TWITTER ETİKET BİRLİKTELİK AĞLARININ MODELLENMESİ VE ANALİZİ

SEMA ARSLAN
161402206

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans
Danışman: Dr. Öğr. Üyesi Volkan TUNALI

İstanbul
T.C. Maltepe Üniversitesi
Fen Bilimleri Enstitüsü
Ocak, 2019

JÜRİ VE ENSTİTÜ ONAYI

JÜRİ VE ENSTİTÜ ONAYI

Sema ARSLAN'ın "Twitter Etiket Birliktelik Ağlarının Modellenmesi Ve Analizi" başlıklı tezi 25.01.2019 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Maltepe Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği"nin ilgili maddeleri uyarınca, Bilgisayar Mühendisliği Anabilim Dalında Yüksek Lisans/Doktora tezi **oy birliğiyle** / ~~**oy çokluğuyla**~~ olarak kabul edilmiştir.

	Unvanı. Adı ve soyadı	İmza
Üye (Tez Danışmanı)	: Dr. Öğr. Üyesi Volkan TUNALI	
Üye (Altınbaş Üniversitesi)	: Dr. Öğr. Üyesi Can RAZBONYALI	
Üye	: Dr. Öğr. Üyesi Mehmet Ali Aksoy TÜYSÜZ	

Prof. Dr. İlter BÜYÜKDIĞAN

Enstitü Müdürü

ETİK İLKE VE KURALLARA UYUM BEYANI

maltepe üniversitesi	ETİK İLKE VE KURALLARA UYUM BEYANI	Doküman No	FR-178
		İlk Yayın Tarihi	01.03.2018
		Revizyon Tarihi	
		Revizyon No	00
		Sayfa	ii/49

Revizyon Takip Tablosu

REVİZYON NO	TARİH	AÇIKLAMA
00	01.03.2018	İlk yayın.

ETİK İLKE VE KURALLARA UYUM BEYANI

25/01/2019

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarından bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilmeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; çalışmanın Maltepe Üniversitesinde kullanılan "bilimsel intihal tespit programı" ile tarandığını ve öngörülen standartları karşıladığını beyan ederim.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

Sema Arslan



İNTİHAL RAPORU

Twitter Etiket Birliktelik Ağlarının Modellenmesi Ve Analizi

ORIJINALLIK RAPORU

%9	%7	%1	%6
BENZERLIK ENDEKSI	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCI ÖDEVLERI

BIRINCIL KAYNAKLAR

1	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	%4
2	m.friendfeed-media.com İnternet Kaynağı	%1
3	oda.hio.no İnternet Kaynağı	%1
4	essay.utwente.nl İnternet Kaynağı	%1
5	angora.baskent.edu.tr İnternet Kaynağı	%1
6	www.sosyalaganalizi.com İnternet Kaynağı	<%1
7	SEÇKİN CODAL, Keziban and COŞKUN, Erman. "SOSYAL AĞ TÜRLERİNİN KARŞILAŞTIRILMASINA İLİŞKİN BİR AĞ ANALİZİ", Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü, 2016.	<%1

TEŞEKKÜR

Tez konusu seçiminde ve tez süreci boyunca destek ve yardımlarını esirgemeyen, hocam Dr. Öğr. Üyesi Volkan Tunalı'ya, eğitim hayatım boyunca her konuda desteğini esirgemeyen hocam Dr. Öğr. Üyesi Ali Akman'a, varlığıyla sonsuz motivasyon kaynağım olan aileme ve Bilal Kaplan'a teşekkürlerimi sunarım.

Sema Arslan

Ocak 2019

ÖZ

Twitter Etiket Birliktelik Ağlarının Modellenmesi ve Analizi

Sema Arslan
Yüksek Lisans Tezi
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans
Danışman: Dr. Öğr. Üyesi Volkan TUNALI
Maltepe Üniversitesi Fen Bilimleri Enstitüsü, 2019

Bu tez çalışmasında, Twitter paylaşımlarının toplanması, paylaşımlardaki etiket birlikteliklerinin ağ modelinin oluşturulması ve analizi hedeflenmiştir. Oluşturulan ağ modelini sosyal ağ analizi teknikleriyle analiz ederek, sonuçları etkin şekilde sunan bir yazılım aracı geliştirilmiştir.

Geliştirilen yazılım aracı, sabit bir veri seti değil, Twitter platformundan anlık olarak çekilen güncel veriler üzerinden işlem yapmaktadır. Uygulama tarafından birliktelik ağı oluşturulan etiketler; merkezilik, topluluk ve yapısal ağ analizi yöntemlerine göre incelenerek ağın en önemli etiketler ve etiket toplulukları ortaya çıkarılmıştır. Elde edilen analiz sonuçları sayesinde, incelenen konu etiketinden yola çıkılarak ulaşılan ilişkili konuların bütüncül bir bakış açısıyla anlaşılması mümkün olmaktadır.

Anahtar Sözcükler: Twitter, Etiket, Birliktelik, Analiz, Ağ Bilimi, Sosyal Ağ Analizi, Ağ Görselleştirme

ABSTRACT

MODELING AND ANALYSIS OF TWITTER HASHTAG CONNECTION NETWORK

Sema Arslan
Master Thesis
Master of Science in Computer Engineering Program
Master of Science in Computer Engineering
Thesis Advisor: Assist. Prof. Dr. Volkan TUNALI
Maltepe University, Graduate School of Science and Engineering, 2019

In this thesis, it was aimed to collect Twitter posts, model and analyze the co-occurrence network of hashtags present in the posts. A new software tool was developed to analyze the network model using social network analysis techniques and to present the analysis results effectively.

The developed tool processes data retrieved online rather than previously stored static data. The hashtag co-occurrence network is analyzed by the tool using centrality, community and structural analysis techniques to reveal the most important hashtags and hashtag communities. It is possible to understand the related concepts from a holistic perspective beginning with a single concept hashtag using the obtained analysis results.

Keywords: Twitter, Hashtag, Network, Analysis, Network Science, Social Network Analysis, Network Visualization

İÇİNDEKİLER

TwItter Etiket Birliktelik Ağlarının Modellenmesi ve Analizi i
JÜRİ VE ENSTİTÜ ONAYI i
ETİK İLKE VE KURALLARA UYUM BEYANI ii
İNTİHAL RAPORU iii
TEŞEKKÜR iv
ÖZ v
ABSTRACT vi
İÇİNDEKİLER vii
TABLOLAR LİSTESİ ix
ŞEKİLLER LİSTESİ x
KISALTMALAR xi
ÖZGEÇMİŞ xii
BÖLÜM 1. GİRİŞ 1
BÖLÜM 2. İLGİLİ ÇALIŞMALAR 3
BÖLÜM 3. AĞ ANALİZİ 5
3.1. Graf Teorisi 5
3.2. Graf Gösterim Şekilleri 8
3.2.1. Komşuluk Matrisi 8
3.2.2. Kenar Listesi 9
3.2.3. İlişki Matrisi 10
3.3. Ağ Ölçütleri 10
3.3.1. Merkezilik Ölçütleri 10
3.3.2. Kümelenme Katsayısı Ölçütleri 14
BÖLÜM 4. TWITTER 17
BÖLÜM 5. YÖNTEM 19
5.1. Veri Toplama 21
5.2. Veri Temizleme 22
5.3. Kullanılan Teknolojiler 23
5.3.1. Asp .NetCore 23
5.3.2. MongoDb 23
5.3.3. Tweetinvi 24
5.3.4. SNAP 24
5.3.5. D3.js 25

5.3.6. Bootstrap 25
5.4. Geliştirilen Sistemin Uygulanması 26
BÖLÜM 6. SONUÇ 33
KAYNAKÇA 34

TABLOLAR LİSTESİ

Tablo 1. Komşuluk Matrisi 9
Tablo 2. Kenar Listesi 9
Tablo 3. Yönsüz Ağda Düğümler Arası En Kısa Yol 11

ŞEKİLLER LİSTESİ

Şekil 1. Zachary’s Karate Club’a Ait Graf 5
Şekil 2. Yönsüz Graf 7
Şekil 3. Yönlü Graf 7
Şekil 4. İşaretli Graf 8
Şekil 5. Zachary’s Karate Club'a ait Derece Merkeziliği 11
Şekil 6. Zachary’s Karate Club'a ait Yakınlık Merkeziliği 12
Şekil 7. Arasındalık Merkeziliğine örnek Graf 13
Şekil 8. Zachary's Karate Club'a ait Arasındalık Merkeziliği 13
Şekil 9. Zachary's Karate Club'a ait Özvektör Merkeziliği 14
Şekil 10. Yönsüz graf 15
Şekil 11. Tweet Yapısı 18
Şekil 12. Sistemin Blok Diyagramı 19
Şekil 13. Akış Diyagramı 20
Şekil 14. API üzerinde kimlik doğrulama 21
Şekil 15. Search API Tweet Sorgulama 21
Şekil 16. Twitter API Sorgu Sonucu 22
Şekil 17. MongoDB Genel Yapısı 24
Şekil 18. Kullanıcı Arayüzü 26
Şekil 19. Genel Analiz 28
Şekil 20. Derece Merkezilik Analizi Ağ Görünümü 28
Şekil 21. Merkezilik Analizi Tablo Görünümü 29
Şekil 22. Topluluk Analizi Ağ Görünümü 30
Şekil 23. Topluluk Analizi Tablo Görünümü 30
Şekil 24. Yapısal Analiz Görünümü 31
Şekil 25. Yapısal Analiz Grafik Detay Görünümü 32
Şekil 26. Yapısal Analiz Tablo Detay Görünümü 32

KISALTMALAR

API : Application Program Interface

CSS : Cascading Style Sheets

DOM : Document Object Model

HTML : Hyper Text Markup Language

SNAP : Stanford Network Analysis Platform

SVG : Scalable Vektor Graphics

RT : Retweet

TT : Trend Topic

ÖZGEÇMİŞ

Sema ARSLAN

Bilgisayar Mühendisliği Anabilim Dalı

Eğitim

Ls.	2009	Maltepe Üniversitesi, Mühendislik ve Doğa Bilimler Fakültesi Bilgisayar Mühendisliği Anabilim Dalı
Lise	2006	Özel İlkem Güneş Koleji

İş/İstihdam

2016 -	Yazılım Uzmanı Doğuş Teknoloji
2014- 2016	Yazılım Uzmanı Logo Yazılım

Kişisel Bilgiler

Doğum yeri ve yılı : Kahramanmaraş 1989 Cinsiyet:Kadın
Yabancı diller : İngilizce
GSM / e-posta : 0541 280 05 18 / semarslan46@gmail.com

BÖLÜM 1. GİRİŞ

Sosyal medya kullanımının son yıllarda artmasıyla insanların günlük yaşantılarında gözle görülür birçok değişiklik meydana gelmiştir. Internet üzerinden yapılan alışverişler, rezervasyonlar, haberleşme vs. birçok etkileşim araçları her geçen gün önemini arttırmakta ayrıca kullanımı yaygınlaşmaktadır. Başlıca sosyal medya platformu olarak Facebook, Instagram, Twitter, Linkedln akla gelmektedir. Facebook, insanların fikirlerini, fotoraflarını paylaşabildiği başlıca sosyal medya platfomudur. Instagram, kısa süreli videolar ve fotoğraf paylaşım platformudur. Twitter, kullanıcıların anlık duygu ve düşüncelerini 280 karakter sınırı ile ifade edebildikleri platformdur. Linkedln, üyelerin iş arama ya da meslektaşları ile ilgili gündemi takip etmek üzere kullanılan sosyal medya platformudur.

Sosyal medyada yapılan paylaşımlar sosyal medya madenciliği yöntemleriyle çeşitli açılardan analiz edilerek önemli çıkarımlar yapmak mümkün olmaktadır. Paylaşımların metin içeriğinin anlamsal olarak analizinin yanı sıra, paylaşımlardan oluşturulan ağ yapıları da önemli bilgiler içerebilmektedir. Çevrimiçi sosyal medya kullanıcılarının hızla artması, klasik analiz metodolojilerinin kullanımı ile analiz edilemeyecek büyüklüğe ulaşmıştır. Bu nedenle, veri madenciliği ve sosyal meyda madenciliği alanında yapılan akademik ve teknik çalışmalar önemini arttırmıştır. Sosyal medya madenciliği birçok çalışma alanında yeni imkanlar sunmaktadır. Satış pazarlama başta olmak üzere birçok sektörde çalışanlar için farklı bakış açıları sunmaktadır.

