

**TÜRKİYE'DEKİ BİLİMSEL İŞBİRLİĞİ AĞI
VE DİNAMİKLERİ**

**2013
DOKTORA TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ**

İlker TÜRKER

TÜRKİYE'DEKİ BİLİMSEL İŞBİRLİĞİ AĞI VE DİNAMİKLERİ

İlker TÜRKER

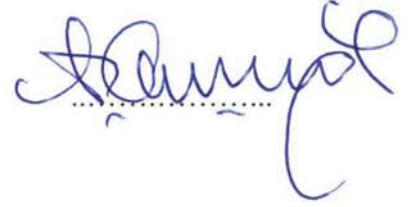
**Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalında
Doktora Tezi
Olarak Hazırlanmıştır**

**KARABÜK
Haziran 2013**

İlker TÜRKER tarafından hazırlanan “TÜRKİYE’DEKİ BİLİMSEL İŞBİRLİĞİ AĞI VE DİNAMİKLERİ” başlıklı bu tezin Doktora Tezi olarak uygun olduğunu onaylarım.

Prof. Dr. Abdullah ÇAVUŞOĞLU

Tez Danışmanı, YBÜ Bilgisayar Mühendisliği Anabilim Dalı



Bu çalışma, jürimiz tarafından oy birliği ile Bilgisayar Mühendisliği Anabilim Dalında Doktora tezi olarak kabul edilmiştir. 20/06/2013

Ünvanı, Adı SOYADI (Kurumu)

İmzası

Başkan : Prof. Dr. Fatih V. ÇELEBİ (YBÜ)



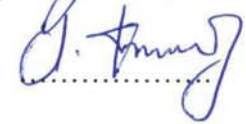
Üye : Prof. Dr. Abdullah ÇAVUŞOĞLU (YBÜ)



Üye : Yrd. Doç. Dr. Muharrem DÜĞENCİ (KBÜ)



Üye : Yrd. Doç. Dr. İsmail KURNAZ (KBÜ)



Üye : Yrd. Doç. Dr. Ergin YILMAZ (BEÜ)

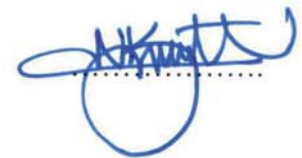


...../...../2013

KBÜ Fen Bilimleri Enstitüsü Yönetim Kurulu, bu tez ile, Doktora derecesini onamıştır.

Prof. Dr. Nizamettin KAHRAMAN

Fen Bilimleri Enstitüsü Müdürü



“Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.”

İlker TÜRKER

ÖZET

Doktora Tezi

TÜRKİYE'DEKİ BİLİMSEL İŞBİRLİĞİ AĞI VE DİNAMİKLERİ

İlker TÜRKER

Karabük Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:

Prof. Dr. Abdullah ÇAVUŞOĞLU

Haziran 2013, 129 sayfa

Bilimsel işbirliği ağları, üniversite derecelendirme, kompleks ağların modellenmesi gibi alanlarda her gün daha fazla ilgi gören bir araştırma konusu olarak dikkat çekmektedir. Web of Science veritabanındaki verilere dayanan çalışmamızda, Türkiye adresli yazarlara ait 1980-2012 yılları arasındaki yayınları baz alan bir yapısal veritabanı elde ederek bir bilimsel işbirliği ağı tanımlanmıştır. Bu kompleks ağın yapısını ve değişimini bilimsel yöntemlerle incelemek için bir takım hesaplama prosedürleri uygulanmıştır. Bunların sonucunda, ülkemiz bilimsel işbirliği ağının ilgili yıllar içerisinde, daha rekabetçi bir yükseköğrenim sistemini hedefleyen milli politikaların da etkisiyle, sürekli büyüme fazında olduğu gözlenmiştir. Elde ettiğimiz sonuçlar, yıllar geçtikçe yayın başına düşen yazar sayısının artmakta olduğunu göstermektedir. Ayrıca small-world konseptine uygun olarak, ağdaki düğümler arası ortalama uzaklığın 4 civarına yakınsadığı görülmektedir. Bu temel gösterge ile birlikte, elde ettiğimiz yüksek kümelenme katsayısı (0.75 civarında), ağımızın iç bağlılığının yüksek seviyede olduğunu ortaya koymaktadır. Ağın topolojisi ile ilgili

bir başka önemli gösterge de derece dağılımıdır. Power-law dağılımına uyumluluk gösteren derece dağılımı, ağımızın scale-free olduğunu göstermektedir. Bahsedilen parametrelerin son değerleri ile birlikte, 33 yıllık süreç içerisinde zamanla değişimleri de çalışmamızda ortaya koyulmuştur. Önceki çalışmalarla uyum içerisinde olan sonuçlarımız, büyüme hızı, kümelenme özellikleri ve ortalama düğüm uzaklığı açısından ilgi çekici farklılıklar da sergilemektedir. Ayrıca seçimli bağlanmanın yeni düğümler için avantaj sağladığı ortaya koyularak literatüre katkı sağlanması amaçlanmıştır. Bununla birlikte çalışmamız, yaklaşık 185.000 düğümden (yazardan) oluşan 33 yıllık disiplinler arası ulusal bilimsel işbirliği veri setinin kullanılmış olması açısından literatüre yenilik getirmiştir.

Anahtar Sözcükler : Kompleks ağlar, scale-free ağlar, bilimsel işbirliği.

Bilim Kodu : 902.1.063

ABSTRACT

Ph. D. Thesis

THE SCIENTIFIC COLLABORATION NETWORK IN TURKEY AND ITS DYNAMICS

İlker TÜRKER

Karabük University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

Thesis Advisor:

Prof. Dr. Abdullah ÇAVUŞOĞLU

June 2013, 129 pages

Networking via co-authorship is an attractive area of research and used in many fields such as ranking of the universities/departments or modeling of the complex networks. Studying on the data supplied by the Web of Science, we constructed a structural database that defines the scientific collaboration network of the authors from Turkey, based on the publications between 1980 and 2012. To uncover the evolution and structure of this complex network by scientific means, we executed some empirical measurements. The Turkish scientific collaboration network is in an accelerating phase in growth, highly governed by the national policies aiming to develop a competitive higher education system in Turkey. As our results suggest the authors tend to make more number of collaborations in their studies over the years. The results also showed that, node separation of the network slightly converges about 4, consistent with the small world phenomenon. Together with this key indicator, the high clustering coefficient, (which is about 0.75) reveals that our network is strongly

interconnected. Another quantity of major interest about such networks is the degree distribution. It has a power-law tail that defines the network as scale-free. Along with the final values, the time evolutions of the above-mentioned parameters are presented in detail with this work. In a good agreement with the recent studies, our network yields some significant differences especially in growing rate, clustering properties and node separation. We also aimed to make contribution to the literature by stating out that preferential attachment provides advantage for the new nodes. Furthermore, studying such a wide dataset of 33 years of national scientific interdisciplinary collaboration data composed of ~185.000 nodes (authors) is unique in the literature.

Key Word : Complex networks, scale-free networks, scientific collaboration.

Science Code : 902.1.063

TEŐEKKÜR

Bu tez alıőmasının planlanmasında, araőtırılmasında, yürütölmesinde ve oluşumunda ilgi ve desteęini esirgemeyen, engin bilgi ve tecrübelerinden yararlandıęım, yönlendirme ve bilgilendirmeleriyle alıőmamı bilimsel temeller ışığında őekillendiren sayın hocam Prof. Dr. Abdullah AVUŐOęLU'ya sonsuz teőekkürlerimi sunarım.

Veri setini elde etmemizde yardımlarını esirgemeyen Thomson & Reuters Ankara bürosu personeline teőekkür ederim.

Sevgili aileme manevi hiçbir yardımcı esirgemeden yanımda oldukları için tüm kalbimle teőekkür ederim.

İÇİNDEKİLER

Sayfa

KABUL	ii
ÖZET.....	iv
ABSTRACT	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER	ix
ŞEKİLLER DİZİNİ.....	xii
ÇİZELGELER DİZİNİ	xiii
BÖLÜM 1	1
GİRİŞ	1
BÖLÜM 2	3
LİTERATÜR	3
BÖLÜM 3	11
KOMPLEKS AĞLAR	11
3.1. KOMPLEKS AĞ TANIMI.....	11
3.2. GERÇEL AĞLAR	15
3.2.1. World Wide Web (WWW).....	15
3.2.2. İnternet	18
3.2.3. Aktör Birlikteliği Ağları	20
3.2.4. Bilimsel İşbirliği Ağları	20
3.2.5. Biyolojik Ağlar	24
3.2.6. Ekolojik Ağlar.....	24
3.2.7. Telefon Ağları.....	26
3.2.8. Bilimsel Atıf Ağları	27
3.2.9. Kelime Ağları	27
3.2.10. Enerji Nakil Hatları ve Sınır Ağları.....	29

3.2.11. Protein Ağları.....	30
3.3. SMALL-WORLD AĞ MODELİ.....	30
3.4. SCALE FREE AĞ MODELİ.....	33
3.5. EVRİLEN AĞLAR TEORİSİ.....	38
3.5.1. Seçimli Bağlanma.....	38
3.5.2. Büyüme.....	40
3.5.3. Diğer Etkenler.....	41
BÖLÜM 4	44
GERÇEKLEŞTİRİLMİŞ OLAN ÇALIŞMALAR	44
4.1. VERİ SETİNİN ELDE EDİLMESİ.....	44
4.1.1. Web of Science Veri Tabanından Verilerin Elde Edilmesi.....	44
4.1.2. Verilerin Yapısal Veri Tabanına Aktarılması.....	47
4.1.3. Veritabanı Oluşturmada Kullanılan Programın Yapısı.....	50
4.2. VERİLERİN İŞLENMESİ VE ELDE EDİLEN SONUÇLAR	54
4.2.1. Yazar Sayılarının Yıllara Göre Değişimi.....	54
4.2.2. Yayın Sayılarının Yıllara Göre Değişimi	54
4.2.3. Yayın Başına Düşen Yazar Sayılarının Oluşma Frekansları.....	57
4.2.4. Yazar Başına Düşen Yayın Sayılarının Oluşma Frekansları.....	57
4.2.5. Yazarların Birbirine Ortalama Uzaklığının Yıllara Göre Değişimi.....	60
4.2.6. Kümelenme Katsayısının Yıllara Göre Değişimi	62
4.2.7. Ortalama Derecenin Değişimi	64
4.2.8. Derece Dağılımı.....	66
4.2.9. Yazar Başına Düşen Ortalama Yayın ve Yayın Başına Düşen Ortalama Yazar Sayılarının Yıllara Göre Değişimleri	69
4.2.10. Ağın Genişleme Hızının İncelenmesi.....	71
4.2.11. Popüler Yazarlarla Yayın Yapmak.....	75
BÖLÜM 5	80
SONUÇLAR VE ÖNERİLER	80
5.1. SONUÇLAR	80
5.2. ÖNERİLER.....	81

	<u>Sayfa</u>
KAYNAKLAR	83
EK AÇIKLAMALAR A. VERİTABANLARI (MS SQL).....	87
EK AÇIKLAMALAR B. PROGRAM KODLARI.....	91
ÖZGEÇMİŞ	129

ŞEKİLLER DİZİNİ

Sayfa

Şekil 2.1. Slovenya fizikçilerinin oluşturduğu sosyal ağ	8
Şekil 2.2. Barabási'nin oluşturduğu, WWW'yi temsil eden ağ yapısı.....	9
Şekil 2.3. NLANR Caching Hierarchy'nin topolojisi ve gelişimi.....	10
Şekil 3.1. 3 farklı ağ için kümelenme katsayısının hesaplanması	13
Şekil 3.2. Poisson dağılımına sahip derece dağılımı grafiği.	14
Şekil 3.3. WWW'nin ağ yapısı. Web sayfaları düğümleri, sayfalar arası bağlantılar (URL'ler) ise linkleri oluşturmaktadır.	16
Şekil 3.4. İnterneti tanımlayan ağ şeması.....	18
Şekil 3.5. İnternet ağına router düzeyinde bakış.	18
Şekil 3.6. İnternet ağına interdomain düzeyde bakış.....	19
Şekil 3.7. Çeşitli gerçel ağlara ait derece dağılımı grafikleri.	22
Şekil 3.8. IRIS konferanslarının tanımladığı bilimsel işbirliği ağı.....	23
Şekil 3.9. Güney Afrika'nın batı sahilindeki Benguela ekosisteminden bir kesit... 26	
Şekil 3.10. Bir dildeki kelimelerin oluşturduğu metine ait muhtemel bir örüntü.....	28
Şekil 3.11. Batı ABD enerji nakil hatları ağı.....	29
Şekil 3.12. Watts-Strogatz ağ modelinde rastgele bağlanmalar arttıkça ağın topolojisinin değişimi.....	31
Şekil 3.13. WS modeli için düğümler arası uzaklık $l(p)$ ve kümelenme katsayısı $C(p)$ değerlerinin değişimi	33
Şekil 3.14. Scale-free ağlara ait power-law derece dağılımı.....	34
Şekil 3.15. Barabási-Albert modeli ile oluşturulmuş, 4 düğümlerle başlayarak her adımda yeni bir düğümün 4 farklı düğüme bağlanması ile 200 düğüme ulaşan oluşan bir ağ yapısı	36
Şekil 3.16. Barabási-Albert modeli ile oluşturulmuş bir ağa ait derece dağılımı grafiği	37
Şekil 3.17. Seçimli bağlanma grafikleri	40
Şekil 3.18. Derece dağılımının power-law'dan sapma eğilimleri: (yaşlanma ve maliyet etkileri).....	42
Şekil 4.1. Web of Science arama prosedürü.....	46
Şekil 4.2. Web of Science arama sonuçları	48
Şekil 4.3. Veritabanı tasarım şeması.	49

Şekil 4.4. Veritabanı oluşturmada kullanılan programın ara yüzü.....	50
Şekil 4.5. Veritabanı oluşturmada kullanılan programın akış diyagramı.....	51
Şekil 4.6. Ağda tanımlı yazar (düğüm) sayısının yıllara göre değişimi	55
Şekil 4.7. Yayın sayısının yıllara göre değişimi.....	55
Şekil 4.8. Yayın başına düşen yazar ve yazar başına düşen yayın sayılarının oluşma frekansları.	58
Şekil 4.9. Yayın başına düşen yazar ve yazar başına düşen yayın sayılarının oluşma frekansları (log-bin uygulanmış ve normalize edilmiş).....	58
Şekil 4.10. Yazarların birbirine ortalama uzaklığı.	60
Şekil 4.11. Kümelenme katsayısının yıllara göre değişimi.	63
Şekil 4.12. Ortalama derecenin yıllara göre değişimi.	65
Şekil 4.13. Ortalama derecenin yıllara göre değişimi (log-lineer çizim).	65
Şekil 4.14. Derece dağılımı.	67
Şekil 4.15. Çeşitli ağlara ait derece dağılımı grafikleri.....	69
Şekil 4.16. Yazar başına düşen ortalama yayın ve yayın başına düşen ortalama yazar sayılarının yıllara göre değişimleri.	70
Şekil 4.17. Türkiye'nin bilimsel işbirliği ağının büyüme istatistikleri.....	72
Şekil 4.18. Link-düğüm uzayındaki değişim grafiği.	74
Şekil 4.19. Ağın genişleme hızının yıllara göre değişimi.	74
Şekil 4.20. Ağa eklenen yeni düğümlerin renk kodlu değişim grafikleri.....	77
Şekil 4.21. 2010 yılı verileri kullanılarak elde edilen Zipf plot.	78
Şekil 4.22. 2010 yılı verileri kullanılarak elde edilen Zipf plot (log-log çizim).	78

ÇİZELGELER DİZİNİ

	<u>Sayfa</u>
Çizelge 3.1. Çeşitli gerçel ağlara ait genel karakteristikler.....	21
Çizelge 3.2. $P(k)$ 'nin power-law dağılımında olduğu çeşitli scale-free ağlara ait parametreler.....	21

BÖLÜM 1

GİRİŞ

Kompleks ağlar, doğada ve toplumdaki bazı birliktelikleri tanımlamak ve dinamiklerini anlamak için önemli modellerdir. Bu alanda en sık atıf yapılan çalışmalar, hücre ağları, kimyasal reaksiyonların oluşturduğu kimyasal ağları ve yönlendirici (router) ve bilgisayarların fiziksel bağlantılarından oluşan internet gibi kompleks ağları kapsamaktadır. Bu sistemler önceleri rastgele graflar şeklinde tanımlanırken, günümüzde bu ve benzeri gerçel ağların topoloji ve gelişimlerinin altında güçlü organizasyon prensiplerinin yatmakta olduğu anlaşılmıştır [1]. Günümüzde kompleks ağların organizasyon ve gelişme eğilimlerinin incelenebildiği önemli alanlardan birisi de bilimsel işbirliği ağlarıdır.

Bilimsel işbirliği ağlarının bu alanda tercih edilme sebepleri, ağı tanımlarken kullanılan yayın bilgilerinin ulaşılabilir olması (Web of Science, Scopus, MEDLINE vs.), her yayının tarihinin belirli olması, dolayısıyla ağın zaman içerisindeki topoloji ve büyüklük açısından gelişiminin de incelenabiliyor olması gibi önemli faktörlerdir. Örneğin internet'i tanımlayan düğümlerin geriye dönük tüm topolojik aktivitelerinin bugün elde edilmesi mümkün değildir. Bu anlamda bilimsel işbirliği ağları, kompleks ağların incelenmesi açısından önemli bir fırsat olma özelliğine sahiptir.

Web of Science elektronik veritabanındaki verilere dayanarak yaptığımız bu kompleks ağ çalışması, genel olarak 4 ana bölümden oluşmaktadır. İlk olarak kompleks ağlarla ilgili yapılmış temel çalışmaların kronolojik olarak gelişimi, literatür bölümünde özetlenmiştir. Ardından gelen bölümde kompleks ağlarla ilgili temel bilgilere yer verilmiştir. 4 alt bölümden oluşan bu bölümde öncelikle kompleks ağ tanımının nasıl ortaya çıktığı anlatılmıştır. Ardından gerçek dünyadaki kompleks ağ örnekleri incelenmiş, bu ağlara ait düğümlerin ortalama uzaklığı, kümelenme katsayısı ve derece dağılımı parametreleri incelenerek bu ağların farklı ve benzer

yönleri ortaya koyulmuştur. Son 2 alt bölümde ise nümerik yöntemlerle ağ tanımlamaya yönelik 2 temel ağ modeli olan small-world (Watts-Strogatz) ve scale free (Barabási-Alberts) ağ modelleri detaylanmıştır. Sonraki ana bölümümüzde gerçekleştirdiğimiz çalışmalar yer almaktadır. Bu bölüm de veri setinin elde edilerek yapısal veritabanına dönüştürülmesi ve bu verilerin işlenmesi ile elde edilen sonuçlar olmak üzere iki alt bölümden oluşmaktadır. Son bölümde ise elde edilen bu sonuçlar yorumlanmış, literatürdeki örneklerine göre gözlenen farklılıkların sebepleri irdelenmiştir.

BÖLÜM 2

LİTERATÜR

Tüm zamanların en üretken matematikçilerinden Paul Erdős, ömründe yaklaşık 1500 matematiksel makale yayınlamış ve çoğunu başkalarıyla ortak çalışma içinde hazırlamıştır. 500'e yakın meslektaşıyla birlikte çalışmış ve matematiksel işbirliğini, çoğu matematikçinin çalışma şeklini değiştirecek türde bir sosyal aktivite haline getirmiştir. Bu eşine az rastlanır üretkenlik, "Erdős sayısı" kavramına ilham vermiştir. Erdős'un kendisi için Erdős sayısı 0 ve doğrudan birlikte çalıştığı kişiler için Erdős sayısı 1 olarak belirlenirken 1 sayısı ıyla nitelenen kişilerin birlikte çalıştıkları kişiler 2 sayısını vs. almıştır [2]. Tüm matematikçiler gibi fizik, ekonomi gibi dallarda çalışma yapan insanların yeterince küçük sayılara sahip olduklarının anlaşılması ile bilim topluluğunun birbirine sıkı sayılabilecek sosyal yakınlık düzeyine sahip olduğu ortaya konmuştur [3].

1967 yılında Milgram Nebraska'da yaşayan 160 kişiye değişik paketler göndermiş, bu paketleri Boston'da arkadaşı olan bir broker'a ulaştırmak üzere göndermelerini tembihlemiştir. Bu paketleri gönderirken Boston'daki adrese teslim edecek ya da daha yakın arkadaşlara iletcek arkadaşlarını aracı yapmalarını istemiştir. Bu deney sonucunda toplumda rastgele seçilen iki insan arasındaki ayrıklık derecesinin "6" olduğu sonucuna varmıştır, çünkü paketler ortalama 6 devir teslim sonucunda ilgili adrese ulaşmışlardır [4]. Milgram'ın bu tespiti yıllarca tartışılmış olsa da 2008'de Leskovec ve Horvitz'in Microsoft'un .NET Messenger servisi üzerinde yapmış oldukları çalışmada kullanıcılar arasındaki ortalama düğüm sayısının 6.6 olduğu belirlenmiştir [5]. Bu değer, Milgram'ın 40 yıl önceki tespitini de doğrular niteliktedir.

Bilim çevresinin birlikteliklerini istatistiksel olarak ortaya koymasının yanında bu türdeki ağ çalışmaları, karmaşık ağların topolojik ve dinamik kurallarının anlaşılması

açısından büyük ilgi görmektedir. World Wide Web ya da genetik ağlar gibi karmaşık sistemler en yaklaşık biçimde kompleks topolojili ağlar ile ifade edilebilmektedir [6]. “Bilimsel işbirliği” verileri bu anlamda herkesin ulaşabileceği en geniş ağ olarak dikkat çekmektedir [3].

Sosyal ağlar genel olarak sosyal bilimlere ait çalışmaların ilgi konusu olmuştur. Bu çalışmaların genel karakteristikleri, göreceli olarak küçük sistemleri ele almaları ve sosyal ağları statik grafikler olarak görmeleridir. Buna karşılık yakın geçmişte yapılan istatistiksel fiziğe dayalı çalışmalar çok daha büyük ağlara odaklanmakta, bu ağların topolojik yapıları ile birlikte gelişimini ifade eden dinamikleri irdelemektedir. Bu çalışmaların sonucu elde edilen teorik ve deneysel sonuçlar, bilgisayar biliminden biyolojiye kadar birçok bilim dalı için sayısız çalışma alanının önünü açmıştır [7,8]. Ayrıca bu çalışmalar World Wide Web’in gelişimi, bilginin yayılımı, kültürel mirasların sürdürülmesi, hastalıkların yayılma biçimi gibi olguların anlaşılmasına da ışık tutmaktadır [9].

Şu ana kadar yapılan çalışmalara ait en önemli üç bulgudan birincisi, çoğu ağın “*small world*” tanımlamasına uyuyor olmasıdır. *Small world* terimi, ağdaki düğümler arasındaki ortalama uzaklığın yeterince küçük olduğu ağlar için kullanılmaktadır. Bir başka deyişle rastgele seçilen iki düğüm arasında kısa bir yol bulunabilmesi ve bu durumun ağdaki düğümlerin çoğu için geçerli olması durumudur [3,10]. Karmaşık ağlar hakkındaki ikinci önemli bulgu, gerçel ağların rassal ağlarda beklenene göre çok daha yüksek “*kümelene katsayısı*”na sahip olmalarıdır. Son önemli bulgu ise “*derece dağılımı*”nın ağın yapısı hakkında önemli bilgiler içeriyor olmasıdır. Bu dağılım, çoğu büyük çaptaki ağlar için scale-free power-law dağılımına uygundur [6,11]. Yani link sayısı arttıkça (*hub* düğümler arttıkça) düğüm sayısı azalmaktadır.

Newman’ın yapmış olduğu, modern ağ terminolojilerinin bilimsel işbirliği veritabanlarına uyarlandığı çalışmalar, karmaşık ağların dinamiklerini öğrenme adına önemli adımlar olmuştur [12-14]. Bu çalışmalarda değişik çalışma alanlarında 5 yıllık periyodu kapsayan bilimsel işbirliği veritabanları kullanılmış, bu veritabanlarının oluşturduğu ağların *small-world* ağların sahip olduğu

karakteristiklerin çoğuna sahip oldukları ortaya koyulmuştur. Elde edilen veriler, şaşırtıcı derecede kısa düğümler arası mesafeler ve büyük kümelenme katsayıları (benzer büyüklükteki bir Erdős-Renyi tipi ağdan bile daha büyük değerde) olmuştur [7]. Aynı zamanda bu ağlardaki derece dağılımını da power-law distribution kuralına uymaktadır.

Bu çalışmayı daha da ileriye götürme adına Barabási ve arkadaşları bilimsel işbirliği ağlarını gelişen ağların bir prototipi olarak düşünmüş, ağın gelişim parametrelerine ve dinamiklerine odaklanan bir çalışma gerçekleştirmişlerdir [7]. Ağın topolojisini ve gelişim dinamiklerini belirleyen en önemli gerçek, ağa sürekli olarak yeni yazarların eklenmesi ya da var olan yazarlar arasında yeni bağlantıların oluşmasıdır. Aslında ağın gelişim parametrelerinin ortaya koyulabilmesi sayesinde, Newman'ın çalışmış olduğu bilimsel işbirliği ağında statik yaklaşımla elde ettiği topolojik özelliklerin de daha iyi anlaşılabilmesi mümkün olmuştur.

Kompleks ağ dinamikleri üzerinde çalışma yapmak için kullanılacak tek veri seti bilimsel işbirliği veritabanları değildir. WWW de aynı zamanda yüksek hızda düğümlerin eklendiği (silindiği), ağ topolojisinin bu büyüme dinamikleri tarafından belirlendiği karmaşık-evrilen bir ağ çeşididir [6,15]. Hollywood aktörlerinin oluşturduğu birliktelik ağı ise bilimsel işbirliği ağı ile benzerlikler taşıyan bir diğer ağ örneğidir. Bu ağ da yeni aktörlerin katılımı ile büyümekte, aynı zamanda var olan aktörler arasında sürekli yeni bağlar oluşmaktadır [3,8]. Benzer biçimde çoğu hücreyel, ekolojik ya da ticari ağın dinamik yapısı benzer özellikler sergiler.

Bu durumda neden karmaşık ağ çalışmalarında bilimsel işbirliği verileri kullanılmaktadır? Bu sorunun cevabını birkaç faktör ile açıklayabiliriz. Öncelikle ağın dinamik gelişimini inceleyebilmek adına, zamana yayılan değişimi görebilmek ilk tercih sebebidir. Yani ihtiyaç duyduğumuz şey sadece ağ topolojisinin bir haritasına ulaşabilmek değil, hangi düğümün ve düğümler arasındaki hangi bağlantının hangi zamanda eklenmiş olduğunun da bilinmesi, ağın değişiminin incelenebilmesi açısından gereklidir. Bu gereklilik, üzerinde çalışılabilecek veri setlerinin sayısını ikiye indirmektedir: aktör ağı ve bilimsel işbirliği ağı. Bu seçeneklerden birincisinde filmlerin vizyona girdiği yıllar, ikincisinde ise bilimsel

yayınların gerçekleştiği yıllar, ağın zamandaki gelişimi için gerekli veriyi sağlayacaktır. Bu iki veri seti içinde de bilimsel işbirliği ağı, aktör ağına göre zamanda evrilen karmaşık bir ağ yapısını modellemek için daha uygun yapıdadır. Çünkü bilimsel çalışma gerçekleştiren bilim adamları, birlikte çalışacakları insanları tamamen kendileri seçerler. Aktörler için ise bu durum genellikle casting direktörünün tercihlerine bağlıdır, yani düğümlerin bir derece üzerinde gerçekleşir. Bu nedenle ağa yeni katılımlar ve ağda var olan düğümler arasında oluşan yeni bağlantılar baz alındığında bilimsel işbirliği ağının WWW ya da diğer sosyal sistemler gibi “evrilen ağlar” için daha uygun bir prototip oluşturduğu ortaya çıkmaktadır.

Barabási ve arkadaşlarının yaptığı çalışma üç aşamadan oluşmaktadır. İlk aşama veritabanı üzerinde sistemin parametrik değerlerinin belirlenmesi ve gelişim mekanizmasının ortaya çıkarılmasını içermektedir. Ölçülen bu değerler, ikinci aşamada bir sistem modeli oluşturmak için kullanılmıştır. Bu model “continuum theory” yardımıyla ağın gelişiminin kestiriminde kullanılarak daha önceki çalışmaların ötesine geçilmiştir. Üçüncü aşamada ise model üzerinde bilgisayar simülasyonları yapılarak (i) continuum theory ile kestirilemeyen büyüklükler irdelenmiş, (ii) önceki aşamada elde edilen veriler doğrulanmış, (iii) ağ parametrelerinin gelişim karakteristikleri daha anlaşılır hale getirilmiştir.

Alanında önemli bir boşluğu doldurmuş olan bu çalışmanın sonuçları da ilginç ve etkileyici olmuştur. Bilimsel işbirliğini konu alması yönüyle de bilim çevrelerinin ilgisini çeken bu çalışmada, aynı zamanda bir karmaşık ağ topolojisi modeli oluşturulmuştur. İlk göze çarpan sonuçlardan birisi, ağ parametrelerinin asimptotik ve ara bölgelerde farklı davranışlar sergilemesidir. Aslında çoğu parametrenin (büyüklük, kümelenme katsayısı, ortalama derece vs.) zamana bağımlı olduğu ve incelenen zaman aralığında doyuma ulaşma eğilimi göstermediği de ortaya çıkmıştır. Dolayısıyla bu parametrelerin herhangi bir andaki değerlerinin bize ağ hakkında çok şey ifade etmemekte olduğu; ağın small world özelliği gösterdiği ve ortalama ayrışma ve kümelenme katsayılarının bir rassal ağda beklenenden yüksek değerde olduğunu ifade etmekten öteye geçemediği ortaya çıkmaktadır.

Zaman içerisinde durağan olduğu düşünölen bir deęer de derece dağılımıdır. Nitekim bu dağılımın sadece asimptotik olarak doęrulanması ve iki farklı rejim sergilediđinin ortaya ıkması da bir diđer ilgi ekici sonutur.

Bilgisayar simölasyonlarının ortaya ıkardığı bir diđer gerek de sınırlı veri seti üzerindeki alıřmaların ortaya koyduđu bazı deęişim trendlerinin, tüm sistem dikkate alındığında sınırlı bir bölgeyi ifade ettiđi ve asıl sistem trendinin gözlenenin tersi yönde olabileceđidir. Kısa zaman dilimleri ile alıřmanın doęurduđu bu artefakt, uzun dönem verileri ya da simölasyon yardımı ile aşılabilmektedir.

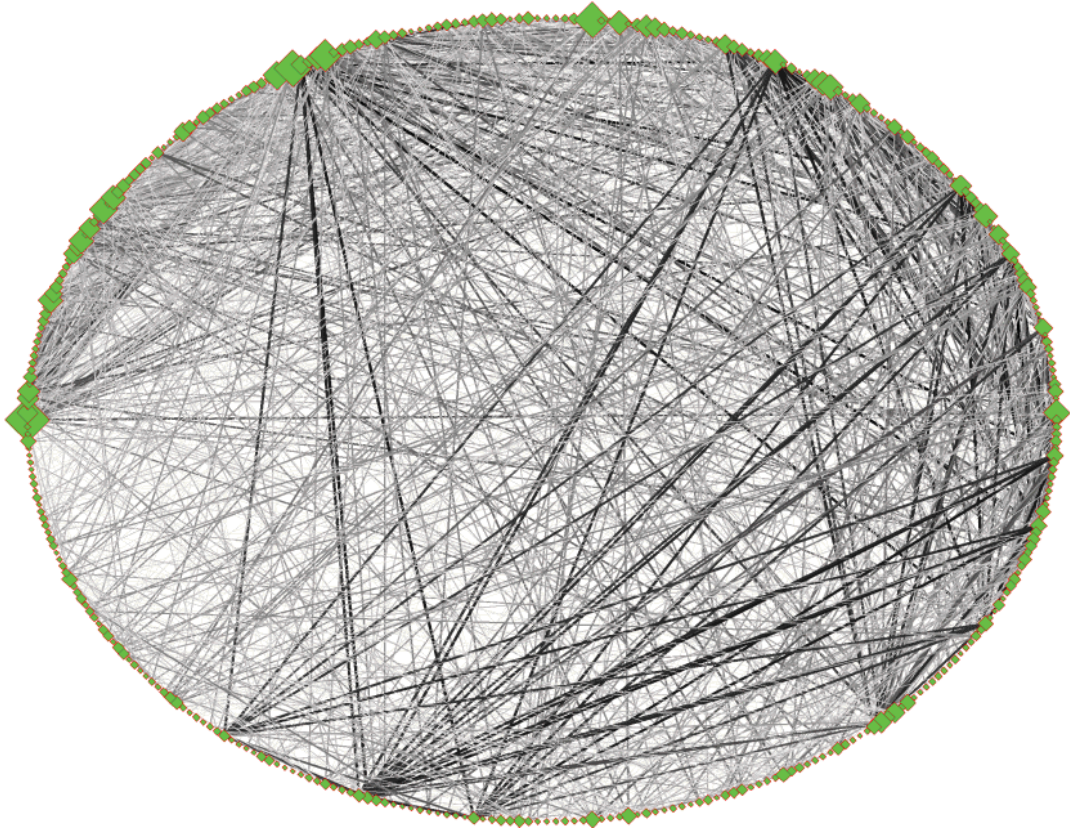
Bu ve benzeri alıřmaların ortaya koyduđu sonular, aslında bir kompleks bilimsel işbirliđi ađı modelinin tasarımı için başlangı noktasını oluřturmaktadır. řüphesiz uzun dönem simölasyonlarına imkan tanıyacak bir model; yaşılanma etkisi, emeklilik, yayın başına varsayılan yazar sayısının stokastik yapısı gibi etkenleri de dikkate alınmalıdır.

Bilimsel işbirliđi ađları üzerine yapılan alıřmalar günümüzde de eşitli perspektiflerde süregelmektedir. Perc'in Slovenya için gerekleřtirmiş olduđu alıřmada yazarlar arasındaki ortalama uzaklıđın zaman ekseninde bir devrilme noktasına sahip olduđu, bu noktadan sonra ortalama uzaklıđın azalmaya başladığı gösterilmiştir [9,16]. Perc, yazar başına düşen işbirliđi sayısında exponansiyel bir artışın yanında, ađın büyüklüğünün de exponansiyel olarak yakın bir artış gösterdiđi sonucuna varmıştır. Ayrıca kümelenme katsayısının zamanla azalmakta olduđu, buna rağmen hala güçlü denebilecek seviyede bulunduđunu ortaya koymuş, son olarak da ađın büyümesinin lineere yakın düzeyde seçmeli bağlanma (preferential attachment) yoluyla gerekleřtiđini saptamıştır. Bu bulgular beklenen derece dağılımına ok iyi yaklaşıklıkla uyum göstermiş, ađa bağlanma (aktif olma) yılının sahip olunan bilimsel ortak sayısını ters orantı ile etkilediđi gözlenmiştir. Genel anlamda gemiş alıřmalarla uyumlu olarak seçmeli bağlanma ve "small world" özellikleri Slovenya'nın bilimsel işbirliđi ađını da karakterize etmektedir [9].

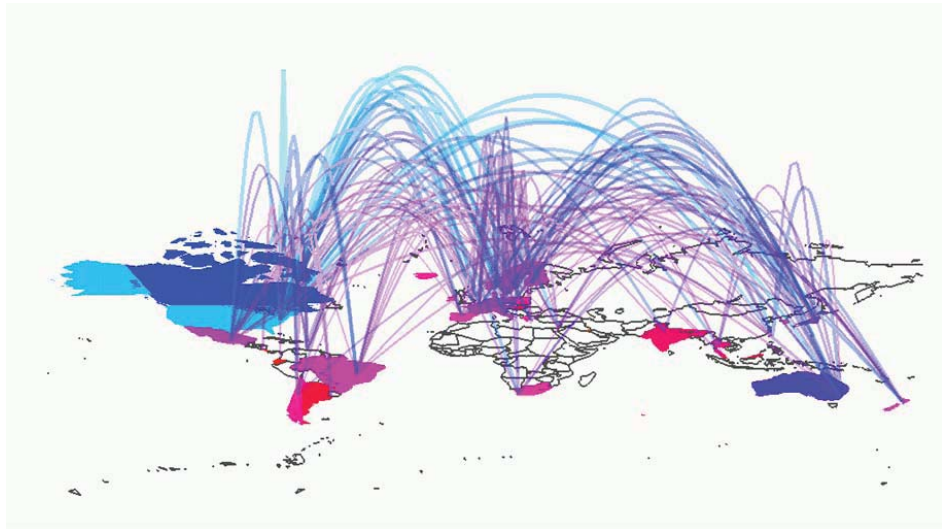
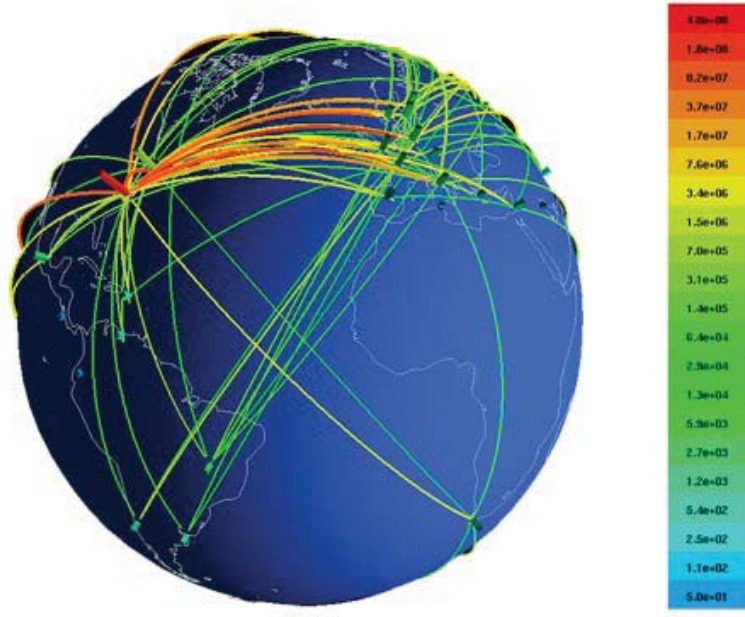
Hoekman ve arkadaşlarının 2010 yılında yayınlanan alıřmalarında, bilimsel ortaklıkların fiziksel mesafelere ve sosyal sınırlara (bölge, ölke, dil) olan bađımlılıđı

incelenmiştir. 33 Avrupa ülkesi ve 313 ayrı bölge için 2000-2007 yılları arasında yapılan bilimsel işbirlikleri baz alınarak yapılan çalışmanın sonucunda, fiziksel yakınlığın yol açtığı kutuplaşmanın zamanla azalmadığı, buna karşılık işbirliğinin sosyal sınırlar içerisinde kalma eğiliminde zamanla düşüş gerçekleştiği gözlenmiştir [17].

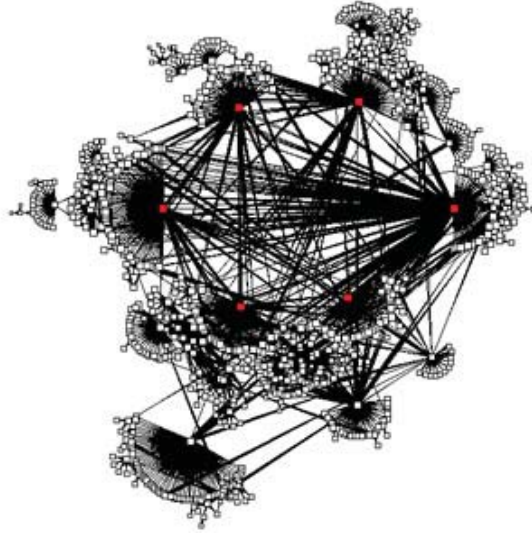
Karmaşık ağ çalışmalarının bir bölümünün çıktısı olarak bu ağların fiziksel gösterimlerinin sunulduğunu görmekteyiz. Bu ilgi çekici çıktılardan bazılarını aşağıda (Şekil 2.1-2.3) yer verilmiştir:



Şekil 2.1. Slovenya fizikçilerinin oluşturduğu sosyal ağ [9].



Şekil 2.2. Barabási'nin oluşturduğu, WWW'yi temsil eden ağ yapısı [18].



Şekil 2.3. NLANR Caching Hierarchy'nin topolojisi ve gelişimi [18].

Türkiye için bilimsel işbirliği alanında yapılmış çalışmalar incelendiğinde, genellikle 5-10 yıllık dönemleri kapsayan bilimsel üretkenlik, yayın türleri incelemeleri, tezlerdeki atıf sayılarının zamanda değişimi gibi büyüklüklerin incelendiği sosyal içerikli çalışmalara rastlanmakta, yeterince uzun dönemi kapsayan ve karmaşık ağ topolojisinin ve parametrelerinin zamandaki değişimini gösteren bir çalışma bulunmamaktadır. Bu çalışmada ilgili alandaki boşluğun da doldurulması hedeflenmektedir.

