İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

GENETİK AĞLARIN TOPOLOJİK ÖZELLİKLERİ

YÜKSEK LİSANS TEZİ

Fizik Müh. Yasemin ŞENGÜN

Anabilim Dalı: FİZİK
Programı: FİZİK MÜHENDİSLİĞİ

HAZİRAN 2005

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

GENETİK AĞLARIN TOPOLOJİK ÖZELLİKLERİ

YÜKSEK LİSANS TEZİ
Fizik Müh. Yasemin ŞENGÜN
(509031113)

Tezin Enstitüye Verildiği Tarih : 9 Mayıs 2005
Tezin Savunulduğu Tarih : 2 Haziran 2005

Tez Danışmanı : **Prof.Dr. Ayşe ERZAN**
Diğer Jüri Üyeleri **Prof.Dr. Ayşe Hümeyra BİLGE (İ.T.Ü.)**
Doç.Dr. Mehmet Nazmi POSTACIOĞLU (İ.T.Ü.)

HAZİRAN 2005

ÖNSÖZ

Bütün çalışmalarım boyunca iyi niyeti ve emeği ile birlikte yanımda olduğu için Ayşe ERZAN'a teşekkür ediyorum.
Ayşe Hümeyra BİLGE ve Mehmet Nazmi POSTACIOĞLU'na tezin kontrolü aşamasında gösterdikleri hassasiyet ve yapıcı eleştirileri ile yaptıkları katkılardan dolayı teşekkür ediyorum.
Fizik bölümünden Duygu BALCAN ve Ali GELİR'e, moleküler biyoloji ve genetik bölümünden Zeynep Neva KONTAK'a hem güzel arkadaşlıkları için hem de bana her konuda yardımlarından dolayı minnettarım.

Haziran 2005 Yasemin Şengün

İÇİNDEKİLER

TABLO LİSTESİ **v**
ŞEKİL LİSTESİ **vi**
SEMBOL LİSTESİ **vii**
ÖZET **viii**
SUMMARY **ix**

1. GİRİŞ **1**
1.1. Ağlar (Networks) 1
1.2. Genetik Ağlar 3

2. MODEL **6**
2.1. İçerik Esaslı Ağ Modeli 6
2.1.1. Modelin Tanımı 6
2.1.2. Modelin Sonuçları 10
2.2. İçerik Esaslı Ağ Modelinin Zamanla Evrimleşme Mekanizmaları 11
2.2.1. Duplikasyon 12
2.2.2. Çeşitli Mutasyonlar 15
2.2.3. Başlangıç koşulları 16
2.2.4. Model tanımlamaları 17

3. SİMÜLASYONLAR **18**
3.1. Dizilimin Zamanla Evrimi 19
3.1.1. Dizilim Uzunluğu (L) 19
3.1.2. Alt Dizi Sayısı (N) 21
3.1.3. Ortalama Alt Dizi Uzunluğu ($\langle l \rangle$) 23
3.1.4. Dizi Uzunluklarının Dağılımı ($u(l)$) 25
3.2. Derece Dağılımları 29
3.2.1. Giren Derece Dağılımı 29
3.2.2. Çıkan Derece Dağılımı 31
3.2.3. Toplam Derece Dağılımı 33
3.3. Çıkan Derece Dağılımının Ölçeklenme Davranışı 35
3.4. Kümelenme Katsayısının Dağılımı ($C(d)$) ve Derece Dağılımlarına l Katkıları 41

4. SONUÇLAR VE TARTIŞMA **45**

KAYNAKLAR **46**

EKLER **50**

ÖZGEÇMİŞ **51**

TABLO LİSTESİ

		Sayfa No
Tablo 2.1.1.	Örnek dizilim için alt dizi gösterimi ve uzunlukları……..	7
Tablo 2.2.4.1.	Modellerin kısa gösterimi……....	17
Tablo 3.1.	Simülasyonda kullanılan sayısal değerler…….....	18
Tablo 3.1.1.1.	Dizilim uzunluğu *L*'nin, zamanla değişim hızlarının değerleri..	21
Tablo 3.1.2.1.	*N*'nin, zamanla değişim hızlarının değerleri.....	22
Tablo 3.1.3.1.	<*l*>'nin M1'li alt modellerde değişim hızları.....	24
Tablo 3.3.1.	Model I için bulunan üsteller	40
Tablo 3.3.2.	Model II için bulunan üsteller	40

ŞEKİL LİSTESİ

Sayfa No

Şekil 1.1.1 : $C(d)$'nin tanımı için örnek şekil 2
Şekil 2.1.1 : $P(x)$ ile oluşturulan örnek dizilim 6
Şekil 2.1.2 : Örnek dizilimdeki alt dizilerin oluşturduğu ağ 7
Şekil 2.2.1.1 : Duplikasyon için 1.a durumunda örnek ağın görünümü 13
Şekil 2.2.1.2 : Duplikasyon için 1.b durumunda örnek ağın görünümü 13
Şekil 2.2.1.3 : Duplikasyon için 2.a durumunda örnek ağın görünümü 13
Şekil 2.2.1.4 : Duplikasyon için 2.b durumunda örnek ağın görünümü 14
Şekil 2.2.1.5 : Duplikasyon için 3.a durumunda örnek ağın görünümü 14
Şekil 2.2.3.1 : $U_g(l)$'nin grafiği 16
Şekil 3.1.1.1 : L'nin M1 altında zamanla değişimi 19
Şekil 3.1.1.2 : L'nin M2 altında zamanla değişimi 20
Şekil 3.1.2.1 : N'nin zamanla değişimi 22
Şekil 3.1.3.1 : M1 kuralı ile $\langle l \rangle$ 'nin zamanla değişimi 23
Şekil 3.1.3.2 : M2 kuralı ile $\langle l \rangle$ 'nin zamanla değişimi 24
Şekil 3.1.3.3 : M2 kuralı ile $\ln(\langle l \rangle)$ 'nin grafiği 25
Şekil 3.1.4.1 : Başlangıç alt dizi uzunluk dağılımları 26
Şekil 3.1.4.2 : Model I için için alt dizi uzunluk dağılımlarının devinimi 27
Şekil 3.1.4.3 : Model I için için alt dizi uzunluk dağılımlarının devinimi..... 27
Şekil 3.1.4.4 : Model I için dizi uzunluk dağılımlarının incelenmesi 28
Şekil 3.1.4.5 : Model II için $\ln(u(l))$'nin grafiği 29
Şekil 3.2.1.1 : t=0'da $U_e(l)$ ve $U_g(l)$ için $n_{giren}(d)$grafiği 30
Şekil 3.2.1.2 : Model I için d-$n_{giren}(d)$ grafikleri 30
Şekil 3.2.1.3 : Model II için d-$n_{giren}(d)$ grafikleri 31
Şekil 3.2.2.1 : t=0'da $U_e(l)$ ve $U_g(l)$ için $n_{çıkan}(d)$ grafiği..... 32
Şekil 3.2.2.2 : Model I için için $n_{çıkan}(d)$ grafikleri 32
Şekil 3.2.2.3 : Model II için $n_{çıkan}(d)$ grafikleri 33
Şekil 3.2.3.1 : t=0'da $U_e(l)$ ve $U_g(l)$ için $n(d)$grafiği 34
Şekil 3.2.3.2 : Model I için d-$n(d)$ grafikleri 34
Şekil 3.2.3.3 : Model II için d-$n(d)$ grafikleri 35
Şekil 3.3.1 : Başlangıç dağılımı için log-log ölçeğinde $n_{çıkan}(d)$ grafiği 36
Şekil 3.3.2 : PM1 ve PDM1 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği 36
Şekil 3.3.3 : UM1 ve UDM1 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği 37
Şekil 3.3.4 : PM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği 38
Şekil 3.3.5 : PDM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği 38
Şekil 3.3.6 : UM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği 39
Şekil 3.3.7 : UDM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği 39
Şekil 3.4.1 : t=400'de PM1 ve UM1 için C(d) grafiği 42
Şekil 3.4.2 : t=400'de PM1 ve UM1 için C(d) grafiği 42
Şekil 3.4.3 : t=400'de PM1 ve UM1 için l katkıları 43
Şekil 3.4.4 : t=400'de PM1 ve UM1 için $n_{çıkan}(d)$ grafiği 44

SEMBOL LİSTESİ

$\boldsymbol{l}$	**:** Alt dizi uzunluğu
$\boldsymbol{u(l)}$	**:** Alt dizi uzunluğu dağılımı
$\mathbf{W}$	**:** Etkileşim matrisi
$\boldsymbol{d_{giren}}$	**:** Giren derece sayısı
$\boldsymbol{d_{çıkan}}$	**:** Çıkan derece sayısı
$\boldsymbol{d}$	**:** Toplam derece sayısı
$\boldsymbol{n_{giren}(d)}$	**:** Giren derece sayısı dağılımı
$\boldsymbol{n_{çıkan}(d)}$	**:** Çıkan derece sayısı dağılımı
$\boldsymbol{n(d)}$	**:** Toplam derece sayısı dağılımı
$\boldsymbol{p}$	**:** Ayraçların seçilme olasılığı
$\boldsymbol{L}$	**:** Dizilim uzunluğu
$\boldsymbol{N}$	**:** Alt dizi sayısı
$\boldsymbol{\mu}$	**:** Mutasyon olasılığı
$\boldsymbol{P(x)}$	**:** {0,1,2} setindeki elemanların seçilme olasılık fonksiyonu
$\boldsymbol{U_e(l)}$	**:** $P(x)$ ile oluşturulan dizilimin uzunluk dağılım fonksiyonu
$\boldsymbol{U_g(l)}$	**:** Gaussien uzunluk dağılım fonksiyonu

GENETİK AĞLARIN TOPOLOJİK ÖZELLİKLERİ

ÖZET

Gen etkileşimlerini modelleyebilecek, ratsgele oluşturulan, içerik esaslı ağ yapısının derece dağılımları daha önce [D. Balcan, A. Erzan, *Eur. Phys. J. B*, 38, 253-260 (2004)] incelenmişti. Simülasyon ve analitik [M. Mungan, A. Kabakcioglu, D. Balcan and A. Erzan, "Analytical Solution of a Stochastic Model of Gene Regulation," q-bio.MN/0406049] çözümler sonunda, *r* tane sembolden oluşan ve ortalama kelime uzunlukları ~ p^{-1} ile verilen içeriksel ağların derece dağılımları için, küçük ve büyük derecelerde, kuvvet yasalarına uygunluk ve sırasıyla $\gamma_1 = \gamma_2 + 1/2$ ile $\gamma_2 = (1/2)[(\ln r + \ln(1-p)]/[\ln r - \ln(1-p)]$ üstelleri bulunmuştu. Nokta mutasyonlar altında değişmeyen bu sonuçlar, genetik düzenleme ağlarının literatürde gözlenen üstellerine benzemektedir. Biz çalışmamızda, evrimde önemli bir yer tutan gen duplikasyonu olgusunu bu modele uyguladığımızda ölçeklenme sonuçlarının değişip değismeyeceği, değişiyorlarsa, nasıl değişeceklerini araştırdık. Bulunan ölçeklenme davranışı, kalitatif olarak, önceden bulunan sonuçlara benzemekte, mutasyon kurallarımızı değiştirmediğimiz takdirde nümerik sonuçlar da ancak çok küçük değişiklikler göstermektedir. Ne var ki, ağ yapısında uzun erişimli korelasyonlara neden olabilen türden nokta mutasyon kuralları seçildiğinde, üstellerin değerleri değişmektedir. Bu durumda, tamamen gelişigüzel kodlar için hesaplanmış olan analitik ifadelere, etkin *p* değerlerinin koyulması ile bulunan ölçeklenme üstelleri de, simülasyonlarımızdan farklılaşan sonuçlar vermektedirler. İçerik temelli bu modellerin verdiği ağ yapıları, Wagner'in gen duplikasyonu ve diverjansı ile genetik etkileşim ağlarını ilişkilendiren şematik modelinden de, sürekli ve ayrık iki değişik ölçeklenme davranışının gözlenmesi açısından, farklılık göstermektedir.

TOPOLOGICAL PROPERTIES OF GENETİC NETWORKS

ABSTRACT

A content-based model for a network of gene interactions was defined and degree distribution of this network was studied by Balcan and Erzan [D. Balcan, A. Erzan, *Eur. Phys. J. B*, 38, 253-260 (2004)]. Simulations and analytical calculations for an alphabet of *r* lettters and a random code of average word length ~ p^{-1} yielded [M. Mungan, A. Kabakcioglu, D. Balcan and A. Erzan, "Analytical Solution of a Stochastic Model of Gene Regulation," q-bio.MN/0406049] power law behaviour with the exponents $\gamma_1 = \gamma_2 + 1/2$ and $\gamma_2 = (1/2)[(\ln r + \ln(1-p)]/[\ln r - \ln(1-p)]$ for small and large degrees, respectively. The model was invariant under point mutations and had exponents similar to those found for genetic networks studied in the literature. In this study we investigate how incorporating the process of "gene duplication" affects the scaling behaviour of the degree distribution. We find scaling behavior qualitatively similar to the original model; if the same mutation rules are employed as in the original model, duplication gives rise to only very small changes in the scaling exponents. However, for mutation rules which seem to give rise to correlations between different nodes, the values of the exponents change appreciably. In this case, even when the sequence length distributions converge to exponential functions, substituting the effective *p* values in the above expressions calculated for purely random sequences does not yield the exponents we obtain from simulations. The degree distributions obtained from both models differ from the network model introduced by Wagner, based on gene duplication and divergence, in that they exhibit both continuous and discrete regimes.

1. GİRİŞ

Gen etkileşimlerini modelleyebilecek içerik esaslı bir ağ yapısı, Duygu Balcan ve Ayşe Erzan tarafından tanımlamış, simülasyonlarla ve analitik olarak incelenen bu modelde ölçeklenme davranışı gözlenmişti [1,2]. Burada evrim açısından motor görevi gördüğü düşünülen ve Andreas Wagner'in matematiksel bir model ile incelediği [3,4], genlerin kopyalanması diyebileceğimiz duplikasyon olayının ve çeşitli mutasyonların, yukarıda bahsedilen içerik esaslı modelde ne gibi sonuçlara yol açacağına ve ölçeklenme davranışını nasıl etkileyeceğine bakacağız. Öncelikle, konunun motivasyonunun kavranması için ağlardan ve bu konuda yapılan çalışma sonuçlarından, modellemenin temelini oluşturan biyolojik içerikten bahsedeceğiz.

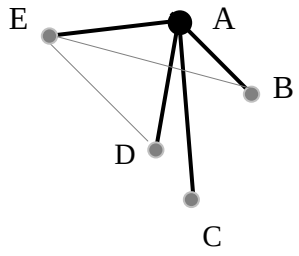
1.1 Ağlar (Networks)

En basit tanımı ile ağ, noktalar ve bu noktaları belirli bir kurala göre birleştiren bağlardan oluşur [5]. Bağlar yönlü veya yönsüz olabilmektedir. Halen aktif olarak çalışılan bu matematiksel problem, soyut veya somut bir çok ilişkiler bütününün sembolik bir temsilini verebilmekte, böylelikle gerçek ağların modellenmesini sağlamaktadır [5]. Gerçek ağlara örnek olarak, internet ve www (world wide web) [6,7], insanlar arasındaki tanışıklık ağları [8,9], ulaşım ağları [10], makalelere yapılan atıfların oluşturduğu ağlar [11-13] ve biyolojik ağlar sayılabilir. Besin ağları [14-16], sinir ağları [17,18], proteinler arası etkileşim ağları [3,19] ve gen etkileşim ağları [4,20,21] da biyolojik ağlara örnek verilebilir. Her ağ için nokta-bağ tanımı ayrı ayrı belirlenmekte, misal olarak, besin ağlarında noktalar hayvanları, aradaki bağlar da iki canlı arasında av ve avcı ilişkisi olduğunu temsil etmektedir. Örneğin bu ağ yapısı incelenerek, ekosistemde kilit rol oynayan bir canlının neslinin tükeneceği öngörülebilmektedir [22].

Yukarıda saydığımız, farklı mekanizmaları içeren alanlarda ortaya çıkan ağ yapılarının incelenen belli başlı global özellikleri, derece dağılımları, kümelenme katsayısı, noktalar arasındaki en kısa yol ve bir büyük kümenin var olup olmadığının tayini olarak sayılabilir [23,24]. Bu özelliklerin tanımlarını sırasıyla verelim.

Ağ içinde, bir noktanın sahip olduğu tüm bağların sayısına, o noktanın derecesi (d) adı verilir, noktalar arasındaki bağlar yönlü olduğunda ise, giren ve çıkan derece sayısı olarak nitelendirilir [23,24]. Tüm noktalar için derece sayıları hesaplandıktan sonra, derece sayılarına karşılık gelen noktaların sayılması ile derece dağılımları ($n(d)$) oluşturulur. Bağlar yönlü olduğunda çıkan-derece dağılımı (out-degree distribution) ve giren-derece dağılımı (in-degree distribution) ayrı ayrı oluşturulur.

Bir noktanın kümelenme katsayısı (C), bağlandığı noktalar arasındaki bağ sayısının, bu noktalar arasında olabilecek maksimum bağ sayısına bölümü olarak tanımlanır. Örneğin d dereceli bir noktanın, bağlandığı noktalar arasındaki bağ sayısı E ise, bu noktanın kümelenme katsayısı $C=E/(d(d-1)/2)$ olur [23,24]. Aşağıdaki örnek ile, şekildeki A noktasının kümelenme katsayısını hesaplayarak tanımı açıklayabiliriz. A noktası B,C,D ve E noktaları ile bağ kurmakta, böylece bu noktanın derece sayısı (d) 4 olmaktadır. Bağ kurulan dört noktadan, B ile E ve D ile E de kendi arasında bağ kurmakta ve A noktasından çıkan iki tane üçgen oluşturmaktadır. Bağlanılan noktaların arasında oluşan bu iki bağ veya A noktasından çıkan üçgenlerin sayısı da E değerini vermektedir. Buradan da bu noktanın kümelenme katsayısı verilen bağıntıdan bulunur. Kümelenme katsayısının, derece sayısına göre dağılımı olan $C(d)$ de hesaplanan diğer büyüklükler arasındadır [25].