Sosyal medya platformları, analiz edilmesi için büyük veri havuzları oluşturmaktadır. Yapılan çalışmalar ile sosyal medya verilerinin güvenilir metotlar aracılığıyla analiz ediliyor olması birçok kaynak tarafından ispatlanmıştır. Sosyal medya verileri ile ekonomi analizler, hastalık yayılımı, spor sonuçları ve haber yayınları gibi analizlerin doğru tahminler verildiği görülmüştür.

Herhangi bir konuda (spor, ekonomi, alışveriş, haber gibi) gelişmeyi anlık olarak sosyal medya platformlarından paylaşılmakta ve diğer takipçiler ile fikir alışverişinde bulunulmaktadır. Ayrıca sosyal medya kullanıcıları arasında istenilen bilgi daha hızlı yayılmaktadır. Günümüzde başta devletler olmak üzere şirketler, topluluklar, kurum ve kuruluşlar alınacak kararları zaman zaman sosyal medya üzerinde oluşan algıya göre

şekillendirmekte ve karar almaktadır. Kullanıcılar kendi fikir ve görüşlerini diğer insanlarla paylaşıp, bilgilendirmektedir.

Twitterda yapılan yazılı paylaşımlarda bir arada geçen etiketlerin (hashtag) birlikteliklerinin ağ yapısı şeklinde modellenmesi ve bu ağ yapısının analizi sayesinde kolayca görülemeyen ilişki ve etkileşimler ortaya çıkartılabilmektedir. Tüm bu işevleri bir arada ve bütüncül bir yapıda sunan kullanışlı araçlar bulunmamakta, mevcut araçlar ise analiz aşamasında ve analiz sonuçlarının kolay bir şekide sunulması aşamasında yetersiz kalmaktadır.

Bu çalışmada twitter paylaşımlarının toplanması, depolanması, bu paylaşımlardaki etiketlerin birliklerinden ağ modellenmesi ve bu ağın sosyal ağ analizi teknikleriyle analiz edilerek analiz sonuçlarını etkin bir şekilde sunan bir yazılım aracı geliştirilmesi amaçlanmaktadır.

BÖLÜM 2. İLGİLİ ÇALIŞMALAR

Bu bölümde Twitter ile ilgili sosyal ağ analizi alanında yapılan çalışmalardan bahsedilmektedir.

A. Köktürk'ün 2013 yılında yaptığı çalışmada, üniversitelerin Twitter kullanım şekillerinin anlaşılması ve üniversiteler arasında bir sıralamanın yapılıp yapılamadığını tespit etmek amacıyla, hâlihazırda etkili hesapları ölçmeye yarayan araçlar, üniversitelerin Twitter kullanım şekilleri ve bu alanda yapılan çalışmalar incelenerek, bir ölçek sistemi geliştirmiştir [1].

E. Beğenilmiş 2011 yılında yaptığı çalışmada, 200 milyon tweet kümesi içerisindeki davranışların "organize" veya "organk" olduğunu anlayabilen, otomatik olarak kullanılabilecek bir sınıflandırma modeli sunmuştur [2].

Y. E. Karabulut'un 2018 yılında yaptığı çalışmada, geliştirilen uygulama yardımıyla Twitter üzerinde kullanıcıların metinlerinde kullandığı etiketler (#hashtag) üzerinden analiz yapılabilmektedir. Yapılan analiz yardımı ile meydana gelen toplumsal bir hareketin ve kullanılan etiketin detaylarına ulaşılabilmektedir. Oluşan etiket kullanımı ile ilgili atılan tweet sayısı, tweet atan gerçek kişi sayısı, etiket kullanımının başladığı zaman dilimi, etiket kullanımını başlatan kullanıcı ve tweet, etkileşim sayısı, atılan tweetler üzerinden tweetlerin olumlu veya olumsuz olma durumunu analiz eden duygu analizi ve tweet atan kullanıcılar üzerinden cinsiyet analizi yapabilmektedir [3].

Hao 2011 yılında yapmış olduğu Twitter'ın oluşturmuş olduğu veri setindeki görsel duyarlılık analizi çalışmasında, Twitter'ın günde yaklaşık 190 milyon tweet atılan bir sosyal paylaşım platformu olduğundan bahsetmiştir. Twitter'ın birçok alan hakkında fikir verebilecek tweetleri içerdiğini belirtmiştir. Yapmış olduğu çalışmada üç yeni zaman bazlı görsel duyarlılık analizi tekniği sunmuştur. Sunmuş olduğu tekniklerden birincisi; müşteri görüşlerini ayıklayan, haritalayan ve ölçen konu bazlı duyarlılık analizi tekniğidir. İkinci teknik; ilginç tweetleri yoğunluklarına, olumsuzluklarına ve etki özelliklerine göre tanımlayan akış analizidir. Üçüncü teknik ise; büyük hacimli veriyi tek bir görünümde görselleştiren piksel hücresitabanlı duyarlılık takvimleri ve yüksek yoğunluklu coğrafi haritaların analizidir [4].

Wang, Can 2012 yılında yapmış oldukları çalışmalarında, 2012 ABD Seçimlerinde Twitter'da açıklanan başkan adaylarına yönelik, kişilerin atıkları tweetler

üzerinden duyguların gerçek zamanlı analizine yönelik bir sistem geliştirmişlerdir. Twitter'ın insanların siyasi partiler ve adaylar hakkında görüşlerini dile getirdiği merkezi bir site ya da sosyal paylaşım platformu olduğundan bahsetmişlerdir. Geleneksel içerik analizinin tamamlanmasının haftalar aldığını, önerilen sistemde ise, seçimle ilgili tüm tweetlerde duygu analizi yapılarak sonuçların anında ve sürekli olarak izlenebildiğini belirtmişlerdir [5].

Bruns 2012 yılında yapmış olduğu çalışmasında, Twitter tabanlı iletişimin analizi için yaygın olarak uygulanabilir, standartlaştırılmış metriklerden oluşan bir kataloğu özetleyerek Twitter üzerindeki iletişimin analiz edilmemesi sorununa değinmiş, özellikle hashtaglere odaklanmıştır. Ayrıca, çeşitli ölçütlerin daha kapsamlı karşılaştırmalarının nasıl elde edebileceğine ilişkin bir gösterge sunarak, bu metriklerin potansiyel kullanımlarını da işaret etmiştir [6].

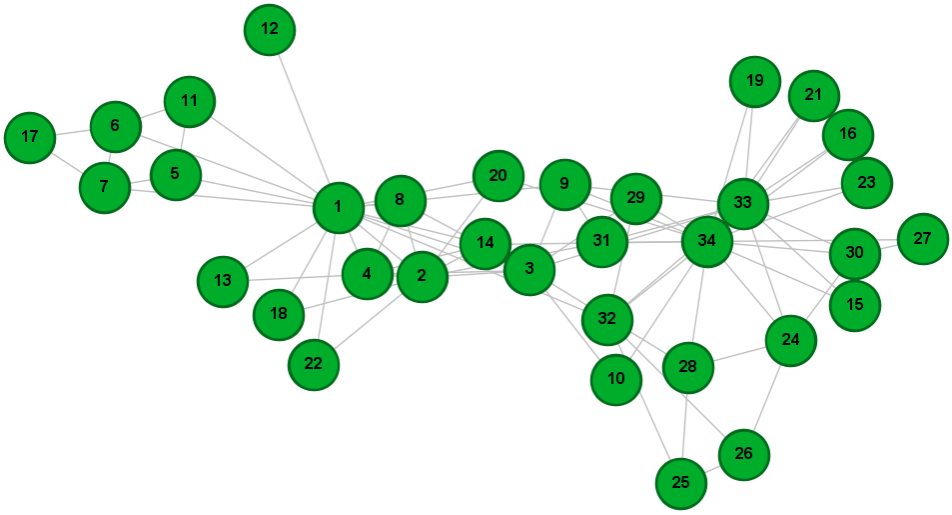
BÖLÜM 3. AĞ ANALİZİ

3.1. Graf Teorisi

Leonard Euler'in 1736 yılında, "Königsberg'in Yedi Köprüsü" adlı yayınladığı makale graf teorisinin (graph theory) başlangıcını oluşturmuştur. James Joseph Slyvester tarafından ilk kez Graf sözcüğü 1822 yılında kullanılmıştır [7].

Graf, düğüm olarak adlandırılan noktalar ve bu noktaları birleştiren bağlantılardan oluşan, sadece düğümler arasındaki ilişkiyi gösteren çizgiler topluluğudur.

Graf teorisi, bilgisayar, matematik ve fizik gibi bilim dallarında yaygın olarak kullanılmıştır. Graf teorisi ile ilgili olarak akla ilk gelen graf Zachary's ait olan Karate Club graf örneğidir. Zachary, üniversitenin karate kulübü üyelerinin kulüp dışında nasıl etkileşim halinde olduğunu gösteren graf örneğini oluşturmuştur [8]. Elektronik devreler, veritabanları, işletim sistemlerinde, ekonomi, yönetim bilimi, satış pazarlama ve navigasyon sistemlerinde graflar kullanılır. Graf teorisi, geniş kapsamlı birçok problemin çözümü için tercih edilmektedir.



Şekil 1 . Zachary's Karate Club'a Ait Graf

Yönsüz bir graf düğümler ve ayrıtlar kümesinden oluşur. G(V,E) gösterilmektedir.

(i) boş olmayan sonlu bir V düğümler kümesi

(ii) solu bir E ayrıtlar kümesi

(iii) bir $\delta : E \rightarrow P(V)$ fonksiyonu öyle ki her bir e ayrıtı için $\delta(e)$ V'nin bir veya iki elemanlı bir alt kümesidir.

Düğüm, köşe veya verteks olarak da adlandırılmaktadır. Graflardaki en temel birim düğümdür. Bir graftan söz edebilmek için en az bir düğüm gereklidir. *n* adet düğümden oluşan bir grafa ait düğümler kümesi Denklem (1)'deki gibi gösterilmektedir [9].

$$V = \{ v_1, v_2, \dots, v_n \} \quad (1)$$

Denklem (1)' deki ifadeki *n* sayısı grafın boyutudur. Graflarda boyut |V|=*n* şeklinde gösterilir.

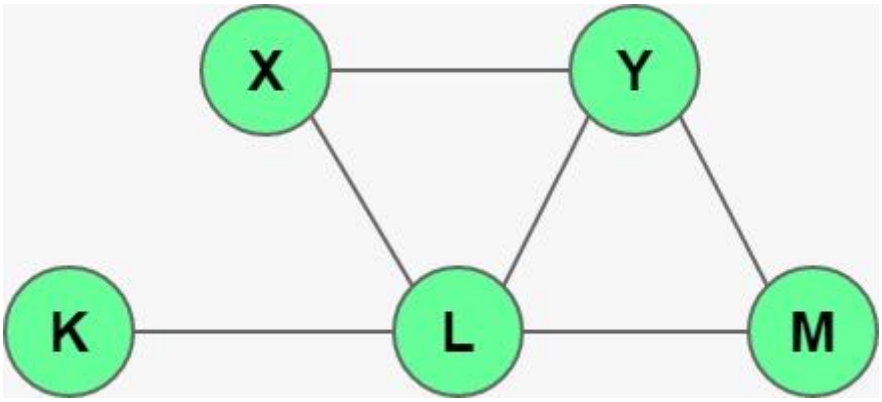
Kenar, düğümleri birbirine bağlayan bağlantılardır. Grafın temsil ettiği yapıya göre ilişki, sosyal bağ veya bağlantı isimleriyle de anılır. Bir grafta kenar zorunluluğu olmasına rağmen düğüm zorunluluğu bulunmamaktadır.