BÖLÜM 3

KOMPLEKS AĞLAR

3.1. KOMPLEKS AĞ TANIMI

Kompleks ağlar, geleneksel olarak graf teorisinin çalışma alanları arasında yer almıştır. İlk zamanlarda düzenli ağlara odaklanan graf teorisi, 1950’li yıllardan itibaren geniş ölçekli ve görünürde bir düzene sahip olmayan ağları random (rassal) ağlar olarak tanımlamıştır. Bu tanım, kompleks ağların en basit yaklaşımı olarak literatüre geçmiştir [1]. Random graflar (rassal ağlar), ilk defa Macar matematikçiler Paul Erdos and Alfred Renyi tarafından ortaya koyulmuştur. Erdos-Renyi modeline göre N düğümden oluşan bir ağdaki düğüm çiftlerinin birbirine bağlanma olasılığı p ise, bu düğümler rastgele dağılımlı $pN(N - 1)/2$ adet bağlantıdan (link) oluşan bir ağ tanımlarlar. Bu model, ortaya koyulduğu yıllardan bu yana kompleks ağlar hakkındaki diğer çalışmalara da yön vermiştir [2]. Yakın zamanda kompleks ağların daha fazla ilgi görmesi ile birlikte bilim adamları bu modelin önerdiği yaklaşımı sorgulamaya başlamıştır. Bunun sonucu olarak da şu soru cevap aramaya başlamıştır: Biyolojik hücreler ya da internet tarafından tanımlanan kompleks ağlar gerçekte rassal ağlar mıdır? Eğer değilse, bu ağların iç organizasyon prensiplerinin izleri bir şekilde ağın topolojisinde gözlenebilir olmalıdır. Rassal ağlardan farklı olan topolojik özelliklerin ve organizasyon prensiplerinin nicel olarak ortaya koyulabilmesi için de bazı araçlara ihtiyaç duyulacağı aşikârdır [1,7].

Geçtiğimiz yıllar, çeşitli paralel gelişmelerin katkısıyla, kompleks ağların incelenebilirliğini üst seviyeye çıkarmıştır. Öncelikle bilgisayarların, özellikle veri işleme ve saklama alanında kullanımının yaygınlaşması, çeşitli gerçel ağların topolojilerini saklayan büyük veritabanlarının da oluşumunu tetiklemiştir. İkinci olarak, bilgisayarların sunduğu yüksek işlem gücü sayesinde milyonlarca düğümden oluşan ağlar da incelenebilir hale gelmiş, bilinmeyen yönleri aydınlatılabilmektedir.

Üçüncü olarak, bilimsel disiplinler arası sınırların yavaş yavaş da olsa yıkılmakta oluşu, daha fazla araştırmacının çeşitli veritabanlarına ulaşarak kompleks ağları incelemesine olanak tanımıştır. Son olarak, geçmişteki kısıtlayıcı yaklaşımın aksine, sistemlerin davranışlarının bir bütün olarak incelenmesi gereğinin gittikçe daha çok seslendirilmeye başlamasıdır.

Bu gerçekler ve gelişmeler ışığında, son birkaç yıl içerisinde ağ topolojisi ve karakteristiklerini tanımlayan yeni ölçüler ve yaklaşımlar önerilmiş ve kullanılmıştır. Kompleks ağlarla ilgili bu yeni yaklaşımlardan üç tanesi kendine çoğu araştırmada baskın biçimde yer bulmuştur. Bu tanımlamalar aşağıda kısaca açıklanmış olup detaylı anlatımları ilerleyen bölümlerde yer almaktadır.

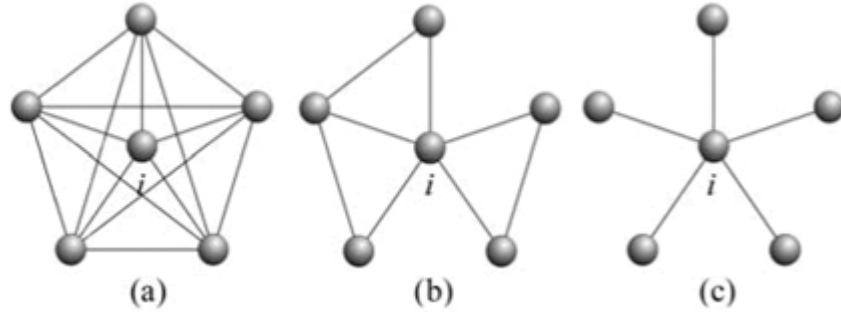
Small Worlds : Small-world konsepti kısaca, büyük boyutlarına rağmen çoğu gerçel ağda, ağı oluşturan düğümlerden herhangi ikisi arasında nispeten kısa bir yolun bulunabiliyor olmasıdır. Düğümler arasındaki mesafe, düğümleri birleştiren en kısa yolu oluşturan linklerin sayısıdır. Small-world konsepti ile ilgili en bilinen çalışma Milgram'ın "6 derece uzaklık" adıyla anılan ve önceki bölümde de detaylarına yer verdiğimiz posta deneyidir [4]. Bu deneyden günümüze kadar yapılmış bir çok kompleks ağ çalışmasında, ağ ne kadar büyük olursa olsun düğümler arasındaki ortalama mesafelerin, Milgram'ın ortaya koyduğu "6" rakamına yakın değerlerde olduğunu görmekteyiz. İlgi çekici bir kavram olmasının yanında, small-world konsepti belirli bir organizasyon prensibinin bir göstergesi olarak algılanmamalıdır. Aslında Erdos-Renyi, tanımladıkları rassal ağdaki düğümler arası ortalama uzaklığın, ağdaki düğüm sayısının logaritması ile ölçeklendiğini göstermişlerdir [2]. Yani rassal ağlar da aslında small-world tanımına uyan yapılardır [1].

Kümelenme (Clustering) : Sosyal ağların bir diğer önemli özelliği de *kamplaşma* gerçeğidir. Bu kamplaşmaların oluşturduğu bir *küme*, içerisinde kalan tüm düğümlerin birbirini tanıdığı bir kapalı çevrim gibi düşünülebilir. Gerçel ağlardaki kümelenme eğilimi 1998 yılında Watts ve Strogatz tarafından tanımlanan kümelenme katsayısı ile nicel olarak incelenmiştir [8]. Kümelenme aslında kökleri sosyolojide olan "geçişli üçgenlerin oranı" tanımına dayanmaktadır. Ağdaki i . düğümün, diğer k_i düğümlerle bağlanmasını sağlayan k_i adet kenar (link) mevcuttur.

Eğer i . düğümün tüm komşuları tam bir kümelenme içinde ise, yani her biri diğer tüm k_i komşu ile bağlantılı ise, tüm bu kamplaşmanın oluşturduğu $k_i(k_i - 1)/2$ adet link vardır. Bu değer, k_i adet komşunun oluşturabileceği maksimum link sayısıdır. i . düğüme ait kümelenme katsayısını hesaplamak için, k_i adet komşunun aralarında gerçekte oluşturmuş oldukları E_i link de dikkate alınarak kümelenme katsayısı aşağıdaki gibi hesaplanabilir.

$$C_i = 2E_i/k_i(k_i - 1) \quad (3.1)$$

Tüm ağın kümelenme katsayısı ise ağdaki tüm düğümlere ait C_i değerlerinin ortalaması olarak hesaplanmaktadır. Herhangi bir düğüme ait C_i değerinin hesaplanması Şekil 3.1’de anlatılmıştır.



Şekil 3.1. 3 farklı ağ için kümelenme katsayısının hesaplanması (Eşitlik 3.1’e bakınız). a) ağında i . düğümün 5 komşusunun kendi aralarında 10 adet link vardır. Bu durumda $C_i = \frac{2E_i}{k_i(k_i-1)} = \frac{2(10)}{5(4)} = 1$ olur. b) ağında i . düğümün 5 komşusunun kendi aralarında 3 adet link vardır. Bu durumda $C_i = \frac{2(3)}{5(4)} = 0.3$ olur. c) ağında ise i . düğümün komşularının kendi aralarında hiç link yoktur. Bu durumda $C_i = \frac{2(0)}{5(4)} = 0$ ’dır.

Bir rassal ağda linkler de rastgele dağılımlı oldukları için, kümelenme katsayısı $C = p$ (p : N düğümden oluşan bir ağdaki düğüm çiftlerinin birbirine bağlanma olasılığı) olarak kabul edilir. Gerçel ağların çoğunda kümelenme katsayısı, benzer boyutlardaki (örn. eşit sayıdaki düğüm ve link sayısına sahip) bir rassal ağa oranla daha büyük değerlerdedir.

Derece Dağılımı (Degree Distribution): Ağdaki bir düğümün diğer düğümlere bağlandığı link sayısı, onun *derecesi* olarak adlandırılır. Bir ağdaki tüm düğümlerin dereceleri birbirine eşit değildir. Düğümlerin dereceleri, rastgele seçilen bir düğümün k adet linke sahip olma olasılığını tanımlayan bir $P(k)$ dağılım fonksiyonu ile karakterize edilebilir. Bir rassal ağda linkler de rastgele dağıldığından, düğümlerin büyük çoğunluğu yaklaşık olarak aynı dereceye sahip olacaktır. Bu değer de ağın ortalama derecesi $\langle k \rangle$ 'ya yakın olması beklenir. Rassal ağların derece dağılımı ise $P(\langle k \rangle)$ değeri civarında pik yapan bir Poisson dağılımıdır (Şekil 3.2).



Şekil 3.2. Poisson dağılımına sahip derece dağılımı grafiği.

Kompleks ağların anlaşılmasındaki en ilgi çeken gelişmelerden birisi, çoğu büyük çaplı gerçel ağda derece dağılımının Poisson dağılımından farklı olduğunun keşfedilmesidir. World Wide Web, internet ya da biyolojik ağlar gibi bazı önemli gerçel ağ türlerinde, derece dağılımı power-law dağılımı ile uyum göstermektedir, yani aşağıdaki karakteristiğe sahiptir [1].

$$P(k) \sim k^{-\gamma} \quad (3.2)$$

Bu karakterdeki derece dağılımına sahip ağlar scale-free olarak adlandırılmaktadır [15]. Bazı gerçel ağlarda exponansiyel kuyruk bölgesi gözlenirse de $P(k)$ 'nin fonksiyonel formu, bir rassal ağda beklenen Poisson dağılımından oldukça farklıdır.

Bu keşifler, yakın geçmişte ağ modelleme ile ilgili çalışmalara hız kazandırmış, 3 ana modelleme anlayışının ortaya çıkmasını sağlamıştır. Birincisi, Erdos-Renyi modelinin türevleri olan rassal ağlar olup çeşitli alanlarda kullanılmıştır ve çoğu

modelleme ya da deneysel çalışma için bir benchmark görevi görmeye devam etmektedir. Kümelenmenin varlığından yola çıkan ikinci grup modeller ise small-world modeller olarak adlandırılmıştır. Bu modeller, kümelenmenin çok yüksek olduğu düzenli kafes yapılar ile rassal ağlar arasında yer almaktadır. Son olarak power-law derece dağılımının keşfinin öncülük yaptığı çeşitli scale-free modeller oluşturulmuştur. Ağ dinamiklerine odaklanan bu çalışmalar, ağdaki değişimleri de içeren evrensel bir model oluşturma amacını gütmektedirler.

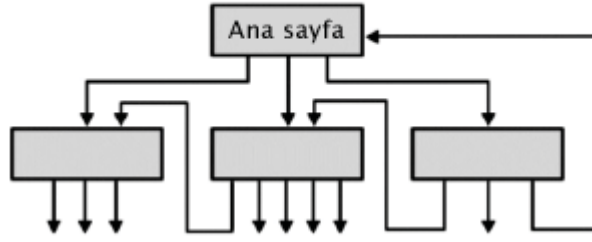
3.2. GERÇEL AĞLAR

Çoğu kompleks ağ çalışması, iletişim ağlarından biyolojik kümelere kadar çeşitli gerçel ağın topolojik yapısını ve gelişim süreçlerini anlama isteğinden yola çıkarak başlamıştır. Dolayısıyla kompleks ağ çalışmaları çeşitli disiplinleri kapsayan bir perspektiftedir. Bu bölümde, literatürde farklı alanlarda yapılmış ağ çalışmalarını kısaca özetleyeceğiz. Erişilebilir veritabanları hakkında bilgi vermenin ötesinde, ağ topolojisi hakkında fikir veren 3 önemli ölçüm üzerinde duracağız: düğümlerin ortalama uzaklığı, kümelenme katsayısı ve derece dağılımı. Bu büyüklükler dışındaki çeşitli parametreler ve katsayılar hakkındaki bilgilere de ilerleyen bölümlerimizde ulaşabilirsiniz.

3.2.1. World Wide Web (WWW)

World Wide Web (dünya çapındaki web sayfası ağı), topolojik bilginin ulaşılabilir olduğu en büyük ağıdır. Bu ağın düğümleri web sayfaları, linkleri ise bu sayfalar arasındaki bağlantılardır (Şekil 3.3).

World Wide Web



Şekil 3.3. WWW'nin ağ yapısı. Web sayfaları düğümleri, sayfalar arası bağlantılar (URL'ler) ise linkleri oluşturmaktadır.

WWW'nin toplam boyutu, günümüz (Mayıs 2013) itibarıyla 4.6 milyar web sayfası civarındadır [19]. WWW'nin ağ topolojisine olan ilgi, Albert ve arkadaşlarının 1999 yılındaki ve Kumar ve arkadaşlarının aynı yıl yaptıkları çalışmalarında, web sayfalarının derece dağılımlarının power-law karakteristiğine sahip olduğunu göstermelerinin de etkisiyle bu tarihten sonra hızlı bir artış eğilimine girmiştir [15,20]. WWW linkleri yönlü oldukları için derece dağılımı için de 2 farklı dağılımdan söz edilebilir: düğümlerden çıkan linklerin tanımladığı $P_{out}(k)$ ve düğümlere gelen linklerin tanımladığı $P_{in}(k)$. WWW üzerinde yapılan çoğu çalışma, bu dağılımların da power-law uyumlu kuyruklara sahip olduğunu göstermiştir.

$$P_{out}(k) \sim k^{-\gamma_{out}} \quad \text{ve} \quad P_{in}(k) \sim k^{-\gamma_{in}} \quad (3.3)$$

Albert ve arkadaşları, WWW'nin 325,729 düğümden oluşan bir alt kümesi üzerinde yaptıkları çalışmalarında $\gamma_{out} = 2.45$, $\gamma_{in} = 2.1$ değerlerini elde etmişlerdir [15]. Kumar ve arkadaşları ise Alexa üzerinden elde ettikleri 40 milyonu aşan düğümden oluşan veri seti üzerinde yaptıkları çalışmalarında $\gamma_{out} = 2.38$, $\gamma_{in} = 2.1$ katsayılarını elde etmişlerdir [20]. Broder ve arkadaşları 2000 yılında 200 milyon civarındaki düğümden oluşan Altavista verileri üzerindeki çalışmalarında ise $\gamma_{out} = 2.72$, $\gamma_{in} = 2.1$ katsayılarına sahip power-law kuyruklu derece dağılımı grafikleri elde etmişlerdir [21]. Adamic ve Huberman ise 2000 yılında farklı bir yaklaşımla gerçekleştirdikleri çalışmalarında her bir domain adını bir düğüm olarak kabul etmişler, iki farklı domain altındaki herhangi birer web sayfası arasında bağlantı mevcutsa bu iki domaini birbirine bağlı kabul etmişlerdir [22]. Aynı domain

altındaki web sayfalarının gruplandığı bu çalışmada da yine $\gamma_{in} = 1.94$ katsayısına sahip power-law kuyruklu derece dağılımı elde etmişlerdir.

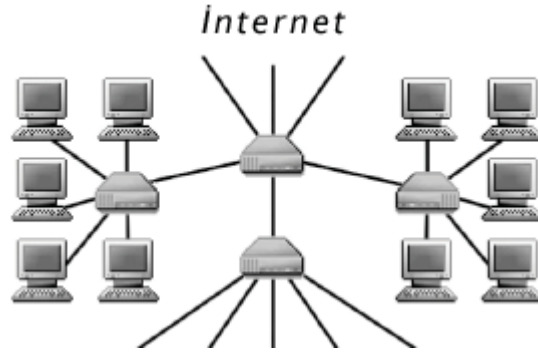
Yukarıda bahsedilen ilk ve son çalışmaların veri setleri arasındaki iki yıllık zaman diliminde WWW 5 kat büyümüş olmasına rağmen elde edilen power-law kuyruklu derece dağılımlarına ait γ_{in} katsayıları neredeyse aynıdır. γ_{out} katsayısı ise üzerinde çalışılan ağın büyüklüğü ile doğru orantılı olarak değişim göstermektedir (Çizelge 3.2).

İki düğüm arasında bulunabilecek en kısa yol, düğümler arası uzaklık olarak tanımlanmaktadır. Büyük ağ boyutuna rağmen World Wide Web ağı, small-world özelliği sergilemektedir. Bu durum ilk olarak Albert ve arkadaşlarının 325,729 düğümden oluşan WWW alt kümesi üzerindeki araştırmaları sonucu buldukları ortalama uzaklık değerinin 11.2 olması ile ortaya çıkmıştır [15]. WWW'nin 1999 yılındaki toplam boyutu olan 800 milyon düğüm için yaptıkları sonlu boyut ölçeklemesi ile bu değeri tüm WWW için yaklaşık 19 olarak tahmin etmişlerdir [15,23]. Ardından Broder ve arkadaşlarının 2000 yılında 50 milyon düğüm için yaptığı çalışmada buldukları 16 değeri, bu tahmini doğrular niteliktedir [21]. Adamic'in domain tabanlı çalışmasında ise bu değer 3.1 olarak bulunmuştur [24]. Aynı domaindeki tüm web sayfalarının tek düğüm kabul edilmesi, ortalama uzaklıkta dramatik bir düşüşü sonuç vermiştir.

WWW linklerinin yönlü yapısı, kümelenme katsayısının eşitlik 3.1 kullanarak hesaplanmasına izin vermemektedir. Bu durumu aşmanın bir yolu, linklerin yön bilgisini göz ardı etmek, yani linkleri yönsüz hale getirmektir. Adamic çalışmasında bu yolu izlemiştir [24]. Alexa veritabanından elde ettiği 279,794 domain içerisindeki 50 milyon web sayfasını içeren veri setini, tek link içeren düğümleri silerek 153,127 domaine indirgemıştır. Bu değişikliğin kümelenme katsayısını artırması zaten bekleniyor olsa da, bulduğu $C = 0.1078$ değerinin aynı büyüklük ve ortalama uzaklık değerine sahip rassal ağ için hesaplanan $C = 0.00023$ değerinden kat kat büyük olması, gerçel ağlarda kümelenmenin önemli derecede var olduğunu ortaya koymuştur.

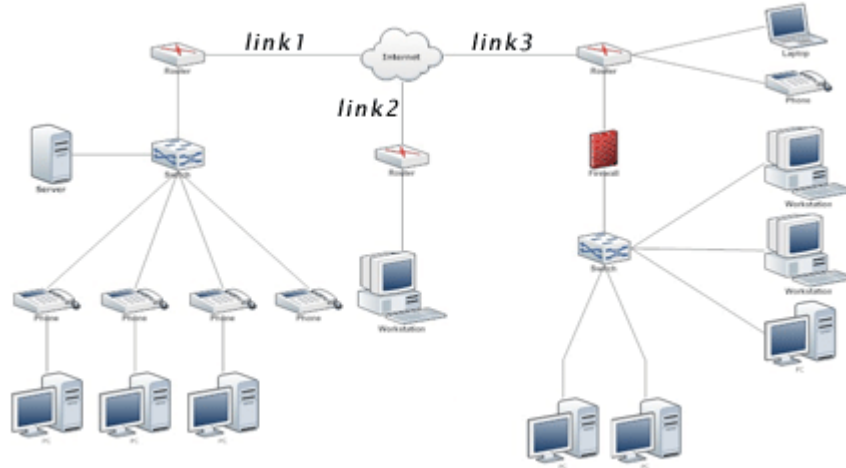
3.2.2. İnternet

İnternet, bilgisayarlar ve diğer iletişim araçları arasındaki fiziksel ya da kablosuz bağlantılardan oluşan bir ağıdır (Şekil 3.4). İnternette düğümler router ve bilgisayarlardır. Linkler ise bu düğümleri bağlayan kablolu ya da kablosuz (RF) bağlantılardır.



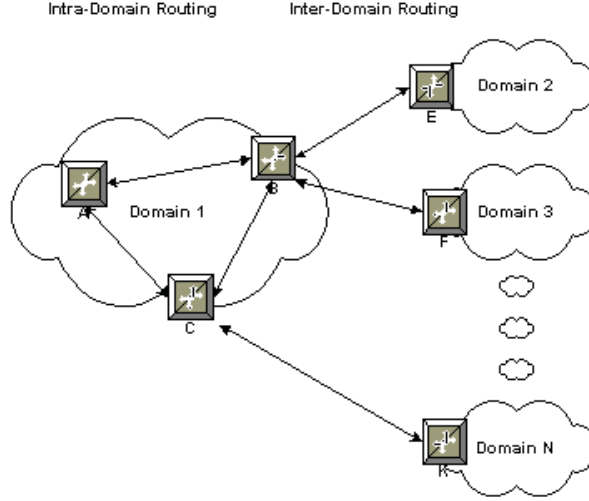
Şekil 3.4. İnterneti tanımlayan ağ şeması.

İnternet topolojisi iki farklı düzeyde değerlendirilebilir. Router düzeyinde incelendiğinde router'lar birer düğüm, aralarındaki fiziksel bağlantılar da link olarak kabul edilir (Şekil 3.5).



Şekil 3.5. İnternet ağına router düzeyinde bakış.

İnterdomain düzeyde incelendiğinde ise yüzlerce router ve bilgisayarlardan oluşan her bir domain bir düğümü, iki domaini birleştiren en az bir yol bulunduğu ve bu yol bir linki ifade etmektedir (Şekil 3.6).



Şekil 3.6. İnternet ağına interdomain düzeyde bakış.

İki düzeyde de yapılan çalışmaların çoğunun ortaya koyduğu derece dağılımları, power-law kuyruğa sahiptir. İnternetin interdomain topolojisi üzerinde 1997 başından 1998 yılı sonuna kadar Faloutsos ve arkadaşları tarafından 3 farklı tarihte yapılmış olan çalışmalara ait derece dağılımı grafikleri, $\gamma = 2.15$ ile $\gamma = 2.2$ arasında katsayılara sahip power-law dağılımları göstermiştir [25]. Router düzeyinde yapılan çalışmalarda da $\gamma = 2.3$ ile $\gamma = 2.48$ arasındaki katsayı değerlerine ulaşılmıştır [1].

Bir ağ olarak internet, kümelenme ve kısa düğümler arası mesafe özellikleri de göstermektedir. Yook ve arkadaşları ile Pastor-Satorras ve arkadaşları 1997-1999 yılları arasında internet üzerinde interdomain düzeyde yapmış oldukları çalışmalarda kümelenme katsayısı C 'yi 0.18 – 0.3 aralığındaki değerlerde tespit etmişlerdir [26,27]. Bu değerler, benzer parametrelere sahip rassal ağlara kıyasla ($C = 0.001$) oldukça yüksektir. Bu da kümelenmenin varlığının açık bir göstergesidir. Bu çalışmalarda interdomain düzeydeki ortalama düğümler arası uzaklık değerleri 3.7 - 3.77 aralığında, router düzeyinde ise 9 civarında bulunmuş, bu da bahsedilen ağların small-world karakteristiğine sahip olduğunu göstermiştir [1,26,27].

3.2.3. Aktör Birlikteliği Ağları

Üzerinde sıkça çalışılan bir başka ağ ise film aktörlerinin birlikteliklerinin oluşturduğu işbirliği ağlarıdır. Bu ağların oluşturulmasında, internetten de erişilebilen 1890'lardan bu yana çekilmiş tüm filmlere ait aktör birlikteliklerine ait veritabanları kullanılmaktadır. Bu ağlarda düğümler aktörlerden oluşmakta, aktörlerin aynı filmdeki birliktelikleri ise linkleri oluşturmaktadır. 2000'li yılların başlarında yaklaşık 500,000 düğümden oluşan ve günümüzde de halen genişlemeye devam eden aktör ağı üzerinde Watts ve Strogatz'ın elde etmiş olduğu ortalama düğümler arası uzaklık değeri (3.65), aynı parametrelere sahip rassal ağa göre (2.9) önemli bir fark sergilememiştir [8]. Ancak kümelenme katsayısının bir rassal ağa göre 100 kat büyük olduğunu gözlemlemişlerdir [1].

Derece dağılımı grafiğinin diğer gerçel ağlardaki gibi power-law dağılımına uyumlu olarak gözlemlendiği bu çalışmalarda $\gamma = 2.3 \pm 0.1$ katsayısı da yukarıda değindiğimiz ağlara ait katsayılar ile uyum içerisindedir.

3.2.4. Bilimsel İşbirliği Ağları

Aktör ağlarının bir benzeri de bilim adamları için oluşturulabilir. Düğümlerin bilimsel makale yazarlarından oluştuğu, yazarlar arasındaki yayın birlikteliklerinin ise birer link olarak tanımlandığı bu ağlar, bilimsel işbirliği ağları olarak adlandırılmaktadır. Bu kompleks ağ türünün topolojisini anlamak için Newman [12,13], 1995-1999 yılları arasındaki 4 farklı disiplindeki (fizik, biyomedikal, yüksek enerji fiziği ve bilgisayar bilimleri) bilimsel yayınların oluşturduğu bilimsel işbirliği ağlarını incelemiştir.

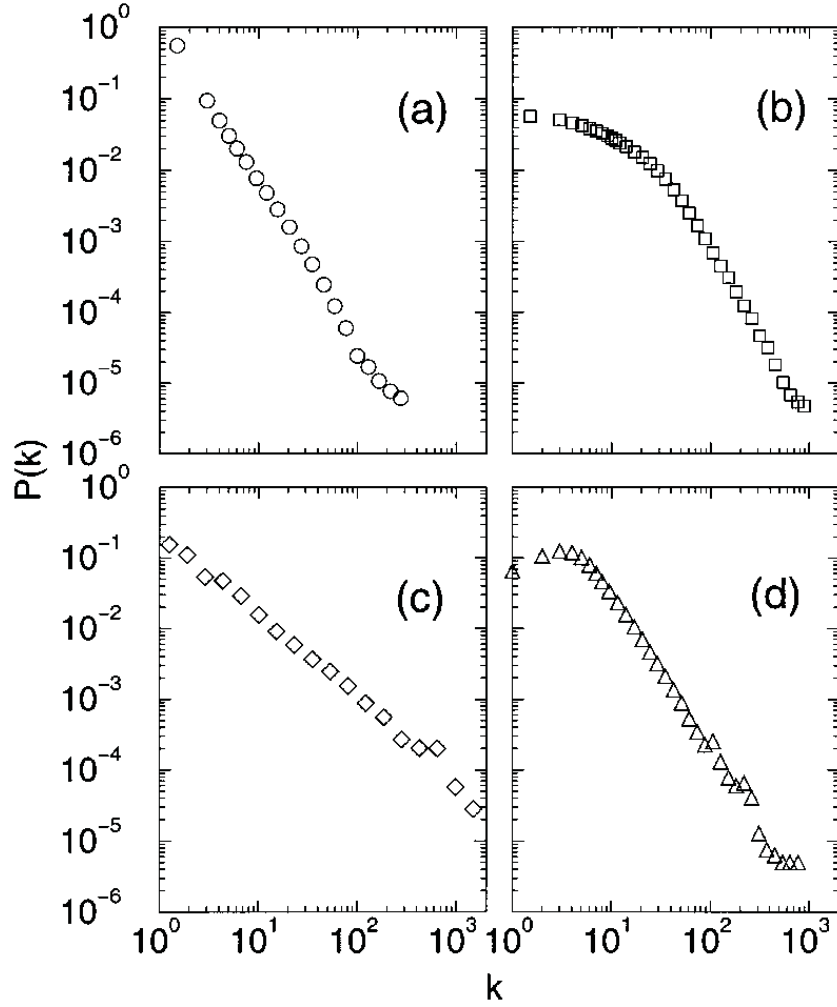
Tüm bu ağlar Çizelge 3.1'de de özetlendiği şekliyle, düşük düğümler arası uzaklık ve yüksek kümelenme katsayısı değerleri sergilemişlerdir. Özellikle yüksek enerji fiziğine ait derece dağılımı incelendiğinde (Şekil 3.7c), $\gamma=1.2$ katsayısı ile kusursuz bir power-law uyumluluğu görülmektedir. Şekil 3.7'deki diğer derece dağılımı grafikleri de daha büyük katsayılar ile yine power-law kuyruğa sahiptirler.

Çizelge 3.1. Çeşitli gerçek ağlara ait genel karakteristikler. Her bir ağ için düğüm sayısı (Size), ortalama derece ($\langle k \rangle$), düğümler arasındaki ortalama uzaklık (l) ve kümelenme katsayısı (C) değerleri verilmiştir. Karşılaştırma amaçlı olarak aynı büyüklüğe ve ortalama derece değerine sahip rassal ağlara ait l_{rand} ve C_{rand} değerleri de verilmiştir [1].

Network	Size	$\langle k \rangle$	l	l_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook <i>et al.</i> , 2001a, Pastor-Satorras <i>et al.</i> , 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási <i>et al.</i> , 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási <i>et al.</i> , 2001	9
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460 902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook <i>et al.</i> , 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Çizelge 3.2. $P(k)$ 'nin power-law dağılımında olduğu çeşitli scale-free ağlara ait parametreler. Her bir ağa ait düğüm sayısı (Size), ortalama derece ($\langle k \rangle$), power-law dağılımının *cutoff* değeri (κ), düğümlere gelen ve düğümlerden çıkan bağlantılara ait derece dağılımları (γ_{in} ve γ_{out} – yönsüz ağlar için birbiri ile aynıdır ve * ile işaretlenmiştir), düğümler arasındaki ortalama uzaklık (l) değerleri verilmiştir. l_{real} , l_{rand} ve l_{pow} sırasıyla aynı boyuttaki gerçek, rassal ve power-law dağılımlı ağlar için ortalama düğüm uzaklığı değerleridir [1].

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	l_{real}	l_{rand}	l_{pow}	Reference	Nr.
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999	1
WWW	4×10^7	7		2.38	2.1				Kumar <i>et al.</i> , 1999	2
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder <i>et al.</i> , 2000	3
WWW, site	260 000				1.94				Huberman and Adamic, 2000	4
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999	5
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999	6
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000	7
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999	8
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b	9
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási <i>et al.</i> , 2001	10
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási <i>et al.</i> , 2001	11
Sexual contacts*	2810			3.4	3.4				Liljeros <i>et al.</i> , 2001	12
Metabolic, <i>E. coli</i>	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong <i>et al.</i> , 2000	13
Protein, <i>S. cerev.</i> *	1870	2.39		2.4	2.4				Jeong, Mason, <i>et al.</i> , 2001	14
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000	14
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000	16
Citation	783 339	8.57			3				Redner, 1998	17
Phone call	53×10^6	3.16		2.1	2.1				Aiello <i>et al.</i> , 2000	18
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001	19
Words, synonyms*	22 311	13.48		2.8	2.8				Yook <i>et al.</i> , 2001b	20

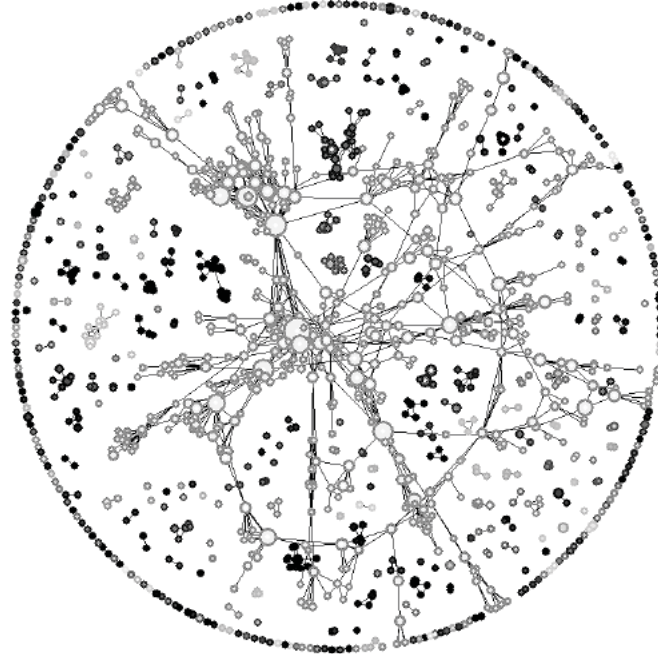


Şekil 3.7. Çeşitli gerçel ağlara ait derece dağılımı grafikleri. a) Router düzeyinde incelenen internet ağı; b) Aktör ağı; c) Yüksek enerji fiziği alanındaki bilimsel işbirliği ağı; d) Biyomedikal alanındaki bilimsel işbirliği ağı [1].

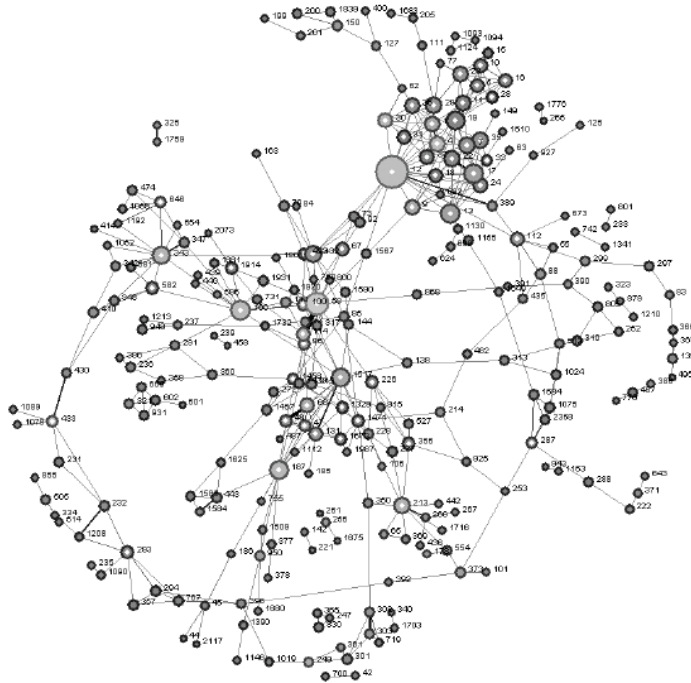
Bu alandaki önemli kilometre taşlarından birisi de Barabási ve arkadaşlarının 2002 yılında, 1991-1998 yılları arasındaki verilere dayanan matematik ve sinirbilim alanındaki bilimsel işbirliği ağları üzerinde yapmış oldukları çalışmadır [7]. Bu çalışmada düğümler arası ortalama uzaklık değerlerini $l_{math} = 9.5$ ve $l_{nsi} = 6$, kümelenme katsayılarını ise $C_{math} = 0.59$ ve $C_{nsi} = 0.76$ olarak bulmuşlardır. Bu ağlara ait derece dağılımları da sırasıyla 2.1 ve 2.5'lik katsayılara sahip power-law dağılımları ile uyum göstermektedir (Şekil 3.7d).

Bilimsel işbirliği ağları ile ilgili yapılan çalışmalardan bir kısmında ağların görselleştirildiğini görmekteyiz. Moltka-Danielsen ve arkadaşlarının 1978'den beri

yapılmakta olan IRIS konferanslarının 1978-2006 yıl aralığını kapsayan bilimsel işbirliği ağı çalışmasına ait görsel çıktılar Şekil 3,8’de görülmektedir [28].



a) Tüm ağa ait örüntü



b) En az 3 yayına katkıda bulunmuş 256 yazardan oluşan alt küme (ortalama derece : 4.182)

Şekil 3.8. IRIS konferanslarının tanımladığı bilimsel işbirliği ağı. [28].

541 yazardan oluşan bu çalışmada ortalama derece 4.3, düğümler arası ortalama uzaklık 6.38 olarak hesaplanmıştır.

3.2.5. Biyolojik Ağlar

Jeong ve arkadaşları, 2000 yılında 3 ana canlı türünü temsil eden 43 farklı organizma üzerinde bir kompleks ağ çalışması yapmıştır.

Bu çalışmada canlı organizmalardaki ATP, ADP gibi enzimler düğümleri, bu enzimlerin yer aldığı kimyasal reaksiyonlar ise düğümler arasındaki yönlü linkleri oluşturmuştur. Düğümlerin derece dağılımlarını incelediklerinde ise tüm organizmalar için yine power-law dağılımları gözlemlemiş, γ katsayılarının ise 2.0 - 2.4 aralığında olduğunu tespit etmişlerdir [29].

Oluşturdukları ağlar yönlü yapıda olduğu için kümelenme katsayısı hesaplanamamıştır. Bununla birlikte bu çalışmanın ilgi çeken sonucu, tüm organizmalar için düğümler arası ortalama uzaklık değerleri yaklaşık 3.3 olarak gerçekleşmiş olmasıdır [30].

Biyolojik sistemler üzerinde yapılmış olan çalışmalardan Çizelge 3.1'de de yer verdiğimiz *Escherichia Coli* bakterisi enerji ve biyosentez metabolizması üzerinde yine diğer çalışmalara benzer derece dağılımı grafiği elde edilmiş, ağın yönsüz olarak çalışılan versiyonunda küçük düğümler arası uzaklık ve yüksek kümelenme katsayısı değerleri elde edilmiştir.

İlerleyen yıllarda protein zincirlerinin oluşturduğu ağlar üzerinde yapılan çalışmalar da yine benzer sonuçlar vermiş, small-world konseptinin canlı organizmaların neredeyse her alanında geçerli bir kural olduğu ortaya koyulmuştur [1].

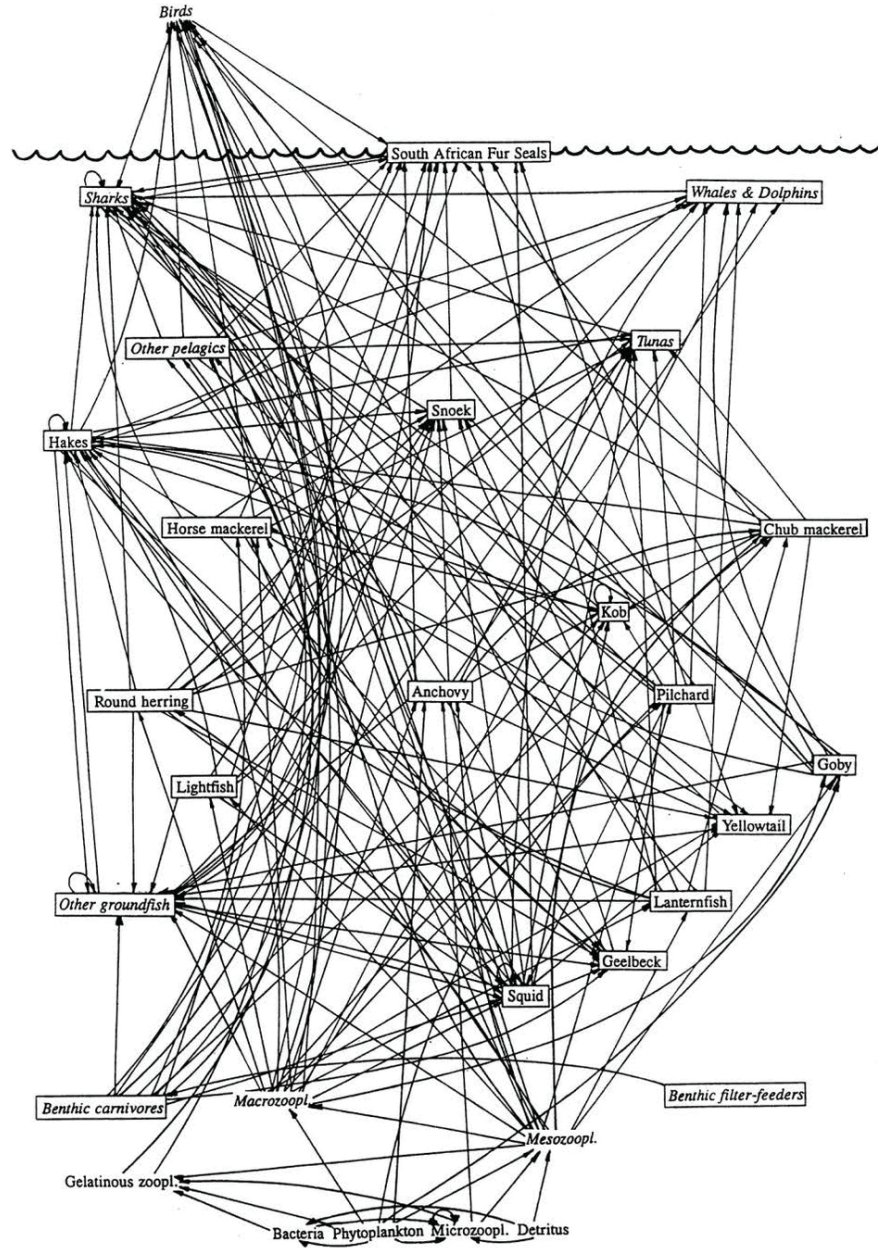
3.2.6. Ekolojik Ağlar

Beslenme ağları, ekoloji biliminde çeşitli canlı türleri arasındaki ilişkilerin ortaya koyulmasında sıklıkla araştırma konusu olmaktadır. Kompleks ağ çalışmalarında ise canlı türleri birer düğüm, bu türler arasındaki avcı-av ilişkileri ise linkleri

tanımlamaktadır. Şekil 3.9'da Güney Afrika'nın batı sahilindeki Benguela ekosisteminden bir kesit gösterilmiştir.

Williams ve arkadaşları, en çok bilinen ve dokümantasyona sahip olan 7 besin ağı üzerinde yaptıkları çalışmada, canlı türü sayısı ve ortalama derece açısından çok farklı yapıdaki bu ağların tümünde, tüm türlerin birbirine 3 ya da daha az sayıda düğüm kadar uzaklıkta olduğunu tespit etmişlerdir [31]. Derece dağılımının elde edilmesinde ise aslında yönlü yapıdaki ağdaki yön bilgileri göz ardı edilmekte, bu şekilde dağılım karakteristikleri elde edilebilmektedir.

