

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
İSTATİSTİK BİLİM DALI

**BÜYÜK VERİ ÜZERİNE ULUSLARARASI LİTERATÜRÜN
BİBLİYOMETRİK ANALİZİ**

Yüksek Lisans Tezi

BURCU KOCARIK

İstanbul, 2018

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
İSTATİSTİK BİLİM DALI

**BÜYÜK VERİ ÜZERİNE ULUSLARARASI LİTERATÜRÜN
BİBLİYOMETRİK ANALİZİ**

Yüksek Lisans Tezi

BURCU KOCARIK

Danışman: PROF. DR. İ. ESEN YILDIRIM

İstanbul, 2018



T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ ONAY BELGESİ

EKONOMETRİ Anabilim Dalı İSTATİSTİK Bilim Dalı TEZLİ YÜKSEK LİSANS öğrencisi BURCU KOCARIK'ın BÜYÜK VERİ ÜZERİNE ULUSLARARASI LİTERATÜRÜN BİBLİYOMETRİK ANALİZİ adlı tez çalışması, Enstitümüz Yönetim Kurulunun 5.07.2018 tarih ve 2018-18/31 sayılı kararıyla oluşturulan jüri tarafından oy birliği / ~~oy çokluğu~~ ile Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi ..20..../...7.../.....2018

Öğretim Üyesi Adı Soyadı

İmzası

Öğretim Üyesi Adı Soyadı	İmzası
1. Tez Danışmanı Prof. Dr. İ.ESEN YILDIRIM	
2. Jüri Üyesi Prof. Dr. ŞAHAMET BÜLBÜL	
3. Jüri Üyesi Prof. Dr. MÜNEVVER TURANLI	

GENEL BİLGİLER

İsim ve Soyadı	: Burcu Kocarı
Anabilim Dalı	: Ekonometri
Programı	: İstatistik
Tez Danışmanı	: Prof. Dr. İ. Esen Yıldırım
Tez Türü ve Tarihi	: Yüksek Lisans - Temmuz 2018
Anahtar Kelimeler	: Büyük Veri, Bibliyometri, Sosyal Ağ Analizi

ÖZET

BÜYÜK VERİ ÜZERİNE ULUSLARARASI LİTERATÜRÜN BİBLİYOMETRİK ANALİZİ

Teknolojinin gelişimine paralel biçimde hacimlenen veri setlerini analiz etmek ve öngörülse ipuçları bulmak, çeşitli sektörlerde karar vericiler için önemli bir sorun haline gelmiştir. Büyük miktardaki veriyi elde etme, depolama ve analiz ederek bu veriler aracılığıyla tahminde bulunma süreçlerinin tamamı “büyük veri (big data)” olarak adlandırılmaktadır. Büyük veri incelemeleri, tüm dünyada bireyler, ticari firmalar, resmi kurumlar ve bilim camiası için önemli bir gündem konusudur. Bu çalışmada, 2013 – 2017 yılları arasında Web of Science veri tabanında büyük veri üzerine yapılan ve “big data” anahtar kelimesi ile taranan uluslararası nitelikteki makaleler; yıl, yazar, ülke, konu, anahtar kelime ve araştırmacının alanı yönlerinden tanımlayıcı nitelikte bibliyometrik analizler uygulanarak veri seti oluşturulmuştur. Ardından bu veriler sosyal ağ analizi yardımı ile görselleştirilmiş ve yıl bazlı karşılaştırmalar yapılmıştır. Sonuçta, büyük verinin çalışma alanlarının gittikçe arttığı görülmüş, ayrıca bu konuda öncü olan yazar ve ülkeler belirlenmiştir. Özellikle ABD, Çin Halk Cumhuriyeti ve BK’nın bu konuda öncü olduğu görülmüştür. Araştırma kapsamında, dünya literatürü içerisinde Türkiye’deki araştırmacılar tarafından gerçekleştirilen büyük veri çalışmalarının mevcut durumu ortaya konulmuştur.

GENERAL KNOWLEDGE

Name and Surname	: Burcu Kocarık
Field	: Econometrics
Programme	: Statistics
Supervisor	: Prof. Dr. İ. Esen Yıldırım
Degree Awarded and Date	: Master - July 2018
Keywords	: Big Data, Bibliometri, Social Network Analysis

ABSTRACT

BIBLIOMETRIC ANALYSIS OF INTERNATIONAL LITERATURE ON BIG DATA

Analyzing data sets that have been volume in parallel with the development of technology and finding predictive clues has become an important challenge for decision makers in various sectors. All of the processes of obtaining large amounts of data, storing and analyzing it and making a prediction it through these data are called “big data”. Big data reviews are an important agenda for individuals, commercial companies, government agencies and the science community all over the world. In this study, a set of data is created applying descriptive bibliometric analyses from scientific articles of international quality which were scanned with the keyword ‘big data’ in Web of Science data base by year, author, country, subject, keyword and researcher’s field, between 2013 and 2017. These data were then visualized with the help of social network analysis and year-based comparisons were made. As a result, it has been seen that the studying field of the big data have been increasing, and also the leading researchers and countries have been identified in this subject. In particular, the United States, The People's Republic of China and the UK have been a pioneer in this matter. The present situation of big data studies carried out by researchers in Turkey has been revealed in the international literature within the scope of the research.

İÇİNDEKİLER

Sayfa No

TABLO LİSTESİ	v
ŞEKİL LİSTESİ	vi
KISALTMALAR	viii
1. GİRİŞ	1
2. BÜYÜK VERİ	4
2.1 Veri Kavramı	4
2.1.1 Veri Elde Etme Yaklaşımları.....	5
2.1.2 Verinin Tarihi ve Evrimi	6
2.2 Büyük Veri Kavramı	9
2.2.1 Büyük Verinin Doğurduğu Sonuçlar	15
2.2.2 Büyük Verinin Özellikleri	19
2.2.3 Büyük Veri Teknolojileri	24
2.2.4 Büyük Veri ve İşletmeler.....	31
3. BIBLİYOMETRİK ANALİZ	34
3.1 Bibliyometrik Analiz İle İlgili Kavramlar	35
3.1.1 Bibliyometri.....	35
3.1.2 Atıf.....	39
3.1.3 Etki Faktörü ve Anındalık Endeksi	41
3.1.4 Atıf Endeksleri.....	41
3.1.5 Ortak Çalışmalar	43
3.2 Bibliyometrik Analiz	44
3.2.1 Bibliyometrik Yasalar	44
3.2.1.1 Bradford Yasası	44
3.2.1.2 Lotka Yasası	45
3.2.1.3 Price Yasası	46
3.2.1.4 Zipf Yasası.....	46
3.2.2 Atıf Analizi	47
3.2.2.1 Bibliyografik Eşleştirme.....	48

3.2.2.2 Ortak Atıf.....	48
3.2.3 Bibliyometrik Haritalama Yöntemi.....	49
4. SOSYAL AĞ ANALİZİ	50
4.1 Ağ ve Özellikleri.....	50
4.2 Sosyal Ağlar	56
4.3 Sosyal Ağ Analizi.....	57
5. UYGULAMA	58
5.1 Çalışmanın Amacı ve Kapsamı.....	58
5.2 Kullanılan Veri Seti ve Tanımlayıcı İstatistikler.....	59
5.3 Sosyal Ağ Analizi.....	66
5.3.1 Makaleler ve Yazarları Arasındaki Sosyal Ağ.....	66
5.3.2 Anahtar Kelimeler Arasındaki Sosyal Ağ.....	77
5.3.3 Alanlar Arasındaki Sosyal Ağ.....	81
5.3.4 Ülkeler Arasındaki Sosyal Ağ.....	86
6. SONUÇ	91
EKLER	95
KAYNAKÇA.....	105

TABLO LİSTESİ

	Sayfa No
Tablo 1: Büyük Veri	13
Tablo 2: Büyük Verinin Tasnif Edilmesi	14
Tablo 3: Büyük Verinin Bileşenleri.....	20
Tablo 4: Veri Büyüklükleri	21
Tablo 5: Yayın Sayısı	59
Tablo 6: Alan Dağılımı Özeti	62
Tablo 7: Yayın Dilleri	63
Tablo 8: Ortak Yazarlık.....	65
Tablo 9: Derece Ölçüleri ve Kümelenme Katsayıları	76

ŞEKİL LİSTESİ

Sayfa No

Şekil 1: Dünyada Büyük Veri Kelimesine Zaman İçinde Gösterilen İlgi.....	11
Şekil 2: Türkiye’de Büyük Veri Kelimesine Zaman İçinde Gösterilen İlgi.....	12
Şekil 3: Büyük Verinin Bileşenleri Arasındaki İlişki.....	20
Şekil 4: Yeni Veri Kaynakları.....	23
Şekil 5: Bulut Bilişim.....	28
Şekil 6: Veriden Bilgi Elde Edilmesi Süreci ve Veri Madenciliği.....	31
Şekil 7: Bibliyometrinin Kapsamı.....	37
Şekil 8: Bibliyografik Eşleştirme ve Ortak Atıf Arasındaki Fark.....	49
Şekil 9 : İki Düğüm Arası Bağlantılar.....	51
Şekil 10: Kuvvet Yasası Dağılımı.....	52
Şekil 11: Sosyal Ağ Örnekleri.....	57
Şekil 12: Yayın Sayısı Dağılımı.....	60
Şekil 13: Yayın Türü Dağılımı.....	60
Şekil 14: WOS Kategorileri Dağılımı	61
Şekil 15: Makale Sayıları.....	63
Şekil 16: Ülkelerin Yayın Sayıları	64
Şekil 17: 2013 Yılında WOS’ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı	67
Şekil 18: 2014 Yılında WOS’ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı	68
Şekil 19: 2015 Yılında WOS’ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı	69
Şekil 20: 2016 Yılında WOS’ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı	70
Şekil 21: 2017 Yılında WOS’ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı	71
Şekil 22: Yıllara Göre WOS’ta Yayınlayan Makalelerin Yazarları Arasındaki	75
Şekil 23: 2013 ve 2014 Yılları Makalelerde Kullanılan Anahtar Kelime Sosyal Ağı	77
Şekil 24: 2015 Yılı Makalelerde Kullanılan Anahtar Kelime Sosyal Ağı	78

Şekil 25: 2016 Yılı Makalelerde Kullanılan Anahtar Kelime Sosyal Ağı	79
Şekil 26: 2017 Yılı Makalelerde Kullanılan Anahtar Kelime Sosyal Ağı	80
Şekil 27: 2017 Yılı Makalelerde Kullanılan Anahtar Kelime Ağ Yoğunluğu	81
Şekil 28: 2013 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı	82
Şekil 29: 2014 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı	82
Şekil 30: 2015 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı	83
Şekil 31: 2016 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı	83
Şekil 32: 2017 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı	84
Şekil 33: 2013 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ.....	86
Şekil 34: 2014 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ.....	87
Şekil 35: 2015 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ.....	88
Şekil 36: 2016 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ.....	89
Şekil 37: 2017 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ.....	90

KISALTMALAR

ABD	Amerika Birleşik Devletleri
AHCI	Art & Humanities Citation Index (Sanat ve Beşeri Bilimler Uluslararası Bilim Endeksi)
BK	Birleşik Krallık
ISI	Institute for Scientific Information (Bilimsel Enformasyon Enstitüsü)
IoT	Internet of Things (Nesnelerin İnterneti)
NLP	Natural Language Processing (Doğal Dil İşleme)
NoSQL	Not Only Structured Query Language (SQL ve Daha Fazlası)
SAA	Sosyal Ağ Analizi
SCI	Scientific Citation Index (Fen Bilimleri Uluslararası Bilim Endeksi)
SSCI	Social Science Citation Index (Sosyal Bilimler Uluslararası Bilim Endeksi)
SQL	Structured Query Language (Yapısal Sorgulama Dili)
TÜBİTAK	Türkiye Bilimsel ve Teknolojik Araştırma Kurumu
WOS	Web of Science
www	World Wide Web (Dünya Çapında Ağ)
UK	United Kingdom (Birleşik Krallık)

1. GİRİŞ

1990'lı yıllarda internet sağlayıcıların yaygınlaşmaya başlaması ve o günden bugüne hızla gelişmesi; son yıllarda dizinlenmiş, sorgulanabilen ve insanoğlunun bu güne kadar görmediği bir hızda büyüyen bir veri havuzu oluşmasına sebep olmuştur. İnternet teknolojilerindeki bu gelişimle birlikte, havuzdaki verinin neredeyse tamamı internet siteleri tarafından üretilmekte, depolanmakta ve işlenmektedir. Son on yılda ise özellikle mobil teknolojilerde meydana gelen ilerlemeler ve sosyal medyanın yaygın olarak kullanımına bağlı olarak kayıt altına alınan verinin miktarı, ölçmekte zorlanılacak boyutlara ulaşmıştır. Kısa bir süre öncesinde, internetle birlikte farklı türdeki ve ölçüdeki, bilimsel olan veya olmayan, karmaşık yapıdaki birçok verinin ortaya çıkmasıyla büyük veri kavramı doğmuştur. Büyük veri, artan teknolojiyle birlikte daha da büyüyen bir unsur olarak dikkat çekmektedir. Genel olarak büyük veri kavramı bilim dünyasının her alanında ciddi bir yere sahip olmayı başarmış, her an kendisine kattığı yeni verilerle sürekli güncellenişi ve gelişmeye açıklığıyla da uzunca bir süre daha önemini koruyacağını göstermektedir. Kişisel depolama hacimlerinin bile terabayt kapasitesine ulaştığı göz önüne alındığında; aynı zamanda gerçekleşen her olayın kişisel ve kurumsal depolama alanlarına kaydedildiği düşünüldüğünde nasıl bir veri havuzu içinde bulunduğu anlaşılabilir.

Son dönemlerde artık dijital teknolojilerin aşırı miktarda verinin depolanabilmesine izin vermesi ve veri toplama araçlarının çoğalması, büyük miktarda veriye çok kolay bir şekilde ulaşma imkanı sağlamaktadır. Öyle ki, verinin boyutundan ötürü veriyi depolamak zaman zaman verinin kendisine ulaşmaktan daha maliyetli bir hale gelmiştir. Zaman içindeki yeni gelişmelerle depolama maliyetleri de aşılmış ve veri, coğrafi olarak farklı bölgelerde bulunabilen çok sayıda sunucunun sağladığı veri ambarlarında depolanabilir ve gerektiğinde saniyeden daha kısa sürelerde geri çağrılıp kullanıma hazır olabilecek hale gelmiştir. Üretilen büyük veri; pazarlama, halkla ilişkiler, sağlık, spor, bankacılık, güvenlik, akademik vb. pek çok alanın yanında araştırmacıların yaptıkları araştırmalarda da kullanılmakta ve kullanıcılara yararlı hale getirilmeye çalışılmaktadır.

Böyle geniş hacimdeki veriyi analiz etmek ise veri madenciliği, metin madenciliği ve makine öğrenmesi gibi ileri yazılım teknikleri ve uygulamaları kullanılarak mümkün olmaktadır. Büyük veri analizi, kaynaklarının etkin kullanılması, toplumsal dinamikler hakkında çok detaylı ve güncel bilgi elde edilmesi ve mevcut durum analizleri yapılabilmesini sağlamaktadır. Toplanan verilerin analizi, geleceğe yönelik daha iyi kararlar almayı sağlayarak, kurum ve işletmeler arasındaki rekabeti arttırmaktadır. Büyük verinin analizinden elde edilen bilgiler sayesinde özellikle e-ticaret siteleri, müşterileri ve pazarı daha iyi anlamayı sağlayacak yeni ürün ve hizmetler geliştirilmektedir. Büyük verinin öneminin anlaşılması ile birlikte dünya çapında büyük veri üzerine eğilen araştırmacıların ve yapılan çalışmalarla birlikte analiz sonuçlarının yayınlanması gibi bilimsel yayın ve faaliyetlerin sayısı da artmaktadır.

Bu çalışmanın amacı, büyük veri ile ilgili olarak yayınlanan uluslararası akademik çalışmaların mevcut durumları hakkında bilgi verebilmek amacıyla bir resim elde edebilmektir. Bunun için öncelikle, uluslararası literatürde büyük veri üzerine yapılmış olan çalışmaların derlenmesi ile araştırmanın veri seti elde edilmiştir. Bu amaçla, ISI tarafından sağlanan Web of Sciences veri tabanı kullanılarak uluslararası dizinde literatür taraması gerçekleştirilmiştir. Tarama esnasında uluslararası literatürde araştırma yapılacağı için anahtar olarak “Big Data” kelimesi kullanılmıştır. Sonuçta elde edilen yayınlar sadece makale olarak filtrelenerek yayınlandıkları yıl, dergi, ülke, kurum, alan ve yazar değişkenleri yönünden Bibliyometrik Analize tabi tutulmuştur. Bibliyometrik Analiz uygulanarak elde edilen veri, Sosyal Ağ Analizi Teknikleri uygulanarak görselleştirilmiştir. Sosyal Ağ Analizi kapsamında Pajek ve VOSviewer programları kullanılmaktadır. Ardından bu çalışmalar arasındaki ağ yapılarına ilişkin sonuçlar belirlenip değerlendirilmiştir. Bu doğrultuda çalışmanın hipotezi, ‘Büyük veri üzerine yapılan çalışmalar bakımından Türkiye’nin durumu Dünya ülkeleri ile aynıdır’ şeklindedir. Böylece bu konuda dünya ülkeleri içerisinde Türkiye’nin konumu araştırılmıştır.

Bu çalışma dört ana bölümden oluşmaktadır. Birinci bölümde, verinin elde edilmesi, verinin tarihi gelişimi, büyük veriye geçiş süreci anlatılacak; büyük verinin

özellikleri ve büyük veri teknolojileri tanımlanacak ve işletmeler açısından büyük veri üzerinde durulacaktır.

Çalışmanın ikinci bölümünde, Bibliyometrik Analiz hakkında bilgi verilecektir. Analiz yönteminin uygulama alanları açıklanacaktır. Çalışmanın üçüncü bölümünde birimler arasındaki ağ analizinde kullanılan Sosyal Ağ Analizi tanıtılacak ve kullanım amaçları anlatılacaktır. Dördüncü bölümde, çalışmanın amacına yönelik olarak uygulama yapılarak, elde edilen bulgular değerlendirilecektir.

Çalışmanın sonuç bölümünde ise analiz sonucunda elde edilen bulgular ışığında çalışmanın genel bir değerlendirmesi yapılarak; uluslararası literatürde büyük veri hakkında yapılmış olan çalışmaların durumları ve Türkiye'nin bunun içindeki konumu belirlenecek ve öneriler sunulacaktır.

2. BÜYÜK VERİ

2.1 Veri Kavramı

Verinin geçmişten günümüze kabul gören belirli tanımları olmakla birlikte veri miktarı arttıkça ve çeşitlendikçe, veri (data) sözcüğünün anlamının ve literatürdeki tanımının değişmekte olduğu gözlenmektedir. Veri, incelenen birimlerin çeşitli özelliklerinden derlenen sembolik değerler olarak tanımlanabilir.¹ Genel olarak, araştırılan konuyla ilgili bilinen veya bir kaynaktan elde edilen ve bir sayı ile ifade edilebilen değerlere veri denilmektedir.

Geleneksel tanımlarda anlatılan veri daha çok nicel yapıda olup sayılar ya da rakamları ifade etmektedir. Günümüzde ise artık nitel araştırma tekniklerinin gelişmesi ile kelimeler, ifadeler, cümleler, çeşitli diğer anlatım biçimleri ve bununla beraber resimler, fotoğraflar, videolardan elde edilen bulgular da sayılar ve kelimelerle raporlandığı için veri tanımı kapsamında değerlendirilmektedir.²

Bunlara ek olarak veri; kavram veya komutların, yorum ve işlem için uygun biçimde ifade edimi olarak da tanımlanabilir. İlgilenilen herhangi bir durum hakkında bilgi edinmek için öncelikle veriye bakılır. Veri bir çeşit bilgi edinme yöntemidir.³ Veri için, sayısal ortamlarda bulunan, işlenen veya taşınan sinyaller; henüz anlamlandırılmamış bitler veya birbiriyle bağlantısı henüz açığa çıkarılmamış bilinenler şeklinde bir tanım yapmakta bilişim teknolojileri açısından uygundur.⁴

Sonuçta verinin anlam ifade edecek şekilde işlenerek var olan belirsizliğinin en aza indirilmesi ile de bilgi elde edilmektedir. Bu süreçte kullanılan tekniklerin bütününe de “istatistik” adı verilir. Veri aşağıdaki gibi gruplara ayrılarak nitelendirilebilir.⁵

¹ Şahamet Bülbül, Tanımlayıcı İstatistik, Gözden Geçirilmiş 3. Baskı, İstanbul: Der Yayınları, 2013, s.19.

² Esen Yıldırım, İstatistiksel Araştırma Yöntemleri, 3.Baskı, Ankara: Seçkin Yayıncılık, 2017, s.101.

³ Viktor Mayer-Schönberger ve Kenneth Cukier, Büyük Veri Yaşama, Çalışma ve Düşünme Şeklimizi Dönüştürecek Bir Devrim, Çev: B. Aydın, 1. Baskı İstanbul: Paloma Yayınevi, 2013, s.27.

⁴ Şeref Sağıroğlu ve Orhan Koç, Büyük Veri ve Açık Veri Analitiği: Yöntemler ve Uygulamalar, 1. Basım, Ankara: Grafiker Yayınları, 2017, s.13.

⁵ Keith Jeffery, “Data is the New Oil” Best Practices for Data Management & Sharing, Dü. The Joint Research Centre (JRC), 2014; Korcan Doğan ve Sacit Arslantekin, Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum, DTCF Dergisi 56(1):15-36, 2016, s.16.

Bunlar;

- Yapılandırılmış, yapılandırılmamış, yarı yapılandırılmış veri,
- Statik, dinamik, akan veri,
- Güvenli/özel/erişimi sınırlı veya halka açık veri,
- Ücretli veya ücretsiz veri,
- Açık hükümet verisi,
- Açık veri,
- Büyük veri şeklindedir.

2.1.1 Veri Elde Etme Yaklaşımları

Veri toplama yöntemleri asıl olarak “birincil” ve “ikincil” veri toplama yöntemleri olarak ikiye ayrılmaktadır. “Birincil” ya da “ikincil” veriye dayanılarak yapılan analizlerin sonuçlarını yeniden veri olarak kullanmak da “üçüncül” veri toplama yöntemi olarak adlandırılmaktadır.⁶

Araştırmacının araştırmasına konu olacak veriyi kendisinin toplayarak yeni veriler oluşturmasına birincil veri; bu tekniğe de birincil kaynaklardan veri toplama tekniği adı verilir. Birincil veri toplama teknikleri;⁷

- Tümdengelimci gözlem teknikleri (Deney vb.),
- Tümevarımcı gözlem teknikleri (Etnografik araştırmalar vb.),
- Tümdengelimci soru sorma teknikleri (Anket vb.),
- Tümevarımcı soru sorma teknikleri (Mülakat ve odak grup görüşmesi vb.),
- Tümdengelimci metin incelemesi (Kantitatif içerik analizi vb.),
- Tümevarımcı metin incelemesi (Kalitatif içerik analizi vb.) şeklindedir.

⁶ Yıldırım, 2017, s.105.

⁷Matthew David ve Carole D. Sutton, *Social Research: The Basics*, Thousand Oaks: Sage Publications, 2004, s.27.

Arařtırmacının arařtırmasına konu olacak veriyi, bařka kiři ya da kurumlar tarafından daha nceden toplanmıř verilerden yani ikincil verilerden oluřturmasına ikincil kaynaklardan veri toplama yntemi adı verilmektedir. İkincil veri kaynaklarına, resmi ya da yarı resmi kurumların yayınları bařta olmak zere birok kuruluřun dzenli olarak topladıkları veriler rnek gsterilebilir.

Teknolojinin geldiđi son noktada, pazarlama alanı bařta olmak zere, insan davranıřlarını inceleyerek bařta tketicim alıřkanlıkları olmak zere insana dair birok Őeyi đrenmek zere geliřtirilen algoritma tabanlı veri toplama teknikleri artık elektronik ortama kaymıřtır. Mřterilerin satın alma bilgileri kadar demografik bilgilerini de ieren bireysel kartlarla (mađaza mřteri kartları, kredi kartları vb.) yapılan tm iřlemler artık byk veri tabanlarında toplanmaktadır. Bu verileri toplayan kurumlar hem kendi amaları dođrultusunda bilgileri deđerlendirebilmekte hem de verileri nc tarafların kullanmasına izin verebilmekte veya satabilmektedir.⁸

Verilerin adlandırılması deđerlikti ve veriler geliřtike, kullanılan teknik ve teknolojiler ile yntemlerde de deđerlikler gerekli olmuřtur. Bu durum “veri bilimi” olarak isimlendirilen yeni bir bilim dalını ortaya ıkarmaktadır.⁹ Veri toplama tekniklerinin geliřmesi, veri miktarının bymesi ve bu verinin anlamlandırılabilmesi iin analizinin gerekmesi veri biliminin geliřmesine katkı sađlamaktadır.

2.1.2 Verinin Tarihi ve Evrimi

İnsan var olduđundan beri bilinli veya bilinsiz, anlamlı veya anlamsız olarak srekli veri retmektedir. Dolayısıyla retilen bu veriyi kaydetmek de insanlık tarihi kadar eskidir. Kaydedilmeyen veri tarih kayıtlarına geememekte ve zaman iinde yok olup gitmektedir. Bilgisayarların olmadıđı dnemlerde veri kaydı el ile yapılmaktaydı. Veriyi kaydetmek kadar iřleyip bilgiye evirmek de ok eskiye dayanmaktadır. Buradan da anlařılacađı zere hem gemiře hem de geleceđe iřik tutması aısından veri olgusu, verinin kaydedilmesi ve analizi olduka nemlidir.

⁸ Ahmet Mete ilingirtrk, İstatistiksel Karar Almada Veri Analizi, 1.Basım, Ankara: Sekin Yayıncılık, 2011, s.31.

⁹ Sađırođlu ve Ko, s.14.

Veri üzerine sayılabilecek ilk kayıtların Mezopotamya’da, Muhasebe’nin ortaya çıkmasından yaklaşık 7000 yıl önce, mahsullerin büyümesini ve sürülerin sayısını kaydetmek amacıyla tutulduğu düşünülmektedir.¹⁰ Hem verinin az olduğu hem de veri analizi yöntemlerinin yetersiz olduğu dönemlerde çetele tutmak ile başlayan kısıtlı veri kayıt yöntemleri, son yıllarda bilgisayar yazılımlarının ve istatistiksel paket programların da yaygınlaşması ile boyut değiştirmiştir. Bilgisayarın icat edilme amacı şüphesiz ki bilgiyi hesaplamak, kaydetmek ve insan eli ve aklı ile yapılamayacak olan işlemleri gerçekleştirmektir. Artık üretilen her şey farklı alanlar söz konusu olarak bir veri değerindedir ve makinelerce kaydedilmektedir.

IBM’in ilk kişisel bilgisayarları piyasaya sürmesi ve 1990’lara gelindiğinde artık bilgisayarların iş yerlerinde, marketlerde ve evlerde yani yaşamın her alanında görülmeye ve kullanılmaya başlanması da verisel sürecin gelişmesinde önemli bir etkiye sahip olmuştur. Buhar makinesinin endüstri devrimine neden oluşu gibi bilgisayar ve internet teknolojileri de veride devrim yapmıştır.¹¹

Eski zamanlarda veriyi toplamak ve analiz etmek çok maliyetli ve oldukça zaman gerektiren bir iş olarak görülmektedir. Buna göre, yeni sorular ya da eklenen yeni verilerle yapılan analizin tekrar yapılması gerekmektedir.¹² Zaman içinde bilgisayarların yaygınlaşması, verinin sayısallaştırılarak bilgisayarlara aktarılması ile bilginin depolanması ve işlenmesi eskiden olduğu kadar uzun zaman almamaktadır.

Yirminci yüzyıl, insanlık tarihine teknolojinin imkanlarını sunmuştur ve teknolojik gelişmeler içinde bulunduğumuz yirmi birinci yüzyılda da hızla gelişerek devam etmektedir. Her yeni teknoloji beraberinde yeni araştırma alanları, yeni bilimsel yöntemler, yeni kuramlar, yeni alışkanlıklar, yeni yaşam tarzları ve bunlara bağlı olarak da hızla artan miktardaki veriyi getirmektedir.

Bilgi ve iletişim teknolojileri kapsamında kabul edilen ve hızla gelişen internet teknolojileri ve elektronik cihazların günlük hayata girmesi sonucunda hem insan kaynaklı hem de makine kaynaklı veri üretimi oldukça artmıştır.

¹⁰ A Short History Of Big Data, <https://datafloq.com/read/big-data-history/239> (14 Ekim 2017).

¹¹ Murat Sezgin ve Lütfiye Talaz, Bilişim Devrimi, Karabük Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 6(2): 559-571, 2016, s.560.

¹² Mayer-Schönberger ve Cukier, s.2.

Web sayfaları, blog ve sosyal ağ kullanımlarının artması, sosyal medya uygulamalarının çok hızlı yaygınlaşması, medya ve sanal dünyanın ürettikleri, sensörler ve veri toplayan cihaz ve uygulamaların her alanda kullanılır olması, e-ticaretin günlük hayatın içinde vazgeçilmez bir noktaya gelmesi gibi sebeplerden ötürü her an bilimsel olan veya olmayan veriler toplanır hale gelmiştir.¹³ Tüm bunlar sonucunda web üzerindeki verinin zettabyte'lar büyüklüğüne eriştiği tahmin edilmektedir.¹⁴

İnsanoğlu, 2000'li yılların başında kurulan, dijital devrim, bilişim devrimi veya veri devrimi adı verilen yaşadığımız tüm dönemlerden farklı olan sanal bir alem ile en büyük değişimini yaşamaktadır.¹⁵ Bilgisayar ve internet (bilişim) teknolojilerinin gelişimiyle üretilen veri miktarı üstel olarak artmaktadır.¹⁶ Üstel artış, veri miktarındaki başlangıçta küçük miktarlarda meydana gelen artışların bir zamandan sonra hızlı bir yükselişle insan aklının alamayacağı miktarlara gelmesi anlamına gelmektedir. Veriler artık mevcut geleneksel bilgisayar programlarıyla analiz yapılabilecek miktardan oldukça fazladır. Dolayısıyla veri miktarındaki bu üstel artış sonucunda yönetilebilen ve analiz edilebilen verinin oranı azalmıştır. Bununla beraber veri saklama ve işlem kapasitelerinde veri miktarının artış hızına yetişebilmek için yeni yöntem ve algoritmalar geliştirilmeye devam edilmektedir. Bu durum veri kavramına yeni bir bakış açısı getirmiş, veri bilimi önem kazanmış ve “büyük veri” terimi doğmuştur.

Büyük veri henüz daha çok yeni, genç bir olgudur. En popüler sosyal medya hesaplarından Facebook'un 2004, Twitter'in 2006 yılında kurulduğu; yine popüler olan bazı akıllı cihaz üreten markalardan iPhone'un 2005, iPad'in 2010 yılında gündeme girdiği göz önüne alındığında, akıllı telefon kullanan insanların ve internete bağlanabilen nesnelerin birer veri üreticisine dönüşmeleri de daha çok yeni sayılabilir.¹⁷ Bu sebeplerle büyük verinin geçmişi çok eskiye dayanmamaktadır.

¹³ Doğan ve Arslantekin, s.20.

¹⁴ *Büyük Veri Kavramı ve Karakteristik Özellikleri*, <http://ab.org.tr/ab16/bildiri/66.pdf> (5 Kasım 2017)

¹⁵ Necmi Gürsakal, *Büyük Veri*, Genişletilmiş 3. Baskı, Bursa: Dora Yayıncılık, 2017, s.1.

¹⁶ Mayer-Schönberger ve Cukier, s.30.

¹⁷ Gürsakal, 2017, s.24.

2.2 Büyük Veri Kavramı

Dünyada her geçen gün her zamankinden daha büyük ölçekte veri üretilmekte ve bunun sonucunda var olan veri miktarı sürekli katlanmaktadır. Oldukça büyük miktardaki veriyi elde etme, depolama ve analiz ederek bu veriler aracılığıyla tahminde bulunma süreçlerinin tamamı “büyük veri” olarak adlandırılmaktadır.¹⁸ Büyük veri, mevcut durumdan yeni içgörüler çıkarmak ya da yeni değer birimleri oluşturmak amacıyla daha küçük ölçekle yapılamayacak ancak büyük bir ölçekle yapılabilecek şeyleri ifade eder.¹⁹

Büyük veri, terabyte’lar ve petabyte’larla ölçülen, her zaman kullandığımız bilgisayarlar ve klasik istatistiksel yöntemlerle analiz edilemeyen veriler olup Excel’in tüm satır ve sütunlarını kullanan veriler bile henüz büyük veriye erişememektedir.²⁰ Bir başka ifade ile büyük veri, hesaplama ve iletişim yeteneklerimize meydan okuyacak kadar büyük bir hızla akan veridir.²¹ Akan veride büyük hacimde ve farklı türdeki veriler hızla yer değiştirmektedir. Bu veri selinde kalite açısından farklılık gösteren, karışık ve tüm dünyadaki çok sayıda sunucu arasında dağıtılmış durumda veriler mevcuttur. McKinsey Global Enstitü tarafından yapılan tanıma göre büyük veri, tipik veri tabanı yazılım araçlarının kaydetme, depolama ve analiz etme yeteneğinin ötesinde olan veri setleridir.²² Literatürde birbirine benzeyen birçok büyük veri tanımı olmakla birlikte üzerinde kanaat kılınmış tek bir tanım bulunmamaktadır. Literatürde yer alan tanımlardan bazıları büyük verinin ne olduğuna; bazıları da ne yaptığına ya da neyi yapamadığına odaklanmışlardır.

Verinin nasıl büyük veri olduğu konusunda kesin bir tarihten ziyade tarihsel bir süreçten bahsetmek daha doğru olacaktır. Büyük veriye doğru akışın ve farkındalığın bundan çok yıllar önce yeni yeni başladığı tahmin edilmektedir. 1940’larda Oxford İngilizce Sözlüğü’nde veri miktarındaki büyüme oranını nitelendirmek için “bilgi

¹⁸ David Feinleib, Big Data Bootcamp, The Big Data Landscape, Apress: Berkeley. CA, 2014, s.21.

¹⁹ Mayer-Schönberger ve Cukier, s.14.

²⁰ Gürsakal, 2017, s.8.

²¹ S. Mutukrishnan, Data Stream: Algorithms and Applications, Foundations and Trends® in Theoretical Computer Science: Vol. 1: No.2:117-236, 2005, s.125.

²² James Manyika ve Diğerleri, Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2011, s.7.

patlaması” teriminin kullanıldığı bilinmektedir.²³ Büyük veri kavramı ilk kez 1970 yılında yayınlanan atmosferin ve okyanusun derinlikleri konulu bir makalede ortaya çıkmıştır.²⁴ Büyük veri terimi ise ilk kez, Ağustos 2000’de Seattle’daki 8. Dünya Ekonometri Kongresi’nde sunulan “Big Data Dynamic Factor Models for Econometric Measurement and Forecasting” adlı bildiriye Francis X. Dainbold tarafından ortaya atılmıştır.²⁵ Dainbold burada büyük veriyi, “ulaşılabilir ve potansiyel olarak bulunan, nicelik bazen de nitelik özelliği gösteren verilerdeki patlama ile veri kayıt ve depolama teknolojisinde büyük gelişmeler oluşmasını sağlayan bir olgu” olarak tanımlamıştır.

Büyük veri ifadesi tek bir sunucuya sığmayacak kadar büyük, yapılandırılmış ya da yapılandırılmamış; resim, video, e-posta, işlem verileri ve sosyal medya etkileşimleri gibi verileri ifade etmekte olup bu veriler sürekli arttığı için statik bir veri ambarına sığmamakta ve geleneksel yöntemlerle analiz edilememektedir.²⁶ Genel olarak büyük veri hem verinin türünü hem de onu depolamak ve işlemek için kullanılan teknolojiyi kapsamaktadır. Bu teknolojiler ise Google, Amazon, Facebook, LinkedIn vb. şirketlerin kendi analistleri tarafından kullanıcılarının sosyal medya verileri ve sitelerinde bıraktıkları her dijital iz ile baş edebilmek amacıyla kendileri için geliştirdikleri teknolojilerden doğmuştur.²⁷ Büyük miktarda veriden bu bilgi çıkarma işlemi ise veri madenciliğine dayanmaktadır.

Büyük veriden önce analizler, daha veriyi toplamadan önce tanımlanan hipotezleri test etmek üzerineydi. Düzenlenmiş (yapısal) ve az sayıda olan veri ile analiz yapmak görece daha kolaydı. Satır ve sütun halinde gösterilebilen bu tarz veriler ikili (binary), nümerik, eşit aralıklı ve programlama yapısına uygun şekilde belirlenen veriler olup, günümüzde çoğunluğu yapısal olmayan verilerle yer değiştirmiştir.²⁸ Verinin tümünü kullanmak bilginin genişliği içinde gizlenen bağlantıları, ilişkileri ve ayrıntıları

²³ Press Gill, 2013, A Very Short of Big Data, <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#32425fcb65a1> (20 Kasım 2017).

²⁴ Gali Halevi ve Henk Moed, The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature, 2012, <https://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/> (25 Kasım 2017).

²⁵ Gürsakal, 2017, s.8.

²⁶ Thomas H. Davenport, Big Data at Work: Dispelling the Myths, Uncovering the Opportunities, Harvard Business Review, 2014, s.10.

²⁷ Doğan ve Arslantekin, s.22.

²⁸ Jules Berman, Principles of Big Data: Preparing, Sharing and Analyzing Complex Information, Elsevier, 1st Edition, USA: MK Publications, 2013, s.2.

fark etmeyi mümkün kılar. Temelde büyük veri, geçmişe göre daha ayrıntılı ölçüm ve sonuçların daha detaylı analiz edilebilmesi ile ilgilidir.