Grafta, düğümleri birbirine bağlayan *m* tane kenar içeren kenar kümesi Denklem (2)' deki gibidir [9] .

$$E = \{ e_1, e_2, \dots, e_m \} \quad (2)$$

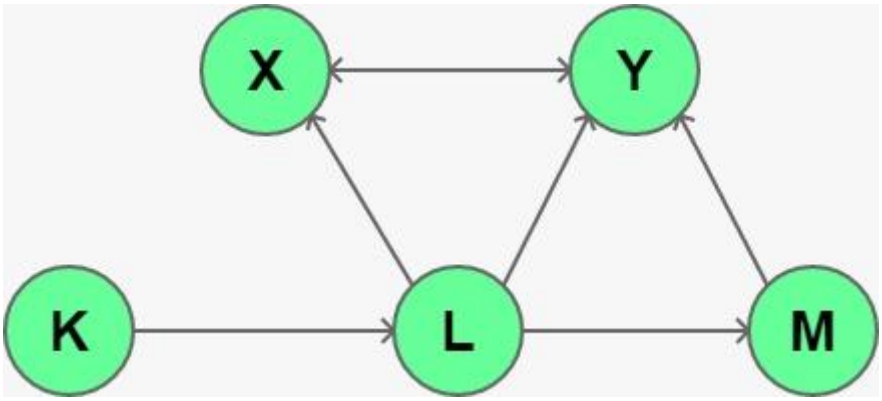
Graflarda kenarlar, düğümler arasındaki ilişkiyi tanımlar. Kenarlar yönlü (directed) veya yönsüz (undirected) olarak tanımlanabilir. Yönlü kenarlar yay olarak da adlandırılır. Yönsüz kenarların çizimi için düz çizgi kullanılırken, yönlü kenarların çiziminde yönü gösterecek şekilde oklu çizgi kullanılmaktadır. Yönsüz kenarlara simetrik bağ, yönlü kenarlara ise asimetrik bağ denilir [9].

Yönsüz graf genellikle Facebook'taki karşılıklı arkadaşlık benzeri durumları temsil etmektedir. *X* kişisi *Y* kişisi ile arkadaş ise *Y*'nin *X* ile arkadaş olduğu da kesindir.



Şekil 2. Yönsüz Graf

Yönlü graf genellikle Instagram'daki takip etme gibi, karşılık beklenmeyen durumları temsil eder. *X* kişisi *Y* kişisini takip edebilir, ancak *Y* kişisinin *X* kişisini takip etme gibi zorunluluk yoktur.



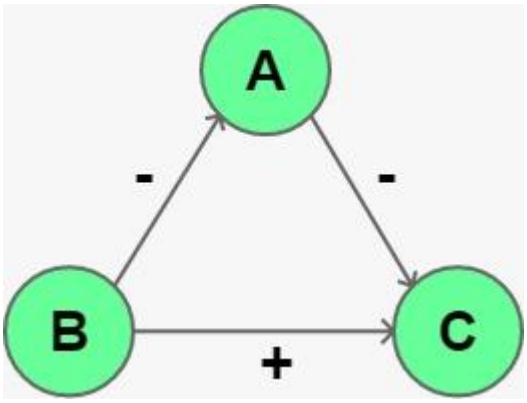
Şekil 3. Yönlü Graf

Garflar genellikle ya yönlü ya da yönsüz kenarlar içeren graflardır. Nadiren de olsa hem yönlü hem de yönsüz kenarlar içeren graf olabilir. Bu tür graflara karışık graf denir.

Graflarda kenarlarla ilgili temel özelliklerden diğeri ağırlıktır. Bazı kaynaklar kenar ağırlığı yerine değer (value) deyimine yer vermektedir. Kenarın değerliği, kenarın birleştirdiği iki düğüm arasında tanımlanan ilişkiyi ifade eden sayısal değerdir. Her kenarın değerliği olmak zorunda değildir. Örneğin, harita üzerindeki şehirleri ve bu şehirleri birleştiren yolları temsil eden bir graf düşünelim. Şehirler arasındaki mesafeyi kenar değerleri temsil etmektedir. Matematiksel olarak, ağırlıklı bir graf G(V,E,W)

şeklinde gösterilir. W sembolü; E'deki her bir kenarın ağırlığının kümesini gösterir ve daima |W|=|E| eşitliği geçerlidir.

Ağırlıklı grafın özel bir hali ise pozitif veya negatif işaretli kenarlara sahip işaretli graflardır. İşaretli kenarlar genellikle düğümler arasındaki zıtlık durumlarını tanımlar.



Şekil 4. İşaretli Graf

3.2. Graf Gösterim Şekilleri

Matematiksel ve bilgisayar programları tarafından Grafların hesaplanabilirliğini kolaylaştırmak amacıyla gösterim şekilleri belirlenmiştir.Temel de kullanımı yaygın olan üç gösterim şekli vardır. İlki kenar listesi (edge list) ikincisi komşuluk matrisi (adjancecy matrix) sonuncusu ilişki matrisi (incidence matrix). Her gösterim şeklinin avantajları ve dezavantajları vardır.

3.2.1. Komşuluk Matrisi

Grafları göstermenin en temel ve basit yöntemi komşuluk matrisi tekniğidir. Komşuluk matrisi bazı kaynaklarda sosyomatris (socimatrix) olarak da ifade edilmektedir. Komşuluk matrisi grafların matematiksel olarak en doğal gösterim şeklidir.

Komşuluk matrisinin hücrelerindeki 1 değeri, o hücreye karşılık gelen satır ve sütundaki düğümler arasında bir bağlantı yani bir kenar olduğunu, 0 değeri ise aralarında herhangi bir bağlantı olmadığını göstermektedir. Ağırlıklı graflarda ise 1 değeri yerine ilgili kenarın sayısal ağırlık değeri kullanılır [9]. Şekil 2' deki grafa ait komşuluk matrisi Tablo 1' de verilmiştir.

Tablo 1. Komşuluk Matrisi

	A	B	C	D	E
A	0	1	0	0	0
B	1	0	1	1	1
C	0	1	0	0	1
D	0	1	0	0	1
E	0	1	1	1	0

3.2.2. Kenar Listesi

Graflarda kenar listesi, her bir satırında birbiriyle bağlantılı iki düğümü içeren listelerdir. Tablo incelendiğinde Şekil 4'teki graf yönlü olduğu için farklı yönleri temsil eden ayrı satırlar kullanılmıştır. Elle oluşturulması kolay ve bilgisayarda işlenebilirliği açısından avantajlıdır. Kenar listesinin diğer avantajı ise her bir satıra o kenarla ilgili etiket, kenar ağırlığı gibi bilgilerinin eklenebilir olmasıdır. Sosyal ağların dosya olarak saklanmasında kenar listesini temel alan dosya formatları oldukça kullanışlıdır. Şekil 2'de ki grafa ait kenar listesi Tablo 2'de gösterilmiştir.

Tablo 2. Kenar Listesi

A,B
B,D
D,B
D,C

3.2.3. İlişki Matrisi

Komşuluk matrisi gibi, büyük grafların temsilinde pek kullanışlı olmayan diğer graf gösterim şekli ilişki matrisidir (incidence matrix). Bu matrisin satırları düğümlerden ve sütunları ise kenarlardan oluşmaktadır.

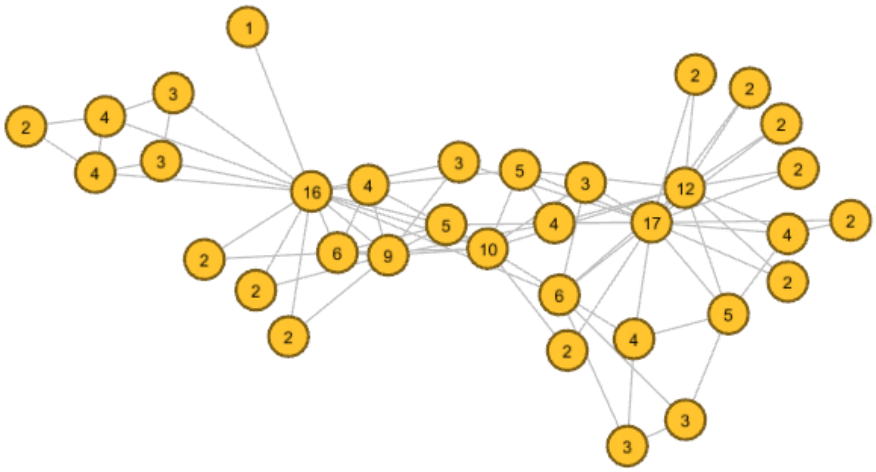
3.3. Ağ Ölçütleri

Merkezilik ölçütlerinin en başında derece, yakınlık, arasındalık ve özvektör merkezilik ölçütleri gibi sıklıkla kullanılan ölçütler gelmektedir. Bu merkezilik ölçütlerinden her birinin en etkili düğümü bulmadaki varsayımı birbirinden farklıdır. Dolayısıyla, her birinin bir ağda herhangi bir düğümü etkili veya merkezi yapan farklı bir bakış açısı vardır.

3.3.1. Merkezilik Ölçütleri

Derece Merkeziliği, bir grafta, bir düğüm ile ilgili en temel ölçüt derecedir. Yönsüz ağlarda bir düğümün derecesi o düğüme bağlı kenarların sayısıdır. Yönlü ağlarda ise iç-derece (in-degree) ve dış derece (out-degree) şeklinde iki derece vardır. Düğüme gelen kenarların sayısı iç derece, düğümden çıkan kenarların sayısı ise dış-derecedir. Yönlü ağlarda iç ve dış derecelerin toplamı düğümün toplam derecesini verir.

En basit merkezilik ölçütü derece merkeziliğidir. Grafta bulunan düğümlerin yalnızca derecelerini kullanarak ağdaki önemini belirlemektedir. Yani bir düğümün derecesi ne kadar yüksek ise o düğüm o kadar önemlidir. Yönlü ağlarda ise iç-derece merkeziliği bir bireyin sosyal ağdaki popülerliğini, dış derece merkeziliği ise bireyin sokulganlık ve sosyalleşebilirliğini (gregariousness, sociability) temsil eder [9]. Zachary's Karate Club ağı için düğüm derece merkeziliği değerleri Şekil 5'te gösterilmiştir.



Şekil 5. Zachary’s Karate Club'a ait Derece Merkeziliği

Yakınlık Merkeziliği, ağdaki bir düğümün diğer tüm düğümlere yakınlığına Yakınlık Merkeziliği denir [9]. Bir düğüm bu ölçüte göre ne kadar merkezi ise, belirlenen düğümden ağdaki diğer düğümlere o kadar çabuk diğer bir deyişle kısa yoldan ulaşabilir denilebilir. Bir düğümün yakınlık merkeziliği hesaplanırken, o düğümün ağdaki diğer tüm düğümlere olan en kısa yol uzunluklarının ortalaması hesaplanır ve bu değerin tersi alınır. Özetle, ortalama en kısa yol ne kadar küçükse yakınlık merkeziliği değeri o kadar yüksektir. Tablo 3 ‘te yönsüz ağda düğümler arasındaki en kısa yol uzunlukları gösterilmiştir.

