

SOSYAL AĞLARDA ETİKET AĞI ANALİZİ

2016
YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ

Eyüb Ekmel SULAK

SOSYAL AĞLARDA ETİKET AĞI ANALİZİ

Eyüb Ekmel SULAK

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

KARABÜK

Haziran 2016

Eyüb Ekmel SULAK tarafından hazırlanan “SOSYAL AĞLARDA ETİKET AĞI ANALİZİ” başlıklı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

Yrd. Doç. Dr. İlker TÜRKER

Tez Danışmanı, Bilgisayar Mühendisliği Anabilim Dalı



Bu çalışma, jürimiz tarafından oy birliği ile Bilgisayar Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir. 14/ 06/ 2016

Ünvanı, Adı SOYADI (Kurumu)

İmzası

Başkan : Yrd. Doç. Dr. Burhan SELÇUK (KBÜ)



Üye : Yrd. Doç. Dr. İlker TÜRKER (KBÜ)



Üye : Doç. Dr. Ergin YILMAZ (BEÜN)



...../...../2016

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

Prof. Dr. Nevin AYTEMİZ

Fen Bilimleri Enstitüsü Müdürü



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

Eyüb Ekmel SULAK

ÖZET

Yüksek Lisans Tezi

SOSYAL AĞLARDA ETİKET AĞI ANALİZİ

Eyüb Ekmel SULAK

Karabük Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:

Yrd. Doç. Dr. İlker TÜRKER

Haziran 2016, 116 sayfa

Tarih boyunca insanlar birbirleriyle iletişim halinde olmuşlardır. Bu iletişim süreci, zaman içinde insanlığın algısı ve devamında teknolojinin gelişimi ile eşzamanlı olarak değişim göstermiştir. Bugün internet tabanlı sosyal ağlar ile insanlar iletişimin en üst noktasına çıkmışlardır. Bu iletişim direk olabildiği gibi, bireylerin yazdıkları ve paylaştıkları materyallerin birbirleri ile olan benzerlikleri ile dolaylı olarak da sağlanmaktadır. Dolaylı iletişimi sağlayan bu veri hazinesinin içinde yapısal ve anlamsal olarak birliktelikler bulunmaktadır. Bu birliktelikler farklı yöntemler ile analiz edilerek, anlamlı sonuçlara ulaşmak amaçlanmaktadır.

Bu çalışmada sosyal ağ verileri kullanılarak etiket ağı analizi yapılması ve sonuçların görselleştirilmesi amaçlanmıştır. Etiket ağı analizi çalışması, mikroblog sitesi özelliğindeki Twitter sosyal ağına ait veriler ile yapılmıştır. Twitter API kullanılarak elde edilen veriler belli kriterlere göre ayrıştırıldıktan sonra birlikte kullanım ve semantik yakınlık kriterlerine göre kompleks ağlara dönüştürülmüştür. Bu iki farklı

kompleks ađ analiz edilerek sonuçları karşılaştırılmış ve görselleştirilmiştir. Bu analiz sonucunda, etiketlerin birlikte kullanımının anlamsal olarak doğruluđu sorgulanmıştır.

Her iki kompleks ađ Gephi yazılımı ile görselleştirilmiş ve temel analizleri yapılmıştır. Ayrıca Gephi çıktıları MATLAB ile işlenerek daha detaylı analizler grafikler halinde sunulmuştur. Böylece Twitter etiketlerinin birlikte kullanım ve anlamsal yakınlık ađları detaylı olarak karşılaştırılmıştır.

Anahtar Sözcükler : Etiket ađı analizi, semantik ađ, kompleks ađ, Twitter.

Bilim Kodu : 902.2.042

ABSTRACT

M. Sc. Thesis

TAG NETWORK ANALYSIS IN SOCIAL NETWORK

Eyüb Ekmel SULAK

Karabük University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

Thesis Advisor:

Assist. Prof. Dr. İlker TÜRKER

June 2016, 116 pages

Throughout history mankind have been in contact with each other. This communication process has changed simultaneously with the perception of humanity and the development of technology in time. Today, people are out on the top spot of communication by the use of internet-based social networks. This communication as well as being direct, indirect materials that individuals are provided with what they write and share similarities with each other. Resulting in the treasury of this data, there are structural and semantic togetherness and various methods are used for this togetherness are analyzed to arrive at meaningful results.

The aim of this study, making the hashtag network analysis using social network data and visualization of results. Hashtag network analyze study was conducted using data from Twitter which is a micro-blog. The data obtained using the Twitter API, disaggregated by certain criteria, is transformed into complex Networks according to the collocation and semantic similarity criteria. These both complex network

analyzed and after that compared the results and visualised. As a consequence of this analysis, questioned semantic precision of collocationof hashtags.

Gephi software was used for visualization and basic analyzation of both networks. Results of GEPHI analyze, analized detailed with MATLAB and shown in graphical format. Thus, collocation and semantic similarity networks of Twitter Hashtags are compared in depth.

Key Word : Analyze of hashtag network, semantic network, complex network, Twitter.

Science Code : 902.2.042

TEŐEKKÜR

Bu tez alıőmasının planlanmasında, araőtırılmasında, yürütölmesinde ve oluşumunda ilgi ve desteęini esirgemeyen, engin bilgi ve tecrübelerinden yararlandıęım, yönlendirme ve bilgilendirmeleriyle alıőmamı bilimsel temeller ışığında őekillendiren sayın hocam Yrd. Do. Dr. İlker TÜRKER'e sonsuz teőekkürlerimi sunarım.

Tez alıőmam ve yazımında yardımlarını ve desteęini esirgemeyen arkadaşlarım Öğr. Gör. Hakan YILMAZ'a ve Arő. Gör. Sait DEMİR'e teőekkür ederim.

Sevgili aileme manevi hiçbir yardımını esirgemedен yanımda oldukları için tüm kalbimle teőekkür ederim.

Tez alıőmasında kullandıęım 140dev yapısını ücretsiz paylaőtıęı ve verdięi hızlı destek için Adam Green'e teőekkür ederim.

İÇİNDEKİLER

| | <u>Sayfa</u> |
|--|--------------|
| KABUL..... | ii |
| ÖZET..... | iv |
| ABSTRACT..... | vi |
| TEŞEKKÜR..... | viii |
| İÇİNDEKİLER | ix |
| ŞEKİLLER DİZİNİ..... | xiii |
| ÇİZELGELER DİZİNİ | xvi |
| SİMGELER VE KISALTMALAR DİZİNİ | xvii |
| | |
| BÖLÜM 1. | 1 |
| GİRİŞ | 1 |
| 1.1. KOMPLEKS AĞ TANIMI | 1 |
| 1.2. GRAF (ÇİZGE) YAPILARI | 2 |
| 1.3. KARMAŞIKLIK (COMPLEXITY)..... | 3 |
| 1.4. KOMPLEKS AĞ YAPILARI | 4 |
| 1.4.1. Ağ Parametreleri..... | 4 |
| 1.4.1.1. Düğüm Dereceleri..... | 5 |
| 1.4.1.2. Derece Dağılımı | 6 |
| 1.4.1.3. Kümelenme (Clustering)..... | 10 |
| 1.4.1.4. En Kısa Yol..... | 11 |
| 1.4.1.5. Çap (Diameter)..... | 12 |
| 1.4.1.6. Modülerite (Modularity) | 12 |
| 1.4.2. Gerçel Ağlar..... | 13 |
| 1.4.2.1. Teknolojik Ağlar | 14 |
| 1.4.2.2. Sosyal Ağlar | 21 |
| 1.4.2.3. Biyolojik Ağlar | 26 |
| 1.4.2.4. Finansal Ağlar | 26 |
| 1.4.2.5. Semantik Ağlar | 27 |
| 1.4.2.6. Diğer Gerçel Ağlar..... | 27 |

| | <u>Sayfa</u> |
|--|---------------------|
| 1.4.3. Sentetik (Model) Ağlar | 28 |
| 1.4.3.1. Rassal (Random) Ağlar | 29 |
| 1.4.3.2. Küçük-Dünya (Small-World) Ağları | 30 |
| 1.4.3.3. Ölçek Bağımsız (Scale-Free) Ağlar | 32 |
| | |
| BÖLÜM 2. | 34 |
| VERİ KAYNAĞI OLARAK SOSYAL AĞLAR..... | 34 |
| 2.1. YÖNTEMLER | 36 |
| 2.1.1. Twitter API..... | 36 |
| 2.1.1.1. Search API | 36 |
| 2.1.1.2. REST API | 37 |
| 2.1.1.3. Streaming API..... | 37 |
| 2.1.2. 140dev Streaming API Yapısı | 38 |
| 2.1.3. Phirehose Kütüphanesi | 39 |
| | |
| BÖLÜM 3. | 41 |
| SEMANTİK | 41 |
| 3.1. BİLGİ KAVRAMI | 41 |
| 3.2. SEMANTİK KAVRAMI | 42 |
| 3.3. SEMANTİK AĞLAR..... | 43 |
| 3.4. SEMANTİK WEB | 47 |
| 3.5. SEMANTİK YAKINLIK HESAPLAMA | 50 |
| 3.5.1. TF (Terim Frekansı) | 50 |
| 3.5.2. IDF (Ters Doküman Frekansı) | 50 |
| 3.5.3. TF-IDF..... | 51 |
| 3.5.4. L2 Normalizasyonu | 51 |
| 3.5.5. Cosine Yakınlığı Hesaplama | 52 |
| | |
| BÖLÜM 4. | 53 |
| UYGULAMA | 53 |
| 4.1. VERİ SETİNİN OLUŞTURULMASI | 53 |
| 4.1.1. Filtreleme..... | 53 |

| | <u>Sayfa</u> |
|--|---------------------|
| 4.1.1.1. Dil Seçimi | 53 |
| 4.1.1.2. Otomatik Tweet Mesajları Engelleme | 53 |
| 4.1.1.3. ReTweet Engelleme | 54 |
| 4.1.2. Ön İşleme..... | 54 |
| 4.1.2.1. URL'lerin Çıkarılması | 54 |
| 4.1.2.2. Karakter Tekrar Sayılarının 2'ye Düşürülmesi..... | 55 |
| 4.1.2.3. Altçizgilerin Kaldırılması | 55 |
| 4.1.2.4. User Mention'ların Temizlenmesi | 55 |
| 4.1.2.5. Smiley'lerin Temizlenmesi..... | 56 |
| 4.1.2.6. Noktalama Ve İmla İşaretlerinin Temizlenmesi..... | 56 |
| 4.1.2.7. Karakter Sayısı 20'den Fazla Olan Kelimelerin Silinmesi..... | 56 |
| 4.1.2.8. Düzensiz Boşlukların Teke Düşürülmesi..... | 56 |
| 4.1.2.9. Küçük Harfe Dönüştürme | 57 |
| 4.1.2.10. Türkçe Karakterlerin Dönüşümü | 57 |
| 4.1.2.11. Tek Karakterlerin Temizlenmesi | 57 |
| 4.1.2.12. Yabancı Alfabelerin Karakterlerinin Temizlenmesi..... | 58 |
| 4.2. AĞLARIN OLUŞTURULMASI | 58 |
| 4.2.1. Kullanım Birlikteliklerine Göre Hashtag Ağı | 58 |
| 4.2.1.1. TekTweet'te Geçen Tag'lerin Silinmesi..... | 58 |
| 4.2.1.2. Nodes Tablosunun Oluşturulması..... | 59 |
| 4.2.1.3. Edges Tablosunun Oluşturulması | 59 |
| 4.2.2. Semantik Yakınlığa Göre Hashtag Ağı | 61 |
| 4.2.2.1. Kelime Setinin Elde Edilmesi..... | 61 |
| 4.2.2.2. Stopword'lerin Çıkarılması..... | 61 |
| 4.2.2.3. 3'ten Az, 12'den Fazla Karakterden Oluşan Kelimelerin Çıkarılması..... | 62 |
| 4.2.2.4. Kelime Setinin Oluşturulması..... | 63 |
| 4.2.2.5. 7'den Az Kullanılan Tag'lerin İptali..... | 70 |
| 4.2.2.6. Semantik Yakınlık Hesaplama..... | 70 |
| 4.2.3. İki Ağ Çalışmasının Ortak Kümesinin İncelenmesi | 76 |

| | <u>Sayfa</u> |
|--|---------------------|
| BÖLÜM 5. | 81 |
| AĞ ANALİZİ SONUÇLARI..... | 81 |
| 5.1. GEPHI İLE AĞ ANALİZİ..... | 81 |
| 5.1.1. Gephi Yazılımı..... | 81 |
| 5.1.2. Gephi Temel Analiz Sonuçları | 83 |
| 5.1.3. Ağ Görselleri | 85 |
| 5.1.3.1. Birliktelik Ağı Görseli | 85 |
| 5.1.3.2. Semantik Ağı Görseli..... | 86 |
| 5.1.3.3. Ortak Küme Ağı Görseli..... | 88 |
| 5.2. DETAYLI DÜĞÜM VE LİNK ANALİZİ | 89 |
| 5.2.1. Düğüm Analizleri | 89 |
| 5.2.1.1. Derece Dağılımlarının Saçılmış Grafikleri (Scatter) | 89 |
| 5.2.1.2. Derece Dağılımlarının Toplanmış Grafikleri (Log-binned Degree Distribution)..... | 91 |
| 5.2.1.3. Ağırlıklı Derece Dağılımlarının Toplanmış Grafikleri (Log-binned Weighted Degree Distribution)..... | 92 |
| 5.2.2. Link Analizleri..... | 93 |
| 5.2.2.1. Link (Edge) Weight Dağılımlarının Scatter Grafikleri..... | 93 |
| 5.2.2.2. Maximum Degree Difference Scatter Grafikleri | 97 |
| 5.2.2.3. 2 Boyutlu Yüzdeli Dağılım Grafikleri (Percentile) | 100 |
| 5.2.3. Birlikte Kullanım Ve Semantik Yakınlık İlişkisinin Sorgulanması | 104 |
| 5.2.3.1. Hashtag’lerin Cosine Yakınlıklarına Göre Birlikte Kullanılma Olasılıkları | 104 |
| 5.2.3.2. Hashtag’lerin Birlikte Kullanılma Sıklıklarına Göre Cosine Yakınlıklarının Yüksek Olma Olasılıkları..... | 105 |
| BÖLÜM 6. | 108 |
| SONUÇ VE ÖNERİLER | 108 |
| 6.1. SONUÇLAR | 108 |
| 6.2. ÖNERİLER..... | 109 |
| KAYNAKLAR | 110 |
| ÖZGEÇMİŞ | 116 |

ŞEKİLLER DİZİNİ

Sayfa

| | |
|---|----|
| Şekil 1.1. Grafların matematiksel ve görsel gösterimleri..... | 3 |
| Şekil 1.2. Rassal ağlarda poisson ve gerçel ağlarda power-law derece dağılımı..... | 7 |
| Şekil 1.3. Rassal ağlar için Poissonian derece dağılımı..... | 8 |
| Şekil 1.4. Scale-free ağlarda power-law derece dağılımı..... | 9 |
| Şekil 1.5. Çeşitli gerçel ağlara ait power-law uyumlu derece dağılımı grafikleri..... | 9 |
| Şekil 1.6. Sosyal ağlarda kümelenme..... | 11 |
| Şekil 1.7. Modülerite ve çap korelasyonu..... | 13 |
| Şekil 1.8. WWW ağının 2007’de gelen linklere göre derece dağılımı..... | 16 |
| Şekil 1.9. WWW ağının 2007’de çıkan linklere göre derece dağılımı..... | 16 |
| Şekil 1.10. WWW ağının görselleştirilmiş görüntüsü..... | 16 |
| Şekil 1.11. İnternet ağı düğüm derece dağılımı..... | 17 |
| Şekil 1.12. İnternet ağının görselleştirilmiş görüntüsü..... | 17 |
| Şekil 1.13. Türkiye güç iletim ağı..... | 18 |
| Şekil 1.14. Dünyadaki havayolu ulaşım ağı..... | 19 |
| Şekil 1.15. Bir kargo firmasının dağıtım ağı..... | 20 |
| Şekil 1.16. E-mail ağının görselleştirilmesi..... | 21 |
| Şekil 1.17. Sosyal ağ görseli..... | 22 |
| Şekil 1.18. Bilimsel işbirliği ağına ait derece dağılım grafiği..... | 23 |
| Şekil 1.19. Bilimsel işbirliği ağı görseli..... | 23 |
| Şekil 1.20. Bilimsel işbirliğine ait görselleştirme..... | 24 |
| Şekil 1.21. 1960, 1972 ve 2007 yıllarına ait film-aktör ağının derece dağılımı..... | 25 |
| Şekil 1.22. Kelime ağı görseli..... | 25 |
| Şekil 1.23. New York Borsası finans ağı görseli..... | 27 |
| Şekil 1.24. Üretilmiş ağ türleri..... | 28 |
| Şekil 1.25. Farklı p değerlerine göre Erdos-Renyi modeli ağ görselleri..... | 29 |
| Şekil 1.26. Watts-Strogatz ağının p değerinin artışına göre görünümü..... | 31 |
| Şekil 1.27. Ölçek bağımsız ağ modeli görseli..... | 33 |
| Şekil 2.1. REST API akış şeması..... | 37 |

| | |
|--|----|
| Şekil 2.2. Streaming API akış şeması..... | 38 |
| Şekil 2.3. 140dev yapısı | 39 |
| Şekil 3.1. Veri ve bilgi çevrimi | 42 |
| Şekil 3.2. İki cümleye ait düğümler ve ilişkileri | 45 |
| Şekil 3.3. Semantik çalışmaya ait düğümlerin mesafeleri..... | 46 |
| Şekil 3.4. Semantik ve normal arama algoritmalarının doğruluk karşılaştırılması .. | 48 |
| Şekil 3.5. Önerilen sistem mimarisi | 49 |
| Şekil 3.6. Semantik ve örtülü semantik ilişki..... | 50 |
| Şekil 4.1. Otomatik Tweet mesajlarının engellenmesi..... | 54 |
| Şekil 5.1. Gephi yazılımı ara yüzü | 82 |
| Şekil 5.2. Birliktelik ağı görseli..... | 85 |
| Şekil 5.3. Birliktelik ağı görseli yakın görünümü | 86 |
| Şekil 5.4. Semantik ağı görseli | 87 |
| Şekil 5.5. Semantik ağı görseli yakın görünüm..... | 87 |
| Şekil 5.6. Birliktelik ağı ile semantik ağı ortak kümesi görseli | 88 |
| Şekil 5.7. Birliktelik ağı dağıtık gösterimi | 90 |
| Şekil 5.8. Semantik ağı dağıtık gösterimi..... | 90 |
| Şekil 5.9. Ortak küme dağıtık gösterimi..... | 91 |
| Şekil 5.10. Birliktelik ağı toplanmış derece dağılımı..... | 92 |
| Şekil 5.11. Semantik ağı toplanmış derece dağılımı | 92 |
| Şekil 5.12. Ortak küme toplanmış derece dağılımı | 92 |
| Şekil 5.13. Birliktelik ağı toplanmış ağırlıklı derece dağılımı | 93 |
| Şekil 5.14. Ortak küme toplanmış ağırlıklı derece dağılımı | 93 |
| Şekil 5.15. Birliktelik ağı linklerin dağıtılmış gösterimi | 94 |
| Şekil 5.16. Semantik ağı linklerine ait cosine değerlerinin frekans dağılımı (scatter)..... | 94 |
| Şekil 5.17. Ortak küme linklerin ağırlık değerlerinin frekans dağılımı (scatter) | 95 |
| Şekil 5.18. Ortak küme linklerine ait cosine değerlerinin frekans dağılımı (scatter)..... | 95 |
| Şekil 5.19. Birliktelik ağı linklerinin ağırlık dağılımı (Log-binned) | 96 |
| Şekil 5.20. Semantik ağı linklerinin ağırlık dağılımı (Log-binned)..... | 96 |
| Şekil 5.21. Ortak küme linklerinin ağırlık dağılımı (Log-binned)..... | 96 |
| Şekil 5.22. Ortak küme linklerinin cosine dağılımı (Log-binned)..... | 97 |

| | |
|---|-----|
| Şekil 5.23. Birliktelik ağına ait linklerin bağladığı düğümlerin, ağın son halindeki maksimum derecelerinin farklarına ait frekans dağılımı (scatter) | 97 |
| Şekil 5.24. Semantik ağına ait linklerin bağladığı düğümlerin, ağın son halindeki maksimum derecelerinin farklarına ait frekans dağılımı (scatter) | 98 |
| Şekil 5.25. Ortak kümeye ait linklerin bağladığı düğümlerin, ağın son halindeki maksimum derecelerinin farklarına ait frekans dağılımı (scatter) | 98 |
| Şekil 5.26. Birliktelik ağına ait linklerin bağladığı düğümlerin ağın son halindeki maksimum derecelerinin farkına ait dağılım grafiği (Log-binned)..... | 99 |
| Şekil 5.27. Semantik ağına ait linklerin bağladığı düğümlerin ağın son halindeki maksimum derecelerinin farkına ait dağılım grafiği (Log-binned)..... | 99 |
| Şekil 5.28. Ortak kümeye ait linklerin bağladığı düğümlerin ağın son halindeki maksimum derecelerinin farkına ait dağılım grafiği (Log-binned)..... | 99 |
| Şekil 5.29. Birliktelik ağına ait link ağırlığı – maksimum derece farkı percentile grafiği | 100 |
| Şekil 5.30. Semantik ağına ait cosine – maksimum derece farkı percentile grafiği | 101 |
| Şekil 5.31. Ortak küme ağına ait link ağırlığı – maksimum derece farkı percentile grafiği | 101 |
| Şekil 5.32. Ortak küme ağına ait cosine – maksimum derece farkı percentile grafiği. | 102 |
| Şekil 5.33. Ortak küme ağına ait link ağırlığı – cosine farkı percentile grafiği | 102 |
| Şekil 5.34. Hashtag'lerin semantik yakınlığa (cosine) bağlı olarak birlikte kullanılma olasılık dağılımı..... | 105 |
| Şekil 5.35. Hashtag'lerin birlikte kullanım ağırlıklarına göre semantik yakınlık kurma olasılıkları. | 105 |

ÇİZELGELER DİZİNİ

| | <u>Sayfa</u> |
|---|---------------------|
| Çizelge 3.1. Semantik benzer cümleler..... | 44 |
| Çizelge 3.2. Öğrencilerin verdikleri cevaplar | 47 |
| Çizelge 4.1. Türkçe karakter dönüşümleri | 57 |
| Çizelge 4.2. Linklerin ağırlık değerleri | 59 |
| Çizelge 4.3. Ağırlığı en yüksek 20 link | 60 |
| Çizelge 4.4. Stopwords | 62 |
| Çizelge 4.5. En çok kullanılan 1000 kelime | 63 |
| Çizelge 4.6. Cosine yakınlık değeri en yüksek 100 hashtag birlikteliği | 72 |
| Çizelge 4.7. Cosine yakınlık değeri en düşük 100 hashtag birlikteliği..... | 74 |
| Çizelge 4.8. Ortak kümenin cosine değeri en yüksek 100 hashtag birlikteliği | 76 |
| Çizelge 4.9. Ortak kümenin weight değeri en yüksek 100 hashtag birlikteliği | 78 |
| Çizelge 5.1. İnteraktif görselleştirme ve veri analiz programı örnekleri | 81 |
| Çizelge 5.2. Ağlara ait düğüm ve link sayıları..... | 83 |
| Çizelge 5.3. Gephi yazılımı analiz sonuçları | 83 |

SİMGELER VE KISALTMALAR DİZİNİ

SİMGELER

- G : Graf
 V : Düğüm Kümesi
 E : Link Kümesi
 n : Düğüm Sayısı
 p : Düşümler Arasındaki Bağlanma Olasılığı
 a_{ij} : İki Düşüm Arasındaki Bağlantı Sayısı
 k_i^{out} : Düşümden Çıkan Linklerin Sayısı
 k_i^{in} : Düşüme Gelen Linklerin Sayısı
 k_i : Toplam Düşüm Derecesi
 k : Ortalama Derece Değeri
 $P(k)$: Derece Dağılım Fonksiyonu
 $P(k_{in})$: Gelen Linkler İçin Derece Dağılım Fonksiyonu
 $P(k_{out})$: Çıkan Linkler İçin Derece Dağılım Fonksiyonu
 C : Kümelenme Değeri
 T_i : Geçişli Üçgenler Arasındaki Üçlülerin Sayısı
 tf : Terim sıklığı
 idf : Dökümandaki tekrar sayısı
 $tfidf$: Terimin tekrar sayısı ve dökümandaki tekrar sayısı

KISALTMALAR

- WWW : World Wide Web
MAC : Media Access Control Adress (Medya Erişim Kontrol Adresi)
IP : Internet Protokol
TUIK : Türkiye İstatistik Kurumu
API : Application Programming Interface (Uygulama Programı Arayüzü)

BÖLÜM 1

GİRİŞ

Çok fazla sayıda düğüm ve linklerin birlikte olduğu ağlar, karmaşık ağlar veya kompleks ağlar (Complex Networks) olarak isimlendirilmektedir. Kompleks ağların bir türü olan ve teknolojinin gelişimiyle ortaya çıkan internet ağı beraberinde www (world wide web) ağını getirmiştir. Bu ağ, sınırları olmayan ve hergün büyüyen bir yapıdadır. Normalde sürekli büyüyen bu veri yığını saklamanın ve incelemenin imkansız olması beklenirken, teknoloji gelişimi beraberinde bilgisayarların saklama alanlarını, işlem hızlarını ve performanslarını arttırdığı için sürekli artan düğüm ve link veri setini işlemeyi ve analiz etmeyi olanaklı hale getirmiştir.

Ağdaki düğüm ve linklerin birbirleriyle olan ilişkilerini açıklamak için Graf teoremi kullanılmaktadır. Graf teoreminin çıkış noktası Leonhard Euler tarafından yayınlanan Königsberg köprüsü problemidir. Bu problemde Königsberg şehrinde herbir köprüden sadece bir kez geçerek şehri gezmek amaçlanmaktadır. Euler, problemi graf yapıları kullanarak analiz etmiştir [1].

Teknolojinin ve beraberinde gelişen bilgisayarların oluşturduğu internet ağı, www ağı ve sosyal ağlar gerçek dünyada (real-world) yer alan ağ yapılarıdır. Bu ağlar graf yapılarıyla ifade edildiğinde daha kolay anlaşılabilir ve analiz edilebilmektedir.

1.1. KOMPLEKS AĞ TANIMI

Kompleks ağ kavramından önce ağ kavramını tanımlarsak; ağ, matematiksel olarak ifade edilebilen ve elemanları arasında ilişki içeren sistemlerin graf yapıları kullanılarak tanımlanmasıdır [2]. Karmaşık ağ olarak da isimlendirilen kompleks ağlar çok büyük yapıdadırlar. Büyüklük ifadesi değişken olmakla beraber düğüm ve link sayısı ile doğru orantılı olarak nitelendirilebilir. Yani bir kompleks ağda düğüm

ve link sayısı ne kadar fazla ise o kompleks ağı o kadar büyük diyebiliriz. Farklı özellikteki kompleks ağlar birbirlerini etkilemektedirler. Örnek olarak sosyal ağ topolojisi bir bilginin yayılım hızı ile doğrudan orantılıdır. Bu bilginin aktarım hızı da internet ağı, www ağı ve telekomünikasyon ağı ile doğrudan alakalıdır. Farklı bilim dalları ağ yapıları ile çalışmalar yaptığından tümünü kapsayan ve eksiksiz bir tanım yapmak zordur. Kompleks ağlar için bilim insanları ve araştırmacılar kendi çalışma alanlarına yönelik lokal tanımlar yapmışlardır [3].

Kompleks ağlar farklı tanımları olmasına rağmen şu şekilde ifade edilebilir: Gerçek dünyadaki ilişkileri modelleyen ve barındırdığı düğüm ve link sayısının büyüklüğü nedeniyle, karmaşık bir topolojiye sahip, büyük hacimli ilişki yapılarıdır. Kompleks ağlarda, düğüm kümesindeki elemanlar birbirlerine linkler ile bağlıdır [4].

1.2. GRAF (ÇİZGE) YAPILARI

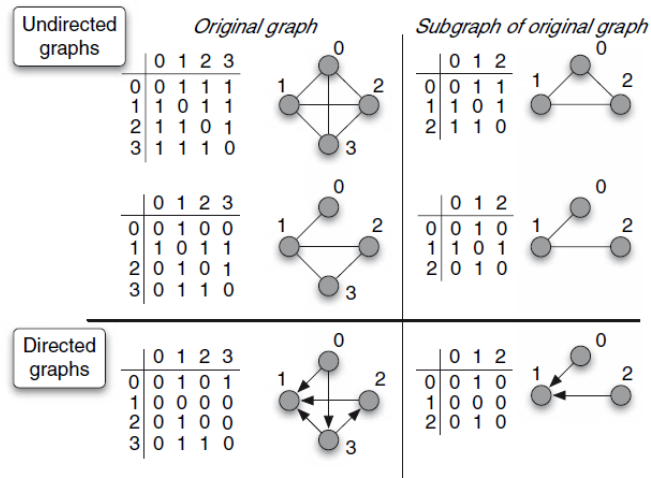
Bir ağıdaki düğüm ve linkler graf yapısı ile gösterilmektedir. Düğümler nokta, kare, üçgen gibi şekillerle; linkler ise çizgisel gösterimlerle ifade edilmektedir. Matematiksel olarak anlaşılması zor olan ağlar graf gösterimi ile anlaşılır hale gelmektedir.

$$G = (V; E) \quad (1.1)$$

Formülde G: Graf, V: düğüm kümesi, E: link kümesi kavramlarını ifade etmektedir. Modellemek istediğimiz nesnelere düğüm (Vertex) (V), bu düğümler arasındaki ilişki ve bağlantılar Linkler (Edge) (E) dir. Düğümler ve linkler arasındaki ilişkinin çizimsel oluşturulması Graf olarak adlandırılır. Sosyal ilişki ağını ele alırsak, düğümleri insanlar, bu insanlar arasındaki arkadaşlık ilişkileri ise linkler olarak isimlendirilir [3]. Graf teoremi ayrık matematiğin incelediği konulardan olmakla beraber ağ çalışmaları sosyal bilimlerde de önemli gelişmeler göstermiştir [1].

Düğümlerin ve linklerin konumları graf yapısında önemli değildir. Düğümlerin ve linklerin grafiksel özellikleri (boyut, şekil vb.) da önemli değildir. Düğümler

arasındaki ilişkinin doğru biçimde tanımlanması önemlidir. D ğ mler arası iliŐki iki eŐtittir: y nl  ve y ns z. İki d ğ m arasında oluŐan link iin y n bilgisi yoksa, y ns z link olarak adlandırılır. Link d ğ m 1 den d ğ m 2 ye dođru gider Őeklinde bilgi var ise y nl  link olarak adlandırılır. Y nl  ve y ns z grafa ait temel fark Őekil 1.1 de g sterilmiŐtir. Y ns z graflarda bađlantı her iki y nl  olabilmektedir. Ancak y nl  graflarda akıŐ sadece kaynak d ğ mden hedef d ğ me dođru olabilmektedir [5].



Őekil 1.1. Grafların matematiksel ve g rsel g sterimleri [6].

1.3. KARMAŐIKLIK (COMPLEXITY)

KarmaŐıklık teorisine g re; d nyada var olan her Őey birbiriyle bađlantılıdır ve bu her Őey bir ađın d ğ mlerinden birisini oluŐurmaktadır [7]. KarmaŐıklık kavramı b nyesinde iki temel tanımı barındırır. İlki, sistemlerin b y kl kleri  lisinde heterojen dađılım g stermeleri, ikincisi ise karmaŐıklıđın sistemi oluŐuran birimler arasındaki etkileŐimlerden kaynaklanmasıdır [5].

KarmaŐıklık gerel ađlarda zamana bađlı olarak ođalmaktadır. Ađların karmaŐıklıđı her yeni eklenen d ğ mde artmaktadır.  rneđin bir  niversite ađında her yeni eklenen b l m,  đrenci ve personel ile ađ dinamik olarak b y mektedir. Gerel ađlar ođunlukla heterojen bir b y me g steriler [5]. Zamanla karmaŐıklıđı artan ađlara  rnek olarak bilimsel iŐbirliđi ađları da verilebilir. Zamanla bilim ilerlemekte ve

ilerleme beraberinde yeni arařtırmacıları ve yayınları ortaya ıkartmaktadır [8]. Dnya zerinde yařayan yaklařık 7 milyar insanın oluřturacađı 7 milyar dđmn link sayısının hesaplanabilmesi ve oluřan sosyal iliřki ađının analizi gnmz teknolojisini zorlayacak veri hacmine sahiptir [7].

1.4. KOMPLEKS Ađ YAPILARI

Gerek dnyada bulunan bilgisayar ađları gibi fiziksel ađların veya sosyal ađlar gibi sanal ađların birbirleriyle iliřkili olarak bir araya getirilmesi kompleks ađları oluřturur.

Dđm ve link sayısı kk olan ađların gsterimi ve anlařılması kolaydır. Ancak kompleks ađlarda dđm ve link sayılarının ok byk olması, grselleřtirme aralarını yetersiz kılar. Dđm ve link sayılarının bymesine yol aan bilgisayar teknolojisi, aynı zamanda bu ađların modellenmesi ve topolojisinin belirlenmesi iinde kullanılır. Kompleks ađlar denilince karřımıza gerel ađlar ıkar. Bir de bu gerel ađlardan retilen retilmiř ađlar, sentetik ađlar veya bir diđer ismiyle model ađlar vardır. Kompleks ađlar; sentetik (model) ađlar ve gerel ađlar olmak zere iki grupta incelenir[7].

1.4.1. Ađ Parametreleri

Kompleks ađların anlamsal olarak nitelik kazanabilmesi iin bazı temel kavram ve tanımların bilinmesi gerekmektedir. Topolojik zellikler incelenerek gerel ve model ađlar analiz edilebilmektedir. Bu temel zellikler ile ađın yapısı ve durumu hakkında bir fikir sahibi olunabilmektedir. Model ađlar ve gerel ađlar ortak zellikler gstermektedirler. Ortak zelliklerin yanında farklı zelliklerde gsterebilirler.

Geliřen bilgisayar teknolojisi ile birlikte veri miktarı bymektedir. Byyen bu verilerin yine geliřen bilgisayar teknolojisi yardımıyla iřlenmesi ile gerek dnyadaki kompleks ađların analizi mmkn hale gelmiřtir. ok sayıda dđmn yer aldıđı ađlar iin, ađ yođunluđu, derece dađılımı ve ortalama yol uzunluđu gibi

bilgilerin elde edilmesi ve işlenebilmesi bu gelişen bilgisayar teknolojisi ile mümkün olmuştur [2].

Teknolojinin gelişimine bağlı olarak analiz edilebilen veri kapasitesi ve yeteneği artmaktadır. Bununla birlikte kompleks ağlar ile alakalı yeni kavram ve analiz türleri de ortaya çıkmaktadır.

1.4.1.1. Düğüm Dereceleri

Düğümler ve linkler bir ağ yapısının en temel öğeleridir. En çok kullanılan kavram ise düğümlerin sahip olduğu derece değerleridir.

Düğüm: Kompleks ağ yapısında bulunan en küçük birime düğüm (*node*) denir. Düğümler bir araya gelerek ağı oluştururlar. İnternet ağında bulunan internet erişimi olan her cihaz bir düğüm noktası olarak örneklenebilir. Twitter’ da kullanılan her bir hashtag yine düğümlere örnek olarak verilebilir.

Link: Kompleks ağ yapısında her bir düğümün bir diğer düğüm ile olan bağlantısına link denilir. Twitter’ da bir Tweet’te kullanılan her bir hashtag düğüm iken, hashtag’lerin aynı Tweet (Twitter mesajı) içerisinde kullanımı ise aralarındaki linkleri tanımlar.

Derece: Kompleks ağ yapısında bir düğümün diğer düğümler ile yapmış olduğu bağlantı (link) sayısı o düğümün derece (degree) değeri olarak isimlendirilir. Özellikle sosyal ağlarda derece kavramı önemlidir çünkü popüler ve etkisi yüksek olan düğümler derece kavramına bakılarak bulunabilir [9]. Sosyal ağlarda genellikle takipleşme ağı bir kompleks ağ olarak değerlendirilir. Bu tür ağlardaki kullanıcılar birer düğüm, kullanıcılar arasındaki takipleşme hareketleri ise link olarak değerlendirilir. Bu takipleşme ağı Twitter’daki gibi yönlü veya Facebook’taki gibi yönsüz linklerden oluşabilir.

N adet düğüm bulunan bir ağda düğümler arasındaki bağlantılar (linkler) $N \times N$ boyutlu bir matris ile gösterilebilir (Eşitlik 1.2). Matris içerisinde her iki düğüm

arasındaki bağlantı sayısı a_{ij} ile gösterilir. Her bir düğüm için derece değeri olan a_{ij} Eşitlik 1.2'deki ilişki matrisinin elemanları olmak üzere aşağıdaki Eşitlik 1.3'deki formül ile hesaplanmaktadır [9].

$$\begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1N} \\ a_{21} & \cdot & \cdot & \cdot & \cdot \\ a_{31} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{N1} & \cdot & \cdot & \cdot & a_{NN} \end{bmatrix} \quad (1.2)$$

$$k_i = \sum_{j \in N} a_{ij} \quad (1.3)$$

Düğümünden çıkan linklerin sayısı Eşitlik 1.4'deki formül ile hesaplanmaktadır. Düğüme gelen linklerin sayısının hesaplanması ise Eşitlik 1.5'de gösterilmiştir. Bazı ağlarda düğümler arasında tanımlan linklerde yön bilgisi mevcuttur. Linklerin yönlü ve yönsüz olmasına göre düğümlerin dereceleri değişmektedir. Eğer ağımızı oluşturan linkler yönlü ise her bir düğümünden dışarı çıkan link sayısı ile düğüme gelen link sayısı toplanarak düğüm derecesi elde edilir [8]. Düğüm derecesi Eşitlik 1.6'da verilen formül ile elde edilir.

$$k_i^{out} = \sum_{j \in N} a_{ij} \quad (1.4)$$

$$k_i^{in} = \sum_{j \in N} a_{ji} \quad (1.5)$$

$$k_i = k_i^{in} + k_i^{out} \quad (1.6)$$

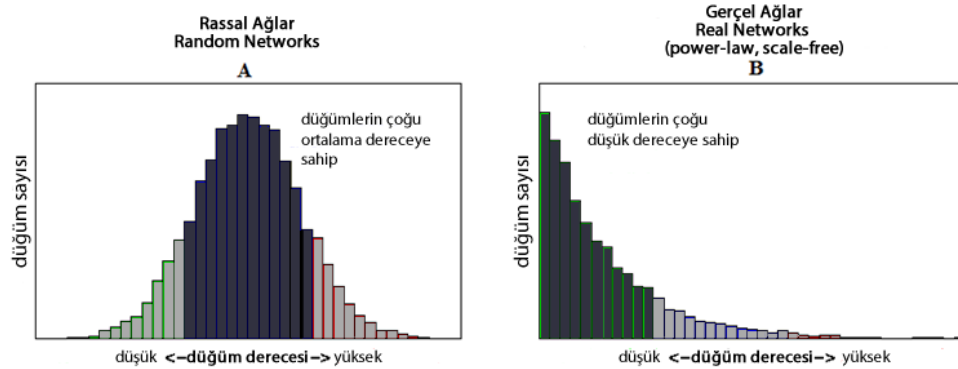
1.4.1.2. Derece Dağılımı

Düğüm derece dağılımı, kompleks ağların özelliğini belirleyen en önemli parametrelerden birisidir. Kompleks ağdaki her bir düğüme ait derece değerleri

bulunarak, bu derece değerlerinin tekrar sayıları elde edilir. Bu tekrar sayıları da normalize edilerek bir olasılık yoğunluk fonksiyonu şeklinde ($P(k)$) grafiğe dönüştürülür. Grafiğin analizini kolaylaştırmak için bazen logaritmik ölçekli eksenler de kullanılabilir.

Gerçek dünyada bulunan kompleks ağlarda düğüm dereceleri genellikle birbirine eşit değildir. Rastgele alınan bir düğümün k adet linke yani bağlantıya sahip olma olasılığını tanımlayan $P(k)$ dağılım fonksiyonu, derece dağılım kavramını açıklar [10]. Yönlendirilmiş graflarda düğümlere hem giren ve hem de çıkan linkler için ayrı ayrı olarak derece dağılımı hesaplanmaktadır. Rastgele seçilen bir düğümün k_{in} ve k_{out} sayısınca linke sahip olma olasılığını, düğüme gelen bağlantılara (linklere) ait $P(k_{in})$ ve düğümden çıkan bağlantılara (linklere) ait $P(k_{out})$ dağılımları belirler [6].

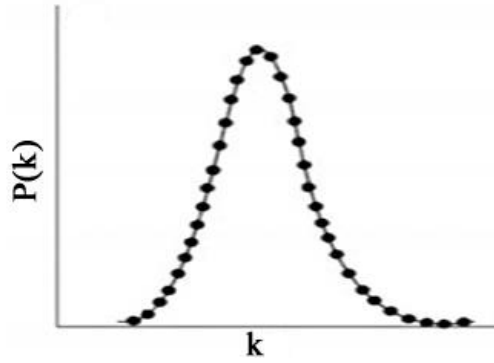
Gerçek ağlar ve model ağlar incelendiğinde bu ağların farklı düğüm derece dağılımlarına sahip oldukları gözlemlenmiştir. Bu gözlem sonucunda kompleks ağların derece dağılımlarının iki farklı şekilde gerçekleştiği görülmüştür. Poisson veya Binom olarak isimlendirilen derece dağılımında, ağ içerisinde bulunan düğümlere ait dereceler belli bir ortalama derece etrafında çan eğrisi şeklinde yaygınlaşmıştır. Diğeri ise, ağ içerisinde bulunan düğümlerin ilişkilerinin heterojen olarak dağıldığı kompleks ağlarda karşılaşılan power-law (güç-yasası) derece dağılımı ve türevleridir [4].



Şekil 1.2. a) Rassal ağlarda Poisson ve b) gerçek ağlarda power-law derecede dağılımı [11].

Poission Derece Dağılımı

Poission derece dağılımı, düğümler arasındaki bağlantıların (linklerin) rastgele dağıtıldığı rassal ağlarda görülmektedir. Düğümler arası linklerin rastgele dağıtıldığı rassal ağlarda bu rastgele dağıtımdan dolayı düğümlerin ortalama derece değerleri birbirine yakın olmaktadır. Kompleks ağdaki derece dağılımı bu ağın ortalama derecesi olan $\langle k \rangle$ değeri seviyesinde yoğunlaşan bir çan eğrisi şeklinde oluşmaktadır [10]. Rassal ağlarda elde edilen ortalama derece değerinin görünümü Şekil 1.3'deki gibi görüntülenmektedir.



Şekil 1.3. Rassal ağlar için Poissonian derece dağılımı [12].

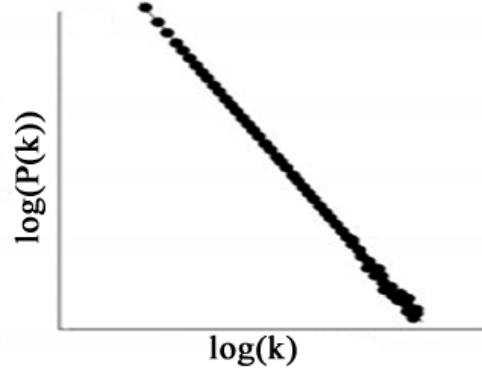
Power-Law (Güç-Yasası) Derece Dağılımı

Gerçek ağlarda Rassal ağlardan farklı olarak ağ içerisinde her ölçekte dereceye sahip düğümler bulunabilmektedir. Bundan dolayı düğümlerin sahip olduğu derece değerleri bir bölgede yoğunlaşmayıp, geniş bir ölçek içerisinde dağılmaktadır. Gerçek ağlarda rastlanan bu dağılım scale-free (ölçeksiz) olarak adlandırılmakta ve matematiksel olarak power-law ilkesine uyum göstermektedir [7]. Ağ analizlerinde power-law derece dağılımı ile ilgili çalışmalar, WWW ağının derece dağılımının Poisson dağılımından farklı olduğunun gösterilmesinden sonra hız kazanmıştır [10].

