



# **SPOR AĞLARINDA BAĞLANTI TAHMİNİ YÖNTEMLERİNİN KARŞILAŞTIRILMASI**

**Mine KELEŞ**

**2022  
YÜKSEK LİSANS TEZİ  
BİLGİSAYAR MÜHENDİSLİĞİ**

**Tez Danışmanı  
Dr. Öğr. Üyesi Emrah ÖZKAYNAK**

**SPOR AĞLARINDA BAĞLANTI TAHMİNİ YÖNTEMLERİNİN  
KARŞILAŞTIRILMASI**

**Mine KELEŞ**

**T.C.  
Karabük Üniversitesi  
Lisansüstü Eğitim Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalında  
Yüksek Lisans Tezi  
Olarak Hazırlanmıştır**

**Tez Danışmanı  
Dr. Öğr. Üyesi Emrah ÖZKAYNAK**

**KARABÜK**

**Ocak 2022**

Mine KELEŞ tarafından hazırlanan “SPOR AĞLARINDA BAĞLANTI TAHMİNİ YÖNTEMLERİNİN KARŞILAŞTIRILMASI” başlıklı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

Dr. Öğr. Üyesi Emrah ÖZKAYNAK .....

Tez Danışmanı, Bilgisayar Mühendisliği Anabilim Dalı

Bu çalışma, jürimiz tarafından Oy Birliği ile Bilgisayar Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir. 07/01/2022

Ünvanı, Adı SOYADI (Kurumu)

İmzası

Başkan : Doç. Dr. Salih GÖRGÜNOĞLU ( KÜ) .....

Üye : Dr. Öğr. Üyesi Emrah ÖZKAYNAK ( KBÜ) .....

Üye : Dr. Öğr. Üyesi Ferhat ATASOY ( KBÜ) .....

KBÜ Lisansüstü Eğitim Enstitüsü Yönetim Kurulu, bu tez ile, Yüksek Lisans derecesini onamıştır.

Prof. Dr. Hasan SOLMAZ .....

Lisansüstü Eğitim 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.”*

Mine KELEŞ

## ÖZET

**Yüksek Lisans Tezi**

### **SPOR AĞLARINDA BAĞLANTI TAHMİNİ YÖNTEMLERİNİN KARŞILAŞTIRILMASI**

**Mine KELEŞ**

**Karabük Üniversitesi**

**Lisansüstü Eğitim Enstitüsü**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Tez Danışmanı:**

**Dr. Öğr. Üyesi Emrah ÖZKAYNAK**

**Ocak 2022, 61 sayfa**

Bağlantı tahmini, karmaşık ağlarda düğümler arasındaki bağlantıları analiz ederek gelecekte oluşabilecek muhtemel yeni bağlantıların tahmininde kullanılmaktadır. Özellikle dinamik ağlarda, düğümler arasındaki ilişkilerdeki değişkenlik yeni bağlantıların ortaya çıkmasıyla ya da mevcut bağlantıların kopmasıyla sonuçlanmaktadır. Bu durum ise düğümler arasındaki ilişkilerin gelecekte nasıl şekillenebileceğini analiz etmenin önemini arttırmaktadır. Sosyal ağlar, bilimsel işbirliği ağları, ticaret ağları gibi varlıkları arasındaki ilişkileri sürekli gelişen ve değişen gerçek dünya ağlarında da gelecekte olası yeni bağlantıların tahmin edilmesi önemlidir. Bağlantı tahmininin sosyal ağlar, bilimsel ağlar ve ticaret ağları gibi uygulandığı gerçek dünya ağlarında başarılı sonuçlar vermiş olması, bağlantı tahmini yöntemlerinin dinamik yapıya sahip farklı karmaşık ağlarda kullanılabilmesini mümkün kılmaktadır. Bununla birlikte farklı alanlardan ağlar oluşturularak bağlantı tahmininin yapılabilirliğini gösteren çalışmaların az olduğu görülmektedir. Bu tez

alışmasında, bağlantı tahmini yöntemlerinin mevcut kullanım alanlarından farklı olarak karmaşık ađ özelliklerine sahip spor ađlarındaki başarısı araştırılmıştır. Çalışmada Avustralya Açık Tenis turnuvalarından, UEFA Avrupa Ligi karşılaşmalarından ve Dünya Snooker Şampiyonası müsabakalarından farklı zaman periyotlarında oluşturulan ađlarda komşuluk tabanlı bağlantı tahmini yöntemleri uygulanmıştır. Komşuluk tabanlı bağlantı tahmini yöntemlerinin, uygulandığı ađlardaki başarısını ölçmek için AUC metriđi kullanılmıştır. Yapılan deneysel çalışmalarda komşuluk tabanlı bağlantı tahmini yöntemlerinin spor ađlarında başarılı sonuçlar verdiğini görülmektedir.