Tablo 3. En kısa yol

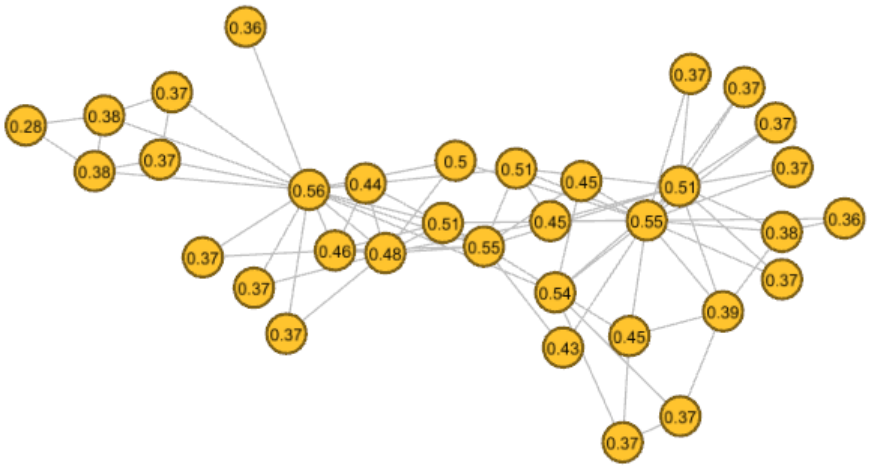
	A	B	C	D	E
A	0	1	2	2	2
B	1	0	1	1	1
C	2	1	0	2	1
D	2	1	2	0	1
E	2	1	1	1	0

Yönsüz bir ağda her bir düğümün yakınlık merkeziliğini hesaplamak için önce düğümler arası en kısa yol uzunlukları Tablo 3 ‘teki gibi bulunur. Daha sonra, her bir düğüm için hesaplama yapılır. Denklem 3’te yakınlık merkeziliği hesaplaması bulunmaktadır.

$$
\begin{aligned}
C_{clo}(A) &= 1 / ((1 + 2 + 2+ 2) / 4) = 1/ 1{,}75 = 0{,}57 \\
C_{clo}(B) &= 1 / ((1 + 1 + 1+ 1) / 4) = 1/ 1 \quad = 1 \\
C_{clo}(C) &= 1 / ((2 + 1 + 2+ 1) / 4) = 1/ 1{,}5 \quad = 0{,}67 \\
C_{clo}(D) &= 1 / ((2 + 1 + 2+ 1) / 4) = 1/ 1{,}5 \quad = 0{,}67 \\
C_{clo}(E) &= 1 / ((2 + 1 + 1+ 1) / 4) = 1/ 1{,}25 = 0{,}8
\end{aligned}
\tag{3}
$$

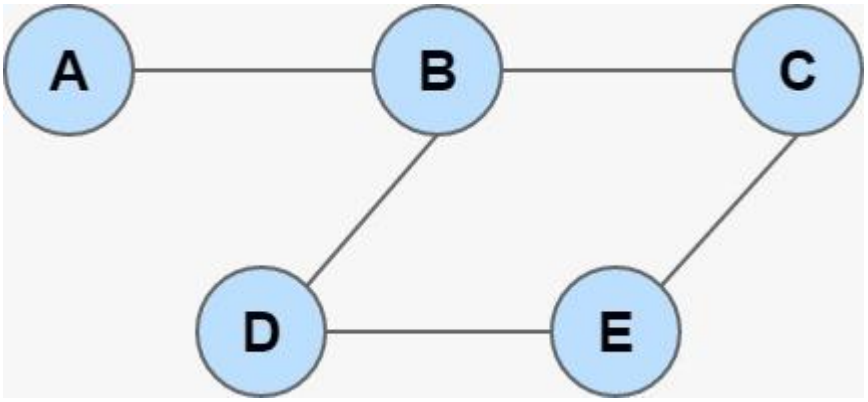
Hesaplamadaki sonuçlara göre en merkezi düğüm B düğümüdür. Yakınlık merkeziliğine göre E, C, D, A olmaktadır.

Zachary's Karate Club üyelerinin yakınlık merkeziliği analizi ile gösterimi Şekil 6'da verilmiştir.



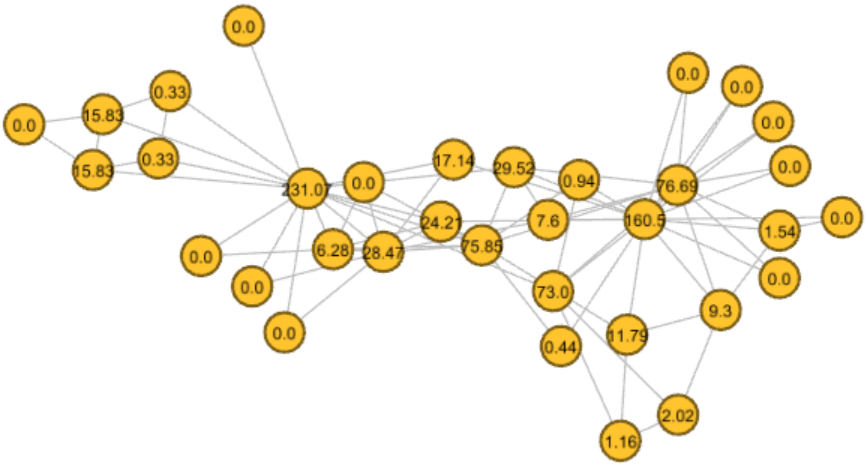
Şekil 6. Zachary's Karate Club'a ait Yakınlık Merkeziliği

Arasındalık merkeziliği, yakınlık merkeziliği gibi en kısa yol tabanlı bir merkeziliğidir [9]. Arasındalık merkeziliği, bir düğümün diğer düğümlere yakınlığıyla ilgili değil, diğer düğüm çiftleri arasındaki en kısa yollar üzerinde yer almasını temel almaktadır. Kısacası, en kısa yolların uzunluklarıyla değil, bunların sayısı temel alınmaktadır. Bazı kaynaklarda Arasındalık Merkeziliği terimi yerine Aradılık Merkeziliği diye de yer almaktadır.



Şekil 7. Arasındalık Merkeziliğine örnek Graf

Zachary's Karate Club ağındaki düğümler için arasındalık merkeziliği değerleri Şekil 8'de gösterilmiştir.



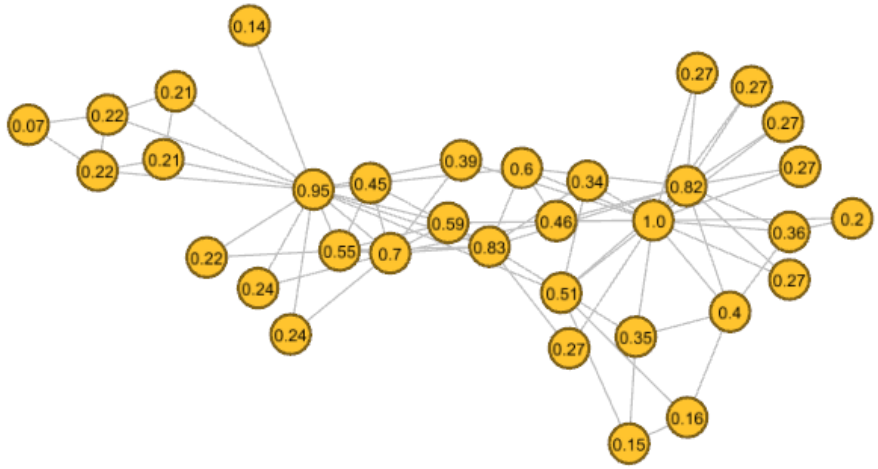
Şekil 8. Zachary's Karate Club'a ait Arasındalık Merkeziliği

Özvektör merkeziliği, derece merkeziliğinin gelişmiş bir diğer ölçütüdür. Özvektör merkezliğinde (eigenvector centrality) diğer ağ merkezliği ölçütlerinde olduğu gibi bir düğümün önemi sadece komşularının sayısına bağlı değildir. Özvektör merkezliğinde bir düğümün önemi komşularının sayısının yanı sıra komşularının da önemine bağlıdır. Ağ ölçütü bu özelliği ile insan toplulukları içerisinde çok sayıda arkadaş edinme değerini yakalar. Bu özelliği ile özvektör merkeziliği, geribildirim tabanlı

merkezilik ölçütleri kategorisindedir. Bazı kaynaklarda özvektör merkeziliği prestij (prestige) olarak da anılmaktadır.

Özvektör merkeziliği için komşuluk matrisi üzerinde özdeğer ve özvektör hesaplamaları yapılır. Özvektör merkeziğili ile benzerlik gösteren diğer geri bildirim tabanlı ölçütler Katz Merkeziliği (Katz Centrality), PageRank Merkeziliği (PageRank Centrality), Otorite ve Göbek Merkeziliği (Authority-Hub Centrality) ağ ölçütleridir [9].

Zachary's Karate Club ağındaki düğümler için özvektör merkeziliği değerleri Şekil 9'da gösterilmiştir.



Şekil 9. Zachary's Karate Club'a ait Özvektör Merkeziliği

3.3.2. Kümelenme Katsayısı Ölçütleri

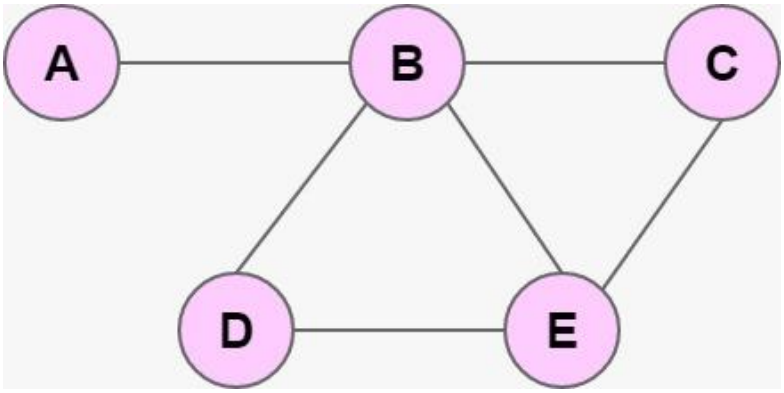
Geçişlilik sosyal ağlardaki bağ oluşturma davranışı üzerinde analiz yapan bir ölçüttür. Geçişlilik ölçütü için en az 3 kenar ve bu kenarların birleştirdiği 3 düğüm gereklidir. Bu kenar ve düğümler bir üçgen oluşturur. X, Y, Z bir ağda herhangi üç düğüm olsun, X-Y ve X-Z kenarları ağda mevcutken bir Y-Z kenarı da varsa bu üç düğüm arasında bir geçişlilik davranışı söz konusudur. Bu davranış "arkadaşımın arkadaşı benim de arkadaşımdır" anlamındadır [9].

Geçişlilik kavramıyla ilgili en önemli özellik ise kümelenme katsayısıdır. Kümelenme katsayısı, grafta ortak bir komşuya sahip düğümlerin kendilerinin de komşu olma olasılığı denilebilir.

Kümelenme katsayısı, düğümler için hesaplandığında yerel kümelenme katsayısı, bütün ağ için hesaplandığında global kümelenme kaysayısı denilmektedir. Kümelenme katsayısı [0,1] aralığında değer alır [9].

Yerel kümelenme katsayısı, geçişliliği ölçerken düğümü temel almaktadır. Yerel kümelenme katsayısını tanımlayacak olursak; bir v düğümünün komşularının kendi aralarında ne kadar sıkı bağlı olduğunun hesaplanmasıdır. Denklem 4'te yerel kümelenme katsayısı formülü yer almaktadır.

$$C_{cc}(v) = \frac{v'yi\ içeren\ üçgenlerin\ sayısı}{v'yi\ merkez\ alan\ bağlı\ üçlülerin\ sayısı} \tag{4}$$



Şekil 10. Yönsüz graf

Denklem 4'e göre, Şekil 9'daki B düğümünün kümelenme katsayısını hesaplamak istediğimizde ;

- ✓ B'nin A, C, D, E olmak üzere 4 tane komşusu vardır.
- ✓ B'nin birbirine bağlı komşu çiftleri 2 tanedir (CE ve DE)
- ✓ B'ye bağlı olup olmaması önemli olmadan var olan komşu çiftleri 6 tanedir. B'nin AC, AD, AE, CD, CE, DE bağlı olmayan komşu çiftleridir.

C_{cc} (B) = 2/6 = 0,33 olarak bulunur.

Yerel kümelenme katsayısındaki bakış açısı ağa uygulanabilir. Bazı kaynaklarda global kümelenme katsayısı yerine ağın global geçişliliği tanımı da yer almaktadır. Bir ağın global kümelenme katsayısı formülü Denklem 5'te verilmiştir [9].

$$C_{cc}(v) = \frac{ağdaki\ üçgenlerin\ sayısı\ x\ 3}{ağdaki\ bağlı\ üçlülerin\ sayısı} \quad (5)$$

Şekil 5'teki yönsüz ağ için global kümelenme katsayısını hesaplamak istediğimizde;

- ✓ Ağda 2 tane üçgen yer almaktadır. BCE, CEB, EBC, BDE, DEB, EBD
- ✓ Birbirine bağlı üçgensiz üçlüler 5 tanedir. ABC, ABD, ABE, DBC, DEC
- ✓ 3 tanesi diğer üçgenden: BDE, DEB, EBD
- ✓ 5 tanesi üçgensiz : ABC, ABD, ABE, DBC, DEC

Denklem 5'te verileri yerine koyup global kümelenme katsayısnı hesapladığımızda; C_{cc} (B) = (2x3) / 11 = 0,55 olarak bulunur.

BÖLÜM 4. TWITTER

Twitter, günümüzde en çok kullanılan sosyal medya platformlarından biridir. Twitter, Jack Dorsey, Noah Glass, Biz Stone, Evan Williams tarafından 21 Mart 2006 yılında kurulmuştur. 2008 yılında Barack Obama'nın kullanması ile üne kavuşmuştur. Twitter 2011 yılının Nisan ayında Türkçe dil desteği ile kullanılmaya başlanmıştır. Twitter, ilk etapta 140 karakteri desteklerken en son güncelleme ile 280 karakterle sınırlandırılmış olup, kullanıcılarının fikir ve görüşlerinin yayınlanmasına imkan sağlamaktadır. Kullanılar kendi profillerini oluşturabilmekte ve kendi yayın kitlesini belirleyebilmektedir.