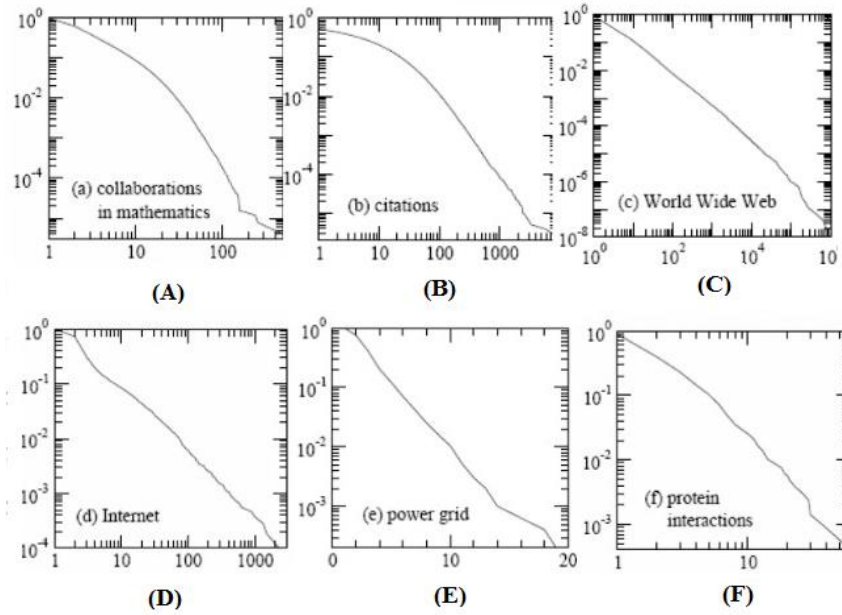
Düğümlere ait derece dağılımlarının heterojen özellik gösterdiği Scale-free özellikli gerçek ağlarda genellikle Power-law benzeri derece dağılımları görülür. Power-law yapıdaki ağlar Eşitlik 1.7'deki denklem ile ifade edilmektedir [5].

$$P(k) = Ak^{-y} \quad (1.7)$$

Power-law dağılımı analitik olarak, iki eksenini de logaritmik ölçeklenmiş bir derece dağılımı grafiğinin doğrusal düşüş gösterdiği durumlarla ilişkilendirilir (Şekil 1.4). Bu doğrusal düşüşün eğimi, Eşitlik 1.7'deki $-y$ katsayısına eşittir [13].



Şekil 1.4. Scale-free ağlarda power-law derece dağılımı [12].



Şekil 1.5. Çeşitli gerçel ağlara ait power-law uyumlu derece dağılımı grafikleri [14].

Şekil 1.5'deki görselde, farklı boyut ve türdeki bilimsel işbirliği ağından güç transfer ağına kadar gerçel ağların düğüm derece dağılımının power-law ilkesine uygun olarak gerçekleştiği görülmektedir. Ancak nadiren power-law prensibine tam uyumlu

(Şekil 1.5c) ağlar görülür, log-log eksenlerdeki doğrunun sol tarafında saturasyon (Şekil 1.5b) ya da sağ tarafında expansiyonel faz (Şekil 1.5e) eğilimli doğrularda sıkça rastlanmaktadır [13]. Power-law dağılımında grafiğin kuyruğu zamanla yavaş bir şekilde sıfır değerine doğru yakınsamaktadır [15].

1.4.1.3. Kümelene (Clustering)

Kümelene kavramı da sosyal ağlarda sık kullanılan terimlerden birisidir. Popüler düğümlerin çok olduğu işbirliği veya aktör ağları gibi ağlarda, kümelene durumu ile sıkça karşılaşılır. Kümelene sonucunda birbirleri ile ilişkili olan kapalı çevirim küçük kümeler oluşur. Örneğin bilimsel işbirliği ağında sürekli birbirleri ile yayın yapan yazarların oluşturduğu kümelene durumu buna örnek olarak gösterilebilir [10].

Bir düğümün komşularının kendi aralarındaki komşuluk ilişkileri kümelene kavramı ile ölçülmektedir. Sosyal ağlarda bir düğüme ait arkadaşların birbirlerini tanıyor olması kümelene değildir. Kümelene değeri bu durumda yüksek olmaktadır. Kümelene değeri yüksek olan bu tarz ağlarda düğümler arasında gerçekleşen üçgenlerin sayısı görece fazla olmaktadır [3].

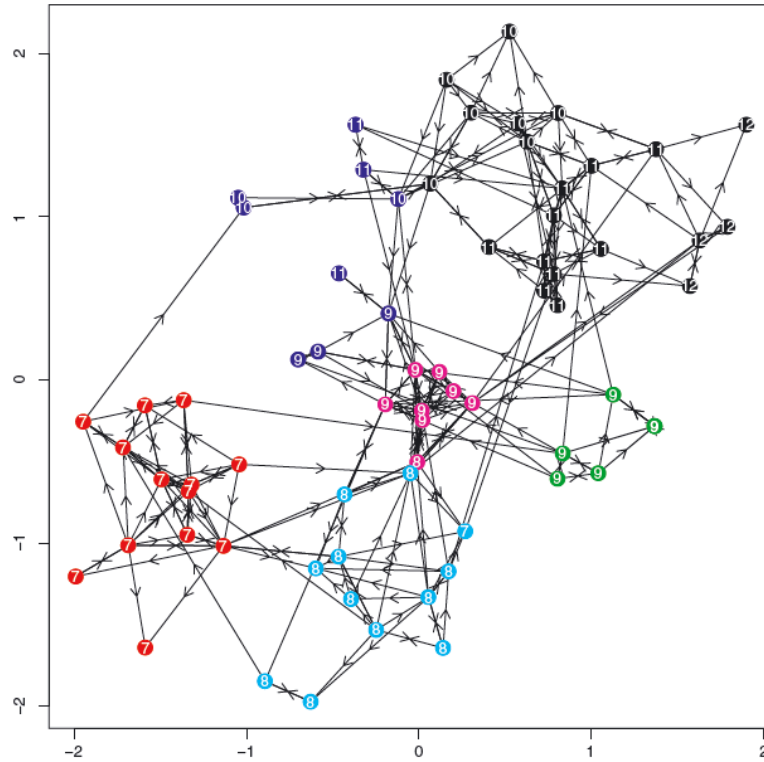
Genel bir grafta, T geçişli üçgenler arasındaki üçlülere ifade etmek üzere, T_i değeri Eşitlik 1.8a kullanılarak hesaplanır. Kümelene değeri olan C Eşitlik 1.8b kullanılarak bulunur [1,7]. C 'nin aldığı değer ağın sahip olduğu toplam kümelene katsayısının değerini ifade etmektedir. Herhangi bir düğüme ait kümelene katsayısının hesaplanması Eşitlik 1.8c'de gösterilmiştir. Eşitlik 1.8c'deki, i .Düğüme ait komşuların sayısı; E_i ise, bu komşuların gerçekte oluşturmuş oldukları link sayısıdır [10].

$$T_i = \frac{\text{Birbilerine üçgen şeklinde bağlı düğüm sayısı}}{\text{Üçgen şeklinde olup birbirlerine bağlı üçlü düğümlerin sayısı}} \quad (1.8a)$$

$$C = \frac{1}{n} \sum_i T_i \quad (1.8b)$$

$$C_i = 2E_i/k_i(k_i - 1) \quad (1.8c)$$

Sosyal ağlarda kümelenme üzerine yapılan bir çalışmada, gençlerin ergenliğin ilk dönemlerinde sağlıklı olmaya yönelik davranışları ve bu davranışların yetişkinliğin ilk dönemlerinde verdiği sonuçlar incelenmiştir [16]. Kurulan ağa ait kümelenme Şekil 1.6'da gösterilmiştir.



Şekil 1.6. Sosyal ağlarda kümelenme [16].

1.4.1.4. En Kısa Yol

İki düğüm arasındaki yollar içerisinde en kısa mesafeye sahip bağlantı yoluna en kısa yol denilmektedir. İki düğüm arasında belirlenen en kısa yol, başlangıç düğümünden çıkılarak bitiş düğümüne ulaşıncaya kadar geçilen link sayısına eşittir. Tüm düğümlerin birbiri ile ilişkili olduğu model ağlarda bu uzunluk her zaman bulunabilmektedir. Ancak bazı gerçek ağlarda tüm düğümler birbiri ile ilişkili olmadığından ve ana kümeden ayrık bir takım izole düğüm kümeleri bulunduğundan her düğüm için mevcut en kısa yol bilgisine ulaşılamamaktadır [8].

Ulaşım ve telekomünikasyon ağlarında en kısa yol kavramı çok önemlidir. Örnek olarak internet ağında bulunan bilgisayarlar arasında paket transferlerinin hızlı bir şekilde gerçekleşmesi için en kısa yolun bilinmesi gerekliliğini gösterebiliriz. Tespit edilen en kısa yol paketlerin hızlı bir şekilde transfer edilmesini ve sistem kaynaklarının maksimum verimle kullanılmasını sağlamaktadır [1]. Coğrafi bilgi içeren ulaşım ağlarında ise iki düğüm arasındaki en kısa yol link sayısı ile değil düğümler arasındaki mesafe bilgisine bakılarak bulunur. Örneğin karayolları ağında kaç tane düğüm sayısı olduğu değil, düğümler arası coğrafi mesafelere bakılarak en kısa yol belirlenir.

Gerçel ağlarda düğümler birbirlerine oldukça yakın mesafelerde bulunurlar. Milgram'ın mektup deneyinde ve Twitter gibi sosyal ağların arkadaşlık ilişkisinin araştırılması gibi çalışmalarda, küçük dünya ağları arasındaki ortalama uzaklık değerinin 3 ile 6 adım arasında değiştiği görülmektedir [10].

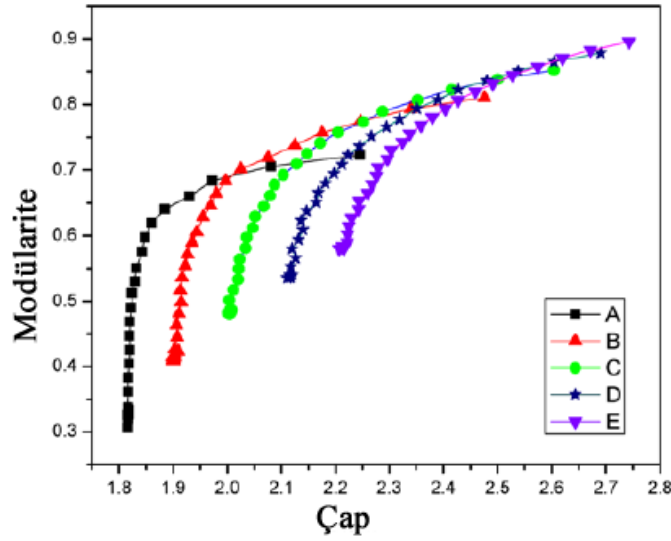
1.4.1.5. Çap (Diameter)

Milgram çalışmasında ortalama uzaklığın 6 adım olarak tespit edilmesi ile birlikte ağlardaki çap uzunluğuna dair önem artmıştır. Bir ağda bulunan tüm düğüm çiftleri arasındaki ortalama asgari yol uzunluğuna çap adı verilir. Rastgele ağlar kısa çap uzunluğuna sahiptir. Kısa çap uzunluğu ağın metabolik iletişim etkinliğini arttırdığından dolayı gerçel ağlar rastgele ağlara göre daha kısa çap uzunluğuna sahiptir [17]. Bir ağa ait topolojik özelliklerin belirlenebilmesi için ağın çap uzunluğunun bilinmesi gerekmektedir. Çap ağda lineer büyüklük ile ifade edilir. Bu büyüklük ağın gerçekte ne kadar büyük olduğunu gösterir. Başka bir deyişle çap ne kadar büyükse, ağ gerçekte o kadar büyüktür.

1.4.1.6. Modülarite (Modularity)

Linklerin kümelenme sayısı ile bağlantıların rastgele olduğu eşdeğer bir ağdaki bağlantı sayısının farkının çarpımsal sabitine modülarite adı verilmektedir [18]. Bir başka deyişle ağların alt gruplara bölünme ihtimaline modülarite denilmektedir. Modülarite pozitif veya negatif olabilir.

Ağ çapı büyüdükçe ağda bulunan düğümler ve linklerin birbirleri ile kümelenme ihtimalleri artacağından modülaritesi artacaktır. Simüle olarak oluşturulan farklı ağlara ait ağ çapları ve modülarite korelasyonu Şekil 1.7’de verilmiştir. Grafikte (Şekil 1.7) görüleceği üzere çap arttıkça modülarite artmaktadır. Çap ve modülarite arasındaki ilişki lineer değildir. Çapın kısa olduğu durumlarda, çaptaki küçük değişim modülaritede önemli artışa yol açmaktadır [17].



Şekil 1.7. Modülarite ve çap korelasyonu [17].

1.4.2. Gerçel Ağlar

Kompleks ağlar, dünya üzerinde mevcut olan gerçek veya sanal olarak varlığı bilinen olgulara ait ilişkilerin toplanarak, bunların birleştirilmesi sonucu elde edilen ağlardır. Bölüm 1.3.1. de incelediğimiz model ağlar, gerçel ağların anlaşılabilirliği ve incelenmesi için geliştirilmiş ağ modelleridir. Gerçel ağlarda düğümleri gerçek objeler ve olgular oluşturur. Örneğin bir arkadaşlık ağında gerçek insanlar düğümleri, bu insanlar arasındaki ilişkiler de linkleri oluşturur. Karmaşık yapıdaki ağların veri setleri elde edilerek gerçel ağlar ile alakalı araştırmalar yapılabilir. Ağ yapılarına ilişkin bilgiler, veri setlerinin bilgisayar yardımıyla metrik analizi ve görselleştirilmesiyle elde edilir. Gerçel ağlar üzerinde yapılan çalışmalar iki temel bölüm altında incelenmektedir [8].

1. Ağ sistemlerine ait düğüm derece dağılımı, kümelenme ve en kısa yol gibi matematiksel ve istatistiksel durumların incelenmesidir [8].
2. Ağ yapılarının görselleştirilmesidir. Görselleştirme çalışmaları ile ağıma ait ilişkiler resim formatına getirilmektedir. Görselleştirme ile soyut ağ yapılar somut olarak sunulmaktadır [8].

Aralarında ilişki bulunan dünya üzerindeki yapılar bir araya gelerek gerçel ağları oluştururlar. Türlerine ve yapılarına göre sınıflandırılan gerçel ağlar literatürde aşağıdaki ana başlıklar altında incelenmektedir [13,19].

1. Teknolojik Ağlar (Technological networks)
2. Sosyal Ağlar (Social networks)
3. Biyolojik Ağları (Biological networks)
4. Biyokimya Ağları (Biochemical networks)
5. Bilgi Ağları (Networks of information)
6. Jeofizik Ağları (Geophysical networks)
7. Ekolojik Ağlar (Ecological networks)
8. Finansal Ağlar (Financial networks)

1.4.2.1. Teknolojik Ağlar

1947 yılında transistörün icadıyla hızlı bir ivme kazanan bilişim sektörü ile birlikte teknolojik ağlar hayatımıza dahil olmuştur. Günlük hayatımızda en çok dahil olduğumuz bu ağlara internet ağı ve www örnek olarak gösterilebilir. Teknolojinin getirdiği elemanlar bu ağlarda düğümleri ve linkleri oluştururlar. İnternet ağına bilgisayarlar ve yönlendiriciler düğümleri, bu düğümler arasında bulunan bağlantılar da linkleri oluşturur. Bu linkler kablolu ve kablosuz olarak bulunabilirler.

Teknolojik ağlara örnek olarak aşağıdakiler verilebilir [1,13,19,20].

1. İnternet ve WWW (World Wide Web) Ağı
2. Telefon Ağları
3. Enerji İletim Ağları (Power grids)

4. Ulaşım (Transportation) Ağları
5. Ulaştırma ve Dağıtım Ağları
6. E-posta Ağı

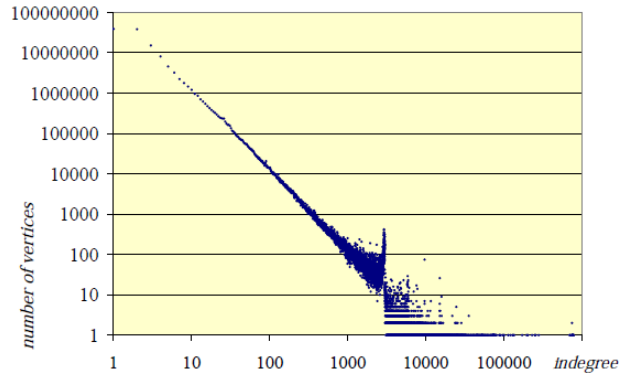
İnternet Ve WWW Ağı

Birbirine karıştırılan WWW ve internet kavramlarını şu şekilde açıklayabiliriz; web sayfalarının oluşturduğu ağ www ağı, birbirleriyle elektronik olarak haberleşebilen fiziksel donanımların oluşturduğu ağ ise internet ağıdır. Özetle internet erişimine açık cihazların birlikte oluşturduğu ağ internet ağı olarak isimlendirilebilir [14]. Bugüne kadar tespit edilebilen en büyük ağ WWW ağıdır. Sürekli yeni eklenen düğüm ve linkler ile devamlı büyüyen WWW ağında düğümleri web sayfaları, düğümler arasındaki bağlantıları (linkleri) ise bu düğümleri yani web sayfalarını birbirlerine bağlayan hyperlink'ler oluşmaktadır [20]. Günümüzde (Nisan 2016) www'nin en az 4.63 milyar adet web sayfasından oluştuğu tahmin edilmektedir [21]. Web sayfalarında yönlendirilmiş bağlantı bilgisi içeren gelen ve giden linkler vardır. Düğüme ait derece değerini her bir web sayfasına gelen ve giden linklerin toplamı oluşturmaktadır [22]. Sürekli eklenen düğüm ve linklerle devamlı büyüyen www ağının gerçek boyutunu tam olarak belirlemek imkânsızdır. Google gibi gelişmiş arama motorlularında dahi bütün web sayfalarına ait veriler indekslenmemektedir [9].

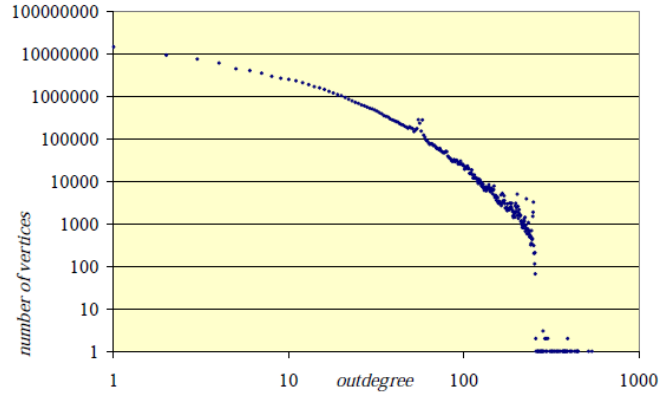
WWW'e ait düğüm derece değerleri Eşitlik 1.9 kullanılarak bulunur. Linkler yönlü olduğu için giren ve çıkan linklerin derece değerler ayrı ayrı hesaplanır [1]. Gelen linklerin dereceleri Eşitlik 1.9a'daki formül ileçıkan linklerin dereceleri Eşitlik 1.9b'deki formül ile bulunur. WWW ağı düğüm derece dağılımı power-law'a uygun olarak gerçekleşmektedir [20].

$$P_{in}(k) = k^{-\gamma_{in}} \quad (1.9a)$$

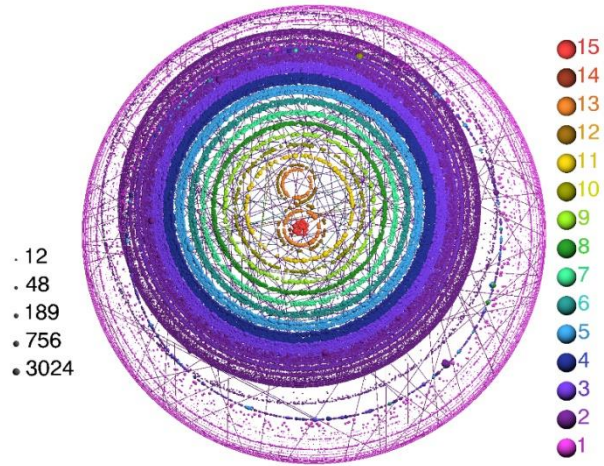
$$P_{out}(k) = k^{-\gamma_{out}} \quad (1.9b)$$



Şekil 1.8. WWW ağının 2007’de gelen linklere göre derece dağılımı [23].



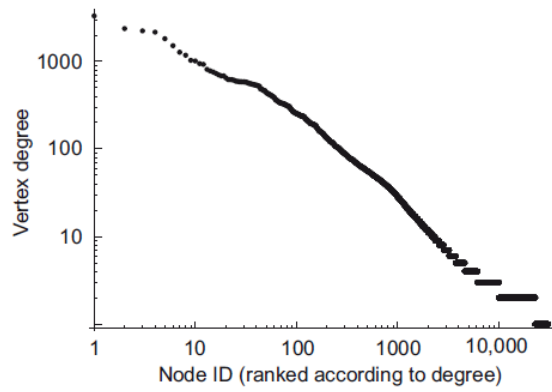
Şekil 1.9. WWW ağının 2007’de çıkan linklere göre derece dağılımı [23].



Şekil 1.10. WWW ağının görselleştirilmiş görüntüsü [24].

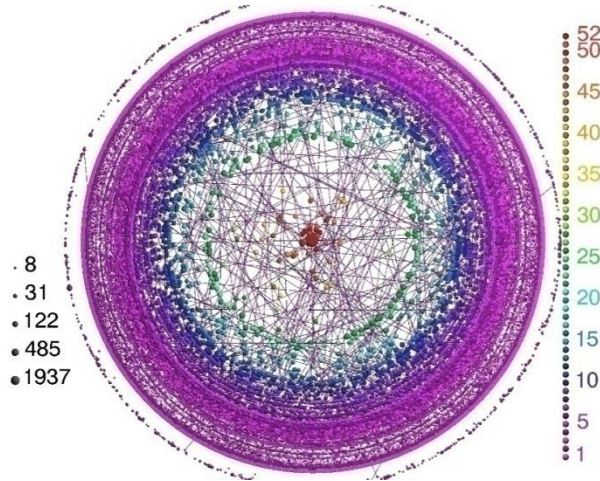
Şekil 1.10'da WWW ağının derece merkeziliği değerine göre renklendirilmiş görseli verilmiştir. Şeklin sağ tarafında merkezilik değerine göre renk ölçeği verilmiştir. Şeklin (Şekil 1.10) sol tarafında ise bazı düğümlere ait dereceler verilmiştir [24].

İnternet ağında düğümler, yönlendirici (router) ve bilgisayarlardan oluşmaktadır. Düğümler arasındaki linkler ise bu fiziksel cihazları birbirlerine bağlayan kablolu ve kablosuz bağlantılardan oluşmaktadır [20]. MAC adresleri ve IP protokolleri gibi sistemler internet ağında iletişim için kullanılmaktadır [9].



Şekil 1.11. İnternet ağı düğüm derece dağılımı [9].

Şekil 1.11'de görüldüğü üzere internet ağının düğüm derece dağılımı power-law özelliği göstermektedir [20]. İnternet ağının görselleştirilmiş görünümü Şekil 1.12'de verilmiştir.



Şekil 1.12. İnternet ağının görselleştirilmiş görüntüsü [25].

Telefon Ağı

Teknolojik ağlardan günümüzde kullanılan en eskisi telefon ağıdır. Telefon ağlarında veriye ulaşmak zor olduğundan bu ağla ilgili çalışma çok yapılmamıştır. Hizmet sağlayıcılar telefon kayıtlarını saklamalarına rağmen özel hayatının gizliliği ilkesinden dolayı araştırmacılar ile paylaşmamaktadırlar [19].

Telefon ağlarında düğümleri telefon numaraları oluşturmaktadır. Sabit veya mobil bir telefona ait numara bir düğümü oluşturur. Bu düğümler arasında yapılan aramalar ise linkleri oluşturur. İki telefon arasındaki görüşme bir düğümden (arayan numaradan) diğer düğüme (aranan numaraya) doğru yönlendirilmiş bir bağlantı (link) oluşturmaktadır. Bundan dolayı derece dağılımları yapılırken giden ve gelen linklere göre incelenmesi gerekmektedir. Telefon ağı için yapılan araştırmalarda giden ve gelen linklerin derece dağılımının power-law uyumlu olduğu tespit edilmiştir [20].

Enerji İletim Ağları

Yüksek voltajlı elektrik akımının ülkeler ve şehirler arasında uzun mesafeler boyunca taşınması için oluşturulan güç aktarım ağlarına enerji iletim hatları (Power Grids) denilmektedir.

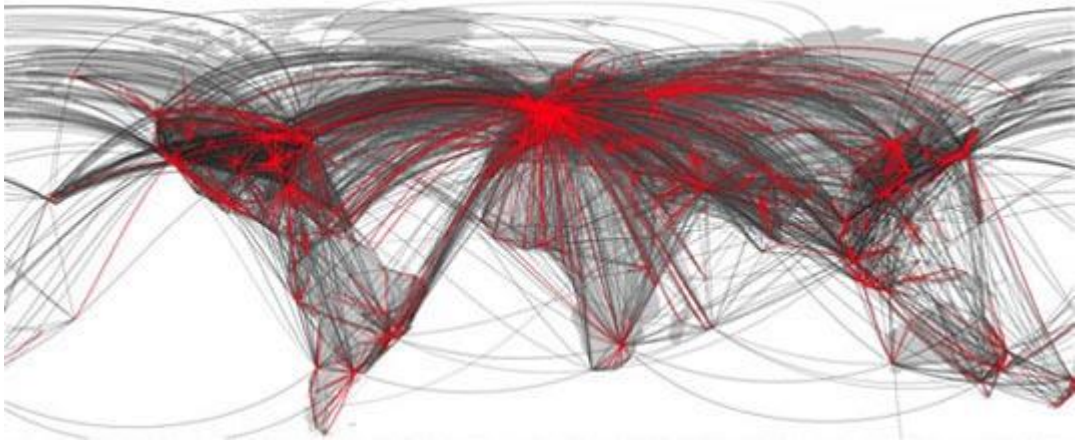


Şekil 1.13. Türkiye güç iletim ağı [26].

Yüksek voltaj elektrik iletim hatlarının oluşturduğu ağ yapıları düğüm ve linklerden oluştuğu için kompleks ağlar içerisinde yer almaktadır. Ağın düğümlerini güç aktarım için kullanılan trafolar (switching substations) ve iletim ağlarında elektrik üretimi yapılan istasyonlar (barajlar ve santraller gibi) oluşturmaktadır. Linkleri ise üretim tesisleri arasındaki kablolardan oluşan elektrik iletim hatları oluşturmaktadır. Elektrik iletim hatlarına dair bilgiler kolay elde edilebildikleri için üzerlerinde araştırmalar yapılabilir [19]. Şekil 1.13'de Türkiye'ye ait elektrik iletim ağının görseli görülmektedir.

Ulaşım Ağları

Ulaşım ağları, uçak, tren veya otobüs rotalarını barındıran ağ çeşididir. Bu ağları tanımlamak çok zor olmamakla birlikte, verileri analiz etmek zahmetli olabilmektedir [19]. Bir başka kompleks ağ çeşidi olan internet ağı ve www ağı gibi ağlar kullanılarak işlevlik kazanan ulaşım ağlarında, trenler için raylardan oluşan, otobüsler için karayollarından oluşan ve uçaklar için havayollarından oluşan bir yapı mevcuttur. Ulaşım ağlarında havaalanları düğüm noktaları bu havaalanları arasındaki uçak yolculukları linkleri oluşturmaktadır.



Şekil 1.14. Dünyadaki havayolu ulaşım ağı [27].

Ulaştırma Ve Dağıtım Ağları

Dağıtım ağları; petrol ve gaz boru hatları, su ve kanalizasyon yolları ve posta ofisleri ile kargo firmalarının kullandığı rotalardan oluşur [19]. Su ve kanalizasyon hatlarında her ev bir düğümü temsil ederken, su kaynakları da bir düğümü temsil etmektedir. Bunların arasındaki yollar ise linkleri oluşturur.



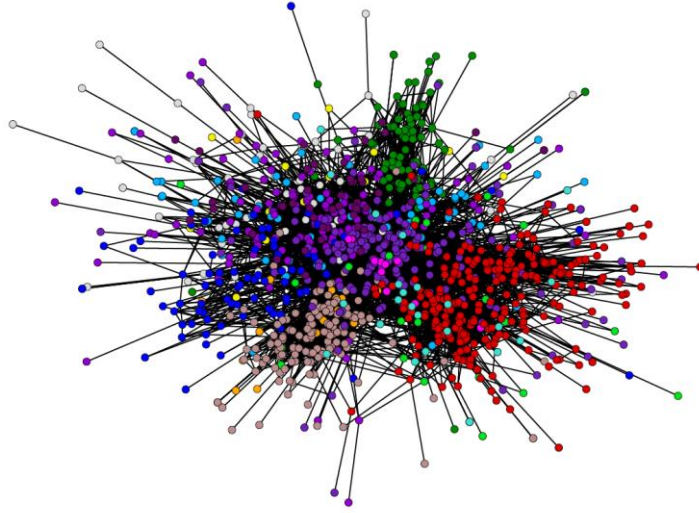
Şekil 1.15. Bir kargo firmasının dağıtım ağı [28].

Kargo dağıtım ağında her bir kargo şubesi bir düğümü temsil ederken arada izlenen rotalar ise linkleri oluşturur. Şekil 1.15’de bir kargo firmasına ait Türkiye sınırlarındaki düğüm noktaları ve linkleri gösterilmiştir.

E-mail (Elektronik Posta) Ağları

Teknolojik gelişmeler ile birlikte popülerliğini kaybeden kâğıt mektupların yerini elektronik postalar almıştır. Elektronik posta kullanımı internet kullanımının yaygınlaşması ile artmıştır. E-posta ağlarında, kompleks ağı oluşturan düğümleri posta adresleri oluşturmaktadır. E-posta hesapları arasında gönderilen mesajlar ise linkleri oluşturmaktadır. Hesaplar arasında iletilen bilgilerin gizliliğinden dolayı araştırmacılar ile paylaşılmamakta ve bu durum bu ağın incelenmesini zor kılmaktadır. E-posta ağında da telefon ağında olduğu gibi düğümler arasında ki

linkler bir noktadan diğere doğru olduğundan yönlendirilmiş bilgi içermektedirler [14].



Şekil 1.16. E-mail ağının görselleştirilmesi[14].

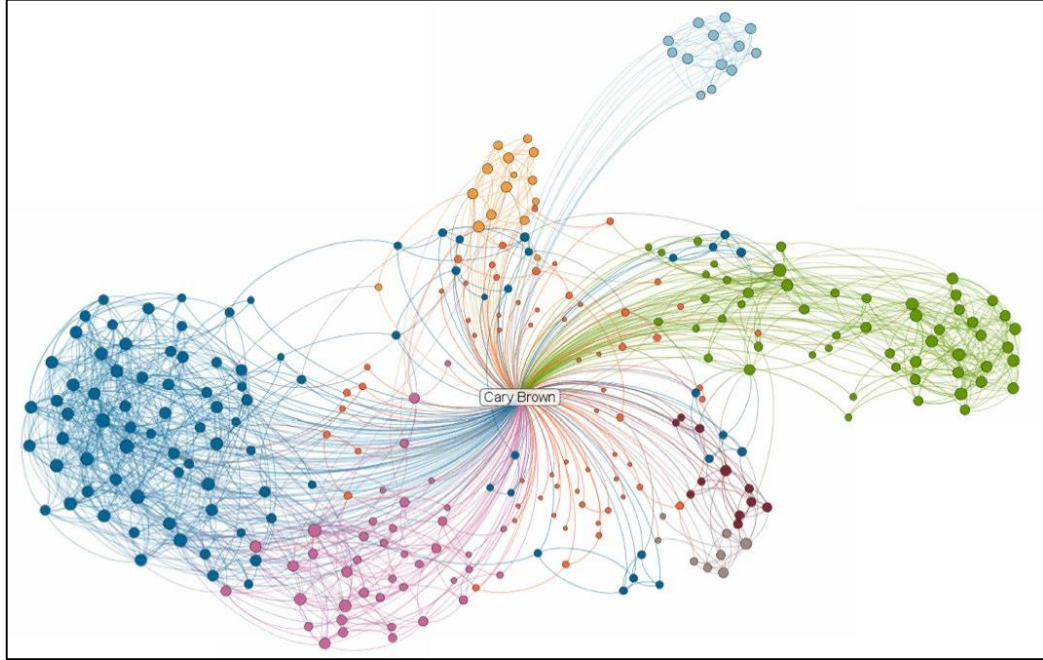
Tarragona Üniversitesine ait e-mail ağının görseli Şekil 1.16’da verilmiştir. Üniversitenin farklı birimleri farklı renkler ile gösterilmektedir. Kullanıcılar arasındaki mail trafiğine göre linkler oluşturulmuştur [14].

1.4.2.2. Sosyal Ağlar

Sosyal ağlar yeryüzündeki gerçek bireylerin (aktör, bilim insanı ve müzisyen vb.) dostluk, bağlılık (affiliation) ve işbirliği gibi sosyal kavramlar ile olan birlikteliklerini tanımlamak için kullanılmaktadır. Gelişen teknoloji ile yeni sosyal ağ türleri ortaya çıkmış ve hayatımıza girmiştir [1]. En bilinen sosyal ağlara bilimsel işbirliği ağları ve film-aktör ağları örnek olarak verilebilir [20]. Son kullanıcı seviyesinde internet kullanımının artması ile Facebook ve Twitter gibi arkadaşlık ağları da popüler olan sosyal ağlardır.

Sosyal ağ kavramı Facebook, Twitter ve Instagram gibi bireylerin sanal ortamda bir araya gelerek sosyal etkileşimde buldukları siteler ile sıkça duyulmaya başlanmıştır. Her ne kadar internet ortamındaki bu tarz siteler sosyal ağlara örnek

olsa da sosyal ağ için internet altyapısına gerek yoktur. İnternet tabanlı bir yapı içermeyen bilimsel işbirliği ağları da sosyal ağa örnek olarak verilebilir. Şekil 1.17’de bir sosyal ağ görseli örneği verilmiştir.



Şekil 1.17. Sosyal ağ görseli [29].

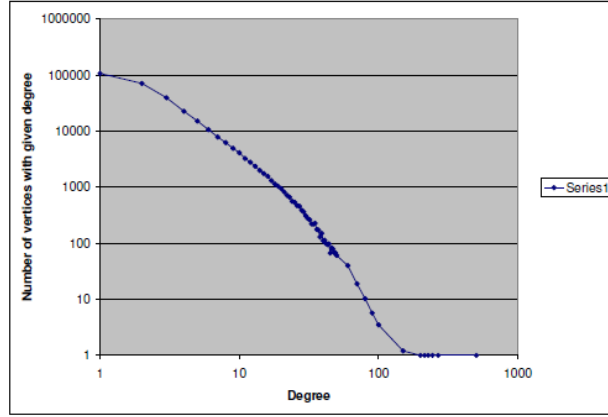
Sosyal ağların en bilinenleri şunlardır[1,10,20]:

1. Bilimsel İşbirliği Ağları (Scientific collaboration networks)
2. Film-Aktör İşbirliği Ağları (Movie actor collaboration networks)
3. Kelime Ağları (Networks in linguistics)
4. Facebook, Twitter ve Instagram sosyal ağları

Bilimsel İşbirliği Ağları

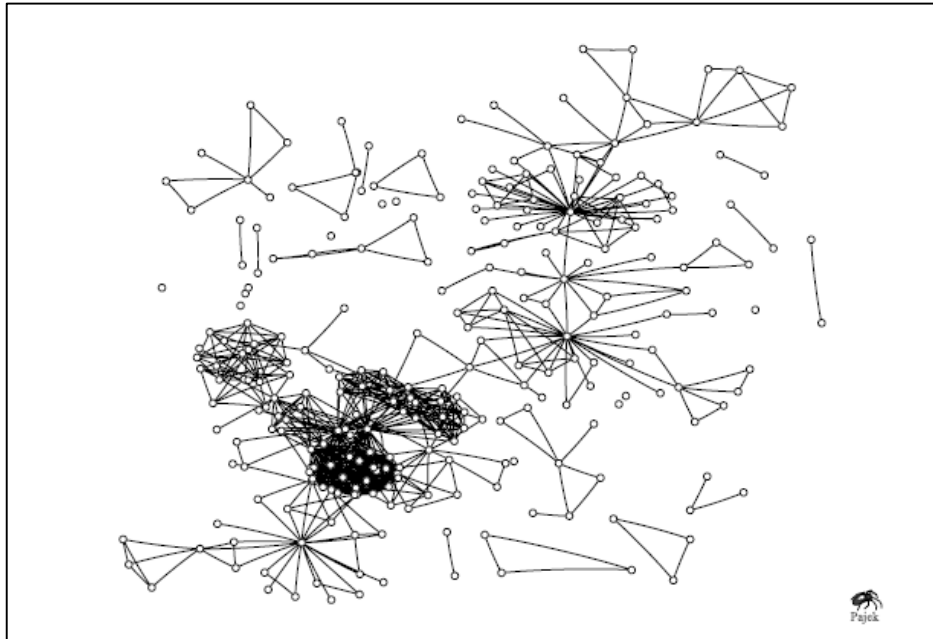
Bilim insanlarından oluşan sosyal ağa bilimsel işbirliği ağları (Scientific collaboration networks) denilmektedir. Bilimsel işbirliği ağı temelde bilim insanlarının birlikte yaptığı çalışmalardan oluşur. Bu birlikte yapılan çalışmalar bir makale haline getirildiğinde bilim insanları (yazarlar) düğümleri, yaptıkları yayındaki ortaklıkları ise linkleri oluşturur [20].

Ortalama yol uzunluđu oldukça kısa olan bilimsel iřbirliđi ađlarında kmelenme katsayısı ok yksek olarak gzlenmektedir. Bilimsel iřbirliđi ađlarında dđm derece dađılımı power-law ilkesine uyumludur (řekil 1.18)[20].



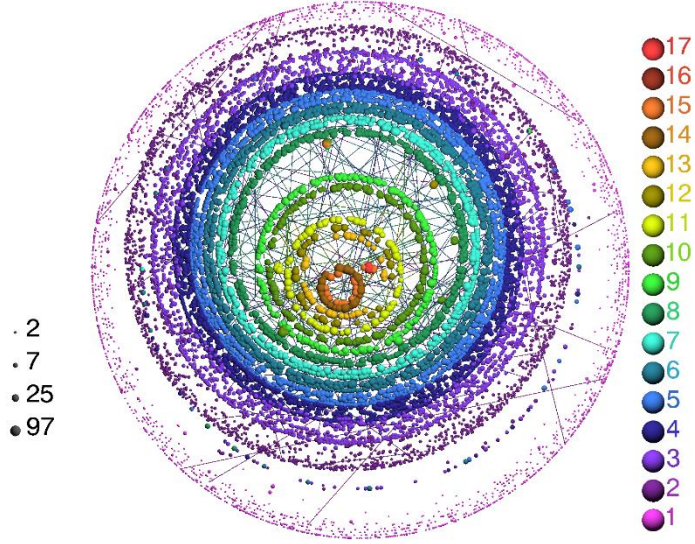
řekil 1.18. Bilimsel iřbirliđi ađına ait derece dađılım grafiđi [3].

řekil 1.19’da bilimsel iřbirliđi ađına rnek basit bir ađ grselleřtirmesi verilmiřtir. Ađ zerinde bulunan ok sayıda kmelenme mevcuttur. Bu farklı kmelenmeler farklı alanlarda faaliyet gsteren bilim adamlarını temsil etmektedir. Ađda izole dđmler ve dđm kmeleri bulunmaktadır [14].



řekil 1.19. Bilimsel iřbirliđi ađı grseli [14].

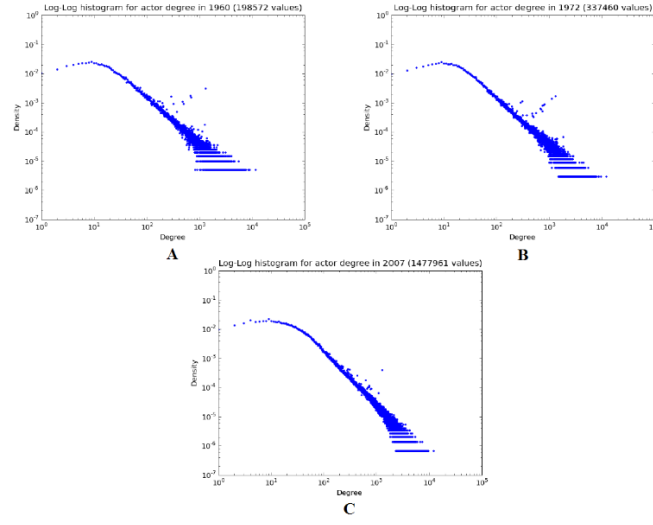
1995 ile 1998 yılları arasında fizik dalında yayın yapan bilim adamlarına ait bilimsel işbirliği ağının renklendirilmiş görseli Şekil 1.20’de verilmiştir [30].



Şekil 1.20. Bilimsel işbirliğine ait görselleştirme [30].

Film-Aktör İşbirliği Ağları

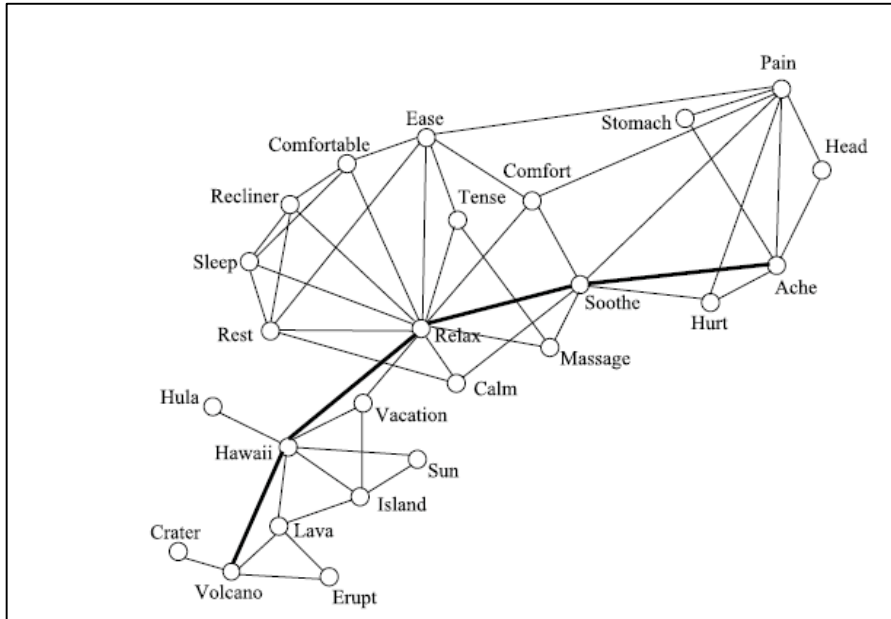
Analiz edilecek veri setinin nasıl elde edileceği kompleks ağlar ile alakalı yapılan çalışmalarda en büyük zorluktur. İnternet ortamında bulunan veri tabanlarında filmlere ait yapım yılı, yönetmen, oyuncular gibi verilere ulaşmanın kolay olması kompleks ağ türü olan film-aktör işbirliği ağları (Movie actor collaboration networks) üzerinde sıklıkla çalışma yapılmasına neden olmaktadır. Bu ağlarda aktörler düğümleri oluşturmaktadır. Aynı filmde yer alan aktörlerin birbirleri ile olan bağlantısı bu filmler olduğu için, filmler ilişkiyi oluşturmaktadır. Film-Aktör işbirliği ağlarında power-law’a uygun derece dağılımı gözlenmektedir [20]. Aktör ağının 1960, 1972 ve 2007 yılları için oluşturulmuş düğüm derece dağılımları Şekil 1.21’de verilmiştir.



Şekil 1.21. a) 1960, b) 1972 ve c) 2007 yıllarına ait film-aktör ağının derece dağılımı [3].

Kelime Ağları

Kelimeler konuşma dilini oluşturmaktadır. Bu kelime (dil) ağları da sosyal ağ türlerindedir. Her bir kelimenin bir düğümü oluşturduğu kelime ağlarında, kelimelerin arka arkaya gelmesi veya bir kelime uzaklıkta olması durumları da kelimeler arası linkleri tanımlamaktadır [10].



Şekil 1.22. Kelime ağı görseli [14].

Düğüm derece dağılımı kelime ağlarında power-law'a uygun olarak gerçekleşmiştir [20]. Volkan (Volcano) kelimesinin ardından gelen kelimelere olduğu kelime ağı görseli Şekil 1.22'de verilmiştir [14].