Geçmişte veri azlığı yaşayan bilim, bugün hızla artan verinin çokluğu sorunu ile karşı karşıya kalmış durumdadır. Jesse, popüler bilim üzerine “Büyük Verinin Zaferi” başlıklı makalesinde, verilerin örneklenecek istatistiksel çıkarım yapılması dünyasından, konuya ilişkin tüm verilerin toplanarak analiz edildiği büyük veri dünyasına geçişin söz konusu olduğunu belirtmektedir. Bu sebeple hipotez destekli bilimden, veri güdümlü, makine öğrenmesine dayanan ve keşfetmeye yönelik bir bilimsel dünyaya doğru dönüşüm olduğu açıktır. Ayrıca büyük verinin örnekleme ile tamsayım arasında kalan, tamsayımına daha yakın olan ancak tam anlamıyla tamsayım da olmayan ara bir veri grubu olduğu düşünülmektedir.²⁹

Zaman içinde büyük veri (big data) kelimesine dünya çapında olan ilgi en sık kullanılan arama motoru Google’ın arama trendlerinden araştırılınca sonuç aşağıdaki grafiklerde görülmektedir. Google’da büyük verinin aranma sayısı gittikçe artmaktadır. Düşey eksendeki sayılar aranan bölge içindeki arama ilgisini grafikteki en yüksek noktaya göreli olarak göstermektedir. Bu anlamda, 100 değeri, aranan terimin en yüksek popülerliğe sahip olması anlamına gelmektedir.



Şekil 1: Dünyada Büyük Veri Kelimesine Zaman İçinde Gösterilen İlgi

Kaynak: <https://trends.google.com/trends/explore?date=2010-01-01%202018-03-15&q=Big%20Data> (15 Şubat 2018).

²⁹ Mayer-Schönberger ve Cukier, s.20.

Şekil 1’de büyük veri kelimesinin dünya genelinde 2010 yılından günümüze kadar olan aranma oranları görülmektedir. Buna göre, 2010 yılından 2012 yılına kadar aranma oranı aynı miktarlarda ve yaklaşık olarak %10 dolaylarında seyrederken, 2012 yılı itibari ile hızla artan bir eğilimde olduğu ve %25 aranma oranına ulaştığı anlaşılmaktadır. Ardından aramalar 2 yıl gibi kısa bir sürede de %75’in üzerine çıkmaktadır. Büyük veri teriminin 2014 yılından itibaren yüksek oranlarda arandığı ve 2017’den itibaren de artık maksimum aranma oranlarına ulaşarak popülerliğini sürdürdüğü görülmektedir.

Şekil 2’de Türkiye’de büyük veriye olan ilginin bir seyri görülmektedir. Bu iki şekil arama sonuçları itibariye karşılaştırıldığında zaman ilerledikçe Türkiye’nin de dünyadaki popüler trendleri takip ettiği söylenebilir.



Şekil 2: Türkiye’de Büyük Veri Kelimesine Zaman İçinde Gösterilen İlgi

Kaynak: <https://trends.google.com/trends/explore?date=2010-01-01%202018-03-15&geo=TR&q=Big%20Data> (15 Şubat 2018).

Küçük veriden büyük veriye, küçük bilimden büyük bilime doğru bir geçiş söz konusudur.³⁰ Büyük veriyle birlikte artık bir hipotezi test etmek için uygulanacak yöntemleri araştırmak yerine, veriye uygulanabilecek analizler araştırma konusu olmaktadır. Bununla beraber büyük verinin boyutlarının satır ve sütunları aşması ve geleneksel analitik yöntemlerde kullanılan yapılandırılmış formlardan farklı ve düzensiz

³⁰ Gürsakal, s.78.

(yapılandırılmamış), geleneksel veri tabanlarının baş edemeyeceği kadar büyük halde bulunması verinin analizinde ve yorumlanmasında zorlukları da ortaya çıkarmaktadır.

Küçük veriden büyük veriye geçiş süreci Tablo 1’de görüldüğü gibi veriyi sağlayan ve saklayan sunucu sayısında da büyük bir artışa sebep olmuştur. Miktarı bir terabayttan büyük olan verinin büyük veri olarak nitelendirilebileceği görülmektedir.

Tablo 1: Büyük Veri

Veri Sınıfı	Veri Büyüklüğü	Veri Ne İle Yönetilir?	Veri Nerede Saklanır?	Veri Büyüklüğüne Dair Örnek
Küçük	< 10 GB	Excel, R	Sadece tek bir makinenin belleği	Binlerce satış sayısı
Orta	10 GB–1 TB	Endekslenmiş dosyalar, Monolitik veri tabanları	Bir makinenin diski	Milyonlarca web sayfası
Büyük	> 1 TB	Hadoop, Dağıtık veri tabanları	Epey çok sayıda makinenin belleği	Milyarlarca web tıklaması

Kaynak: Winning with Big Data, Slide Share, 2010.

Lord Kelvin’in “Ölçmek bilmektir” ve Francis Bacon’un “Bilgi güçtür” sözlerinden hareketle verinin doğru toplanması, depolanması ve analiz edilmesinin ne kadar önemli olduğunun altı çizilebilir.³¹ Ölçülemeyen bilgi yönetilememektedir. Günümüzde ölçülmesi gereken verinin miktarı oldukça yüksektir. Burada veriyi doğru analiz edebilen, veri akışını yakalayabilen organizasyonların daha avantajlı konumda olacağı açıktır.

IBM’in duyurduğu sonuçlara göre 2020 yılında yaklaşık 40 zetabayt veya 43 trilyon gigabayt veri üretileceği öngörülmektedir, bu miktar 2005 yılında tüm dünyada

³¹ Mayer-Schönberger ve Cukier, s.41.

üretilen verinin tam 300 katına denk geliyor.³² Bir başka ifade şekliyle, 2005'ten 2020'ye kadar, dijital dünyada veri miktarı 130 exabayttan, 40 bin exabayta (bu, 2020 yılında her erkek, kadın ve çocuk için 5,200 gigabayttan daha fazla veriye tekabül etmektedir) kadar büyüyecek demek oluyor.³³ Bu bilgilerden yola çıkarak dijital dünyanın her iki yılda bir yaklaşık olarak kendini ikiye katladığı belirtilebilir. Bu da büyümenin çarpıcı boyutlarını göz önüne sermektedir. Dolayısıyla bu büyümeyle baş edebilmek için üretilen, dağıtılan ve tüketilen tüm dijital verilerin ölçüsü olarak nitelenen dijital evrenin alt yapısının da büyüyüp güçlenmesi gerekmektedir. Büyük veri setleri Tablo 2' de görüldüğü gibi birbirinden farklı şekillerde oluşabilmektedir.

Tablo 2: Büyük Verinin Tasnif Edilmesi

Veri Türü	Meta Veri, Ana Veri, Geçmiş Veri, İşlemsel Veri
Veri Biçimi	Yapısal Veri, Yarı Yapısal Veri, Yapısal Olmayan Veri
Veri Kaynağı	Web Siteleri, Sosyal Medya, Makine veya İnsan Kaynaklı, Nesnelerin İnterneti (IoT), Dâhili Kaynaklar, İşlem Verisi, Veri Sağlayıcıları
Veri Frekansı	İsteğe Bağlı, Sürekli, Gerçek Zamanlı, Zaman Serileri
Veri Saklama	Sütun Tabanlı, Grafik Tabanlı, Doküman Tabanlı
Veri Tüketicileri	İnsan, İş Süreci, Kurumsal Uygulamalar, Veri Ambarları
Veri Kullanımı	Endüstri, Akademi, Devlet, Araştırma Merkezleri
Analiz Türü	Toplu, Akan, İnteraktif
İşleme Amacı	Tahmine Dayalı, Analitik, Sorgu ve Raporlama, Modelleme
İşleme Metodu	Yüksek Performanslı Hesaplama, Dağıtık, Paralel, Kümeleme

Kaynak: Sağiroğlu ve Koç, Büyük Veri ve Açık Veri Analitiği, 2017, s.18

³² Ogan Özdoğan, Büyük Veri Denizi, İş ve Yönetim Serisi, 72. 1.Basım, Ankara: Elma Yayınevi, 2016, s.12.

³³ John Gantz ve David, Reinsel, The Digital Universe In 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East, 2012, <http://www.emc.com/leadership/digital-universe/index.htm> (2 Aralık 2017).

Sosyal ağlardan, web sunucularına yapılan bağlantılardan, web sayfalarının içeriklerinden, ATM'ler ve bankacılık işlemlerinden, finansal piyasalardan, akıllı telefonlardan, kamera ve mikrofonlardan, YouTube'a yüklenen görüntü ve videolardan, anten ya da radyo frekansı kullanarak nesnelere bilgilerini ileten teknolojilerden, GPS izlerinden, otomobillerin telemetrelerinden, uzay teleskoplarından, havayı algılayan meteoroloji teknolojilerinden, radyo frekansı okuyuculardan, CERN'de yapılan deneylerden ve dijitalliğin içinde bulunduğu daha bir sürü yerden dalgalar halinde artarak veri gelmektedir. Bu araçlar, hızlı ve kolay bir şekilde veri oluşturup kaydedilmektedir. Öyle ki sağlık sektörü, geniş ölçekli e-ticaret uygulamaları, eğitim sektörü vb. her alanda büyük veri ile karşılaşmak mümkündür.

2.2.1 Büyük Verinin Doğurduğu Sonuçlar

Büyük verinin yaygınlaşma sebepleri sıralandığında bunlardan ilki, yüksek sayıda kullanıcı profiline sahip olan sosyal medya şirketlerinin büyük veri kullanımını arttırmalarıdır. Diğer bir sebep, bu sosyal medya hesaplarına ilk kuruldukları dönemlerde kısıtlı sayıda kullanıcı tarafından erişilebiliyorken daha sonraları halka açık hale gelmeleridir. Üçüncü ve son sebep ise, daha küçük şirketlerin teknolojinin getirdiği yeniliklere dönük olarak kendilerini geliştirmeyi ve büyük verinin kullanımına sahip olmayı hedeflemeleridir.³⁴ Tüm bu gelişmelere bakıldığında 2000'li yıllarla beraber yavaş yavaş başlayan büyük veri etkisinin sınırları aşır tüm dünyaya yayılarak günümüze ulaştığı ve bunun sadece başlangıç olduğu sonucu çıkarılabilir.

2012 yılında Davos'ta gerçekleşen Dünya Ekonomik Forumu'nda "Büyük Veri, Büyük Etki" başlıklı raporla büyük verinin önemi vurgulanmıştır. Yine o yıl Amerika Birleşik Devletleri (ABD), "Büyük Veri Girişimi"ni başlatmıştır. Bu girişimle ABD, büyük veri ile ilgili çalışmalara 200 milyon dolar bütçe ayırdığını ilan etmiştir.³⁵ 2012 yılı Nisan ayı, The American Mathematical Society, the American Statistical Association, the Mathematical Association of America, and the Society for

³⁴ Feinleib, s.22-25.

³⁵ *Press Release: Obama Administration Unveils "Big Data" Initiative: Announces \$200 Million In New R&D Investments*, 2012, <https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unveils-big-data-initiative-announces-200> (10 Aralık 2017).

Industrial and Applied Mathematics’’ tarafından ‘‘Matematik, İstatistik ve Veri Seli İçin Farkındalık Ayı’’ olarak ilan edilmiştir.³⁶

Büyük veri için dönüm yılı 2012 olarak değerlendirilmektedir. Bu yıl, büyük veri çağının adeta başlangıcı kabul edilmektedir.³⁷ Tüm bunların sonucunda ilgi çekmiş olan büyük veriyi her geçen yıl önceki yıllardan daha hızlı ve dolayısıyla daha fazla bir şekilde üretilir hale gelmiştir.

2017 yılında internette bir dakika kadar kısa bir süre içinde gerçekleşen olaylar ve ortalama meydana geliş miktarları aşağıda sıralanmıştır.³⁸

- 156 milyon mail gönderilmiş,
- 16 milyon kısa mesaj gönderilmiş,
- Youtube’da 4.1 milyon video izlenmiş,
- Google’da 3.5 milyon tarama yapılmış,
- Snapchat ile 1.8 milyon snap paylaşılmış,
- Facebook’a 900 bin bağlantı (giriş) sağlanmış,
- Twitter’dan 452 bin tweet atılmış,
- Google Play ya da App Store’dan 342 bin uygulama indirilmiş,
- Instagram’da 46 bin 200 fotoğraf paylaşılmış,
- LinkedIn’da 120 yeni hesap açılmış,
- İnternet alışverişinde yaklaşık 752 bin dolar harcanılmış,
- Netflix’te 70017 saat film seyredilmiştir.

Sadece 60 saniyede meydana gelen bu çıktılar, birkaç dakika ya da saatte bile ne kadar çok miktarda ve karmaşık veri üretildiğini anlamak açısından önemlidir. Burada büyük verinin doğru bir şekilde işlenerek bulgularının doğru kullanılması ve

³⁶ *Mathematics Awareness Month*, 2012, <http://www.mathaware.org/mam/2012/> (15 Aralık 2017).

³⁷ Steve Lohre, *How Big Data Became So Big*, 2012, <http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html> (15 Aralık 2017).

³⁸ *Here’s Everything That Happens in One Minute on the Internet*, <http://www.businessinsider.com/everything-that-happens-in-one-minute-on-the-internet-2017-9> (20 Aralık 2017).

yorumlanması ile veriden öğrenilenler ve bunların faydalarının büyük veriyi kullananlara ne kadar büyük katkılar sağlayacağı açıktır.

Sağlayabileceği faydalar kadar büyük veri hususunda karşılaşılabilecek bazı sorunlar da bulunmaktadır. Başta depolama alanlarının yetersiz kalması ve gerçek zamanlı anlık analiz yapabilecek tekniklerin henüz büyük verinin tamamını analiz edemiyor olması gibi sorunlar mevcuttur. Özellikle kişisel internet kullanım mahremiyeti sorunu çok dikkat edilmesi gereken bir konudur. Çok sayıda insan farkında bile olmadan ürettiği veriler dolayısıyla internette bilinmeyen taraflarca izlenebilmektedir. Kullanıcıların internet sitelerinde gezinirken bıraktıkları şifreler, tıklamalar, girilen menüler gibi dijital izler kişisel bilgi ve özellikleri barındırmaktadır. Bunların kullanılan web sayfaları dışında ikincil kullanımları ya da üçüncü şahısların eline geçmeleri sonucunda güvenlik sorunları yaşanmaktadır. Bu sebeple siteler, mahremiyet hukuku ve etik kurallar gereğince, kişisel bilgilerin işleme tabi tutulup tutulmayacağına ya da kimler tarafından tutulacağına karar vermelerine imkan tanıyarak kontrolü bireylerin kendilerine bırakmaktadır. Ancak bu bir ideal olup ne kadar tam anlamı ile sağlanabildiği bilinmemektedir. Birçok işletme müşterilerinden topladıkları verileri diğer taraflarla paylaşabilmekte ya da satabilmektedir. Diğer bir büyük veri sorunu ise büyük verinin kime ait olduğudur. Üretilen bu büyük miktardaki verinin hangi taraflara ait olduğu henüz kesin olmamakla birlikte şimdilik veriyi elde eden verinin sahibi olarak görülmektedir. Hangi verilerin toplandığı ve nasıl kullanıldığı gibi soruların cevapları açıkça belirtilmemekte ve bu konudaki tartışmalar sürmektedir.³⁹⁴⁰

Mobil cihazlara indirilen ve kurulan uygulamaların arka planda çalışırken hangi verileri topladıkları kesin olarak bilinmemektedir. Dolayısıyla verilen erişim izinleri sonucunda uygulamaların nelere erişebildikleri de tam olarak bilinmemektedir. Bu durum güvenlik açıklarına sebep olmakta ve güvenlik sorunları doğurmaktadır. Avrupa Birliği Komisyonu verilerine göre, Avrupalıların yüzde 90'ı mobil uygulamalar aracılığıyla şirketlerin izinsizce kendi verilerini toplamalarından rahatsızlık duymakta

³⁹Gürsakar, s.40

⁴⁰Finley Klint, Why Everyone (Not Just Geeks) Should Care About Big Data, 2012, <https://www.wired.com/2012/09/human-face-of-big-data/> (5 Ocak, 2018).

ve bu verilerin ifşa edilmesinden kaygılanmaktadır.⁴¹ 2013 yılında İngiltere merkezli kurulan ve tüm dünya genelinde faaliyet gösteren bir tür danışmanlık ve veri analiz firması olan Cambridge Analytica'nın, yaklaşık 87 milyon Facebook kullanıcı profiline ait kişisel verilerinden oluşan büyük bir veriyi izinsiz bir şekilde topladığı ve veri analizi için kendisine başvuran müşterilerinin özel amaçları doğrultusunda kullandığı konusu gündeme gelmiştir. Şirket, çeşitli uygulama şekilleri ile verileri analiz edip, davranış bilimini kullanarak tüketici, takipçi, seçmen davranışlarını değiştirmek isteyen kuruluşların pazarlama araçlarıyla hedef kitlesindeki kişileri internet reklamcılığı aracılığıyla eşleştirebilmektedir.⁴² Cambridge Analytica, Facebook'a bağlı bir uygulama üzerinden izinsiz toplanan dijital verileri 2016 yılında, ABD Başkanlık Seçimleri'ni manipüle etmek ve İngiltere'de gerçekleşen AB referandumunun sonuçlarını etkilemek için kullanmakla suçlanmış ve soruşturma altına alınmıştır. Aynı zamanda Federal Ticaret Komisyonu (FTC), Facebook'un kullanıcı verilerini Cambridge Analytica'ya kasıtlı sunduğu şüphesi ile şirket hakkında soruşturma başlatmıştır.⁴³

Yaklaşık 2 milyar kullanıcısı olan Facebook aynı zamanda siyasetçilerin seçmenleriyle bulunduğu yerlerden birisi konumunda yer almaktadır.⁴⁴ Bu dijital verilerin elde edilmesi ile sosyal medyada kişiye özel kampanyalar yürütülüp rakip adaylar hakkında olumsuz ifadeler kullanılabilen ve bazı adayların desteklenmesine imkan sağlanabilmektedir. Türkiye, 50 milyondan fazla kullanıcı ile dünyada Facebook'u en çok kullanan ülkelerden birisidir.⁴⁵ Bu sebeple dijital verilerin güvenliği açısından bu olayın sonuçlarının Türkiye'yi de yakından ilgilendirdiği kabul edilebilir.

Soruşturmalar sonucunda verilerin izinsiz bir şekilde toplandığı ortaya çıkmıştır. Skandalın ortaya çıkmasıyla dünyanın en büyük sosyal paylaşım sitesi Facebook kısa süre içerisinde önemli miktarda maddi değer kaybetmiştir. Cambridge Analytica, müşteri kayıpları gerekçesi ile faaliyetlerini durdurma kararı almış ve iflas

⁴¹ *Facebook'un Veri Skandalı Regülatörleri Harekete Geçirdi*, 2018, <http://avrupaforum.org/facebookun-veri-skandalı-regulatorleri-harekete-gecirdi/> (15 Nisan 2018).

⁴² Facebook skandalı 87 milyon kullanıcıyı etkiledi, 2018, <https://www.bbc.com/turkce/haberler-dunya-43649116> (30 Nisan 2018).

⁴³ *Facebook'un Veri Skandalı*, 2018, <http://www.turkiyegazetesi.com.tr/teknoloji/551691.aspx> (20 Nisan 2018).

⁴⁴ *Facebook skandalı 87 milyon kullanıcıyı etkiledi*, 2018, <https://www.bbc.com/turkce/haberler-dunya-43649116> (30 Nisan 2018).

⁴⁵ Banu Güven, Facebook'un İmtihani, 2018, <http://www.dw.com/tr/banu-guven-facebook-imtihan/a-43290135> (5 Mayıs 2018).

sürecine girmiştir.⁴⁶ Facebook – Cambridge Analytica skandalı sonrasında açılan davalar sonucunda Facebook'un federal yasaları çiğnediği, haksız rekabette bulunduğu ve sorumluluklarını yerine getirmediği iddia edilmiştir.⁴⁷ Bunun üzerine Facebook tarafından, usulsüz veri kullanımına dair iddiaların gözden geçirileceği, Cambridge Analytica'nın faaliyetlerinin bilişim uzmanları tarafından inceleneceği açıklanmıştır. İngiltere Parlamentosu Dijital, Kültür, Medya ve Spor (DCMS) Komisyonu, iddialarla ilgili ifade vermek üzere Facebook'un kurucusu ve yöneticisi olan Mark Zuckerberg'i ifade vermeye davet etmiştir.⁴⁸ Benzer şekilde Facebook'un CEO'su Mark Zuckerberg iddialar üzerine ABD Kongresi'nde ifade vermiş ve verilerin kötüye kullanılması nedeniyle ortaya çıkan skandaldan dolayı özür dilemiştir. Facebook, bu skandaldan sonra tekrar kullanıcı güvenini kazanmaya çalışmaktadır.⁴⁹

2.2.2 Büyük Verinin Özellikleri

Verinin hacmi, türü, hızı ve karmaşıklığı inanılmaz ölçüde artmış durumdadır. Verinin karmaşıklığını tanımlamak ve veri yönetimini kolaylaştırmak için bugün kullanılan 3V modeli ilk kez Doug Laney tarafından 2001'de, "3D Data Management, Controlling Data Volume, Velocity and Variety" başlıklı makalesinde açıklanmıştır. Buna göre, büyük veriyi kapasite/hacim/miktar, hız ve tür/çeşitlilik olarak üç boyutta düşünmek mümkündür. Bu bileşenlerin hepsi sürekli olarak birbiriyle etkileşim halindedir.

Yapılan yeni çalışmalarda, bu bileşenlerin sayısının arttığı görülmektedir. Laney'in bu üç boyut tanımına, iki boyut; geçerlik ve değer de eklenerek büyük verinin bileşenleri beş boyutlu hale gelmiştir.⁵⁰ Şekil 3'te büyük verinin bileşenleri arasındaki ilişki görülmektedir.

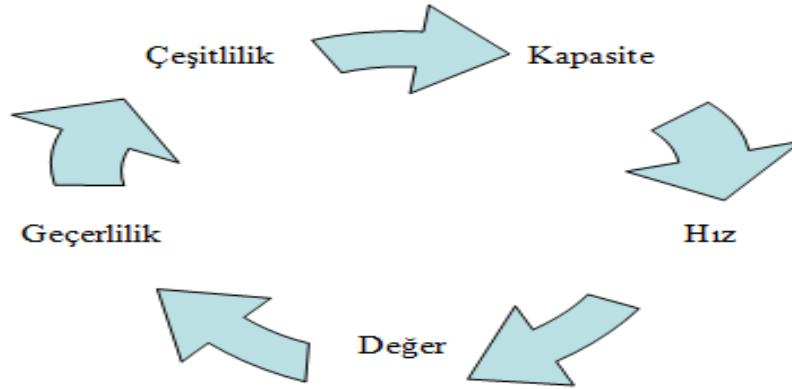
⁴⁶ Cambridge Analytica Şirketi Kapanıyor, 2018, <http://www.dw.com/tr/cambridge-analytica-sirketi-kapaniyor/a-43630357>. (10 Mayıs 2018).

⁴⁷ Facebook - Cambridge Analytica veri skandalı hakkında bilmeniz gerekenler, 2018, <http://www.haberturk.com/facebook-cambridge-analytica-veri-skandal-hakkinda-bilmeniz-gereken-her-sey-1885728-ekonomi> (10 Mayıs 2018).

⁴⁸ Facebook skandalı büyüyor, 2018, <http://www.haberturk.com/facebook-skandal-bu-yuyor-1884885#> (15 Mayıs 2018).

⁴⁹ Facebook skandalı 87 milyon kullanıcıyı etkiledi, 2018, <https://www.bbc.com/turkce/haberler-dunya-43649116> (20 Mayıs 2018).

⁵⁰ Mohammed A. F., Humbe V. T., ve Chowhan, S. S., A Review of Big Data Environment and its Related Technologies, In Information Communication and Embedded Systems (ICICES), International Conference on (pp. 1-5). IEEE, 2016, s.1.



Şekil 3: Büyük Verinin Bileşenleri Arasındaki İlişki

Kaynak: <http://andressilvaa.tumblr.com/post/87206443764/big-data-refers-to-5vsvolume> (20 Mart 2018).

Bu kavramlar, literatürde İngilizce karşılıklarından ötürü kısaca 5V (Volume, Velocity, Variety, Veracity, Value) olarak adlandırılmaktadır. Bu beş bileşen genel olarak şu şekilde özetlenebilir:

Tablo 3: Büyük Verinin Bileşenleri

5V Bileşen	Özellik
Volume(Büyükölç/Hacim/Kapasite/Miktar)	Terabyte, Petabyte, Exabyte, Zetabyte
Velocity (Hız)	Batch (birimi), Gerçek Zamanlı, Etkileşimli, Akan Veri
Variety (Çeşitlilik, Tür)	Yapılandırılmış, Yarı Yapılandırılmış, Yapılandırılmamış
Veracity (Geçerlik)	Tutarlık, Güvenilirlik, Doğruluk
Value (Değer, Kalite)	İstatistiksel, Gizli, Bilinmeyen

Bu kavramlara ilişkin detaylı açıklamalara aşağıda yer verilmektedir:

Büyüklik:

Büyük verinin en çok bilinen ve verinin miktarının nitelendiği özelliğidir. Verinin bilgisayarda ya da sistemde ya da bulutta kapladığı alan olarak da tanımlanabilmektedir.

Tablo 4: Veri Büyüklükleri

Birim	Büyüklik	Karşılık
Bit (b)	1 ya da 0	“Binary digit” kısaltmasıdır, veriyi işlerken bilgisayarların kullandığı ikili koddur.
Byte (B)	8 bits	Bilgisayarda, bir karakteri ya da sayıyı oluşturmak için yeterli bilgidir.
Kilobyte (KB)	1000 ya da 2 ¹⁰ bytes	Yunanca “bin” kelimesinden gelmektedir. Örneğin, bir sayfalık bir metin 2KB’dır.
Megabyte (MB)	1000 KB; 2 ¹⁰ bytes	Yunanca “büyük” kelimesinden gelmektedir. Örneğin, Shakespear’in çalışmalarının tümü 5MB’tır. Normal bir pop şarkısı yaklaşık 4MB civarındadır.
Gigabyte (GB)	1000 MB; 2 ³⁰ bytes	Yunanca “dev” kelimesinden gelmektedir. Örneğin, iki saatlik bir film 2GB’a sıkıştırılabilir.
Terabyte (TB)	1000 GB; 2 ⁴⁰ bytes	Yunanca “kocaman” kelimesinden gelmektedir. Amerikan Kongre kütüphanesindeki tüm sınıflandırılmış kitaplar toplamda 15TB’dır.
Petabyte (PB)	1000 TB; 2 ⁵⁰ bytes	Google saat başı yaklaşık 1PB işlem yapmaktadır.
Exabyte (EB)	1000 PB; 2 ⁶⁰ bytes	Örneğin, ekonomistin 10 milyar kopyası yaklaşık 1EB’dır.
Zettabyte (ZB)	1000 EB; 2 ⁷⁰ bytes	2010 yılındaki verinin toplam varlığının, 1.2 ZB olduğu tahmin edilmektedir.
Yottabyte (YB)	1000 ZB; 2 ⁸⁰ bytes	Hayal edilemeyecek kadar büyüktür.

Kaynak: Cukier, K., Data, Data Everywhere, The Economist Newspaper, 2010, s.2

Tablo 4’de yer aldığı gibi bir petabyte bir katrilyon byte’a eşdeğerdir ve exabyte bir milyar gigabyte’a karşılık gelmektedir. Bugün internetin her bir saniyede karşılaştığı veri miktarı, bundan 25 yıl önce depolanan tüm veriden daha fazladır ve bu da tek bir veri seti ile değil, petabytelar boyutunda veriden sonuç elde etme olanağı vermektedir.⁵¹

Hız:

Büyük verinin hızı ile kastedilen sürekli yeni verilerin üretilmesi ile veri setinin artan bir dinamizme sahip olmasıdır. Ağ ya da internet ortamında bulunan bir verinin iki nokta arasında iletilmesi sırasında bir saniyede geçen bit ya da byte sayısı verinin hızını vermektedir. Birçok şirket için veri üretme hızı büyüklükten daha önemlidir, çünkü neredeyse gerçek zamanlı bilgiler bir şirketin rakiplerinden çok daha girişken olmasını sağlar.⁵² Veri ne kadar hızlı üretilirse, verinin işlem hızı ve çeşitliliği de o oranda artmaktadır.⁵³ Dolayısıyla akan veri ile birlikte gerçek zamanlı analizlerin önem kazanmakta olduğu söylenebilir.

Çeşitlilik:

Büyük verinin çeşitliliği veri kaynaklarının farklılığından kaynaklanmaktadır. Sıklıkla kullanılan Facebook, Twitter, LinkedIn gibi sosyal medya kaynakları, e-postalar, YouTube’daki paylaşımlar, audio ve video dosyaları, tüm hareketlerin tek tek kayıt altına alındığı log dosyaları, internet siteleri, akıllı telefon veya IoT (internete bağlanabilen) cihazlar ile çeşitli makineler tarafından üretilen sinyal ve sensörler önemli veri kaynaklarını oluşturmaktadır. Bu veri kaynakları her an artan bir şekilde ve çok hızlı veri üretmektedir. İnternetin yaygınlaşması, her türden verilerle karşı karşıya kalınmasına neden olmaktadır. Dolayısıyla, verilerin elde edildiği kaynakların artmasıyla da verilerin çeşitlendiği kabul edilebilir. Veri kaynakları IBM tarafından aşağıdaki şekilde özetlenmiştir.

⁵¹ Andrew McAfee ve Erik Brynjolfsson, Big Data: The Management Revolution, Harvard Business Review, Harvard Business School Publishing Corporation, 2012, s.4.

⁵² McAfee ve Brynjolfsson, s.5.

⁵³ Bülent Yılmaz, Samet Bülbül ve Murat Atik, Büyük Verinin (Big Data) Muhasebe Üzerindeki Etkisi Ve Muhasebeye Sağladığı Katkıların İncelenmesi, Kara Harp Okulu Bilim Dergisi, Haziran, 27(1). 79-112, 2017, s.90.

Üretilen veriler yapısal, yapısal olmayan veya yarı yapısal formlarda oluşmaktadır. Klasik veri tabanı sistemleri küçük boyuttaki düzenli veriler için tasarlanmıştır. Bu veriler yapısal/yapılandırılmış veri olarak adlandırılmaktadır. Ancak tüm veriler belirli bir düzen içinde üretilmemekte ya da düzenlenememekte ve her türlü düzen ya da yapı içerisinde bulunabilmektedir.⁵⁴ Kısacası büyük veri karmaşık bir yapıda üretilmektedir. Resim, ses, video, metin dosyaları, konum ve mekan verileri, uydu görüntüleri, internet üzerinden sağlanan veriler ve sosyal medya verileri yapısal olmayan/yapılandırılmamış verilerdir.



Şekil 4: Yeni Veri Kaynakları

Kaynak: IBM Offering Information, Making Sence of Big Data, 2017

Yarı yapısal veriler ise kendine has bir düzen ve yapı içerisinde bulunabilen veriler olup XML dosyaları (Genişletilebilir İşaretleme Dili; Hem insanlar hem de bilgi işlem sistemleri tarafından kolayca oluşturulabilecek dökümanlar elde etmeye yarayan formattır) ve işlem tabloları örnek gösterilebilir. Büyük veri genellikle yapılandırılmamış veriden oluşmaktadır.⁵⁵ Bu kadar farklı yapıdaki verinin bir arada kullanılıyor olması, ölçeklenebilirlik ve veri bütünlüğü ile verinin niteliğini bozmadan analiz edebilme aşamalarında problem oluşturabilmektedir.

⁵⁴ Mayer-Schönberger ve Cukier, s.105.

⁵⁵Şemsettin Çiğdem ve Halil İbrahim Seyrek, İşletmelerde Büyük Veri Uygulamaları: Bir Literatür Taraması, 2015, s.3. <https://www.researchgate.net/publication/293439800> (25 Şubat 2018).

Değer ve Geçerlik:

Veriyi anlamlandırabilmek için veriden üretilen değer, verinin içeriğine, üretilme amacına, uygulama alanına vb. faktörlere göre değişiklik göstermektedir.⁵⁶ Bu sebeple sahip olunan verilerin analiz çalışmasının yapılması ve değer katması gerekmektedir. Verinin değer bileşeni, verinin ortaya çıkardığı maddi değer olarak da ifade edilebilmektedir. Verinin üretildiği ve depolandığı kurum için de bir artı değer yaratıyor olması gerekir, aksi halde katlanılan maliyet elde edilen değerden büyük hale gelebilmektedir.⁵⁷

Geçerlik, büyük verinin güvenli olması durumuyla ilgilidir. Günümüzde güvenilirliği şüpheli olan pek çok farklı kaynaktan veri üretilmektedir. Yüksek hacimde, çeşitli ve çok hızlı akarak sisteme giren verinin doğru ve güvenilir olması geçerli olmasında etkili olmaktadır.⁵⁸ Bu durum verinin kalitesine ve analiz sonuçlarının doğruluğuna da etki etmektedir. Verilerin geçerli veya güvenilir olmadığı durumlarda, örneğin işletmeler geleceğe dönük yanlış kararlar alabileceklerdir. Bu durumda elde edilen verilerin güvenilirliğinin test edilmesi önemli bir husus olarak değerlendirilebilmektedir.

2.2.3 Büyük Veri Teknolojileri

Büyük verinin ortaya çıkmasıyla birlikte, elde edilen verilerin saklanması ve analiz edilmesi konusu gündeme gelmiştir. Büyük veri teknolojilerinden öncelikle büyük verinin temel yapısını oluşturan MapReduce ve Hadoop ardından da NoSQL, nesnelerin interneti, doğal dil işleme, bulut bilişim, makine öğrenmesi gibi yazılımlar ve veri madenciliği büyük verinin hem kaydedilmesi hem de işlenmesinde temel oluşturmaktadır.⁵⁹ Özellikle Hadoop (Dağıtık Veri Tabanı), Veri Merkezleri (Data Center), Bulut Bilim (Cloud Computing) ve İnternet gibi teknolojiler elde edilen büyük miktardaki verinin depolanmasında hizmet sunmaktadırlar.

⁵⁶ Sağıroğlu ve Koç, s.21.

⁵⁷ James Manyika ve Diğerleri, s.27.

⁵⁸ Gürsakal, 2017, s.59.

⁵⁹ Çiğdem ve Seyrek, s.6.

MapReduce

Veri hacminin çok büyümesiyle beraber bilgisayarların işlem için kullandıkları bellekler yeterli kalmayıp analiz araçlarının yenilenmesi gerekmiştir. Bu sebeple Google, MapReduce (Eşle İndirge) adında büyük veriyi depolayan yeni bir işletim teknolojisi geliştirerek bilgisayarların işlem güçlerini arttırmayı amaçlamıştır.⁶⁰ Veriler işlenirken iki aşama kullanılmaktadır. Map (Eşle) aşamasında, büyük veri ana düğümden (sunucudan) alınıp yönetilebilir daha küçük görevlere bölünerek tekrar çok sayıda düğüme (sunuculara) dağıtılır. Küçük düğümlerdeki veri analiz edildikten sonra elde edilen sonuçlar ana düğümlere geri gönderilir. Reduce (İndirge) aşamasında ise tamamlanan işler birleştirilerek sonuç elde edilmektedir. Bu açıdan Eşleİndirge, ‘‘Böl ve Yönet’’e benzemektedir.⁶¹ Çok büyük bir veriyi işleyebilmek için çok yüksek donanıma sahip sunucular kullanmak yerine, paralel olarak birlikte çalışan ve sıradan sunuculardan oluşan bir küme üzerinde MapReduce yardımıyla aynı işlem çok daha etkin bir şekilde gerçekleştirilmektedir.⁶²

Hadoop

Hadoop, çok büyük veri kümelerini, paralel olarak çalışan binlerce bilgisayar düğümüne (işlemci) işlemek için tasarlanmış yüksek oranda ölçeklenebilir bir dağıtık dosya sistemi ile MapReduce çerçevesinde, Java ile geliştirilmiş açık kaynak kodlu bir depolama platformudur. Aynı zamanda bir programlama modeli olan MapReduce, dağıtık yapı alan Hadoop’un temel yapısı olup çok sayıda sunucu arasında büyük ölçeklenebilirlik sağlamaktadır.⁶³ Gerekli verinin hızla ulaşılabilir şekilde amaca uygun bir şekilde saklanması ve gerektiğinde saklandığı yerden hızla geri çağrılabilmesi gerekmektedir. Böylece Hadoop, büyük veri hacimleri için yüksek maliyet gerektirmeyen depolama çözümleri sunmaktadır. Burada bahsedilen açık kaynak kodlu yazılım ile anlatılmak istenen ise yazılımın kaynak kodunun/kaynağının erişilebilir,

⁶⁰ Apache Mapreduce, <https://www.ibm.com/analytics/hadoop/mapreduce> (2 Mart 2018).

⁶¹ Gürsakal, 2017, s.212.

⁶² MapReduce Nedir?, <http://devveri.com/hadoop/mapreduce-nedir> (14 Mart 2018).

⁶³ Mapreduce, <https://www.ibm.com/analytics/hadoop/mapreduce> (2 Mart 2018).

tekrarlanabilir ve doğrulanabilir ya da kullanıcı tarafından yazılımın değiştirilebilir olması amacıyla açık yani paylaşılabılır olması durumudur.⁶⁴

NoSQL

Geleneksel veri tabanlarına erişmede, önceden belirlenmiş bir kayıt yapısına sahip olan SQL (Structured Query Language – Yapısal Sorgulama Dili) kullanılmakta iken son yıllarda söz konusu yapılandırılmamış veriden oluşan büyük veri olunca ilişkisel olmayan veri tabanları öne çıkmaktadır. NoSQL (Not Only Structured Query Language - Not Only SQL – SQL ve Daha Fazlası) geleneksel veri tabanlarından farklı olarak, sabit tablo şemaları gerektirmeyen, genellikle yatay olarak ölçeklenen, bir SQL arayüzüne sahip olmayan ve açık kaynaklı olabilen dağıtık bir veri tabanıdır.⁶⁵ Bu özelliğiyle farklı büyüklük, tip ve karakterdeki verinin depolanmasına ve gerektiğinde kolaylıkla aranmasına olanak sağlamaktadır. NoSQL veri depolama sistemleri hız ve boyut açısından yüksek performans sunmak ve fiyat uygunluğuyla yüksek kullanılabilirlik sağlamak amacıyla oluşturulmuştur.⁶⁶ Bu amacına geniş ölçeklenebilirlik ve kullanım kolaylığı gibi nedenlerle büyük verinin saklanması ve yönetiminde yaygın hale gelerek ulaşmaktadır.

Nesnelerin İnterneti

Nesnelerin İnterneti (Internet of Things - IoT); her şey için ve nesneler için internet veya endüstriyel internet olarak tanımlanabilen, makineler ve cihazların küresel bir ağ içinde birbirleriyle iletişim ve etkileşim yeteneğine sahip olabildiğini öngören teknolojidir.⁶⁷ Teknoloji araştırmaları yapan Gartner şirketinin çalışmalarına göre, internete ve dolayısıyla birbirlerine bağlanabilen akıllı nesnelerin sayısının 2009 yılında 0.9 milyardan daha fazla olduğu, 2015 yılında 4.9 milyar, 2017 yılında 8.4 milyar olduğu ve bu rakamlardan yola çıkarak 2020 yılına kadar ise dünya çapında 20.4 milyar birime ulaşacağı tahmin edilmektedir. Ayrıca 2017’ de bu konudaki toplam harcamanın

⁶⁴ Gürsakal, 2017, s.220.