Twitter, sosyal paylaşımlar için kullanıcıların bilgilerini paylaşma ve paylaşmama imkanı sağlamaktadır. Kilitli hesap, kullanıcıların sadece kendi belirledikleri arkadaş grupları ile paylaştığı hesaplardır. Açık hesap ise paylaşımların Twitter'da üye olan ve kullanıcıyı takip eden diğer bireyler arasında paylaşıldığı hesaplardır.

Twitter' da kullanıcıların paylaşımlarına tweet adı verilir. Tweetler üzerinden kullanıcı bazı durum etiketlerini paylaşarak özel konu belirleyebilmektedir. Tweetler üzerindeki konu etiketlerine hashtag adı verilir. Hashtag, içeriğin konusunu belirlemek için kullanılan özel kelimelerdir. Hashtagler insanlar üzerinde merak ve ilgi uyandırmaktadır. Sosyal medyada paylaşlan tweetlerin görünürlüğünü arttırır. Ayrıca hashtagler üzerinden geri dönüş sağlanmasını, insanların en çok hangi konuda paylaşım yapıldığı bilinir. Blog yazarları, bloğuna hashtag yardımıyla yönlendirme yapabilir, böylece ziyaretçi sayısı veya takipçileri artmış olacaktır. Hashtagler globalleşmeyi sağlamaktadır. Tweet paylaşılırken, paylaşımın herhangi bir kısmında diyez "#" kullanılması Twitter kullanıcısının hashtag eklediğini belirtmektedir. Büyük/küçük harf kullanılmasının önemi olmadan tek kelime ya da birden fazla kelimenin birlikte yazılması ile hashtag oluşturulur.

İnsanlar tarafından en fazla paylaşılan hashtag sözcüklerine trend topic (TT) adı verilmektedir. Trend topic, en çok kullanılan 10 konunun/hashtagin listesidir. Trend topic olan hashtagler insanların ilgisini diğer hashatglere göre daha çok çekmektedir. Twitter da TT olmuş hashtagler üzerinden gündemin takip edilmesi kolaylaşır.

Twitter kullanıcıları beğendikleri tweetlerin üstünde yer alan Retweetle butonu ile kendi sayfasında yayınlayabilmekte ve takipçileri ile paylaşabilmektedir. Retweet yapmak kısaca RT olarak adlandırılmıştır. Kullanıcılar önemli olduğunu düşündükleri konuları (kan bağışı, kayıp aranıyor vs.) retweet yaparak ve hashtag kullanarak yaygınlaşmasını sağlamaktadır. Şekil 11'de tweet ve hashtag yapısı için örnek verilmiştir.

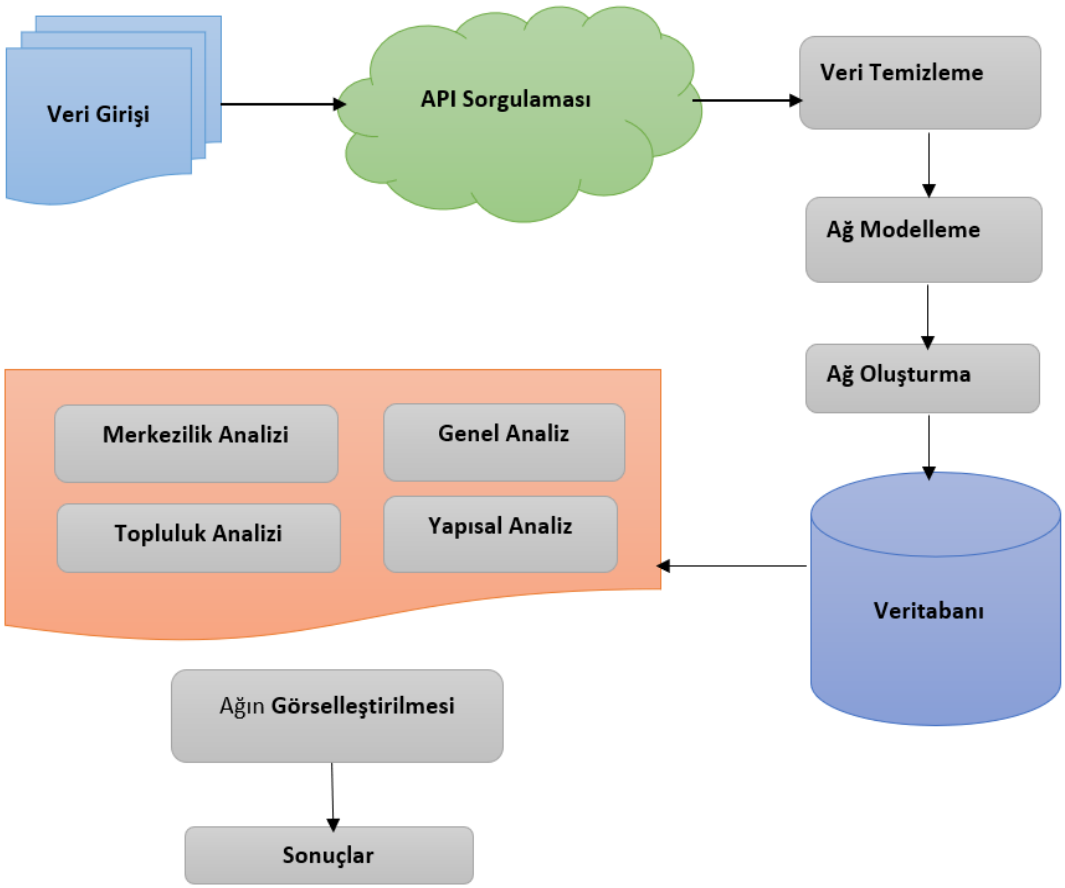


Şekil 11. Tweet Yapısı

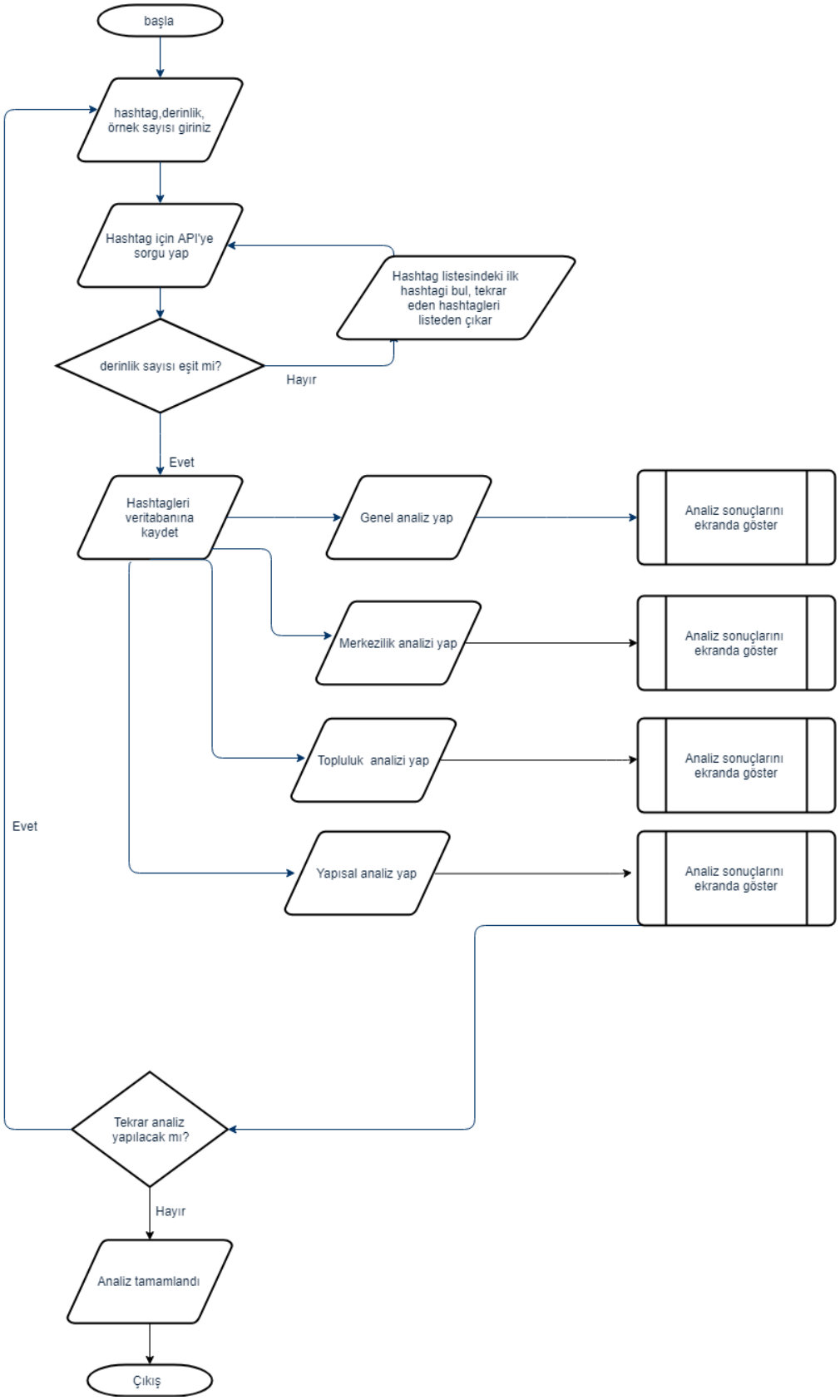
BÖLÜM 5. YÖNTEM

Bu bölümde, analiz yapılacak ağın elde edilmesi, elde edilen ağın analiz edilmesi için gerekli olan veri toplama, veri temizleme, ağın modellenmesi ve oluşturulması süreçleri anlatılmaktadır. Uygulanan yöntem için sistemin blok diyagramı Şekil 12'de, uygulamanın akış diyagramı Şekil 13'te verilmiştir.

Ağ analizinde, ilk olarak veri toplama amacıyla API sorgulaması yapılarak veri havuzu oluşturulur. Sorgu sonucunda elde edilen veri üzerinde analiz işleminde kullanılmayacak veriler temizlenir. Elde edilen veri seti kullanılarak ağ oluşturulur. Oluşturulan ağ modeli veritabanına kayıt edilir. Bu işlem genel analiz olarak adlandırılmaktdır. Veritabanındaki veriler ilgili araçlar ile merkezilik, topluluk, yapısal analizleri yapılır. Ağ oluşturulduktan sonra ağın görselleştirilmesi yapılır.



