları: Önerilen yöntem için üç farklı amaç fonksiyonu belirlenmiş ve bu doğrultuda üç farklı amaç fonksiyonu için ağyapı çıkarımı yapılmıştır:

i) Hata değeri: Gerçek çözümü temsil eden kromozomun, sistemi en az hata payı ile sağlayacağı temel alınarak amaç fonksiyonu olarak minimum hata değeri kullanılmıştır.

ii) Hem hata değeri hem de sıfır sayısını içeren ortak bir fonksiyon: Hata değerinin tek başına kullanılmasına ek olarak daha hızlı bir şekilde sonuca ulaşmak için hücrelerin ağyapılarının seyrek aralıklı olduğu göz önünde tutularak hem hata değerini hem de sıfır sayısını içeren ortak bir fonksiyon tanımlanmıştır:

$$\text{Uygunluk Değeri} = \lambda \times \text{Sıfır Sayısı} + \sigma \quad (3.14)$$

Burada sigma (σ) hata değerinin 10 tabanında logaritmasının negatif değeri, lamda (λ) ise ağırlık fonksiyonudur.

Öncelikle, bu iki amaç fonksiyonu için yani hata değeri ve hem hata değeri hem de sıfır sayısını içeren ortak bir fonksiyon için küçük bir gen modeli üzerinde hesaplamalar yapılmıştır. Bu amaçla McClean ve ark. tarafından geliştirilen *S.cerevisiae*'ye ait 6 protein ve 10 reaksiyondan oluşan mitojen aktive eden protein (MAP) kinaz yolağı kullanılmıştır [McClean ve ark., 2007]. Böylelikle hem yöntem için kullanılacak amaç fonksiyonlarının başarısı hem de kullanılacak parametreler belirlenmiştir.

Her iki amaç fonksiyonu da gerçek ağyapıyı hem yönlü hem de yönsüz olarak başarılı bir şekilde ortaya koymuştur. Sonuçlar EK 4'te yer almaktadır. Fakat bu iki amaç fonksiyonu aynı başarıyı büyük modeller için göstermediğinden dolayı GGM sonucunu temel alan üçüncü bir amaç fonksiyonu tanımlanarak büyük modellerde hesaplamalar ayrıca bu amaç fonksiyonu için de yapılmıştır.

iii) Hem GGM sonucunu hem de hata değerini içeren ortak bir fonksiyon:

İstatistiksel ağyapı çıkarımında yönsüz olarak en iyi sonuç GGM yöntemi ile elde edilmiştir. Bu nedenle amaç fonksiyonu olarak içsel değişkenlik modellenerek üretilen *in silico* metabolom verilerine GGM yönteminin uygulanması ile elde edilen ağyapı temel alınarak mümkün olduğunca GGM sonucuna yakın ve aynı zamanda Lyapunov Denklemini sağlayan bir çözüm aranmıştır. Bu amaçla GGM çözümü 0 ve 1 sayılarından oluşan bir kromozomla temsil edilmiş, üretilen kromozomların elemanları GGM’i temsil eden bu kromozomun elemanları ile değer olarak karşılaştırılmış ve aynı olanların sayısı belirlenmiştir. Amaç fonksiyonu olarak bu sayı ve hata payından oluşan ortak bir fonksiyon tanımlanmıştır:

$$\text{Uygunluk kriteri} = \lambda \times \text{Kesişen Eleman Sayısı} + \sigma \quad (3.15)$$

Burada sigma (σ) hata payının 10 tabanında logaritmasının negatif değeri, lamda (λ) ise ağırlık fonksiyonudur.

Çözümlerin Kodlanması: j vektörünün boyutu ile aynı yani $n^2 \times 1$ boyutunda 0 ve 1’den oluşan kromozomlar üretilmiştir. Üretilen kromozomların sıfır sayısı belirli bir aralıkta olacak şekilde ayarlanmıştır. Jacobian matriste köşegenler her zaman sıfırdan farklı bir değer almaktadır. Bu bilgi ışığında kromozomların sıfır sayısı, en fazla jacobian vektörünün köşegenleri hariç yani $(n^2 - n) \times 1$ boyutu kadar olacak şekilde ve Jacobian matrisinde köşegenlere karşılık gelen kromozomlar üzerindeki indislerin değerleri 1 olacak şekilde üretilmiştir.

İlk popülasyonun üretimi: İlk popülasyon olarak 600 tane kromozom üretilmiştir. Üretilen kromozomlar kontrol mekanizmasından geçirilerek belirlenen şartları sağlamayan kromozomlar atılmıştır.

Kontrol mekanizması ve Uygunluk değerinin (‘fitness’) Hesaplanması : Kontrol mekanizması olarak en küçük kareler yöntemi kullanılmıştır. Bu yöntem, artık-tanımlanmış sistemlerin çözümünde kullanılan standart bir yaklaşımdır. Yapılan çalışmada en küçük kareler yöntemi, Eşitlik 2.23’i birebir sağlayacak çözüm yerine $A_j + 2d$ ifadesinin karelerinin toplamının değeri minimum tutacak şekilde esneklik

katarak j vektörü için çözümler sunmuştur. Yani amaç fonksiyonu olarak Eşitlik 3.14'ü kullanmıştır:

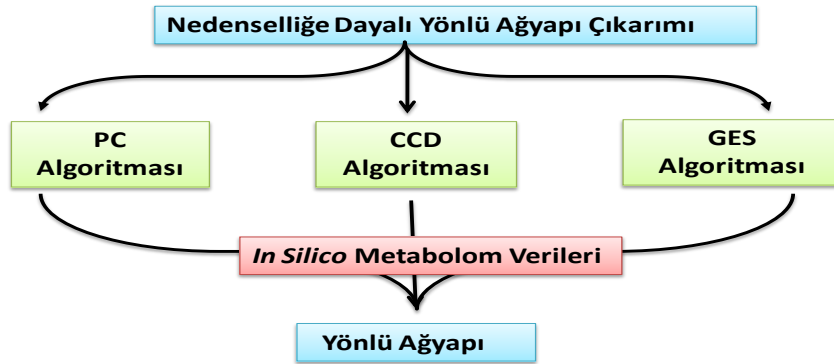
$$\min \|A_j + 2d\| \quad (3.16)$$

Üretilen her bir kromozom için üzerindeki sıfırlar kısıt olarak kullanılmış ve Jacobian matriste o indeksin değeri 0'a eşitlenerek en küçük kareler yöntemi ile çözdürülmüştür. Amaç fonksiyonu değeri yani hata değeri ('residual') 0,005'in üstünde olanlar ve içerdiği sıfır sayısı serbestlik derecesinden bir fazlasına eşit veya küçük olanlar atılarak üretilen popülasyon kontrol edilmiştir. Geriye kalan kromozomlar ve bunların hata değerleri üzerinden algoritmaya devam edilmiştir. Uygunluk değeri olarak amaç fonksiyonlarında tanımlanan üç farklı kriter kullanılmıştır: (i) Hata değeri, (ii) Hem hata değeri hem de sıfır sayısını içeren bir fonksiyon, (iii) Hem GGM sonucunu hem de hata değerini içeren bir fonksiyon. Üç farklı uygunluk değeri için de ayrı ayrı hesaplamalar yapılmıştır.

Yeni Popülasyonun Üretilmesi: Algoritmada bir önceki popülasyona biyolojik evrimi örnek alan Bölüm 2.4.5'te anlatıldığı gibi çeşitli yöntemler uygulanmış ve bunların sonuçlarından elde edilen kromozomlar uygunluk değerlerine göre sıralanarak belirli oranlarda bir sonraki popülasyona aktarılmıştır. Bu amaçla öncelikle üretilen kromozomlara çapraz değişim uygulanmıştır. Çapraz değişim için McClean ve ark. tarafından geliştirilen model üzerinde yapılan çalışmalarda çift nokta çapraz değişime göre daha hızlı sonuca ulaşan tek nokta çapraz değişim tipi kullanılmış ve ortaya çıkan yeni kromozomların %74 ü bir sonraki popülasyona aktarılmıştır. Daha sonra çapraz değişime uğramış tüm kromozomlara basit mutasyon uygulanmıştır. Mutasyon oranı olarak 0,4 kullanılmıştır. Mutasyona uğramış kromozomların % 1'i de aynı şekilde yeni popülasyona aktarılmıştır. Bir önceki popülasyondan gelen kromozomlardan en iyi uygunluk değerine sahip olanların % 5'i ve buna ek olarak yeni üretimin %20'si rastgele yeniden üretilen kromozomlarla sağlanmıştır. Her popülasyon için toplam 200 kromozom olacak şekilde algoritma ayarlanmıştır. Her popülasyonda bu işleme devam edilmiş ve algoritma yeterli iterasyon sayısına ulaşınca kadar çalıştırılarak hesaplamalar yapılmıştır. Elde edilen Jacobian matrisleri değerlendirilerek yönlü ağıyapı çıkarımı yapılmıştır.

3.4 Nedenselliğe Dayalı Yönlü Ağyapı Çıkarımı Tekniklerinin *In Silico* Metabolom Verilerine Uygulanması

Yapılan çalışmada *Escherichia coli* ve *Saccharomyces cerevisiae* mikroorganizmalarına ait *in silico* metabolom verilerine farklı nedensellik yöntemleri uygulanmış ve gerçek ağyapıya ulaşılmaya çalışılmıştır. Bu amaç doğrultusunda skor ve kısıt bazlı 3 farklı algoritma kullanılmış ve elde edilen sonuçlar karşılaştırılarak algoritmaların analiz kabiliyeti ortaya konmuştur. Bu amaçla PC, CCD ve GES algoritmaları kullanılmıştır (Şekil 3.9).



Şekil 3.9 Nedenselliğe Dayalı Yönlü Ağyapı Çıkarımında Kullanılan Yöntemler

Bu algoritmalar hakkında detaylı bilgi Bölüm 2.5'te yer almaktadır. Yapılan hesaplamalar için PC, CCD ve GES algoritmalarının da içinde yer aldığı Carnegie-Mellon Üniversitesinden bir grup (Spirtes, Glymour ve Scheines) tarafından geliştirilen Java tabanlı TETRAD 4 yazılımı kullanılmıştır [Spirtes et al., 1998]. Üç yöntem için de kısmi korelasyon katsayısı olarak birinci dereceden kısmi korelasyon katsayısı ve hata payı olarak 0,01 değeri kullanılmıştır. Bölüm 2.5'te anlatıldığı gibi bu yöntemler tüm ilişkilerin yönünü tespit edememektedir. Elde edilen yarı yönlü yapılar gerçek ağyapı ile karşılaştırılarak değerlendirildi.

3.5 Elde Edilen Sonuçların Değerlendirilmesi

Tez çalışmasında farklı ağyapı çıkarımı tekniklerinin *in silico* metabolom verilerine uygulanması ile elde edilen ağyapılar, dinamik modellere ait gerçek ağyapı bilgisine sahip Jacobian Matris ile karşılaştırılarak değerlendirildi. Bu amaçla

kullanımı yaygın olan istatistiksel Doğru Pozitif Oranı ('True Pozitive Rate') ve Yanlış Pozitif Oranı ('False Pozitive Rate') kavramları kullanıldı. Bu kavramların temelini Doğru Pozitif, Yanlış Pozitif, Doğru Negatif ve Yanlış Negatif terimleri oluşturmaktadır. Bu tür testlerde, Bölüm 2.3.5'te anlatılan yokluk hipotezinin doğru olması Negatif bir sonuç olarak değerlendirilmektedir. Yani metabolitlerin arasında ilişki olmaması Negatif, ilişki olması ise Pozitif sonuç olarak değerlendirilir.

Eğer gerçek ağyapıda bir metabolit çifti arasında bir ilişki varsa yani gerçek sonuç Pozitif ise, ağyapı çıkarımı sonucunda elde edilen ağyapıda da bu metabolit çiftleri arasında bir ilişki varsa gerçek sonuç doğru olarak bulunduğundan sonuç 'Doğru Pozitif', bir ilişki yok ise gerçek sonuç yanlış bulunduğundan sonuç 'Yanlış Pozitif' olarak adlandırılır. Eğer gerçek ağyapıda metabolit çiftleri arasında ilişki yoksa yani gerçek sonuç Negatif ise, ağyapı çıkarımı sonucunda elde edilen ağyapıda da bu metabolit çiftleri arasında bir ilişki yoksa gerçek sonuç doğru olarak bulunduğundan sonuç 'Doğru Negatif', bir ilişki bulunuyor ise gerçek sonuç yanlış bulunduğundan sonuç 'Yanlış Negatif' olarak adlandırılır. Doğru Pozitif Oran (TPR), bulunan ağyapının gerçek ağyapı ile karşılaştırılması sonucu bulunan Doğru Pozitif sayısının toplamının, gerçek ağyapıdaki toplam Pozitif sayısına oranıdır. Yani, aralarında gerçekten bağlantı olan metabolit çiftlerinin kaçının doğru tahmin edilebildiğinin oranıdır. Yanlış Pozitif Oranı (FPR) ise, bulunan ağyapının gerçek ağyapı ile karşılaştırılması sonucu bulunan Yanlış Pozitif sayısının toplamının, gerçek ağyapıdaki toplam Negatif sayısına oranıdır. Yani FPR, aralarında gerçekte bağlantı olmadığı (tepkime gerçekleşmediği) halde aralarında bağlantı olduğu yanlışlıkla bulunan çiftlerin oranıdır. Ağyapı çıkarımı sonuçlarında TPR'nin mümkün olduğunca yüksek, FPR'nin ise düşük olması istenmektedir.

4 *Saccharomyces cerevisiae* MİKROORGANİZMASI İÇİN YAPILAN ÇALIŞMALAR

Yapılan tez çalışmasında metabolom verilerine ağyapı çıkarımı tekniklerini uygulayarak hücrenin biyolojik olarak amacını belirlemek ve gerçek ağyapıya ulaşmak amaçlanmıştır. Bu amaç doğrultusunda öncelikli olarak metabolom verileri *in silico* olarak bilgisayar ortamında üretilmiştir. *S.cerevisiae* için *in silico* olarak üç farklı tip veri üretilmiştir:

- 1.Enzimatik Değişkenliğin Modellenmesi ile Üretilen Metabolom Verisi
- 2.İçsel Değişkenliğin Modellenmesi ile Üretilen Metabolom Verisi
- 3.Hem İçsel Hem de Çevresel Değişkenliğin Birlikte Modellenmesi ile Üretilen Metabolom Verisi

Bu üç farklı veri tipine öncelikle korelasyon yöntemleri uygulanarak yönsüz ağyapı çıkarımı yapılmış ve sonuçlar istatistiksel olarak değerlendirilmiştir.

4.1 İstatistiksel Ağyapı Çıkarımı Tekniklerinin *Saccharomyces cerevisiae* Mikroorganizmasına ait *In Silico* Metabolom Verilerine Uygulanması

Bölüm 3.1’de anlatıldığı gibi 3 farklı tipte *in silico* metabolom veri üretilmiş ve her bir metabolit için 1000 adet kararlı hal (‘steady-state’) etrafında değişen veri toplanmıştır. Elde edilen sonuçların güvenilirliğini ve tekrarlanabilirliğini test etmek için bu şekilde 3 farklı veri tipinin her biri için 10 ayrı veri seti toplanmış ve yapılan işlemlerin tümünde bu 10 tekrarın ortalaması alınmıştır. Her bir veri setine Pearson korelasyonu (PC), birinci dereceden kısmi korelasyon (PPC1), ikinci dereceden kısmi korelasyon (PPC2) ve (n-2)’nci dereceden kısmi korelasyon (GGM) yöntemleri uygulanmış ve sonuçlar gerçek ağyapı bilgisini içeren modele ait Jacobian matris ile karşılaştırılarak istatistiksel olarak değerlendirilmiştir.

Enzimatik deęiřkenlięin modellenmesi ile retilen *in silico* metabolom verilerine bu yntemler uygulanmıř ve sonular istatistiksel olarak deęerlendirilmiřtir. Anlam testi iin hata payı ('cut-off') 0,01 olarak seilmiřtir. Sonular Tablo 4.1'de yer almaktadır.

Tablo 4.1 *S.cerevisiae* mikroorganizmasına ait enzimatik deęiřkenlięin modellenmesi ile retilen *in silico* metabolom verilerine uygulanan istatistiksel aęyapı ıkarımı yntemlerinin istatistiksel olarak deęerlendirme sonuları

	PC	PPC1	PPC2	GGM
Doęru Negatif	10,4	42,8	54,5	45,4
Doęru Pozitif	19,2	12,8	10,2	14,7
Yanlıř Negatif	1,8	8,2	10,8	6,3
Yanlıř Pozitif	46,6	14,2	2,5	11,6
Doęru Pozitif Oranı	0,91	0,61	0,49	0,70
Yanlıř Pozitif Oranı	0,82	0,25	0,04	0,20

Tablo 4.1'de grldę gibi, enzimatik deęiřkenlik kullanılarak retilen deneysel verilere Pearson korelasyonu uygulandıęında Doęru Pozitif Oranı (TPR) 0,91 ve Yanlıř Pozitif Oranı (FPR) 0,82 olarak elde edilirken, GGM uygulandıęında 0,70 TPR ve 0,20 FPR elde edilmiřtir. *S.cerevisiae* metabolik modelinin oluřturduęu gerek aęyapıda birbirleriyle iliřkili toplam 21 metabolit ifti vardır. Yani Pearson korelasyonu bu 21 iliřkiden %91'ini (yaklařık 19 tane) bulmuřtur. Ancak te yandan, aralarında iliřki olmaması gereken 57 iftin de %82'sini (47 tane) arasında iliřki var olarak bulmuřtur. Dięer yntemlere de bakıldıęında PPC1; TPR 0,61 ile GGM sonularına yaklařırken, FPR'si GGM'den daha yksektir. PPC2'de FPR'nin dřk olmasına karřılık dięer yntemlere gre TPR deęeri daha dřktr. Bu nedenle, nisbeten dřk FPR deęeri veren, ancak te yandan TPR deęeri de ok dřk olmayan tek yntem GGM'dir. Yani, incelenen yntemlerden en iyisi GGM'dir. GGM ile 21 iliřkinin yaklařık 15 tanesi bulunurken 57 adet aralarında gerekte iliřkili olmayan iftin de yaklařık 12 tanesinin arasında iliřki var olarak bulunmuřtur.