1.4.2.3. Biyolojik Ağlar

Çok geniş bir alana yayılan biyoloji bilimi birçok farklı disiplin ile ilişki içerisindedir. Kendi içerisinde biyolojik nesnelere ilişkilerinin anlaşılabilmesi ve diğer disiplinlerle olan ilişkilerinin incelenebilmesi için ağ yapıları kullanılmaktadır. Örnek olarak elementler arasındaki kimyasal reaksiyonların nasıl meydana geldiğini anlamak için moleküler biyoloji ağ yapıları kullanılmaktadır. Bir başka örnek olarak beyin hücrelerinin birbirlerine nasıl bağlandıklarını modellemek için nörologlar kompleks ağ yapılarını kullanmaktadırlar [19].

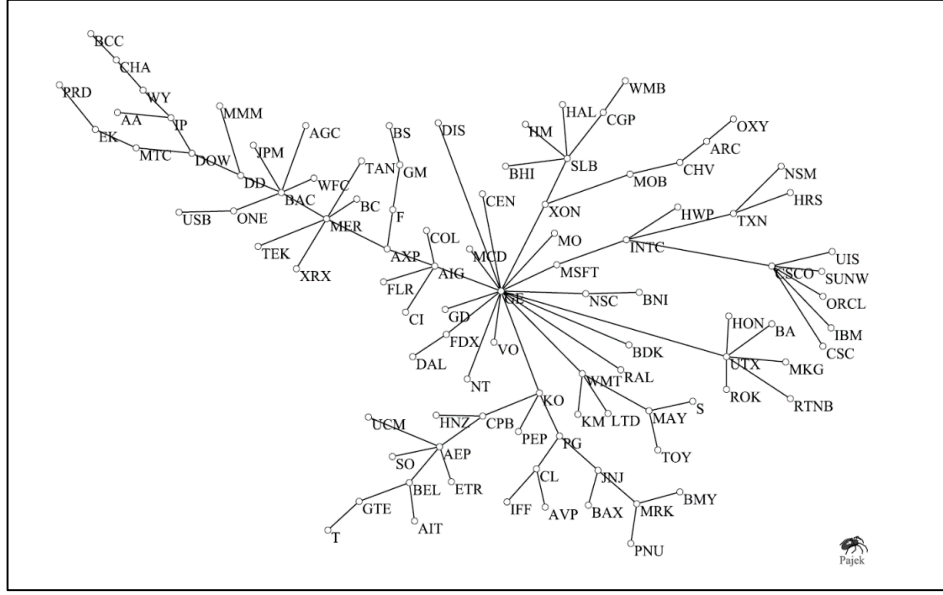
Günümüz biyoloji bilimi sadece elementlerin, moleküllerin ve hücrelerin fonksiyonları ile değil, bunların birbirlerine nasıl bir ağ ile bağlandığını incelemektedir. Bu sebepten dolayı biyolojik yapılardan oluşan ağın topolojisinin ve özelliklerinin de araştırılması gerekmektedir. Fakat biyolojik ağlar çok karmaşık yapıdadır. Bir hücre içerisinde bulunan binlerce molekül ve bu moleküllerin birbirleriyle ilişki halinde olması bu karışık yapının temel sebebidir [1].

Biyolojik ağ yapıları şu temel başlıklar altında incelenmektedirler [19];

1. Metabolik Ağlar (Metabolic networks)
2. Protein Etkileşim Ağları (Protein-protein interaction networks)
3. Genetik Ağlar (Genetic networks)
4. Beyin Ağı (Brain networks)

1.4.2.4. Finansal Ağlar

Finansal ağlar (*Financial networks*) mevcut finans sistemleri modellenerek tanımlanmaktadır. Genellikle şirketler arası ilişkileri tanımlamak için sosyal ağ özelliği gösteren finansal ağlar kullanılmaktadır.



Şekil 1.23. New York Borsası finans ağı görseli [14].

Hisse senedi işlemleri ve şirketler arasındaki ilişkiler modellenerek finansal ağlar meydana getirilmektedir. Örnek olarak hammadde üreten bir firma, bu hammaddeyi işleyen bir başka firma ve oluşan yeni ürünü pazarlayan bir diğer firma arasındaki ilişkiler kompleks bir ağ yapısı oluşturmaktadır. Bu kompleks ağ finansal ağlar ile açıklanmaktadır. Finans sisteminde yer alan bankalar, şirketler ve bireylerin birbirleri ile olan ilişkileri finansal ağlar yardımı ile tanımlanmaktadır [14].

1.4.2.5. Semantik Ağlar

Semantik ağlar, cümlelerin ve barındırdıkları kelimelerin anlamsal olarak analizinin yapıldığı ve ilişkilerin anlama dayalı olarak kurulduğu ağ yapılarıdır. Belirlenen kurallara bağlı olarak makine öğrenmesi ve yazılımı sonucu anlamlandırmanın oluşturulduğu bu ağ çeşidinde birden fazla yaklaşım bir arada kullanılabilir. Semantik ağlar ile daha detaylı açıklama Bölüm 3’de yapılmıştır.

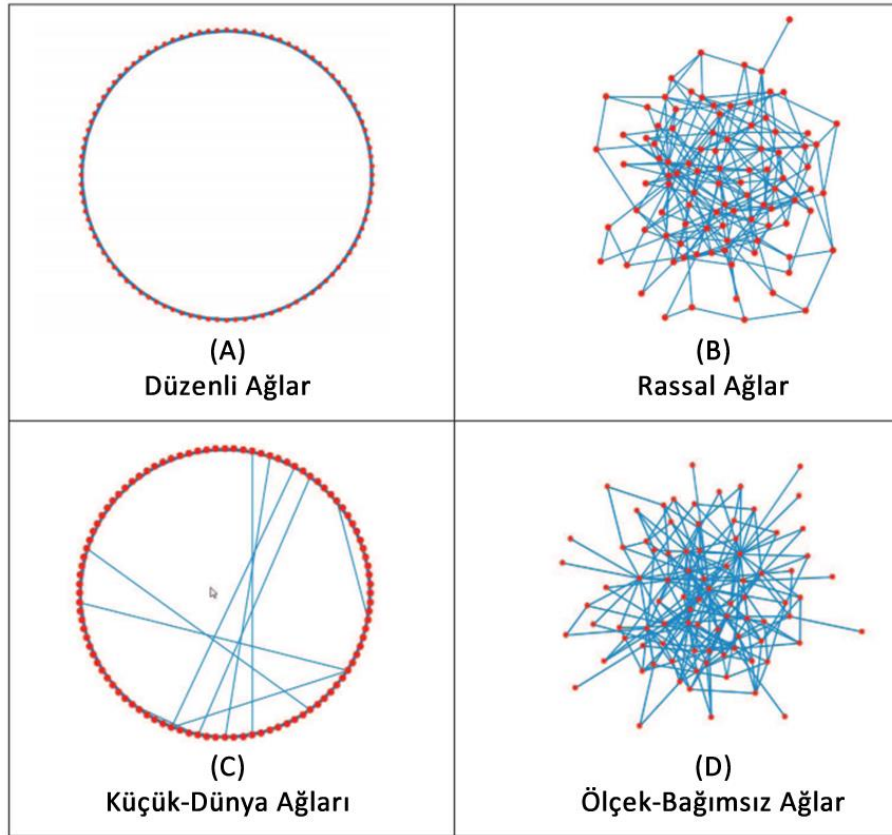
1.4.2.6. Diğer Gerçek Ağlar

Gerçek veya sanal ilişkileri barındıran çok sayıda gerçek ağ tanımlanabilmektedir. Büyük ölçekte veri içeren gerçek dünyada bulunan bazı kompleks ağ çeşitleri aşağıda verilmiştir [14,19].

1. Bilgi Ağları (Networks of information)
2. Jeofizik Ağları (Geophysical networks)
3. Ekolojik Ağlar (Ecological networks)

1.4.3. Sentetik (Model) Ağlar

Günlük hayatımızda var olan kompleks ağlara gerçel ağlar adı verilir. Bu ağları daha anlaşılır hale getirmek için üretilen yapay ağlara sentetik (model) ağlar denilmektedir. Bu yapay yapı, ağa ait parametrelerin çeşitli algoritmalar kullanılarak değiştirilmesi sonucu oluşur. Gerçel ağlara ait verilerin daha anlaşılır hale getirilmesi, model ağların üretilmesindeki temel amaçtır. Model ağlarda düğüm ve link sayısının istenildiği kadar tanımlanabilmesi ile istenilen büyüklükte yapay ağlar oluşturulabilir. Düğüm ve linkler arasındaki bağlanma ilişkileri çeşitli parametrelere bağlıdır [7].



Şekil 1.24. Üretilmiş ağ türleri [7].

En bilinen model ağ türleri şunlardır [1].

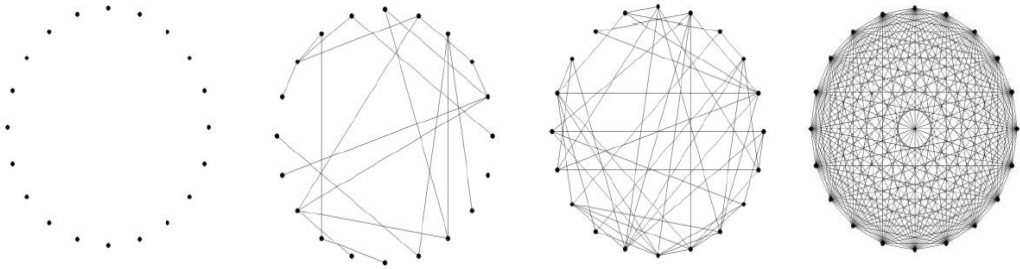
1. Rassal Ağlar (Random Networks)
2. Küçük-Dünya Ağları (Small-Worlds Networks)
3. Ölçek Bağımsız Ağlar (Scale-Free Networks)

En yaygın olan yapay ağ modellerine ait gösterimler Şekil 1.24'te verilmiştir. Parametreleri değiştirilerek model ağlar için farklı topolojilerde ağlar elde edilebilir.

1.4.3.1. Rassal (Random) Ağlar

Rassal ağlar ilk olarak 1959 yılında Macar matematikçiler Paul Erdős ve Alfred Renyi tarafından incelenmiştir [4]. Linklerin düğümler arasında rastgele olarak dağılması rassal veya diğer adıyla random ağları oluşturur. Rassal ağlarda, düğümler arasında oluşan linklere yönelik bir algoritma veya model bulunmadığından, hangi düğümün hangi düğümle bağlantı kuracağına rastgele kadar verilir. Erdős ve Renyi'nin çalışması kompleks ağ çalışmalarında sıklıkla kullanılmaktadır [20].

Erdős-Renyi modeline göre yapay bir ağ oluşturulurken n düğüm sayısını ve p düğümler arasındaki bağlanma olasılığını göstermek üzere, $0 < p < 1$ olması durumunda $n(n-1)p/2$ adet rastgele dağıtılmış linkden oluşan bir ağ elde edilir. Erdős-Renyi modelinde aynı sayıda düğümden oluşan ağlarda p 'nin değerine göre farklı link sayıları olabilir [8].



Şekil 1.25 .Farklı p değerlerine göre Erdős-Renyi modeli ağ görselleri [5].

Şekil 1.25’de 20 adet düğümden oluşan Erdos-Renyi rassal ağında p değerinin 0’dan (soldan) 1’e (sağa) doğru artırılması sonucunda oluşan ağlar görülmektedir [5].

1.4.3.2. Küçük-Dünya (Small-World) Ağları

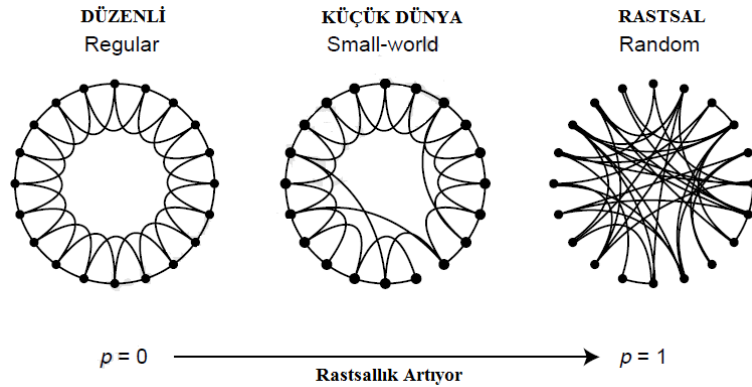
1967’de Stanley Milgram’ın yaptığı bilimsel araştırma Küçük-Dünya Ağları (Small-World Networks) model ağının temelini oluşturmaktadır. Milgram, rastgele seçilen iki insanın birbirleri arasında bağlantı olup olmasını ve birbirlerini tanıma ihtimallerinin ne olduğunu incelemiştir. Bu araştırmayı yaparken iki insanın birbiri ile direk bağlantısından farklı olarak ortak arkadaşlar üzerinden birbirleri ile olan dolaylı bağlantıyı tespit etmeye çalışmıştır. Bunun için iki insanın birbirinden ne kadar uzakta olduğu bilgisine ihtiyaç duyulmuştur. Uzaklık veya adım kavramı, iki kişi arasındaki bağlantıda, bu iki kişiden (düğümden) farklı olarak arada bulunan insan sayısıdır. Bir insandan (başlangıç düğümü) diğer insana (bitiş düğümü) ortalama kaç adımda varılacağı sorusuna odaklanılarak insanların birbirlerini tanıma ilişkileri ortaya konulmaya çalışılmıştır. Bugün kullanılan sosyal ağlara ait ilk örnek olarak Milgram tarafından yapılan çalışma gösterilebilir. Meşhur mektup deneyi Milgram tarafından gerçekleştirilmiştir. Bu deneyde başlangıç kişisine verdiği mektubu, tanıyorsa hedef kişisine, tanıyorsa hedef kişisini tanıma olasılığı en yüksek kişiye vermesi istenmiştir. Mektubun her geçtiği kişiye ait isimler mektuba not edilmiş ve hedefe ulaştığında mektubun geçtiği adımlar incelenmiştir. Sonuç olarak başlangıç kişisi ile hedef kişisi arasındaki ortalama uzaklığın altı (6) adım olduğu sonucuna varılmıştır [9].

Özetle küçük-dünya ağları, ağdaki düğümler arasında bulunan kısa yolun varlığının gösterilebiliyor olmasıdır. Bu durum, ağın boyutundan bağımsız olarak çoğu gerçel ağ için tanımlanabilmektedir. Ağ yapılarında düğümleri birbirine bağlayan link sayısı düğümler arasındaki en kısa mesafedir. Küçük dünya özelliği gösteren ağlarda iki düğüm arası ortalama uzaklığın 6 adım olduğu kabul edilir [10].

Milgram’ın çalışması, 2001 ve 2008 yıllarında bilgisayar teknolojisi kullanılarak iki farklı çalışmada test edilmiştir. 2001 yılında Duncan Watts tarafından yapılan çalışmada, mektuplar yerine insanlara birbirlerine paket taşıyan elektronik postalar

(e-mail) gönderilmesi istenilmiştir. 2008 yılında ise Jure Leskovec ve Eric Horvitz tarafından Microsoft .NET Messenger servisi kullanılarak benzer bir çalışma yapılmıştır. Her iki çalışmada da düğümler arasındaki ortalama uzaklığın yaklaşık 6 adım olduğu sonucuna ulaşılmıştır [3,31]. Teknolojik gelişmeler kullanılarak yapılmış olan bu analizler Milgram'ın çalışması doğrulamıştır.

Watts-Strogatz modeli: 1998 yılında Milgram'ın küçük dünya deneyi temel alınarak Watts ve Strogatz tarafından oluşturulmuş bir ağ modeli olup gerçel ağların düşük düğümler arası uzaklıklarını taklit edebildiği için small-world model olarak da adlandırılmaktadır. Rastgele seçilmiş iki düğümün birbirine, komşu düğümlerden oluşan daha kısa bir yol ile bağlanabileceğini söylenir [22]. Watts-Strogatz ağında, düzenli bir ağda rastgele seçilen az sayıda link koparılarak kaynak düğüm sabit kalacak şekilde hedef düğümü tamamen rassal olarak yeniden belirlenir [7]. Model ağda rassallığı sağlamak için 0 ile 1 arasında bir Prewire (yeniden bağlanma olasılığı) değeri tanımlanmıştır. Bu Prewire değeri düşük tutularak ağın kümelenme özelliği de korunmaya çalışılmış, böylece hem yüksek kümelenme, hem de düşük düğümler arası uzaklıklara sahip bir ağ modeli elde edilmiştir [10].



Şekil 1.26. Watts-Strogatz ağının p değerinin artışına göre görünümü [32].

Şekil 1.26'da gösterildiği gibi p değeri 0'dan 1'e doğru arttıkça ağ düzenli ağdan rassal ağa doğru geçiş yapmaktadır.

1.4.3.3. Ölçek Bağımsız (Scale-Free) Ağlar

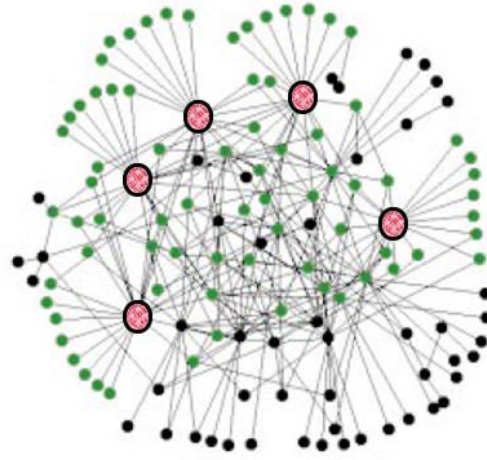
Model ağlar içerisinde, gerçek dünyadan üretilen ağlara en çok benzeyen yapay ağ modeli ölçek bağımsız ağlardır [7]. Watts-Strogatz modeli gerçek dünya ağlarının tüm özelliklerini yansıtamamaktadır. Bu nedenle 1999 yılında Albert-Laszlo Barabasi ve Reka Albert tarafından gerçel ağların anlaşılabilirliğini yükseltmek için, ölçek-bağımsız (*scale-free*) ağlar kavramı ortaya çıkarılmıştır [9]. Rassal ve küçük dünya ağlarından farklı olarak, ölçek bağımsız ağlarda, bir düğüm ağa eklendiğinde mevcut düğümlerden birine rastgele bağlanmayı değil popülaritesi yüksek hub olarak isimlendirilen düğümlere bağlanmayı (link oluşturmayı) daha çok tercih ederler. Bu durum, ağ içerisinde popülaritesi yüksek hub düğümlerinin sayısının azlığına, popülaritesi düşük düğüm sayısının da çok olmasına yol açmaktadır. Ölçek bağımsız ağlarda yayılım hızlıdır. Örnek olarak veri ve haber yayılımları gösterilebilir. Yayılım yapan şey, istenen ya da istenmeyen bir olgu olabilir. İstenilmeyen durum olarak hastalıkları örnek verebiliriz. Bir virüs eğer yüksek dereceli bir hub'a bulaşırsa, ağın geneline daha kolay yayılacaktır [10]. Ölçek bağımsız ağlar; sabit ölçek bağımsız ağ ve gelişen ölçek bağımsız ağ olarak iki başlığa ayrılmaktadır. Sürekli değişim ve gelişim içinde olmayan ağlar sabit ölçek bağımsız ağ (static scale-free network), dinamik değişim gerçekleşen ağlara ise gelişen ölçek bağımsız ağ (evolving scale-free network) adı verilir [1].

Barabasi-Albert Modeli: Bu modelde, küçük-dünya ve rassal ağlardaki ağların topolojilerinin modellenmesi yaklaşımından farklı olarak, gerçel ağların oluşum ve büyüme süreçlerinin incelenmesi, aşamalarının gözden geçirilmesi ve modellenmesi gerektiği fikri önerilmiştir. Bir kompleks ağ modellenirken ağın özelliklerinin iyi bilinmesi gerekmektedir. Gerçel dünyada yer alan ve sürekli düğüm ve linklerin eklenip çıkarıldığı dinamik bir yapıya sahip olan ağlarda, bu ağın değişimine etki eden etkenler bulunabilirse model ağ daha doğru bir topoloji ile gerçekleştirilebilecektir [9]. Barabasi-Albert modelinde, gerçel ağlarda olduğu gibi power-law düğüm derece dağılımı olan ağlar oluşturulması amaçlanmaktadır [1].

Barabasi-Albert modelinde, ortaya konulan iki temel özellik vardır; bunlar büyüme ve tercihli bağlantı yaklaşımlarıdır [5].

Büyüme (growth): Ağ sistemlerine zaman içinde yeni düğümler eklenerek ağın genişlemesine büyüme denilir. Az sayıda düğüm ile başlayan ve zamanla yeni düğümlerin link oluşturması ile büyüyen ağlar, müdahale edilmediği sürece büyümeye devam eder. WWW ağı sürekli büyüyen ağlara örnek olarak verilebilir. Bu sürekli büyüme, her yeni web sayfasının yayına girmesi ile gerçekleşmektedir [16]. Bilimsel işbirliği ağları da zaman içerisinde hızla büyüyen ağlardandır [8].

Tercihli bağlantı (preferential attachment): Ağa yeni eklenen düğümün, mevcut bulunan düğümlerden hangisine bağlanacağını tanımlayan kavrama tercihli bağlantı denilir. Yeni eklenen düğümün, popüler bir düğüme bağlanma ihtimalinin daha yüksek olması durumuna tercihli bağlantı denir. Popülerite kavramı düğümün derecesi ile doğru orantılıdır. Örnek olarak bilimsel atıf ağlarında çok atıf alan ve dolayısı ile derecesi yüksek bir yayının atıf alma olasılığı, ağa yeni giren ve dolayısı ile derecesi düşük olan bir yayının atıf alma olasılığından yüksektir [10].



**Ölçek Bağımsız Ağ
(scale-free)**

Şekil 1.27. Ölçek bağımsız ağ modeli görseli [33].

Şekil 1.27’de gösterilen ölçek bağımsız ağ örneğinde, düğüm sayısı 130 , link sayısı 215’dir. Derecesi yüksek olan popüler düğümler diğerlerinden daha büyük olarak gösterilmiştir. Büyük düğümler en popüler hub noktalarıdır [33].

BÖLÜM 2

VERİ KAYNAĞI OLARAK SOSYAL AĞLAR

İlk insan ile başlayan sosyal ağ kavramı ilk olarak 1954 yılında Barnes tarafından dile getirilmiştir. Barnes, sosyal ağ kavramını bireyin çevresindeki diğer insanlarla olan ilişkilerini tanımlamak için kullanmıştır [34]. Bireyin dahil olduğu topluluklar büyüdükçe sosyal iletişim ve birliktelik değişmiş, mevcut duruma ayak uydurmuştur. Teknolojinin gelişmesi ile bu sosyal iletişim değişmeye başlamış, internet ile başka bir formata bürünmüştür. Teknoloji ve internet bugün hayatımızın her anında kullandığımız kavramlar haline gelmiştir. İnternet ile birlikte başlayan “sanal” kavramı, sosyal ağları sanal sosyal ağ kavramına dönüştürmüş ve ilk başlarda insanlar çekimser davransa da zamanla ve sosyal ağların yaygınlaşması ve bunları kullanan insanların sayısının artması ile bu sanal ortama adım atmışlardır. İlk başlarda sadece bir fikir almak veya sohbet etmek şeklinde başlayan bu eylemler her geçen gün hayatımızdaki tüm gerçeklikleri barındırmaya başlamıştır. Artık yaptığımız tatiller, yediğimiz yemekler, doğum ve vefat gibi haberlerimiz, iş toplantıları, ders notları vb gibi tüm gerçek yaşantımız bu sanal sosyal ağlar üzerinden yürümektedir. İnsanların iletişim aracı olarak bu sosyal ağları ilk sırada görmesi ve işlerin ağırlıklı olarak bu ağlar üzerinden yürütülmesi, diğer insanları da bu ağları kullanmaya mecbur kılmaktadır.

2004 yılında kurulan Facebook ile sosyal ağ kavramı dünya çapında hızla yayılmaya başlamıştır. Dünyada en çok kullanılan sosyal ağ sitesi haline gelen Facebook 2016 yılının ilk çeyreğinde 1,6 milyarın üstünde kullanıcıya erişmiştir [35]. Türkiye 32 milyonun üzerinde kullanıcı sayısı ile dünyada 7. Sıradadır [36]. Facebook ile başlayan bu sosyal ağ kullanımı patlaması beraberinde başka fikirler ortaya çıkarmıştır. Facebook ardından en çok kullanılan sosyal ağ olan Twitter bu fikirlerin başındadır. 2006 yılında kurulan Twitter 140 karakterlik metin paylaşım mantığı ile ortaya çıkmıştır. Kişiler o anki görüşlerini türkçede cıvılda manasına gelen

Tweetlerle paylaşmakta ve bu Tweetlere hashtag ekleyerek paylaşmaktadırlar. 2011 yılında Tweet metinleri türkçe dil desteği sağlamaya başlamıştır. 2006 yılının verilerine göre 320 milyon aylık aktif kullanıcıya sahiptir [37]. Türkiyede Twitter kullanımını türkçe desteği ile artışa geçmiştir. Sosyal ağlardaki her geçen gün artan kullanıcı sayısı ve bu kullanıcıların oluşturduğu büyük miktardaki veriler çeşitli analizlere imkan sağlamaktadır.

Kompleks ağ analiz çalışmalarının konu aldığı gerçel ağlar genellikle “self-organization” olarak da bilinen, düğümlerin bağlanma tercihleri ile şekillenen yapıya sahiptirler. Ancak düğümlerin bağlanma tercihleri dışındaki etkenlerle şekillenen gerçel ağlar da vardır. Örneğin aktör ağları, aktörlerin filmlerde birlikte çalışması ile şekillenen ağlardır. Filmde yer alan aktör kadrosunun belirlenmesinde aktörlerin etkisi olsa da “cast director” tercihleri de etkin rol oynamaktadır. Ya da power-grid ağlarında trafo yerleşimlerinde trafoların tercihleri söz konusu olmayıp özellikle coğrafi ve demografik yapının yol açtığı enerji ihtiyaçları haritası, ağın şekillenmesinde önemli pay sahibidir.

Bu açıdan bakıldığında düğümlerin “self-organization” tercihlerinin baskın olduğu ağ türlerinin başında bilimsel işbirliği ağları, sosyal ağlar, telefon ağları gibi türler gelmektedir. Bu ağ örneklerinde link oluşum süreçleri, düğümlerin kendi tercihleri ile şekillenmektedir. Bunun sonucu olarak da”scale-free”,”clustering” ve “small-world” özellikleri daha saf olarak gözlenebilmektedir.

Teknolojik ağ türleri “self-organization” olgusunun baskın olduğu bir ana başlıktır. Twitter sosyal ağı üzerinde gerçekleştirmiş olduğumuz etiket (hashtag) birliktelik ağı çalışması, etiket birlikteliğini doğuran süreçlerin doğallığı ve kullanıcı tercihleri dışındaki etkenlere kapalı oluşu nedeniyle iyi bir “self-organization” örneğidir. Aynı zamanda birlikte kullanılan etiketlerin tanımlandığı linklerin semantik doğruluğunun da Twitter’ın sağladığı geniş veri seti sayesinde sorgulanabilir olması, çalışmaya derinlik katmapotansiyeli olan bir olgudur.

Türkiye’de 10 milyona yaklaşan kullanıcı sayısı ile Twitter, günde 8 milyon Tweet verisine sahiptir [38]. Bu rakamlara bakıldığında Twitter’ın veri seti sağlamak

açısından büyük bir hazine olduğu görülmektedir. Bu verilerin elde edilmesi için çeşitli yöntemler kullanılmaktadır. Bir sonraki bölümde bu yöntemler hakkında bilgi verilecektir.

2.1. YÖNTEMLER

Bu tez çalışması kapsamında Twitter sosyal ağından sağlanan veriler üzerinde etiket ağ analizi yapılmış, aynı etiketlerin semantik yakınlıklarının tanımladığı bir diğer ağ çalışması ile etiket birlikteliği ağ çıktıları karşılaştırılmıştır. Bu çalışmada veri kaynağımız olan Twitter ortamından veri sağlamak için kullanılan Twitter Streaming API, diğer veri sağlama yöntemleri ile birlikte bu bölümde ele alınmıştır.

2.1.1. Twitter API

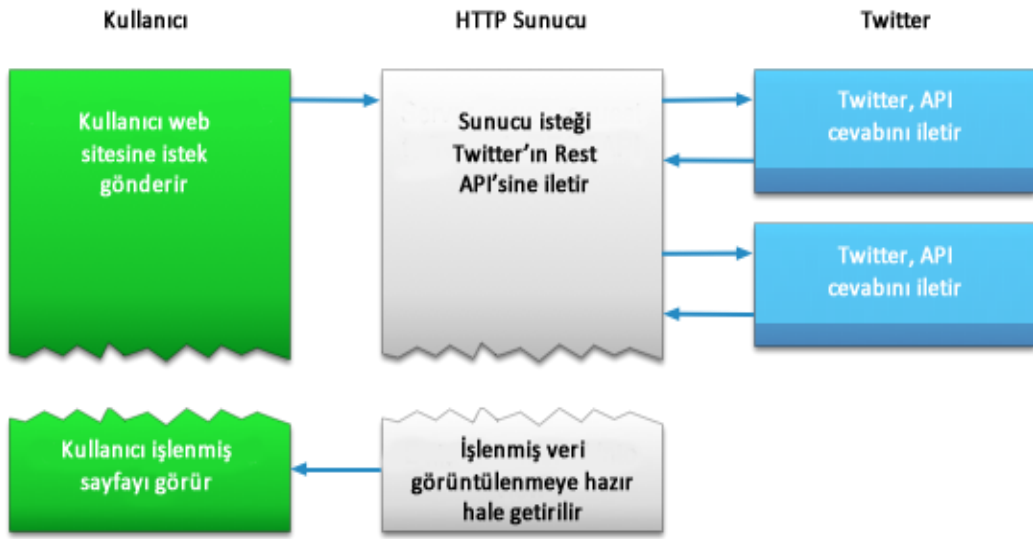
Twitter, barındırdığı veri setine API'ler aracılığıyla kısıtlı ölçüde erişime izin vermektedir. Twitter API; farklı özelliklere sahip üç farklı API'den oluşmaktadır. Bunlar Search API, REST API ve Streaming API'dir [39]. Bu API'ler devamlı geliştirilmekte ve güncellenmektedir. Yapılacak çalışmalara yönelik olarak Twitter'ı bir veri kaynağı olarak kullanmamızı sağlayan bu API'ler geliştiricilere veri merkezli uygulamalar geliştirme imkanı sağlamaktadır.

2.1.1.1. Search API

Belirlenmiş bir arama sorgusuna yönelik olarak Twitter içeriğine ulaşma ihtiyacı duyan geliştiriciler için tasarlanmıştır. Bu arama sorgusu; belirli anahtar kelimeler, belirli bir kullanıcıya ait Tweetler veya belirli bir kullanıcıdan bahseden Tweetler gibi sorgular olabilir. Arama kriteri belirlendiği için sonuçlar kısıtlanacak ve dolayısıyla sonuçlarda eksikler olabilecektir. Bütün sonuçların eksiksiz dönmesi istenilirse Streaming API kullanılmalıdır. Search API, REST API v1.1'in bir parçasıdır [40].

2.1.1.2. REST API

REST API zaman çizelgeleri, kullanıcı bilgileri, durum güncellemeleri gibi Twitter veri setinde mevcut olan daha özel bilgilere ulaşmak için kullanılmaktadır. Ayrıca REST API ile kullanıcı Tweet oluşturma, gönderme, favori ekleme ve cevap yazma gibi Twitter ile etkileşimli işlemlerde de kullanılabilir. REST API akış şeması Şekil 2.1'de gösterilmektedir.



Şekil 2.1. REST API akış şeması [41].

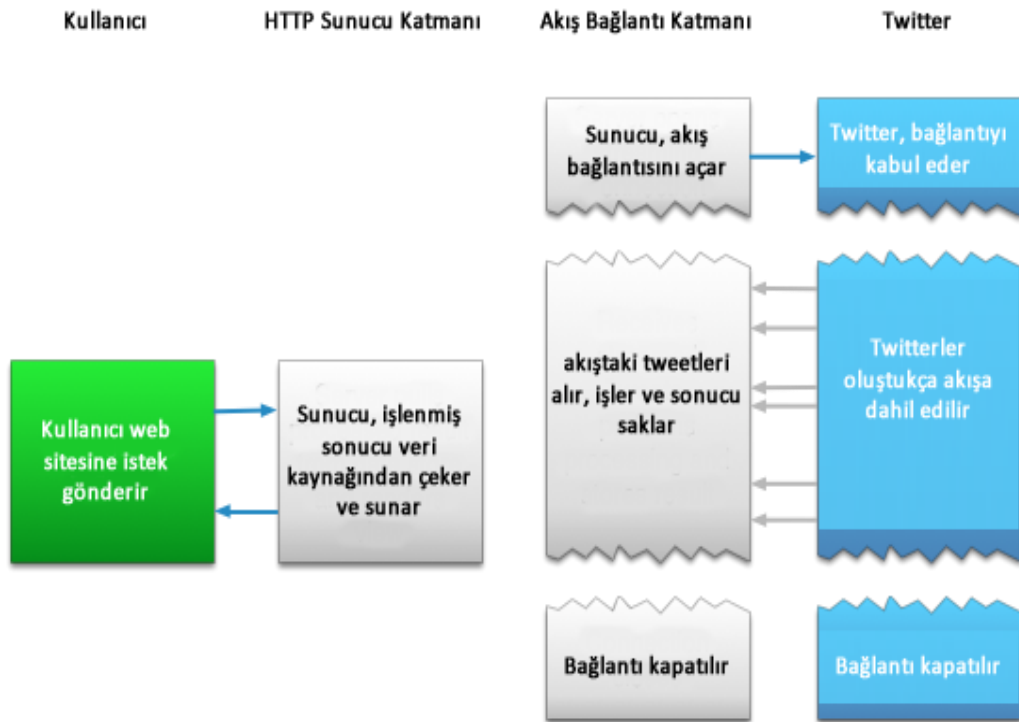
2.1.1.3. Streaming API

Twitter veri akışına anlık olarak erişim imkanı sağlar. Çok fazla veriye ihtiyaç duyan geliştiriciler için kullanışlıdır. Arama kriterlerini etkin biçimde kullanabildiğimiz Search API'ye göre Streaming API ile daha fazla veri elde edebiliriz. Streaming API'de arama kriterlerini sadece belli kelime veya kullanıcı bazında sınırlandırabiliriz.

Twitter üzerinde verilere eş zamanlı erişim Firehose olarak isimlendirilmektedir. Twitter Streaming API aracılığıyla bu Firehose erişiminin sadece %1 lik dilimini geliştiricilere ücretsiz olarak sunmaktadır [42]. %1 lik dilimden daha fazla veri elde

etmek isteyen geliştiriciler bu veriler için ücret ödemek zorundadırlar. Ancak veri hacmi çok yüksek olduğu için %1'lik dilim arařtırmalar için önemli bir kaynaktır.

Streaming API akıř řeması řekil 2.2'de görölmektedir. Kullanıcı istemde bulunur ve bu istem sunucu tarafından deęerlendirilir. Bu iřlem için sürekli açık bir http baęlantısı olması gerekmektedir. Baęlantı kurulduktan sonra kullanıcı istemi neticesinde Tweet'ler bu isteme cevap olarak sunucu tarafından kullanıcıya aktarılır. Bu döngü kurulmuř olan baęlantı sonlandırılana kadar devam eder.

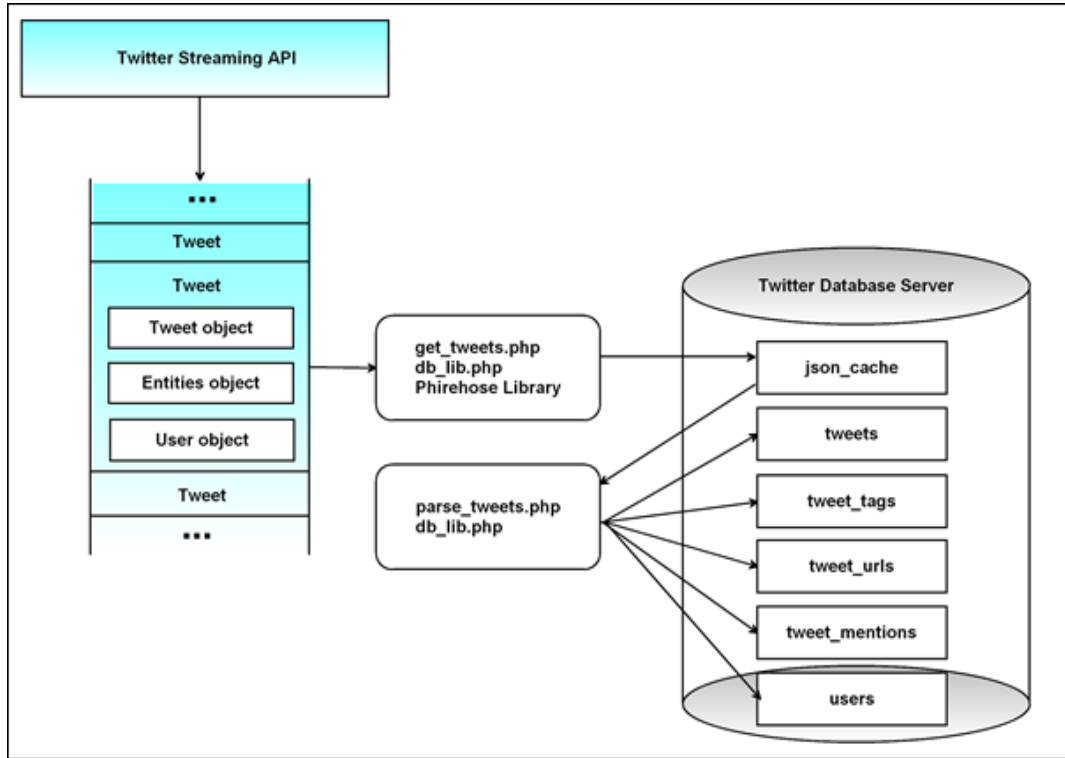


řekil 2.2. Streaming API akıř řeması [41].

2.1.2. 140dev Streaming API Yapısı

140dev Streaming API Adam Green tarafından oluřturulmuř ve GPL lisansı ile ücretsiz yayınlanan bir kütüphanedir. Twitter Streaming API'ye kolay ve etkili biçimde, basit bir arayüz ile baęlantı saęlamayı amaçlamaktadır.

140dev, PHP ve Javascript dilleri kullanılarak geliştirilmiştir. Depolama için MySQL veritabanını kullanmaktadır. Veritabanı üzerindeki veriler etkileşim içindedir ve modüller arası etkileşim veritabanı üzerinden gerçekleşir. Bu sayede bir başka programlama dilinde yazılmış MySQL veritabanına sahip modüller bu yapıya dahil edilebilirler [43]. Şekil 2.3’de Adam Green tarafından geliştirilen 140dev yapısının çalışma şekli görülmektedir.



Şekil 2.3. 140dev yapısı [44].

2.1.3. Phirehose Kütüphanesi

140dev yapısı ile birlikte yüklenmiş olan Phirehose kütüphanesi, Twitter Streaming API ile Tweet alma işleminin büyük kısmını gerçekleştirir. Phirehose Fenn Bailey tarafından geliştirilmiş güncel sürümünün 140dev yapısı ile birlikte çalışmasına izin verilmiştir. Phirehose, 140dev yapısından ayrı olarak açık kaynak kodlu bir projedir. Twitter Streaming API, eşzamanlı veri alma sırasında sürekli açık bir bağlantıya ihtiyaç duymaktadır. Phirehose, bağlantıyı kurma, veri alma ve bağlantının kopması durumlarında tekrar bağlanma gibi işlerin hepsini gerçekleştirmektedir. Ayrıca veri

toplamak için gerekli olan kullanıcı id ve arama kelimelerinin otomatik güncellenmesi işlemleri de Phirehose tarafından gerçekleştirilir [45].

Çalışma kapsamında büyük veri hazinesine sahip olduğundan dolayı veri kaynağı olarak Twitter mikroblog sitesi, veri sağlama yöntemi olarak ise veri akışına anlık erişim imkanı sağlamasından dolayı Streamin API kullanılmıştır. 140dev Streaming API ve bağlantıyı sürekli açık tutan Phirehose kütüphanesi açık kaynak kodlu olmasından dolayı veri sağlama yönteminde tercih edilmiştir.

BÖLÜM 3

SEMANTİK

İnsanlık tarihi boyunca bilgi kavramı yaşamın devamı ve gelişimi için en önemli kavram olmuştur. Bilgiye ulaşmaya çalışırken verileri değiştiren insanoğlu, ulaştığı her yeni bilgiden başka bilgiler çıkarma yollarını aramıştır. Bu arayış neticesinde nitel ve nicel gözlem ve çalışmaların ürünü olarak elde edilen veriye bilgi denilmektedir. Bu bilgiye ulaşım dolaylı veya dolaysız olabilir. Yapılan analizin niteliği, bilginin soyutlama seviyesini belirler. Semantik analiz, doğrudan ölçülebilen bilgi niteliklerinden dolayı, daha üst seviyede yer alan bilgi niteliklerine erişilmesini sağlayan yöntemlerdir ve semantik analiz sonucunda erişilen bilgi de semantik bilgidir [46]. Öğrenmenin en önemli tarafı, hedef tarafından bilginin değerlendirilmesi olarak kabul edilebilir. İnsan yapısından kaynaklanan subjektiflik ve sınırlı veri işleme sorunlarından dolayı objektif değerlendirme ve büyük verilerin işlenmesinde bilgisayar kaynaklı öğrenme tercih edilmektedir [47]. Ancak bilgisayarın sentetik yapıya sahip olmasından dolayı kelimeler kurallar bütünü içerisinde verilmelidir [48].

3.1. BİLGİ KAVRAMI

Teknolojideki hızlı gelişim ile birlikte düne kıyasla bugün daha fazla bilgi elde etmekteyiz. Bu bilgi yığınının organize edilmesi ve işlenebilmesi her geçen gün daha zor hale gelmektedir. Bilginin ham haline ve yapılmış veya yapılacak işlemlerin belirli bir formatta kayıt altına alınmış haline veri denilmektedir. Genellikle tek başına bir anlamı olmayan bu veriler, birlikte kullanıldıklarında veya bir işlemde geçirildiklerinde anlam kazanarak bilgi adını almaktadırlar. Yani veriler analiz edilerek yorumlandığında bilgi haline dönüşürler. Şekil 3.1’de veri ve bilgi çevrimi görülmektedir. Analiz ve işlem basamağı, bilgisayar yazılımları, metin madenciliği, önerme çalışmaları gibi birçok adımı barındırmaktadır.



Şekil 3.1. Veri ve bilgi çevrimi.

Mevcut bulunan veriler kullanılarak, kullanışlı bilgiye ulaşılmasına veri madenciliği denilmektedir. Metin madenciliği ise, bilgisayar tarafından yazılmış dokümanlarda bulunan verilerden bilinmeyen bilgiye ulaşılmasıdır. Veri madenciliğinde veriler veritabanlarında düzgün bir formatta tutulurken, metin madenciliğinde genellikle doğal dil kalıplarının içinde bulunmaktadır [49].

3.2. SEMANTİK KAVRAMI

Semantik (anlambilim) ögenin içindeki veya ögeler arasındaki anlamları inceler. Ögeler arasındaki anlamlandırma üç seviyede gerçekleşebilir [50]:

1. Sözdizimsel (syntactic): Tamamen otomatik olarak çalışan makineye özgü anlamlandırma seviyesidir. Karakterlerin sıralamasından ve bir araya gelişlerinden anahtar oluşturularak bu anahtarlar yardımı ile anlamlandırma gerçekleşir.
2. Anlamsal (semantic): Sadece makine değil insanında etkin olduğu anlamlandırma seviyesidir. İnsan tarafından belirlenen kurallara bağlı olarak makine tarafından değerlendirilir.
3. İşlevsel (pragmatic): Anlam net değildir. Ne anlatılmak istendiğinin bilinmesi gerekir. İnsan yapısı anlamlandırabilir. Bundan dolayı makine tarafında bir model tasarlanamamıştır.

Her geçen gün daha fazla alanda semantik yakınlık kullanılmaktadır. Bu sayede verilerin ve bilgilerin birbirleri ile yakınlığı tespit edilerek daha verimli ve işe yarar bilgi çıkarımları elde edilebilmektedir. Örnek olarak, “Türkiye Cumhuriyeti Cumhurbaşkanı” ile “Recep Tayyip Erdoğan” arasında semantik bir yakınlık vardır.