Şekil 1.1.1 A , B, C, D ve E noktalarından oluşan ağ.

Ağ içindeki iki nokta arasındaki toplam bağ sayısının en küçük değeri ise, o iki nokta arasındaki en kısa yol olarak adlandırılır. Ortalam en kısa yolun küçük olduğu

durumda, kümelenme katsayısı büyük ise, ağın küçük-dünya (small world) özelliğine sahip olduğu söylenmektedir [23,24].

Ağlar ilk defa Paul Erdös ve Alfred Renyi tarafından incelendi. Gelişigüzel bir ağ yapısı olan Erdös-Renyi modelinde, *n* tane nokta arasındaki bağlar *p* olasılığı ile kurulur, mümkün bağ sayısı n(n-1)/2 tane olduğundan, n noktalı, *n*(*n*-1)*p*/2 kenarlı bir ağ oluşur. Her bağ iki noktayı birleştirdiğinden, bir noktanın ortalama derece sayısı <*d*>=n*(*n-*1)p/*n bulunur [26]. Nokta sayısının, çok büyük değerlerinde <*d*>=*np* ve derece dağılımını $P(d) \sim e^{-np} (np)^d/d!$ biçiminde bir Poisson dağılımı olarak bulunur. Erdös-Renyi modelinde noktalar arasındaki ortalama en kısa yolun log(*n*) biçiminde, nokta sayısı ile yavaş büyüdüğü küçük-dünya özeliği görülmektedir [26].

Bir diğer model, Barabasi-Albert n_0 tane nokta ile başlayıp, her zaman adımında bir tane noktayı ağa ekleyerek, "prefential attachment" da denilen, mevcut noktaların kenar sayısının bir fonksiyonu olan, $p_i(d_i) = d_i / \sum d_i$ olasılığı ile, eklenen nokta ile diğer noktalar arasında bağlar kurmaktadırlar. Olasılık ifadesinde d_i i. noktanın derecesi ve toplam ise yeni nokta eklenmeden önceki tüm noktalar üzerinden yapılmaktadır [27]. Bu modelde derece dağılımı $P(d) \sim d^{-3}$ biçiminde bir kuvvet yasası olarak, yani ölçekten bağımsız bulunmaktadır [27]. Sosyal ya da biyolojik sistemlerde ortaya çıkan etkileşim ağlarının birçoğunun kuvvet yasası türünden derece dağılımları verdiği gözlenmiştir [23,24]. Burada anlatılan ağ yapısı Barabasi ve Albert'in, bu tür dedece dağılımlarının nasıl ortaya çıkabileceğine dair örnek olarak önerdikleri bir modeldir. Bu modelde de, noktalar arasındaki yolun ln(*n*)/ln(ln(*n*)) biçiminde değiştiği, küçük-dünya özeliği görülmektedir [27].

1.2 Genetik Ağlar

Gen etkileşimlerini daha iyi açıklamak için, öncelikle gerekli bazı genetik kavramlardan bahsedelim. Kalıtımın ve bir hücre için yaşamsal faaliyetlerin merkezi hücre çekirdeğindeki DNA (deoksiribonükleik asit) denilen moleküldür, DNA'dan oluşan tüm genetik malzemeye genom adı da verilmektedir. DNA, karşılıklı duran ve birbirinin tamamlayıcısı olan, iki zincirin oluşturduğu, çift sarmal da denilen, helezonik bir yapıdır. Bir çok canlıda DNA, bir nevi üst üste katlanarak kromozom

adı verilen yapılar oluştururlar. DNA' daki her bir zincir nükleotid adı verilen, fosfat grubu, şeker ve baz üçlüsünden oluşmaktadır. Dört çeşit olan bazlar adenin (A), timin (T), guanin (G) ve sitozin (C) olup, 'Watson-Crick' baz eşleşmesine göre, iki zincirde A-T ve C-G daima karşılılı durmaktadır. Bu zincirdeki baz dizilimleri, hücresel işlevleri yerine getirebilecek asıl moleküller olan proteinleri şifrelemektedir. Her üçlü baz dizilişi, proteinlerin yapıtaşı olan aminoasitlere karşılık gelmektedir. Örneğin AAG dizisi fenilalanin adlı aminoasiti ifade etmektedir. Belli bir proteini kodlayan kısa DNA parçasına da gen adı verilmektedir. Basit bir anlatımla, DNA'ya benzeyen ama kısa ve tek zincirden oluşan RNA (ribonükleik asit) yardımı ile gerçekleşen, protein kodlanmasını ve üretimini şöyle açıklayabiliriz. DNA'nın iki zinciri açılarak, kodlanacak proteine ait bölgenin (gen), elçi RNA (messenger RNA) denilen kopyası çıkarılır (transkripsiyon) ve proteini üretmek için hücre içinde ribozom denilen organele giderek, elçi RNA üzerindeki üçlü nükleotitlere karşılık gelen aminoasitler, transfer RNA aracılığı önce taşınıp, sonra birbirine bağlanarak (translasyon) protein oluşturulur [28-30].

Çok hücreli canlıların, her hücresinde, canlı için tüm proteinleri üretme kapasitesine sahip, aynı DNA dizilimi bulunmaktadır. Ancak, hem hücrelerde her an her protein üretilmez, hem de hücreler farklı işlevleri yerine getirirler, dolayısıyla her hücrede aynı proteinler de üretilmez [28-30]. Proteinlerin ne zaman üretilip, ne zaman üretilmeyeceğine dair kontrol mekanizmaları mevcuttur. Bu kontrol mekanizmaları, genlerin birbirleri ile olan ilişkilerine, yani aralarında oluşturdukları ağın yapısına bağlıdır. Aslında gen denilen DNA parçası, hem işlevsel hem de düzenleyici proteinler kodlayabilir, aynı zamanda geni açma-kapama düğmesi sayılabilecek dizilimleride içinde barındırır [28-31]. Bir genin ürettiği düzenleyici protein (transcription factor-TF), diğer bir genin üzerindeki düzenleyici bölgeye (geni açma-kapama bölgesine) kimyasal bağ ile bağlanarak, o genin ürettiği proteinin kodlanmasına ya da kodlanmamasına neden olabilmektedir [28-31]. Bu etkileşim biçimi nedeni ile gen etkileşimleri ağ yapıları ile modellenmektedir.

Balcan-Erzan'ın gen etkileşimleri modelinin temelinde, bir genin ürettiği proteinin, diğer genin çalışmasını durdurması gözönüne alınmıştır. Oluşturulan ağ modelinde, noktalar belli sayıdaki harf kümesinden, rastgele seçilerek oluşturulan kelimeler diyebiliriz. Bu iki nokta arasında bağ kurulmasının kuralı ise, noktalardan birindeki kelimenin harf dizilişinin, diğer noktadaki kelimenin harf dizilişi arasında kesintisiz

olarak bulunmasıdır. Burada oluşan ağ modelinde, hem simülasyon hem de analitik hesaplar sonunda, çıkan-derece dağılımı küçük derecelerde bir kuvvet yasası biçiminde ($d^{-\gamma_1}$), büyük derece sayılarında ise logaritması salınan bir davranış göstermekte, ama bir kuvvet yasası ($d^{-\gamma_2}$) ile zarfı sınırlandırılabilmektedir [1,2]. Ölçeklenme davranışı gösteren bu modelin nokta mutasyonlar altında kararlı bir davranış sergilediği de gözlenmiştir [1,2].

Bir organizmada kalıtımsal karmaşıklık, DNA'nın ne kadar fazla baz içerdiğinden çok, bazların oluşturduğu genler arasındaki ilişkilerde, yani gen etkileşim ağı ve onun mümkün kıldığı farklı kombinatorik oluşumlarda kendini göstermektedir. Bu etkileşim ağı, bir nevi genetik öğrenme-algılama mekanizması olarak da görülebilir [32].

Gen duplikasyonları, DNA içinde, gerekli olmamasına rağmen fazladan aynı baz dizilerin oluşmasına neden olur. Ancak uzun zaman içinde, bu fazlalık genler çeşitli mutasyonlarla yeni işlevler kazanıp, daha kompleks biçimlerin oluşmasına, evrim süreci içinde daha gelişmiş canlılara geçişi sağlamaktadır [33-35]. Örneğin, genomun, hücre çekirdeğinin olmadığı prokaryotlarda %50'sinin, hücre çekirdeğinin olduğu ökaryotlarda %90'nın duplikasyon sonucundaki genlerden oluştuğu düşünülmektedir [36].

Tezin 2. bölümünde Balcan–Erzan modelini tanıtıp, duplikasyon ve çeşitli mutasyonların buraya nasıl uygulandığını anlatacağız. Modeli tanıtırken kullanacağımız alt dizi ifadesi genel olarak gen içindeki düzenleyici bölgeye, dizilim ifadesi ise tüm genoma karşılık gelecektir. Daha sonra 3. bölümde simülasyonları ve 4. bölümde sonuçları vereceğiz.

2. MODEL

İlk olarak burada kullandığımız modeli tanıtacağız, daha sonra bu modele uyguladığımız evrimleşme mekanizmalarından bahsedeceğiz.

2.1 İçerik Esaslı Ağ Modeli

Bu bölümde, genlerin birbirini baskılamalarını gözönünde bulunduran Balcan-Erzan'ın modelinin (BE) tanımını, sonra bu model sonunda bulunan sonuçları kısaca vereceğiz. Bölüm 2.1'in tamamı Balcan-Erzan çalışmasının özeti olacaktır, bundan dolayı bölüm içinde sürekli referans vermeyeceğiz.

2.1.1 Modelin Tanımı

Uzun bir dizilim, {0,1,2} setindeki elemanlar, x bu set elemanlarından biri olmak üzere, belirli bir p değeri için aşağıdaki olasılıklar ile seçilerek oluşturulsun.

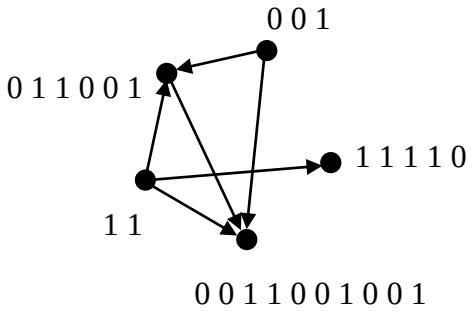
$$P(x) = p\delta(x-2) + \frac{(1-p)}{2}[\delta(x-1) + \delta(x)] \qquad (2.1.1.1)$$

Yani 2 elemanı p olasılığı ile, 0 ve 1 elemanları ise (1-p)/2 olasılıkları ile seçilmektedir. Modelde ağı oluşturan noktalar {0,1} setindeki karakterlerden oluşan alt dizilerdir, alt dizileri birbirinden ayıran, yani alt diziyi hem başlatan hem de sonlandıran karakter olarak 2 kullanılır. Böylece aşağıdaki gibi bir dizilim oluşturulabilir.

2 0 0 1 2 0 1 1 0 0 1 2 1 1 2 0 0 1 1 0 0 1 0 0 1 2 1 1 1 1 0

Şekil 2.1.1.1 $P(x)$ ile oluşturulan örnek dizilim.

Dizilim uzunluğu *L* ile gösterilirse, bu dizilim için *L*=31'dir. Buradaki alt diziler ağdaki noktalara karşılık gelmektedir. Noktaların bağlanma kuralı, dizilerin biribirini içinde barındırması, yani birebir eşleşmeleridir. Başka bir deyişle, *i.* ve *j.* alt diziler için, bağlanma kuralı *i.* alt dizinin, *j.* alt dizinin bir alt kümesi olması, yani *j.* dizinin *i.* diziyi içinde barındırmasıdır. Buradaki bağlar yönlü olup, *i'*den *j'* ye doğrudur, yani *i.* diziden başlayan bir okun ucu içinde bulunduğu *j.* diziye doğrudur. Yukarıdaki dizilim için, alt diziler arasındaki yönelimli bağlarla oluşan ağın görünümü Şekil 2.1.1.2 de gibi olur. Örneğin bu ağda 0 0 1 alt dizisinden, içinde bulunduğu 0 1 1 0 0 1 alt dizisine doğru bir ok çizilmiştir.



Şekil 2.1.1.2 Örnek dizilimdeki alt dizilerin oluşturduğu ağ.

Dizilimde *i.* alt dizi G_i, bu alt dizinin uzunluğu l_i ile gösterilirse, yukarıdaki örnek için aşağıdaki tabloyu oluşturabiliriz.

Tablo 2.1.1.1 Örnek dizilim için alt dizi gösterimi ve uzunlukları.

i	G_i	l_i
1	0 0 1	3
2	0 1 1 0 0 1	6
3	1 1	2
4	0 0 1 1 0 0 1 0 0 1	10
5	1 1 1 1 0	5

Bir alt dizinin, örneğin l_i uzunluğundaki *i.* alt dizinin elemanları (2.1.1.2)'deki biçimde gösterilebilir. Burada $x_{i,\lambda} \in \{0,1\}$ ve $\lambda = 1,2,.....,l_i$ değerlerini almaktadır.

$$G_i = x_{i,1}, x_{i,2}, x_{i,3}....., x_{i,l_i} \qquad (2.1.1.2)$$

Alt dizi sayısını *N* ile, uzunluğu *l* olan alt dizilerin sayısını *u*(*l*) ve en büyük alt dizi uzunluğunu da l_{maks} ile gösterirsek, toplam alt dizi uzunluğu ve alt dizi sayısı için (2.1.1.3)'deki bağıntılar yazılabilir.

$$\sum_{i=0}^{N} l_i = L - N\,, \qquad \sum_{l=0}^{l_{maks}} u(l) = N \qquad (2.1.1.3)$$

Uzunluğu *l* olan bir dizinin oluşma olasılığı, *l*+2 bölmeli bir kutunun, ilk ve son bölmesine 2 koyma olasılıkları ile aradaki *l* tane bölmeye {0,1} setinden birini koyma olasılıklarının çarpımına eşittir. Bu olasılığı dizilim uzunluğu ile çarpınca, *l* uzunluğunda ortalama kaç alt dizi olduğu yani $\langle u(l) \rangle$ değeri bulunmuş olur. Dizilimi oluştururken, 2 seçme olasılığı p, 0 veya 1 seçme olasılığı ((1-*p*)/2+(1-*p*)/2) olduğundan ortalama değer için,

$$\langle u(l) \rangle = Lp^2(1-p)^l = Lp^2 e^{-l|\ln(1-p)|} \qquad (2.1.1.4)$$

bulunur. Yani uzunluk dağılımı $u(l):\ e^{-l|\ln(1-p)|}$ biçiminde eksponansiyel olmakta, $|\ln(1-p)|$ değeri de dağılımı karakterize eden üs olmaktadır. Buradan $\langle l \rangle$ küçük bir hesapla L' nin sonsuza gittiği limitte aşağıdaki şekilde bulunur.

$$\langle l \rangle = \sum_{l=1}^{l_{maks}} l \frac{u(l)}{N} = \frac{1-p}{p} \qquad (2.1.1.5)$$

Ağlar etkileşim matrisi (adjacency matrix) adı verilen bir matris ile kolayca ifade edilebilmektedir. Tüm ağı, etkileşim matrisi W ile verdiğimizde, bu matrisin w_{ij} elemanı, *i.* ve *j.* noktalar arasında bağ varsa, yani *j.* alt dizi *i.* alt diziyi içeriyorsa 1,

diğer durumlarda 0 değerini alacaktır. Böylece tüm matris elemanları (2.1.1.6)'daki koşula göre oluşturulur.

$$w_{ij} = \begin{cases} 1, \ \mathrm{G_i} \subseteq \mathrm{G_j} \\ 0, diğer \end{cases} \tag{2.1.1.6}$$

Yukarıdaki örnek ağ için, *W* matrisi (2.1.1.7)'deki gibi yazılır.

$$W = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.1.1.7}$$

W matrisi ile, asıl araştırılmak istenen giren, çıkan ve toplam derece sayıları (bağ sayıları) bir *i* noktası için, sırasıyla (2.1.8)'deki eşitliklerle bulunur. Bir noktanın toplam derece sayısı, bağların yöneliminden bağımsız olarak kaç tane nokta ile bağlı olduğunun sayısıdır.

$$\begin{aligned} d_{giren}(i) &= \sum_{j \neq i} w_{ji} \\ d_{çıkan}(i) &= \sum_{j \neq i} w_{ij} \\ d(i) &= d_{giren}(i) + d_{çıkan}(i) - \sum_{j \neq i} w_{ij} w_{ji} \end{aligned} \tag{2.1.1.8}$$

Mesela $G_2 = 0,1,1,0,0,1$ dizisi için (2.1.8) eşitlikleri kullanılarak ve şekil 2.1.2'de de görüldüğü gibi $d_{giren}(2)=2$, $d_{çıkan}(2)=1$ ve $d(2)=3$ bulunur. Tüm dizilimin istatistiğini yaparken, yani ağın topolojik özelliklerini bulurken, tüm alt diziler üzerinden hesap yapılarak derece sayılarının dağılımları bulunacaktır. Bu dağılımların hesabı aşağıdaki eşitlikler kullanılarak oluşturulur.

$$n_{giren}(d) = \sum_{i=1}^{N} \delta(d - d_{giren}(i))$$
$$n_{çıkan}(d) = \sum_{i=1}^{N} \delta(d - d_{çıkan}(i)) \qquad (2.1.1.9)$$
$$n(d) = \sum_{i=1}^{N} \delta(d - d(i))$$

Yani çıkan (giren) derece dağılımını oluşturmak için, belli bir *d* çıkan (giren) derece sayısına sahip kaç tane nokta olduğu sayılmaktadır.

2.1.2 Modelin Sonuçları

Önceki bölümde anlatılan şekilde dizilim oluşturulduktan sonra, her elemanı belli bir olasılıkla nokta mutasyonlara tabi tutularak zaman içindeki evrimine bakılmıştır. Böylelikle (2.1.1.1) deki dağılım ile oluşturulan dizilimin, hesaplanan global özelliklerinin mutasyon altında değişip değişmeyeceği incelenmiştir. Belirlenen mutasyon kuralı,

- $0 \leftrightarrow 1$ değişimi (değiş tokuşu),
- alt dizileri sınırlayan 2'lerin rastgele sağa veya sola hareketleri,

olarak kabul edilmiştir.