İsel deęiřkenlięin modellenmesi ile retilen *in silico* metabolom verilerine PC, PPC1, PPC2 ve GGM yntemleri uygulanmıř ve sonular istatistiksel olarak deęerlendirilmiřtir. Sonular Tablo 4.2'de yer almaktadır.

Tablo 4.2 *S.cerevisiae* mikroorganizmasına ait içsel değişkenliğin modellenmesi ile üretilen *in silico* metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı yöntemlerinin istatistiksel olarak değerlendirme sonuçları.

	PC	PPC1	PPC2	GGM
Doğru Negatif	11,6	42,6	53,2	46,7
Doğru Pozitif	19,5	14,3	11,1	15,9
Yanlış Negatif	1,5	6,7	9,9	5,1
Yanlış Pozitif	45,4	14,4	3,8	10,3
Doğru Pozitif Oranı	0,93	0,68	0,53	0,76
Yanlış Pozitif Oranı	0,80	0,25	0,07	0,18

İçsel değişkenlik modellenerek üretilen *in silico* verilere uygulanan ağyapı çıkarımı Enzimatik değişkenlik modellenerek üretilen *in silico* verilere uygulanan ağyapı çıkarımı sonuçlarına benzer niteliktedir. Yine en iyi sonucu tüm değişkenlerin dolaylı etkilerinin çıkarıldığı GGM yöntemi vermiştir.

Hem içsel hem de çevresel değişkenliğin modellenmesi ile üretilen *in silico* metabolom verilerine PC, PPC1, PPC2 ve GGM yöntemleri uygulanmış ve sonuçlar istatistiksel olarak değerlendirilmiştir. Sonuçlar Tablo 4.3'te yer almaktadır.

Tablo 4.3 *S.cerevisiae* mikroorganizmasına ait hem içsel hem de çevresel değişkenliğin modellenmesi ile üretilen *in silico* metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı sonuçlarının istatistiksel olarak değerlendirme sonuçları.

	PC	PPC1	PPC2	GGM
Doğru Negatif	14,5	43,6	53	46,5
Doğru Pozitif	19,2	14,7	10,6	15,6
Yanlış Negatif	1,8	6,3	10,4	5,4
Yanlış Pozitif	42,5	13,4	4	10,5
Doğru Pozitif Oranı	0,91	0,70	0,50	0,74
Yanlış Pozitif Oranı	0,75	0,24	0,07	0,18

Genel olarak *S.cerevisiae* mikroorganizmasına ait üç veri tipi için yapılan istatistiksel ağyapı çıkarımı sonuçlarında GGM yönteminin başarı açısından ön planda olduğu

görülmüştür. Bu nedenle GGM yöntemi sonuçlarını iyileştirecek bazı çalışmalar yapılmıştır:

i) Bonferoni Düzeltmesinin Uygulanması: İstatistiğe dayalı ağyapı çıkarımı hesaplamalarının tümünde hata payı olarak 0.01 değeri kullanılmıştır. Aynı anda birden fazla durum için anlam testi uygulandığında, istatistiksel olarak kullanılan hata payı değerinin aynı anda gerçekleştirilen anlam testi sayısına göre düzenlenmesi literatürdeki yaygın uygulamalardandır [Sokal and Rohlf, 1995]. Bu amaçla, literatürde Bonferoni düzeltmesi olarak bilinen düzeltmenin kullanılmasının GGM sonuçlarını iyileştirmedeki etkisine bakılmıştır. Bu doğrultuda, seçilen hata payı değeri *S.cerevisiae* modeli için toplam muhtemel ikili çift sayısına bölünerek 0,01/78 olarak kullanılmıştır. Bu şekilde üç veri tipi için de hesaplamalar tekrar edilmiştir. İçsel değişkenlik için yapılan hesaplamalara bakıldığında Bonferoni düzeltmesiyle 21 gerçek ilişkiden doğru olarak bulunan metabolit çifti sayısı 16'dan 14'e düşerken bunun yanında ilişki olmadığı halde ilişkili bulunan metabolit çifti sayısı 10'dan 6'ya düşmüştür. Üç farklı veri tipi için elde edilen sonuçlar Tablo 4.4'te yer almaktadır.

ii) Belirli Bir Korelasyon Katsayısı Değerinin Altındaki İlişkilerin Yok Sayılması: Bilindiği gibi iki değişken arasındaki korelasyon katsayısı o değişkenler arasındaki ilişkinin gücünü gösterir. Korelasyon katsayısı, -1 ve +1 arasında değer alır ve korelasyon katsayısı +1 veya -1'e ne kadar yakınsa ilişki o oranda güçlüdür, 0'a ne kadar yakınsa ilişki o oranda zayıftır. İstatistiksel olarak değerlendirmeler sonucunda aralarında ilişki bulunan çiftlerin korelasyon katsayılarına bakılmıştır. Korelasyon katsayıları 0,15'ten küçük olan ilişkilerin gücü çok zayıf olarak düşünülmüş, bu değer altında korelasyon katsayısı değerine sahip metabolit çifti arasında istatistiksel olarak ilişki bulunsa dahi bu çiftler arasında ilişki olmadığı kabul edilmiştir. Ve buna göre tekrar hesaplamalar yapılmıştır.

Enzimatik değişkenlik için GGM sonuçlarına bakıldığında; 21 ilişkiden doğru olarak bulunan 15 metabolit çifti 11'e düşerken bunun yanında gerçekte arasında ilişki olmadığı halde ilişkili bulunan 12 metabolit çifti sayısı 7'e düşürülmüştür. Bu yöntemdeki en önemli problem FPR'i düşürülürken, TPR'nin de düşürülmesi olmuştur. Üç farklı veri tipi için elde edilen sonuçlar Tablo 4.4'te yer almaktadır.

iii) Budama(‘Pruning’): Budama, dolaylı ilişkileri tespit edip eleyebilen alternatif bir yöntemdir [Çakır ve ark., 2009]. Bu yöntem, uygulanan korelasyon analizi sonucu birbirine tamamen bağlı gözükten üç metabolit arasındaki korelasyon katsayılarının karşılaştırılması esasına dayanmaktadır [Margolin et al. 2006]. Yani, basitçe, bu yöntemle en düşük korelasyon katsayına sahip olan iki metabolit arasındaki ilişki ağyapıdan kaldırılır. Buradaki varsayım şudur; üç metabolit de aralarında bir ilişki varsa, bunlar arasında en düşük korelasyon gösteren çift, dolaylı etkilerden dolayı ilişki göstermektedir. Eğer Eşitlik 4.1 sağlanıyorsa bu ilişki ağyapıdan kaldırılır.

$$|R_{xz}| \leq \min(|R_{yz}|, |R_{xy}|) \times (1 - \tau) \quad (4.1)$$

τ tolerans değeridir. Tolerans değeri, çok katı budama yapılmasını engellemek için kullanılır. Ayrıca daha yüksek dereceden budama kontrolü yapılabilir. Yapılan tez çalışmasında 3’lü kontrole ek olarak birbirleriyle ilişkili 4 metabolit arasındaki ilişkiler de kontrol edilmiştir. Bu kontrol yapılırken Eşitlik 4.2’deki işaret kontrolü de yapılmıştır. Yani, örneğin üçlü kontrolde, çıkartılacak olan bağlantının işareti, diğer iki bağlantının işaretlerinin çarpımına eşit olmalıdır. Başka bir ifade ile, diğer iki bağlantının işaretleri birbirinin aynı olmalıdır [Klampt et al., 2010].

$$\text{işaret}(R_{xz}) = \text{işaret}(R_{yz} R_{xy}) \quad (4.2)$$

Buna göre bu iki şartı sağlayan metabolit çiftleri arasındaki ilişkiler ağyapıdan kaldırılmıştır. Bu hesaplamalar yapılırken tolerans değeri 0,80 alınmıştır. Yani %20 esneklik verilmiştir.

Enzimatik değişkenlik için GGM sonuçlarına bakıldığında; 21 ilişkiden doğru olarak bulunan 15 metabolit çifti 12’e düşerken bunun yanında gerçekte arasında ilişki olmadığı halde ilişkili bulunan 12 metabolit çifti sayısı 8’e düşürülmüştür. Tablo 4.4’te görüldüğü gibi, bu yaklaşım ile istenildiği gibi çok düşük FPR değerleri elde edilmesine rağmen, TPR de yapılan analizden dikkate değer derecede olumsuz olarak etkilenmiştir. Üç farklı veri tipi için elde edilen sonuçlar Tablo 4.4’te yer almaktadır.

iv) GGM’e dayalı Korelasyon Katsayısının Pearson Korelasyon Katsayısı ile Kontrol Edilmesi: Pearson korelasyon katsayısı dolaylı etkilerin çıkarılmamasından dolayı; her zaman değer olarak GGM korelasyon katsayısından daha yüksek olmalıdır. Bu nedenle bunun kontrolü yapılarak her bir metabolit çifti arasındaki Pearson korelasyon katsayısı ile GGM’e dayalı korelasyon katsayısı karşılaştırılmış, Pearson korelasyon katsayısının değeri GGM’e dayalı korelasyon katsayısının değerinden küçük ise bu çift arasında ilişki yok olarak kabul edilmiştir. Bu kontrol yapılırken yaklaşık aynı olan korelasyon değerine sahip ilişkilerin atılmasını önlemek için Pearson korelasyon katsayısı 1,5 ile çarpılarak yapılan uygulamaya esneklik sağlanmıştır. Enzimatik değişkenlik için GGM sonuçlarına bakıldığında; 21 ilişkiden doğru olarak bulunan 15 metabolit çifti 14’e düşerken bunun yanında gerçekte arasında ilişki olmadığı halde ilişkili bulunan 12 metabolit çifti sayısı 11’e düşürülmüştür. Üç farklı veri tipi için elde edilen sonuçlar Tablo 4.4’te yer almaktadır.

Tablo 4.4 *S.cerevisiae* mikroorganizmasına ait üç farklı veri tipi için GGM sonuçlarını iyileştirmek için yapılan denemelerin sonuçları.

DENENEN YÖNTEMLER	Enzimatik		İçsel		İçsel+ Çevresel	
	TPR	FPR	TPR	FPR	TPR	FPR
Orijinal Durum	0,70	0,20	0,76	0,18	0,74	0,18
Bonferoni düzeltmesi	0,63	0,15	0,70	0,11	0,69	0,10
Korelasyonu 0.15’ten Küçük İlişkilerin Yok Sayılması	0,53	0,12	0,63	0,07	0,61	0,06
“Pruning”	0,55	0,13	0,56	0,05	0,54	0,06
Pearson korelasyon katsayısı ile kontrol	0,66	0,18	0,75	0,17	0,73	0,17

4.2 Optimizasyona Dayalı Ağyapı Çıkarımı Tekniklerinin *Saccharomyces cerevisiae* Mikroorganizmasına ait *In Silico* Metabolom Verilerine Uygulanması

Optimizasyona dayalı ağyapı çıkarımının temelini oluşturan Lyapunov eşitliği içsel değişkenlikten modellenerek üretilen veriyi kullanır. Bu nedenle optimizasyona dayalı yapılan çalışmalarda sadece içsel değişkenliğin modellenmesi ile üretilen *in silico* deneysel verilere ağyapı çıkarımı teknikleri uygulanmış ve elde edilen sonuçlar

gerçek ağyapı bilgisini içeren Jacobian matris ile karşılaştırılarak sonuçlar yönlü ve yönsüz ağyapı çıkarımı açısından istatistiksel olarak değerlendirilmiştir. Bu amaçla hücrenin minimum enzim üretimini (seyrek aralıklı olduğunu) yansıtan ve Bölüm 3.2.1’de anlatılan 3 farklı amaç fonksiyonu için farklı optimizasyon yöntemleri ile hesaplamalar gerçekleştirilmiştir.

4.2.1 Karesel Programlama ile Yönlü Ağyapı Çıkarımı

Bu yöntem için, Bölüm 3.3.1’de de anlatıldığı gibi amaç fonksiyonu olarak, Jacobian matrisinin köşegen olmayan elemanlarının karelerinin toplamının minimizasyonu kullanılmıştır. Buna göre içsel değişkenliğin modellenmesi ile elde edilen *in silico* metabolom verilerine belirtilen amaç fonksiyonu için karesel programlama uygulanmıştır. Elde edilen ağyapı, yönlü ve yönsüz ağyapı çıkarımı açısından değerlendirilmiştir. Sonuçlar Tablo 4.5’te yer almaktadır.

Tablo 4.5 *S.cerevisiae* mikroorganizması için karesel programlama ile ağyapı çıkarımı sonuçları.

	Yönlü Ağyapı	Yönsüz Ağyapı
Doğru Negatif	0	0
Doğru Pozitif	20	21
Yanlış Negatif	1	0
Yanlış Pozitif	57	57
Doğru Pozitif Oranı	0,95	1
Yanlış Pozitif Oranı	1	1

Ağyapıda bulunan bütün ilişkiler yönsüz olarak doğru bulunurken, gerçekte olmayan 57 metabolit çifti arasında ise ilişki bulunmuştur. Aralarında yönsüz olarak ilişki bulunan çiftlerin yön olarak 20 tanesi doğru olarak bulunmuştur. Yönsüz olarak tüm metabolit çiftleri arasında ilişki olan (‘full connected’) bir ağyapı elde edilmiştir. Sonuçlar ağyapı çıkarımı açısından iyi değildir.

4.2.2 Doğrusal Programlama ile Yönlü Ağyapı Çıkarımı

Amaç fonksiyonu olarak, Bölüm 3.3.2’de anlatıldığı gibi Jacobian matrisinin köşegen olmayan elemanlarının mutlak değerleri toplamının minimizasyonu

kullanılmıştır. Buna göre içsel değişkenliğin modellenmesi ile elde edilen *in silico* metabolom verilerine belirtilen amaç fonksiyonu için doğrusal programlama uygulanmış ve bu yöntem ile elde edilen ağyapı yönlü ve yönsüz ağyapı çıkarımı açısından değerlendirilmiştir. Sonuçlar Tablo 4.6'da yer almaktadır.

Tablo 4.6 *S.cerevisiae* mikroorganizması için doğrusal programlama ile ağyapı çıkarımı sonuçları.

	Yönlü Ağyapı	Yönsüz Ağyapı
Doğru Negatif	2	2
Doğru Pozitif	1	21
Yanlış Negatif	20	0
Yanlış Pozitif	55	55
Doğru Pozitif Oranı	0,05	1
Yanlış Pozitif Oranı	0,96	0,96

Gerçekte varolan metabolitler arasındaki ilişkilerin tümü yönsüz olarak bulunmasına rağmen, hem FPR değerinin 1'e çok yakın olması hem de yönsüz olarak bulunan ilişkilerin sadece bir tanesinin yönünün tahmin edilmesi açısından sonuçlar ağyapı çıkarımı açısından iyi değildir.

4.2.3 Karesel ve Doğrusal Programlama Sonuçlarının İyileştirilmesi

Hem doğrusal programlama ile hem de karesel programlama ile ağyapı çıkarımı açısından iyi sonuçlar elde edilememiştir. Bu amaçla bu yöntemlerin sonuçlarını iyileştirmeye yönelik birtakım çalışmalar gerçekleştirilmiştir:

i) Belli Bir Eşik Değerinin Kullanılması: Jacobian matrisindeki değerler sadece ilişkinin olup olmadığını değil aynı zamanda o ilişkinin gücünü göstermektedir. Bu bilgiye dayanarak sistemin karesel programlama ile çözümünden elde edilen Jacobian matriste değeri belirli bir eşik değerinin altında olan elemanlar sıfıra sabitlenerek tekrar karesel programlama ile çözdürülmüştür. Hesaplamalara yeterli sayıda iterasyona (sonuçlar değişmeyinceye kadar) ulaşıncaya kadar devam edilmiştir. Eşik değeri olarak 0,01 kullanılmış, elde edilen sonuçlar değerlendirilmiş

ve sonuçların her iki yöntem için de değişmediği görülmüştür. Sonuçlar Tablo 4.7’de yer almaktadır.

ii) Korelasyon Katsayına Ait Bilgilerin Kullanılması: Graphical Gaussian Model (GGM) korelasyonuna ait yapılan değerlendirmeler sonucu en düşük 5 korelasyon katsayısına ait ilişkilerin gerçekte ilişkili olmadığı bulunmuştur. Bu bilgi sisteme eklenerek tekrar hesaplamalar gerçekleştirilmiştir. Bu amaçla, *S.cerevisiae* için daha önceden hesaplanan GGM korelasyon katsayıları baz alınarak en düşük korelasyon katsayısına ait çiftler bulunmuştur ve bu çiftler arasında ilişki olmadığı bilindiği için sistemde bu 5 çiftin alt ve üst sınırı sifıra eşitlenerek öncelikle karesel programlama ile çözdürülmüştür. Elde edilen sonuçlar orijinal duruma göre daha iyi olmasına rağmen yine de ağyapı çıkarımı açısından yeterli değildir. Aynı yöntem kullanılarak hesaplamalar doğrusal programlama için tekrar edilmiştir. Fakat sonuçlar değişmemiştir. Sonuçlar Tablo 4.7’de yer almaktadır.