Semantik yakınlığın çoğunlukla kullanıldığı alanlar şunlardır [51]:

1. Doğal dil işleme
2. Bilgi elde etme
3. Kelime anlamı belirsizliklerin giderilmesi
4. Metin parçalamaları
5. Soru cevaplama
6. Öneri sistemleri
7. Bilgi çıkarımı

Kelimeler arasındaki semantik yakınlık çalışmaları doğal dil işleme ve bilgi çıkarım çalışmalarının bir parçası haline gelmiştir [52]. Kelimelerin anlamlarının değişmesi, değişen ve gelişen hayat ile birlikte kelimelere yeni anlamlar yüklenmesi bu semantik çalışmaların önemini artırmaktadır.

3.3. SEMANTİK AĞLAR

Düğüm ve linklerden oluşan ağlarda bu düğümler birbirleri ile yönlü veya yönsüz ilişki içinde bulunurlar. Bir ağda bulunan düğümlerin birbirleri ile olan anlamsal ilişkilerinin incelenmesine semantik ağ, birbirleri ile olan yakınlık seviyelerine ise semantik yakınlık ismi verilir. Semantik ağlar yönlü veya yönsüz olarak kurulabilirler.

En çok bilinen semantik ağ çeşitleri şunlardır [53]:

1. Tanımsal (Definitional) Ağlar: Tanımsal ağlar, alt anlam veya kavramlar ile alt anlamlar arasındaki ilişkileri inceler. Bu alt anlamların oluşturduğu genel anlamın özelliklerinin hiyerarşik olarak saklandığı ve genellendiği ağlardır. Bu ağlarda, bilgiler genellikle doğru olarak kabul edilir.
2. İddiasal (Assertional) Ağlar: İddiasal ağlar, önermelere yönelik olarak tasarlanmıştır. Tanımının aksine, iddiasal ağlarda, bir operatör tarafından belirtilmediği sürece bilginin bir ihtimal doğru olduğu varsayılır.

3. Sezdirimsel (implicational) Ağlar: Sezdirimsel ağlarda, düğümleri bağlamak için birincil olarak imalar kullanılır. Bunlar inanışlar, nedensellik ve sonuç kavramlarını içermektedir.
4. Çalıştırılabilir (Executable) Ağlar: Çalıştırılabilir ağlar, işaretleyici veya çıkarım yapmak, mesaj aktarmak, kalıplar ve birlikteliklerini aramak gibi bağlı prosedürleri içermektedir.
5. Öğrenme (Learning) Ağları: Öğrenme ağları, örneklerden bilgi edinme yolu ile ağı oluşturur veya mevcut ağları genişletir. Yeni bilgiler varolan düğümleri ve linkleri silerek veya yeni düğümler ve linkler ekleyerek ağı değiştirir. Yeni bilgiler sayısal değerleri değiştirerek düğümlerin ve ilişkilerin ağırlıklarını değiştirerek de ağ yapısında değişimlere neden olurlar.
6. Hibrid (hybrid) Ağlar: Hibrid ağlar, önceki tekniklerden 2 veya daha fazlasını bir ağda çalıştırarak etkileşimli ağlar ortaya çıkarır. Genellikle çalışmalarda tanımsal, çalıştırılabilir ve öğrenme ağları birlikte kullanıldığından bu semantik ağlar hibrid ağlar ismi ile anılır.

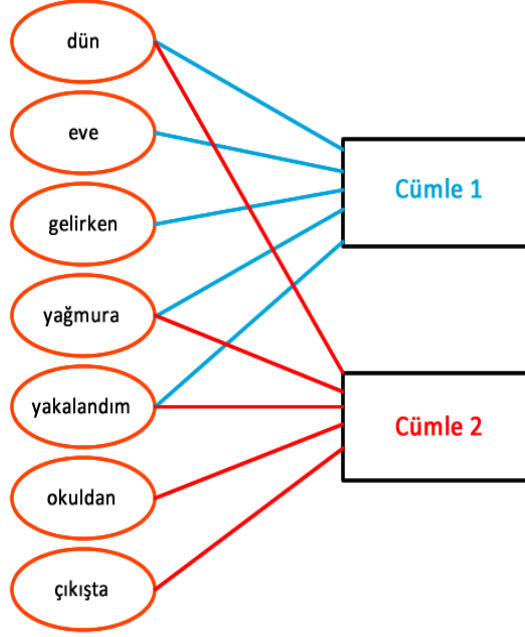
Çizelge 3.1. Semantik benzer cümleler.

| | |
|---------|---|
| Cümle 1 | Dün eve gelirken yağmura yakalandım. |
| Cümle 2 | Dün okuldan çıkışta yağmura yakalandım. |

Çizelge 3.1’de semantik yaklaşıma örnek olarak iki cümle verilmiştir. İlk cümle “dün,eve,gelirken,yağmura,yakalandım” kelimelerinden, ikinci cümle ise “dün,okuldan,çıkışta,yağmura,yakalandım” kelimelerinden oluşmaktadır. İki cümlenin de beşer kelimedenden oluştuğu görülmektedir. Bu beş kelimenin üç tanesinin her iki cümlede de ortak olarak kullanıldığı görüldüğüne göre, hiçbir ağırlıklandırma yapmadan ve kelime seti kullanmadan analiz edersek %60 oranında kelimelerin eşleştiği ve bu cümlelerin %60 oranında benzer olduğu söylenebilir. Toplam yedi adet kelimedenden oluşan düğüm tablomuzun olduğu bu ağda, ortak kullanılan 3 kelime olduğunu göz önünde bulundurursak 3/7 oranında bir eşleşme var gibi gözükmemektedir. Ancak düğüm tablomuzdaki her kelimenin ağırlığının da aynı olmadığını unutmamalıyız. Kelimelerin kullanım sayıları da hesaba katıldığında bu 3 kelimenin toplam ağırlığının 6, düğüm kümesinin ise 10 olduğu görülmektedir. 6/10

yani %60 eşleşme sonucuna ulaşılmaktadır. Semantik ilişkilendirmede, daha büyük kelime setleri ve fazla sayıda veriler ile daha doğru sonuçlara ulaşmak mümkündür.

Bu kurulan ağda kelimeler düğümleri, bu kelimelerin birbirleri ile olan bağı cümleler ise ilişkileri oluşturmaktadır. Şekil 3.2’de düğümler ve ilişkileri görülmektedir.

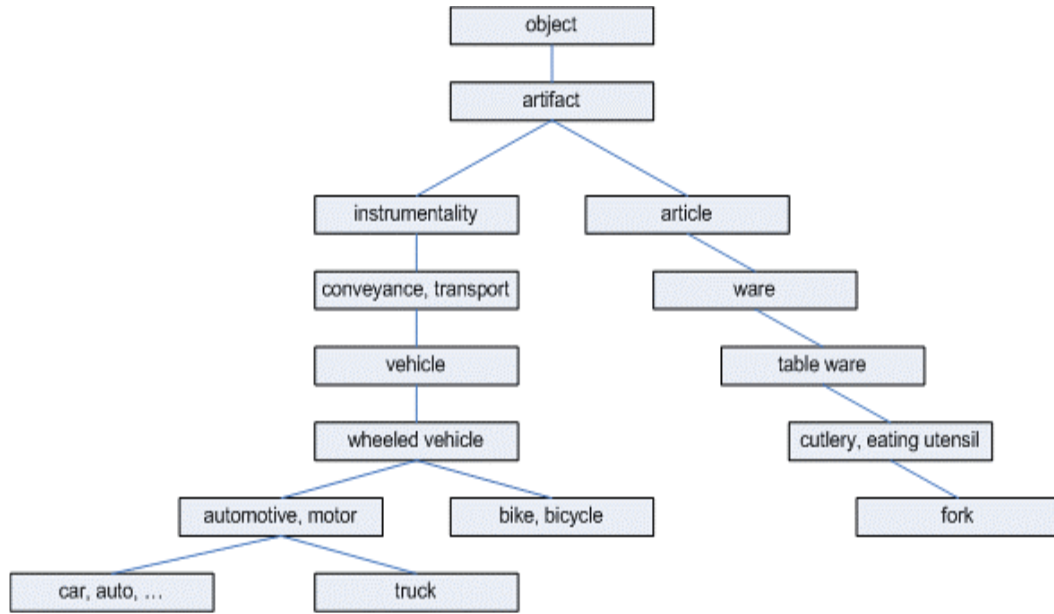


Şekil 3.2. İki cümleye ait düğümler ve ilişkileri.

Semantik analiz çalışması yapılmadan önce veri seti üzerinde önışleme çalışması yapılması gerekmektedir. Dao ve Simpson tarafından yapılan çalışmada, iki cümlenin semantik ilişkisini incelemek için izlenmesi gereken adımlar şu şekilde verilmiştir [54]:

1. İlk olarak cümleler parçalanır.
2. Belirsizlikler giderilir.
3. Kelime ekleri temizlenir.
4. Her kelime için en uygun anlam belirlenir.
5. Son olarak, kelimelerin anlamsal yakınlığına göre cümlelerin anlamsal yakınlıkları hesaplanır.

Bu aşamalar dikkate alınarak yapılacak bir semantik analiz çalışmasında öncelikle cümleler kelimelere bölünür ve ardından cümle içerisinde sık kullanılan fakat anlamsal olarak zayıf olan stopword'ler temizlenir. Belirsizliklerin giderilmesi aşamasında, kelimenin yapısının ne olduğu (isim, fiil, zarf.. vb.) belirlenir. Kelimeler eklerinden arındırılarak, kök haline getirilir. Kök haline getirilen kelimelerin sözlük anlamları üzerinden benzerlikleri ölçülür. Eş anlamlı kelimeler arasında ilişki semantik (anlamsal), kelimelerin anlamları arasındaki ilişki ise lexical (sözcüksel) olarak isimlendirilir. Şekil 3.3'de yapılan çalışmaya ait görsel görülmektedir. Bir düğümden diğer düğüme mesafe ne kadar kısa ise o iki düğüm birbiri ile o kadar alakalıdır [54].



Şekil 3.3. Semantik çalışmaya ait düğümlerin mesafeleri [54].

Tekerlekli araçların iki tekerli ve daha fazla tekerlerden oluşan iki düğüme ayrıldığı şekilde görülmektedir. “otomobil” ve “kamyon” bir düğüm ile birbirlerine bağlı iken, “otomobil” ve bisiklet” iki düğüm ile birbirlerine bağlıdır. Bundan yola çıkarak, “otomobil” “kamyon” ile “bisiklet” ten daha yakın ilişki içindedir denilebilir.

Mohler and Mihalcea, tarafından yapılan çalışmada Kuzey Texas Üniversitesi'nde 30 öğrencinin veri yapıları sınavında verdikleri cevaplar değerlendirilerek semantik ilişkiler hesaplanmış ve bilgisayara 1 ile 5 arasında otomatik puan verdirilmesi

sağlanmıştır. Çizelge3.2’de sorular ve öğrencilerin cevapları gösterilmektedir. İki kişi tarafından gerçekleştirilen değerlendirilme sonuçları da puan bölümünde ayrı ayrı gösterilmektedir.

Çizelge 3.2. Öğrencilerin verdikleri cevaplar [47].

| Sample questions, correct answers, and student answers | Grade |
|--|-------|
| <i>Question: What is the role of a prototype program in problem solving?</i> | |
| <i>Correct answer: To simulate the behavior of portions of the desired software product.</i> | |
| <i>Student answer 1: A prototype program is used in problem solving to collect data for the problem.</i> | 1, 2 |
| <i>Student answer 2: It simulates the behavior of portions of the desired software product.</i> | 5, 5 |
| <i>Student answer 3: To find problem and errors in a program before it is finalized.</i> | 2, 2 |
| <i>Question: What are the main advantages associated with object-oriented programming?</i> | |
| <i>Correct answer: Abstraction and reusability.</i> | |
| <i>Student answer 1: They make it easier to reuse and adapt previously written code and they separate complex programs into smaller, easier to understand classes.</i> | 5, 4 |
| <i>Student answer 2: Object oriented programming allows programmers to use an object with classes that can be changed and manipulated while not affecting the entire object at once.</i> | 1, 1 |
| <i>Student answer 3: Reusable components, Extensibility, Maintainability, it reduces large problems into smaller more manageable problems.</i> | 4, 4 |

Bilgisayar analizleri farklı algoritmalarla gerçekleştirilmiştir. Kullanılan başlıca algoritmalar LSA (örtülü semantik analiz) ve ESA (açık semantik analiz) algoritmalarıdır. Doğrudan değil dolaylı olarak anlamsal yakınlıklara bakılan LSA algoritması, soru başına 0,5099 korelasyon ve sınav kağıdı başına 0,6735 korelasyon ile en başarılı sonucu vermiştir[47]. Bu çalışma kelimelerin direk anlamlarının yanında örtülü anlamlarının da semantik çalışmalarda önemli bir etken olduğunu göstermiştir.

3.4. SEMANTİK WEB

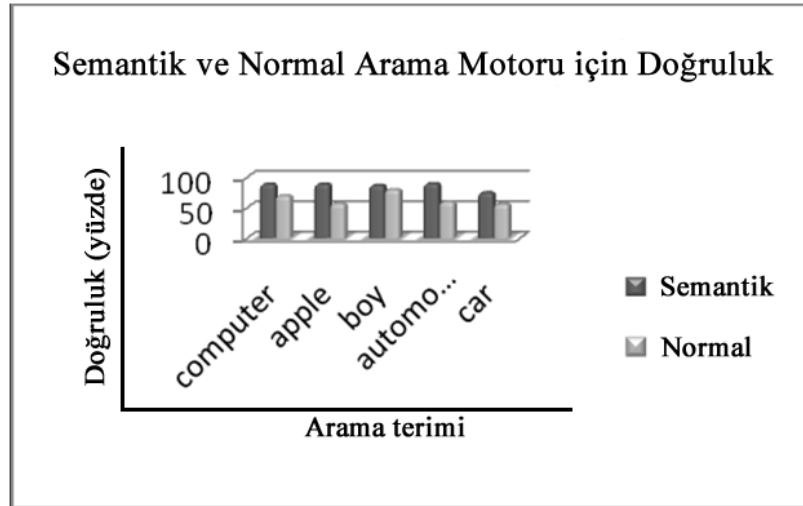
Bireylerin bilgiye ulaşmak için kullandıkları yöntemlerin başında bugün internet gelmektedir. Türkiye İstatistik Kurumu (TÜİK) tarafından yapılan araştırmaya göre 2007 yılında 16-74 yaş grubu bireylerin %30 civarı internet kullanırken, 2015 itibari ile bu oran %55’e yükselmiştir [55]. İnternet kullanım seviyesindeki bu artış, web tabanlı yazılımların kullanıcıya yönelik olmasının önemini arttırmıştır. İnsanların bilgiye ulaşmasını kolaylaştırmaya yönelik çalışmalara ağırlık verilmiştir. Dünyanın farklı noktalarındaki insanlar aynı anda aynı aramaları farklı yazımlarla ve dillerde yapmaktadırlar [56]. Anlamsal olarak aynı fakat yazımsal olarak farklı olan bu arama cümlelerinin bağdaştırılmasında semantik analizlere başvurulmaktadır. Kullanıcılar tarafından yapılan aramalarda kelimelerin anlamlarının yanında kullanımları önem

kazanmaktadır. Örnekle olarak “Blackberry” araması yapıldığında bu büyük ihtimalle meyveden ziyade, telefon markası olarak kastedilmektedir [57].

İnternet kullanıcılarının arama sorguları ile her geçen gün daha büyük bilgi topluluğu haline gelen www ağında bu sorgularda kullanılan kelimelerin semantik analizlerine yönelik çalışmalar artmaktadır. Semantik analize örnekle olarak, google arama motorunun “bunu mu demek istediniz” şeklindeki önermelerini gösterilebilir. Bu kavram semantik web olarak isimlendirilmektedir. Arama sorguları ile normalde bir ilişkisi olmayan bilgiler gizli olarak gözükme ve bu bilgilerin kullanıcılara gösterilebilmesi için semantik yakınlık kavramı önem kazanmaktadır [58].

Bu bağlamda arama motorlarında semantik analizler devreye sokulmuştur. En çok kullanılan semantik yakınlık ölçme algoritmaları şunlardır [48]:

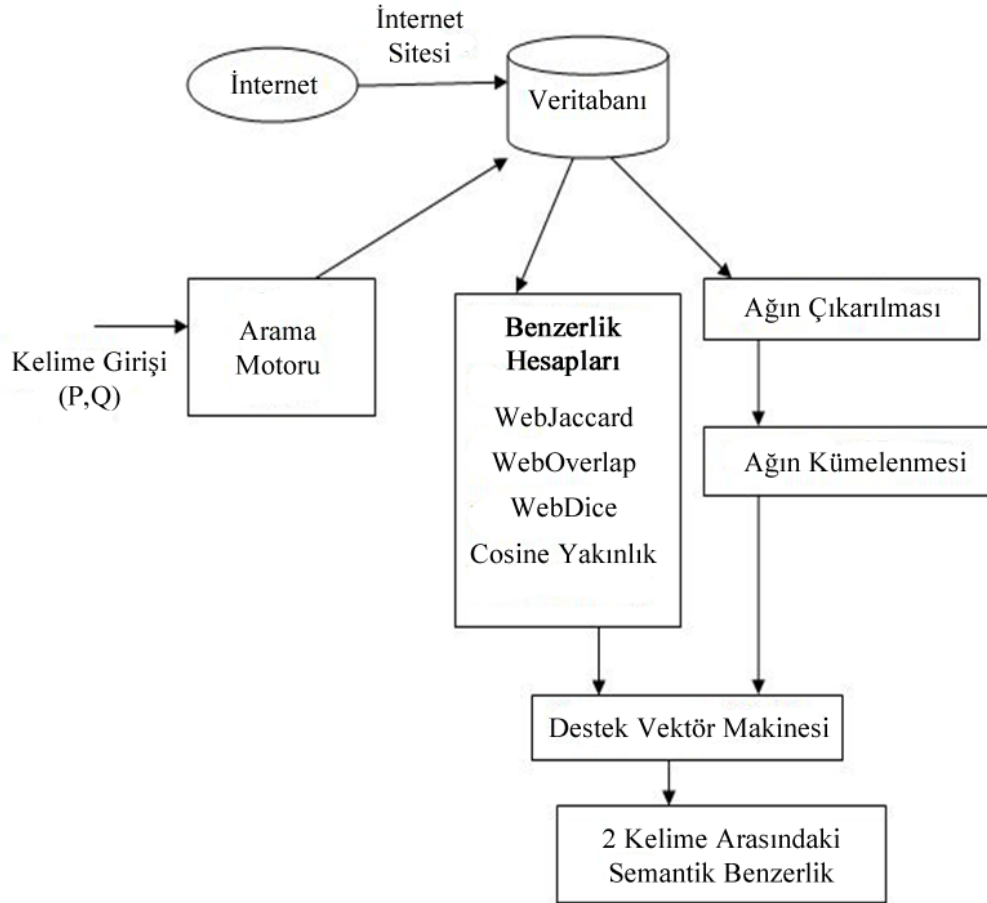
1. Simple matching (Basit eşleşme)
2. Dice
3. Jaccard
4. Overlap
5. Cosine coefficient (Cosine katsayısı)



Şekil 3.4. Semantik ve normal arama algoritmalarının doğruluk karşılaştırılması [58].

Şekil 3.4’de normal ve semantik arama algoritmalarının doğruluk sonuçlarının yüzde cinsinden karşılaştırılması görülmektedir. Semantik algoritmaların kullanıldığı aramaların daha doğru sonuçlar geri döndürdüğü görülmektedir.

Arama sorgusunun nasıl çalıştığına dair önerilen sistem mimarisi Şekil 3.5’de görülmektedir.



Şekil 3.5. Önerilen sistem mimarisi [57].

Google tarafından, anlamsal (semantic) ilişki ve örtülü anlamsal indeksleme (latent semantic indexing) değerlendirilerek sonuçlar kullanıcılara gösterilmektedir. Örtülü anlamsal indeksleme, bir metin yığını içinde kelimelerin anlamsal olarak yakınlıklarına bakılarak istatistiksel hesaplamalar sonucu elde edilen yöntemin adıdır [59]. Semantik ve örtülü semantik ilişkiye örnek Şekil 3.6’da gösterilmiştir.

Örnek’de (Şekil 3.6) görüldüğü üzere uçak ile yolcu ve tren ile yolcu anlamsal olarak doğrudan ilişki içindedir. İlişki içinde olduğu yolcu kelimesinden dolayı uçak ve tren kelimeleri de bir dolaylı ilişkiye sahiptir. Bu ilişkiye örtülü anlamsal ilişki adı verilmektedir.



Şekil 3.6. Semantik ve örtülü semantik ilişki.

Bir diğer çalışma örneği ise e-ticaret sitelerindeki semantik çalışmalardır. 2009 ile 2014 arasında e-ticaret %290 oranında artmıştır [60]. Firmalar kullanıcı profillerini doğru analiz ettikleri takdirde müşteri memnuniyetini üst düzeye çıkarabileceklerdir. Bunun için de metin madenciliği ve semantik analiz çalışmalarını yerinde ve verimli kullanabilmeleri önem taşımaktadır.

3.5. SEMANTİK YAKINLIK HESAPLAMA

3.5.1. TF (Terim Frekansı)

Veri setinde bulunan tüm dökümanlar ve bu dökümanlar içindeki kelimeler semantik yakınlık analizi yaparken önem teşkil etmektedir. Her kelimenin anlamsal yakınlık kurulurken oluşturacakları önem miktarı değişiklik göstermektedir. Bu önem miktarını en çok etkileyen faktör, kelimenin doküman içinde kaç defa kullanıldığıdır. TF (term frequency) belirlenen terimin dökümanda kaç defa tekrar ettiğinin sayısıdır [48]. Bir kelimenin TF değeri arttıkça semantik yakınlığa etkisi de artmaktadır.

3.5.2. IDF (Ters Döküman Frekansı)

Kelimelerin tekrar sayıları TFdeğerlerini vermektedir. Ancak tekrar sayısı kadar kaç farklı dökümanda kullanıldığı da önemlidir. IDF (inverse document frequency)

verisi, her bir kelime için toplam düğüm sayısının, o kelimenin kaç düğüm ile bağı olduğuna bölünerek, sonucun logaritması alınarak bulunur [61].

$$idf(w) = \log\left(\frac{|D|}{DF(w)}\right) \quad (3.1)$$

Eşitlik 3.1'deki formülde $|D|$ toplam düğüm sayısını, $DF(w)$ ise kelimenin kaç düğüm ile bağlantılı olduğunu göstermektedir[62]. IDF, her kelime için global değere sahiptir, yani hashtag'e göre değişim göstermez.

3.5.3. TF-IDF

“TF” ile “IDF” çarpımı ile “TF-IDF” verisi elde edilmiştir. TF-IDF verisi ile kaç defa geçtiğinin yanında kaç farklı dokümanda geçtiği de önem kazanmıştır. Eşitlik 3.2'de TF-IDF'in denklemi görülmektedir [62].

$$tfidf(w) = tf(w) \times idf(w) \quad (3.2)$$

Formülde de (Eşitlik 3.2) görüldüğü gibi kelimelerin tekrar sayıları ile kaç farklı dokümanda geçtikleri çarpılarak, hem TF hem de IDF'in önemli olduğu TF-IDF verisi bulunmaktadır.

3.5.4. L2 Normalizasyonu

Bir kelimeye ait TF-IDF vektörü, o kelimenin, kelime uzayındaki semantik örüntüsünü temsil eder. Birden fazla kelimeyi semantik olarak birbiri ile karşılaştırabilmek için bu vektörler kullanılır. Ancak bu vektörlerin uzunlukları çeşitli sebeplerden dolayı çok düşük ya da yüksek olabilir. Bu da karşılaştırma işlemlerini olumsuz etkiler. Semantik karşılaştırma genellikle bu vektörlerin aralarındaki açının tespiti ile yapılmaktadır. Bu açının tespitinde ise vektörlerin birim vektöre normalize edilmesi önem arz etmektedir. Bu nedenle TF-IDF vektörleri L2 normalizasyonuna tabi tutulmuştur.

L2 normalizasyonu yapılırken öncelikle n boyutlu uzayda vektörün tüm boyutlardaki uzunluklarının kareleri toplamının karekökü alınarak uzunluğu bulunur (öklit uzunluğu). Sonra vektörün tüm elemanları bu uzunluk değerine bölünerek normalize edilir. Böylece tüm kelimelerin TF-IDF vektörlerinin uzunluklarının 1 olması sağlanmış olur (Eşitlik 3.3 ve 3.4).

$$uzunluk = \sqrt{tfidf_1^2 + tfidf_2^2 + tfidf_3^2 + \dots + tfidf_n^2} \quad (3.3)$$

$$L2_x = tfidf_x / uzunluk \quad (3.4)$$

3.5.5. Cosine Yakınlığı Hesaplama

İki vektörün birbirleri ile yakınlıklarının incelenmesi yöntemlerinden birisi cosine yakınlık fonksiyonudur. Cosine yakınlık incelenirken vektörler arasındaki açıya bakılır. İki vektör arasındaki açı büyüdükçe cosine değeri sıfıra yaklaşacağından benzerlik azalacak, açı azaldıkça cosine değeri bire yaklaşacağından benzerlik artacaktır [63]. İlişkisi incelenecek iki düğüm için cosine değeri, Eşitlik 3.5’de görüldüğü üzere L2 normalizasyonuna tabi tutulmuş TF-IDF vektörlerinin skalar çarpımları ile bulunur.

$$cosine = L2_1 * L2_2 \quad (3.5)$$

BÖLÜM 4

UYGULAMA

4.1. VERİ SETİNİN OLUŞTURULMASI

Bu çalışmada 140dev Streamin API yapısı kullanılarak Twitter mikroblog sitesinden 01.10.2015 – 20.01.2016 tarihleri arasında veri çekilmiştir. Elde edilen 9 320567 adet Tweet'in 7 917 949 adetinde etiket bulunmayıp, 1 402 618 adet Tweet'te etiket mevcuttur. Ayrıca 1952237 adet etiket elde edilmiştir. Bu etiket kümesi ise 104492 adet tekil tag'den oluşmaktadır.

4.1.1. Filtreleme

Veri çekme işlemi sırasında veritabanının boyutunun gereksiz artmasını önlemek için 3 farklı filtreleme uygulanmıştır. Bunlar dil seçimi, otomatik Tweet mesajlarının önlenmesi ve reTweet engelleme filtreleridir.

4.1.1.1. Dil Seçimi

Türkçe Tweetler ve tagler arasındaki ilişkilerin incelenmesi amaçlanan bu uygulamada diğer dillerdeki Tweetlerin alınmaması gerekmektedir. Streaming API ile dönen Tweet nesnesinde bulunan dil alanı Türkçe olarak seçilmiştir. Bu sayede yapılacak çalışmada Tweet metinlerinin Türkçe olması sağlanmıştır.

4.1.1.2. Otomatik Tweet Mesajları Engelleme

Twitter mikroblog sitesinde direk Tweet göndermenin 3. parti uygulamalar üzerinden de otomatik mesajlar gönderilebilmektedir. "I'm at", "I'm in" veya "Facebook'ta yeni bir" şeklinde başlayan bu mesajlar anlamsal olarak bir değer taşımamaktadır.

Anlamsal olarak bir değeri olmadığı için uygulamada kullanılamayacaktır. Veritabanında boyut arttırmaması için bu mesajlar Tweet'ler parse edilirken Şekil 4.1'de verilen kod ile otomatik olarak engellenmiştir.

```
53 $tweet_text = $oDB->escape($tweet_object->text);
54 //Otomatik Tweet engelleme
55 if (0 != strpos($tweet_text, "I'm at") ) {continue;}
56 if (0 != strpos($tweet_text, "I'm in") ) {continue;}
57 if (0 != strpos($tweet_text, "Facebook'ta yeni bir") ) {continue;}
58
```

Şekil 4.1. Otomatik Tweet mesajlarının engellenmesi.

4.1.1.3. ReTweet Engelleme

Bir başkasına ait Tweet mesajının diğer bir kişi tarafından paylaşılmasına reTweet adı verilmektedir. Mevcut olan Tweet metninin tekrar kopyalanması, veri setine yeni bir mesaj ve anlam kazandırmayacağı için bu reTweet mesajları engellenmiştir.

4.1.2. Ön İşleme

Twitter Streaming API yöntemi ile elde edilen veri setindeki ham haldeki Tweet verileri, yapılacak olan çalışmada kullanılabilmek için bazı ön işleme aşamalarından geçirilmiştir. Bu ön işlemenin amacı Tweet metinlerinin kelimelere ayrılarak anlam kazandırılabilmesi, bu sayede semantik analizde kullanılabilmesidir. Semantik analizde, kelimelerin birlikte kullanım istatistikleri önem taşımaktadır. Aşağıdaki işlemler sırasıyla uygulanarak ham veri seti uygulamada kullanılabilir hale dönüştürülmüştür.

4.1.2.1. URL'lerin Çıkartılması

Kullanıcılar Tweet metinleri içerisinde bir başka siteye ait URL (web bağlantıları) paylaşımı da yapmaktadırlar. Ayrıca Twitter kullanıcılarına metnin yanında fotoğraf ve video paylaşımı yapma imkanı da sunmaktadır. Bu video ve fotoğraflar gönderilerde URL olarak yer alır. Hem gönderi linkleri hem de dış kaynaklara

verilen URL'ler anlamsal olarak bir deęer taşımadığı için Tweet metinlerinden temizlenmiştir.

4.1.2.2. Karakter Tekrar Sayılarının 2'ye Düşürülmesi

Herhangi bir imla denetim mekanizmasına sahip olmayan Twitter platformunda, kullanıcıların gönderilerinde imla hataları yer almaktadır. Kullanıcılardan kaynaklanan birçok yanlış yazım meydana gelmektedir. Bütün bu yanlışları kontrol edebilmek ve düzeltebilmek mümkün değildir.

Bazı yanlışları ortadan kaldırabilmek için yapılan bu ön işleme çalışmalarında Türkçe sözcüklerde bir karakterin ikiden fazla tekrar etmediği gerçeği göz önünde bulundurularak, ikiden fazla tekrar eden karakter sayısı ikiye düşürülmüştür. Örnek olarak bu işlem sonunda “gonullu” kelimesi “gonullu” haline dönüştürülmüştür.

4.1.2.3. Altçizgilerin Kaldırılması

Kullanıcıların yazım alışkanlıkları aynı metni farklı yazım şekillerinde karşımıza çıkarmaktadır. Hashtag kullanımında bazı kullanıcılar “1kasimsecimi” şeklinde kullanırken, bazı kullanıcılar kelimelerin arasına altçizgi koyarak “1_kasim_secimi” şeklinde kullanmışlardır. Bu farklılığı ortadan kaldırmak için altçizgiler silinmiştir. Böylece “1_kasim_secimi” şeklinde yazılmış bir hashtag “1kasimsecimi” şekline dönüştürülmüştür.

4.1.2.4. User Mention'ların Temizlenmesi

Bazı Tweet metinlerinde user mention'lar bulunmaktadır. Türkçe karşılığı “bahsetmek” olan mention kelimesi Twitter'da metnimizin içinde bir kişiyi etiketlemek istediğimizde kullandığımız özelliktir. Bu user mention'lar herhangi bir kelime değeri ve anlamı olmadığı için metinlerden temizlenmiştir.

4.1.2.5. Smiley'lerin Temizlenmesi

Kullanıcılar Tweet gönderilerinde smiley adını verdiğimiz ifadelerle yer vermektedirler. Smiley'ler belli karakter kombinasyonları ile yüz ifadelerini belirtmenin kolay yoludur. Bu karakter kombinasyonlarına örnek olarak “:), :-), :(“ gösterilebilir. Bu ifadeler anlamsal olarak bir bağ kurulmasına yardımcı olmayacağından metinlerden temizlenmiştir.

4.1.2.6. Noktalama Ve İmla İşaretlerinin Temizlenmesi

Bütün metinlerde olduğu gibi Tweet metinlerinde de noktalama işaretleri kullanılmaktadır. Noktalama işaretlerinin kelimelerle bitişik kullanıldığı gözlemlenmiştir. Örneğin “Bencil insanların,mutlu olduğu konusunda hemfikirim” metninde “insanların” ve “mutlu” kelimeleri “,” işaretiyle birleşik kullanılmıştır. Kelimelerin ayrılmasını sağlamak için bu noktalama işaretleri, boşluk yani “ ” karakteriyle değiştirilmiştir.

4.1.2.7. Karakter Sayısı 20'den Fazla Olan Kelimelerin Silinmesi

Tez çalışmamızda, Tweet metinlerinde kelimelerin ayrışmasını sağlayarak anlamlandırma yapmak amaçlanmaktadır. Ancak bazı kullanıcıların kelimeler arasında boşluk kullanmadan metinler oluşturmaları bu kelimeleri ayrıştırmayı ve anlamlandırmayı imkansız kılmaktadır. Karakter sayısı 20'den fazla olan kelimelerde ve hashtag'lerde kelime öbeklerinin birleşik yazıldığı tespit edilmiştir. Bu amaçla “bencil insanların mutlu olduğu konusunda hemfikirim” şeklindeki yazımlar Tweet metinlerinden çıkarılmıştır.

4.1.2.8. Düzensiz Boşlukların Teke Düşürülmesi

Kullanıcılar metinlerini yazarken herhangi bir kontrol mekanizması olmadığından kelimeler arasında birden fazla boşluk bırakabilmektedir. Ayrıca ön işleme aşamalarında bazı karakterlerin boşluk ile değiştirilmesi sonucu da metinler içinde

düzensiz boşluklar oluşmuştur. Tüm bu düzensiz boşluklar teke düşürülerek kelimeler arası tek boşluk olması sağlanmıştır.

4.1.2.9. Küçük Harfe Dönüştürme

Bir yazım kuralı olmadığı için Tweet metinlerinde büyük ve küçük harf kullanımı farklılıkları ortaya çıkmaktadır. Kelimelerin eşleşmesi belli bir kurala göre yapılacağından bütün karakterler küçük harfe dönüştürülmüştür. Bu işlem sonrasında örnek olarak “Bencil insanların MUTLU olduğu konusunda HEMFİKİRİM” ifadesi “bencil insanların mutlu olduğu konusunda hemfikirim” şekline dönüştürülmüştür.

4.1.2.10. Türkçe Karakterlerin Dönüşümü

Kullanıcıların kullandıkları işletim sistemi ve klavye dili tercihleri aynı metni farklı şekilde yazmalarına sebep olmaktadır. Kontrol edilemediği için kişilerin yazım alışkanlıkları da bu farklılıkları doğurmaktadır. Örnek olarak “insanların” ifadesi bir başkası tarafından “insanların” şeklinde yazılmakta ve bu iki yazım aynı olmasına rağmen farklı değerlendirilmektedir. Bu farklılığı önlemek için bütün Türkçe karakterler İngiliz alfabesinde bulunan karakterlerle eşlenerek dönüşümü yapılmıştır. Bu dönüşüm sonunda hangi karakterin hangi karaktere dönüştürüldüğü Çizelge 4.1’de verilmiştir.

Çizelge 4.1. Türkçe karakter dönüşümleri.

| Eski hali | Dönüştürüldüğü karakter |
|-----------|-------------------------|
| ç | c |
| ğ | g |
| ı | i |
| ö | o |
| ş | s |
| ü | u |

4.1.2.11. Tek Karakterlerin Temizlenmesi

Bütün yapılan ön işleme aşamalarının sonrasında oluşan tek karakterler ve metin içerisinde yer alan “a dostlar” ifadesindeki “a” gibi tek karakterler anlamsal bir değer taşımadığından dolayı metinlerden temizlenmiştir.

4.1.2.12. Yabancı Alfabe Karakterlerinin Temizlenmesi

Dil Türkçe seçilmesine rağmen, kullanıcıların Türkçe Tweet içinde gerek Tweet metni gerekse hashtag’ler içinde başka alfabe karakterleri ile yazımları tespit edilmiştir. Örnek olarak “sonunda Twitter Arap alfabesini desteklemeye başladı” Tweet ine arap alfabesi ile hashtag eklenmiştir. Bu yabancı alfabe ile yazılmış Tweet metinleri ve hashtag’ler temizlenmiştir.

4.2. AĞLARIN OLUŞTURULMASI

Tez çalışmamızda Twitter mikroblog sitesinden elde edilen veri seti kullanılarak etiket ağı analizi yapılmıştır. İki farklı yöntem ile bu etiketlerin birbirleri ile olan ilişkileri incelenmiştir. Bu yöntemlerden birincisi kullanım birlikteliklerine göre hashtag ağı analizi, diğeri ise semantik yakınlığa göre hashtag ağı analizidir. Hashtag’lerin birlikte kullanımının anlamsal olarak hangi ölçüde doğru olduğu ve anlamsal ilişkilerin tanımladığı ağ yapılarının ne derece uyumlu olduğunun ortaya konulması amacıyla bu iki yöntemin eşleşen sonuçları karşılaştırılmıştır.

4.2.1. Kullanım Birlikteliklerine Göre Hashtag Ağı

4.2.1.1. TekTweet’te Geçen Tag’ların Silinmesi

Twitter kullanıcıları bazen Twitter’da mevcut tag’lardan birisini Tweet’lerinde kullanmakta, bazen de kendi tag’larını oluşturup popüleştirmeyi denemektedirler. Ancak çoğu zaman bu tag denemeleri başarısız olmakta, sadece o kullanıcıya has kalmaktadır. Bu tür tekrarsız tag’lar ile birliktelik analizi yapılamayacağından, sadece bir Tweet’te geçen tag’lar silinmiştir. Sonraki adımda ise her bir Tweet için

eğer bir Tweet'te sadece bir hashtag var ise, yine hashtag birlikte kullanım analizi yapılamayacağı için o hashtag ve Tweetde silinmiştir. Bu işlem sonunda 1 855 802 adet hashtag kümesinden 1 084 407 adeti silinerek 771395 adet hashtag kalmıştır.

4.2.1.2. Nodes Tablosunun Oluşturulması

Birliktelik ağımızda hashtag'ler düğümü, iki düğümün aynı Tweet'te kullanılması da linkleri temsil etmektedir. Düğüm tablosu oluşturmak için hashtag'ler tekil hale getirilmiştir. Bunun sonucunda 34111 adet tekil hashtag ten oluşan düğüm tablosu elde edilmiştir.

4.2.1.3. Edges Tablosunun Oluşturulması

Aynı Tweet içerisinde birlikte kullanılan, dolayısıyla veritabanımızda aynı Tweet id'ye sahip olan hashtag'ler birbiri ile alakalı kabul edildiğinden bunlar arasında yönsüz bir ilişki kurulmuş ve edges tablosu oluşturulmuştur. Bu işlem sonunda toplam 197675 adet link oluşmuştur. İki adet hashtag'in tekrar birlikte kullanımlarında yeni edge oluşturulmayıp, edge tablosunda mevcut ilişkinin ağırlık (weight) değeri bir arttırılmıştır. Çizelge4.2'de görüldüğü üzere, ağırlığı az olan link sayısı fazladır. Ağırlık değeri yükseldikçe bu değere sahip olan link sayısı azalmaktadır.

Çizelge 4.2. Linklerin ağırlık değerleri.

| Ağırlık (weight) | Edge sayısı |
|------------------|-------------|
| 1 | 119833 |
| 2 | 30468 |
| 3 | 12886 |
| 4 | 7630 |
| 5 | 5123 |
| 6 | 3313 |
| 7 | 2676 |
| 8 | 1986 |
| 9 | 1495 |
| >10 | 12265 |

Çizelge 4.3'de ağırlığı en yüksek 20 link listelenmiştir. Bu listeye göre en çok birlikte kullanılan hashtag'ler “acil” ve “kan” hashtag'leridir. Listenin 5. sırasında da

“acil” ve trombosit” kelimeleri görmekteyiz. “kan” kelimesinin günlük hayatta da “acil” lazım olan bir çağrışım yaptığını göz önünde bulundurursak, ağırlık ne kadar fazla ise iki hashtag’in o kadar ilişkili olduğu sonucuna varabiliriz. “haber” hashtag’inin “Türkiye” ve “İstanbul” hashtagleri ile olan kuvvetli birlikteliğini de yine tablodan görebilmekteyiz. İstanbulun türkiyede ki nüfus ve diğer açılardan konumunu düşündüğümüzde haber kelimesi ile Türkiye ve hemen ardından İstanbul kelimelerinin eşleşmesi günlük hayatta olduğu gibi çalışma sonucunda da ortaya çıkmıştır.

Çizelge 4.3. Ağırlığı en yüksek 20 link.

| source | target | weight |
|----------|--------------|--------|
| acil | kan | 4898 |
| dirilis | haber | 4613 |
| istanbul | kan | 4146 |
| haber | turkiye | 3976 |
| acil | trombosit | 3602 |
| dirilis | gazete | 3339 |
| gazete | haber | 3332 |
| dirilis | turkiye | 3325 |
| gazete | turkiye | 3325 |
| guncel | haber | 3316 |
| gazete | guncel | 3299 |
| dirilis | guncel | 3299 |
| guncel | turkiye | 3295 |
| takip | takipleselim | 3258 |
| acil | istanbul | 3044 |
| ass | video | 2567 |
| istanbul | trombosit | 2513 |
| sex | video | 1940 |
| ass | dul | 1907 |
| ankara | trombosit | 1892 |

Bu bölümde elde etmiş olduğumuz hashtag birliktelik ağının analizi 5. Bölümde detaylı olarak sunulacaktır. Tez çalışmamızın önemli çıktılardan birisi de hashtag’lerin birlikte kullanım sıklıklarının anlamsal olarak ne kadar doğru, gerçek hayatla ne kadar ilişkili olduğunun ortaya çıkarılmasıdır. İkinci ağ çalışmamız olan

semantik ağ ile hashtag'lerin anlamsal yakınlığı incelenecek olup, birlikte kullanımlarının gerçekçilik oranı nicel olarak da sorgulanabilecektir.

4.2.2. Semantik Yakınlığa Göre Hashtag Ağı

Tez çalışmamızın bu bölümünde hashtag'lerin birlikte kullanımı değil, semantik (anlamsal) yakınlığı göz önünde bulundurularak bir kompleks ağ tanımlaması yapılacaktır. İki hashtag'in semantik yakınlığını tespit edebilmek, birlikte kullanıldıkları kelime kümelerinin benzerliklerinin belirlenmesi ile mümkündür. Bu nedenle ikinci ağ çalışmamız, hashtag'lerin birlikte kullanıldığı kelimelerin analizini gerektirmektedir.

Bu analiz esnasında, bir tweet içerisinde birlikte kullanılan hashtag'ler aynı kelime kümesine sahip olacağından dolayı semantik hesaplama üzerinde gürültü etkisi oluşturacaktır. Bu nedenle birden fazla hashtag içeren 771 395 adet tag veri setinden temizlenmiş, sadece tek hashtag içeren 1 084 407 adet Tweet ile semantik analiz gerçekleştirilmiştir.

Semantik analiz kendine özgü bir takım kelime işleme prosedürleri gerektirmektedir. Bu bölümde bu adımlar izah edilecektir.

4.2.2.1. Kelime Setinin Elde Edilmesi

Hashtag'lerin semantik olarak anlamlandırılması ve bağ kurulması için birlikte kullanılmış oldukları kelimeler birer nitelik olarak ele alınacaktır. Bu nedenle her bir Tweet'te geçen kelimeler ayrıştırılarak kelime tablosuna tekil olarak kaydedilmiştir. Tweet setinde 372937 adet farklı kelime bulunmaktadır.

4.2.2.2. Stopword'lerin Çıkarılması

Tüm diller, cümle içerisinde sık kullanılan fakat anlamsal etkisi düşük olan kelimeler barındırırlar. Bu kelimeler semantik analizde stopword olarak adlandırılırlar ve kelime setinden çıkarılırlar. Çizelge 4.4'de Türkçe Tweet'lerde tespit edilen bu

kelimeler listelenmiştir. Semantik bağ kurulurken sonucu etkilemesini önlemek için bu kelimeler, kelime listesinden temizlenmiştir. Bu işlem sonucunda kelime tablosunda 372756 adet kelime kalmıştır.