Bu mutasyonlar sadece bir gelişigüzel diziyi bir diğeri ile değiştirdiklerinden ağ yapısının davranışının bu mutasyon kuralı altında değişmez kaldığı gözlenmişir. Giriş bölümünde de belirtidiği gibi "çıkan derece" dağılımları için belli aralıklarda davranışın başat özeliğinin kuvvet yasasına benzediği bulunmuş, büyük ve küçük derece değerleri için iki farklı ölçeklenme üsteli, hem simülasyonlarla bulunmuş hem de modelin analitik çözümü ile doğrulanmıştır. Bu sonuçlar 3. bölümde daha ayrıntılı olarak tartışılacaktır. Burada yapılan analitik hesaplar sonucunda küçük ve büyük dereceler için bulunan ölçeklenme üstellerinin ifadelerini ispatsız vereceğiz.

$$\gamma_2 = \frac{1}{2}\frac{\ln(1/2) - \ln(1-p)}{\ln(1/2) + \ln(1-p)}$$
$$\gamma_1 = \gamma_2 + 0.5 \qquad (2.1.2.1)$$

Bizim hesaplarımızda sadece 0 ve 1'lerden oluşan alt diziler söz konusu olduğu için bu ifadede alfabe uzunluğu 2 olarak alınmıştır.

2.2 İçerik Esaslı Ağ Modelinin Zamanla Evrimleşme Mekanizmaları

İlk bölümde de bahsettiğimiz gibi duplikasyon ve sonrasındaki mutasyonların evrimleşme mekanizmaları olduğu düşünülmektedir. Evrim içinde tek hücreli organizmalardan, kompleks yapıdaki çok hücreli organizmalara geçişte, sadece, farklı allel genlerin doğal seçilimle (natural selection) seçilmesinin yeterli olamayacağı görüşü vardır [37]. Allel gen, kromozom üzerinde belli bir özelliğin, seçilme şansı olan değişik varyasyonlarıdır denebilir [38]. Doğal seçilim için, duplikasyon sonunda oluşan iki kopya, çeşitliliğe yol açacak biçimde işlenebilir malzeme olarak düşünülebilir, böylelikle genom içinde yeni işlevlere sahip genler oluşabilecektir [37]. Duplikasyon sonucundaki kopya gen, ya hemen kopyalanan genin ardından (tandem), ya kopyalanan genin ardından ama ters sırada (reverse-tandem) ya da kromozom üzerinde herhangi bir yere (any point) yerleşebilmektedir [38].

Buradaki modelin içerik esaslı olduğu gözönünde bulundurulduğunda, Wagner modelini içeriklendirmiş olduğumuz düşünülebilir. Wagner'in gen duplikasyonu ve diverjansı ile genetik etkileşim ağlarını ilişkilendiren şematik modelinde ağ soyut noktalardan oluşmaktadır. Wagner, bu soyut noktalardan rastgele seçtiği birini duplikasyona uğratıp ağa ekliyor ve kopya nokta da aynı bağlara sahip oluyor. Eklenen noktanın bir bağını δ olasılığı ile keserken, α olasılığı ile de ağdaki diğer bir nokta ile eklenen nokta arasında bağ kuruyor. Tüm derece sayıları için tek kuvvet yasası veren bir derece dağılımı buluyor.

Genom içinde gerçekeleşen bazı mutasyonlar kromozomun büyük bir kısmını veya sadece bir gendeki birkaç bazı etkileyenler olarak sınıflandırılmaktadır [39]. Bunlara sırası ile kromozomal ve nokta mutasyonlar denmektedir. Duplikasyon kromozomal bir mutasyon olarak sayılmaktadır [39]. Böylece, modelde hem nokta mutasyonları hemde kromozomal mutasyonları kullandığımızı söyleyebiliriz.

Şimdi modelde duplikasyonu ve mutasyonları nasıl uyguladığımızı anlatalım.

2.2.1 Duplikasyon

Her zaman adımında rastgele olarak seçilen bir alt dizi duplikasyona uğratılmaktadır, yani dizilim içine kopyalanmaktadır. Bu kopyalanma aşağıdaki üç farklı şekilde olabilmektedir.

1. Ard arda kopyalama ('Tandem duplication') (1/3 olasılıkla)

   a. Kopyalanan ve kopya alt dizi yan yana, arada 2 var.

   b. Kopyalanan ve kopya alt dizi yan yana, arada 2 yok.

2. Ard arda ters sırada kopyalama ('Reverse-Tandem duplication') (1/3 olasılıkla)

   a. Kopyalanan ve ters sıradaki kopya alt dizi yan yana, arada 2 var.

   b. Kopyalanan ve ters sıradaki kopya alt dizi yan yana, arada 2 yok.

3. Herhangi bir yere kopyalama ('Any-Point duplication') (1/3 olasılıkla)

   a. Kopya alt dizi, dizilim üzerinde rastgele seçilen bir yere yerleştirilir.

Tanımını verdiğimiz duplikasyon işleminin nasıl yapıldığını örnek dizilim üzerinde görelim. Alt dizi olarak "1. altdizi" seçilmiş olsun. Duplikasyondan önce dizilim:



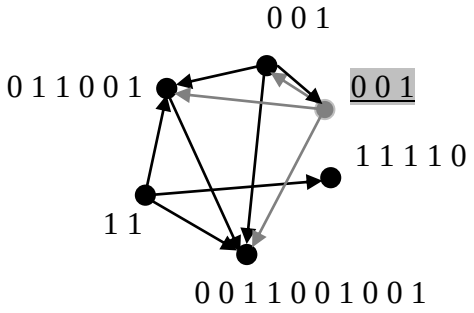
1. alt dizi

2 0 0 1 2 0 1 1 0 0 1 2 1 1 2 0 0 1 1 0 0 1 0 0 1 2 1 1 1 1 0

Seçilen alt diziyi kopyalıyoruz, dizilime yerleştirirken 1/3 olasılıkla aşağıdakilerden birini seçip onu uyguluyoruz.

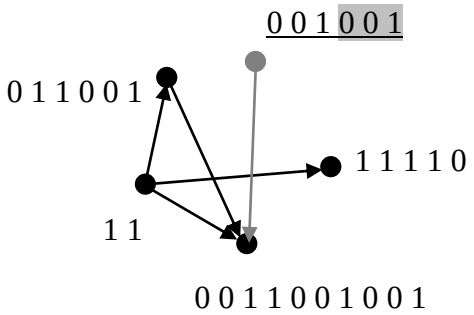
1. Kopya alt diziyi, seçilen alt dizinin sağ yanına yerleştirirken,

   a. 1/2 olasılıkla araya gen ayracı koyuyoruz, aslında bu durumda kopyanın dizilimin neresine yerleştirildiğinin ağ yapısı için bir farkı yok. Ağa bir tane daha nokta eklenmiş olup, bağ sayısını arttırıyor.



Şekil 2.2.1.1 Duplikasyon için 1.a durumuna verilen örnekte ağın görünümü.

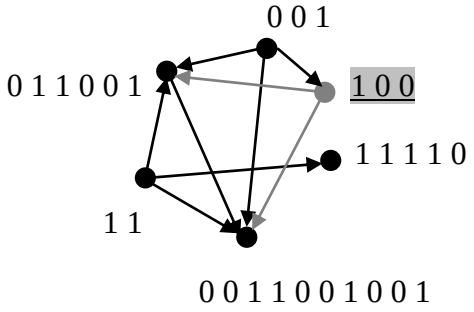
b. 1/2 olasılıkla araya gen ayracı koymayıp, kopya alt diziyi seçilen diziye ekliyoruz, böylece daha uzun bir alt dizi elde ediyoruz. Bu durum, noktalar arasındaki bağları yeniden düzenliyor, ya yeni bağlar kuruluyor ya da önceki bağlar koparılıyor.



Şekil 2.2.1.2 Duplikasyon için 1.b durumuna verilen örnekte ağın görünümü.

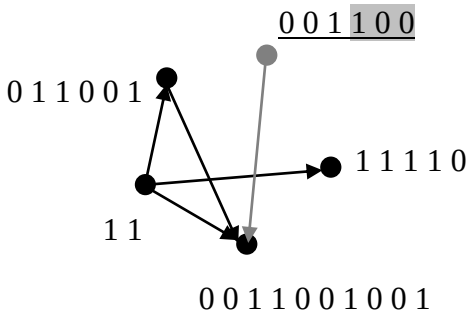
2. Kopya alt dizi ters çevrilerek seçilen alt dizinin sağ yanına yerleştirirken,

a. 1/2 olasılıkla araya gen ayracı koyuyoruz, kopya ile aynı uzunlukta yeni bir alt dizi oluşuyor. Ağ içindeki bağ sayısını arttırıyor.



Şekil 2.2.1.3 Duplikasyon için 2.a durumuna verilen örnekte ağın görünümü.

b. 1/2 olasılıkla araya gen ayracı koymayıp, kopya alt diziyi seçilen diziye ekleyerek daha uzun bir alt dizi elde ediyoruz. Bu durum, noktalar arasındaki bağları yeniden düzenliyor, ya yeni bağlar kuruluyor ya da önceki bağlar koparılıyor.



Şekil 2.2.1.4 Duplikasyon için 2.b durumuna verilen örnekte ağın görünümü.

3. Kopya alt diziyi, dizilim üzerinde rastgele seçilen bir yere yerleştiriyoruz. Örneğin,

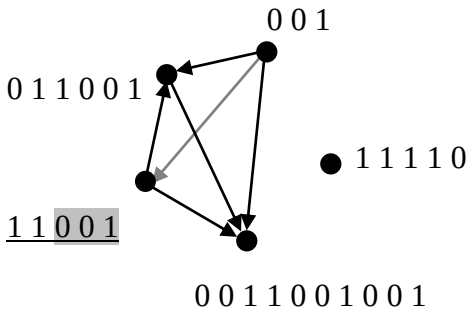
Rastgele seçilen yer

2 0 0 1 2 0 1 1 0 0 1 2 1 1 2 0 0 1 1 0 0 1 0 0 1 2 1 1 1 1 0

ise en son aşağıdaki hali alır.

2 0 0 1 2 0 1 1 0 0 1 2 1 1 0 0 1 2 0 0 1 1 0 0 1 0 0 1 2 1 1 1 1 0

Kopya alt dizi araya girdiği diziyi parçalayarak, ondan daha uzun yeni bir alt dizinin ağa katılmasına neden olmaktadır.



Şekil 2.2.1.5 Duplikasyon için 3.a durumuna verilen örnekte ağın görünümü.

2.2.2 Çeşitli Mutasyonlar

Çalışmamızda iki farklı mutasyon seti kullanacağız. Bunlar Balcan-Erzan'ın oluşturdukları modele uyguladığı, burada M1 ile göstereceğimiz mutasyon seti ve M2 ile göstereceğimiz yeni bir mutasyon seti olacak. Mutasyonları nasıl uyguladığımızı şöyle anlatabiliriz. Dizilim üzerindeki elemanları 1'den, dizilim uzunluğu L'ye kadar numaralandırdığımızı düşünelim. Bir t zamanında sıra ile i=1 den baslayarak, $i=L$ ye kadar tüm dizilim mutasyon için taranıyor. Bulunduğumuz nokta, *k.* eleman ise, önce μ olasılığı ile mutasyonun uygulanıp, uygulanmayacağına karar veriliyor. Eğer uygulanmayacaksa, *k.* eleman hiçbir işleme tabi tutulmuyor, bir sonraki elemana geçiliyor. Eğer uygulanacaksa belirlenen sete göre mutasyon uygulanıyor. Mutasyon setlerini k.eleman üzerinde aşağıdaki şekilde anlatabiliriz.

M1 mutasyon seti:

- *k.* eleman 0 ise 1 ile, 1 ise 0 ile değiş tokuş yapılıyor.
- *k.* eleman 2 ise rastgele olarak sağ veya solundaki eleman ile yer değiştiriyor.

M2 mutasyon seti:

Rastgele olarak aşağıdaki üç işlemden biri seçilir ve o uygulanır.

a. Ekleme (insertion)

- *k.* elemanın sağına 0,1 ve 2 den rastgele seçilen biri yerleştirilir.

b. Silme (deletion)

- *k.* eleman silinir.

c. Değiştirme (replacement)

- 0 ,1 veya 2 seçilmiş ise bu değer ,rastgele olarak seçilen 0 veya 1 ile değistirilir.

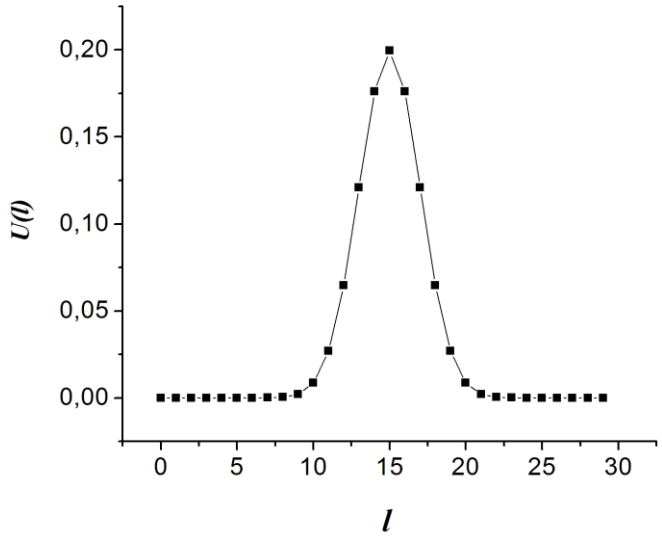
M1 ve M2 mutasyon setinden bahsederken kimi zaman sadece M1 ve M2 diyeceğiz.

2.2.3 Başlangıç Koşulları

Burada (2.1.1.1)'deki seçimle dizilimleri oluşturma yanında farklı bir dağılımla da dizilimleri oluşturduk. Bunu yaparken düzenleyici dizilerin (regulatory sequences) 5 – 26 arasında nükleotide sahip olduğunu [40] gözönünde bulundurarak dizi uzunlukları için aşağıdaki Gaussien dağılımını da başlangıç dizilimini oluşturmada kullanacağız.

$$U_g(l) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{1}{2}}} e^{\frac{-(l-15)^2}{2\sigma^2}} \qquad (2.2.3.1)$$

Simülasyon hesaplarında $\sigma = 2$ aldık, aşağıda bu değer için dağılımın grafiğini görüyoruz.



Şekil 2.2.3.1 $\sigma = 2$ için $U_g(l)$'nin grafiği.

Tüm dizilimi oluştururken, her alt dizi için, $U_g(l)$ dağılımına göre l alt dizi uzunluğunu seçtik ve rastgele seçilen 0 ve 1 elemanları ile diziyi oluşturduk. Dizileri birbirinden ayırt etmek için aralara yine "2" sembolünü yerleştirdik.

2.2.4 Model Tanımlamaları

Elimizde iki farklı başlangıç koşulu, evrimleşme mekanizmaları olarak da iki farklı mutasyon seti ve duplikasyon var. Bunların etkilerini tam olarak görebilmek için seçilen mutasyon setine göre Model I ve Model II olarak, sekiz tane alt modeli bunlar altında gruplandırdırarak hepsinin ne sonuçlar verdiğine baktık. Bunların tanımları ve kısa gösterimleri tablo 2.2.4.1'deki gibi olmaktadır.

Tablo 2.2.4.1 Modellerin kısa gösterimi.

Başlangıç koşulu	Mutasyon seti	Duplikasyon	Kısa gösterimi
Eksponansiyel	M1	Yok	PM1
Gaussien	M1	Yok	UM1
Eksponansiyel	M1	Var	PDM1
Gaussien	M1	Var	UDM1
Eksponansiyel	M2	Yok	PM2
Gaussien	M2	Yok	UM2
Eksponansiyel	M2	Var	PDM2
Gaussien	M2	Var	UDM2

Böylelikle Model I'in içinde PM1, UM1, PDM1ve UDM1 alt modelleri bulunmakta, Model II'in içinde de PM2, UM2, PDM2 ve UDM2 alt modelleri bulunmaktadır.

3. SİMÜLASYONLAR

İkinci bölümde tanımlamış olduğumuz modellerin sayısal benzeştirim (simülasyon) aracılığı ile incelenmesi sonucunda elde ettiğimiz davranışlarını bu bölümde sunacağız. Bir önceki bölümde anlattığımız alt modellerin simülasyonunu yaparken p=0.05 (dizilim oluşturulurken $P(x)$ olasılık fonksiyonu ile {0,1,2} elemanlarının seçilme olasılıklarını belirleyen değer), μ=0.05 (dizilimdeki bir elemanın mutasyona uğrama olasılığı) ve üzerinden hesap yapılan gerçekleşim sayısı ise 500 alındı. İki başlangıç koşulumuz vardı, bunlardan birinde dizilim uzunluğu L'nin ve diğerinde alt dizi sayısı N'in başlangıç değerleri (yani t=0'da) sırasıyla L_0=15000 ve N_0=700 olarak alındı. Aşağıdaki tabloda simülasyonda kullanılan bu sayısal değerleri birarada görebiliriz.

Tablo 3.1 Simülasyonda kullanılan sayısal değerler.