Şekil 12 . Sistemin Blok Diyagramı



Şekil 13 . Akış Diyagramı

5.1. Veri Toplama

Bu tez çalışmasında, varolan bir veri setini analiz etmek yerine veri seti sistem içerisinde oluşturulmuştur. Twitter [10] üzerinden elde edilen hashtagler veri setini oluşturmaktadır. Twitter'ın Standard Search API [11] kullanılarak kelime bazında sorgulanan tweetler, analiz edilecek veri seti oluşturmaktadır. Twitter Search API kullanmak için geliştirici hesabı oluşturulur. Bu hesap üzerinde tanımlanan uygulama kimlik bilgileri ve hesap bilgileri birlikte kullanılarak API üzerinde kimlik doğrulama ve yetkilendirme işlemleri gerçekleştirilir. Kimlik doğrulama ve yetkilendirme işlemi için uygulamada kullanılan kod parçacığı örnek Şekil 14'te verilmiştir.

```
1 reference
public void Authenticate(string consumerKey, string consumerSecret, string accessToken, string accessTokenSecret)
{
    // Get a AuthenticatedUser from a specific set of credentials
    var userCredentials = Auth.CreateCredentials(consumerKey, consumerSecret, accessToken, accessTokenSecret);

    // Use the user credentials in your application
    Auth.SetCredentials(userCredentials);

}
```

Şekil 14 . API üzerinde kimlik doğrulama

API aracığılıyla sorgulama yapmak için Şekil 15'teki method kullanılmaktadır.

```
1 reference
public List<ITweet> SearchTweeetsByHasTag(string hashTag, int resultCount)
{
    var searchParameter = new SearchTweetsParameters(hashTag)
    {
        Lang = LanguageFilter.Turkish,
        SearchType = SearchResultType.Recent,
        MaximumNumberOfResults = resultCount,
        Filters = TweetSearchFilters.Hashtags,
        TweetSearchType = TweetSearchType.OriginalTweetsOnly
    };

    return Search.SearchTweets(searchParameter).ToList<ITweet>();
}
```

Şekil 15 . Search API Tweet Sorgulama

5.2. Veri Temizleme

Twitter data kaynağı olarak oldukça zengindir. Kullanıcılara, atılan tweetler ile ilgili birçok meta veri sunmaktadır. Yayınlanma tarihi, içeriği, içerisindeki hashtagler, atılan tweetin tüm kullanıcılar tarafından görülüp görülmemesi, hesap sahibinin profil bilgileri gibi birçok bilgi API ile sunulmaktadır. Şekil 16'da Twitter API üzerinden gerçekleştirilen örnek sorgu sonucunun içeriği gösterilmiştir. Bu tez çalışmasında tweetlerin hashtagleri alınarak diğer tweet bilgileri temizlenmiştir.


```
"created_at": "Mon Feb 26 01:07:05 +0000 2018",
"id": 967928974960545793,
"id_str": "967928974960545793",
"text": "Congratulations to #Olympics athletes who won gold! Neutron stars like
  the one at the heart of the Crab Nebula may… https://t.co/vz4SnPupe2",
"truncated": true,
"entities": {
    "hashtags": [
        {
            "text": "Olympics",
            "indices": [
                19,
                28
            ]
        }
    ],
    "symbols": [],
    "user_mentions": [],
    "urls": [
        {
            "url": "https://t.co/vz4SnPupe2",
            "expanded_url": "https://twitter.com/i/web/status/967928974960545793",
            "display_url": "twitter.com/i/web/status/9…",
            "indices": [
                116,
                139
            ]
        }
    ]
},
"metadata": {
    "result_type": "popular",
    "iso_language_code": "en"
},
```


Şekil 16 . Twitter API Sorgu Sonucu

5.3. Kullanılan Teknolojiler

Bu tez çalışmasında yazılım dili olarak C#, platform olarak ASP.Net Core [12] tercih edilmiştir. Twitter'lerin işlenmesi ile elde edilen analiz sonuçlarının saklanması için veritabanı olarak MongoDB [13] kullanılmıştır. .Net Core ile Twitter API'sini kullanmak için güncel bir kütüphane olan Tweetinvi [14] kullanılmıştır. Twitter API yardımıyla elde edilen tweetlerin hashtagleri ayrıştırılarak, SNAP [15] kütüphanesinin sunduğu exeler yardımıyla analiz edilmiştir. Analizler sonucu elde edilen veriler D3.js [16] kütüphanesi ile görselleştirilmiştir. Uygulamanın ara yüzü Bootstrap [17] frameworkü ile oluşturulmuştur.

5.3.1. Asp .NetCore

Asp.Net Core web uygulamaları, servisler, kütüphaneler ve uygulama geliştirmek için open source modüler bir .NET platformudur. Esnek geliştirme olanağı ile ön plana çıkmaktadır. Microsoft tarafından Windows, MacOs ve Linux üzerinden çalıştırılabilir [18]. Yeni bir proje oluşturma, derleme ve dağıtma işlemleri cmd (Command-Line) üzerinden yapılabilmektedir. Asp .Net Core Xamarin, Mono, .NET Framework ve .Net Standard Library ile uyumlu çalışmaktadır. .Net Core açık kaynaklı olmasının yanı sıra .NET Foundation'ın projesi olması ile de ön plana çıkmaktadır.

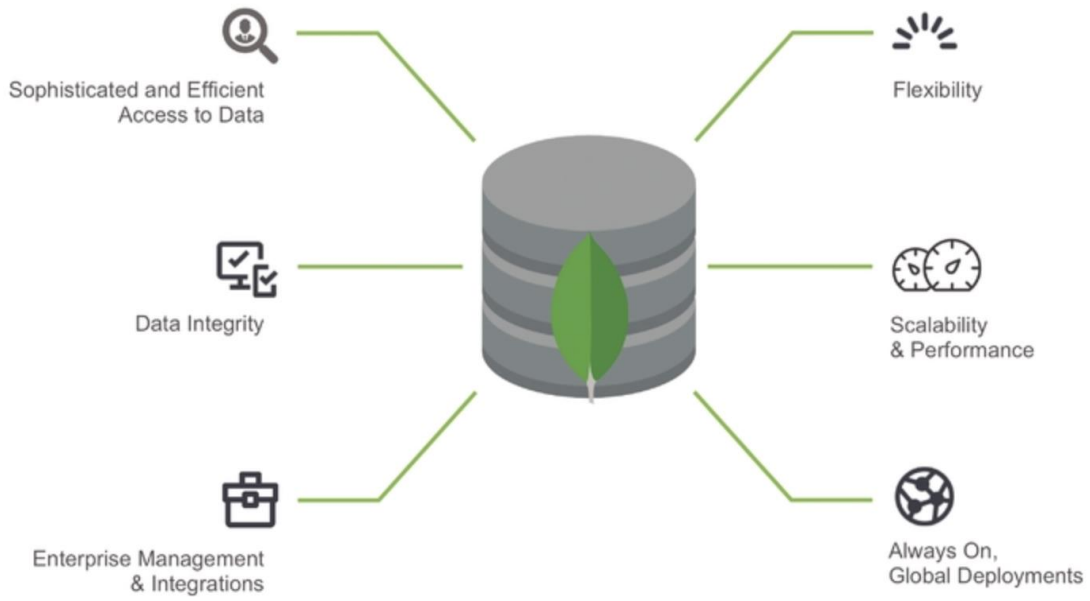
Bu tez çalışmasında, uygulamanın web tabanlı olması için tercih edilen platform Asp.Net Core platformudur.

5.3.2. MongoDb

MongoDb, NoSQL datalarını belge biçiminde saklayabilmek ve son kullanıcının anlık sorgulama yapabilmesi için ölçeklenebilir bir veritabanı biçimidir [19].

MongoDb kullanımında, veri boyutunun beklenenden fazla olması durumunda ya da performans açısından sıkıntı yaşanması halinde yeni makine eklenebilmektedir. Veriler JSON formatında ve belge biçminde tutulur. Verilerin JSON şeklinde tutulması veri yapısı değişse bile kaydetmeden değiştirme imkanı sunmaktadır. Veri kaybı yaşanmaz, verilerin birden fazla kopyası saklanabilir.

Bu tez çalışmasında, ağ görselleştirmesi için JSON veri formatı kullanılmaktadır. MongoDb verileri saklamak için, JSON veri formatını kullanır, bu nedenle tercih edilmiştir.



Şekil 17 . MongoDB Genel Yapısı

5.3.3. Tweetinvi

C# ile Twitter API'si kullanmak için hazırlanmış tweet atma, tweetleri okuma olmak üzere tüm beklentileri karşılayan en güncel kütüphane Tweetinvi ön plana çıkmaktadır.

5.3.4. SNAP

SNAP (Stanford Network Analysis Platform), Stanford Ağ Analizi Platformu olarak Türkçe'ye çevrilmektedir. C++ dilinde yazılmış, açık kaynaklı bir ağ analizi ve grafik madenciliği projesidir. Birçok düğüm ve milyonlarca kenar ile büyük ağları kolaylıkla ölçeklendirilebilir. Büyük veri setlerini verimli şekilde işler, düzenli ve rastgele grafikler üretir, düğümler ile kenarlarındaki öznitelikleri destekler ve yapısal özellikleri hesaplar [20].

Bu tez çalışmasında, SNAP projesi kaynak kodu derlenerek merkezilik, topluluk ve yapısal analiz işlemlerinden kullanılacak uygulamalar oluşturulmuştur. Uygulamalar sırası ile centrality.exe, community.exe ve netstat.exe olarak isimlendirmiştir. API aracılığı ile elde edildikten sonra işlenen veri seti, bu uygulamlar ile analiz edilmiştir.

Projenin güncel versiyonu olan 4.1 versiyonunu derlemek için Microsoft Visual C++ Compiler 2015 uygulamasına ihtiyaç duyulmaktadır. Bu uygulamanın eksik olması durumunda projenin derlenmesi sırasında derleyici eksikliği hatası ile karşılaşılmaktadır. Hatanın düzeltilmesi için Visual Studio 2017 sürümü güncellenmelidir.

SNAP projesi grafik çizimleri için Gnuplot [21] adındaki komut satırı temelli grafik çizim uygulamasına ihtiyaç duymaktadır. SNAP projesi içisindeki yapısal analiz için kullanılan netstat uygulaması, sistemde gnupilot uygulamasının eksik olamsı durumunda çıktı üretmeyecektir. Bu nedenle uygulamanın kullanıldığı sistem üzerinde gnuplot yüklü olmalıdır.

Sistem üzerinde genel analiz gerçekleştirildikten sonra hashtag düğümleri numaralandırılır ve text dosyasına yazılır. Bu dosya SNAP projesi ile üretilen uygulamalara girdi olarak verilir.

centrality**,** bu uygulama grafiğin düğümleri için merkezilik ölçülerini hesaplar. Parametre olarak düğüm ilişkilerinin yazılı olduğu text dosyası ve çıktı olarak üretilecek dosya ismini alır.

community**,** bu uygulama ağ üzerinde topluluk analizi yapmayı sağlar.Parametre olarak düğüm ilişkilerinin yazılı olduğu text dosyası, çıktı olarak üretilecek dosya ismini ve hesaplama algoritmasını alır. Bu uygulama için algoritma olarak Clauset-Newman-Moore algoritması kullanılmıştır.

netstat**,** bu uygulama ağ istatistik verilerinin grafik olarak oluşturulmasını sağlar. Parametre olarak düğüm ilişkilerinin yazılı olduğu text dosyası, çıktı olarak üretilecek dosya ismini, girdi verilerinin yönlü grafa ait olup olmadığına dair anahtar ve hangi analizlerin gerçekleştirileceğini alır. Bu uygulama için tüm analizler kullanılmıştır.

5.3.5. D3.js

D3.js, web standartlarına uygun, veri görselleştirme için geliştirlen bir Javascript kütüphanesidir. Dataların HTML, SVG ve CSS kullanarak görselleştirilmesine yardımcı olur. Tüm tarayıcılar üzerinde DOM manipülasyonuna dayalı yaklaşımı yönüyle güçlü bir görselleştirme kütüphanesidir.

Verilerin grafiklere dökülmesi için birçok Javascript kütüphanesi vardır. Bu kütüphanelerin en yeteneklisi D3.js'tir. Dökümanlardaki dataları otomatik hesaplamayı, orantılar yapmayı, görselleştirilmesi ve interaktivite gibi birçok avantajı sağlamaktadır.

5.3.6. Bootstrap

Bootstrap, açık kaynak kodlu ücretsiz bir önyüz çatısıdır. Bootstrap kütüphanesi ile tüm uygulama ve tarayıcılarda çalışması uygun tasarımlar yapılabilmektedir [22].