Çizelge 4.4. Stopwords.

| | | | | | |
|----------|-----------|----------|-----------|------------|----------|
| acaba | altı | ama | ancak | artık | asla |
| aslında | az | bana | bazen | bazı | bazıları |
| bazısı | belki | ben | beni | benim | beş |
| bile | bir | birçoğu | birçok | birçokları | biri |
| birisi | birkaç | birkaçı | birşey | birşeyi | biz |
| bize | bizi | bizim | böyle | böylece | bu |
| buna | bunda | bundan | bunu | bunun | burada |
| bütün | çoğu | çoğuna | çoğunu | çok | çünkü |
| da | daha | de | değil | demek | diğer |
| diğeri | diğerleri | diye | dokuz | dolayı | dört |
| elbette | en | fakat | falan | felan | filan |
| gene | gibi | hala | hangi | hangisi | hani |
| hatta | hem | henüz | hep | hepsi | hepsine |
| hepsini | her | her biri | herkes | herkese | herkesi |
| hiç | hiç kimse | hiçbiri | hiçbirine | hiçbirini | için |
| içinde | iki | ile | ise | işte | kaç |
| kadar | kendi | kendine | kendini | ki | kim |
| kime | kimi | kimin | kimisi | madem | mı |
| mi | mu | mü | nasıl | ne | ne kadar |
| ne zaman | neden | nedir | nerde | nerede | nereden |
| nereye | nesi | neyse | niçin | niye | on |
| ona | ondan | onlar | onlara | onlardan | onların |
| onların | onu | onun | orada | oysa | oysaki |
| öbürü | ön | önce | ötürü | öyle | rağmen |
| sana | sekiz | sen | senden | seni | senin |
| siz | sizden | size | sizi | sizin | son |
| sonra | şayet | şey | şeyden | şeye | şeyi |
| şeyler | şimdi | şöyle | şu | şuna | şunda |
| şundan | şunlar | şunu | şunun | tabi | tamam |
| tüm | tümünü | üç | üzere | var | ve |
| veya | veyahut | ya | ya da | yani | yedi |
| yerine | yine | yoksa | zaten | zira | |

4.2.2.3. 3'ten Az, 12'den Fazla Karakterden Oluşan Kelimelerin Çıkarılması

“a”, “e” gibi tek karakterden oluşan veya “aa”, “eh”, “oh” gibi iki karakterden oluşan kelimeler anlamsal yönden güçlü olmayıp çok sık kullanılmaktadır. Yine belli bir uzunluktan fazla olan kelimeler ya anlamsız olmakta ya da birden fazla kelimenin birleşiminden oluştuğu için anlamsız hale gelmektedir. Veri seti incelendiğinde uzunluk sınırı için 12 karakter eşik değerinin uygun olduğu görülerek, uzunluk sınırı 12 olarak kabul edilmiştir. 3'ten az ve 12'den fazla karakterden oluşan 59962 adet kelime silinerek kelime tablosunda kelime sayısı 312794 adete indirilmiştir.

4.2.2.4. Kelime Setinin Oluşturulması

Kelimelerin ağırlıklarındırılması ve ardından bu kelimelerin anlam birlikteliğindenitelik olarak kullanılabilmesi için bir kelime setine ihtiyaç duyulmaktadır. Bu kelime seti, semantik analizin hesaplama prosedürlerini de etkilediği için belli bir sayı ile sınırlı olmak durumundadır. Çalışmamızda 500 kelime bir nitelik kümesi kullanımı uygun görülmüştür. Kelime tablosunda en çok kullanılan 1000 kelime Çizelge 4.5’de listelenmiştir.

Çizelge 4.5. En çok kullanılan 1000 kelime.

| | | | | | | | | |
|------|------------|--------|----|----------|-------|-----|-------------|------|
| sıra | kelime | tekrar | 35 | karsi | 10313 | 70 | bolum | 6566 |
| 1 | olsun | 28578 | 36 | dunya | 10232 | 71 | hesaplar | 6565 |
| 2 | yok | 25702 | 37 | takip | 10147 | 72 | islam | 6520 |
| 3 | iyi | 24010 | 38 | olmak | 10128 | 73 | ayni | 6381 |
| 4 | guzel | 20508 | 39 | pkk | 9869 | 74 | isid | 6362 |
| 5 | ilk | 19745 | 40 | guclu | 9366 | 75 | vatan | 6344 |
| 6 | olan | 19224 | 41 | saat | 9197 | 76 | kadin | 6325 |
| 7 | allah | 18780 | 42 | htt | 8955 | 77 | polis | 6294 |
| 8 | gun | 18682 | 43 | takipci | 8587 | 78 | dogru | 6112 |
| 9 | yeni | 18570 | 44 | adam | 8555 | 79 | aksam | 5977 |
| 10 | turkiye | 18405 | 45 | istanbul | 8294 | 80 | sayin | 5963 |
| 11 | ask | 18177 | 46 | akp | 8218 | 81 | sehit | 5951 |
| 12 | bugun | 17530 | 47 | egitim | 8196 | 82 | bekliyoruz | 5946 |
| 13 | tek | 17221 | 48 | yil | 7922 | 83 | adalet | 5940 |
| 14 | enes | 16839 | 49 | eden | 7818 | 84 | kimse | 5904 |
| 15 | insan | 16505 | 50 | 2015 | 7732 | 85 | acik | 5882 |
| 16 | zaman | 14864 | 51 | cocuk | 7708 | 86 | teror | 5876 |
| 17 | buyuk | 14700 | 52 | yarin | 7688 | 87 | gore | 5808 |
| 18 | hadis | 14541 | 53 | insanlar | 7609 | 88 | kasim | 5755 |
| 19 | nbavote | 13820 | 54 | mutlu | 7527 | 89 | haydi | 5697 |
| 20 | kanter | 13735 | 55 | devlet | 7469 | 90 | yapan | 5680 |
| 21 | oldu | 13525 | 56 | diyen | 7435 | 91 | gunaydin | 5663 |
| 22 | destek | 13395 | 57 | gunu | 7337 | 92 | hdp | 5630 |
| 23 | fenerbahce | 13108 | 58 | baris | 7241 | 93 | olsa | 5524 |
| 24 | tahirelci | 12803 | 59 | eger | 7136 | 94 | basina | 5477 |
| 25 | oldugu | 12599 | 60 | amp | 7094 | 95 | ozel | 5433 |
| 26 | kisi | 12163 | 61 | cocuklar | 7073 | 96 | paris | 5415 |
| 27 | sadece | 11929 | 62 | can | 7055 | 97 | 10kasim | 5384 |
| 28 | olur | 11847 | 63 | bin | 6955 | 98 | periscopeta | 5367 |
| 29 | turk | 11825 | 64 | sahip | 6932 | 99 | hak | 5366 |
| 30 | lutfen | 11760 | 65 | olacak | 6756 | 100 | kurt | 5329 |
| 31 | devam | 11429 | 66 | ali | 6686 | 101 | rus | 5313 |
| 32 | olarak | 11098 | 67 | baska | 6638 | 102 | olmaz | 5309 |
| 33 | canli | 10684 | 68 | ediyor | 6608 | 103 | nin | 5287 |
| 34 | istiyoruz | 10528 | 69 | haber | 6580 | 104 | gercek | 5244 |

Çizelge 4.5. (devam ediyor).

| | | |
|------------|--------------------|-------------|
| 105 | agir | 5225 |
| 106 | oluyor | 5175 |
| 107 | gece | 5139 |
| 108 | degildir | 5131 |
| 109 | basladi | 5129 |
| 110 | allaha | 5114 |
| 111 | sokaga | 5028 |
| 112 | falkutusu | 5016 |
| 113 | yerde | 4993 |
| 114 | tam | 4991 |
| 115 | atama | 4991 |
| 116 | onemli | 4980 |
| 117 | bekliyor | 4978 |
| 118 | hicbir | 4966 |
| 119 | oldugunu | 4946 |
| 120 | sondakika | 4877 |
| 121 | allahin | 4858 |
| 122 | geri | 4842 |
| 123 | ulke | 4841 |
| 124 | yer | 4817 |
| 125 | gelecek | 4816 |
| 126 | hayirli | 4762 |
| 127 | 2016 | 4734 |
| 128 | verin | 4727 |
| 129 | geldi | 4717 |
| 130 | vardir | 4713 |
| 131 | cikti | 4712 |
| 132 | serif | 4710 |
| 133 | yeter | 4673 |
| 134 | isi | 4658 |
| 135 | etmek | 4653 |
| 136 | kismetseolur | 4638 |
| 137 | hadi | 4618 |
| 138 | insallah | 4616 |
| 139 | savas | 4612 |
| 140 | hayir | 4610 |
| 141 | secim | 4587 |
| 142 | ses | 4580 |
| 143 | acil | 4552 |
| 144 | demirtas | 4528 |
| 145 | turkmen | 4500 |
| 146 | musluman | 4484 |
| 147 | turkiyenin | 4463 |
| 148 | fazla | 4441 |
| 149 | olmus | 4435 |
| 150 | birlikte | 4432 |
| 151 | all | 4427 |
| 152 | kan | 4425 |
| 153 | erdogan | 4424 |
| 154 | sonucu | 4419 |
| 155 | aciklama | 4412 |
| 156 | dua | 4395 |
| 157 | besiktas | 4383 |
| 158 | gelen | 4352 |
| 159 | katil | 4350 |
| 160 | etti | 4333 |
| 161 | baskani | 4332 |
| 162 | ulkede | 4308 |
| 163 | varsa | 4304 |
| 164 | yardim | 4300 |
| 165 | bak | 4295 |
| 166 | milli | 4292 |
| 167 | tesekkurler | 4284 |
| 168 | geliyor | 4283 |
| 169 | hayat | 4281 |
| 170 | secim2015 | 4263 |
| 171 | kaldi | 4244 |
| 172 | asik | 4235 |
| 173 | chp | 4231 |
| 174 | tarafindan | 4195 |
| 175 | arkadaslar | 4183 |
| 176 | parti | 4180 |
| 177 | tweet | 4171 |
| 178 | terorist | 4091 |
| 179 | biraz | 4089 |
| 180 | ogretmen | 4084 |
| 181 | kiralikask | 4060 |
| 182 | ankara | 4056 |
| 183 | rusya | 4028 |
| 184 | hafta | 4020 |
| 185 | zor | 3999 |
| 186 | millet | 3990 |
| 187 | lazim | 3989 |
| 188 | aturturk | 3987 |
| 189 | evet | 3986 |
| 190 | tahir | 3981 |
| 191 | halk | 3977 |
| 192 | nba | 3972 |
| 193 | sevgi | 3925 |
| 194 | insanin | 3914 |
| 195 | cikma | 3909 |
| 196 | izlemek | 3908 |
| 197 | mtvstars | 3906 |
| 198 | tag | 3894 |
| 199 | ihbar | 3866 |
| 200 | olun | 3859 |
| 201 | bigbrothertr | 3849 |
| 202 | dedi | 3841 |
| 203 | saygi | 3840 |
| 204 | cocugu | 3834 |
| 205 | mac | 3826 |
| 206 | yapanlar | 3750 |
| 207 | allahim | 3722 |
| 208 | yeniden | 3722 |
| 209 | onunde | 3721 |
| 210 | sabah | 3719 |
| 211 | kez | 3717 |
| 212 | aniyoruz | 3703 |
| 213 | diyor | 3685 |
| 214 | israil | 3653 |
| 215 | 100 | 3643 |
| 216 | genel | 3632 |
| 217 | insanlarin | 3632 |
| 218 | izmir | 3607 |
| 219 | oykullan | 3591 |
| 220 | ver | 3590 |
| 221 | istiyor | 3590 |
| 222 | soz | 3570 |
| 223 | gerek | 3562 |
| 224 | dun | 3562 |
| 225 | edin | 3550 |
| 226 | ulkenin | 3549 |
| 227 | gelir | 3541 |
| 228 | huzur | 3528 |
| 229 | insanlik | 3526 |
| 230 | para | 3498 |
| 231 | olum | 3493 |
| 232 | egitimi | 3473 |
| 233 | aski | 3472 |
| 234 | kadro | 3469 |
| 235 | kotu | 3455 |
| 236 | olmayan | 3444 |
| 237 | veren | 3435 |
| 238 | sosyal | 3427 |
| 239 | yalniz | 3415 |
| 240 | yapilan | 3398 |
| 241 | beyaz | 3395 |
| 242 | hizmet | 3385 |
| 243 | yap | 3379 |
| 244 | subatta | 3371 |
| 245 | kiz | 3369 |
| 246 | beden | 3364 |
| 247 | ettigi | 3363 |
| 248 | 1dtr | 3359 |
| 249 | dunyanin | 3343 |
| 250 | kuran | 3335 |
| 251 | dakika | 3332 |

Çizelge 4.5. (devam ediyor).

| | | |
|-----|--------------------------|-------------|
| 252 | bunlar | 3328 |
| 253 | yaninda | 3326 |
| 254 | yoktur | 3326 |
| 255 | bence | 3321 |
| 256 | galatasaray | 3308 |
| 257 | dolu | 3304 |
| 258 | aralik | 3268 |
| 259 | bos | 3266 |
| 260 | aci | 3245 |
| 261 | gecen | 3219 |
| 262 | ahmet | 3203 |
| 263 | kabul | 3188 |
| 264 | lan | 3179 |
| 265 | bizde | 3163 |
| 266 | instagram | 3156 |
| 267 | eder | 3146 |
| 268 | abi | 3145 |
| 269 | karsilikli | 3121 |
| 270 | ortak | 3116 |
| 271 | gel | 3113 |
| 272 | tesekkur | 3110 |
| 273 | yapiyor | 3099 |
| 274 | gormek | 3097 |
| 275 | edildi | 3044 |
| 276 | diyarbakir | 3038 |
| 277 | turkiyede | 3036 |
| 278 | verdigi | 3027 |
| 279 | yol | 3016 |
| 280 | cagatay | 3008 |
| 281 | olmasi | 2992 |
| 282 | yalan | 2985 |
| 283 | yuzde | 2984 |
| 284 | dogdun | 2982 |
| 285 | adi | 2967 |
| 286 | tekrar | 2965 |
| 287 | kalp | 2961 |
| 288 | bakin | 2945 |
| 289 | anda | 2944 |
| 290 | dunyada | 2942 |
| 291 | ogretmenle ri | 2939 |
| 292 | silah | 2937 |
| 293 | istiyorum | 2935 |
| 294 | omer | 2928 |
| 295 | genc | 2925 |
| 296 | yuzden | 2918 |
| 297 | bizler | 2913 |
| 298 | mucadele | 2909 |
| 299 | ozgur | 2902 |
| 300 | insanlari | 2902 |

| | | |
|-----|-------------------|-------------|
| 301 | elci | 2901 |
| 302 | twitter | 2900 |
| 303 | hocam | 2896 |
| 304 | mustafa | 2894 |
| 305 | yere | 2892 |
| 306 | tebrikler | 2886 |
| 307 | insani | 2885 |
| 308 | sandiga | 2882 |
| 309 | gelsin | 2880 |
| 310 | yaz | 2875 |
| 311 | medya | 2861 |
| 312 | satın | 2856 |
| 313 | yapın | 2855 |
| 314 | keske | 2845 |
| 315 | tarih | 2839 |
| 316 | cagri | 2838 |
| 317 | beraber | 2836 |
| 318 | 00da | 2835 |
| 319 | selam | 2833 |
| 320 | takiplesir | 2833 |
| 321 | turkce | 2826 |
| 322 | silvan | 2825 |
| 323 | genclik | 2823 |
| 324 | yasagi | 2822 |
| 325 | tiklayın | 2811 |
| 326 | cuma | 2807 |
| 327 | yolda | 2804 |
| 328 | belli | 2796 |
| 329 | dedigi | 2793 |
| 330 | demektir | 2788 |
| 331 | rtbegeni | 2786 |
| 332 | ayse | 2785 |
| 333 | umut | 2784 |
| 334 | ilan | 2782 |
| 335 | takipclub | 2779 |
| 336 | basin | 2776 |
| 337 | star | 2771 |
| 338 | halki | 2770 |
| 339 | yapti | 2754 |
| 340 | almak | 2751 |
| 341 | goz | 2738 |
| 342 | ediyorum | 2733 |
| 343 | bas | 2732 |
| 344 | olalim | 2732 |
| 345 | iyiki | 2728 |
| 346 | harika | 2718 |
| 347 | kutlu | 2702 |
| 348 | milyon | 2696 |
| 349 | ediyoruz | 2693 |

| | | |
|-----|------------------|-------------|
| 350 | yasamak | 2691 |
| 351 | yayın | 2688 |
| 352 | kasimda | 2678 |
| 353 | kahve | 2673 |
| 354 | olmali | 2671 |
| 355 | tweeti | 2664 |
| 356 | spor | 2660 |
| 357 | oangeliyor | 2660 |
| 358 | hakkimiz | 2659 |
| 359 | asil | 2652 |
| 360 | mhp | 2652 |
| 361 | altında | 2636 |
| 362 | saglik | 2631 |
| 363 | hakki | 2624 |
| 364 | uzun | 2609 |
| 365 | vardi | 2594 |
| 366 | din | 2583 |
| 367 | arasında | 2583 |
| 368 | suriye | 2573 |
| 369 | sekilde | 2564 |
| 370 | kirazmevsi mi | 2554 |
| 371 | cocuklari | 2554 |
| 372 | yaptigi | 2553 |
| 373 | sandik | 2545 |
| 374 | diyener | 2542 |
| 375 | dur | 2534 |
| 376 | gencler | 2526 |
| 377 | kitap | 2526 |
| 378 | rabbim | 2510 |
| 379 | magdur | 2507 |
| 380 | ozgurluk | 2505 |
| 381 | dikkat | 2502 |
| 382 | gerçekten | 2502 |
| 383 | hemen | 2500 |
| 384 | rahmet | 2493 |
| 385 | aciklamasi | 2492 |
| 386 | iman | 2490 |
| 387 | duyun | 2488 |
| 388 | ulan | 2478 |
| 389 | cizre | 2477 |
| 390 | kucuk | 2474 |
| 391 | ilgili | 2472 |
| 392 | sessiz | 2452 |
| 393 | yada | 2447 |
| 394 | ates | 2447 |
| 395 | imzala | 2442 |
| 396 | tarihi | 2437 |
| 397 | istikrar | 2426 |
| 398 | ailesi | 2412 |

Çizelge 4.5. (devam ediyor).

| | | | | | | | | |
|------------|-------------------|-------------|------------|------------------|-------------|------------|-------------------|-------------|
| 399 | halde | 2407 | 448 | goren | 2231 | 497 | basliyor | 2084 |
| 400 | vermek | 2405 | 449 | olmasin | 2224 | 498 | turlu | 2083 |
| 401 | dunyaya | 2387 | 450 | inadina | 2220 | 499 | suan | 2081 |
| 402 | zulum | 2375 | 451 | sur | 2217 | 500 | bende | 2079 |
| 403 | dizi | 2374 | 452 | engel | 2215 | 501 | basbakan | 2078 |
| 404 | ugruna | 2372 | 453 | engelli | 2208 | 502 | dershane | 2071 |
| 405 | istanbulda | 2365 | 454 | 300 | 2198 | 503 | bolumuyle | 2069 |
| 406 | yasa | 2364 | 455 | zamani | 2196 | 504 | sucu | 2067 |
| 407 | karar | 2363 | 456 | yasinda | 2194 | 505 | hakkinda | 2062 |
| 408 | amk | 2352 | 457 | aldigimiz | 2192 | 506 | fatih | 2060 |
| 409 | ayaga | 2350 | 458 | ceza | 2190 | 507 | mektup | 2060 |
| 410 | hersey | 2349 | 459 | ihanet | 2186 | 508 | dogum | 2060 |
| 411 | askin | 2344 | 460 | yazip | 2177 | 509 | uzerine | 2054 |
| 412 | mehmet | 2344 | 461 | verildi | 2173 | 510 | sonu | 2052 |
| 413 | fav | 2333 | 462 | giden | 2172 | 511 | kurdistan | 2052 |
| 414 | hediyeye | 2326 | 463 | hesap | 2168 | 512 | hazir | 2048 |
| 415 | ister | 2324 | 464 | erkek | 2167 | 513 | davutoglu | 2040 |
| 416 | yapmak | 2317 | 465 | yazik | 2164 | 514 | suc | 2039 |
| 417 | 17aralik | 2308 | 466 | gurur | 2161 | 515 | nesil | 2035 |
| 418 | ders | 2307 | 467 | baba | 2156 | 516 | cocuklarin | 2030 |
| 419 | alan | 2307 | 468 | calisan | 2155 | 517 | unutma | 2024 |
| 420 | icine | 2300 | 469 | iktidar | 2150 | 518 | tikla | 2024 |
| 421 | karsisinda | 2292 | 470 | kemal | 2146 | 519 | gozaltina | 2024 |
| 422 | yuksekk | 2291 | 471 | bol | 2139 | 520 | muebbet | 2024 |
| 423 | eski | 2290 | 472 | formasyon | 2132 | 521 | fragmani | 2021 |
| 424 | misiniz | 2289 | 473 | edip | 2130 | 522 | merkezi | 2017 |
| 425 | sayisi | 2288 | 474 | cikan | 2129 | 523 | yanlis | 2017 |
| 426 | delibal | 2288 | 475 | boyun | 2126 | 524 | cevap | 2016 |
| 427 | herkesin | 2288 | 476 | gonul | 2124 | 525 | emre | 2006 |
| 428 | amed | 2287 | 477 | cozum | 2120 | 526 | oyunu | 2006 |
| 429 | bati | 2287 | 478 | cennet | 2116 | 527 | buyurdu | 2002 |
| 430 | issiz | 2284 | 479 | farkli | 2115 | 528 | unutmayin | 1995 |
| 431 | ulusoy | 2283 | 480 | ihtiyaci | 2114 | 529 | isin | 1995 |
| 432 | istemiyoruz | 2282 | 481 | gelecegi | 2114 | 530 | anne | 1993 |
| 433 | aziz | 2278 | 482 | hava | 2113 | 531 | ortaya | 1992 |
| 434 | rahat | 2278 | 483 | cay | 2113 | 532 | one | 1990 |
| 435 | adami | 2268 | 484 | yasindaki | 2112 | 533 | orgutu | 1988 |
| 436 | paralel | 2264 | 485 | ulkeyi | 2110 | 534 | hedef | 1988 |
| 437 | olursa | 2262 | 486 | verilen | 2107 | 535 | saglikli | 1987 |
| 438 | izin | 2262 | 487 | ziyaret | 2106 | 536 | askla | 1987 |
| 439 | sene | 2261 | 488 | rahmetle | 2102 | 537 | milletin | 1986 |
| 440 | gunde | 2259 | 489 | olup | 2101 | 538 | oyku | 1981 |
| 441 | gidiyor | 2249 | 490 | yola | 2095 | 539 | yolu | 1979 |
| 442 | binlerce | 2242 | 491 | yayalim | 2095 | 540 | yakin | 1978 |
| 443 | okul | 2242 | 492 | isteyen | 2089 | 541 | olduk | 1977 |
| 444 | sende | 2239 | 493 | yaa | 2088 | 542 | sonunda | 1976 |
| 445 | olabilir | 2237 | 494 | den | 2087 | 543 | bizden | 1970 |
| 446 | kampanyayi | 2236 | 495 | hayatini | 2087 | 544 | bilgi | 1970 |
| 447 | dundar | 2232 | 496 | pazar | 2085 | 545 | isler | 1967 |

Çizelge 4.5. (devam ediyor).

| | | |
|------------|----------------------|-------------|
| 546 | etsin | 1950 |
| 547 | olen | 1950 |
| 548 | diyoruz | 1948 |
| 549 | masum | 1946 |
| 550 | akpnin | 1946 |
| 551 | film | 1941 |
| 552 | verdi | 1936 |
| 553 | yildir | 1935 |
| 554 | gecti | 1932 |
| 555 | oglu | 1931 |
| 556 | akitelanet | 1931 |
| 557 | defne | 1927 |
| 558 | yilinda | 1922 |
| 559 | dini | 1921 |
| 560 | sesimizi | 1921 |
| 561 | sesi | 1918 |
| 562 | demokrasi | 1918 |
| 563 | imtihanday iz | 1914 |
| 564 | insanlara | 1910 |
| 565 | 112 | 1906 |
| 566 | ozgecanasla n | 1905 |
| 567 | asker | 1901 |
| 568 | adim | 1899 |
| 569 | yasin | 1896 |
| 570 | kara | 1894 |
| 571 | zafer | 1893 |
| 572 | fuat | 1893 |
| 573 | deger | 1890 |
| 574 | islami | 1889 |
| 575 | rasulullah | 1889 |
| 576 | dost | 1884 |
| 577 | geceler | 1884 |
| 578 | fransa | 1884 |
| 579 | lider | 1880 |
| 580 | seviyorum | 1871 |
| 581 | birlik | 1871 |
| 582 | dogu | 1869 |
| 583 | esit | 1868 |
| 584 | erdem | 1865 |
| 585 | ayca | 1861 |
| 586 | abd | 1860 |
| 587 | gitti | 1857 |
| 588 | versin | 1856 |
| 589 | kisa | 1850 |
| 590 | olanlar | 1850 |
| 591 | hakkini | 1846 |
| 592 | ederiz | 1844 |
| 593 | pkknin | 1844 |
| 594 | gunler | 1843 |
| 595 | devletin | 1842 |
| 596 | askvegurur | 1838 |
| 597 | kursun | 1834 |
| 598 | halkin | 1830 |
| 599 | iran | 1826 |
| 600 | taga | 1826 |
| 601 | merak | 1819 |
| 602 | gundur | 1819 |
| 603 | basinda | 1818 |
| 604 | git | 1818 |
| 605 | aldi | 1818 |
| 606 | koca | 1815 |
| 607 | teog | 1812 |
| 608 | canim | 1810 |
| 609 | ana | 1808 |
| 610 | ici | 1806 |
| 611 | istiklal | 1806 |
| 612 | edecek | 1803 |
| 613 | gulen | 1801 |
| 614 | gereken | 1801 |
| 615 | ilce | 1794 |
| 616 | yasam | 1794 |
| 617 | neler | 1794 |
| 618 | baskan | 1790 |
| 619 | muhammed | 1790 |
| 620 | kemalsunal | 1787 |
| 621 | ikinci | 1781 |
| 622 | helal | 1780 |
| 623 | ipek | 1780 |
| 624 | bosa | 1779 |
| 625 | kadinlar | 1779 |
| 626 | acikladi | 1770 |
| 627 | gec | 1763 |
| 628 | hayati | 1756 |
| 629 | aciask | 1755 |
| 630 | sahibi | 1754 |
| 631 | twit | 1752 |
| 632 | putin | 1750 |
| 633 | katliam | 1749 |
| 634 | sesini | 1745 |
| 635 | olmaya | 1743 |
| 636 | oldum | 1743 |
| 637 | saygiyla | 1739 |
| 638 | iyilik | 1737 |
| 639 | sadakat | 1733 |
| 640 | yolunda | 1732 |
| 641 | net | 1731 |
| 642 | kardesim | 1729 |
| 643 | tutuklandik | 1729 |
| 644 | kurtler | 1728 |
| 645 | kar | 1728 |
| 646 | olmesin | 1728 |
| 647 | gunun | 1725 |
| 648 | 500 | 1722 |
| 649 | dan | 1720 |
| 650 | kalan | 1718 |
| 651 | ankarada | 1716 |
| 652 | bitti | 1715 |
| 653 | namaz | 1711 |
| 654 | sanki | 1703 |
| 655 | ihtiyac | 1703 |
| 656 | misin | 1703 |
| 657 | acilan | 1701 |
| 658 | sonrasi | 1698 |
| 659 | bakalim | 1698 |
| 660 | hain | 1697 |
| 661 | alip | 1695 |
| 662 | osmanli | 1695 |
| 663 | imza | 1694 |
| 664 | cizrede | 1691 |
| 665 | ocak | 1687 |
| 666 | turkiyeye | 1685 |
| 667 | seyyid | 1685 |
| 668 | sonsuz | 1682 |
| 669 | hayal | 1682 |
| 670 | video | 1681 |
| 671 | cezasi | 1681 |
| 672 | beyazshow | 1681 |
| 673 | gununde | 1679 |
| 674 | siteleri | 1676 |
| 675 | tatil | 1668 |
| 676 | yasasin | 1667 |
| 677 | kere | 1659 |
| 678 | kolay | 1654 |
| 679 | ozur | 1652 |
| 680 | yuregi | 1652 |
| 681 | justin | 1652 |
| 682 | askina | 1652 |
| 683 | burasi | 1652 |
| 684 | der | 1651 |
| 685 | ismail | 1647 |
| 686 | olduruldu | 1647 |
| 687 | oylomasinda | 1645 |
| 688 | surda | 1644 |
| 689 | takibi | 1639 |
| 690 | hakaret | 1636 |
| 691 | oglum | 1636 |

Çizelge 4.5. (devam ediyor).

| | | |
|------------|-------------------------|-------------|
| 692 | 00de | 1635 |
| 693 | dunyayi | 1635 |
| 694 | ogretmenle r | 1635 |
| 695 | sart | 1632 |
| 696 | ara | 1632 |
| 697 | hamenei | 1630 |
| 698 | biliyor | 1629 |
| 699 | cikar | 1629 |
| 700 | dava | 1624 |
| 701 | diyorum | 1624 |
| 702 | esi | 1623 |
| 703 | resmi | 1619 |
| 704 | yuz | 1618 |
| 705 | selin | 1617 |
| 706 | aille | 1613 |
| 707 | maci | 1610 |
| 708 | adina | 1608 |
| 709 | dustu | 1608 |
| 710 | yuzunden | 1607 |
| 711 | dahi | 1606 |
| 712 | dostlar | 1606 |
| 713 | gunes | 1606 |
| 714 | bisey | 1605 |
| 715 | silvanda | 1601 |
| 716 | ucagi | 1599 |
| 717 | russia | 1597 |
| 718 | imtihan | 1596 |
| 719 | eglenceli | 1594 |
| 720 | kelepce | 1594 |
| 721 | olamaz | 1591 |
| 722 | sizde | 1590 |
| 723 | sakin | 1590 |
| 724 | gunluk | 1586 |
| 725 | saldiri | 1583 |
| 726 | olma | 1583 |
| 727 | oyun | 1581 |
| 728 | cumhuriyet | 1579 |
| 729 | sorun | 1579 |
| 730 | nusaybin | 1574 |
| 731 | ele | 1573 |
| 732 | deyip | 1571 |
| 733 | iside | 1569 |
| 734 | hayata | 1568 |
| 735 | gul | 1566 |
| 736 | diyerek | 1563 |
| 737 | mal | 1560 |
| 738 | uzerinde | 1558 |
| 739 | allahtan | 1557 |
| 740 | hepimiz | 1556 |
| 741 | sira | 1554 |
| 742 | seninle | 1553 |
| 743 | adamlar | 1552 |
| 744 | gol | 1549 |
| 745 | terk | 1546 |
| 746 | askim | 1542 |
| 747 | istifa | 1540 |
| 748 | teslim | 1540 |
| 749 | edilen | 1540 |
| 750 | tercih | 1539 |
| 751 | atururke | 1539 |
| 752 | atanmak | 1539 |
| 753 | acin | 1536 |
| 754 | tespit | 1535 |
| 755 | degilim | 1534 |
| 756 | kazanmak | 1533 |
| 757 | olunca | 1532 |
| 758 | ucak | 1531 |
| 759 | peki | 1527 |
| 760 | ekmek | 1527 |
| 761 | hoott | 1524 |
| 762 | ocaka | 1524 |
| 763 | lanet | 1520 |
| 764 | hesaptan | 1517 |
| 765 | onlari | 1517 |
| 766 | uzak | 1515 |
| 767 | bolumu | 1514 |
| 768 | basi | 1513 |
| 769 | ederim | 1511 |
| 770 | alinan | 1510 |
| 771 | suresi | 1505 |
| 772 | aydin | 1499 |
| 773 | kisinin | 1499 |
| 774 | demis | 1499 |
| 775 | hergun | 1498 |
| 776 | pazartesi | 1498 |
| 777 | yeri | 1494 |
| 778 | etme | 1491 |
| 779 | aciga | 1491 |
| 780 | katili | 1490 |
| 781 | suriyede | 1489 |
| 782 | kullandigi | 1489 |
| 783 | reis | 1489 |
| 784 | mutlaka | 1489 |
| 785 | davasinda | 1489 |
| 786 | resmen | 1487 |
| 787 | hilesi | 1486 |
| 788 | gelin | 1483 |
| 789 | baziinsanlar | 1482 |
| 790 | haklari | 1480 |
| 791 | islamin | 1478 |
| 792 | stara | 1477 |
| 793 | kesin | 1476 |
| 794 | yapar | 1476 |
| 795 | mukemmel | 1474 |
| 796 | sevgili | 1474 |
| 797 | olmadigi | 1472 |
| 798 | olacaktır | 1468 |
| 799 | kadina | 1467 |
| 800 | hesabini | 1465 |
| 801 | kiralikaskin | 1463 |
| 802 | dusman | 1456 |
| 803 | hale | 1455 |
| 804 | olu | 1453 |
| 805 | mavi | 1453 |
| 806 | yapalim | 1452 |
| 807 | hukumet | 1448 |
| 808 | yas | 1448 |
| 809 | ardindan | 1448 |
| 810 | ayri | 1445 |
| 811 | defa | 1442 |
| 812 | birakin | 1442 |
| 813 | kitabı | 1440 |
| 814 | muhtesem | 1440 |
| 815 | kimsenin | 1438 |
| 816 | dogan | 1437 |
| 817 | birliği | 1436 |
| 818 | mutluluk | 1435 |
| 819 | bey | 1433 |
| 820 | nur | 1432 |
| 821 | isik | 1432 |
| 822 | karari | 1427 |
| 823 | yigit | 1427 |
| 824 | begen | 1427 |
| 825 | insana | 1425 |
| 826 | puan | 1424 |
| 827 | bilir | 1423 |
| 828 | tane | 1423 |
| 829 | ustune | 1423 |
| 830 | nefret | 1422 |
| 831 | bomba | 1421 |
| 832 | mesaj | 1419 |
| 833 | alin | 1418 |
| 834 | sivil | 1418 |
| 835 | varken | 1417 |
| 836 | guc | 1417 |
| 837 | evde | 1415 |
| 838 | iskence | 1414 |

Çizelge 4.5. (devam ediyor).

| | | |
|------------|-------------------|-------------|
| 839 | dusen | 1413 |
| 840 | fark | 1410 |
| 841 | hapis | 1404 |
| 842 | etmeyin | 1403 |
| 843 | verir | 1402 |
| 844 | basarilar | 1401 |
| 845 | cocugun | 1401 |
| 846 | teпки | 1400 |
| 847 | hakim | 1400 |
| 848 | soguk | 1399 |
| 849 | kurban | 1399 |
| 850 | oncesi | 1398 |
| 851 | budur | 1396 |
| 852 | gider | 1395 |
| 853 | gelecege | 1393 |
| 854 | hakkimizi | 1393 |
| 855 | kimligi | 1392 |
| 856 | bari | 1391 |
| 857 | beri | 1390 |
| 858 | sozleri | 1389 |
| 859 | kontenjan | 1386 |
| 860 | rte | 1385 |
| 861 | cik | 1384 |
| 862 | olsaydi | 1384 |
| 863 | tagi | 1383 |
| 864 | avrupa | 1382 |
| 865 | gazi | 1381 |
| 866 | gunku | 1379 |
| 867 | ogretmeni | 1377 |
| 868 | ait | 1374 |
| 869 | atan | 1374 |
| 870 | yuzunu | 1374 |
| 871 | gundem | 1373 |
| 872 | serbest | 1373 |
| 873 | hakli | 1372 |
| 874 | kis | 1371 |
| 875 | bakan | 1370 |
| 876 | ummet | 1365 |
| 877 | veriyor | 1365 |
| 878 | zulme | 1364 |
| 879 | yazdi | 1364 |
| 880 | burda | 1363 |
| 881 | katledildi | 1359 |
| 882 | yana | 1358 |
| 883 | kisiye | 1355 |
| 884 | arkadas | 1354 |
| 885 | super | 1353 |
| 886 | cehennem | 1350 |
| 887 | gunlerden | 1349 |
| 888 | halka | 1349 |
| 889 | kanal | 1348 |
| 890 | davet | 1342 |
| 891 | devleti | 1342 |
| 892 | sozu | 1342 |
| 893 | elcinin | 1341 |
| 894 | tehdit | 1340 |
| 895 | agliyor | 1339 |
| 896 | onune | 1339 |
| 897 | icim | 1338 |
| 898 | yilda | 1336 |
| 899 | milleti | 1336 |
| 900 | december | 1336 |
| 901 | askeri | 1335 |
| 902 | deniz | 1329 |
| 903 | yayinda | 1329 |
| 904 | yukseldi | 1327 |
| 905 | agac | 1325 |
| 906 | hirsiz | 1323 |
| 907 | doguda | 1320 |
| 908 | kufur | 1319 |
| 909 | yapip | 1319 |
| 910 | durum | 1318 |
| 911 | kor | 1317 |
| 912 | kardes | 1311 |
| 913 | ulkeye | 1310 |
| 914 | akin | 1310 |
| 915 | hareket | 1309 |
| 916 | hukuk | 1307 |
| 917 | alim | 1307 |
| 918 | millite | 1306 |
| 919 | yilin | 1305 |
| 920 | dilekdogan | 1305 |
| 921 | sinan | 1301 |
| 922 | savasa | 1301 |
| 923 | arada | 1300 |
| 924 | etmis | 1299 |
| 925 | gomez | 1297 |
| 926 | tayyip | 1296 |
| 927 | sizce | 1295 |
| 928 | dedim | 1293 |
| 929 | nasip | 1293 |
| 930 | zamanda | 1293 |
| 931 | izle | 1292 |
| 932 | anadolu | 1292 |
| 933 | nefes | 1291 |
| 934 | turkmenler | 1291 |
| 935 | yasamini | 1289 |
| 936 | tahirelcinin | 1289 |
| 937 | isim | 1288 |
| 938 | topraga | 1287 |
| 939 | gazeteci | 1285 |
| 940 | fransiz | 1284 |
| 941 | katledilen | 1283 |
| 942 | sicak | 1282 |
| 943 | yapsin | 1281 |
| 944 | onur | 1281 |
| 945 | olmadan | 1276 |
| 946 | elinde | 1275 |
| 947 | ataturkun | 1274 |
| 948 | istedigi | 1273 |
| 949 | sozde | 1272 |
| 950 | haberi | 1272 |
| 951 | cocuga | 1270 |
| 952 | van | 1266 |
| 953 | ciddi | 1266 |
| 954 | kirmizi | 1265 |
| 955 | kul | 1263 |
| 956 | kiniyoruz | 1263 |
| 957 | inat | 1262 |
| 958 | oyuncu | 1261 |
| 959 | isleri | 1260 |
| 960 | edelim | 1258 |
| 961 | tvde | 1254 |
| 962 | cumartesi | 1254 |
| 963 | alindi | 1254 |
| 964 | cicek | 1253 |
| 965 | soru | 1253 |
| 966 | yillar | 1252 |
| 967 | hoca | 1251 |
| 968 | girdi | 1251 |
| 969 | edenler | 1250 |
| 970 | adil | 1249 |
| 971 | gazetesi | 1248 |
| 972 | bayrak | 1245 |
| 973 | amin | 1241 |
| 974 | yasak | 1241 |
| 975 | dik | 1241 |
| 976 | inadinaask | 1241 |
| 977 | peygamber | 1241 |
| 978 | basini | 1241 |
| 979 | lideryaoal | 1241 |
| 980 | kiralik | 1237 |
| 981 | kesinlikle | 1237 |
| 982 | silopi | 1236 |
| 983 | laf | 1234 |
| 984 | isterdimki | 1233 |
| 985 | yarali | 1231 |

Çizelge 4.5. (devam ediyor).

| | | |
|-----|-----------|------|
| 986 | sarki | 1230 |
| 987 | yargi | 1230 |
| 988 | baskanlik | 1228 |
| 989 | bayan | 1228 |
| 990 | olmadi | 1227 |
| 991 | aska | 1225 |

| | | |
|-----|-----------------|------|
| 992 | rahatsiz | 1223 |
| 993 | benden | 1222 |
| 994 | insaallah | 1222 |
| 995 | vefatinin | 1221 |
| 996 | altina | 1219 |
| 997 | muminedur us | 1216 |

| | | |
|------|---------|------|
| 998 | the | 1215 |
| 999 | vermeye | 1215 |
| 1000 | iftira | 1212 |

Çizelge 4.5’de ki kelimeler incelenerek Türkçe dilinde kelime kullanımları ve anlamsal etkileri göz önünde bulundurularak 500 kelimedenden oluşan bir kelime seti oluşturulmuştur. Çizelge 4.5’de seçilen 500 kelime seti koyu olarak gösterilmiştir.

4.2.2.4. 7’den Az Kullanılan Tag’lerin İptali

Hashtag’lerin kullanım sayısı arttıkça bağlı oldukları Tweet metinleri ve bu metinlerde geçen kelime sayısı da artacaktır. Belli bir sayının altında kullanılan hashtagler sınırlı sayıda kelime ile eşleşeceğinden anlamlandırılması zorlaşacaktır. Yazılımsal olarak çeşitli eşik değerleri belirlenerek çalışmalar yapılmış ve bu çalışma sonuçlarına göre eşik değeri 7 olarak belirlenmiştir. 7’den az tekrarla kullanılan hashtagler ve bağlı olduğu Tweetler temizlenmiştir. Bu işlem sonucunda 1032812 adet hashtag ve Tweet verisi kalmıştır. Hashtag’ler düğümleri oluşturacağı için tekilleştirilmiştir. Tekil olarak belirlenen 10169 adet hashtag, semantik ağın düğümlerini oluşturmuştur. Her bir hashtag birlikte kullanıldığı kelimeler ile eşleştirilmiş ve hashtag-kelime birlikteliği tablosu oluşturulmuştur. Bu ilişki de ağın linklerini tanımlamaktadır. Bu prosedür sonucunda 235650 adet link oluşmuştur.

4.2.2.5. Semantik Yakınlık Hesaplama

TF

Düğüm tablosunu oluşturan her bir hashtag’in hangi kelimeler ile birlikte kullanıldığı incelenerek, bu birlikteliklerden tag_kelime birliktelik tablosu oluşturulmuştur. Hashtag ve kelime birlikte kullanım tekrarlarında, tekrar sayıları arttırılmıştır. Böylece her bir hashtag ve kelime birlikteliğinin veri setinde kaç defa geçtiğine dair

sayılar bulunmuştur. Bu tekrar sayıları Bölüm 3’de anlatıldığı gibi TF verisini vermektedir. Bu çalışmada bir hashtag ile birlikte kullanılan kelimelerin tekrar sayıları, o kelimeye ait TF değerini vermektedir. Kelimeler, farklı hashtag’leri için farklı TF değerlerine sahiptir. Hesaplanan TF değerleri tag-kelime tablosunda oluşturulan TF sütununa yazdırılmıştır.

IDF

Bölüm 3’de anlatıldığı gibi kelimelerin tekrar sayılarının yanında, kelimelerin kaç farklı dökümanda kullanıldığına da önemli olmasından dolayı IDF verisi semantik yakınlık hesaplamasında önemlidir. Çalışmamızda her bir hashtag ile ilişkili Tweet kümesi bir doküman olarak kabul edilmiştir. Eşitlik 3.1’deki formül ile IDF verileri hesaplanmış ve kelime tablosunda IDF sütunu oluşturularak her kelime için hesaplanan değerler veritabanına kaydedilmiştir.

TF-IDF

Kelimelerin kaç defa geçtiğinin yanında kaç farklı dökümanda olduğunun da önem kazandığı TF-IDF verisi Bölüm 3’de anlatılan TF-IDF hesaplama formülasyonuna uygun olarak hesaplanmıştır. Her hashtag’a ait kelime listesinin yanında TFIDF sütunu oluşturularak hesaplanan değerler bu hücrelere yerleştirilmiştir.

L2 Normalizasyonu

Bir hashtag’e ait TF-IDF vektörünün Bölüm 3’de anlatıldığı gibi normalize edilmesi gerekmektedir. Normalizasyon için, hashtag’e ait vektörün tüm boyutlardaki uzunluklarının karelerinin toplamının karekökü alınarak uzunluk bulunur. Tüm elemanlar bu uzunluğa bölünerek normalize edilmiş olur.

Cosine Yakınlık Hesaplama

İki hashtag’in cosine yakınlıkları incelenirken, Bölüm 3’de anlatılan yöntem ile cosine değeri hesaplanır. Bu değer 1’e yaklaştıkça hashtag’ler arasındaki semantik

yakınlık artacaktır. Cosine değeri, hashtag vektörlerine ait L2 normalizasyon değerleri çarpılarak bulunur.