Gerçekleşim sayısı	L_0	N_0	p	μ
500	15000	700	0.05	0.05

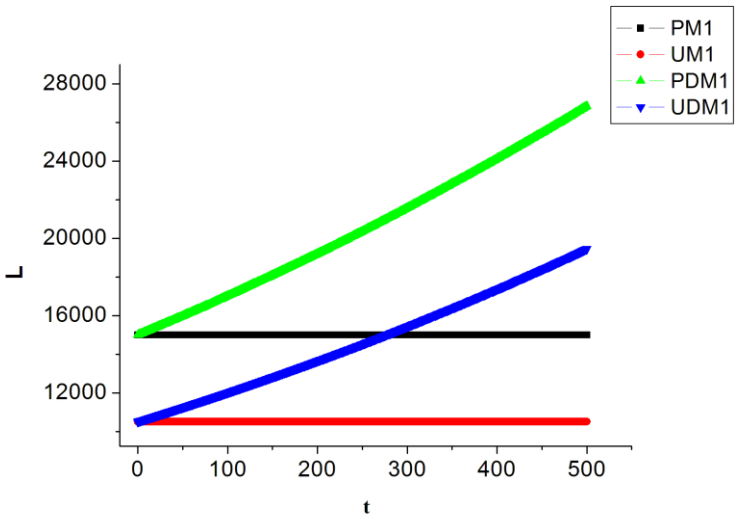
Eksponansiyel ve Gaussien olmak üzere iki farklı başlangıç koşulu simulasyonlarda şöyle uygulanmaktadır. Eksponansiyel başlangıç koşulu ile, t=0'da ilk dizilimi oluştururken L_0'ın (dizilim uzunluğunun) verilmiş olması gerekiyor. Böylelikle dizilimi oluşturacak olan L_0 tane eleman, {0,1,2} setinden $P(x)$ olasılık dağılımına göre seçiliyor. Yani p olasılığı ile 2, (1- p)/2 olasılığı ile 0 ve yine (1- p)/2 olasılığı ile 1 seçiliyor. Gaussien başlangıç koşulu olduğunda, N_0 (alt dizi sayısı) veriliyor. Örneğin i. alt dizi oluşturulurken $U_g(l)$ dağılımından bu alt dizinin uzunluğu, yani l_i belirleniyor. Daha sonra {0,1} setindeki elemanlar rastgele seçilerek (1/2 olasılıkla) belirlenen uzunluktaki alt dizi oluşturuluyor. Bu şekilde oluşturulan herbir alt dizi arasına da 2 koyuluyor.

3.1 Dizilimin Zamanla Evrimi

Bu bölümde dizilim uzunluğunun (L), alt dizi sayısının (N), ortalama alt dizi uzunluğunun ($\langle l \rangle$) ve alt dizilerin uzunluk dağılımının ($\langle u(l) \rangle$) zamanla nasıl değiştiğine bakılacaktır.

3.1.1 Dizilim Uzunluğu (*L*)

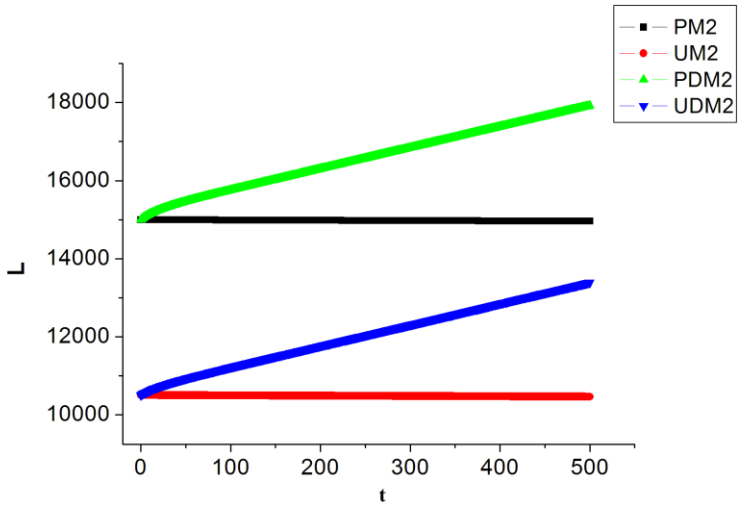
Model I'e bakıldığında, M1 dizilim üzerindeki 0 olan elemanı 1 ve 1 olan elemanı 0 değeri ile değiş tokuş yapıyor, ayrıca gen ayraçları olan 2'leri sağdaki veya soldaki elemanla yer değiştiriyordu. M1, içeriği dolayısıyla toplam uzunluk değişimine neden olacak bir faktör içermediğinden, şekil 3.1.1.1'de görüldüğü gibi, PM1 ve UM1 zamanla *L*'de değişime neden olmuyor.



Şekil 3.1.1.1 Dizilim uzunluğu *L*'nin M1 altında zamanla değişimi.

M2 mutasyon kuralında, değiştirme işlemi dizilim uzunluğunda zaten bir etki yapmıyor. Ekleme ve silme işlemleri ise eşit olasılıkla (μ/3) uygulandığından ortalamada L'yi değiştirmeyecektir. Ancak PM2 ve UM2 için simülasyon sonuçlarında, gerçekleşim sayısının belki de yetersiz olmasından L'de çok küçük de olsa bir azalma olmaktadır. Zaten şekil 3.1.1.2' ye baktığımızda bu küçük değer

değişimlerini ayırt edemiyoruz ama tablo 3.1.1.1'de PM2 ve UM2 için, L'nin değişim hızlarının değerlerini görebiliriz.



Şekil 3.1.1.2 Dizilim uzunluğu *L*'nin M2 altında zamanla değişimi.

Alt diziler kopyalandığında, yani duplikasyona uğratıldıklarında, *L* uzunluğu elbette her iki mutasyon seti için de değişiyor. Bu durumda M1, başlangıç koşuluna, yani t=0 da başladığımız dağılıma göre, PDM1 ve UDM1'de *L* için farklı artma hızları veriyor. Şekil 3.1.1.1'de görüldüğü gibi lineer artmadan biraz sapıyor olsalar da, bir doğruya fit edilebiliyorlar. Alt dizi uzunluklarının başlangıçta $U_g(l)$ ile gösterdiğimiz Gaussien bir dağılımdan seçildiği durum ile, dizi ayraçlarının *p* olasılığı ile gelişigüzel serpiştirildiği, ve böylece dizi uzunluğunun $U_e(l) = \exp(-l/\xi)$, $\xi = 1/|\ln(1-p)|$ olduğu durumlar arasında doğal olarak geçici bir süre için bile olsa farklar gözleniyor. Artış hızlarına bakıldığında eksponansiyel dağılım için daha büyük bir hız olduğunu görüyoruz. Yani PDM1 alt modelinde *L*'nin artma hızı UDM1'e göre daha fazla oluyor, çünkü PDM1'de duplikasyona uğrayan alt diziler, UDM1'dekilere göre çok daha uzundurlar.

Duplikasyonla birlikte M2 mutasyon seti, her iki başlangıç dağılımı içinde L'nin zamanla davranışında, ilk zaman adımlarında lineer davranıştan sapıyor olsa da, daha sonraki zaman adımlarında, aynı artma hızı ile büyümeye neden olmaktadır.

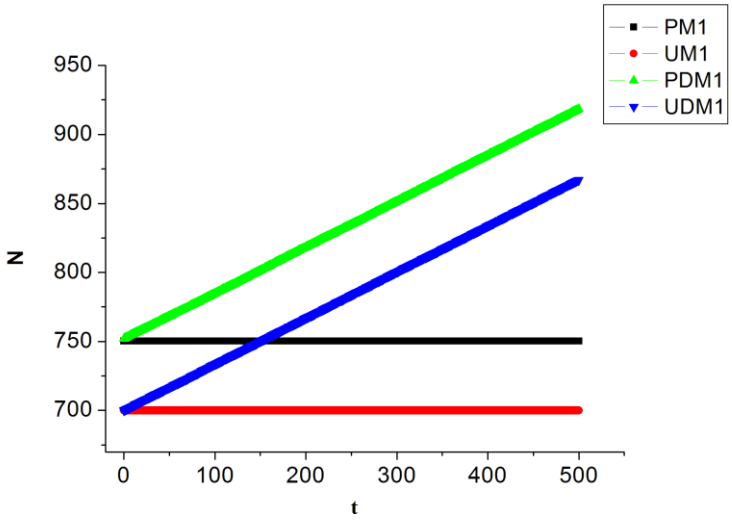
Tablo 3.1.1.1 Dizilim uzunluğu *L*'nin , zamanla değişim hızlarının değerleri.

	PM2	UM2	PDM1	UDM1	PDM2	UDM2
dL / dt	-0.078	-0.081	23.71	17.86	5.51	5.51

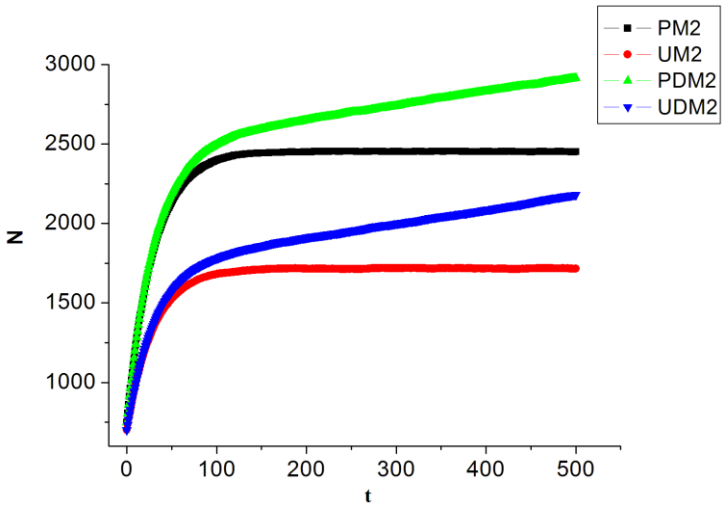
3.1.2 Alt Dizi Sayısı (*N*)

Alt dizi sayısı, ayraç olan 2'lerin dizilimdeki sayısı idi. Duplikasyon olmadığında, M1 ayraçların sayısını değiştirecek bir etken içermediğinden şekil 3.1.2.1 (a)'da da görüldüğü gibi, her iki başlangıç dağılımı için, yani PM1 ve UM1'de alt dizi sayısı sabit kalmaktadır. Duplikasyon işlemi ile, tablo 3.1.2.1'den de görüldüğü gibi, PDM1 ve UDM1'de alt dizi sayısı aynı hızla artıyor. Bu iki alt modelde, M1 mutasyonu alt dizi sayısını değiştirmediğinden, burada sadece duplikasyonun alt dizi sayısına olan etkisini görebiliyoruz.

Diğer mutasyon kuralı M2'nin etkisi, şekil 3.1.2.1 (b)'de görüldüğü gibi, daha farklı olarak, hem her iki başlangıç dağılımı için, hem de duplikasyon olup olmamasından bağımsız olarak, t=150'ye kadar alt dizi sayısını hızla arttırmak oluyor. Bu zaman adımına kadar duplikasyondan bağımsız olarak, PM2 ve PDM2 aynı artışı göstermektedir. Benzer durum UM2 ve UDM2 içinde geçerli olmaktadır. Bu zamandan sonra, duplikasyon olmadığında alt dizi sayısı çok çok küçük bir hızla artmaya devam ederken, duplikasyon olduğunda, daha hızlı biçimde *N* artmaya devam ediyor. Artma hızı PDM2 ve UDM2 için de hemen hemen aynı olmaktadır. Artma hızlarının t=150'den sonrası için hesaplanan değerlerini Tablo 3.1.2.1'den okuyabiliriz.



(a)



(b)

Şekil 3.1.2.1 Dizilim uzunluğu *N*'nin zamanla değişimi.

Tablo 3.1.2.1 Alt dizi sayısı *N*'nin, zamanla değişim hızlarının değerleri.

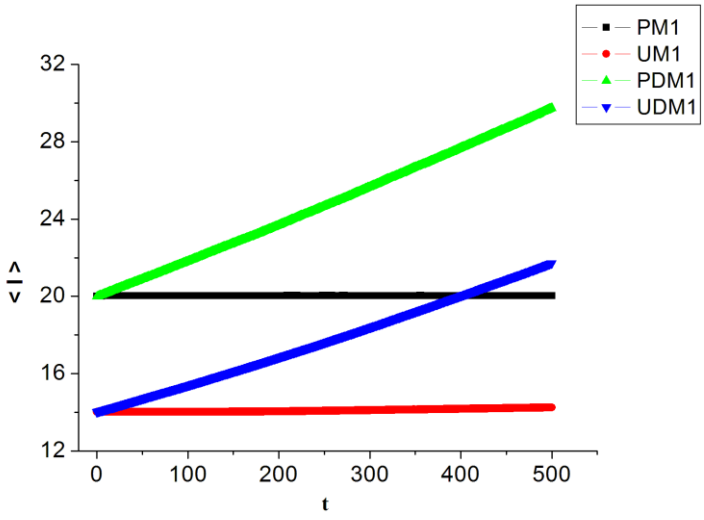
	PDM1	UDM1	PM2	UM2	PDM2	UDM2
dN / dt	0.332	0.332	0.008	0.005	0.899	0.898

3.1.3 Ortalama Alt Dizi Uzunluğu ($\langle l \rangle$)

Zaman içinde alt dizilerin ortalama uzunlukları, duplikasyon ve mutasyonlar sonunda $\langle l \rangle = (1-p)/p$ aynı kalmayıp değişmektedir. Belli bir gerçekleşim için, ortalama alt dizi uzunluğunu, tüm alt dizilerin uzunluklarını toplayıp, sıfır uzunluğunda olmayan alt dizilerin sayısına bölerek (3.1.3.1)'de de görüldüğü gibi hesaplıyoruz.

$$\langle l \rangle = \frac{\sum_{i=1}^{N} l_i}{N - N_{l=0}} \qquad (3.1.3.1)$$

Simülasyona baslamadan önce, t=0 anı için, eksponansiyel ve Gaussien dizi uzunluğu dağılımları için analitik olarak $\langle l \rangle$ değerlerini biliyoruz. $P(x)$ ile oluşturulan dizilim için $\langle l \rangle = (1-p)/p$'den $\langle l \rangle_{U_e(l)}$=19, $U_g(l)$ ile oluşturulan dizilim için de, dağılımın kendisinden gördüğümüz gibi $\langle l \rangle_{U_g(l)}$=15 olur. Simülasyon ile ilk dağılımları oluşturduğumuzda, şekil 3.1.3.1 ve şekil 3.1.3.2 den t=0 zamanına baktığımızda $U_e(l)$ ve $U_g(l)$ için ortalama alt dizi uzunluklarının, analitik değerlere yakın olarak, sırası ile ~20 ve ~14 olduğu görülmektedir.

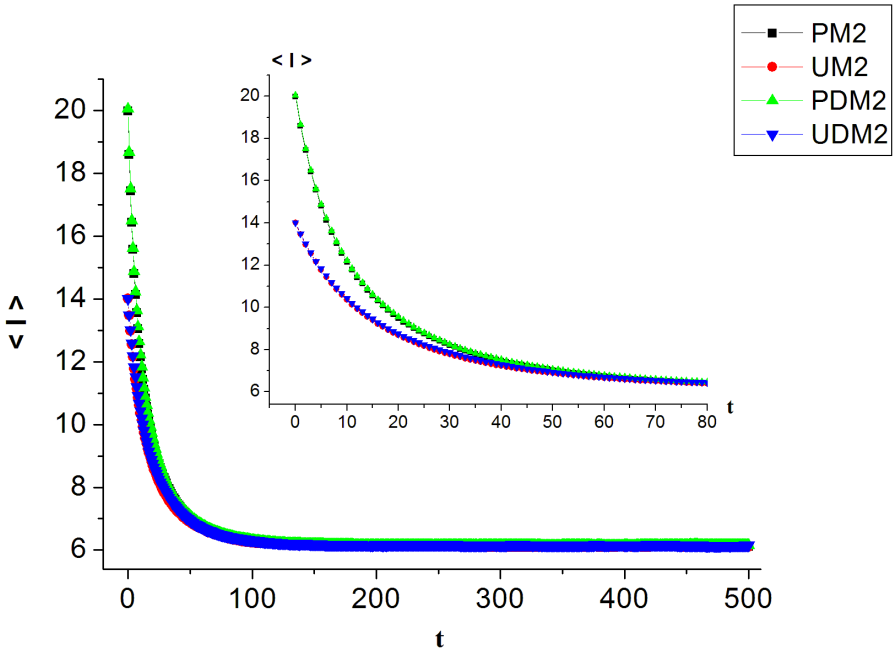


Şekil 3.1.3.1 M1 kuralı ile $\langle l \rangle$'nin zamanla değişimi.

Ortalama alt dizi uzunluğunun, yani $\langle l \rangle$'nin zamanla değişiminde bakıldığında, iki mutasyon kuralının belirleyici bir rol oynamakta olduğunu görüyoruz. M1, duplikasyon olmadığında, yani PM1 ve UM1 durumlarında $\langle l \rangle$'nin değerini zamanla değişmiyor diyebiliriz. Sadece $N_{l=0}$ sayısındaki ufak değişimler nedeni ile, PM1 ve UM1 için çok küçük bir değişim hızı bulunuyor. M1 mutasyon kuralı ile birlikte, duplikasyon uygulandığında, beklenildiği gibi, $\langle l \rangle$ için, PDM1'deki artış, UDM1'e göre daha hızlı (yaklaşık iki katı) oluyor. Çünkü t=0'da $U_e(l)$ dağılımı, $U_g(l)$'ye nazaran daha uzun alt diziler içeriyor ve dizilime katılan daha uzun alt diziler $\langle l \rangle$ 'yi arttırıyor.

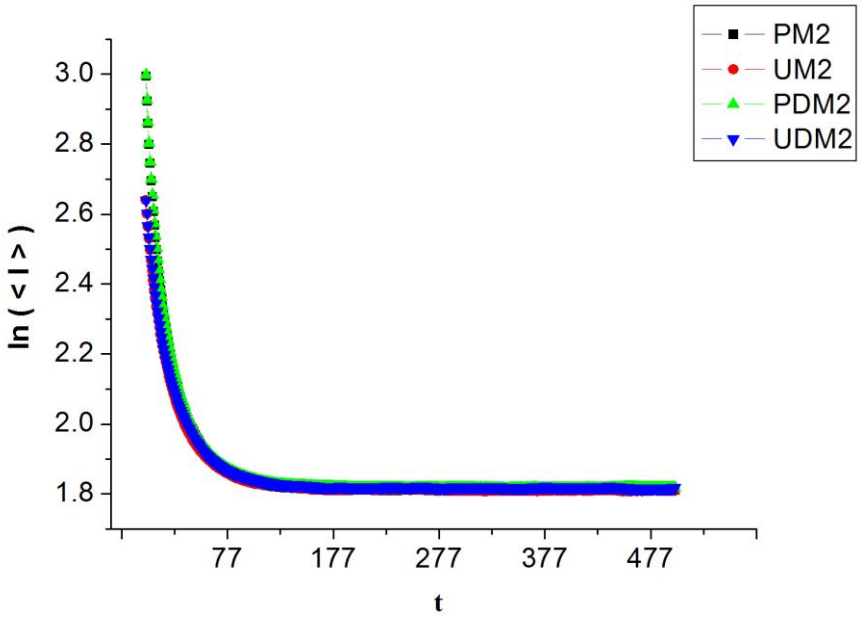
Tablo 3.1.3.1 Ortalama alt dizi uzunluğunun, $\langle l \rangle$, M1'li alt modellerde, zamanla değişim hızları.

