

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



SOSYAL AĞLARDAN EĞİTİM DÜZEYİ TAHMİNİ

YÜKSEK LİSANS TEZİ

Rudi Miran Babayi

(Y1413.010040)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Ali GÜNEŞ

Mart, 2018









T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

Yüksek Lisans Tez Onay Belgesi

Enstitümüz Bilgisayar Mühendisliği Ana Bilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı **Y1413.010040** numaralı öğrencisi **Rudi Miran BABAYI**' nin "**SOSYAL AĞLARDAN EĞİTİM DÜZEYİ TAHMİNİ**" adlı tez çalışması Enstitümüz Yönetim Kurulunun 28.02.2018 tarih ve 2018/04 sayılı kararıyla oluşturulan jüri tarafından **gözetim** ile Tezli Yüksek Lisans tezi olarak **Kabul** edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 05/03/2018

1) Tez Danışmanı: Prof. Dr. Ali GÜNEŞ

.....
Ali Güneş

2) Jüri Üyesi : Prof. Dr. Zafer ASLAN

.....
Zafer Aslan

3) Jüri Üyesi : Yrd. Doç. Dr. Ferdi SÖNMEZ

.....
Ferdi Sönmez

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form **imzalanacaktır**. Aksi halde geçersizdir.



YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum “Sosyal Ağlardan Eğitim Düzeyi Tahmini” adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim.

Mart,2018

Rudi Miran BABAYİ







Aileme ve Destek Olan Herkese,



ÖNSÖZ

İlk olarak tez çalışmamın hazırlanmasında her türlü yardımı esirgemeyen ayrıca değerli görüş ve yorumları, rehberliği ve desteği için değerli danışman hocam sayın Prof. Dr. Ali GÜNEŞ'e Teşekkür ederim. Bu tez çalışma süresi boyunca sabrı, anlayışı ve desteği ile bana yardımcı olan ailem ve sevgilime sonsuz teşekkür ederim.

Mart,2018

Rudi Miran BABAYİ





İÇİNDEKİLER

	<u>Sayfa</u>
ÖNSÖZ	xi
İÇİNDEKİLER	xiii
KISALTMALAR	xv
ÇİZELGE LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
ABSTRACT	xxiii
1.GİRİŞ	1
2. SOSYAL AĞLAR	3
2.1 Sosyal Ağ Kavramı.....	3
2.1.1 Sosyal Ağların Özellikleri	5
2.1.2 Sosyal Ağ Analizi	7
2.1.3 Sosyal Ağlarda Veri Madenciliği.....	9
2.1.4 Sosyal Ağ Verilerine Erişim	12
2.2. Sosyal Ağ Teorileri	12
2.3. Sosyal Ağ Analizi Değerlendirmelerinde Kullanılan Ölçütler	14
2.3.1. Yoğunluk (Density)	14
2.3.2. Büyüklük (size)	15
2.3.3. Genel kümelenme katsayısı (Global Clustering Coefficient)	15
2.3.4. Derece Merkeziliği.....	15
2.3.5. Özvektör Merkeziliği (Eigenvector Centrality)	16
2.3.6. Arasındalık Merkeziliği (Betweenness Centrality)	16
2.3.7. Yakınlık Merkeziliği (Closeness Centrality)	17
2.3.8. Kümelenme Katsayısı (Clustering Coefficient)	17
2.4. Sosyal Ağ Analizi Verilerinin Sınıflandırılması	18
3.SOSYAL AĞLARDA BAĞLANTI DURUMUNUN BELİRLENMESİ	21
3.1. Sosyal Ağlarda Bağlantı Tahmini	21
3.1.1. Sosyal Ağlarda Bağlantı Tahmini Problemleri	22
3.2. Matematiksel Yöntemler	24
3.3. Öncel Çalışmalar	26
3.4. Sosyal Ağlarda İlişki ve Ağ Türleri	27
3.5. Sosyal Ağlarda Benzerlik İlişkileri	28
3.6. Sosyal Ağlarda Bağlantı Tahmininde kullanılan Yöntemler ve Benzerlik İndeksleri.....	30
3.6.1 Yerel Benzerlik İndeksleri	33
3.6.2 Genel Benzerlik İndeksleri	35
4.SOSYAL AĞLARDA EĞİTİM DÜZEYİNİN VERİ MADENCİLİĞİ İLE ANALİZİ	37
4.1. Veri Seti.....	37
4.1.1 Eğitim Düzeyi Tespiti için Filtre Kelimelerin Belirlenmesi	38
4.2. Twitter Kullanıcılarının Eğitim Düzeylerinin Analizi.....	39
4.2.1. Uygulama Kapsamında Kullanılan Programlar.....	44

5.SONUÇ VE DEĞERLENDİRMELER.....	47
KAYNAKÇA	49
EKLER.....	53
ÖZGEÇMİŞ.....	57



KISALTMALAR

SNA	: Social Network Analysis
API	: Application Programing Interface
V	: Vertex
E	: Edge
LHN2	: Leitch Holme Newton Indexing
RWR	: Yeniden Başlamalı Rastgele Yürüyüş
LRW	: Yerel Rastgele Yürüyüş
SRW	: Çakıştırılmış Rastgele Yürüyüş



ÇİZELGE LİSTESİ

Sayfa

Çizelge 3.1: Sosyal Ağ Araştırmalarında Karşılaşılan Bağ Türleri28





ŞEKİL LİSTESİ

Sayfa

Şekil 2.1: Sosyal Ağların Bileşenleri	5
Şekil 2.2: Sosyal Ağlarda Yapılabilecek İşlemlerin Sınıflandırılması	9
Şekil 2.3: Örnek Çizge Gösterimi ve Komşuluk Matrisi	14
Şekil 2.4: Derece Merkeziliği Örnek Gösterimi	16
Şekil 3.1: Bir Sosyal Ağa Ait Çizge Yapısı	21
Şekil 3.2: Bir Sosyal Ağın Zaman İçindeki Değişimi	25
Şekil 3.3: İki Mod İlişkisi Matrisi	29
Şekil 3.4: İki moddan tek moda dönüşümle bireyler arası ilişkilerin bulunması.....	30
Şekil 3.5: Bağlantı Tahmini Yöntemleri	31
Şekil 3.6: Ağ Yapısının Zaman İçinde Gösterdiği Değişimden Geleceğe Yönelik Tahmin Yapmak	32
Şekil 4.1: Twitter'a Ait Veri Tabanından Tweetlere Erişim Görseli	38
Şekil 4.2: Yüksek Öğrenime İlişkin Tweetlerin Tüm Tweetler İçerisindeki Aylık Yüzdesel Dağılımı (2016 yılı).....	42
Şekil 4.3: Yüksek Öğrenimli Olarak Tahmin Edilen Kullanıcıların Twitter'daki Toplam Kullanıcı Sayısı ile Kıyaslanması (2016 yılı).....	43



SOSYAL AĞLARDAN EĞİTİM DÜZEYİ TAHMİNİ

ÖZET

Bu çalışmada sosyal ağlardan kullanıcıların eğitim düzeylerinin tahmin edilmesi konusu ele alınmıştır. Sosyal ağlara ait verilerden farklı bilgiler çıkarılmasına ihtiyaç duyulma nedenleri, çok değişik gerekçelere dayanabileceği gibi güvenlik ve ekonomik gerekçeler bunlardan ön plana çıkanlardır. Sosyal ağlar üzerine veri madenciliği (özellikle fikir madenciliği) yöntemi ile yapılan çalışmalar bilhassa yabancı literatürde oldukça yaygınlaşmaktadır. Ancak bu konu üzerine Türkçe literatürde yapılan çalışma sayısı artsa da yeterli olmadığı değerlendirilmektedir. Bu çalışmada yapılan analiz neticesinde Twitter kullanıcılarının eğitim düzeyleri hakkında bulgulara ulaşılmıştır. Çalışmanın son bölümünde ise ulaşılan bu bulgulardan hareketle bir değerlendirme yapılarak, gelecekte yapılacak benzer çalışmalar için bu çalışmadan elde edilen tecrübeler tavsiye niteliğinde paylaşılmıştır.

Anahtar Kelimeler: *Sosyal Ağlar, Veri Madenciliği, Twitter*



ESTIMATION OF EDUCATIONAL LEVEL IN SOCIAL NETWORKS

ABSTRACT

In this study, estimation of the educational level of social network users are discussed. Different information from data of social networks can be based on many different reasons. Security and economic reasons are the forefront. Data mining (especially opinion mining) methods on social networks have been widely used in foreign literature. However, it is observed that the number of works are not enough in Turkish literature about this subject. As a result of the analysis made in this study, findings about the education levels of Twitter users have been reached. In the last part of this study, an evaluation can be made with the movement from this assessment. This assessment is shared as experience and advice for similar work to be done in the future.

Keywords: *Social Networks, Data Mining, Twitter*



1. GİRİŞ

İnsanlar için iletişim ihtiyacı daima var olmuştur. Özellikle son dönemlerde internet erişiminin yaygınlaşması ve internet tabanlı geliştirilen uygulamalarla iletişim ağırlıklı bu alana kaymaya başlamıştır. Teknolojik gelişmelerin artmasıyla birlikte veriler dijital ortamlarda tutulmaya başlanmıştır. Bu durum verilerin depolanabilmesi için gerekli hacimlerde ve yeni veri tabanı sistemlerine duyulan ihtiyaçta büyük bir artış yaşanmasına neden olmuştur [1].

İnternet tabanlı geliştirilen uygulamaların tek amacı insanlar arası iletişim ihtiyacını karşılamak olmamıştır. Bu ihtiyaca cevap vermelerinin yanı sıra, kişilerin iyi vakit geçirmelerine imkan tanıyacak birer sosyal platformlar olarak işlev görmektedirler. Bu özelliklere sahip platformlara literatürde genel olarak “sosyal ağlar” adı verilmektedir. Sosyal ağlar kişilere dünyanın her yerinden insanlarla kolaylıkla iletişim kurma imkanı sağlamaktadır. Bu ve benzeri siteler oldukça çeşitli niteliklerde olabilmektedir. Bloglar, insanların karşılıklı bilgi alışverişinde buldukları forum siteleri, online oyun oynamaya imkan tanıyan siteler vb. de bu gruba dahil etmek mümkündür.

Günümüzde en popüler sosyal platformlar olan Facebook, Twitter ve Youtube ülkemizde ve Dünya’da büyük bir kitle tarafından kullanılmaktadır. Bu ve benzeri uygulamalarda insanlar istedikleri paylaşımı yapabilmektedirler. Ancak bu durum veri tabanlarında oldukça büyük miktarlarda veri birikmesine yol açmaktadır. Sosyal ağların bu büyük veri tabanları, genel olarak dışarıdan erişime izin verilen türden veri tabanları olmaktadır. Özellikle son yıllarda sosyal ağların veri tabanlarında yer alan bu bilgilerin analiz edilerek kullanıcılar hakkında çıkarımlar yapılabilmesi üzerine çalışmalar yapılmaya başlanmıştır. Nitekim sosyal ağların ham verilerinin analiz edilerek bilgiye dönüştürülebilmesi için çeşitli programlar da geliştirilmeye başlanmıştır. Sosyal ağların veri tabanlarının analiz edilerek istenilen bilgilerin çıkarılması için geliştirilen yöntemlerden birisi de “veri madenciliği” olarak adlandırılan yöntemdir. Bu yöntemle sosyal medya kullanıcılarının profilleri analiz

edilerek mevcut sosyal gruplarla ilişkileri tespit edilebilmektedir. Benzer şekilde kullanıcılar tarafından yapılan paylaşımlardan hareketle herhangi bir konuda kullanıcının ne düşünebileceğine ilişkin çıkarım yapılabilmektedir.

Bu çalışma kapsamında da ele alınan “Twitter” gibi sosyal platformlarda ise yukarıda değinilen çıkarımların yapılabilmesi için, kullanıcılar tarafından çeşitli konularda yazılmış kısa yazıların değerlendirilebilmesi gerekmektedir. Bu durumda da literatürde sıklıkla kullanılan yöntem olarak veri madenciliği karşımıza çıkmaktadır. Bu nedenle yapılan bu çalışma kapsamında, veri madenciliği yöntemi esas alınarak twitter sosyal paylaşım sitesi kullanıcılarının yapmış oldukları paylaşımlardan hareket edilerek kullanıcıların eğitim düzeylerine ilişkin çıkarımlar yapılmaya çalışılmıştır. Sadece Türkçe paylaşımların dikkate alındığı bu çalışmada kişilerin yaptıkları paylaşımlar belirlenen filtre kelimeleri kullanma durumlarına göre analiz edilerek değerlendirilmiştir.

2. SOSYAL AĞLAR

2.1. Sosyal Ağ Kavramı

Bilişim kavramları sözlüğünde “ağ” ın tanımı şöyledir. [2].

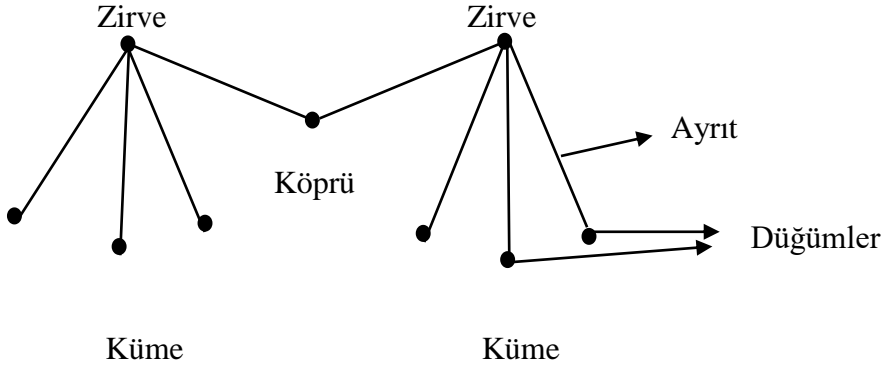
“Ağ, çok sayıda nokta ile bunlar arasındaki bağlantıların gösterildiği bir dizgeye ilişkin yapıdır.” Sosyal Ağ kavramı ise literatüre ilk olarak 1954 yılında J. A. Barnes’in çalışmasıyla girmiştir. Sosyal ağlar; her biri düğüm olarak adlandırılan bireyler, organizasyonlardan oluşmuştur. Bu düğümleri aralarındaki etkileşimlere göre birbirine bağlayan yapıya verilen isim de sosyal ağlardır. Her birisi sosyal aktörler olarak da adlandırılabilir olan bahse konu bu düğümlerin birbirlerine bağlanması farklı şekillerde olabilmektedir. Bu düğümler arası ilişkiler; kişiler arasındaki yakınlık, benzerlik, akrabalık, tanışıklık, ticari ilişki vb. olabilmektedir. David Liben-Nowell tarafından yapılan bir başka tanıma göre ise sosyal ağlar; kişiler arasında sosyal anlamdaki ilişkileri ve her türlü etkileşimi gösteren bir yapıdır[3].

İnternet imkanlarının ve erişilebilirliğinin her geçen gün artması sebebiyle iletişim alanında da internet tabanlı pek çok yeni uygulama ortaya çıkmaya başlamıştır. Bu durum bireylerin iletişim tercihlerinin internet tabanlı bu uygulamalara kaymasına neden olmaktadır. Çünkü bu ve benzeri uygulamaların özellikleri, kullanan kişilerin iletişim ihtiyaçlarını karşılamaktan ibaret değildir. Günümüzde bu uygulamalar çoğu zaman eğlenmenin ve iyi vakit geçirmenin de bir aracı haline gelmiştir. Bu kategoride yer alan ve literatürde genel itibariyle “sosyal ağlar” olarak adlandırılan bahse konu uygulamalar, çok geniş kitleler arasında internet üzerinden kolaylıkla iletişim kurulabilmesine olanak sağlamaktadır. İnsanlar bu sosyal ağlar aracılığıyla pek çok konu (siyasi, resmi, gayri resmi, ailesel, bölgesel ya da herhangi bir başka ortak özellik üzerinden) üzerine sosyal gruplar oluşturarak, bunlar aracılığıyla bilgi paylaşımı yapabilmektedir [4].

Başka bir deyişle sosyal ağlar, insanlar arası ilişkilerin/paylaşımın sanal ortam üzerinden yapıldığı ve yönlendirildiği platformlardır [5]. Bilhassa son dönemde

internet ve bilişim teknolojilerinde yaşanan süratli gelişim nedeniyle bu ve benzeri sosyal platformların sayısında ve kullanımında büyük bir artış yaşanmıştır. İnsanlar kendilerini gerçek hayata nazaran daha rahat ifade edebildikleri bu sosyal platformları her geçen gün daha yoğun olarak kullanmaktadır. Nitekim literatürde, insanların vakitlerinin büyük çoğunluğunu bu ve benzeri uygulamalar başta olmak üzere internette geçirdikleri üzerine çalışmalar yapılmaya başlanmıştır. Sosyal ağların kişilere sunduğu imkanların başında, sanal ortamdan resim, müzik, video vb. paylaşımlar yapılabilmesi gelmektedir. Bu ve benzeri paylaşımlarla insanlar sanal bir ağ üzerinden sosyallik imkânı yakalamış olmaktadır. Buna ek olarak çok farklı ülke ve coğrafyalardan insanların da çok kolay bir şekilde iletişim sağlayabilmeleri konusunda büyük bir imkan sunmaktadır [6].