Cosine Eşiğini Belirleme

Hashtag çiftlerinin normalize TFIDF vektörlerinin skalar çarpımı ile elde edilen Cosine değerleri göz ile incelenmiş, anlamsal yakınlığın yaklaşık 0,5 değeri üzerinde daha doğru eşleşmeler sunduğu sonucuna varılmıştır. Bu nedenle semantik bağların varlığı için Cosine > 0,5 eşik değeri kabul edilmiştir. Cosine 0,5'in üstünde olan edge'lerin semantik bağı bulunduğu kabul edilmiştir. 67111 adet edge bu eşik değerinin üstündedir. Çizelge 4.6'da cosine yakınlık değeri en yüksek 100 hashtag birlikteliği görülmektedir.

Çizelge 4.6. Cosine yakınlık değeri en yüksek 100 hashtag birlikteliği.

| sıra | source_node | target_node | cosine |
|------|--|--|----------|
| 1 | abluka38ncigununde | abluka40ncigununde | 1 |
| 2 | abluka37ncigununde | ablukalarkalksin | 1 |
| 3 | abluka39ncugununde | abluka40ncigununde | 1 |
| 4 | abluka37ncigununde | abluka38ncigununde | 1 |
| 5 | abluka37ncigununde | abluka39ncugununde | 1 |
| 6 | abluka39ncugununde | ablukalarkalksin | 1 |
| 7 | abluka40ncigununde | ablukalarkalksin | 1 |
| 8 | abluka37ncigununde | abluka40ncigununde | 1 |
| 9 | abluka38ncigununde | abluka39ncugununde | 1 |
| 10 | abluka38ncigununde | ablukalarkalksin | 1 |
| 11 | followingclup | takipgirl | 0.999581 |
| 12 | habar | simditekistedigim | 0.998695 |
| 13 | grammys | videomtv2015 | 0.998181 |
| 14 | hayatsevince | sevgisaygibarishuzur | 0.997789 |
| 15 | tatilinsongunutumfamdomlartakiplesiyor | tatilkeyfinicikaranfandomlartakiplesiyor | 0.997755 |
| 16 | memursinavsorulari | uykumualamadigimda | 0.997669 |
| 17 | kahvaltihakberleri | sagliklimutluhuzurlu | 0.997281 |
| 18 | hayatimda1defa | memursinavsorulari | 0.997159 |
| 19 | 11ocaksaglikcilargununde | dinkulturuveimamhatip500talepediyor | 0.997124 |
| 20 | hayatimda1defa | uykumualamadigimda | 0.997037 |
| 21 | kizimiistiyorum | seyahatetmekozgurluktur | 0.996662 |
| 22 | takipclub | takipedenintakipederim | 0.996228 |
| 23 | formasyonahayir | formasyonlular2ncioncelikliatansin | 0.996174 |
| 24 | boniboncuqilefandomkardeslertakiplesiyor | yarinyazilisolanperisanfandomlartakiplesiyor | 0.995829 |
| 25 | aktakipbasladi | yenidendirilis | 0.995770 |
| 26 | denizverosalindailefandomlartakiplesiyor | tatilkeyfinicikaranfandomlartakiplesiyor | 0.994763 |
| 27 | evlenecekleremesajim | habar | 0.994569 |
| 28 | gucenme | mafya | 0.994378 |
| 29 | tatilkeyfinicikaranfandomlartakiplesiyor | taylanvealeyfandomtakiplesmesi | 0.994315 |
| 30 | evlenecekleremesajim | simditekistedigim | 0.994287 |
| 31 | emojiliuzayilemubarekfandomlartakiplesiyor | tatilkeyfinicikaranfandomlartakiplesiyor | 0.994187 |
| 32 | denizverosalindailefandomlartakiplesiyor | tatilinsongunutumfamdomlartakiplesiyor | 0.994070 |
| 33 | seymvepikooilefandomlartakiplesiyor | tatilkeyfinicikaranfandomlartakiplesiyor | 0.994034 |

Çizelge 4.6. (devam ediyor).

| | | | |
|----|--|---|----------|
| 34 | seymvepikooilefandomlartakiplesiyor | tatilinsongunutumfamdomlartakiplesiyor | 0.993887 |
| 35 | ketiketesrailetumtirrekfandomlartakiplesiyor | pazargunuietumfondomlartakiplesiyor | 0.993737 |
| 36 | dinkulturuveimamhatip500talepediyor | issizisvemeslekdanismanlari | 0.993696 |
| 37 | denizverosalindailefandomlartakiplesiyor | seymvepikooilefandomlartakiplesiyor | 0.993587 |
| 38 | enpishuyum | usandimartik | 0.993363 |
| 39 | 70bintibbisekretercozumubekliyor | tibbisekreterlerinkadrohakki | 0.993263 |
| 40 | tibbisekreterlerasmdekadrohakki | tibbisekreterlererecozum | 0.993076 |
| 41 | giyimogretmenleriyuksekpuanlarlaatamabekli | halk | 0.992892 |
| 42 | ugurmumcu | vicdanlardamahkumsunuz | 0.992826 |
| 43 | enpishuyum | sabahkeske | 0.992765 |
| 44 | saglicitibbisekreterlersesleniyor | tibbisekreterlerhaklariniariyor | 0.992675 |
| 45 | 1kasimbenimicin | gucenme | 0.992343 |
| 46 | dualarlasunnetiseniyeyihya | ertugrulgazi | 0.992325 |
| 47 | sabahkeske | usandimartik | 0.992244 |
| 48 | larryveziamshipperlartakiplesiyor | pazargunufandomlartakiplesiyor | 0.992205 |
| 49 | denizverosalindailefandomlartakiplesiyor | taylanvealeyfandomtakiplesmesi | 0.992181 |
| 50 | khaleesivezgiiletumfandomlartakiplesiyor | tatilinsongunutumfamdomlartakiplesiyor | 0.992101 |
| 51 | bizimyuvamiz | dikkat | 0.992069 |
| 52 | tatilinsongunutumfamdomlartakiplesiyor | taylanvealeyfandomtakiplesmesi | 0.991970 |
| 53 | saatinkacoldugucomezemeyenfandomlartakiplesi | yarinyazilisolanperisanfandomlartakiplesiyor | 0.991820 |
| 54 | potaninkraliceleri | sarimelekler | 0.991807 |
| 55 | pazargunuietumfondomlartakiplesiyor | pofudukornitorenkilefandomlartakiplesiyor | 0.991595 |
| 56 | tibbisekreterlercozumubekliyor | tibbisekreterlerinkadrohakki | 0.991576 |
| 57 | emojiliuzayilemubarekfandomlartakiplesiyor | tatilinsongunutumfamdomlartakiplesiyor | 0.991463 |
| 58 | 11ocaksaglikcilargununde | issizisvemeslekdanismanlari | 0.991349 |
| 59 | 70bintibbisekretercozumpesinde | tibbisekreterleremeginiistiyor | 0.991315 |
| 60 | tibbisekreterlerasmdekadrohakki | tibbisekreterlerinkadrohakki | 0.991199 |
| 61 | ketiketesrailetumtirrekfandomlartakiplesiyor | saatinkacoldugucomezemeyenfandomlartakiplesiyor | 0.991098 |
| 62 | 70bintibbisekretercozumuistiyor | tibbisekreterlererecozum | 0.990971 |
| 63 | nissiletumfandomlarkazaniyor | sinavlardansikilanfandomlartakiplesiyor | 0.990805 |
| 64 | seymvepikooilefandomlartakiplesiyor | taylanvealeyfandomtakiplesmesi | 0.990773 |
| 65 | 70bintibbisekretercozumubekliyor | tibbisekreterlerasmdekadrohakki | 0.990758 |
| 66 | saatinkacoldugucomezemeyenfandomlartakiplesi | seymvepikooilefandomlartakiplesiyor | 0.990709 |
| 67 | saatinkacoldugucomezemeyenfandomlartakiplesi | taylorindogumgunundefandomlartakiplesiyor | 0.990659 |
| 68 | pazargunuietumfondomlartakiplesiyor | pazargunutumkekofandomlartakiplesiyor | 0.990557 |
| 69 | stadyum11 | stadyumeniyi | 0.990479 |
| 70 | tasariverildisozsizde | tavsiye | 0.990457 |
| 71 | noronlamutlupazarlaravedaeredektakiplesiyor | taylanvealeyfandomtakiplesmesi | 0.990420 |
| 72 | enpishuyum | gicikolum | 0.990406 |
| 73 | ketiketesrailetumtirrekfandomlartakiplesiyor | taylorindogumgunundefandomlartakiplesiyor | 0.990339 |
| 74 | 1kasimbenimicin | mafya | 0.990313 |
| 75 | saripiveogumceklegvagilefandomlartakiplesi | seymvepikooilefandomlartakiplesiyor | 0.990222 |
| 76 | amsterdam | matchday | 0.990070 |
| 77 | denizverosalindailefandomlartakiplesiyor | emojiliuzayilemubarekfandomlartakiplesiyor | 0.989729 |
| 78 | tibbisekereterlerinsesiniduyun | tibbisekreterlerinkadrohakki | 0.989632 |
| 79 | bayangrangerilepazargunutumfandomlartakip | pofudukornitorenkilefandomlartakiplesiyor | 0.989535 |
| 80 | tibbisekereterlerinsesiniduyun | tibbisekreterlercozumubekliyor | 0.989473 |
| 81 | benlisede | sabahkeske | 0.989457 |
| 82 | ozelegitimalandisiatamaistemiyor | ozelegitimalandisinakarsi | 0.989432 |
| 83 | ketiketesrailetumtirrekfandomlartakiplesiyor | noronlamutlupazarlaravedaeredektakiplesiyoruz | 0.989400 |
| 84 | benlisede | enpishuyum | 0.989395 |
| 85 | benlisede | usandimartik | 0.989209 |
| 86 | emojiliuzayilemubarekfandomlartakiplesiyor | sutlumilqaciqvenutellaylacikolatasevenfandomlar | 0.989114 |
| 87 | benceteror | enpishuyum | 0.989090 |
| 88 | kazimkoyuncu | kazimkoyuncu44yasinda | 0.988933 |
| 89 | emojiliuzayilemubarekfandomlartakiplesiyor | taylanvealeyfandomtakiplesmesi | 0.988902 |
| 90 | reyelquyivenisaiselsecimgunufandomlartakipl | saripiveogumceklegvagilefandomlartakiplesiyor | 0.988900 |

Çizelge 4.6. (devam ediyor).

| | | | |
|-----|---|---|----------|
| 91 | bayangrangeriletumfandomlartakiplesiyor | okuldanbikanfandomlartakiplesiyor | 0.988774 |
| 92 | gicikolurum | usandimartik | 0.988718 |
| 93 | noronlamutlupazarlaravedaeredektakiplesiyor | quyinciqilefandomlarseritakiplesiyor | 0.988537 |
| 94 | kerimfrei | velikavlak | 0.988421 |
| 95 | hallowendanoncetumfandomlartakiplesiyor | seymvepikooilefandomlartakiplesiyor | 0.988376 |
| 96 | hzmevlana | imtihan | 0.988296 |
| 97 | askdrew | dondumvededimki | 0.988266 |
| 98 | bugeceyuregimderki | enpishuyum | 0.988259 |
| 99 | quyinciqilefandomlarseritakiplesiyor | taylanvealeyfandomtakiplesmesi | 0.988255 |
| 100 | kirezimukotatilderemmyhazelbucurilefando | saripipiveogumceklegvagilefandomlartakiplesiyor | 0.988238 |

Çizelge 4.7’ de ise 0,5 eşik değeri uygulandıktan sonra en düşük cosine değerine sahip 100 hashtag birlikteliği gösterilmiştir. Eşik değere yaklaştıkça anlamsal yakınlığın azaldığı, ancak halen hashtag’ler arasında doğru ya da dolaylı bazı benzerlikler olduğu gözlenmektedir.

Çizelge 4.7. Cosine yakınlık değeri en düşük 100 hashtag birlikteliği.

| sıra | source_node | target_node | cosine |
|------|--|--------------------------------|----------|
| 1 | draynurdagdemir | italyan | 0.500002 |
| 2 | hafiza2015 | soma | 0.500004 |
| 3 | beyazshowyilbasiozel | oyunusevensaysin | 0.500006 |
| 4 | acil | pendik | 0.500009 |
| 5 | buyukturkiyeguclugetecek | milliyayinahaber | 0.500010 |
| 6 | ttdestekolmayageldi | weloveyouselena | 0.500014 |
| 7 | parkedekazandikmasadakaybetmeyecegiz | tecavuz | 0.500020 |
| 8 | mutlulukdediginbence | turkiyedunyagundemine | 0.500021 |
| 9 | 5ocakadanankurtulusu | insanhaklari | 0.500024 |
| 10 | vururyuzeifadesihamzaaydogdusosyalinbitane | w20 | 0.500027 |
| 11 | 2015detribunde | komurudurdur | 0.500031 |
| 12 | bedenegitimicandir300atamahaktir | dunyafelsefegunu | 0.500032 |
| 13 | googledoodle | yanindayizortac | 0.500034 |
| 14 | cumhurbaskani | sakagibi | 0.500039 |
| 15 | cocuklaricinhemenbaris | encokbegendigimkapak | 0.500040 |
| 16 | babaevi | potaninkartallari | 0.500043 |
| 17 | adana2017yegiriyor | olmazdeme | 0.500044 |
| 18 | akpartiyapardedikmerkeziatamaya20binkadroi stedik | arenadabudamavar | 0.500045 |
| 19 | kurtlarvadisipusu270 | twitpalas | 0.500046 |
| 20 | 4buyuklersalonturnuvasi | stadyumonhafta | 0.500048 |
| 21 | kilicdaroglukonusuyor | worldtoiletday | 0.500055 |
| 22 | benceask | soylesenesevgili | 0.500064 |
| 23 | barisiedirmeyiz | hellobtchzitscl | 0.500066 |
| 24 | bsl | efsanetarafaraefsanetarayici | 0.500067 |
| 25 | mutlulukdenince | serifyusuf | 0.500074 |
| 26 | g20dagitilacak | geziyihatirla | 0.500079 |
| 27 | 25kasimkadinasiddetehayir | siddetekarsidur | 0.500081 |
| 28 | erdemgul | yilbasibence | 0.500081 |
| 29 | duyugumenkotuespri | ziyaince | 0.500082 |
| 30 | akpisrailliseviyorcunku | hatirlatirim | 0.500082 |
| 31 | iyikidogdungalcinergul | kardesim | 0.500083 |
| 32 | rusyasavassucuisliyor | uyumatuerkiyesahipciktuerkmene | 0.500085 |

Çizelge 4.7. (devam ediyor).

| | | | |
|----|--|---|----------|
| 33 | bittigiman | ikramiyebanaciksa | 0.500089 |
| 34 | favorimigeriver | hantoluyumugibisibirdahagelmez | 0.500091 |
| 35 | gidamuhendisialiminezaman | mahkumasartlisaliverme | 0.500093 |
| 36 | ahaberimedokunma | allahimaf | 0.500096 |
| 37 | psikolojibozulankucukayilartakiplesiyor | secimdensikilanfandomlartakiplesiyor | 0.500097 |
| 38 | birgaripask | mayiskralicesi | 0.500099 |
| 39 | burakyilmazsatilsin | stadyumgoller | 0.500102 |
| 40 | askateslim | paramparca | 0.500103 |
| 41 | egitimajansi | enazindan | 0.500105 |
| 42 | psikolojibozulankucukayilartakiplesiyor | yilbasimerkeztakiplesiyorherkes | 0.500108 |
| 43 | iyikivarsinnilayduru | iyisenelerysegulaldinc | 0.500110 |
| 44 | delinindelisiolsam | hayalimdekiask | 0.500110 |
| 45 | cagatayulusoylabesyil | inegolspor | 0.500111 |
| 46 | acaydimgollarimioppacedaldiyedimdiyenerler ekpopfandomutakiplesiyor | quyinciqiletumfandomlarunfollowsuztakiplesiyor | 0.500113 |
| 47 | aycaemre | buyukaskanes | 0.500113 |
| 48 | beceriyunlari | yaratincinotlar | 0.500115 |
| 49 | bayirbucakturkiyedir | lazkiye | 0.500115 |
| 50 | ahiskasoykirimim71yasinda | calintiiktidar | 0.500116 |
| 51 | defomsahneleriartsin | yildiztilbe | 0.500120 |
| 52 | seymakorkmaz | turkiyedekadinolmak | 0.500122 |
| 53 | burdayimde | kizlarinayarolduguerkektipi | 0.500122 |
| 54 | kotukediserafetin | yinegeceolunca | 0.500122 |
| 55 | burcuozberk | directionersarelisteningtoperfect | 0.500127 |
| 56 | bilgi | giresunspor | 0.500129 |
| 57 | cagrimerkeziolumune500atama | cagrimerkezihizmtleri haziranda500atama | 0.500130 |
| 58 | islambes | mcfc | 0.500130 |
| 59 | belkianlarsin | gsbyazarbulusmalari | 0.500131 |
| 60 | hesabisorulur | kursunlucamii | 0.500131 |
| 61 | askimbenialdatirsan | kimisidesaniyorki | 0.500132 |
| 62 | bulunamadi | merkeziatamakadrolarincekisiyormezunlarferyat ediyor | 0.500136 |
| 63 | azizgueler | kimyaya200iletakviyekurslariboskalmasin | 0.500136 |
| 64 | gtyapangiller | mamatopluyoruz | 0.500138 |
| 65 | busabah | chpliderinesoruyorum | 0.500139 |
| 66 | duayenlerle1gun | nihatlamuhabbet | 0.500142 |
| 67 | defomsahneleriartsin | emrecondeyince | 0.500150 |
| 68 | asikolamasam | uyandigimdailkis | 0.500154 |
| 69 | ozelegitimealandisiatamayahayir | yfyi2015 | 0.500154 |
| 70 | aslindaask | karanlikask2 | 0.500156 |
| 71 | merve | yay | 0.500159 |
| 72 | ihslimemurmuisimi | orkunasorun | 0.500164 |
| 73 | haddinizibilinvatanmilliyet | iddaa | 0.500165 |
| 74 | kardes | silahlarinsusmasiicin | 0.500166 |
| 75 | birturluogrenemedin | hayalimdekiask | 0.500167 |
| 76 | flash | idlip | 0.500169 |
| 77 | azizgulerezulmun40incigunu | cumhurbaskanimiz | 0.500174 |
| 78 | sontopakadar | trtbencedevmac | 0.500176 |
| 79 | bjk | hedef1903cocuk | 0.500177 |
| 80 | 1hayatdiliyoruz | istanbulescort | 0.500177 |
| 81 | gencgelisimsempozyumu4 | kanavice | 0.500181 |
| 82 | icimdengeceniyaziyorum | vururyuzeifadesiilkmata4binbitanesi | 0.500185 |
| 83 | kurdsinfreedomwar | ozyonetimleozgurkurdistan | 0.500190 |
| 84 | degisiklikolsun | licedavasi | 0.500191 |
| 85 | isvemeslekdanismanlari | isvemeslekdanismanlari3yildir | 0.500191 |
| 86 | kuralsiz | uykulobisi | 0.500192 |
| 87 | adanalicocugukadikoyedavetedin | android | 0.500195 |
| 88 | firat | yuva | 0.500200 |

Çizelge 4.7. (devam ediyor).

| | | | |
|-----|---|-----------------------------------|----------|
| 89 | arshisessiz | serkanozel | 0.500205 |
| 90 | ilkdokunus | yenidenkalbimegedekaldi | 0.500207 |
| 91 | hayatsevince | sevdamevsimindenkalan | 0.500210 |
| 92 | pembemetrobustiyoruz | yilmakyokdirenisvar | 0.500210 |
| 93 | muhammed | muhammedaltunbas | 0.500214 |
| 94 | aytil | benimutluetmeyeyeter | 0.500219 |
| 95 | beyazfutbolsizinle | taftaryandextenersunyanaliistiyor | 0.500220 |
| 96 | acaydingollarimioppacediyeyimdiyenerler | amacsizsayfa | 0.500221 |
| 97 | direnisibuyutuyoruz | kurdistandireniyor | 0.500225 |
| 98 | gurur | singforyou | 0.500227 |
| 99 | benimicinannem | karlihavalarda | 0.500228 |
| 100 | dunya | russasirmasabrimizitasirma | 0.500230 |

4.2.3. İki Ağ Çalışmasının Ortak Kümesinin İncelenmesi

İki ağ çalışmasında farklı yöntemler denendiği ve farklı ön işleme prosedürleri uygulandığı için farklı düğüm setleri ile analiz yapılmıştır. İkisinin karşılaştırılabilmesi için ortak bir düğüm küme setine ihtiyaç vardır. Bu amaçla her iki çalışmanın düğüm setleri incelenmiş ve her ikisinde de bulunan hashtag'lerden ortak bir düğüm seti kümesi çıkarılmıştır.

Düğüm seti, her iki çalışmada da bulunan düğümlere indirildiğinde, ortak düğüm kümesi 7647 adet düğümden oluşmaktadır. Bu düğümler ile oluşmuş olan edge sayıları;

1. çalışmada (birlikte kullanım) 61612
2. çalışmada (semantik yakınlık) 42270 adet olarak tespit edilmiştir.

Her iki çalışmada da aynı source ve target'e sahip bağlantı sayısı ise 2247 adet olmuştur. Ortak kümedeki linkler'den en yüksek cosine değerine sahip ilk 100 linkçizelge 4.8'de görülmektedir.

Çizelge 4.8. Ortak kümenin cosine değeri en yüksek 100 hashtag birlikteliği.

| sıra | source_node | target_node | weight | cosine |
|------|--|-------------------------------------|--------|----------|
| 1 | grammys | videomtv2015 | 2 | 0.998181 |
| 2 | memursinavsorulari | uykumualamadigimda | 8 | 0.997669 |
| 3 | denizverosalindailefandomlartakiplesiyor | seymvepikooilefandomlartakiplesiyor | 1 | 0.993587 |
| 4 | enpishuyum | usandimartik | 2 | 0.993363 |
| 5 | potaninkraliceleri | sarimelekler | 1 | 0.991807 |
| 6 | enpishuyum | karsicinsimdenricam | 2 | 0.987604 |
| 7 | karsicinsimdenricam | usandimartik | 8 | 0.985714 |

Çizelge 4.8. (devam ediyor).

| | | | | |
|----|--|---|----|----------|
| 8 | vasippazartesitatildesene | vasipsahin | 19 | 0.983976 |
| 9 | bildigingibi | sabahkeske | 3 | 0.981808 |
| 10 | erkencitayfa | gunaydin | 56 | 0.981453 |
| 11 | ketiketesrailetumtirrekfandomlartakiplesiyor | pazargunutumkekofandomlartakiplesiyor | 18 | 0.980222 |
| 12 | pazargunuiiletumfondomlartakiplesiyor | saatinkacoldugucomezemeyenfandomlartaki | 1 | 0.980144 |
| 13 | sarayinsultanlari | yenilmezarmada | 16 | 0.979736 |
| 14 | dondumvededimki | muminleresoruyorum | 7 | 0.977386 |
| 15 | cumartesi gunutumkekofandomlartakiplesiyor | limovevillyiletumfondomlartakiplesiyor | 13 | 0.974053 |
| 16 | hadisis | hadisiserif | 45 | 0.973578 |
| 17 | hayatimda l defa | uzulmekistemiyorsan | 17 | 0.973401 |
| 18 | bakanaydin | karsicinsimdenricam | 4 | 0.971745 |
| 19 | fenerinmacivar | matchday | 13 | 0.969590 |
| 20 | khalesiveezgiiletumfondomlartakiplesiyor | secimdensikilanfandomlartakiplesiyor | 1 | 0.967353 |
| 21 | benceteror | coksevdikcunku | 2 | 0.967325 |
| 22 | haftasonununmutlulugunuyasayanfandomlarta kiplesiyor | larryveziamskipperlartakiplesiyor | 1 | 0.966975 |
| 23 | osh radyo | rad yoders | 6 | 0.966907 |
| 24 | benherzaman | sevdimcunku | 1 | 0.966404 |
| 25 | quyinci qiletumfandomlarunfollowsuztakiplesiyor | reyelquyinvenisaisilesecimgunufandomlartakiplesiyor | 7 | 0.965501 |
| 26 | a9tv | sevgi | 4 | 0.963256 |
| 27 | besiktas | eurocup | 9 | 0.962717 |
| 28 | larryveziamskipperlartakiplesiyor | pofudukornitorenkilefandomlartakiplesiy | 6 | 0.962627 |
| 29 | 1kasimda | formasyonahayir | 22 | 0.962346 |
| 30 | hadisi | hadisiserif | 45 | 0.960070 |
| 31 | aleyilekardesfandomlartakiplesiyor | denizverosalindailefandomlartakiplesiyor | 16 | 0.959462 |
| 32 | hadisiser | hadisiserif | 18 | 0.959214 |
| 33 | besiktas | besiktaslilarlayeniyltakiplesmesi | 1 | 0.959164 |
| 34 | hadis | ufkayolculuk4 | 18 | 0.958623 |
| 35 | hadisiserif | kuran | 2 | 0.958311 |
| 36 | 3aralik | 3aralikdunyaengellilergunu | 3 | 0.957922 |
| 37 | 15binikabuletmiyoruzok | ekformasyonkontenjani | 5 | 0.957652 |
| 38 | 50binformasyonistiyoruz | yokformasyonsesimiziduy | 1 | 0.956995 |
| 39 | art | iran | 4 | 0.956701 |
| 40 | formasyonbekliyoruzok | formasyonyoksaisteyok | 1 | 0.955495 |
| 41 | gorevdehaksizyükselme | gorevdeyükselmehaksizligi | 4 | 0.953706 |
| 42 | pazargunutumkekofandomlartakiplesiyor | quyinci qilefandomlarseritakiplesiyor | 19 | 0.951722 |
| 43 | fenerinmacivar | rakibimizantalya | 1 | 0.951400 |
| 44 | besiktas | hedef1903cocuk | 34 | 0.950940 |
| 45 | ekformasyonkontenjani | formasyonasinirkoyulamaz | 2 | 0.950776 |
| 46 | haftasonununmutlulugunuyasayanfandomlarta kiplesiyor | pofudukornitorenkilefandomlartakiplesiy or | 3 | 0.950090 |
| 47 | hadisiserif | mucizebu | 2 | 0.949451 |
| 48 | enaz50binformasyon | yokformasyonsesimiziduy | 1 | 0.948762 |
| 49 | dursunozbekistifa | yonetimistifa | 24 | 0.948639 |
| 50 | tatil | vasipsahin | 5 | 0.947370 |
| 51 | fenerinminmacivar | fenerinmacivar | 1 | 0.946463 |
| 52 | euroleague | fenerbahce | 47 | 0.946113 |
| 53 | egitim | yds | 2 | 0.945486 |
| 54 | ifsahane | turk | 46 | 0.945426 |
| 55 | namaz | namazuykudanhayiridir | 2 | 0.945209 |
| 56 | istifa | yonetimistifa | 1 | 0.944568 |
| 57 | istenirseonlenir | kadincinayetlerionlenebilir | 9 | 0.943837 |
| 58 | besiktas | besiktaslilarbutagda | 5 | 0.943008 |
| 59 | acilyuva | yuva | 1 | 0.942771 |
| 60 | mutluluk | mutlulukheryerde | 2 | 0.942245 |
| 61 | 4buyuklersalonturnuvasi | acunturnuvasi | 1 | 0.940007 |
| 62 | aleyvecatlakfandomlartakiplesmesi | bellatrixilemukofandomlartakiplesiyor | 2 | 0.939931 |

Çizelge 4.8. (devam ediyor).

| | | | | |
|-----|--|--|-----|----------|
| 63 | dursunozbek | yonetimistifa | 1 | 0.939776 |
| 64 | gorevdehaksizyükselme | torpilimyok | 4 | 0.938166 |
| 65 | fenerinmacivar | saldirfener | 29 | 0.937959 |
| 66 | hadisiserif | yusuflarozgurolsun | 4 | 0.937840 |
| 67 | balerinheriivepofuduqkatyiletumfandomlartak iplesiyor | pandaliceileamacsizpazartakiplesmesi | 13 | 0.931991 |
| 68 | fenerbahce | fenerinmacivar | 657 | 0.931829 |
| 69 | katliamakarsidirenis | simditekistedigim | 19 | 0.931478 |
| 70 | a9tv | sevgisevgisevgi | 2 | 0.931135 |
| 71 | iran | sunnikatiliiranvesuud | 3 | 0.930816 |
| 72 | yeniyilagenbakis | yilinsongunundenben | 1 | 0.929957 |
| 73 | quyinciqiletumfandomlarunfollowsuztakiplesi | secimdensikilanfandomlartakiplesiyor | 1 | 0.928824 |
| 74 | formasyonasinirkoyulamaz | hakkimiziveryok | 1 | 0.928798 |
| 75 | isci | taseronakadro | 2 | 0.928716 |
| 76 | nehaltiyedigibilmeyenfandomlartakiplesiyor | nesquikqueeniletakiples | 2 | 0.928715 |
| 77 | furkanilepazargunusikilanlartakiplesiyor | pazargunutumkekofandomlartakiplesiyor | 15 | 0.928137 |
| 78 | fenerinmacivar | saldirkanarya | 12 | 0.926945 |
| 79 | furkanilepazargunusikilanlartakiplesiyor | ketiketesrailetumtirrefandomlartakiplesi | 3 | 0.926079 |
| 80 | ozgecanaslan | ozgecanlaricinadalet | 23 | 0.925931 |
| 81 | gorevdeyükselmehaksizligi | torpilimyok | 168 | 0.924637 |
| 82 | gthb20bin | ziraatmuhendisleri | 72 | 0.924464 |
| 83 | nesquikqueeniletakiples | rosalindavedenizilefandomlartakiplesiyor | 1 | 0.924372 |
| 84 | dirensrdf | espyalnizdegildir | 2 | 0.921707 |
| 85 | euroleague | fenerinmacivar | 42 | 0.921508 |
| 86 | cop21 | paris | 10 | 0.921311 |
| 87 | hadisi | yusuflarozgurolsun | 2 | 0.921235 |
| 88 | formasyonasinirkoyulamaz | formasyonverdirmeartikmeb | 1 | 0.920964 |
| 89 | aycabegen | ayem | 6 | 0.920405 |
| 90 | isci | iscinayeti | 2 | 0.918707 |
| 91 | 50binformasyonunutma | yokformasyonsesimiziduy | 1 | 0.918458 |
| 92 | formasyonaekkontenjanistiyoruz | formasyonasinirkoyulamaz | 5 | 0.917971 |
| 93 | hadisi | mucizebu | 1 | 0.917956 |
| 94 | akit | rtuk | 1 | 0.917288 |
| 95 | 1kasimdaoyumhdpye | unutmayanlarhdpdiyor | 5 | 0.917174 |
| 96 | stopgenocideofturkmen | turkmendagindakatliamvar | 91 | 0.916619 |
| 97 | 3aralikdunyaengellilergunu | engelolmadestekol | 1 | 0.916425 |
| 98 | 4kartalyuvasitarftarorganizasyonu | besiktas | 2 | 0.916388 |
| 99 | fenerbahce | sarimelekler | 10 | 0.916349 |
| 100 | ozgecan | ozgecandavasi | 1 | 0.915676 |

Çizelge 4.9'daise weight değeri en yüksek 100 kayıt listelenmektedir.

Çizelge 4.9. Ortak kümenin weight değeri en yüksek 100 hashtag birlikteliği.

| Sıra | source_node | target_node | weight | cosine |
|------|----------------------------------|-------------------|--------|----------|
| 1 | kan | trombosit | 1692 | 0.534923 |
| 2 | felsefegrubuderslerizorunluolsun | felsefeye200atama | 978 | 0.699474 |
| 3 | bulunamadi | kan | 867 | 0.934924 |
| 4 | istanbul | kayipkopek | 778 | 0.581675 |
| 5 | enguzelveda | kirazmevsimi | 746 | 0.688602 |
| 6 | fenerbahce | fenerinmacivar | 657 | 0.931829 |
| 7 | haberler | sondakika | 565 | 0.567821 |
| 8 | ayem | kismetseolur | 529 | 0.542444 |
| 9 | ios | sondakika | 509 | 0.625632 |
| 10 | katildevlet | tahirelci | 431 | 0.799151 |

Çizelge 4.9. (devam ediyor).

| | | | | |
|----|----------------------------|--------------------------|-----|----------|
| 11 | oykullan | sandikbekcisiyiz | 419 | 0.741754 |
| 12 | haber | teknoloji | 417 | 0.512513 |
| 13 | besiktas | liderbesiktas | 416 | 0.805977 |
| 14 | halep | suriye | 398 | 0.672756 |
| 15 | ayseogretmenyalnizdegildir | beyazshow | 397 | 0.784369 |
| 16 | gunesinkizlari | kirgincicekler | 368 | 0.733586 |
| 17 | kirazmevsimi | onsuzhayat | 351 | 0.804346 |
| 18 | 1kasimdaoyumhdpye | inadinasandiga | 344 | 0.790663 |
| 19 | dikkathirsizvar | oykullan | 342 | 0.655703 |
| 20 | haydioylarhdpye | oykullan | 315 | 0.693133 |
| 21 | silvan | silvanasesver | 311 | 0.854268 |
| 22 | fakirelink | fenerinmacivar | 308 | 0.636742 |
| 23 | kirazmevsimi | olurmoulur | 301 | 0.822944 |
| 24 | silvan | silvandakatliamvar | 294 | 0.751603 |
| 25 | izmir | izmirescort | 285 | 0.969175 |
| 26 | oykullan | secim2015 | 276 | 0.661902 |
| 27 | istanbul | yuva | 272 | 0.845655 |
| 28 | gunesinkizlari | hoscakal | 264 | 0.523281 |
| 29 | fenerbahce | uel | 263 | 0.668108 |
| 30 | gundem | sondakika | 256 | 0.621562 |
| 31 | polislikpuani60olsun | pomem60puanolsun | 250 | 0.641501 |
| 32 | silvanasesver | silvandakatliamvar | 246 | 0.751010 |
| 33 | gunesinkizlari | yenidenbaslayalim | 244 | 0.637711 |
| 34 | emeklilikteyasatakilanlar | eyt | 232 | 0.545640 |
| 35 | akpartigeliyor | istikraricinoylarakolsun | 229 | 0.676332 |
| 36 | oykullan | oyver | 221 | 0.556416 |
| 37 | ikiannearasinda | kirazmevsimi | 220 | 0.834359 |
| 38 | bendensakladigin | kirazmevsimi | 219 | 0.841205 |
| 39 | dikkathirsizvar | sandikbekcisiyiz | 214 | 0.554955 |
| 40 | gunesinkizlari | hicvazgecmedim | 210 | 0.630215 |
| 41 | destek | kan | 205 | 0.542339 |
| 42 | mtvstars | videomtv2015 | 203 | 0.542631 |
| 43 | benceask | gunesinkizlari | 200 | 0.563247 |
| 44 | besiktas | besiktasinmacivar | 198 | 0.891653 |
| 45 | 10kasim | aturkdemek | 192 | 0.624625 |
| 46 | tahirelci | tahirelciolumsuzdur | 192 | 0.584371 |
| 47 | haydibismillah | secim2015 | 191 | 0.652151 |
| 48 | konyapostasi | sondakika | 190 | 0.668196 |
| 49 | gunesinkizlari | kaybedinceanladim | 183 | 0.580042 |
| 50 | kirgincicekler | seninicin | 179 | 0.596105 |
| 51 | oyveotesi | oyver | 174 | 0.728052 |
| 52 | halep | rus | 172 | 0.670021 |
| 53 | 10kasim | benimicinaturk | 172 | 0.521288 |
| 54 | haberk | sondakika | 171 | 0.620506 |
| 55 | destek | trombosit | 171 | 0.564847 |
| 56 | gorevdeyükselmehaksizligi | torpilimyok | 168 | 0.924637 |
| 57 | delibal | delibalvizyonda | 166 | 0.879865 |
| 58 | gsbgm | yenibirnefes | 165 | 0.866928 |
| 59 | katilpkk | tahirelci | 165 | 0.632318 |
| 60 | bayirbucakturkmenleri | turkmen | 160 | 0.891473 |
| 61 | hadis | ufkayolculuk | 160 | 0.671894 |
| 62 | kiralikask | kirgincicekler | 152 | 0.567650 |
| 63 | bulunamadi | destek | 144 | 0.507550 |
| 64 | madaya | suriye | 142 | 0.537630 |
| 65 | 10kasim | aniyoruzariyoruz | 141 | 0.748568 |
| 66 | nusaybin | nusaybinesesver | 137 | 0.785667 |
| 67 | alanlaradirenise | cizreyesilopiyesesver | 136 | 0.528207 |

Çizelge 4.9. (devam ediyor).

| | | | | |
|-----|----------------------------|--------------------------|-----|----------|
| 68 | fakirelink | fenerbahce | 135 | 0.581370 |
| 69 | acil | dualarimizoykuyle | 133 | 0.642581 |
| 70 | 10kasim | izinsilinmez | 132 | 0.763063 |
| 71 | kararsenin | sandigason2 | 129 | 0.673029 |
| 72 | akpyeoyvermemcunku | oyveroyunlariniboz | 129 | 0.538051 |
| 73 | kirgincicekler | paramparca | 127 | 0.911588 |
| 74 | hilafethaberleri | islamdevleti | 127 | 0.741461 |
| 75 | mahkumakapsamliaf | mahkumunumuduenveryilmaz | 127 | 0.721910 |
| 76 | 10kasim | unutulmaz | 125 | 0.505166 |
| 77 | ask | love | 124 | 0.631946 |
| 78 | kurdistanelcisiniugurluyor | tahirelci | 122 | 0.693477 |
| 79 | besiktas | gururlan | 115 | 0.845125 |
| 80 | gunesinkizlari | paramparca | 114 | 0.678513 |
| 81 | bigbrothertr | bigbrotherturkiye | 113 | 0.596171 |
| 82 | haydibismillahturkiye | tekbasinaisbasina | 113 | 0.593766 |
| 83 | berkoktay | iliskidurumukarisik | 112 | 0.831906 |
| 84 | oyver | secim2015 | 110 | 0.527799 |
| 85 | 10kasim | atamizindeyiz | 109 | 0.813521 |
| 86 | idilekalpolalim | kalbimizseninle | 109 | 0.633599 |
| 87 | izmir | karsiyaka | 108 | 0.784168 |
| 88 | acil | samsun | 107 | 0.755687 |
| 89 | filistin | israil | 105 | 0.837944 |
| 90 | delibal | sinema | 103 | 0.639291 |
| 91 | bayirbucak | turkmendagi | 99 | 0.673277 |
| 92 | iliskidurumukarisik | kertenkele | 97 | 0.506515 |
| 93 | silvanasesver | silvankobanedir | 97 | 0.502790 |
| 94 | askyeniden | isteaskyeniden | 96 | 0.614009 |
| 95 | haydibismillah | tekbasinaisbasina | 95 | 0.645107 |
| 96 | herseyimoldun | inadinaask | 94 | 0.563012 |
| 97 | dikkathirsizvar | hirsizakpoycaliyor | 93 | 0.686132 |
| 98 | stopgenocideofurkmen | turkmendagindakatliamvar | 91 | 0.916619 |
| 99 | sur | surasesver | 91 | 0.901303 |
| 100 | galatasaray | sarayinsultanlari | 91 | 0.674890 |

Çizelge 4.8 ve 4.9’da, hashtag’lerin birlikte kullanımları ile semantik yakınlıklarının belli oranda ilişkili oldukları görülmekle birlikte bu ilişkinin nasıl bir korelasyona sahip olduğu 5. Bölüm’de detaylı olarak incelenecektir.

Çalışmamızın bu bölümünde hedeflenen hashtag birliktelik ve semantik yakınlık ağları oluşturulmuş, aynı zamanda bu iki ağın ortak düğüm (hashtag) ve linklerinin tanımlandığı bir de kesişim ağı elde edilmiştir. Bu kesişim ağı, hem birliktelik ağından gelen “weight” değerlerini, hem de semantik ağdan gelen “cosine” değerlerini içerdiğinden hashtag’lerin birlikte kullanımlarının ne derece doğru ve anlamlı olduğu ile ilgili de önemli bilgiler taşımaktadır. Bu 3 farklı ağ ile ilgili analizler bir sonraki bölümde detaylandırılacaktır.

BÖLÜM 5

AĞ ANALİZİ SONUÇLARI

5.1. GEPHI İLE AĞ ANALİZİ

5.1.1. Gephi Yazılımı

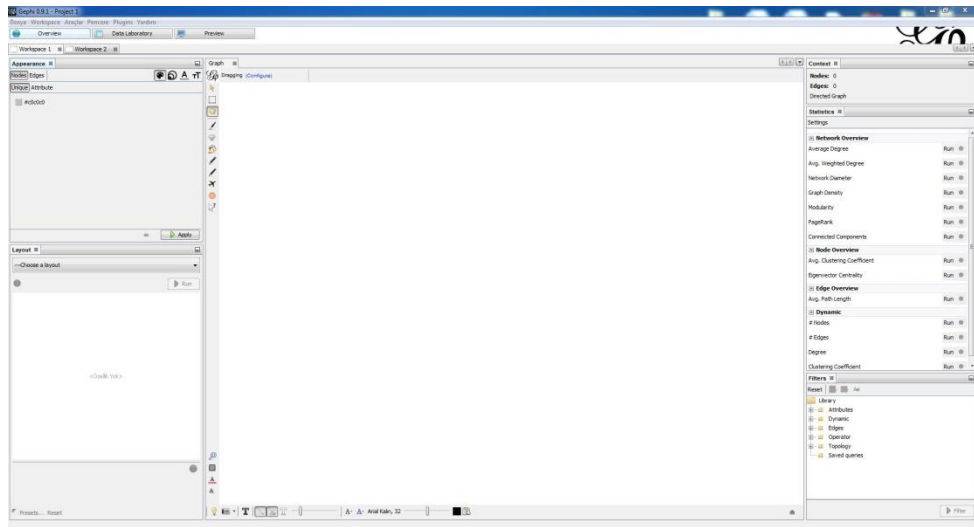
Gerçek kompleks ağlar genellikle çok sayıda düğüm ve linkten oluşurlar. Bu ağların düğüm ve link yapılarının analizi bazı karmaşık işlemler bütünü ile mümkündür. Bu işlemlerin başında düğümler arasında düğümler arasında en kısa yolların bulunması, kümelenme katsayısının tespiti, derece dağılımlarının hesaplanıp grafiğe dökülmesi gibi tekdüze olmayan fonksiyonlar gelmektedir. Aynı zamanda düğüm ve link sayısı oldukça fazla olan bu ağların görselleştirilmesi de algoritmik ve grafiksel açıdan farklı yaklaşımları gerektirir.

Çizelge 5.1. İnteraktif görselleştirme ve veri analiz programı örnekleri [9].

| Yazılım | Platform | Lisans | Görselleştirme | Analiz | Açıklama |
|-----------|--------------------------|----------------------------|----------------|--------|---|
| Gephi | Windows, Linux, Mac OS X | Açık Kaynak Kodlu Ücretsiz | VAR | VAR | Her türlü ağı görselleştirebilir. Yönlü ve yönsüz linkleri işleyebilir. Metrik analizler yapılabilir. 1 milyon adet düğüm ve linki görselleştirebilir. ForceAtlas ve GeoLayout gibi yerleştirme algoritmalarını kullanılır. 2 ve 3 boyutlu görselleştirme yapabilir. Java tabanlı olarak çalışmaktadır. |
| Pajek | Windows | Ücretsiz | VAR | VAR | Büyük boyutlu ağlar analiz edebilir ve görselleştirebilir. Bilimsel işbirliği ağları, internet ağı ve atf ağları gibi ağların işlenmesi ve görselleştirilmesi için özel araçları vardır. |
| Graphviz | Windows, Linux, Mac OS | Ücretsiz | VAR | YOK | Açık kaynak kodlu Graf ve Network görselleştirme yazılımıdır. Görselleştirme işlemi için farklı yerleştirme algoritmaları kullanılır |
| R | Windows, Linux, Mac OS | Ücretsiz | VAR | VAR | Açık kaynak kodlu bir programlama dili ve görselleştirme yazılımıdır. Ara yüz yerine kullanıcı tarafından verilen kodlar ile çalışır. |
| CytoScape | Windows, Linux, Mac OS | Ücretsiz | VAR | VAR | Moleküler etkileşim ve biyolojik ağları görselleştirmek için geliştirilen ancak her türlü kompleks ağları görselleştirebilen açık kaynak kodlu, Java tabanlı bir görselleştirme ve analiz yazılımıdır. |

Gelişen ağ bilimi, beraberinde bir takım ağ analiz yazılımlarının da ortaya çıkmasını sağlamıştır. En çok bilinen yazılımlara örnek olarak Gephi, CytoScape, Pajek, Graphviz, Rve CytoScape programları gösterilebilir [9]. Çizelge 5.1’de bu progamlara ait temel özellikler gösterilmiştir.