	PM1	UM1	PDM1	UDM1
d $\langle l \rangle$ / dt	5.08×10^{-5}	5.08×10^{-5}	0.029	0.014



Şekil 3.1.3.2 M2 kuralı ile $\langle l \rangle$'nin zamanla değişimi.

Şekil 3.1.3.2'de görüldüğü gibi, M2 mutasyonu, duplikasyonlu ya da duplikasyonsuz, t=0 da her iki dağılımla da başlayan koşullarda, ortalama alt dizi uzunluğunun değerini t≈150'e kadar eksponansiyel bir biçimde azaltıp, daha sonraki tüm zaman adımlarında $\langle l \rangle \approx 6$ değerinde sabit tutmaktadır. Bunların doğal logaritmasını aldığımızda (şekil 3.1.3.3) $\langle l \rangle$'deki azalmanın eksponansiyel biçimde olmadığını görüyoruz.



Şekil 3.1.3.3 ln($\langle l \rangle$)'nin grafiği.

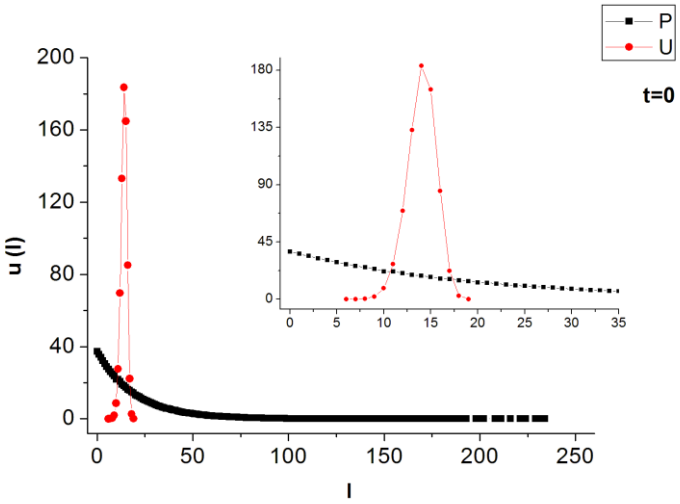
M1 mutasyonu alt dizilerin ayraç sayısında bir değişim yapmadığından, duplikasyon herhangi bir l uzunluğundaki alt dizinin, 2l uzunluklu olmasına neden olabiliyor ve $\langle l \rangle$ artıyor. Buna karşılık M2 uzunluk değişimlerini dengeleyerek $\langle l \rangle$'nin zaman içinde sabit kalmasına neden oluyor.

3.1.4 Dizi Uzunluklarının Dağılımı ($u(l)$)

Aşağıda, şekil 3.1.4.1'de t=0 anında, $U_e(l)$ ve $U_g(l)$ ile oluşturduğumuz alt dizilerin, simülasyonlar sonunda bulunan $u(l)$ grafiklerini görüyoruz. Eksponansiyel uzunluk dağılımı için (2.1.1.4) denleminde analitik olarak hesaplanan $u(l)$, ayraçların

yerleştirilme olasılığına yani *p*'ye bağlı idi. $U_e(l)$ olasılık dağılımı için zaten p'yi baştan biliyoruz.

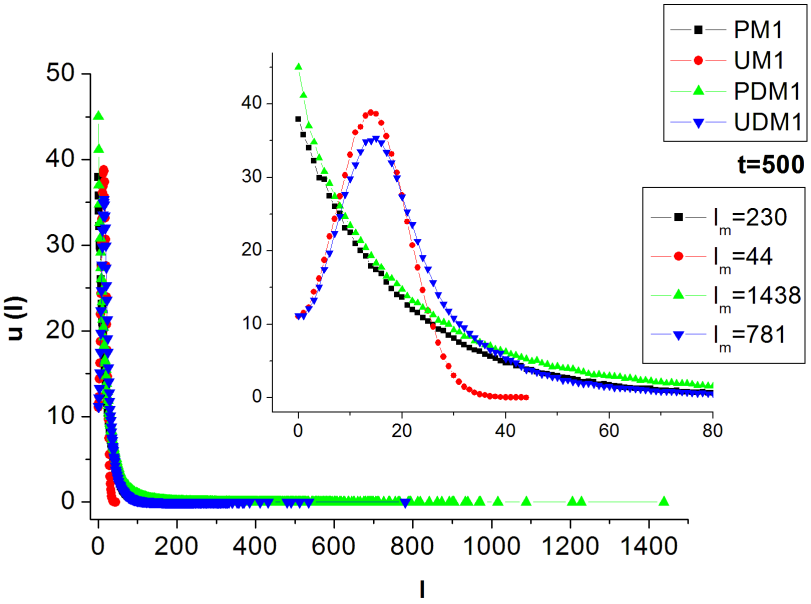
İkinci bölümdeki (2.1.1.1) bağıntısındaki gibi, verilen bir *p* olasılığına bağlı olarak dizilimi oluşturmasak da, elde edilen uzunluk dağılımı eksponansiyel olduğu takdirde bu grafiğin doğal logaritması alındığında çıkan eğimden, bir tür etkin *p* değeri hesaplamamız mümkün olmaktadır.



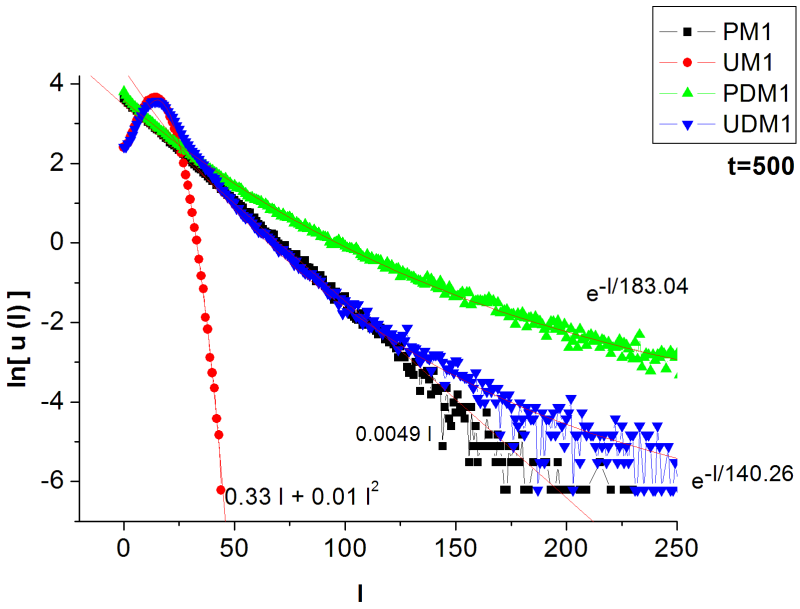
Şekil 3.1.4.1 Başlangıç alt dizi uzunluk dağılımları.

Ağ yapısının evrimleşme kuralları altında uzun bir süre devindirilmesinden sonra, alt dizilerin uzunluk dağılımları, her iki model için de eninde sonunda eksponansiyel bir dağılıma gitmektedir. M1 bunu M2'ye göre daha yavaş yapmakla birlikte, çok daha uzun zaman adımı sonrasında, her iki mutasyon kuralının da Gaussien dağılımı eksponansiyel bir dağılıma ulaştırdığını gördük. Şekil 3.1.4.2'de görüldüğü gibi, UM1 ve UDM1 için t=500'de Gaussien dağılımın yüksekliği azalmış ve eksponansiyel bir kuyruk oluşmaya başlamıştır. Duplikasyon ile daha uzun alt diziler oluştuğundan UDM1'deki eksponansiyel kuyruk, UM1'e göre daha büyük *l* değerlerine doğru uzanabilmiştir.

Model I'de eksponansiyel uzunluk dağılımına gitme yavaş olduğundan etkin bir p değeri, t=500 zaman adımı sonrasında, UM1, PDM1 ve UDM1 alt modelleri için elde edilememekle birlikte şekil 3.1.4.3'de $u(l)$'nin nasıl davrandığı incelenmeye çalışılmıştır.

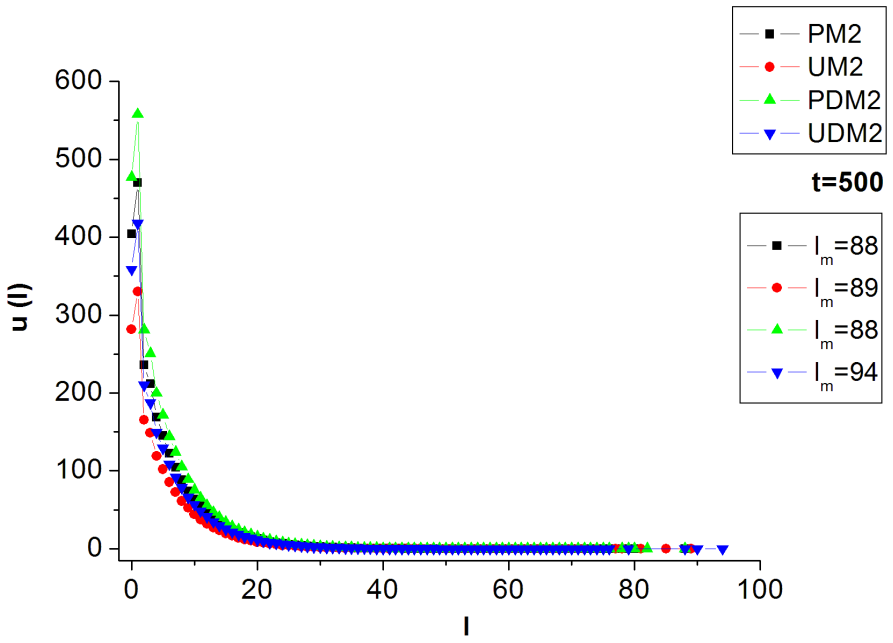


Şekil 3.1.4.2 Model I için alt dizi uzunluk dağılımlarının devinimi, t=500.

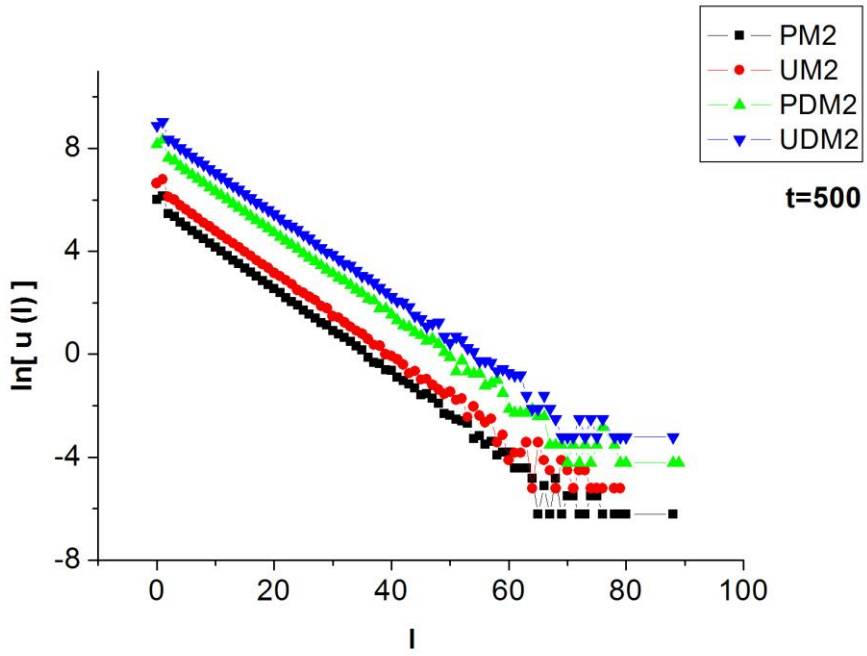


Şekil 3.1.4.3 Şekil 3.1.4.2'te görülen fonksiyonların kuvadratik ve eksponansiyel foksiyonlar aracılığı ile incelenmesi.

Bu zaman adımı sonrasında Şekil 3.1.4.4'de de görüldüğü gibi Model II'deki tüm alt modeller $e^{-0.15l}$ olan aynı eksponansiyel dağılıma ulaştığından (şekil 3.1.4.5), etkin bir p değeri hesaplama olanağı vermektedir. Şekil 3.1.4.4 için p_{etkin}=0.14 olmaktadır. $P(x)$ dağılımına nazaran yaklaşık üç kat daha fazla olasılıkla ayraçları yerleştirmekte, $<l>$'nin azalma davranışını da doğrulamaktadır. Aşağıdaki Model I ve Model II sonuçlarını birer grafikte gösterdik, tam seçilemeyen l'nin aldığı en büyük değerleri belirttik.



Şekil 3.1.4.4 Model II için alt dizi uzunluk dağılımlarının devinimi, t=500.



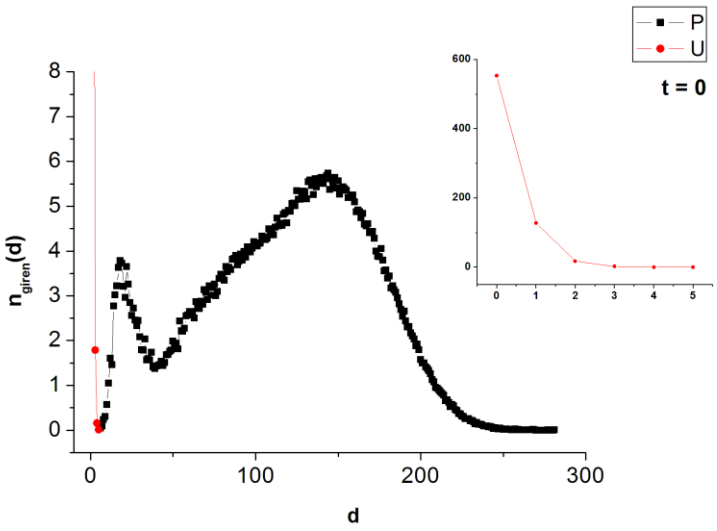
Şekil 3.1.4.5 Model II için ln($u(l)$)'nin grafiği. (Bulunan ln($u(l)$) sonuçları UM2 için bir, PDM2 için iki ve UDM2 için üç birim y ekseni yönünde kaydırılmıştır.)

3.2 Derece Dağılımları

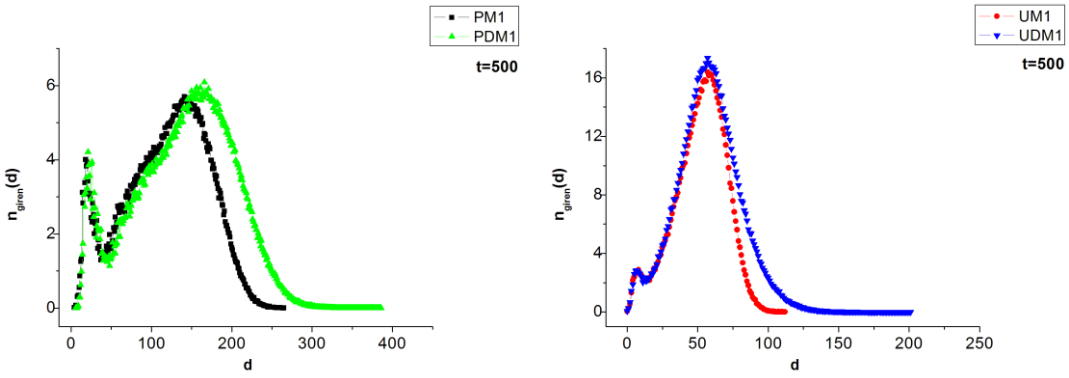
Dizilimdeki bir alt dizi diğerinin içinde aynen varsa, bu alt diziden içinde olduğu alt diziye doğru yönlü bir bağ kuruluyordu. Bir alt diziden çıkan bağların sayısı, onun çıkan derece sayısını ve giren bağlarının sayısı da giren derece sayısını veriyordu. Bağların yönelimden bağımsız olarak sayılması ile de toplam derece sayısı bulunuyordu. Her bir derece sayısına, kaç tane alt dizinin sahip olduğu sayılarak da ilgili derece dağılımları hesaplanır. Bu bölümde toplam, giren ve çıkan derece dağılımlarının simülasyonlar sonunda bulunan davranışlarına bakacağız.

3.2.1 Giren Derece Dağılımı

Şekil 3.2.1.1'de t=0 anında iki başlangıç koşulu ile elde edilen giren derece dağılımlarını görüyoruz.



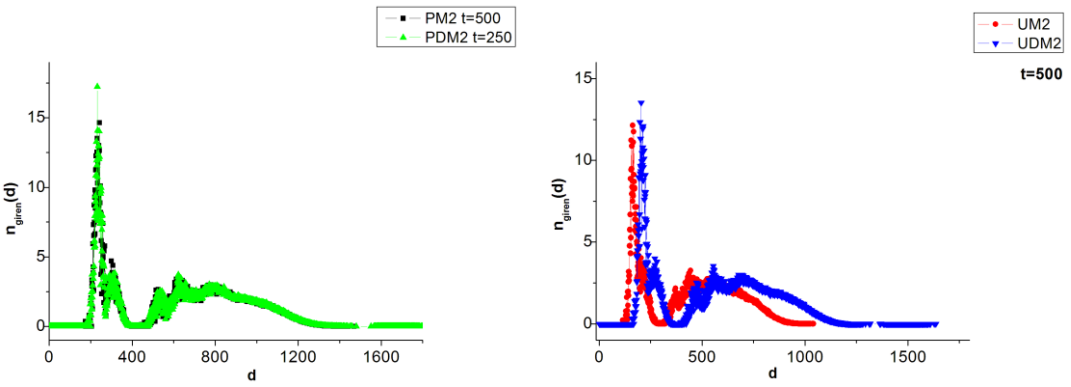
Şekil 3.2.1.1 t=0'da $U_e(l)$ ve $U_g(l)$ için $n_{giren}(d)$grafiği.