Ekolojik ağlar konusundaki diğer çalışmalar da small-world konseptine uygun sonuçlar vermiştir. Ancak ekolojik ağları tanımlayan veritabanlarının kısıtlı yapısı ve bu ağları oluşturan düğüm sayısının da sınırlı olması nedeniyle, derece dağılımı grafikleri de her ne kadar power-law dağılımında olsalar da küçük γ katsayıları ile (~ 1.1) diğer çalışmalardan ayrılmaktadır. Daha geniş veri setleri ile bu farklılığın giderileceği düşünülmekte iken, veri toplama prosesinin nispeten yavaş ilerlemesi nedeniyle yakın zamanda yeterli büyüklükte bir ağ çalışması yapılması beklenmemektedir [1].



Şekil 3.9. Güney Afrika'nın batı sahilindeki Benguela ekosisteminden bir kesit [32].

3.2.7. Telefon Ağları

Telefon numaralarının düğümleri, bu numaralar arasındaki aramaların ise linkleri oluşturduğu telefon ağları, yönlü yapısıyla kompleks ağ çalışmaları için bir başka çalışma alanı oluşturmaktadır. Aslında çok geniş bir ağ tasarımına imkan vermekle birlikte veri işleme zorluklarını azaltmak adına bir ya da birkaç günlük telefon

aramalarının oluşturduğu ağlar incelenmiş, $\gamma_{out} = \gamma_{in} \cong 2.1$ civarında katsayılara sahip power-law dağılımına uyumlu derece dağılımları elde edilmiştir [1].

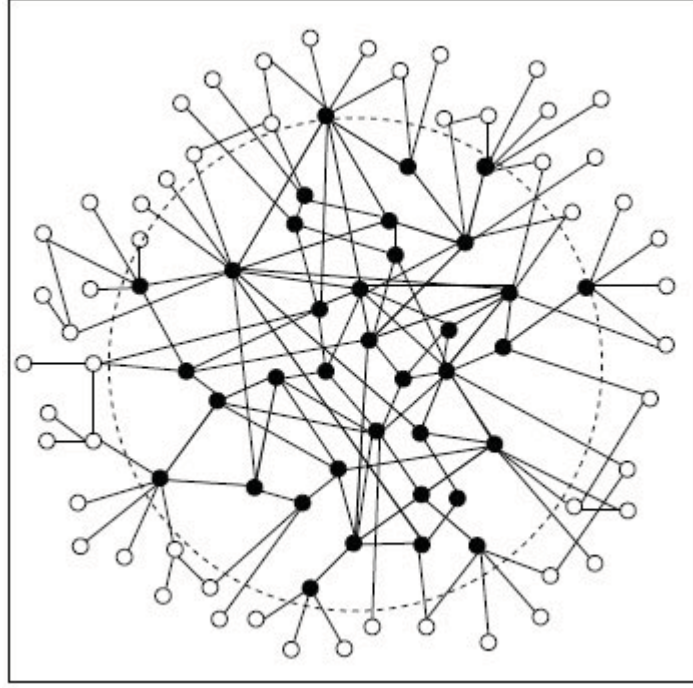
3.2.8. Bilimsel Atıf Ağları

Yukarıdaki örneklere göre daha kompleks yapıdaki bir ağ türü olan bilimsel atıf ağlarında bilimsel yayınlar birer düğüm olarak kabul edilirken, her yayından kendisinden önce yapılmış diğer yayınlara yapılan atıflar (referanslar) ise ağdaki linkleri oluşturmaktadır. Redner, Institute for Scientific Information tarafından indekslenmiş 783,339 adet yayın ve *Physical Review D* dergisinde 1975-1994 yılları arasında yapılmış olan 24,296 yayının atıf dağılımı üzerinden oluşturduğu kompleks ağda, bir yayına k defa atıf yapılmış olması ihtimaline ait dağılımın (derece dağılımı) $\gamma_{cite} = 3$ katsayısına sahip bir power-law eğrisi ile uyum içinde olduğunu bulmuştur [33].

Sonraki yıllarda bir yayına gelen atıflarla birlikte bir yayından çıkan atıfların da dikkate alındığı yönlü ağ çalışmasında Vazquez, yayından çıkan atıfların derece dağılımının exponansiyel dağılıma sahip olduğunu tespit etmiştir [34].

3.2.9. Kelime Ağları

Dillerin kompleks yapısı, kompleks ağ tanımlama ve çalışmalarına uygun bir zemin teşkil etmektedir. Ferrer i Cancho ve Sole, 2001 yılında *British National Corpus*'u baz alarak İngiliz dili için bir kelime ağı tanımlamışlardır.



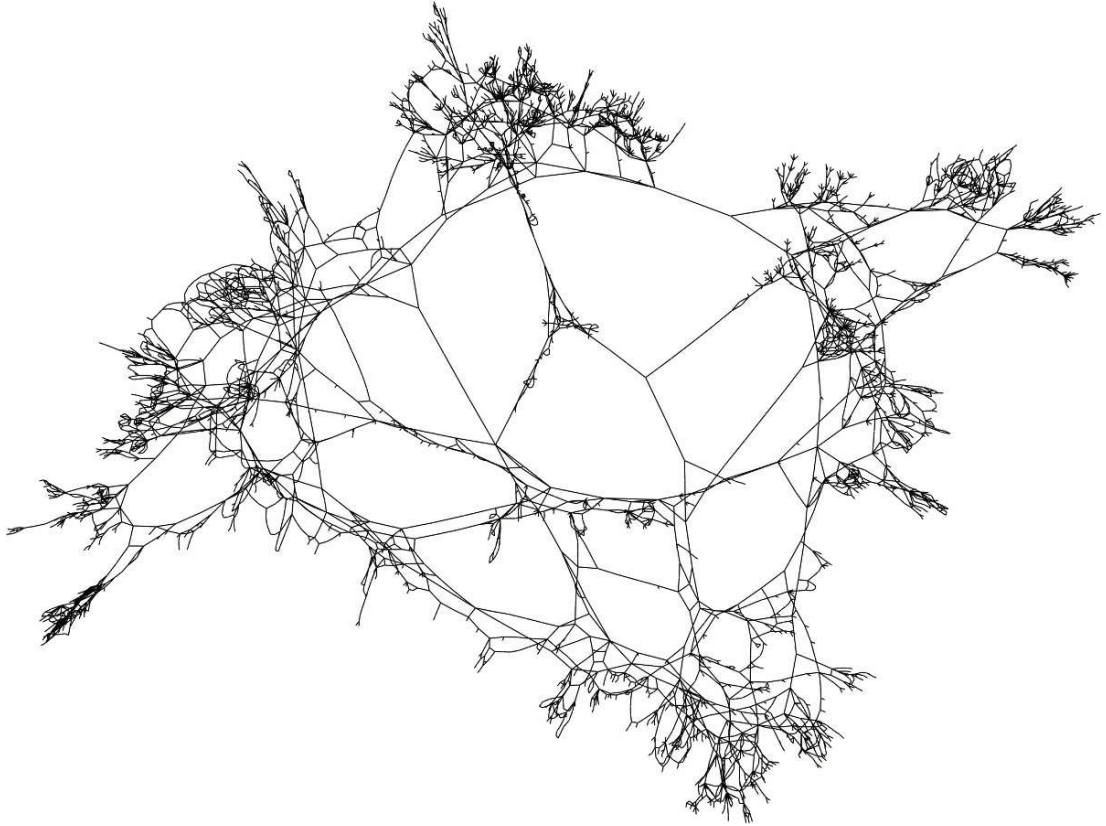
Şekil 3.10. Bir dildeki kelimelerin oluşturduğu metine ait muhtemel bir örüntü. Koyu renkli düğümler sık kullanılan kelimeleri, açık renkli düğümler ise nadir kullanılan kelimeleri temsil etmektedir. Birlikte kullanılan iki kelime arasında linkler tanımlanmıştır [35].

Her bir kelimenin bir düğümü temsil ettiği bu ağda, bir cümlede yan yana ya da bir kelime uzaklıktaki kelimeler arasında linkler oluşturulmuştur. Sonuç olarak 440,902 kelimedenden oluşan bu ağın küçük düğümler arası mesafe ($l = 2.67$) ve yüksek kümelenme katsayısı ($C=0.437$) değerleri ile birlikte iki rejimli bir power-law derece dağılımına sahip olduğunu bulmuşlardır. Derecesi $k \leq 10^3$ olan düğümler $\gamma \leq 1.5$ katsayısı, derecesi $10^3 < k < 10^5$ olan düğümler $\gamma \geq 2.7$ katsayısına sahip power-law eğrileri ile uyum göstermiştir [35]. Şekil 3.10'da bu ağa ait bir alt küme görülmektedir.

Kelimeleri üzerinde yapılan farklı çalışmalarda da (örn. eşanlımlı kelimeler arasında linklerin tanımlanması ile) bu ağların organizasyon prensiplerinin diğer ağ yapılarındakilere benzer özellikler gösterdiği ortaya koyulmuştur [1].

3.2.10. Enerji Nakil Hatları ve Sinir Ağları

Watts ve Strogatz, batı ABD enerji nakil hatlarını bir kompleks ağ olarak tanımlarken, tüm jeneratör, transformator ve trafoları birer düğüm kabul etmiş, bu düğümler arasındaki iletim hatlarını da linkler olarak tanımlamışlardır (Şekil 3.11). Bu ağdaki toplam düğüm sayısı 4941 olurken, düğümlerin ortalama derecelerini ise $\langle k \rangle = 2.67$ olarak bulmuşlardır. *Nematod kurdu*'nu oluşturan küçük sinir ağını (N=282) da inceleyerek bir kompleks ağ tanımlamışlar, her bir sinir hücrelerini düğüm, bu hücreler arasındaki sinaps ya da boşluklu bağlantıları da birer link olarak kabul etmişlerdir. Oluşturdukları her iki ağda da ortalama düğümler arası uzaklığın, aynı boyut ve ortalama dereceye sahip rassal ağlarla yaklaşık aynı değerde olduğunu bulmuşlardır. Bununla birlikte kümelenme katsayısı ise çok daha yüksek değerde gözlenmiştir (Çizelge 3.1). enerji nakil hattına ait derece dağılımı eksponansiyel dağılıma sahipken, inceledikleri sinir ağının ortalama bir k değerinde zirve yaptıktan sonra eksponansiyel düşüş gösterdiği gözlenmiştir [8].



Şekil 3.11. Batı ABD enerji nakil hatları ağı [8].

3.2.11. Protein Ağları

Proteinler, yapılarını belirleyen katlanma sürecinde ardışıl dönüşümler geçirirler. Bu dönüşümler arasındaki her bir durumları bir düğüm olarak kabul edilirse, tek bir dönüşüm hareketi ile birbirine dönüşebilen komşu durumlar arasında tanımlanacak linkler ile bir kompleks ağ yapısı tanımlanabilir. Scala ve arkadaşları iki boyutlu bir kafes polimerin dönüşüm proseslerinin tanımladığı kompleks ağ üzerinde yaptıkları çalışmalarında ağın small-world karakteristiğine sahip olduğunu gözlemlemiştirler [36]. Spesifik olarak, polimerin boyutu (dolayısıyla ağ boyutu) arttıkça, düğümler arasındaki ortalama uzaklığın rassal ağlardakine benzer biçimde logaritmik olarak arttığını ortaya koymuşlardır. Bununla birlikte kümelenme katsayısının bir rassal ağdakine göre çok büyük olduğunu ve rassal ağlardan farklı olarak ağın boyutu arttıkça artma eğiliminde olduğunu göstermişlerdir.

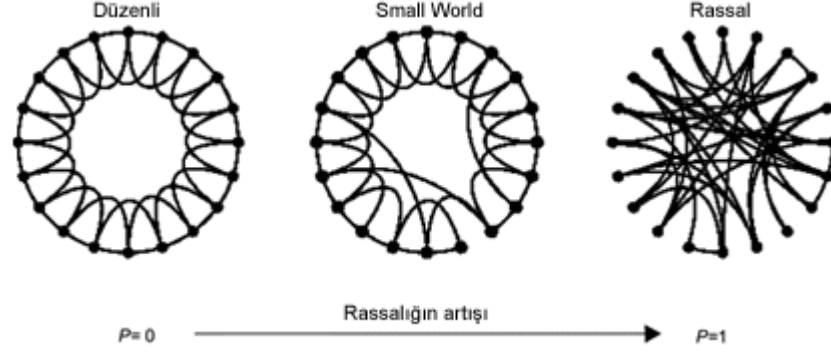
Yukarıda bahsedilen tüm veritabanları ve bu veritabanları yapılmış olan kompleks ağ çalışmaları, gerçel ağların topolojik özelliklerin keşfedilmesi açısından birer ilham kaynağıdır. Bu çalışmaların ortaya koyduğu ağ parametrelerinden yola çıkarak bir takım ağ modelleri oluşturulmaya çalışılmış, çeşitli teorik yaklaşımlar geliştirilmiştir. Erdos-Renyi modeli, Watts-Strogatz modeli, genelleştirilmiş rassal ağlar, small-world ağlar, scale-free ağlar, evrilen ağlar bunlardan sadece bir kaçıdır [1].

3.3. SMALL-WORLD AĞ MODELİ

Small-world terimi, ağdaki düğümler arasındaki ortalama uzaklığın yeterince küçük olduğu ağlar için kullanılmaktadır. Yani rastgele seçilen iki düğüm arasında görece kısa bir yol bulunabilmesi ve bu durumun ağdaki düğümlerin çoğu için geçerli olması durumudur [3,10]. Yukarıda bahsi geçen gerçel ağ çalışmaları, aynı topolojik özelliklere sahip ağların modelleme yolu ile de tasarlanabilmesi için birer referans olmuşturlardır.

Gerçel ağların rassal ağlardaki gibi small-world özellikleri sergiledikleri, ancak rassal ağlara göre daha büyük kümelenme katsayılarına sahip oldukları

gözlenmektedir. Aynı zamanda bu çalışmalardan bazıları, kümelenme katsayısının ağıın büyüklüğünden bağımsız olduğunu da ortaya koymaktadır.



Şekil 3.12. Watts-Strogatz ağ modelinde rastgele bağlanmalar arttıkça ağıın topolojisinin deęişimi [8].

Örneęin Şekil 3.12'deki tek boyutlu grid yapıyı inceleyelim. Her düęüm, kendisine en yakın K adet düęüme baęlandığında düęümler arasında çok sayıda ortak komşular yer alacaęından dolayı yüksek kümelenme katsayısına sahip bir ağ ortaya çıkar. Böyle bir grid yapı için kümelenme katsayısı şu şekilde hesaplanabilir:

$$C = \frac{3(K - 2)}{4(K - 1)} \quad (3.3)$$

Bu eşitlikten anlaşılacaęı üzere düşük boyutlu düzenli ağlar için K deęeri arttıkça kümelenme katsayısı $3/4$ 'e yakınsamaktadır. Ancak düęümler arası uzaklık deęeri gerçel ağlardaki kadar küçük olmamaktadır.

Düşük düęümler arası uzaklık ve yüksek kümelenme katsayısı deęerlerine sahip bir ağ oluşturmak için yapılan ilk başarılı ağ modeli, Watts ve Strogatz'ın 1998 yılındaki çalışmasıdır [8]. Kendisinden sonra gelen dięer modellerin de dayanaęı olan bu modelin detaylarına biraz göz atalım.

Watts Strogatz modeli : Watts ve Strogatz, 1998 yılında, sonlu büyüklükteki düzenli grid yapı ile rassal ağların arasında yer alacak tek parametrelili bir model önermişlerdir. Bu modele ait algoritma şu şekildedir:

- i) *Düzenli kafes yapıyı oluştur:* Her düğümün kendisine en yakın K adet komşusuna bağlandığı, N adet düğümden oluşan, halka şeklindeki bir düzenli kafes yapı oluştur. Seyrek fakat birbirine yeterince bağlı bir ağa sahip olmak için $N \gg K \gg \ln(N) \gg 1$ 'e bağlı kal.
- ii) *Rassallaştır:* Ağdaki mevcut linkleri p olasılığı ile değiştirerek, rastgele başka düğümlere linkler tanımla. Tekrarlı linkleri ve kendi kendine bağlanan linkleri iptal et. Bu işlem, düğümleri i . aşamada komşu olamayacakları uzak düğümlere bağlayan $pNK/2$ yeni linki sisteme eklemektedir. p değeri değiştirilerek ağın düzenli ($p=0$) ve rassal ($p=1$) yapı arasındaki değişimi kolaylıkla izlenebilir.

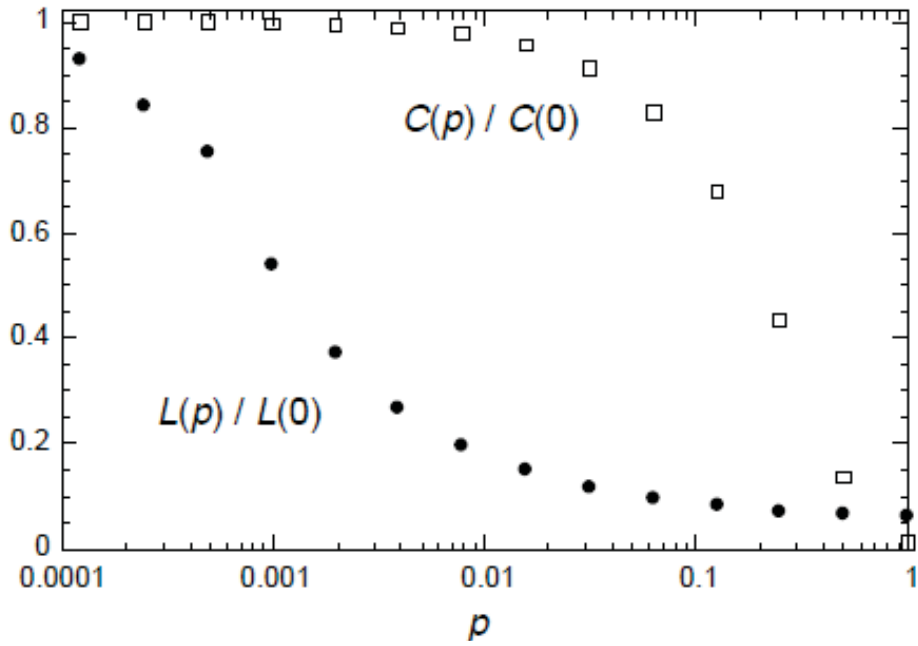
Bu modelin kökleri aslında çoğu insanın yakın komşuları ve onların komşuları ile arkadaş olduğu sosyal sistemlerde mevcuttur. Komşuluk ilişkisi aynı sokakta yaşayan, aynı işyerinde çalışan ya da aynı arkadaş çevresine mensup olan insanlar arasında olabilir. Bununla birlikte, her insanın bulunduğu çevreden çok daha uzakta (başka şehirde ya da ülkede) bir ya da birkaç arkadaşı vardır. Bu arkadaşlık ilişkileri ise Watts-Strogatz (WS) modelindeki *ii*. aşamada oluşturulan rastgele linklerle temsil edilmektedir [8].

Düşük düğümler arası uzaklık ve düşük kümelenme katsayılarının birlikte oluşumunu anlayabilmek için Watts ve Strogatz, bu parametrelerin yeniden bağlanma olasılığı p 'ye bağlı değişimlerini incelemişlerdir (Şekil 3.12). Şekilde de görülebileceği gibi, düzenli kafes yapı için ($p=0$), $l(0) \cong N/2K \gg 1$ ve $C(0) \cong 3/4$ olacaktır. Bu durumda l , ağın büyüklüğü ile doğru orantılı, C ise yüksek bir değerdedir. Öte yandan $p \rightarrow 1$ durumu için model, $l(1) \sim \ln(N)/\ln(K)$ ve $C(1) \sim K/N$ 'e yakınsayan bir rassal ağa dönüşmektedir. Bu durumda ise l , N 'e logaritmik olarak bağlı, C ise N ile ters orantılıdır.

Bu limit durumları, büyük l 'nin her zaman büyük C ile, küçük l değerlerinin ise küçük C ile bağlantılı olduğunu akla getirebilir. Ancak Watts ve Strogatz, Şekil 3.12'de de görülebileceği gibi, $l(p)$ 'nin $l(1)$ 'e yakın olduğu, ancak $C(p)$ 'nin $C(1)$ 'den çok büyük olduğu geniş bir aralığın varlığını keşfetmişlerdir. Bu aralık, aslında

$l(p)$ 'nin küçük p değerlerinden itibaren gösterdiği hızlı düşüş ve aynı bölgede $C(p)$ 'nin neredeyse sabit kalması ile açıklanabilir [8].

Kümelene katsayısı yüksek fakat düğümler arası mesafeleri düşük olan ağlar da bu aralıktaki p değerlerine sahip ağlardır. Önceki bölümde incelediğimiz gerçel ağların çoğu da bu tanıma uymaktadır. Şekil 3.12 ve Şekil 3.13'ün orta bölgeleri ile temsil edilen bu ağlar *small-world* olarak adlandırılırlar.



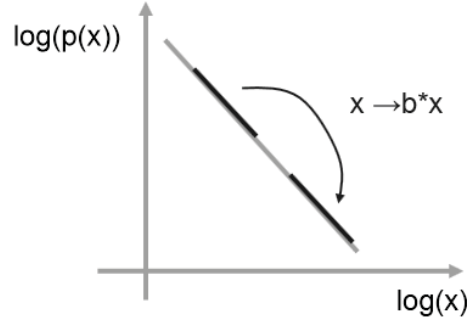
Şekil 3.13. WS modeli için düğümler arası uzaklık $l(p)$ ve kümelene katsayısı $C(p)$ değerlerinin değişimi. Parametreler, düzenli kafes yapılarıdaki değerleri $l(0)$ ve $C(0)$ ile normalize edilmişlerdir [8].

3.4. SCALE FREE AĞ MODELİ

Scale free ağlar, rassal ağlar ve small world ağlara oranla bilginin, hastalıkların, virüslerin daha hızlı yayıldığı bir modeldir [11]. Bu modelde ağa yeni bağlanan düğümler, bağlanmak için rastgele düğümler seçmek yerine hub olarak adlandırılan popüler düğümlere bağlanmayı tercih etmektedirler. Bu durum da ağda az sayıda hub düğümlerin ve çok sayıda normal düğümlerin varlığını destekler. Örneğin viral bir hastalığın toplumda yayılması örneğinde, virüs ağdaki hub'lardan birine denk gelirse

çok daha hızlı yayılacaktır. Small world ve rassal ağlarda ise bu tip düğümler çok daha az ihtimalle bulunur [1].

Scale-free, terim olarak power- dağılıma sahip olmanın matematiksel bir ifade edilmiş biçimidir. Derece dağılım grafiği log-log ekseninde çizdirildiğinde, grafiğin hangi kısmına ($k=[1-5]$ aralığına ya da $k=[20-100]$ aralığına) baktığımızdan bağımsız olarak, power-law uyumlu bir eğri aynı karakteristiğe (eğime) sahip olacaktır (Şekil 3.14). Bu durumu ifade etmek için scale free yani “ölçeksiz” tanımlaması tercih edilmiştir.



Şekil 3.14. Scale-free ağlara ait power-law derece dağılımı. x 'i b çarpanı ile ötelemek, dağılımın şeklini değiştirmemekte, sadece grafiğin daha sağ taraftaki değerlerine karşılık gelecek öteleme anlamına gelmektedir ($p(bx) = (bx)^{-\alpha} = b^{-\alpha} x^{-\alpha}$).

Bölüm 3.2'de incelenen gerçel ağların çoğu scale free sınıflandırmasına dahildir. Yani büyük k değerleri için derece dağılımları power-law karakteristiğine sahiptirler. Bununla birlikte, $P(k)$ 'nin eksponansiyel kuyruğa sahip olduğu ağlarda bile derece dağılımı grafiği, Poisson dağılımından oldukça farklı yapıdadır. Rassal graf teorisi ve WS modeli, bu dağılım karakteristiğini karşılamaktan uzaktır [1,8]. Bu eksikliği gidermek için başvurulan yaklaşımlarda birisi, rassal graf teorisinde bazı değişiklikler yaparak $P(k)$ 'nin eksponansiyel kuyruğa sahip olmasını sağlamaktır. *Genelleştirilmiş rassal ağlar* olarak adlandırılan bu model, giriş parametresi olarak sadece derece dağılımını almakta, bunun dışında kalan tüm yönlerden rassallığı korumaktadır. Bir başka deyişle, linkler yine rastgele seçilmiş düğümleri bağlayacak, fakat bu düğümlerin seçiminde, derece dağılımının power-law karakteristiğinde olması ölçütü gözetilecektir.

Yarı-rassal olarak da adlandırılabilir bu model, gerçel ağların derece dağılımını taklit edebiliyor olsa da, bir sorunun cevabını geciktirmekten öteye gidememektedir: Scale-free ağların oluşumun altında yatan mekanizma tam olarak nedir? Bu sorunun cevabını Barabási-Albert modelinde bulmak mümkündür.

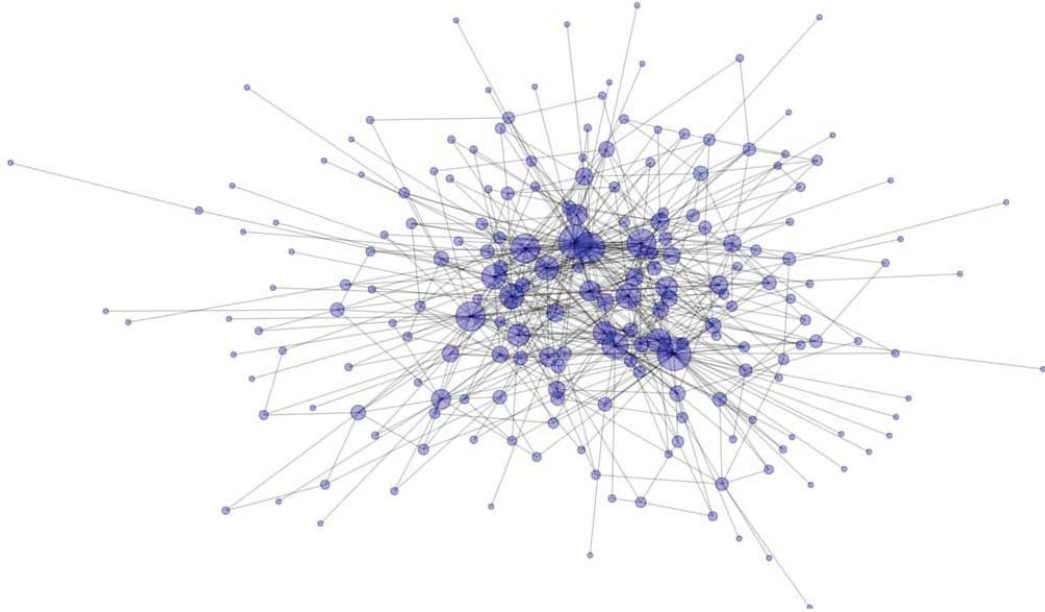
Barabási-Albert Modeli : Gerçel ağların sahip oldukları scale-free derece dağılımının sebeplerini anlayabilmek için ağ topolojisinin modellenmesi yerine, ağın oluşum ve gelişiminin modellenmesi yaklaşımına geçiş yapmak gerekmektedir. Aslında rassal ağlar ve small-world ağların modellenmesindeki anlayıştan tamamen farklı olmamakla birlikte, power-law derece dağılımına sahip bir ağ tasarlayabilmek için bakış açısında önemli bir değişikliğe gitmek gerekir. Önceki modellerde amaç, istenen topolojik özelliklere sahip bir ağı tasarlayabilmek iken, scale-free ağ modeli oluşturabilmek için ağın dinamiklerine hakim olmak gerekmektedir. Yani günümüzün dinamik ve evrilen ağlarını sonuç veren sebepler tam olarak tespit edilebilirse, bu kriterler ile oluşturulacak model zaten istenen topolojiye sahip olacaktır. Öyle ise bu modelleme yaklaşımının sürücü faktörü ağ dinamikleri, sonucu ise topoloji olacaktır.

Gerçel ağlardaki power-law derece dağılımını ilk defa ortaya koyan isimler Barabási ve Albert'tir [6,7]. 1999 yılında yaptıkları öncü çalışmada, gerçel ağların scale-free yapısına uygun bir model tanımlamışlardır. Bu modelin tanımlanmasında gerçel ağların oluşumunda etkin rol alan iki önemli mekanizma üzerinde durmuşlardır.

Öncelikle, Barabási ve Albert'in scale free ağ modeli çalışmasına kadar öne sürülmüş tüm ağ modelleri sabit düğüm sayısı (N) varsayımı altında, sadece ağdaki düğümleri bağlayan linklerin modifiye edildiği modellerdir. Fakat gerçek ağların büyük çoğunluğu, yeni düğümlerin katılımı ile sürekli büyüyen ağlardır. Birkaç çekirdek düğüm ile başlayan bu ağlara ömürleri boyunca birçok yeni düğüm eklenmektedir. Örneğin WWW, yeni web sayfalarının ağa eklenmesi ile eksponansiyel olarak büyümektedir. Aynı şekilde araştırma literatürü de yeni yayınlar ile sürekli genişlemektedir.

İkinci önemli nokta ise önceki modellerde iki düğümün rastgele bağlanması işleminin, düğümlerin o anki derecelerinden bağımsız olmasıdır. Yani düğümler arasındaki linkler eşit olasılıkla oluşturulmaktadır. Ancak çoğu gerçel ağda *seçimli bağlanma* olgusu vardır. Yani bir düğüme bağlanma olasılığı, o düğümün derecesi ile doğru orantılıdır. Örneğin bir web sayfasının, derecesi yüksek olan popüler web sayfalarına link verme ihtimali daha yüksektir çünkü derecesi yüksek web sayfalarının bulunmaları daha kolaydır ve daha iyi bilinirler. Ya da yeni bir bilimsel yayının, iyi bilinen ve çok atıf almış yayınlara atıf içermesi ihtimali, daha az atıf almış dolayısıyla daha az bilinen yayınlara atıf içermeye ihtimalinden daha yüksektir.

Gerçel ağ topolojilerini etkileyen bu iki özellik (büyüme ve seçimli bağlanma), ilk defa power-law derece dağılımına sahip bir ağ tasarlayabilecek modelin oluşturulmasında Barabási ve Albert'e ilham kaynağı olmuştur.



Şekil 3.15. Barabási-Albert modeli ile oluşturulmuş, 4 düğümlle başlayarak her adımda yeni bir düğümün 4 farklı düğüme bağlanması ile 200 düğüme ulaşan oluşan bir ağ yapısı [37].

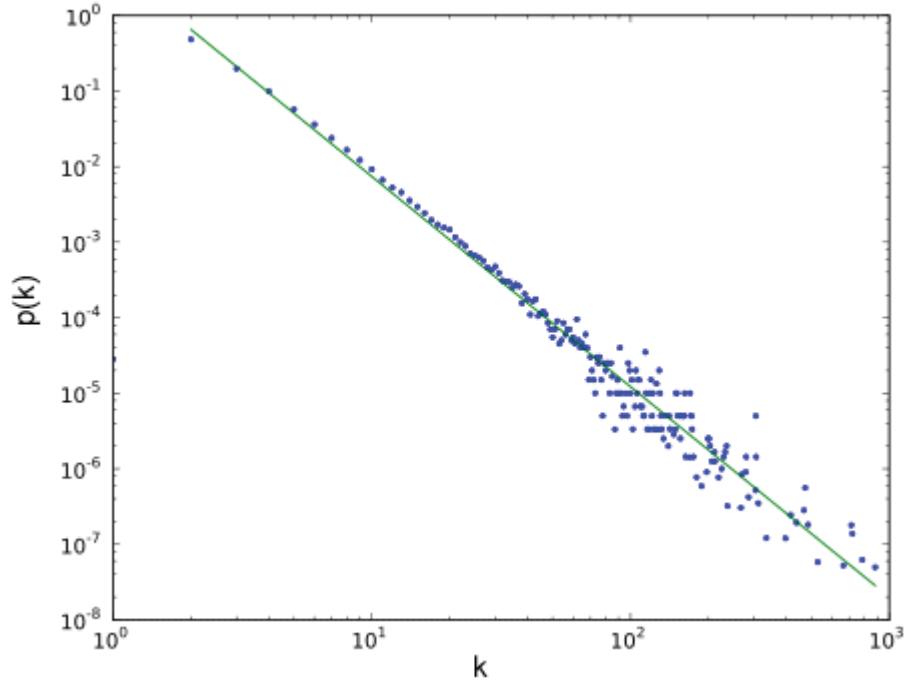
Barabási-Albert modelinin algoritması şu şekildedir:

- i) *Büyüme*: Küçük bir çekirdek düğüm sayısı (m_0) ile başlayarak her adımda ağa yeni eklenen bir düğümün $m < m_0$ adet bağlantı ile önceki düğümlere bağlanmasını sağla.

- ii) *Seçimli bağlanma*: Yeni düğümlerin bağlanacağı eski düğümleri belirlerken, yeni düğümün i . düğüme bağlanma olasılığı $\frac{k_i}{\sum_j k_j}$ değerinin, i . düğümün derecesi olan k_i 'ye aşağıdaki eşitlik ile bağlı olmasını sağla.

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (3.4)$$

Bu algoritma t adım ilerletildiğinde, $N = t + m_0$ düğüm ve mt linkten oluşan bir ağ ortaya çıkacaktır. Nümerik simülasyonlar, bir örneği Şekil 3.15'te de görülen bu ağın Şekil 3.16'daki gibi scale free derece dağılımına sahip bir topolojiye sahip olduğunu, aynı zamanda bu power-law uyumlu eğrinin γ katsayısının, modelin tek parametresi olan m değerinden bağımsız olduğunu göstermektedir [1].



Şekil 3.16. Barabási-Albert modeli ile oluşturulmuş bir ağa ait derece dağılımı grafiği [8].

Barabási-Albert modeli, power-law derece dağılımına sahip bir ağı modelleyebilen, gerçel ağ oluşum ve gelişim mekanizmasını içeren bir minimal modeldir. Gerçel ağlara kıyasla bazı kısıtlamaları da mevcuttur. Örneğin bu modelle tasarlanan bir ağın derece dağılımı, sabit katsayılı bir power-law kuyruğa sahiptir. Oysa bu katsayı

gerçel ağlar için 1-3 aralığında değişim göstermektedir. Bununla birlikte gerçel ağların derece dağılımları, grafiğin sağına doğru exponansiyel cut-off, ya da soluna doğru (örn. $k=1,2$ için) saturasyon bölgeleri içermektedir. Bu kısıtlamaların aşılabilmesi için yapılan çalışmalar ise “Evrilen Ağlar Teorisi”nin ortaya çıkmasına öncülük etmiştir.

3.5. EVRİLEN AĞLAR TEORİSİ

Gerçel ağlar ile Barabási-Albert modelinin farklılıkları, bilim dünyasını farklı arayışlara itmiştir. Öncelikle cevap bekleyen sorular, power-law katsayısının nasıl değiştirilebileceği, gerçel ağlarda olduğu bilinen çeşitli mikroskobik proseslerin ağ topolojisine etkilerinin neler olduğu, derece dağılımı haricinde ağ karakteristiklerini yansıtan başka dağılımların olup olmadığı gibi sorulardır. Bu soruların cevaplanması, gerçel ağların bilinmezliklerini de içeren bir evrilen ağ modelinin tanımlanabilmesi açısından kritik öneme sahiptir.

Bu soruların cevaplanmasına odaklanan bilim adamları, bazı önemli sonuçlara ulaşmışlardır. Bu sonuçlar, ağ gelişimi ve topolojisine eşsiz bakış açıları getiren “Evrilen Ağlar Teorisi”nin ayrı bir alan olarak doğmasını sağlamıştır.

3.5.1. Seçimli Bağlanma

Scale-free ağların oluşturulmasında kullanılan tüm ağ modellerinin temel içeriklerinin en başında seçimli bağlanma olgusu gelmektedir. Seçimli bağlanma, yeni düğümlerle link oluşturma eğiliminin, bir düğümün derecesi ile doğru orantılı olmasıdır. Barabási-Albert modeli, bir düğümün i . düğümlerle bağlantı oluşturma eğiliminin, i . düğümün derecesi olan k ile orantılı olduğunu kabul eder (bkz. eşitlik 3.4). Bu kabul iki sonuç doğurmaktadır: öncelikle rassal ağların link oluşturma olasılığını ifade eden $\Pi(k_i) = p$ eşitliğinin aksine $\Pi(k_i)$ 'nin k 'ya bağımlı olması, bununla beraber $\Pi(k_i)$ 'nin fonksiyonel formunun k 'ya göre lineer olmasıdır. Önceki çalışmaların, derece dağılımının önemli oranda $\Pi(k_i)$ 'ya bağımlı olduğunu ortaya koyması ile birlikte $\Pi(k_i)$ için hassas bir fonksiyonel form tanımlanması, önemli bir akademik gündem oluşturmuştur.

$\Pi(k_i)$ 'nin fonksiyonel formunun tanımlanabilmesi, ancak hangi düğümün hangi zamanda ağa bağlandığının bilindiği ağlarda mümkündür [27]. Bu veriler ise sadece birkaç ağ türü için ulaşılabilir durumdadır: bilimsel işbirliği ağları, yayın atıf ağları, aktör birlikteliği ağları ve domain seviyesindeki internet ağı.

Bu tür ağlarda $\Pi(k_i)$ 'nin hesaplanması için şu adımlar izlenir:

- i) Ağ gelişiminin herhangi bir anında “eski” düğümlerin sayısını ve bu düğümlerin derecelerini kaydet.
- ii) Ağın yaşına göre çok düşük bir T aralığı için “eski” düğümlerin derecelerindeki artışı ölçüp kaydet.
- iii) Her düğümün önceki derecesi k_i 'ye bağımlı olarak derecesindeki göreceli artış $\Delta k_i/\Delta k$, o düğüm için $\Pi(k_i)$ fonksiyonunu verecektir. Burada Δk , ΔT zaman aralığında ağa katılan yeni düğümlerin sayısıdır.

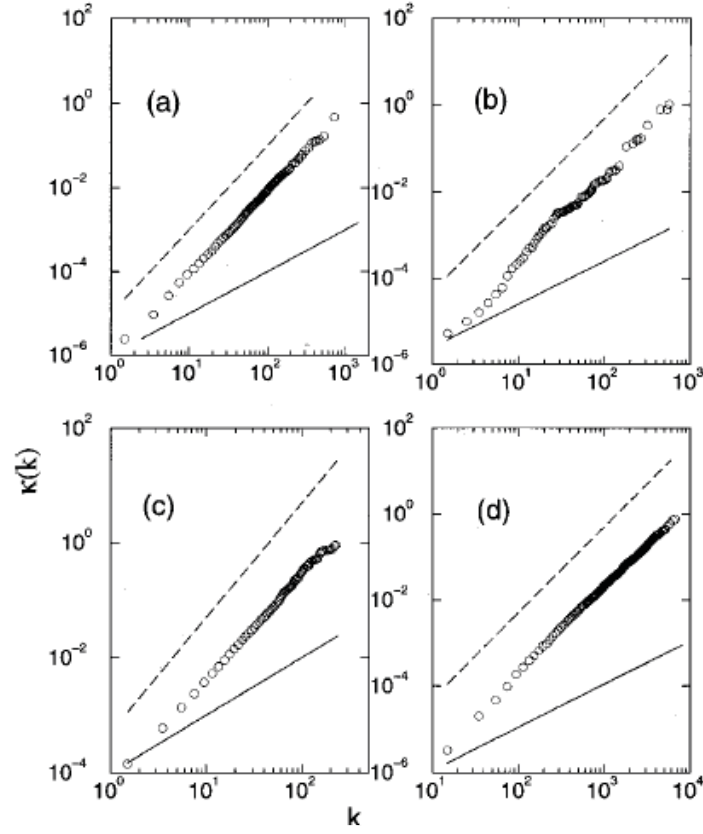
Bu yol ile elde edilen verideki dalgalanmalar, aşağıdaki gibi birikimli dağılımın elde edilmesi ile azaltılabilir:

$$\kappa(k) = \sum_{k_i=0}^k \Pi(k_i) \quad (3.5)$$

Barabási ve arkadaşlarının 4 farklı ağ için yukarıda açıklanan şekliyle elde etmiş oldukları $\Pi(k)$ grafikleri (Şekil 3.17), gerçek ağlarda seçimli bağlanmanın varlığını göstermektedir [7,1]. Bunun da ötesinde, Şekil 3.17'deki grafiklerde $\Pi(k)$ 'nin eşitlik 3.6'daki gibi power-law karakteristiğine bağlı olduğu görülmektedir.

$$\Pi(k) \sim k^{-\alpha} \quad (3.6)$$

İnternet, yayın atıf ağı, Medline ve Los Alamos arşivi gibi bazı ağlar için $\alpha \approx 1$ değerleri elde edilmiş, yani $\Pi(k)$, Barabási-Albert modelindeki gibi k 'ya lineer bağımlılık göstermiştir [13,27,38]. Sinirbilim alanındaki bilimsel işbirliği ağı ve aktör ağı gibi diğer ağlar için ise bu bağımlılık altdoğrusal niteliktedir ($\alpha=0.8 \pm 0.1$) [38].



Şekil 3.17. Seçimli bağlanma grafikleri: a) bilimsel atıf ağı; b) internet; c) sinirbilim alanındaki bilimsel işbirliği ağı; d) aktör ağı. Tüm grafiklerde kesikli çizgiler lineer seçimli bağlanmayı, düz çizgiler ise seçimli bağlanmanın olmadığı durumu ifade etmektedir [7].