Bu yazılımlar arasından, ücretsiz olarak sunulması, görselleştirme seçeneklerinin daha gelişmiş olması ve akademik çalışmalarda daha çok tercih edilmesi nedeniyle, çalışmamızda Gephi yazılımı kullanılmıştır.



Şekil 5.1. Gephi yazılımı ara yüzü.

Şekil 5.1’de Gephi yazılımının ara yüzü görülmektedir. Gephi yazılımında işlemler temel üç pencere altında gerçekleştirilmektedir. Bunlar overview penceresi, data laboratory penceresi ve preview penceresidir.

Overview Penceresi: Bu bölümde kompleks ağa ilişkin genel bilgiler yer alır. Temel görselleştirme ayarları bu bölümde yapılır. Yine bu bölümde yerleştirme algoritmasının seçimi işlevi ve metrik analizler yer alır. Partition, Ranking, Layouts, Context, Statistics ve Filters bölümleri bulunur.

Data Laboratory Penceresi: Bu bölümde düğümlere ve linklere ait veritabanı bulunur. Ağa yeni düğümlerin ve linklerin eklenmesi bu bölümde gerçekleşir. Yine var olan düğümlerde yapılmak istenilen değişiklikler bu bölümde yapılır. Veriyi

dışarıdan csv, gexf gibi çeşitli formatlarda almak, ya da csv formatında dışa aktarmak mümkündür.

*Preview Penceresi:*Bu bölümde detaylı görselleştirme ayarları ve ağ görseli mevcuttur. Kurulmuş olan ağa ait görüntü üzerinde değişiklikler ve ince ayarlar (büyüklük, renk, şekil vb.) bu bölümde yapılır.

5.1.2. Gephi Temel Analiz Sonuçları

Çalışmamızda birliktelik ve semantik olmak üzere iki yaklaşım sergilenmiş ve iki farklı ağ kurulmuştur. Bu iki ağa ait veriler analiz edildikten sonra, iki ağın ortak kümesine ait veriler de üçüncü bir alt ağ olarak analiz edilmiştir. Çizelge 5.2’de iki ağa ve ortak kümelerine ait düğüm ve link sayıları verilmiştir. Bu düğüm ve linkler kullanılarak yapılan Gephi analiz sonuçları Çizelge 5.3’de görülmektedir.

Çizelge 5.2. Ağlara ait düğüm ve link sayıları.

| | Düğüm sayısı | Link sayısı |
|-----------------|--------------|-------------|
| Birliktelik ağı | 34 111 | 197 675 |
| Semantik ağ | 10 169 | 67 111 |
| Ortak küme | 7 647 | 2 247 |

Çizelge 5.3. Gephi yazılımı analiz sonuçları.

| | Birliktelik ağı | Semantik ağ | Ortak küme |
|--------------------------------|-----------------|-------------|------------|
| Average degree | 11,59 | 17,493 | 2,613 |
| Average weighted degree | 52,597 | 17,493 | 53,059 |
| Network diameter | 12 | 14 | 15 |
| Modularity | 0,728 | 0,872 | 0,948 |
| Average clustering coefficient | 0,559 | 0,626 | 0,479 |
| Eigenvector centrality | 0,215 | 0,459 | 0,092 |
| Average Path length | 3,821 | 5,818 | 4,23 |

Temel analiz sonuçlarında öncelikle semantik ağın ortalama derece (average degree) değerinin birliktelik ağına göre yüksek olduğu görülmektedir. Bu durum, semantik olarak ilişkili olan bazı tag'ların Tweet'lerde birlikte kullanılmamış olduğunun bir göstergesidir. Bunun temel sebeplerinden bir tanesi bu tag'ların aynı zaman penceresine ait olmamaları olabilir. 4 aya yakın süren veri toplama sürecinin başında ve sonunda kullanılmış iki farklı tag semantik olarak bağlantılı olsa da birlikte kullanılması mümkün olmamaktadır. Dolayısıyla semantik ağ zaman kavramından bağımsız, birlikte kullanım ağı ise zamana bağımlı bir yapıda olduğundan ortalama derecelere yansıyan bu sonuç beklenen bir durumdur. Bununla birlikte semantik ağda yapısı gereği link ağırlıkları olmayıp birlikte kullanım ağı için ayrıca "weighted degree" değerleri söz konusudur.

Ağın çapı, en kısa mesafelerin en uzunudur. Bu değer düşük olması, ağın iç bağlılığının daha yüksek olduğunun bir göstergesidir. Diameter değerleri incelendiğinde semantik ağın yüksek ortalama dereceye rağmen daha büyük çapa sahip olduğu görülmektedir. Birliktelik ağında düşük ortalama dereceye rağmen çapın daha düşük olması, anlamsal ilişki olmamasına rağmen birlikte kullanılmış tag'ların bir sonucudur. Böylece anlamsal olarak ilgisiz tag'lar arasında da bağlar kurulmuş, bu da ağın çapını düşürücü etki yapmıştır. Bu durum Çizelge 5.3'ün sonunda yer alan average path degree (ortalama yol uzunluğu) değerlerinde de etkisini göstermiş, en düşük ortalama yol uzunluğu değeri birliktelik ağında elde edilmiştir.

Ortalama yol uzunluğu değerinin tüm ağlarda düşük olduğu görülmektedir. Gerçek ağların önemli özelliklerinden olan "small-world" özelliği, hashtag ağlarında da dikkat çekici biçimde küçük düğümler arası uzunluklar ile kendini göstermektedir.

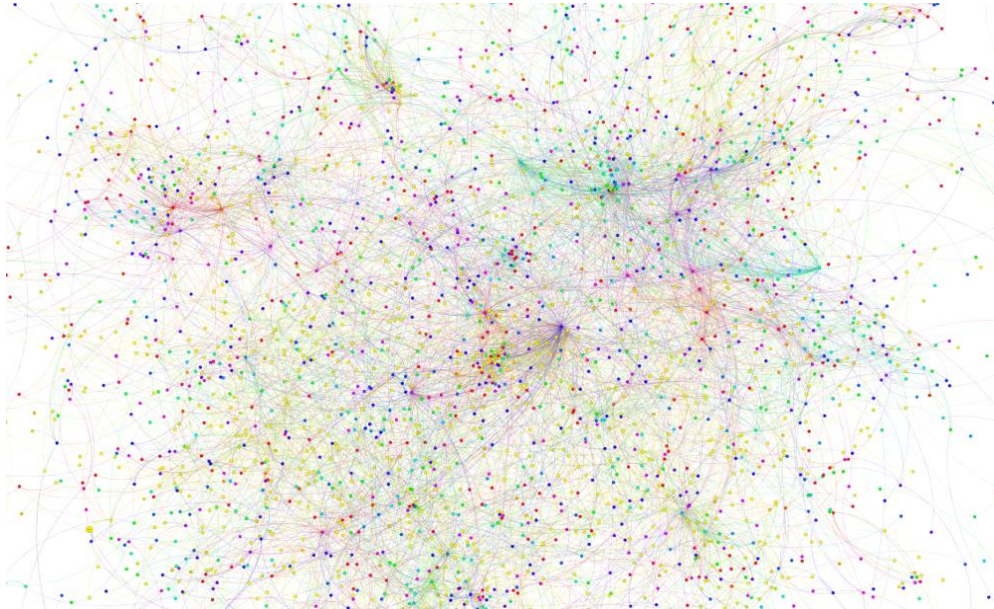
Yine gerçek ağların önemli özelliklerinden ikisi yüksek kümelenme ve modülerliktir. Bu kavramlar 1. Bölüm'de açıklanmıştır. Semantik ağın birliktelik ağına göre daha yüksek kümelenme ve modülerlik sergilediği görülmektedir. Bu durum üzerinde de semantik ağın zamandan bağımsız oluşu etkili olmuştur. Zaman bağımlılığına rağmen birliktelik ağının da semantik ağa yakın kümelenme ve modülerlik değerleri sergilemesi ilgi çekicidir. Literatürdeki gerçek ağ çalışmaları ile uyumlu olan bu

özellikler, hashtag ağlarının da iyi bir “self-organization” sistem prototipi olduğunu göstermektedir.

5.1.3. Ağ Görselleri

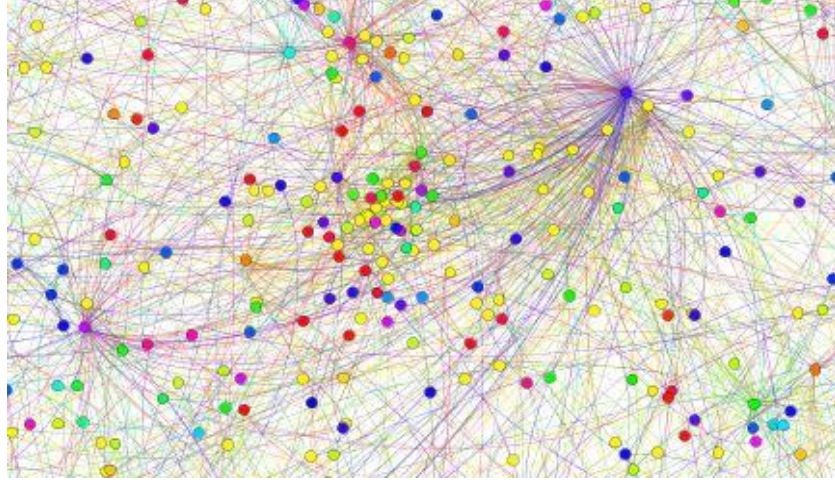
5.1.3.1. Birliktelik Ağı Görseli

Hashtag'lerin birlikte kullanımlarına göre oluşturulan ağda 34 111 adet düğüm ve 197 675 adet link bulunmaktadır. Düğüm ve link sayısının çok fazla olması sebebiyle Gephi'de oluşturulan ağ görsel olarak çok karışık olduğundan dolayı düğüm ve linklere Gephi yazılımında filtre uygulanmıştır. Ağırlık (weight) değeri 10'dan küçük linkler devre dışı bırakılarak link sayısı 12 149 adete indirilmiştir. Bunun ardından düğüm derece değeri filtresi uygulanmıştır. Bu filtre uygulanırken düğüm derece değeri 10'dan küçük olan düğümler ve bu düğümlerin linkleri devre dışı bırakılmıştır. Bu işlemin ardından düğümlerin %25,2 adeti görselleştirilerek 8 595 adet düğüm, linklerin %4,53 adeti görselleştirilerek 8949 adet link ile Şekil 5.2'de verilen ağ görseli oluşturulmuştur.



Şekil 5.2. Birliktelik ağı görseli.

Şekil 5.2’de verilen ağ görselindeki düğüm ve linklere ait bir bölümden yakın görünüm Şekil 5.3’te verilmiştir.

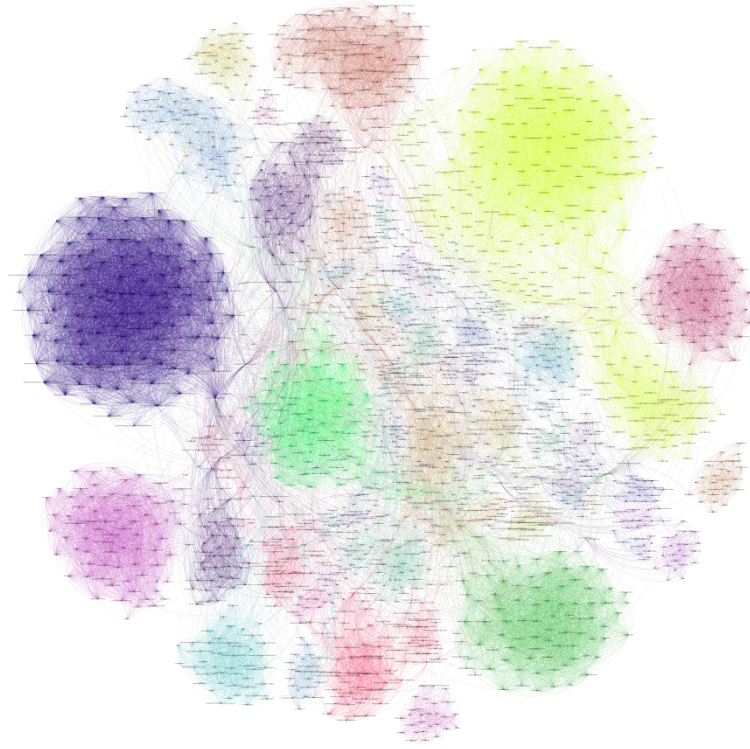


Şekil 5.3. Birliktelik ağı görseli yakın görünümü.

5.1.3.2. Semantik Ağı Görseli

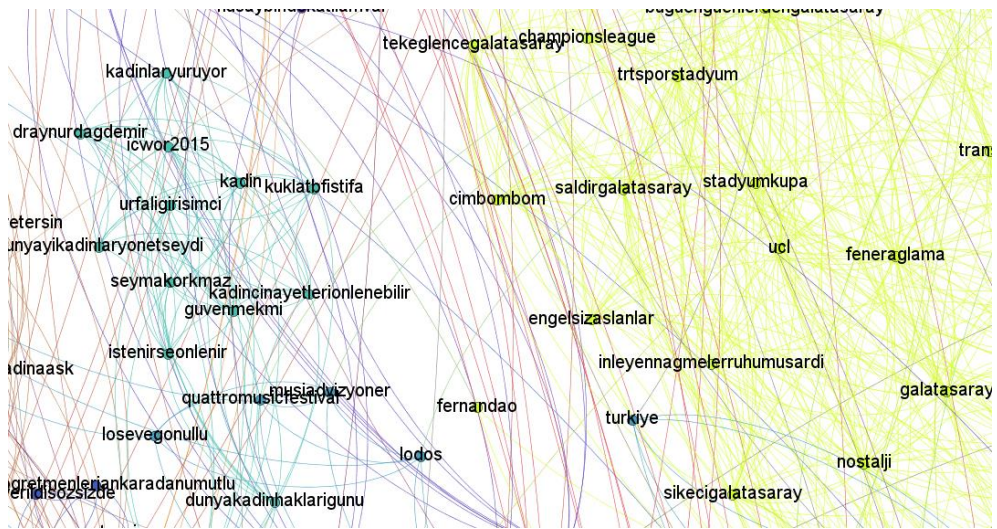
Hashtag’lerin semantik yakınlığına göre oluşturulan ağda 10 169 adet düğüm ve 67 111 adet link bulunmaktadır. Görselleştirme yapılırken düğüm derece filtresi 20’den küçük olanların devre dışı bırakılması olarak uygulanmıştır. Filtreleme işlemi ardından 2105 düğüm (%20,7) ve 36311 link (%54,11) ile 5.4’de verilen ağ görseli oluşturulmuştur.

Semantik yakınlığa göre analiz edilerek oluşturulan bu ağa ait görselde görüldüğü gibi kümelenmeler belirgin olarak oluşmuştur. Kümelenmenin bu ağda çok yüksek olmasının en temel sebebi semantik yakınlığın zamandan bağımsız bir yapıda olması olarak gösterilebilir.



Şekil 5.4. Semantik ağı görseli.

Bu ağıın yakın görünümünün gösterildiği Şekil 5.5’de görüleceği gibi, sarı bölüm “saldırgalatasaray”, ”stadyumkupa”, “sikecigalatasaray”, “cimbombom” gibi futbol ile alakalı ve semantik yakınlığı bulunan kelimelerden oluşmaktadır. Mavi bölüm ise “kadin”, “kadincinayetlerionlenebilir”, “istenirseonlenir”, “dunyakadinhaklarigunu” gibi yine semantik olarak yakınlığı bulunan kelimelerin kümelendiği bölümdür.

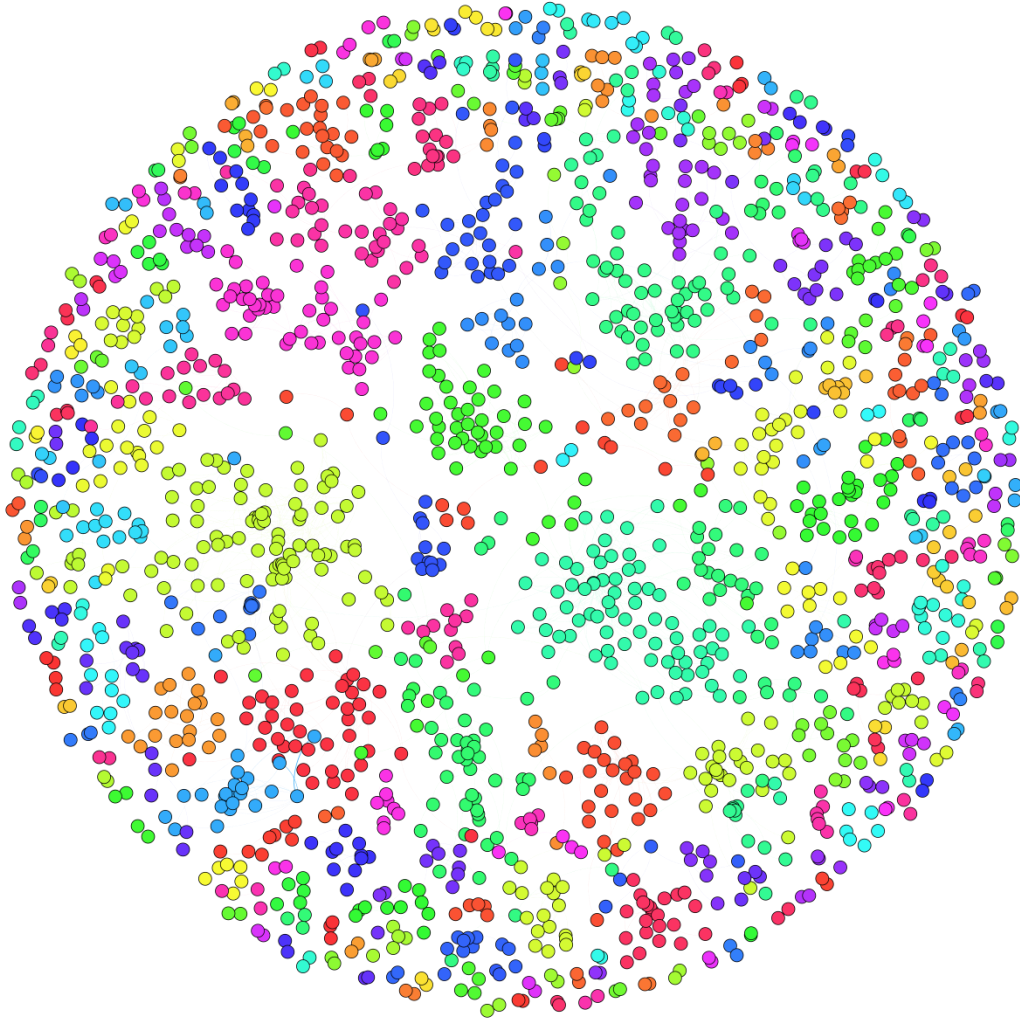


Şekil 5.5. Semantik ağı görseli yakın görünümü.

5.1.3.3. Ortak Küme Ağı Görseli

Hashtag'lerin birlikte kullanımları ve semantik yakınlığına göre oluşturulan ağların ortak kümesinde 7 647 adet düğüm ve 2 247 adet link bulunmaktadır. Ağın görselleştirilmesi yapılmadan önce düğüm derecesi 0 olan düğümler devre dışı bırakılarak 1 785 düğüm ve 2 247 adet link ile Şekil 5.6'da verilen ağ görseli oluşturulmuştur.

Düğüm derece değerlerine göre renklendirme uygulanmıştır. Düğüm derece değerleri, birbirine yakın olan düğümlerin kümelenmeleri Şekil 5.6'da gözlemlenmektedir.



Şekil 5.6. Birliktelik ağı ile semantik ağı ortak kümesi görseli.

5.2. DETAYLI DÜĞÜM VE LİNK ANALİZİ

Gephi’de yapılan analizlerin sonucunda Çizelge 5.3’teki çıktılar ile birlikte derece dağılım grafiği elde edilebilmektedir. İleri istatistiksel analizler ve derece dağılımının logaritmik eksenlerde çizimi, eğri uydurma gibi çalışmalar için MATLAB gibi bir analiz ortamına ihtiyaç duyulmaktadır. Bu nedenle çalışmamızın Gephi ayağından elde edilen işlenmiş veri (nodes ve edges tabloları) MATLAB’a aktarılarak, ileri hesaplama ve analiz prosedürleri gerçekleştirilmiştir.

5.2.1. Düğüm Analizleri

Bu bölümde düğümlerin tekil ve ağırlık derecelerine ait dağılım grafikleri çizdirilmiştir. Power-law uyumunun anlaşılabilmesi için eğri uydurma çalışmalarına da yer verilmiştir.

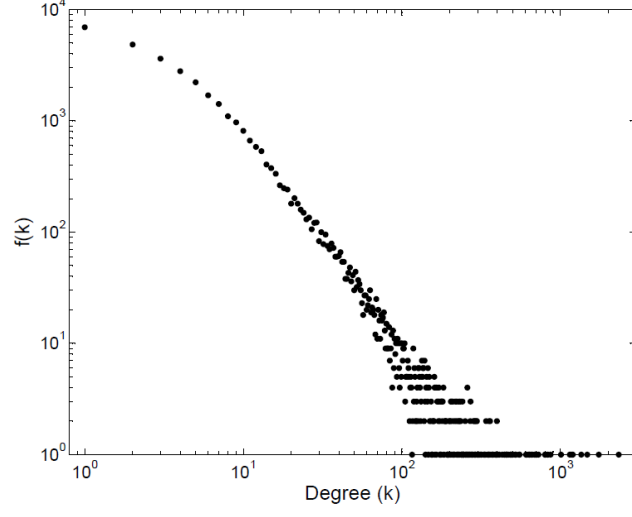
5.2.1.1. Derece Dağılımlarının Saçılmış Grafikleri (Scatter)

Düğüm derecelerinin gerçekleşme frekanslarına ait scatter dağılımlar Şekil 5.7, 5.8 ve 5.9’da görülmektedir. Grafiklerdeki her nokta, ağda ilgili dereceden kaç adet düğüm bulunduğunu ifade etmektedir. Gerçek ağlar genellikle scale-free özellik sergilediğinden bu durumun anlaşılabilmesi için çizimlerde logaritmik (log-log) eksenler kullanılmıştır.

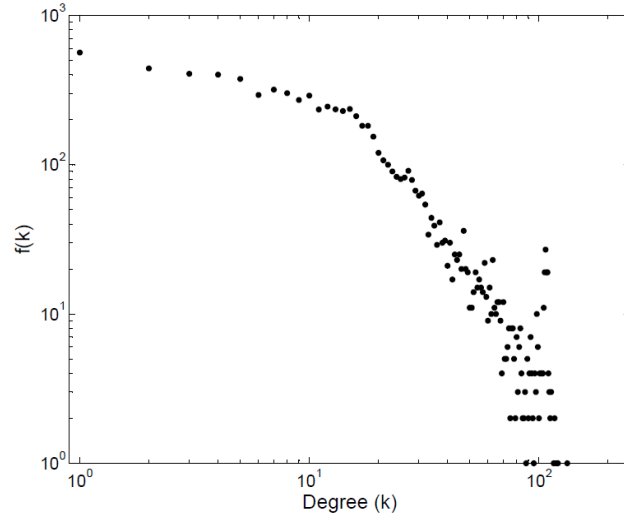
Saçılmış frekans grafikleri, aslında birer olasılık dağılım fonksiyonu değildirler ve sağ tarafa doğru sıklaşan yapıları nedeniyle power-law uyumları çok net biçimde teşhis edilemeyebilir. Ancak yine de log-log ekseninde doğru ile uyumlu düşüş karakteristikleri, power-law uyumu hakkında fikir verebilir.

Şekil 5.7’deki birliktelik ağı scatter dağılımı ve Şekil 5.9’daki ortak ağ scatter dağılımı power-law uyumlu oldukları izlenimi vermektedirler. Şekil 5.8’deki semantik ağa ait grafik ise bir düşük k saturasyon bölgesinin ardından power-law uyumlu kuyruk rejimine sahip izlenimi vermektedir. Bu izlenimlerin doğrulanması

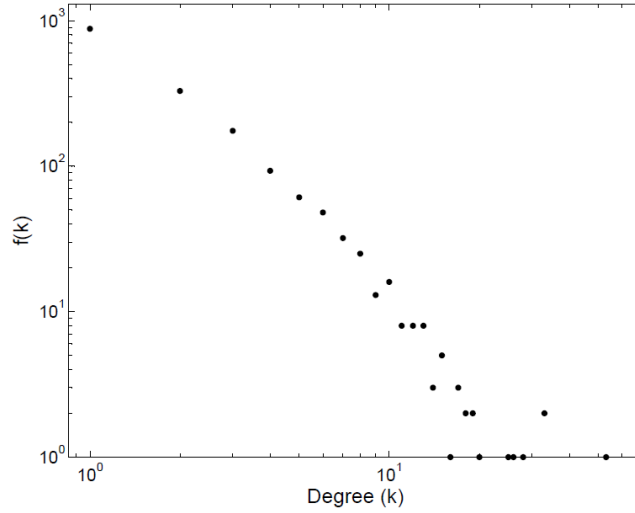
için sonraki bölümde log-binning uygulanmış olasılık yoğunluk fonksiyonlarından faydalanılacaktır.



Şekil 5.7. Birliktelik ağı dağıtık gösterimi.



Şekil 5.8. Semantik ağı dağıtık gösterimi.



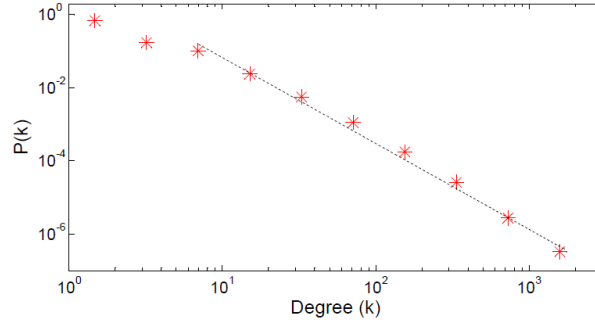
Şekil 5.9. Ortak küme dağıtık gösterimi.

5.2.1.2. Derece Dağılımlarının Toplanmış Grafikleri (Log-binned Degree Distribution)

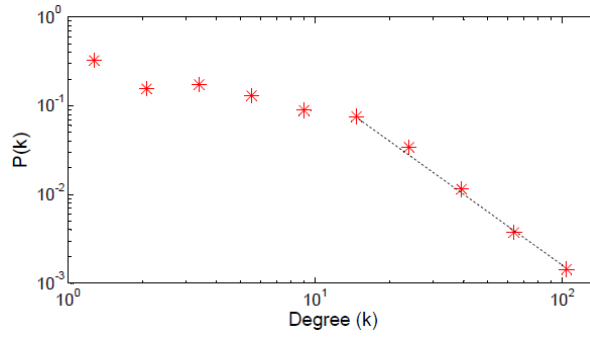
Önceki bölümdeki scatter çizimlerin olasılık yoğunluk fonksiyonuna dönüştürülmesi için logaritmik eksenle eşit aralıklı noktalar tespit edilip, bu noktalar etrafında sık olarak yoğunlaşmış noktaların ortalamasının tek bir nokta ile ifade edilmesi sağlanmıştır. Bu işleme log-binning adı verilmektedir. Log-binning uygulandıktan sonra elde edilen noktalar, toplamları 1 olacak şekilde normalize edildiğinde ise artık grafikler olasılık yoğunluk fonksiyonuna dönüşmüş olur. Şekil 5.10, 5.11 ve 5.12’de üç farklı ağımıza ait log-binning uygulanmış ve normalize edilmiş derece dağılımları gösterilmiştir. Ayrıca bir önceki bölümde varlıklarını sorguladığımız power-law uyumlu bölgelere eğri uydurularak scale-free topolojileri doğrulanmıştır.

Şekil 5.10’deki birliktelik ağı ve Şekil 5.12’deki ortak küme ağının derece dağılımları büyük ölçüde power-law uyumludur. Eğri eğimleri sırasıyla 2,35 ve 2,5’dir.

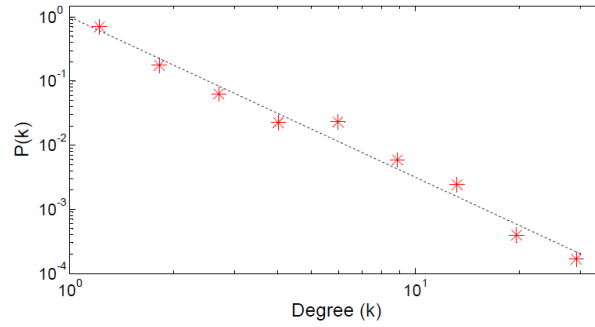
Şekil 5.11’deki semantik ağ ise soldaki düşük k satürasyon bölgesinin devamında 2,0 eğimine sahip bir power-law uyumlu bölgeye sahiptir. Dolayısıyla her 3 ağında scale-free olduğu sonucuna varılmıştır.



Şekil 5.10. Birliktelik ağı toplanmış derece dağılımı.



Şekil 5.11. Semantik ağı toplanmış derece dağılımı.

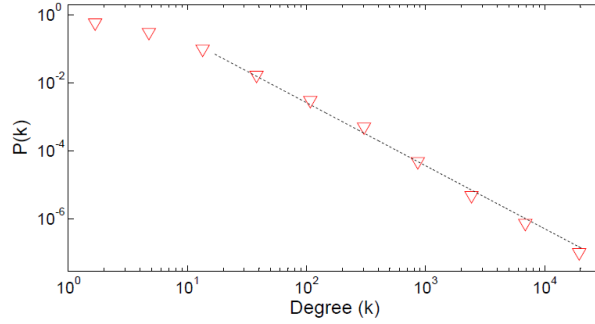


Şekil 5.12. Ortak küme toplanmış derece dağılımı.

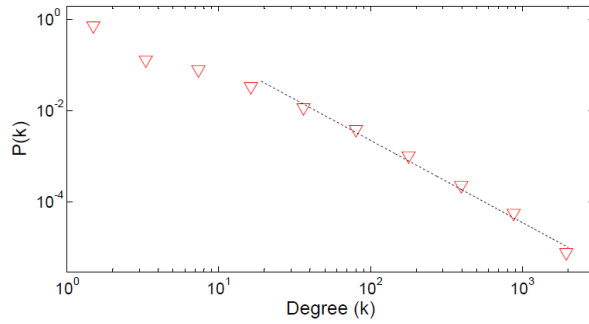
5.2.1.3. Ağırlıklı Derece Dağılımlarının Toplanmış Grafikleri (Log-binned Weighted Degree Distribution)

Tez çalışmamızda yer alan birliktelik ağı ve bu ağın ortak ağıdaki uzantısı bir ağırlıklı ağıdır (weighted network). Ağırlıklı ağlarda aynı düğüme doğru yapılan tekrarlı bağlantılar da dikkate alınarak o düğüme ait “weighted degree” değeri 1 arttırılır. Ağırlıksız ağlarda ise aynı düğüme yapılan tekrarlı bağlantılar göz ardı edilir.

Birliktelik ağı ve ortak ağ üzerinde yaptığımız ağırlıklı derece dağılımı analizlerinin sonuçları Şekil 5.13 ve 5.14'deki grafiklerde sunulmuştur. Bu dağılımlar da sırasıyla 1,85 ve 1,8 eğimlerine sahip olup scale-free uyumludur.



Şekil 5.13. Birliktelik ağı toplanmış ağırlıklı derece dağılımı.



Şekil 5.14. Ortak küme toplanmış ağırlıklı derece dağılımı.

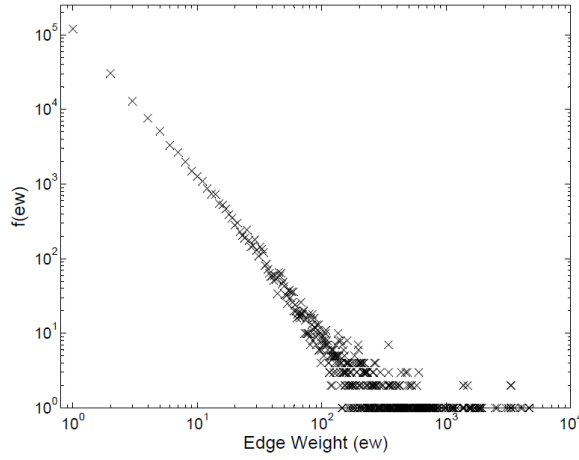
5.2.2. Link Analizleri

Düğümler arası kurulan linklerin özelliklerinin daha anlaşılır kılınması amacıyla mevcut veriler çeşitli analizlerden geçirilerek görsel hale getirilmiştir.

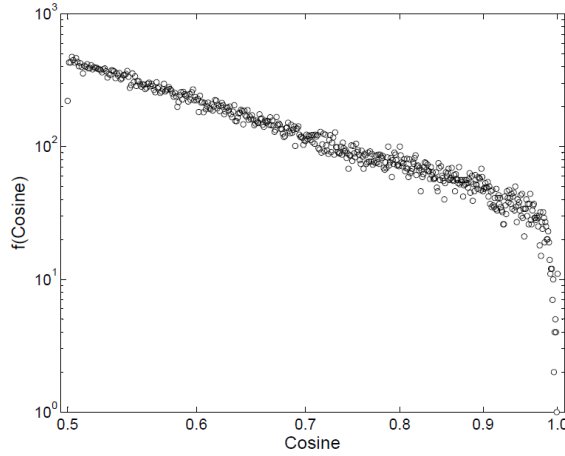
5.2.2.1. Link (Edge) Weight Dağılımlarının Scatter Grafikleri

Kompleks ağlarda tekrarlanan linkler de ağırlıklandırılarak değerlendirilebilir. Tez çalışmamızdaki hashtag birliktelik ağı linkleri ağırlıklandırılarak değerlendirildiği için link ağırlık dağılımı da (scatter) hesaplanmış ve Şekil 5.15'de çizdirilmiştir. Aynı dağılım ortak ağ için de hesaplanarak Şekil 5.17'de çizdirilmiştir.

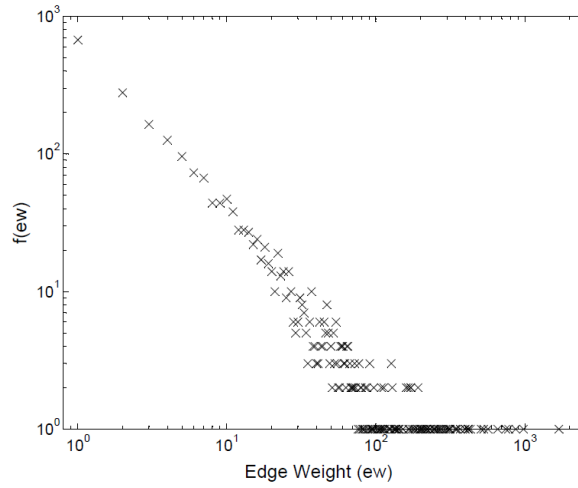
Semantik ağda linklerin tekrarı söz konusu olmadığından ağırlık değerleri yoktur. Ancak her link, 0,5 eşik değerinden büyük bir cosine benzerlik değerine sahiptir. dolayısıyla linklerin cosine değerlerinin de scatter dağılımı çizdirilmiş, Şekil 5.16'da sunulmuştur. Ortak küme düğümleri, hem birlikte kullanım açısından gelen ağırlık değerlerine, hem de semantik ağdan gelen cosine değerlerine sahiptir. Dolayısıyla bu ağ için de cosine dağılımı çizdirilebilir. Söz konusu dağılım scatter olarak Şekil 5.18'de sunulmuştur.



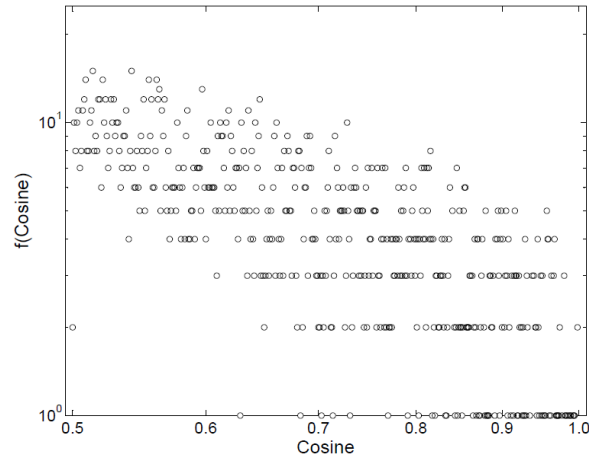
Şekil 5.15. Birliktelik ağırlıklarının dağılımı gösterimi.



Şekil 5.16. Semantik ağırlıklarına ait cosine değerlerinin frekans dağılımı (scatter).



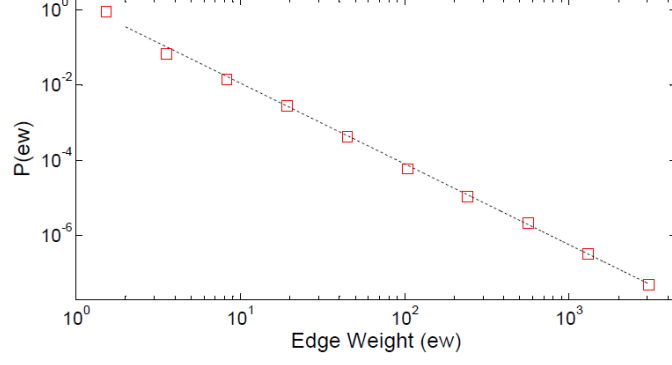
Şekil 5.17. Ortak küme linklerin ağırlık değerlerinin frekans dağılımı (scatter).



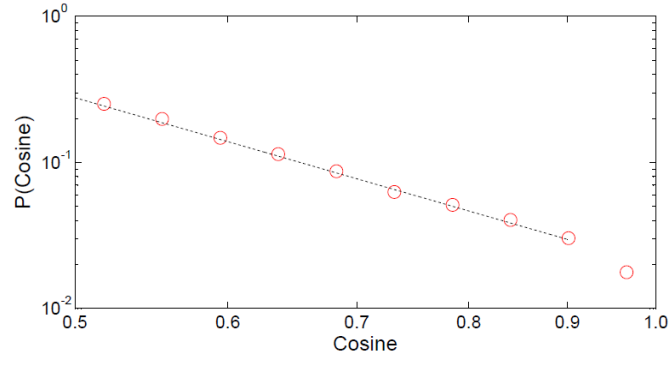
Şekil 5.18. Ortak küme linklerine ait cosine değerlerinin frekans dağılımı (scatter).

Edge weight ve cosine değerlerine ait scatter grafikleri incelendiğinde yine power-law uyumu dikkat çekmektedir. Ancak bu uyumun tam olarak ortaya koyulması için saçılmış çizimlerin olasılık yoğunluk fonksiyonuna dönüştürülmesi gerekmektedir. Bu nedenle bu grafikteki verilere log-binning ve normalizasyon uygulanarak dağılım grafikleri elde edilmiş ve Şekil 5.19, 5.20, 5.21 ve 5.22’de sunulmuştur. Şekil 5.19’daki birliktelik ağı, Şekil 5.20’deki semantik ağ ve Şekil 5.21’deki ortak kümenin weight değerine ait grafiklerin power-law ilkesine uyumlu oldukları gözlemlenmektedir. Eğri eğimleri sırasıyla 2,15, 3,8 ve 1,9’dur.

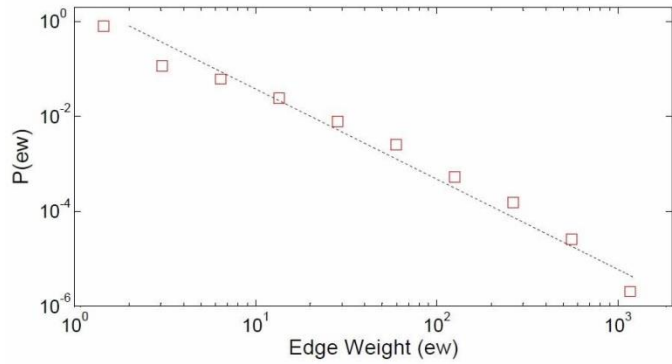
Ortak küme ağının cosine dağılımı Şekil 5.22’de gösterilmiştir. Sağdaki saturasyon bölgesinden önce 2,2 eğime sahip bir power-law uyuma sahiptir.



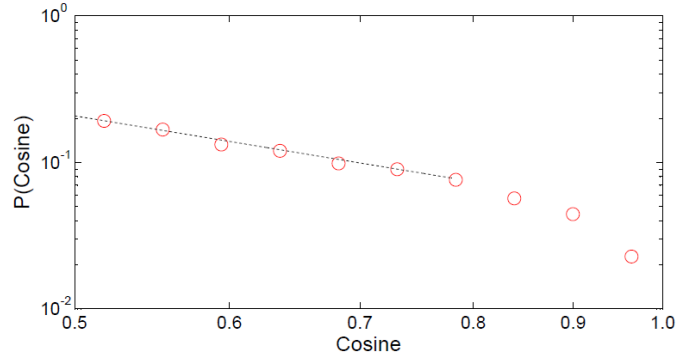
Şekil 5.19. Birliktelik ağı linklerinin ağırlık dağılımı (Log-binned).



Şekil 5.20. Semantik ağı linklerinin ağırlık dağılımı (Log-binned).



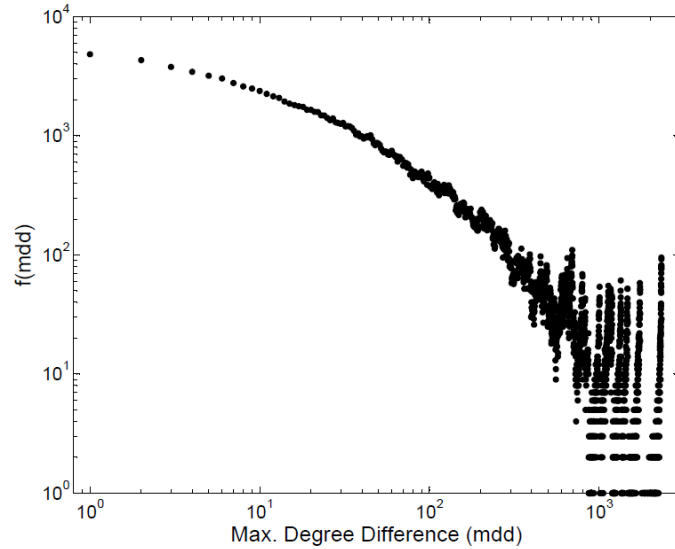
Şekil 5.21. Ortak küme linklerinin ağırlık dağılımı (Log-binned).



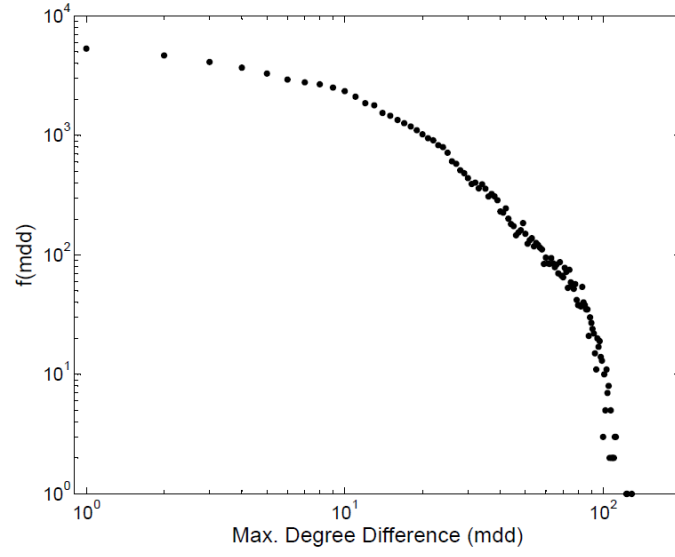
Şekil 5.22. Ortak küme linklerinin cosine dağılımı (Log-binned).