(a) (b)

Şekil 3.2.1.2 Model I için d-$n_{giren}(d)$ grafikleri, t=500.

Model I için elde edilen "giren derece" dağılımları, Balcan-Erzan modeli için elde edilenlerle uyum göstermektedir. Burada da giren derece dağılımları ölçeklenme davranışı göstermemektedir.



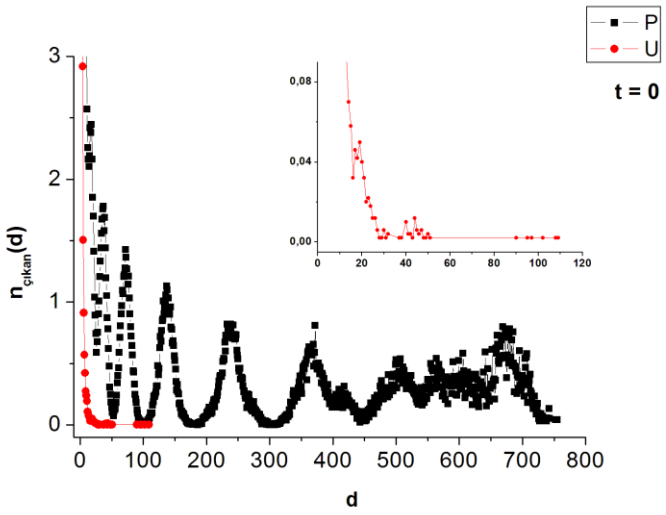
(a) (b)

Şekil 3.2.1.3 Model II için d-$n_{giren}(d)$ grafikleri.

Model II'de giren derece dağılımı için elde ettiğimiz dağılımlardaki ilk tepecikler dizisi, daha sonraki nispeten daha geniş tepeye nazaran daha büyük önem kazanmış, bu yayvan tepecikten daha belirgin bir biçimde ayrılmış bulunmaktadır.

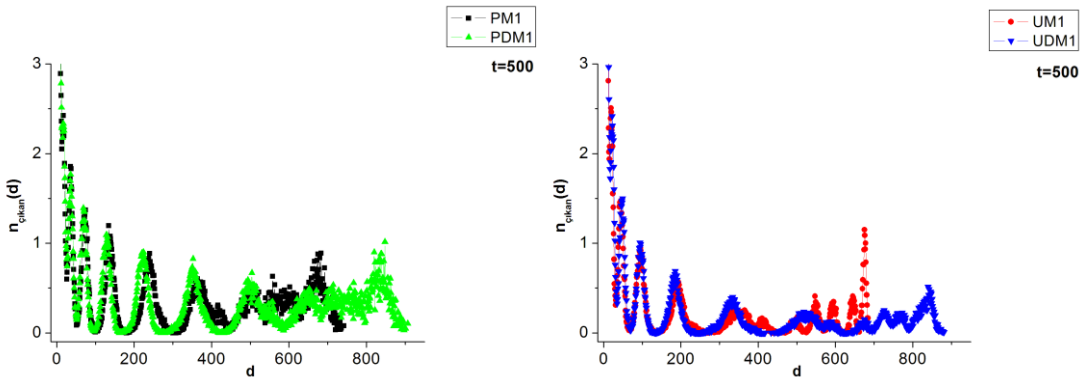
3.2.2 Çıkan Derece Dağılımı

Dizilim ilk olarak $U_e(l)$ ve $U_g(l)$ dağılımları ile oluşturulduğunda, t=0'da, çıkan derece sayıları birbirinden farklı olmaktadır. $U_e(l)$ dağılımı sonucunda çıkan derece sayıları belli noktalarda maksimumlar veren periyodik bir yapı sergilemekte ve alt dizilerin yaklaşık olarak yarısı, diğer alt dizilerin içinde bulunmaktadır ($n_{çıkan}(d=0)$=363 ve N=750). Buna karşılık $U_g(l)$ dağılımı sonucunda, büyük d değerlerine sahip alt diziler hiç olmayıp küçük d değerlerinde de çok az alt dizi bulunmaktadır ($n_{çıkan}(d=0)$=606 ve N=700). Beklenilen bu sonuç, eksponansiyel uzunluk dağılımında l küçük değerler alabilirken (1 ve 2 gibi), Gaussien uzunluk dağılımında en küçük l değerinin 5 olmasından kaynaklanmaktadır. Çünkü büyük d değerlerine katkılar her alt dizi içinde bulunabilen kısa alt dizilerden gelmektedir.



Şekil 3.2.2.1 t=0'da $U_e(l)$ ve $U_g(l)$ için $n_{çıkan}(d)$ grafiği.

Yukarıdaki şekillerde eksponansiyel başlangıç dağılımları ile oluşturulan diziler için sonuçlar, [1] 'de bulunanlarla karşılaştırıldığında aynı oldukları görülecektir.

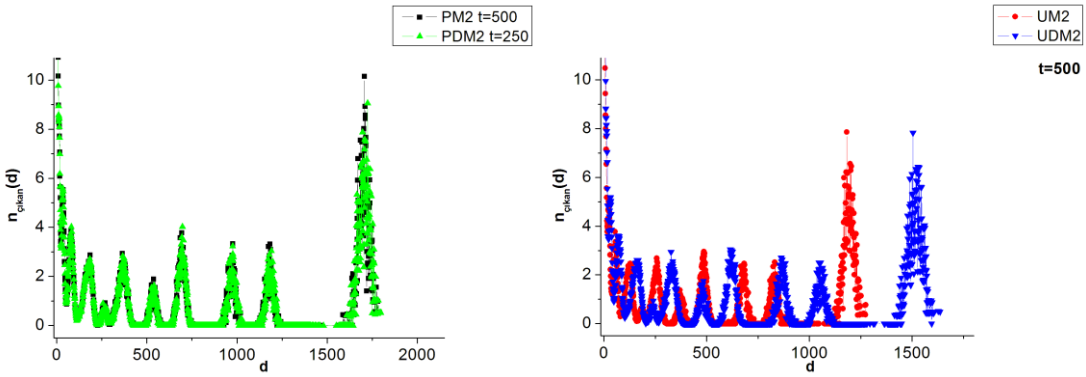


(a) (b)

Şekil 3.2.2.2 Model I için için $n_{çıkan}(d)$ grafikleri, t=500.

Şekil 3.2.2.2'te Model I için çıkan derece dağılımları görülmektedir. Eksponansiyel ve Gaussien başlangıç dağılımları 500 zaman adımından sonra kalitatif olarak benzer davranış göstermektedirler. Dikkat edilecek olursa, (b) şıkkında duplikasyonların olmadığı (UM1) grafiğinde, büyük derece değerleri için çok keskin görünen tepecikler, L'nin sonlu olduğu durumlarda bekleyeceğimizden [2] daha belirgindirler. Bu, hala eksponansiyel olmayan mevcut dağılımda, görece uzun dizilerin sonlu L içinde de yeterince çok sayıda örneklendiğini, böylece gerçekleşimden gerçekleşime

dalgalanmaların daha düşük olduğunu göstermektedir. Duplikasyonlar ise (UDM1) gerçekleşimler arası dalgalanmaları arttırarak bu tepeciğin yayılmasına neden olmaktadırlar.



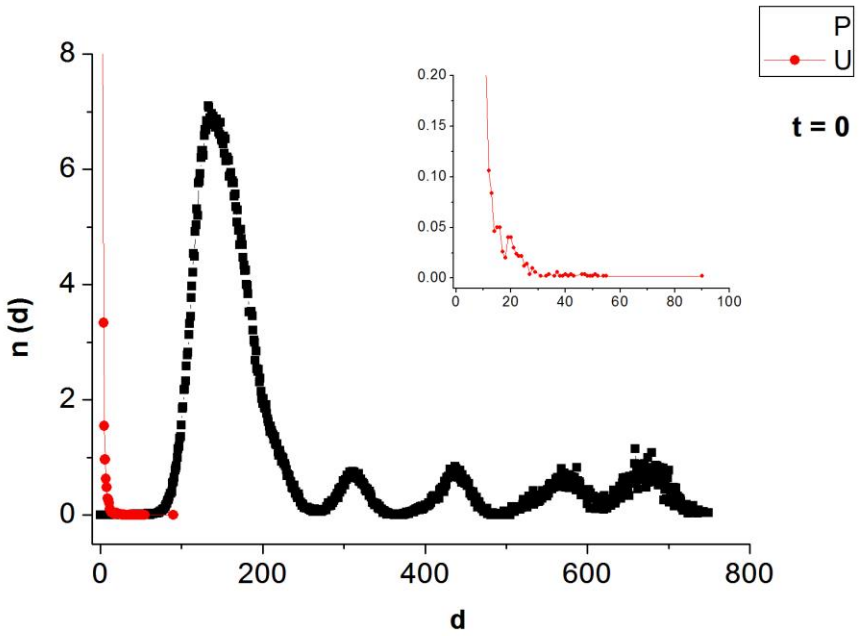
(a) (b)

Şekil 3.2.2.3 Model II için $n_{çıkan}(d)$ grafikleri.

Bu grafiklerde, yukarıda sözü edilen "sonlu gerçekleşim etkisi"ndeki [2] azalma daha da belirgin olarak ortaya çıkmaktadır. Her bir tepeciğin farklı bir *l* uzunluğunda alt dizinin diğerleri içinde bulunmasından kaynaklandığını [1,2] ve en sağdaki tepeciğin de 1 uzunluğundaki alt dizilerden geldiğini göz önüne alacak olursak, bu tepecikleri herhangi sonlu gerçekleşim etkisi içermeyen, sonsuz uzunluktaki bir "ideal " gerçekleşim için beklentilerimiz ile karşılaştırmakta yarar vardır. [39] Sola doğru gidildiğinde, şekil 3.2.2.3'te gözlenen tepeciklerin, şekil 3.2.2.1 ile kıyaslandığında "shift-match number" (kaydırma-çakışım sayıları)na göre birden çok tepeciğe yarılmaları söz konusudur (Bkz. Ek A).

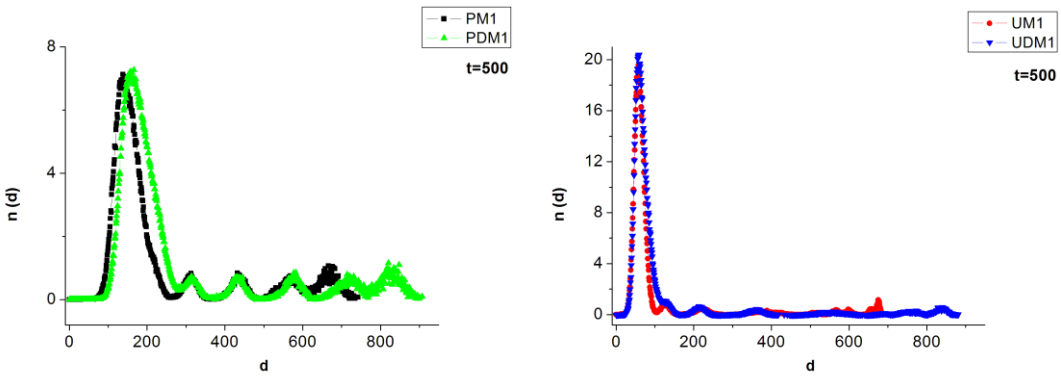
3.2.3 Toplam Derece Dağılımı

Ağ üzerindeki noktaların toplam derece sayılarının dağılımı t=0'da şekil 3.2.1.1'deki gibi olmaktadır. Başlangıç koşulu Gaussien olduğunda pek yapı gözlenmezken, eksponansiyel başlangıç koşulu sonunda, toplam derece sayısı *d*~150 civarında en büyük değerini almaktadır. Sonraki derece sayılarında maksimum ve minimumların birbirini izlediği periyodik yapı görülmektedir.

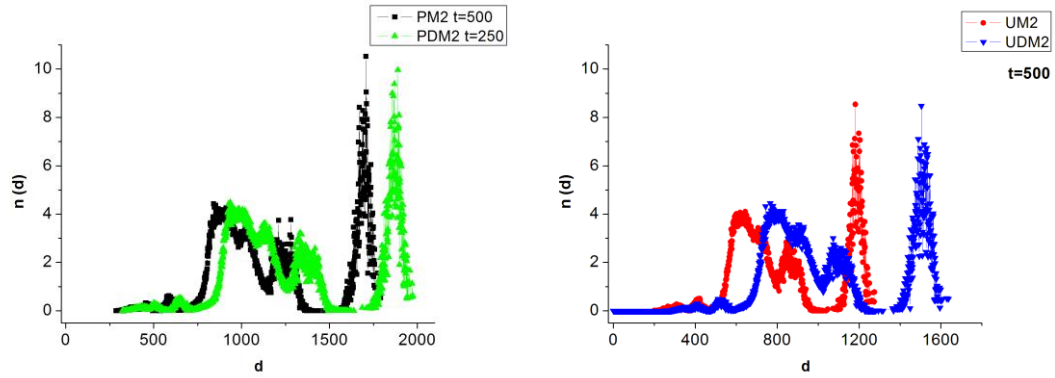


Şekil 3.2.3.1 t=0'da $U_e(l)$ ve $U_g(l)$ için $n(d)$grafiği.

Şekil 3.2.3.2 ve şekil 3.2.3.3'de görüldüğü gibi giren derece dağılımları büyük *d* bölgesinde pek yapı göstermediğinden, bu bölgede toplam derece sayılarına katkıların çıkan derece sayılarından geldiği görülmektedir.



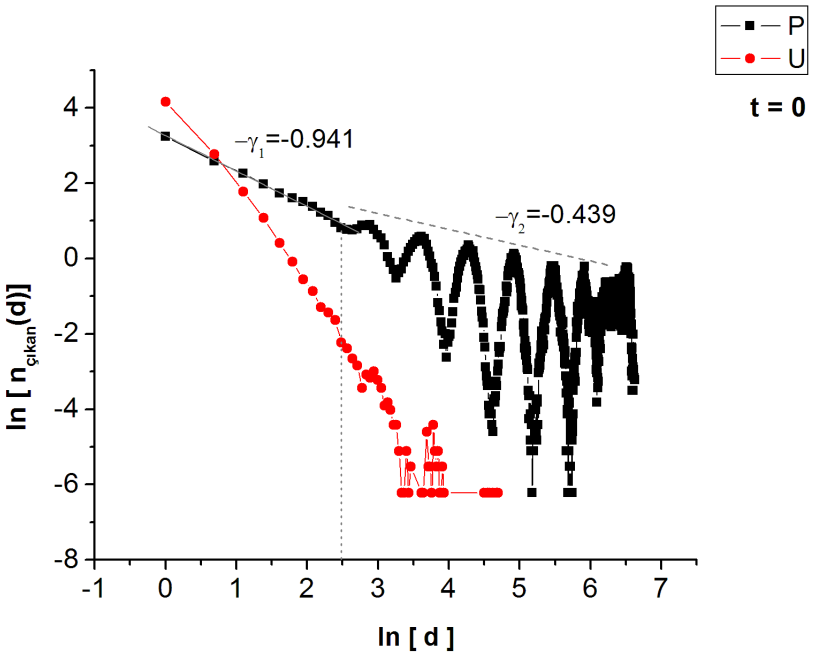
Şekil 3.2.3.2 Model I için *d*-$n(d)$ grafikleri, t=500.



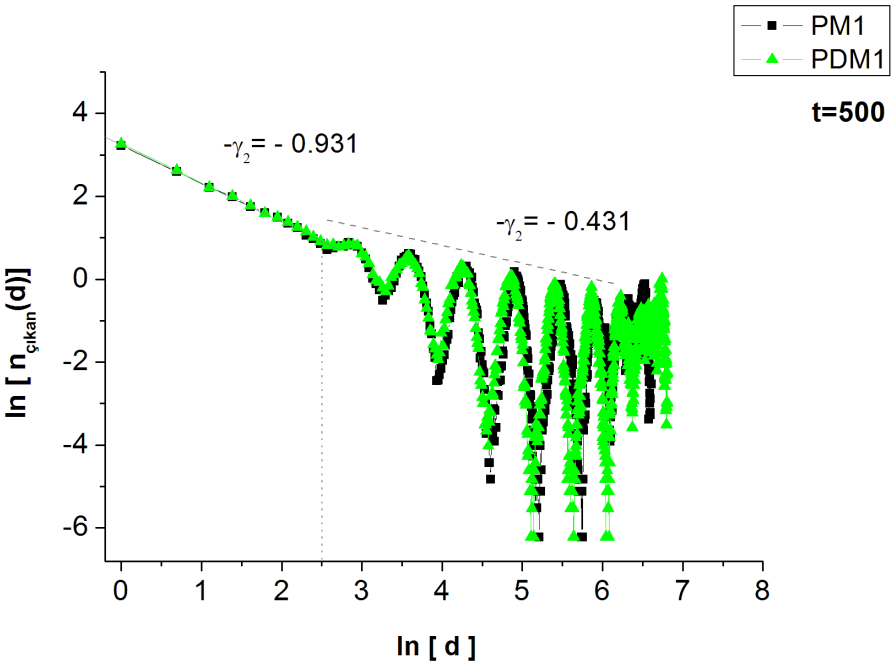
Şekil 3.2.3.3 Model II için *d*-*n*(*d*) grafikleri, t=500.

3.3 Çıkan Derece Dağılımının Ölçeklenme Davranışı

Bu modelde, sadece çıkan derece dağılımları bir ölçeklenme göstermektedir. Grafiklerin doğal logaritması alındığında bu kolayca görülebilmektedir. Aşağıda t=0 zamanında, eksponansiyel başlangıç koşulu için küçük *d*'lerin olduğu bölgede bulunan dağılım bir üstelle ölçeklenmektedir ($n_{çıkan}(d) \sim d^{-0.941}$). Büyük *d* bölgesindeki periyodik davranış da başka bir üstelle sınırlandırılabilmektedir ($n_{çıkan}(d) \sim d^{-0.439}$). Bu sonuçlar, alınan parametre değerleri için Balcan-Erzan çalışmasının [1] sonuçlarını doğru olarak tekrar vermektedirler. Gaussien dağılımla başlandığı durumda ise kısa alt dizilerin az olması nedeni ile büyük derece sayılarındaki kesikli davranışın gözlendiği bölgede çok az yapı sergilenmektedir. Hatta tüm dağılım $\gamma_1 \approx 2$, gibi tek bir üstelle ölçeklenebilir görünmektedir.