İnternet teknolojisinin insanlığın hizmetine sunulmasıyla birlikte kişiler arasındaki iletişimlerde zaman ve mesafe kısıtını kaldıran birçok uygulama ortaya çıkmıştır. Bu uygulamalar sayesinde günümüzdeki iletişimin büyük çoğunluğu bu uygulamalar aracılığıyla gerçekleştirilmeye başlanmıştır. Ancak bahse konu uygulamaların tek amacı insanlar arası iletişim ihtiyacını karşılamak da değildir. Bu ihtiyacın karşılanmasının yanı sıra insanların eğlenceli vakit geçirmesini de sağlayan bu uygulamalara da genel olarak sosyal ağlar denilmektedir. Bugün sosyal ağ denildiğinde ilk olarak, kişilerin sosyal platformlar üzerinden çeşitli içeriklerde (resim, müzik, video vb.) paylaşımlarda buldukları iletişim amaçlı kullanılan araçlar akla gelmektedir. Genç kuşak başta olmak üzere her yaşta insanın günlük hayatının merkezinde yer almaya başlayan sosyal ağlara en yaygın örnekler; Facebook ve Twitter'dır. İnsanlar arasındaki siyasi, resmi/gayri resmi, ailevi vb. herhangi bir bağ sosyal ağlardaki ilişkileri oluşturmaktadır. Özet olarak, kişilerin birbirleri arasındaki ilişkilerini sanal olarak yaptıkları paylaşımlarla gerçekleştirmelerine olanak sağlayan yapılara "sosyal ağlar" denilmektedir. Sosyal ağlar temel olarak bazı bileşenlerden oluşmaktadır. Bu bileşenler Şekil 2.1'de gösterilmektedir.



Şekil 2.1: Sosyal Ağların Bileşenleri

Şekil 2.1 incelendiğinde görüleceği üzere daha önce ifade edilen sosyal ağ elemanları olan düğümler görülmektedir. Bu düğümler arasındaki ilişkiyi gösteren sosyal ağ çizgilerine “ayrıt” adı verilmektedir. Kümeler benzer düğümlerden oluşmuş gruplardır. “Zirve” kavramı ise birden fazla düğüm ya da kümenin etkileşim içinde olduğu düğümlerdir.

Sosyal ağlar genellikle kullanıcı merkezlidir. Sundukları uygulama ve hizmetlerin çoğunluğu ücretsiz hizmetlerdir. E-posta, mesajlaşma, oyunlar, videolar, dosya/fotoğraf paylaşımları, bloglar gibi çeşitli hizmetler sosyal ağlar vasıtasıyla sunulmaktadır. Bu ve benzeri pek çok farklı uygulama ile kullanıcıların iletişim ve ilişkileri kolaylaşır. Bu yapılırken kullanıcılara ait bilgilerin saklandığı bir veri tabanı da oluşturulur. Bu veri tabanı sayesinde kullanıcıların kişisel tercihlerine, ilgilendikleri konulara ilişkin önerilerle de karşılaşılabilir. Bunun anlamı kişilerin ilgi alanları ve paylaşımlarının bir nevi veri toplama işlemine tabi tutulmasıdır. Kişiler kendilerine ait oluşturdukları profillerle paylaştıkları bilgilerin gizlilik ayarlarını ve erişim kısıtlarını belirleyebilme imkanlarına da sahiptir.

2.1.1 Sosyal ağların özellikleri

Yaygın olarak kullanılan sosyal ağların büyük çoğunluğunda temel özellik kullanıcı odaklıdır. Bunun amacı bahse konu sosyal platformda kullanıcıların daha fazla vakit geçirmelerine olanak sağlamaktır.

Sosyal ağların genel olarak öne çıkan diğer özelliği ise sunulan hizmetlerin ve uygulamaların büyük çoğunlukla ücretsiz olmalarıdır. Elektronik posta, mesajlaşma,

video ve oyunlar, blog oluřturma, dosya ve doküman paylaşımı, fotoğraf paylaşımı vb. pek çok hizmet ücretsiz olarak kullanıcılara sunulmaktadır. Bu hizmetler ücretsiz olarak sunularak daha yaygın bir kullanım ve etkileşim sağlanmaktadır. Kullanıcılara ve yapmış oldukları işlemlere ait tüm bilgilerin yer aldığı bir veri tabanı vasıtasıyla sosyal ağlar; kullanıcılara muhtemel arkadaş/arkadaş grupları, ilgi alanına girebilecek konu ve paylaşımlar önerebilmektedir. Sosyal platformların sunduğu özelliklerden bir diğeri ise kişilere kendilerine özel bir profil oluřturabilme imkanı sunmasıdır. Oluřturulan bu profillerde kişiler isteklerine göre paylaşım ve beğeni yapabilmenin yanı sıra bu paylaşımların görülebileme kısıtlarını da (gizlilik ayarları) belirleyebilmektedir [4,5]. Ancak bu veriler başka kişilere kapalı olabilse de sosyal platformun veri tabanına kaydedilmesi dolayısıyla sosyal ağ yönetiminin bilgisi dahilindeki verilerdir. İleride daha ayrıntılı açıklanacağı üzere sosyal ağların analizi zaten veri tabanına kaydedilen bu bilgiler üzerinden yapılabilmektedir.

Sosyal ağ analizlerinin temelini oluřturan sosyal ağlarda iletişim olgusu ise genel olarak kullanıcılar tarafından gönderilen istekler ve karşı tarafın bu isteği kabul etmesiyle başlamaktadır. Ancak farklı sosyal ağlar için farklı öneri sistemleri de mevcuttur. Kişinin oluřturduğu profile, beğenilerine benzer özelliklerdeki kişi ve grupların önerilmesi şeklinde işleyen iletişim süreçleri de bulunmaktadır. Başka bir iletişime başlama yöntemi ise, tek taraflı başlanabilecek iletişimlerdir. Bu tip iletişimlerde hayranlar/takipçiler ile takip edilenler arasındaki iletişim şekli örnek olarak verilebilmektedir.

Sosyal ağlardaki kullanıcılar arası etkileşimlere “arkadaşlık” adı verilmektedir. Ancak bu terim gerçek anlamından ziyade sosyal ağlarda birbirleri arasında iletişim olan kişileri ifade etmek için kullanılmaktadır

Sosyal ağ uygulamaları konusunda literatürde öne çıkan çalışmalarda sosyal ağ uygulamaları ikiye ayrılmaktadır. Bunlar [3]:

1. Çevrimiçi sosyal ağ uygulamaları,
2. Kurum içi sosyal ağ uygulamalarıdır.

Çevrimiçi sosyal ağ uygulamalarından kasıt, kullanıcılar tarafından oluřturulan profillerin internet bağlantısı temelinde diğerkullanıcılarla etkileşimine olanak tanıyan web tabanlı bir hizmet şeklidir. Kurum içi sosyal ağ uygulamaları ise belirli kuruluşlara yönelik hazırlanmış olan “intranet (dahili ağ)” ağ tabanında diğerkurum

içi kullanıcılarla yerel ağ bağlantısı üzerinden etkileşimlerine olanak tanıyan bir ağ modelidir. Bu uygulamalar arasındaki temel fark, çevrimiçi olan sosyal ağlara internet erişimine sahip her kullanıcı ulaşabilecek durumda iken, kurum içi sosyal ağlardaki ön şart belirtilen kurumun mensubu olma zorunluluğudur [3].

2.1.2 Sosyal ağ analizi

Sosyal ağ analizi (Social Network Analysis (SNA)), sosyal ağlar aracılığıyla belirlenen ilişkilere ait edinilen bilgilerin, birtakım yöntemler kullanılarak yapılan analizleri neticesinde anlamlı sonuçların üretilmesi işlemidir. Diğer bir tanıma göre, sosyal yapıdaki ilişki ağları ve etkileri üzerinde yapılan incelemeye sosyal ağ analizi adı verilmektedir [5]

Sosyal ağ yapılarının analiz edilmesiyle birlikte sosyal ağ yapılarında düğüm olarak ifade edilen elemanlar arasında normal şartlarda görülemeyen ilişkilerin tespit edilebilmesi mümkün hale gelmektedir. Sosyal ağ analizi yöntemi sosyoloji bilimi temel dayanağı olacak şekilde gelişim göstermektedir. Sosyal ağ denildiğinde günümüzde her ne kadar internet tabanlı sosyal platform uygulamalarını anlasak da aslında sosyal ağ ve sosyal ağ analizi kavramları çok daha köklü kavramlardır.

Sosyal ağ analizi, sosyoloji, sosyal psikoloji ve antropoloji bilim dallarında yapılan çalışmalar sonucu ortaya çıkan bir kavramdır. Bugün sosyal ağ analizlerinde kullanılan kavramların büyük çoğunluğu geçmiş yıllara dayanmaktadır. Örneğin, sosyal grup kavramı 1940'lı yıllara dayanmaktadır. 1950'lerin sonlarına doğru gerçek hayattaki ilişkilerin ağlarla sembolize edilerek analiz edilmesine yönelik ilk çalışmalar yapılmaya başlanmıştır. Bu çalışmaların sonunda Moreno tarafından, birbirleriyle ilişkili olan düğüm noktalarının gösterimi için sosyal ilişki çizgeleri (sociograms) kavramı ortaya atılmıştır. Sayılan bu çalışmalar sosyal ağ analizi ile ilgili olarak yapılan ilk çalışmalardır. Günümüzde sosyal ağların incelenmesi ile ilgili (gelişen teknolojinin de yardımıyla) çok sayıda çalışma bulunmaktadır.

Sosyal ağ yapılarını oluşturan sosyal aktörler ve aralarındaki ilişkileri gösteren sosyal ilişki çizgeleri çoğu zaman sistemli ve düzenli değildir. Bu sebeple anlaşılabilirliği ve analiz edilmeleri oldukça zordur. Bu karmaşık ve anlaşılabilirliği zor olan sosyal ağ ilişkilerinin daha anlaşılabilir olmasını, son yıllarda yaşanan teknolojik gelişmelerle birlikte bu ilişkilerin dijital ortamlarda gerçekleşmeleri sağlamıştır. Çünkü bu

ilişkilerin sayısal veri tabanlarına dijital olarak kaydedilebilmeleri aralarında gözle tespiti zor olan ilişkilerin bilgisayarlar aracılığıyla gerçekleştirilebilmesine olanak sağlamıştır. Bu durum mevcut sosyal ağ koşullarına ait analizlerin yapılabilmesinin yanı sıra geleceğe yönelik tahminlerin dahi yapılabilmesini sağlamaktadır.

Sosyal ağ analizinde ağda mevcut aktörlere değil bu aktörler arası ilişkilere yoğunlaşılır. Kişiler arasındaki etkileşimlerin sayısal veriler haline getirilerek aralarındaki ilişkilerin bilimsel olarak araştırılması demek olan sosyal ağ analizleri, genel olarak belli organizasyonların meydana getirdiği ağların ilişkilerini sayısallaştırmakta kullanılır. Sosyal ağ analizlerinin yapıldığı programlara veri girişleri yapılarak elde edilecek olan çıktılarla organizasyonel ilişkiler test edilmektedir.

Sosyal ağ analizleri bugün pek çok farklı alanda kullanılmaktadır. Bunlar arasında öne çıkan alanlar şunlardır:

- Sosyolojik açıdan kişisel ve grupsal davranışlarının incelenmesi
- İnternet tabanlı ticaret yapan şirketler için (e-ticaret), potansiyel müşteri profillerinin ve satın alma eğilimlerinin tespit edilmesi,
- Hedef müşteri grubuna uygun reklam kampanyaları önerme,
- Büyük bir veri yükünün olduğu (medya takip analizi, akademik yayın analizi vb.) alanlarda ortak özelliklerine göre veri gruplandırma işlemleri,
- Bazı ülkelerde firmaların ekonomik bağlarını belirleyerek sektör analizlerinin yapılması,
- Özellikle ABD’de kişisel bazda her çeşit ikili veya çoklu ilişkileri belirleyebilmek amacıyla,
- Facebook, Twitter, Linked-in vb. sitelerde kurulan sosyal ağ gruplarında bilginin kişiler arasında nasıl ve ne yönde aktığının tespit edilmesi gibi alanlarda sosyal ağ analizleri sıklıkla kullanılmaktadır.

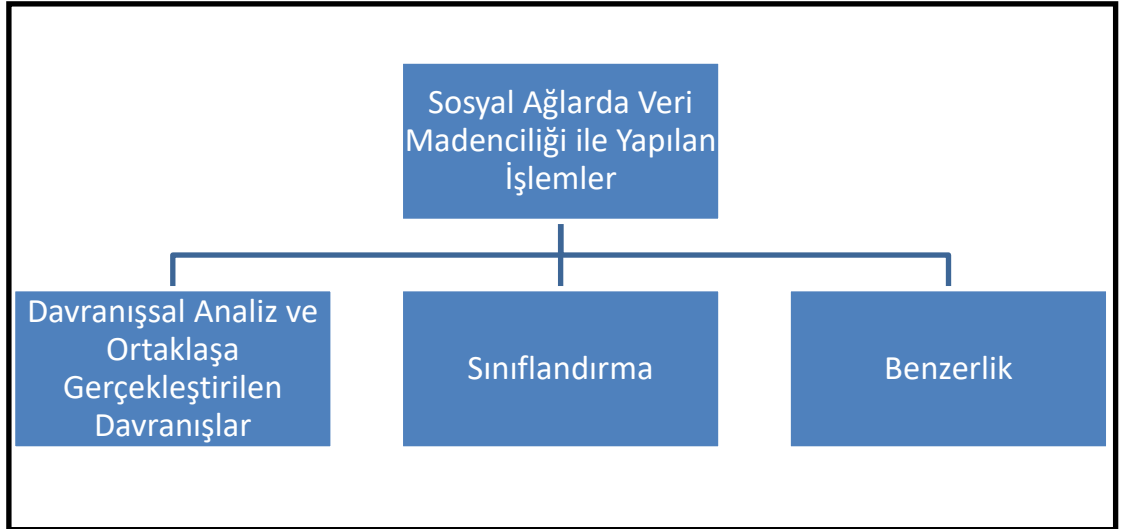
Yukarıda sayılan sosyal ağ analizi uygulamaları dışında coğrafya, işletme, biyoloji vb. pek çok alanda bu analizden faydalanılmaktadır. Ancak çalışmamız itibariyle ilişkili konular yukarıda sunulmuştur.

Sosyal ağlarda yapılan analizler, literatürde sıklıkla veri madenciliği kavramı ile anlatılmaktadır. Bu sebeple çalışma kapsamında bu kavrama değinilmesi gerektiği değerlendirilmektedir.

2.1.3 Sosyal ağlarda veri madenciliği

Sosyal ağların analizinin yapılabilmesi için, bahse konu ağlardan elde edilen veriler kullanılarak sosyal ağlar hakkında birtakım çıkarımlar yapılabilmesi gerekmektedir. Mevcut verilerden hareketle yeni çıkarımlara ulaşabilme işlemine ise literatürde “veri madenciliği” adı verilmektedir [6].

Bir sosyal ağda yer alan bireyler tarafından pek çok işlem gerçekleştirilmektedir. Bu işlemler aynı anda olabileceği gibi birbirlerinden etkilenme yoluyla da gerçekleşebilmektedir. Dolayısıyla bu işlemler arasında belirli yönlerden benzerlikler olması durumu söz konusudur. Bu işlemlerin analiz edilebilmesi için her bir işlemin özellikleri baz alınarak doğru bir şekilde sınıflandırılabilmesi büyük önem taşımaktadır. Aşağıda sunulan Şekil 2.2’de bir sosyal ağda yer alanların yapmış oldukları yorum ve yazılarından hareketle elde edilen verilerle yapılabilecek işlemler sınıflandırılmıştır.



Şekil 2.2: Sosyal Ağlarda Yapılabilecek İşlemlerin Sınıflandırılması

Sosyal ağ kullanıcılarının yapmış oldukları işlemlerden elde edilen verilerle, veri madenciliği (yeni bilgilere ulaşmak) kapsamında yapılabilecek olan uygulamalar yukarıdaki şekilde sunulmuştur. Bu işlemlere ait genel özellikler aşağıda detaylandırılmıştır.

- **Davranışsal analiz ve ortaklaşa gerçekleştirilen davranışlar;**

Bu grupta, kişilerin sosyal ağlardaki davranışları baz alınmaktadır. Ortaklaşa davranışlar, bireysel davranışlardan ziyade sosyal ağda gerçekleşen birbirleriyle bağlantılı ortaklaşa davranışları nitelendirmektedir. Başka bir deyişle, aynı sosyal ağda yer alan her bir kişi tarafından yapılan davranışların diğer kişileri etkilemesi ve bu etki zincirinin devam etmesi durumudur. Bu duruma literatürde kullanıcılar arasındaki “davranışsal ilişki” adı verilmektedir [7].