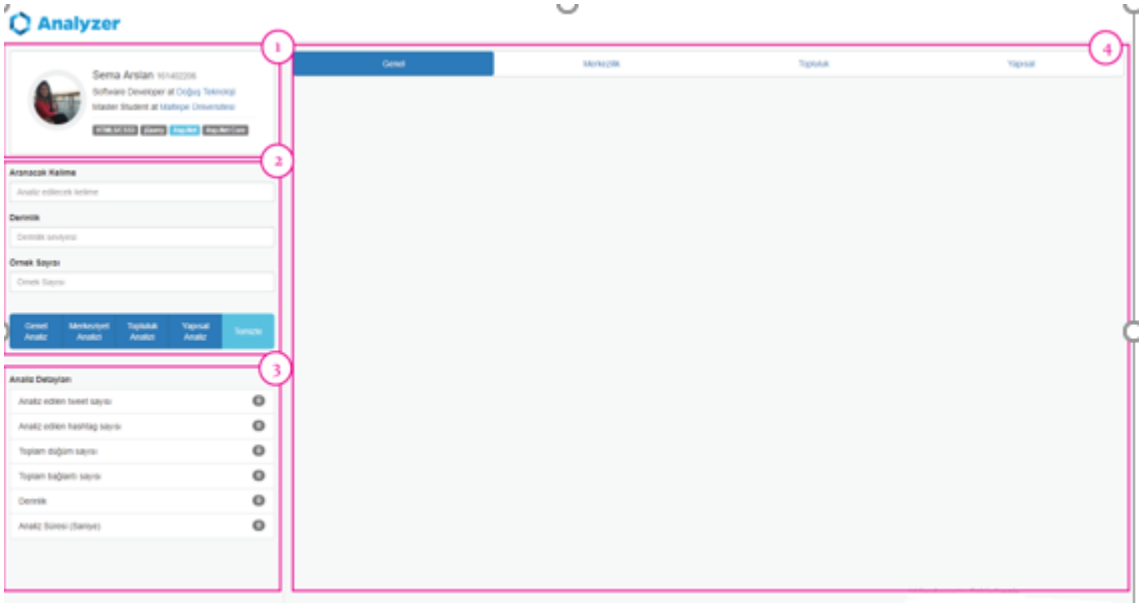
Bootstrap, uygulamalarda kullanıma hazırdır. Dünyanın en popüler mobil uyumlu önyüz çatısıdır. Farklı ekran boyutlarına uyum sağlayabilen adaptif yapısı ile ön plana çıkmaktadır. Bootstrap içerisindeki hazır tablolar, butonlar, açılan menüler, navigasyon bar ve daha fazla özelliği ile web sitesi yapımında tercih edilmektedir. Bu hazır bileşenlere Bootstrap içerisindeki CSS kodları ile istenilen tüm değişiklikleri uygulamak mümkündür. Bootstrap'i diğer frameworklerden ayıran temel özelliği grid yapısıdır.

5.4. Geliştirilen Sistemin Uygulanması

Yapılan tez çalışması sonucunda, Twitter platformundaki paylaşımlar üzerinden bağlantılı konu analizi yapılmasını sağlayan bir uygulama geliştirilmiştir.

Sisteme Genel Bakış

Kullanıcı arayüzüne genel bakış;



Şekil 18 . Kullanıcı Arayüzü

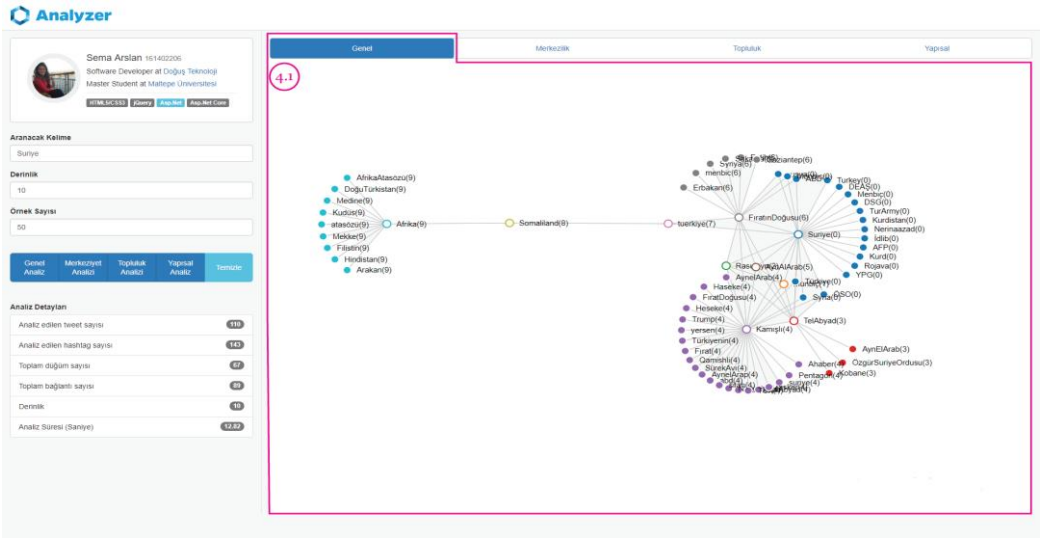
1 Uygulamaya giriş yapan kullanıcı adı bilgilerinin gösterildiği alandır. Ayrıca, uygulamanın geliştirildiği teknolojiler belirtilmiştir.

2 Analyzer uygulaması 3 parametre ile çalışmaktadır. Giriş paneli; aranacak kelime, derinlik ve örnek sayısı olmak üzere 3 alandan oluşmaktadır. Aranacak kelime alanına, analiz edilmesi gereken hashtag girilmektedir. Derinlik alanı, analiz yapılacak olan hashtag parametresi için derinliği belirlemektedir. Son olarak örnek sayısı alanı, Twitter'dan çekilecek olan tweet sayısını belirlenmektedir. Giriş panelinde 5 buton yer almaktadır. Genel analiz, merkezcilik analizi, topluluk analizi, yapısal analiz ve temizle butonlarıdır. Analiz butonları, isimlendirmeleri yönünde amacına yönelik analizler yapılmaktadır. Temizle butonu yardımıyla, giriş paneline girilen parametreler ve analiz sonuçları silinmektedir.

3 Giriş paneline girilen parametreler ile Twitter API'sine sorgulama yapılır. 3. bölüm analiz sonuçlarını göstermektedir. Analiz sonuçları; analiz edilen tweet sayısı, analiz edilen hashtag sayısı, toplam düğüm sayısı, toplan bağlantı sayısı, derinlik ve analiz edilen süre değerlerini içermektedir.

4 Analiz paneli olarak isimlendirilmiştir. Kullanıcılar,ilgili parametreleri girdikten sonra dört farklı analiz yapılır.Analiz sonuçlarına göre, hashtagler ile grafik çiziminin gösterildiği bölümdür.

4.1 Analiz yapmak için hashtag olarak "Suriye" parametresi ve derinlik olarak 10 değeri girilmiştir. Genel analiz tabına tıklandığında, "Suriye" hashtagi ile bağlantılı hashtagler ile grafik çizilmektedir. Genel analiz tabında, hashtagler tekrarsız olarak görüntülenmektedir.



Şekil 19 . Genel Analiz

4.2 Merkezilik analizi sonuçlarının gösterildiği alandır. Girilen hashtag ile ilgili Tweet API'den dönen bağlantılı hashtaglerin derece merkezilik analizine göre ağı gösterilir.



Şekil 20 . Derece Merkezilik Analizi Ağ Görünümü

4.2.1 Merkezilik analizi grafiğinin gösterildiği butondur. Merkezilik analizi ile ilgili birden çok değer elde edilmektedir. Merkezilik tabında kalındığında, tablo görünümüne geçiş yapılabilir. 4.2.1 butonu ile istenildiğinde tekrar grafik görünümüne geçilebilir.



Şekil 21 . Merkezilik Analizi Tablo Görünümü

4.2.2 Girilen hashtag ile ilgili bağlantılı hashtagler tablo ile gösterilmektedir. Butona tıklandığında, ağdaki diğer düğümler ile farklı analiz değerleri elde edilmektedir. Butona tıklandığında ağdaki düğümler için derece, yakınlık, arasındalık, özvektör, ağ kısıtı, kümelenme, pagerank, göbek, otorite analiz sonuç değerleri elde edilir.

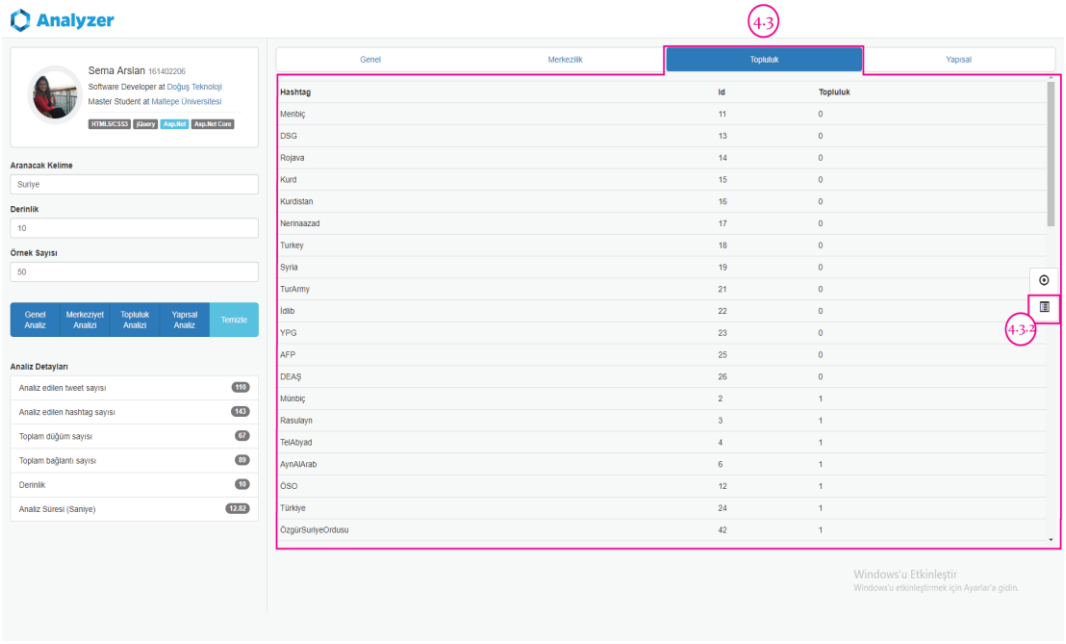
4.3 Topluluk analizi sonucunun gösterildiği bölümdür. Hashtag ile bağlantılı düğümlerin topluluk analizi sonucunda tablo ve ağ görünümünün elde edildiği bölümdür. Topluluk analizi ile ilgili analiz sonuçları yer almaktadır.

4.3.1 Topluluk analizi ile ilgili grafik sonuçlarının gösterildiği bölümdür. Girilen hashtag ile bağlantılı düğümler, topluluk analizine göre analiz edilmiştir.



Şekil 22. Topluluk Analizi Ağ Görünümü

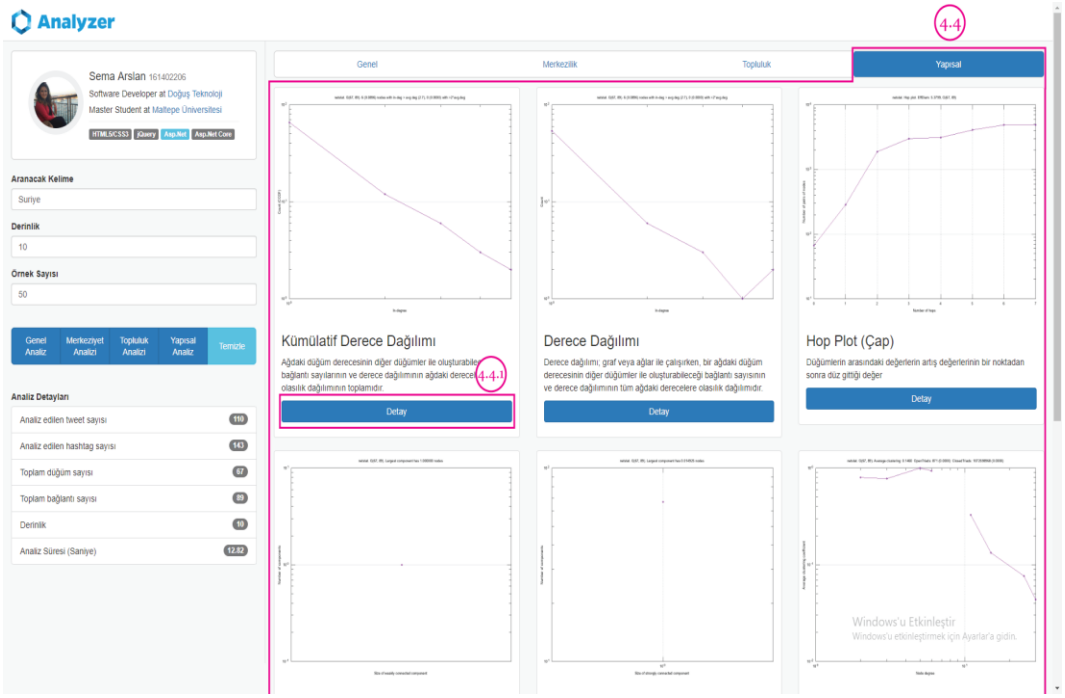
4.3.2 Topluluk analizi sonucunda ağdaki düğümlerin topluluk analizinin tablo görünümü elde edilmektedir.