Daha önce yapılan hesaplamalara dayanarak yapılan incelemelerde korelasyon katsayıları 0,35’ten büyük olan ikililer arasında ilişki olduğu bilinmektedir. Bu ilişkiler işaret olarak gerçek Jacobian matriste karşılık gelen elemanların işaretleri ile aynıdır. Bu bilgi hem karesel programlamaya hem de doğrusal programlamaya bazı formülasyonlarla eklenerek bu şekilde hesaplamalar tekrarlanmıştır. Bu formülasyonlar kısaca açıklanırsa; eğer iki değişken arasındaki korelasyon katsayısı 0,35’ten büyükse bu ikilinin korelasyon katsayısı matrisindeki işareti bu metabolit çiftinin indislerine karşılık gelen jacobian matrisin elemanları ile işaret olarak uyuşmalıdır. GGM korelasyon katsayısının işareti negatif veya pozitif olabilir. Bu nedenle iki tür eşitsizlik kısıtı tanımlanmıştır (Eşitlik 4.3-4.4).

$$\text{GGM Korelasyon Katsayısı Pozitif ise: } j(a,b)+j(b,a) \geq 0 \quad (4.3)$$

$$\text{GGM Korelasyon Katsayısı Negatif ise: } j(a,b)+j(b,a) \leq 0 \quad (4.4)$$

Bu kısıtlar sisteme eklenerek karesel programlama ve doğrusal programlama ile sistem çözdürülmüş ve elde edilen sonuçlar istatistiksel olarak değerlendirilmiştir. Hem karesel programlama hem de doğrusal programlama ile yapılan hesaplamalarda sonuçlar değişmemiştir. Sonuçlar Tablo 4.7’de yer almaktadır.

iii) Değişkenlik Analizi: Yapılan hesaplamalarda değişkenlik analizi kullanılarak sistemin çözüm aralığının daraltılması belirlenmesi ve daraltılması amaçlandı. Değişkenlik analizi kısıtların sağlandığı durumda her bir değişkenin alabileceği maksimum ve minimum değerleri ortaya çıkaran bir yaklaşımdır. Buna bağlı olarak her bir değişkenin alabileceği maksimum ve minimum değerler hesaplandı. Eğer bir değişkenin alacağı hem maksimum hem de minimum değeri negatif ise sistemde bu değişkenin sınırları negatif olacak şekilde, pozitifse sınırları pozitif olacak şekilde ayarlanarak hesaplamalar tekrar edilmiştir. Fakat hem doğrusal programlama hem de karesel programlama için sonuçlar değişmemiştir. Sonuçlar Tablo 4.7’de yer almaktadır.

Tablo 4.7 *S.cerevisiae* mikroorganizması için doğrusal ve karesel programlama ile elde edilen ağyapı çıkarımı sonuçlarını iyileştirmek için yapılan denemelerin sonuçları.

DENENEN YÖNTEM	Karesel Programlama				Doğrusal Programlama			
	Yönlü		Yönsüz		Yönlü		Yönsüz	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
Orijinal Durum	0,05	0,96	1	0,96	0,95	1	1	1
Eşik Değeri 0,15'ten küçük	0,05	0,96	1	0,96	0,95	1	1	1
Korelasyon Katsayısı Küçük	0,14	0,88	1	0,88	0,95	1	1	1
Değişkenlik Analizi	0,05	0,96	1	0,96	0,95	1	1	1
Korelasyon katsayısı 0,35'den büyük	0,05	0,96	1	0,96	0,95	1	1	1

4.2.4 Karma Tamsayılı Programlama ile Yönlü Ağyapı Çıkarımı

Amaç fonksiyonu olarak, Jacobian vektörünün köşegen olmayan elemanlarının sıfır olmayanlarının sayısının minimizasyonu kullanılarak hesaplamalar gerçekleştirilmiştir (Bölüm 3.3.3). Buna göre içsel değişkenliğin modellenmesi ile elde edilen *in silico* metabolom verilerine belirtilen amaç fonksiyonu için Genetik Algoritmada karma tamsayılı programlama uygulanmış ve bu yöntem ile elde edilen ağyapılar yönlü ve yönsüz ağyapı çıkarımı açısından değerlendirilmiştir. Son popülasyonda yer alan en iyi beş sonuca ait TPR ve FPR Tablo 4.8’de yer almaktadır.

Tablo 4.8 *S.cerevisiae* mikroorganizması için genetik algoritma tabanlı karma tamsayılı doğrusal programlama sonuçları: Uygunluk Değerleri, TPR ve FPR'ler.

En iyi Uygunluk Değerine Kromozom Sırası	Yönlü Ağyapı		Yönsüz Ağyapı		Uygunluk Değeri
	TPR	FPR	TPR	FPR	
1	0,19	0,82	0,71	0,82	78
2	0,29	0,82	0,81	0,82	78
3	0,24	0,74	0,76	0,74	78
4	0,19	0,75	0,62	0,75	78
5	0,29	0,82	0,67	0,82	78

Son popülasyondaki en iyi uygunluk değerine sahip beş kromozom için genel olarak sonuçlara bakıldığında hem yönlü hem de yönsüz ağyapı çıkarımı için iyi değildir. Fakat doğrusal programlamaya ve karesel programlamaya göre kullanılan amaç fonksiyonu yani hücrenin minimum enzim üretme (seyrek aralıklı olması) amacını daha iyi yansıtmıştır. En iyi uygunluk değerine sahip ilk 25 kromozom için sonuçlar EK'tedir.

4.2.5 Önerilen Genetik Algoritma Tabanlı ile Yönlü Ağyapı Çıkarımı

Bölüm 3.3.4'te anlatıldığı gibi sistemi artık-tanımlanmış olarak çözmeye imkan sağlayan en küçük kareler yöntemi kullanılarak geliştirilen genetik algoritma tabanlı yöntem ile yapılan hesaplamalarda McClean modelinde yönlü ve yönsüz olarak gerçek ağyapıya ulaşılmıştır (EK 4). Bu nedenle önerilen yöntem hem *S.cerevisiae* hem de *E.coli* mikroorganizmalarına uygulanarak gerçek ağyapıya ulaşmaya çalışılmıştır. Bu amaçla, *Saccharomyces cerevisiae* mikroorganizmasına ait *in silico* metabolom verilerine önerilen yöntem uygulanarak Bölüm 3.3.4'te tanımlanan 3 farklı amaç fonksiyonu için hesaplamalar yapılmıştır.

Öncelikle uygunluk değeri olarak, hata değeri kullanılmış ve bu doğrultuda *Saccharomyces cerevisiae* mikroorganizması için hesaplamalar gerçekleştirilmiştir. Elde edilen sonuçlar gerçek Jacobian Matris ile karşılaştırılarak yönlü ve yönsüz olarak ağyapı çıkarımı için değerlendirilmiştir. Son popülasyondaki en iyi uygunluk değerine sahip beş kromozoma ait sonuçlar Tablo 4.9'dadır.

Tablo 4.9 *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak hata değerinin kullanıldığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR’ler

	Uygunluk Değeri	Sıfır Sayısı	Yönlü		Yönsüz	
			TPR	FPR	TPR	FPR
1	$1,29 \times 10^{-13}$	80	0,19	0,67	0,81	0,67
2	$9,96 \times 10^{-9}$	81	0,19	0,67	0,81	0,67
3	$4,06 \times 10^{-6}$	85	0,10	0,61	0,90	0,61
4	$5,58 \times 10^{-6}$	83	0,38	0,65	0,81	0,65
5	$1,94 \times 10^{-5}$	84	0,14	0,65	0,86	0,65

Elde edilen sonuçlara bakıldığında yönsüz ağyapı çıkarımı için TPR açısından iyi sonuçlar elde edilmesine karşın, düşük olması istenen FPR’nin de yüksek olduğu görülmektedir. Bu yöntem ile bulunan ilişkilerin yönlerinin az bir kısmının doğru olarak tahmin edildiği gözlenmektedir. Son popülasyondaki en iyi uygunluk değerine sahip ilk 25 kromozom için sonuçlar EK 5’tedir.

İkinci kriter olarak Eşitlik 3.14’te tanımlanan ve hem hata değerini hem de sıfır sayısını birlikte değerlendiren ortak fonksiyon kullanılmıştır. Burada (λ) değeri hesaplamalarda 0,025 olarak kullanılmış ve *S.cerevisiae* mikroorganizması için hesaplamalar yapılarak sonuçlar değerlendirilmiştir. En son popülasyonda en iyi uygunluk değerine sahip beş kromozoma ait sonuçlar Tablo 4.10’dadır.

Tablo 4.10 *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak sıfır sayısı ve hata değerinden oluşan ortak fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR’ler.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	12,73	$1,85 \times 10^{-11}$	80	0,29	0,79	0,71	0,79
2	12,08	$8,27 \times 10^{-11}$	80	0,24	0,81	0,71	0,81
3	11,83	$1,47 \times 10^{-10}$	80	0,24	0,79	0,76	0,79
4	11,37	$4,27 \times 10^{-10}$	80	0,24	0,77	0,76	0,77
5	11,34	$4,61 \times 10^{-10}$	80	0,29	0,77	0,67	0,77

En iyi uygunluk derecesine sahip 5 kromozom için sonuçlara bakıldığında ağyapı çıkarımı açısından iyi olmadığı ve sadece hata değerinin kullanılması ile elde edilen çözümlere göre yönsüz ağyapı çıkarımının daha kötü olduğu, fakat ilişkilerin yönünü tahmin gücünün daha iyi olduğu ortaya çıkmıştır. Son popülasyondaki en iyi uygunluk değerine sahip ilk 25 kromozom için sonuçlar EK 5'tedir.

Uygunluk değeri için Eşitlik 3.14'ün kullanıldığı algorithmada ek olarak istatistiksel ağyapı çıkarımında elde edilen incelemeler doğrultusunda elde edilen "GGM korelasyon katsayısı 0,35'den büyük olan çiftlerin arasında ilişki vardır" varsayımı kullanılmıştır. İçsel değişkenliğin modellenmesi ile elde edilen *in silico* veriler üzerinden GGM korelasyon katsayısı 0,35'ten büyük ilişkiler tespit edilmiş ve bunların kromozomlar üzerindeki yerleri 1'e eşitlenerek *Saccharomyces cerevisiae* mikroorganizması için hesaplamalar yapılmıştır. Son popülasyondaki en iyi uygunluk değerine sahip beş kromozoma ait sonuçlar Tablo 4.11'dedir.

Tablo 4.11 *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak hem sıfır sayısı hem de hata değerinden oluşan ortak fonksiyonun kullanıldığı ve ek olarak GGM sayılarının dikkate alındığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR'ler.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	15,29	$5,08 \times 10^{-14}$	80	0,62	0,58	0,90	0,58
2	14,71	$1,94 \times 10^{-13}$	80	0,67	0,54	0,90	0,54
3	14,54	$2,91 \times 10^{-13}$	80	0,62	0,60	0,90	0,60
4	14,33	$4,66 \times 10^{-13}$	80	0,67	0,56	0,90	0,56
5	14,22	$6,08 \times 10^{-13}$	80	0,62	0,60	0,90	0,60

Elde edilen sonuçlara yönsüz olarak bakıldığında TPR'nin oldukça iyi olmasına karşın, FPR'in de yüksek çıktığı görülmektedir. Fakat sonuçlar; diğer optimizasyon yöntemlerine göre daha iyidir. Ayrıca bulunan ilişkilerin yönünü doğru olarak tahmin etme gücü de oldukça yüksektir. Son popülasyondaki en iyi uygunluk değerine sahip ilk 25 kromozom için sonuçlar EK 5'tedir.

Üçüncü bir kriter olarak da hem GGM yöntemi ile elde edilen sonucu ve hem de sıfır sayısını dikkate alan Eşitlik 3.15'teki ortak fonksiyon kullanılmıştır. Lamda (λ) değeri hesaplamalarda 0,025 olarak kullanılarak, bu uygunluk kriterinde *Saccharomyces cerevisiae* mikroorganizması için hesaplamalar yapılmış ve sonuçlar yönlü ve yönsüz olarak değerlendirilmiştir. En son popülasyonda en iyi uygunluk değerine sahip beş kromozoma ait sonuçlar Tablo 4.12'dedir.

Tablo 4.12 *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak hem GGM sonucunu hem de hata değerini içeren bir ortak fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR'ler.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	16,44	$1,01 \times 10^{-13}$	80	0,38	0,63	1	0,63
2	15,58	$1,06 \times 10^{-12}$	80	0,38	0,58	1	0,58
3	15,50	$9,91 \times 10^{-13}$	80	0,29	0,65	1	0,65
4	15,49	$1,29 \times 10^{-12}$	80	0,38	0,60	1	0,60
5	15,45	$9,93 \times 10^{-13}$	80	0,33	0,65	1	0,65

İçsel değişkenlik modellenerek üretilen *in silico* metabolom verilerine GGM yönteminin uygulanması ile elde edilen ağyapının TPR'si 0,76, FPR'si ise 0,18'dir. Elde edilen sonuçlara yönsüz olarak bakıldığında TPR'nin oldukça iyi olmasına karşın, FPR daha da yükselmiştir. Yönlü olarak ise genel olarak düşük bir TPR'ye sahiptir, yani bulunan ilişkilerin yönünü tahmin etme konusunda bu yöntem ile başarılı sonuçlar elde edilememiştir. Son popülasyondaki en iyi uygunluk değerine sahip ilk 25 kromozom için sonuçlar EK 5'tedir.

4.3 Nedenselliğe Dayalı Ağyapı Çıkarımı Tekniklerinin *Saccharomyces cerevisiae* Mikroorganizmasına ait *In Silico* Metabolom Verilerine Uygulanması

Yapılan tez çalışmasında *S.cerevisiae* mikroorganizmasına ait *in silico* olarak üretilen üç farklı veri tipine öncelikle istatistiksel ağyapı çıkarımı teknikleri

uygulanmış ve içsel değişkenlik ile hem içsel hem de çevresel değişkenliğin birlikte modellenerek üretildiği *in silico* verilerin sonuçlarının birbirine çok yakın olduğu görülmüştür. Bu nedenle nedenselliğe dayalı ağyapı çıkarımı enzimatik ve içsel değişkenlik modellenerek üretilen verilere uygulanmıştır. Bu verilere PC, CCD ve GES algoritması uygulanarak yarı yönlü ağyapılar elde edilmiş ve bu ağyapılar gerçek ağyapı ile karşılaştırılarak yöntemlerin analiz kabiliyetleri ortaya konmuştur.

4.3.1 PC Algoritmasının Uygulanması

Enzimatik ve içsel olmak üzere iki farklı değişkenlik modellenerek üretilen *in silico* verilere skor bazlı bir algoritma olan PC algoritması uygulanmıştır. Hata payı olarak yapılan hesaplamalarda 0,01 kullanılmış ve elde edilen yarı yönlü yapılar karşılaştırılarak sonuçlar değerlendirilmiştir. Sonuçlar Tablo 4.13'te yer almaktadır.

Tablo 4.13 *S.cerevisiae* mikroorganizmasına ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen *in silico* verilere PC algoritmasının uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi

Veri Türü	Yönsüz		Yönü Doğru Bulunan İlişki Sayısı
	TPR	FPR	
Enzimatik	0,48	0,11	1
İçsel	0,52	0,07	0

Elde edilen sonuçlara göre her iki veri türü içinde yönsüz olarak iyi sonuçlara ulaşılmıştır. Fakat bulunan ilişkinin yönünün tahmin edilmesi bakımından PC algoritması kötü sonuçlar vermiştir. Enzimatik değişkenlik modellenerek üretilen *in silico* verilere ait sonuçlara bakıldığında, toplam gerçek ağyapıda bulunan 21 ilişkiden 10 tanesini bulmuş fakat bunlardan sadece 1 tanesinin yönünü doğru olarak bulabilmiştir. Aynı şekilde içsel değişkenlik modellenerek üretilen *in silico* verilere ait sonuçlarda da durum aynı şekildedir.

4.3.2 CCD Algoritmasının Uygulanması

Enzimatik ve içsel olmak üzere iki farklı değişkenlik modellenerek üretilen *in silico* verilere CCD algoritması uygulanmış ve elde edilen yarı yönlü yapılar

karşılaştırılarak sonuçlar değerlendirilmiştir. Hata payı olarak yapılan hesaplamalarda 0,01 kullanılmıştır. Sonuçlar Tablo 4.14'te yer almaktadır.

Tablo 4.14 *S.cerevisiae* mikroorganizmasına ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen *in silico* verilere CCD algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi

Veri Türü	Yönsüz		Yönü Doğru Bulanan İlişki Sayısı
	TPR	FPR	
Enzimatik	0,43	0,12	0
İçsel	0,52	0,07	0

Yönsüz açıdan sonuçlar PC algoritmasına benzerdir. Yine PC algoritmasında olduğu gibi yönü belirlenen ilişkilerin hiçbirinin yönü doğru olarak bulunamamıştır.