Burada sözü edilen davranışlara; belirli gruplara katılım, diğer bireylerle iletişim kurmaya yatkınlık, reklamlara tıklama durumu vb. örnek göstermek mümkündür. Ortaklaşa davranışların başka bir tanımına göre ise bu kavram, sosyal medya kullanıcılarının gerçekleştirdikleri ve maruz kaldıkları işlemlerin tümüne verilen addır. Kullanıcılar tarafından gerçekleştirilen işlemlerin en önemli özelliği ise, taklit yöntemiyle yani başka işlemlerden etkilenme yoluyla gerçekleştirilmiş olmalarıdır. Örnek olarak, bir kişinin belirli bir reklama tıkladıktan sonra alışveriş yapmasının başka bir kişi tarafından daha gerçekleştirilmiş olması durumudur. Ya da bir konu hakkında daha önce yapılan yorumların paylaşılması durumu da bu duruma örnek olarak verilebilir. Bahse konu tüm davranışlar sosyal ağların etkisiyle şekillenmektedir [7].

Yukarıda örnekleri verilen davranışsal analizlerin incelenmesi neticesinde özellikle kişilerin beğenileri üzerine fikir elde edilmesi mümkündür. Bu bilgiye dayanarak kişilere özel uygulamalar sunulması da mümkün hale gelmektedir.

- **Sınıflandırma;**

Sınıflandırmada ise özellikle sosyal ağlardaki kullanıcı gruplarına odaklanılmaktadır. Bu durumun temelinde de aslında sosyal ağlardaki karşılıklı etkileşimler rol oynamaktadır. Sosyal ağlarda belirli konular hakkında olumlu (beğenme) ya da olumsuz (beğenmeme) ilişkiler baz alınarak o grupta yer alan kişilerin belirlenen bir faaliyeti destekleyip desteklemediklerinin anlaşılması böylece mümkün olabilmektedir [8]. Bunun sosyal ağ analizindeki diğer bir adı da “etiketleme”dir.

Sosyal ağ kullanıcıları yorum yaparken ya da paylaşımlarda bulunurken, belirli sözcükleri kullanma eğiliminde olabilmektedirler. Yukarıda açıklanan etiketleme

işlemi ile bu anahtar kelimeler üzerinden gidilerek çıkarımlar yapılması sağlanmaktadır [9]. Benzer şekilde sosyal ağlarda yayınlanan konulara ilişkin beğeni durumları da analiz yapılabilmesine olanak sağlamaktadır. Burada odaklanılan durum kişilerin beğendikleri konular ya da kullandıkları kelimelerden yola çıkılarak belirli sınıflara ayrılması işleminin gerçekleştirilmesidir [6].

- **Benzerlik**

Sosyal ağlardaki bir baka veri madenciliği işlemi ise benzerlikler üzerinden gerçekleştirilmektedir. Benzerlik alanına pek çok konu girebilmektedir. Grupların, blogların, yorum ya da paylaşımların benzerliği üzerine yapılan işlemler literatürde yer bulmaktadır [6]. Örneğin ele aldığı konular bakımından benzerliklere sahip olan blog ya da web siteleri gruplandırılarak, kullanıcılara ilgi alanlarına göre birer alternatif olarak sunulabilmektedir [10]. Böylece sosyal ağları kullanan kişilere benzer konularda paylaşımlar yapan siteleri bulmalarında yardımcı olunabilmektedir [11].

Ancak benzerlikler esas alınarak yapılan değerlendirmelerde detaylı bir gruplandırma yapılmaması durumunda istenilen sonuçların elde edilememesi ihtimali bulunmaktadır. Örneğin bloglarda/sitelerde çeşitli konularda içerik paylaşımları yapılabildiğinden bunların detaylı olarak tanımlanmaması, istenilen konudaki benzer paylaşımların gözden kaçırılmasına neden olabilecektir. Bu sıkıntıyla karşılaşmamak için benzerlik işlemlerinin kategorik temelde yapılması yöntemi geliştirilmiştir [11].

Benzerlik temelli yapılan çalışmalardan bir diğeri ise, kişilerin profillerinin benzerliği üzerinden hareket edilmesidir. Kişiler, üye oldukları gruplara, okul, iş arkadaş gruplarına vb. göre etiketlenmektedir. Bu grupların benzerliğinden hareketle kişilere ilgili olabileceği alternatiflerin sunulması da mümkün hale gelebilmektedir [12]. Bazı grupların benzer olarak sınıflandırılması ise kullanıcı profillerinden kaynaklanabilmektedir.

Sosyal ağlarda grupları, ağdaki kullanıcılar arası ilişkileri görebilmek amacıyla kullanılan en kullanışlı yöntem olarak çizge yöntemi öne çıkmaktadır. Sosyal ağlarda arkadaş olan kullanıcıların genel itibarıyla benzer gruplara üyeliklerinin olması durumu söz konusudur. Örneğin bir kullanıcının gittiği gitar kursundan bir arkadaşıyla üye oldukları benzer gruplar gitar eğitimi ile ilgili olmaktadır. Bu durum, benzerliğin bir alt dalı olarak “grupların benzerliği” olarak adlandırılmaktadır.

Benzerlik konusunda uygulanan bir başka yöntem ise, gruplar üzerinden değil kullanıcılar üzerinden hareket edilen benzerliklerdir. Burada üye olunan gruplar farklı olabilmekte hatta ele alınan kişiler birbirlerini de tanımayabilmektedir. Ancak burada önemli olan kişilerin benzer beğenilere sahip olmaları konusudur. Bu duruma da literatürde “kullanıcıların benzerliği” adı verilmektedir [13].

2.1.4 Sosyal ağ verilerine erişim

Önceki bölümlerde ayrıntıları ifade edilen sosyal ağ analizine ilişkin işlemleri gerçekleştirebilmenin en önemli aşaması, sosyal ağlardan veri elde edilebilmesidir. Sosyal ağlardan ya da bloglardan bazılarının veri tabanları kullanıma açık olup herhangi bir erişim kısıntısı bulunmamaktadır. Özellikle bloglara ilişkin IceMetric [14] adı verilen veri seti yaygın olarak kullanılmaktadır. Buna benzer şekilde “The Unofficial Apple Weblogs” [15] de bloglar hakkında pek çok veriye ulaşabilme imkânı sunmaktadır. Twitter ile ilgili verilere ulaşabilmek için ise Twitter Api [16] kullanımı oldukça yaygındır. Bu öne çıkan uygulamalarla sosyal ağlara ilişkin verilere ulaşılabilir. Bunlar dışında kalan internet üzerinden yapılan diğer paylaşımlara ait veri setlerine ise geliştirilen çeşitli yazılımlarla ulaşabilmek mümkündür.

2.2 Sosyal Ağ Teorileri

Sosyal ağ analizlerinde toplanan verilerden belirli bir bilgiye ulaşabilmeyi sağlayan birtakım teknikler bulunmaktadır. Bu teknikler genel olarak şunlardır [8]:

- İstatiksel işlemler

Elde edilen verilerin tanımlanmasında en iyi teknikler istatiksel olanlardır. Tanımlayıcı istatistik analizleri bu maksatla kullanılır.

- Birliktelik kuralları (Association Rules)

Buradaki temel mantık, birlikte kullanılacak olan verilerin tespit edilerek birbirleriyle ilişkilendirilmesi işlemidir. (Online alışveriş sitelerinde bu ürünü alanlar şu ürünü de aldılar gibi bilgilendirmeler bu kapsamdaki işlemlerdir.)

- Kümeleme (Clustering)

Bu analizde benzer özelliklere sahip olan verilerin tespit edilerek aynı gruplara dahil edilmelerini sağlamaktır. Bu teknik sosyal ağ analizinde en sık kullanılan yöntemdir.

- **Sınıflandırma (Classification)**

Kümelemede olduğu gibi verilerin çeşitli yönlerden sınıflandırılarak kullanımını kolaylaştırma tekniğidir.

- **Sıralı örüntüler**

Oluşturulan veri grupları arasındaki benzerliklerin tespit edilmesini sağlamak için kullanılan tekniklerdir.

- **Bağımlılık modellemesi**

Ağ üzerindeki değişkenler arasındaki bağımlılıkların tespitini sağlamak maksadıyla kullanılan tekniklerdir.

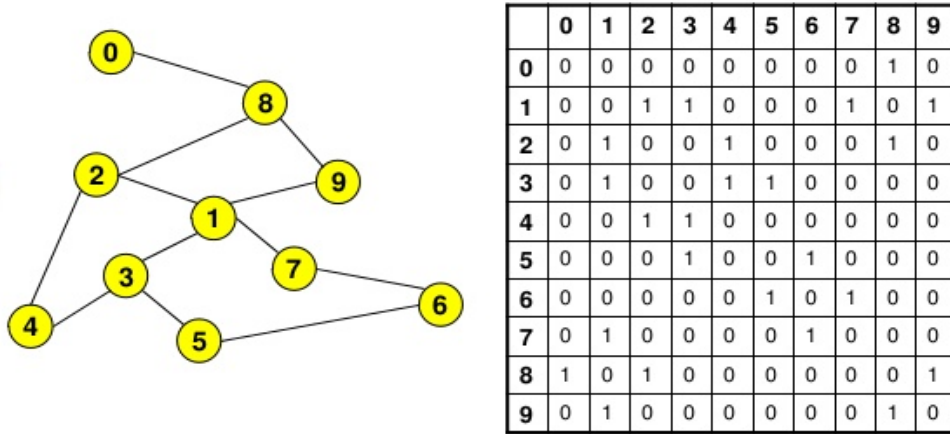
Yukarıda sunulan ve sosyal ağ analizinde kullanılan tekniklerden faydalanılarak söz konusu ağ yapılarının analiz edilmesi konusunda kullanılan teorilerden biri de çizge teorisi (Graph Theory)'dir [17]. Bu teoriye göre analizi yapılacak olan sosyal yapının tamamı bir çizge olarak kabul edilmektedir. Sosyal yapıdaki aktörler çizgenin düğümleri, aktörler arasındaki ilişkiler ise çizgenin ayrıtları olarak ifade edilmektedir. Bu şekilde çizgelerle oluşturulan bir sosyal yapıdan elde edilen matrisler matematiksel hesaplamalara tabi tutularak düğümler arası ilişkiler matematiksel olarak hesaplanmaktadır. Çıkan sonuçlar bahse konu sosyal yapı hakkında daha sağlıklı yorumlar yapılmasına olanak sağlamaktadır [17]. Bu hesaplamaların yapılmasını sağlayan bazı sosyal ağ analizi araçları bulunmaktadır. Bu araçlar bir çeşit istatistiksel yazılımlardır. Bunların başlıcaları; UCI-NET [10], NetMiner, Pajek, ORA, Stat-Net, SocNet-V, InFlow ve Keyhubs'tır.

Çizge Teoreminin genel gösterimi $G = (V, E)$ 'dir. Bu ifadedeki V (vertex) terimi düğüm kümesini, E (edge) terimi ayrıt kümesini göstermektedir. Daha önce de ifade edildiği üzere ayrıtlar birbirleriyle ilişkili olan düğümleri birbirlerine bağlama görevini üstlenirler. Sosyal ağ aktörleri (kişiler, gruplar vb.) çizge teoreminde düğümlerle ifade edilirken, ilişki çeşitleri aktörleri birbirlerine bağlayan çeşitli bağlar (tanışıklık, dostluk vb.) çizge teoreminde ayrıtlarla ifade edilmektedirler. İki düğüm arasında tek bir ayrıt bulunma zorunluluğu bulunmamaktadır. İki düğüm arası birden fazla ayrıtın olduğu çizgelere "multigraphs" denilmektedir. Ayrıtların da çeşitleri (yönlü-yönsüz) bulunmaktadır.

Çizge teoreminin sosyal ağ yapılarının gösteriminde kullanımı için temel olarak iki farklı metot kullanılmaktadır:

- Komşuluk Listesi (Adjacency List) Metodu
- Komşuluk Matrisi (Adjacency Matrix) Metodu

Komşuluk Listesi (Adjacency List) Metodu'nda bulunan her bir düğümün komşu olduğu düğümlerin listesinin oluşturulmasıdır. Komşuluk Matrisi (Adjacency Matrix) Metodu'nda ise n sayıda düğüm için $n*n$ 'lik bir matris oluşturulur. Matris oluşturulurken birbirine komşu olan noktalar "1" olarak, birbiriyle komşu olmayan noktalar ise "0" olarak yazılarak matris oluşturulur. Şekil 2.3'de çok temel düzeyde bir sosyal ağ örneği ve komşuluk matrisinin gösterimi bulunmaktadır.



Şekil 2.3: Örnek Çizge Gösterimi ve Komşuluk Matrisi

Şekil 2.3'de de görüldüğü üzere birbiriyle doğrudan bağlantısı olmayan düğümler arası ilişkiye 0 değeri, birbirine komşu düğümlere 1 değeri verilerek matris tamamlanmıştır.

2.3 Sosyal Ağ Analizi Değerlendirmelerinde Kullanılan Ölçütler

2.3.1 Yoğunluk (Density)

Sosyal ağda mevcut olan tüm bağlantıların, olması muhtemel tüm bağlantılara oranına yoğunluk adı verilmektedir. Eğer bir sosyal ağ yapısında tüm aktörler (düğümler) birbirine bağlı ise bu ağda söz konusu yoğunluk değeri 1'e eşittir.

Yoğunluk değerinin artması aktörler arası bağların güçlenmesi anlamına gelmektedir [18].

Yönlü çizgelerde yoğunluk hesabı $y=m/n(n-1)$ olarak ölçülürken, yönsüz ayrıtların olduğu ağlarda $y=m/n(n-1)/2$ olarak ölçülmektedir. Bir ağın bütününe ilişkin yapılan yoğunluk ölçümleri ile elde edilen bilgiler fazla değildir. Bir sosyal ağa ait yoğunluk ölçümleri farklı zamanlarda yapılırsa iki ölçüm arasındaki değişimlerden fikir elde etmek mümkündür. Farklı sosyal kümelerin kıyaslanmasında yoğunluk ölçüsünün kullanılması uygun değildir. Diğer bir deyişle yoğunluk ölçütü, büyüklükleri değişik olan grupların karşılaştırılması için uygun değildir.

2.3.2 Büyüklük (Size)

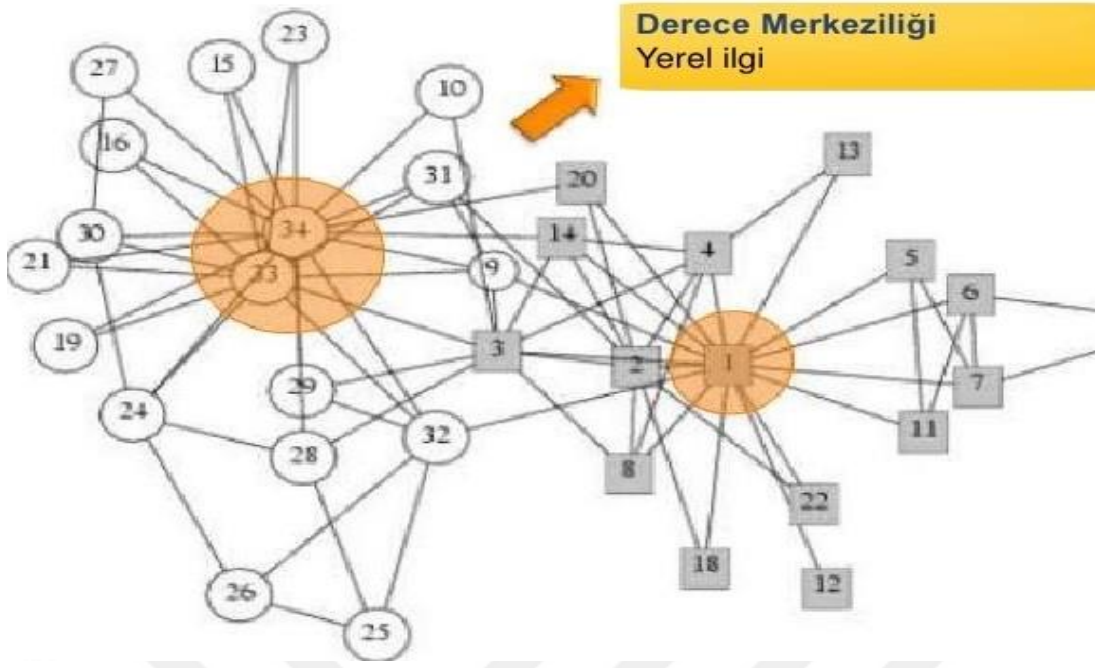
Sosyal ağlardaki düğümlerin sayısı arttıkça o sosyal ağın büyüklüğü artmaktadır. Ağda bulunan düğüm sayısı arttıkça bu düğümlerin birbirleri ile ilişkili olma ihtimalleri azalmaktadır. Diğer bir ifadeyle bir sosyal ağ, az sayıda düğüme (sosyal aktöre) sahipse bu aktörlerin birbiri ile ilişkide olma olasılıkları yüksek, çok sayıda düğüme sahipse mevcut düğümlerin birbirleri ile ilişkide olma ihtimalleri daha düşüktür. Böyle durumlarda bazı düğümler arasında hiç ilişki dahi bulunmayabilmektedir [19].

2.3.3 Genel kümelenme katsayısı (Global Clustering Coefficient)

Genel kümelenme katsayısı sosyal ağdaki üçgen ağ sayısının, kapalı veya açık bütün üçlü ağlara oranına verilen addır. Birtakım literatürde geçişkenlik olarak da adlandırılmaktadır. Bu ölçüt, mevcut ağ yapısının kümelenmeye hangi oranda uygun olduğunun tespiti için kullanılmaktadır [20].