5.2.2.2. Maximum Degree Difference Scatter Grafikleri

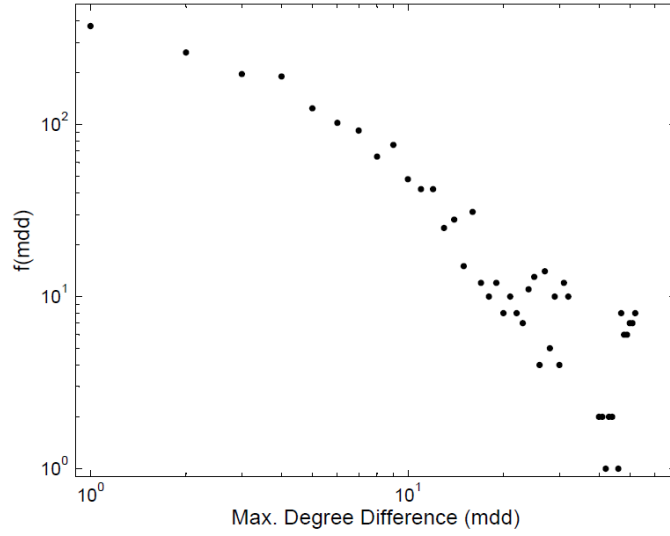
Standart dağılım prosedürleri dışında tezimiz kapsamında düğümlerin bağlanma tercihleri de incelenmiştir. Bu bölümdeki analizde, her linkin iki ucundaki düğümlerin derece farklarına ait dağılım, öncelikle saçılmış scatter grafikleri ile, sonra da logaritmik eş aralıklı dağılım grafikleri ile ifade edilmiştir. Linklerin bağladığı düğümlerin derecelerinin hesaplanmasında ağın son hali, dolayısıyla düğümlerin nihai (maksimum) dereceleri baz alınmıştır. Şekil 5.23, 5.24 ve 5.25’de scatter çizimler, Şekil 5.26, 5.27 ve 5.28’de ise dağılım grafikleri sunulmuştur.



Şekil 5.23. Birliktelik ağına ait linklerin bağladığı düğümlerin, ağın son halindeki maksimum derecelerinin farklarına ait frekans dağılımı (scatter).

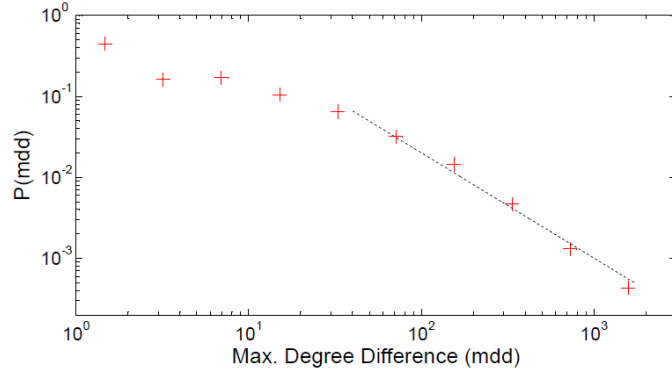


Şekil 5.24. Semantik ağına ait linklerin bağladığı düğümlerin, ağın son halindeki maksimum derecelerinin farklarına ait frekans dağılımı (scatter).

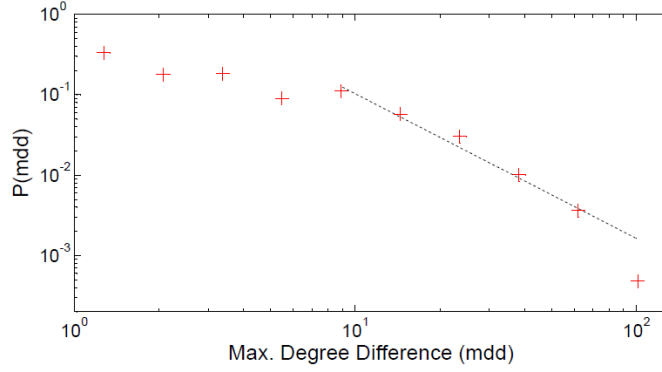


Şekil 5.25. Ortak kümeye ait linklerin bağladığı düğümlerin, ağın son halindeki maksimum derecelerinin farklarına ait frekans dağılımı (scatter).

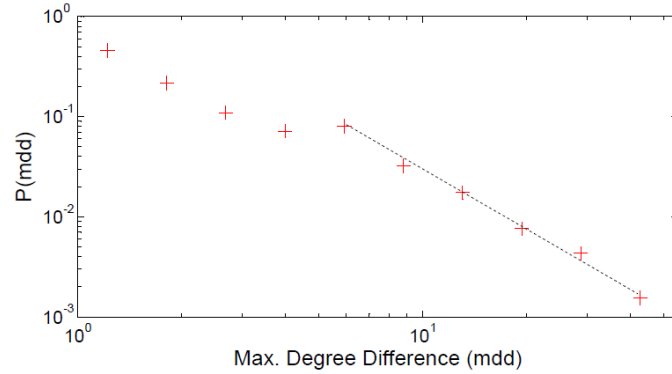
Şekil 5.23, 5.24 ve 5.25'deki scatter çizimler, Şekil 5.26, 5.27 ve 5.28'dekigibi eş aralıklı olasılık yoğunluk fonksiyonuna dönüştürüldüklerinde 3 ağımızın da sol tarafta oluşmuş olan bir saturasyon bölgesinin ardından power-law uyumlu hale geldiği görülmektedir. Birliktelik ağı, semantik ağ ve ortak küme ağına ait eğimler sırasıyla 1,3, 1,8 ve 2,0'dır.



Şekil 5.26. Birliktelik ağına ait linklerin bağladığı düğümlerin ağıın son halindeki maksimum derecelerinin farkına ait dağılım grafiği (Log-binned).



Şekil 5.27. Semantik ağına ait linklerin bağladığı düğümlerin ağıın son halindeki maksimum derecelerinin farkına ait dağılım grafiği (Log-binned).



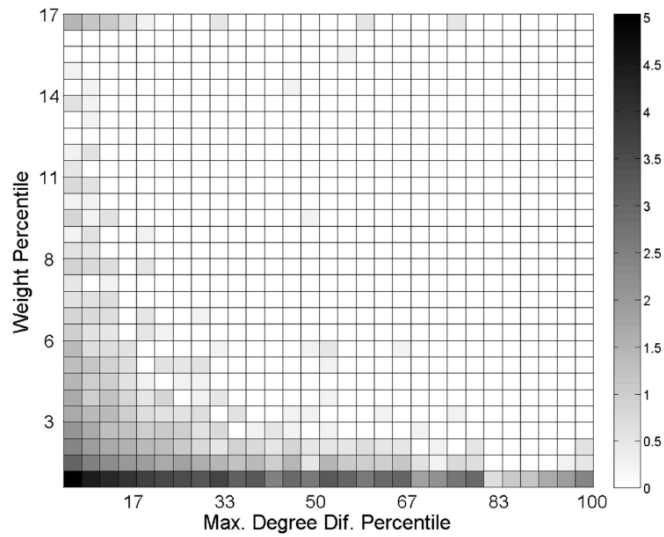
Şekil 5.28. Ortak küme ağına ait linklerin bağladığı düğümlerin ağıın son halindeki maksimum derecelerinin farkına ait dağılım grafiği (Log-binned).

Derece dağılımı haricindeki parametrelere ait dağılımlarda da power-law uyumuna rastlanması, gerçel ağlarda rastlanan bir durumdur [64]. Çalışmamızda derece dağılımı haricinde link ağırlık ve cosine dağılımları ile birlikte linklerin bağladığı

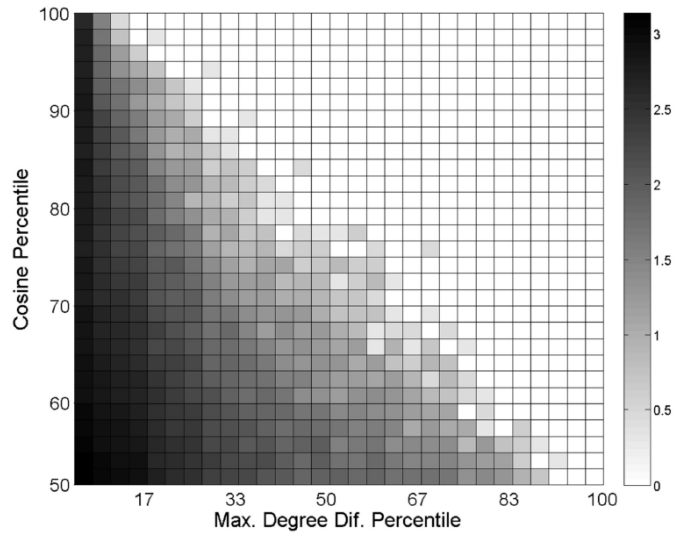
düğümlemlerin derece farklarına ait dağılımlar da power-law uyumlu çıkmış, bu uyumun gerçel sistemlerin çoğu istatistiğine yansıyan universal bir kural olduğu bir kez daha ortaya koyulmuştur.

5.2.2.5. 2 Boyutlu Yüzdeli Dağılım Grafikleri (Percentile)

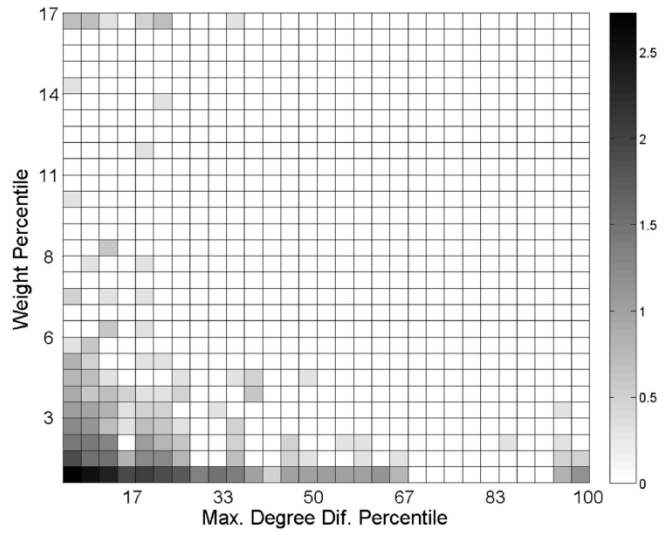
Ağlarda bulunan linklerin ağırlıklarının ve linklerin bağladığı düğümlerin maksimum derece farklarının yüzde dağılımı cinsinden görselleştirilmesi Şekil 5.29, 5.30, 5.31, 5.32 ve 5.33'te verilmiştir. Percentile olarak isimlendirilen bu grafiklerde [65] her iki eksenle ilgili değerler, o parametreye ait maksimum değerin yüzdesi şeklinde ifade edilmiştir. Grafikteki noktaların renkleri, yanlarındaki renk skalası gibi koyulaştıkça artan yoğunluğu ifade etmekte, ancak renk skalasındaki değerlere ($\sim 0-5$ aralığı) göre üstel olarak (10^x) ölçeklenmektedir. Yani örnek olarak Şekil 5.29'da yatay eksenle %17, dikey eksenle %3 skalalarına karşılık gelen noktanın koyuluk değeri $\sim 1,5$ ise, ilgili derece farkı – link ağırlığı ikilisinin ağda gerçekleştirdiği $\sim 10^{1,5}$ linkin var olduğu anlaşılmaktadır.



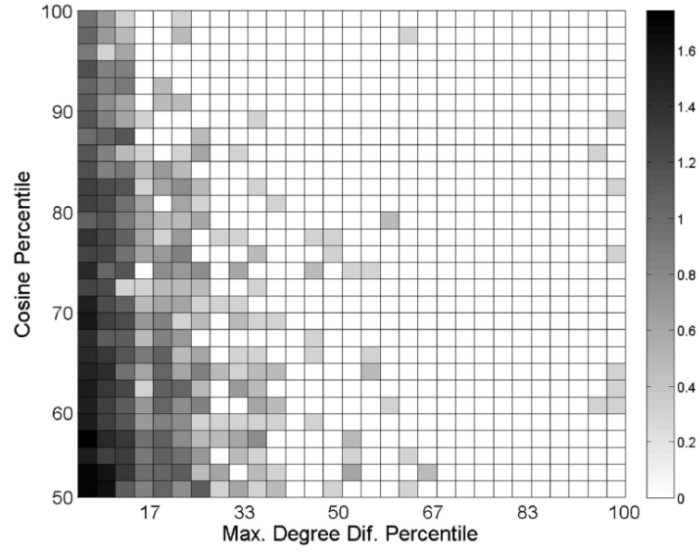
Şekil 5.29. Birliktelik ağına ait link ağırlığı – maksimum derece farkı percentile grafiği.



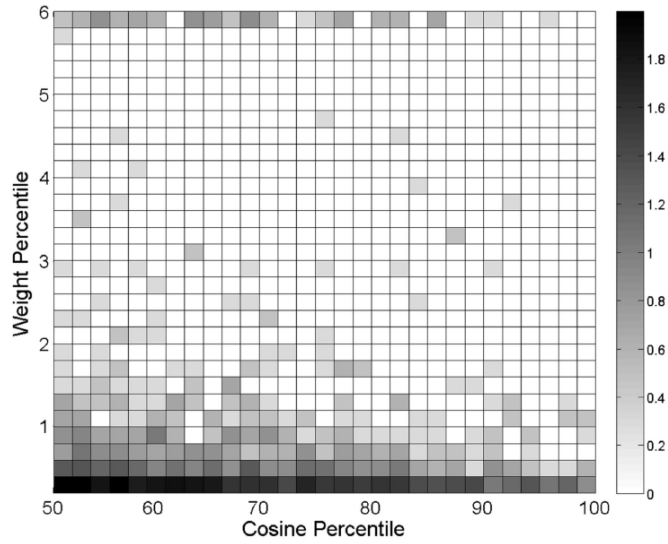
Şekil 5.30. Semantik ağına ait cosine – maksimum derece farkı percentile grafiği.



Şekil 5.31. Ortak küme ağına ait link ağırlığı – maksimum derece farkı percentile grafiği.



Şekil 5.32. Ortak küme ağına ait cosine – maksimum derece farkı percentile grafiği.



Şekil 5.33. Ortak küme ağına ait link ağırlığı – cosine farkı percentile grafiği.

Percentile grafiklerinin yorumlanmasına Şekil 5.29 ile başlayacak olursak, bu grafikte link ağırlıklarının ilk %6'lık, özellikle de %1'lik aralıkta yoğunlaştığı, buna karşılık linklerin bağladığı düğümlerin derece farkı skalasının her aralığında (% 0-100) düğüm çiftlerine rastlandığı görülmektedir. Birliktelik ağındaki baskın durumun, ağırlığı çok düşük linklerin (~%1) her aralıkta dereceye sahip düğümleri birbirine bağladığı bir yapı olduğu anlaşılmaktadır. Link ağırlığı arttığında ise bu linklerin bağladığı düğümlerin daha çok birbirine yakın dereceye sahip düğümler

olduğu görülmektedir. Derece farkı yüksek olan eşleşmelerin, orta-yüksek ağırlığa sahip olamadığı, hep çok düşük (<%3) ağırlıkta rastlandığı anlaşılmaktadır.

Şekil 5.30'da ise semantik ağda linklere ait cosine değerlerinin derece farkları ile ilişkisi görülmektedir. Bu grafiğin verdiği ilk izlenim, her iki parametrenin de her ölçüğünde (% 0-100) linklere rastlandığı yönündedir, ancak ağdaki linklerde bu iki parametrenin ters orantılı bir şekilde oluşum gösterdiği görülmektedir. Yani linklerin cosine değeri arttıkça, bağladığı düğümlerin derece farklarının azaldığı, derece farkı arttıkça da daha düşük cosine değerlerine sahip olma eğiliminin baskın olduğu bir tablo mevcuttur. Cosine eksenini üzerindeki yoğunluk dağılımı daha homojen iken, derece farkı ekseninde sol kısmın (<%50) daha yoğun olduğu görülmektedir.

Birliktelik ve semantik ağlarına ait Şekil 5.29 ve 5.30'da göze çarpan temel fark, linklerin cosine dağılımı daha homojen iken ağırlık dağılımının çok düşük ağırlıklarda sıkışmış olmasıdır. Bu durumu etkileyen temel neden olarak önceki bölümde bahsedilen, semantik ağın zamandan bağımsız olarak anlamsal olarak yakın olan tüm hashtag'leri bağlaması, birliktelik ağının ise zaman bağımlı olması nedeniyle eş zamanlı hashtag'ler arasında ömrü çabuk biten, yerini sürekli yenilenen yeni hashtag'lerin yine kısa süreli ve sınırlı birlikteliklerine bırakan yapıda olması gösterilebilir.

Şekil 5.31'de ise, birliktelik ağı ve semantik ağda ortak olarak kullanılan düğüm ve linklere ait ağırlık-derece farkı percentile grafiği görülmektedir. Bu grafik Şekil 5.29'daki birliktelik ağı percentile grafiği ile çok benzer özelliklere sahiptir, ancak iki ağın kesişim kümesini temsil ettiği için çok daha az sayıda düğüm ve linkten oluşmaktadır. Birliktelik ağının semantik olarak da doğrulanmış alt kümesinin, birliktelik ağı ile benzer özellik sergilemesi, birliktelik ağı yapısının organizasyon prensipleri açısından semantik alt kümesi ile benzerlik taşıdığını göstermektedir. Yani semantik olarak doğrulanmamış çok sayıdaki link de aslında semantik alt küme (kesişim kümesi) ile aynı prensipler ile organize olmaktadır.

Şekil 5.32'deki cosine-derece farkı percentile grafiği ise, yine birliktelik ve semantik ağlarının ortak kümesinden üretilmiş olup Şekil 5.30'daki semantik ağa ait grafik ile

karşılaştırılmalıdır. Bu grafik, Şekil 5.30'daki homojen ve kurallı dağılımın aksine daha rassal ve derece farkı skalasının sol tarafında daha çok kümelenmiş karakterdedir. Cosine eksenini boyunca düzgün dağılım yine gözlenmektedir. Ancak derece fark ekseninin sağ kısmının seyrekleşmiş olması, semantik ağın birlikte kullanımını da içeren alt kümesinin daha çok yakın dereceli (<%50) düğüm birlikteliklerinden oluştuğunu göstermektedir. Buradan da anlaşıldığı gibi saf semantik ağ her alanda daha fazla homojen dağılım sergilemektedir.

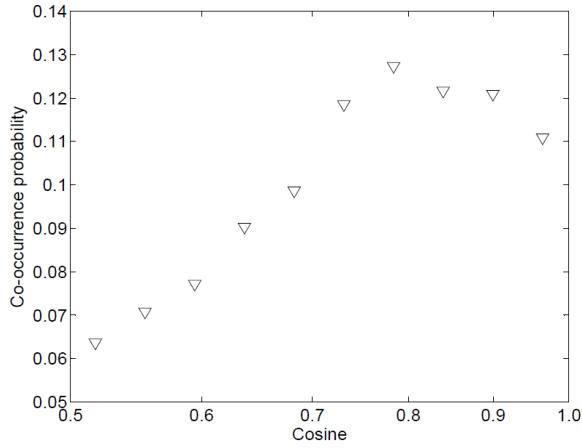
Ortak küme ağındaki linkler, iki ağın kesişimini ifade ettiğinden hem cosine, hem de weight değerlerine sahiptirler. Bu ağdaki semantik yakınlık belirtisi olan cosine ile birlikte kullanım eğiliminin göstergesi olan weight değerlerinin birbiri ile ilişkisi Şekil 5.33'de gösterilmiştir. Yine zaman bağımsız yapının gereği olan homojen cosine dağılımı ile birlikte düşük yüzdelere (<%2) sıkışmış weight dağılımı dikkat çekmektedir. Grafiğin dikkat çekici bir özelliği de çok yüksek cosine değerlerinden ziyade, 0,5'e daha yakın olan (~%50-70) cosine skalasının biraz daha yüksek weight değerleri ile eşleşme eğilimi göstermesidir.

5.2.3. Birlikte Kullanım Ve Semantik Yakınlık İlişkisinin Sorgulanması

Bu bölümde oluşturulmuş olan birliktelik ağı ve semantik ağın ortak kümelerinde bulunan linkler dikkate alınarak birbirleri ile olan ilişkileri analiz edilmiştir. Bu analizler hashtag'lerin cosine yakınlıklarına göre birlikte kullanılma ihtimalleri ve hashtag'lerin ağırlıklarına göre cosine yakınlık kurma ihtimalleridir.

5.2.3.1. Hashtag'lerin Cosine Yakınlıklarına Göre Birlikte Kullanılma Olasılıkları

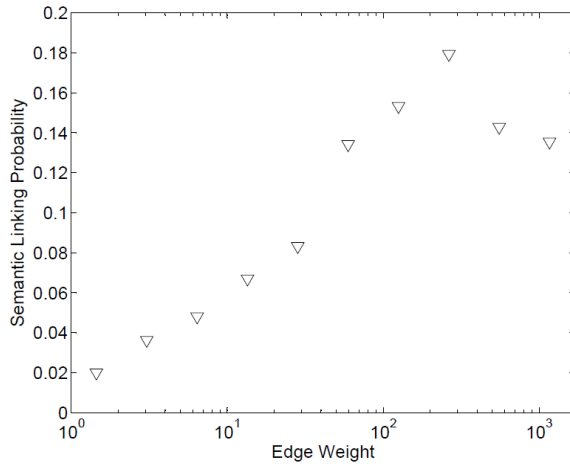
Yapılan analiz sonucunda cosine yakınlığı yüksek olan linklerin birlikte kullanılma ihtimallerinin de yüksek olduğu, yani aralarında doğru orantı olduğu görülmüştür. Bu analize ait grafik şekil 5.34'te görülmektedir.



Şekil 5.34. Hashtag'lerin semantik yakınlığa (cosine) bağlı olarak birlikte kullanılma olasılık dağılımı.

5.2.3.2. Hashtag'lerin Birlikte Kullanılma Sıklıklarına Göre Cosine Yakınlıklarının Yüksek Olma Olasılıkları

Yapılan analiz sonucunda birlikte sık kullanılan linklerin yani ağırlığı yüksek olan linklerin cosine yakınlıklarının da yüksek olma ihtimali olduğu görülmüştür. Bu analize ait grafik şekil 5.35'de görülmektedir.



Şekil 5.35. Hashtag'lerin birlikte kullanım ağırlıklarına göre semantik yakınlık kurma olasılıkları.

Şekil 5.34 ve 5.35'deki grafikler sırasıyla semantik yakınlığa bağlı olarak birlikte kullanım ve birlikte kullanım eğilimine bağlı olarak semantik yakınlık dağılımlarını göstermekte, bu dağılımlar ise yüksek oranda benzerlik sergilemektedir. Bu

benzerliđi nicel olarak ifade etmek için bu dađılımların korelasyonu hesaplanmış ve 0,9728 olarak bulunmuştur.

Gerek her iki grafikte de deđişkenlerin birbiri ile dođru orantılı deđişimleri, gerekse iki dađılımın yüksek oranda korelasyona sahip olması, sosyal ađlarda hashtag'lerin birlikte kullanımının anlamsal olarak da yüksek dođrulukta gerçekteştiđinin bir göstergesidir. Ancak bu durum, iki ađın kesişim kümesi olan nispeten çok küçük bir veri seti için ortaya koyulmuştur.

Birliktelik ve semantik ađları tek başına deđerlendirildiklerinde kesişim ađına göre çok fazla düđüm ve link içermektedir (Çizelge 5.2). Bu durumun oluşmasında semantik sebepler dışında, her iki ađın oluşturulması esnasında farklı ön işleme proseslerinden geçirilmesi de önemli bir etkidir. Bu nedenle birliktelik ađı ve semantik ađ aslında önemli bir miktarda ortak olmayan düđüm de içermektedir. Bu nedenle hashtag'lerin birlikte kullanımının semantik dođruluđu ancak iki ađın kesişim kümesi olan 7 647 düđüm ve 2 247 link için en sağlıklı biçimde sorgulanabilir. Öte yandan bu iki ađın kesişim kümesini oluşturan 7 647 düđümün her iki ađ üzerindeki izdüşümleri de ayrıca incelendiğinde, bu düđümlerin birliktelik ve semantik ađlarda sırasıyla 61 612 ve 42 270 adet linke sahip olduđu görülmekte, bu iki izdüşüm de aslında 4. Ve 5. alt ađlar olarak çalışmamıza farklı bir derinlik getirmektedir. Bu ađlar sırasıyla indirgenmiş birliktelik ve indirgenmiş semantik ađlar olarak isimlendirilebilir.

İndirgenmiş birliktelik ađı linkleri, ortak ađ linkleri ile karşılaştırıldığında (düđüm kümeleri ortak) bu birlikte kullanımların hangi oranda semantik dođruluđu sahip olduđunu tanımlayabilir. Bu amaçla indirgenmiş birliktelik ađı linkleri incelendiğinde 61 612 adet linkin, ortak ađda yer alan 2 247 tanesinin semantik olarak deđerlendirildiđi sonucuna ulaşılmıştır. Dolayısıyla Twitter ortamında birlikte kullanılan hashtag'lerin istatistiksel olarak semantik olarak dođruluk oranı % 3,64 olarak kabul edilebilir.

Benzer biçimde indirgenmiş semantik ađ linkleri ile ortak ađ linkleri ile karşılaştırıldığında (ortak düđümler) yakın anlamdaki hashtag'lerin ne oranda birlikte

kullanıma sahip olduđu hakkında bilgi sahibi olunabilir. İndirgenmiş semantik ağdaki 42 270 adet linkin 2 247 tanesinin ortak ağda doğrulanmış (birlikte kullanılmış) olduđu görülmekte, bu da %5,31 oranında birlikte kullanıma sahip olduklarını ifade etmektedir. Ancak unutulmamalıdırki birlikte kullanım ağı zaman bağımlı olduğundan tüm semantik yakınlıkların birlikte kullanıma dönüşebilmesi beklenemez.

Hashtag'lerin birlikte kullanımlarının %3,64 oranında semantik doğruluklarının olması ve semantik yakınlıklarının %5,31 oranında birlikte kullanıma yatkın olması verilerine bakılarak, semantik yakınlığın birlikte kullanıma göre daha yüksek doğrulukta olduđu söylenebilir. Bunda daha önce belirtildiği gibi semantik ağın zamandan bağımsız bir yapıda olup, anlamsal yakınlık kurması gösterilebilir. Hem kurulmuş ağın görsel çıktılarında semantik ağdaki belirgin kümelenme hem de detaylı analizlerinden yola çıkılarak semantik yakınlığın, birlikte kullanıma göre daha doğru sonuç verdiği sonucuna ulaşılabilir.

BÖLÜM 6

SONUÇ VE ÖNERİLER

6.1. SONUÇLAR

Twitter mikroblog sitesinden elde edilen veriler kullanılarak hashtag etiket ağı analizi yapılan bu çalışmada 2 farklı yaklaşım uygulanmıştır. Bunlar; hashtag birlikte kullanım ağı ve hashtag semantik yakınlık ağıdır. İki ağ birbirinden bağımsız olarak ve farklı yöntemler ile analiz edilmiştir.

Üzerinde çalışılan iki ağın birbirinden farklı özelliklere sahip olması sebebi ile veri işleme aşamasında farklı yöntemler kullanılmıştır. Bu farklılıkların en başında ağ kurulurken izlenen yol gelmektedir. Birliktelik ağı oluşturulurken birden fazla hashtag'in birlikte kullanıldığı Twitter metinleri dikkate alınırken, semantik ağın oluşturulmasında anlamsal olarak analiz yapılacağından, sadece bir hashtag kullanılmış Tweet metinleri dikkate alınmıştır.

Birliktelik ağı incelendiğinde hashtag'lerin birlikte kullanımlarının zamana ve olaylara bağlı olduğu görülmüştür. Ancak semantik olarak incelenen semantik ağın zamandan bağımsız olması sebebi ile linklerin tanımlanmasında hashtag'lerin hangi zaman diliminde kullanıldığına bir önemi olmadığı gözlemlenmiştir.

Oluşturulmuş iki ağın ortak düğüm ve linklerinden ortak küme ağı oluşturulmuş ve incelenmiştir. Gerek birliktelik ağı, gerekse semantik ağına ait derece dağılımlarında tipik "self-organization" karakteristiği olan güç yasası uyumluluğu gözlenmektedir. Bu uyum ortak ağda daha kusursuz yapıdadır. Bu da semantik olarak da doğrulanmış birlikte kullanımların daha saf organizasyona sahip olduğunun bir göstergesidir.

Derece dağılımları ile birlikte, ağlara ait link ağırlık dağılımları da (edge weight) güç yasası uyumlu olarak gözlemlenmiştir. Semantik ağda ise ağırlık yerine cosine yakınlık değeri incelenmiştir. Linklerin cosine yakınlığına ait dağılım da benzer biçimde güç yasası uyumludur. Bu durum literatürde ilk defa bu tez çalışmasında deneysel olarak ortaya koyulmuştur. Linklerin her iki ucundaki düğümlerin derece farklarına ait dağılım da yine literatürde örneği olmayan bir dağılım çalışmasıdır. Bu dağılımlar da gerek birliktelik, gerek semantik, gerekse de ortak ağlarda güç yasası uyumludur.

Ortak ağ analiz sonuçlarına göre, hem semantik yakınlığı olan hashtag'lerin birlikte kullanılma potansiyeli sahip oldukları, hem de birlikte kullanılan hashtag'lerin anlamsal olarak yakın oldukları sonucuna varılmıştır.

Birliktelik ağının ve semantik ağının ortak küme ağındaki düğümlere indirgenmesi ile oluşturulan indirgenmiş alt ağlar incelenerek o ağa ait istatistikî veriler elde edilmiştir. İndirgenmiş birliktelik ağının %3,64 oranında semantik yakınlık doğruluğuna sahip olduğu, indirgenmiş semantik ağın ise %5,31 oranında birlikte kullanımistatistiği sergilediği gözlenmiştir. Bu rakamlarda yola çıkarak Twitter kullanıcılarının birlikte tag kullanım eylemlerinin sadece anlamsal yakınlığa bağlı olmadığı anlaşılmaktadır. Bu durumun sözlük tabanlı semantik yakınlık haricinde de kullanıcıların bilgisi dahilinde olan örtülü anlamsal ilişkilerden, ya da popüler taglar altında görüntülenerek Twitter'da popülerleşme amacı güdüyor olmasından kaynaklandığı değerlendirilmektedir.

6.2. ÖNERİLER

Çalışmada önışleme aşamalarında kelimeler olabildiğince arındırılmıştır. Tükçe dil yapısından kaynaklanan, kelimelerin kökünü çıkarma işleminin zorluğu çalışmanın doğruluğunu olumsuz etkilemiştir. Kelimeler kök haline getirilerek incelendiğinde aynı köke fakat farklı çekimlere sahip kelimeler de aynı parametre ile ifade edileceğinden semantik olarak daha doğru sonuçlar elde edilebilecektir. Bununla birlikte Twitter'daki karakter kısıtlamasının sebep olduğu bazı olumsuzlukları

gidermek için de daha ileri dil işleme yordamları kullanılarak semantik hassasiyet artırılabilir.

Semantik analiz yapılırken veri setindeki en çok kullanılan 1000 kelime içinden günlük dil kullanımını da göz önünde bulundurularak 500 kelimelik bir parametre seti oluşturulmuştur. Parametre seti daha fazla sayıda kelimedenden oluşturulduğunda ve kelimelerin seçiminde daha fazla kriter göz önünde bulundurulduğunda hashtag'lerin semantik yakınlıklarını hesaplamada daha doğru sonuçlar alınacaktır.

Öte yandan bu tez çalışmasında hashtag'lerin Türkçe yazılmış metinler üzerinden birliktelik analizi yapılmıştır. Ön işleme çalışmaları ve kelime seti seçimi farklı dillere yönelik yapıldığında bu çalışma diğer dillere de uygulanabilir. Çeşitli diller arasında da özellikle hashtag'lerin birlikte kullanım ağı istatistikleri karşılaştırılabilir.

KAYNAKLAR

1. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D. U., “Complex networks: structure and dynamics”, *Physics Reports*, 424: 175-308 (2006).
2. Barrat, A., Barthélemy, M. and Vespignani, A., “Dynamical Processes on Complex Networks,1st ed.”, *Cambridge University Press*, New York, 1-3, 11-13, 26, 27 (2008).
3. Van der Hofstad, R., “Random Graphs and Complex Networks, Vol. I”,*Department of Mathematics and Computer Science, Eindhoven University of Technology*, Eindhoven, Netherlands, 1, 2, 16, 17, 21, 22, 27, 28, 33, 34 (2014).
4. Estrada, E., “Graph and Network Theory,Mathematical Tools for Physicist”, 2nd d., Grinfeld, M., *Wiley-Vch*, Weinheim, 134, 135, 137-139 (2015).
5. Caldarelli, G. and Vespignani, A., “Preliminaries and basic definitions in network theory”, Large scale structure and dynamics of complex networks from information technology to finance and natural science,Caldarelli, G. and Vespignani, A.,*World Scientific Publishing Co.*, Singapore, 2, 3, 5, 6 ,12-15, 205-207 (2007).
6. Barrat, A., Barthélemy, M. and Vespignani, A., “Dynamical Processes on Complex Networks,1st ed.”, *Cambridge University Press*, New York, 1-3, 11-13, 26, 27 (2008).
7. Akal, Ş., “Gerçek ve model ağların karakteristik özelliklerinin karşılaştırılması”, *Marmara Üni. Öneri Dergisi.*, 11 (41): 251-272 (2014).
8. Van Steen, M., “An Introduction to Graph Theory and Complex Networks,1st ed.”,*Maarten van Steen*,11,12, 133-150, 166,167, 172,173, 187-195, 212-214, 225-227 (2010).
9. Demir, S., “Türkiye’de mühendislik alanındaki bilimsel işbirliği ağı haritasının analizi ve görselleştirilmesi”, Yüksek Lisans Tezi, *Karabük Üniversitesi Fen Bilimleri Enstitüsü*, Karabük, 5, 8, 10-11, 15-17, 21, 49-50 (2015).
10. Türker, İ., “Türkiye’deki bilimsel işbirliği ağı ve dinamikleri”, Doktora Tezi, *Karabük Üniversitesi Fen Bilimleri Enstitüsü*, Karabük, 11-15, 27, 28, 30-38, 48-50, 60-63 (2013).
11. İnternet: Network Science, “Node Degree Distribution Random Networks Versus Scale-Free Power-Law Networks”, http://www.network-science.org/powerlaw_scalefree_node_degree_distribution.html (2014).

12. Costa, L., Rodrigues, F., and Cristino, A., "Complex networks: the key to systems biology", *Genetics and Molecular Biology*, 31 (3): 591-601 (2008).
13. Clauset, A., Shalizi, C. R. and Newman, M. E. J., "Power-law distributions in empirical data", *SIAM Review*, 51(4): 661-703 (2009).
14. İnternet: Odd's Blog, "Eight Properties of Social Network", <http://oddllee.blogspot.com.tr/2007/09/eight-properties-of-social-network-14.html>(2014).
15. Caldarelli, G., "Scale-Free Networks Complex Webs in Nature and Technology, 1st ed.", *Oxford University Press*, New York, 39-41, 130-250, 185-205, 208-210, 214, 215, 227, 228, 231-234 (2007).
16. Handcock, M., Raftery, A., Tantrum, J., "Model-based clustering for social Networks", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170 (2): 301-354 (2007).
17. Zhang, Z., Zhang, J., "A big world inside small-world Networks", *Plos one*, 4 (5): e5686 (2009).
18. Newman, M., "Modularity and community structure in networks", *Proceedings of the national academy of sciences*, 103 (23): 8577-8582 (2006).
19. Newman, M. E. J., "Networks:An Introduction", *Oxford University Press*, New York, 15-100,28-32,36-40,63-67, 79-82, 168-222 (2010).
20. Albert, R. and Barabási, A. L., "Statistical mechanics of complex networks", *Reviews of Modern Physics*, 74: 47-97 (2002).
21. İnternet: World Wide Web Size, "The Size of the World Wide Web (The Internet)",<http://www.worldwidewebsite.com> (2016).
22. Huang, C.Y., Sun, C.T., and Lin, H. C., "Influence of local information on social simulations in small-world network models",*Journal of Artificial Societies and Social Simulation*, 8 (4): 1-26 (2005).
23. Donato, D., Laura, L., Leonardi, S., and Millozzi, S., "The web as a graph: How far we are", *ACM Transactions on Internet Technology*, 7(1): 4 (2007).
24. İnternet: Large Networks Visulization Tool, "World Wide Web", <http://lanet-vi.fi.uba.ar/gallery.php>(2016).
25. İnternet: Large Networks Visulization Tool, "Internet Routers", <http://lanet-vi.fi.uba.ar/gallery.php> (2016).

26. İnternet: Hurriyet Daily News “Explained: How 76 Million People Were Hit By Turkey’s Worst Blackout Since 1999”, <http://www.hurriyetaailynews.com/explained-how-76-million-people-were-hit-by-turkeys-worst-blackout-since-1999.aspx?pageID=238&nID=80442&NewsCatID=374> (2016).
27. İnternet: Northwestern University, “Researchers Show Complex Networks Have Some Skeletons”, http://www.mccormick.northwestern.edu/news/articles/2012/06/article_brockmann-networks-skeletons.html (2016).
28. İnternet: KargoTurk, “Dağıtım Ağı”, <http://www.kargoturk.com.tr/dagitim-agi/>(2016).
29. İnternet: Strategic Informatics, “Social Network Data Visualization”, <http://www.strategicinformatics.com/?p=804> (2016).
30. İnternet: Large Networks Visulization Tool, “Scientific Collaborations”,<http://lanet-vi.fi.uba.ar/gallery.php> (2016).
31. Leskovec, J. and Horvitz, E., “Planetary-scale views on an instant-messaging network”, *WWW '08 Proceeding of the 17th International Conference on World Wide Web*, Beijing, 915-924 (2008).
32. Watts, D. J. and Strogatz, S.H., “Collective dynamics of ‘small-world’ networks”, *Nature*, 393 (4): 440-442 (1998).
33. Albert, R., Jeong, H. and Barabási, A. L., “Error and attack tolerance of complex networks”, *Nature*, 406(27): 378-382 (2000).
34. Aksüt M., Ateş, S., Balaban, S., “Lise öğrencilerinin sosyal paylaşım sitelerine ilişkin tutumları”, *XVI. Türkiye’de İnternet Konferansı*, Ege Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, İzmir (2011).
35. İnternet: Statista, “Number of monthly active Facebook users worldwide as of 1st quarter 2016 (in millions)”,<http://www.statista.com/statistics/264810/number-of-monthly-active-Facebook-users-worldwide> (2016).
36. İnternet: World Knowing,” Top 10 Most Facebook User Country In The World”,<http://worldknowing.com/top-10-most-Facebook-user-country-in-the-world/> (2016).
37. İnternet: Twitter,”Şirket”,<https://about.Twitter.com/tr/company> (2016).
38. İnternet: TeknolojiOku, “Rakamlarla Türkiye’de Twitter istatistikleri”,<http://www.teknolojioku.com/foto-galeri/rakamlarla-turkiyede-Twitter-istatistikleri-2387.html?p=5> (2016).
39. İnternet: Twitter, “Getting Started | Twitter Developers”, <https://dev.Twitter.com/start> (2016).

40. İnternet: Twitter, "Using the Twitter Search API | Twitter Developers", <https://dev.Twitter.com/docs/using-search> (2016).
41. İnternet: Twitter, "The Streaming APIs | Twitter Developers", <https://dev.Twitter.com/docs/api/streaming> (2016).
42. İnternet: "Twitter API Tutorial – Aggregating Tweets: Search API vs. Streaming API", <http://140dev.com/Twitter-api-programming-tutorials/aggregating-Tweets-search-api-vs-streaming-api/> (2016).
43. İnternet: 140DEV, "140dev Streaming API Framework ", <http://140dev.com/free-Twitter-api-source-code-library/> (2016).
44. İnternet: 140DEV, "Free Source Code – Twitter Database Server: Code Architecture", <http://140dev.com/free-Twitter-api-source-code-library/Twitter-database-server/code-architecture/>(2016).
45. İnternet: 140DEV, "Free Source Code – Twitter Database Server: Phirehose Library", <http://140dev.com/free-Twitter-api-source-code-library/Twitter-database-server/phirehose/> (2016).
46. Tekir, S., "Semantik bilginin analizi ve modellenmesi", Doktora Tezi, *Ege Üniversitesi Fen Bilimleri Enstitüsü*, İzmir, 9-10 (2010).
47. Mohler, M., and Mihalcea, R., "Text-to-text semantic similarity for automatic short answer grading" *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics*, Singapore,567-575 (2009).
48. Takale, Sheetal A., and Sushma S. Nandgaonkar. "Measuring semantic similarity between words using web documents." *International Journal of Advanced Computer Science and Applications (IJACSA)*, 1(4) (2010).
49. İlhan, S., Duru, N., Karagöz, Ş., ve Sağır, M., "Metin madenciliği ile soru cevaplama sistemi", *ELECO 2008*, Bursa, 356-359 (2008).
50. Daconta, M., Obrst, L., and Smith, K., "The semantic web: a guide to the future of XML, web services, and knowledge management",*John Wiley & Sons*, Indianapolis, 1-4 (2003).
51. Meng, L., Huang, R., Gu, J., "A review of semantic similarity measures in wordnet", *International Journal of Hybrid Information Technology*, 6 (1): 1-12 (2013).
52. Bollegala, D., Matsuo, Y., and Ishizuka, M.,"An integrated approach to measuring semantic similarity between words using information available on the web", *NAACL HLT 2007 - Human Language Technologies 2007: The Conference of the North American Chapter of the Association for*

- Computational Linguistics*, Proceedings of the Main Conference, Rochester, NY, 340-347 (2007).
53. Sowa, J., "Principles of semantic networks", *Morgan Kaufmann*, California, 1-10 (1991).
 54. Dao, T., Simpson, T., "Measuring Similarity between sentences." *WordNet. Net, Tech. Rep* (2005).
 55. İnternet: Türkiye İstatistik Kurumu, "Hanehalkı Bilişim Teknolojileri Kullanım Araştırması, 2015", <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=18660> (2016).
 56. Martinez-Gil, J., and Aldana-Montes, J., "Semantic similarity measurement using historical google search patterns", *Information Systems Frontiers*, 15 (3): 399-410 (2013).
 57. Kamatchi, M., and Sundaram, N., "Measuring semantic similarity between words and improving word similarity by augmenting pmi", *International Journal of Emerging Trends & Technology in Computer Science*, 3 (6): 167-171 (2014).
 58. Karthiga, M., and Sankarananth, S., "A semantic search engine using semantic similarity measure between words", *International Journal of Scientific and Engineering Research*, 4 (5): 82-86, (2013).
 59. Landauer, T., Foltz, P., and Laham, D., "An introduction to latent semantic analysis", *Discourse processes*, 25 (2-3): 259-284 (1998).
 60. İnternet: eTicaretMag, "Türkiye E-Ticaret Perakende Pazar Büyüklüğü [İnfografik]", <http://eticaretmag.com/turkiye-e-ticaret-perakende-pazar-buyuklugu-infografik> (2016).
 61. Albitar, S., Fournier, S., and Espinasse, B., "An effective TF/IDF-based text-to-text semantic similarity measure for text classification", *Web Information Systems Engineering–WISE*, Springer International Publishing, 105-114 (2014).
 62. Joachims, T., "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization", *In Proceedings of ICML-97, 14th International Conference on Machine Learning*, Nashville, 143-151 (1997).
 63. Rajendra, L., Wang, Q., and Raj, J., "Recommending News Articles using Cosine Similarity Function", *SAS Global Forum*, Washington, 1886-1893 (2014).
 64. Li, C., and Chen, G., "Network connection strengths: Another power-law?.", *arXiv preprint cond-mat/0311333* (2003).
 65. Onnela, J., Arbesman, S., Gonzalez, M., Barabasi, A., Christakis, N., "Geographic constraints on social network groups", *PLoS one*, 6(4): e16939(2011).

ÖZGEÇMİŞ

Eyüb Ekmel SULAK 1984 yılında Ereğli’de doğdu; ilk, orta ve lise öğrenimini aynı şehirde tamamladı. Ereğli Yabancı Dil Ağırlıklı Lise'den 2002 yılında mezun oldu. Aynı yıl Marmara Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi bölümü Bilgisayar ve Kontrol Teknolojisi’nde öğrenime başlayıp 2009 yılında orta derece ile mezun oldu. 2008 ile 2010 yılları arasında özel sektörde çeşitli görevlerde yer aldıktan sonra 2010 yılında Karabük Üniversitesi Meslek Yüksekokulunda öğretim görevlisi olarak göreve başladı. Halen Karabük Üniversitesi’ndeki görevine devam etmektedir. Evli ve 1 çocuk babasıdır.

ADRES BİLGİLERİ

Adres : Karabük Üniversitesi
Türkiye Odalar ve Borsalar Birliği Teknik Bilimler Meslek
Yüksekokulu
Balıklarkayası Mevkii / KARABÜK

Tel : (505) 4058669

E-posta : ekmelsulak@karabuk.edu.tr