4.3.3 GES Algoritmasının Uygulanması

Enzimatik ve içsel olmak üzere iki farklı değişkenlik modellenerek üretilen *in silico* verilere kısıt bazlı bir algoritma olan GES algoritması uygulanarak ve elde edilen yarı yönlü yapılar karşılaştırılarak sonuçlar değerlendirilmiştir. Hata payı olarak yapılan hesaplamalarda 0,01 kullanılmıştır. Sonuçlar Tablo 4.15'te yer almaktadır.

Tablo 4.15 *S.cerevisiae* mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen *in silico* verilere GES algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi

Veri Türü	Yönsüz		Yönü Doğru Bulanan İlişki Sayısı
	TPR	FPR	
Enzimatik	0,67	0,21	0
İçsel	0,76	0,26	0

Yönsüz olarak GGM sonuçlarına benzer sonuçlar elde edilmiştir. Yönü belirlenen ilişkilerin hiçbirinin yönü doğru olarak bulunamamıştır. Genel olarak nedenselliğe dayalı ağyapı çıkarımı teknikleri yönsüz olarak iyi sonuç vermelerine, rağmen ilişkilerin yönünü belirlemede başarısız kalmaktadırlar.

5 *Escherichia coli* MİKROORGANİZMASI İÇİN YAPILAN ÇALIŞMALAR

Tezin bu bölümü *Escherichia coli* mikroorganizması için yapılan çalışmaları içermektedir. *Saccharomyces cerevisiae* mikroorganizması için yapılan ve 4. Bölüm’de ayrıntılı olarak anlatılan tüm hesaplamalar paralel olarak *Escherichia coli* mikroorganizması için de tekrarlanmıştır. Bu doğrultuda öncelikli olarak bilgisayar ortamında enzimatik, içsel ve içsel-çevresel değişkenlikler modellenerek metabolom verileri *in silico* olarak üretilmiştir.

In silico olarak üretilen verilere istatistik, optimizasyon ve nedenselliğe dayalı yöntemler uygulanarak yönlü ve yönsüz ağyapı çıkarımı yapılmıştır. Elde edilen ağyapılar gerçek ağyapı ile karşılaştırılarak, sonuçlar istatistiksel olarak değerlendirilmiştir.

5.1 İstatistiksel Ağyapı Çıkarımı Tekniklerinin *Escherichia coli* Mikroorganizmasına ait *In Silico* Metabolom Verilerine Uygulanması

İstatistiğe dayalı ağyapı çıkarımı için 3 farklı *in silico* metabolom verisi üretilmiş ve her bir metabolit için 1000 tane kararlı hal etrafında değişen veri toplanmış ve elde edilen sonuçların güvenilirliğini ve tekrarlanabilirliğini test etmek için amacıyla bu şekilde 3 farklı veri tipinin her biri için 10 ayrı veri seti toplanmıştır. İstatistiğe dayalı yapılan tüm hesaplamalarda her bir veri tipi için elde edilen sonuçlar, bu 10 tekrarın ortalamasıdır.

Her bir veri setine PC, PPC1, PPC2 ve GGM yöntemleri uygulanmıştır. Elde edilen ağyapılar gerçek ağyapı bilgisini içeren modele ait Jacobian matris ile karşılaştırılarak istatistiksel olarak değerlendirilmiştir. Enzimatik değişkenliğe ait veri ile elde edilen sonuçlar Tablo 5.1’de verilmiştir.

Tablo 5.1 *E.coli* mikroorganizmasına ait enzimatik değişkenliğin modellenmesi ile üretilen *in silico* metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı yöntemlerinin istatistiksel olarak değerlendirme sonuçları

	PC	PPC1	PPC2	GGM
Doğru Negatif	12,6	104,2	112,5	97,4
Doğru Pozitif	34,5	19,6	12,9	25,8
Yanlış Negatif	4,5	19,4	26,1	13,2
Yanlış Pozitif	101,4	9,8	1,5	16,6
Doğru Pozitif Oranı	0,88	0,50	0,33	0,66
Yanlış Pozitif Oranı	0,89	0,09	0,01	0,15

Enzimatik değişkenlik kullanılarak üretilen *in silico* metabolom verilerine uygulanan korelasyon yöntemleri içinde GGM'nin daha iyi sonuç verdiği görülmektedir. Gerçek ağyapıda bulunan toplam 39 ilişkinin yaklaşık 26 tanesini bulmaktadır. Aralarında ilişki olmaması gereken 115 çiftin 17 tanesi arasında ilişki var olarak bulunmuştur. PPC1 ve PPC2 yöntemlerinin FPR'si daha düşük olmasına karşın GGM'e göre daha az ilişki ortaya açığa çıkartmaktadırlar. Hiçbir dolaylı etkileşimin ihmal edilmediği Pearson korelasyonu yöntemi ise beklenildiği gibi yine kötü sonuç vermiştir.

İçsel değişkenliğin ve hem içsel hem de çevresel değişkenliğin modellenmesi ile üretilen *in silico* metabolom verilerine ait ağyapı çıkarımı sonuçları Tablo 5.2'de ve Tablo 5.3'te yer almaktadır.

Tablo 5.2 *E.coli* mikroorganizmasına ait içsel değişkenliğin modellenmesi ile üretilen *in silico* metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı yöntemlerinin istatistiksel olarak değerlendirme sonuçları.

	PC	PPC1	PPC2	GGM
Doğru Negatif	13,4	101,4	111,9	102,3
Doğru Pozitif	35,5	17,2	14,3	23,1
Yanlış Negatif	3,5	21,8	24,7	15,9
Yanlış Pozitif	100,6	12,6	2,1	11,7
Doğru Pozitif Oranı	0,91	0,44	0,37	0,59
Yanlış Pozitif Oranı	0,88	0,11	0,02	0,10

Tablo 5.3 *E.coli* mikroorganizmasına ait hem içsel hem de çevresel değişkenliğin modellenmesi ile üretilen *in silico* metabolom verilerine uygulanan istatistiksel ağyapı çıkarımı sonuçlarının istatistiksel olarak değerlendirme sonuçları.

	PC	PPC1	PPC2	GGM
Doğru Negatif	12,9	101,5	111,1	103,3
Doğru Pozitif	35,9	16,4	13,8	24,2
Yanlış Negatif	3,1	22,6	25,2	14,8
Yanlış Pozitif	101,1	12,5	2,9	10,7
Doğru Pozitif Oranı	0,92	0,42	0,35	0,62
Yanlış Pozitif Oranı	0,89	0,11	0,03	0,09

Genel olarak *S.cerevisiae* mikroorganizması için yapılan istatistiksel çalışmalarda da ortaya çıktığı gibi üç veri tipi için en iyi sonuçlar yine GGM yöntemi ile elde edilmiştir. Bölüm 4.1’de detaylı olarak anlatılan *S.cerevisiae* mikroorganizmasına ait GGM sonuçlarını iyileştirmek için yapılan çalışmalar *E.coli* mikroorganizması için de tekrar edilmiştir: (i) Bonferoni Düzeltmesinin Uygulanması (0,01/153) ,(ii)Belirli Bir Korelasyon Katsayısı Değerinin Altındaki İlişkilerin Yok Sayılması ,(iii)Budama ,(iv)GGM Korelasyon Katsayısının Pearson Korelasyon Katsayısı ile Kontrolü. Yapılan çalışmaların sonuçları Tablo 5.4’te yer almaktadır.

Tablo 5.4 *E.coli* mikroorganizmasına ait üç farklı veri tipi için GGM sonuçlarını iyileştirmek için yapılan denemelerin sonuçları

DENENEN YÖNTEMLER	Enzimatik		İçsel		İçsel+ Çevresel	
	TPR	FPR	TPR	FPR	TPR	FPR
Orijinal Durum	0,66	0,15	0,59	0,10	0,62	0,09
Bonferoni düzeltmesi	0,63	0,10	0,52	0,04	0,54	0,03
Korelasyonu 0.15’ten Küçük İlişkilerin Yok Sayılması	0,58	0,06	0,45	0,01	0,48	0,01
“Pruning”	0,47	0,04	0,48	0,06	0,50	0,06
Pearson korelasyon katsayısı ile kontrol	0,58	0,14	0,46	0,09	0,51	0,08

GGM sonuçlarını iyileştirmek için yapılan denemeler içinden TPR’e göre FPR’yi daha fazla düşüren Bonferoni düzeltmesi ön plana çıkmaktadır. Bonferoni düzeltmesinin uygulanması ile gerçekte varolan 39 ilişkiden doğru olarak aralarında ilişki tespit ettiği metabolit çifti sayısı 26’dan 24’e düşerken, aralarında ilişki

bulunmadığı halde aralarında ilişki tespit edilen metabolit çifti sayısı ise 17'den 11'e düşürülmüştür.

5.2 Optimizasyona Dayalı Ağyapı Çıkarımı Tekniklerinin *Escherichia coli* Mikroorganizmasına ait *In Silico* Metabolom Verilerine Uygulanması ile Yönlü Ağyapı Çıkarımı

Saccharomyces cerevisiae mikroorganizmasında olduğu gibi *Escherichia coli* için de optimizasyona dayalı yapılan çalışmada sadece içsel değişkenliğin modellenmesi ile üretilen *in silico* deneysel verilere ağyapı çıkarımı teknikleri uygulanmış ve elde edilen sonuçlar gerçek ağyapı bilgisini içeren Jacobian matris ile karşılaştırılarak sonuçlar değerlendirilmiştir. Bu amaçla hücrenin minimum enzim üretimini yansıtan 3 farklı amaç fonksiyonu için farklı optimizasyon yöntemleri ile hesaplamalar gerçekleştirilmiş ve sonuçlar istatistiksel olarak değerlendirilmiştir.

5.2.1 Karesel Programlama ile Yönlü Ağyapı Çıkarımı

Amaç fonksiyonu olarak, Jacobian matrisinin köşegen olmayan elemanlarının karelerinin toplamının minimizasyonu kullanılarak içsel değişkenliğin modellenmesi ile elde edilen *in silico* metabolom verilerine karesel programlama uygulanmış ve elde edilen sonuçlar yönlü ve yönsüz olarak iki açıdan değerlendirilmiştir. Sonuçlar Tablo 5.5'te yer almaktadır.

Tablo 5.5 *E.coli* mikroorganizması için karesel programlama ile ağyapı çıkarımı sonuçları.

	Yönlü Ağyapı	Yönsüz Ağyapı
Doğru Negatif	21	21
Doğru Pozitif	9	37
Yanlış Negatif	30	2
Yanlış Pozitif	93	93
Doğru Pozitif Oranı	0,23	0,95
Yanlış Pozitif Oranı	0,82	0,82

Ağyapıda gerçekte bulunan 39 ilişkiden 37 tanesi yönsüz olarak doğru bulunurken, gerçekte aralarında ilişki olmayan 114 metabolit çiftinden 93 tanesini yanlış olarak bulmuştur. İlişkili bulduğu 37 ilişkinin de 7 tanesinin yönü doğru olarak bulunmuştur. Elde edilen sonuçların yönsüz olarak TPR'sinin yüksek olmasına karşın FPR'sinde çok yüksek olması nedeniyle sonuçlar ağyapı çıkarımı için iyi değildir.

5.2.2 Doğrusal Programlama ile Yönlü Ağyapı Çıkarımı

Amaç fonksiyonu olarak, Jacobian matrisinin köşegen olmayan elemanlarının mutlak değerleri toplamının minimizasyonu kullanılmış ve buna göre içsel değişkenliğin modellenmesi ile elde edilen *in silico* metabolom verilerine doğrusal programlama uygulanarak hesaplamalar yapılmış ve sonuçlar değerlendirilmiştir. Sonuçlar Tablo 5.6'da yer almaktadır.

Tablo 5.6 *E.coli* mikroorganizması için doğrusal programlama ile ağyapı çıkarımı sonuçları.

	Yönlü Ağyapı	Yönsüz Ağyapı
Doğru Negatif	0	0
Doğru Pozitif	33	39
Yanlış Negatif	6	0
Yanlış Pozitif	114	114
Doğru Pozitif Oranı	0,85	1
Yanlış Pozitif Oranı	1	1

Sonuçlara yönsüz olarak bakıldığında tüm metabolitler arasında bir ilişkinin olduğu ('fully connected') bir ağyapı ortaya çıkmıştır. Ağyapı çıkarımı açısından doğrusal programlama ile iyi bir sonuç elde edilememiştir.

5.2.3 Karesel ve Doğrusal Programlama Sonuçlarının İyileştirilmesi

S.cerevisiae mikroorganizmasında da olduğu gibi *E.coli* için de hem doğrusal programlama ile hem de karesel programlama ile ağyapı çıkarımı açısından iyi sonuçlar elde edilememiştir. Bu amaçla yöntemlerin sonuçlarını iyileştirmeye

yönelik *S.cerevisiae* mikroorganizması için yapılan çalışmalar tekrar edilmiştir. Sonuçlar Tablo 5.7’de verilmiştir.

Tablo 5.7 *E.coli* mikroorganizması için doğrusal programlama ve karesel programlama ile yapılan ağyapı çıkarımı sonuçlarını iyileştirmek için yapılan denemelerin sonuçları.

DENENEN YÖNTEM	Karesel Programlama				Doğrusal Programlama			
	Yönlü		Yönsüz		Yönlü		Yönsüz	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
Orijinal Durum	0,85	1	1	1	0,23	0,82	0,95	0,82
Eşik Değeri 0,15'ten küçük	0,85	1	1	1	0,23	0,82	0,95	0,82
Korelasyon Katsayısı Küçük	0,85	1	1	1	0,26	0,82	1	0,82
Değişkenlik Analizi	0,85	1	1	1	0,23	0,82	0,95	0,82
Korelasyon katsayısı 0,35'den büyük	0,85	1	1	1	0,23	0,82	0,95	0,82

Hem *E.coli* hem de *S.cerevisiae* mikroorganizmasına ait verilere uygulanan doğrusal programlama ve karesel programlama ile elde edilen sonuçların iyileştirilmesi için denenen yöntemlerden olumlu bir sonuç alınamamıştır.

5.2.4 Önerilen Genetik Algoritma Tabanlı ile Yönlü Ağyapı Çıkarımı

Saccharomyces cerevisiae mikroorganizması için yapılan hesaplamalarda daha iyi sonuç verdiği görülen iki uygunluk değeri için *E.coli* hesaplamaları tekrar edilerek sonuçlar değerlendirilmiştir:

1. Hem hata değeri hem de sıfır sayısını içeren ortak bir fonksiyon (GGM korelasyon katsayılarının 0,35'ten büyük olduğu durum dahil)
2. Hem GGM sonucunu hem de hata değerini içeren ortak bir fonksiyon

E.coli mikroorganizması için öncelikle hesaplamalar uygunluk değeri hem hata değeri hem de sıfır sayısını içeren ortak fonksiyon için gerçekleştirilmiştir (Eşitlik 3.14) ve ek olarak aynı algoritma içinde GGM sonuçları 0,35'ten büyük olan

ilişkilerin kromozomlarda karşılık geldikleri yerlerin değerleri 1'e eşitlenerek en iyi sonucu verecek kromozomlar aranmıştır. Lamda (λ) değeri yine *S.cerevisiae* de kullanıldığı gibi 0,025 olarak kullanılarak hesaplamalar gerçekleştirilmiştir. Son popülasyondaki en iyi uygunluk değerine sahip beş kromozoma ait sonuçlar Tablo 5.8'dedir.

Tablo 5.8 *E.coli* mikroorganizması için uygunluk değeri kriteri olarak hem sıfır sayısı hem de hata değerinden oluşan ortak fonksiyonun kullanıldığı ve ek olarak GGM sayılarının dikkate alındığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR'ler.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	10,07035	$1,20 \times 10^{-3}$	286	0,15	0,12	0,28	0,12
2	10,07034	$1,20 \times 10^{-3}$	286	0,15	0,11	0,33	0,11
3	10,07032	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
4	10,07032	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
5	10,07028	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11

Elde edilen sonuçlara yönsüz olarak bakıldığında FPR'lerin düşük olmasına karşın, TPR'ler de düşük değerlere sahiptir. Fakat sonuçlar; diğer optimizasyon yöntemlerine göre daha iyidir. Son popülasyondaki en iyi uygunluk değerine sahip ilk 25 kromozom için sonuçlar EK 6'dadır.

S. cerevisiae mikroorganizması için diğer yöntemlere nispeten daha iyi sonuç veren diğer bir uygunluk değeri kriteri de hem GGM sonucunu hem de hata değerini içeren ortak bir fonksiyondur (Eşitlik 3.15). Bu uygunluk değeri kriteri kullanılarak *E.coli* mikroorganizması için de hesaplamalar tekrar edilmiş ve sonuçlar değerlendirilmiştir. Lamda değeri hesaplamalarda 0,025 olarak kullanılmıştır. Son popülasyondaki en iyi uygunluk değerine sahip beş kromozoma ait sonuçlar Tablo 5.9'dadır.