2.3.4 Derece merkeziliği

Mevcut bir ağdaki her bir aktörün derecesi, kendisiyle olan bağlantıların sayısı ile ölçülmektedir. Aktörün ağdaki öneminin bir nevi göstergesidir. Genel kabule göre, bir ağda bir düğümün (aktörün) onunla ilintili bağ sayısı arttıkça bahse konu aktörün o ağ için önemi artmaktadır. Ağlarda, bağlantı sayısı çok olan aktörler aktif üyeler olarak nitelendirilmektedir. Eğer sosyal ağda kullanılan ayrıtlar yönlü ayrıtlar ise, giren (in-degree) ayrıt sayısı ve çıkan (out-degree) ayrıt sayısı ayrı ayrı hesaplanır. Bu açıklamalara ait örnek bir gösterim Şekil 2.4'te sunulmuştur.



Şekil 2.4: Derece Merkeziliği Örnek Gösterimi

2.3.5 Özvektör merkeziliği (Eigenvector Centrality)

Derece hesaplarında göz önüne alınmakta olup, aktöre ait bağlantıların eşit olmadığını göstermektedir. Başka bir deyişle bir ağdaki her bağlantının önem derecesi aynı değildir. Ağdaki önemli aktörlerle olan bağlantı diğer bağlantılara göre daha değerli olabilmektedir. Önemli aktörlerle kurulan bağlantı o aktörün de merkezileşmesini sağlamaktadır. Yani bu faktörün hesabında bağlantıda bulunan aktörlerin merkezilik dereceleri toplanarak hesaplama yapılır.

2.3.6 Arasındalık merkeziliği (Betweenness Centrality)

Bu merkezilik ölçütü hesaplanması en karışık olan ölçütlerdendir. Kısaca, ağdaki herhangi bir düğümden geçen en kısa yolların oranı ile ölçülmektedir. İlk olarak bir ağdaki mevcut düğümler arası en kısa mesafeler (geodesics) tespit edilir, daha sonra bu mesafelerin hangilerinde bahse konu düğümün yer aldığı oranlanması yapılarak arasındalık ölçütü hesaplanır. Ağ büyüklüğü arttıkça bu ölçütün hesaplanması zorlaşmakta, maliyetleri artmaktadır. Bir ağda arasındalık merkeziliği derecesi yüksek olan aktörlerin ağ içindeki önem dereceleri de yüksektir. Bir aktörün bu ölçütünün derecesi yükseldikçe ağ içerisindeki olaylara daha hâkim olduğu anlamı çıkarılmaktadır [13].

2.3.7 Yakınlık merkeziliği (Closeness Centrality)

Ağ üzerinde herhangi bir düğüm seçilerek bu düğümden, ağ üzerindeki diğer tüm düğümlere olan en kısa mesafelerin (geodesic distance) toplamı bulunarak hesaplanır. Ayrıtlar yönlü ise, en kısa mesafeler için yollar tespit edilirken bu yönlere dikkat edilir. Böyle durumlarda düğüme giriş ve düğümden çıkış için iki farklı yakınlık merkeziliği ölçütü hesaplaması yapılır. Bir ağda, bu ölçüt değeri yüksek olan düğümler ağda en merkezi konumda olan düğümlerdir. Bu düğümlerden ağın diğer düğümlerine kısa mesafelerle ulaşılabilir. Dolayısıyla daha az zamanda bu ulaşımın gerçekleştirilebilir [21].

2.3.8 Kümelene katsayısı (Clustering Coefficient)

Bu ölçüt, ağ üzerindeki bir aktör ile komşu düğümleri arasındaki bağlantının gücü hakkındaki ölçüttür. Ölçütün hesaplanması, aktörün komşuları arasındaki gerçek bağlantı sayısının muhtemel bütün bağlantılarının sayısına oranlanmasıyla hesaplanmaktadır. Bir ağ eğer tam bağlı (complete) bir ağ ise, diğer bir deyişle düğümler arası muhtemel tüm ayrıtların gerçekte de olduğu bir durum söz konusu ise, bu ağda her bir aktör için kümelene ölçütü “1” dir [39].

Ölçütlere ilişkin yukarıda ifade edilenler dışında ölçütlere ve hesaplamalarına dair birtakım terimlerin bilinmesi ağ analizlerinin anlaşılması için önemlidir. Bu sebeple, ağ analizi değerlendirmelerinde kullanılan ölçütlere dair öne çıkan terimlerden daha önce açıklaması yapılmayanlar aşağıda sunulmuştur [22]:

- **Köprü:** Ağdan kaldırıldığında ağı bir noktaya ulaşımın mümkün olmadığı düğümlere verilen addır. Ağ grupları arasında yer alan alternatifsiz bağlantılardır. Ortadan kaldırılması durumunda gruplar arası bağlantı kesilmiş olmaktadır.
- **Yol Uzunluğu:** Herhangi iki düğüm arası mesafedir. Seçilen iki düğüm arasındaki mesafenin kısa olması, düğümler arasında yer alan diğer düğüm ya da ayrıtlar sayısının daha az olduğu anlamına gelmektedir. Yol uzunluğu kısa olan düğümler arasında ilinti olma ihtimali artmaktadır.
- **Bütünlük:** Birbirleriyle güçlü bağlar oluşturan düğümleri ölçen kavramdır. Aralarında güçlü bağlar olan düğümlerin benzerlikleri de artmaktadır.

- **Lokal Köprü:** Ağın bitiş noktasında komşu bulunmuyorsa bu kenara verilen addır.
- **Prestij:** Ayrıtların yönlü olması durumunda düğümün merkeziliğini belirtmek için kullanılan kavramdır.
- **Radiality:** Ağdan yeni bilgiler elde edilebilmesinin ölçüsüdür.
- **Ulaşım:** Ağda yer alan herhangi bir aktörün derecesinden diğer aktörlere ulaşabilme durumudur.
- **Yapısal Bütünlük:** Ağdan çıkarıldıklarında buldukları grupla olan bağlantının kesildiği asgari düğüm sayısıdır.

2.4 Sosyal Ağ Analizi Verilerinin Sınıflandırılması

Sosyal ağ analizlerinde sınıflandırma çalışmaları çok geniş bir alanda faaliyet göstermektedir. Sosyal ağlarda meydana gelen etkileşimlerin izlenerek, kişilerin beğenilerinin tespit edilebilmesi mümkün hale gelmektedir. Bahse konu etkileşimler olumlu ya da olumsuz olabilmektedir. Kişiler herhangi bir durumla ilgili olarak beğeni durumlarını oluşturulan linklere tıklayarak belirtebilmektedirler [8]. Bu tür eylemlere başka bir deyişle etiketleme denilmektedir. Kullanıcılar tarafından gerçekleştirilen etiketleme veya daha ayrıntılı görüş paylaşımı olan yorumlama benzeri yollarla ağ tarafından veriler elde edilmektedir. Günümüzde internet kullanımının boyutları göz önüne alındığında, herhangi bir durumla ilgili olarak toplanabilecek olan verilerin büyüklükleri de tahmin edilebilecektir. Bu kadar büyük miktarda verinin elde edildiği durumlarda bu verilerin kullanılarak analizler yapılabilmesi oldukça güçleşmektedir. Dolayısıyla çeşitli yöntemlerle elde edilen büyük miktarlardaki verilerin, benzerliklerine göre sınıflandırılmaları büyük önem arz etmektedir. Kişilerin yaptıkları yorumlarda belirli kelimeleri kullanmaları durumunda belirli gruplara dahil edilmelerinde örneğinde olduğu gibi sınıflandırmalar yapılması mümkündür.

Sosyal ağlara ait verilerin sınıflandırılmaları bir tür veri madenciliği çalışmasıdır. Veri madenciliğinin sosyal ağlardan elde edilen verilerin analizinde, birkaç kullanım alanı bulunmaktadır. Bunlar [23]:

1. Dügüm Tabanlı Sınıflandırmalar
 - 1.1. Bağlantı Tabanlı Popülerlik
 - 1.2. Bağlantı Tabanlı Sınıflandırma
 - 1.3. Dügüm Kümeleme
 - 1.4. Dügüm Tanımlama

2. Bağlantı Tabanlı Sınıflandırmalar
 - 2.1. Bağlantı Tahmini

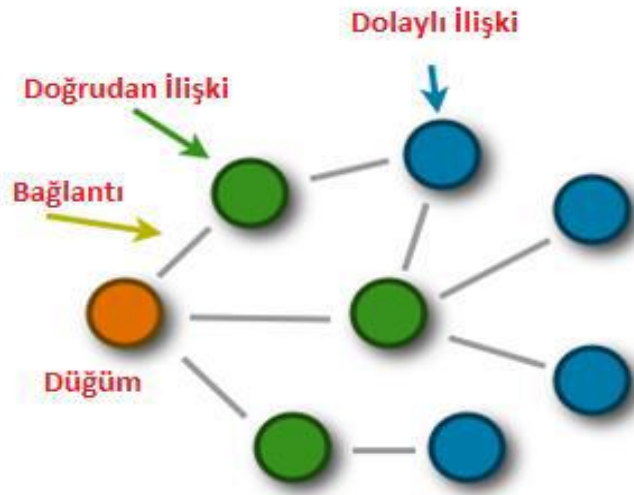
3. Çizge Tabanlı Sınıflandırmalar
 - 3.1. Alt Çizge Keşfi
 - 3.2. Çizge Sınıflandırma
 - 3.3. Üretken Modeller



3. SOSYAL AĞLARDA BAĞLANTI DURUMUNUN BELİRLENMESİ

3.1 Sosyal Ağlarda Bağlantı Tahmini

Kişilerin birbirleri arasında iletişim kurmaları sonucunda bir yapı oluşmaktadır. Bir karayolu haritası görünümüne sahip olan bu yapıdaki veriler çizge kuramına göre oluşturulmaktadır. Bu sosyal ağda yer alan her bir kişi birer düğüm olarak simgelenmektedir. Aşağıda Şekil 3.1’te gösterildiği üzere düğümler (kişiler) arası ilişkiler ise kenar olarak simgelenmektedir. Sosyal ağ yapısında hangi bireylerin kimlerle ilişkili olduğu ise, $G=(V,E)$ bağıntısı üzerinden ifade edilmektedir.



Şekil 3.1: Bir Sosyal Ağa Ait Çizge Yapısı

Sosyal ağların birtakım özellikleri bulunmaktadır. Bunlar; dinamik olma, seyrek olma ve kolektifliktir. Bu sebeple sosyal ağlarla ilgili olarak olasılık hesapları ve isabetli tahminlerde bulunabilmek zordur. Sosyal ağlarda öne çıkan (popülaritesi olan) düğüm noktalarının ağın ilk evrelerindeki bu durumunun devam edip etmeyeceğini tespit edebilmek oldukça zordur.

Herhangi bir sosyal ağda mevcut olan tüm bağlantı ve ilişkilerin tespit ve tahmininin zor olmasının yanı sıra, bahse konu sosyal ağ üzerinde mevcut yapılanlar dışında ne

tür tahminlerin yapılabileceğinin tespiti de zor bir husustur. Günümüzde artan iletişim teknolojilerinin de katkısıyla, herhangi bir sosyal ağ incelenerek ağda yer alan kişilerin sosyal ilişkileri, eğilimleri, davranışları vb. pek çok değerlendirmenin yapılması mümkün hale gelmektedir [10].

3.1.1 Sosyal ağlarda bağlantı tahmini problemleri

Sosyal ağlarda bağlantı tahmini probleminden kasıt, ağın yapısına dair tahminler yapılırken yaşanan problemlerdir. Ağ yapılarının tahmini temel olarak düğüm bilgileri ve ağın yapısal bilgilerinden meydana gelmektedir. Bir sosyal ağ yapısındaki tüm kişiler ve bunlar arasındaki ilişkiler daha önce ifade edilen çizge yapısında ifade edilmelidir. Sosyal ağdaki veriler/kişiler düğüm, ilişkiler ise kenarlar/bağlantı çizgileri olarak adlandırılmaktadır. Sosyal ağ modellerindeki düğümler vektörel olarak yapılandırılmış olan veriler içerebilmektedir. Bahse konu verileri tablolaştırarak (sütunlar→özellikleri, satırlar→değerleri) ağın analizi yapılabilmektedir. Bir sosyal ağdaki bütün düğüm noktalarının taşıdığı verilerin bilinmesi ve yapıdaki bazı ilişkilere ait bilgilere sahip olunması durumunda diğer bağlantıların/ilişkilerin de ortaya çıkarılabilmesi mümkün hale gelmektedir. Ya da benzer şekilde yeni düğümlere ve bağlantılara ilişkin bazı bilgilere sahipsek bu bilgilerden de yola çıkarak farklı düğüm ve bağlantılara ait yeni çıkarımlar yapabilmek mümkündür.

Ancak halihazırda var olan bir ilişkinin daima var olacağına dair tahmin yapabilmek de önemlidir. Çünkü sosyal ağın analizi yapılırken o anki kişi ve ilişkilere dair çıkarımlar yapılmaktadır. Ağdaki mevcut durumun ileride de geçerli olup olmamasına dair tahminlerde bulunmak da oldukça zor problemlerdendir. Başka bir deyişle dinamik ve değişken bir yapıda olan sosyal ağın analizi yapılırken belirli bir andaki fotoğrafı çekilmektedir. Sosyal ağın geleceğine dair çıkarımlar yapabilmek için, gelecekte bahse konu sosyal ağa yeni kişilerin (düğümler) ya da yeni ilişkilerin eklenmesi durumuna dair bilgilere ihtiyaç duyulmaktadır. Geleceğe ilişkin ihtiyaç duyulan bu bilgiler ise, tahmin yöntemiyle elde edilmeye çalışılmaktadır.

Bir sosyal ağa ait elimizdeki mevcut bilgilerden yola çıkarak bilinmeyen bilgilere ulaşabilmek önemli bir problemdir. Burada önemli olan, eldeki bilgilerin kullanım şekillerinin yapılacak tahminlere sunacağı katkının iyi öngörülebilmesidir. Sosyal ağ

bağlantılarının tahmin edilmesi her yönüyle oldukça zor bir işlemdir. Ancak bağlantı tahminleri sürecinde yaşanan zorlukları üç gruba ayırmak mümkündür [24]:

1. Sosyal ağın topolojik bilgisine, her bir düğüm için ek bilgilere ve eş değişkenlerin varlığına duyulan ihtiyaç,
2. Sosyal ağ yapısı itibariyle dengesiz olabilmektedir. Bunun anlamı, bilinen bağlantı sayısının olması gerekenden az olmasıdır. Bu durum dengeli verilerde kullanılan modellerin etkinliği üzerinde olumsuz bir etki yaratmaktadır.
3. Geniş sayıdaki düğüm ve kenarlar (bağlantılar) çizge olarak ölçeklendiğinde hesaplama için kullanılan modelin verimli olması bir gerekliliktir.

Bağlantı tahmininin işaret ettiği problemler şunlardır:

- Bağlantının oluşumu
- Bağlantının türü
- Bağlantının ağırlığı
- Bağlantının önemi

Literatür incelendiğinde görüleceği üzere bağlantı tahmini problemleri üzerine yapılan araştırmaların odak noktasında “bağlantı oluşumu” bulunmaktadır. Başka bir deyişle sosyal ağlarda mevcut olan iki düğüm arasında ileride bir bağ kurulması ihtimali üzerine çalışılmaktadır. Literatürde yoğun olarak bağlantı oluşumu üzerine çalışma yapılmasının temel sebebi, bağlantı oluşumu probleminin bir yönüyle diğer problemleri de kapsamasındadır. Diğer problem türlerinden olan bağlantı türü, düğümler arası bağlantıların türünü anlamaya; bağlantının ağırlığı, bağlantının etkisini hesaplamaya; bağlantının önemi ise, benzer düğümler arası bağlantının var olma durumunu tespit etmeye çalışılmaktadır [23].

Bağlantı tahmini amacıyla kullanılan modellerde özetle yapılan işlem, bilinenden hareketle bilinmeyenlerin tahmin edilmesidir. Bu durumu, bir grup öğrencinin herhangi bir dersten almış oldukları vize notlarına bakarak finalde alacakları notları tahmin etmeye benzetmek mümkündür.