⁶⁵ Bogdan G. Tudorica ve Cristian Bucur, A comparison between several NoSQL databases with comments and notes, RoEduNet International Conference 10th Edition: Networking in Education and Research, Iasi. pp. 1-5, 2011, s.2.

⁶⁶ List of NoSQL Databases, <http://nosql-database.org> (8 Mart 2018).

⁶⁷ In Lee and Kyoochun Lee, Kelley School of Business Indiana University, Science Direct, Published by Elsevier Inc, All rights reserved, Business Horizons, 58. 431-440, 2015, s.1.

2 trilyon dolara ulaştığı tahmin edilmektedir.⁶⁸ Bu kadar çok sayıda nesnenin ürettiği ve gelecekte üreteceği veri sayısının oldukça yüksek miktarda olacağı açıktır.

Bulut Bilişim (Cloud Computing)

Sanal ve tek bir sunucuda yani bulutta toplanan bilgi, belge, sunum ve dosyaların kısaca sahip olunan tüm uygulama, program ve verilerin depolanması ve internetin var olduğu her yerden kolayca ulaşım sağlanabilmesi hizmetler bütününe Bulut Bilişim veya Bulut Teknoloji denilmektedir.⁶⁹ Burada “bulut” sözcüğü, internet kaynakları kümesi anlamına gelmektedir.⁷⁰ Bulut bilişim, bilişim teknolojilerinin sağladığı olanakların dağıtıldığı ve gerçek zamanlı internet teknolojilerinin bir servis olarak kullanıldığı bir çeşit programlamadır. Bu sayede şirketler veya kişiler; diledikleri yerden, servis aldıkları servis sağlayıcı üzerinden, ofislerine, kurumsal ya da kişisel bilgilerine ulaşabilmektedirler.⁷¹ Bu işlemler için, bir internet ağı aracılığıyla çoklu sunucu bağlantıları sağlanmaktadır.

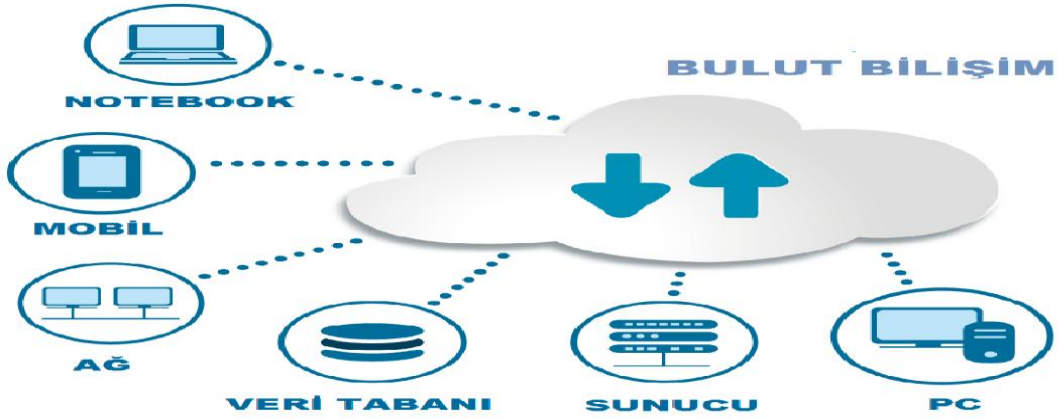
Özetle, depolanan bilgilerin ve verilerin internet ortamında saklanması işlemine verilen isim bulut bilişimdir. Başka bir ifadeyle, eskiden fiziki aletler olan hard disklerde depolanan verilerin günümüzde internet ortamındaki sanal sunucularda saklanması işlemi olarak da adlandırılabilir. Bu açıdan bulut bilişim, büyük ölçekli sanal bir hard diske benzetilebilir.

⁶⁸Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016, <https://www.gartner.com/newsroom/id/3598917> (4 Nisan 2018).

⁶⁹ Bulut Bilişim Nedir?, <http://www.yazilimnet.com/tr/blog/38/bulut-bilisim-nedir-> (8 Nisan 2018).

⁷⁰ E. B. Dudin ve Y. G. Smetanin, A Review Of Cloud Computing, Scientific and Technical Information Processing, 38(4), 280–284, 2011, s.280.

⁷¹Türkiye'nin Dijital Ekonomiye dönüşümü; Bulut ve Veri, http://www.tubisad.org.tr/Tr/Library/Analizler/bulut_bilisim_dosyasi.pdf (10 Nisan 2018).



Şekil 5: Bulut Bilişim

Kaynak: <http://www.yazilimnet.com/tr/bulut-bilisim> (8 Nisan 2018)

Şekil 5’te görüldüğü gibi, bulut bilişim dosya saklama ve dosyalara ulaşım gibi problemlerin ortadan kalkmasını sağlayarak bize bulut ile daha fazla depolama alanı, hızlı veri transferi, maliyet tasarrufu yapabilme gibi olanaklar tanımaktadır. Bulut bilişimin faydaları olarak; işgücünden ve zamandan tasarruf etmeyi sağlaması, maliyetleri düşürmesi, altyapı karmaşasını ortadan kaldırması, çalışma alanını genişletmesi, verileri korumasıyla dosya ve veri kayıplarının ortadan kalması ve verilerin güvende olmasına olanak sunması, kişisel bilgisayarların yükünü azaltması, istenilen zamanda ucuz pahalı fark etmeksizin internete bağlanabilen herhangi bir cihazla bilgiye, kişisel veriye ulaşma imkanı vermesi gibi nedenler sayılabilmektedir.⁷²

İlk gerçek bulut bilişim hizmeti olan Amazon S3, internet devi Amazon’un 2006 yılında veri merkezlerini modernize edip bulut bilişimin gelişmesinde anahtar bir rol oynamasıyla gerçek hayatta hizmete girmiştir.⁷³ Ardından Google Drive, Dropbox, SkyDrive, iCloud, Yandex, Disk, Microsoft Bulut gibi hizmetler de kurulan diğer bulut bilişimin ve depolama hizmetlerine örnek olarak verilebilir. Bulut teknolojisi büyük şirketler veya üniversiteler gibi kuruluşlar tarafından kurulup paylaşılmaktadır. Bulut bilişim başta büyük şirketler olmak üzere her kurum ve artık neredeyse herkes tarafından verilerin depolanması için kullanılmaktadır. Örneğin günümüzde hemen

⁷² Neden Bulut Bilişim?, <http://www.dia.com.tr/neden-bulut-bilisim-farki-ne-ayricaliği-ne/> (10 Mayıs 2018).

⁷³ Bulut Bilişim (Cloud Computing) Nedir?, <http://www.endustri40.com/bulut-bilisim-cloud-computing-nedir/> (10 Mayıs 2018).

herkes tarafından en az birisi kullanılan sosyal ağlardaki birçok veri bu sitelerin kendi bulutlarında depolanmaktadır. Bu bağlamda günlük hayatta bile kişiler tarafından dolaylı yollardan da olsa bulut bilişim kullanılmaktadır.

Bir bulut bilişimi kurmak donanımsal gereksinimlerin karşılanması açısından maliyetli olduğu için bu hizmetten kiralama şeklinde yararlanılması da son zamanlarda oldukça artmakta olan bir eylemdir. Firma ve kişilere esnek imkanlar sağlayan bulut bilişim ücretli veya ücretsiz ve bilgilerin herkese açık ya da yalnızca yetkili kişilerin ulaşabildikleri yapıya sahip olması seçenekleriyle hayatı kolaylaştıran teknolojiler arasında yerini almaktadır.⁷⁴

Doğal Dil İşleme

Doğal Dil İşleme (Natural Language Processing - NLP), büyük verinin analiz edilmesinde kullanılan yöntemlerden birisidir. Çok büyük metin verilerine sahip olan kurum ve kuruluşların bu verilerinin desenini belirleme, anlamlı yeni bilgi ve öngörüler çıkarma gereği duymaları sonucunda gelişen bu yöntem, makineler tarafından dil kurallarının yapısının çözümlenerek anlaşılması ve yeniden üretilmesine dayanmaktadır.⁷⁵ Örneğin, yazım esnasında kelime hatalarını düzelten ve doğal dilin yapısını taşıyan bir imla düzeltme aracı, bir doğal dil işleme yazılımıdır. Benzer şekilde, komut anlama ve otomatik konuşma gibi doğal dil ile ilgili eylemleri gerçekleştiren yazılımlar da bu uygulamanın bir sonucudur. Bu yazılımlarla kendi dilini makinelere öğreten insanoğlu ile makineler arasındaki dil engeli de gittikçe ortadan kalkacağı benziyor.

Makine Öğrenmesi (Machine Learning)

İnsanlar verilerdeki belli bir faaliyet örüntüsünün bir olayın gerçekleşme olasılığıyla ilişkisini fark edemeyebilmektedir ancak bir ilişki varsa bilgisayarlarda oluşturulan algoritmalar bunu yakalamaktadır.⁷⁶ Bilgisayarlar sadece kodlama dilinde kendilerine söyleneni yapmaktadırlar. Bilgisayarlara takip edecekleri algoritmaları

⁷⁴ *Neden Bulut Bilişim?*, <http://www.dia.com.tr/neden-bulut-bilisim-farki-ne-ayricaliği-ne/> (11 Mayıs 2018).

⁷⁵ Doğan ve Arslantekin, s.28.

⁷⁶ Bernard Marr, *Büyük Veri İş Başında: 45 yıldız Şirket Büyük Veriyi Nasıl Kullandı?* Birinci Basım, İstanbul: MediaCat Kitapları, 2017, s.5.

vererek onlara ne yapacaklarını öğretmek mümkün olmaktadır. Bir veri kümesine en iyi uyan modeli hızla bulmak için kullanılan yazılımlar makine öğrenmesi olarak ifade edilmektedir.⁷⁷ Makine öğrenmesi tekniği büyük veri sayesinde daha çok gelişmekte ve ihtiyaçlara daha fazla karşılık vermektedir.

Makine öğrenmesi, verilen bir problemi ortamdan edindiği bilgiye göre modelleyen Yapay Zeka disiplininin bir alt dalıdır. Makine öğrenmesi yöntemleri, denetimli ve denetimsiz öğrenme tekniklerinden oluşmaktadır. Denetimli (öğreticili - gözetimli) öğrenme, önceden gözlemlenmiş ve sonuçları bilinen bir başka ifade ile etiketlenmiş verileri kullanarak bu verileri ve sonuçları kapsayan bir fonksiyon oluşturmayı amaçlayan makine öğrenimi metodudur. Örneğin, Yapay Sinir Ağları insan sinir sistemini taklit ederek öğrenmeyi hedefleyen denetimli bir makine öğrenmesi yöntemidir. Denetimsiz (öğreticisiz - gözetimsiz) öğrenme ise etiketlenmemiş verideki gizli yapıyı bulma, gözle görülemeyen bağıntıları ortaya çıkarma işlemidir.⁷⁸

Veri Madenciliği (Data Mining)

Bilişim teknolojisi ve dolayısıyla veri toplama kaynakları oldukça gelişmiş ve yaygınlaşmıştır. Bu sayede her kurumun elinde çok ciddi büyüklüklerde veritabanları oluşmaya başlamıştır. Bilgisayar sistemleri ile kaydedilen veriler tek başlarına ve çıplak gözle bakıldığında herhangi bir anlam ifade etmemektedir. Büyük miktarlardaki verinin (veri madeni) içindeki ilişkileri inceleyerek veriler aralarındaki bağlantıyı bulmaya yardımcı olan ve veri tabanı sistemleri içerisinde gizli kalmış bilgilerin tespit edilmesini sağlayan anlamlı ve yararlı bağlantı ve kuralların bilgisayar programları ile aranmasına veri madenciliği denilmektedir.⁷⁹ Buradan yola çıkarak veri madenciliğinin bir veri analiz tekniği olduğu ve elde edilen sonuçların gelecekle ilgili tahminlerde kullanılmakta olduğu söylenebilir. Veri madenciliği büyük miktarda veri inceleme amacı üzerine kurulu olduğundan dolayı veri tabanları ile yakından ilişkilidir.⁸⁰ Birden

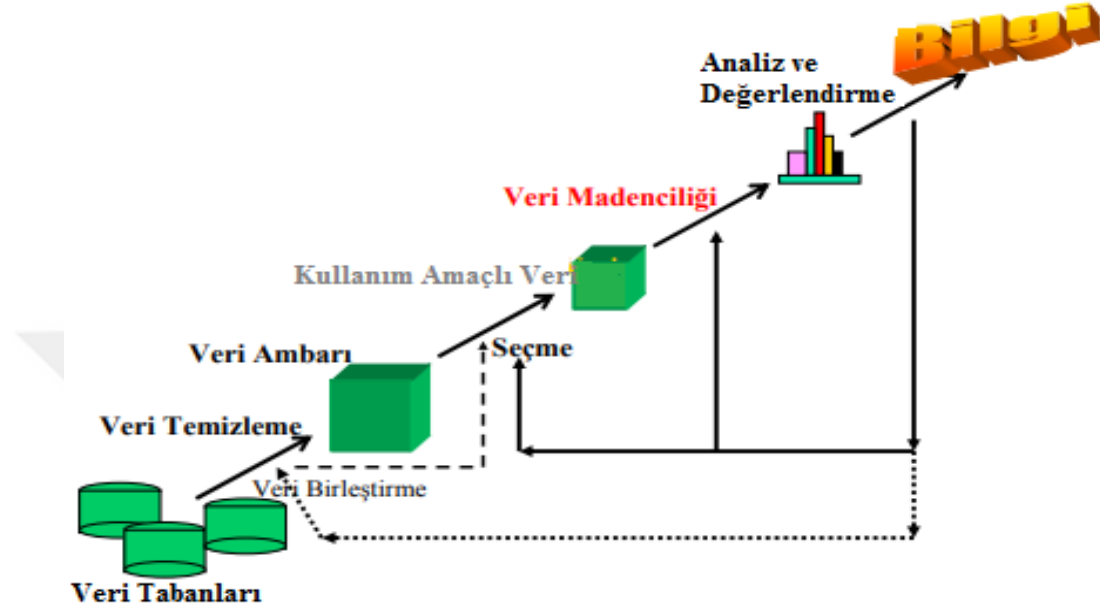
⁷⁷ Davenport, s.5.

⁷⁸ Hatice Nizam ve S. Sıla Akın, Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması, XIX. Türkiye'de İnternet Konferansı Bidiri Kitabı, Kasım 2014, s.130.

⁷⁹ Anarbak Kalkıov, “Veri Madenciliği ve Bir E-Ticaret Uygulaması”, (Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, 2006), s.8.

⁸⁰ Abdullah Baykal, Veri Madenciliği Uygulama Alanları. D.Ü. Ziya Gökalp Eğitim Fakültesi Dergisi, 7, 95-107, 2006, s.2.

fazla veri tabanının birleştirilmesiyle oluşan yapılara veri ambarı adı verilmektedir. Kısaca veri madenciliği, veri yığımından nitelikli bilgi elde edebilme sürecinde kullanılan tekniklere denilmektedir. Bu durum aşağıdaki şekilde görülmektedir.



Şekil 6: Veriden Bilgi Elde Edilmesi Süreci ve Veri Madenciliği

Kaynak: <http://web.itu.edu.tr/~sgunduz/courses/verimaden/slides/d1.pdf> (15 Mart 2018).

2.2.4 Büyük Veri ve İşletmeler

Veri konusunda başta iş ve bilişim dünyası olmak üzere toplumun tüm kesimine etki eden bir değişim sürecine şahit olunmaktadır. Her şey hakkında artık daha fazla veriye sahip olan şirketler depolama ve analiz etme becerilerini geliştirerek bunu kendi gelişimleri adına bir fırsata çevirmektedirler. Büyük verinin üretimi arttıkça değeri de artmakta ve nitelikli büyük veri tüketimi ortaya çıkmaktadır.

Büyük veri üreten ve kullanan şirketlerden bazılarına ilişkin örneklere aşağıda yer verilmektedir:⁸¹

⁸¹ Marr, s.14,21,79,149,255,261,287.

- Dünyanın en büyük fiziki ürün, e-kitap gibi sanal ürün, video ve web hizmetleri satıcısı olan Amazon, büyük veriyi kullanarak ne istediğinizi ve onu ne zaman istediğinizi tahmin etmeyi hedefleyen tahmin motorları ve öneri sistemleri geliştirmektedir. Büyük veriyle ağdaki ilişkilerin yapısını çözüp, ortak filtrelemeler sonucu müşterilerinin satın alma potansiyelini arttırarak kendine değer katmaktadır.
- Teknoloji devi Apple'ın cihazları milyonlarca kişi tarafından kullanılmaktadır. Apple, kullanıcı dostu arayüzlerinin yanı sıra kullanıcı verileri incelenerek geliştirilen uygulamalarla çok çeşitli veriler toplamaktadır. Akıllı telefonların popüler hale gelmesi için tasarlanan günlük uygulamalar (haritalar, navigasyon ve konum bilgisi, ses tanıma vb), kullanıcı verilerinin izlenmesine ve paylaşılmasına ve de bilgilerin kendi veri merkezlerinde depolanarak analiz edilmesine dayanmaktadır.
- Dünyanın en büyük medya kuruluşlarından biri olan BBC, üretiminin büyük bölümünü artık online hizmetler aracılığı ile yapmaktadır. Dijital ortamda daha fazla veri üretmekte ve toplamakta, bunları da izleyicilerine sundukları hizmetleri uyarlamak için kullanmaktadır.
- Avrupa Nükleer Araştırma Merkezi, CERN'de, Büyük Hadron Çarpıştırıcısı deneyini yürüten uluslararası bir bilimsel kurumdur. Bu deney, saniyede 40 TB veri üretilmektedir.⁸² CERN deneyleri evrenin işleyişiyle ilgili sırları büyük veri ile çözmeye çalışırken oldukça yüksek miktarlarda veri üretmektedirler. İçinde bulunduğumuz büyük veri dönemini başlatan internetin gelişiminde ana etken olan Dünya Çapında Ağ (World Wide Web)'in oluşturulması da, 1990'larda araştırmacılar arasında iletişim sağlamak amacıyla CERN'de gerçekleşmiştir.
- Facebook, Twitter ve LinkedIn gibi sosyal medya devleri, arkadaşlarla bağlantıda kalmak ve yeni bağlantılar elde etmek, özel olayları ve anları paylaşmak, sosyal etkinlikleri organize etmek, haberleri okumak, oyun oynamak

⁸² Kenneth Cukier, Data, Data Everywhere: A Special Report on Managing Informatio, The Economist Newspaper. pp 1-13, 2012, s.5.

gibi çeşitli amaçlarla hemen herkes tarafından kullanılmaktadırlar. Bu esnada bu hesaplar da kullanıcıları hakkında geniş çaplı ve kapsamlı kişisel bilgi toplamakta, beğenilip paylaşılan içerikleri analiz etmektedir. Bu şirketler büyük veriyi müşterilerini anlamak ve sosyal medyadaki görünürlüğünü arttırmak için kullanılmaktadırlar. Böylece anlaştıkları firmaların reklamlarını kendileri ile ilgilenebilecek potansiyel müşterilere göstererek gelir elde etmektedirler.

- Arama motoru Google'da arama yaptığımızda büyük veriyi kullanılmaktadır. Google aradığımız konu ile ilgili bir şeyler bulmamızı kolaylaştırmak için internetteki tüm bilgileri bir araya getirmeye çalışmaktadır. Google'ın bulabildiği tüm web sayfaların arşivinden oluşan endeksin boyutunun yaklaşık olarak 100 milyon gigabayt olduğu tahmin edilmektedir.⁸³ İnternetin büyük ve geniş bir alana yayılması ile bilgiler dünyanın herhangi bir yerinde bulunabilen sunuculara yükleniyor ve bu da kendilerine sunulan verilere göz atan kişilerin kimi zaman birbirinden binlerce kilometre uzakta olan bilgisayarlara bağlandıklarını göstermektedir. Google'ın büyük veri kullanım becerileri de düşünüldüğünde bu alanda öncü bir güce sahip olduğu açıktır. Burada Google, arama sonucunda sayfalarca veri verirken Sayfa Sırası (PageRank) prensibini kullanılmaktadır. Buna göre, belli bir sayfaya ne kadar çok sayfa bağlanırsa, o sayfanın sırası o kadar ilerde olmaktadır çünkü bağlanan sayfalar bir nevi o sayfaya atıf vermektedir. Başka bir ifade ile bu işlem, yapılandırılmamış veri olarak niteleyebileceğimiz web sayfası içeriklerini, bilgiyi yararlılığı açısından derecelendirmek için gereken yapılandırılmış veriye dönüştürmeyi içermektedir. Google'ın sunucuları arama taleplerini cevaplarırken ve kullanıcı profiline uygun reklamlar sunarken her gün yaklaşık 20 petabyte büyüklüğünde veri işlemektedir.
- Microsoft, makineler arası akıllı iletimler için oluşturulan Microsoft Azure ve Nesnelerin İnterneti (IoT) projelerini makine öğrenmesi adı verilen çalışma alanlarında geliştirmiştir. İnternet vasıtasıyla birbirine bağlanabilen, kendi kendine ve birbirinden öğrenebilen, olasılık ilkesi ile çalışan, doğru olma

⁸³ *Total Number of Pages Indexed by Google*, <https://www.statisticbrain.com/total-number-of-pages-indexed-by-google/> (16 Mayıs 2018).

olasılıđına göre sıralanmış en muhtemel cevaplar üzerinden hareket eden sistemlerle karşı karşıya olduğumuz bilinmektedir. Bu sistemler tarafından da günlük hayata ilişkin bir hayli veri üretilmektedir.

- Dünyanın en büyük perakendecilerinden biri olan Walmart, büyük veriyi gerçek zamanlı analizlerle süpermarket performansını arttırmak için kullanmaktadır. Şirket kendi çapında yeni veri tabanlı girişimleri arařtırmak ve uygulamak amacıyla @WalmartLabs ve Hızlı Büyük Veri Ekibi'ni kurmuřtur.
- Uber, bir yere gitmek isteyen yolcularla onları götürmek isteyebilecek sürücülerini buluřturan akıllı telefon yardımıyla kullanılan bir taksi rezervasyon hizmetidir. Şirket büyük veriyi kullanarak trafik kořulları ve yolculuk sürelerini gerçek zamanlı hesaplayan algoritmalar geliřtirmektedir.

3. BİBLİYOMETRİK ANALİZ

Bilimsel ilerleme, yapılan bilimsel yayın ve bilimsel projelerin, arařtırma geliřtirme (AR-GE) faaliyetlerine ayrılan kaynakların ve alınan patentlerin sayısı gibi bazı belli bařlı deđerlerle ölçülmektedir. Bilimde ilerlemek isteyen her toplumun bu ölçütlerin niceliđini ve niteliđini artırabilecek girişimlerde bulunması gerekmektedir. Bu süreçte en iyi yol gösterici üretilen yayınların istatistiksel yöntemlerle incelendiđi bibliyometrik arařtırmalar olmaktadır.⁸⁴ Bibliyometrik arařtırmalarda bilimsel ürünlerin niceliklerinden yola çıkarak niteliklerine ulařılabilecek pek çok etmen incelenmekte ve deđerlendirilmektedir. Bu anlamda, bibliyometrik çalıřmalar hem herhangi bir konudaki en verimli çalıřmaları ve arařtırmacıları belirlenirken hem de bunlar arasındaki etkileřimlerin boyutlarını gösterdiđi için bilim dünyasına ve ülkelere bilimsel çalıřmaların geleceđi hakkında yön veren arařtırmalardır.

⁸⁴Umut Al, ‘‘Türkiye'nin Bilimsel Yayın Politikası: Atıf Dizinlerine Dayalı Bibliyometrik Bir Yaklařım’’ (Doktora Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, 2008), s.18.

3.1 Bibliyometrik Analiz İle İlgili Kavramlar

Bibliyometri olgusu ile ilgili temel kavramlar, bibliyometri, atıf, atıf endeksi ve ortak yazarlı çalışmalar şeklinde belirtilebilir. Aşağıda bu kavramlarla ilgili bilgilere yer verilmektedir.

3.1.1 Bibliyometri

Bibliyometri kavramının geçmişi eskilere dayanmakla birlikte, başlarda istatistiksel bibliyografi olarak değerlendirilen bir kavram olmuştur. İstatistiksel bibliyografi kavramı ise ilk kez Wyndham Hulme tarafından 1922'de Cambridge Üniversitesi'nde ders verirken kullanılmıştır. Aynı zamanda Hulme bu kavramı yazılı iletişimin çeşitli yönlerini sayarak ve analiz ederek bir disiplin hakkında bilgi toplamak olarak değerlendirmiştir.⁸⁵ İstatistiksel bibliyografi ile bibliyometrinin birbirinden ayrılması ise ilk olarak Pritchard'ın çalışmasında görülmektedir. Pritchard bibliyometriyi, matematiksel ve istatistiksel tekniklerin yayınlanmış dergi, kitaplar ya da diğer bilgi paylaşım ve iletişim yöntemlerine uygulanabilmesi yöntemi olarak tanımlamıştır.⁸⁶ Pritchard'ın istatistiksel bibliyografi kavramını yeniden tanımlaması yeni bir araştırma yöntemi olan bibliyometrik analizin doğmasına olanak sağlamıştır.

Sengupta (1992), içerik olarak ilk bibliyometrik çalışmanın Campbell tarafından 1896'da yayınlanmış olan ve ele alınan yayınların konu dağılımlarını istatistiksel bir yöntemle analiz eden "*Theory of the National and International Bibliography*" isimli çalışma olduğunu belirtmektedir. Buradan bibliyometri kavramının kullanım itibariyle 1800'lerin sonuna kadar uzanmakta olduğu görülmektedir. Güncel literatürden anlaşıldığı üzere günümüzde çalışma alanlarının da çeşitlenmesi ile bibliyometri araştırmalarına olan ilgi hala ve giderek artarak devam etmektedir.

Bibliyometri kelimesinin, kitap (biblio) ve ölçüm bilimi (metric) sözcüklerinin birleşiminden oluşan ve bilimsel dergilerin, makalelerin ve araştırma kuruluşlarının

⁸⁵ Wyndham E. Hulme, *Statistical Bibliography In Relation To The Growth Of Modern Civilization*, Grafton, London, 1923, s.9.

⁸⁶ Alan Pritchard, *Statistical Bibliography Or Bibliometrics?* *Journal Of Documentation*, 25 (4), 348-349, 1969, s.348.

etkilerini belirlemeye yarayan bir ölçüm birimi olduğunu belirtilebilir.⁸⁷ Bibliyometri, bilimsel araştırma ve çalışmaların disiplin, konu, yazar, yayın bilgisi, atıf yapılan yazar ve atıf yapılan kaynaklar, kurum, ülke gibi verilerin istatistiksel olarak incelenmesi ile ilgilenmekte elde edilen sonuçlar belirli bir disipline ait genel bir bakış açısı vermektedir.⁸⁸ Genel olarak ise bibliyometri kavramı belli bir bilim dalı veya konuda yapılmış çalışmalar bütünü olarak değerlendirilmektedir.⁸⁹ Bibliyometri, bilimsel iletişimin geliştirilmesine dönük yöntemler içermektedir.⁹⁰

Nicholas ve Ritchie (1978) bibliyometriyi iki farklı şekilde ele almaktadır.⁹¹ İlki; makale, kitap, yazı ve bunları içeren dergilerin miktarının sayılması ve bunun ülke, zaman ve alan boyutunda yapıldığında betimleyici bibliyometri olduğu; ikincisi ise, alanyazının parçaları arasında kurulan ilişkileri ortaya çıkaran referanslar, atıf alan yazarlar gibi konularda yapıldığında değerlendirici bibliyometri olduğu şeklindedir.

Bibliyometrik araştırmalar ise bilimsel yayınların kaynakçalarında bulunan kaynakların ayrıntılı olarak incelendiği araştırmalardır.⁹² Bibliyometrik yöntemler uygulanarak belirli bir alanda, belirli bir ülkede çalışılan konu ve konu başlıkları, bu konuları çalışan yazarlar, yazarlar arası işbirliği ve ortaklık, çalışma yapılan konu başlıklarının sayıca çokluğu belirlenebilmektedir.⁹³

19. yüzyılın sonlarına doğru yayın sayılarındaki artış sonucunda, bu yayınları denetlemenin de gerekliliğinden kaynaklı, yayınları ölçmeye ve yayınlar hakkında bilgi elde etmeye yönelik olarak istatistiksel inceleme amaçlı bazı kavramlar doğmuştur. Bu

⁸⁷ Tullio Basaglia, What Is Bibliometrics And Why We Should Care About It?, CERN Library, 2014, <http://Slideplayer.Com/Slide/6213346/> (18 Mart 2018).

⁸⁸ Gülden F. Önal, Bildirilerin Bibliyometrik Profili, IX. Ulusal Müzik Eğitimi Sempozyumu, KEFAD, Cilt 18, Sayı 3. (1079-1097), 2017, s.1080.

⁸⁹ Esen Yıldırım ve Özlem Ergüt, Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu'nun Bibliyometrik Analizi, Namık Kemal Üniversitesi, Sosyal Bilimler Metinleri, 2014, s.1.

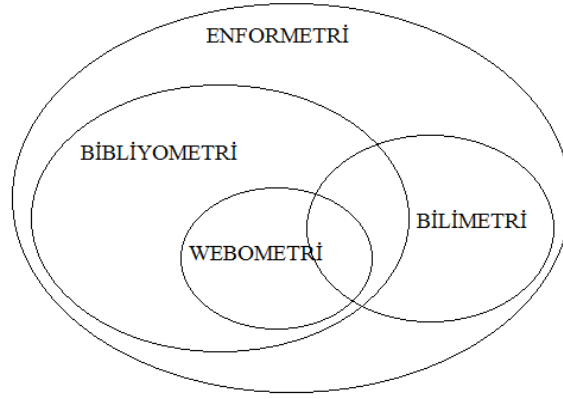
⁹⁰ Christine Borgman ve Jonathan Furner, Scholarly Communication And Bibliometrics, In B. Cronin (Ed.), Annual Review Of Information Science And Technology, Vol 36, Pp 3-72, 2002, s.2.

⁹¹ David Nicholas, ve Maureen Ritchie, Literature and Bibliometrics London: Clive Bingley, 1978.

⁹² Umut Al ve İrem Soydal, Bilgibilim Alanında Kendine Atıf Üzerine Bir Çalışma, Bilgi Dünyası. 11(2) 349-364, 2010, s.350.

⁹³ Danny P. Wallace, Bibliometrics And Citation Analysis. J. N. Olsgaard (Yay. Haz.). Principles And Applications Of Information Science: For Library Professionals İçinde (Ss.10-26), Chicago And London: American Library Association, 1989, s.10-11.

doğrultuda, Şekil 7’de kapsamaları gösterilen bibliyometri, enformetri, bilimetri ve webometri terimleri ve bunların birbirleri ile ilişkisi aşağıda açıklanmaktadır.



Şekil 7: Bibliyometrinin Kapsamı

Kaynak: <http://ww2.lib.metu.edu.tr/ek/BilgiMerkezleriVeBibliyometri.pps> (17 Nisan 2018)

Enformetri: Bilginin ölçülmesi ile bilgi teorisi oluşturmayı amaçlayan bir kavram olan enformetri, belgeler ve bibliyografyalarla beraber her türlü formdaki bilginin sayısal durumunu inceler, nicel analizini yapar.⁹⁴ Enformetrinin, bilimetri ve bibliyometriyi kapsadığı düşünülmektedir.⁹⁵

Bilimetri: Bilim dallarının analizinde, bilim sosyolojisi ve bilimin örgütlenmesi konularında bilimetriciden yararlanılırken, bilimsel yayınlara ve dağılımlarına ilişkin değerlendirmelerde bibliyometriciden yararlanılmaktadır.⁹⁶

Webometri: Webometri, enformetrik yöntemlerin web bilgi kaynaklarının yapıları ve kullanımlarının analizine uygulanmasıdır.⁹⁷ Webometri, World Wide Web (Dünya Çağında Ağ)’de belli bir konu hakkında bibliyometrik ve enformetrik yaklaşımlarla nicel analiz anlamına gelmektedir.

⁹⁴Burcu Umut Zan, “Türkiye’de Bilim Dallarında Karşılaştırmalı Bibliyometrik Analiz Çalışması” (Yayınlanmamış Doktora Tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Ankara 2012), s.72.

⁹⁵ Lennart Björneborn, “Small-World Link Structures Across An Academic Web Space: A Library And Information Science Approach, (Yayınlanmamış Doktora Tezi, Royal School Of Library And Information Science, Copenhagen, 2004), s.14.

⁹⁶ Björneborn, s.14.

⁹⁷ Al, 2008, s.25.

Bilimsel araştırma yayınlarının türlü özellikler açısından sayısal analizinin yapılıp araştırma sonuçlarının değerlendirilerek ilgili alanda geleceğe ışık tutması amacıyla kullanılan bibliyometri, analiz sonucunda bazı sorulara yanıt bulmayı beklemektedir. Bu sorular genel olarak şu şekilde belirtilebilir:⁹⁸

- Belirli bir bilim dalında yayın yapan araştırmacıların dağılımı nasıl olmaktadır?
- Belirli bir bilim dalında yıllara göre yayın sayıları nasıl değişmektedir?
- Bilim dallarına sağladıkları katkı bakımından ülkelerin durumu nedir?
- Belirli bir bilim dalına hangi yazarlar ne açıdan daha fazla katkı sağlamıştır?
- Hangi bilim alanında hangi dil daha yoğun olarak kullanılmaktadır?
- Belirli bir bilim dalında hangi belgeler ne kadar kullanılmaktadır?
- Belirli bir bilim dalında araştırma makaleleri, hangi dergilerde ne ölçüde yer almaktadır?
- Belirli bir bilim dalında kişi başına atıf sayısı dağılımı nasıl değişmektedir?

Bibliyometrik çalışmalar, bilimsel yayınların çeşitli faktörler bakımından etkinliği konusunda bilgi sahibi olunmasını sağlamaktadır. Bibliyometrik araştırmalarla bir yandan herhangi bir konudaki en verimli araştırmalar ve araştırmacılar belirlenirken, diğer yandan bunlar arasındaki etkileşim ve ağların şekli ortaya çıkarılmaktadır. Benzer bir yaklaşımla, bibliyometrik araştırmalar, çeşitli konularda ülkeler arasında, kurumlar arasında ya da kişiler ve konu başlıkları arasında karşılaştırmalar yapılmasına da olanak sağlamaktadır.⁹⁹ Bibliyometrik araştırma yapan araştırmacıların temel iki hedefi vardı. Birincisi belli bir konudaki araştırmaları belirlemek ve ikincisi bunlar arasında var olan etkileşim ve ilişkileri tespit etmektir. Literatürden de anlaşıldığı üzere özellikle gelişmiş ülkelerde, bilimsel performansların ve yakın geleceğe dönük politikaların belirlenmesinde bibliyometrik çalışmalar sıklıkla yürütülmektedir.

⁹⁸ Wallace, 1989, s.11.

⁹⁹ Wallace Koehler, Information Science As "Little Science": The Implications Of A Bibliometric Analysis Of The Journal Of The American Society For Information Science, Scientometrics, 51(1): 117-132, 2001, s.120.

Türkiye’de de bilimle ilgili ulusal politika ve girişimlerin belirlenmesinde bibliyometrik çalışmalar kaynak olarak kabul edilmektedir.¹⁰⁰ Türkiye’de veri tabanlarına erişimin yaygınlaşması sayesinde ülkenin, gerek kişi ve kurum olarak gerek genel olarak bilimsel yayın performansının ölçülmesinde, hem kurumsal hem de bireysel düzeylerde bibliyometrik analizler gerçekleştirilmektedir.

3.1.2 Atıf

Atıf konusu, bibliyometrinin ilgilendiği konuların başında gelmektedir. Atıf yapma, temelde bir düşüncenin dayanağını açıklama; başka bir deyişle atıf yapan belge ile atıf yapılan belge arasında bir ilişki kurma mantığına dayanmaktadır.¹⁰¹ Atıf ile ilgili temel varsayımda makale(1)’e, makale(2)’de atıf verildiyse, bu şu demektir: makale(2)’nin yazarı, makale(1)’in yazarından akademik düşünce olarak etkilenmiştir.¹⁰² Başka bir anlatımla, atıf verilen yazar, atıf veren yazarı akademik görüş ve düşünce anlamında etkilemiştir.

Burada değinmekte fayda olan başka bir konu ise atıflara temel oluşturan terimler olan alıntı ile referans arasındaki ayrımıdır. Şöyle ki, eğer bir makale(1) çalışması bibliyografik dipnot kullanarak başka bir makale(2) çalışmasını tanımlamakta ise, burada makale(2), makale(1)’den bir atıf almakta ve makale(1) de makale(2)’yi referans (kaynak) göstermektedir.¹⁰³ Narin (1976)’e göre ‘‘*atıf, bir belgenin başka bir belgeye verdiği bilgiyi, alıntı ise bir belgenin başka bir belgeden aldığı bilgiyi temsil eder*’’¹⁰⁴ Bu doğrultuda alıntı ve referansın (kaynak) birbiriyle oldukça ilişkili olduğu açıktır. Bilimsel bir çalışma, kendi alanında daha önce yapılan çalışma veya çalışmalarla belli bir ölçüde benzerlik göstermeli; bu benzerliklerle birlikte ilgili alana yenilikler de getirmelidir. Burada önemli olan kendi alanında daha önce yapılan çalışmalara gösterdiği benzerlikler ve alıntılar için bilimsel yayın kaynak gösterme

¹⁰⁰Ertunç Ukşul, ‘‘Türkiye’de Eğitimde Ölçme ve Değerlendirme Alanında Yapılmış Bilimsel Yayınların Sosyal Ağ Analizi ile Değerlendirilmesi: Bir Bibliyometrik Çalışma’’ (Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, 2016), s.4.

¹⁰¹Güleda Doğan, Bibliyometri ve Atıf Analizi, Araştırma Yöntemleri, 2017, http://www.acikders.net/pluginfile.php/6131/course/section/2834/hafta-11_bibliyometri-atif-analizi.pptx
<http://Slideplayer.Biz.Tr/Slide/10155906/> (2 Mart 2018).