3.5.2. Büyüme

Barabási-Albert modelinde her zaman adımında ağı bağlanan yeni düğüm sayısı sabittir. Buna bağlı olarak ağın büyüklüğü lineer artış göstermekte, ortalama derece değeri ise sabit kalmaktadır. Buna karşılık gerçel evrilen ağlarda değişik büyüme örüntüleri izlenmektedir [1]. Örneğin Faloutsos ve arkadaşları, internet ağı üzerinde yaptıkları araştırmada 1997 yılı kasım ayı için buldukları ortalama derece değeri olan 3.42'nin 1998 yılı aralık ayında 3.96'ya çıktığını göstermişlerdir [25]. Benzer biçimde Broder ve arkadaşları, World Wide Web'in ortalama derece değerinin 5 ay içerisinde 7.22'den 7.86'ya çıktığını gözlemlemişlerdir [21]. Diğer gerçel ağ çalışmalarının sonuçları da bu durumu destekler niteliktedir [1]. Tüm bu sonuçlar, 2007 yılında Smith ve arkadaşlarının yapmış oldukları çalışmada “accelerating

networks” (hızlanan ağlar) tanımının ortaya çıkması ile daha çok anlam kazanmıştır [39].

Barabási ve arkadaşlarının önerdiği büyüme modelinde, ağa her zaman adımında b adet yeni düğüm eklenmekte, bu b adet düğüm ağda mevcut olan b adet eski düğüme aşağıdaki eşitliğe bağlı olarak seçimli bağlanmaktadır [7].

$$P_i = b \frac{k_i}{\sum_j k_j} \quad (3.7)$$

Buna ek olarak her zaman adımında, lineer olarak artan sayıda düğüm arasında (a : ağdaki toplam düğüm sayısının bir sabit oranı olacak şekilde), aşağıdaki eşitliğe bağlı olarak yeni bağlantılar oluşmaktadır.

$$P_{ij} = \frac{k_i k_j}{\sum_{s,l} k_s k_l} N(t) a \quad (3.8)$$

Burada $N(t)$ ağdaki düğüm sayısını ifade etmektedir. Gerek ağa yeni eklenen düğümler, gerekse mevcut düğümler arasında tanımlanan yeni bağlantılar sayesinde ağın ortalama derece değeri gerçel ağlarda olduğu gibi sürekli olarak artmaktadır.

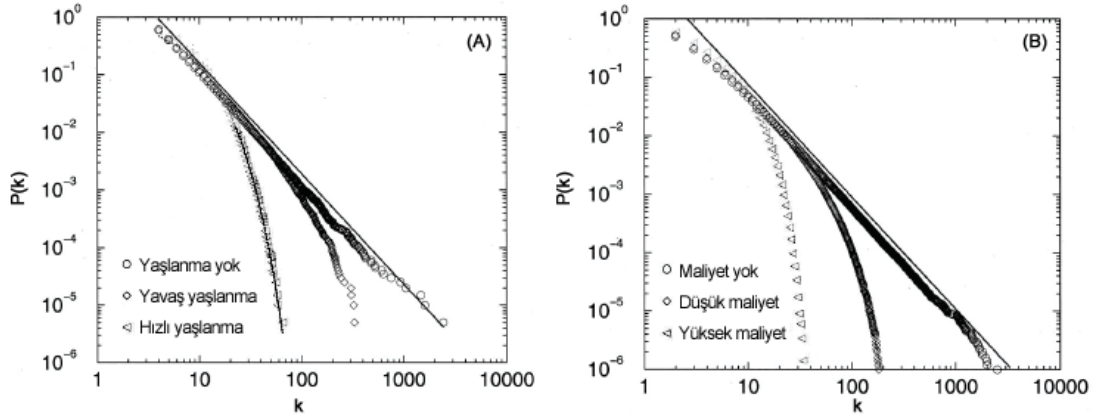
3.5.3. Diğer Etkenler

Evrilen ağ teorisinin diğer önemli bileşenleri arasında yerel olaylar, büyüme kısıtları, rekabet etkisi gibi maddeler sayılabilir. Bu bileşenler, ağın topolojik değişimini içten içe etkileyen önemli dinamiklerdir.

Barabási-Albert modeli, ağın büyümesini tek bir mekanizma ile sağlamaktadır: ağa yeni katılan düğümlerin mevcut düğümlere bağlanması. Ancak gerçel ağların gelişimi, ağ içinde yeni bağlantıların oluşması, mevcut düğüm ya da bağlantıların silinmesi gibi bir dizi mikroskopik olaydan da etkilenmektedir. Yakın geçmişte, ağın scale-free karakteristiğine etki yapan bu aksiyonları da içeren yeni modeller üzerinde bazı çalışmalar yapılmıştır [1,7,11]. Ağ topolojisi üzerindeki bir yerel değişim 4 ana

unsurun kombinasyonu ile gerçekleşmektedir: düğüm ekleme ve çıkartma, bağlantı ekleme ve çıkartma. Gerçekte ise bu eylemlerden bir kaçı kombine olarak gelişebilir. Örneğin bir düğümün “yeniden bağlanma” prosedürü, düğüm çıkartma ve ekleme eylemlerinin birleşimidir. Gerçekçi bir ağ modelinin scale-free karakteristiğine katkıda bulunan tüm bu prosesleri içermesi gerekmektedir.

Ağ topolojisini etkileyen diğer bir etken de büyüme kısıtlarıdır. Büyüme kısıtları, ağın düğüm ve bağlantı bazında büyümesini kısıtlayan düğüm ömrü (örneğin sosyal ağlar), maliyet etkisi ve sonlu bağlantı kapasitesidir (örneğin bir internet router'ının sınırlı bağlantı kapasitesi). Bu kısıtların derece dağılımına etkileri üzerine bazı çalışmalar yapılmıştır.



Şekil 3.18. Derece dağılımının power-law'dan sapma eğilimleri: (a) yaşlanma etkisi (b) maliyet etkisi [40].

Amaral ve arkadaşları, Şekil 3.18'de görüldüğü gibi yaşlanma ve maliyet etkilerinin, ağın derece dağılımında power-law karakteristiğinden sapmalara yol açtığını göstermişlerdir [40]. Bu çalışma, gerçel ağların derece dağılımlarının sağ tarafındaki eksponansiyel cutoff bölgesini açıklar niteliktedir.

Barabási-Albert modelinin bir diğer eksik yanı ise gerçel ağlardaki rekabet şartlarını içermiyor olmasıdır. Barabási-Albert modeline göre her bir düğümün derecesi, sabit katsayıya sahip ($\beta = 1/2$) power-law bağımlı bir artış göstermektedir [1]. Bunun sonucu olarak derecesi en yüksek olanlar, ağdaki en eski düğümlerdir. Halbuki gerçel ağlarda düğümlerin dereceleri sadece yaş faktörüne bağlı değildir. Örneğin

WWW’de bazı sayfalar, iyi bir içerik ve reklam kombinasyonu ile kısa zamanda çok fazla “gelen link” oluşturabilmektedir [22]. Ya da bazı bilimsel yayınlar, eş zamanlı oldukları diğer yayınlardan çok daha fazla atıf alabilmektedir.

Bianconi ve Barabási gerçel ağların rekabetçi bir yanının olduğunu, her düğümün bağlantı oluşturmak için diğer düğümlerle rekabet içerisinde olduğunu öngörmüşlerdir [41]. Bu öngörü neticesinde, ağ akatılan her düğüm için katılma anında atanan ve değeri sonradan değişmeyecek olan bir kondisyon (fitness) parametresi (η_i) içeren büyüme modeli tanımlamışlardır. Bu modele göre her zaman adımında ağa η_j kondisyonuna sahip bir j düğümü eklenirken, bu η_j parametresi bir $\rho(\eta)$ dağılımına göre hesaplanacaktır. Her yeni düğüm, addaki m düğüme bağlanacak, i . düğüme bağlanma olasılığı ise o düğümün derecesine ve kondisyonuna aşağıdaki eşitlik ile bağlı olacaktır:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \quad (3.9)$$

Bu genelleştirilmiş seçimli bağlanma prensibi, gerçel ağlarda olduğu gibi ağa yakın zamanda bağlanmış bir düğümün de kısa zamanda yüksek derecelere sahip olabildiğini sağlamaktadır.

Gerçel evrilen ağların derece dağılımı karakteristiklerini etkileyen etkenler yukarıda anlatılanlarla sınırlı değildir. Günümüzün ağ modelleri, kademeli yaşlanma, bağlantıların kalıtımı, ağda yürüme gibi birçok etkiyi de bünyesinde barındırarak gerçel ağlara en yakın topolojik yapıyı sağlamayı amaçlamaktadırlar [11,34,42].

Bu tez çalışmasının önemli çıktılarında birisi de evrilen ağ modellerine referans oluşturacak geniş bir gerçel ağ veri setinin sunulması olacaktır.

BÖLÜM 4

GERÇEKLEŞTİRİLMİŞ OLAN ÇALIŞMALAR

Tez kapsamında gerçekleştirilmiş olan çalışmalar, veri setinin elde edilip yapısal veritabanına aktarılması ve bu verilerin işlenerek nicel büyüklüklerin elde edilmesi olarak iki aşamaya ayrılmaktadır.

4.1. VERİ SETİNİN ELDE EDİLMESİ

Çalışmamızda öncelikle Web of Science veritabanındaki Türkiye adresli yayın verileri elde edilmiş, bu veriler yapısal veritabanı biçimine (SQL) dönüştürülmüştür. Ardından veri setimiz üzerinde bir ağın dinamiklerini ve gelişimini tanımlayan çeşitli parametrelerin yıllara göre değişimi incelenmiştir. Bu parametreler sırasıyla: yıllara göre birikimli yayın sayıları ve yazar sayıları, yazar başına düşen yayın sayılarının oluşma frekansları, yayın başına düşen yazar sayılarının oluşma frekansları, yazarların birbirine ortalama uzaklığının yıllara göre değişimi, kümelenme katsayısının yıllara göre değişimi, ortalama derecenin yıllara göre değişimi, derece dağılımı, yazar başına düşen ortalama yayın ve yayın başına düşen ortalama yazar sayıları, ağın genişleme hızının yıllara göre değişimi ve bunlarla iniltili büyüklüklerdir.

Veri setinin oluşturulması işleminin aşamaları aşağıdaki gibidir.

4.1.1. Web of Science Veri Tabanından Verilerin Elde Edilmesi

Institute for Scientific Information (ISI), merkezi Amerika Birleşik Devletleri (ABD)'de bulunan ve Science Citation Index (SCI), Science Citation Index-expanded, Social Sciences Citation Index (SSCI) ve Arts and Humanities Citation Index (AHCI) gibi indeksleri kullanan, düzenli olarak binlerce bilimsel dergiyi

tarayarak içeriğini okuyucularına duyuran bir sistemdir. İnternete geçişle beraber ISI'nin önemi daha da artmıştır. ISI bu hizmeti çok çeşitli şekillerde okuyucularına sunmaktadır. İnternet ortamında Web of Science (WoS) ve Web of Knowledge (WoK) isimleri ile, kendisine üye kurum ya da kişilere açık olan arama hizmetini sunmaktadır. WoS, yapılan arama sonuçları üzerinde çalışma yapmak isteyen kişilere, bu arama sonuçlarını her seferinde en fazla 500 kayıt içerecek biçimde izin vermektedir. Arama sonuçları 1980 yılı ve sonrasında yapılan yayınları içermektedir. Kullanıcılar BibTeX, HTML, Plain Text, Tab delimited seçeneklerinden birini seçerek kayıt formatını belirleyebilmektedir (Şekil 4.1).

Türkiye'nin bilimsel işbirliği ağı üzerinde yapmayı düşündüğümüz çalışma için WoS'u internet ortamına sunmakta olan Thomson Reuters firmasının Ankara temsilciliği ile fikir alışverişi yapıldı. Veri sağlama konusunda yapılan öneriler doğrultusunda WoS arama ara yüzündeki yıl aralığını 1980-2010 olarak belirleyip "address" alanına "Turkey or Turkiye or Türkiye" yazarak ihtiyaç duyulan veri setine ulaşıldı. Bunun dışında bir kısıtlama yapmadan gerçekleştirilen aramada yaklaşık 240 bin bilimsel yayın elde edildi. Arama sonuçlarını kaydetmek için sistemin bir seferde izin verdiği en yüksek değer olan 500 değeri kullanıldığında tüm kayıtların kaydedilmesi prosedürünün yaklaşık 476 defada gerçekleşeceği görüldü (Bu çalışmanın başladığı 2011 yılından 2 yıl sonra, 2011 ve 2012 yıllarına ait verilerin de eklenmesiyle güncellediğimiz veritabanımız, 292,895 yayın ve 185,886 yazarı kapsayacak biçimde genişlemiş oldu). Bu kayıt işlemi yaparken, aramamızın listelendiği sayfanın (Şekil 4.1a) alt kısmındaki kaydetme aralığının (Şekil 4.1b) her seferinde 500 artırılarak ilerlenmesi gerekiyordu. Bununla birlikte her seferinde, tercih ettiğimiz kayıt formatı olan "HTML"nin de seçili hale getirilmesi, ardından "Save" butonuna tıkladıktan sonra tarayıcıda açılan diyalog penceresinde kaydedilecek dosya için farklı bir isim verilmesi gerekiyordu.

Web of Science Additional Resources

Search Author Finder Cited Reference Search Advanced Search Search History

Web of ScienceSM

Results Address=(turkey or turkiye or türkiye)
Timespan=1900-2010. Databases=SCI-EXPANDED, SSCI, CPCL-S, CPCL-SSH.
Lemmatization=On

Results: 237.838 Page 1 of 10.000 Go

Sort by: Publication Date

Refine Results

Search within results for

Web of Science Categories Refine

SURGERY (19,009)

PEDIATRICS (10,033)

ENGINEERING ELECTRICAL ELECTRONIC (9,250)

CLINICAL NEUROLOGY (8,696)

PHARMACOLOGY PHARMACY (7,889)

more options / values...

Document Types Refine

ARTICLE (189,430)

MEETING ABSTRACT (19,250)

PROCEEDINGS PAPER (18,986)

1. Title: NEW AND KNOWN CONSTITUENTS FROM IRIS UNGUICULARIS AND THEIR ANTIOXIDANT ACTIVITY
Author(s): Atta-ur-Rahman; Hareem Sumaira; Choudhary M. Iqbal; et al.
Source: HETEROCYCLES Volume: 82 Issue: 1 Pages: 813-824 DOI: 10.3987/COM-10-S(E)6 Published: DEC 31 2010
Times Cited: 0 (from Web of Science)
[Full Text] [View abstract]

2. Title: 1,3-Propanediol production potential of Clostridium saccharobutylicum NRRL B-643
Author(s): Gungormusler Mine; Genen Cagdas; Ozdemir Guven; et al.
Source: NEW BIOTECHNOLOGY Volume: 27 Issue: 6 Pages: 782-788 DOI: 10.1016/j.nbt.2010.07.010 Published: DEC 31 2010
Times Cited: 1 (from Web of Science)
[View abstract]

3. Title: Morphometric effect of nandrolone decanoate used as doping in sport on femur of rats in puberty period
Author(s): Lok Sefa; Yalcin Hakan
Source: ARCHIVES OF BUDO Volume: 6 Issue: 4 Pages: 217-220 Published: DEC 31 2010
Times Cited: 0 (from Web of Science)
[View abstract]

a) Web of Science arama sonuçları arayüzü

Results: 237.837 Show 10 per page Page 1 of 10.000 Go

Output Records

Step 1:

Selected Records on page

All records on page

Records to

Step 2:

Authors, Title, Source

plus Abstract

Full Record

plus Cited References

Step 3: [How do I export to bibliographic management software?]

Save to: EndNote Web EndNote ResearcherID

Save to other Reference Software

Save to BibTeX

Save to HTML

Save to Plain Text

Save to Tab-delimited (Win)

Save to Tab-delimited (Mac)

237.837 records matched your query of the 39.448.558 in the data limits you selected.

View in: 简体中文 English 日本語

b) Arama sonuçlarının kaydedilmesi

Şekil 4.1. Web of Science arama prosedürü.

Web sayfası, bu işlemleri URL değişkenleri ile değil, form değişkenleri ile gerçekleştirdiği için aramaya ve kaydetmeye ait parametreler URL içerisinde verilememektedir, yani her seferinde kaydetme formu üzerindeki veriler güncellenip bu şekilde ilerlemek gerekmektedir.

Tüm bu kayıt prosedürünü hızlandırmak adına “iMacros for Firefox” isimli tarayıcı eklentisi kullanıldı. Bu eklentinin kullanımına kısaca değinmek gerekirse, “Record” tuşuna bastıktan sonra tarayıcıda yapılan tüm işlemleri “Stop” tuşuna basılana kadar kendi formatında kodlayarak saklamaktadır. Ardından “Save” butonu ile kaydedilebilen bu aksiyonlar dizisi ileride “Play” tuşu ile tekrar çalıştırılabilmektedir. Bu eklenti aktifleştirildikten sonra kayıt prosedürünün bir aşamasını (1-500 aralığındaki sonuçları kaydetme) gerçekleştirip aksiyonlarımızı iMacros üzerinde kaydettik. iMacros’un kodlama dili döngü ya da şart ifadeleri

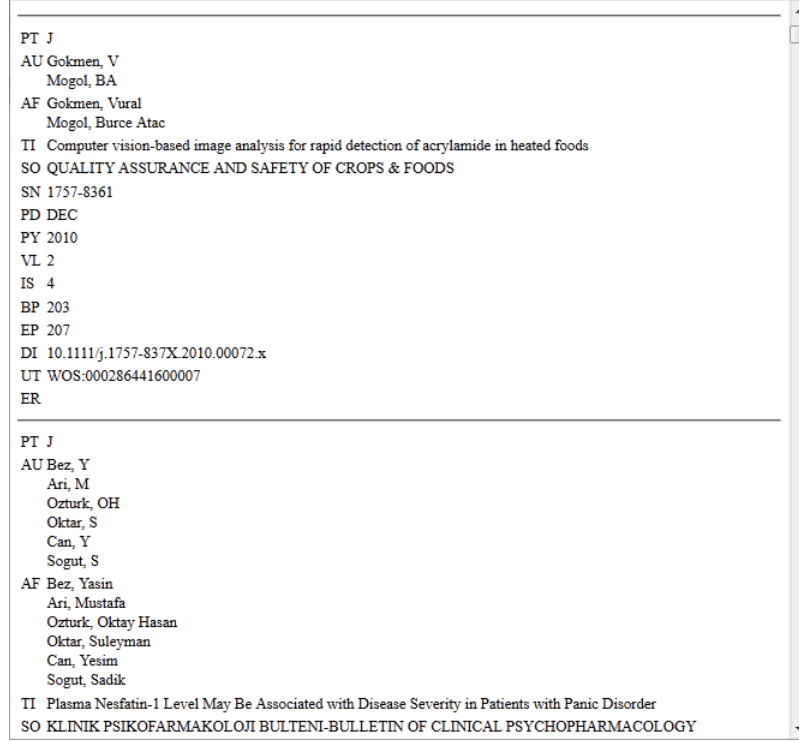
içermemektedir. Bu nedenle kayıt aralıklarını değiştirmek için gerekli kod bloğu C# içerisinde basit bir döngü ile elde edilmiştir. Yani 476 defa çalışacak bir döngü yerine iMacros'un kod penceresine aynı işi yapan ancak kayıt aralığının sürekli 500 artırıldığı 476 kod bloğu yapılandırılmıştır. Kaydedilen dosyaların isimlendirmesinde otomatik olarak kayıt anındaki "time" değeri kullanılmıştır. Bu şekilde yaklaşık 240 bin kaydı içeren 476 HTML sayfası (2011-2012 yıllarını kapsayan genişletme çalışmalarımızda 111 HTML sayfasının daha eklendiği ham veritabanımız toplamda 587 HTML sayfasından oluşmaktadır), 6-7 saat zarfında bir klasör içerisinde kaydedilmiş oldu. Yine C# ortamında yazılan bir dosya isimlendirme programı yardımı ile bu klasördeki tüm dosyalar "wos00x.html" formatında (wos000 – wos475.html) isimlendirildi (Bkz. Ek Açıklamalar B).

4.1.2. Verilerin Yapısal Veri Tabanına Aktarılması

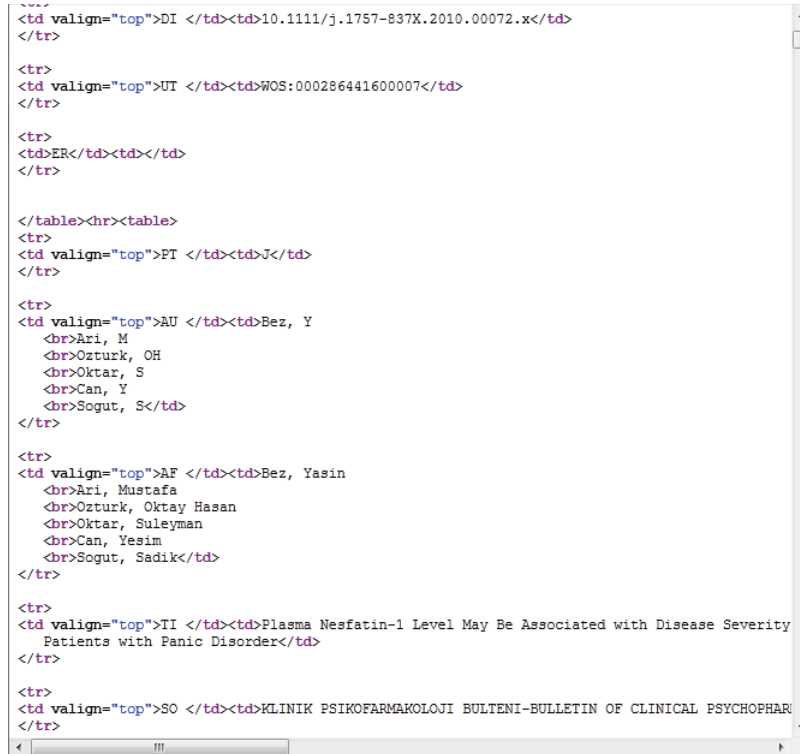
Verilerin HTML dosyaları halinde kaydedilmesinin ardından bu çalışmanın önemli aşamalarından birisi başladı. Bu aşama, verileri bir yapısal veritabanına dönüştürmek için öncelikle uygun bir veritabanı modelinin oluşturulması, ardından da bir program yardımı ile HTML formatındaki verilerin işlenerek bu veritabanına aktarılması adımlarını içermektedir.

Öncelikle WoS'tan elde edilen veri formatı (Şekil 4.2) incelenerek çalışmamızda gerekli olabilecek veri alanları belirlendi. Bilimsel işbirliği ağını oluşturabilmek için bu veri alanlarından sadece yazar isimleri (AU : Yazar soyadı, adının baş harfleri ya da AF : Yazar soyadı ve tam adı) ve yayın yılı (PY) alanları yeterli iken, farklı çalışmalarda da kullanılabilmesi açısından bu alanlarla birlikte yayın türü (PT), yayın başlığı (TI), yayınlanan kaynak (SO) alanları da dikkate alındı.

Verileri işleyip yapısal bir veritabanına aktarma aşamasında verilerin işleneceği programlama dili olarak Microsoft Visual Studio 2010 bünyesindeki C#.Net, veritabanı olarak da Microsoft SQL Server seçildi. Bilimsel işbirliği çalışmasının işleyişine uygun olarak Şekil 4.3'teki veritabanı tasarımı yapılarak SQL Server'da tablo ve ilişkiler tanımlandı (Bkz. Ek Açıklamalar A).

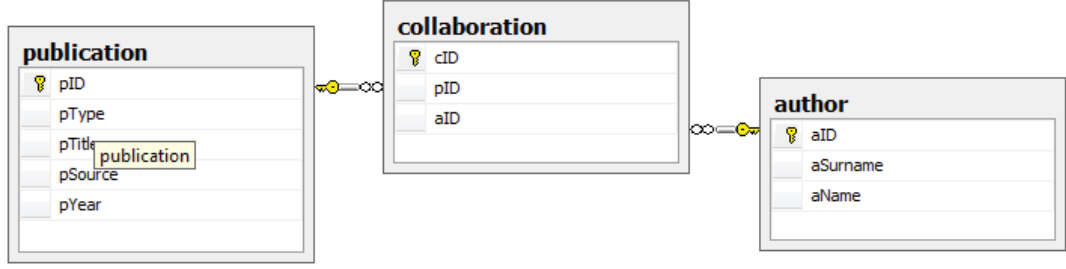


a) HTML formatılda kaydedilen arama sonuçları



b) Sayfanın kaynak kodu

Şekil 4.2. Web of Science arama sonuçları.



Şekil 4.3. Veritabanı tasarım şeması.

Şekil 4.2’deki veri yapısı incelendiğinde yazarların tam isimlerini içeren AF alanının yanı sıra “Soyadı, Adının ilk harfleri” formatındaki AU alanını da görmekteyiz. Ancak bu veri formatı 2008 yılından bugüne kadar hatasız olarak uygulanabilmiş olsa da daha önceki yıllara ait yayınlarda AF alanının da AU alanı gibi sadece ismin baş harflerini içerdiği görülmektedir. Bu da veri setimizi oluşturan 587 dosyadan ilk 320’sinin yazarların tam ismini içermemesi anlamına gelmektedir. Bu nedenle çalışmada veri setinin başından sonuna kadar AU alanının kullanılmasına karar verilmiştir.

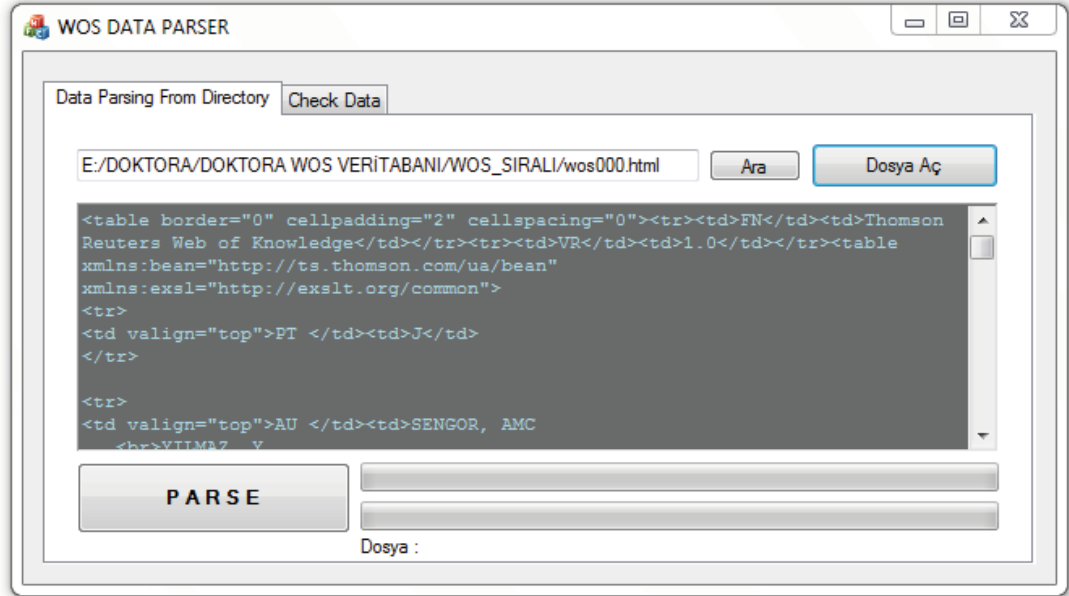
Bu alanda çalışma yapan bütün araştırmacıların ortak sorunu olan yazar ayrıştırma hatalarına Barabási şu şekilde değinmiştir: Öncelikle, veritabanında yazarlar soyadları ile birlikte adlarının baş harfleri ile ifade edildiği için bu durum bazı yazarların ayrıştırılmasında hatalara yol açacaktır. Soyadları ile birlikte adlarının baş harfleri aynı olan iki farklı yazar, veritabanında aynı yazar olarak görüneceklerdir. Bu durum özellikle Çin ve Japon yazarlar için önemli derecede bilgi sapması oluşturacaktır. İkinci olarak, nadir de olsa bazı yazarlar bazı yayınlarında sadece ilk isimlerinin baş harfini kullanırken bazı yayınlarında ise ikinci isimlerinin baş harflerini de kullanmaktadırlar. Bu durumda ise söz konusu yazar iki farklı düğüm olarak veritabanında yer alacaktır. Newman bu problemlerin yol açtığı hatanın düşük bir yüzde değerini geçmeyeceğini göstermiştir [12].

HTML dosyalarındaki verilerin gerçek bir veritabanına dönüştürülmesi işlemi kabaca 5 aşamadan oluşmaktadır. Bir sonraki başlıkta ayrıntılı olarak tanıtılacak olan algoritmanın temel yapısı şu aşamalarda oluşmaktadır:

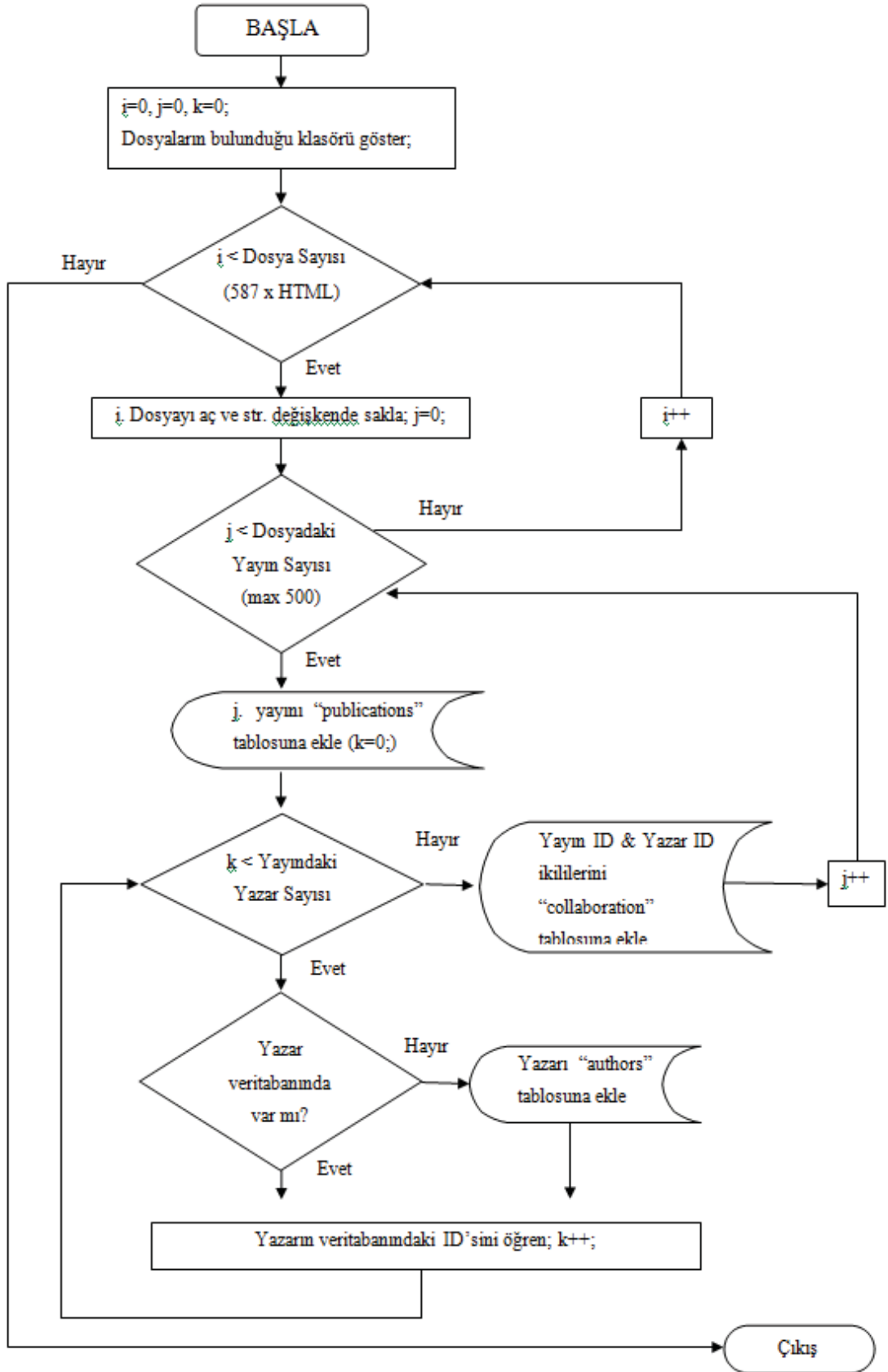
- i) Sırası gelen HTML dosyasını aç
- ii) Dosya içindeki verileri yayınlara ayırıştır (Her dosyada max. 500 adet)
- iii) Yayına özgü verileri “publication” tablosuna ekle.
- iv) Her yayını yazarlarına ayırıştır. Bu yazarlardan veritabanına eklenmemiş (yeni yazar) olanlar varsa “author” tablosuna ekle.
- v) Yazar ID’lerini yayın ID’si ile birlikte ayrı ayrı “collaboration” tablosuna ekle.

4.1.3. Veritabanı Oluşturmada Kullanılan Programın Yapısı

Veri tabanının oluşturulmasında kullanılan program ara yüzü Şekil 4.4’te görüldüğü gibidir. Önceki bölümde ana hatlarıyla sunulan program algoritmasına ait akış diyagramı ise Şekil 4.5’te sunulmuştur.



Şekil 4.4. Veritabanı oluşturmada kullanılan programın ara yüzü.



Şekil 4.5. Veritabanı oluşturmada kullanılan programın akış diyagramı.

Programın algoritmik akışını biraz daha detaylandırmak gerekirse,

- i) HTML dosyaları sıralı olarak isimlendirilmiş olduklarından bu dosyaların açılması bir ana döngü ile gerçekleştirilmiştir.
- ii) Her HTML dosyası 500 (son dosya istisna olarak) yayından oluşmaktadır. Her yayının arasında bulunan <HR> kodu kullanılarak tüm HTML kodunu içeren string, 500 elemanlı başka bir string dizisine ayrıştırılarak her yayına ait HTML kodlarının ayrıca ele alınması sağlanmıştır. Bu kısım, önceki tarafından kapsanan 2. döngüyü oluşturmaktadır.
- iii) Her yayına ait HTML kodu içerisinde yayına özgü alanlar bulunarak veritabanındaki “publication” tablosuna aktarılmıştır. Bu alanlara ait verileri ayrıştırırken yayın türü için "PT <", yayın başlığı için "TI <", yayının kaynağı için "SO <") ve yayın yılı için "PY <" karakterlerinden sonra gelen ve </TD> karakterleri ile sonlanan karakter katarları alınmıştır.
- iv) Her yayına ait HTML kodu incelenerek yazarlar ad ve soyad bilgileri ayrı olmak kaydıyla ayrıştırılmıştır. Eğer bir yayın birkaç yazardan oluşmakta ise, yazarların arasında “Satır atlama” +
 kodları bulunmaktadır. Son yazardan sonra ise </TD> kodu bulunmaktadır. Yayın tek yazarlı ise sadece </TD> kodu ile yazar listesi bitmektedir. Yazarların soyadları ve adlarının ilk harfleri arasında ise “virgül + boşluk” karakterleri bulunmaktadır.
- v) Ayrıştırılan her yazar için veritabanındaki “author” tablosunda bu yazarın var olup olmadığı sorgulanmıştır. Veritabanında kayıtlı olmayan yazarlar veritabanına eklenmiştir. Listedeki tüm yazarların veritabanında kayıtlı olmaları sağlandıktan sonra ID’leri belirlenerek bir dizide saklanmıştır. Yazarların incelendiği bu prosedür, içteki 3. Döngü bloğunu oluşturmaktadır.
- vi) İncelenen yayının ID’si ile yayında ortaklık kurmuş olan yazarların ID’leri çiftler halinde [pID, aID] “collaboration” tablosuna eklenmiştir.

Böylece “publication” tablosuna yayınlara özgü bilgiler, “author” tablosuna sadece sisteme yeni dahil olan yazarlar, “collaboration” tablosuna ise yayın-yazar ikilileri kaydedilmiş, çalışmamızın ilerleyen aşamalarının işlerliğini akıcı hale getirecek bir veritabanı elde edilmiştir.

Programın çalışma seyri 2 adet *progressbar* ile takip edilebilmektedir. Birincisi açılan dosya sayısının, ikincisi ise dosya içerisindeki incelenmiş yayın sayısının ilerleyişini bildirmekte, bu sayede programın ilerleyişi gözlenebilmektedir. Çalıştırdıktan sonraki ilk dakika ve saatlerde gayet hızlı ilerleyen ayrıştırma işlemleri, veritabanının gittikçe dolması ve dolayısıyla “author” tablosundaki SELECT sorgularının gittikçe daha fazla zaman alması sebebiyle ilerleyen saatlerde yavaşlama göstermiştir. İlk aşamadaki (1980-2010) 476 adet dosyada bulunan toplam 237.409 yayın, bu yayınlara katkı yapmış 151.745 yazar ve bu yayınlardaki tüm birliktelikleri temsil eden 967.086 adet kayıttan oluşan “collab” veritabanımızın oluşması toplam 19 saat gibi bir zaman diliminde gerçekleşmiştir. Ancak tahmin dışı bazı etkenler yüzünden kesintisiz olarak programın çalışması birkaç denemeden sonra mümkün olmuştur. Program kodlarını da değiştirmemize neden olan bu nedenlerden bazıları:

- i) Yazar isminin boş olması,
- ii) Yayın isimlerinde kullanılmış olan tek tırnakların SQL sorgularını bölmesi,
- iii) Yayın başı yazar sayısı için düşünülen max. 350 değerinin, ardından 500 değerinin bazı yayınlarda aşılması sonucu dizi değişkenlerin yetersiz kalması,
- iv) Yazar soyadlarında ve bazılarının adlarının baş harfleri arasında tek tırnak karakteri olması ve SQL sorgularının bölünmesi.

Bu beklenmedik durumlar, dizi değişkenlerin sınırlarının artırılması ve SQL sorgularına giren tüm parametrelerin içinde geçen sakıncalı karakterlerin başka karakterler ile değiştirilmesi yoluyla aşılmıştır. Ancak gerek birkaç defa program çalışmasının yarıda kalması, gerekse kodlardaki string kontrolleri ve replace fonksiyonlarının artması ya da dizi değişkenlerin fazlaca geniş tutulması nedeniyle zaman kayıpları meydana gelmiştir.

Veritabanı oluşturulduktan sonra verilerin doğruluğunun kontrolünü sağlamak adına HTML dosyaları ile karşılaştırmalar yapılmış, veritabanının ham verilerle uyum içerisinde olduğu doğrulanmıştır. Sadece yapılan kontrol sorguları neticesinde 417 kayıt içerisinde aynı yazara ait [soyadı, adının ilk harfleri] verilerinin tekrarlandığı

gözlenmiştir. Bu tekrarları irdelemek için HTML dosyalarında bu yayınlar bulunmuş, birkaçında görülen evli çift ya da kardeş durumu dışında ana yazar adının diğer yazarlar arasında tekrarlandığı gözlenmiştir. WoS veritabanından kaynaklanan bu hatalı kayıtlar “collaboration” tablosundan silinmiştir.

4.2. VERİLERİN İŞLENMESİ VE ELDE EDİLEN SONUÇLAR

4.2.1. Yazar Sayılarının Yıllara Göre Değişimi

Tıpkı karmaşık topolojili diğer ağlar (özellikle de WWW) gibi bilimsel işbirliği ağı da sürekli ağa yeni düğümlerin katılması ile genişlemektedir. Bilimsel işbirliği ağlarının öne çıkan en büyük avantajı, her yeni düğümün ağa eklenme tarihinin de geçmişe doğru bilinmekte olmasıdır. Üzerinde kompleks ağ çalışması yapılan ağ çeşitlerinin sadece bir kaçında (bilimsel işbirliği ağları, bilimsel atıf ağları, telefon ağları vs.) ulaşılabilen bu veriler, ağın genişleme dinamiklerinin izlenebilmesi için çok önemlidir.

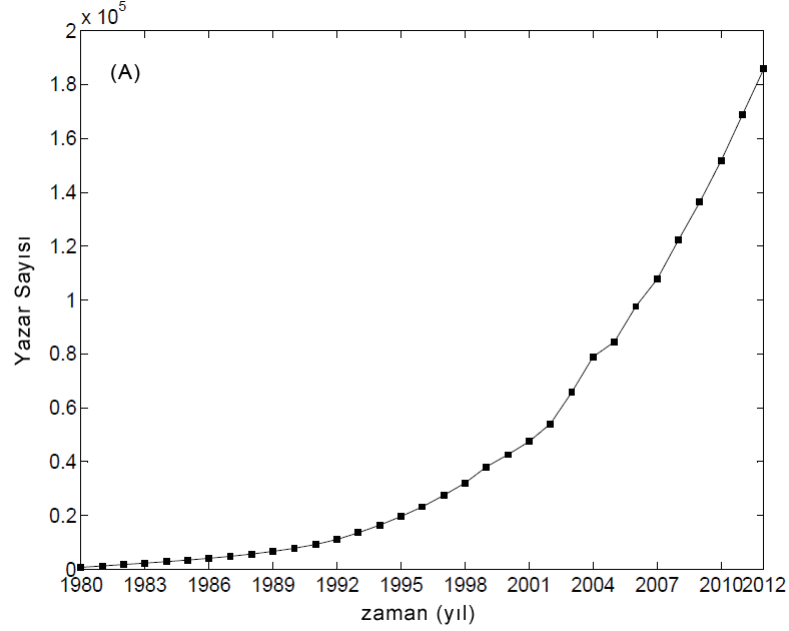
Türkiye adresli yayınlardan oluşan veritabanımız üzerinde yaptığımız analizde ağın düğüm (yazar) bazında nicel büyüklüğünün gelişimini Şekil 4.6a'daki gibi elde ettik. Şekil 4.6b'de ise her yıl ağa katılan yeni düğüm sayılarının yıllara göre değişimi görülmektedir.