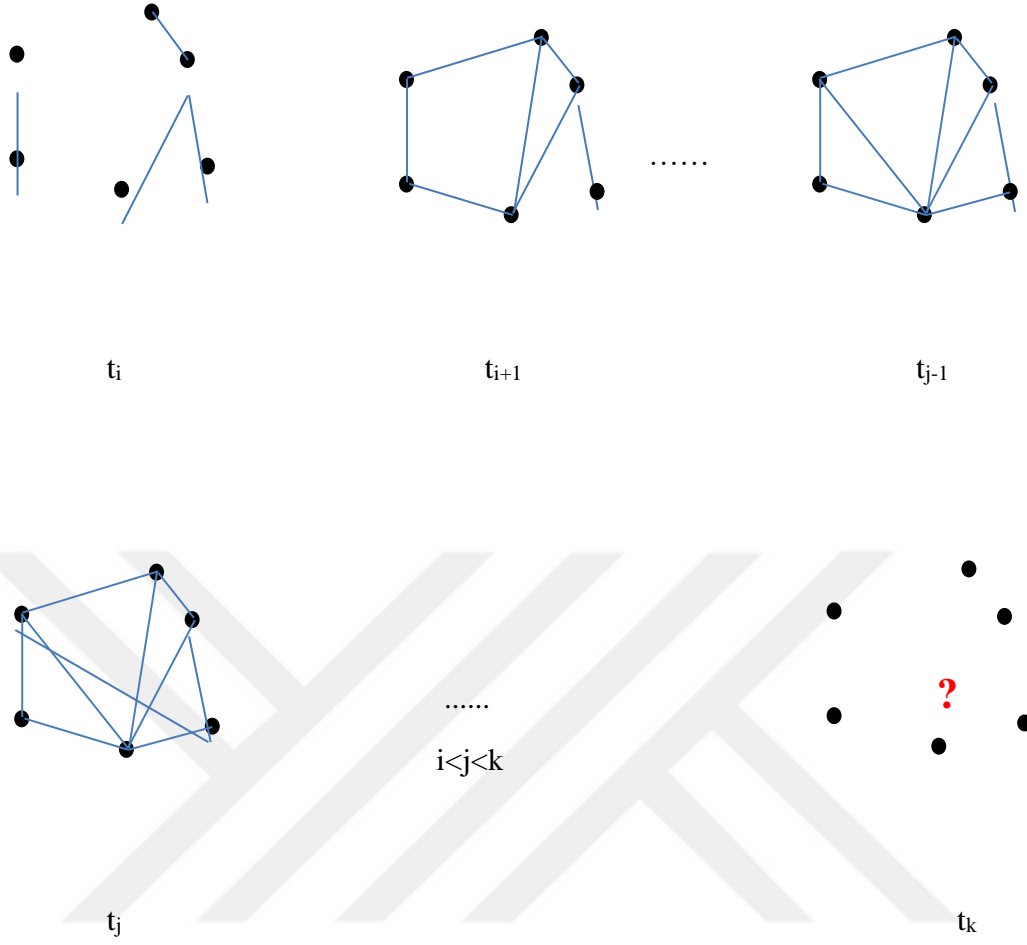
3.2 Matematiksel Yöntemler

Sosyal ağların matematiksel gösterimi aşağıdaki şekildedir. (3.1)

$$G=(V,E)$$

(3.1)

Gösterimdeki “ V ” veri örneklerini, $V= \{V_i\}_{i=1}^n$ E ise veriler arasındaki ilişkilerin gösterimidir. Her bir düğüm için düğüm çiftleri (v_i, v_j) bulunmaktadır. Bu düğüm çiftlerinden aralarında bağlantının oluşmadığı çiftler için tahminler yapılmaktadır. Oluşturulan düğüm çiftleri arasında yer alan bağlantıların önemi ise skor fonksiyonu (x,y) ile ifade edilmektedir. Buradan elde edilen fonksiyon değerleri büyükten küçüğe doğru önem sırasına göre sıralanmaktadır. Bahse konu tüm bu işlemlerin yapılmasının temel amacı, herhangi iki düğüm arasındaki benzerlik ve ilişki durumlarının tespitini sağlayabilmektir. Sosyal ağın bugünkü durumunun analizi bu işlemlerle yapıldıktan sonra, ağın geleceğine dair tahminler ise düğümler arası yollar ve bağlantı şekilleri göz önüne alınarak gerçekleştirilmektedir [10]. Sosyal ağın geleceğine dair tahminlerin doğruluğu ağın zaman içindeki değişimi göz önüne alınarak ölçülmektedir. Şekil 3.2’te bir sosyal ağın zaman içindeki değişimi gösterilmektedir [25].



Şekil 3.2: Bir Sosyal Ağın Zaman İçindeki Değişimi

Şekil 3.2 incelendiğinde görüleceği üzere sosyal ağların geçmiş dönemlerde yaşadığı değişimler göz önüne alınarak geleceğe yönelik çıkarımlar yapılabilmesi mümkündür. Bir sosyal ağda önceki dönemlerde oluşan ilişki ağları, geleceğe yönelik muhtemel ağlara ilişkin tahminler yapılmasını sağlamaktadır.

Sosyal ağlarda geleceğe dönük ilişki tahmini yapabilmek için sıklıkla kullanılan iki yöntem bulunmaktadır. Bunlar:

- Denetimli Yöntemler,
- Denetimsiz Yöntemlerdir.

3.3 Öncel Çalışmalar

Bu bölümde sosyal ağların analiz edilmesi konusundaki literatür incelenecektir. Literatürün incelenmesi, çalışma hakkında bilgi edinmek ve çıkan sonuçların literatürde öne çıkan görüşlerle uyum düzeyini tespit edebilmek açısından önemlidir. Literatürdeki bu çalışmalar, temel konu başlıkları açısından şöyledir [30]:

- Sosyal ağlarda düğümler arası ilişkileri keşfetme, benzerliklerine göre sınıflandırma,
- Bağlantı/ilişki durumlarından hareket ederek düğümlerin özelliklerinin tahmin edilmesi,
- Düğüm özellikleri bilinen kişilerin bağlantı türlerinin tahmin edilmesi,
- Sosyal ağlardaki mevcut ilişkileri analiz ederek potansiyel bağlantıların tahmin edilmesidir.

Yukarıda ana konu başlıkları sunulmuş olan literatürdeki çalışmalardan öne çıkan bazı örnekler aşağıda sunulmuştur [26]:

- Seglen tarafından yapılan çalışmada belirlenen dergilerde yer alan makalelerin genel eğilimlerinin analizi yapılmıştır [27].
- Vinkler ve Davidson tarafından yapılan araştırmada popüleritesi artan dergilerde yer alan makalelere atıf yapılma ihtimallerinin arttığı tespit edilmiştir [28].
- Liben-Nowell ve Kleinberg tarafına gerçekleştirilen çalışmada yazarlar arası bağlantılar üzerine tahmin modellemesi yapılmıştır [29].
- Popescul ve Ungar ise çalışmalarında ilişkisel öğrenme üzerine bağlantı tahmini yapılması konusunda yeni bir model sunmuşlardır [30].
- Clauset vd. tarafından yapılan çalışmada sosyal ağ yapılarının hiyerarşik durumları göz önüne alınarak, ağlarda bilinmeyen bağlantılara ait özellikler büyük bir doğrulukla tespit edilmiştir [31].
- Caragea vd. tarafından yerel bir sosyal ağ platformu üzerinde kümelenme yaklaşımı kullanılarak, sosyal ağdaki muhtemel arkadaşlıkları doğrulukla tahmin edebilen bir yazılım geliştirilmiştir [31].

- Lu tarafından yapılan çalışmada ise iki kişi arasında bir bağlantı/ilişki kurulması olasılığı tahmin edilmiştir [31].
- Davis vd. tarafından yapılan çalışmada herhangi bir hastanın geçmiş bilgilerinden yararlanılarak hastanın gelecekte yakalanması muhtemel hastalıklar belirlenmiştir [26,27].

Yukarıda sunulan çalışmalara benzer konuları baz alan pek çok çalışma literatürde bulunmaktadır. Ancak yukarıda sunulmuş olanlar öne çıkan çalışmalar olup sosyal ağ analizi konusundaki çalışmalar hakkında genel bir bilgi sahibi olunması amacıyla verilmiştir.

3.4. Sosyal Ağlarda İlişki ve Ağ Türleri

Günümüzde sosyal ağlar üzerinde mevcut olan oldukça geniş skala ilişki türleri bulunmaktadır. Sevme, takip etme, arkadaşlık, profesyonel iş ilişkisi vb. ilişki türleri bunlardan bir kaçıdır. Bu alandaki ilişki türlerinin sayısının oldukça fazla olması sebebiyle ilişki türleri bazı özellikleri baz alınarak sınıflandırılmıştır. Bahse konu sınıflandırma Çizelge 3.1’de sunulmuştur.

Çizelge 3.1: Sosyal Ağ Araştırmalarında Karşılaşılan Bağ Türleri

Bağ türü	Örnekler	Çizge/veri seti özellikleri
Benzerlikler	Aynı yerde yaşama Aynı kulüp/etkinliğe gitme Aynı cinsiyetten olma Benzer alışkanlıkları olma	İki mod ve yönlü kullanmadan önce dönüştürme gerekir dönüştürme sonucu yönsüz ilişki çoğu zaman ağırlıksız
İlişkiler	Akrabalık, evlilik, arkadaşlık, iş ilişkisi, Sevme/sevmeme	Çoğu zaman yönsüz çoğu zaman ağırlıksız
Etkileşimler	Görüşme, yardım etme, tavsiye verme	Genellikle yönlü ve ağırlıksız
Akışlar	Bilgi akışı, Şirketler arası personel akışı, ülkeler arası ticaret	Genellikle yönlü ve ağırlıklı

Çizelge 3.1’de görüldüğü üzere herhangi bir sosyal ağda mevcut olabilecek olan tüm bağlantı türleri dört sınıfa ayrılmıştır. Bu sınıflandırmanın önemli bir avantajı, analizi yapılacak olan sosyal ağda yer alan bağlantı türlerine göre hangi veri setinin kullanılması gerektiğinin belirlenmiş olmasıdır. Yukarıda yer alan tablonun büyük bir kısmı Borgatti vd. tarafından yapılan akademik çalışma sonucunda şekillenmiştir. Yapılan bu ve benzeri sınıflandırmaların sosyal ağ analizleri çalışmalarına büyük katkısı bulunmaktadır [32].

3.5. Sosyal Ağlarda Benzerlik İlişkileri

Benzerlik ilişkileri olarak adlandırılan ilişkiler aslında gerçek ilişkiler değildir. Ancak benzerlik ilişkilerinin özellikle sosyal araştırma yöntemlerinde kullanışlı olması sebebiyle kullanımı yaygındır. Aynı sosyo-kültürel çevreden gelme, benzer kötü alışkanlıklara sahip olma vb. özellikleri yönüyle benzerlikler taşıyan aktörler böylece aynı sınıflarda yer alırlar. Benzerlik ilişkileri genel olarak, arkadaşlık gibi gerçek ilişkilerle kurulu sosyal ağlarla ne ölçüde uyum sağladıklarını kontrol etme

amacıyla kullanılmaktadır. Özellikleri itibariyle bazı benzer ilişkiler gerçek ilişkilere daha yakındır. Örneğin “aynı kulübe üye olma” benzer ilişkisine sahip olan kişiler arasında gerçek ilişkiler olma olasılığı daha yüksektir. Aynı kulübe üye olma benzer ilişkisi üzerine yapılan bu çıkarım literatürde Galaskiewicz tarafından yapılan çalışmaya dayanmaktadır [33]. Yapılan bu çalışmada şirket yöneticisi ve eşlerinden oluşan bir örneklem ele alınmıştır. Bu kişilerin şirket yönetim kurullarına ve şehir kulüplerine üyelik durumları üzerine araştırmalar yapılmıştır. Çalışma sonucunda aynı kulübe üye olma durumunun, kişiler arasında gerçek ilişkilere karşılık gelme ihtimalini oldukça arttırdığı tespit edilmiştir.

Literatürde ilk olarak bu çalışmanın veri setinde kullanılan teknik bir özellik bulunmaktadır. “İki-mod” adı verilen bu özellik birinci modda kişilerin, ikinci modda kulüplerin bulunması üzerine kurulmuştur. Böylece bahse konu ilişkide daima birinci moddan ikinci moda yönelen bir ilişkinin varlığı söz konusudur. Farklı çalışmalarda üçlü mod kullanımları da bulunmaktadır. Ancak sosyal ağ üzerine yapılan analizler için en uygun modun, tekli mod olduğu ifade edilmektedir. Bu sebeple bazı sosyal ağ analizlerinde veri setinin iki-mod şeklinde olması durumunda tek-mod’a dönüştürülme işlemi gerçekleştirilmektedir [21].

Galaskiewicz tarafından yapılan bu çalışmanın esas amacı, kulüp üyelikleri ilişkilerini inceleyerek bireyler arasındaki gerçek ilişkilere ulaşabilmektir. Bu inceleme ise temel olarak matris yöntemiyle gerçekleştirilmektedir. Aşağıdaki şekilde bu konuya örnek teşkil eden temel düzeyde bir matris ilişkisi görülmektedir.

	Kulüp1	Kulüp2
Birey1	1	1
Birey2	1	0
Birey3	0	1

Şekil 3.3: İki Mod İlişkisi Matrisi

Yukarıdaki şekilde görüldüğü üzere kişilerin kulüp üyelikleri üzerinden ilişkileri matris halinde sunulmaktadır. Birey 1 olarak adlandırılan kişinin, birey 2 ile ilişkisi kulüp 1 üzerinden, birey 3 ile ilişkisi ise kulüp 2 üzerindedir. Daha önce ifade edildiği üzere yukarıda örnek iki-mod’a ait bir örnektir. Başka bir deyişle birey-kulüp ilişkisini gösteren bir ilişkidir. Birey-kulüp ilişkisinden birey-birey ilişkisine

geçebilmek için ise yukarıdaki 3*2 lik matrisi, tersi olan 2*3 lük matrisle çarparak 3*3 lük birey-birey matrisi elde edilmektedir. Başka bir deyişle kişilerin üye oldukları kulüplere ait bilgilerden yola çıkılarak kişiler arası ilişkiler hakkında bilgilere ulaşılma gerçekleştirilmektedir. Bahse konu sürece ait matris işlemleri aşağıda sunulmuştur.

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 0 \\ \hline 0 & 1 \\ \hline \end{array} \times \begin{array}{|c|c|c|} \hline 1 & 1 & 0 \\ \hline 1 & 0 & 1 \\ \hline \end{array} \rightarrow \begin{array}{|c|c|c|} \hline 2 & 1 & 1 \\ \hline 1 & 1 & 0 \\ \hline 1 & 0 & 1 \\ \hline \end{array}$$

	Birey1	Birey2	Birey3
Birey1	0	1	1
Birey2	1	0	0
Birey3	1	0	0

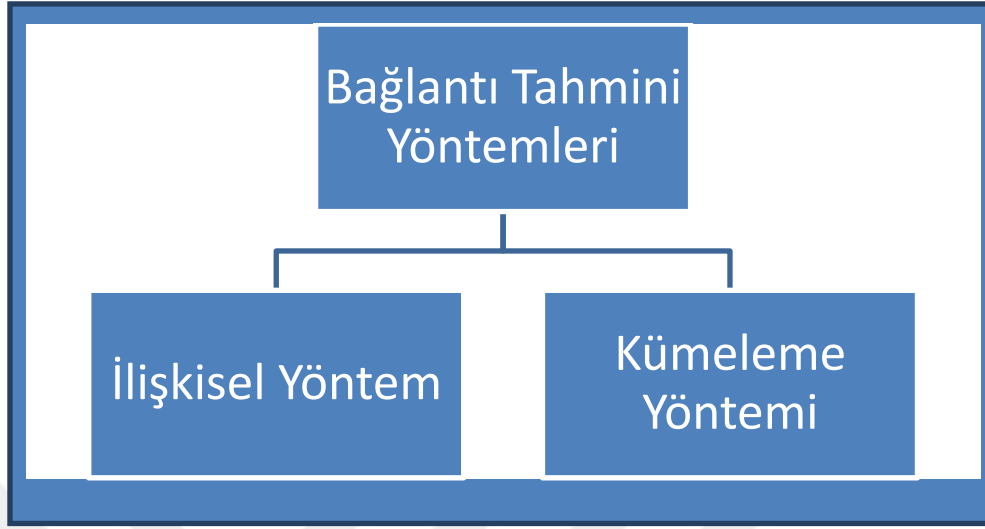
Şekil 3.4: İki moddan tek moda dönüşümle bireyler arası ilişkilerin bulunması

Şekil 3.4'de gösterilen süreç bireylerin kulüp üyelikleri ilişkilerinden bireysel ilişkilerinin çıkarımının yapıldığı sürecin matematiksel gösterimidir. Matris ve matrisin tersinin çarpımı sonucunda elde edilen 3X3'lük matris bireyler arası ilişkinin bulunduğu matristir. Ancak bu matriste her bireyin kendisiyle ilişkili olduğunu gösteren kısımlar düzeltilerek nihai tabloya ulaşılmıştır. Böylece üç kişiden oluşan bu grupta kimin kimlerle ilişkili olduğunu net olarak görmek mümkündür [32].

3.6. Sosyal Ağlarda Bağlantı Tahmininde kullanılan Yöntemler ve Benzerlik İndeksleri

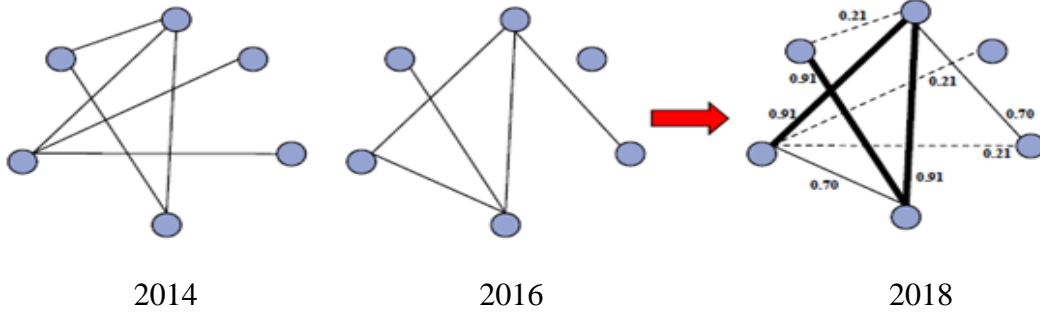
Sosyal ağların analizlerinin yapılmasında ve genel olarak bağlantı analizlerinin tahmin edilmesinde veriler incelenirken kullanılan yöntemlerin esası, verilerin tamamının ilişkisel bir tablo olarak görülmesidir. Verilerin bu şekilde gösterilmesi, müşterilere ürün satışlarında yeni önermeler yapılabilmesine olanak sağlamaktadır. Bağlantı tahmini uygulamaları amacıyla oluşturulan ilişkisel tablolardan aktörler

arası yeni ilişkilerin bulunabilmesi de mümkün hale gelmektedir. Aşağıda sunulan Şekil 3.5’de bağlantı tahmini yöntemlerinin çeşitleri gösterilmiştir.