Şekil 23.Topluluk Analizi Tablo Görünümü

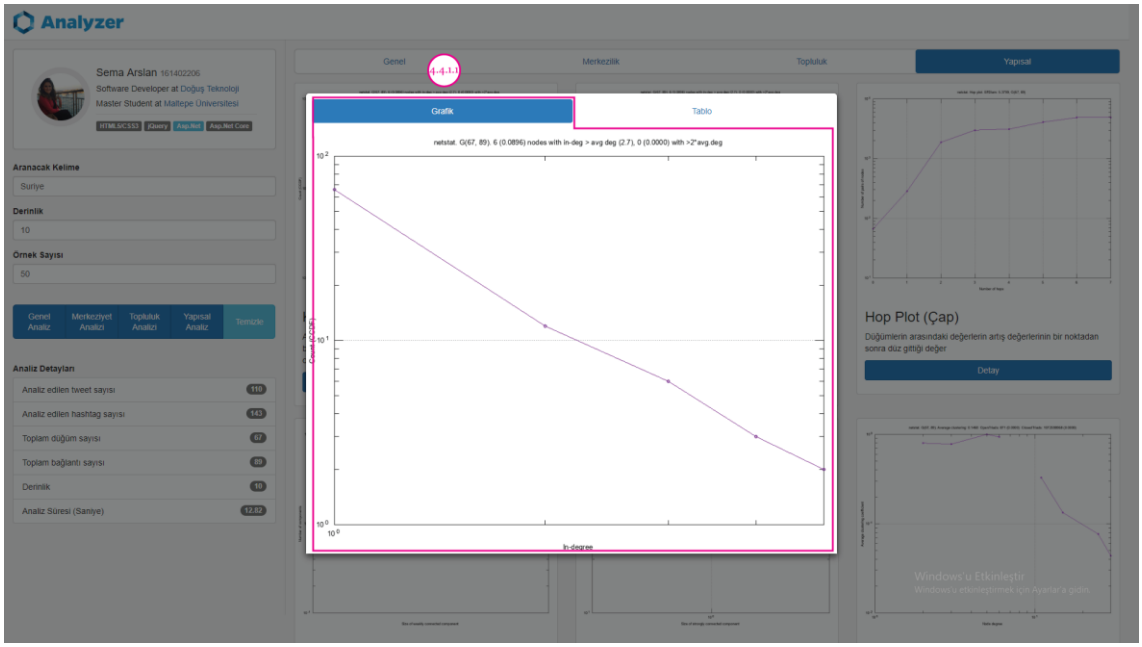
4.4 Analiz edilmesi için girilen hashtag ile ilgili bağlantılı hashtaglerin yapısal analiz sonuçlarının gösterildiği tabdır. Yapısal analiz sonucu, kümülatif derece dağılımı, derece dağılımı, hotplot çapı, sonucu elde edilmektedir.

4.4.1 Yapısal analiz sonucunda ilk elde edilen sonuç kümülatif derece dağılımıdır. 4.4.1 butonunun detayına tıklandığında, grafik ve tablo görünümü elde edilmektedir.



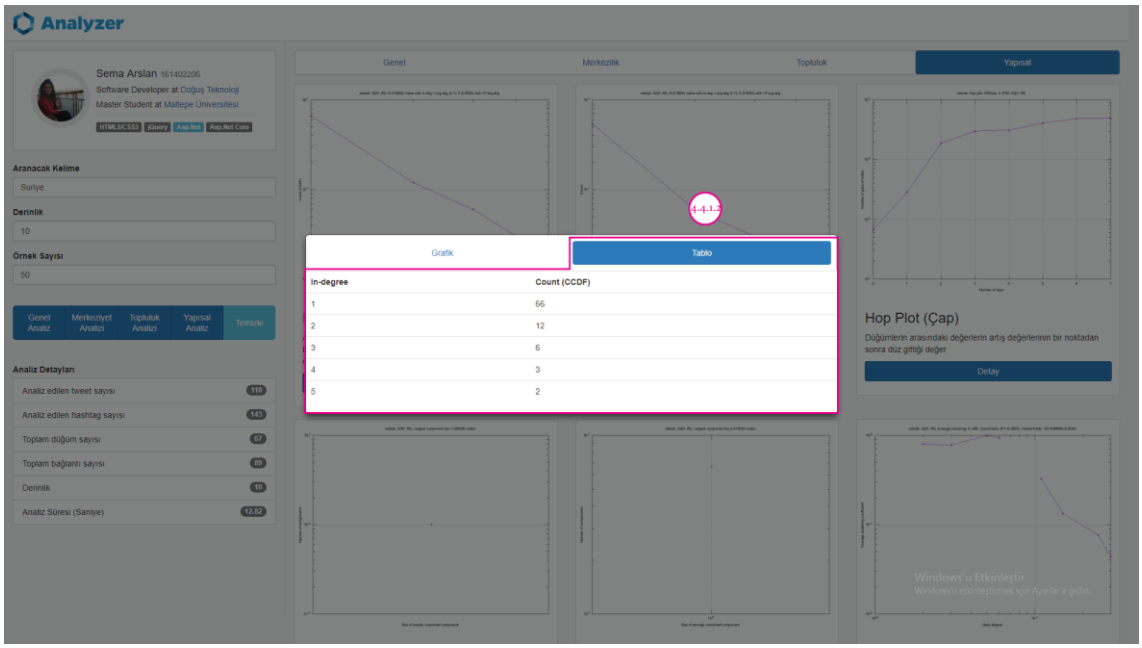
Şekil 24. Yapısal Analiz Görünümü

4.4.1.1 Detay butonu, ilgili analiz sonucunun grafik detayı verilmiştir. Detay butonu ile, elde edilmek istenen değerlerin grafik sonuçları gözlemlenir. Şekil 25'te kümülatif derece dağılımının grafik görünümü yer almaktadır.



Şekil 25.Yapısal Analiz Grafik Detay Görünümü

4.4.1.2 Detay butonu, ilgili analiz sonucunun tablo görünümü verilmiştir. Detay butonu ile, elde edilmek istenen değerlerin tablo sonuçları gözlemlenir. Şekil 26'da kümülatif derece dağılımının tablo görünümü yer almaktadır.



Şekil 26 .Yapısal Analiz Tablo Detay Görünümü

BÖLÜM 6. SONUÇ

Sosyal medya üzerinden yapılan paylaşımların veri madenciliği yöntemleriyle analiz edilmesi, büyük anlamsız veri kalıpları içerisindeki saklı bilgilerin insanlar için anlamlı hale gelmesini sağlar. Bu veriler üzerinden önemli çıkarımlar yapmak mümkündür. Paylaşımların metin içeriğinin anlamsal olarak analizinin yanı sıra, paylaşımlardan oluşturulan ağ yapıları da önemli bilgiler içermektedir.

Bu tez çalışmasında, Twitter paylaşımlarının toplanması ve paylaşımlardaki etiket birlikteliklerinin ağ modeli oluşturulmuştur. Oluşan ağ modeli, sosyal ağ analizi teknikleriyle analiz edilerek, sonuçları etkin şekilde sunan yazılım aracı geliştirilmiştir. Geliştirilen yazılım aracı, sabit bir veri seti değil anlık Twitter platformundan veri setlerini alarak güncel veriler üzerinden işleme yapmaktadır. Etiketlerin analiz anındaki ilişkilerine erişebilmesi bu tez çalışmasını önemli hale getirmektedir.

Bu uygulamada kullanılan veri setlerinden, karmaşık ağ analizi yöntemleri kullanılarak anlamlı sonuçlar elde edilmiştir. Uygulama tarafından sosyal ağı oluşturulan etiketler; merkezilik, topluluk ve yapısal ağ analizi yöntemlerine göre incelenerek ağın en güçlü aktörleri ortaya çıkarılmıştır. Elde edilen aktörler, ağın incelenmesi ve doğru yorumlanmasında önemlidir.

Uygulamada "seçim" etiketi için 5 derinlikli 50 tweet örneği ile gerçekleştirilen analiz sonucunda, seçim etiketi ile bağlantılı olan etiketler merkezilik ölçütü sırası ile, ekonomi, kitap, yazar, şiir olarak tespit edilmiştir. Elde edilen verilere bakıldığında, bu ağ için en değerli aktörün ekonomi olduğu görülmektedir. Bu örnek için, seçim kazanmak isteyen bir siyasi partinin öncelikle ekonomi alanında gelişmeler yapması sonucu ortaya çıkmaktadır. Dolayısıyla, bu tez çalışmasında ortaya konulan uygulama kullanılarak seçmenin anlık eğilimleri belirlenebilir ve seçim stratejileri geliştirilebilir.

Bu tez çalışması sonucunda, uygulama aracılığıyla elde edilen analiz sonuçları birçok farklı alanda ve işlevde kullanılmak üzere fikir vermektedir. İşletmelere ekonomi, pazarlama, finans, sanayi başta olmak üzere sektöre özel, yeni pazar vizyonu katmaktadır. İncelenen konu etiketinin hangi etiketlerle bağlantılı olduğunun ortaya çıkması bilgisi, yeni pazar arayışındaki ticari işletmelere kullanıcı eğilimleri hakkında yararlı bilgiler sağlamaktadır.

KAYNAKÇA

[1] A. Köktürk, "Sosyal Ağ Kullanım Performansına Göre Üniversite Değerlemesi," Yayımlanmış Yüksek Lisans Tezi, Ankara: Başkent Üniversitesi FBE, 2013.

[2] E. Beğenilmiş, "Detection Of Organized Behaviours On Twitter," İstanbul, 2017.

[3] Y. E. Karabulut ve E. U. Küçüksille, "Twitter Profesyonel İzleme ve Analiz Aracı,"Teknik Bilimleri Dergisi, cilt 8, no. 2, pp. 17-24, 2018.

[4] M. Hao, C. Rohrdantz, H. Janetzko, U. Dayal, D. A. Keim, L.-E. Haug ve M.-C. Hsu, "Visual Sentiment Analysis On Twitter Data Streams," IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 275-276, 2011.

[5] H. Wang, D. Can, A. Kazemzadeh, F. Bar ve S. Narayanan, "A System for Real-time Twitter Sentiment Analysis of 2012 US Presidential Election Cycle," pp. 115-120, 2012.

[6] A. Bruns ve S. Stieglitz, "Towards More Systematic Twitter Analysis: Metrics for Tweeting Activities," International Journal of Social Research Methodology, cilt 2, no. 16, pp. 91-108, Mart 2013.

[7] F. Mete, İ. Ceylan, "Türkçe Kurallı Cümle Yapısının Graf Teori ile Gösterilmesi," International Periodical for the Languages, Literature and History of Turkish or Turkic, vol 10/16, pp. 953-968, 2015.

[8] W. W. Zachary, "An Information Flow Model for Conflict and Fission in Small Groups," Journal of Anthropological Research, vol. 33, pp. 452-473, 1977.

[9] V. Tunalı, Sosyal Ağ Analizine Giriş, Ankara: Nobel Akademik Yayıncılık, 2016.

[10] Twitter, "Twitter Developer Platform," https://developer.twitter.com/, 2019.

[11] Twitter,"Twitter API Document", https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets", 2019.

[12] Microsoft, "Microsoft .Net Core," https://docs.microsoft.com/en-us/dotnet/core/tutorials/, 2019.

[13] MongoDB, "MongoDB," https://docs.mongodb.com/, 2019.

[14] NuGet, "TweetinviAPI,". https://www.nuget.org/packages/TweetinviAPI/, 2019.

[15] Stanford University, "Download SNAP," http://snap.stanford.edu/snap/releases.html, 2018.

[16] M. Bostock, "D3.js," https://d3js.org/, 2018.

[17] Bootstrap, "Get Started," https://getbootstrap.com/docs/4.1/getting-started/introduction/. 2018.

[18] . Net Core, ".Net Core", https://dotnet.microsoft.com/, 2018.

[19] MongoDb, "MongoDb", https://docs.mongodb.com/manual/, 2018..

[20] SNAP, "SNAP ", http://snap.stanford.edu/, Stanford University, 2018.

[21] Gnuplot, "Gnuplot", http://www.gnuplot.info/, 2018

[22] Bootstrap, "Browsers Devices,": http://getbootstrap.com/docs/4.1/getting-started/browsers-devices/, 2018.