Tablo 5.9 *E.coli* mikroorganizması için uygunluk değeri kriteri olarak hem GGM sonucunu hem de hata değerini içeren ortak bir fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR’ler

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	13,71	$1,87 \times 10^{-4}$	158	0,36	0,68	0,85	0,68
2	13,54	$3,99 \times 10^{-6}$	158	0,33	0,68	0,82	0,68
3	13,43	$1,63 \times 10^{-4}$	158	0,31	0,70	0,82	0,70
4	13,39	$6,90 \times 10^{-6}$	159	0,33	0,68	0,85	0,68
5	13,35	$1,32 \times 10^{-6}$	158	0,31	0,68	0,87	0,68

E.coli modelinde içsel değişkenlik modellenerek üretilen *in silico* metabolom verilerine GGM yönteminin uygulanması ile elde edilen ağyapının TPR’si 0,59, FPR’si ise 0,10’dur. Elde edilen sonuçlara yönsüz olarak bakıldığında bu amaç fonksiyonu için yapılan hesaplamalarda *S.cerevisiae* sonuçlarında da olduğu gibi TPR’nin oldukça iyi olmasına karşın, FPR, TPR’ye nazaran daha da yükselmiştir. Yönlü olarak ise yine *S.cerevisiae* mikroorganizması için yapılan hesaplamalarda olduğu gibi genel olarak düşük bir TPR’ye sahiptir, yani bulunan ilişkilerin yönünü tahmin etme konusunda her iki mikroorganizma için de bu yöntem ile başarılı sonuçlar elde edilememiştir. Son popülasyondaki en iyi uygunluk değerine sahip ilk 25 kromozom için sonuçlar EK 6’dadır.

5.3 Nedenselliğe Dayalı Ağyapı Çıkarımı Tekniklerinin *Escherichia coli* Mikroorganizmasına ait *In Silico* Metabolom Verilere Uygulanması

Yapılan tez çalışmasında *S.cerevisiae* mikroorganizmasına için yapılan nedenselliğe dayalı ağyapı çıkarımı *E.coli* mikroorganizması için de tekrar edilmiştir. *E.coli* mikroorganizmasına ait enzimatik ve içsel değişkenlik modellenerek üretilen verilere nedenselliğe dayalı ağyapı çıkarımı tekniklerinden PC, CCD ve GES algoritması uygulanmış ve elde edilen ağyapılar gerçek ağyapı ile karşılaştırılarak sonuçlar değerlendirilmiştir.

5.3.1 PC Algoritmasının Uygulanması

S.cerevisiae mikroorganizma için yapılan çalışma *E.coli* için de tekrar edilerek enzimatik ve içsel olmak üzere iki farklı değişkenlik modellenerek üretilen *in silico* verilere PC algoritması uygulanmıştır. Hata payı olarak yapılan hesaplamalarda 0,01 olarak kullanılmıştır. Sonuçlar Tablo 5.10'te yer almaktadır.

Tablo 5.10 *E.coli* mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen *in silico* verilere PC algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi

Veri Türü	Yönsüz		Yönü Doğru Bulanan İlişki Sayısı
	TPR	FPR	
Enzimatik	0,36	0,04	3
İçsel	0,36	0,07	3

Elde edilen sonuçlar yönsüz olarak değerlendirildiğinde FPR'nin her iki veri tipi için de çok düşük olduğu gözlenmektedir. Fakat TPR'si düşük kalmıştır. Her iki veri tipi için de önerilen ağyapıda aralarında gerçekte ilişki bulunan 39 metabolit çiftinin sadece 3 tanesinin yönü doğru olarak tespit edebilmiştir.

5.3.2 CCD Algoritmasının Uygulanması

Enzimatik ve içsel olmak üzere iki farklı değişkenlik modellenerek üretilen *in silico* verilere CCD algoritması uygulanmış ve elde edilen yarı yönlü yapılar karşılaştırılarak sonuçlar değerlendirilmiştir. Elde edilen sonuçlar Tablo 5.11'te yer almaktadır. Hata payı olarak yapılan hesaplamalarda 0,01 kullanılmıştır.

Tablo 5.11 *E.coli* mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen *in silico* verilere CCD algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi

Veri Türü	Yönsüz		Yönü Doğru Bulanan İlişki Sayısı
	TPR	FPR	
Enzimatik	0,36	0,04	0
İçsel	0,36	0,07	1

S.cerevisiae'de olduğu gibi yönsüz açıdan sonuçlar PC algoritmasına benzerdir. Fakat PC algoritmasına göre yönü doğru belirlenen ilişki sayısı daha kötüdür.

5.3.3 GES Algoritmasının Uygulanması

Enzimatik ve içsel olmak üzere iki farklı değişkenlik modellenerek üretilen *in silico* verilere GES algoritması uygulanarak ve elde edilen yarı yönlü yapılar karşılaştırılarak sonuçlar değerlendirilmiştir. Hata payı olarak 0,01 kullanılmıştır. Sonuçlar Tablo 5.12'de yer almaktadır.

Tablo 5.12 *E.coli* mikroorganizması ait içsel ve enzimatik değişkenliğin modellenmesi ile elde edilen *in silico* verilere GES algoritmanın uygulanması sonucunda ortaya çıkan ağyapının değerlendirilmesi

Veri Türü	Yönsüz		Yönü Doğru Bulanan İlişki Sayısı
	TPR	FPR	
Enzimatik	0,49	0,14	0
İçsel	0,54	0,14	0

Yönsüz olarak elde edilen sonuçlara bakıldığında enzimatik değişkenlik veri tipi için gerçekte yönü belirlenen ilişkilerin hiçbirinin yönü doğru olarak bulunamamıştır. Genel olarak nedenselliğe dayalı ağyapı çıkarımı teknikleri yönsüz olarak iyi sonuç vermelerine rağmen ilişkilerin yönünü belirlemede başarısız kalmaktadırlar.

6 TARTIŞMA VE YORUM

Bu tez çalışmasındaki genel amaç, Yukarıdan Aşağıya Sistem Biyolojisi yaklaşımlarını kullanarak literatürde en yaygın çalışılan *Escherichia coli* ve *Saccharomyces cerevisiae* mikroorganizmalarına ait metabolik ağyapıyı çıkararak hücrenin biyolojik amacını anlayabilmektir. Bu amaç doğrultusunda, her iki mikroorganizma için de dinamik modeller kullanılarak bu modeller üzerinde 3 farklı biyolojik değişkenlik modellemesi yapıldı. Böylelikle bilgisayar ortamında 3 farklı metabolom veri seti oluşturuldu. Üretilen bu deney setine farklı ağyapı çıkarımı teknikleri uygulanarak yönlü ve yönsüz olarak ağyapı elde edilmeye çalışıldı. Tezin amacı kapsamında yapılan ağyapı çıkarımı hesaplamaları kısaca özetlenirse;

(i) İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımı Hesaplamaları

- Kullanılan Yöntemler: PC, PP1, PPC2, GGM
- Yöntemlerin Uygulandığı Veri Tipleri: (i) Enzimatik değişkenlik, (ii) İçsel değişkenlik, (iii) Hem içsel değişkenlik hem de çevresel değişkenlik modellenmesi ile üretilen *in silico* metabolom verisi

(ii) Optimizasyona Dayalı Yönsüz Ağyapı Çıkarımı Hesaplamaları

- Kullanılan Yöntemler: Doğrusal Programlama, Karesel Programlama, Karma Tamsayılı Doğrusal Programlama ve Genetik Algoritma ve 6 farklı amaç fonksiyonu
- Yöntemlerin Uygulandığı Veri Tipleri: İçsel değişkenlik modellenmesi ile üretilen *in silico* metabolom verisi

(iii) Nedenselliğe Dayalı Yönsüz Ağyapı Çıkarımı Hesaplamaları

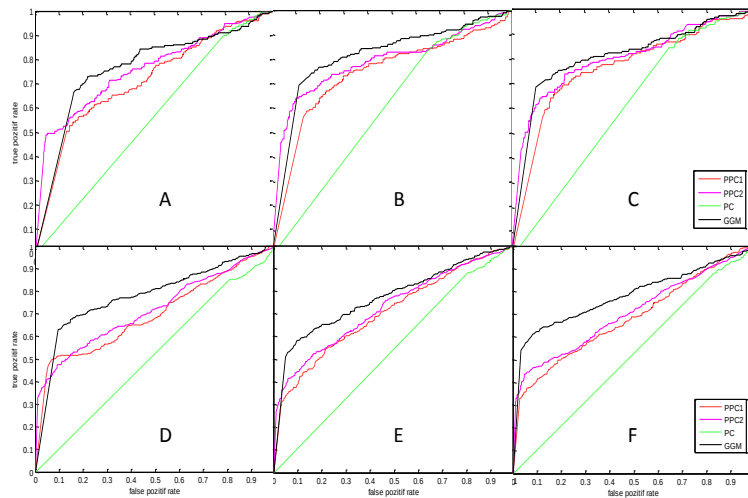
- Kullanılan Yöntemler: PC, CCD ve GES Algoritması
- Yöntemlerin Uygulandığı Veri Tipleri: (i) İçsel değişkenlik, (ii) Enzimatik değişkenlik modellenmesi ile üretilen *in silico* metabolom verisi

Belirtilen yöntemlerle iki farklı mikroorganizmaya ait yönlü ve yönsüz ağyapı çıkarımı yapılmış ve sonuçlar gerçek ağyapı ile karşılaştırılarak her bir yöntemin avantaj ve dezavantajı ortaya çıkarılmıştır. Öncelikle üç farklı veri tipi için istatistiğe

dayalı yönsüz ağyapı çıkarımı yapılmıştır. Yapılan bu hesaplamalarda aynı zamanda üç farklı veri tipi için karşılaştırmalı değerlendirme yapma imkanı bulunmuş ve deney setini en iyi açığa vuran biyolojik değişkenlik modellenmesi belirlenmiştir. Yapılan değerlendirmeler sonucu her iki mikroorganizma için de en iyi sonuçların içsel değişkenlik modellenmesi ile üretilen *in silico* metabolom verisi ile elde edildiği görülmüştür. Yani deneysel sonuçları en iyi yansıtan biyolojik değişkenlik türünün içsel değişkenlik olduğu ortaya çıkarılmıştır.

6.1 İstatistiğe Dayalı Yönsüz Ağyapı Çıkarımı Sonuçlarının Değerlendirilmesi

S.cerevisiae ve *E.coli* mikroorganizmalarına ait gerçekleştirilen yönsüz ağyapı çıkarımında *in silico* verilere 4 farklı korelasyon yöntem uygulanmıştır. Bu bağlamda yapılan çalışmalarda hem *S.cerevisiae* hem de *E.coli* mikroorganizmaları için GGM metodunun iyi sonuçlar verdiği görülmektedir. Ayrıca istatistiksel yöntemleri değerlendirmede yaygın olarak kullanılan Receiver Operating Characteristic (ROC) eğrisi [Metz, 1978] uygulanarak istatistiksel yöntemlerin analiz kabiliyetleri ortaya çıkarılmıştır. Bunun için 0-1 arası tüm hata değeri değerleri ('cut-off') 0.01 aralıklarla taranarak her durumda her iki mikroorganizma için TPR ve FPR değerleri hesaplanmış, ve bu değerler grafiksel olarak çizdirilerek ROC eğrileri elde edilmiştir (Şekil 6.1).



Şekil 6.1 *S.cerevisiae* ve *E.coli* mikroorganizmaları sonuçlarına ait ROC eğrisi. *S.cerevisiae* mikroorganizması için (A) enzimatik değişkenlik (B) içsel değişkenlik,

(C) hem içsel değişkenlik hem de çevresel değişkenlik modellenmesi ile, *E.coli* mikroorganizması için (D) enzimatik değişkenlik, (E) içsel değişkenlik, (F) hem içsel değişkenlik hem de çevresel değişkenlik modellenmesi ile üretilen *in silico* metabolom verilerine uygulanan yöntemler için ROC eğrileri.

Şekil 6.1’de görüldüğü gibi 45 derece çizgisine en uzak grafiğe sahip olan GGM, *S.cerevisiae* ve *E.coli* mikroorganizmalarına ait üç veri tipi için de metabolik ağyapıyı yönsüz olarak tahmin etmede en başarılı korelasyon yöntemidir. Grafiklerde görüldüğü gibi, her iki mikroorganizma için de 45 derece çizgisine en yakın grafiğe sahip Pearson korelasyon metodu, ağyapı çıkarımı için kötü sonuçlar vermektedir. Grafiklerde birbirine yakın sonuçlar veren birinci ve ikinci dereceden kısmi korelasyon yöntemleri ise Pearson korelasyonuna göre daha iyi sonuç vermelerine rağmen yine de her iki yöntem de yönsüz ağyapı çıkarmada GGM yönteminin başarısını yakalayamamaktadır.

Yöntemlerin analiz kabiliyetlerini ROC eğrileri üzerinden belirlemek için ayrıca literatürde eğri altında kalan alan (Area Under Curve, AUC) olarak bilinen bir değerlendirme kriteri kullanılmaktadır [Soranzo et al., 2007; Werhli et al., 2006]. Bu amaç doğrultusunda, her bir yönteme ait ROC eğrisinin altında kalan alanlar hesaplanmıştır. ROC eğrilerinin altında kalan alanlar nümerik olarak Trapez kuralı ile hesaplanmış ve bu amaçla MATLAB’ın *trapz* fonksiyonu kullanılmıştır. Sonuçlar Tablo 6.1’dedir.

Tablo 6.1 *E.coli* ve *S.cerevisiae* mikroorganizmaları için üç farklı biyolojik değişkenliğe ait ROC eğrilerinin AUC değerleri

	S.cerevisiae			E.coli		
	Enzimatik	İçsel	İçsel-Çevresel	Enzimatik	İçsel	İçsel-Çevresel
PC	0,55	0,60	0,60	0,52	0,52	0,52
PPC1	0,71	0,74	0,76	0,70	0,70	0,68
PPC2	0,76	0,79	0,80	0,72	0,75	0,71
GGM	0,77	0,81	0,81	0,78	0,78	0,78

Eğer AUC değeri 0,5’e yakınsa rastgele bir tahmine karşılık gelmektedir. Yani bu kullanılan yöntemin kötü olduğunu göstermektedir. AUC değerinin 0,7’den küçük

olması yöntemin ağyapı çıkarımı açısından yetersiz olduğunun, 0,7 ile 0,8 arasında olması yöntemin kabul edilebilir olduğunun, 0,8'den büyük olması ise yöntemin ağyapı çıkarımı açısından analiz kabiliyetinin iyi olduğunun göstergesidir [Soranzo et al., 2007].

Her iki mikroorganizmaya ait ROC eğrilerinin altında kalan alanların değerlerine bakıldığında üç farklı değişkenlik türü için de GGM yönteminin genel olarak ya 0,8'e çok yakın ya da 0,8'den büyük değerler verdiği görülmektedir. Bu, GGM yönteminin başarısını başka bir açıdan ortaya koymaktadır. Diğer yöntemlere bakıldığında ise, 0,7-0,8 arasında değerler veren PPC1 ve PPC2 yöntemleri kabul edilebilirdir. Diğer değişkenlerin dolaylı etkileri barındıran PC metodunun ise 0,5'e yakın sonuçlar verdiği ve yöntemin ağyapı çıkarımı açısından başarısız olduğu görülmektedir. PC yönteminin kötü sonuçlar verdiği bir kez daha doğrulanmıştır.

Hem *E.coli* hem de *S.cerevisiae* mikroorganizmalarının yönsüz olarak ağyapılarını ortaya koymada GGM metodu istatistiksel yöntemler arasında en başarılısıdır. Her iki mikroorganizma için de GGM sonuçları karşılaştırıldığında, *E.coli* mikroorganizmasında GGM yöntemi genel olarak üç veri tipi için de *S.cerevisiae*'ye göre daha düşük bir FPR vermiştir. Fakat *S.cerevisiae* mikroorganizmasında GGM yöntemi ile TPR açısından daha iyi sonuçlar gözlenmektedir. Ayrıca, her iki mikroorganizmanın GGM sonuçlarını iyileştirecek yöntemler denenmiş ve Bonferoni düzeltmesi kullanılarak GGM sonuçlarının iyileştirilebileceği gösterilmiştir.

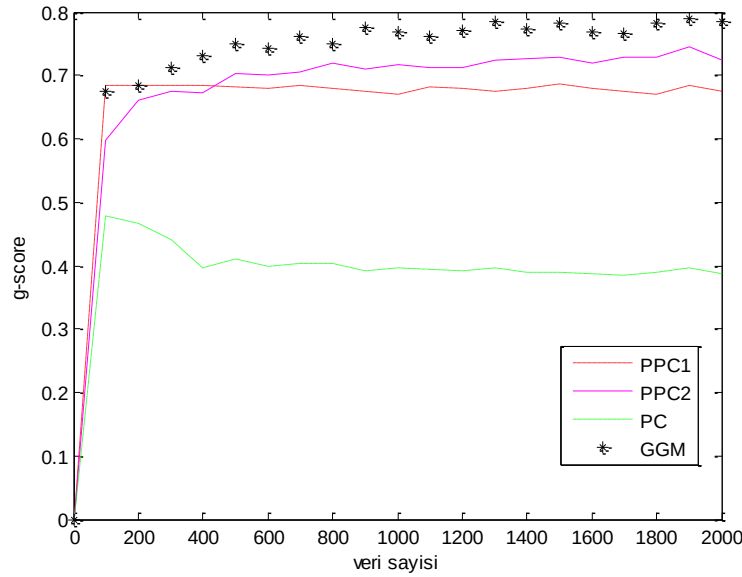
İstatistiğe dayalı yapılan ağyapı çıkarımında korelasyon katsayıları kullanılarak iki değişken arasında doğrusal ilişkiler değerlendirilmiştir. Doğrusal olmayan ilişkileri değerlendirmek için de literatürde yöntemler mevcuttur fakat bu yöntemler ağyapı çıkarımı için hemen hemen aynı sonuçları verdiği önceki çalışmalarda gösterilmiştir [Çakır ve ark., 2009]. Bu nedenle yapılan tez çalışmasında yalnızca farklı korelasyon yöntemleri kullanılmıştır.