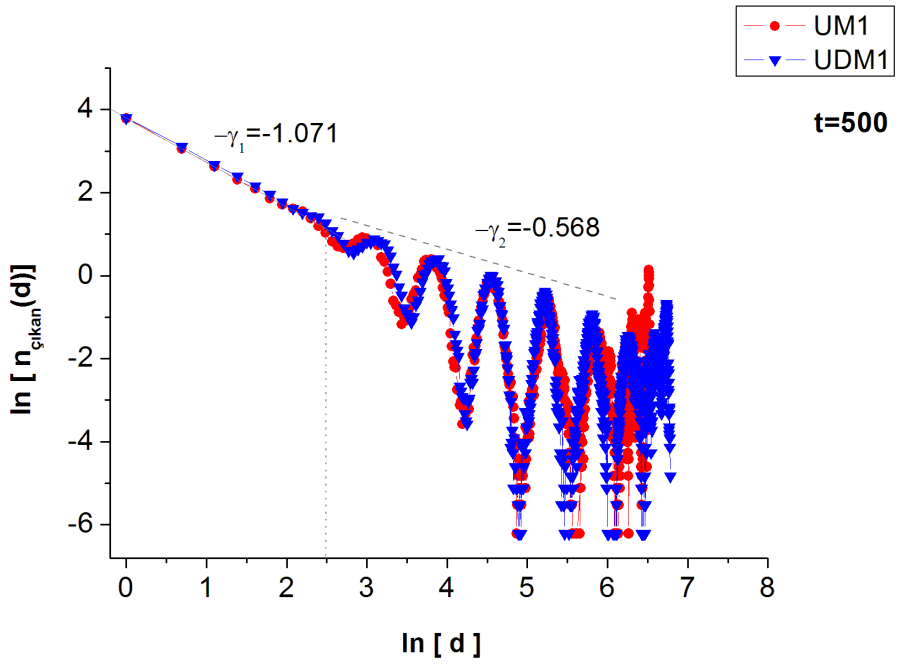
Şekil 3.3.1 t=0'da iki başlangıç dağılımı için log-log ölçeğinde $n_{çıkan}(d)$ grafiği.



Şekil 3.3.2 t=500'de PM1 ve PDM1 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği.

Şekil 3.3.2 de, daha önce bulunmuş olduğu gibi M1 nokta mutasyonları altında Referans [1]'in sonuçlarının değişmediği gibi, duplikasyonlar eklendiğinde de sonuçların, hata payları içinde aynı kaldıkları gözlenmektedir.

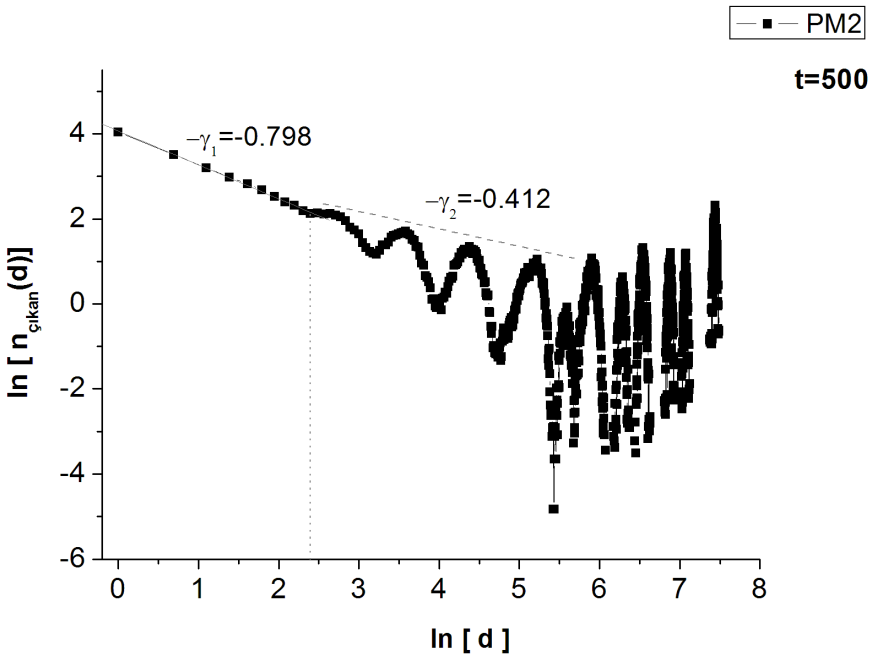
Tablo 3.2.2.1'den görüleceği gibi, Gaussien başlangıç dağılımı için elde edilen eğimler eksponansiyel dağılımdan elde edilene sayısal olarak çok yakın çıkmaktadır.



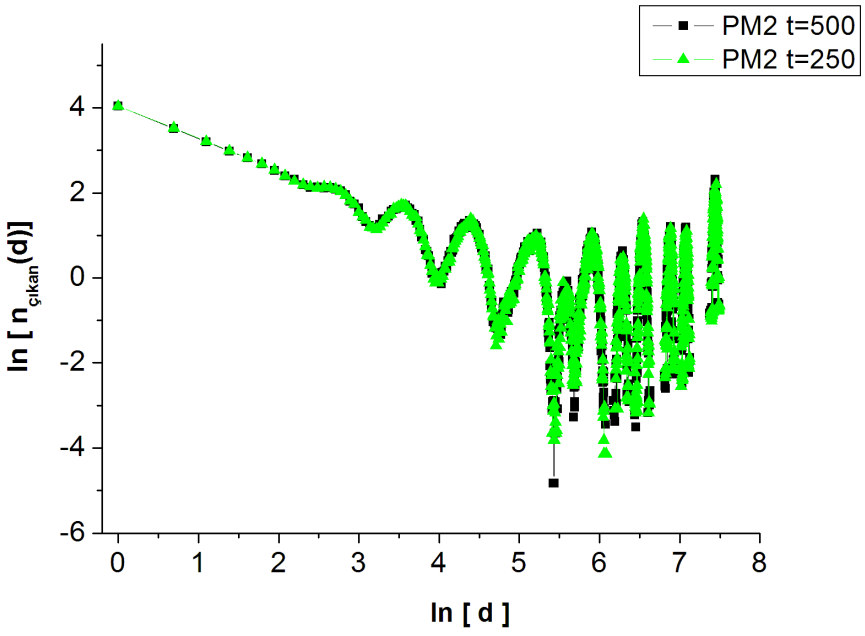
Şekil 3.3.3 t=500'de UM1 ve UDM1 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği.

Balcan-Erzan modelinin [1] sonuçları, başlangıç dağılımı olarak eksponansiyel değil Gaussien bir dağılım alındığınıda da, duplikasyona izin verilsin verilmesin, $\gamma_1 \cong 1$, $\gamma_2 \cong 1/2$ değerlerini korumaktadırlar.

Model II'ye geçildiğinde, gayet kararlı bir biçimde $\gamma_1 \cong 1$ civarında olan başlangıç eğiminin belirgin bir biçimde azaldığı, küçük derecelerde, ayrık derece dağılımının gözlendiği bölgede ise, en sağdaki tepeciklerin sönümlenmek yerine *d* ile büyüme eğilimlerinin arttığını görüyoruz.

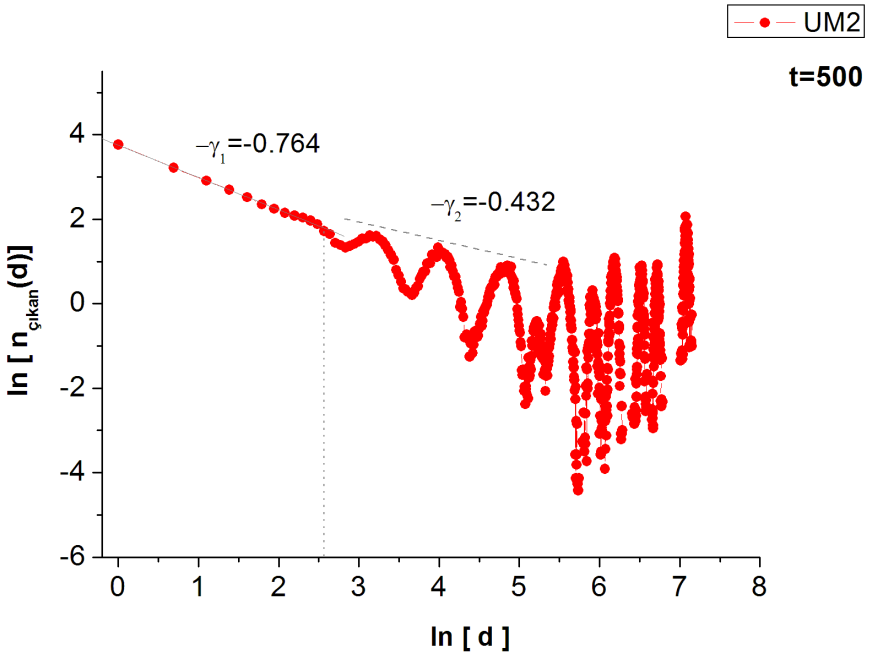


Şekil 3.3.4 t=500'de PM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği.

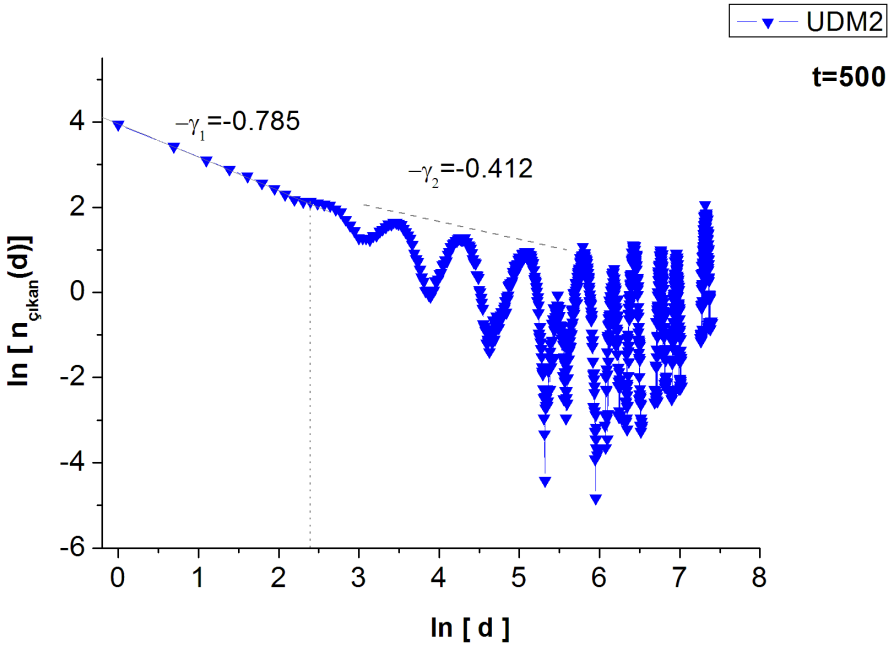


Şekil 3.3.5 t=500'de PDM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği.

Model II'de M2 mutasyon protokolu ve duplikasyonlara karşın, sürekli ve ayrık ölçümlenme bölgelerinin göreli uzunlukları muhafaza olunmaktadır.



Şekil 3.3.6 t=500'de UM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği.



Şekil 3.3.7 t=500'de UDM2 için log-log ölçeğinde $n_{çıkan}(d)$ grafiği.

Yukarıdaki şekillerden görüldüğü gibi derece dağılımının ölçeklenme davranışı duplikasyona izin verilip verilmemesinden çok, mutasyon kurallarına bağlı çıkmaktadır.

Aşağıdaki iki tabloda Model I ve Model II için, uzun zaman sonra vardıkları kestirilen eksponansiyel uzunluk dağılımlarından hesaplanan etkin *p* değerlerini ilk sütunda, (2.1.2.1)'deki bağıntılarda yerine koyarak hesaplanan üstelleri 2. ve 3. sütunlarda, sağdaki sütünlarda ise simülasyonlar sonucunda bulunan üstelleri görüyoruz. BE modelinde bulunan [1,2] üstellerin sayısal değerlerinden çok uzak olmamakla birlikte, etkin *p* değerleri ile hesaplanan üstellerle simulasyon sonuçları örtüşmemektedirler. Demek ki (2.1.2.1) denkleminin çıkarsanmasında kullanılan, dizilerin istatistiksel olarak birbirinden bağımsız olduğu varsayımı, Model II'de geçerliliğini yitirmektedir.

Tablo 3.3.1 Model I için t=500'de analitik hesap ve simülasyon sonunda bulunan üsteller.

	p_{etkin}	γ_1^{etkin}	γ_2^{etkin}		$\gamma_1^{sim.}$	$\gamma_2^{sim.}$
PM1 (BE)	0.05	0.93	0.43		0.931	0.431
UM1					1.071	0.568
PDM1	0.0049	1.007	0.507		0.941	0.431
UDM1					1.071	0.568

Tablo 3.3.2. Model II için t=500'de analitik hesap ve simülasyon sonunda bulunan üsteller.

	p_{etkin}	γ_1^{etkin}	γ_2^{etkin}		$\gamma_1^{sim.}$	$\gamma_2^{sim.}$
PM2	0.147	0.813	0.313		0.798	0.412
UM2	0.141	0.819	0.319		0.764	0.432
PDM2	0.145	0.815	0.315		0.798	0.412
UDM2	0.14	0.821	0.321		0.785	0.412

3.4 Kümelenme Katsayısının Dağılımı ($C(d)$) ve Derece Dağılımlarına l Katkıları

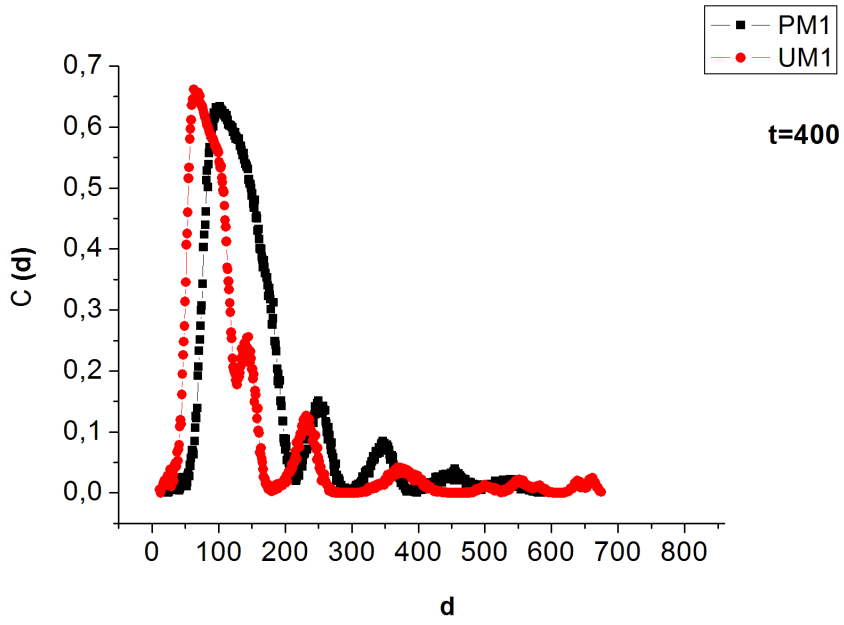
Çıkan derece sayılarının ölçeklenme üstellerinin, $\gamma_1 \cong 1$ ve $\gamma_2 \cong 1/2$ yaklaşık değerlerini koruduğu, aşağıdaki sayısal değerlerle diğer bir sümülasyon hesabı yaptık. Buradaki simülasyonlarlar bir önceki bölümdeki sonuçlar ile uyum içinde olmakta, mutasyon olasılığının farklı alınması bu PM1 ve UM1 için bulunan sonuçlarda değişikliğe yol açmamıştır. Burada kümelenme katsayısının, derece sayısına göre dağılımı olan $C(d)$'nin davranışının nasıl olduğuna ve derece sayılarına katkıların geldiği alt dizi uzunluklarının dağılımına baktık.

Tablo 3.4.1 Alınan sayısal değerler.

Gerçekleşim sayısı	L_0	N_0	p	μ
400	12000	700	0.05	0.1

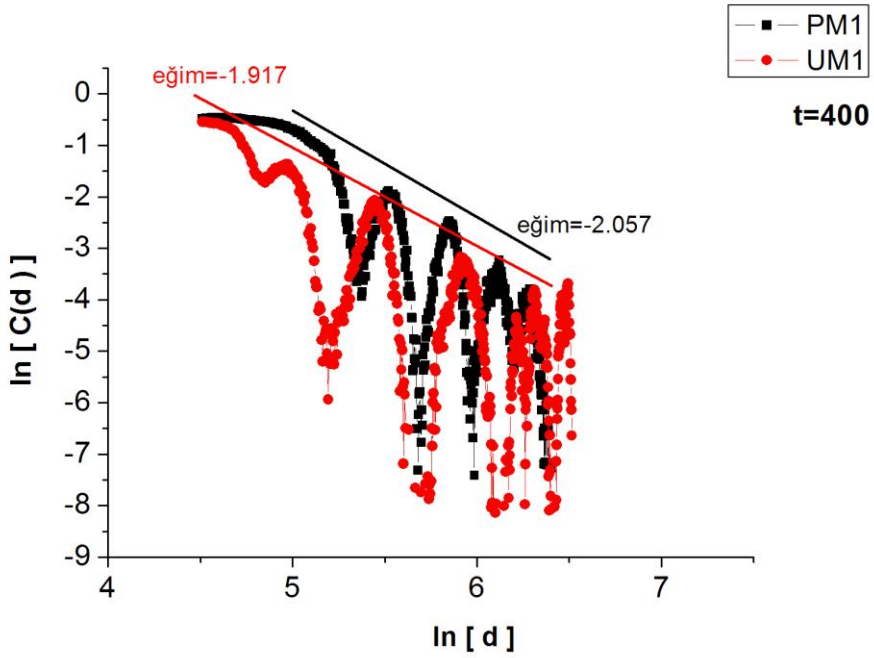
Ağ üzerindeki herhangi bir noktanın kümelenme katsayısı, bu noktanın derece sayısı d ve bağlandığı noktalar arasındaki bağ sayısı E ise $C=E/(d(d-1)/2)$ olarak hesaplanıyordu. Burada E sayısı için seçili noktadan çıkan üçgenlerin sayısıdır diyebiliriz. Simülasyonlarda toplam derece sayısına göre (yani sanki bağlar yönlü değilmiş gibi) ağ üzerindeki her bir noktanın kümelenme katsayısı hesaplandı. Daha sonra kümelenme katsayılarının derece sayısına (d) göre dağılımı bulundu.