Ağın genişleme hızının (*network velocity*) incelendiği ilerleyen bölümlerimizde bu grafiklerin anlamları da irdelenecektir.

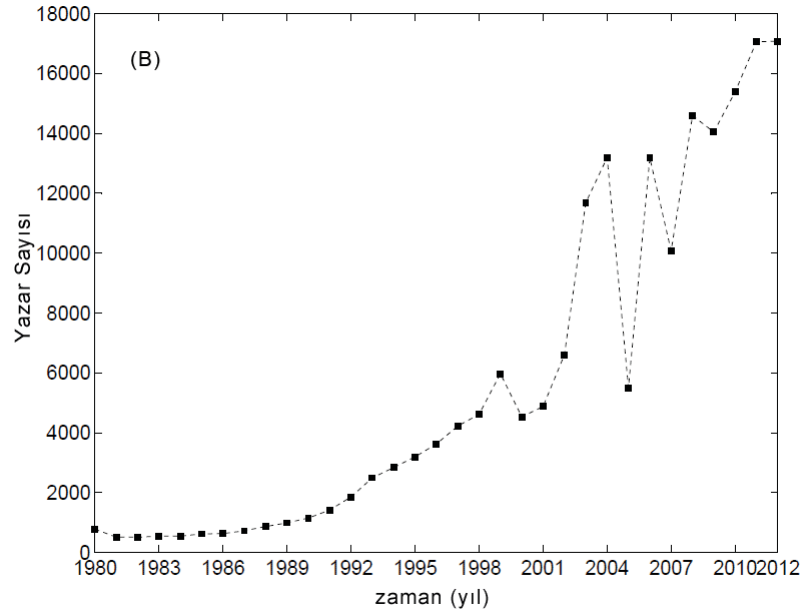
4.2.2. Yayın Sayılarının Yıllara Göre Değişimi

Karmaşık ağlarda gerek ağa yeni düğümlerin eklenmesi, gerekse mevcut düğümler arasında yeni bağlar kurulması ile ağı yapısı sürekli değişim göstermektedir. Bilimsel işbirliği ağında bu değişimin birimi bir bilimsel yayındır. Her bilimsel yayın, ağıma yeni düğüm ekleme ya da eski/yeni düğümler arasında bağlantılar tanımlama fonksiyonlarını yerine getirmektedir. Bilimsel yayın sayılarının nicel olarak

izlenmesi de ülkemiz bilim adamlarının üretkenlikleri ve bu üretkenliğin yıllar bazında gelişimini izlememize imkan tanımaktadır.

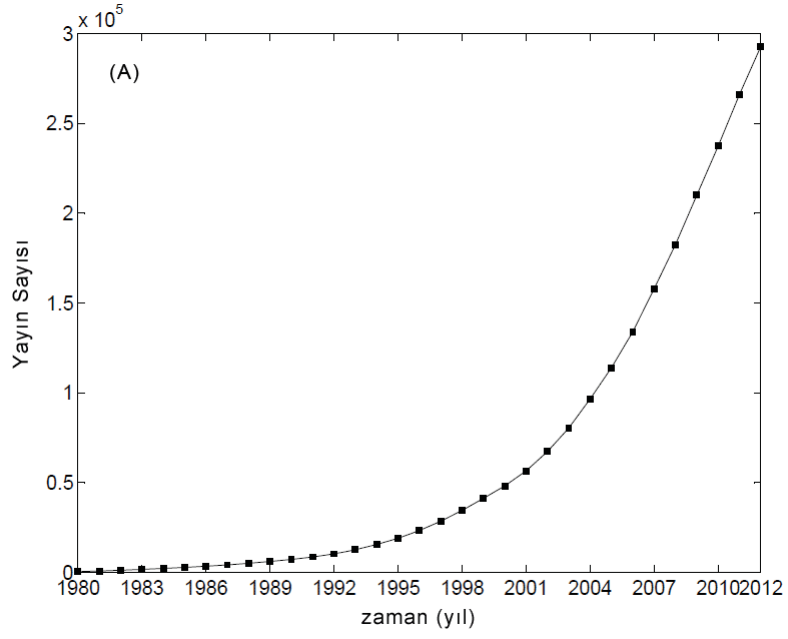


a) Ağda tanımlı yazar (düğüm) sayısının yıllara göre değişimi.

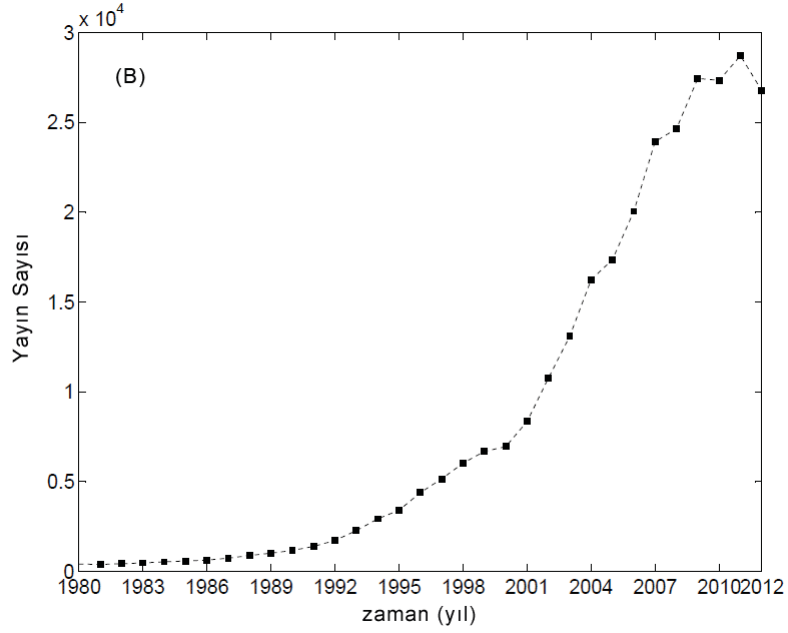


b) Ağda her yıl yeni katılan yazar sayısının yıllara göre değişimi.

Şekil 4.6. Ağda tanımlı yazar (düğüm) sayısının yıllara göre değişimi.



a) Yayın sayısının yıllara göre değişimi (birikimli).



b) Her yıl yapılan yeni yayın sayısının yıllara göre değişimi.

Şekil 4.7. Yayın sayısının yıllara göre değişimi.

Çalışmamızda bilimsel yayınların yıllara göre değişimi Şekil 4.7a'daki gibi elde edilmiştir. Şekil 4.7b'de ise her yıl yapılan yeni yayın sayılarının yıllara göre değişimi görülmektedir.

4.2.3. Yayın Başına Düşen Yazar Sayılarının Oluşma Frekansları

Çalışmamızın bu aşamasında veri tabanımızdaki yayın başına düşen yazar sayıları dikkate alınarak bu yazar sayısı değerlerinin tüm ağdaki oluşma frekansları incelenmiş, Şekil 4.8a'daki grafik elde edilmiştir.

4.2.4. Yazar Başına Düşen Yayın Sayılarının Oluşma Frekansları

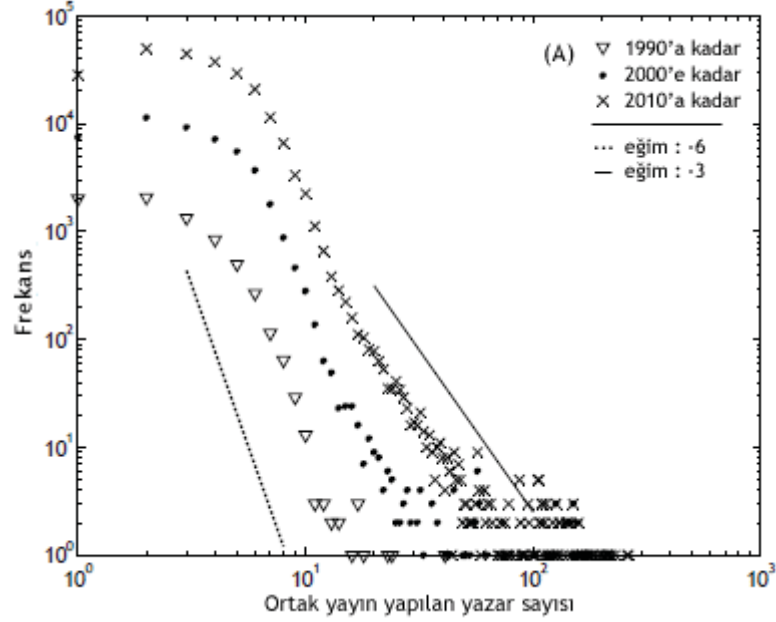
Veri tabanımızdaki yazar başına düşen yayın sayısı değerlerinin tüm ağdaki oluşma frekansları incelenmiş, Şekil 4.8b'deki grafik elde edilmiştir.

Elde etmiş olduğumuz yayın başına düşen yazar sayılarının ve yazar başına düşen yayın sayılarının oluşma frekansları grafikleri log-binned ve normalize hale getirilerek power-law fit uygulanmış, Şekil 4.9a ve 4.9b'deki dağılım grafikleri elde edilmiştir.

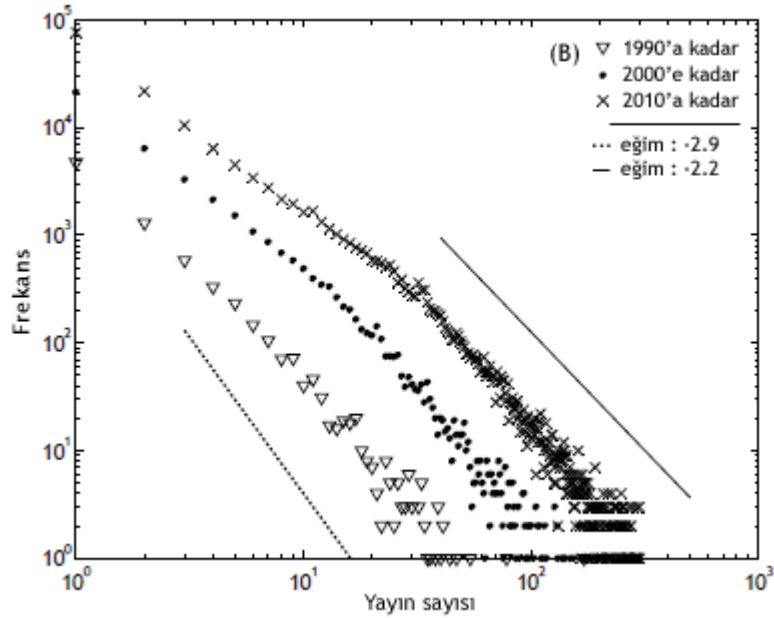
Grafikte dikkati çeken durum, 1990 yılına kadarki dönemde 1 ve 2 yazarlı yayınların diğerlerinden çok daha fazla sayıda oluşu, ancak 2010'a doğru gelindiğinde 2, 3, 4 ve 5 yazarlı yayın sayılarının 1 yazarlı yayın sayılarının sayıca üzerine çıkmış olmasıdır. Bu da aslında ağdaki düğümlerin birbirine uzaklığını azaltan, iç bağlantıları artıran bir etkiye sahiptir. Nitekim ilerleyen bölümlerimizdeki Şekil 4.12'de görüldüğü gibi ortalama derecenin yıllara göre artış trendinde olması, aynı zamanda Şekil 4.16b'de yayın başına düşen ortalama yazar sayısının da 2012 yılına doğru artış göstermesi bu durumu destekler niteliktedir. Tüm bu grafiklerin ortak çıktısı, bilim adamlarının git gide daha çok işbirliği eğiliminde olması, her geçen gün daha çok kişinin katkı sağladığı yayınlar üretiyor olmalıdır.

Şekil 4.8a'da aynı zamanda verileri sınırlayan (katsayıları 3 ve 6 olan) power-law eğrileri de sunulmuştur. Ancak power-law karakteristiği hakkında daha sağlıklı analiz yapabilmek için bu verilere ait dağılım grafiğinin elde edilmesi gerekmektedir. Bu nedenle Şekil 4.8a'daki verilere log-binning uyguladıktan sonra normalize ederek 4.9a'daki dağılım grafiğini elde ettik. Bu grafiğe uyguladığımız $\gamma=4.3$ katsayısına sahip power-law fit, her ne kadar alt bölgelerinde bir exponansiyel cut-off gösterse

de Newman'ın çalışmasında bulunduğu Medline veritabanı için 6.2, Los Alamos Arşivi için 3.34, NCSTRL veritabanı için 4.6 ve SPIRES veritabanı için 2.18 katsayısına sahip power-law eğrileri ile uyum göstermektedir [13].

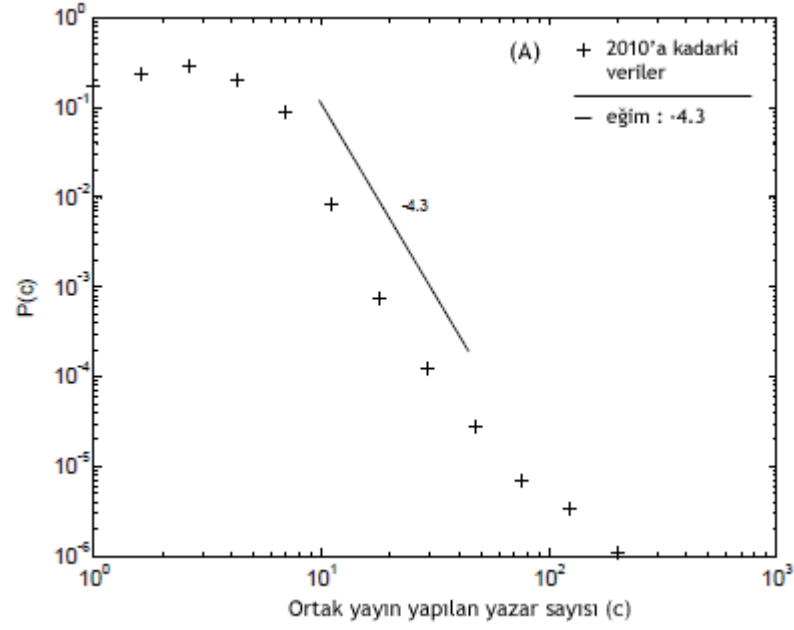


a) Yayın başına düşen yazar sayılarının oluşma frekansları.

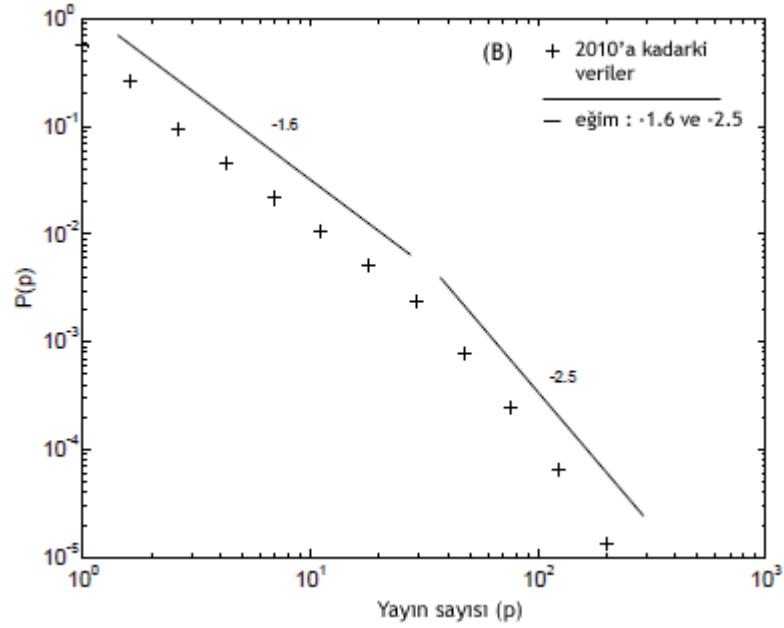


b) Yazar başına düşen yayın sayılarının oluşma frekansları.

Şekil 4.8. Yayın başına düşen yazar ve yazar başına düşen yayın sayılarının oluşma frekansları.



a) Yayın başına düşen yazar sayılarına ait dağılım grafiği ve uygulanan power-law fit.



b) Yazar başına düşen yayın sayılarına ait dağılım grafiği ve uygulanan power-law fit.

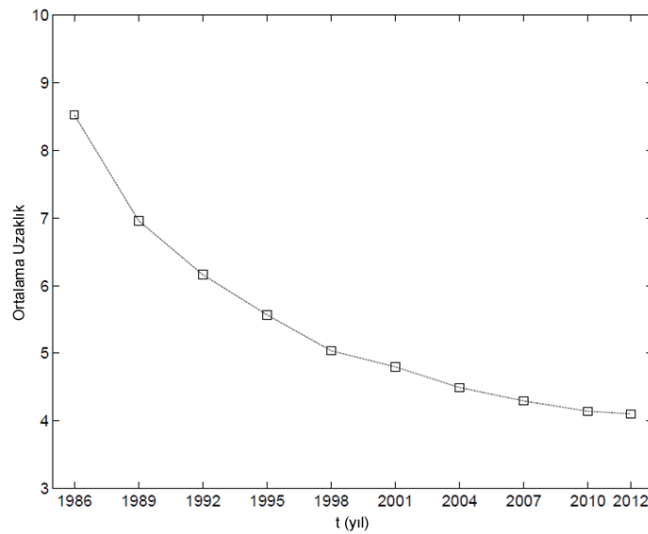
Şekil 4.9. Yayın başına düşen yazar ve yazar başına düşen yayın sayılarının dağılım grafikleri (log-bin uygulanmış ve normalize edilmiş).

Şekil 4.8b ve 4.9b grafiklerinde sunulan yazar başına düşen yayın verileri de power-law karakteristiği sergilemektedir. Şekil 4.8b'deki scatter grafiğini sınırlayan $\gamma=2.2$

ve $\gamma=2.9$ katsayılı power-law eğrilerinin hissettirdiği power-law uyumlu yapı, Şekil 4.9b'deki log binning uygulanıp normalize edilmiş dağılım grafiğindeki çift fazlı power-law yapı (sınırlayan $\gamma=1.6$ ve $\gamma=2.5$) ile daha anlaşılır hale gelmiştir. Bu veriler de Newman'ın çalışmasındaki Medline veritabanı için 2.86, NCSTRL veritabanı için 3.41 katsayılı power-law eğrileri ile benzeşmekte, Lothka'nın çalışmasında elde ettiği $\gamma=2$ katsayılı dağılıma büyük uyum göstermektedir [13,43].

4.2.5. Yazarların Birbirine Ortalama Uzaklığının Yıllara Göre Değişimi

Çalışmamızın en çok hesaplama zamanı gerektiren prosedürlerinden “ortalama uzaklık” değeri, bir ağı tanımlayan en önemli parametrelerden biridir. Ağdaki iki düğümün (i ve j) birbiri ile bağlantı kurabilme kabiliyeti, aralarındaki en kısa yolun uzunluğuna bağlıdır. Bu değer, ağdaki tüm düğüm çiftlerinin arasındaki uzaklıkların bulunması ve bu uzaklık değerlerinin ortalamalarının alınması ile hesaplanmakta ve $d = \langle l_{ij} \rangle$ ile gösterilmektedir. Ortalama uzaklık değeri, bir ağın ayrıklığını ya da iç bağlılığını ifade etmektedir. Geniş ağlar, şaşırtıcı derecede küçük ayrıklık değeri gösterebilmektedir. Literatürde *small-world* olarak anılan bu tür ağlar, Milgram'ın posta deneyinde ya da Twitter gibi sosyal ağlar üzerinde yapılan çalışmalarda gibi 3-6 aralığında ortalama uzaklık değerleri gösterirler ve bu değerler ağı oluşturan düğümlerin birbirine fazlasıyla yakın olduğu anlamına gelir [4,44].



Şekil 4.10. Yazarların birbirine ortalama uzaklığı. Değerler, ilgili yıla kadar olan birikimli veri üzerinde hesaplanmıştır.

Ağdaki tüm düğüm çiftleri arasında gerçekleştirilmesi çok uzun zaman gerektiren hesaplama prosedürü, genellikle ağın bütününden homojen biçimde daha az sayıda düğüm seçilmesi ve hesaplamanın bu alt küme üzerinden yapılması ile makul zaman dilimlerinde gerçekleştirilebilmekte ve bu yöntem ile kabul edilebilir yaklaşıklıkta sonuçlar elde edilebilmektedir [8].

İki düğüm arasındaki mesafeyi bulmak için çeşitli algoritmalar kullanılmaktadır. Bakhshandeh ve arkadaşları, bu algoritmalar üzerinde yaptıkları karşılaştırmalı çalışmanın sonucunda “Bidirectional Search: Breadth-first (optimal)” algoritmasının optimum işlem yükü ile gerçekçi değerler elde edebildiği sonucuna varmışlardır [44]. Bu algoritmaya göre, aralarındaki mesafe bulunacak iki düğüm için iki ayrı “komşu” listesi oluşturulacaktır. Bu listeler öncelikle bu düğümlerin 1. komşuları ile doldurulacaktır. Algoritma, iki liste arasında dönüşümlü olarak çalışmaktadır. Her listedeki düğümlerin komşuları sırayla listeye eklenmekte, yeni liste diğer düğüme ait liste ile karşılaştırılmaktadır. Listeler arasında eşleşme bulunduğu anda iki düğüm arasındaki en kısa yol da bulunmuş olmaktadır. Bu esnada her iki listenin de kaçınıcı komşuluğu tutmakta olduğu bilindiğinden, aradaki düğüm sayısı bu komşuluk dereceleri toplanarak bulunabilmektedir.

Şekil 4.10’da veri tabanımızın yıllara göre sergilediği ortalama uzaklık değerleri görülmektedir. Ortalama uzaklığın 1980’deki 8.52 gibi bir değerden 2012 yılında 4.1’e kadar düştüğü görülmektedir. Elde edilen değerler literatürdeki çalışmalarla karşılaştırıldığında Barabási ve arkadaşlarının 1998 yılı için dünya çapında sinirbilim ve matematik alanları için elde ettikleri 6 ve 9 civarındaki değerler (çalışmamızda 5.03), Perc’in 2010 yılı için Slovenya veri tabanında elde ettiği 4.7 civarındaki değer (çalışmamızda 4.14), Newman’ın 1999 yılı için MEDLINE, Los-Alamos, SPIRES ve NCSTRL veri tabanları üzerinde elde ettiği 4.6, 5.9, 4.0 ve 9.7 değerleri (çalışmamızda 1998 yılı için 5.03) ve bu değerlerin zaman içerisinde gösterdikleri değişimler ile kısmen benzerlik gösterdiği görülmektedir [7,9]. Nitekim genellikle 5-10 yıl aralığındaki zaman dilimlerini içeren veri tabanlarının incelendiği bu çalışmalardan daha geniş bir veri seti (33 yıl) üzerinde çalışmış olmamız, sınırlı veri setlerinin sebep olduğu artifaktlardan daha az etkilenmiş sonuçlar elde etmemizi

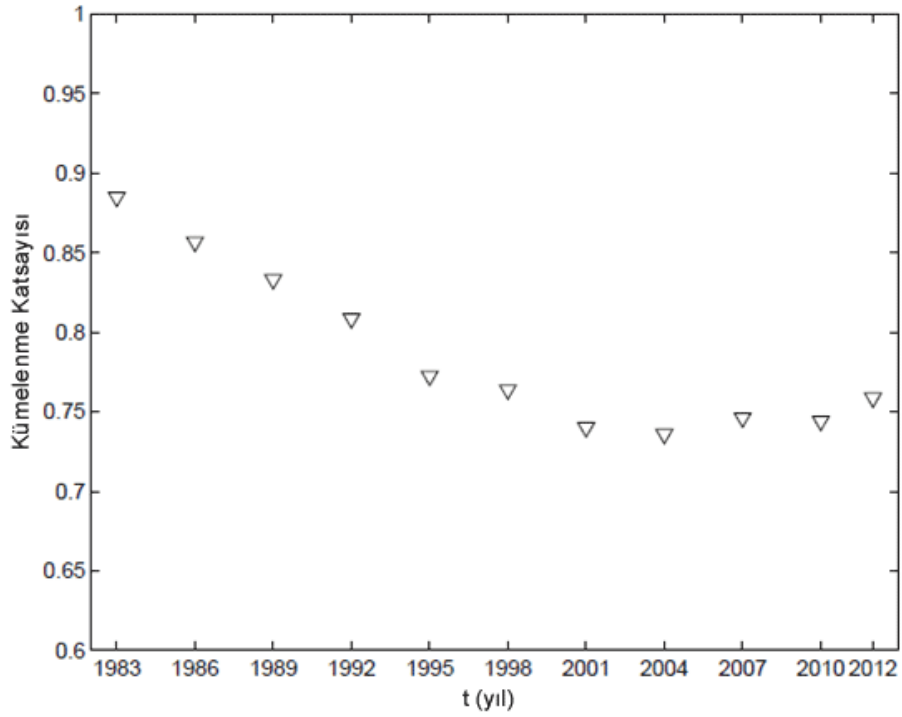
sağlamıştır. Bunun bir göstergesi de ortalama uzaklık eğrisinin 2012 yılına yaklaştıkça yataya yakın bir seyir izlemekte oluşudur.

Ancak Bölüm 4.1.2’de de yer verdiğimiz gibi, bilimsel ağ çalışmalarında yazarların soyadları ile birlikte isimlerinin baş harfinin indekslenmiş olması, ağda aynı baş harf ve soyadına sahip birkaç düğümün aynı düğüm olarak temsil edilmesine yol açmaktadır. Aynı zamanda aynı aileden gelen yazarların yol açtığı bu tür benzerlikler de bu hata faktörünü artırıcı etki yapmaktadır. Bu da, farklı disiplinler arasındaki yazarlar arasında bağlantılar oluşturmakta, bu da disiplinler arasında yalancı geçişler ve yakınlıklar tanımlamaktadır. Bu durumun, ağda rastgele seçilen iki düğümün arasındaki uzaklığın hesaplanmasında, uzaklık değerini düşürücü etki oluşturduğunu düşünmekteyiz. Newman’ın ağın bütününe yüzde birkaç derecelik etki öngördüğü bu duruma rağmen, ağımızın small-world konsepti ile fazlasıyla uyumlu olduğu, büyük boyutuna rağmen sergilediği düşük düğümler arası uzaklık değerlerinden anlaşılmaktadır.

4.2.6. Kümelenme Katsayısının Yıllara Göre Değişimi

Kümelenme katsayısı, gerçel ağların tamamen rassal Erdős-Renyi modelinden farkını karakterize eden önemli bir parametredir. Bir düğüm için kümelenme katsayısı C ’nin hesaplanması, şu algoritma ile gerçekleştirilmektedir:

- i) Ağ içerisinde k_i düğüm ile bağlı bir i düğümü seç.
- ii) Eğer bu k_i düğümün tamamı, diğer her biri ile bağlı olsaydı aralarında $k_i(k_i-1)/2$ bağlantı bulunacaktı. Fakat gerçekte bundan daha az sayıda düğüm bulunur. Gerçekte bu komşuların aralarındaki düğüm sayısı E_i ’yi bul,
- iii) i . düğüm için kümelenme katsayısını $C_i = 2E_i/k_i(k_i-1)$ olarak hesapla.
- iv) Kümelenme katsayısını ağdaki tüm düğümler için bul ve bu C_i değerlerinin ortalamasını alarak tüm ağın kümelenme katsayısını bul.



Şekil 4.11. Kümelenme katsayısının yıllara göre değişimi.

Bilimsel işbirliği ağında kümelenme katsayısının anlamı, bir yazarın (birlikte yayın yaptığı) komşularının kendi aralarında yayın yapma eğilimidir. Aynı zamanda bu komşulardan ikisinin birlikte yayın yapma olasılığı olarak da düşünülebilir. Topolojik olarak ise bir düğümün komşu iki düğümü ile birlikte oluşturduğu üçgenlerin yoğunluğu olarak ifade edilir.

Veri setimiz üzerinde yıllar bazında o yıla kadarki birikimli kayıtlar üzerinden yapılan hesaplamaların sonucunda Şekil 4.11'deki grafik elde edilmiştir. Ağımız için 0.75 gibi dikkat çekici bir değere yakınsayan kümelenme katsayısı, düğümler arasındaki bağlılığın üst seviyede olduğunun önemli bir göstergesidir. Bu değer, Barabási ve arkadaşlarının Neuroscience veritabanı için elde ettikleri ~ 0.76 değeri haricindeki hiçbir çalışmada gözlenmemiştir [7].

Ağ tanımlamalarının bir ucu, kümelenmenin maksimum olduğu ancak düğümler arası mesafenin çok büyük olduğu düzenli graflar iken, diğer uç olan rassal ağlarda kümelenme katsayısı ve düğümler arası mesafe çok düşüktür. Sergilediği düşük

düğümler arası uzaklık ve yüksek kümelenme katsayısı ile bu iki tanımın ortasında yer alan ağımız, small-world tanımlamasıyla fazlasıyla uyum içindedir.

4.2.7. Ortalama Derecenin Değişimi

Ortalama derecenin bir başka anlamı da düğüm başına düşen ortalama bağlantı sayısıdır. Ağ yeni düğümlerin eklenmesi ile sürekli büyürken, mevcut düğümler arasında oluşan yeni bağlantıların sayısı da sürekli artmaktadır. Ağın iç bağlılığını tanımlayan önemli büyüklüklerden birisi olan ortalama derece, bu bağlantıların sayısının toplam yazar sayısına bölünmesi ile bulunmaktadır. Ancak düğümler arasındaki bir bağ, bağlı olduğu 2 düğümün derecesine de katkıda bulunduğu için ortalama derece şu şekilde hesaplanmaktadır:

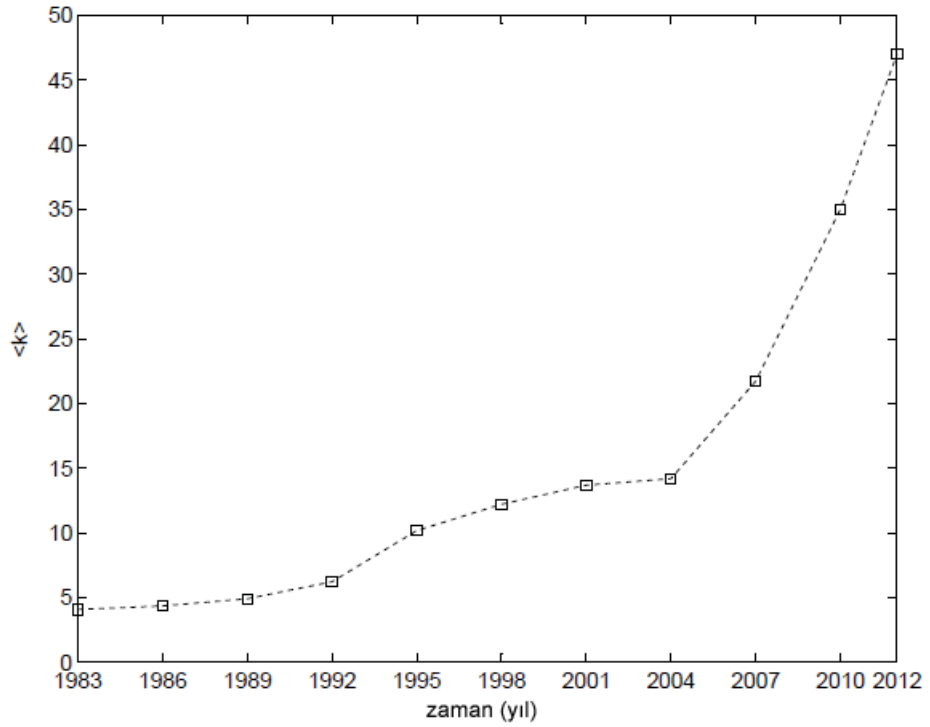
$$\langle k \rangle = \frac{2 \times \text{toplam bağlantı sayısı}}{\text{toplam düğüm sayısı}} \quad (4.1)$$

Ortalama derecenin yıllara göre değişimi Şekil 4.12’de, log-lineer ölçekteki çizimi (exponansiyel fit uygulanmış biçimde) Şekil 4.13’te sunulmuştur.

Ortalama derece grafiğinin özellikle 2004 yılından sonra hızlı bir artış trendine girmesi, yakın geçmişimizde ülkemizin bilimsel işbirliği ağının iç bağlılığının önemli ölçüde arttığını göstermektedir.

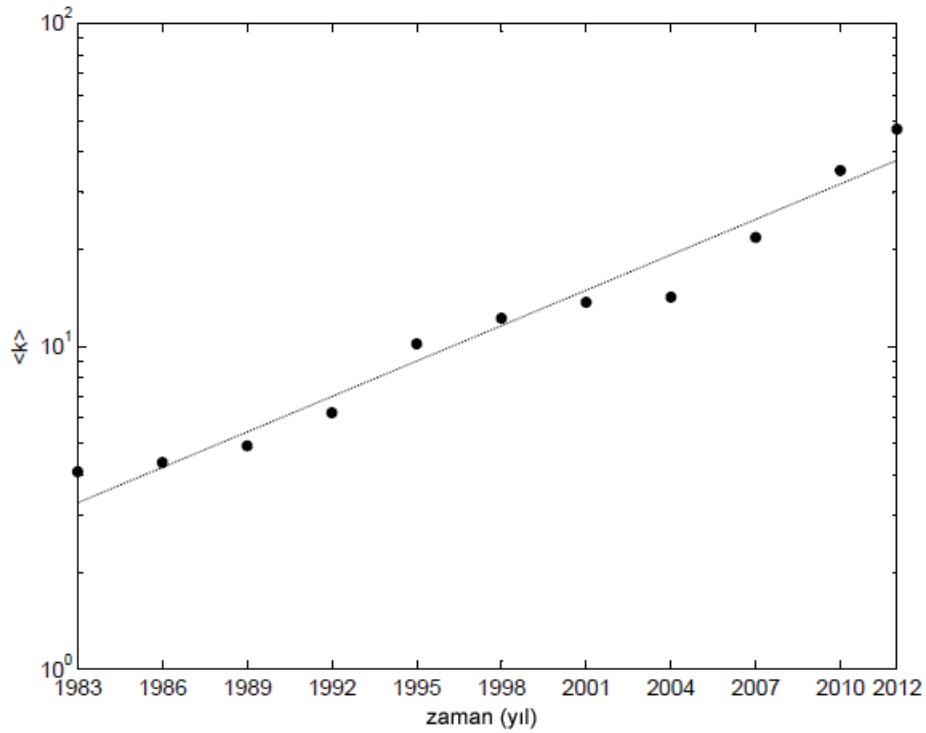
Literatürdeki çalışmaların kapsadığı yıllar da göz önünde bulundurulduğunda ağımızdaki düğümlerin ortalama derece değerinin (2012 yılında 47), MEDLINE (1999 yılında 18.1), Los Alamos (1999 yılında 9.7), NCSTRL (1999 yılında 3.59), Neuroscience (1998 yılında ~12), Mathematics (1998 yılında ~4), Slovenya ulusal veritabanı (2010 yılında 10.7), İsviçre ulusal veritabanı (2006 yılında ~4) ve Mathematical Reviews (1999 yılında 2.94) veritabanlarına oranla çok yüksek olduğu görülmektedir [9,13,14,45,46]. Bu durumun özellikle Türkiye’de yükseköğrenime yönelik milli politikaların son yıllarda bilimsel çalışmaları daha çok destekleyen bir yapıya bürünmesi ve spesifik olarak da son yıllarda çok yazarlı araştırma makalelerinin sayısında artış olmasıdır. Ağımızı tanımlarken bir yayın için

varsaydığımız maksimum yazar sayısı olan 500 değerine ulaşan, hatta bu değeri geçtiği için yazarların bir kısmını dışarıda bırakmak zorunda kaldığımız yayınların sayısı son yıllarda artış göstermiştir. Normal bir araştırma makalesinin pratikte ulaşamayacağı bu yazar sayıları, özellikle uygulamalı fizik ve tıp alanında yürütülen projelerde yer alan tüm araştırmacıların o yayına ait yazar listesine eklenmesi ile ortaya çıkmaktadır. Çalışmamızın kapsamı dışında bırakmadığımız bu çalışmaların sayısındaki artış da ortalama derecedeki hızlı artışın önemli bir nedenidir.



Şekil 4.12. Ortalama derecenin yıllara göre değişimi.

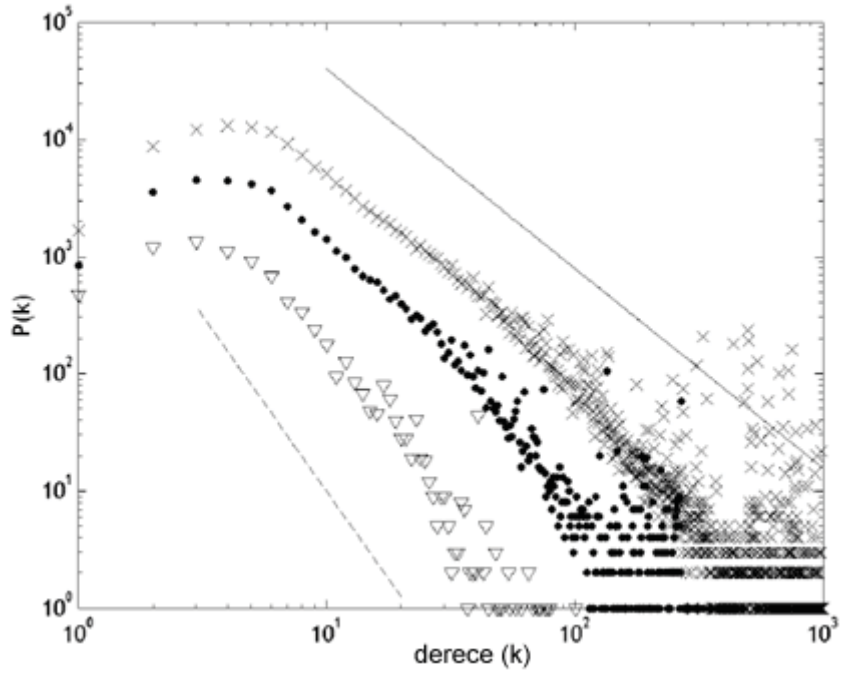
Şekil 4.13'te logaritmik ekseninde çizdirdiğimiz ortalama derece değişim grafiğinin uyguladığımız eksponansiyel fite uyumlu yapısı, ortalama derece eğrisinin eksponansiyel artış fazında olduğunu göstermektedir.



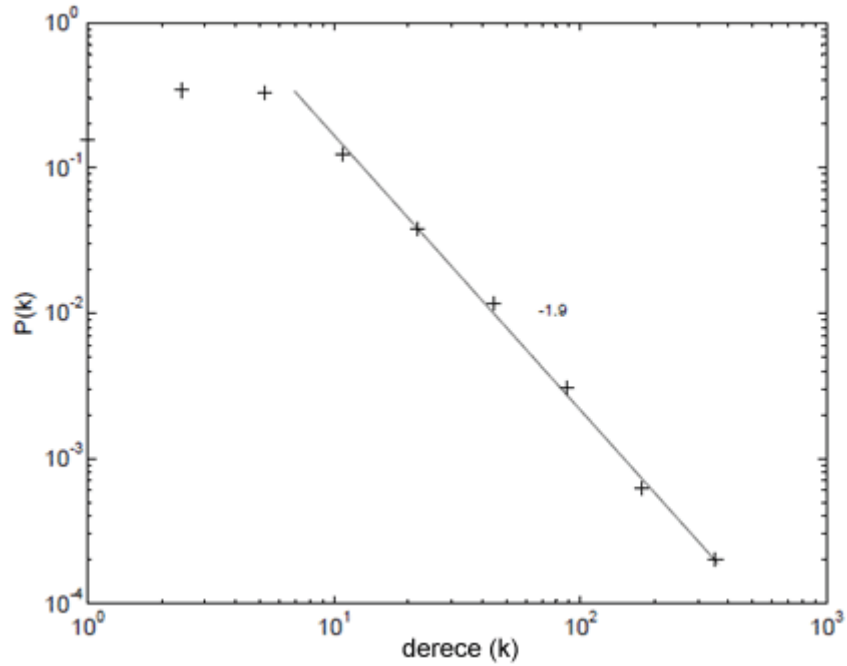
Şekil 4.13. Ortalama derecenin yıllara göre değişimi (log-lineer çizim). Uygulanan eksponansiyel fit : $y(x)=ae^{bx}$ olmak üzere, $a=1.05 E-0.72$, $b=0.084$.

4.2.8. Derece Dağılımı

Ağ topolojisi hakkında fikir veren en önemli parametrelerden olan derece dağılımı (P_k), ağda rastgele seçilen bir düğümün k adet bağlantıya sahip olma ihtimali olarak ifade edilmektedir [7]. Ağ üzerinde yaptığımız çalışmada derecelerin oluşum sıklıklarını ve derece dağılımını Şekil 4.14'teki gibi elde ettik. P_k 'nin power-law dağılımına sahip olması, ilgili ağın scale-free olarak sınıflandırılabilceği anlamı taşımaktadır [6,7]. Şekil 4.14a'da sunulmuş olan derece oluşum sıklıkları grafiğinde iki taraftan uygulanmış power-law fit doğruları (log-log ölçekte doğru olarak ifade edilirler) dağılımın power-law olduğunu hissettirse de bunun somut olarak ispatlanması için verilere log-binning ve uygun bir power-law fit uygulanması gerekmektedir. Aynı zamanda veri setinin bir "dağılım" olarak sunulabilmesi için normalize edilmiş olması gerekmektedir. Bu işlemlerden geçirmiş olduğumuz veri setimize ait dağılım grafiği elde edilmiş, Şekil 4.14b'de sunulmuştur.