Şekil 3.5: Bağlantı Tahmini Yöntemleri [37]

Bağlantı tahmini yapmanın farklı yöntemleri bulunmaktadır. Bu yöntemler ağ yapısı ve problemin tanımlanma şekline göre farklılık göstermektedir. Çeşitli veriler kullanılarak bağlantı tahmini yapılmaya çalışılması durumunda analiz yapılacak ağda iki varlığın olduğu kabul edilerek bu varlıklar arası ilişki olup olmadığı incelenir. Daha önce açıklanan çizge tabanlı ilişkiler kurularak bu inceleme gerçekleştirilmektedir. Çünkü çizge yapısına göre oluşturulan ağ ilişkilerinde diğer ilişki gösterimlerine oranla ilişkilerin daha rahat görülebilmesini ve olasılıksal tahminlemelerin daha isabetli yapılmasını sağlamaktadır. Bu yapıdaki gösterimlerde köprüler (bireyler) arası en kısa yol mesafesi, komşuluk yoğunluğu ve nispi önemi gibi kavramların değerlendirilmesi ön plana çıkmaktadır. Buna ek olarak çizge ilişkilerde, geçen süre zarfında oluşan ilişkiler baz alınarak geleceğe yönelik isabetli ilişki tahminleri yapılabilmektedir. Buna yönelik örnek gösterim Şekil 3.6’da sunulmuştur.



Şekil 3.6: Ağ Yapısının Zaman İçinde Gösterdiği Değişimden Geleceğe Yönelik Tahmin Yapmak [38]

Sosyal ağlar üzerinden ilişki tahmini yapabilmenin literatürde öne çıkan iki yöntemi bulunmaktadır. Bunlar denetimli yöntemler ve denetimsiz yöntemlerdir. Denetimli yöntemlerde esas olarak ağın yapısal özelliklerinin ayıklanıp eşleme fonksiyonu yardımıyla analiz yapılmaktadır. Böylece ağa ait elde edilmiş olan verilerden hareketle bilinmeyenlere ulaşılmaya çalışılmaktadır. Belli bir düğüm çifti baz alınarak bunlar arasındaki bağlantının varlığı, ikili bir yöntemle tahmin edilmeye çalışılmaktadır. Denetimsiz yöntemlerde ise temel olarak benzerlik yöntemlerinden yararlanılmaktadır. Buna ek olarak bağlantıların mevcudiyeti araştırılırken karar ağaçları, en yakın komşu, destek vektör makineleri gibi algoritmalar da faydalanılmaktadır. Literatürdeki ağırlıklı görüşe göre bağlantı tahminlerinde en etkili yöntem, daha az sistem kullanımına gerek duyulan yöntemlerdir. Dolayısıyla daha doğru sonuçlar verdiği kabul edilen benzerlik algoritmalarının kullanımı literatürdeki analizlerde öne çıkmaktadır. Bu sebeple bu çalışmada benzerlik tabanlı algoritmalara değinilecektir. Ancak temel olarak bağlantı tahmini yaparken işleyen süreç şöyledir:

Seçilen iki düğüm çifti arasındaki bağlantı durumuna göre bir skor atanmaktadır ($skor_{xy}$). Atanan bu skorun temel amacı x ve y olarak adlandırılan düğümler arasındaki benzerliği ölçmektir. Bu işlem ağda mevcut olan tüm düğüm çiftleri için gerçekleştirilir. Böylece ağdaki tüm düğüm çiftlerine ait benzerlik skorları sıralanmaktadır. Düğüm çiftleri arasındaki skorlar arttıkça benzerliğin ve dolayısıyla bahse konu düğümler arasındaki bağlantı ihtimalinin artacağı ifade edilmektedir. Bağlantı tahmini yöntemlerinin kullanıldığı çalışmalar literatürde genel olarak oldukça zorlu çalışmalar olarak nitelendirilmektedir. Bu konuda karşılaşılan en

büyük sıkıntı düğümler arasındaki benzerlik skorlarının atanmasıdır. Kimi zaman analiz yapılacak ağın niteliğine göre kullanılacak benzerlik indeksleri oldukça karmaşık olabilmektedir. Dolayısıyla benzerlik indekslerinin karmaşıklığı, analiz sonuçlarından beklenen verimin alınmasına engel teşkil edebilecektir.

Ağda yer alan düğümlerin çeşitli faktörler üzerinden benzerlik kıyaslamaları yapılabilmektedir. Böylece düğümlerin benzerliklerinin ölçümünün kolaylaştırılması mümkün ve istenilen bir durumdur. Bazı durumlarda düğümlerin özelliklerinin gizli olması söz konusudur. Böyle zamanlarda “yapısal benzerlik” adı verilen benzerlik indekslerinin kullanımı daha uygun olmaktadır. Literatürde yaygın bir kullanıma sahip olan bu indeksler çeşitli sınıflara ayrılmıştır [34]. Bu benzerlik indeksleri ve uygulamaları sonraki bölümlerde açıklanmaktadır.

3.6.1 Yerel benzerlik indeksleri

Ağ analizlerinde kullanılan “ortak komşu” kavramı, yerel benzerlik indekslerindedir. Ağ üzerinde seçilen x ve y noktaları olduğunu varsayarsak, $\Gamma(x)$ ile gösterilen küme x ile komşu olan düğümlerin kümesi olarak, $\Gamma(y)$ ise y ile komşu olan düğümlerin kümesi olarak nitelendirilsin. Böyle bir durumda hem $\Gamma(x)$ hem de $\Gamma(y)$ kümesinin elemanı olan düğümler bulunması durumunda, başka bir deyişle x ve y düğümlerinin her ikisinin de komşusu olan düğümlerin olması durumunda, bahse konu iki düğüm arasında yüksek ihtimalle bağlantı kurulacağı söylenebilir. Bu durumun matematiksel gösterimi aşağıda sunulmuştur:

$\Gamma(x) = x$ Düğüm noktasının komşularının kümesi,

$\Gamma(y) = y$ Düğüm noktasının komşularının kümesi olmak üzere,

$|\Gamma(x) \cap \Gamma(y)| \rightarrow$ Boş kümeden farklı bir küme ise x ve y düğüm noktaları arasında bağlantı kurulması muhtemeldir.

Ortak komşu kavramı kullanılarak, baz alınan düğüme komşu olan düğümlerin sayısı hesaplanıp, bu düğümler arasındaki potansiyel ortak özellikler belirlenebilmektedir. Bu yöntem kullanılarak pek çok ağ analizi çözümlemesinin yapılması mümkündür. Ortak komşuluğun kullanıldığı farklı normalizasyon metotları da bulunmaktadır.

- **Salton İndeksleme;** k_x x düğümünün derecesini, k_y ise y düğümünün derecesini göstermek üzere aşağıda sunulan formül sonucunda bulunan elemanlar ortak komşu elemanlarıdır. Bu gösterime literatürde kosinüs benzerliği adı verilmektedir. (3.2)

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{k_x + k_y}} \quad (3.2)$$

- **Jaccard İndeksleme;** Ortak komşular kullanılarak başka bir benzerlik bulma metodudur. Aşağıda formüle edilmiş hali gösterilen durumda ortak komşular, toplam komşulara bölünerek elde edilmektedir. (3.3)

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (3.3)$$

- **Sorensen İndeksleme;** Bu yöntem ekolojik topluluk verileri için kullanılmaktadır. (3.4)

$$S_{xy} = \frac{2|\Gamma(x) \cap \Gamma(y)|}{k_x + k_y} \quad (3.4)$$

- **Hub Promoted İndeksleme;** Bu metot daha çok metabolik ağlarda yüzey çiftlerinin örtüşme oranlarını hesaplamak amacıyla kullanılmaktadır. Paydanın minimum olması sebebiyle merkeze daha yakın olan düğümlerin bağlantı olasılıkları yüksek olmaktadır. (3.5)

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\min\{k_x, k_y\}} \quad (3.5)$$

- **Hub Depressed İndeksleme;** Hub Promoted İndekslemenin tersi mantıkla çalışmaktadır. Ağın merkezinden uzaklaştıkça düğümlerin bağlantı olasılıkları azalmaktadır. (3.6)

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\max\{k_x, k_y\}} \quad (3.6)$$

- **Leicht-Holme-Newman İndeksleme;** Bu metoda göre ortak komşu sayısı çok fazla olan düğüm çiftlerinin benzerlik oranı yüksek çıkmaktadır. Ancak muhtemel bağlantı oranı komşu sayısından daha az olacaktır.(3.7)

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{k_x \times k_y} \quad (3.7)$$

- **Tercihli Bağlantı İndeksleme;** Analizi yapılacak olan ağın ölçeksiz büyümesi durumunda kullanılmaktadır. Herhangi bir düğümle bağlantı kurulma ihtimali “ k_x ” ile orantılı ise, başka bir düğümle bağlantı kurulma ihtimali ise “ k_y ” ise bahse konu bu iki düğümün kendi aralarında yeni bir ilişki kurma olasılıkları da “ $k_x \times k_y$ ” ile doğru orantılı olmaktadır.

Başka bir deyişle “ S_{xy} ” boş kümeden farklı bir küme olmak üzere;(3.8)

$$S_{xy} = k_x \times k_y \text{’dir.}$$

(3.8)

- **Adamic-Adar İndeksleme;** Bu yöntem ise ağ yapısındaki bağlantıları daha az olan düğümlerin ortak komşularını belirlemektedir.
- **Kaynak Tahsisli İndeksleme;** Karmaşık ağ yapılarında tercih edilir. Aralarında direkt bağlantı olmayan düğüm çiftlerinin ilişki durumlarını ölçmeye yaramaktadır. Birbiri ile doğrudan bağlantısı olmayan düğümlerin ilişkisi ortak komşu düğümleri üzerinden sağlanmaktadır.

Yukarıda kısaca açıklanan yerel benzerlik indeksleri üzerine literatürde çok sayıda çalışma bulunmaktadır. Bu yöntemlerin hangi koşullar altında daha etkin sonuçlar vereceğine dair çalışmalar yapılmış olup, analizi yapılacak olan ağın özellikleri dikkate alınarak bahse konu indekslerin kullanılması gerektiği ifade edilmektedir.

3.6.2. Genel benzerlik indeksleri

Yerel benzerlik indekslerinde olduğu gibi çeşitli yöntemler kullanılmaktadır. Katz İndeksleme olarak ifade edilen yöntemde, bağlantı yollarının gruplandırılması yöntemi tercih edilmektedir. Dolayısıyla düğümler arası kısa mesafelerin benzerlik oranları daha yüksek olmaktadır. Bu indekslerin matematiksel gösterimlerinden ziyade çalışma mantıklarının verilmesinin daha uygun olduğu değerlendirilmektedir.

Leicht-Holme-Newman indeksleme (LHN2) olarak literatüre girmiş olan genel benzerlik indeksi ise, Katz indekslemesine benzer bir mantıkla hareket edilmektedir. Buradaki temel mantık, eğer benzerlikleri ölçülecek iki düğümün komşuları benzerse kendilerinin de benzer olma ihtimalleri yükselmektedir.

Benzerlik indekslerinde bir diğer yöntem, düğümler arası geliş gidişte harcanan süre baz alınarak düğümlerin benzerliklerinin tespitinin yapılmasıdır. Bu yöntem literatürde “ortalama işlem süresi” adı verilmektedir.

Düğümler arası mesafeden benzerlik tespitinin bir başka çeşidi de “kosinüs benzerliği” olarak adlandırılmaktadır. Analiz edilecek iki düğüm arası kosinüs uzaklığı dikkate alınmaktadır. Bu değer 0-1 aralığında değerler almaktadır. Aldığı değerler 0’dan 1’e yaklaştıkça benzerlik oranının arttığı ifade edilmektedir.

Yeniden başlamalı rastgele yürüyüş (RWR) yöntemine göre ise herhangi bir düğümden yola çıkılıp rastgele hareket edilerek komşu düğümlerde ya da tekrar aynı düğümden sonlanması ihtimalleri üzerinden benzerlik tespiti yapılmaktadır.

Simrank adı verilen yöntemi, yeniden başlamalı rastgele yürüyüş yöntemi ile benzerlikler göstermektedir. Buradaki temel anlayış ise, farklı düğümler eğer benzer düğümlere bağlanıyorsa bu düğümlerin de benzerlikleri yüksektir.

Genel benzerlik indekslerinde son olarak Matrix Forest İndeksleme yer almaktadır. Buradaki mantık ise aynı düğümden kaynaklanan yeni düğümlerin benzer olacağıdır. Bu durum indeksin isminden de anlaşılacağı üzere aynı kökten gelen iki ağacın birbirine benzetilmesi şeklinde ifade edilmektedir.

Bunlar haricinde literatürde yarı yerel benzerlik indeksleri olarak yer alan yöntemler de bulunmaktadır. Bunlar:

1. Yerel Yol İndeksleme
2. Yerel Rastgele Yürüyüş (LRW)
3. Çakıştırılmış Rastgele Yürüyüş (SRW)

4. SOSYAL AĞLARDA EĞİTİM DÜZEYİNİN VERİ MADENCİLİĞİ İLE ANALİZİ

4.1. Veri Seti

Sosyal ağların analizinde temel olarak mevcut bilgilerden çıkarımlar yapılması yer almaktadır. Bu çıkarımlarda sıklıkla kullanılan unsur ise kelime ve cümlelerdir [4]. Sosyal ağların kullanımının her geçen gün yaygınlaşmasıyla birlikte sosyal ağlara ait veri tabanlarında mevcut olan bilgi miktarında büyük artışlar görülmeye başlamıştır. Bu durum, böyle büyük miktarlardaki veri arasından çıkarınlar yapılabilmesine olanak sağlayan veri madenciliği ve fikir madenciliği yöntemlerinin ön plana çıkmasını sağlamıştır. Literatürde bu yöntemler kullanılarak yapılan çok sayıda çalışma bulunmaktadır. Bahse konu çalışmalarda sosyal ağ kullanıcılarına ait bilgilerden hareketle duygusal, siyasal vb. çok çeşitli alanlarda analiz yapılabilmesi mümkündür.

Bu çalışmada sosyal ağlar içerisinde kelime ve cümle kullanımıyla öne çıkan popüler uygulama “twitter” ele alınmaktadır. Twitter, belirli konu başlıkları altında kullanıcıların görüşlerini belirttikleri gerçek zamanlı bir sosyal bilgi ağı olarak ifade edilmektedir. Kullanıcılar bu uygulama sayesinde, istedikleri konularda bir sohbet ortamı rahatlığında görüş bildirebilme imkanına sahiptir. Yazılabilecek olan her bir tweet 140 karakterle sınırlandırılmış olup böylece kullanıcıların kısa ve öz olarak görüş bildirmeleri sağlanmaktadır. Twitter sosyal ağının kullanıcıları sadece gerçek kişilerden değil kurumsal şirketlerin hesaplarından da oluşmaktadır. Firmalar bu sayede müşterileri ile olan iletişimlerini artırabilmekte, bilgilendirme amaçlı paylaşımlarını yapabilmektedirler.

Twitter’da yer alan veriler herkes tarafından ulaşılabilenler ve gizliler olmak üzere gruplandırılmaktadır. Twitter verilerin herkese açık olan kısımlarını API yardımıyla kamuoyuyla paylaşmaktadır [35]. API kullanımıyla istenilen filtre kelimeler girilerek mesajlar arasında arama yapılabilmesi imkanı bulunmaktadır.



Şekil 4.1: Twitter'a Ait Veri Tabanından Tweetlere Erişim Görseli

Twitter'da twitter kullanıcıları tarafından girilen tweetlerin yer aldığı MYSQL veri tabanı bulunmaktadır. Yazılan tweetlere ait link vb. diğer içerikler de burada kayıtlıdır. Twitter API yardımıyla belirlenen tarih aralıkları için twitter kullanıcıları tarafından girilen mesajlara ulaşılabilmesi mümkündür. Çalışmanın bu bölümünde twitter veri tabanındaki Türkçe tweetler veri olarak kullanılmıştır. Twitter'daki mesaj ve paylaşım yoğunluğu göz önüne alınarak uzun dönemli bir çalışma yerine yıllık kapsamda bir analiz gerçekleştirilmiştir. Bu sebeple 2016 yılı twitter paylaşımları esas alınmıştır.

Bu çalışmada araştırılan olgu, literatürde sıklıkla kullanılan bir yöntem olan kelime ve cümle temelli veri madenciliğinin kullanılmasıyla sosyal ağ kullanıcılarının eğitim düzeylerinin tahmin edilebilmesini sağlamaktır. Çalışma kapsamında ele "Twitter" sosyal ağının baz alınması ise verilerine olan ulaşım kolaylığı olmuştur. Bu kapsamda sonraki bölümlerde eğitim düzeyi tespiti için belirlenen kriterlere ve analizin yapılışına değinilmiştir.