Simülasyonlar sonunda, $C(d)$ büyük d bögesinde, çıkan derece dağılımına benzer biçimde davranmaktadır. Kümelenme katsayısı, PM1 ve UM1 için de, $d \cong 100$ civarında derece sayısına sahip olan dizilerde en büyük değerini almaktadır.



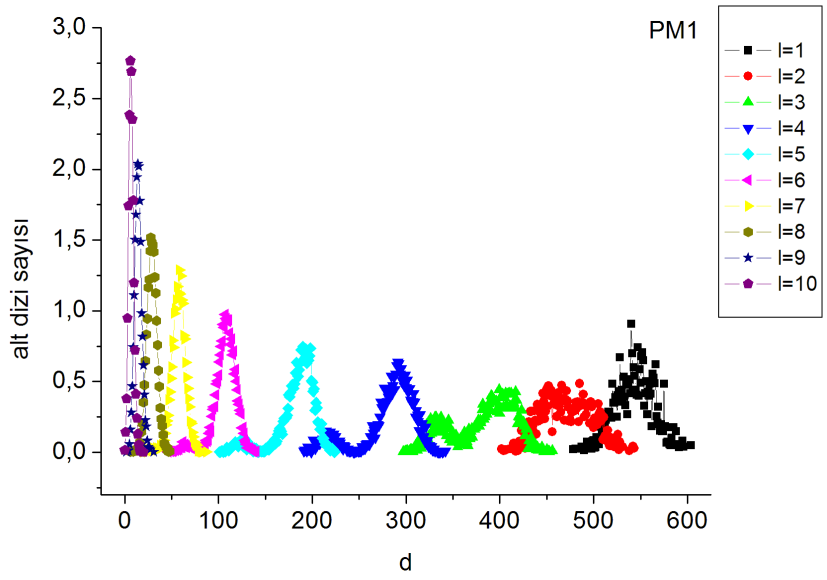
Şekil 3.4.1 t=400'de PM1 ve UM1 için C(*d*) grafiği.

PM1 ve UM1 için, büyük *d* bölgesinde periyodik davranış gösteren bölgenin zarfı, eğimi yaklaşık -2 gibi bir doğru ile sınırlandırılabilmektedir, dolayısıyla bu bölge için ölçeklenme davranışı gösterdiği söylenebilir.

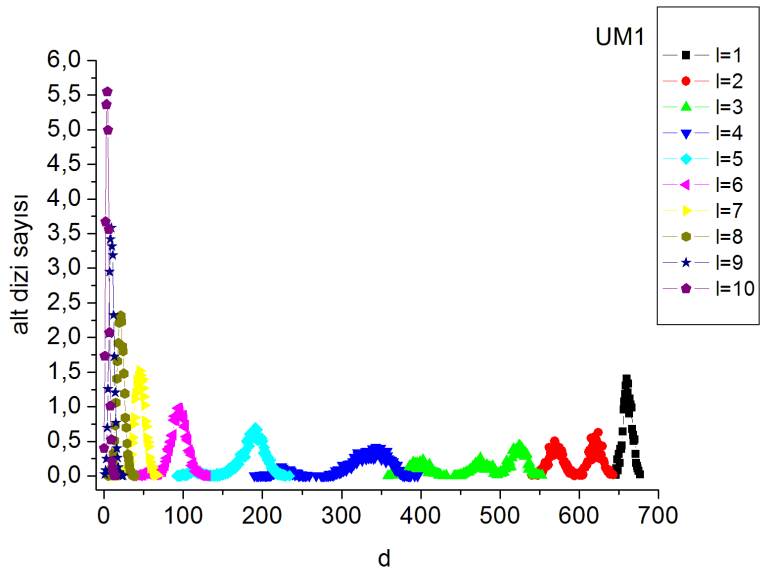


Şekil 3.4.2 t=400'de PM1 ve UM1 için C(*d*) grafiği.

Derece dağılımlarında, her bir çıkan derece sayısına (*d*) l uzunluğundaki alt dizilerin kaç tanesinden katkı geldiğine baktığımızda daha önce bahsedilen "shift-match number" ına göre yarılmaları şekil 3.4.3'de görebiliriz. Aşağıdaki simülasyon hesaplarında t=400 adımında, *d* çıkan derece sayısına, l uzunlugundaki alt dizilerden kaç tanesinden katkı geldiği sayılıyor, sonra bunların gerçekleşim sayısı üzerinden ortalaması alınıyor.



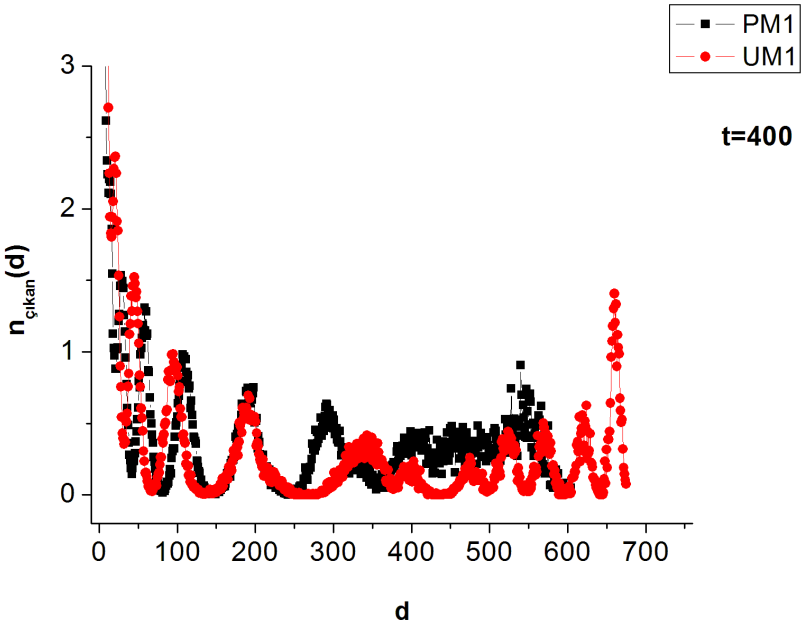
(a)



(b)

Şekil 3.4.3 t=400'de PM1 ve UM1 için l katkıları.

UM1'de l=1,2 ve 3 için yarılmalar çok daha net görünmektedir. Bu iki modelin çıkan derece dağılımına baktığımızda, şekil 3.4.3'de, yarılmalardan kaynaklanan tepecikler biraz kaynaşsada yapının bu yarılmalardan dolayı oluştuğunu anlayabiliyoruz.



Şekil 3.4.4 t=400'de PM1 ve UM1 için $n_{çıkan}(d)$ grafiği.

4. SONUÇLAR VE TARTIŞMA

Simulasyon sonuçları gösteriyor ki, duplikasyonlar ve uygulanan mutasyonlar, BE modelinde bulunan ölçeklenme davranışını kalitatif olarak değiştermese de, belli tür mutasyonlar, dizilimleri başlangıçtaki gibi gelişigüzel olmaktan çıkartıp, ağ yapısının nodları (ya da alt diziler) arasında korelasyonlara neden olmaktadır. Bu da aslında biyolojik olarak DNA dizilerinde gözlenen gerçeğe daha uygun bir durum oluşturmaktadır [40]. Ancak bir ölçüde şaşırtıcı olan şey, BE modelinden sayısal olarak en büyük sapmalara alt dizilerin duplikasyonlarının değil, ağ yapısında nodların birleşmesine ya da ayrışmasına neden olabilen M2 mutasyon protokolunun yol açmasıdır. Farklı mutasyon olasılıklarında, sonuçların hata payları içinde aynı kaldığı gözlenmiştir.

Ref. [1]'de önerilmiş olan içeriksel ağ yapılarının herhangi parametreye bağımlılık açısından çok kararlı olduklarını (p parametresine üstellerin bağlılıklarının çok zayıf olduğunu), ayrıca duplikasyon ve nokta mutasyonlar altında topolojik yapılarını koruduklarını görüyoruz. Öte yandan, farklı alt dizilerin birleşmesi ya da ayrışmasına neden olabilecek türden mutasyonlar (M2), doğrudan doğruya ağ yapısına gelişigüzel müdahalelere neden olmakta ve ölçeklenme davranışına etkin olabilmekteler. Model II'de gözlenen bu davranış, bu çalışmada tek bir mutasyon olasılığı için simüle edilmiş olsa da, bu değer değiştiğinde üstellerin değerlerinin de değişebileceğini söylemek mümkün.

KAYNAKLAR

[1] **Balcan, D. and Erzan, A.,** 2004. Random model for RNA interference yields scale free network, Eur. Phys. J. B **38**, 253-260; **Balcan, D.,** 2003. Gen İfadesinin Şebeke Modeli, *Yüksek Lisans Tezi,* İ.T.Ü. Fen Bilimleri Enstitüsü, İstanbul.

[2] **Mungan, M., Kabakçıoğlu. A., Balcan, D., Erzan, A.,** 2005. Analytical Solution of a Stochastic Content Based Network Model, q-bio.MN/0406049.

[3] **Wagner, A.,** 2001. The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes, *Mol. Biol. Evol.* 18(7),1283-1292.

[4] **Wagner, A.,** 1994. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization, *Proc Natl. Acad. Sci.* USA Vol. 91, 4387-4391.

[5] **Bollobas, B.,** 2001. Random Graphs, Cambridge University Press, United Kingdom.

[6] **Albert, R., Jeong, H. and Barabasi, A.-L.,** 1999. The diameter of the world-wide web, *Nature*, **401**, 130, cond-mat/9907038.

[7] **Huberman, B.A. and Adamic, L.A.,** 1999. Growth dynamics of the World-Wide Web, *Nature*, **401**, 131, cond-mat/9901071.

[8] **Milgram, S.,** 1967. The small world problem, *Psychology today*, **2**, 60.

[9] **Scott, J.,** 1991, Social network analysis: A handbook, SAGE Publications, London.

[10] **Banavar, J.R., Maritan, A. and Rinaldo, A.,** 1999. Size and form in efficient transportation networks, *Nature*, **99**, 130.

[11] **Shockley, W.,** 1957. On the Statistics of Individual Variations of Productivity

in Research Laboratories, *Proceedings of the IRE*, **45**, 279.

[12] **Garfield, E.,** 1972. Citation analysis as a tool in journal evaluation, *Science*, **178**, 471.

[13] **Garfield, E.,** 1979. Citation Indexing: Its Theory and Application in Science, Wiley, New York.

[14] **Camacho, J. Guimera, R. and Amaral, L.A.N.,** 2002, Analytical solution of a model for complex food webs, Phys Rev E **65**, 030901 R.

[15] **Montoya, J.M. and Sole,** R. V., 2000. Small world patterns in food webs, cond-mat/0011195.

[16] **Camacho, J. Guimera, R. and Amaral**, **L.A.N.,** 2002. Robust patterns in food web structure, *Phys. Rev. Lett* **88**, 228102.

[17] **Hopfield, J. J.,** 1982. Neural networks and physical systems with emergent collective computational abilities, *Proc. Nat. Acad. Sci.* U.S.A., 79, 2554.

[18] **Amit, D. J.,** 1989. Modeling Brain Function: The World of Attractor Neural Networks, Cambridge University Press, Cambridge.

[19] **Jeong, H., Mason, S. P., Barab_asi, A.-L. and Oltvai, Z. N.,** 2001. Lethality and centrality in protein networks, *Nature*, 411, 41.

[20] **Guet, C., Elowitz, M.B., Hsing W., Leibler, S.,** 2002. Combinatorial Synthesis of Gene Networks, *Science*, **296**, 1446-1470.

[21] **Becskei, A. and Serrano, L.,** 2000. Engineering Stability in Gene Networks by autoregulation, *Nature*, **405**, 590.

[22] **Jordan, F., Lıu, W.C. and Wyatt, T.,** Topological constraints on the dynamics of wasp-waist ecosystems, *J. Marine Systems*, basılacak.

[23] **Albert, R., Barabasi, A.L.,** 2004. Statistical Mechanics of Complex Network, *Reviews of Modern Physics* **74**, 47-49.

[24] **Dorogovtsev, S.N. and Mendes, J.F.F.,** 2002. Evolution of Networks, *Adv. Phys.* 51, 1079.

[25] **Vazquez, A., Dobrin, R., Sergi, D., Eckmann, J.-P., Oltvai, Z. N. and Barabasi, A.L.,** 2004. *PNAS* **101**, 17940 (cond-mat/0408431).

[26] **Erdos, P. and Renyi, A.,** 1960. On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.*, 5, 17.

[27] **Barabasi, A.-L. and Albert R.,** 1999. Emergence of Scaling in Random Networks, *Science* **286**, 509.

[28] **Lewin, B.,** 1997. Genes, Oxford University Press, Inc., New York.

[29] **Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, D.W.,** 1994. Molecular Biology of The Cell, Garland Publishing, Inc., New York.

[30] **Griffiths, Anthony J.F., Miller, Jeffrey H., Suzuki, David T., Lewontin, Richard C., Gelbart and William M.,** 1996**.** Introduction to Genetic Analysis, W.H. Freeman & Company, USA.

[31] **Sole, R.V. and Pastor-Satorras, R.,** 2002. Complex Networks in Genomics and Proteomics, S. Bornhodth and H.Schuster eds., Handbook of Graphs and Networks, Wiley-VCH Verlag, Berlin.

[32] **E. Szathmary, F. Jordan, and C. Pál,** 2001. Science (*Perspective*), **292**, 1315.; **Stewart, J.,** 1997. *J. Soc. Ecol. Syst.* **20**, 53.

[33] **Smith, J.M. and Szathmary, E.,** 1999. The Origins of Life, Oxford University Press, Inc., New York.

[34] **Spring, J.,** 2003. Major Transition in Evolution by Genome Fusions: from Prokaryotes to Eukaryotes, Metazoans, Bilaterians and Vertebrates, *Journal of Structural and Functional Genomics* **3**, 19-25.

[35] **Smith, J.M. and Szathmary, E.,** 1995. The Major Transitions in Evolution, Oxford University Press, Inc., New York.

[36] **Teichmann, S.A. and Babu, M.M.,** 2004. Gene Regulatory Network Growth by Duplication, *Nature* Genetics, **36**, 492-496.

[37] **Ohno, S.,** 1970. Evolution by Gene Duplication, Springer-Verlag Publishing Company, New York.

[38] **Griffiths, Anthony J.F.; Miller, Jeffrey H., Suzuki, D.T., Lewontin, Richard C. and Gelbart, William M.,** 2000**.** Introduction to Genetic Analysis, W. H. FREEMAN, New York.

[39] **Pai, A.C.,** 1986. Foundation of Genetics, McGraw-Hill Book Company, Kefford Press, Singapure.

[39] **Bilge, A.,Erzan, A., Balcan, D.,** 2004. The shift-match number and string matching probabilities for binary sequences, preprint. q-bio.GN/0409023

[40] **Harbison CT, *et al.*,** 2004**.** Transcriptional regulatory code of a eukaryotic genome, *Nature* 431(7004), 99-104.

EK A

“SHIFT-MATCH NUMBER” (KAYDIRMA-ÇAKIŞIM SAYILARI)

Bilge, Erzan ve Balcan {0,1}’lerden oluşan çeşitli uzunluktaki dizilerin, birbirinin içinde bulunma olasılıklarının, “shift-match number” (s) denilen denklik sınıfına (equivalence class) göre belirlendiğini bulmuşlardır [39]. “Shift-match number” ın ne anlama geldiğini anlamak için nasıl hesaplandığına basit bir örnek üzerinde bakalım.

Dizi 0 1 0 1 0 1 olsun. Her defasında dizinin sağından bir elemanı silerek bir eleman sağa kaydırılır ve kaydırılan elemanlar asıl dizi ile tamamen aynı ise 1 değeri aksi durumda 0 değeri atanır. Atanan elemanların oluşturdurduğu dizi, seçili dizinin “Shift-match number” ını oluşturur.

		Atanan değer
0 1 0 1 0 1	:	1
0 1 0 1 0	:	0
0 1 0 1	:	1
0 1 0	:	0
0 1	:	1
0	:	0

Yukarıda görüldüğü gibi 0 1 0 1 0 1 dizisinin “shift-match number” ı s=1 0 0 0 1 0 olur. Dizilerin rastgele oluşturulduğu durumda [39]’da yapılan analitik hesaplar ile dizilerin birbirini içinde barındırma olasılıklarının s sayıları ile ilişkili olduğu görülmüştür. Buradaki derece dağılımlarında da, l=2 uzunluğundaki dizilerin 2 maksimuma, l=3 uzunluğundaki dizilerin 3 maksimuma vs. ayrıldığı davranışlar görülmüştür..

ÖZGEÇMİŞ

1979 yılında İstanbul'da doğdu. Beylerbeyi Lisesi'nden 1996 yılında mezun olduktan sonra, 1997 yılında İstanbul Teknik Üniversitesi Fen-Edebiyat Fakütesi Fizik Mühendisliği bölümüne girdi. Bir yıl İ. T. Ü. Dil-İnkilap Fakültesi'nde okuduğu ingilizce hazırlık sınıfından sonra başladığı lisans öğrenimini 2003 yılında tamamladı. Aynı yıl İ. T. Ü. Fen Bilimleri Enstitüsü Fizik Mühendisliği Programı Fizik Mühendisliği Anabilimdalı'nda yüksek lisans öğrenimine başladı. Mart 2005'den itibaren İ. T. Ü. Fen-Edebiyat Fakütesi Fizik Bölümü'nde araştırma görevlisidir.