¹⁰²Erhan Erkut, Bibliyometrik Analiz, 2013, <http://eresearch.ozyegin.edu.tr/bitstream/handle/10679/285/Bibliyometrik%20Analiz.pdf?sequence=1> (3 Mart 2018)

¹⁰³Jack Yuri Gómez-Morales, Citations and Scientific Indexing. Published Online: Wiley Online Library, 2017, s.1.

¹⁰⁴ Ukşul, s.31.

ilkesi gereğince kaynakça oluşturmak gerektiğidir. Bilimsel çalışmaların kaynakçalarında, konuyla ilgili benzerlik gösteren çalışmalar yer almakta; bu da düşüncesine başvurulana yazara ya da yayına atıf olarak değerlendirilmektedir.

Atıf yapma olgusunun çok uzun yıllardan bu yana devam etmektedir. Bilimsel çalışma yapan bilim insanlarının çalışma yapılan alanlarla ilgili daha önceki yayınlara atıf yapmasının 19. yüzyıldan itibaren süregeldiği görülmektedir.¹⁰⁵ Garfield (1965), tarafından yapılan çalışmada, atıf yapma amacının iki yayın arasında yalnızca bir ilişki kurmak olmadığı; bununla birlikte aşağıdaki gibi farklı amaçlar da içerebileceği belirtilmektedir. Bu amaçlar;

- Çalışma konusuna saygı gösterilmesi,
- Bir düşünce veya kavramla ilgili orijinal kaynağın tanıtılması,
- Daha önceki yıllarda konuyla ilgili çalışma yapmış kişilere saygı gösterilmesi,
- Konuyla ilgili olarak daha önce yapılmış yayınların okunmasının sağlanması,
- Araştırmada kullanılan verinin doğruluk ve gerçekliğinin kanıtlanması
- Araştırma yöntemine ilişkin bilgi verilmesi,
- Konu ile ilgili iddiaları doğrulanması veya başka çalışmalara yönelik eleştirilerde bulunulması,
- Önceden yapılan yeterince bilinmeyen yayınların tanıtılması, şeklindedir.

Atıf konusunda ele alınan farklı bir boyut ise kendine atıftır. Kendine atıf birkaç türden oluşmaktadır. Bunlar içerisinde en çok üzerinde durulan yazarın kendine atıfı olmaktadır. Diğer kendine atıf türleri ise, dergi kendine atıfı, kurum kendine atıfı hatta ülke kendine atıfıdır.¹⁰⁶ Yazarın yeni çalışmasında daha önceden yayımlanmış olan eski çalışmalarından alıntı yapması, yazar kendine atıfı; bir dergide yayımlanan makalelerin, o dergideki diğer yayımlanmış makalelerden alıntı yapması derginin kendine atıfı; bir kurum tarafından yapılan araştırmalar sonucunda yayımlanan

¹⁰⁵ Leo Egghe ve Ronald Rousseau, Introduction To Informetrics: Quantitative Methods In Library Documentation And Information Science, Amsterdam: Elsevier Science Publishers, 1990, s.204.

¹⁰⁶ Al ve Soydal, s.355.

çalışmaların yine aynı kurum tarafından daha önceden yayımlanmış çalışmalardan alıntı yapması ise kurum kendine atfı olarak açıklanabilir.

3.1.3 Etki Faktörü ve Anındalık Endeksi

Etki faktörü, bir dergide son iki yılda yayımlanan makale sayısına, yine bu makalelere bulunulan yıl içerisinde yapılan toplam atıf sayısının oranı dikkate alınarak elde edilen bir değerdir.¹⁰⁷ Bu açıklamadan etki faktörünün, aldığı atıflardan yola çıkarak dergileri değerlendirilmede kullanıldığı açıktır. Dergiler arasında göreceli bir önem sırası oluşturan en bilinen değerlendirme ölçütüdür.¹⁰⁸

Garfield (1976)'a göre anındalık endeksi bir dergide bir yılda yayımlanan yayınlara o yıl içerisinde yapılan atıf sayısının, o yıl ki toplam yayın sayısına bölünmesi ile elde edilen bir değerdir. Etki faktörü ve anındalık endeksi dergilerin bilimsel etkinliğini değerlendiren birer kalite değeri olarak görülmektedirler.¹⁰⁹

3.1.4 Atıf Endeksleri

Atıf endeksleme (citation indexing) kavramı, 1800'lü yılların sonlarından itibaren görülmeye başlanmıştır. Günümüzde akademik olarak kabul gören kriter, bilimsel çalışmalara yapılan atıf sayıdır. Atıf sayısı, yayının ne kadar önemli bilgi içerdiğinin ve kalitesinin bir göstergesi olarak görülmektedir. Atıf endeksleri literatürde atıf dizinleme görevini yapmakta olan araçlar olarak karşımıza çıkmaktadır.

Atıfları bir değerlendirme kriteri olarak kabul edip yayınları atıf sayılarına göre sınıflandıran bir kuruluş olan Bilimsel Enformasyon Entitüsü (Institute for Scientific Information - ISI) tarafından, uluslararası bir bilim endeksi olan The Science Citation Index (SCI) uygulamaya konulmuştur. Amerikalı bir kütüphaneci olan Eugene Garfield (1960), SCI başta olmak üzere diğer atıf endeksleri Current Contents, Journal Citation Reports ve Index Chemicus'un da bulucusudur.¹¹⁰ Yaygın olarak kullanılan ve ISI

¹⁰⁷ Al, 2008, s.24.

¹⁰⁸ Doğan, 2017, s.44.

¹⁰⁹ Al, 2008, s.25.

¹¹⁰ Nuri Özgirgin, Özgirgin, Uluslararası İndeksler Neden Önemli? Sağlık Bilimlerinde Süreli Yayıncılık, (27-43), 2010 <http://Uvt.Ulakbim.Gov.Tr/Toplantı/Uay10/Nozgirgin.Pdf> (18 Mart 2018).

tarafından sağlanan uluslararası bilimsel atıf endeksleri; fen bilimleri için, Science Citation Index (SCI), sosyal bilimler için, Social Science Citation Index (SSCI) ve sanat ve beşeri bilimler için, Art & Humanities Citation Index (AHCI)'dir. ISI tarafından sağlanan Web of Sciences; Elsevier ürünü olan Scopus; Google Scholar ve Tübitak destekli olan ULAKBİM gibi bilimsel yayın veri tabanları da atıfların internet üzerinden taranarak otomatik olarak dizinlenmesine olarak sağlamaktadırlar.

Referans gösterme (kaynakça), düzenlenmiş ve yayınlanmış metinlerin sonunda bulunan teknik bir standarttır ve okuyucu tarafından rahatlıkla görülebilmektedir. Ancak alıntılar okuyucu tarafından doğrudan görülemezler. Bu durumun sebebi, referanslar yayınlanan metnin sonunda bir kısımda yer alırken; alıntılarla yayımlanmış literatürün içinde karşılaşıyor olmasıdır. Alıntılar, literatürün bibliyometrik analizi yapıldığında okuyucu için görünür hale gelmektedir. Bir atıf endeksinin (alıntı endeksi/alıntı dizini) yapmaya çalıştığı şey de tam olarak budur.¹¹¹ Buradan anlaşılacağı üzere, atıf endeksleri, bibliyografik referansların düzenlenmesi ve anlamlı bir bütün halinde gösterilmesi için kullanılmaktadır.

Ülkelerin bilim alanında dünya ülkeleri arasında konumlandırılmasında, ülkelerin ya da üniversitelerin bilimsel niteliklerinin karşılaştırılmasında ve bilim insanlarının akademik performanslarının değerlendirilmesinde üç ölçüt dikkate alınmaktadır. Bu ölçütler sırasıyla, uluslararası bilimsel dergilerde yayınlanan yayın sayısı, yayınların bilim endekslerince taranan bilimsel dergilerde yayınlanması ve yayınlara yapılan atıfların sayılarıdır.¹¹² Buradan da anlaşılacağı gibi atıf endeksleri herhangi bir konudaki bilimsel gelişmeyi de takip etmede bir araç olarak kullanılmaktadır. Burada bahsedilen sayıların belirlenmesinin de bibliyometri ile yapılacağı açıktır. Yayınların bibliyometrik analiz sonucunda elde ettiği sıralamalar, ülkelerin bilimsel gelişmişliği ve üniversitelerin akademik kalitesinin bir göstergesi olarak değerlendirilmektedir.

¹¹¹ Gómez-Morales, s.1.

¹¹² Mehmet Zeki Ak ve Ahmet Gülmez, Türkiye'nin Uluslar Arası Yayın Performansının Analizi, Akademik İncelemeler Dergisi, 1 (1). 21-50, 2006, s.22.

3.1.5 Ortak Çalışmalar

Bilimsel arařtırmalar sonucunda ortaya ıkan yayınlar, tek yazarlı ve birden fazla yazarlı olmak zere ikiye ayrılmaktadır. Birden fazla yazarın katkı sađladığı alıřmalar son dnemlerde sıklıkla karřılařılan bir durumdur. Hatta bazı disiplinlerde tek yazarlı yayın sayısı nadiren grlmektedir.¹¹³ Bilimsel iletiřimde ok yazarlılıđa dođru bir geiř olduđu da aıktır. Ortak yapılan alıřmalar, belli bir konuda yođunlařılmasına neden olabilmekte; aynı zamanda daha derinlemesine bilgiler edinilebilmektedir. Artık birden fazla uzmanın grřne ihtiya duyan bilimsel alıřma konuları da ortak yazarlılıđı gerekli kılmaktadır.

Ortak alıřmaların yapılmasında birden fazla ama bulunmektedir. Bu amalar;¹¹⁴

- Daha fazla uzmanlařabilme ve iřblm,
- Daha fazla kaynađa ulařabilme,
- Finansal kaynađı daha kolay elde edebilme,
- Daha hızlı akademik geliřim sađlayabilme,
- Daha fazla bilimsel arařtırma ortaya koyabilme,
- Belirli bir konuda daha fazla merak duygusunu ortaya koyabilme

řeklinde dir.

İnternetin aracılıđı sayesinde uluslararası alanlarda farklı lkelerdeki arařtırmacıların iřbirliđi sonucunda ortak yazarlı alıřmalar yapılabilmekte ve yayınlar retilmektedir.

¹¹³ AI, 2008, s.29.

¹¹⁴Donald Beaver, Scientometrics, 52. Pp.365, 2001, <https://Doi.Org/10.1023/A:1014254214337> (19 Mart 2018).

3.2 Bibliyometrik Analiz

Belli bir alandaki yayınlar, betimleyici bibliyometri, değerlendirici bibliyometri ve atıf analizi üzere üç farklı yolla analiz edilebilir.¹¹⁵ Betimsel bibliyometri literatürün ülke, yazar, yayın yılı, konu, dil gibi bileşenlere göre gösterdiği dağılımları ortaya çıkarmayı ve tanımlayıcı istatistikleri belirlemeyi sağlamaktadır. Bilimsel verimlilik betimsel bibliyometri yoluyla değerlendirilebilir. Bibliyometrik haritalandırma tekniği ile bulgular görselleştirilerek sunulabilmektedir. Bibliyometrik yasalar genel olarak bu dağılımlar hakkında öngörude bulunmaktadır. Değerlendirici bibliyometri ise, yazarlar, yayınlar ve ülkeleri arasında var olan ilişkilerin, araştırmacıların birbirinden etkilenme ve birbirine atıf verme gibi durumlarının tespit edilerek ortaya çıkarılmasını sağlamaktadır. Yayınlar yapılan atıfların istatistiksel olarak analiz edilmesi ise atıf analizidir.¹¹⁶

3.2.1 Bibliyometrik Yasalar

Bibliyometrik yasalar; Bradford, Lotka, Price ve Zipf Yasaları olarak sıralanabilir. Bu yasalara ilişkin bilgilere aşağıda yer verilmektedir.

3.2.1.1 Bradford Yasası

Bradford Dağılım Yasası olarak da bilinen Bradford Yasası, *“belirli bir konuda yapılan yayınların dergilere göre dağılımı olarak tanımlanmaktadır.”*¹¹⁷

Bradford Yasası, Bradford'un 1934 yılında yayınladığı bir makalesine dayanmaktadır. Bu çalışmada makaleler üç grupta ele alınmıştır. Bu gruplar şu şekilde ifade edilmektedir; *“Bilimsel bir dergideki makalelerin, azalan verimlilikler kanununa göre sıralanması gerekirse, öncelikle aynı konudaki yayınlardan bir çekirdek grup oluşturulmak üzere, bu çekirdek grupla aynı sayıda makale içeren yayınlardan oluşan azalan verimlilik de bulunan gruplar oluşur.”*¹¹⁸ Bu yasaya göre ilgili alandaki tüm dergiler yayın sayılarına göre azalan oranda sıralanıp yaklaşık olarak eşit sayıda yayın

¹¹⁵Esen Yıldırım ve Özlem ErgütYıldırım, Researches on the Subject of “Violence Against Women”: A Bibliometric Analysis Since 2000. MÜ İktisadi ve İdari Bilimler Dergisi. 38(2) pp.311-333, 2016, s.313.

¹¹⁶Doğan, 2017, s.9.

¹¹⁷Eugebe Garfield, Bradford's Law And Related Statistical Patterns. Essays Of An Information Scientist, 4, 476-483, 1980, s.476.

¹¹⁸Egghe ve Rousseau, s.205.

içerecek biçimde belirli sayıda sınıflara ayrılır.¹¹⁹ Buna göre Bradford Yasası, belli bir alanda yer alan dergileri;

- Dergilerin en verimli fakat az sayıda olduğu temel grup (çekirdek sınıf),
- Dergilerin daha fazla olduğu fakat verimin azaldığı ikincil grup ve
- Çok daha fazla sayıda dergilerin olduğu fakat verimin daha da azaldığı grup,

olmak üzere üç sınıfta sınıflandırmaktadır. Bradford grupların sahip oldukları dergilerin sayılarını sırasıyla, $1:n:n^2$ ile gösterilebilecek şekilde formüle etmiştir.¹²⁰

Özetle, belli bir alanda yayınlanmış 300 makale olduğunu varsayarsak; küçük bir çekirdek grup dergi, yayımlanan makalelerin önemli bir kısım olan 100'ünü içermektedir. Daha sonra ikinci ve daha fazla sayıda dergi içeren bir grup, makalelerin diğer 100'ünü içermektedir. Son olarak, çok daha fazla sayıda dergi içeren çok daha büyük bir dergi grubu da makalelerin kalan 100 'ü içermektedir."¹²¹

3.2.1.2 Lotka Yasası

Lotka yasası: “Belirli bir alanda n sayıda yazı yazan yazarların sayısı, bir yazı yazan yazarların sayısının $1/n^2$ 'si kadardır. Yazı yazanların tümü içinde sadece bir yazı yazanların oranı ise yaklaşık %60'tır.” biçimindedir.¹²² Yani, bu yasaya göre, belirli bir alanda iki yazı yazanların sayısı, bir yazı yazanların sayısının $1/4$ ' ü; üç yazı yazanların sayısı, bir yazı yazanların sayısının $1/9$ 'u kadardır. O halde, belli bir alanda 100 tane yazardan 60'ı sadece bir yazı yazmakta, geriye kalan 40 tane yazardan 15 ($60/2^2$) tanesi iki yazı, 7 (yaklaşık $60/3^2$) tanesi üç yazı ve 4 (yaklaşık $60/4^2$) tanesi de 4 yazı yazmaktadır.

Bu yasa, Lotka tarafından 1926'da yayınlanan “Bilimsel Üretimin Frekans Dağılımı” isimli makalede açıklanmıştır. Lotka, bu çalışmasında yaptığı analizlerle belli

¹¹⁹ Virgil Diodato, Dictionary of Bibliometrics, Binghamton, NY: Haworth Press, 13904-1580, ISBN-1-56024-852-1, 1994, s.45.

¹²⁰ Garfield, Eugene, 1980, s.480.

¹²¹ Garfield, Eugene, 1980, s.481.

¹²² Alfred Lotka, The Frequency Distribution Of Scientific Productivity. Journal of the Washington Academy of Sciences Vol. 16 (12) pp. 317-323, 1926, s.319.

bir konudaki arařtırmacıların ilgili bilim alanına ne ölçüde katkı sağladığını bilimsel olarak ölçmeye çalışmış ve elde ettiği dağılımın sonucunda Lotka yasasını duyurmuştur.¹²³ Özetle bu yasanın, belirli bir konuda yapılan yayın sayısının önemli bir miktarının az sayıda yazar tarafından yapıldığını öngördüğü açıktır.

3.2.1.3 Price Yasası

Price yasası, “Belirli bir alandaki toplam yazıların yarısı, o alandaki toplam yazar sayısının karekökü kadar yazar tarafından yazılmaktadır.” ve “Toplam dergi sayısının karekökü kadar dergi, toplam makalelerin yarısını içermektedir.” biçimindedir.¹²⁴ Örneğin, bu yasaya göre, bir alanda eğer 100 yazı ve 25 yazar olduğu varsayılırsa, bu 100 yazının 50 (100/2) tanesi sadece 5 ($\sqrt{25}$) yazar tarafından üretilmiştir. Kalan 50 tanesi de kalan 20 yazar tarafından üretilmiştir.

Price, karakök yasası olarak da adlandırılan bu yasayı, etkin bilim insanlarının sayılarının tahmin edilebilmesi için oluşturmuştur. Başka bir deyişle, karekök yasası; bilimsel yazıların yarısının, tüm bilim insanlarının sayısının kareköküne eşit sayıda yazar tarafından yazılmakta olduğunu söyler.¹²⁵

Lotka ve Price yasalarına göre, bir alanda bilimsel üretkenliği yüksek olan az sayıda yazar vardır.¹²⁶ Bu doğrultuda, bir alanda yapılan çalışmaların büyük bir bölümünün bilimsel üretkenliği fazla ancak sayısı az olan bir grup yazar tarafından oluşturulduğu söylenebilir.

3.2.1.4 Zipf Yasası

Zipf'e göre bir sözcüğün uzunluğu, bu sözcüğün kullanılma derecesi ile oldukça ilişkilidir. Buna göre, bir sözcük az harften oluşuyorsa kullanım derecesi artmaktadır. Zipf, İrlandalı yazar James Joyce'un “Ulysses” başlıklı kitabında geçen 29.899 adet sözcüğü, kullanım sıklığına göre büyükten küçüğe doğru sıralamıştır. Her sözcük için 1'den 29.899'a kadar bir sıra (r) atanmıştır. Ardından her sıra sayısını

¹²³ Lotka, s.320.

¹²⁴ Derek De Sol Price, Big Science, Little Science, Columbia University, New York, 119-119, 1963, s.119.

¹²⁵ I.N. Sengupta, Bibliometrics, Informetrics, Scientometrics And Librametrics: An Overview. Libri, 42(2):75-98, 1992, s.79.

¹²⁶ Uksul, 2016, s.22.

atfettiği kelimenin kullanım sıklığı sayısı (f) ile çarpmıştır. Zipf, bu çarpım sonucunda her sözcük için c sabit değerini elde etmiştir. Buna göre Zipt yasası:

$$\log(r) + \log(f) = \log(c)$$

şeklinde formüle edilmektedir.¹²⁷

Zipf yasasıyla yapılan bir çalışmada kullanılan kelimelerin hangi sıklıkta geçtiğinin tespit edilerek, kütüphane ve bilgi biliminde dizinleme için kullanılabilir.¹²⁸ Ayrıca, çalışma başlıklarında geçen kelimelerin frekans analizleri yapılabilir, böylece bir alandaki yapılan çalışmaların hangi konularda daha fazla hangi konularda ise daha az olduğu tespit edilebilir ve eksik görülen alanlara yönelim sağlanabilir.

3.2.2 Atıf Analizi

Atıf analizi, bilimsel bir çalışmayla ilgili yapılan atıfların sayısal olarak incelenmesi olarak tanımlanabilir. Garfield'e göre atıf analizi, atıflar kullanılarak yapılan bir tür analitik araç şeklinde değerlendirilmiştir.

Atıf analizi atıf dizinleriyle ilgilidir. Atıf analizleri, tarihsel süreç içerisinde ele alındığında ilk olarak ISI tarafından 1960'lı yıllarda geliştirildiği görülmektedir. Bununla birlikte, zamanla akademik alanda araştırılan konuların artmasıyla, atıf analizlerinde kolaylık sağlayan atıf ve bilim endeksleri olarak SCI; 1961'de, SSCI; 1966'da ve A&HCI; 1976'da ortaya çıkmışlardır.¹²⁹ Adı geçen bu atıf endeksleri gibi endeksler kullanılarak çalışmalar üzerine yapılan atıflar hakkında nicel ve nitel veriler elde edebilmek için atıf analizleri yapılmaktadır.

Çalışma sahibi yazarları çalışmalarını yaparken etkilendikleri çalışma ve yazarlara gösterdikleri atıf, birbirleriyle etkileşim içinde olan yazarların durumunu ortaya koymaktadır.¹³⁰ Aynı zamanda bir yayın, yazar veya bilimsel kurum hakkında

¹²⁷ George Kingsley Zipf, Human Behavior and the Principle of least Effort, AddisonWesley, Cambridge, USA. Reprinted: Hafner, New York, USA,1949, s.11.

¹²⁸ Murat Yılmaz, "Kütüphane ve Bilgibilimi Açısından Bibliyometrinin Önemi", (Yayınlanmamış Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü,1999), s.45.

¹²⁹ Farideh Osareh, Bibliometrics, Citation Analysis and Co-citation Analysis: A Review of Literature I, Libri, 46:149-158, 1996, s.154.

¹³⁰ Ak ve Gülmez, s.23.

yapılan atıf analizi, bunların bilimsel dünyadaki etkilerini ölçmekte kullanılmaktadır. Atıf analizleri ile çok sayıda atıf alan değerli bilgiler de ortaya çıkmaktadır. Yani bu sayede, özgün ve değerli bilgilere daha kolay erişim sağlandığı düşünülebilir.

3.2.2.1 Bibliyografik Eşleştirme

Bir yayına farklı iki kaynaktan atıf yapılması bibliyografik eşleştirme şeklinde tanımlanmaktadır.¹³¹ Bir diğer ifade ile bir atfın iki farklı yayın tarafından kullanılmasıyla bibliyografik olarak eşleşme meydana gelmektedir. Bu kavram yayınların ve çalışma konularının birbiriyle bağlantı kurmasını sağlayabilecek aradaki ilişkiyi ortaya çıkarabilecek bir unsurdur.

Bibliyografik eşleştirmenin niceliksel değeri, yayınların kaynakçalarındaki benzer yayın künyeleri sayısına bağlı olarak gösterilmektedir. Örneğin, iki farklı makalenin kaynakçalarında ne kadar benzer künye varsa, bu iki makalenin eşleme gücü o kadar olarak belirlenmektedir.¹³²

3.2.2.2 Ortak Atıf

Bibliyografik eşleştirmenin aksine, ortak atıf, farklı iki yayına tek bir kaynaktan atıf verilmesi şeklinde tanımlanabilir. Yani, ortak atıf, iki farklı bibliyografik künyeye, başka bir yayın tarafından birlikte atıfta bulunulmasıdır.¹³³ Ortak atfın niceliği ise iki belgenin ortak olarak atıf verilme yoğunluğu şeklinde değerlendirilmektedir.

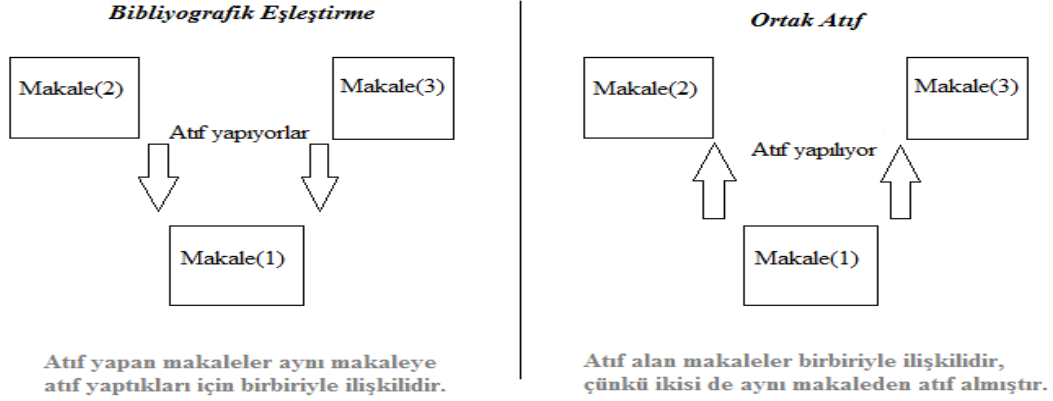
Bibliyografik eşleştirme ve ortak atıf; bibliyografik eşleştirmenin eşleşen iki kaynak arasındaki ilişkiyi, ortak atfın ise atıf yapılan dokümanlar arasındaki ilişkiyi ortaya koyuyor olmasından dolayı birbirinden farklılık göstermektedir. Bibliyografik eşleştirme ve ortak atıf yapısını basit bir şekilde göstermek gerekirse bu durum Şekil 8'de gösterilmektedir.¹³⁴

¹³¹ Umut Al ve Yaşar Tonta, Atıf Analizi: Hacettepe Üniversitesi Kütüphanecilik Bölümü Tezlerinde Atıf Yapılan Kaynaklar. Bilgi Dünyası, 5(1):19-47, 2004, s.23.

¹³² Yılmaz, s.58.

¹³³ Yılmaz, s.59.

¹³⁴ Eugene Garfield, Essays of an Information Scientist, 1985, s.20.



Şekil 8: Bibliyografik Eşleştirme ve Ortak Atıf Arasındaki Fark

3.2.3 Bibliyometrik Haritalama Yöntemi

Bibliyometrik analiz yöntemlerinden biri olan haritalama yöntemi özellikle son yıllarda bilimsel yayınlara yönelik olarak daha fazla kullanılmaktadır. Bilimsel haritalama yöntemi, veri tabanlarından elde edilen yayın bilgilerine ait veri kümelerinin oluşturulması, bu bibliyometrik verinin analiz edilmesi, mevcut ağ yapılarının haritalandırılarak oluşturulması ve sonuçların uzman kişiler tarafından değerlendirilmesi şeklinde ifade edilmektedir.^{135 136}

Analiz sonuçlarını görselleştirerek sunduğu avantajlar açısından önem arz eden bibliyometrik bilimsel haritalandırma yöntemi, birçok farklı yazılım tarafından gerçekleştirilebilmektedir. Özellikle bibliyometrik analizle elde edilen verilere SAA uygulamak için kullanılan Pajek, CiteSpace, VOSViewer, Netdraw, Publish or Perish, Science of Science, UCINET, VIVO, HistCite gibi yazılımlardan elde edilen görsel çıktılar oldukça işe yaramaktadır.¹³⁷ Her bir programın farklı görsel avantajlar sunması hepsinin de tercih edilme olasılığını neredeyse eşit kılmaktadır.¹³⁸

¹³⁵Cobo López-Herrera, E. Herrera-Viedma ve F. Herrera, Science Mapping Software Tools: Review, Analysis, And Cooperative Study Among Tools. Journal Of The American Society For Information Science And Technology 62(7) pp.1382-402, 2011, s.385.

¹³⁶Levent Kıdak ve Diğerleri, Bilimsel Haritalama Yöntemiyle Hastane Yönetimi Alanındaki Tematik Değişimin Analizi, Genel Tıp Dergisi 2017. 27(2):43-50, 2017, s.45.

¹³⁷ Yıldırım ve Ergüt, 2014, s.3.

¹³⁸ Cobo vd., s.386.

4. SOSYAL AĞ ANALİZİ

İnternetin yaygınlaşması ve mobil cihazlarla internete erişimin kolaylaşması insanları birbirine gözle görünmeyen ağlarla daha hızlı ve daha fazla bağlamaktadır. Artık sürekli bir ağa bağlanılır durumuna gelinmiştir. Kurulan bu bağlantılar oluşan sosyal ağlarda geniş çaplı bir yayılım yaratmıştır. Bu bağlantılar insanların davranışlarını birçok yönden etkileyebilmektedir. İnsanlar arasında meydana gelen sosyal ağların pek çok oluşum sebebi ve şekli bulunmaktadır. Bunlardan internet ağının oluşturduğu sosyal medya ve YouTube gibi kullanıcı içerikli topluluklarla oluşan sosyal ağlar sonucunda insanlar arasında küresel düzeyde işbirliği söz konusu olmaktadır. Ağların insanları sosyal davranış, sosyal pozisyon, sosyal ilişkiler ve sosyal bağlantılar yönünden etkilediği açıktır.

4.1 Ağ ve Özellikleri

Ağlar, ‘‘sistemi oluşturan parçaların karşılıklı etkileşimi ile beliren karmaşık sistemleri göstermenin paradigmatik bir yolu’’ olarak tanımlanmaktadır.¹³⁹ Ağ kavramının kökleri antropoloji, sosyoloji ve topolojiye dayanmakla birlikte, son dönemlerde sosyal ağların yapılarını analiz edebilmek için geliştirilmiş olan tekniklerle birçok farklı disiplinde de bu yaklaşımın ilgi çekmesini sağlamıştır.¹⁴⁰ Ağ Bilimi ise bağlantıları inceleyen yeni bir bilim dalı olarak sayılmakta, bilimsel yöntem ve araştırmalara dayanarak elde edilen ağlara ilişkin ağ davranışlarının ortak ilkelerini araştırarak organize bilgi elde etme şeklinde tanımlanmaktadır.¹⁴¹ Ağ analizinin alt dalları arasında insanlar ve onların aralarındaki akrabalık, iş ortaklığı, arkadaşlık gibi sosyal bağlardan oluşan ağları niteleyen, ilişki biçimlerini ortaya koyan sosyal ağ analizi yer almaktadır. Bununla beraber birbiri ile herhangi bir şekilde ilişki içinde olan aktörler de sosyal ağları ortaya çıkarmaktadır.

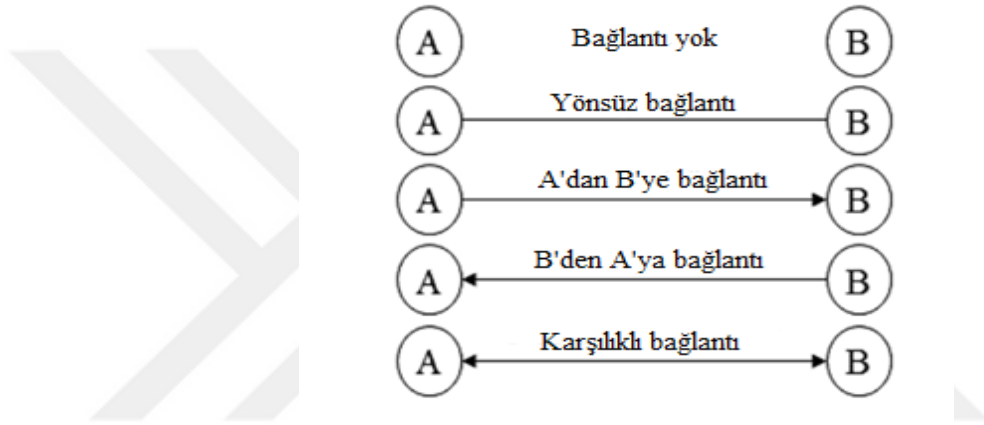
Ağ yapısını anlayabilmek için çok sayıda parçadan oluşan bu ağların bileşenleri arasındaki karşılıklı etkileşimlerin araştırılması, karmaşıklığının anlaşılması

¹³⁹ Zoltan Toroczkai, Complex Networks The Challenge of Interaction Topology, Los Alamos Science Number 29, 2005, s.40.

¹⁴⁰ Elizabeht Bott, Family and Social Networks, London:Tavistock, 1957, s.60.

¹⁴¹ Necmi Gürsakal, Sosyal Ağ Analizi, 1.Baskı, Bursa: Dora Yayıncılık, 2009, s.57.

ve sayısallaştırılması gerekmektedir. Bu ağları daha açık, anlaşılır ve sade hale getirmek için görselleştirme tekniklerini kullanmak, başvuru bir ağ analizi yöntemidir. Ağlar, bir nevi karmaşıklığın görselleştirilmesi olarak ifade edilmektedir.¹⁴² Bilgisayarlar aracılığı ile oluşturulan bu çizimlerden sonra ağın istatistiksel özellikleri belirlenerek diğer ağlarla karşılaştırmalar yapılabilmektedir. Ağlar ilişkileri daha yakından anlamayı ve analiz etmeyi sağlamaktadır. Bilişim teknolojileri, matematik ve istatistikten destek alan sosyal ağ analizi de bu amacı gütmektedir. Bireyler veya gruplara yönelik çıkarımlar yaparak bilgi elde etmek amacıyla sosyal ağ analizi kullanılmaktadır.^{143 144}



Şekil 9 : İki Düğüm Arası Bağlantılar

Bir ağda temel bileşen aktör (birim, düğüm, node, vertex, ego)dür. Aynı zamanda kişilerin sosyal ağdaki temsilcileri aktördür. Aktörler arasındaki ilişki ve bağlantıları gösteren sosyal birliktelikleri sağlayan ise bağ/bağlantılar (edge, kenar)dır. Şekil 9'da yer aldığı gibi A ve B gibi iki aktör arasındaki bağlantı ya yoktur ya yönsüz olarak ya tek yönlü olarak ya da çift yönlü olarak vardır. Bağlantı yönünden ötürü A'dan B' ye veya B'den A'ya ya da karşılıklı olarak birbirileri ile iletişim durumundan söz edilebilir.

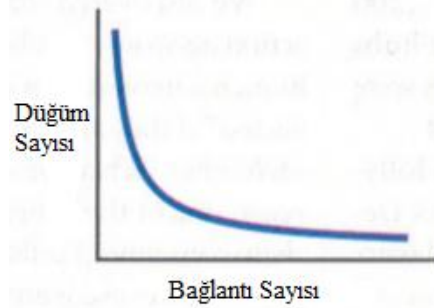
¹⁴² Gürsakar, 2009, s.40.

¹⁴³ Stanley Wasserman, ve Katherine Faust, Social Network Analysis: Methods and Applications. Press 4. Cambridge: Cambridge University, 1994, s.35.

¹⁴⁴ Linton C. Freeman, The Development of Social Network Analysis: A Study in the Sociology of Science, Vancouver, BC: Empirical Press, 2004, s.14.

Ağdaki toplam bağlantı sayısı maksimum bağlantı sayısına yakınsa bu ağa yoğun ağ; az sayıda bağlantı içeren ağlara ise seyrek ağ adı verilmektedir. Bir ağın yoğunluğunu formüle etmek gerekirse, bu ağda bulunan bağlantı sayısının o ağın bir tam ağ olması durumunda sahip olacağı maksimum bağlantı sayısına oranı şeklinde ifade edilebilir. Bu değer, ağda gerçekleşen bağlantı miktarının yüzde olarak oranını verilmektedir. Herhangi iki düğüm arasındaki en kısa yola en kısa patika denilmektedir ve en kısa patikalar üstünde başka düğümler de yer alabilmektedir. Bir ağın yarıçapı ise, ağdaki bütün ikililer arasındaki en kısa patikaların en uzun olanıdır. Yarıçapın uzunluğu, ağdaki bilgi yayılımının hızını ve süresini etkilemektedir.¹⁴⁵

Bir ağda düğümlerin sahip oldukları bağlantı sayıları derece dağılımı adı verilmekte olan bir dağılıma sahiptir. Rassal bir ağda derece sayıları birbirine yakın olup ağın derece dağılımı normaldir. Bağlantılar birbirinden bağımsız olarak p olasılığı ile gerçekleşir. Diğer bir derece dağılımı türü ise “kuvvet yasası” dağılımıdır. Şekil 10’da görüldüğü gibi bu dağılımda ise, az sayıda bağlantıya sahip çok sayıda düğüm ve çok sayıda bağlantıya sahip az sayıda düğüm bulunur.



Şekil 10: Kuvvet Yasası Dağılımı

Dolayısıyla bu dağılımda büyük değerlerden ötürü kalın bir kuyruk mevcuttur. Ağların derece dağılımları karşılaştırılırken düğüm sayısından kaynaklanan farklılığı gidermek için yine normalize edilmiş derece dağılımlarını dikkate almak gerekir. Ağdaki bağlantıların yönlerine göre girdi (gelen doğrular) ve çıktı (giden doğrular) derece dağılımları ve bunların toplamından da ağ için “bütün derece dağılımı” elde edilmektedir.

¹⁴⁵ Gürsakal, 2009, s.43.

Ağ analizinde aykırı/sapan değerleri gösteren giden bağlantıları çok olan düğümlere hub, gelen bağlantıları çok olan değerlere ise otorite adı verilmektedir. Derece dağılımının varyansı yani homojenliği ise ağdaki merkezileşmemenin bir ölçüsüdür. Bir birim diğer birimlerle birçok bağlantıya sahip ise bu birim ağda merkezi bir durumdur. Kolay ulaşılabilen çok sayıda en kısa patikası bulunan birim merkezi bir birimdir. Burada ağdaki merkezilik kavramı önemlilik, popülerite, prestij ve güç anlamlarına gelmektedir. Merkezilik derecesi, farklı sayıda birimlere sahip ağların merkeziliğinin karşılaştırılabilmesi amacı ile normalize edilerek 0-1 aralığında değerlere dönüştürülmektedir. Sıklıkla kullanılan merkezilik indeksleri farklı yöntemlerle hesaplanabilmekte olup bunlardan bazıları aşağıda açıklanmıştır.^{146 147}

Derece Merkeziliği (Degree Centrality – C_D):

Ağda bir birimin diğer aktörlerle olan ilişkisini yani kaç tane bağlantıya sahip olduğunu göstermektedir. Burada bağlantı sayısına derece denilmektedir. Her düğüm yönlü veya yönsüz olmak üzere farklı sayıda bağlantıya sahip olabilir. Derecesi yüksek olan bir birim ağda merkezi bir durumdur.

$C_D(x) = x$ düğümünün derecesi

$$C_D = \frac{\sum (En \text{ yüksek derece} - C_D(x))}{n - 2}$$

İndeksin 0 olması demek tüm birimlerin merkeziliğinin eşit olması tam tersi indeksin 1 olması demekse bir birimin diğer tüm birimlerle bağlantıda olması anlamına gelmektedir.

¹⁴⁶ Gürsakal, 2009, s.90-99.

¹⁴⁷ Martin Everett ve Stephen Borgatti, The Centrality of Groups and Classes, Journal of Mathematical Sociology. 23(3): 181-201, 1999, s.190.