Veri sayısının ağyapı çıkarımı üzerinde önemli olduğu bilinmektedir. Bu nedenle *S.cerevisiae* mikroorganizması için farklı veri sayılarında yapılan çalışmalar tekrar edilmiş elde edilen sonuçlar karşılaştırılmıştır. Bu amaçla 100-2000 veri sayısı

arasında 100'er aralıklarla veri sayısı kullanarak eğriler çizdirilmiştir. Eğriler oluşturulurken g-skor (geometrik ortalama skoru) terimi kullanılmıştır [Çakır ve ark., 2009]. Bu sayı TPR ve (1-FPR)'nin geometrik ortalamasının alınması ile hesaplanır:

$$g - skor = \sqrt{TPR \times (1 - FPR)} \quad (6.1)$$

Bu skor 0 ile 1 arasında değer almaktadır. Değer 1'e ne kadar yakınsa ağyapı çıkarımı sonucu o kadar iyi, 0'a ne kadar yakınsa ağyapı çıkarımı sonucu o kadar kötüdür. Eğriler veri sayısına karşılık gelen g-skorlar için çizdirilmiştir (Şekil 6.2).



Şekil 6.2 Farklı veri sayıları için *S.cerevisiae* sonuçlarına ait eğriler. 100-2000 veri sayısı arası 100'er aralıkla taranarak her durumda g-skor değerleri hesaplanmış ve bu değerler grafiksel olarak çizdirilerek eğriler elde edilmiştir.

Şekil 6.2'de de görüldüğü gibi veri sayısı 500'e kadar düşse bile yaklaşık olarak birbirine benzer sonuçlar elde edilebilmektedir. Yani ağyapı çıkarımı için 1000 veri yerine 500 veri de kullanılarak da benzer sonuçlara ulaşılabilir. Fakat veri sayısı daha da azaldıkça sonuçların kötüye doğru bir eğilim çizdiği görülmektedir. Veri sayısının azalması ağyapı çıkarımı sonuçlarında FPR'nin azalışına nispeten daha fazla TPR'de azalmaya neden olmakta ve sonuçları olumsuz yönde

etkilemektedir. Veri sayısının artması ise ağyapı çıkarımı sonuçlarını belirli bir aralıktan sonra etkilememekte ve birbirine benzer sonuçları vermektedir.

Ayrıca belirli bir standart sapma etrafında üretilen *in silico* veriler ile *E.coli* ve *S.cerevisiae*'ye ait deneysel olarak ölçülmüş metabolom verisi karşılaştırılarak gerçeğe uygun bir veri seti kullanılıp kullanılmadığının kontrolü yapılmak istenmiştir. Fakat literatürde bu mikroorganizmalara ait tekrar sayısı fazla olan deneysel veri seti bulunmadığından dolayı insan hücresine ait 1020 tekrarlı bir metabolom veri seti [Illig et al., 2010] ele alınmış ve deneysel verilerin standart sapmalarına bakılmıştır. İnsan hücresine ait deneysel veri setinde genel olarak ölçümlerin standart sapma değerleri 0,25 civarında toplanırken, tez çalışmasında her iki mikroorganizma için de enzimatik değişkenlik modellenerek üretilen *in silico* deneysel verilerin standart sapmaları 0,05 etrafındadır. Gerçekte, insana ait bir hücrenin daha karmaşık bir sisteme sahip olmasından dolayı diğer mikroorganizmalara göre tekrarlanan ölçümler arasındaki farklılıkların daha fazla olması beklenmektedir. Bunu göz önünde bulundurarak, *E.coli* mikroorganizması için verilerin standart sapmaları 0,15 olacak şekilde enzimatik değişkenlik modellenerek tekrar *in silico* metabolom veri üretilmiş ve GGM yöntemi ile tekrar ağyapı çıkarımı yapılmıştır. Elde edilen sonuçlara bakıldığında standart sapmanın 0,15 olarak ayarlandığı durumda g-skor 0,74 olarak bulunmuştur. Yapılan tez çalışmasında standart sapmanın 0,05 olarak ayarlandığı durumda ise g-skor 0,71'dir. Yani yapılan tez çalışmasında kullanılan deneysel verilerin sonuçları ile standart sapma açısından gerçeğe daha uygun olarak üretilen deneysel verilerin sonuçları birbirine benzerdir. Buna dayanarak kullanılan *in silico* deneysel veri seti parametrelerinin uygun olduğu kanıtlanmıştır.

6.2 Optimizasyona Dayalı Yönlü Ağyapı Çıkarımı Sonuçlarının Değerlendirilmesi

Hem *S.cerevisiae* hem de *E.coli* mikroorganizmasına ait içsel değişkenlik modellenerek üretilen *in silico* verilere farklı optimizasyon yöntemleri uygulanarak ağyapılar hem yönlü hem de yönsüz olarak elde edilerek sonuçlar değerlendirildi. Öncelikle minimum enzim üretimi prensibinin yansıtıldığı üç amaç fonksiyonuyla

$A_j = -2d$ sistemi için tek bir çözüm veren doğrusal programlama, karma tamsayılı doğrusal programlama ve karesel programlama kullanılarak ağyapı çıkarımı yapıldı. Elde edilen sonuçlar hem yönlü olarak hem de yönsüz olarak gerçek ağyapı ile karşılaştırıldı ve yöntemlerin analiz kabiliyetleri için değerlendirmeler yapıldı.

Her iki mikroorganizma için de hem yönlü olarak hem de yönsüz olarak bu yöntemler kötü sonuçlar vermektedir. Bu sonuçların iyileştirilmesi adına Bölüm 4.2’de anlatılan bazı denemeler gerçekleştirildi fakat hiçbirisinden olumlu sonuç alınamadı. Bu bağlamda bu yöntemlerin metabolik ağyapı çıkarımı için kötü sonuçlar vermesinin nedenleri araştırıldı. Karesel programlama için kullanılan amaç fonksiyonu Jacobian matrisinin köşegen elemanlar haricindeki elemanlarının karelerinin toplamının minizasyonudur. Fakat zaten bu amaç fonksiyonunun gerçek Jacobian matrisindeki değeri *S.cerevisiae* modelinde $2,78 \times 10^{12}$ ’dir. Yani, bu da karesel programlamanın bu tür büyük metabolik modeller için iyi sonuç vermemesini açıkça göstermiştir. Doğrusal ve karma tamsayılı doğrusal programlama sonuçları incelendiğinde ise elde edilen çözümlerin her iki model için de serbestlik derecesi kadar sıfır içerdikleri görülmüştür. Yani *E.coli* için doğrusal programlama ile elde edilen ağyapıda bu modelin serbestlik derecesi kadar (153) sıfır içerdikleri görülmüştür. Aynı şekilde *S.cerevisiae* için de yapılan karma tamsayılı ve doğrusal programlama sonuçlarında da modelin serbestlik derecesi kadar (78) sıfır içerdikleri gözlenmiştir. Fakat metabolik ağyapılar seyrek aralıklı bir yapıya sahiptir [Yeung et al., 2002]. Yani gerçekte Jacobian matris yapılarında serbestlik derecesinden daha fazla sıfır olduğu bilinmektedir. Bu nedenle sistemin, serbestlik derecesinden daha fazla sıfır sayısı içerecek bir çözüm verecek şekilde çözümüne olanak sağlayan yeni bir genetik algoritma tabanlı bir yöntem önerildi. Bu sayede elde edilen çözümlerin, serbestlik derecesinden daha fazla sıfır sayısı içerdikleri görülmüştür. Fakat yine de elde edilen çözümler serbestlik derecesi etrafında toplanmıştır. Bu problemin metabolik ağyapı sistemlerine özgü olabileceği düşünülmektedir.

Önerilen algoritmada (i) hata değeri, (ii) hem hata değeri hem de sıfır sayını içeren bir fonksiyon, (iii) hem GGM sonucunu hem de hata değerini içeren bir fonksiyon olmak üzere 3 farklı amaç fonksiyonu kullanılarak ağyapı çıkarımı gerçekleştirildi. Ayrıca GGM sonuçlarının incelenmesi ile elde edilen GGM korelasyon katsayısının 0,35’ten büyük olduğu durumlarda gerçekte bir ilişki olduğu

bilgisi hem hata değeri hem de sıfır sayısını amaç fonksiyonu olarak içeren algoritmaya eklenerek de bir hesaplama yapıldı.

S.cerevisiae mikroorganizması için yukarıda anlatılan tüm bu amaç fonksiyonları için hesaplamalar yapılarak gerçek ağyapıyı en iyi şekilde ortaya çıkaracak amaç fonksiyonu belirlenmeye çalışıldı. Bu bağlamda yapılan hesaplamalarda, ağyapı çıkarımı açısından yönlü ve yönsüz olarak en iyi sonuç, amaç fonksiyonu olarak hem hata değeri hem de sıfır sayısını içeren fonksiyonunun (GGM korelasyon katsayısının 0,35'ten büyük olduğu durum da dahil) kullanıldığı durumda elde edilmiştir.

Hem hata değeri hem de sıfır sayısını içeren fonksiyonun (GGM korelasyon katsayısının 0,35'ten büyük olduğu durumda dahil) kullanıldığı durumda yapılan hesaplamalarda yönsüz olarak diğer optimizasyon yöntemlerine nazaran daha düşük bir FPR ve 0,90 gibi daha yüksek bir TPR elde edilmiştir. Bulunan ilişkiler yönlü olarak değerlendirildiğinde ise, yine diğer optimizasyon yöntemlerine göre daha iyi sonuçlar vermiş, gerçek ağyapıda bulunan toplam 21 ilişkiden 13 tanesini yönleri ile birlikte doğru bulma başarısı göstermiştir.

E.coli için de hem hata değeri hem de sıfır sayısını içeren(GGM korelasyon katsayısının 0,35'ten büyük olduğu durumda dahil) amaç fonksiyonu kullanılarak hesaplamalar gerçekleştirilmiş ve sonuçlar değerlendirilmiştir. Elde edilen sonuçlar *S.cerevisiae* için elde edilen sonuçlara göre daha kötü olmasına karşın, yine de bu amaç fonksiyonu ile hem yönlü hem de yönsüz olarak diğer optimizasyon yöntemlerine nazaran daha iyi sonuçlar elde edilmiştir. Ayrıca diğerlerine göre nispeten daha iyi sonuç veren hem GGM sonucunu hem de hata değerini içeren amaç fonksiyonu da kullanılarak hesaplamalar yapılmıştır. *S.cerevisiae* mikroorganizması için ağyapı çıkarımı hesaplamalarında olduğu gibi en iyi sonuçlar hem hata değeri hem de sıfır sayısını içeren fonksiyonunun (GGM korelasyon katsayısının 0,35'ten büyük olduğu durum da dahil) kullanılması ile elde edilmiştir.

Önerilen yöntem ile diğer optimizasyon yöntemlerine göre daha iyi sonuçlara ulaşılmasına rağmen, gerçek ağyapıya ulaşılammamıştır. Bunun nedenlerini araştırmak amacıyla, *S.cerevisiae* mikroorganizması ele alınmış, gerçek ağyapı bir

kromozom ile temsil edilerek uygunluk değerleri kriterlerinde kullanılan hata değerine bakılmıştır. Gerçek çözümü temsil eden kromozomun hata değeri $2,012 \times 10^{-30}$ 'dur. Bu kromozom üzerinde birer genin değeri rastgele değiştirilerek (0 ise 1, 1 ise 0 yapılarak) gerçek çözüm etrafında 50 farklı kromozom üretilmiş ve bu kromozomlar en küçük kareler yöntemi ile çözdürülerek elde edilen sonuçlar değerlendirilmiştir. Böylelikle gerçek çözüme çok yakın çözümler veren kromozomların hata değerleri kontrol edilmiştir. Sonuçlara bakıldığında, sadece bir genin değerinin değiştirilmesinin hata değerinde büyük sapmalara neden olduğu görülmüştür. Genel olarak gerçek çözümü temsil eden kromozom üzerinde bir genin değerinin değiştirilmesi hata değerini 10^{-5} - 10^{-6} 'lara kadar düşürmüştür. Fakat gerçek çözümden uzak çözümler veren kromozomların içerisinde çok daha iyi hata değerine sahip kromozomlar mevcuttur. Bu nedenle genetik algoritmada yapılan çalışmalarda sonuç olarak bu kromozomlar bulunmakta ve gerçek çözüme yaklaştıracak çözümlerin hata değerlerinin çok kötü olması nedeniyle iterasyon sayısı ne kadar olursa olsun gerçek ağyapı elde edilememektedir.

Genel olarak optimizasyon yöntemleri ile yapılan ağyapı çıkarımında yönlü ve yönsüz olarak başarılı sonuçlar elde edilemedi. Fakat önerilen genetik algoritma tabanlı yöntem ile diğer yöntemlere nazaran daha iyi sonuçlar elde edilmiş ve yön bulma konusunda başarılı bir yöntem olduğu gösterilmiştir.

6.3 Nedenselliğe Dayalı Yönlü Ağyapı Çıkarımı Sonuçlarının Değerlendirilmesi

Nedenselliğe dayalı ağyapı çıkarımı çalışmalarında 3 farklı yöntem izlendi ve elde edilen sonuçlar değerlendirilerek yöntemlerin analiz kabiliyetleri ortaya konmaya çalışıldı. Her iki mikroorganizma için yapılan çalışmalara bakıldığında kullanılan yöntemlerin metabolik ağyapılara ait ilişkilerin yönünü bulma konusunda başarısız oldukları görüldü. Bunun nedeninin metabolik ağyapıda yer alan reaksiyonların tersinir oluşundan kaynaklandığı düşünülmektedir. Bu algoritmalar (CCD hariç) yönlü döngüsüz ağyapı çıkarımı için tasarlanmıştır.

Fakat bu yöntemlerle elde edilen sonuçlar yönsüz olarak değerlendirildiğinde, daha iyi sonuçlar elde edildi. Özellikle yönsüz olarak GES algoritmasının sonuçları (Tablo 4.15 ve 5.12) diğerlerine göre daha iyidir ve her iki mikroorganizma için de bulunan sonuçlar GGM sonuçlarına yakındır. Nedenselliğe dayalı olarak yapılan ağyapı çıkarımında yönsüz olarak ağyapı bulunmasında GES algoritmasının öne çıktığı görülmektedir.

7 SONUÇ

S.cerevisiae ve *E.coli* mikroorganizmalarına ait *in silico* metabolom verileri için yapılan ağyapı çıkarımı çalışmalarından elde edilen genel sonuçlar şunlardır:

Yönsüz olarak yapılan çalışmalarla seyrek yapılı bir ağyapı verme eğiliminde olan GGM yöntemi ile başarılı sonuçlar elde edilmesi hücrenin biyolojik amacının minimum enzim üretimi olduğunu kanıtlar niteliktedir.

Yönlü olarak ise başarılı sonuçlar elde edilememiştir. Fakat metabolik ağyapıların yapısını ortaya koymak için yapılacak olan özellikle de optimizasyona dayalı ağyapı çıkarımı çalışmalarında karşılaşılabilecek olan problemlere dikkat çekilmiş ve çözüm önerisi olarak yeni bir yöntem önerilmiştir. Bu yöntem ile küçük bir model için yapılan çalışmalarda çok başarılı sonuçlar elde edilmiş yönlü olarak gerçek ağyapıya ulaşılmıştır. Önerilen yöntem, *E.coli* ve *S.cerevisiae* mikroorganizmaları için yapılan ağyapı çıkarımı çalışmalarında diğer yöntemlere nazaran bulunan ilişkilerin yönlerini doğru şekilde tahmin etmede başarılı sonuçlar göstermiştir. Ancak, küçük modelde elde edilen başarı yakalanamamıştır.

Yönlü olarak ayrıca metabolik ağyapı çıkarımında nedensellik yöntemlerinin başarısızlıkları ortaya konmuştur. Fakat bu yöntemlerden özellikle GES algoritma yönsüz olarak iyi sonuçlar vermektedir. Yönlü olarak bu yöntemlerin başarısızlığına, metabolik ağyapılarda yer alan tersinir reaksiyonların neden olduğu düşünülmektedir.

Yönlü yöntemlerin sonuçlarını iyileştirmek için yapılan ayrıntılı çalışmalar dikkate alındığında, bu tez çalışmasından çıkan en önemli sonuçlardan biri de, metabolom verilerine dayalı ağyapı çıkarımı çalışmalarında yönsüz ağyapı çıkarımı yöntemlerinin tercih edilmesinin daha iyi olduğu gerçeğidir.

8 ÖNERİLER

Yapılan tez çalışması kapsamında *E.coli* ve *S.cerevisiae* mikroorganizmaları ele alınmış ve bu mikroorganizmalara ağyapı çıkarımı yöntemleri uygulanarak hücrenin biyolojik amacı incelenmiştir. Bu doğrultuda *in silico* deneysel veriler yerine gerçek deneysel veriler kullanılarak yapılan çalışmalar tekrar edilebilir. Bu sayede yöntemlerin gerçek durumda başarıları tespit edilebilir. Ayrıca literatürde yer alan farklı mikroorganizmalar ile de bu çalışmalar tekrarlanabilir ve elde edilen sonuçlar karşılaştırılabilir. Böylelikle tez çalışmasında dikkat çekilen ağyapı çıkarımındaki problemler daha kapsamlı bir şekilde incelenebilir ve bu doğrultuda hücrenin biyolojik amacını en iyi yansıtacak ağyapı çıkarımı yöntemi belirlenebilir.