- a) Ağdaki düğümlerin derecelerinin oluşum sıklıkları. Değerler 1990 (∇), 2000 (\bullet) ve 2010 (\times) yıllarına kadar olan birikimli veri üzerinden hesaplanmıştır. Verilere uygun eğrilerin eğimleri sırasıyla -3 (kesikli) ve -1.7'dir (düz).



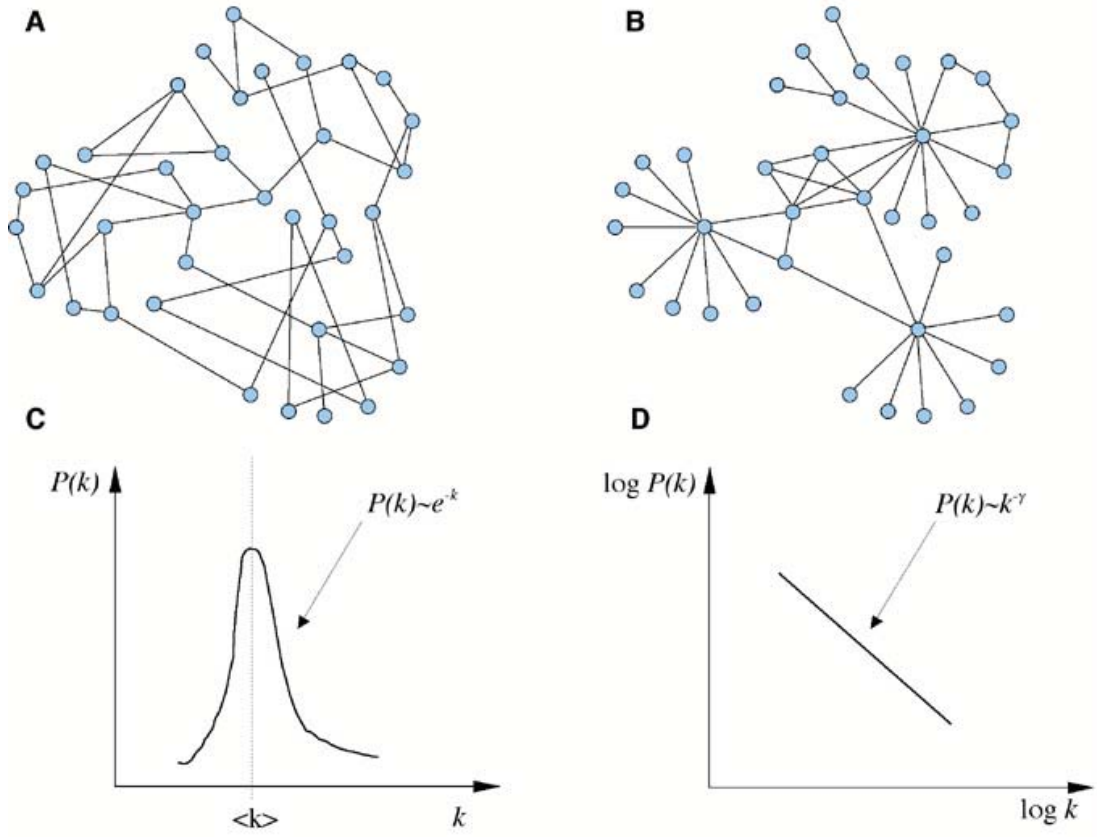
- b) 2010 yılına kadarki verilerin log-binning ve normalize edilmiş hali (derece dağılımı). Düz çizgi, eğimi -1.9 olan power-law fit'i temsil etmektedir.

Şekil 4.14. Derece dağılımı.

Scale free ağlarda k ve P_k arasındaki ilişki $P(k) \sim ck^{-\gamma}$ şeklinde olup γ tipik olarak $2 < \gamma < 3$ aralığında değer almaktadır (bu çalışmada -1.9 civarında elde edilmiştir). Bununla birlikte ağda, ağ ortalamasının çok üzerinde dereceye sahip düğümler (hub) bulunmaktadır (Şekil 4.8, Şekil 4.14a). Bu düğümlerin varlığı, ağda seçimli bağlanmanın varlığının da bir göstergesidir. Scale-free özelliği aynı zamanda ağın başarısı ile de ilişkilidir. Bu tür ağlarda popüler düğümler daha az popüler olanlar tarafından takip edilirken, bu düğümler de derecesi çok daha düşük olanlar tarafından takip edilmektedir. Bu hiyerarşik yapı, ağda hata düzeltici etkiye sahiptir. Oluşan rassallıkların bir hub'a olan etkisi ihmal edilebilecek düzeydedir. Bir hub çokse bile ağ, diğer hub'lar sayesinde iç bağlılığını korumaya devam edecektir. Öte yandan süper-hub'lardan birkaçı ağdan uzaklaştırılırsa, ağ izole kümeler setine dönüşür. Dolayısıyla hub'lar bir ağın hem güçlü, hem de zayıf noktalarıdır [47]. Poisson ve power-law derece dağılımları, Şekil 4.15'teki gibi rassal ve scale-free ağ topolojilerinin birer sonucudur [48,49].

Scale-free ağların bir diğer karakteristik özelliği de ortalama derece gittikçe artarken kümelenme katsayısının azalmasıdır [47,50]. Ağımızın ortalama derece ve kümelenme katsayısı değişimleri de bu karakteristikleri taşımaktadır (Şekil 4.11 ve 4.12).

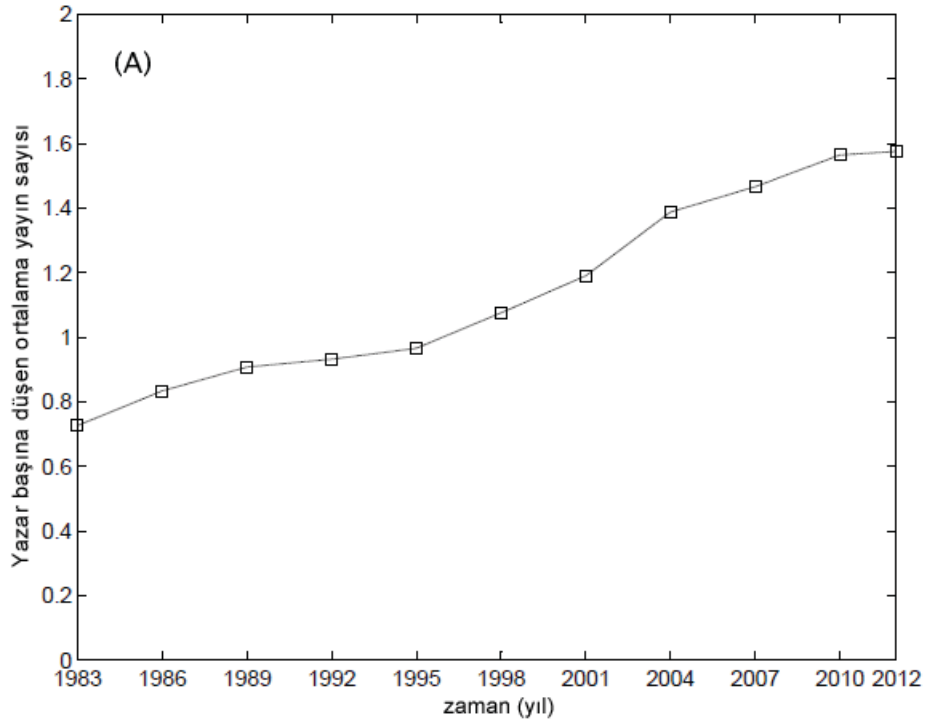
Derece oluşum sıklıklarının 10'ar yıllık aralıklarla çizdirildiği Şekil 4.14a'da, derece oluşum frekanslarının zamana bağlı bir karakter gösterdiği görülmektedir. 1990'dan 2010 yılına doğru gelindikçe grafiği sınırlayan power-law eğrilerinin γ katsayıları 3'ten 1.7'ye doğru düşüş göstermektedir. Power-law uyumunun izlenmesini zorlaştıran, derece sıklık grafiğinin kuyruk kısmındaki saçılmalar, Şekil 4.11b'de elde edilen dağılım grafiği ile bertaraf edilmiş olup, bu grafiğin yüksek regresyon değerine sahip power-law uyumu, ağımızın scale free özelliğinin kusursuz yapıda olduğunun önemli bir göstergesidir.



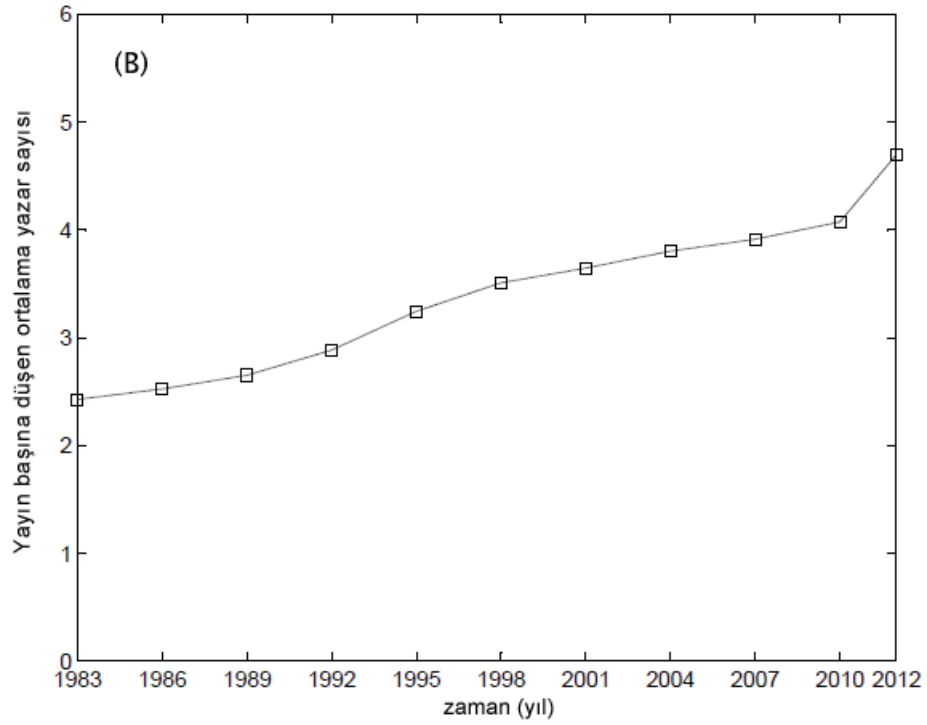
Şekil 4.15. a) Rassal ağ; b) Scale-free ağ; c) Rassal ağa ait derece dağılımı; d) Scale-free ağa ait derece dağılımı [49].

4.2.9. Yazar Başına Düşen Ortalama Yayın ve Yayın Başına Düşen Ortalama Yazar Sayılarının Yıllara Göre Değişimleri

Yayın ve yazar sayıları arasındaki ilişkileri incelemek için Şekil 4.16'da görülmekte olan yazar başına düşen ortalama yayın ve yayın başına düşen ortalama yazar sayılarının yıllara göre değişim grafiklerini elde ettik.



a) Yazar başına düşen ortalama yayın sayısı



b) Yayın başına düşen ortalama yazar sayısı

Şekil 4.16. Yazar başına düşen ortalama yayın ve yayın başına düşen ortalama yazar sayılarının yıllara göre değişimleri.

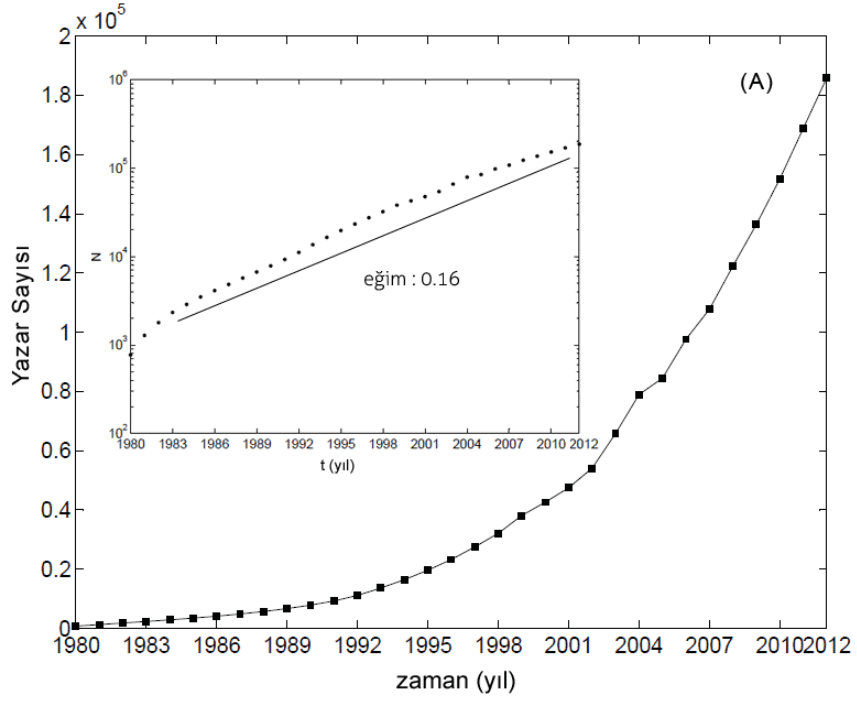
Şekil 4.16a incelendiğinde, yazarların üretkenliklerinin zaman içerisinde artış gösterdiği görülmektedir. 1983 yılı için 0.73 olan yazar başına düşen yayın değeri, 2012 yılında 1.575'e ulaşmıştır. Bununla birlikte Şekil 4.16b'de yayın başına düşen ortalama yazar sayısının da artış trendi gösterdiği, bu değer aynı yıl aralığında 2.43'ten 4.7'ye çıktığı görülmektedir. Bu grafikten yola çıkarak yazarların işbirliği eğiliminin gittikçe arttığı söylenebilir.

Çalışmamızda kullandığımız veri setinin Web of Science tarafından taranan tüm bilim dallarını ayırma yapmadan içeriyor olmasının, bilimsel işbirliğine ilişkin ortalama değerleri etkileyen önemli bir etken olduğunu düşünüyoruz. Çünkü tıp ve uygulamalı fizik gibi bazı dallarda yürütülen araştırma projelerinin sonucu olarak ortaya çıkan bazı yayınlarda, projelerde yer alan yüzlerce araştırmacı da yazar olarak yer almaktadır. Bu yayınlar, aynı zamanda Şekil 4.8a ve 4.14a'nın sağ tarafındaki saçılmaların da başlıca sebebidir. Bu saçılmalar, özellikle Şekil 4.14a'daki derece dağılımı dikkate alındığında, ilk 10 ve 20 yıllık verilerden çok 30 yıllık birikimli veri üzerinde daha fazla izlenmektedir. Bu da çok yazarlı araştırma projelerinin özellikle son 10 yılda yoğunlaşmış olduğunun bir göstergesidir.

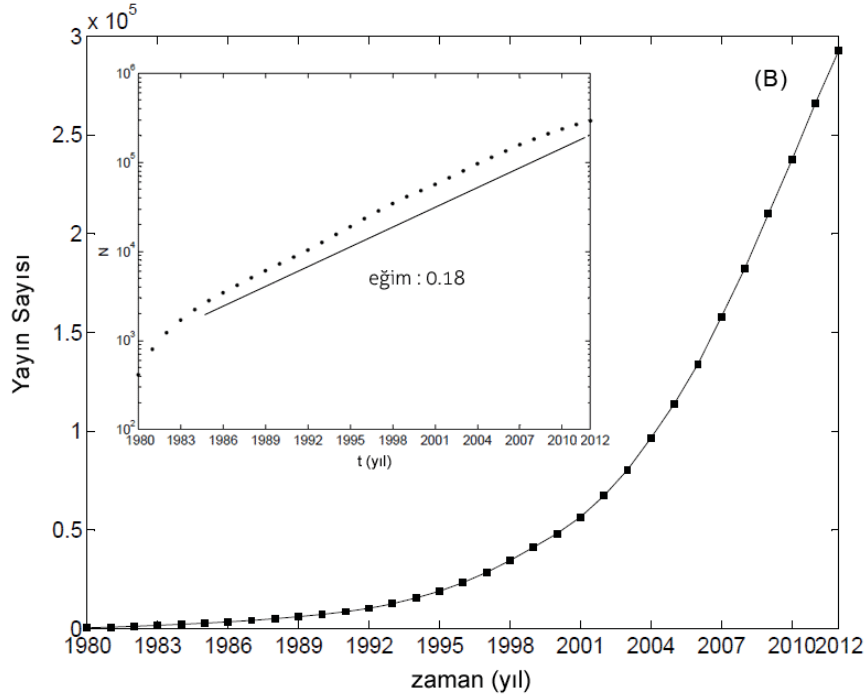
4.2.10. Ağın Genişleme Hızının İncelenmesi

Şekil 4.6 ve 4.7'de görselleştirilmiş olan verilere dayanarak yapmış olduğumuz exponansiyel büyüme tahminini somutlaştırabilmek için bu şekilde yer alan 2 grafiğin de log-lineer ölçekte çizdirilmesi gerekmektedir. Log-lineer ölçekte çizdirilen verilerin doğrusal bir görüntü vermesi durumunda exponansiyel büyümeden söz edilebilmektedir [23,39]. Ağın genişleme hızı ile ilgili somut bir sonuca varabilmek için Şekil 4.6 ve 4.7'deki veriler log-lineer ölçekte yeniden çizdirilmiş, Şekil 4.17'deki grafik elde edilmiştir.

Şekil 4.17'deki log-lineer ölçekteki çizimler ve uygulanan exponansiyel fit'ler incelendiğinde, ağın zaman içerisinde exponansiyel büyüme karakterine sahip olduğu anlaşılmaktadır.



a) Ağda tanımlı yazar (node) sayısının yıllara göre değişimi. İçteki grafikte log-lineer ölçekte çizdirilmiş ve $a \cdot \exp(bx)$ formunda exponential fit uygulanmıştır ($b=0.16$).



b) Yayın sayısının yıllara göre değişimi. İçteki grafikte log-lineer ölçekte çizdirilmiş exponential fit uygulanmıştır ($b=0.18$).

Şekil 4.17. Türkiye'nin bilimsel işbirliği ağının büyüme istatistikleri.

Fiziksel ağların büyüme hızları ile ilgili Smith ve arkadaşları, ağın hızını tanımlayan “network velocity” kavramını ortaya atmışlardır [39]. Bu yaklaşım, yukarıda incelemiş olduğumuz düğüm sayısında zamanla eksponansiyel büyüme kavramını da içine almakla birlikte, düğümlerden çıkan linklerin sayısındaki değişimin düğüm sayısındaki değişime oranını o ağa ait “hız” olarak tanımlamaktadır. Bununla birlikte network hızının zaman eksenindeki eğimini de ağın büyüme ivmesi olarak adlandırmaktadır.

Herhangi bir t anında ağdaki düğümler arası toplam link sayısını $M(t)$, toplam düğüm (yazar) sayısını da $N(t)$ olarak kabul ettiğimizde, ağın büyüme hızı $v(t)$ aşağıdaki gibi tanımlanabilir:

$$v(t) \equiv \frac{dM(t)}{dN(t)} = \frac{m(t)}{n(t)} \approx \frac{M(t) - M(t-1)}{N(t) - N(t-1)} \quad (4.2)$$

Bu denklemde $m(t)$ ve $n(t)$ sırasıyla ağa link ve düğüm eklenme oranlarını ifade etmektedir. Bu denklemin türevi alındığında ise aşağıdaki ivme denklemi elde edilmektedir.

$$a(t) = \frac{dv(t)}{dt} \quad (4.3)$$

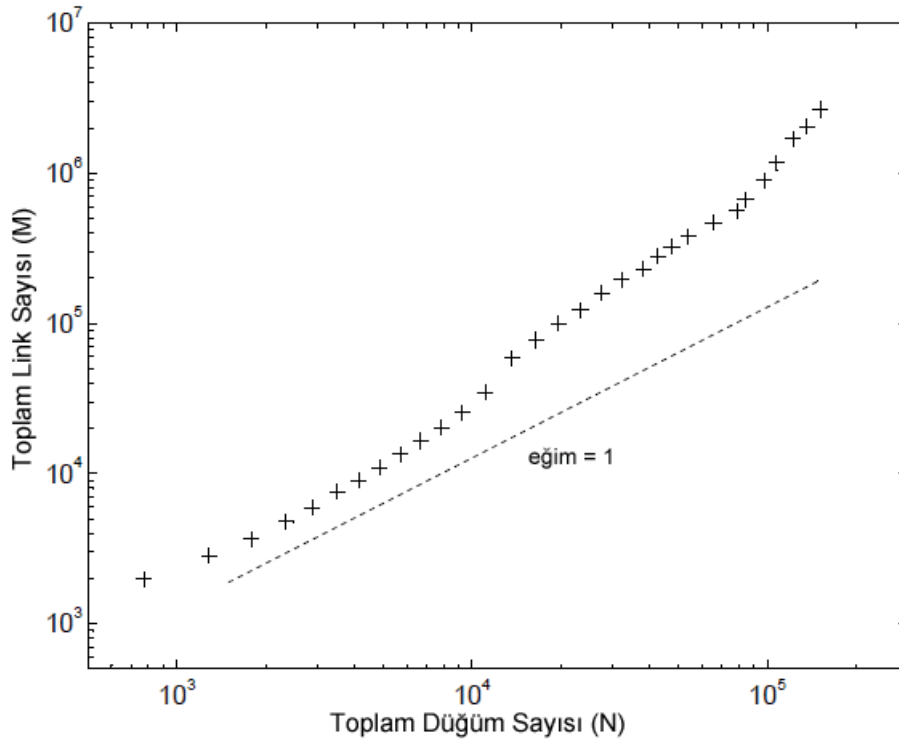
Bu denklemde $a(t)$, ağın büyüme ivmesini ifade etmektedir.

Veritabanımız üzerinde link-düğüm uzayını inceleyerek yıllara göre değişim karakterlerini Şekil 4.18’deki gibi elde ettik.

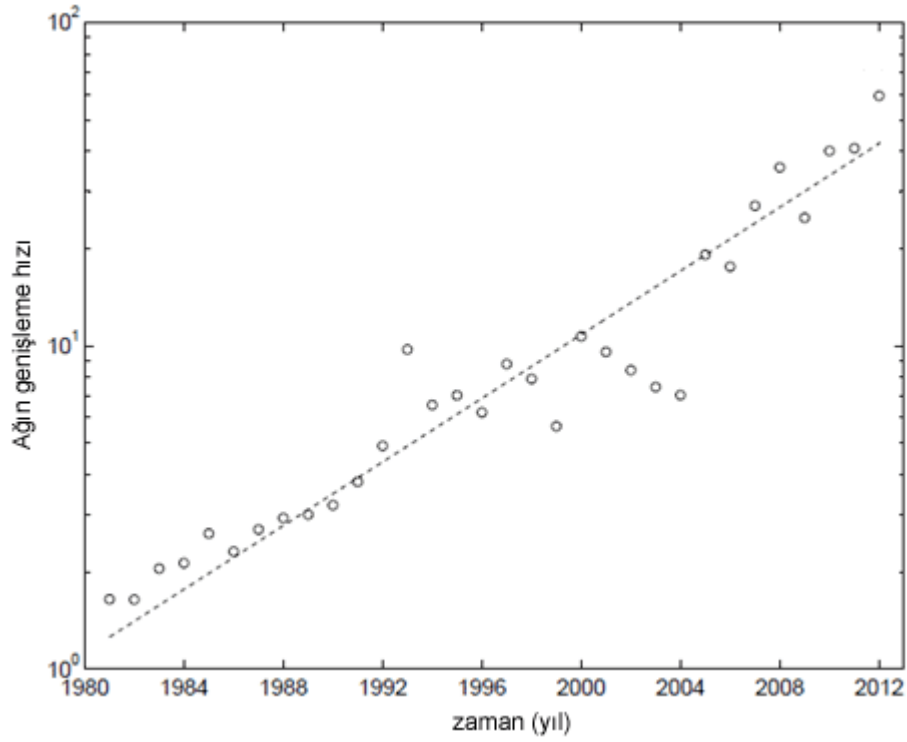
Şekil 4.18 incelendiğinde, toplam düğüm ve link sayılarının sürekli olarak artış gösterdiği görülmekte, bu nedenle ağın gelişim hızının hiçbir zaman negatif olmadığı anlaşılmaktadır. Bu durum, Smith ve arkadaşlarının çalışması ile benzerlik göstermektedir [39]. Bununla birlikte, Smith ve arkadaşlarının incelemiş olduğu 3 farklı ağda ortaya koymuş oldukları, başlangıçta hızlanan fakat zamanla büyüme hızının yatay seyre geçtiği gelişim karakterlerinin aksine, ülkemiz bilimsel işbirliği ağının sürekli olarak pozitif ivme ile büyümekte olduğu Şekil 4.19’daki “network

velocity” grafiğinden anlaşılmaktadır. Yani veritabanımızın tanımladığı ağın, diğer fiziksel ağlar gibi büyüme karakteri ortaya koymakla birlikte, bu karakteri sürdürmekte olması açısından farklılık gösterdiği görülmektedir.

Ağımızın pozitif ivmeyi sürdürdüğü büyüme karakterinin, Türkiye’nin ekonomik gelişimi, yükseköğretimde uygulanan reformlar ve bilimsel çalışmalara ayrılan bütçedeki artış tarafından desteklendiğini düşünmekteyiz. Özellikle son 10 yılda üniversite sayısının 76’dan 170’e çıkmış olması, aynı zaman diliminde akademik personel sayısının 76,090’dan 111,495’e çıkmış olması, bu düşüncemizi destekleyen verilerdir [51,52]. Ağımızın yüksek büyüme hızına rağmen yazar başına düşen ortalama yayın sayısının (2012 yılında 1.58) hala düşük olduğu da söylenebilir.



Şekil 4.18. Toplam link ve düğüm sayılarının link-düğüm uzayındaki değişim grafiği.



Şekil 4.19. Ağın genişleme hızının yıllara göre değişimi (network velocity).

4.2.11. Popüler Yazarlarla Yayın Yapmak

Gerçel ağların önemli bir kısmında seçimli bağlanma olgusu mevcuttur. Yani ağa yeni bağlanan bir düğüm genellikle çok sayıda bağlantısı olan (derecesi yüksek) bir düğüm üzerinden ağa bağlanmayı tercih eder [6]. Seçimli bağlanmanın varlığını sorgulayan birçok çalışma yapılmıştır. Bu çalışmalardan bir kısmı düğüm düzeyinde bağlanma oranlarına odaklanırken, bir kısmı da ağın derece dağılımı karakteristikleri üzerinden değerlendirmelerde bulunmuşlardır [6,7,9,53,54].

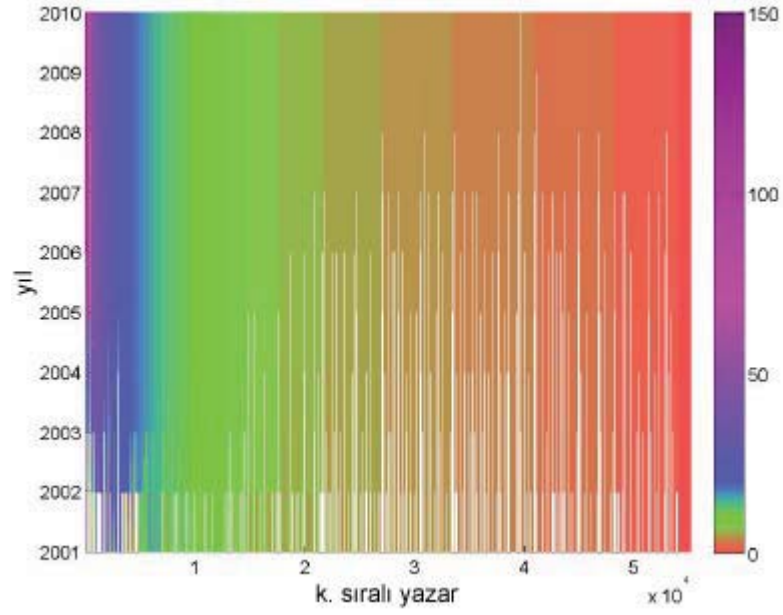
Veritabanımızın tanımlamış olduğu bilimsel işbirliği ağının derece dağılımını gösteren Şekil 4.14b'nin, üzerine uygulamış olan power-law fit eğrisi ile son derece uyumlu yapısı, ağımızda seçimli bağlanma olgusunun varlığını göstermektedir.

Çalışmamızın bu kısmında, ağa yeni bağlanan düğümlerin seçimli bağlanma tercihlerinin kendilerine avantaj sağlayıp sağlamadığını göstermeyi amaçladık. Böyle bir avantajın varlığını sorgulayabilmek için, öncelikle ağımıza 2001-2010 yılları arasında katılan yeni düğümleri, ağdaki popüler düğümlere bağlananlar veya daha az

popüler düğümlere bağlananlar olmak üzere 2 sınıfa ayırdık. Bu durumda, popüler düğüm olma ölçüsünü, “derecesi en az 35 olmak” (2010 yılına ait ortalama derece değeri) olarak belirledik. Bu 10 yıllık süreç içerisinde popüler düğümlerle herhangi bir zamanda (sadece ilk bağlantı olarak değil) bağlantı kuran (ortak yayın yapan) bir yeni düğümü “seçimli bağlanan” olarak işaretledik. Seçimli bağlanan ya da bağlanmayan yazarların yıllar bazındaki derece değişimlerini Şekil 4.20’deki gibi görselleştirebilmek için, her düğümün (yazarın) yıllık çözünürlükte derece değişimlerini içeren bir veritabanı tablosu oluşturduk. Bu veriler doğrultusunda, 2 sınıfa ayırdığımız yeni yazarların 10 yıllık süreç içerisindeki derece değişimlerini Şekil 4.20’deki gibi renk kodlu olarak görselleştirdik.

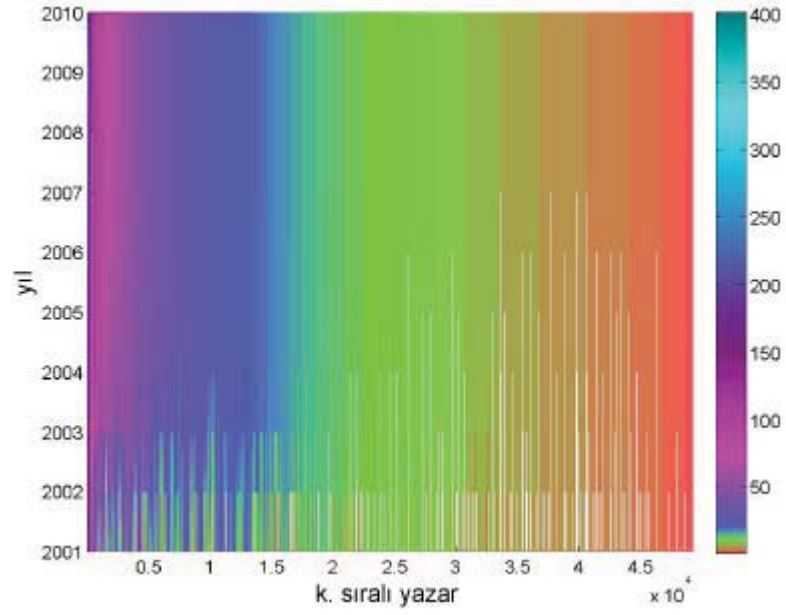
Şekil 4.20’deki 2 grafik incelendiğinde, popüler yazarlarla bağlantı kurmuş olan yeni düğümlerin (sağdaki grafik), zaman içerisinde daha yüksek derecelere ulaştığı görülebilmektedir. Bu durum, aynı zamanda yazarların derecelerinin Zipf-plot olarak çizdirildiği Şekil 4.21 ve Şekil 4.22’de de açıkça görülebilmektedir [55].

Yüksek dereceli düğümlerle bağlantı kuran yeni düğümlerin derecelerinin düz çizgi ile, diğer yeni düğümlerin derecelerinin ise kesikli çizgi ile ifade edildiği bu grafik, gerçek bir kompleks ağda popüler düğümlerle link oluşturmanın, yeni düğümler için zamanla daha popüler olması sonucunu doğurduğunu açıkça ortaya koymaktadır. Yüksek dereceli düğümlere bağlanan yeni düğümlerin 2010 yılı itibariyle ortalama derecelerinin 18.45, diğer yeni düğümlerin ortalama derecelerinin ise 7.43 olması, bu durumu sayısal olarak desteklemektedir.



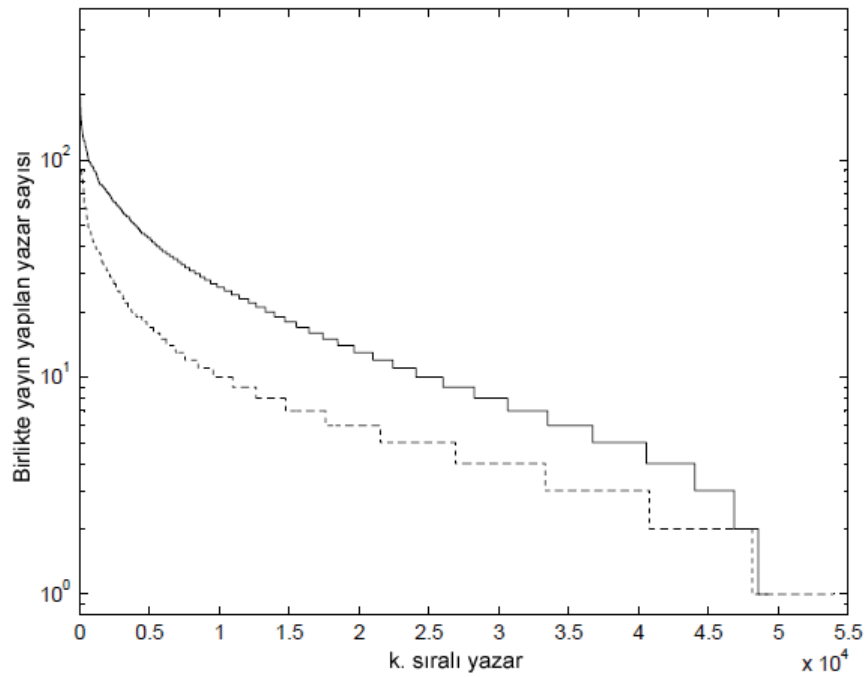
a) Yüksek dereceli yazarlarla yayın yapmamış olan 55,144 yazara ait renk kodlu değişim grafiği.

Şekil 4.20. 2001-2010 yılları arasında ağa eklenen yeni düğümlere ait renk kodlu değişim grafikleri. Her bir düşey renkli çizgi 1 yazarı ifade etmektedir. Yazarların rank (sıralama) değerleri ise 2010 yılı verilerine göre yüksek dereceli olandan düşük dereceli olana doğru sıralama şeklinde yapılmıştır. Yani rank değeri 1 olan yazar, 2010 yılında en yüksek dereceye sahip olan düğümü ifade etmektedir. Her bir düşey çizginin başladığı nokta, o çizginin ifade etmekte olduğu yazarın ilk bilimsel birlikteliğinin (ortak yayın) gerçekleştiği yıldır. Çizginin renk değişimi ise ait olduğu yazarın derecesinin değişimini ifade etmektedir. Hangi rengin hangi derece (bilimsel ortak sayısı) değerine karşılık geldiği, grafiklerin yanındaki renk skalalarından okunabilir. Karşılaştırılabilirlik açısından iki grafikde de aynı renkler aynı derecelere karşılık gelmektedir, ancak maksimum derece değerlerinin daha yüksek olmasından dolayı sağ taraftaki grafiğin spektrumu biraz daha geniştir.

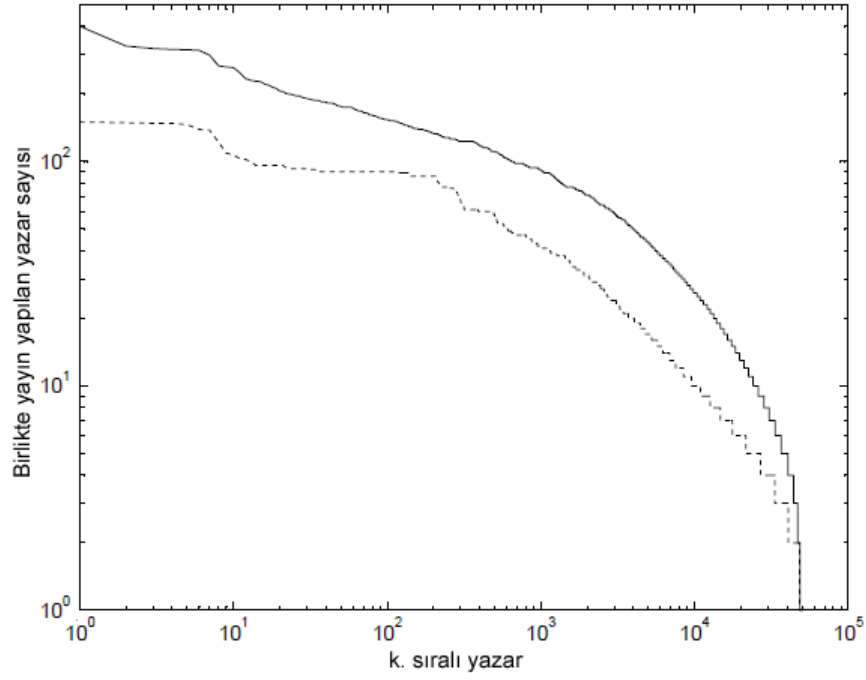


b) Yüksek dereceli yazarlarla bağlantı kurmuş olan 49,246 yazara ait renk kodlu dağılım grafiği.

Şekil 4.20. (devam ediyor).



Şekil 4.21. 2010 yılı verileri kullanılarak elde edilen derece değerlerinin Zipf plot gösterimi. Üstte yer alan düz çizgi, derecesi yüksek olan yazarlarla yayın yapmış olan yeni yazarlara, altta yer alan kesikli çizgi ise bu yazarlarla yayın yapmamış yazarlara aittir.



Şekil 4.22. 2010 yılı verileri kullanılarak elde edilen derece değerlerinin Zipf plot gösterimi (log-log çizim). Üstte yer alan düz çizgi, derecesi yüksek olan yazarlarla yayın yapmış olan yeni yazarlara, altta yer alan kesikli çizgi ise bu yazarlarla yayın yapmamış yazarlara aittir.

Popüler olarak tanımlayabileceğimiz yazarlarla link oluşturmanın sağladığı avantajı gözle görünür hale getiren son 3 grafik, önceden kompleks ağlar üzerinde yapılmış olan ve seçimli bağlanmanın varlığını ispatlayan çalışmalara yani bir bakış açısı ile katkı sağlamış, seçimli bağlanmanın bir gerçel kompleks ağda yeni düğümler için önemli bir avantaj sağlayıcı etken olduğunu göstermiştir.

BÖLÜM 5

SONUÇLAR VE ÖNERİLER

5.1. SONUÇLAR

Bilimsel işbirliği ağları, ağı oluşturan düğümlerin (yazar) diğer düğümlerle gerçekleştirdikleri bağlarda (yayın) kendi kararları ile hareket etmeleri sebebiyle gelişmekte olan kompleks ağlar için iyi bir prototip teşkil etmektedirler. Düğümlerin ağa bağlanma zamanlarının ve bağlantı oluşturma zamanlarının zaman içerisinde kolayca takip edilebilmesini sağlayan bilimsel yayın veritabanları, bu ağların gelişmelerinin incelenmesini de mümkün kılmaktadır. Bu çalışmamızda, Türkiye adresli yazarların ortaya koymuş olduğu yayınlara ait veriler kullanılarak bir ulusal bilimsel işbirliği ağı oluşturulmuş ve bu ağ üzerinde karşılaştırmalı analizler gerçekleştirilmiştir. Ağın son durumuna odaklanmak yerine, ağı tanımlayan parametrelerin 33 yıllık gelişimleri incelenmiştir.

Öncelikle ağa ait yazar ve yayın sayıları, ortalama derece, yazar başına düşen ortalama yayın ve yayın başına düşen ortalama yazar gibi istatistiksel büyüklüklerin yıllık değişimleri incelenmiştir. Yazar ve yayın sayılarının değişimini gösteren ilk 2 grafik, ağın yazar ve yayın bazında eksponansiyel büyüme fazında olduğunu göstermiştir. Takip eden link-düğüm uzayı ve network velocity grafikleri ise ağın büyüme hızının da eksponansiyel artış içerisinde olduğunu ortaya koymuştur. Diğer çalışmaların aksine büyüme hızının zamanla yavaşlayıp sabitlenmemesi durumunu, Türkiye'nin içinde bulunduğu sosyo-ekonomik gelişim trendine yükseköğrenime ayrılan milli kaynaklardaki artışa bağlamaktayız. Bu bağlamda son 10 yıllık periyot göz önüne alındığında, üniversite sayısının 76'dan 170'e, akademik personel sayısının ise 76,090'dan 111,495'e çıktığını görmekteyiz [52]. Ağın sergilediği bu yüksek büyüme hızına rağmen yazar başına düşen yayın sayısının hala yetersiz seviyede olduğu da görülmüştür (2012 yılında 1.57).

Evrilen ağ modellerinin aksine bilimsel işbirliği ağları lineer artış gösteren ortalama derece değerlerine sahiptirler [7,45]. Ağımızın yapısı da bu durumu destekler niteliktedir. Bununla birlikte, yıllar geçtikçe ağımızdaki yazarlar bir yayın için daha fazla yazarla birliktelik kurma eğilimi gösterirken, aynı zamanda ortalama olarak daha fazla yayında pay sahibi olmuşlardır.