Yakınlık Merkeziliği (Closeness Centrality – C_C):

Ağdaki diğer aktörlerle olan yakınlığın derecesidir. Bağlantıların zayıf ya da kuvvetli olması bu ölçü ile belirlenir. Yakınlık, bilgiye erişim hızı olarak da ifade edilmektedir. Yüksek bir yakınlığa sahip olmak için yüksek bir dereceye sahip olmak gerekmemektedir.¹⁴⁸ Bir birimin ağdaki diğer birimlere olan dolaylı ya da dolaysız en kısa uzaklıklarının terslerinin toplamı yakınlık ölçüdür.

$$C_C(x) = \frac{1}{\sum d(x,y)} ; d(x,y): \text{birimler arasındaki en kısa uzaklıklar}$$

$$C_C = \frac{(2n-3)(\sum \text{En yüksek yakınlık} - C_C(x))}{(n-1)(n-2)}$$

Arasındalık Merkeziliği (Betweenness Centrality – C_B):

Bir aktörün ağda diğer aktörler arasında bulunma derecesidir. Birbiri ile doğrudan bağlantılı olmayan düğümlerle bağlantı halinde olarak onlar arasındaki bilgi aktarımında köprü görevi görürler. Bu şekilde önemli bir koordinasyona sahip olan ve anahtar görevi gören düğümlerin yüksek bir arasındalık derecesine sahip olacağı açıktır.¹⁴⁹

$$C_B(x) = \frac{\sum (y \text{ ile } z \text{ arasında } x' \text{ den geçen en kısa patika sayısı})}{y \text{ ile } z \text{ arasında en kısa patika sayısı}}$$

¹⁴⁸Ko Kilkon, Jun Lee Kyoung ve Park Chisung, Rethinking Preferential Attachment Scheme: Degree centrality versus closeness centrality, Connections, Volume 28, Issue 1.p.4, 2008, s.4.

¹⁴⁹ Prased Balkundi ve Martin Kilduff, The Ties That Lead: A Social Network Approach to Leadership, The Leadership Quarterly 16, 941-961, 2004, s.946.

$$C_B = \frac{\sum (En \text{ yüksek arasındalık} - C_B(x))}{n-1}$$

Bir ağın sosyal yoğunluğunun sıkı ya da gevşek oluşu arasındalık derecesi ile nitelendirilmektedir.

Bütün merkezilik ölçüleri ile pozitif ve anlamlı ilişki içinde olan *karşılıklık* (reciprocity) değişkeni birimler arasındaki bağlantıların çift yönlü olup olmadığıdır. İki yönlülük artarsa, merkezilik ölçüleri arasındaki fark azalmaktadır.¹⁵⁰

Bir ağda bazı birimler kendi aralarında yüksek grupları dışında düşük sayıda bağlantıya sahiptirler. Bu durumda birimlerin komşuları ile arasındaki bağlantıların ölçüsünü veren “tabakalanma katsayısı” hesaplanmaktadır. Bu katsayı;

$$C_i = k_i / n ; k_i: \text{bir } i \text{ düğümünün komşuları ile olan bağlantı sayısı}$$

olup bu değer aynı zamanda iki düğümün bağlantılı olma olasılığı p’yi vermektedir.¹⁵¹ Burada bahsi geçen aktörler arasındaki bağlantılar ağ analizi için veri olarak alınmakta ve bundan elde edilen ölçüler çerçevesinde ağın sistemi içinde konumlandırılacakları yerler belirlenmektedir.

Bir ağın ölçekten bağımsızlık özelliğini tanımlamak gerekirse, şeklin parçalarının şeklin bütünü ile benzer olması anlamında kullanılan “kendi kendine benzerlik”ten bahsetmek gerekir. Bu durum, sürecin bir parçasına odaklanıldığında elde edilen görüntünün sürecin tamamı ele alındığında elde edilen görüntü ile oldukça benzer olması yani sürecin ölçekten bağımsız olarak aynı kalması anlamına gelmektedir. Ölçekten bağımsızlığa örnek olarak kuvvet yasası dağılıma sahip dağılımlar verilebilir. Kuvvet yasası dağılımına örnek olarak da Google, Yahoo, Facebook gibi sitelerin popülerlikleri sonucu her geçen gün daha fazla bağlantı çekiyor olmaları verilebilir. Gerçek hayattaki ağlar genellikle ölçeksizdirler.¹⁵²

¹⁵⁰ Thomas W. Valente ve Diğerleri, How Correlated Are Network Centrality Measures?, *Connections*, Volume 28, 1,s.24, 2008, s.12.

¹⁵¹ Gürsakal, 2009, s.101.

¹⁵² Albert-Laszio Barabasi, İş Hayatında ve Günlük Yaşamda Bağlantılar, Çev: Nurettin Elhüseyni, *Optimist Yayınları*, İstanbul, 2010; Gürsakal, 2009, s.123.

Ölçekten bağımsız ağlar birer küçük dünya ağlarıdır. Küçük dünya ağları gerçek dünyaya daha uygun ağlardır.¹⁵³ Küçük dünya ağlarında her düğüme birkaç sıçrama sonucunda erişilebilir. Özellikle sosyal ağlar ve internet küçük dünya ağı özelliği gösterir. Miligram (1967) ortaya attığı “altı adım” hipotezi ile herkesin birbirine birkaç bağlantı, en fazla da 6 kişi uzaklıkta olduğunu öne sürmüştür. *Küçük dünya ağları (small world networks)* olarak da bilinen bu hipotezi Granovetter’in (1973) yayımlanan *Zayıf bağların gücü teorisi* destekleyici niteliktedir. Granovetter, zayıf bağların kuvvetli bağlara göre bilgi çeşitliliği sağlamada ve ağın geri kalanına bağlanmada önemli bir rol üstlendiğini savunmaktadır.

Burada bahsedilen rassal ağlar, ölçekten bağımsız ağlar ve küçük dünya ağları ağ türlerini ve özelliklerini ortaya koymaktadır. Ağları daha iyi anlamak için onları yakından tanımak ve yapılarını bilmek gerekmektedir.

4.2 Sosyal Ağlar

İlk kez J.A Barnes (1954) tarafından kullanılan sosyal ağ kavramı, düğümleri genellikle bireyler olan ve bunların arasındaki bağlantılardan oluşan bir yapı olarak tanımlanmaktadır.¹⁵⁴ Küreselleşme yoluyla da ağlar artık dünya çapına genişleyebilmektedir. Böylece sosyal ağ sistemleri birimlerini daha büyük gruplanmaların içine dahil etmektedir.

Sosyal ağlar, bireyler ve topluluklar için sosyal sermaye oluşturur. Sosyal sermaye genel olarak, katılımcıların hareketini sağlayan ağlar, normlar ve güveni içeren sosyal yaşam özellikleri olarak ifade edilebilir.¹⁵⁵ Bu iki kutup, birey ve toplumsal yapı, arasında birleştirici rol üstlenmekte ve mikro düzeyden makro düzeye geçişi incelemektedir. Kısaca; sosyal ağ, birbirinden bağımsız olmayan bu bireylerin etkileşimini içermektedir. Sosyal ağ analizi ise; sosyal yapıyı ve onun etkilerini incelemektedir.¹⁵⁶ Sosyo-spikolojik etkenlerden oluşan sosyal çevreleri formüle etmeye

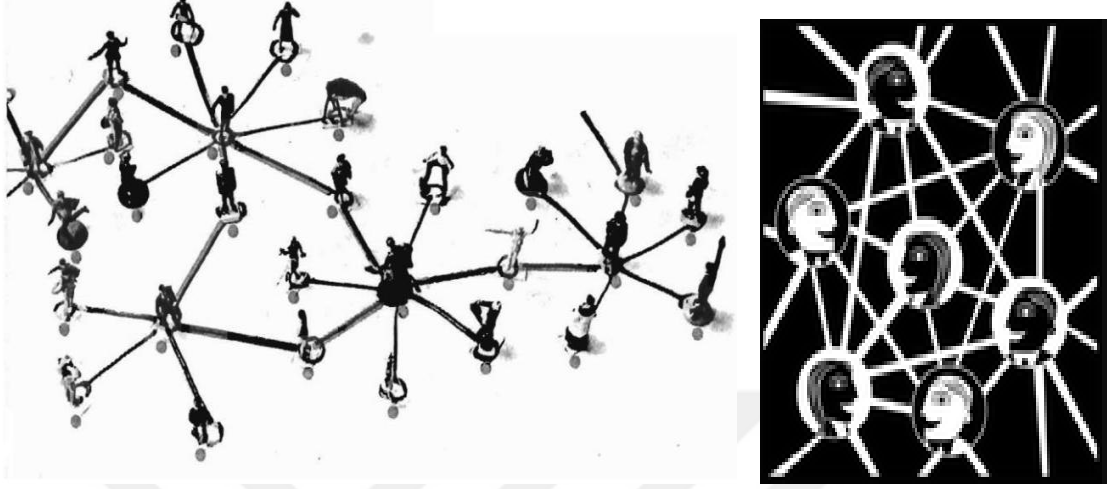
¹⁵³Nunes L.A. Amaral ve Diğerleri, Classes of Small-World Networks. Proceedings of the National Academy of Sciences, 97(21):11149-52, 2000, s.50.

¹⁵⁴ <https://www.webopedia.com> (11 Mayıs 2018).

¹⁵⁵ Robert D. Putnam, Tuning in, Tuning out: the Strange Disappearance of Social Capital in America, Political Sciences and Politics, Vol: 28, No: 4, pp. 664-683, 1995, s.670.

¹⁵⁶ David B. Tindall ve Barry Wellman, Canada as Social Structure: Social Network Analysis and Canadian Sociology, 26 (2), s.1-2, 2001, s.2.

çalışmak amacı ile matrisler ve grafiksel yöntemler kullanılmaktadır. Şekil 11’de sosyal ağ örnekleri verilmiştir.



Şekil 11: Sosyal Ağ Örnekleri

Kaynak: Öztürk, 2014 ; Serrat, 2017

4.3 Sosyal Ağ Analizi

Büyük veri kavramının doğurduğu karmaşık ilişkiler yumağı, ağ biliminde derinleşmeyi ve yeni yaklaşımlar ortaya koymayı gerektirmektedir. Karmaşık bir yapı içinde bulunan sosyal etkileşimleri analiz etmek için SAA geliştirilmiştir. Buradan anlaşılacağı üzere sosyal ağın analizi kişisel özelliklerin etkileri kadar farklı sosyal ağ üyeleri arasındaki etkileşimle birlikte sosyal yapının özelliklerinin de etkilerini ortaya koymaktadır. Bu yolda SAA, sosyal ağ yapısını betimlemeyi, görselleştirmeyi, istatistiksel olarak modellemeyi ve son olarak da ağdan bilgi üretmeyi amaçlamaktadır.¹⁵⁷ Disiplinler arası bir teknik olan sosyal ağ analizi; sosyoloji, ekonomi ve biyoloji gibi pek çok alanda uygulama yapma olanağı bulmaktadır.

SAA belirli bir alanda araştırma konuları, yazarlar ve kurumlar arasındaki ilişkilerin incelenmesi ile bilgi bağlantılarının ve bilgi geçmişlerinin tanımlanması ve analiz edilerek yorumlanmasına kullanılan bir yaklaşımdır.¹⁵⁸ SAA'nin diğer araştırma

¹⁵⁷ Gürsakal, 2009, s.184.

¹⁵⁸ John P. Scott, Social Net Work Analysis A Handbook. Second Edition. Sage Publitions, 2000, s.20.

yaklaşımlarından önemli bir farkı bulunmaktadır. Bu fark, sosyal aktörler olarak tanımlanan kavramların bireysel davranışları yerine bu sosyal varlıkların diğer varlık ya da aktörlerle etkileşimine odaklanarak bu etkileşim sonucunda oluşan yapıyı incelemektir.¹⁵⁹ Kısaca bu durum aktörler arasındaki sosyal ilişkilerin de analiz edilmekte olduğu şeklinde değerlendirilebilir.

SAA’de aktörlerin ağ içerisindeki yerlerini tespit etmek, aralarındaki ilişkileri belirlemek ve bir bütün olarak oluşan ağın tanımlanması ile karmaşıklığın daha iyi bir şekilde anlaşılması, ağın türünün belirlenmesi ve farklı ağların karşılaştırılabilmesi amacıyla ağa ilişkin belirli ölçüler hesaplanmaktadır.¹⁶⁰ Bu temel ölçümlerin sonucunda aktörlerin bağlantılarının durumu hakkında bilgi elde edilebilmektedir.

5. UYGULAMA

5.1 Çalışmanın Amacı ve Kapsamı

Çalışmanın bu bölümünde SAA’ne zaman boyutu da dahil edilerek yazarlar, kurumlar, makale konuları ve atıflar yönünden tespit edilen ağ yapıları ile ilgili görsel analizler sunulmaktadır. Bu amaçla, Web of Science (WOS) veri tabanında uluslar arası dizinde “big data” anahtar kelimesi ile taranan makale çalışmaları gözlem birimi olarak kabul edilmiş ve onlara ait bilgiler veri olarak derlenmiştir. Taranan makalelerin derlenerek veri setinin oluşturulmasında Bibexcel yardımı ile Bibliyometrik Analiz yapılmıştır. Ardından SAA için Pajek ve VOSViewer programları kullanılmıştır.

Bu bölümde aynı zamanda büyük veri üzerine yazılan makalelerin yazarları arasındaki sosyal ağ yapısı ortaya çıkarılmış, araştırmacılar arasındaki ortak yazarlık durumu ve makalelerin konu bazındaki dağılımları da belirlenmiştir. Zaman içindeki büyük veri konusunda yapılan çalışmalar için oluşan ağ yapılaşması genel olarak değerlendirilmiştir. Yöntemin tercih edilme nedeni ilişkilerin bir ağa benzemesi ve bu ağın da en iyi görselleştirilerek sunulabilecek olmasıdır. Çalışma kapsamında büyük

¹⁵⁹ C.L. Streeter ve David F. Gillespie, Social Network Analysis, Quantitative Methods in Social Work: State of the Art, s201-222, Haworth Press, Inc, 1992, s.204.

¹⁶⁰ Everett, 1999; Borgatti vd, 2002, s.16.

veri literatürüne en fazla katkı sağlayan yazarlar, kurumlar ve bunlar arasındaki sosyal ağ yapısı ortaya çıkarılmaya çalışılmıştır.

5.2 Kullanılan Veri Seti ve Tanımlayıcı İstatistikler

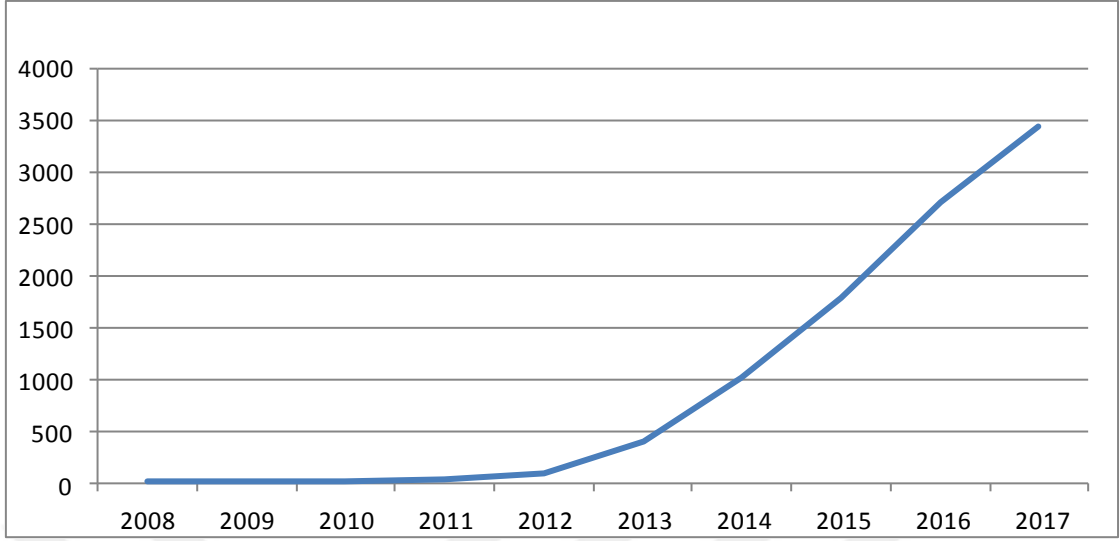
Öncelikle büyük veri üzerine yapılan tüm yayınlar hakkında genel bilgiler verilmekte olup ardından sadece makale çalışmalarının tanımlayıcı istatistikleri bilgisi verilecektir. Büyük veri hakkında ilk çalışmanın yapıldığı 1970 yılından bu yana büyük veri ile ilgili yapılan yayın sayısındaki değişim aşağıdaki gibidir:

Tablo 5: Yayın Sayısı

Yıl	Yayın	Toplam	Yıl	Yayın	Toplam	Yıl	Yayın	Toplam
1970	1	1	2001	1	10	2010	12	65
1974	1	2	2002	1	11	2011	26	91
1986	1	3	2003	3	14	2012	99	190
1993	1	4	2004	2	16	2013	407	597
1994	1	5	2005	4	20	2014	1020	1617
1995	1	6	2006	4	24	2015	1791	3408
1998	1	7	2007	4	28	2016	2704	6112
1999	1	8	2008	13	45	2017	3441	9553
2000	1	9	2009	8	53			

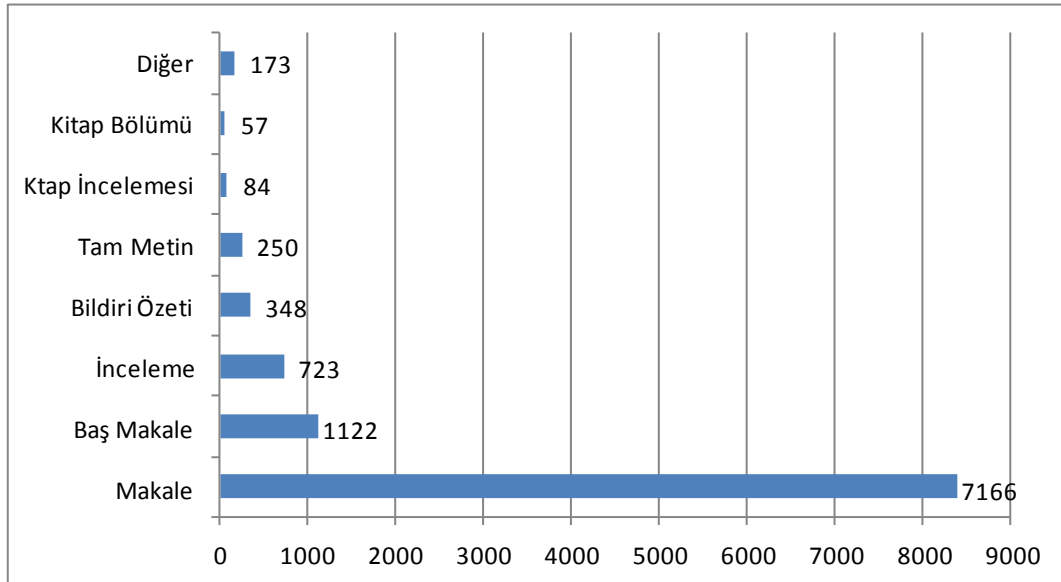
Tablo 5’te görüldüğü gibi, 1970 yılından 2002 yılına kadar her yıl birer yayın yapılmıştır. 2003 - 2007 yıllarında bir miktar artış gözlenirse de çift haneli sayılara ulaşamamıştır. 2012 yılına kadar toplam yayın sayısı 200’e yaklaşmış 2013 yılında ise geçmiş yayın sayısını üçe katlayarak 600’e yaklaşmıştır.

Burada bir sıçrama olduğu açıktır. 2017 yılına kadar da yayın sayısında hızlı bir artış olduğu görülmektedir. Son 10 yılda büyük veri ile ilgili yapılan yayın sayısı değişimin dağılım grafiği Şekil 12’de görülmektedir.



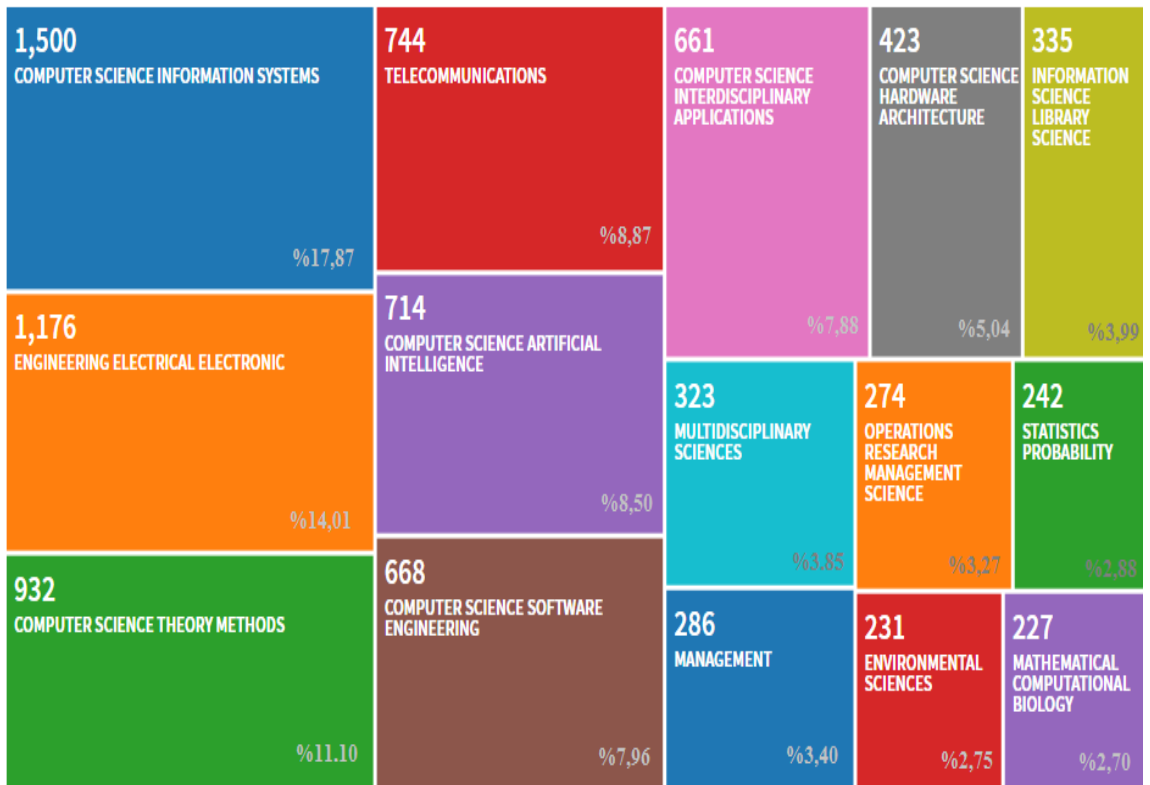
Şekil 12: Yayın Sayısı Dağılımı

Yayınların yıllara göre dağılımından kırılmanın büyük veri yılı olarak atfedilen 2012 yılından itibaren yaşandığı grafikten de açıkça görülmektedir. Bu sebeple veri setinde yayın sayısının artışa geçtiği 2013 yılı ve sonrası detaylı olarak incelemeye alınmıştır. 2013 yılından itibaren yapılan çalışma tiplerinin dağılımı Şekil 13'te gösterilmektedir. Şekil 13'te en çok yayın sayısının makale türünde olduğu görülmektedir.



Şekil 13: Yayın Türü Dağılımı

2013 ve sonrası sadece ‘‘Makale’’ alıřmaları iin bilimsel alan bazında WOS Kategorileri incelendiĐinde ařaĐıdaki haritada ‘‘big data’’ kelimesini kullanan makale sayılarına gre ilk on beřte yer alan kategoriler grlmektedir. Analize dahil edilen 7166 makaleden 1500 tanesi bilgisayar bilimleri ve enformatik sistemler kategorisinde, 1176 tanesi elektrik elektronik mhendisliĐi kategorisinde, 932 tanesi bilgisayar bilimlerinde teorik metotlar kategorisinde alıřılmıř olup bunu 744 adetle telekomnikasyon kategorisi izlemektedir.



řekil 14: WOS Kategorileri DaĐılımı

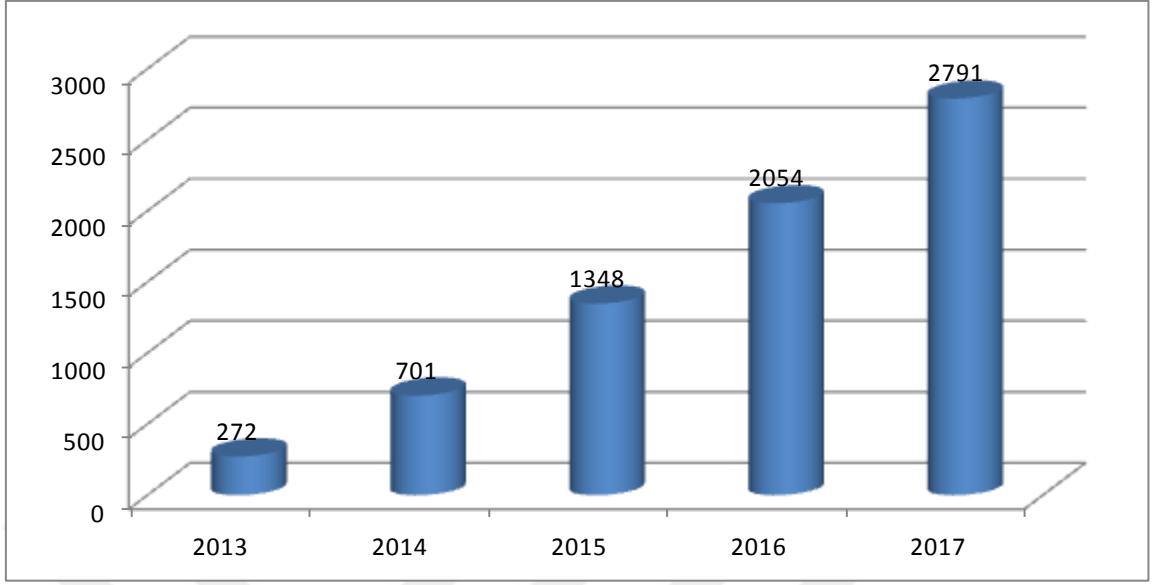
Byk veri konusuna sahip 233 farklı bilimsel kategoride makale alıřması bulunmakta olup bunlar bařta bilgisayar bilimleri olmak zere; teorik bilimlerden uygulamalı bilimler, disiplinler arası bilimler, mhendislik, telekomnikasyon, bilgi bilimleri ve ktphanecilik, iřletme, evre bilimleri, istatistik ve olasılık haritada yer almayan diĐer %1’lik kısımda kalan saĐlık bilimleri ve tıp, biyoloji, matematik, otomasyon, ekonomi, sosyoloji, hukuk, fizik, kimya, biyoloji, tarih, coĐrafya, bitki bilimleri, politik bilimler ve kltrel arařtırmalar olmak zere grldĐi gibi pek ok

bilim dalı ve alana konu olmaktadır. Genel olarak özetlendiğinde araştırma alanına göre makale sayıları aşağıdaki gibidir:

Tablo 6: Alan Dağılımı Özeti

Araştırma Alanı	Makale Sayısı	Araştırma Alanı	Makale Sayısı
Bilgisayar Bilimleri	3369	Yönetim	90
Mühendislik	1859	Eczacılık	81
Telekomünikasyon	744	Eğitim Araştırmaları	81
İşletme	559	Ziraat	59
Matematik	420	Şehir Çalışmaları	46
Bilgi Sistemleri ve Kütüphanecilik	335	Edebiyat	17
Sağlık Bilimleri	171	Kültürel Çalışmalar	16
Haberleşme	137	Spor Bilimleri	8
Psikoloji	131	Diğer	146
Ulaşım/Taşımacılık	124		

2013 yılından itibaren yayın yılına göre “Makale” sayıları aşağıdaki grafikte verilmiştir. WOS veri tabanlarında 2012 yılında 58 tane makale çalışmasının yayınlandığı bilgisi altında 2013 yılında makale sayısının 2012’deki yaklaşık 4.7 katı arttığı söylenebilir. 2015 yılında 2014 yılındakinin yaklaşık 2,5 katı; 2016 yılında 2015 yılındakinin yaklaşık 1,5 katı; 2017 yılında ise 2016’dakinin yaklaşık 1,35 katı arttığı görülmektedir. Büyük veri konusuna olan ilginin artışı bu konudaki makale çalışması sayısındaki artışta da açıkça görülmektedir. Son beş yılda makale sayısındaki artış azalarak arttığı görülmektedir.



Şekil 15: Makale Sayıları

WOS'ta taranan 7166 adet makalenin yaklaşık %98,2'i İngilizce dilinde yazılmıştır. Diğer sıklıkla çalışma yapılan diller Almanca, İspanyolca, Çince ve Fransızca'dır. Yayın dillerine göre dağılım aşağıdaki gibidir.

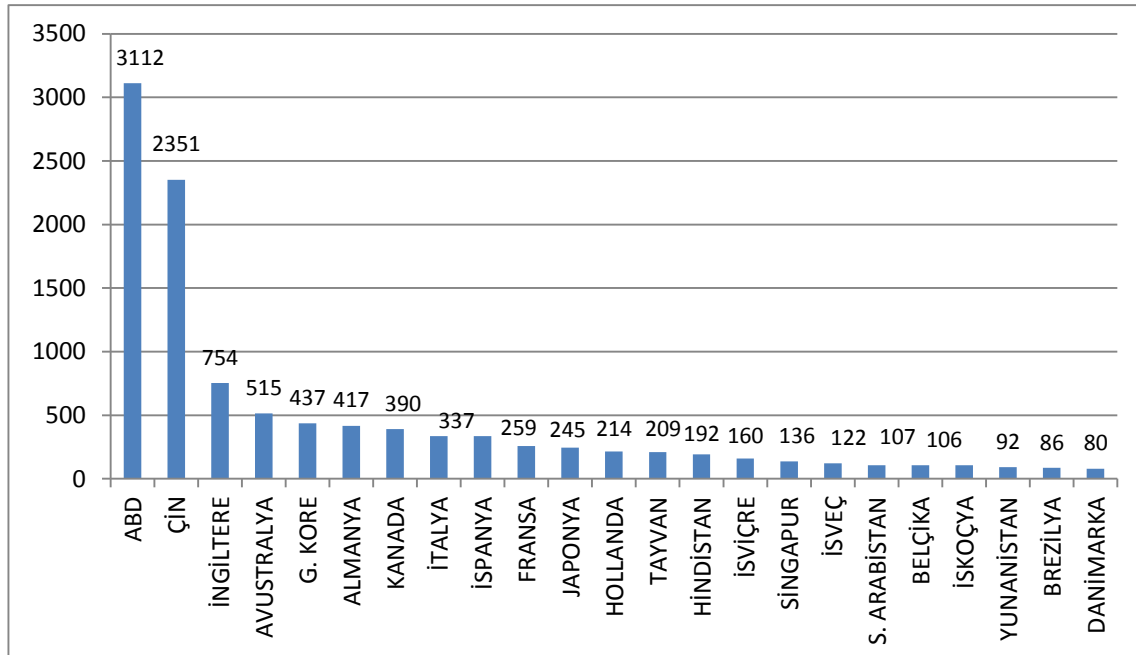
Tablo 7: Yayın Dilleri

Dil	Makale Sayısı	Dil	Makale Sayısı
İngilizce	7018	Portekizce	5
Almanca	54	Slovençe	2
İspanyolca	34	Çekçe	1
Çince	18	Flemenkçe	1
Fransızca	15	Japonca	1
Rusça	9	Korece	1
Macarca	7		

Bu makalelerin yayınlanmasını sağlayan toplamda 5762 üniversite ve kuruluş bulunmaktadır. Bunlar arasından büyük veri konusunda en fazla makalenin yayınlandığı yaklaşık ilk 100 üniversite ve kuruluş Ek 1'de listelenmektedir. En çok yayın sayısına

sahip ilk 10 kurum sırası ile California Üniversitesi (305), Çin Bilim Akademisi (274), Harvard Üniversitesi (145), Texas Üniversitesi (142), Londra Üniversitesi (136), Tsinghua Üniversitesi (132), Pennsylvania Commonwealth Yüksek Eğitimi (100), Huazhong Bilim ve Teknoloji Üniversitesi (98) ve Stanford Üniversitesi (97)'dir. Yayın çoğunluğunun ABD (California Üniversitesi, Harvard Üniversitesi, Texas Üniversitesi, Pennsylvania Commonwealth Yüksek Eğitimi ve Stanford Üniversitesi) ve Çin Halk Cumhuriyeti (Çin Bilim Akademisi, Tsinghua Üniversitesi, Huazhong Bilim ve Teknoloji Üniversitesi) tarafından sağlanmakta olduğu görülmektedir.

WOS'ta taranan 7166 makalenin yazıldığı 104 ülke içerisinde en çok makale sayısına sahip olan ülkeler Şekil 16'da verilmiştir. 3112 ile ilk sırada ABD yer alırken, 2351 ile ikinci sırada Çin Halk cumhuriyeti yer almaktadır. Türkiye 60 makale ile 30. sırada yer almaktadır. Farklı yazarlar tarafından yazılan bu makalelerin çoğunlukla uluslararası yayıncılar aracılığı ile yayınlanmakta olduğu görülmektedir. Bu açıdan Türkiye, makale sayıları 63 olan İrlanda, Polonya ve Norveç ile benzerlik göstermektedir. Bulgaristan, Ukrayna, Irak, Küba, Kuveyt, Bosna Hersek, Nepal, Senegal, Jamaika, Kazakistan, Kırgızistan, Makedonya gibi ülkelerin ise beş ve daha az sayıda makalesi bulunmaktadır.



Şekil 16: Ülkelerin Yayın Sayıları

Uluslararası literatürde büyük veri konusunda makale sayısı bakımından en fazla katkı sağlayan yazarlar 24083 yazar içinden listelenmiştir. En çok çalışma frekansına sahip olan yazarların bilgileri Tablo Ek 2’de verilmiştir. Sıralamada büyük veri literatürüne en yüksek frekansta makale yayını yapan yazarların Çinli yazarlar Yin Zhang, Yang Li ve Yunhao Liu olduğu görülmektedir.

Makalelerin yazar sayısına ilişkin dağılımı değerlendirildiğinde makalelerin oranları birbirine yaklaşık değerler olup büyük veri konusunda yazarların ortak çalışmalara da önem verdiği görülmektedir.

Makaledeki yazar sayısına ilişkin dağılım incelendiğinde son beş yıldaki (2013-2017) toplam makale sayısı 7166 olmak üzere, makalelerin yaklaşık yarısının tek, iki ve üç yazarlı kalan yarısının da dört ve üstü sayıda yazara sahip olduğu görülmektedir. En yüksek sayıda olanın %20 oranı ile üç yazarlı makale sayısı olduğu belirlenmiştir.

Tablo 8: Ortak Yazarlık

Yazar Sayısı	2013	2014	2015	2016	2017	Ortalama	Oran
1	52	110	195	235	338	186	0,13
2	54	114	271	353	499	258	0,18
3	51	147	262	425	563	290	0,20
4	41	119	201	395	476	246	0,17
5	24	63	157	259	390	179	0,13
5’ten daha fazla	50	148	262	387	525	274	0,19

WOS veri tabanında taranan büyük veri konusu ile ve Türkiye’deki üniversiteler ve kurumlarda bulunan yazarlar tarafından yazılan eserlerin çoğunlukla Elsevier, Springer, IEEE, TÜBİTAK gibi yayıncılar tarafından yayınlandığı görülmektedir. Çoğunluğu ortak yazarlıkla oluşturulmuş çalışmalarda da sıklıkla

yabancı yazarlar ortaklığının da olduğu görülmüştür. İlgili tablo Ek 3'te verilmiştir. Yabancı ortak yazarlarla yapılan çalışmalar * ile belirtilmiştir.

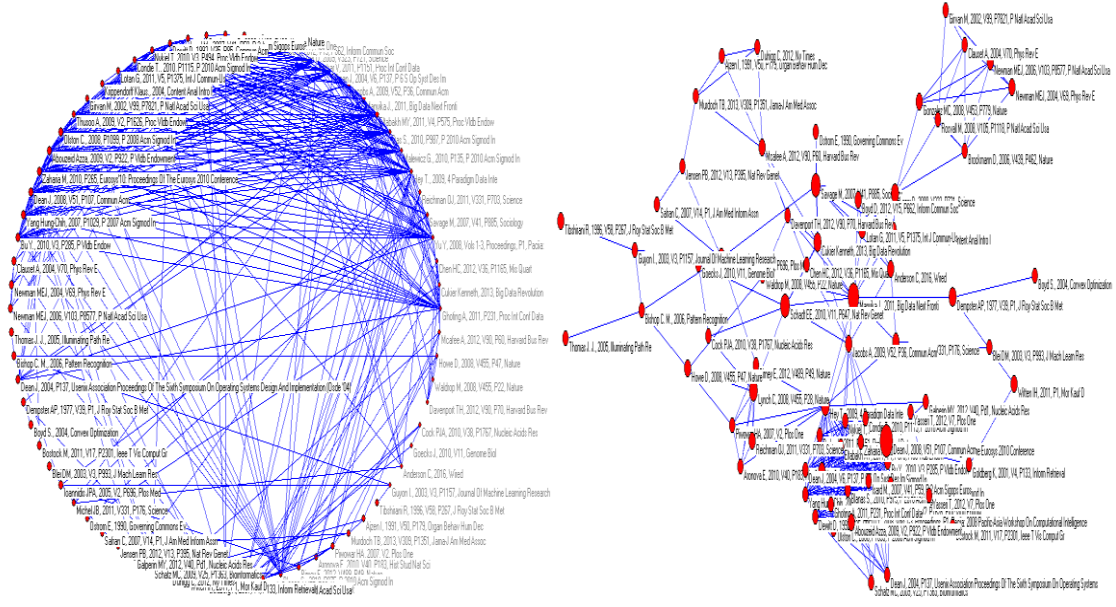
Son beş yılda Türkiye'de WOS veri tabanında taranan büyük veri konulu makale çalışmalarının 13 tanesi Bilkent Üniversitesi'nden, 6 tanesi Sabancı Üniversitesi'nden, 4 tanesi Türk Telekom'dan, 4 tanesi TÜBİTAK'dan, 3 tanesi Fatih Üniversitesi'nden ve 3 tanesi ise Koç Üniversitesi'nden çıkmıştır. Boğaziçi, İstanbul Teknik, İstanbul, Ortadoğu Teknik, İzmir Ekonomi, Kocaeli Üniversitelerinden de ikişer tane çalışma bulunmaktadır. Yayın yılı bakımından ise 2013 ve 2014 yıllarından üçer tane makale taranmakta iken; 2015 yılında bu sayı 13'e yükselmiş, 2016'da sayı sabit kalmış ve 2017'de 23'e çıkmıştır.