Daha da öte, literatürde yer alan farklı amaç fonksiyonları ağyapı çıkarımı yöntemlerinde modellenerek yöntemler tekrarlanabilir. Böylelikle literatürde önerilen hücrenin biyolojik amaçlarını ağyapı çıkarımı yöntemleri ile tekrar incelenme fırsatı bulunabilir.

KAYNAKLAR

1. Klipp, E., Herwig, R., Kowald, A., Wierling, C. Lehrach, H., (2005), *Systems Biology in Practice*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.
2. Bansal, M., Belcastro, V., Ambesi-Impimbato, A., di Bernardo, D., (2007), *How to infer gene networks from expression profiles*, *Molecular Systems Biology*, 3: 78.
3. Soranzo, N., Bianconi, G., Altafini, C., (2007), *Comparing association network algorithms for reverse engineering of large-scale gene regulatory networks: Synthetic versus real data*, *Bioinformatics*, 23: 1640–1647.
4. Nemenman, I., Escola, G.S., Hlavacek, W.S., Unkefer, P.J., Unkefer, C.J., Wall, M.E., (2007), *Reconstruction of metabolic networks from high-throughput metabolite profiling data: In silico analysis of red blood cell metabolism*, *Annals of the New York Academy of Sciences*, 1115: 102–115.
5. Çakır, T., Hendriks, M.M.W.B., Westerhuis, J.A., Smilde, A.K., (2009), *Metabolic network discovery through reverse engineering of metabolome data*, *Metabolomics*, 5: 318-329.
6. Demain, A. L., (1980), *Microbial production of primary metabolites*, *Naturwissenschaften*, 67:582-587.
7. Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., (2002), *Dynamic modeling of the central carbon metabolism of Escherichia coli*, *Biotechnology and Bioengineering*, 79: 53–73
8. Teusink B, Passarge J, Reijenga CA., Esgalhado E., van der Weijden CC., Schepper M., Walsh MC., Bakker BM., van Dam K., Westerhoff HV., Snoep JL., (2000), *Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry*, *European Journal of Biochemistry*, 267: 5313–5329.
9. Shinto H., Tashiro Y., Yamashita M., Kobayashi G., Sekiguchi T., Hanai T., Kuriya Y., Okamoto M, Sonomoto K., (2007), *Kinetic modeling and sensitivity analysis of acetone-butanol-ethanol production*, *Journal of Biotechnology*, 131: 45-56.
10. Kriete, A., Eils,R.,(2005), *Computational Systems Biology*, Elsevier Academic Press.

11. Kauffman, K. J., Prakash, P., Edwards, J.S., (2003), *Advances in flux balance analysis*, Current Opinion in Biotechnology, 14:491-496.
12. Markowitz, F., Spang, R., (2007), *Inferring cellular Networks-A review*, BMC Bioinformatics, 8:5.
13. Çakır, T., (2010), Hücrenin Metabolik Davranışlarının Ardında Yatan Biyolojik Amacın 'Aşağıdan-Yukarı' Ve 'Yukarıdan-Aşağı' Sistem Biyolojisi Yaklaşımlarıyla Bütünleşik İncelenmesi: İnsan Ve Mikroorganizma Metabolizmalarına Uygulamaları, *Tübitak 110M464 Proje Gelişme Raporu*.
14. Husmeier, D., (2003), *Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian Networks*, Bioinformatics, 19: 2271-2282.
15. Thomas, R., Mehrotra, S., Papoutsakis, E.T., Hatzimanikatis V., (2004), *A model-based optimization framework for the inference on gene regulatory networks from DNA array data*, Bioinformatics, 20: 3221-3235.
16. Wang, Y., Joshi, T., Zang, X-S., Xu, D., Chen, L., (2006), *Inferring gene regulatory networks from multiple microarray datasets*, Bioinformatics, 22: 2413-2420.
17. Dempster, A. P., (1972), *Covariance selection*, Biometrics, 28:157-175.
18. Ulucan, A., (2004), *Yöneylem Araştırması*, Siyasal Kitabevi, 1. Baskı, Ankara.
19. Patil, K.R., Rocha, I., Förster, J., Nielsen, J., (2005), *Evolutionary programming as a platform for in silico metabolic engineering*, BMC Bioinformatics, 6: 308.
20. Cheng R., Gen M.,Tosawa,T., (1995), *Genetic Algorithms for Desingning Loop Layout Manufacturing Systems*, Computer & Industrial Engineering, 31:587-591
21. Steuer, R., Kurths, J., Fiehn, O., Weckwerth, W., (2003), *Observing and interpreting correlations in metabolomic networks*. Bioinformatics Oxford, England, 19, 1019–1026.
22. Meek, C., (1995b), *Causal inference and causal explanation with background knowledge*, Uncertainty in Artificial Intelligence, 11: 403–410.
23. Spirtes, P., Glymour, C., Scheines, R., (2000), *Causation, Prediction, and Search*. The MIT Press, 2nd edition.

24. Richardson, T. (1996). *Discovering cyclic causal structure* , Pittsburgh, PA: Philosophy Department, Carnegie Mellon University.
25. Chickering, D. M., (2002), *Optimal structure identification with greedy search* (Technical Report MSR-TR-2002-10), Microsoft Research. Submitted to *JMLR*.
26. Picchini, U. (2007). *Sde toolbox: Simulation and estimation of stochastic differential equations with matlab*.: <http://sdetoolbox.sourceforge.net>.
27. Yeung, M.K., Tegner, J., Collins, J.J., (2002), *Reverse engineering gene networks using singular value decomposition and robust regression*, Proceedings of the National Academy of Sciences USA, 99: 6163-8.
28. Scheines, R., Spirtes, P., Glymour, C., Meek, C., and Richardson, T., (1998). *The TETRAD Project: Constraint Based Aids to Causal Model Specification*, Multivariate Behavioral Research., 33: 65-117.
29. Sokal, R.R. and Rohlf, F.J., (1995), *Biometry: the principles and Practice of Statistics in Biological Research*, Freeman, New York.
30. Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al., (2006)., *Aracne: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context*, BMC Bioinformatics, 7:7.
31. Klamt, S., Flassig, R. J., Sundmacher, K., (2010), *Inferring cellular networks with transitive reduction*, Bioinformatics, 26: 2160-2168.
32. McClean, M.N., Mody, A., Broach, J.R., Ramanathan, S., (2007), *Cross-talk and decision making in MAP kinase pathways*, Nature Genetics, 39:409-414
33. Metz, C.E., (1978), *Basic principles of ROC analysis*. Seminars in nuclear medicine, 8: 283-298.
34. Werhli, A.V., Grzegorzcyk, M., Husmeier, D., (2006), *Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian Networks*, Bioinformatics , 22: 2523-2531.
35. Illig, T., Gieger, C., Zhai, G., Römisch-Margl, W., Wang-Sattler, R., Prehn C., Altmaier, E., Kastenmüller, G., Kato B.S., Mewes, H., Meitinger, T., Hrabé de Angelis, M., Kronenberg, F., Soranzo, N., Wichmann, H., Spector, T., Adamski, J., Suhre, K., (2010), *A genome-wide perspective of genetic variation in human metabolism*, Nature Genetics, 42:137-141.

ÖZGEÇMİŞ

Adı Soyadı : Melik ÖKSÜZ
İletişim Bilgileri : melikoksuz@gmail.com
Doğum Tarihi : 01.01.1986
Mesleği : Kimya Mühendisi
Öğrenim Durumu :Yüksek Lisans Mezunu

Derece	Alan	Üniversite	Yıl
Lisans	Kimya Mühendisliği	Gazi Üniversitesi Çorum Mühendislik Fakültesi	2004-2008
Yüksek Lisans	Kimya Mühendisliği	Gebze Yüksek Teknoloji Enstitüsü	2010-2012

EK 1: *S.cerevisiae* MİKROORGANİZMASINA AİT DİNAMİK MODEL

Saccharomyces cerevisiae mikroorganizmasına ait Teusink ve ark. tarafından geliştirilen 13 metabolit ve 18 reaksiyondan oluşan glikoliz yoluna ait dinamik modelin zamana bağlı konsantrasyon değişimlerini içeren diferansiyel eşitlikleri aşağıdadır [Teusink et al., 2000].

$$\begin{aligned}
 d[\text{Glc}_{\text{in}}]/dt &= v_{\text{transport}} - v_{\text{HK}} \\
 d[\text{G6P}]/dt &= v_{\text{HK}} - v_{\text{PGI}} - 2v_{\text{trehalose}} - v_{\text{glycogen}} \\
 d[\text{F6P}]/dt &= v_{\text{PGI}} - v_{\text{PFK}} \\
 d[\text{F1, 6bP}_2]/dt &= v_{\text{PFK}} - v_{\text{ALD}} \\
 d[\text{Trio - P}]/dt &= 2v_{\text{ALD}} - v_{\text{GraPDH}} (-v_{\text{glycerol}}) \\
 d[\text{BPG}]/dt &= v_{\text{GraPDH}} - v_{\text{PGK}} \\
 d[3\text{GriP}]/dt &= v_{\text{PGK}} - v_{\text{PGM}} \\
 d[2\text{GriP}]/dt &= v_{\text{PGM}} - v_{\text{ENO}} \\
 d[\text{phosphoenolpyruvate}]/dt &= v_{\text{ENO}} - v_{\text{PYK}} \\
 d[\text{PYR}]/dt &= v_{\text{PYK}} - v_{\text{PDC}} \\
 d[\text{AcAld}]/dt &= v_{\text{PDC}} - v_{\text{ADH}} (-2v_{\text{succinate}}) \\
 d\text{P}/dt &= -v_{\text{HK}} - v_{\text{PFK}} + v_{\text{PGK}} + v_{\text{PYK}} - v_{\text{ATPase}} \\
 &\quad (-v_{\text{trehalose}} - v_{\text{glycogen}} - 4v_{\text{succinate}}) \\
 d[\text{NADH}]/dt &= v_{\text{GraPDH}} - v_{\text{ADH}} (-v_{\text{glycerol}} + 3v_{\text{succinate}}) \\
 d[\text{NAD}]/dt &= -d[\text{NADH}]/dt
 \end{aligned}$$

EK 2: *E.coli* MİKROORGANİZMASINA AİT DİNAMİK MODEL

Escherichia coli mikroorganizmasına ait Chassagnole ve ark. tarafından geliştirilen 18 metabolit ve 30 reaksiyondan oluşan merkezi karbon metabolizmasına ait dinamik modelin zamana bağlı konsantrasyon değişimlerini içeren diferansiyel eşitlikleri aşağıdadır [Chassagnole et al., 2002].

$$\begin{aligned} \frac{dC_{\text{glc}}^{\text{extracellular}}}{dt} &= D(C_{\text{glc}}^{\text{feed}} - C_{\text{glc}}^{\text{extracellular}}) + f_{\text{pulse}} - \frac{C_x r_{\text{PTS}}}{\rho_x} \\ \frac{dC_{\text{g6p}}}{dt} &= r_{\text{PTS}} - r_{\text{PGI}} - r_{\text{G6PDH}} - r_{\text{PGM}} - \mu C_{\text{g6p}} \\ \frac{dC_{\text{f6p}}}{dt} &= r_{\text{PGI}} - r_{\text{PFK}} + r_{\text{TKb}} + r_{\text{TA}} - 2r_{\text{MurSynth}} - \mu C_{\text{f6p}} \\ \frac{dC_{\text{fdp}}}{dt} &= r_{\text{PFK}} - r_{\text{ALDO}} - \mu C_{\text{fdp}} \\ \frac{dC_{\text{gap}}}{dt} &= r_{\text{ALDO}} + r_{\text{TIS}} - r_{\text{GAPDH}} + r_{\text{TKa}} + r_{\text{TKb}} - r_{\text{TA}} + r_{\text{TrpSynth}} - \mu C_{\text{gap}} \\ \frac{dC_{\text{dhap}}}{dt} &= r_{\text{ALDO}} - r_{\text{TIS}} - r_{\text{G3PDH}} - \mu C_{\text{dhap}} \\ \frac{dC_{\text{pgp}}}{dt} &= r_{\text{GAPDH}} - r_{\text{PGK}} - \mu C_{\text{pgp}} \\ \frac{dC_{\text{3pg}}}{dt} &= r_{\text{PGK}} - r_{\text{PGluMu}} - r_{\text{SerSynth}} - \mu C_{\text{3pg}} \\ \frac{dC_{\text{2pg}}}{dt} &= r_{\text{PGluMu}} - r_{\text{ENO}} - \mu C_{\text{2pg}} \\ \frac{dC_{\text{pep}}}{dt} &= r_{\text{ENO}} - r_{\text{PK}} - r_{\text{PTS}} - r_{\text{PEPCxylase}} - r_{\text{DAHPS}} - r_{\text{Synth1}}^a - \mu C_{\text{pep}} \\ \frac{dC_{\text{pyr}}}{dt} &= r_{\text{PK}} + r_{\text{PTS}} - r_{\text{PDH}} - r_{\text{Synth2}}^b + r_{\text{MetSynth}} + r_{\text{TrpSynth}} - \mu C_{\text{pyr}} \\ \frac{dC_{\text{6pg}}}{dt} &= r_{\text{G6PDH}} - r_{\text{PGDH}} - \mu C_{\text{6pg}} \\ \frac{dC_{\text{ribu5p}}}{dt} &= r_{\text{PGDH}} - r_{\text{Ru5P}} - r_{\text{R5PI}} - \mu C_{\text{ribu5p}} \\ \frac{dC_{\text{xyl5p}}}{dt} &= r_{\text{Ru5P}} - r_{\text{TKa}} - r_{\text{TKb}} - \mu C_{\text{xyl5p}} \\ \frac{dC_{\text{sed7p}}}{dt} &= r_{\text{TKa}} - r_{\text{TA}} - \mu C_{\text{sed7p}} \\ \frac{dC_{\text{rib5p}}}{dt} &= r_{\text{R5PI}} - r_{\text{TKa}} - r_{\text{RPPK}} - \mu C_{\text{rib5p}} \\ \frac{dC_{\text{e4p}}}{dt} &= r_{\text{TA}} - r_{\text{TKb}} - r_{\text{DAHPS}} - \mu C_{\text{e4p}} \\ \frac{dC_{\text{glp}}}{dt} &= r_{\text{PGM}} - r_{\text{GIPAT}} - \mu C_{\text{glp}} \end{aligned}$$

EK 3: RASTGELE SIRALAMA TEST FONKSİYONU

Bölüm 3.2.4'te ayrıntılı olarak anlatılan rastgele sıralama testinin MATLAB dosyası formatındaki matematiksel ifadesi aşağıda verilmiştir.

```
function p_value=randimization_test(R,Cveri);
%R: GGM korelasyon matrisi
%Cveri= GGM korelasyon matrisinin hesaplandığı veri matrisi
% Değişken sayının
de=length(R);
ag=1000;
%p-value hesaplanması için yukarıda elde edilen verilerin
%sütunlardaki değerlerin yerlerinin 1000 kez değiştirilerek 1000 tane
%farklı veri matrisin elde edilmesi
for m=1:ag;
    for j=1:ms;
        K(:,j,m)=Cveri(randperm(ag),j);
    end
end
%1000 tane veri matrisinin herbirinin GGM korelasyon katsayılarının hesaplanması
for n=1:ag;
    [Rhesap(:,n),zz(:,n)]=corr(K(:,n));
    %n-2. dereceden korelasyon matrisinin ggm metodu ile hesaplanması
    A(:,n)=inv(Rhesap(:,n));
    g(:,n)=diag(A(:,n));
    Rggm(:,n)=(-1*A(:,n))./(sqrt(g(:,n))*(g(:,n))));
end
%Bulunan korelasyon katsayılarının karşılaştırılması sonucu pvalue nin
%hesaplanması
for a=1:ms;
    for b=1:ms;
        %Elde edilen 1000 farklı GGM korelasyon katsayı matrisinin
        %gerçek GGM korelasyon katsayı matrisi ile
        %karşılaştırılarak korelasyon katsayısının gerçek duruma
        %göre daha yüksek olan GGM korelasyon katsayısının
        %tespit edilmesi
        h(a,b)=length(find(abs(R(a,b,ag))<=abs(Rggm(a,b,:))));
    end
end
%p-value değerleri
p_value(:,ag)=h/ag;
end
```

EK 4: M.CLEAN MODELİ İÇİN ÖNERİLEN YÖNTEM İLE YAPILAN HESAPLAMALAR

McClean ve ark. tarafından geliştirilen *S.cerevisiae*'ye ait 6 protein ve 10 reaksiyondan oluşan mitojen aktive eden protein (MAP) kinaz yolağı kullanılarak Bölüm 3.3.4'te ayrıntılı olarak anlatılan önerilen yöntem ile iki farklı uygunluk değeri için yapılan hesaplamalara ait sonuçlar aşağıda verilmiştir.

Tablo E4-1 *M.Clean* modeli için uygunluk değeri kriteri olarak hata değerinin kullanıldığı genetik algoritma tabanlı önerilen yöntemin uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 10 kromozoma ait sonuçlar.