4.1.1 Eğitim düzeyi tespiti için filtre kelimelerin belirlenmesi

Öncelikle Türkçe tweet atan twitter kullanıcılarının attıkları tweetler ve diğer tüm paylaşımları dikkate alınarak bu kişilerin eğitim düzeylerinin tahmini için analizde kullanılacak bir kelime seti belirlenmiştir. Twitter sosyal platformunun genel olarak günlük konuşma dilinde paylaşımların yapıldığı bir ağ olduğu göz önüne alınarak, filtre kelimeler arasında akademik kelimelerin ağırlığı çalışma sonuçlarının sağlıklı olması açısından sınırlı tutulmuştur. Filtre kelimeler belirlenirken özellikle yüksek öğrenim ve üstü eğitim seviyesindeki bireylerin ilgi alanına giren sınavlarla (YDS,

TOEFL, KPSS, ALES, TUS, DUS, Kurum Sınavları vb.) ilgili paylaşımlar esas alınmıştır. Bu durumun sınav esaslı bir eğitim sistemine sahip olan ülkemiz özelinde kullanışlı bir metot olduğu değerlendirilmektedir. Bunun yanı sıra yine konuşma dili kullanımı göz önüne alınarak bazı kelimelerin çekimli halleri de veri tabanına eklenmiştir. Analiz kapsamında kullanılan filtre kelimelerin listesi Ekler bölümünde sunulmuştur.

4.2 Twitter Kullanıcılarının Eğitim Düzeylerinin Analiz

Twitter kullanıcılarının eğitim düzeylerinin ne olduğu konusunda, kullanıcıların yaptıkları paylaşımlardan çıkarım yapılabilmesi amacıyla gerçekleştirilen bu çalışmada daha önce ifade edilen twitter tarafından paylaşılan veri tabanı kullanılmıştır. Twitter kullanıcılarının eğitim düzeylerinin tespit edilebilmesi kullanıcıların daha iyi tanınabilmesi bakımından oldukça önemlidir. Çünkü kullanıcı profili bilinen kitlelere sunulabilecek pazarlama ürünlerinin belirlenebilmesi de kolaylaşmaktadır.

Literatürdeki bu ve benzeri türde çalışmalar incelendiğinde, kullanıcıların istenilen özelliklerinin belirlenebilmesi için çeşitli “filtre” kelimeler kullanıldığı görülmektedir. Örneğin sosyal ağ kullanıcılarının mutluluk düzeylerinin tahmin edilmesine yönelik olarak yapılan çalışmada kişiler (1,0) mantığı baz alınarak mutlu ya da mutsuz olarak nitelendirilmektedir. Bu yapılırken de kişileri “mutlu” olarak nitelendirilebilecek filtre kelimeler belirlenerek, paylaşımlarında bu kelimeleri kullanan kişiler mutlu olarak kabul edilmiştir. Bahse konu çalışmalarda öne çıkan bir diğer husus ise, analiz için ele alınan özelliğin var olması ya da var olmaması şeklinde ikili bir değerlendirme yapılmış olmasıdır. Kelime ve cümlelerden hareketle yapılan veri madenciliği kapsamında yapılan çalışmalarda bu sonucun normal bir durum olduğu değerlendirilmektedir. Nitekim kişilerin mutluluk seviyelerinin (az mutlu, çok mutlu vb.) belirlenebilmesi bir çalışma için çok daha fazla girdiye ihtiyaç duyulacaktır. Bu sebeple literatürde de hâkim olan araştırma şekli ele alınan durumun var olması ile olmamasının tespitine yönelik olmuştur.

Yukarıda ifade edilen gerekçelerle bu çalışmada da twitter kullanıcılarının eğitim düzeylerine yönelik tahmin işlem, gerçekleştirilirken belirli bir eğitim seviyesi belirlenerek kullanıcıların bu seviyenin üstünde ya da altında bir eğitimde

olduklarının tahmini yapılmaya çalışılmıştır. Çalışma kapsamında belirlenen seviye “yüksek öğrenim” eğitim seviyesi olarak alınmıştır. Dolayısıyla bu araştırmada, Türkiye’de yer alan twitter kullanıcılarının belirli süreler boyunca yapmış oldukları paylaşımlar esas alınarak bu kullanıcının eğitim seviyesinin yüksek öğrenim (üniversite) ve üzerinde mi altında mı olduğu konusunda çıkarımlar yapılmaya çalışılmıştır. Bu çıkarımlar yapılırken analiz için daha önceden belirlenmiş olan filtre kelimelerin kullanıcıların paylaşımlarında geçip geçmemesi baz alınmıştır. Belirlenen kelimelerden herhangi birinin kullanıcı tarafından yapılan paylaşımlarda yer alması durumunda bahse konu kullanıcının yüksek öğrenim ve üzeri eğitim düzeyine sahip olduğu (1) varsayılacaktır. Bu kelimelerle ilgili bir paylaşımı bulunmayan twitter kullanıcıları ise yüksek öğrenim altında eğitim seviyesine sahip olduğu (0) varsayılacaktır.

Eğitim seviyesi analizi kapsamında, atılan her bir tweette belirlenmiş olan kelimelerin yer alıp almaması durumuna göre analiz yapılabilmektedir. Eğitim seviyesi için baz alınan “yüksek öğrenim düzeyinde” eğitime sahip olup olmama durumlarına göre kullanıcı değerlendirilmesi yapılmıştır. Bu durumların gösterimi için vektör modeli kullanımı tercih edilmiştir [34]. Bu modele göre her bir kullanıcı için ikili (binary) matris oluşturularak matris elemanlarını belirlenen nitelikler belirlemektedir. Bu yöntem aslında yoğunlukla benzerlik çalışmalarında kullanılmaktadır. Fakat çalışmamızda da olduğu şekilde benzerlik tipi araştırmalarda da kullanımı söz konusudur.

Analiz için her bir tweet vektörel olarak ifade edilmektedir. Bu duruma ait vektörel gösterim aşağıda sunulmuştur:

$V = \{\text{Yüksek öğrenim ve üzeri}, \text{Yüksek öğrenim altı}\}$ şeklinde ikili bir şekilde gösterilmektedir.

Seçilen bir tweet’in (1,0) şeklinde ifade edilmesi bahse konu kullanıcının en az yüksek öğrenim düzeyinde, (0,1) olması yüksek öğrenim altında bir eğitim seviyesinde olmasına karşılık gelmektedir.

Örneğin,

“Pazar’ın güzelliği’ diyemiyorum. Çünkü ders çalışmak zorunda olduğum bir gün #ales”

Yukarıdaki örnek tweet incelendiğinde filtre kelimelerden “ales” in kullanılmış olduğunu görülmektedir. Bu kapsamda bu tweeti atan kullanıcı bu çalışma kapsamında $V = (1,0)$ olarak yani yüksek öğrenim ve üzeri düzeyde eğitime sahip olan gruba dahil edilmektedir. Veriler bu şekilde kodlandıktan sonra buradan çeşitli çıkarımlar yapılması mümkündür. Örneğin, bu ve benzeri şekilde filtre kelimeler kullanılarak atılan tweetler, belirli dönem boyunca atılan tüm tweetlerin sayısına oranlanarak hedef kitle kullanıcıların sosyal ağı kullanım yoğunlukları hakkında fikir edinebilmek mümkündür. Buradaki hedef kitle yüksek öğrenimli kullanıcılar olduğundan, yüksek öğrenime sahip kişilerin sosyal medyayı kullanım sıklıkları hakkında genel de olsa bir değerlendirmeye gidilmesi mümkündür. Bu duruma ait matematiksel gösterim aşağıda sunulmuştur.(4.1)

$$\frac{\sum_{i=0}^k \text{Filtre Kelimelerin kullanıldığı tweet sayısı}}{\sum_{i=0}^n \text{Toplam tweet sayısı}}$$

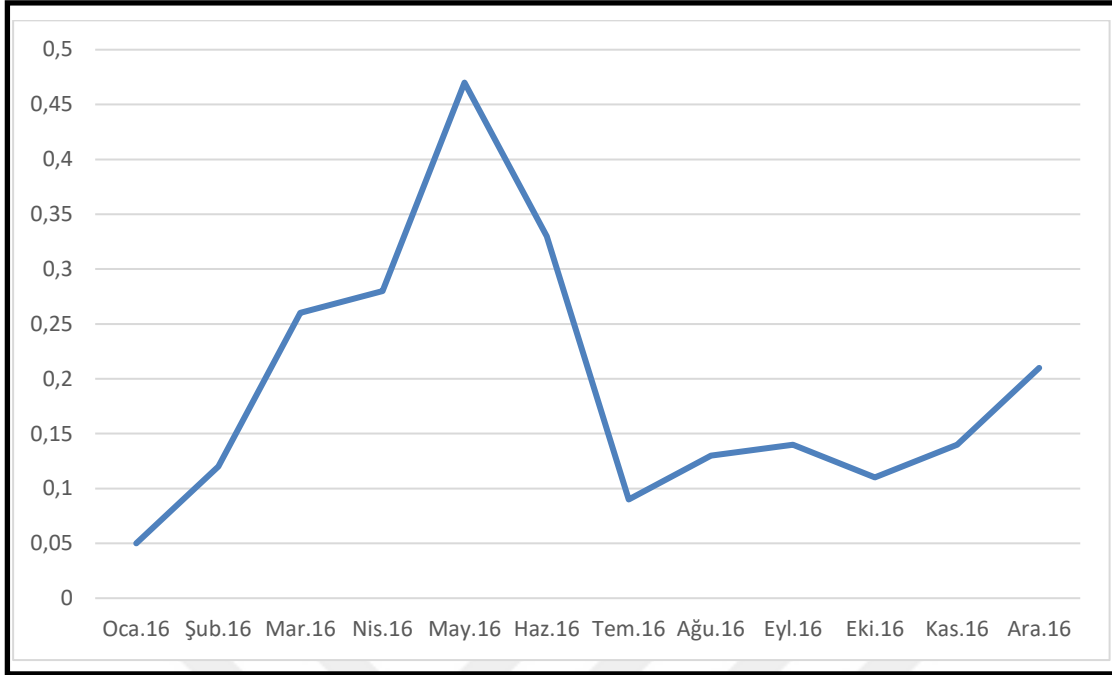
(4.1)

Yukarıdaki matematiksel gösterimde;

k: Filtre Kelimelerin kullanıldığı tweet sayısını,

n: Toplam tweet sayısını ifade etmektedir.

Bu kapsamda ele aldığımız çalışma dönemi olan 2016 yılı göz önüne alındığında yıl boyunca yüksek öğrenime ilişkin kelimelerin kullanıldığı/paylaşıldığı tweetlerin tüm tweetlere oranları aşağıdaki şekilde sunulmuştur.

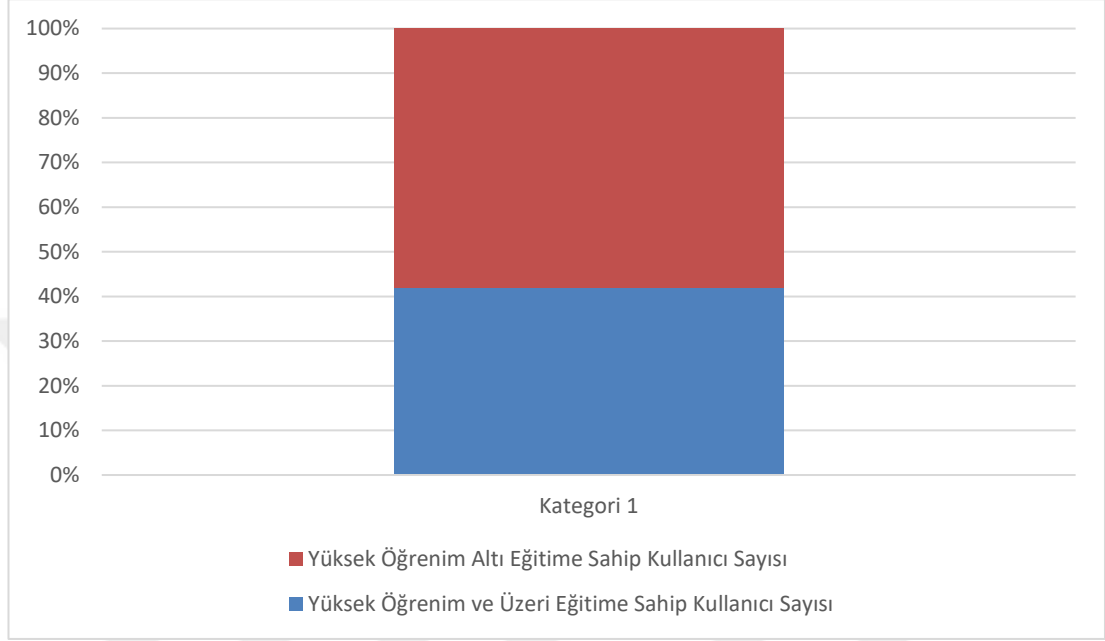


Şekil 4.2: Yüksek Öğrenime İlişkin Tweetlerin Tüm Tweetler İçerisindeki Aylık Yüzesel Dağılımı (2016 yılı) [38]

Şekil 4.2 incelendiğinde 2016 yılında atılan tweetlerden yüksek öğrenime ilişkin kelimeleri içeren tweetlerin tüm tweetlere oranlarının aylara göre değişimi görülmektedir. Yüksek öğrenime ilişkin tweetlere genel olarak bakıldığında genel itibariyle düşük yüzdelerde ilerlediğini söylemek mümkündür. En yüksek oranlara ulaştığı Mayıs, Haziran döneminde, genel aylık ortalamanın yarısını oluşturacak seviyelere yaklaşabildiği görülmektedir. Genel itibariyle tüm bir yıla bakıldığında tweetlerin yaklaşık yüzde 10'u civarının hedef konu üzerinde yoğunlaştığı söylenebilmektedir.

Şekildeki dağılımların 2016 yılı boyunca yapılan sınavlarla paralel bir biçimde şekillendiğini söylemek mümkündür. Nitekim 2016 yılı mayıs ayında gerçekleşen ALES ve KPSS sınavları üniversite mezunu adaylarca girilen sınavlar olarak twitter gündeminde yoğunlaştığı görülmektedir. Benzer şekilde Nisan ve Eylül aylarında yapılan Tıpta Uzmanlık Sınavları (TUS), Diş Hekimliği Uzmanlık Sınavları (DUS) da ALES, KPSS seviyesinde olmasa da nispeten etkili olmuşlardır. Yıl sonuna doğru yeniden belirlenen içerikte tweetlerin sayısının artmaya başlamasını ise, özellikle yıl sonunda Sayıştay, Adalet Bakanlığı vb. kurum sınavlarının yoğunlaşmasıyla açıklamak mümkündür. Genel itibariyle sınav tarihlerine göre şekillenen bir konu

trafiği olduğu değerlendirmesinin isabetli olacağı değerlendirilmektedir. Ancak bu değerlendirmeler, Twitter’da yüksek öğrenim içerikli tweetlerin yoğunluğunu göstermektedir. İçerik analizi neticesinde bu ve benzeri paylaşımları yapan kişilerin Twitter’ın toplam kullanıcı sayısının ne kadarını oluşturduğu konusu önem kazanmaktadır. Bu konudaki istatistik ise aşağıdaki Şekil 4.3’de sunulmuştur



Şekil 4.3: Yüksek Öğrenimli Olarak Tahmin Edilen Kullanıcıların Twitter’daki Toplam Kullanıcı Sayısı ile Kıyaslanması (2016 yılı)

Bu çalışma kapsamında yapılan analize göre Twitter’ın toplam kullanıcı sayısının yaklaşık olarak %42si yüksek öğrenime sahip olarak bulunmuştur. Şekil 4.2’deki içerik analizine göre daha yüksek bir oran elde edildiği dikkat çekmektedir. Ancak bu sonuçlar gerçekteki verilerle örtüşmemektedir. Günümüzde başta sosyal medya platformları olmak üzere, dijital dünya ile ilgili pek çok alanda düzenli olarak istatistik açıklayan uluslararası güvenilir kuruluşlar bulunmaktadır. Bu kuruluşların açıkladıkları istatistiklerin de birbirleriyle yaklaşık değerlerde olması sebebiyle literatürde göz önüne alınmakta olan itibarlı değerlendirmelerdir. Bu kuruluşlardan “wearesocial” tarafından açıklanan rapor (Digital in 2016 Global Overview) göz önüne alındığında kelime odaklı veri madenciliği yöntemi ile yapılan bu çalışmanın sonuçlarının açıklanan verilerle uyuşmadığı görülmektedir.