5.3 Sosyal Ağ Analizi

5.3.1 Makaleler ve Yazarları Arasındaki Sosyal Ağ

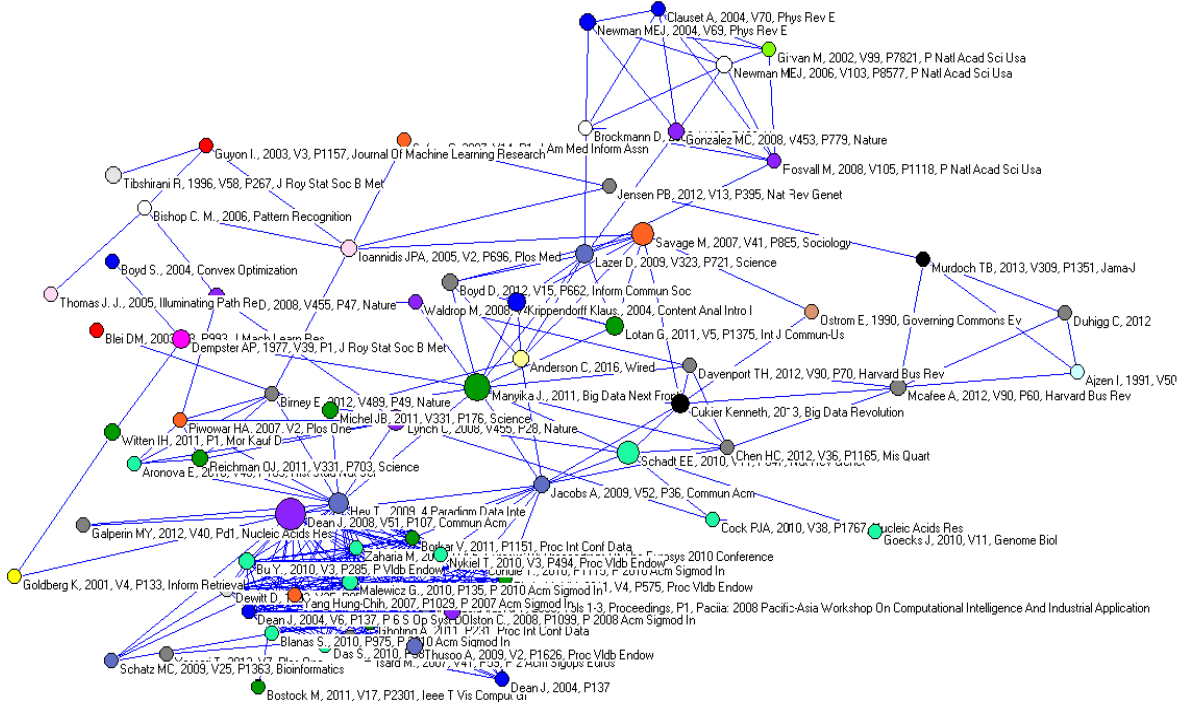
Gerçek dünyadaki ağların çoğu yapısı itibari ile ölçekten bağımsız ağ olarak yorumlanmaktadır. Yapısı düzensiz, karmaşık ve zaman içinde dinamik bir şekilde değişen ağlar da karmaşık ağlar olarak nitelendirilmektedir. Büyük veri üzerine yapılan tarama sonucunda elde edilen makalelerin ve yazarlardan oluşturulan sosyal ağ yapılarının, ölçekten bağımsız ve yıllar itibari ile karmaşık bir ağ yapısında olduğu söylenebilir.

Aşağıda ilgili ağ ve vektör grafikleri yıllar (2013-2017) bazında gösterilmektedir.



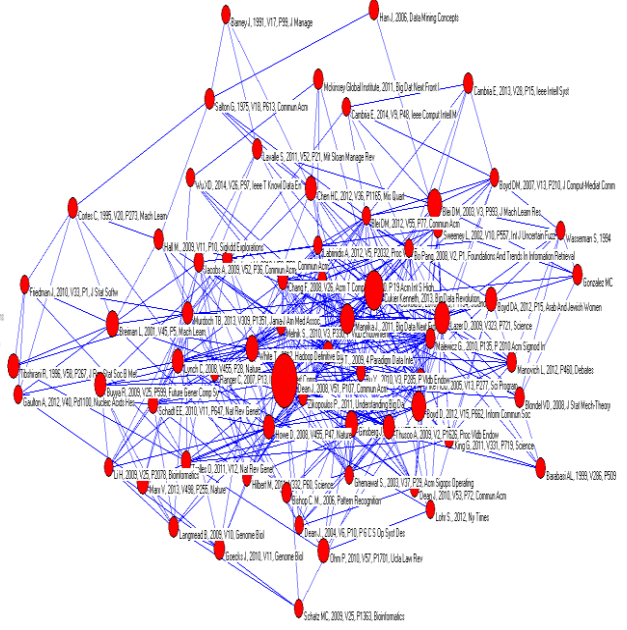
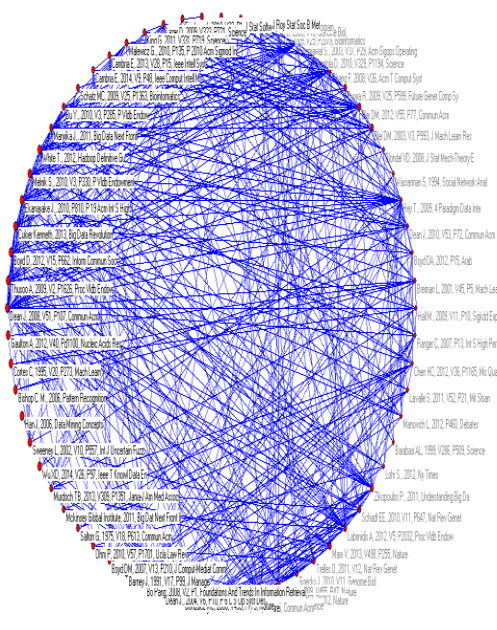
Ağ Grafığı

Vektör Grafığı



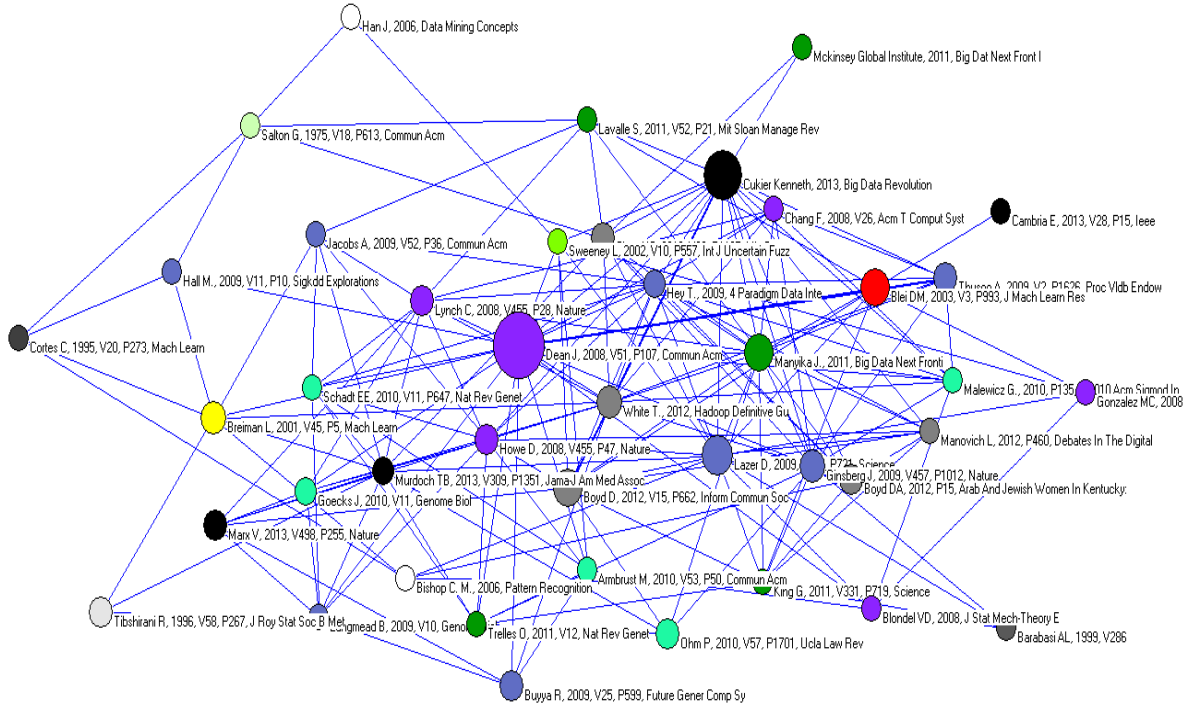
Atıf Miktarlarına Göre Vektör Grafığı

Şekil 17: 2013 Yılında WOS'ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağ Yapısı



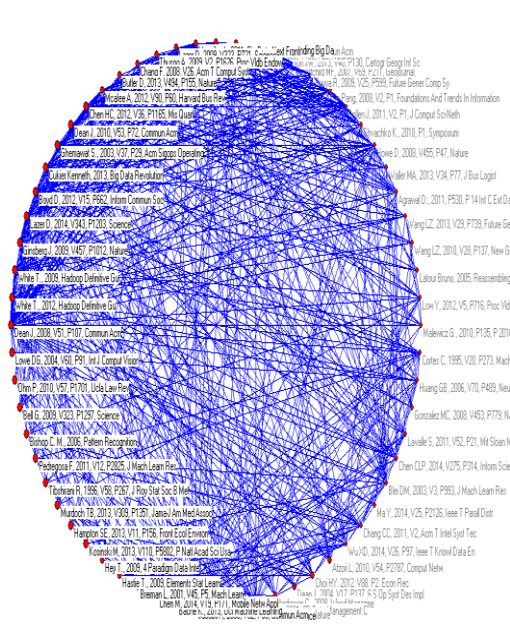
Ağ Grafığı

Vektör Grafığı

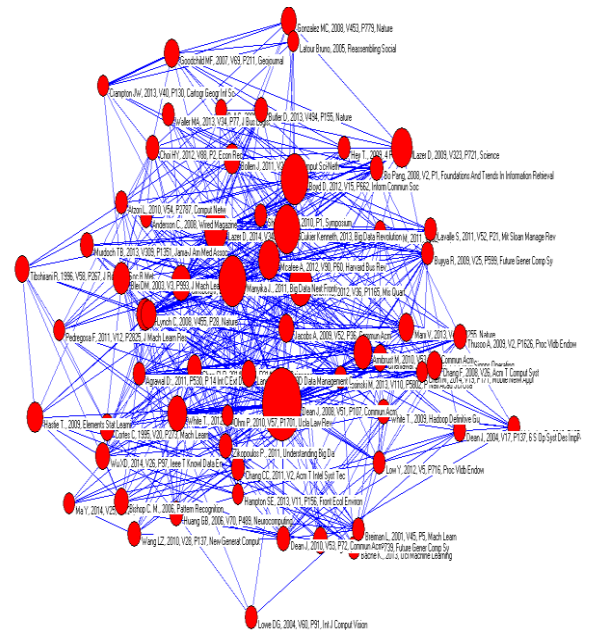


Atıf Miktarlarına Göre Vektör Grafığı

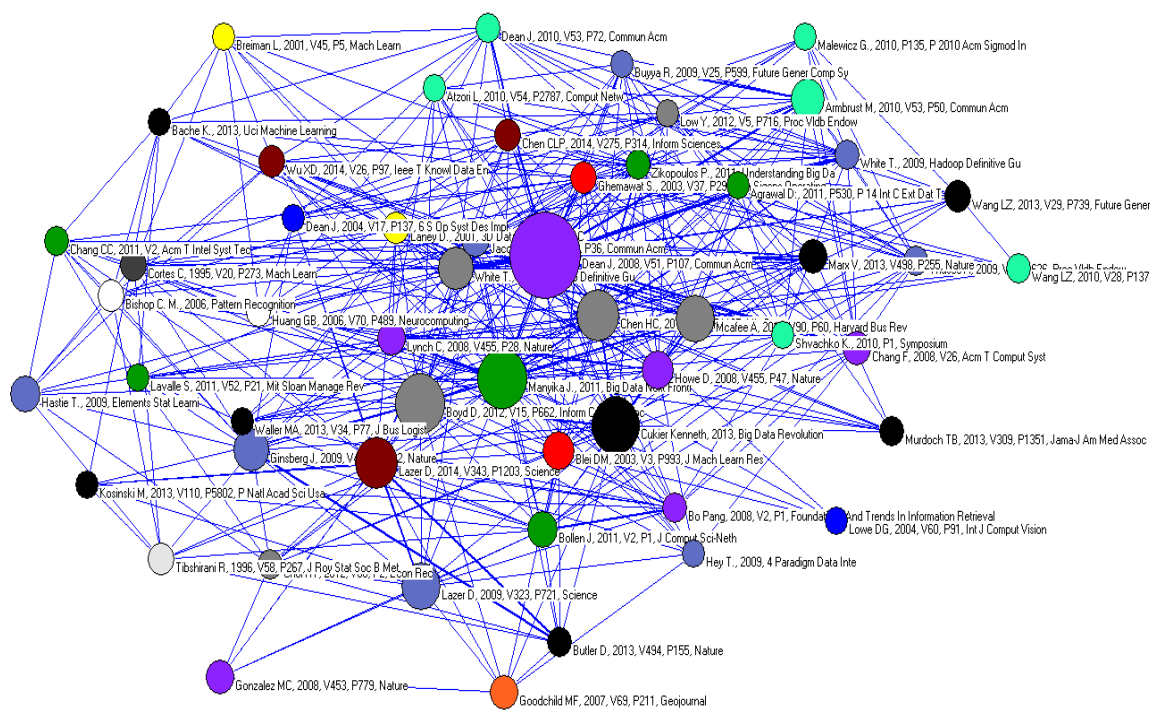
Şekil 18: 2014 Yılında WOS'ta Yayımlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı



Ağ Grafığı

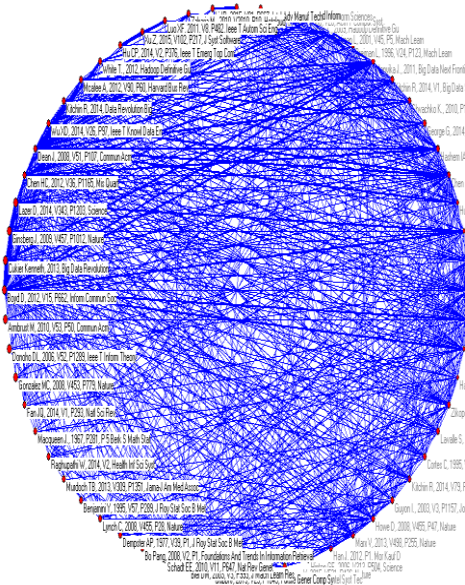


Vektör Grafığı

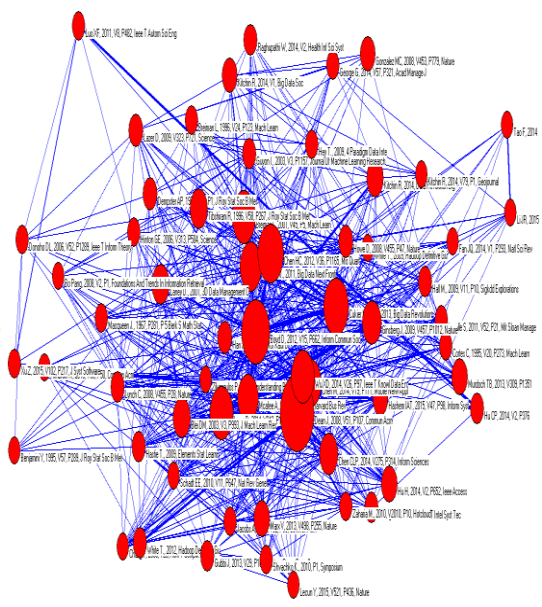


Atıf Miktarlarına Göre Vektör Grafığı

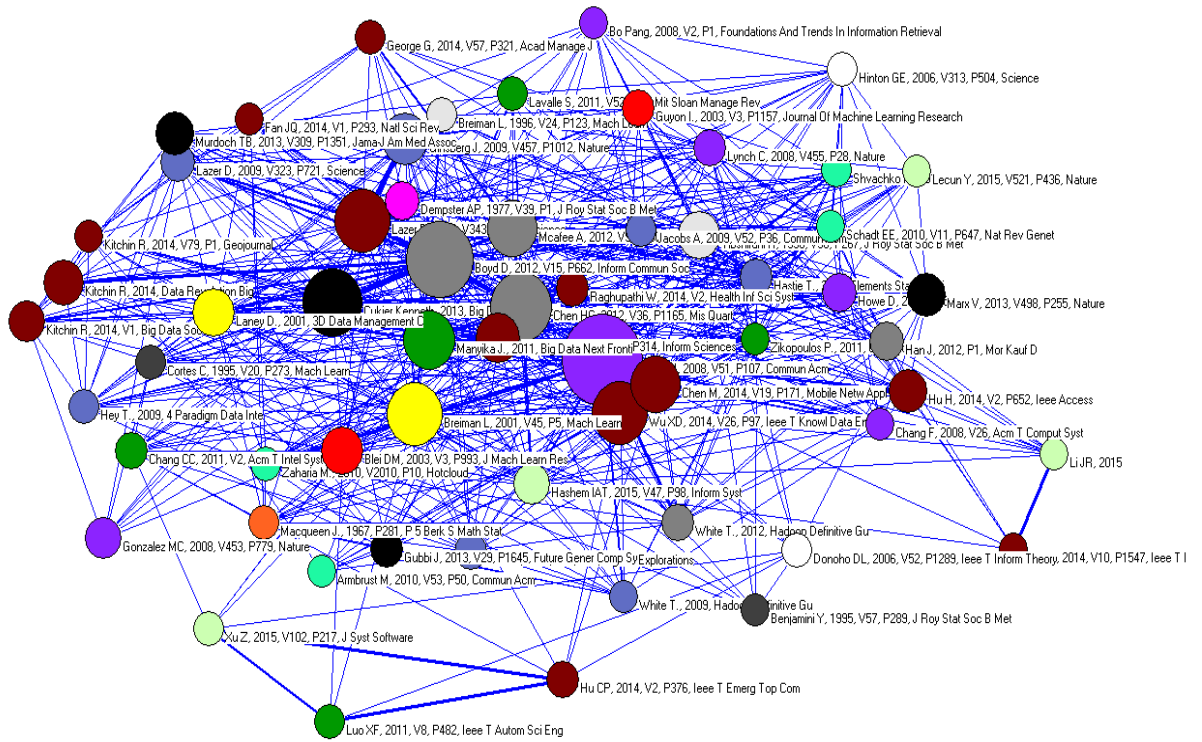
Şekil 19: 2015 Yılında WOS'ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı



Ağ Grafiği

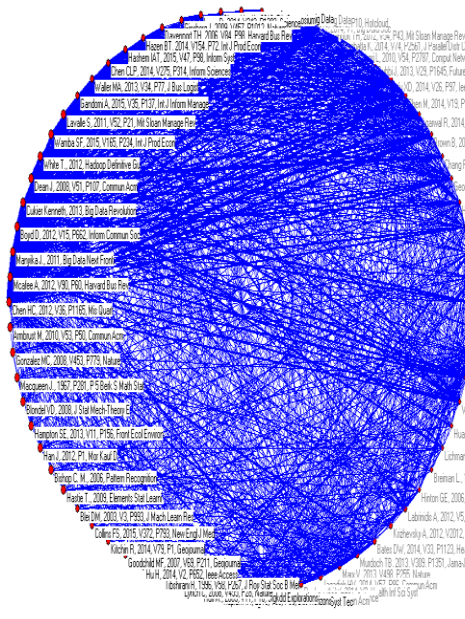


Vektör Grafiği

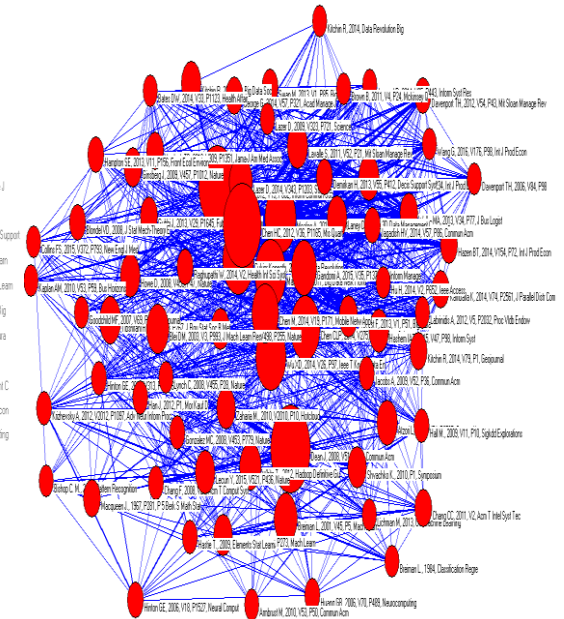


2016 Yılı Atıf Miktarlarına Göre Vektör Grafiği

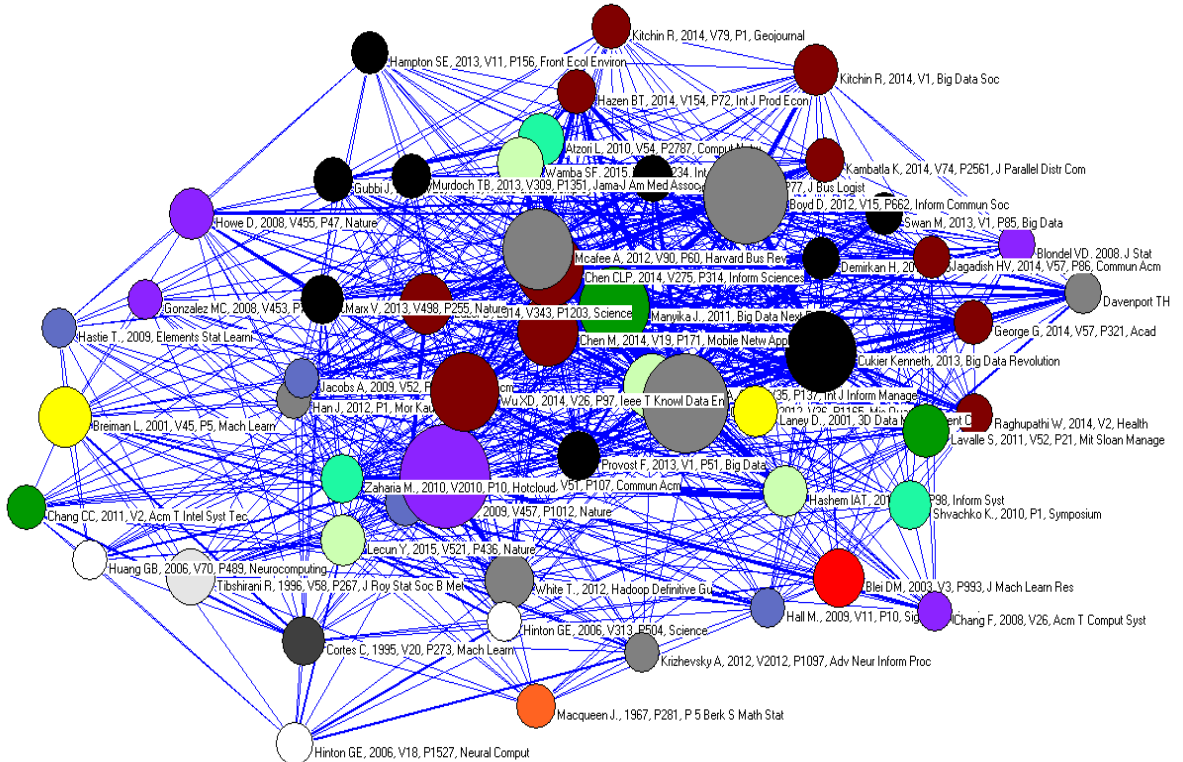
Şekil 20: 2016 Yılında WOS'ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı



Ağ Grafiği



Vektör Grafiği



Atıf Miktarlarına Göre Vektör Grafiği

Şekil 21: 2017 Yılında WOS'ta Yayınlanan Makaleler Arasındaki Ortak Atıf Ağı Yapısı

Şekil 17, 18, 19, 20 ve 21’de elde edilen ağların yapısı incelendiğinde;

2013 yılında yazarlar arasındaki sosyal ağın zayıf olduğu görülmektedir. 2013 yılında görece daha seyrek bağlantılaşma, gevşek bir ağ yapısı, bağlantıları çok az olan izole (bağlantısız) olmaya yakın düğümler bulunmaktadır. 2013 yılından itibaren grafikler incelendiğinde ise her yıl ağ yapılarının daha da sıkılaştığı, bağlantı sayılarının arttığı gözlenmektedir. Yıllar itibari ile yazarlar arasındaki bilgi alışverişinde ilişki ve iletişimlerin arttığı yorumu yapılabilir. 2017 yılına gelindiğinde ise önceki yıllara kıyasla oldukça sıkı ve yoğun bir ağ yapısının mevcut olduğu görülmektedir. 2017 yılında ağın çekirdeğinde yoğun bağlantılar mevcuttur. Atıf miktarlarına göre vektör grafiklerinde düğümlerdeki dairenin büyüklüğü atıf sayısı ile orantılı olmakta olup, büyük dairesel düğümlerdeki yazarların diğer yazarlara göre daha çok sayıda atıf alındığını ifade etmektedir.

2013 yılında yayınlanan çalışmalar incelendiğinde ilk yazarı Jeffrey Dean (2008) olan “*MapReduce: Simplified Data Processing on Large Clusters*” başlıklı çalışma 14 kez ve ilk yazarı James Manyika (2011) olan “*Big Data: The Next Frontier for Innovation, Competition, and Productivity*” başlıklı çalışma 11 kez atıf olarak en çok atıf alan iki yazar olmuştur. İlk yazarı Eric E. Schadt (2010) olan “*Computational Solutions to Large-Scale Data Management and Analysis*” başlıklı çalışma 7 kez atıf almıştır. Sonraki yıllarda çok atıf alan yazarlar arasında olan ve ilk yazarı Kenneth Cukier (2013) olan “*Big Data: A Revolution That Will Transform How We Live, Work, and Think*” başlıklı çalışma ise çalışmanın henüz yayınlandığı 2013 yılında 5 kez atıf almıştır. Şekil 17’de görüldüğü gibi genel olarak yazarlar arasında zayıf ağların olduğu sadece bazı yazarlar arasında ağın sıkılaşma eğiliminde olduğu ve bağlantı sayılarının yazarların çoğu arasında 4 ve daha az sayıda olduğu söylenebilir.

2014 yılında yapılan yayın sayısının 2013 yılında yapılan yayın sayısının yaklaşık 3 katına çıkmasıyla ağın yoğunluğunda da bir miktar artış olduğu Şekil 18’deki grafiklerde görülmektedir. Jeffrey Dean (2008)’in atıf sayısındaki artış da benzer şekilde gerçekleşmiş ve 49 kez atıf almıştır. Ardından Kenneth Cukier (2013) 27 kez atıf almış, ilk yazarı David Lazer (2009) olan “*The Parable of Google Flu: Traps in Big Data Analysis*” başlıklı çalışma 19 kez, James Manyika (2011) 17 kez, ilk yazarı

Danah Boyd (2012) olan “Critical Questions for Big Data” başlıklı çalışma 16 kez atıf almış ve en çok atıf alan yazarlar olmuşlardır.

2015 yılında Jeffrey Dean (2008) 98, James Manyika (2011) ve Danah Boyd (2012) 50’şer ve Kenneth Cukier (2013) 44 kez en çok atıf almıştır. İlk yazarı Chiang H. Chen (2012) olan “*Business Intelligence and Analytics: From Big Data to Big Impact*” başlıklı çalışma 36; David Lazer (2009) 30 ve ilk yazarı Tom White (2012) olan “*Hadoop: The Definitive Guide*” başlıklı çalışma 24 atıf almıştır. Önceki iki yıla göre ağdaki sıkılaşıma ve yayılımın artmış olduğu Şekil 19’da görülmektedir.

2016 yılında Jeffrey Dean (2008) 109; Danah Boyd (2012) 75 ve Chiang H. Chen (2012) 63 kez, Kenneth Cukier (2013) 61 kez, David Lazer (2009) 53 kez atıf alarak en çok atıf alan yazarlar olmuşlardır. James Manyika (2011) ise 44 kez atıf almıştır. Ağda her yıl artmakta olan yoğunlaşmanın sonuçları Şekil 20’de görülmektedir.

2017 yılında ise Jeffrey Dean (2008) 133; Chiang H. Chen (2012) 119; Danah Boyd (2012) 115 kez atıf alarak en çok atıfı almaktadırlar. James Manyika (2011) ve ilk yazarı Andrew McAfee (2012) olan “Big Data: The Management Revolution” başlıklı çalışma 83 kez ve Kenneth Cukier (2013) 82 kez atıf almaktadır. Önceki yıllara göre ağın en yoğun halde olduğu, yazarlar arasındaki bağlantıların en sık olduğu ve yayın sayısının da en fazla bu yılda olduğu Şekil 21’de açıkça görülmektedir.

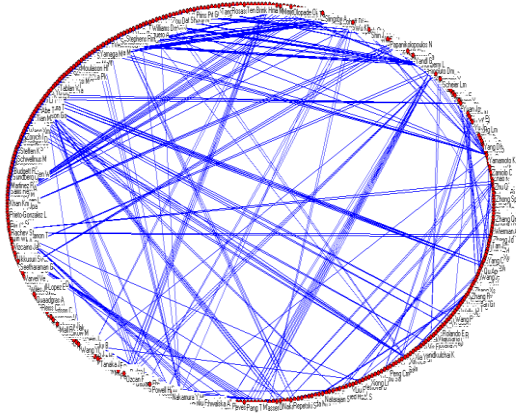
Buradan yola çıkarak bu beş yılda (2013-2017) her yıl Jeffrey Dean (2008)’in Mapreduce (Eşle-İndirge)’u anlatan kaynağı ile en çok atıf alan ortak yazar olduğu söylenebilir. Görünen o ki, büyük veri konusunda yapılan çalışmaların pek çoğu Jeffrey Dean (2008)’a atıf vermekte ve kaynak olarak bu çalışmadan yararlanmaktadır. Daha önceki başlıklarda da bahsedildiği üzere Mapreduce büyük veriyi depolayan işletim teknolojilerinin geliştirilmesi ile ilgili büyük verinin depolanma zorluğu sebebiyle önemli bir konudur. Burada alınan atıflardaki fazlalıkla da bu konunun önemliliği ve büyük veri konusunda Jeffrey Dean (2008)’in önemli bir kaynak olduğu sonucuna ulaşılmaktadır. Benzer şekilde sırasıyla büyük veri hakkındaki soruları cevaplayan Danah Boyd (2012), büyük veri devriminden bahseden Kenneth Cukier (2013) ve

büyük verinin sonraki sınırlarını inceleyen James Manyika (2011) da en çok atıf alan, her yıl en çok faydalanılan ve literatüre içerdikleri bilgiler açısından en fazla katkıda bulunan ortak yazarlardandır.

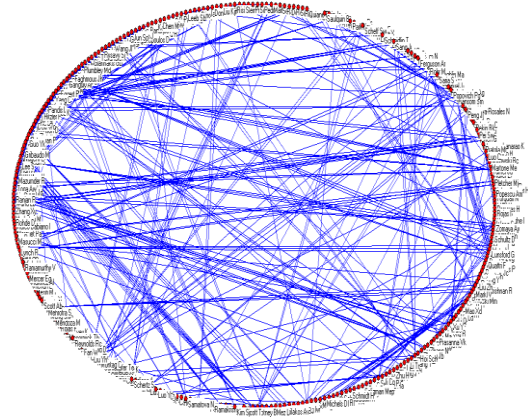
Burada belirtmekte fayda olan diğer bir husus ise makalenin yazım dilinin önemli olduğudur. Bir makaleyi daha sık kullanılan bir dilde yazmak ağın daha fazla etkisinde kalmayı ve ağın içine girmeyi sağlamaktadır. Benzer şekilde daha az kullanılan dillerde yazmak ise ağın dışında kalmaya sebep olabilir. Dolayısıyla çalışmada kullanılan dil, ağın içinde kalabilmek için önemli bir etkidir. Makaleler incelendiğinde ağın çekirdeğinde bulunan yazarların büyük veri üzerine yazılan makalelerde en çok kullanılan dil olan İngilizce'yi kullandıkları görülmüştür.

Burada uluslararası literatürde büyük veri konusunda en çok makale sayısına sahip olan yazarlarla (Yin Zhang, Yang Li ve Yunhao Liu), büyük veri konusunda en çok atıf alan çalışmaların yazarlarının farklı olduğu da elde edilen bir diğer sonuçtur. Burada büyük veri literatürüne sayıca en fazla katkıyı sağlayan yazarların Çinli yazarlar olduğu görülmektedir.

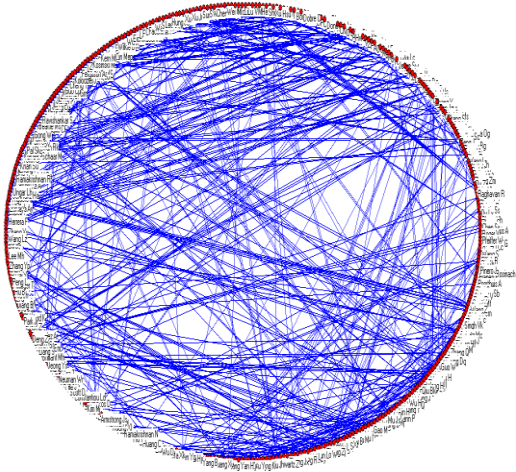
Yazarlar arasındaki sosyal ağ yapısı incelendiğinde Şekil 22'de olduğu gibi ölçekten bağımsız bir ağ yapısı olduğu görülmektedir. 2013 yılında gevşek bir yapıda olan ağ da 2017 yılına doğru sıkılaşıma meydana geldiği görülmektedir. Yıllar ilerledikçe yazarlar arası iletişimin ve sosyal ilişkilerin artmakta olduğu söylenebilir.



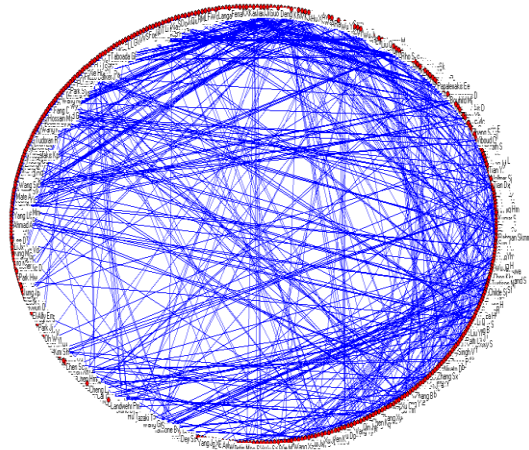
2013



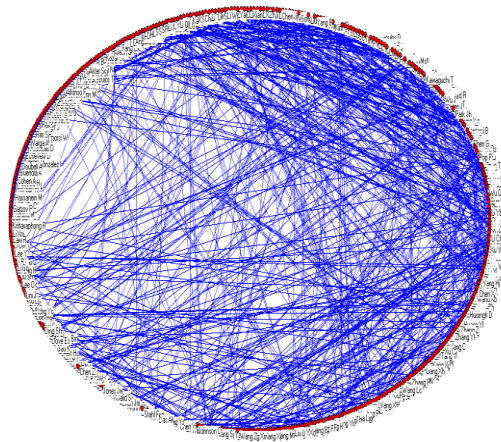
2014



2015



2016



2017

Şekil 22: Yıllara Göre WOS'ta Yayınlayan Makalelerin Yazarları Arasındaki Sosyal Ağ Yapısı

Yazarlar ağı için hesaplanan derece ölçüleri Tablo 9’da verilmiştir. Genel olarak merkezilik derecelerindeki düşük ölçü değerleri yazarlar arasındaki sosyal ağ yapısının görece gevşek, yüksek ölçü değerleri ise görece sıkı olduğunu göstermektedir. Burada yazarlar arasında 2013 yılından 2017 yılına artan ve orta derece yoğunlukta ağlar söz konusudur. Ağ yoğunluğunun en yüksek (%55) olduğu yıl 2017 yılıdır.

Tablo 9: Derece Ölçüleri ve Kümelenme Katsayıları

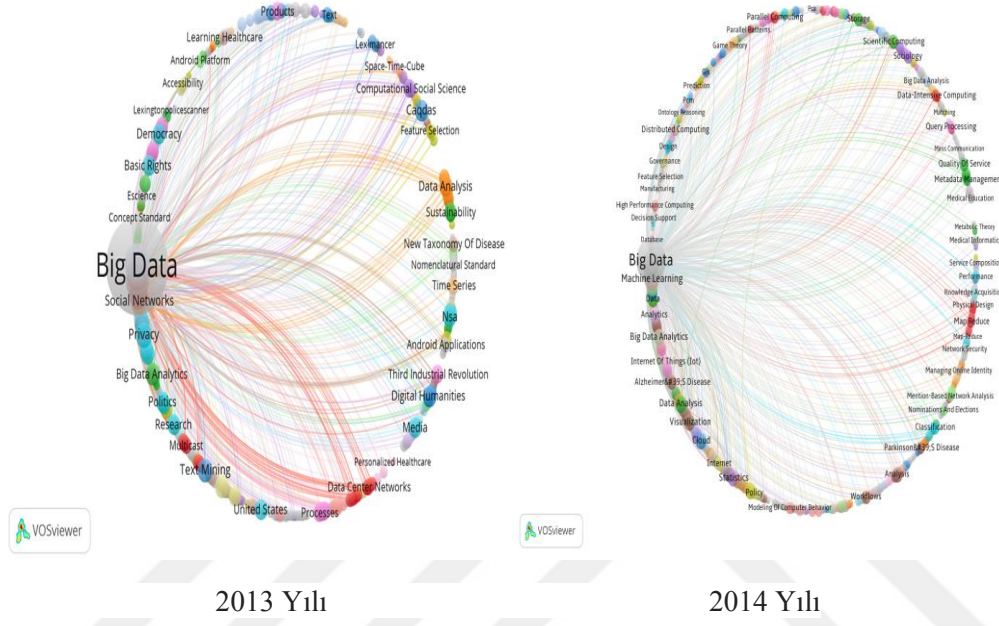
Yıl	Merkezilik Dereceleri	Arasındalık Dereceleri	Kümelenme Katsayıları	
			Watts-Strogatz	Transitivity
2013	0,35682	0,09718	0,40123	0,38568
2014	0,39153	0,15291	0.46273	0.42051
2015	0,47756	0,09646	0.54877	0.48016
2016	0,49983	0,04852	0.60153	0.55893
2017	0,55489	0,18795	0,64786	0,61258

Arasındalık merkeziliği dereceleri incelendiğinde, elde edilen değerlerin düşük olması sebebiyle yazarlar arasında bağlantı sağlayan köprü görevinde bulunan çok sayıda yazar olmadığı söylenebilir. Bu sayının 2014 ve 2017 yıllarında diğer yıllara göre daha fazla olabileceği görülmektedir.