	Uygunluk Değeri	Sıfır Sayısı	Yönlü		Yönsüz	
			TPR	FPR	TPR	FPR
1	$2,72 \times 10^{-32}$	20	1	0	1	0
2	$3,40 \times 10^{-32}$	20	1	0	1	0
3	$3,66 \times 10^{-32}$	20	1	0	1	0
4	$4,22 \times 10^{-32}$	20	1	0	1	0
5	$4,80 \times 10^{-32}$	20	1	0	1	0
6	$5,36 \times 10^{-32}$	20	1	0	1	0
7	$5,62 \times 10^{-32}$	20	1	0	1	0
8	$5,65 \times 10^{-32}$	20	1	0	1	0
9	$5,99 \times 10^{-32}$	20	1	0	1	0
10	$6,06 \times 10^{-32}$	20	1	0	1	0

Tablo E4-2 *M.Clean* modeli için uygunluk değeri kriteri olarak sıfır sayısı ve hata değerinden oluşan ortak fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 10 kromozoma ait sonuçlar.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	70,14	$7,22 \times 10^{-31}$	20	1	0	1	0
2	63,31	$4,88 \times 10^{-4}$	30	0	0	0	0
3	61,41	$3,91 \times 10^{-4}$	29	0,1	0	0,2	0
4	61,40	$3,95 \times 10^{-4}$	29	0,1	0	0,2	0
5	61,40	$3,99 \times 10^{-4}$	29	0,1	0	0,2	0
6	61,39	$4,09 \times 10^{-4}$	29	0	0,05	0	0,1
7	61,39	$4,10 \times 10^{-4}$	29	0	0,05	0	0,1
8	61,38	$4,18 \times 10^{-4}$	29	0	0,05	0	0,1
9	61,38	$4,20 \times 10^{-4}$	29	0	0,05	0	0,1
10	61,37	$4,23 \times 10^{-4}$	29	0,1	0	0,2	0

Tablo E4-1'e bakıldığında ilk 10 kromozom (aday çözüm vektörü) ile yönlü ve yönsüz olarak gerçek ağyapıya ulaşılmıştır. Tablo E4-2'de ise birinci kromozom ile aynı şekilde yönlü ve yönsüz olarak gerçek ağyapıya ulaşılmıştır.

EK 5: *S.cerevisiae* MİKROORGANİZMASI İÇİN ÖNERİLEN YÖNTEM İLE YAPILAN HESAPLAMALAR

Bölüm 4.2.5'te ayrıntılı olarak anlatılan önerilen yöntem ile *S.cerevisiae* mikroorganizması için üç farklı uygunluk değeri için yapılan hesaplamalara ait sonuçlar aşağıda verilmiştir.

Tablo E5-1. *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak hata değerinin kullanıldığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 25 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR.

	Uygunluk Değeri	Sıfır Sayısı	Yönlü		Yönsüz	
			TPR	FPR	TPR	FPR
1	$1,29 \times 10^{-13}$	80	0,19	0,67	0,81	0,67
2	$9,96 \times 10^{-9}$	81	0,19	0,67	0,81	0,67
3	$4,06 \times 10^{-6}$	85	0,10	0,61	0,90	0,61
4	$5,58 \times 10^{-6}$	83	0,38	0,65	0,81	0,65
5	$1,94 \times 10^{-5}$	84	0,14	0,65	0,86	0,65
6	$4,23 \times 10^{-5}$	83	0,19	0,74	0,71	0,74
7	$5,35 \times 10^{-5}$	85	0,14	0,63	0,81	0,63
8	$5,49 \times 10^{-5}$	84	0,14	0,70	0,76	0,70
9	$6,12 \times 10^{-5}$	84	0,19	0,68	0,76	0,68
10	$6,45 \times 10^{-5}$	85	0,19	0,65	0,71	0,65
11	$7,23 \times 10^{-5}$	89	0,19	0,65	0,62	0,65
12	$7,87 \times 10^{-5}$	85	0,14	0,63	0,76	0,63
13	$8,07 \times 10^{-5}$	82	0,10	0,63	0,71	0,63
14	$9,07 \times 10^{-5}$	82	0,10	0,68	0,71	0,68
15	$1,01 \times 10^{-4}$	85	0,19	0,70	0,67	0,70
16	$1,08 \times 10^{-4}$	82	0,24	0,68	0,71	0,68
17	$1,20 \times 10^{-4}$	82	0,24	0,67	0,67	0,67
18	$1,24 \times 10^{-4}$	83	0,24	0,67	0,76	0,67
19	$1,28 \times 10^{-4}$	88	0,14	0,63	0,81	0,63
20	$1,33 \times 10^{-4}$	88	0,19	0,67	0,71	0,67
21	$1,44 \times 10^{-4}$	87	0,19	0,70	0,67	0,70
22	$1,47 \times 10^{-4}$	84	0,19	0,77	0,71	0,77
23	$1,51 \times 10^{-4}$	94	0,10	0,65	0,52	0,65
24	$1,56 \times 10^{-4}$	93	0,24	0,65	0,71	0,65
25	$1,60 \times 10^{-4}$	91	0,24	0,60	0,81	0,60

Tablo E5-2. *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak sıfır sayısı ve hata değerinden oluşan ortak fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin in silico metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 25 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	12,73	$1,85 \times 10^{-11}$	80	0,29	0,79	0,71	0,79
2	12,08	$8,27 \times 10^{-11}$	80	0,24	0,81	0,71	0,81
3	11,83	$1,47 \times 10^{-10}$	80	0,24	0,79	0,76	0,79
4	11,37	$4,27 \times 10^{-10}$	80	0,24	0,77	0,76	0,77
5	11,34	$4,61 \times 10^{-10}$	80	0,29	0,77	0,67	0,77
6	11,20	$6,35 \times 10^{-10}$	80	0,24	0,79	0,71	0,79
7	11,11	$7,81 \times 10^{-10}$	80	0,29	0,79	0,67	0,79
8	10,96	$1,09 \times 10^{-10}$	80	0,29	0,77	0,71	0,77
9	10,84	$1,43 \times 10^{-10}$	80	0,24	0,77	0,67	0,77
10	10,84	$1,45 \times 10^{-10}$	80	0,29	0,79	0,71	0,79
11	10,76	$1,74 \times 10^{-9}$	80	0,29	0,79	0,71	0,79
12	10,74	$1,83 \times 10^{-9}$	80	0,24	0,79	0,67	0,79
13	10,72	$1,92 \times 10^{-9}$	80	0,24	0,79	0,71	0,79
14	10,71	$1,94 \times 10^{-9}$	80	0,29	0,77	0,71	0,77
15	10,70	$2,01 \times 10^{-9}$	80	0,24	0,77	0,71	0,77
16	10,69	$2,07 \times 10^{-9}$	80	0,24	0,79	0,71	0,79
17	10,68	$2,10 \times 10^{-9}$	80	0,24	0,81	0,67	0,81
18	10,67	$2,15 \times 10^{-9}$	80	0,24	0,75	0,71	0,75
19	10,66	$2,32 \times 10^{-9}$	81	0,24	0,79	0,71	0,79
20	10,66	$2,19 \times 10^{-9}$	80	0,24	0,79	0,71	0,79
21	10,65	$2,24 \times 10^{-9}$	80	0,24	0,77	0,76	0,77
22	10,64	$2,29 \times 10^{-9}$	80	0,24	0,79	0,71	0,79
23	10,63	$2,33 \times 10^{-9}$	80	0,24	0,77	0,71	0,77
24	10,62	$2,39 \times 10^{-9}$	80	0,24	0,77	0,71	0,77
25	10,62	$2,42 \times 10^{-9}$	80	0,24	0,79	0,71	0,79

Tablo E5-3 *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak hem sıfır sayısı hem de hata değerinden oluşan ortak fonksiyonun kullanıldığı ve ek olarak GGM sayılarının dikkate alındığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	15,29	$5,08 \times 10^{-14}$	80	0,62	0,58	0,90	0,58
2	14,71	$1,94 \times 10^{-13}$	80	0,67	0,54	0,90	0,54
3	14,54	$2,91 \times 10^{-13}$	80	0,62	0,60	0,90	0,60
4	14,33	$4,66 \times 10^{-13}$	80	0,67	0,56	0,90	0,56
5	14,22	$6,08 \times 10^{-13}$	80	0,62	0,60	0,90	0,60
6	14,16	$6,98 \times 10^{-13}$	80	0,62	0,58	0,90	0,58
7	14,15	$7,11 \times 10^{-13}$	80	0,67	0,56	0,90	0,56
8	14,07	$8,42 \times 10^{-12}$	80	0,62	0,58	0,90	0,58
9	13,94	$1,15 \times 10^{-12}$	80	0,62	0,56	0,90	0,56
10	13,90	$1,26 \times 10^{-12}$	80	0,67	0,56	0,90	0,56
11	13,88	$1,32 \times 10^{-12}$	80	0,62	0,58	0,90	0,58
12	13,86	$1,37 \times 10^{-12}$	80	0,67	0,56	0,90	0,56
13	13,84	$1,44 \times 10^{-12}$	80	0,62	0,58	0,90	0,58
14	13,76	$1,75 \times 10^{-12}$	80	0,67	0,58	0,90	0,58
15	13,73	$1,85 \times 10^{-12}$	80	0,62	0,60	0,90	0,60
16	13,72	$1,92 \times 10^{-12}$	80	0,62	0,58	0,90	0,58
17	13,69	$2,03 \times 10^{-12}$	80	0,62	0,60	0,90	0,60
18	13,66	$2,17 \times 10^{-12}$	80	0,67	0,56	0,90	0,56
19	13,62	$2,43 \times 10^{-12}$	80	0,62	0,60	0,90	0,60
20	13,56	$2,76 \times 10^{-12}$	80	0,62	0,58	0,90	0,58
21	13,54	$2,90 \times 10^{-12}$	80	0,67	0,58	0,90	0,58
22	13,53	$2,98 \times 10^{-12}$	80	0,67	0,58	0,90	0,58
23	13,51	$3,11 \times 10^{-12}$	80	0,62	0,58	0,95	0,58
24	13,49	$3,21 \times 10^{-12}$	80	0,62	0,60	0,90	0,60
25	13,48	$3,32 \times 10^{-12}$	80	0,62	0,58	0,90	0,58

Tablo E5-4 *S.cerevisiae* mikroorganizması için uygunluk değeri kriteri olarak hem GGM sonucunu hem de hata değerini içeren bir ortak fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR’ler

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	16,44	$1,01 \times 10^{-13}$	80	0,38	0,63	1	0,63
2	15,58	$1,06 \times 10^{-12}$	80	0,38	0,58	1	0,58
3	15,50	$9,91 \times 10^{-13}$	80	0,29	0,65	1	0,65
4	15,49	$1,29 \times 10^{-12}$	80	0,38	0,60	1	0,60
5	15,45	$9,93 \times 10^{-13}$	80	0,33	0,65	1	0,65
6	15,43	$1,04 \times 10^{-12}$	80	0,38	0,61	1	0,61
7	15,38	$1,32 \times 10^{-12}$	80	0,38	0,63	1	0,63
8	15,32	$1,50 \times 10^{-12}$	80	0,33	0,63	1	0,63
9	15,32	$1,52 \times 10^{-12}$	80	0,38	0,63	1	0,63
10	15,32	$1,53 \times 10^{-12}$	80	0,38	0,60	1	0,60
11	15,30	$1,99 \times 10^{-12}$	80	0,33	0,61	1	0,61
12	15,28	$1,84 \times 10^{-12}$	80	0,38	0,63	1	0,63
13	15,28	$1,68 \times 10^{-12}$	80	0,38	0,60	1	0,60
14	15,27	$2,12 \times 10^{-12}$	80	0,38	0,61	1	0,61
15	15,26	$1,94 \times 10^{-12}$	80	0,38	0,61	1	0,61
16	15,26	$1,95 \times 10^{-12}$	80	0,38	0,60	1	0,60
17	15,26	$1,96 \times 10^{-12}$	80	0,33	0,63	1	0,63
18	15,24	$1,80 \times 10^{-12}$	80	0,38	0,61	1	0,61
19	15,23	$2,35 \times 10^{-12}$	80	0,38	0,61	1	0,61
20	15,23	$2,10 \times 10^{-12}$	80	0,38	0,60	1	0,60
21	15,22	$2,12 \times 10^{-12}$	80	0,38	0,63	1	0,63
22	15,22	$1,71 \times 10^{-12}$	80	0,38	0,63	1	0,63
23	15,21	$2,46 \times 10^{-12}$	80	0,33	0,61	1	0,61
24	15,20	201×10^{-12}	80	0,38	0,63	1	0,63
25	15,19	257×10^{-12}	81	0,38	0,61	1	0,61

EK 6: *E.coli* MİKROORGANİZMASI İÇİN ÖNERİLEN YÖNTEM İLE YAPILAN HESAPLAMALAR

Bölüm 5.2.4'te ayrıntılı olarak anlatılan önerilen yöntem ile *S.cerevisiae* mikroorganizması için iki farklı uygunluk değeri için yapılan hesaplamalara ait sonuçlar aşağıda verilmiştir.

Tablo E6-1 *E.coli* mikroorganizması için uygunluk değeri kriteri olarak hem sıfır sayısı hem de hata değerinden oluşan ortak fonksiyonun kullanıldığı ve GGM sayılarının dikkate alındığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	10,0704	$1,20 \times 10^{-3}$	286	0,15	0,12	0,28	0,12
2	10,0703	$1,20 \times 10^{-3}$	286	0,15	0,11	0,33	0,11
3	10,0703	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
4	10,0703	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
5	10,0703	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
6	10,0703	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
7	10,0702	$1,20 \times 10^{-3}$	286	0,15	0,12	0,28	0,12
8	10,0702	$1,20 \times 10^{-3}$	286	0,15	0,11	0,33	0,11
9	10,0702	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
10	10,0701	$1,20 \times 10^{-3}$	286	0,15	0,11	0,33	0,11
11	10,0701	$1,20 \times 10^{-3}$	286	0,15	0,11	0,28	0,11
12	10,0701	$1,20 \times 10^{-3}$	286	0,15	0,11	0,33	0,11
13	10,0701	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
14	10,0700	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
15	10,0700	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
16	10,0700	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
17	10,0699	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
18	10,0699	$1,20 \times 10^{-3}$	286	0,15	0,11	0,28	0,11
19	10,0699	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
20	10,0698	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
21	10,0698	$1,20 \times 10^{-3}$	286	0,15	0,11	0,33	0,11
22	10,0698	$1,20 \times 10^{-3}$	286	0,15	0,11	0,28	0,11
23	10,0697	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
24	10,0697	$1,20 \times 10^{-3}$	286	0,15	0,11	0,31	0,11
25	10,0697	$1,20 \times 10^{-3}$	286	0,15	0,11	0,33	0,11

Tablo E6-2 *E.coli* mikroorganizması için uygunluk değeri kriteri olarak hem GGM sonucunu hem de hata değerini içeren ortak bir fonksiyonun kullanıldığı genetik algoritma tabanlı önerilen yöntemin *in silico* metabolom verisine uygulanması ile elde edilen en iyi uygunluk değerine sahip ilk 5 kromozoma ait sonuçların değerlendirilmesi: Uygunluk Değerleri, TPR ve FPR.

	Uygunluk Değeri	Hata Payı	Sıfır Sayısı	Yönlü		Yönsüz	
				TPR	FPR	TPR	FPR
1	13,71	$1,87 \times 10^{-4}$	158	0,36	0,68	0,85	0,68
2	13,54	$3,99 \times 10^{-6}$	158	0,33	0,68	0,82	0,68
3	13,43	$1,63 \times 10^{-4}$	158	0,31	0,70	0,82	0,70
4	13,39	$6,90 \times 10^{-6}$	159	0,33	0,68	0,85	0,68
5	13,35	$1,32 \times 10^{-6}$	158	0,31	0,68	0,87	0,68
6	13,30	$4,60 \times 10^{-6}$	159	0,28	0,69	0,85	0,69
7	13,29	$2,54 \times 10^{-6}$	158	0,31	0,70	0,82	0,70
8	13,28	$4,83 \times 10^{-6}$	158	0,36	0,68	0,82	0,68
9	13,26	$1,15 \times 10^{-6}$	158	0,33	0,69	0,79	0,69
10	13,19	$5,09 \times 10^{-4}$	158	0,33	0,68	0,85	0,68
11	13,17	$2,31 \times 10^{-6}$	159	0,36	0,68	0,85	0,68
12	13,12	$3,74 \times 10^{-6}$	158	0,31	0,69	0,85	0,69
13	13,12	$1,26 \times 10^{-6}$	159	0,33	0,68	0,85	0,68
14	13,12	$1,40 \times 10^{-6}$	159	0,33	0,68	0,85	0,68
15	13,11	$2,85 \times 10^{-6}$	158	0,31	0,69	0,82	0,69
16	13,09	$3,36 \times 10^{-6}$	158	0,28	0,70	0,82	0,70
17	13,08	$3,66 \times 10^{-6}$	158	0,28	0,69	0,87	0,69
18	13,07	$1,50 \times 10^{-6}$	159	0,31	0,68	0,85	0,68
19	13,05	$5,39 \times 10^{-6}$	158	0,28	0,69	0,85	0,69
20	13,03	$5,79 \times 10^{-5}$	158	0,28	0,68	0,87	0,68
21	13,03	$1,34 \times 10^{-6}$	159	0,33	0,69	0,79	0,69
22	13,02	$2,78 \times 10^{-7}$	158	0,28	0,69	0,82	0,69
23	13,01	$4,07 \times 10^{-5}$	158	0,31	0,70	0,82	0,70
24	13,01	$2,57 \times 10^{-6}$	159	0,31	0,69	0,85	0,69
25	13,00	$1,39 \times 10^{-6}$	159	0,31	0,68	0,85	0,68