Yukarıda bahsedilen istatistiksel parametrelerin paralelinde, ağ üzerindeki hesaplamalı ölçümler de zamanla değişim gösteren sonuçlar göstermiştir. Düğümler arası ortalama uzaklık 4 civarına yakınsarken kümelenme katsayısı ise 0.75 civarında denge bulma eğilimi göstermiştir. Bu küçük mesafe ve yüksek kümelenme karakteristiği, ağımızın iç bağlılığının çok yüksek seviyede olduğunu ortaya koymuştur. Yukarıda bahsedilen ortalama derece karakteristiği de bu durumu destekler niteliktedir. Tüm bu özellikler ağımızın small-world sınıfına dahil olduğunu gösterirken, power-law dağılımına uyum gösteren derece dağılımı da ağımızın scale-free olarak nitelendirilebileceğini göstermiştir. Derece dağılımı grafiğinin power-law uyumluluğu aynı zamanda ağda seçimli bağlanmanın varlığına işaret etmektedir.

Son olarak popüler bir yazara bağlanmanın ileride daha popüler olma sonucunu verip vermeyeceğinin test edildiği çalışmamızda, ağa yeni bağlanan düğümler, popüler yazarlarla bağlantı kuranlar ve kurmayanlar olarak iki sınıfa ayrılmıştır. Yazar aktivitelerinin görselleştirildiği renk kodlu grafiklerin ve Zipf grafiğinin ortaya koyduğu şekliyle, popüler yazarlara bağlanmayı tercih eden yeni yazarların ileride daha yüksek derecelere sahip oldukları görülmüştür. Gerçek zaman ağlarında seçimli bağlanmanın varlığını irdeleyen çalışmalardan farklı olarak, çalışmamızda seçimli bağlanmanın düğümlere avantaj sağladığı ortaya koyulmuştur.

5.2. ÖNERİLER

Türkiye adresli bilimsel yayınların geniş bir zaman aralığında taranarak kompleks ağ olarak tanımlandığı bu çalışmanın, veritabanının büyüklüğü açısından dünya literatürüne, Türkiye adresli yayınlardan oluşması nedeniyle de ülkemiz bibliyografik literatürüne katkı sağlayacağını ümit ediyoruz. Bu çalışmanın daha ileriye götürülmesi açısından veri setinin Web of Science'dan indirilmesi prosesine,

yayınlarnın ait olduđu üniversite ve şehir bilgilerinin de dahil edilmesi ile birlikte Türkiye'nin kendi içindeki bilimsel etkinlik haritasının çıkarılması, ülke literatürüne önemli katkı sağlayacaktır. Aynı zamanda hesaplama prosedürlerini zorlaştıracığı için göz ardı ettiğimiz “yayın türü” ve “bilim dalı” bilgilerinin de eklenmesi ile yapılacak geliştirmelerin, bu tarz bir çalışmaya değer katacağını düşünmekteyiz.

KAYNAKLAR

1. Albert, R. and Barabási, A. L., “Statistical mechanics of complex networks”, *Rev. Mod. Phys.*, 74: 47-97 (2002).
2. İnternet: Oakland University, “Erd Number Project”, <http://www.oakland.edu/enp> (2013).
3. Watts, D. J., “Small worlds: the dynamics of networks between order and randomness”, *Princeton University Press*, Princeton, 128-133 (1999).
4. Milgram S., “The small world problem”, *Psychology Today*, 2: 60-67 (1967).
5. Leskovec J. and Horvitz E., “Planetary-scale views on an instant-messaging network”, *WWW '08 Proceeding of the 17th international conference on World Wide Web*, Banff, 915-924 (2008).
6. Barabási, A. L. and Albert, R., “Emergence of scaling in random networks”, *Nature*, 286: 509-512 (1998).
7. Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A. and Vicsek, T., “Evolution of the social network of scientific collaborations”, *Physica A*, 311: 590-614 (2002).
8. Watts, D. J. and Strogatz, S. H., "Collective dynamics of 'small-world' networks", *Nature*, 393 (6684): 440–442 (1998).
9. Perc, M., “Growth and structure of Slovenia's scientific collaboration network”, *Journal of Informetrics*, 4: 476-482 (2010).
10. Kochen, M., “The Small World”, *Norwood*, 46-92 (1989).
11. Dorogovtsev, S. N. and Mendes, J. F. F., “Scaling behaviour of developing and decaying Networks”, *Europhys. Lett.*, 52: 33-39 (2000).
12. Newman, M. E. J., “The structure of scientific collaboration networks”, *Proc. Natl. Acad. Sci. USA*, 98: 404-409 (2001).
13. Newman, M. E. J., “Scientific collaboration networks: I. Network construction and fundamental results”, *Phys. Rev. E*, 64 (1): 016131/1-016131/8 (2001).
14. Newman M. E. J., “Coauthorship networks and patterns of scientific collaboration”, *Proc. Natl. Acad. Sci, USA*, 101: 5200–5205 (2004).
15. Albert, R., Jeong, H. and Barabási, A. L., “Diameter of the world wide web”, *Nature*, 401: 130-131 (1999).

16. Perc, M., “Zipf’s law and log-normal distributions in measures of scientific output across fields and institutions: 40 years of Slovenia’s research as an example”, *Journal of Informetrics*, 4: 358-364 (2010).
17. Hoekman, J., Frenken, K. and Tijssen, R. J. W., “Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe”, *Research Policy*, 39: 662-673 (2010).
18. Internet: Barabasilab by Barabási, A. L., “BarabásiLab Gallery” <http://barabasilab.com/gallery/g1.php> (2013).
19. Internet: Daily Estimated Size of the World Wide Web, “The Size of the World Wide Web (The Internet)” <http://www.worldwidewebsite.com> (2013).
20. Kumar, R., Raghavan, P., Rajalopagan, S. and Tomkins, A., “Extracting large-scale knowledge bases from the web”, *Proceedings of the 9th ACM Symposium on Principles of Database Systems*, Rome, 1-8 (1999).
21. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajalopagan, S., Stata, R., Tomkins, A., and Wiener, J., “Graph structure in the web”, *Comput. Netw.*, 33: 309 (2000).
22. Adamic, L. A., and Huberman, B. A., “Power-law distribution of the World Wide Web”, *Science*, 287: 2115-2129 (2000).
23. Clauset, A., Shalizi, C. R. and Newman, M. E. J., “Power-law distributions in empirical data”, *SIAM Review*, 51: 661-703 (2009).
24. Adamic, L. A., “The small world Web”, *Proceedings of the Third European Conference, ECDL’99*, Berlin, 443-448 (1999).
25. Faloutsos, M., Faloutsos, P. and Faloutsos, C., “On power-law relationships of the internet topology”, *Comput. Commun. Rev.*, 29: 251-255 (1999).
26. Yook, S., Jeong, H., Barabási, A. L., and Tu, Y., “Weighted evolving networks”, *Phys. Rev. Lett.*, 86: 5835-5844 (2001).
27. Pastor-Satorras, R., Vazquez, A. and Vespignani, A., “Dynamical and correlation properties of the internet”, *Physical Review Letters*, 87 (25): 258701/1-258701/9 (2001).
28. Moltka-Danielsen, J., Trier, M., Shlyk, V., Bobrik, A. and Nurminen, M. I., “IRIS (1978-2006) historical reflection through visual analysis”, *IRIS30 Preceedings*, Tampere, 121-128 (2007).
29. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., Barabási, A. L. , “The large-scale organization of metabolic networks”, *Nature*, 407: 651-655 (2000).
30. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A. L., “The large-scale organization of metabolic networks”, *Nature*, 407: 651 (2000).

31. Williams, R. J., Martinez, N. D., Berlow, E. L., Dunne, J. A. and Barabási, A. L., “Two degrees of separation in complex food webs”, *Proceedings of the National Academy of Sciences*, 12913-12916 (2002).
32. Yodzis, P., “Must top predators be culled for the sake of fisheries?”, *Trends in Ecology & Evolution*, 16 (2): 78–84 (2001).
33. Redner, S., “How popular is your paper? An empirical study of the citation distribution”, *Eur. Phys. J. B.*, 4: 131-142 (1998).
34. Vazquez, A., “Statistics of citation networks”, *Science*, 292: 13-16 (2001).
35. Ferrer i Cancho, R., and Sole, R. V., “The small world of human language”, *Proceedings of the Royal Society B: Biological Sciences*, 268 (1482): 2261-2265 (2001).
36. Scala, A., Amaral, L. A. N., and Barthelemy M., “Small-world networks and the conformation space of a short lattice polymer chain”, *Europhys. Lett.*, 55: 594-612 (2000).
37. Internet: Sourceforge.net, “Clustered Barabási-Albert Network”, http://pagmo.sourceforge.net/pagmo/classpagmo_1_1topology_1_1clustered__ba.html (2013).
38. Jeong, H., Neda, Z., and Barabási, A. L., “Measuring preferential attachment in evolving networks”, *Europhysics Letters*, 61: 567-572 (2003).
39. Smith, D. M. D., Onnela, J. P. and Johnson, N. F., “Accelerating networks”, *New Journal of Physics*, 9: 181-188 (2007).
40. Amaral, L. A. N., Scala, A., Barthelemy, M. and Stanley, H. E., “Classes of small-world networks.”, *Proc. Natl. Acad. Sci. U. S. A.*, 97: 11149-11152 (2000).
41. Bianconi, G., and Barabási, A. L., “Competition and multiscaling in evolving Networks”, *Europhys. Lett.*, 54: 436-442 (2001).
42. Dorogovtsev, S. N. and Mendes, J. F. F., “Scaling behaviour of developing and decaying Networks”, *Europhys. Lett.*, 52: 33-39 (2000).
43. A. J. Lotka, “The frequency distribution of scientific productivity”, *Journal of the Washington Academy of Sciences*, 16: 317-323 (1926).
44. Bakhshandeh, R., Samadi, M., Azimifar, Z. and Schaeffer, J., “Degrees of separation in social networks”, *Proceedings, The Fourth International Symposium on Combinatorial Search*, Barcelona, 18-23 (2011).
45. Tomassini M. and Luthi L., “Empirical analysis of the evolution of a scientific collaboration network”, *Physica A*, 385: 750–764 (2007).

46. Grossman J. W., “Patterns of collaboration in mathematical research”, *SIAM*, 35 (9): 8-9 (2002).
47. Tolba, A., “Scale Free Networks; A Literature Review”, *International Conference on Complex Systems*, Cambridge, 121-129 (2007).
48. Newman, M. E. J., Strogatz, S. H. and Watts, D. J., “Random graphs with arbitrary degree distributions and their applications”, *Phys. Rev. E*, 64 (1) 026118/1-026118/17 (2001).
49. İnternet: Alexei Kurakin, “Network Lectures” <http://www.alexeikurakin.org/main/lecture4Ext.htm> (2013).
50. Barabási, A. L., Albert, R. and Jeong, H., “Mean-field theory for scale-free random networks”, *Physica A*, 272: 173-187 (1999).
51. İnternet: Türkiye İstatistik Kurumu, “2012 Eğitim İstatistikleri”, http://www.tuik.gov.tr/VeriBilgi.do?alt_id=14 (2013).
52. İnternet: Türkiye İstatistik Kurumu, “1923-2011 Ulusal İstatistikler”, http://www.tuik.gov.tr/IcerikGetir.do?istab_id=158 (2013).
53. Newman, M. E. J., “Clustering and preferential attachment in growing Networks”, *Physical Review E*, 64: 025102/1-025102/13 (2001).
54. Jeong, H., Neda, Z. and Barabási, A. L., “Measuring preferential attachment in evolving networks”, *Europhysics Letters*, 61: 567-572 (2003).
55. Zipf, G. K., “Human Behavior and the Principle of Least-Effort”, *Addison-Wesley*, Cambridge, 204-2011 (1949).

EK AÇIKLAMALAR A.

VERİTABANLARI (MS SQL)

1. Veritabanı-1 (collab)

a. “author” tablosu

```
CREATE TABLE [dbo].[author](
    [aID] [int] IDENTITY(1,1) NOT NULL,
    [aSurname] [nchar](30) NOT NULL,
    [aName] [nchar](30) NOT NULL,
    CONSTRAINT [PK_Author] PRIMARY KEY CLUSTERED
(
```

b. “collaboration” tablosu

```
CREATE TABLE [dbo].[collaboration](
    [cID] [int] IDENTITY(1,1) NOT NULL,
    [pID] [int] NOT NULL,
    [aID] [int] NOT NULL,
    CONSTRAINT [PK_collaboration] PRIMARY KEY CLUSTERED
(
```

c. “publication” tablosu

```
CREATE TABLE [dbo].[publication](
    [pID] [int] IDENTITY(1,1) NOT NULL,
    [pType] [nchar](5) NULL,
    [pTitle] [nchar](1000) NULL,
    [pSource] [nchar](1000) NULL,
    [pYear] [smallint] NOT NULL,
    [numAut] [int] NULL,
    CONSTRAINT [PK_publication] PRIMARY KEY CLUSTERED
(
```

2. Veritabanı-II (collab_results)

a. “authorPlus” tablosu

```
CREATE TABLE [dbo].[authorPlus](
    [aID] [int] NOT NULL,
    [aSurname] [nchar](30) NOT NULL,
    [aName] [nchar](30) NOT NULL,
    [d1980] [int] NULL, [d1981] [int] NULL, [d1982] [int] NULL,
    [d1983] [int] NULL, [d1984] [int] NULL, [d1985] [int] NULL, [d1986]
    [int] NULL, [d1987] [int] NULL, [d1988] [int] NULL, [d1989] [int]
    NULL, [d1990] [int] NULL, [d1991] [int] NULL, [d1992] [int] NULL,
    [d1993] [int] NULL, [d1994] [int] NULL, [d1995] [int] NULL, [d1996]
    [int] NULL, [d1997] [int] NULL, [d1998] [int] NULL, [d1999] [int]
    NULL, [d2000] [int] NULL, [d2001] [int] NULL, [d2002] [int] NULL,
    [d2003] [int] NULL, [d2004] [int] NULL, [d2005] [int] NULL, [d2006]
    [int] NULL, [d2007] [int] NULL, [d2008] [int] NULL, [d2009] [int]
    NULL, [d2010] [int] NULL, [d2011] [int] NULL, [d2012] [int] NULL,
    [ranking] [int] NULL) ON [PRIMARY]
```

b. “AuthorsPerPaperY” tablosu

```
CREATE TABLE [dbo].[AuthorsPerPaper_Y] (
    [ID] [int] IDENTITY(1,1) NOT NULL,
    [Anum] [float] NULL,
    [Afreq] [float] NULL,
    [upTo] [int] NULL
) ON [PRIMARY]
```

c. “clusteringCf” tablosu

```
CREATE TABLE [dbo].[clusteringCf] (
    [ccID] [int] IDENTITY(1,1) NOT NULL,
    [aID] [int] NOT NULL,
    [year] [smallint] NOT NULL,
    [cc] [float] NOT NULL
) ON [PRIMARY]
```

d. “DegreeDistribution” tablosu

```
CREATE TABLE [dbo].[DegreeDistribution] (
    [aID] [int] NULL,
    [aDeg] [float] NULL,
    [upTo] [int] NULL
) ON [PRIMARY]
```

e. “DegreeDistributionFreq” tablosu

```
CREATE TABLE [dbo].[DegreeDistributionFreq] (
    [Deg] [int] NULL,
    [Freq] [int] NULL,
    [Year] [int] NULL
) ON [PRIMARY]
```

f. “distances” tablosu

```
CREATE TABLE [dbo].[distances] (
    [dID] [int] IDENTITY(1,1) NOT NULL,
    [dYear] [smallint] NOT NULL,
    [distance] [smallint] NOT NULL,
    [y1] [int] NULL,
    [y2] [int] NULL
) ON [PRIMARY]
```

g. “newAuthor01” tablosu

```
CREATE TABLE [dbo].[newAuthor01] (
    [aID] [int] NOT NULL,
    [aSurname] [nchar](30) NOT NULL,
    [aName] [nchar](30) NOT NULL,
    [d2001] [int] NULL, [d2002] [int] NULL, [d2003] [int] NULL,
    [d2004] [int] NULL, [d2005] [int] NULL, [d2006] [int] NULL, [d2007]
```



```
[int] NULL, [d2008] [int] NULL, [d2009] [int] NULL, [d2010] [int]
NULL, [d2011] [int] NULL, [d2012] [int] NULL,
    [ranking] [int] NULL,
    [super] [int] NULL
) ON [PRIMARY]
```

h. “PapersPerAuthor_Y” tablosu

```
CREATE TABLE [dbo].[PapersPerAuthor_Y] (
    [ID_] [int] IDENTITY(1,1) NOT NULL,
    [Pnum] [float] NULL,
    [Pfreq] [float] NULL,
    [upTo] [int] NULL
) ON [PRIMARY]
```

i. “supernode00” tablosu

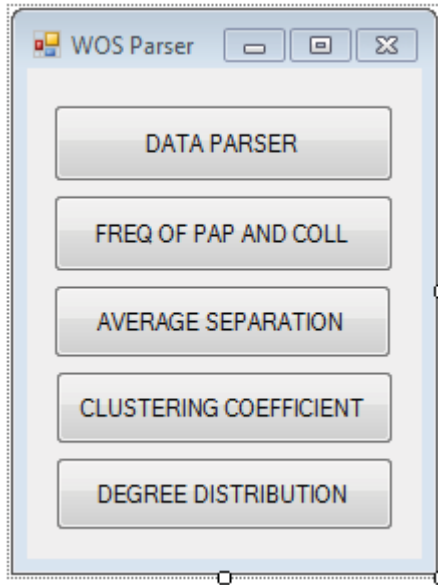
```
CREATE TABLE [dbo].[supernode00] (
    [aID] [int] NOT NULL,
    [aSurname] [nchar](30) NOT NULL,
    [aName] [nchar](30) NOT NULL,
    [d2000] [int] NULL,
    [d2010] [int] NULL,
    [d2012] [int] NULL,
    [ranking] [int] NULL
) ON [PRIMARY]
```

EK AÇIKLAMALAR B.

PROGRAM KODLARI

1. Başlangıç Formu

(Başlangıç formu)



```
public partial class Form2 : Form
{
    public Form2 ()
    {
        InitializeComponent ();
    }

    private void
button1_Click(object sender, EventArgs
e)
    {
        Form f=new Form1 ();
        f.Show ();
    }

    private void
button2_Click(object sender, EventArgs
e)
    {
        Form f = new Form3 ();
        f.Show ();
    }

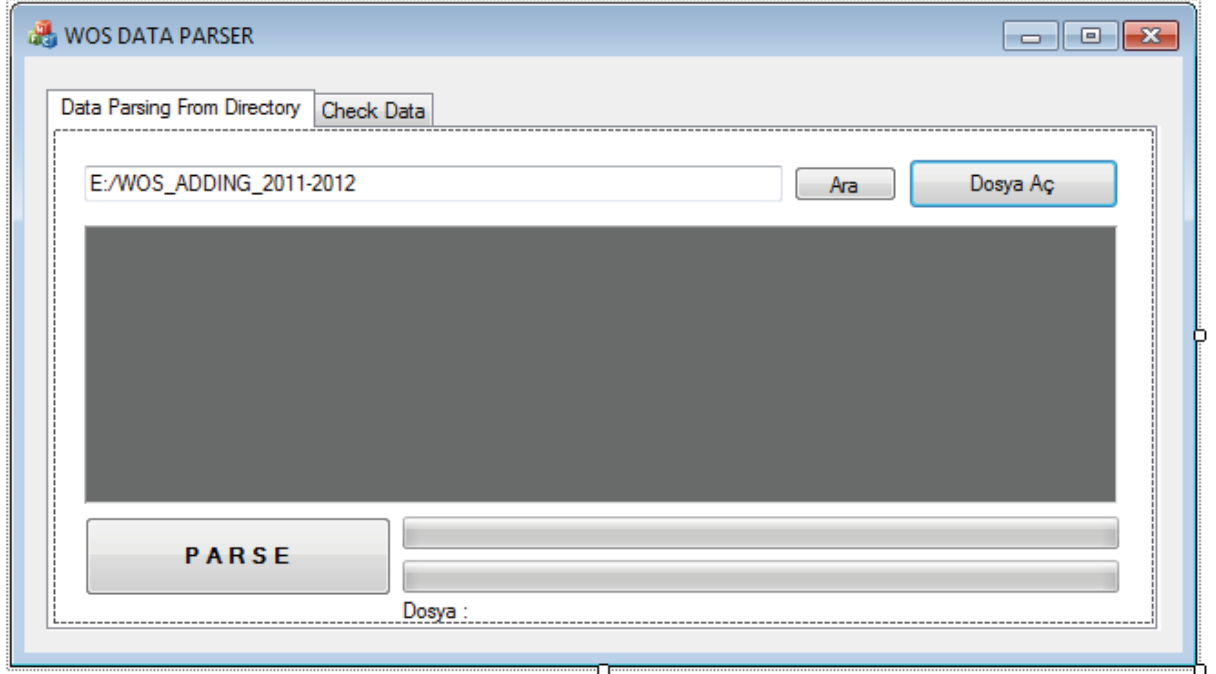
    private void button3_Click(object sender, EventArgs e)
    {
        Form f = new Form4 ();
        f.Show ();
    }

    private void button4_Click(object sender, EventArgs e)
    {
        Form f = new FormCC ();
        f.Show ();
    }

    private void button5_Click(object sender, EventArgs e)
    {
        Form f = new FormDD ();
        f.Show ();
    }
}
```

2. Form1

(Aynı klasördeki HTML formatındaki verileri işleyerek yazar ve yayın bilgilerini ve birlikteliklerini ilgili tablolara aktarmaktadır)



```
public partial class Form1 : Form
{
    string[] parca = new string[501]; //bir html dosyasındaki
500 yayını ayrı ayrı taşıyacak

    string butun,aranan; //bir html dosyasındaki
500 yayının tümü (butun)
    int j = 0, k = 0, i = 0; //k: karakter arama
başlangıç noktası, j bitiş noktası
    int DosyadakiYayınSayisi;

    // yazar ayırma değişkenleri
    kisi[] x = new kisi[500];
    string bitisik;
    string[] adsoyad = new string[2]; // 1 yazara ait ad ve
soyad bilgilerini tutuyor.
    int[] jj = new int[2]; // yazar adının bitişini
kontrol ediyor. "\n" YA DA "</td>
    string Enter = "\n";
    int id, id_; // yazarın
veritabanından gelen aID'si (AID fonk), id_:pID

    // connection değişkenleri
```

```

        string bag =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab";
        string Sorgu, SQLim;

        // dosya deęişkenleri
        string Klasor = "E:/WOS_ADDING_2011-2012";

        public Form1()
        {
            InitializeComponent();
        }

        public void Authorate(string Soyadi, string Adi) //Author
ekleme prosedürü
        {
            SqlConnection bagSQL = new SqlConnection(bag);
            SQLim = "select * from author where aSurname='" + Soyadi
+ "'" and aName='" + Adi + "'" ;
            SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
            SqlDataAdapter sda = new SqlDataAdapter(cmd);
            DataSet ds = new DataSet();

            try
            {
                bagSQL.Open();
                sda.Fill(ds);
                //if (ds.Tables.Count > 0)
                //{
                    if (ds.Tables[0].Rows.Count == 0) //yazar
veritabanında yoksa
                    {
                        Sorgu = "INSERT INTO author(aSurname,aName)
values('" + Soyadi + "', '" + Adi + "')";
                        SqlCommand Komut = new SqlCommand(Sorgu,
bagSQL);

                        Komut.ExecuteNonQuery();
                    }
                //}
            }

            catch (Exception ex)
            {
                MessageBox.Show("Oluşan Hata:" + ex.Message);
            }

            finally
            {
                bagSQL.Close();
            }
        }
}

```

```

        public void Publicate(string pType, string pTitle, string
pSource, string pYear)          //Publication ekleme prosedürü
        {

            SqlConnection bagSQL = new SqlConnection(bag);
            try
            {
                bagSQL.Open();
            }
            catch (Exception ee)
            {
                Console.WriteLine(ee.ToString());
            }

            Sorgu = "INSERT INTO
publication(pType,pTitle,pSource,pYear) values ('" + pType + "','" +
pTitle.Replace("'", "'") + "','" + pSource.Replace("'", "'") + "','"
+ Convert.ToInt32(pYear) + ")";
            SqlCommand Komut = new SqlCommand(Sorgu, bagSQL);

            Komut.ExecuteNonQuery();
            bagSQL.Close(); //
        }

        public void Collaborate(int px, int ax)
//Collaboration ekleme prosedürü
        {
            SqlConnection bagSQL = new SqlConnection(bag);
            try
            {
                bagSQL.Open();
            }
            catch (Exception ee)
            {
                Console.WriteLine(ee.ToString());
            }

            Sorgu = "INSERT INTO collaboration(pID,aID) values (" +
px + "," + ax + ")";
            SqlCommand Komut = new SqlCommand(Sorgu, bagSQL);
            Komut.ExecuteNonQuery();
            bagSQL.Close();

        }

        public int AID(string Soyadi, string Adi)
        {
            SqlConnection bagSQL = new SqlConnection(bag);
            try
            {
                bagSQL.Open();
                SqlDataReader Okuyucu = null;

```

```

        SqlCommand Komutum = new SqlCommand("select aID from
author where aSurname='" + Soyadi + "' and aName='" + Adi + "'"
,bagSQL);
        Okuyucu = Komutum.ExecuteReader();

        while (Okuyucu.Read())
        {
            //a = (Okuyucu["aSurname"].ToString()) ;
            //b = (Okuyucu["aName"].ToString()) ;
            id = Convert.ToInt32(Okuyucu["aID"]);
        }
    }
    catch (Exception zed)
    {
        MessageBox.Show("Oluşan Hata:" + zed.Message);
    }

    finally //
    {
        bagSQL.Close();
    }

return id;
}

public int PID() //son eklenen yayının ID'sini bulur
{
    SqlConnection bagSQL = new SqlConnection(bag);
    try
    {
        bagSQL.Open();
        SqlDataReader Okuyucu = null;
        SqlCommand Komutum = new SqlCommand("select top 1
pID from publication order by pID desc", bagSQL);
        Okuyucu = Komutum.ExecuteReader();

        while (Okuyucu.Read())
        {
            //a = (Okuyucu["aSurname"].ToString()) ;
            //b = (Okuyucu["aName"].ToString()) ;
            id_ = Convert.ToInt32(Okuyucu["pID"]);
        }
    }
    catch (Exception zed)
    {
        MessageBox.Show("Oluşan Hata:" + zed.Message);
    }

    finally
    {
        bagSQL.Close();
    }
}

```

```

        return id_;
    }

private void button5_Click(object sender, EventArgs e)
{
    DosyaOku(textBox5.Text);
    richTextBox1.Text = butun;
}

public struct kisi
{
    public string ad, soyad;
    public int aid;
}

public string Bul(String TamMetin, String Kod)
{
    int k = TamMetin.IndexOf(Kod)+12;
    int j = TamMetin.IndexOf("</td>", k + 1);
    return TamMetin.Substring(k, j - k);
}

public void DosyaOku(string DosyaYolu) //yolu verilen
dosyayı okuyup "butun" değişkenine atar
{
    //veriyolu=textBox5.Text;
    if (File.Exists(DosyaYolu) == true)
    // txt dosyamız şayet varsa
    {

        StreamReader dosyaOku = new StreamReader(DosyaYolu,
Encoding.UTF8);

        // StreamReader oluşturuyoruz ve txt dosyamızın
yolunu belirterek bağlanıyoruz.

        butun = dosyaOku.ReadToEnd();
        const string quote = "\"";
        butun = butun.Replace(" xmlns:set=" + quote +
"http://exslt.org/sets" + quote, "");

        //Sonrasındaa dosyaOku StreamReader'imizi
ReadToEnd() yaparak baştan sona oku ve RichTextBox'ı doldur diyoruz.
        dosyaOku.Close();
    }

    else
        butun = "";
}

public void PARSEr(string DosyaYolu) //VERİ İŞLEME

```



```

{
    DosyaOku(DosyaYolu);    // HTML dosyası OKUNDU

    if (butun == "")
        return;
    //-----HTML'Yİ YAYINLARA BÖLME İŞLEMİ-----
    aranan = "<hr>";
    //butun = richTextBox1.Text;
    k = 0;
    i = 0;
    while (true)    // <HR> leri kullanarak HTML'yi 500
yayına böl : parca[500]
    {
        j = butun.IndexOf(aranan, k + 1);
        if (j == -1) break;
        parca[i] = butun.Substring(k, j-k);
        k = j;
        i++;
    }
    DosyadakiYayinSayisi = i;
    progressBar1.Maximum = DosyadakiYayinSayisi;
    //-----HTML 500'E BÖLÜNDÜ-----
    // yazar ayıklama prosedürü
    for (int y = 0; y < DosyadakiYayinSayisi; y++)
// y:yayın indexini tutuyor (bir dosya için 500 adet)
    {
        progressBar1.Value = y + 1;
        Publishate(Bul(parca[y], "PT <"), Bul(parca[y], "TI
<"), Bul(parca[y], "SO <"), Bul(parca[y], "PY <"));
        // Veritabanına yayını ekliyoruz (Publication tablosuna)
        int PubID = PID();
        aranan = ">AU <";
        k = parca[y].IndexOf(aranan, 1) + 13;
        // k başlangıç noktası
        i = 0;

        while (true)
        {
            jj[0] = parca[y].IndexOf(EnteR, k + 1);
            jj[1] = parca[y].IndexOf("</td>", k + 1);
            j = Math.Min(jj[0], jj[1]);
            bitisik = parca[y].Substring(k, j - k);
            if (bitisik.IndexOf(",") == -1)
                bitisik = bitisik + ", ?";
            adsoyad = bitisik.Split(',');
            x[i].soyad =
adsoyad[0].ToUpperInvariant().Replace("'", "");

            x[i].ad = adsoyad[1].Trim().ToUpperInvariant();
            Authorate(x[i].soyad, x[i].ad);
//veritabanında yoksa yazarı ekle
            x[i].aid=AID(x[i].soyad, x[i].ad);

```

```

//BU YAYINA AİT YAZARLARIN aID'LERİNİ VERİTABANINDAN AL
        Collaborate(PubID, x[i].aid);
        if (jj[0] > jj[1] || i==499) break;
        i++;
        k = j + 8;
        // YAYIN YAZARLARINA AYRILDI, x[].ad, x[].soyad,
x[].aid ŞEKLİNDE ERİŞİLEBİLİYOR
    }
}

private void button9_Click(object sender, EventArgs e)
{
    string uza;
    Klasor = textBox5.Text;
    for (int sayac = 1; sayac <476; sayac++)
    {
        progressBar2.Value = sayac;
        if(sayac<10)
            uza="00" + sayac.ToString();
        else if(sayac<100)
            uza="0" + sayac.ToString();
        else
            uza=sayac.ToString();

        uza = uza + ".html";
        PARSER(Klasor + "/" + uza);
    }
    labell1.Text = "Bitiş saati : " +
DateTime.Now.ToString();
}

public void Sil(int pIDdel, int aIDdel)
{
    SqlConnection bagSQL = new SqlConnection(bag);
    int silinecek=200000;
    try
    {
        bagSQL.Open();
        SqlDataReader Okuyucu = null;
        SqlCommand Komutum = new SqlCommand("select cID from
collaboration where pID=" + pIDdel + " and aID=" + aIDdel, bagSQL);
        Okuyucu = Komutum.ExecuteReader();

        while (Okuyucu.Read())
        {
            //a = (Okuyucu["aSurname"].ToString()) ;
            //b = (Okuyucu["aName"].ToString()) ;
            id = Convert.ToInt32(Okuyucu["cID"]);
            silinecek = id;
        }
    }
}

```

```

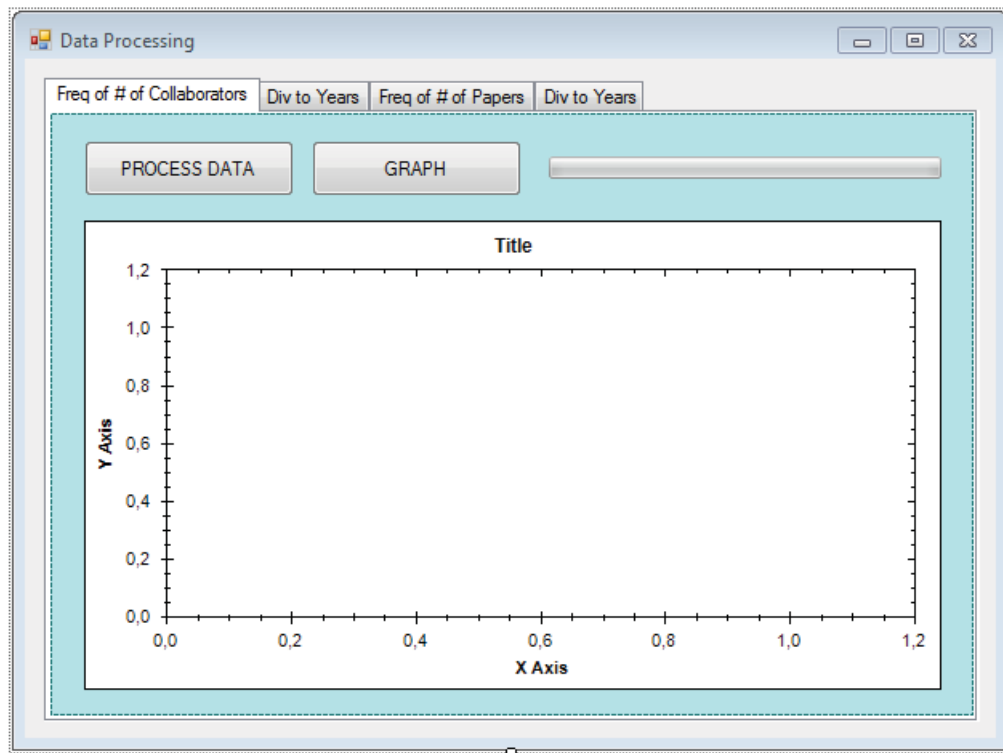
        bagSQL.Close();
        bagSQL.Open();
        Sorgu = "DELETE FROM collaboration WHERE cID=" +
silinecek;

        SqlCommand Komut = new SqlCommand(Sorgu, bagSQL);
        Komut.ExecuteNonQuery();
    }
    catch (Exception zed)
    {
        MessageBox.Show("Oluşan Hata:" + zed.Message);
    }
    finally //
    {
        bagSQL.Close();
    }
}

private void button6_Click(object sender, EventArgs e)
{
    if (folderBrowserDialog1.ShowDialog() ==
DialogResult.OK)
    {
        textBox5.Text = folderBrowserDialog1.SelectedPath;
    }
}
}

```

3. Form 3 (Graphs)



```

public partial class Form3 : Form
{
    string bag =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab";
    string bagR =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab_results";
    string Sorgu, SQLim;
    double[,] veri = new double[50,2];
    int VeriSayac = 0;

    public Form3()
    {
        InitializeComponent();
        //this.Load += new System.EventHandler(this.Form3_Load);
    }

    public void AuthorsPerPaper(int A) //FREQ OF # OF
COLLABORATORS per paper FUNCTION
    {
        int threshold = 10;
        SqlConnection bagSQL = new SqlConnection(bag);
        if (A > threshold)
            SQLim = "SELECT pID, COUNT(aID) AS NumOccurrences
FROM collaboration GROUP BY pID HAVING (COUNT(aID) BETWEEN " + (A-1)
+ " AND " + (A+1) + ")";
        else
            SQLim = "SELECT pID, COUNT(aID) AS NumOccurrences
FROM collaboration GROUP BY pID HAVING (COUNT(aID)=" + A + ")";
        //MessageBox.Show("SELECT pID, COUNT(aID) AS
NumOccurrences FROM collaboration GROUP BY pID HAVING (COUNT(aID)
BETWEEN " + A + " AND " + (A + (Delta - 1)) + ")");
        SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
        SqlDataAdapter sda = new SqlDataAdapter(cmd);
        DataSet ds = new DataSet();

        try
        {
            bagSQL.Open();
            sda.Fill(ds);
            SqlConnection bagSQLR = new SqlConnection(bagR);
            bagSQLR.Open();

            if (A > threshold)
                Sorgu = "INSERT INTO AuthorsPerPaper (Anum,Afreq)
VALUES(" + A + ", " +
Math.Round(Convert.ToDouble(ds.Tables[0].Rows.Count)/3,2).ToString()
.Replace(",", ".")) + ")";
            else
                Sorgu = "INSERT INTO AuthorsPerPaper (Anum,Afreq)
VALUES(" + A + ", " +

```

```

Math.Round(Convert.ToDouble(ds.Tables[0].Rows.Count),
2).ToString().Replace(",", ".") + ")";
        SqlCommand Komut = new SqlCommand(Sorgu, bagSQLR);
        Komut.ExecuteNonQuery();
        bagSQLR.Close();
    }

    catch (Exception ex)
    {
        MessageBox.Show("Oluşan Hata:" + ex.Message);
    }

    finally
    {
        bagSQL.Close();
    }
}

public void AuthorsPerPaper_Y(int A, int Y) /**
{
    //int threshold = 10;
    SqlConnection bagSQL = new SqlConnection(bag);
    //if (A > threshold)
    //    SQLim = "SELECT pID, COUNT(aID) AS NumOccurrences
FROM collab_join_pYear WHERE pYear<=" + Y + " GROUP BY pID HAVING
(COUNT(aID) BETWEEN " + (A - 1) + " AND " + (A + 1) + ")";
    //else
        SQLim = "SELECT pID, COUNT(aID) AS NumOccurrences
FROM collab_join_pYear WHERE pYear<=" + Y + " GROUP BY pID HAVING
(COUNT(aID)=" + A + ")";
        //MessageBox.Show("SELECT pID, COUNT(aID) AS
NumOccurrences FROM collaboration GROUP BY pID HAVING (COUNT(aID)
BETWEEN " + A + " AND " + (A + (Delta - 1)) + ")");
        SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
        SqlDataAdapter sda = new SqlDataAdapter(cmd);
        DataSet ds = new DataSet();

    try
    {
        bagSQL.Open();
        sda.Fill(ds);

        SqlConnection bagSQLR = new SqlConnection(bagR);
        bagSQLR.Open();

        //if (A > threshold)
        //    Sorgu = "INSERT INTO
AuthorsPerPaper_Y(Anum,Afreq,upTo) VALUES(" + A + "," +
Math.Round(Convert.ToDouble(ds.Tables[0].Rows.Count) / 3,
2).ToString().Replace(",", ".") + "," + Y + ")";
        //else

```

```

        Sorgu = "INSERT INTO
AuthorsPerPaper_Y (Anum,Afreq,upTo) VALUES (" + A + "," +
Math.Round(Convert.ToDouble(ds.Tables[0].Rows.Count),
2).ToString().Replace(",", ".")) + "," + Y + ")";
        SqlCommand Komut = new SqlCommand(Sorgu, bagSQLR);
        Komut.ExecuteNonQuery();
        bagSQLR.Close();
    }

    catch (Exception ex)
    {
        MessageBox.Show("Oluşan Hata:" + ex.Message);
    }

    finally
    {
        bagSQL.Close();
    }
}

public void PapersPerAuthor(int A) //FREQ OF # OF PAPERS
per author FUNCTION
{
    int threshold = 10;
    SqlConnection bagSQL = new SqlConnection(bag);
    if (A > threshold)
        SQLim = "SELECT aID, COUNT(pID) AS NumOccurrences
FROM collaboration GROUP BY aID HAVING (COUNT(pID) BETWEEN " + (A-1)
+ " AND " + (A + 1) + ")";
    else
        SQLim = "SELECT aID, COUNT(pID) AS NumOccurrences
FROM collaboration GROUP BY aID HAVING (COUNT(pID)=" + A + ")";

    SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
    SqlDataAdapter sda = new SqlDataAdapter(cmd);
    DataSet ds = new DataSet();

    try
    {
        bagSQL.Open();
        sda.Fill(ds);
        SqlConnection bagSQLR = new SqlConnection(bagR);
        bagSQLR.Open();

        if (A > threshold)
            Sorgu = "INSERT INTO PapersPerAuthor (Pnum,Pfreq)
VALUES (" + A + "," +
Math.Round(Convert.ToDouble(ds.Tables[0].Rows.Count) / 3,
2).ToString().Replace(",", ".")) + ")";
        else
            Sorgu = "INSERT INTO PapersPerAuthor (Pnum,Pfreq)
VALUES (" + A + "," +

```

```

Math.Round(Convert.ToDouble(ds.Tables[0].Rows.Count),
2).ToString().Replace(",", ".") + ")";
        SqlCommand Komut = new SqlCommand(Sorgu, bagSQLR);
        Komut.ExecuteNonQuery();
        bagSQLR.Close();
    }

    catch (Exception ex)
    {
        MessageBox.Show("Oluşan Hata:" + ex.Message);
    }

    finally
    {
        bagSQL.Close();
    }
}

public void PapersPerAuthor_Y(int A, int Y) /**
{
    //int threshold = 10;
    SqlConnection bagSQL = new SqlConnection(bag);
    //if (A > threshold)
    //    SQLim = "SELECT aID, COUNT(pID) AS NumOccurrences
FROM collab_join_pYear WHERE pYear<=" + Y + " GROUP BY aID HAVING
(COUNT(pID) BETWEEN " + (A - 1) + " AND " + (A + 1) + ")";
    //else
        SQLim = "SELECT aID, COUNT(pID) AS NumOccurrences
FROM collab_join_pYear WHERE pYear<=" + Y + " GROUP BY aID HAVING
(COUNT(pID)=" + A + ")";

    SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
    SqlDataAdapter sda = new SqlDataAdapter(cmd);
    DataSet ds = new DataSet();

    try
    {
        bagSQL.Open();
        sda.Fill(ds);
        SqlConnection bagSQLR = new SqlConnection(bagR);
        bagSQLR.Open();
        Sorgu = "INSERT INTO
PapersPerAuthor_Y(Pnum,Pfreq,upTo) VALUES (" + A + "," +
Math.Round(Convert.ToDouble(ds.Tables[0].Rows.Count),
2).ToString().Replace(",", ".") + "," + Y + ")";
        SqlCommand Komut = new SqlCommand(Sorgu, bagSQLR);
        Komut.ExecuteNonQuery();
        bagSQLR.Close();
    }

    catch (Exception ex)