Kümelenme katsayılarının her yıl bir önceki yıla göre artmakta olduğu görülmektedir. Buna göre yazarlar arasında en fazla işbirliğinin ve ortaklığın 2017 yılında olduğu söylenebilir. Buradan hareketle çok yazarlı makale sayısının bu yıl da daha fazla olduğu sonucuna ulaşılabilir. Tablodaki bu değerler yazarlar arasında oluşan tabakalanma yapılarının var olma derecesini göstermektedir. 2013 yılında %40 ve altında seyreden kümelenme oranının, 2017 yılında %60’ların üstüne çıktığı da açıkça görülmektedir.

5.3.2 Anahtar Kelimeler Arasındaki Sosyal Ağ

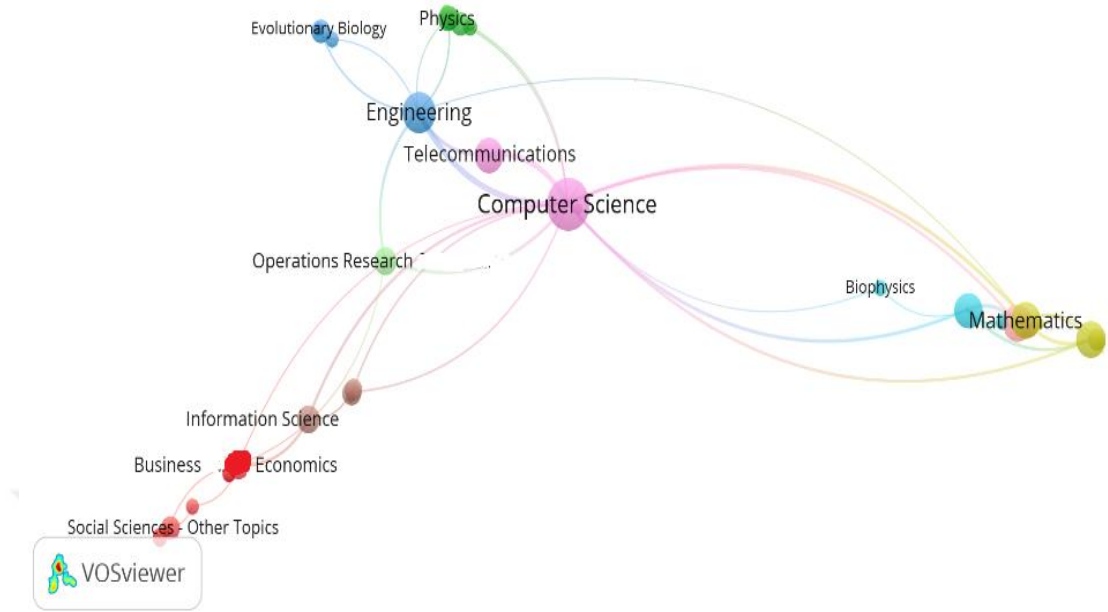
Büyük veri üzerine yapılan tarama sonucunda elde edilen makalelerin anahtar kelimeleri arasında oluşan sosyal ağ yapıları aşağıda gösterilmektedir.



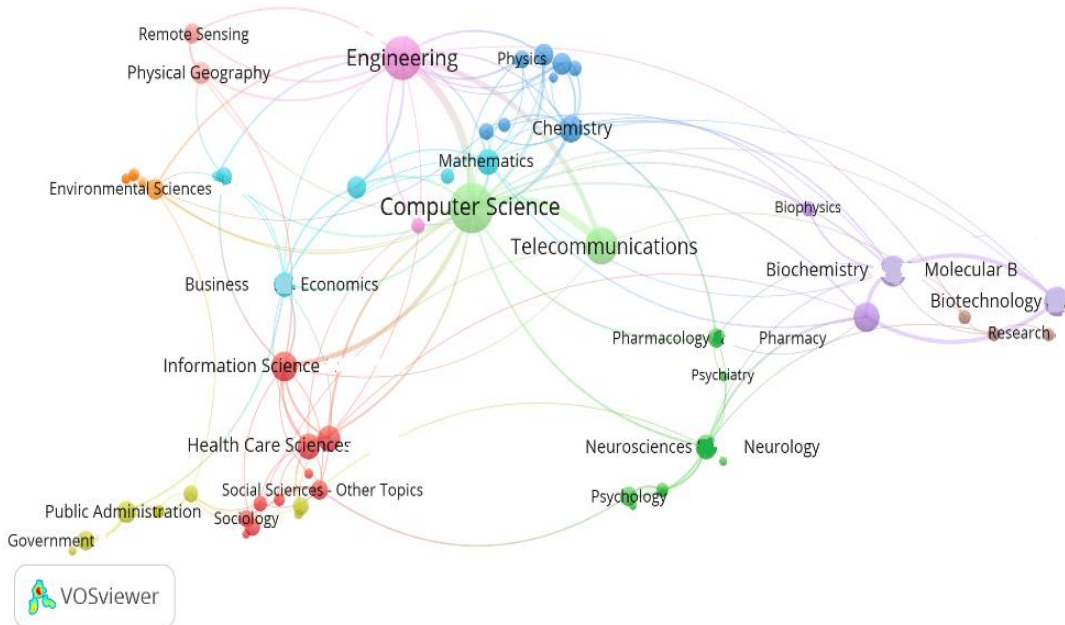
Şekil 23: 2013 ve 2014 Yılları Makalelerde Kullanılan Anahtar Kelime Sosyal Ağı

2012 yılında büyük veri farkındalığının başlamasının ardından 2013 yılında büyük veri üzerine yayınlanan makalelerin sayısı hızla artmaya başlamıştır. Şekil 23'te, 2013 yılı için makalelerde kullanılan anahtar kelimeler incelendiğinde büyük veri (big data) kelimesi ile birlikte veri analizi (data analysis), büyük veri analizi (big data analytics), medya (media), sosyal ağlar (social networks), gizlilik (privacy), 3. sanayi devrimi (third industrial revolution), sustainability (sürdürülebilirlik), android uygulamalar (android applications), metin madenciliği (text mining), veri merkezi ağları (data center network), ebilim (escience), demokrasi (democracy) ve temel haklar (basic rights) kelimelerinin kullanılmaya başlanıldığı görülmektedir.

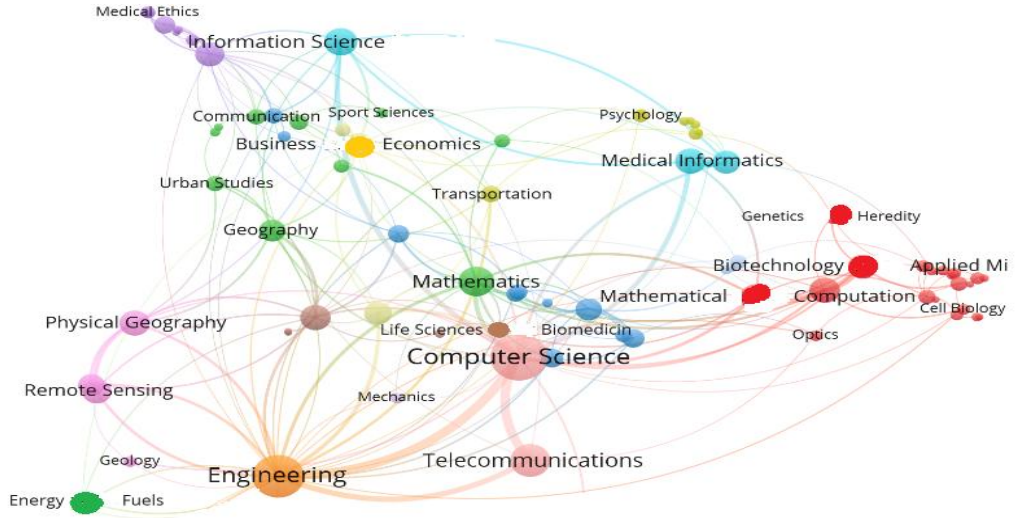
Şekil 23'te, 2014 yılında ise büyük veri ile eşleşen makine öğrenmesi (machine learning), nesnelerin interneti (internet of things), görselleştirme (visualization), bulut (cloud), veritabanı (database), oyun teorisi (game theory), storage (depolama), internet, eşle indirge (mapreduce), metaveri yönetimi (metadata management), ağ güvenliği



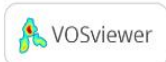
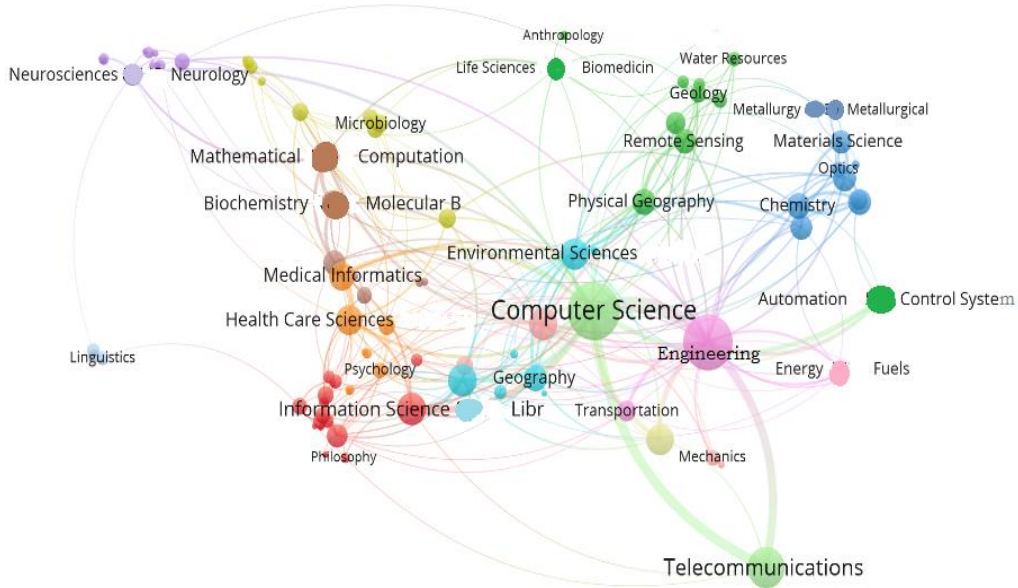
Şekil 28: 2013 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı



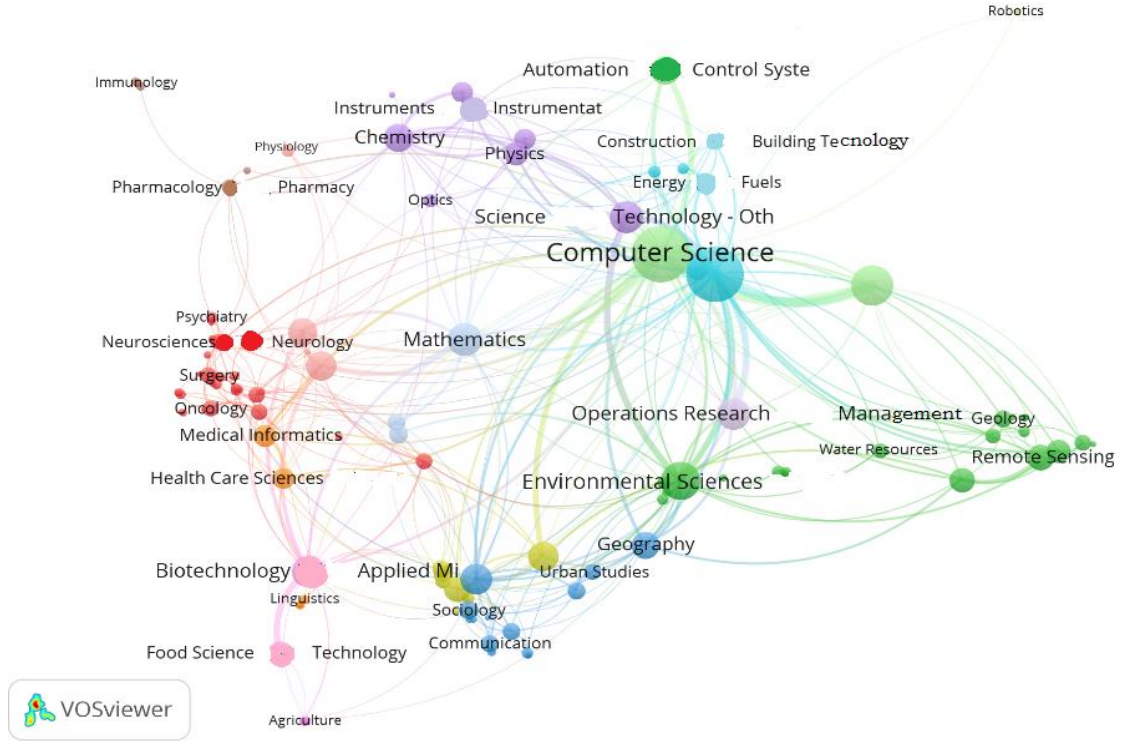
Şekil 29: 2014 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı



Şekil 30: 2015 Yılında Yayınlanan Makalelerin Alan Bazında Kümeleme Yapısı



Şekil 31: 2016 Yılında Yayınlanan Makalelerin Alan Bazında Kümeleme Yapısı



Şekil 32: 2017 Yılında Yayınlanan Makalelerin Alan Bazında Kümelenme Yapısı

Şekil 28’de görüleceği gibi, 2013 yılında büyük veri konusu kısıtlı sayıda alanda kullanılmış olup en çok bilgisayar bilimleri (computer science) alanında çalışma bulunmaktadır. Bunu bilgisayar bilimleri ile de ilişkili olan mühendislik (engineering), telekomünikasyon (telecommunications), matematik (mathematics) ve bilim ve teknoloji izlemektedir. Neredeyse tüm alanlar gerek direkt olarak gerek köprü alanlar vesilesi ile ağın merkezindeki bilgisayar bilimleri ile işbirliği içerisinde.

Şekil 29’da 2014 yılında, 2013 yılındaki sonuçları kapsayıcı daha geniş bir ağ gözlemlemekteyiz. Burada renkler de alanlar arasındaki kümelenme yapısını göstermektedir. Ağ içinde kırmızı renkteki sağlık bilimleri ve sosyal bilim alanları bir küme oluştururken; mor renkteki biyokimya ve moleküler biyoloji, biyofizik ve biyoteknoloji alanları bir küme oluşturmaktadırlar. Psikoloji, psikiyatri, nöroloji eczabilim yeşil renklerle bir küme oluşturmaktadırlar. Fizik, kimya ve matematik ise mavi renklerle başka bir küme oluşturuyorlar. Bu bağlamda büyük verinin daha çok bilimsel alanlarda kullanıldığı kabul edilebilir.

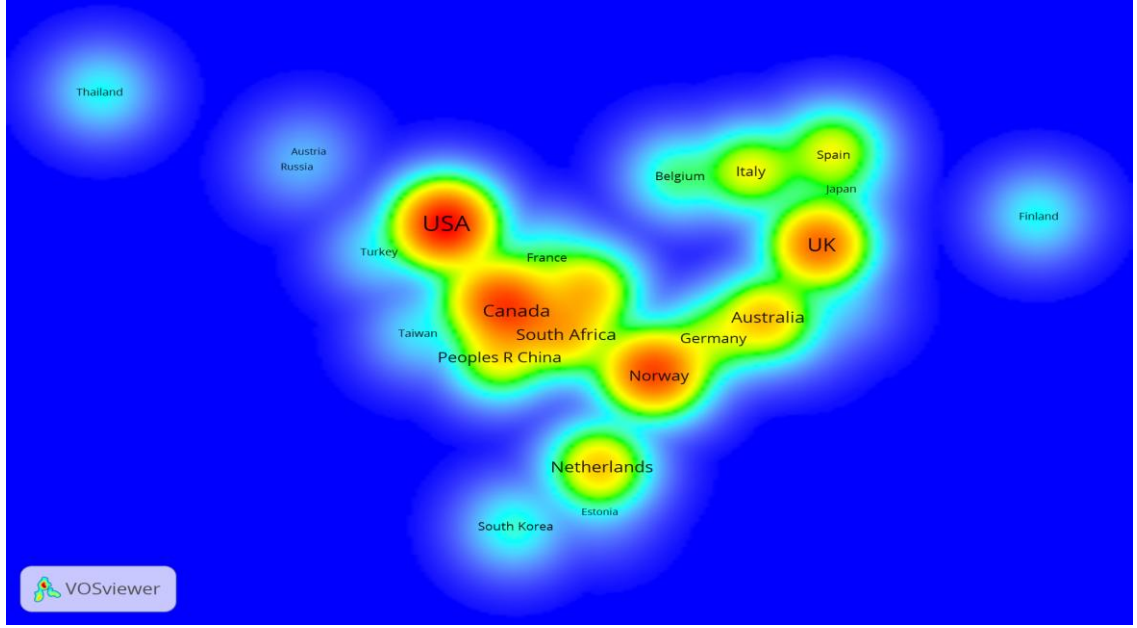
Şekil 30’da, 2015 yılında artan alan çeşitliliği ile birlikte gruplar arasında bir iç içe girme eğilimi gözlenmektedir. Alanlar arasındaki ilişkinin artması sonucunda ağın yoğunlaşmaya başladığı açıkça görülmektedir. Burada ağa katılan kentsel çalışmalar (urban studies) dikkat çekmektedir. Şekil 31’de, 2016 yılında alanlar ağının daha da yoğunlaşmakta olduğu otomatik kontrol sistemlerinin de sürece katıldığını görmekteyiz. Çevresel bilimler, fiziksel coğrafya, su kaynakları ve jeoloji alanlarının ağa katıldığı dikkat çekmektedir. Burada yapılan çalışmalardan çevre bilincinin artmakta olduğu sonucu çıkarılabilir. Şekil 32’ye göre, 2017 yılında ise robotik, uzaktan algılama (remote sensing) gibi alanlardaki çalışmalar önceki yıllara ek olarak ön plana çıkmaktadır.

Büyük veri üzerine yazılan makalelerin analize dahil edilen tüm yıllarda (2013-2017) en çok bilgisayar bilimleri (computer science) alanında yayınlanmış olduğu görülmektedir. Ardından mühendislik (engineering), telekomünikasyon (telecommunication), biyoloji (biology), matematik (mathematic) ve sağlık (medical) başta olmak üzere birçok farklı alanda çalışmalar bulunmaktadır. Bu sonuç göstermektedir ki, büyük veri sadece teknoloji ve bilişimle ilişkili değil, aynı zamanda disiplinlerarası bir konumdadır.

Son beş yıllık tablo incelendiğinde, birçok bilim dalı her geçen gün büyük veriye daha fazla ihtiyaç duymakta ve büyük veriyi daha fazla kullanmaktadır. Dolayısı ile bilimsel çalışmaların toplumsal ihtiyaçlardan kaynaklandığı düşüncesi ile yola çıkarak, toplumsal hayatta artan büyük veri ilgisinin yapılan makale çalışmalarına da yansdığı açıkça görülmektedir.

5.3.4 Ülkeler Arasındaki Sosyal Ağ

Büyük veri üzerine yapılan tarama sonucunda elde edilen makalelerin hangi ülkelerdeki araştırmacılara ait olduğunu incelemek üzere sosyal ağ yapıları yıllar (2013-2017) bazında aşağıda gösterilmektedir.

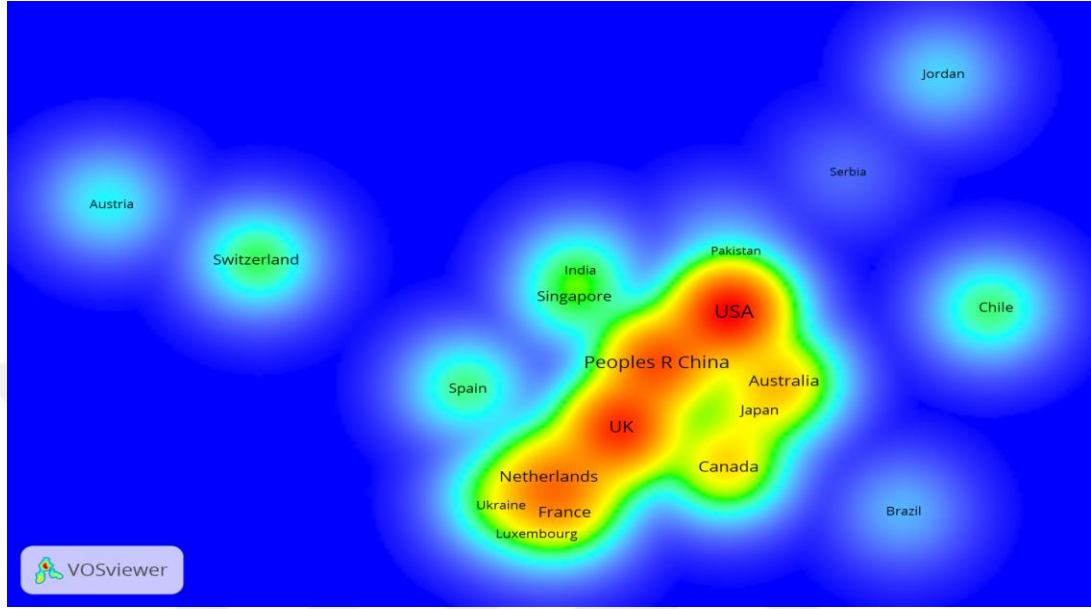


Şekil 33: 2013 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ

2013 yılında ülkeler ağının merkezinde bulunan ve en yoğun kırmızı bölgeyi oluşturan ABD (USA) en çok (142) makale yayımına sahiptir. Ardından diğer kırmızı bölgeler olan Çin Halk Cumhuriyeti (Peoples R China), (40), Kanada (Canada), (16), BK (UK), (31) ve Norveç (Norway) bölgesi yoğunluk göstermektedir. Türkiye'den 3 çalışma mevcuttur. Bu çalışmaların Amerikalı yazarlarla iş birliği içinde olduğu görülmektedir.

Taylant (Thailand), Avusturya (Austria) ve Rusya (Russia)'nın, ABD (USA) tarafında ağa katıldıkları için, ABD ile işbirliği içinde çalışmalar yapmakta oldukları açıktır. Finlandiya (Finland) ve Güney Kore (South Korea) gibi ağa yeni katılan diğer ülkeler de henüz yeni yeni bu alanda çalışmalar vermeye başlamışlardır. Şekil 33'e

göre, solda ve sağda iki farklı temel gruplaşma oluşturan ağ, Norveç dolaylarında birbiri ile birleşmektedir. Bu noktada ağdaki etkileşimin artmakta olduğu söylenebilir.

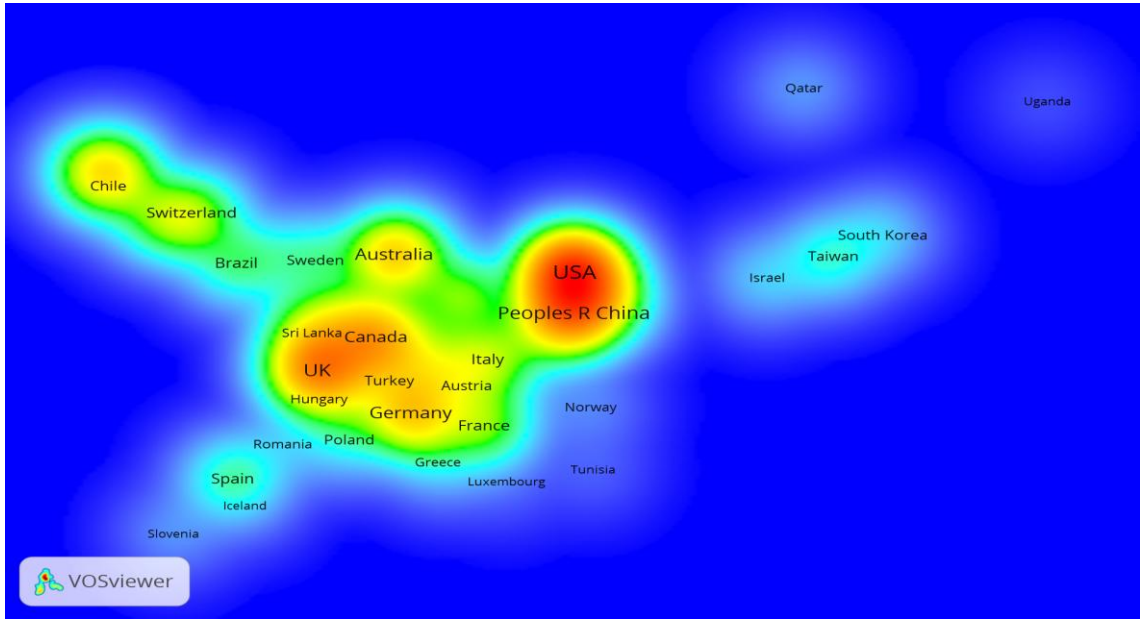


Şekil 34: 2014 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ

2014 yılında yayınlanan makalelerin yine çoğunluğu ABD'nden (330) çıkmış olup onu yine Çin Halk Cumhuriyeti (139) ve BK (66) izlemektedir. Şekil 34'e göre, ağ genel olarak ABD yakınlarında kümelenmiştir. Avustralya (Australia), 41, Fransa (France) 21 ve Hollanda (Netherlands) 18 yayın ile ağa katılmaktadırlar. 2013 yılında iki parçalı olarak gözlenen ağ yapısı 2014 yılına gelindiğinde ülkeler arası ilişkilerin artması ile tek bir parça üzerinde birleşmiştir. Bu yıl da Türkiye'den 3 makale çalışması vardır. Türkiye, Avusturya ve İsviçre (Switzerland) ile bir küme oluşturmaktadır.

Şekil 35'de, 2015 yılında ülkeler ağının kapsamını ve boyutlarını arttırmakta olduğu gözlenebilir. Burada yine birincilik ABD'nde (573) bulunmakta olup onu yine, Çin Halk Cumhuriyeti (318), BK (133) ve Kanada (62) izlemektedir. Türkiye ise yayın sayısını 13'e çıkarmıştır. Şile (Chile), İsviçre, Brezilya (Brazil)'nın kendi aralarında oluşturdukları ağ yukarı kısımdan Avustralya (Australia) ve ABD'ne; aşağı kısımdan da BK'a bağlanmaktadır.

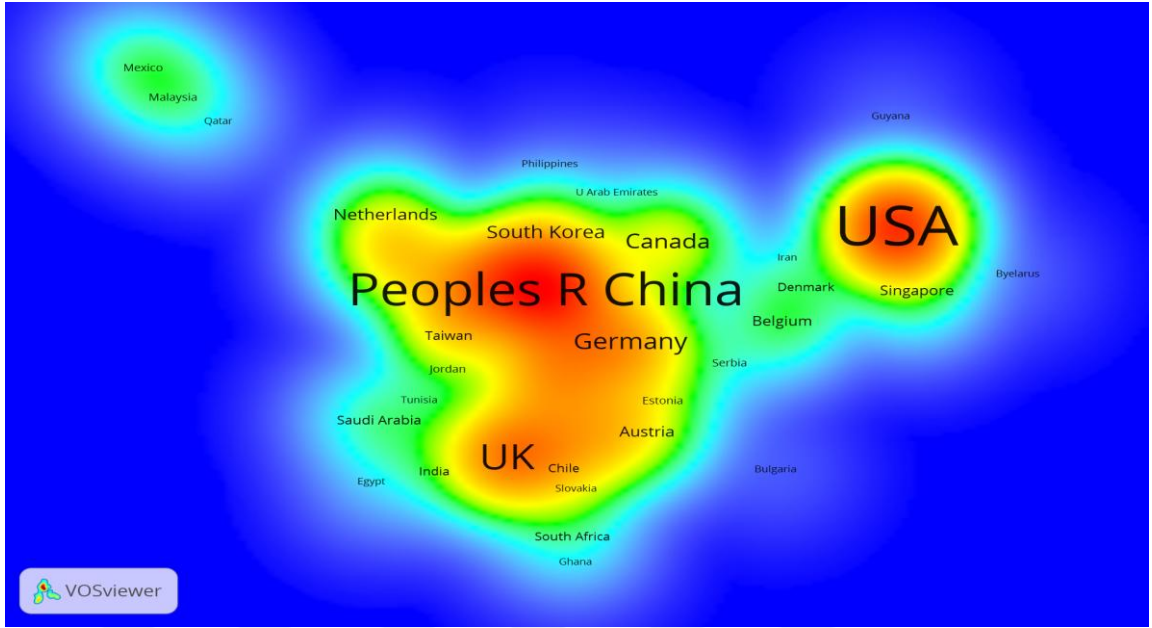
İsrail (Israel), Tayvan (Taiwan), Güney Kore (South Korea), Katar (Qatar) ve Uganda'nın ağda buldukları bölgeden ötürü ABD ile işbirliği içinde yayın üretmekte oldukları açıktır. ABD, ağın sağından Katar, İsrail, Tayvan ve Güney Kore hatta Uganda ile solundan ise Şile, İsviçre, Brezilya ve Avusturalya ile etkileşim içindedir. Ağın alt kısımlarında ise Türkiye'nin, etrafında olan BK, Almanya, Kanada ve İtalya gibi ülkeler ile işbirliği içinde yayınlar ürettiği görülmektedir.



Şekil 35: 2015 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ

Şekil 36'da, 2016 yılına gelindiğinde ağın genişlediği görülmektedir. Makale konusunda ABD'ne (758) önemli bir rakip olan Çin Halk Cumhuriyeti (552)'ndeki büyüme de gözle görülür derecededir. Burada açıktır ki ABD ağdaki gruptan ayrılmaya ve kendi etrafındaki ağını yoğunlaştırmaya başlamıştır. Çin Halk Cumhuriyeti ise BK (228) ve Almanya (107) ile yeni bir gruplanma içindedir.

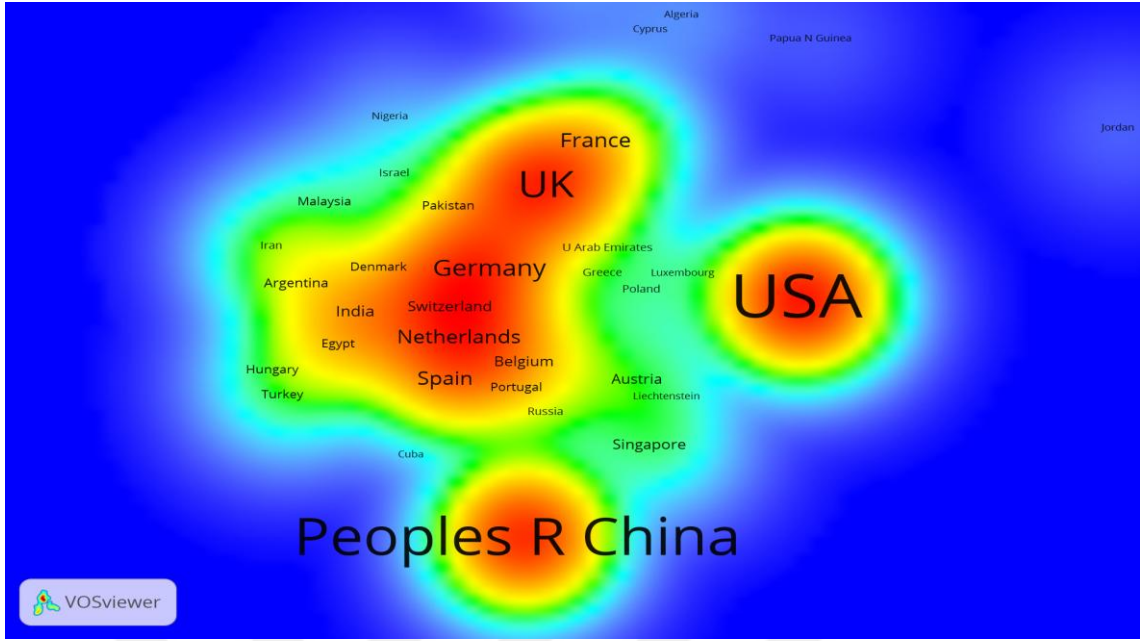
Türkiye 2016 yılında 12 yayına sahiptir. Yayın sayısı diğer ülkelerin oldukça altında kaldığı için ağda görünmemektedir. Türkiye; Estonya, Belgium (Belçika) ve Sırbistan (Serbia)'ın olduğu kümede yer almaktadır.



Şekil 36: 2016 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ

Şekil 37’de, 2017 yılında ise artık ağ en geniş ve en kapsayıcı haline ulaşmıştır. Ağda dikkat çekici olan diğer bir durum ise ABD (925) ile Çin Halk Cumhuriyeti (885)’arasındaki yayın farkı bu yıl neredeyse kapanmıştır. Dolayısı ile artan yayın sayısından ötürü Çin Halk Cumhuriyeti de ağdaki gruptan ayrılmakta ve kendi etrafında yoğun bir ağ oluşturmaktadır. Ancak her iki ülkenin de ağdaki diğer ülkeler ile iletişimleri yine yoğun olarak devam etmektedir.

Buradan büyük veri konusunda en fazla akademik ilerlemelerin bu iki ülkede olduğu sonucu çıkarılabilir. BK (281) ağda Almanya, Fransa ve Hollanda ile birlikte diğer bir yoğunluk bölgesini teşkil etmektedir. Türkiye ise 23 makale yayınına sahiptir ve Macaristan (Hungary), Mısır (Egypt) ve Hindistan (India) ile aynı kümede bulunmaktadır.



Şekil 37: 2017 Yılında Yayınlanan Makalelerin Ülkeleri Arasındaki Sosyal Ağ

Genel olarak ülkeler ağı değerlendirildiğinde, konuya olan ilginin de artması ile birlikte ülkeler arasındaki işbirliği de artmıştır. Aynı zamanda çok sayıda ülke büyük veri konusuna eğilmeye başlamış ve bu konuda yaptığı çalışmalarla ağa katılmayı ve ağda aktif olarak bulunmayı başarmıştır. Büyük veri üzerine yapılan makale yayınlarında öncü ülkelerin ise sırası ile ABD, Çin Halk Cumhuriyeti ve BK olduğu söylenebilir.

6. SONUÇ

Teknolojinin gelişmesi ve internetin yaygınlaşmasıyla birlikte büyük bir hızda büyüyen bir veri havuzu oluşmuştur. Hatta ölçümü gittikçe zorlaşan bu havuzdaki verinin neredeyse tamamı internet siteleri tarafından üretilmekte, depolanmakta ve işlenmektedir. Mobil teknolojilerin de işin içine girmesiyle gerçek zamanlı veri kayıtları birçok araştırmacı tarafından ilgi çekici hale gelmiştir. Bir süre öncesinde farklı tür ve ölçüde karmaşık yapıdaki hacimli verinin ortaya çıkmasıyla doğan büyük veri kavramı günümüzde artık bu verinin analiz edilerek geleceğe dair tahminde bulunulması süreçlerinin tamamını kapsamaktadır. Büyüyen bir unsur olarak dikkat çeken büyük veri kavramı bilim dünyasının her alanında ciddi bir yere sahip olmayı başarmış; her an kendisine kattığı yeni verilerle sürekli güncellenişi, gelişmeye açıklığı ve dijital teknolojiler sayesinde depolanabilirliğinin gelişmesiyle de uzunca bir süre daha önemini koruyacağını benzetmektedir.

Toplanan verilerin analizi mevcut durum bilgisini ortaya koyduğu için geleceğe yönelik daha iyi kararlar almayı sağlamakta ve kurum ve işletmeler arasındaki rekabeti ve hizmet kalitesini arttırmaktadır. Büyük veriyi depolama ve analiz etme süreçlerinde veri madenciliği, metin madenciliği ve makine öğrenmesi gibi ileri yazılım tekniklerine ihtiyaç duyulmaktadır. Büyük verinin öneminin anlaşılması ile birlikte dünya çapında büyük veri üzerine eğilen araştırmacıların ve yapılan çalışmalarla birlikte analiz sonuçlarının yayınlanması gibi bilimsel yayın ve faaliyetlerin sayısı da artmaktadır.

Bu çalışmada, büyük veri ile ilgili olarak yayınlanan uluslararası çalışmaların mevcut durumları hakkında bilgi verilmektedir. Bu amaçla öncelikle, uluslararası literatürde büyük veri üzerine yapılmış olan çalışmaların derlenmesi ile araştırmanın veri seti elde edilmiştir. Büyük veri yılı olarak adlandırılan 2012 yılından itibaren büyük veri yayınlarının sayısında meydana gelen hızlı artış sebebi ile araştırmaya 2013 ve 2017 yılları arası dönem dahil edilmiştir. ISI tarafından sağlanan WOS veri tabanı kullanılarak “Big Data” anahtar kelimesi ile uluslararası dizinde tarama gerçekleştirilmiştir. Ardından elde edilen yayınlar sadece makale olarak filtrelenerek

yayınlandıkları yıl, atıf, alan, yazar, ülke ve kurum değişkenleri yönünden Bibliyometrik Analize uygulanmıştır.

Elde edilen veri, SAA Teknikleri ile görselleştirilmiştir. SAA kapsamında Pajek ve VOSviewer programları kullanılmıştır. Bu çalışmalar arasındaki ağ yapılarına ilişkin elde edilen bulgular ve görsel bilgiler incelenmiş ve Türkiye'nin dünya ülkeleri arasındaki konumu sorgulanmıştır.

SAA sonucunda elde edilen bulgulara göre, araştırmaya dahil edilen dönemde (2013-2017) WOS'da yayınlanan makaleler arasında her yıl daha fazla artan sosyal ağda, yine her yıl en fazla atıf alan makalenin Dean, J. (2008) vd. tarafından gerçekleştirilen "MapReduce: Simplified Data Processing on Large Cluster" başlıklı çalışma olduğu görülmüştür. MapReduce, büyük veriyi depolayan işletim teknolojilerinin geliştirilmesi ile ilgilenen bir konudur. Büyük veri ile ilgili en çok ilgi gösterilen kaynağın bu konuda olması, büyük veriyi depolama konusunda endişeler olabileceğini göstermekte; dolayısıyla bu konunun önemine dikkat çekmektedir. Ardından ikinci sırada en çok atıf alan Boyd, D. (2012) vd. tarafından yapılan "Critical Questions for Big Data" başlıklı çalışma da büyük veri ile ilgili bilinmeyenlerin ilgi uyandırdığını ve çözülmeye çalışıldığını göstermektedir. Burada, en çok atıf alan yazarlarla büyük veri literatürüne en çok sayıda makale ile katkı sağlayan yazarların farklı olduğu da elde edilen başka bir sonuçtur.