Digital in 2016 Global Overview raporunda, 2016 yılı itibariyle Türkiye’de aktif şekilde Twitter kullanan kişi sayısı yaklaşık 15 milyon olarak açıklanmıştır. Bunların

ise %63'ü yüksek öğrenim ve üzeri eğitim seviyesinde kişilerden oluşmaktadır. Diğer sosyal ağlara oranla Twitter'da yüksek öğrenimli kişi oranı oldukça yüksektir. Ancak bu çalışma kapsamında elde edilen oran gerçekteki sonuçların altında kalmıştır. Bu durumun; analiz yöntemi, verilerin ele alınışı vb. pek çok faktörden kaynaklanma ihtimali bulunmaktadır. Nitekim kelime ve cümle odaklı olan bu veri madenciliği yöntemi kullanışlı ve basit bir yöntem olmasına rağmen analize ilişkin birtakım dezavantajlar barındırmaktadır. Kelimelerin tweetler içerisinde geçip geçmemesine göre yapılan basit frekans analizi kimi durumlarda yetersiz kalmaktadır. Çünkü bir tweetin içerdiği kelimelerden yola çıkılarak analiz yapılması her ne kadar basit bir yöntem olsa da analiz sonuçlarının daha sağlıklı olabilmesi için daha detaylı değerlendirilmelere ihtiyaç duyulabilmektedir. Ancak elde edilen sonuçlara da bakıldığında bu yöntemle gerçekleştirilen analizlerde de genel eğilimin tespit edilebilmesinin mümkün olduğu görülmektedir. Analize dair genel değerlendirmeler, eksik yönleri ve yapılması gerekenler kapsamında öneriler çalışmanın sonuç ve değerlendirmeler bölümünde yer almaktadır.

4.2.1 Uygulama kapsamında kullanılan programlar

Kullanıcılar tarafından yapılan Twitter paylaşımlarının incelenmesi maksadıyla şu programlar kullanılmıştır:

- Microsoft Visual Studio,
- STATA 12,
- Mysql Veri Tabanı.

Mysql veri tabanından elde edilen veriler form uygulaması ile işlenerek analize hazır hale getirilmiştir. Analiz için bu çalışmada STATA 12 paket programı kullanılmıştır.

• Form Uygulaması İle Verilerin Analize Uygun Hale Getirilmesi

Twitter' dan elde edilen verilerin işlenmesi ve gerekli dönüşümlerin yapılabilmesi maksadıyla Microsoft Visual Studio geliştirme ortamı, verilerin dönüşümü ve analiz işlemleri için csharp dili kullanılmıştır. Form ekranında analiz öncesi yapılan işlemler şunlardır:

1. Veri Dönüşümü,
2. Tarihsel Analizler,
3. Seçilen Kelimelerin Frekanslarının hesaplanması

Twitter'a ait verilerin (paylaşımların) düz yazı şeklinde olması sebebiyle veri madenciliğine ilişkin işlemler için önce verilerin dönüşümü işlemine ihtiyaç duyulmuştur. Bu dönüşümden kasıt çalışmanın önceki bölümlerinde değinilen ve belirlenen anahtar kelimeler baz alınarak matris oluşturulması işlemidir. Tarihsel analiz yapılması işleminde ise, analiz için ihtiyaç duyulan sql kodlarının yaratılması gerçekleştirilmiştir. Bu bölümde twitter paylaşımlarının tarihsel olarak değerlendirilmesi yapılmıştır. Son olarak, belirlenen filtre kelimeler dikkate alınarak bu kelimelerin kullanılma frekanslarının hesaplanmasına geçilmiştir.

Analize uygun hale getirilen veriler daha sonra STATA paket programına aktarılmıştır. Twitter paylaşımları mysql veri tabanına kaydedilmiştir. Veri tabanında tweetler ile beraber kullanıcılara ait aşağıdaki bilgiler kaydedilmiştir:

- Tweet text
- Tarih
- Kullanıcı ismi
- Takipçi sayısı
- Arkadaş sayısı
- Kaynak
- Profil resmi linki
- Paylaşılan linkler
- Twitter kullanımına başlama zamanı
- En son güncelleme zamanı

Analiz işlemleri için tweet text alanının ve tarih kısmının kullanılarak verinin dönüştürülmesi ile analiz için aşağıdaki alanlar kullanılmıştır:

- Filtre Kelimeler
- Saat
- Gün
- Ay
- Yıl

Analiz işlemleri için saat, gün, ay ve yıl kısımlarının kullanılmasıyla sql sorgusu zaman esas alınarak yapılabilmektedir. Gerçekleştirilen veri dönüşümü işlemiyle birlikte, kullanıcılar tarafından yapılan paylaşımların içeriklerinde filtre kelimelerin her defasında aranması işlemine gerek kalmamaktadır. Bu durum işlemleri kolaylaştıran önemli bir unsurdur.



5. SONUÇ VE DEĞERLENDİRMELER

Literatürde sosyal ağlar üzerine son yıllarda yapılan çalışmalara bakıldığında kişiler tarafından yapılan paylaşımlar, kişiler arası ilişkiler vb. verilerin kullanılarak yeni bilgiler çıkarıldığı görülmektedir. Başka bir deyişle insanların sosyal platformları kullanırken yapmış oldukları her türlü işlemde kullanıcıya ilişkin başkaca bilgilere ulaşabilmek mümkündür. Hatta bu konu üzerine yoğunlaşan çalışmalar neticesinde sosyal veri tabanlarından çekilen ham verilerden hareketle çeşitli bilgilere ulaşabilen yazılımlar da geliştirilmeye başlanmıştır.

Sosyal medya kullanıcılarının tanınabilmesi için çıkarımlar yapılması konusu pek çok açıdan önem taşımaktadır. Güvenlik gerekçesi bunların önde gelenlerindedir. Nitekim suça eğilimli kişilerin bu sayede önceden tespit edilerek kontrol altına alınabilmesi sağlanmaktadır. Diğer gerekçelerin büyük çoğunluğu ise ticaridir. Özellikle pazarlama alanındaki faaliyetler için hedef ve mevcut müşteri kitlesinin ve bunların taleplerinin belirlenebilmesi büyük önem arz etmektedir. Bu bilgilere göre kitleleri daha iyi tanıma olanağı bulan firmalar pazarlama stratejilerini daha doğru oluşturabilmektedir.

Daha önce ifade edildiği üzere twitter ve benzeri düz yazı şeklinde paylaşımların yapıldığı sosyal platformlar için uygun analiz tekniklerinin başında fikir madenciliği gelmektedir. Bu yöntemle yapılan analizler sırasında karşılaşılan en büyük sorun, dillerin yapısı ve çeşitliliğinden kaynaklanmaktadır. Bu konudaki zorlukların aşılabilmesini konu alan farklı diller arasında ortak çalışma yapılabilecek yeni yöntemlerin geliştirilmesi için çeşitli araştırmalar yapılmaktadır. Ancak yine de yapılan çalışmalar sonucunda elde edilen bulgular değerlendirildiğinde fikir madenciliği ile yapılan çalışmalarla başarılı sonuçlar elde edildiği söylenebilmektedir. Ancak özellikle Türkçe literatürde fikir madenciliği yöntemi kullanılarak yapılan çalışma sayısının oldukça az olduğu belirlenmiştir. Bu çalışmanın bu özelliği sebebiyle Türkçe literatüre bir katkı olabilmesi amaçlanmaktadır.

Twitter kullanıcılarının eğitim düzeylerinin tahmin edilebilmesi amacıyla yapılan bu çalışmada fikir madenciliği analizi için tweetlerin incelenmesinde esas alınacak filtre kelimeler belirlenerek, frekans yöntemi kullanılmıştır. Ancak literatürdeki diğer benzer çalışmalarda da olduğu gibi bu yöntemin kullanıldığı çalışmalarda belirli bir olgunun var olup olmaması sorgulanabilmektedir. Bu sebeple kullanıcıların eğitim düzeyleri için belirlenen bir sınırın (yüksek öğrenim) üzeri ve altında eğitim seviyesinde olmaları değerlendirilmiştir.

Analiz sonucunda elde edilen bulgular değerlendirildiğinde bu yöntemin twitter kullanıcılarının eğitim düzeylerinin tahmin edilebilmesinde genel itibariyle kullanışlı bir yöntem olduğunu söylemek mümkündür. Ancak, çalışma sonucunda elde edilen bulgular değerlendirildiğinde twitter kullanıcıları arasında yüksek öğrenimli kişilerin oranının sapmalı olarak bulunduğu görülmüştür. Bu duruma neden olan pek çok etmen olması mümkündür. Analizin temelinde kelimeler olması sebebiyle analize esas filtre kelimelerin belirlenmesi ve değerlendirmesi diğer analiz yöntemlerine göre daha zordur. Bu kapsamda sonuçlarda ortaya çıkan sapma oranına belirlenen filtre kelimelerin yetersiz oluşu, analiz için ele alınan dönem vb. faktörler de bu sonucu doğurabilecektir. Fakat bu çalışma, özellikle Türkçe literatüre sağlayacağı katkı bakımından anlamlı sonuçlara sahiptir. Analiz sonuçlarının doğruluğunun artırılması için bu yöntemin literatürde yer alan diğer yöntemlerle (kişiler, gruplar arası sınıflandırmalar, bağlantı tahmini yöntemleri vb.) birlikte kullanılmasının gelecekte yapılacak çalışmalarda için yararlı olacağı değerlendirilmektedir.

KAYNAKÇA

- [1]. **Köksal, A.** (1981). Bilişim Terimleri Sözlüğü. Türk Dil Kurumu Yayınları.
- [2] **Gürsakal N.**(2009) Sosyal Ağ Analizi
- [3] **Berendt, B. , Hotho, A., Stumme, G.,** (2010) “Bridging the gap - Data mining and Social network analysis for Integrating Semantic Web and Web 2.0”, *Journal of Web Semantics*, 8(2-3): 95-96.
- [4]. **W. Koehler,** (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.
- [5] **Kuduğ, H.,** (2011). “Sosyal ağ analizi ölçütlerinin iş alanlarına uyarlanması, Yüksek lisans tezi, *Ege Üniversitesi Fen Bilimleri Enstitüsü.*
- [6] **Mazman, S. G.,** (2009) “Sosyal ağların benimsenme süreci ve eğitsel bağlamda kullanımı”, Yüksek lisans tezi, *Hacettepe Üniversitesi Fen Bilimleri Enstitüsü,* Ankara.
- [7] **Folino, F. ve Pizzuti, C.,** (2012). “Link Prediction Approaches for Disease Networks”, ITBAM 2012, 99-108.
- [8] **Alpaydın, E.,** (2000). “Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri”, Bilişim Semineri 2000, 1-5.
- [9] **O. Boy, A. Karcı,** (2011)“Sosyal Ağların Web Madenciliği Teknikleri ,le Analizi ve Ortak Atıf Analizi ile Benzerlik Tahmini”, 154-161, FEEB-11, Fırat Üniversitesi, Elâzığ.
- [10] **G. Xu, Y. Zhang, L. Li,** (2011)“Web Mining and Social Networking”, Web Information Systems Engineering and Internet Technologies Book Series Volume 6, 89-107.
- [11] **D. Greene, J. Freyne, B. Smyth and P. Cunningham,** (2008)“An Analysis of Research Themes in the CBR Conference Literature”, 18–43
- [12] **Volkova, S.,** (2009). “Link Prediction in Social Networks”, KDD Lab Course Project.
- [13] **Carminati, B., Ferrari, E. ve Perego, A.,** (2006). “Rule-Based Access Control for Social Networks”, *Lecture Notes in Computer Science*, 4278, 1734-1744.
- [14] **Gastner, M.T.,** (2011). “Networks: Theory and Applications”, Imperial College.
- [15] **P. Mika,** (2007), *Social Networks and the Semantic Web, Semantic Web and Beyond Computing for Human Experience*, Jain,R. and Sheth, A.(Eds.), New York, Springer Science and Business Media, 234.

- [16] **Huisman, M., ve van Duijn, M.A.J.**, (2011). “The SAGE Handbook of Social Network Analysis”, 578-600.
- [17] **Huisman, M., ve van Duijn, M.A.J.**, (2004) “Software For Statistical Analysis Of Social Networks”, University of Groningen.
- [18] **Ting, I.H.**, (2008) “Web Mining Techniques for On-line Social Network Analysis”, IEEE, 696-700.
- [19] **S. Fortunato**, Community detection in graphs. 0906.0612 (2009).
- [20] **J. Galaskiewicz and S. Wasserman**, (1994) Advances in Social Network Analysis: Research in the Social and Behavioral Sciences, Introduction: Advances in the Social and Behavioral Sciences from Social Network Analysis, Sage Focus Editions, 171, Thousand Oaks, California, 299
- [21] **Mika, P.**, (2007). “Social Networks and the Semantic Web, Semantic Web and Beyond Computing for Human Experience”, Jain,R. and Sheth, A.(Eds.), New York, Springer Science and Business Media, 234
- [23] **M.E.J. Newman**, (2008)Mathematics of Networks, 12., The New Palgrave Dictionary of Economics, Durlauf, S.N. and Blume, L.E. (Eds.), Palgrave Macmillan.
- [24] **S. Fortunato**, (2009) Community detection in graphs. 0906.0612.
- [25] **Getoor, L. ve Diehl C.P.**, (2005). “Link Mining: A Survey”, SIGKDD Explorations, 7(2), 3-12.
- [26] **Menon, A.K. ve Elkan, C.**, (2011). “Link Prediction via Matrix Factorization”, ECML/PKDD 2, 437-452.
- [27] **Scott, J.**, (2000). “Social Network Analysis: A Handbook,” Sage Publications, 209.
- [28] **Newman, M.E.J.**, (2008). “Mathematics of Networks”, 12pp., The New Palgrave Dictionary of Economics, Durlauf, S.N. and Blume, L.E. (Eds.), Palgrave Macmillan.
- [29] **Schank, T. ve Wagner, D.**, (2005). “Approximating clustering-coefficient and transitivity”, Journal of Graph Algorithms and Applications, 9: 265–275.
- [30] **Hanneman, R.A. ve Riddle M.**, (2005). “Introduction to social network methods”, Riverside, CA: University of California, Riverside.
- [31] **Getoor, L. ve Diehl C.P.**, (2005). “Link Mining: A Survey”, SIGKDD Explorations, 7(2), 3-12.
- [32] **T. Schank and D. Wagner**, (2005), Approximating clustering-coefficient and transitivity, Journal of Graph Algorithms and Applications, 265– 275.
- [33] **Menon, A.K. ve Elkan, C.**, (2011). “Link Prediction via Matrix Factorization”, ECML/PKDD 2, 437-452.
- [34] **Shibata, N., Kajikawa, Y., ve Sakata, I.**, (2012). “Link Prediction in Citation Networks”, Journal of The American Society for Information Science and Technology, 63(1), 78-85.

[35] **Lü, L., Zhou, T.**, (2011). “Link Prediction in Complex Networks: A survey”, *Physica A* 390, 1150-1170.

[36] **Liben-Nowell, D. ve Kleinberg, J.**, (2003). “The link prediction problem for social networks”, *Proceedings of the 12th International Conference on Information and Knowledge Management (CIKM)*, New York: ACM Press, 556–559.

[37] **Vartak, S.**, (2008). “A survey on Link Prediction”, State University of New York.

[38] **Volkova, S.**, (2009). “Link Prediction in Social Networks”, KDD Lab Course Project.

İNTERNET KAYNAKLARI

[22] **SONIVIS**, (2017)"Social Network Analysis Global Clustering Coefficient Metrics", http://sonivis.org/wiki/index.php/Global_Clustering_Coefficient, Erişim tarihi: 24 Eylül.

[38] <https://dev.twitter.com> (2017) Erişim Tarihi: 16.10.2017



EKLER

EK A: Eğitim Düzeyi Tespitinde Kullanılan Filtre Kelimeler

EK B: “Sosyal Ağlardan Eğitim Düzeyi Tahmini” adlı tez çalışmasının PDF halini CD’de teslim edilecektir.





EK-A

YDS	TOEFL	KPSS	TUS
DUS	Kurum Sınavları	Sınav	Not
Kontenjan	Mülakat	Başvuru	Belge
İngilizce	Skor	Denetmenlik	Bilgi
Öğrenim	Öğretim	Yetkinlik	Üniversite
Fakülte	Yüksek Lisans	Lisans	Doktora
Phd	Kariyer	Yeterlilik	Kitap
Profesör	Doçent	Doktor	Asistan
Eğitim	Yüksek öğrenim	Uzmanlık	Hakimlik
Ön Mülakat	Analiz	Etüt	Akademi
Akademik	Burs	KYK	Kamu
Yurt	Makale	Proje	Bitirme projesi
Bütünleme	Kampüs	Yerleşke	Dekanlık
Yaz okulu	Dekan	Öğretim görevlisi	Öğretim elemanı
Rektörlük	Rektör	Amfi	



ÖZGEÇMİŞ

Rudi Miran BABAYİ, 04.05.1990 tarihinde İstanbul'da doğdu. 2004'de Karagözyan Ermeni İlköğretim Okulu'ndan mezun oldu. 2007 yılında Maçka Akif Tuncel Endüstri Meslek Lisesi Bilgisayar Yazılım bölümünden mezun oldu. 2009 yılında Maltepe Üniversitesi Meslek Yüksekokulu Bilgisayar Teknolojileri ve Programlama bölümünden mezun oldu. 2013 Yılında Anadolu Üniversitesi İşletme Fakültesi bölümünden mezun oldu. 2015 yılında İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği'nde yüksek lisans eğitimine başladı. 2007 yılından bu yana özel sektörde sistem, yazılım ve bilgisayar ağları alanlarında çalışmalarına devam etmektedir. Orta derecede İngilizce, İyi derecede Ermenice dili bilmektedir. İlgi alanlarından bazıları, Bilgisayar Ağları, Sosyal Ağlar ve Ağ ve Sistem Güvenliğidir.