```

```

        {
            MessageBox.Show("Oluşan Hata:" + ex.Message);
        }

        finally
        {
            bagSQL.Close();
        }
    }

    public void VeriAl(string myTable, string Field1, string
Field2)
    {
        SqlConnection bagSQL = new SqlConnection(bagR);
        try
        {
            bagSQL.Open();
            SqlDataReader Okuyucu = null;
            SqlCommand Komutum = new SqlCommand("SELECT * FROM "
+ myTable, bagSQL);
            Okuyucu = Komutum.ExecuteReader();
            VeriSayac = 0;
            while (Okuyucu.Read())
            {
                //a = (Okuyucu["aSurname"].ToString()) ;
                //b = (Okuyucu["aName"].ToString()) ;
                veri[VeriSayac, 0] = (double)(Okuyucu[Field1]);
                veri[VeriSayac, 1] = (double)(Okuyucu[Field2]);
                VeriSayac++;
            }
        }
        catch (Exception zed)
        {
            MessageBox.Show("Oluşan Hata:" + zed.Message);
        }

        finally //
        {
            bagSQL.Close();
        }
    }

    public void VeriAl_Y(string myTable, string Field1, string
Field2, int Y)
    {
        SqlConnection bagSQL = new SqlConnection(bagR);
        try
        {
            bagSQL.Open();
            SqlDataReader Okuyucu = null;
            SqlCommand Komutum = new SqlCommand("SELECT * FROM "
+ myTable + " WHERE upTo=" + Y, bagSQL);

```



```

Okuyucu = Komutum.ExecuteReader();
VeriSayac = 0;
while (Okuyucu.Read())
{
    //a = (Okuyucu["aSurname"].ToString()) ;
    //b = (Okuyucu["aName"].ToString()) ;
    veri[VeriSayac, 0] = (double) (Okuyucu[Field1]);
    veri[VeriSayac, 1] = (double) (Okuyucu[Field2]);
    VeriSayac++;

}
}
catch (Exception zed)
{
    MessageBox.Show("Oluşan Hata:" + zed.Message);
}

finally //
{
    bagSQL.Close();
}
}

private void Form3_Load(object sender, System.EventArgs e)
{
    CreateGraph(zedGraphControll1);
    SetSize();
}

private void button1_Click(object sender, EventArgs e)
//FREQ OF # OF COLLABORATORS TRIGGER
{
    progressBar1.Visible = true;
    progressBar1.Maximum = 13;
    progressBar1.Value = 0;

    int[] s = new
int[14]{1,2,3,4,6,10,16,25,40,63,100,159,251,398};
// eşit logaritmik aralıklar

    for (int i = 0; i < 14; i++)
    { AuthorsPerPaper(s[i]);
    progressBar1.Value = i;
    }
    progressBar1.Visible = false;
}

private void button4_Click(object sender, EventArgs e)
//FREQ OF # OF PAPERS TRIGGER
{
    progressBar2.Visible = true;
    progressBar2.Maximum = 13;

```

```

        progressBar2.Value = 0;
        int[] s = new int[14] { 1, 2, 3, 4, 6, 10, 16, 25, 40,
63, 100, 159, 251, 398 };
        for (int i = 0; i < 14; i++)
        {
            PapersPerAuthor(s[i]);
            progressBar2.Value = i;
        }
        progressBar2.Visible = false;
    }

    private void button3_Click(object sender, EventArgs e)
        // 2. grafik
    {
        VeriAl("PapersPerAuthor", "Pnum", "Pfreq");
        CreateGraph(zedGraphControl2, "Freq of # of Papers per
Author", "# of Papers", "Frequency", "Freq of # of Collaborators per
Paper");
        SetSize();
    }

    private void button2_Click(object sender, EventArgs e)
        // 1. grafik
    {
        VeriAl("AuthorsPerPaper", "Anum", "Afreq");
        CreateGraph(zedGraphControl1, "Freq of # of
Collaborators per Paper", "# of Collaborators", "Frequency", "Freq
of # of Collaborators per Paper");
        SetSize();
    }

    private void Form3_Resize(object sender, EventArgs e)
    {
        SetSize();
    }

    private void SetSize()
    {
        zedGraphControl1.Location = new Point(10, 60);
        zedGraphControl1.Size = new Size(ClientRectangle.Width -
52, ClientRectangle.Height - 115);
        zedGraphControl2.Location = new Point(10, 60);
        zedGraphControl2.Size = new Size(ClientRectangle.Width -
52, ClientRectangle.Height - 115);
        zedGraphControl3.Location = new Point(10, 60);
        zedGraphControl3.Size = new Size(ClientRectangle.Width -
52, ClientRectangle.Height - 115);
        zedGraphControl4.Location = new Point(10, 60);
        zedGraphControl4.Size = new Size(ClientRectangle.Width -
52, ClientRectangle.Height - 115);
    }

```

```

private void CreateGraph(ZedGraphControl zgc, string myTitle,
string Xtitle, string Ytitle, string dotTitle)
{
    GraphPane myPane = zgc.GraphPane;
    myPane.Title.Text = myTitle;
    myPane.XAxis.Title.Text = Xtitle;
    myPane.YAxis.Title.Text = Ytitle;
    myPane.XAxis.Type = AxisType.Log;
    myPane.YAxis.Type = AxisType.Log;

    PointPairList list1 = new PointPairList();
    for (int i = 0; i < VeriSayac; i++)
    {
        list1.Add(veri[i,0], veri[i,1]);
    }

    LineItem myCurve = myPane.AddCurve(dotTitle, list1,
Color.Black, SymbolType.TriangleDown);
    zgc.AxisChange();
}

private void button5_Click(object sender, EventArgs e)
{
    for(int Yr=1990;Yr<=2010;Yr=Yr+10)
    {
        VeriAl_Y("AuthorsPerPaper_Y", "Anum", "Afreq",Yr);
        CreateGraph(zedGraphControl3, "Freq of # of
Collaborators per Paper", "# of Collaborators", "Frequency", "Up To"
+ Yr);
    }
    SetSize();
}

private void button6_Click(object sender, EventArgs e) //
{
    progressBar3.Visible = true;
    progressBar3.Maximum = 300 ;
    progressBar3.Value = 0;

    for (int Yr = 1990; Yr <= 2010; Yr = Yr + 10)
    {
        for (int i = 1; i <= 300; i++)
        {
            AuthorsPerPaper_Y(i, Yr);
            progressBar3.Value = i;
        }
    }
    progressBar3.Visible = false;
}

private void button7_Click(object sender, EventArgs e)
{

```

```

        for (int Yr = 1990; Yr <= 2010; Yr = Yr + 10)
        {
            VeriAl_Y("PapersPerAuthor_Y", "Pnum", "Pfreq", Yr);
            CreateGraph(zedGraphControl4, "Freq of # of Papers
per Author", "# of Papers", "Frequency", "Up To " + Yr);
        }
        SetSize();
    }
    private void button8_Click(object sender, EventArgs e)
    {
        progressBar4.Visible = true;
        progressBar4.Maximum = 300;
        progressBar4.Value = 0;
        //int[] s = new int[14] { 1, 2, 3, 4, 6, 10, 16, 25, 40,
63, 100, 159, 251, 398 }; // sample points for log binning
        for (int Yr = 1990; Yr <= 2010; Yr = Yr + 10)
        {
            for (int i = 1; i <= 300; i++)
            {
                PapersPerAuthor_Y(i, Yr);
                progressBar4.Value = i;
            }
        }
        progressBar4.Visible = false;
    }
}
}

```

4. Form4 (Average Separation)

```

public partial class Form4 : Form
{
    string bag =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab";
    string bagR =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab_results";
}

```

```

string Sorgu, SQLim, SQLim2, SQLimY;
static int YAZARSAYI = 185886;
static int PERIMETER = 12;
static int BRANCH = 100000;
static int REPEAT = 3000;// her yıl dilimi için sim. tekrarı

//double[,] veri = new double[50, 2];

int y1, y2; // yazar node 1,2
int i = 0, j = 0, jj = 0;
bool limit = false;
int cMax = 0; // her yıl için max cID'yi tutacak

bool zed = true; // GEÇİCİ TETİKÇİ OLARAK KULLANILABİLİR

DataTable DT1;
DataTable DT2;

DataTable DT1next;
DataTable DT2next;

DataTable DT1full;
DataTable DT2full;

public Form4()
{
    InitializeComponent();
}

private int DistanceDT(int p, int q, int year)
{
    int kontrol = 0;
    SqlConnection bagSQL = new SqlConnection(bag);
    SqlConnection bagSQL2 = new SqlConnection(bag);

    progressBar1.Maximum = PERIMETER;
    progressBar2.Maximum = BRANCH;
    progressBar3.Maximum = BRANCH;

    if ((p > 0) && (q > 0))
    {
        y1 = p;
        y2 = q;
    }
    else
        YAZARSEC(year);

    DT_DEFINE();

    DT_CLEAR();
    DTnext_CLEAR();
}

```

```

DTfull_CLEAR();

DataRow DR1;
DR1 = DT1.NewRow();
DR1["aID"] = y1;
DT1.Rows.Add(DR1);

DataRow DR2;
DR2 = DT2.NewRow();
DR2["aID"] = y2;
DT2.Rows.Add(DR2);

DT1full.Merge(DT1);
DT2full.Merge(DT2);

for (j = 0; j < PERIMETER; j++)
{
    limit = false;
    progressBar1.Value = j+1;
    int aralik = 0;

    for (jj = 0; jj < DT1.Rows.Count; jj++)
    {
        DataTable DT1temp = new DataTable();
        SQLim = "SELECT DISTINCT aID from collaboration
WHERE (pID IN (SELECT pID FROM collaboration where aID=" +
Convert.ToInt32(DT1.Rows[jj][0]) + ") AND aID<>" +
Convert.ToInt32(DT1.Rows[jj][0]) + " AND cID<=" + cMax + ")";
        SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
        SqlDataAdapter sda = new SqlDataAdapter(cmd);
        //DataSet ds = new DataSet();
        bagSQL.Open();
        sda.Fill(DT1temp);
        bagSQL.Close();

        DT1next.Merge(DT1temp);

        //DataView dv = new DataView(DT1next); //
TEKRARLI satırları silmek için
        //string[] strColumns = { "aID" };
        //DT1next = dv.ToTable(true, strColumns);
// true: tekrarları sil

        int l=
Convert.ToInt32(Math.Round(Convert.ToDecimal(Convert.ToInt32(DT1next
.Rows.Count) / 5000),0));
        if ( l > aralik )
        {
            kontrol=Include1();
            if (kontrol > 0) return kontrol;
            aralik+=(l-aralik);
        }
    }
}

```

```

        if (DT1next.Rows.Count >= BRANCH)
        {
            progressBar2.Value = BRANCH;
            limit = true;
            break;
        }
        else progressBar2.Value = DT1next.Rows.Count;

        if ((limit == true || (DT2.Rows.Count < 2 * (j +
1) && DT1next.Rows.Count > BRANCH / 2))) return 0;
    }

    DT1next = Difference(DT1next, DT1full);

    if (DT1next.Rows.Count < (j + 1)) return 0;
    // kısır node varsa çıkış
    if ((DT1next.Rows.Count < (2 * (j + 1)) && j >= (2*
PERIMETER / 3))) return 0; // sonu baştan belliyse çıkış

    // next >> full'e kopyalanıyor////////////////////////////////////
    DT1full.Merge(DT1next);
    //DataView dv_i = new DataView(DT1full); //
TEKRARLI satırları silmek için
    //string[] strColumns_i = { "aID" };
    //DT1full = dv_i.ToTable(true, strColumns_i);
// true: tekrarları sil

    kontrol = Includelend();

    if (kontrol > 0) return kontrol;
    //varsa distance değerini döndürüyor

    ////////////////////////////////////// 2. liste

    limit = false;
    aralik = 0;

    for (jj = 0; jj < DT2.Rows.Count; jj++)
    {

        DataTable DT2temp = new DataTable();
        SQLim2 = "SELECT DISTINCT aID from collaboration
WHERE (pID IN (SELECT pID FROM collaboration where aID=" +
Convert.ToInt32(DT2.Rows[jj][0]) + ") AND aID<>" +
Convert.ToInt32(DT2.Rows[jj][0]) + " AND cID<=" + cMax + ")";
        SqlCommand cmd2 = new SqlCommand(SQLim2,
bagSQL2);

        SqlDataAdapter sda2 = new SqlDataAdapter(cmd2);
        bagSQL2.Open();
        sda2.Fill(DT2temp);
        bagSQL2.Close();

```

```

DT2next.Merge (DT2temp);

int l =
Convert.ToInt32 (Math.Round (Convert.ToDecimal (Convert.ToInt32 (DT2next
.Rows.Count) / 5000), 0));
if (l > aralik)
{
//MessageBox.Show ("/");
kontrol = Include2 ();
if (kontrol > 0) return kontrol;
aralik += (l - aralik);
}

if (DT2next.Rows.Count >= BRANCH)
{
progressBar3.Value = BRANCH;
limit = true;
break;
}
else progressBar3.Value = DT2next.Rows.Count;

if ((limit == true || (DT1next.Rows.Count < 2 *
(j + 1) && DT2next.Rows.Count > BRANCH / 2))) return 0; // ||
(DT1next.Rows.Count < 2*(j+1) && DT2next.Rows.Count > BRANCH / 2)
}

DT2next = Difference (DT2next, DT2full);

if (DT2next.Rows.Count < (j + 1)) return 0; //
kısır node varsa çıkış
if ((DT2next.Rows.Count < (2 * (j + 1)) && j >=
(2*PERIMETER / 3))) return 0; // sonu baştan belliyse çıkış

// next >> full'e kopyalanıyor////////////////////////////////////
DT2full.Merge (DT2next);
//DataView dv_ii = new DataView (DT2full); //
TEKRARLI satırları silmek için
//string[] strColumns_ii = { "aID" };
//DT2full = dv_ii.ToTable (true, strColumns_ii);
// true: tekrarları sil

kontrol = Include2end ();

if (kontrol > 0) return kontrol; //varsa
distance değerini döndürüyor

DT_CLEAR ();
DT1 = DT1next.Copy ();
DT2 = DT2next.Copy ();
DTnext_CLEAR ();

```



```

    }
    return 0;          // hiç eşleşme yoksa 0 döndürüyor
}

private void YAZARSEC()
{
    Random Rastgele = new Random();
    y1 = Rastgele.Next(1, YAZARSAYI); //rasgele ID üretir
(YAZAR)
    y2 = Rastgele.Next(1, YAZARSAYI); //rasgele ID üretir
(YAZAR)
}

private void YAZARSEC(int year)
{
    SqlConnection bagSQLY = new SqlConnection(bag);
    SQLimY = "SELECT MAX(aID) AS Enbuyuk FROM collaboration
WHERE (pID <= (SELECT MAX(pID) FROM publication where pYear<=" +
year + "))";
    SqlCommand cmdY = new SqlCommand(SQLimY, bagSQLY);
    SqlDataAdapter sdaY = new SqlDataAdapter(cmdY);
    DataSet dsY = new DataSet();
    bagSQLY.Open();
    sdaY.Fill(dsY);
    bagSQLY.Close();
    sdaY.Dispose();
    dsY.Dispose();
    cmdY.Dispose();
    bagSQLY.Dispose();

    object Oo = Convert.ToInt32(dsY.Tables[0].Rows[0][0]);
    int lim = Convert.ToInt32(Oo);

    Random Rastgele = new Random();
    y1 = Rastgele.Next(1, lim); // 1-lim aralığında rasgele
sayı üretir
    y2 = Rastgele.Next(1, lim); // 1-lim aralığında rasgele
sayı üretir
}

private int Includel()
{
    for (int w = 0; w < DT1next.Rows.Count; w++)
        for (int h = 0; h < DT2full.Rows.Count; h++)

if(Convert.ToInt32(DT1next.Rows[w][0])==Convert.ToInt32(DT2full.Rows
[h][0]))
    {
        return 2 * j + 1;
    }
}

```

```

        return 0;
    }

private int Include2()
{
    for (int w = 0; w < DT1full.Rows.Count; w++)
        for (int h = 0; h < DT2next.Rows.Count; h++)
            if (Convert.ToInt32(DT2next.Rows[h][0]) ==
Convert.ToInt32(DT1full.Rows[w][0]))
                {
                    return 2 * j + 2;
                }

    return 0;
}

private int Includelend()
{
    for (int w = 0; w < DT1full.Rows.Count; w++)
        for (int h = 0; h < DT2full.Rows.Count; h++)
            if (Convert.ToInt32(DT1full.Rows[w][0]) ==
Convert.ToInt32(DT2full.Rows[h][0]))
                {
                    return 2 * j + 1;
                }

    return 0;
}

private int Include2end()
{
    for (int w = 0; w < DT1full.Rows.Count; w++)
        for (int h = 0; h < DT2full.Rows.Count; h++)
            if (Convert.ToInt32(DT2full.Rows[h][0]) ==
Convert.ToInt32(DT1full.Rows[w][0]))
                {
                    return 2 * j + 2;
                }

    return 0;
}

private void DT_CLEAR()
{
    DT1.Clear();
    DT2.Clear();
}

private void DTnext_CLEAR()
{
    DT1next.Clear();
    DT2next.Clear();
}

```

```

    }

private void DTfull_CLEAR()
{
    DT1full.Clear();
    DT2full.Clear();
}

private void DT_DEFINE()
{
    DT1= new DataTable("Tablo" + i);
    DT2= new DataTable("Tablo" + i);
    DT1.Columns.Add("aID", typeof(int));
    DT2.Columns.Add("aID", typeof(int));

    DT1next = new DataTable("TabloN" + i);
    DT2next = new DataTable("TabloN" + i);
    DT1next.Columns.Add("aID", typeof(int));
    DT2next.Columns.Add("aID", typeof(int));

    DT1full = new DataTable("TabloF" + i);
    DT2full = new DataTable("TabloF" + i);
    DT1full.Columns.Add("aID", typeof(int));
    DT2full.Columns.Add("aID", typeof(int));

}

private void button1_Click(object sender, EventArgs e)
{
    int sonuc = DistanceDT(0,0,2010);
    MessageBox.Show(y1.ToString() + " | " + y2.ToString() +
" | " + sonuc.ToString());
}

private void button2_Click(object sender, EventArgs e)
{
    int sonuc = DistanceDT(Convert.ToInt32(textBox1.Text),
Convert.ToInt32(textBox2.Text), 2010);
    MessageBox.Show(y1.ToString() + " | " + y2.ToString() +
" | " + sonuc.ToString());
}

private void button3_Click(object sender, EventArgs e)
// process over years //
{
    int sonuc = 0;
    progressBar5.Maximum = REPEAT;

    for (int year = 2012; year >= 1986; year+=3)
    {
        progressBar4.Value = year;
    }
}

```

```

// her yıl için max collaboration ID'yi 1 defa
hesaplamak için ///////////////////////////////////

        SqlConnection bagSQLY = new SqlConnection(bag);
        SQLimY = "SELECT MAX(cID) AS Enbuyuk FROM
collaboration WHERE (pID = (SELECT MAX(pID) FROM publication where
pYear<=" + year + "))";
        SqlCommand cmdY = new SqlCommand(SQLimY, bagSQLY);
        SqlDataAdapter sdaY = new SqlDataAdapter(cmdY);
        DataSet dsY = new DataSet();
        bagSQLY.Open();
        sdaY.Fill(dsY);
        bagSQLY.Close();

        object Oo =
Convert.ToInt32(dsY.Tables[0].Rows[0][0]);
        cMax = Convert.ToInt32(Oo);

        for (int t = 0; t < REPEAT; t++)
        {
            if (zed == true) //!!!!!!!!!!!!!!!!!!!!!!
            {
                t = 0; // ARADAN BAŞLAMAK İÇİN ÖRN 82
                zed = false;
            }
            progressBar5.Value = t;
            sonuc = DistanceDT(0,0,year);
            SqlConnection bagSQLR = new SqlConnection(bagR);
            bagSQLR.Open();
            Sorgu = "INSERT INTO
distances(dYear,distance,y1,y2) VALUES(" + year + "," + sonuc + "," +
+ y1 + "," + y2 + ")";
            SqlCommand Komut = new SqlCommand(Sorgu,
bagSQLR);
            Komut.ExecuteNonQuery();
            bagSQLR.Close();
        }
    }

private void button4_Click(object sender, EventArgs e)
// ONCE - DT
{
    progressBar4.Value = 1983; // değiştir

    // her yıl için max collaboration ID'yi 1 defa
hesaplamak için

        SqlConnection bagSQLY = new SqlConnection(bag);

```

```

        SQLimY = "SELECT MAX(cID) AS Enbuyuk FROM collaboration
WHERE (pID = (SELECT MAX(pID) FROM publication where pYear<=" + 1983
+ "))";

        SqlCommand cmdY = new SqlCommand(SQLimY, bagSQLY);
        SqlDataAdapter sdaY = new SqlDataAdapter(cmdY);
        DataSet dsY = new DataSet();
        bagSQLY.Open();
        sdaY.Fill(dsY);
        bagSQLY.Close();

        object Oo = Convert.ToInt32(dsY.Tables[0].Rows[0][0]);
        cMax = Convert.ToInt32(Oo);

        int sonuc = DistanceDT(Convert.ToInt32(textBox1.Text),
Convert.ToInt32(textBox2.Text), 1983);
        MessageBox.Show(y1.ToString() + " | " + y2.ToString() +
" | " + sonuc.ToString());
    }

    public static DataTable Difference(DataTable First,
DataTable Second)
    {
        // Empty Table
        DataTable table = new DataTable("Difference");

        //Must use a Dataset to make use of a DataRelation
object
        using (DataSet ds = new DataSet())
        {
            //Add tables
            ds.Tables.AddRange(new DataTable[] { First.Copy(),
Second.Copy() });
            //Get Columns for DataRelation
            DataColumn[] firstcolumns = new
DataColumn[ds.Tables[0].Columns.Count];

            for (int i = 0; i < firstcolumns.Length; i++)
            {
                firstcolumns[i] = ds.Tables[0].Columns[i];
            }

            DataColumn[] secondcolumns = new
DataColumn[ds.Tables[1].Columns.Count];

            for (int i = 0; i < secondcolumns.Length; i++)
            {
                secondcolumns[i] = ds.Tables[1].Columns[i];
            }

            // DataRelation
            DataRelation r = new DataRelation(string.Empty,
firstcolumns, secondcolumns, false);

```

```

        ds.Relations.Add(r);

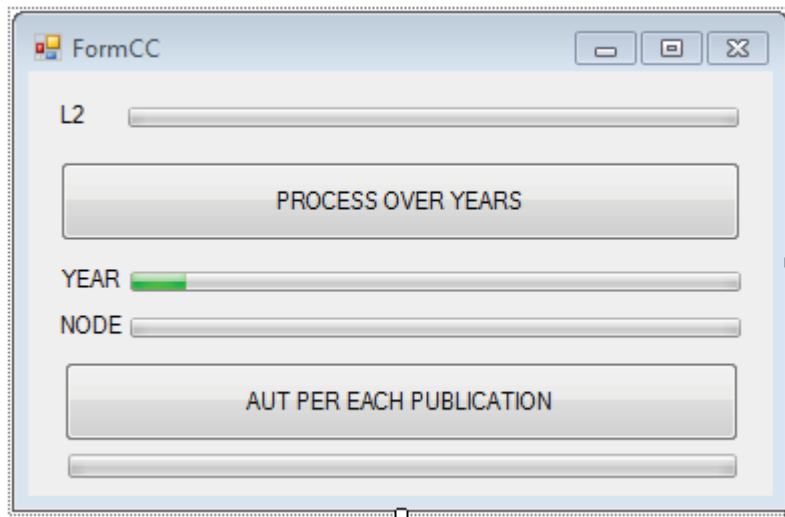
        //Define columns for return table
        for (int i = 0; i < First.Columns.Count; i++)
        {
            table.Columns.Add(First.Columns[i].ColumnName,
First.Columns[i].DataType);
        }

        //If First Row not in Second, Add to return table.
        table.BeginLoadData();

        foreach (DataRow parentrow in ds.Tables[0].Rows)
        {
            DataRow[] childrows = parentrow.GetChildRows(r);
            if (childrows == null || childrows.Length == 0)
                table.LoadDataRow(parentrow.ItemArray,
true);
        }
        table.EndLoadData();
    }
    return table;
}
}
}

```

5. FormCC (Clustering Coefficient)



```

public partial class FormCC : Form
{
    string bag =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab";
    string bagR =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab_results";
    string Sorgu, SQLim, SQLim2, SQLimY;
}

```

```

int jj = 0;
int aMax = 0, cMax=0; // her yıl için max aID, cID'yi
tutacak
bool zed = true; // GEÇİCİ TETİKÇİ
public FormCC()
{
    InitializeComponent();
}

private Single ClusteringCf(int y, int year)
{
    Single Ni = 0;
    SqlConnection bagSQL = new SqlConnection(bag);
    //SqlConnection bagSQL2 = new SqlConnection(bag);
    DataTable DT1temp = new DataTable();
    SQLim = "SELECT DISTINCT aID from collaboration WHERE
(pID IN (SELECT pID FROM collaboration where aID=" + y + ") AND
aID<>" + y + " AND cID<=" + cMax + ")";
    SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
    SqlDataAdapter sda = new SqlDataAdapter(cmd);
    bagSQL.Open();
    sda.Fill(DT1temp);
    bagSQL.Close();
    Single ki=DT1temp.Rows.Count;
    if (ki > 200) return 0;
//----- !
    progressBar3.Maximum =Convert.ToInt32(ki);

    for (jj = 0; jj < DT1temp.Rows.Count; jj++)
    {
        progressBar3.Value = jj + 1;
        DataTable DT1temp2 = new DataTable();
        SQLim2 = "SELECT DISTINCT aID from collaboration
WHERE (pID IN (SELECT pID FROM collaboration where aID=" +
Convert.ToInt32(DT1temp.Rows[jj][0]) + ") AND aID<>" +
Convert.ToInt32(DT1temp.Rows[jj][0]) + " AND cID<=" + cMax + ")";
        SqlCommand cmd2 = new SqlCommand(SQLim2 + "
INTERSECT " + SQLim, bagSQL);
        cmd2.CommandTimeout = 1000;
        SqlDataAdapter sda2 = new SqlDataAdapter(cmd2);
        bagSQL.Open();
        sda2.Fill(DT1temp2);
        bagSQL.Close();

        Ni += DT1temp2.Rows.Count;
    }

    Single f=(Ni / (ki * (ki - 1)));
    //MessageBox.Show(f.ToString() + " " +
Ni.ToString() + " " + ki.ToString());

```

```

        if (ki > 1)
            return Ni / (ki * (ki - 1));
        else
            return 0;
    }

    private void AuthorsPerEachPaper(int p)
    {
        SqlConnection bagSQL = new SqlConnection(bag);
        //SqlConnection bagSQL2 = new SqlConnection(bag);
        bagSQL.Open();
        DataTable DT1temp = new DataTable();
        SQLim = "SELECT DISTINCT aID from collaboration WHERE
        pID=" + p + "";
        SqlCommand cmd = new SqlCommand(SQLim, bagSQL);
        SqlDataAdapter sda = new SqlDataAdapter(cmd);
        sda.Fill(DT1temp);
        int ki = DT1temp.Rows.Count;
        Sorgu = "UPDATE publication SET numAut=" + ki + ",
        numLink=" + (ki*(ki-1)/2) + " WHERE pID=" + p + "";
        SqlCommand Komut = new SqlCommand(Sorgu, bagSQL);
        Komut.ExecuteNonQuery();
        bagSQL.Close();
        DT1temp.Dispose();
    }

    private void button3_Click(object sender, EventArgs e)
    {
        Single sonuc = 0; // burada CC için kullanılacak
        for (int year = 2012; year < 2013; year += 3)
        {
            progressBar4.Value = year;

            SqlConnection bagSQLY = new SqlConnection(bag);
            SQLimY = "SELECT MAX(aID) AS aMax, MAX(cID) AS cMax
            FROM collaboration WHERE (pID <= (SELECT MAX(pID) FROM publication
            where pYear<=" + year + "))";
            SqlCommand cmdY = new SqlCommand(SQLimY, bagSQLY);
            SqlDataAdapter sdaY = new SqlDataAdapter(cmdY);
            DataTable dtY = new DataTable();
            bagSQLY.Open();
            sdaY.Fill(dtY);
            bagSQLY.Close();

            object Oo = Convert.ToInt32(dtY.Rows[0][0]);
            aMax = Convert.ToInt32(Oo);
            int increment = Convert.ToInt32(aMax/1000);
            // 1000 samples
            if (year >= 2010) increment *= 2; // GEREKTİĞİNDE

            object oO = Convert.ToInt32(dtY.Rows[0][1]);
            cMax = Convert.ToInt32(oO);
        }
    }

```



```

        progressBar5.Maximum = aMax;
        for (int t = increment; t < aMax; t+=increment)
            // farklı noktadan örnek almak için +99! !!! ama zed ile
            birlikte kullanılırsa zed'deki geçerli
            {
                if (zed == true)          //!!!!!!!!!!!!!!!!!!!!!!
                {
                    t = 149;           // ARADAN BAŞLAMAK İÇİN
                    zed = false;
                }

                progressBar5.Value = t+1;
                sonuc = ClusteringCf((t+1), year); // t+1 ID'li
yazar

                SqlConnection bagSQLR = new SqlConnection(bagR);
                bagSQLR.Open();
                Sorgu = "INSERT INTO clusteringCf(aID,year,cc)
VALUES (@A, @B, @C)";
                SqlCommand Komut = new SqlCommand(Sorgu,
bagSQLR);

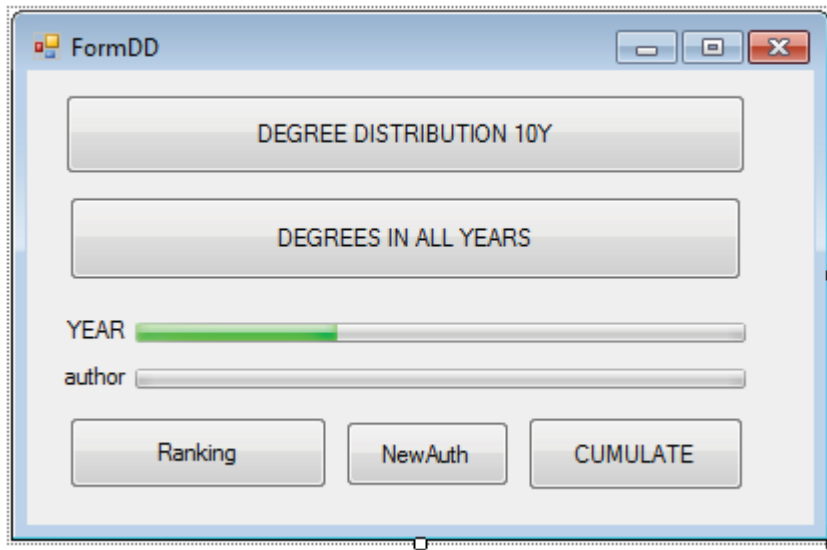
                SqlParameter param1 = new SqlParameter("@A",
System.Data.SqlDbType.Int);
                param1.Value = (t+1);
                Komut.Parameters.Add(param1);
                SqlParameter param2 = new SqlParameter("@B",
System.Data.SqlDbType.SmallInt);
                param2.Value = year;
                Komut.Parameters.Add(param2);
                SqlParameter param3 = new SqlParameter("@C",
System.Data.SqlDbType.Float);
                param3.Value = Math.Round(sonuc, 4);
                Komut.Parameters.Add(param3);

                Komut.ExecuteNonQuery();
                bagSQLR.Close();
            }
        }

        private void button1_Click(object sender, EventArgs e)
        //Her yayının kaç yazarlı olduğu (buradan link sayısı hesaplanacak)
        {
            int maxP=292895;
            progressBar1.Maximum=maxP;
            for (int j = 1; j <= maxP; j++)
            {
                progressBar1.Value = j;
                AuthorsPerEachPaper(j);
            }
        }
    }
}

```

6. FormDD (Degree Distribution)



```
public partial class FormDD : Form
{
    string bag =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab";
    string bagR =
"server=(local)\\SQLEXPRESS;uid=sa;pwd=11;database=collab_results";
    string Sorgu, SQL1;

    public FormDD()
    {
        InitializeComponent();
    }

    private void button3_Click(object sender, EventArgs e) //
Degree Dist over Years
    {
        int[] aIDmax=new int[] {7849,42599,151745};
        for (int Year = 1990; Year < 2011; Year += 10)
        {
            progressBar4.Value = Year;
            int amx=aIDmax[(Year-1990)/10];
            progressBar5.Maximum=amx;
            for (int a = 1; a <= amx; a++)
            {
                progressBar5.Value = a;
                Degree(a, Year);
            }
        }
    }

    public void DegreeFreq(int D, int Year) /**
    {
```

```

        SqlConnection bagSQL = new SqlConnection(bagR);
        SQL1 = "SELECT COUNT(*) FROM DegreeDistribution WHERE
aDeg=" + D + " AND upTo=" + Year;
        SqlCommand cmd = new SqlCommand(SQL1, bagSQL);
        SqlDataAdapter sda = new SqlDataAdapter(cmd);
        DataTable dtD = new DataTable();

        try
        {
            bagSQL.Open();
            sda.Fill(dtD);

            SqlConnection bagSQLR = new SqlConnection(bagR);
            bagSQLR.Open();

            Sorgu = "INSERT INTO
DegreeDistributionFreq(Deg,Freq,Year) VALUES(" + D + "," +
Convert.ToInt32(dtD.Rows[0][0]) + "," + Year + ")";
            SqlCommand Komut = new SqlCommand(Sorgu, bagSQLR);
            Komut.ExecuteNonQuery();
            bagSQLR.Close();

        }

        catch (Exception ex)
        {
            MessageBox.Show("Oluşan Hata:" + ex.Message);
        }

        finally
        {
            bagSQL.Close();
        }
    }

    public void Degree(int aID, int Y) /**
    {
        SqlConnection bagSQL = new SqlConnection(bag);
        SQL1 = "SELECT COUNT(aID)-1 from (SELECT DISTINCT aID
FROM collab_join_pYear WHERE (pID IN (SELECT pID from
collab_join_pYear WHERE aID=" + aID + ")) AND pYear<=" + Y + ") AS
TABLO";

        SqlCommand cmd = new SqlCommand(SQL1, bagSQL);
        SqlDataAdapter sda = new SqlDataAdapter(cmd);
        DataTable dtD = new DataTable();

        try
        {
            bagSQL.Open();
            sda.Fill(dtD);
            SqlConnection bagSQLR = new SqlConnection(bagR);
            bagSQLR.Open();

```

```

        Sorgu = "INSERT INTO
DegreeDistribution(aID,aDeg,upTo) VALUES(" + aID + "," +
Convert.ToInt32(dtD.Rows[0][0]) + "," + Y + ")";
        SqlCommand Komut = new SqlCommand(Sorgu, bagSQLR);
        Komut.ExecuteNonQuery();
        bagSQLR.Close();
    }

    catch (Exception ex)
    {
        MessageBox.Show("Oluşan Hata:" + ex.Message);
    }

    finally
    {
        bagSQL.Close();
    }
}

public void DegreeAllYears(int aID, int Y) /**
{
    SqlConnection bagSQL = new SqlConnection(bag);
    SQL1 = "SELECT COUNT(aID)-1 from (SELECT DISTINCT aID
FROM collab_join_pYear WHERE (pID IN (SELECT pID from
collab_join_pYear WHERE aID=" + aID + ")) AND pYear<=" + Y + ") AS
TABLO";

    SqlCommand cmd = new SqlCommand(SQL1, bagSQL);
    SqlDataAdapter sda = new SqlDataAdapter(cmd);
    DataTable dtD = new DataTable();

    try
    {
        bagSQL.Open();
        sda.Fill(dtD);
        SqlConnection bagSQLR = new SqlConnection(bagR);
        bagSQLR.Open();
        Sorgu = "UPDATE authorD SET d" + Y + "=" +
Convert.ToInt32(dtD.Rows[0][0]) + " WHERE aID=" + aID + """;
        SqlCommand Komut = new SqlCommand(Sorgu, bagSQLR);
        Komut.ExecuteNonQuery();
        bagSQLR.Close();
    }

    catch (Exception ex)
    {
        MessageBox.Show("Oluşan Hata:" + ex.Message);
    }

    finally
    {
        bagSQL.Close();
    }
}

```

```

    }
}

private void button1_Click(object sender, EventArgs e)
// cumulative data for generating the graphs
{
    progressBar5.Maximum = 1000;
    for (int Year = 1990; Year < 2011; Year += 10)
    {
        progressBar4.Value = Year;

        for (int a = 1; a <= 1000; a++)
        {
            progressBar5.Value = a;
            DegreeFreq(a, Year);
        }
    }
}

private void button2_Click(object sender, EventArgs e)
{
    int aIDmax = 185886;
    progressBar4.Maximum = 2012;
    progressBar5.Maximum = aIDmax;
    for (int a = 1; a <= aIDmax; a++)
    {
        progressBar5.Value = a;
        int FYear = 1980;
        FYear = FirstYearOfAuthor(a);
        for (int Year = FYear; Year <= 2012; Year++)
        {
            progressBar4.Value = Year;
            DegreeAllYears(a, Year);
        }
    }
}

private int FirstYearOfAuthor(int aID)
{
    SqlConnection bagSQLz = new SqlConnection(bag);

    SQL1 = "SELECT min(pYear) as Yil FROM collab_join_pYear
WHERE aID=" + aID + ";
    SqlCommand cmd = new SqlCommand(SQL1, bagSQLz);
    SqlDataAdapter sda = new SqlDataAdapter(cmd);
    DataTable dtD = new DataTable();
    bagSQLz.Open();
    sda.Fill(dtD);
    return Convert.ToInt32(dtD.Rows[0][0]);
}

private void button4_Click(object sender, EventArgs e)

```

```

    {
        progressBar5.Maximum = 185886;
        progressBar5.Value = 0;

        SqlConnection bagSQL = new SqlConnection(bagR);
        SQL1 = "SELECT aID from authorPlus ORDER BY d2012 DESC,
aSurname ASC";
        SqlCommand cmd = new SqlCommand(SQL1, bagSQL);
        SqlDataAdapter sda = new SqlDataAdapter(cmd);
        DataTable dtD = new DataTable();
        int[] A=new int[185900];

        try
        {
            bagSQL.Open();
            sda.Fill(dtD);

            for (int tt = 1; tt <= 185886; tt++)
            {
                A[tt] = Convert.ToInt32(dtD.Rows[tt-1][0]);
                progressBar5.Value = tt;
            }

            bagSQL.Close();

            SqlConnection bagSQLR = new SqlConnection(bagR);
            bagSQLR.Open();

            for (int jey = 1; jey <= 185886; jey++)
            {
                progressBar5.Value = jey;
                Sorgu = "UPDATE authorPlus SET ranking=" + jey +
" WHERE aID=" + A[jey] + "";
                SqlCommand Komut = new SqlCommand(Sorgu,
bagSQLR);

                Komut.ExecuteNonQuery();
            }
            bagSQLR.Close();
        }
        catch (Exception ex)
        {
            MessageBox.Show("Oluşan Hata:" + ex.Message);
        }
    }
}

```

7. ConsoleApplication

(Aynı klasörde bulunan rastgele isimlendirilmiş dosyaları belli bir kurala göre isimlendirip başka klasöre taşıyan program. Yukarıdaki programlar çalıştırılmadan önce bu isimlendirme prosedürü uygulandı)

```
namespace FileRename
{
    class Class1
    {
        [STAThread]
        static void Main(string[] args)
        {
            try
            {
                DirectoryInfo di = new
DirectoryInfo(@"E:\WOS_ADD_2011-2012");

                FileInfo[] fiArr = di.GetFiles("*.html");
                // Sayaç
                int i = 0;
                string path;

                // Dosya isimlerini al
                foreach (FileInfo fri in fiArr)
                {

                    Console.WriteLine(fri.Name);        // pasif
yapılabilir
                    path = @"E:\WOS_SIRALI\" + "wos" +
i.ToString("D3") + ".html";
                    fri.MoveTo(path);
                    i++;
                }
                Console.WriteLine("Bitti");
            }
            catch (Exception e)
            {
                Console.WriteLine("İşlem başarısız: {0}",
e.ToString());
            }
        }
    }
}
```

ÖZGEÇMİŞ

İlker TÜRKER 1980 yılında Karabük'te doğdu; ilk, orta ve lise öğrenimini aynı şehirde tamamladı. Karabük Anadolu Lisesi'nden 1997 yılında mezun oldu. Aynı yıl İstanbul Teknik Üniversitesi Elektrik Elektronik Fakültesi Elektronik ve Haberleşme Mühendisliği'nde öğrenime başlayıp 2001 yılında iyi derece ile mezun oldu. 2002 yılında Zonguldak Karaelmas Üniversitesi Karabük Meslek Yüksekokulu'nda Bilgisayar Teknolojisi ve Programlama programında öğretim görevlisi olarak göreve başladı. Daha sonra Karabük Üniversitesi bünyesinde aynı kurumda görevine devam etti. 2003 yılında ZKÜ Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Anabilim Dalı'nda başlamış olduğu yüksek lisans programını 2006 yılında tamamladı. Halen Karabük Üniversitesi'nde görevine devam etmektedir.

ADRES BİLGİLERİ

Adres : Karabük Üniversitesi
Meslek Yüksekokulu
Balıklarkayası Mevkii / KARABÜK

Tel : (531) 793 6507

E-posta : iturker@karabuk.edu.tr