Son beş yılda yazarlar arasında ölçekten bağımsız bir ağ yapısı meydana gelmiş ve ağın her yıl daha da sıkılaştığı görülmüştür. Buna göre, yıllar ilerledikçe yazarlar arasındaki iletişim ve sosyal ilişkiler artmakta; bilgi alışverişine önem verilmekte ve kendi içinde gruplaşarak işbirliği yapan yazar kümelenmeleri oluşmaktadır.

Büyük veri konusunda yapılan çalışmaların anahtar kelimeleri incelendiğinde öne çıkan ve büyük veri ile eşleşen "Google", "medya", "sosyal medya", "Twitter", "Facebook", "sosyal ağlar", "mobil cihazlar", "çevrimiçi öğrenme"; "veri madenciliği", "metin madenciliği", "görsel analiz", "gerçek zamanlı analiz", "doğal dil işleme"; "bulut ve bulut bilişim", "makine öğrenmesi", "oyun teorisi", "yapay

zeka”, “mapreduce”, “hadoop”, “nesnelerin interneti”, “sensör ağları”, “uzaktan algılama”; “ağ güvenliği”, “veri güvenliği”, “güvenilirlik” ve “gizlilik”dir. Bu kelime ve kelime grupları büyük verinin üretilmesi kadar kaydedilmesi, depolanması, saklanması, analiz edilmesinin ve hatta korunması ile ilgilidir. Son dönemde “dijital sağlık” kelimesinin büyük veri ile kullanılmakta olması sağlık alanında da büyük veri analizinin önem kazandığını göstermektedir. Eldeki sonuçlar doğrultusunda bu kapsamdaki çalışmaların bilgisayar bilimleri başta olmak üzere, mühendislik, telekomünikasyon, bilim ve teknoloji, matematik ve sağlık alanları başta olmak üzere birçok alanla ilişkili bir kavram olduğu kabul edilebilir. Yıllar itibariyle büyük veri ile birlikte öne çıkan kelimelerde gizlilik kavramının ön plana çıktığı görülmüştür. Buradan veri gizliliği farkındalığının ve öneminin arttığı kabul edilebilir.

Son beş yıl incelendiğinde birçok bilim dalı her geçen gün büyük veriye daha fazla ihtiyaç duymakta ve büyük veriyi daha fazla kullanmaktadır. Büyük veri incelemeleri, tüm dünyada bireyler, ticari firmalar, resmi kurumlar ve bilim camiası için önem kazanan bir konudur. Gelişen teknolojiyle birlikte büyük veriye olan ilginin, yapılan makale çalışmalarının sayısında da artışa neden olduğu bu bağlamda, büyük verinin hemen her alanda kullanılmaya başlandığı görülmektedir.

ABD, Çin Halk Cumhuriyeti ve UK, büyük veri konusunda yapılan çalışmalar bakımından en fazla sayıda çalışmaya sahip ülkeler olup birbirleri ile yoğun bir ağ içerisinde. 2017 yılında ülkeler ağı önceki yıllara göre en kapsayıcı halini almış, ülkeler arası işbirliği artmıştır. Ele alınan yıllar itibari ile ilk yıllarda ABD büyük veri konusunda en fazla yayına sahipken, son yıllarda Çin Halk Cumhuriyeti'nin de bu konuda artan bir ivmeye sahip olduğu görülmektedir. Bu bağlamda büyük veri konusunda en fazla akademik ilerlemenin bu iki ülkede olduğu söylenebilir. Ayrıca Türkiye'deki araştırmacılar tarafından büyük veri konusunda gerçekleştirilen WOS'ta taranan henüz çok sayıda makalenin bulunmadığı görülmüştür.

Türkiye'nin büyük veri konusunda diğer ülkeleri yakalayabilmesi için büyük veri konusunda uluslararası literatürde taranan çalışma miktarını arttırması gerekmektedir. Bu anlamda büyük veri konusunda farkındalığın artması gerekmektedir. Birçok alanı etkileyen büyük veri kavramı akademik anlamda da daha fazla kullanılır

hale gelmelidir. Elde edilen bulgular büyük veri konusunda çalışma yapmayı planlayan arařtırmacılar için fikir verici nitelikte olabilir. Öncelikle en fazla atıf alan ve büyük veri konusunda temel eserlerden sayılabilecek olan çalışmalar incelenebilir. Gelecek çalışmalarda büyük veriyi kullanan işletmeler üzerine örnek olay çalışmaları yapılarak büyük verinin farklı alanlarda nasıl kullanıldığına ilişkin karşılařtırmalı analizler yapılabilir. Böylelikle akademik anlamda Türkiye’de büyük veri konusunda yapılan çalışmaların sayısının artması beklenmektedir.





EKLER

Ek 1: Üniversite ve Kurumlar

Üniversite	Makale Sayısı	Üniversite	Makale Sayısı
California Üniversitesi	305	Zhejiang Üniversitesi	55
Çin Bilim Akademisi	274	Hong Kong Politeknik Üniversitesi	54
Harvard Üniversitesi	145	London College Üniversitesi	54
Texas Üniversitesi	142	Toronto Üniversitesi	54
Londra Üniversitesi	136	Wisconsin Madison Üniversitesi	54
Tsinghua Üniversitesi	132	Indiana Üniversitesi	53
Pennsylvania Commonwealth Yüksek Eğitimi	100	Va Boston Healthcare System	53
Huazhong Bilim Ve Teknoloji Üniversitesi	98	Central South Üniversitesi	50
Stanford Üniversitesi	97	Penn State Üniversitesi	50
Wuhan Üniversitesi	95	Melbourne Üniversitesi	50
Georgia Üniversitesi	94	Beijing Teknoloji Enstitüsü	49
Florida Eyalet Üniversitesi	93	South China Teknoloji Üniversitesi	48
North Carolina Üniversitesi	87	Maryland College Park Üniversitesi	48
Amerika Birleşik Devletleri Enerji Bakanlığı (DOE)	86	Dalian Teknoloji Üniversitesi	47
Massachusetts Teknoloji Enstitüsü (MIT)	84	Shenzhen Üniversitesi	47
New York Suny Eyalet Üniversitesi	81	Çin Elektronik Bilimler ve Teknoloji Üniversitesi	47
Oxford Üniversitesi	78	Duke Üniversitesi	46
California Los Angeles Üniversitesi	77	Helmholtz Alman Araştırma Merkezleri Topluluğu	46

Shanghai Jiao Tong Üniversitesi	75	Hong Kong Şehir Üniversitesi	45
Centre National De La Recherche Scientifique (CNRS)	74	Tongji Üniversitesi	45
International Business Machines (IBM)	73	Cambridge Üniversitesi	45
Michigan Üniversitesi	72	Xidian Üniversitesi	45
Illinois Üniversitesi	71	Nanjing Üniversitesi	44
Nanyang Teknoloji Üniversitesi	70	Rutgers Üniversitesi	44
Minnesota Üniversitesi	70	Sun Yat-Sen Üniversitesi	44
Washington Üniversitesi	67	Colorado Üniversitesi	43
Wisconsin Üniversitesi	66	Manchester Üniversitesi	43
Maryland Üniversitesi	66	Xi An Jiaotong Üniversitesi	43
Çin Bilim Akademi Üniversitesi	65	Commonwealth Scientific Industrial Research Organisation (CSIRO)	42
New York Üniversitesi	64	King Saud Üniversitesi	42
Beijing Posta Ve Telekomünikasyon Üniversitesi	62	Microsoft	42
Peking Üniversitesi	62	Illinois at Urbana-Champaign Üniversitesi	42
California Berkeley Üniversitesi	62	Kore Bilim Ve Teknoloji Üniversitesi (UST)	42
Sydney Üniversitesi	60	Imperial College London	41
Washington Seattle Üniversitesi	60	Massachusetts Üniversitesi	41
Beihang Üniversitesi	59	Consiglio Nazionale Delle Ricerche (CNR)	40
California San Diego Üniversitesi	59	Georgia Teknoloji Enstitüsü	

Arizona Eyalet Üniversitesi	58	Deakin Üniversitesi	39
China Geosciences Üniversitesi	58	King Abdulaziz Üniversitesi	39
National Institutes Of Health (NIH)	58	Singapur Ulusal Üniversitesi	39
Chicago Üniversitesi	58	Northwestern Üniversitesi	39
Sydney New South Wales Üniversitesi	58	California San Francisco Üniversitesi	39
Southern California Üniversitesi	58	Columbia Üniversitesi	38
Pennsylvania Üniversitesi	56	Milano Politeknik Üniversitesi	38
Sydney Teknoloji Üniversitesi	55	Carnegie Mellon Üniversitesi	37

Ek 2: Yazarlar

Yazar	Frekans	Yazar	Frekans	Yazar	Frekans
Zhang Y.	64	Wu J.	27	Li D.	20
Li Y.	51	Yang L.T	26	Ahmad A.	19
Liu Y.	49	Chang V.	25	Liu P.	19
Li X.	36	Chen J.J.	25	Liu X.	19
Liu J.	34	Kim J.	25	Song H.B.	19
Wang Y.	34	Choo K.K.R.	24	Wang C.	19
Zhang J.	32	Wang H.	24	Wang K.	19
Li J.	30	Chen M.	23	Zhang X.Y.	19
Ranjan R.	30	Chen Y.	23	Li D.	20
Wang X.	30	Li K.Q.	23	Chen L.	18
Wang L.	29	Yang J.	23	Li L.	18
Zomaya A.Y.	29	Herrera F.	22	Paul A.	18
Lee S.	27	Liu C.	22	Chen J.	17
Wang J.	27	Zhang L.	22	Kim S.	17
Wang L.Z.	27	Li P.	21	Li G.	17

Ek 3: Türkiye'deki Genel Durum

Yıl	Üniversite/ Kurum	Yayıncı Kurum/ Kuruluş	Alan	Yazar Sayısı	Anahtar Kelime Sayısı	Toplam Atıf Sayısı
2018	Bahçeşehir Üniversitesi	Elsevier	Bilgi Bilimi; Kütüphanecilik Bilimi	2	5	40
2018	Ortadoğu Teknik Üniversitesi	Elsevier	Bilgisayar Bilimleri	2	5	29
2018	Yıldırım Beyazıt Üniversitesi	Springer	Mühendislik	2	4	24
2018	Altınbaş Üniversitesi	Elsevier	Bilgisayar Bilimleri; Mühendislik	3	6	18
2018	Bilkent Üniversitesi	IEEE	Bilgisayar Bilimleri; Mühendislik	6*	4	30
2018	İstanbul Kültür Üniversitesi	TÜBİTAK	Bilgisayar Bilimleri; Mühendislik	2	6	34
2017	Bilkent Üniversitesi	IEEE	Bilgisayar Bilimleri; Mühendislik	2	4	27
2017	Marmara Üniversitesi	Elsevier	Bilgisayar Bilimleri; Mühendislik	3*	4	74
2017	Munzur Üniversitesi	MDPI	Bilim ve Teknoloji	2	4	52
2017	İzmir Ekonomi Üniversitesi	Elsevier	İşletme; Ekonomi; Bilgisayar Bilimleri	5*	13	180
2017	İzmir Ekonomi Üniversitesi	IEEE	Bilgisayar Bilimleri; Telekomünikasyon	2	6	97

2017	Boğaziçi Üniversitesi	Elsevier	Bilgisayar Bilimleri; Mühendislik	5	3	44
2017	Mimar Sinan Güzel Sanatlar Üniversitesi	BMJ	Medikal; Romatoloji	42*	9	27
2017	Yıldız Teknik Üniversitesi	ACES	Mühendislik; Telekomünikasyon	4*	4	14
2017	TÜBİTAK BİLGEM	Springer	Bilgisayar Bilimleri	1	4	53
2017	Koç Üniversitesi	SAGE	İletişim	2	5	75
2017	Lefke Avrupa Üniversitesi (Kıbrıs)	Springer	Bilgisayar Bilimleri; Mühendislik	6*	3	28
2017	Bilkent Üniversitesi	ASSOC	Bilgisayar Bilimleri	3*	1	76
2017	Bilkent Üniversitesi	Springer	Mühendislik; Tasarım; Fotoğraf Teknolojileri	4	4	15
2017	Bilkent Üniversitesi	IEEE	Bilgisayar Bilimleri; Mühendislik	4	5	41
2017	Bilkent Üniversitesi	Springer	Mühendislik; Tasarım; Fotoğraf Teknolojileri	4	6	16
2017	Bilkent Üniversitesi	Hindawi	Bilgisayar Bilimleri	2*	1	19
2017	Muğla Sıtkı Koçman Üniversitesi	TÜBİTAK	Bilgisayar Bilimleri; Mühendislik	1	3	33
2017	Türkiye Kardiyoloji Derneği	VIA	Medical; Kardiyoloji	48*	9	39

2017	Kahramanmaraş Sütçü İmam Üniversitesi	TÜBİTAK	Bilgisayar Bilimleri; Mühendislik	3*	7	20
2017	Sabancı Üniversitesi	IEEE	Bilgisayar Bilimleri	4*	4	48
2017	İstanbul Üniversitesi	PSP	Çevre Bilimleri & Ekoloji	4	4	43
2017	Ege Üniversitesi	MDPI	Fizik	4	5	28
2017	Gaziantep Üniversitesi	Springer	Medikal; Nörojoli Bilimi	10*	6	4
2016	Türk Telekom	Elsevier	Bilgisayar Bilimleri; Mühendislik; Telekomünikasyon	6*	4	36
2016	Bilkent Üniversitesi	IEEE	Bilgisayar Bilimleri; Mühendislik	4*	7	40
2016	Bilkent Üniversitesi	IEEE	Mühendislik	5	5	50
2016	Bilkent Üniversitesi	Elsevier	Bilgisayar Bilimleri	3	10	45
2016	İstanbul Medeniyet Üniversitesi	Wiley	Bilgisayar Bilimleri	3*	6	31
2016	Türk Telekom	IEEE	Mühendislik; Telekomünikasyon	6*	1	15
2016	İstanbul Teknik Üniversitesi	Elsevier	Bilgi Bilimi; Kütüphanecilik Bilimi	4*	4	32
2016	Sabancı Üniversitesi	Springer	Bilgisayar Bilimleri	3*	5	22
2016	Konda Araştırma ve	Elsevier	Şehir Çalışmaları	2*	6	41

	Danışmanlık					
2016	İstanbul Teknik Üniversitesi	IEEE	Bilgisayar Bilimleri; Mühendislik; Telekomünikasyon	8*	6	84
2016	TÜBİTAK	TÜBİTAK	Bilgisayar Bilimleri; Mühendislik	2	5	26
2016	Fatih Üniversitesi	IEEE	Bilgisayar Bilimleri	4*	8	27
2015	AveaLabs	KICS	Bilgisayar Bilimleri; Telekomünikasyon	7	5	34
2015	Özyeğin Üniversitesi	Sloan Management Review	İşletme & Ekonomi	2*	6	25
2015	TÜBİTAK	OXFORD	Biyokimya; Moleküler Biyoloji; Bilgisayar Bilimleri	3	2	9
2015	Bilkent Üniversitesi	Elsevier	Bilgisayar Bilimleri	5*	6	22
2015	Boğaziçi Üniversitesi	Elsevier	İşletme; Ekonomi; Kamu Yönetimi	2	5	47
2015	Ankara Üniversitesi	Elsevier	Kardiyoloji; Surgery; Organ Nakli	19*	-	4
2015	Sabancı Üniversitesi	PLOS	Bilim & Teknoloji	3*	2	35
2015	Sabancı Üniversitesi	Mary Ann Liebert, Inc	Biyoteknoloji; Mikrobiyoloji; Genetik; Kalıtım	2	10	45
2015	Sakarya Üniversitesi	Elsevier	Mühendislik; Radyoloji; Nükleer Tıp	5*	5	43

2015	Kocaeli Üniversitesi	IEEE	Mühendislik; Fiziksek Coğrafya; Fotoğraf Teknolojileri	1	5	27
2015	Bilkent Üniversitesi	SAGE	Uluslar arası İlişkiler	1	4	2
2015	Fatih Üniversitesi	Springer	Bilgisayar Bilimleri	3*	6	24
2015	Fırat Üniversitesi	Hindawi	Mühendislik; Enstrümantasyon	3	3	59
2014	Ortadoğu Teknik Üniversitesi	Mary Ann Liebert, Inc	Biyoteknoloji; Mikrobiyoloji; Genetik	3	3	35
2014	Kocaeli Üniversitesi	IEEE	Mühendislik; Fiziksek Coğrafya; Fotoğraf Teknolojileri	5	5	25
2014	Okan Üniversitesi	BioMed Central	Biyokimya; Moleküler Biyoloji; Biyoteknoloji	4*	10	46
2013	Fatih Üniversitesi	Elsevier	Bilgisayar Bilimleri	3*	4	33
2013	Bilkent Üniversitesi	IBM CORP	Bilgisayar Bilimleri	11*	2	37
2013	Koç Üniversitesi	Elsevier	İşletme; Ekonomi; Yöneylem Araştırmaları	2	5	46

*Yabancı Ortak Yazarlı Makaleler

KAYNAKÇA

Kitaplar

- Barabási, Albert-Laszio (2010). *Bağlantılar; İş Hayatında, Bilimde ve Günlük Yaşamda*. Çev. Nurettin Elhüseyini. 1. Baskı: Optimist Yayınları.
- Berman, J. Jules. (2013). *Principles of Big Data: Preparing, Sharing and Analyzing Complex Information*. Elsevier. 1st Edition. USA: MK Publications.
- Bülbül, Şahamet. (2013). *Tanımlayıcı İstatistik*. Gözden Geçirilmiş 3. Baskı. İstanbul: Der Yayınları.
- Çilingirtürk, Ahmet, Mete. (2011). *İstatistiksel Karar Almada Veri Analizi*. 1.Basım. Ankara: Seçkin Yayıncılık.
- David, Matthew and Sutton, Carole, D. (2004). *Social Research: The Basics*. Thousand Oaks: Sage Publications.
- Feinleib, David. (2014). *Big Data Bootcamp. The Big Data Landscape*. Apress. Berkeley. CA. eBook Business and Economics. SpringerLink. pp:15-34. Erişim Adresi: https://link.springer.com/chapter/10.1007/978-1-4842-0040-7_2
- Garfield, Eugene. (1979). *Citation Indexing: Its Theory And Application In Science, Technology And Humanities*. New York: Wiley.
- Gürsakal, Necmi. (2009). *Sosyal Ağ Analizi*. 1.Baskı. Bursa: Dora Yayıncılık.
- Gürsakal, Necmi. (2017). *Büyük Veri*. Genişletilmiş 3. Baskı. Bursa: Dora Yayıncılık.
- Hulme, E., Wyndham. (1923), *Statistical Bibliography In Relation To The Growth Of Modern Civilization*, Grafton, London.
- Marr, Bernard. (2017). *Büyük Veri İş Başında: 45 yıldız Şirket Büyük Veriyi Nasıl Kullandı? Birinci Basım*. İstanbul: MediaCat Kitapları.
- Mayer-Schönberger, Viktor ve Cukier, Kenneth. (2013). *Büyük Veri Yaşama, Çalışma ve Düşünme Şeklimizi Dönüştürecek Bir Devrim*. Çev: B. Aydın. 1. Baskı İstanbul: Paloma Yayınevi.
- Nicholas, David ve Ritchie, Maureen. (1978). *Literature and Bibliometrics* London: Clive Bingley.
- Özdoğan, Ogan. (2016). *Büyük Veri Denizi*. İş ve Yönetim Serisi 72. 1.Basım. Ankara: Elma Yayınevi.
- Sağiroğlu, Şeref ve Koç, Orhan. (2017). *Büyük Veri ve Açık Veri Analitiği: Yöntemler ve Uygulamalar*. 1. Basım. Ankara: Grafiker Yayınları.

- Scott, John P. (2000). *Social Net Work Analysis A Handbook*. Second Edition. Sage Publications.
- Thomas H. Davenport. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. Harvard Business Review.
- Wasserman, Stanley ve Faust, Katherine. 1994, *Social Network Analysis: Methods and Applications*. Press 4. Cambridge: Cambridge University.
- Yıldırım, Esen. (2017). *İstatistiksel Araştırma Yöntemleri*. 3.Baskı Ankara: Seçkin Yayıncılık.

Makaleler ve Tez Kaynakları

- Al, Umut ve Tonta, Yaşar. (2004). Atıf Analizi: Hacettepe Üniversitesi Kütüphanecilik Bölümü Tezlerinde Atıf Yapılan Kaynaklar. *Bilgi Dünyası*, 5(1):19-47.
- Al, Umut. (2008). “Türkiye’nin Bilimsel Yayın Politikası: Atıf Dizilerine Dayalı Bibliyometrik Bir Yaklaşım.” Doktora Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü. Ankara, 2008.
- Al, Umut ve Soydal, İrem. (2010). Bilgibilim Alanında Kendine Atıf Üzerine Bir Çalışma. *Bilgi Dünyası*. 11(2) 349-364.
- Ak, Mehmet, Zeki Ve Gülmez, Ahmet. (2006). Türkiye’nin Uluslar Arası Yayın Performansının Analizi. *Akademik İncelemeler Dergisi*. 1 (1). 21-50.
- Amaral, L. A. Nunes, Scala, A. Barthe’le’my M., Stanley, H. E. (2000). Classes of Small-World Networks. *Proceedings of the National Academy of Sciences*. 97(21):11149-52.
- Balkundi, Prased and Kilduff, Martin (2005). The Ties That Lead: A Social Network Approach to Leadership. *The Leadership Quarterly* 16, (2005), 941-961.
- Borgatti, P., Everett, G. and Freeman, C. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- Basaglia, Tullio (2014). What Is Bibliometrics And Why We Should Care About It?, CERN Library. Erişim Adresi: <http://Slideplayer.Com/Slide/6213346/>
- Baykal, Abdullah. (2006). Veri Madenciliği Uygulama Alanları. *D.Ü.Ziya Gökalp Eğitim Fakültesi Dergisi* 7, 95-107.
- Beaver, Donald, Deb. (2001). *Scientometrics* 52. Pp.365. Erişim Adresi: <https://Doi.Org/10.1023/A:1014254214337>

- Borgman, Christine And Furner, Jonathan. (2002). Scholarly Communication And Bibliometrics. In B. Cronin (Ed.). Annual Review Of Information Science And Technology. Vol 36. Medford, NJ: Information Today, Pp 3-72.
- Bott, Elizabeth. (1957). Family and Social Networks. London: Tavistock.
- Björneborn, Lennart. (2004). Small-World Link Structures Across An Academic Web Space: A Library And Information Science Approach. Yayımlanmamış Doktora Tezi, Royal School Of Library And Information Science, Copenhagen.
- Cobo M, López-Herrera, A., Herrera-Viedma, E. ve Herrera F. (2011). Science Mapping Software Tools: Review, Analysis, And Cooperative Study Among Tools. Journal Of The American Society For Information Science And Technology 62(7) pp.1382-402.
- Cukier, Kenneth. (2010). "Data, Data Everywhere: A Special Report on Managing Information", The Economist Newspaper. pp 1-13.
- Diebold, Franchis X. (2003). "Big Data" Dynamic Factor Models for Macroeconometric Measurement and Forecasting. Eighth World Congress of the Econometric Society. Cambridge: Cambridge University Press, 115-122.
- Diodato, Virgil. (1994). Dictionary of Bibliometrics. Binghamton, NY: Haworth Press.
- Doğan, Korcan ve Arslantekin, Sacit. (2016). Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum. DTCF Dergisi 56(1):15-36.
- Doğan, Güleda. (2017). Bibliyometri ve Atıf Analizi. Araştırma Yöntemleri. Erişim Adresi: http://www.acikders.net/pluginfile.php/6131/course/section/2834/hafta-11_bibliyometri-atif-analizi.pptx & <Http://Slideplayer.Biz.Tr/Slide/10155906/>
- Dudin, E. B. ve Smetanin, Y. G. (2011). A Review Of Cloud Computing. Scientific and Technical Information Processing, 38(4), 280–284.
- Garfield, Eugebe. (1965). Can Citation Indexing Be Automated? In Statistical Association Methods For Mechanized Documentation, Symposium Proceedings Pp.189-192.
- Garfield, Eugebe. (1980). Bradford's Law And Related Statistical Patterns. Essays Of An Information Scientist, 4, 476-483
- Gómez - Morales, Yuri, Jack. (2007) Citations and Scientific Indexing. Published Online: Wiley Online Library.
- Egghe, Leo ve Rousseau, Ronalt. (1990). Introduction To Informetrics: Quantitative Methods In Library Documentation And Information Science. Amsterdam: Elsevier Science Publishers

- Erkut, Erhan. (2013). Bibliyometrik Analiz. Erişim Adresi: <http://eresearch.ozyegin.edu.tr/bitstream/handle/10679/285/Bibliyometrik%20Analiz.pdf?sequence=1>
- Freeman, Linton C. (2004). The Development of Social Network Analysis: A Study in the Sociology of Science. Vancouver, BC: Empirical Press.
- Everett, Martin G. and Borgatti, Stephen P., (1999). The Centrality of Groups and Classes. Journal of Mathematical Sociology. 23(3): 181-201 9.
- Halevi, Gali and Moed, Henk. (2012). The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature. Erişim Adresi: <https://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/>
- Jeffery, Keith. (2014). “Data is the New Oil.” Best Practices for Data Management & Sharing. Dü. The Joint Research Centre (JRC). Ispra, Italy.
- Jesse, Lenz. (2011). The Glory Of Big Data. Popular Science. Technology. Erişim Adresi: <https://www.popsci.com/technology/article/2011-10/glory-big-data>
- John, Gantz and David, Reinsel. (2012). The Digital Universe In 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Sponsored by EMC Corporation Content for this paper is excerpted directly from the IDC iView. <http://www.emc.com/leadership/digital-universe/index.htm>.
- Kalikov, Anarbek (2006). Veri Madenciliği ve Bir E-Ticaret Uygulaması. Yüksek Lisans Tezi. Gazi Üniversitesi. Fen Bilimleri Enstitüsü.
- Koehler, Wallace. (2001). Information Science As “Little Science”: The Implications Of A Bibliometric Analysis Of The Journal Of The American Society For Information Science. Scientometrics, 51(1): 117-132.
- Ko Kilkon, Jun Lee Kyoung, Park Chisung. (2008). Rethinking Preferential Attachment Scheme: Degree centrality versus closeness centrality. Connections, June 2008, Volume 28, Issue 1.s.4.
- Kıdak, Levent, Bekir, Demir, Hüseyin ve Özdemir, Esin. (2017). Bilimsel Haritalama Yöntemiyle Hastane Yönetimi Alanındaki Tematik Değişimin Analizi. Genel Tıp Dergisi 2017. 27(2):43-50.
- Lee, In and Lee, Kyoochun. (2015). Kelley School of Business Indiana University. Science Direct. Published by Elsevier Inc. All rights reserved. Business Horizons. 58. 431-440.
- Lotka, Alfred. (1926). The Frequency Distribution Of Scientific Productivity. Journal of the Washington Academy of Sciences Vol. 16 (12) pp. 317-323.

- Manyika, James, Chui, Michael, Brown, Brad, Bughin, Jacques, Dobbs, Richard, Roxburgh, Charles and Byers, Angela, Hung. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
- McAfee, Andrew and Brynjolfsson, Erik. (2012). Big Data: The Management Revolution. Harvard Business Review. Harvard Business School Publishing Corporation. Eriřim Adresi: <http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf>
- Mohammed, A. F., Humbe, V. T., ve Chowhan, S. S. (2016). A Review of Big Data Environment and its Related Technologies. In Information Communication and Embedded Systems (ICICES), International Conference on (pp. 1-5). IEEE.
- Mutukrishnan, S. (2005). Data Stream: Algorithms and Applications. Foundations and Trends® in Theoretical Computer Science: Vol. 1: No.2:117-236.
- Nizam, Hatice ve Akın, Saliha, Sıla. (2014). Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. XIX. Türkiye'de İnternet Konferansı.
- Osareh, Farideh. (1996). Bibliometrics, Citation Analysis and Co-citation Analysis: A Review of Literature I, Libri, 46:149-158.
- Önal, Gülden, Filiz. (2017). Bildirilerin Bibliyometrik Profili. IX. Ulusal Müzik Eğitimi Sempozyumu. KEFAD Cilt 18, Sayı 3. (1079-1097).
- Özgirgin, Nuri. (2010). Uluslararası İndeksler Neden Önemli? Sağlık Bilimlerinde Süreli Yayıncılık. (27-43). Eriřim Adresi: <http://Uvt.Ulakbim.Gov.Tr/Toplanti/Uay10/Nozgirgin.Pdf>
- Price, Derek. De Sol (1963). Big science, little science. Columbia University, New York, 119-119.
- Pritchard, Alan. (1969). Statistical Bibliography Or Bibliometrics? Journal Of Documentation, 25 (4), 348-349.
- Press, Gill. (2013). A Very Shhort of Big Data. Eriřim Adresi: <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#32425fcb65a1>
- Sengupta, I.N. (1992). Bibliometrics, İnformetrics, Scientometrics And Librametrics: An Overview. Libri, 42(2):75-98.
- Sezgin, Murat ve Talaz, Lütfiye. (2016). Biliřim Devrimi. Karabük Üniversitesi Sosyal Bilimler Enstitüsü Dergisi 6(2): 559-571.
- Stanley, Milgram. (1967). "The Small-World Problem", Psychology Today, Vol.1, no.1, pp 61-67.

- Streeter, C.L. ve Gillespie, David F. (1992). *Social Network Analysis. Quantitative Methods in Social Work:State of the Art.* (s201-222). Haworth Press, Inc.
- Putnam, Robert D. (1995a). "Tuning in, Tuning out: the Strange Disappearance of Social Capital in America", *Political Sciences and Politics*, Vol: 28, No: 4, pp. 664-683.
- Tindall, David ve Wellman, Berry. (2001). *Canada as Social Structure: Social Network Analysis and Canadian Sociology.* 26 (2), s.1-2.
- Tomas, C. Almind, Peter Ingwersen, (1997) "Informetric Analyses On The World Wide Web: Methodological Approaches To 'Webometrics'", *Journal Of Documentation*, Vol. 53 Issue: 4, Pp.404-426. Erişim adresi: <https://Doi.Org/10.1108/EUM0000000007205>
- Toroczka, Zoltan. (2005). *Complex Networks The Challenge of Interaction Topology.* Los Alamos Science Number 29.
- Tudorica, Bogdan, G. and Bucur, Cristian. (2011). A comparison between several NoSQL databases with comments and notes. 2011 RoEduNet International Conference 10th Edition: Networking in Education and Research, Iasi. pp. 1-5
- Valente, Thomas, W., Coronges K., Lakon C., and Costenbader E. (2008). How Correlated Are Network Centrality Measures?. *Connecions*, June 2008, Volume 28, 1,s.24.
- Ukşul, Ertunç. (2016). *Türkiye’de Eğitimde Ölçme Ve Değerlendirme Alanında Yapılmış Bilimsel Yayınların Sosyal Ağ Analizi İle Değerlendirilmesi: Bir Bibliyometrik Çalışma.* Yüksek Lisans Tezi. Antalya 2016.
- Wallace, Danny. (1989). *Bibliometrics And Citation Analysis.* J. N. Olsgaard (Yay. Haz.). *Principles And Applications Of Information Science: For Library Professionals İçinde* (Ss.10-26). Chicago And London: American Library Association.
- Yıldırım, Esen ve Ergüt, Özlem. (2014). *Uluslararası Ekonometri, Yöneylem Araştırması Ve İstatistik Sempozyumu’nun Bibliyometrik Analizi.* Namık Kemal Üniversitesi, Sosyal Bilimler Metinleri.
- Yıldırım, Esen ve Ergüt, Özlem. (2016). *Researches on the Subject of "Violence Against Women": A Bibliyometric Analysis Since 2000.* MÜ İktisadi ve İdari Bilimler Dergisi. 38(2) pp.311-333.
- Yılmaz, Murat. (1999). *Kütüphane ve Bilgibilimi Açısından Bibliyometrinin Önemi.* İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Yayınlanmamış Yüksek Lisans Tezi, İstanbul.

- Yılmaz, Bülent, Bülbül, Samet ve Atik, Murat. (2017). Büyük Verinin (Big Data) Muhasebe Üzerindeki Etkisi Ve Muhasebeye Sağladığı Katkıların İncelenmesi. Kara Harp Okulu Bilim Dergisi. Haziran, 27(1). 79-112.
- Zan, Burcu, Umut. (2012). Türkiye’de Bilim Dallarında Karşılaştırmalı Bibliyometrik Analiz Çalışması. Yayınlanmamış Doktora Tezi. Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Ankara 2012.
- Zipf, George, Kingsley (1949). Human Behavior and the Principle of least Effort. AddisonWesley, Cambridge, USA. Reprinted: Hafner, New York, USA, 1965.

İnternet Kaynakları

A Short History Of Big Data, <https://datafloq.com/read/big-data-history/239>

Apache Mapreduce, <https://www.ibm.com/analytics/hadoop/mapreduce>

Bulut, Canan. (2016). Bulut Bilişim (Cloud Computing) Nedir? Erişim Adresi: <http://www.endustri40.com/bulut-bilisim-cloud-computing-nedir/>

Bulut Bilişim Nedir?, <http://www.yazilimnet.com/tr/blog/38/bulut-bilisim-nedir->

Bulut Bilişim (Cloud Computing)Nedir?, <http://www.endustri40.com/bulut-bilisim-cloud-computing-nedir/>

Büyük Veri Kavramı ve Karakteristik Özellikleri, <http://ab.org.tr/ab16/bildiri/66.pdf>

Cambridge Analytica Şirketi Kapanıyor, 2018, <http://www.dw.com/tr/cambridge-analytica-sirketi-kapaniyor/a-43630357>

Çiğdem, Şemsettin ve Seyrek, İbrahim Halil. (2015). İşletmelerde Büyük Veri Uygulamaları: Bir Literatür Taraması. Erişim Adresi: <https://www.researchgate.net/publication/293439800>

Facebook’un Veri Skandalı Regülatörleri Harekete Geçirdi, 2018, <http://avrupaforum.org/facebookun-veri-skandalı-regulatorleri-harekete-gecirdi/>

Facebook - Cambridge Analytica veri skandalı hakkında bilmeniz gerekenler, 2018, <http://www.haberturk.com/facebook-cambridge-analytica-veri-skandalı-hakkında-bilmeniz-gereken-her-sey-1885728-ekonomi>

Facebook skandalı büyüyor, 2018, <http://www.haberturk.com/facebook-skandalı-buyuyor-1884885#>

Facebook skandalı 87 milyon kullanıcıyı etkiledi, 2018, <https://www.bbc.com/turkce/haberler-dunya-43649116>

Facebook'un Veri Skandalı, 2018,
<http://www.turkiyegazetesi.com.tr/teknoloji/551691.aspx>

Finley, Klint. (2012). Why Everyone (Not Just Geeks) Should Care About Big Data. Erişim Adresi: <https://www.wired.com/2012/09/human-face-of-big-data/>

Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016, <https://www.gartner.com/newsroom/id/3598917> (4 Nisan 2018).

Güven, Banu. (2018). Facebook'un İmtihanı. Erişim Adresi: <http://www.dw.com/tr/banu-guven-facebook-imtihan/a-43290135>

Mapreduce, <https://www.ibm.com/analytics/hadoop/mapreduce>

Here's Everything That Happens in One Minute on the Internet, <http://www.businessinsider.com/everything-that-happens-in-one-minute-on-the-internet-2017-9>

MapReduce Nedir?, <http://devveri.com/hadoop/mapreduce-nedir>

Mathematics Awareness Month, 2012, <http://www.mathaware.org/mam/2012/>

Neden Bulut Bilişim?, <http://www.dia.com.tr/neden-bulut-bilisim-farki-ne-ayricaligi-ne/>

Library, 2014, <http://Slideplayer.Com/Slide/6213346> Öğüdücü, Ş. Gündüz. (2012) Veri Madenciliği Giriş. Erişim Adresi: <http://web.itu.edu.tr/~sgunduz/courses/verimaden/slides/d1.pdf>

Öztürk, K. (2014). Sosyal Ağların Küçük Dünyası. Erişim Adresi: <http://www.acikbilim.com/2014/04/dosyalar/sosyal-aglarin-kucuk-dunyasi-elden-aktaralim-lutfen.html>

Press Gill, 2013, A Very Short of Big Data, Erişim Adresi: <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#32425fcb65a1>

Press Release: Obama Administration Unveils "Big Data" Initiative: Announces \$200 Million In New R&D Investments, 2012, <https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unveils-big-data-initiative-announces-200>

Serrat, O. (2017). Social Network Analysis. Knowledge Solutions pp 39-43.

Total Number of Pages Indexed by Google, <https://www.statisticbrain.com/total-number-of-pages-indexed-by-google/>

Steve Lohre, How Big Data Became So Big, 2012, <http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html>

Tullio Basaglia, What Is Bibliometrics And Why We Should Care About It?, CERN
Türkiyenin Dijital Ekonomiye dönüşümü; Bulut ve Veri,
http://www.tubisad.org.tr/Tr/Library/Analizler/bulut_bilisim_dosyasi.pdf

Yapıcı, Cenk. (2010). Bulut Bilişim Dosyası. Erişim Adresi:
http://www.tubisad.org.tr/Tr/Library/Analizler/bulut_bilisim_dosyasi.pdf

Zan, Burcu, Umut. (2012). Bilgi Merkezleri ve Bibliyometri. Erişim Adresi:
<http://ww2.lib.metu.edu.tr/ek/BilgiMerkezleriveBibliyometri.pps>

<https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

<https://www-01.ibm.com/software/data/bigdata/>

<https://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

<http://www.businessinsider.com/everything-that-happens-in-one-minute-on-the-internet-2017-9>

<http://www.slideshare.net/dataspora/s-4455027>

<https://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf>

<http://andressilvaa.tumblr.com/post/87206443764/big-data-refers-to-5vs-volume>

<http://Cabim.Ulakbim.Gov.Tr/Bibliyometrik-Analiz/Bibliyometrik-Analiz-Sikca-Sorulan-Sorular/>

<http://www.dw.com/tr/cambridge-analytica-şirketi-kapanıyor/a-43630357>