**Anahtar Sözcükler :** Bağlantı Tahmini, Karmaşık Ađlar, Sosyal Ađlar, Spor Ađları.

**Bilim Kodu** : 92429

## **ABSTRACT**

**M. Sc. Thesis**

# **COMPARISON OF LINK PREDICTION METHODS IN SPORTS NETWORKS**

**Mine KELEŞ**

**Karabük University  
Institute of Graduate Programs  
Department of Computer Engineering**

**Thesis Advisor:**

**Dr. Emrah ÖZKAYNAK**

**January 2022, 61 pages**

Link prediction is used to predict possible new links that may occur in the future by analyzing links between nodes in complex networks. Especially in dynamic networks, the variability in the relations between nodes results in the emergence of new connections or the breaking of existing connections. This situation increases the importance of analyzing how the relations between nodes can be shaped in the future. It is also important to predict possible new connections in the future in real-world networks such as social networks, scientific cooperation networks, trade networks, whose relationships are constantly evolving and changing. The successful results in real-world networks where link prediction is applied, such as social networks, scientific networks and trade networks, make it possible to use link prediction methods in different complex networks with dynamic structure. However, it is seen that there are few studies showing the feasibility of connection estimation by creating networks from different areas. In this thesis, the success of link estimation methods in sports

networks with complex network properties, different from the current usage areas, was investigated. In the study, neighborhood-based connection estimation methods were applied in networks created in different time periods from Australian Open Tennis tournaments, UEFA Europa League matches and World Snooker Championship competitions. The AUC metric is used to measure the success of the neighborhood-based link prediction methods in the networks where it is applied. Especially in dynamic networks, the variability in the relations between nodes results in the emergence of new connections or the breaking of existing connections. In experimental studies, it is seen that neighborhood-based connection estimation methods give successful results in sports networks.

**Key Word** : Link Prediction, Complex Networks, Social Networks, Sports Networks.

**Science Code** : 9242



## TEŐEKKÜR

Bu tez alıŐmasının planlanmasında, araŐtırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteęini esirgemeyen, engin bilgi ve tecrübelerinden yararlandıęım, yönlendirme ve bilgilendirmeleriyle alıŐmamı bilimsel temeller ışığında şekillendiren sayın hocam Dr. Öğr. Üyesi Emrah ÖZKAYNAK'a sonsuz teşekkürlerimi sunarım.

Sevgili aileme manevi hiçbir yardımını esirgmeden yanımda oldukları için tüm kalbimle teşekkür ederim.

## İÇİNDEKİLER

	<b><u>Sayfa</u></b>
KABUL.....	ii
ÖZET.....	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER .....	ix
ŞEKİLLER DİZİNİ.....	xi
ÇİZELGELER DİZİNİ .....	xii
SİMGELER VE KISALTMALAR DİZİNİ .....	xiii
BÖLÜM 1 .....	1
GİRİŞ .....	1
1.1. LİTERATÜR TARAMASI.....	3
BÖLÜM 2 .....	10
GRAF TEORİSİ.....	10
2.1.GRAF ÇEŞİTLERİ .....	11
BÖLÜM 3 .....	14
KARMAŞIK AĞLAR.....	14
3.1.SOSYAL AĞLAR.....	15
3.2.SOSYAL AĞ ANALİZİ .....	16
3.3.SOSYAL AĞ ANALİZİNDEKİ İŞLEM BASAMAKLARI.....	17
BÖLÜM 4 .....	19
BAĞLANTI TAHMİNİ.....	19
4.1. PROBLEM OLARAK BAĞLANTI TAHMİNİ.....	19
4.2. BAĞLANTI TAHMİN YÖNTEMLERİ.....	20
4.2.1 Benzerlik Tabanlı Bağlantı Tahmin Yöntemleri .....	20

**Sayfa**

4.2.1.1. Yerel Benzerlik Tabanlı Bağlantı Tahmin Yöntemleri.....	21
BÖLÜM 5 .....	25
DENEYSEL ÇALIŞMA .....	25
5.1. VERİ SETİ .....	26
5.2. ZAMAN PERİYODUNA BAĞLI OLARAK AĞLARIN OLUŞTURULMASI .....	28
5.3. DENEY SONUÇLARI.....	29
5.3.1. 2004-2011 Arası Oluşturulan Ağlar .....	29
5.3.2. 2007-2014 Arası Oluşturulan Ağlar .....	31
5.3.3. 2007-2017 Arası Oluşturulan Ağlar .....	33
5.3.4. 2004-2017 Arası Oluşturulan Ağlar .....	35
BÖLÜM 6 .....	38
SONUÇLAR .....	38
KAYNAKLAR .....	40
ÖZGEÇMİŞ .....	40

## ŞEKİLLER DİZİNİ

### Sayfa

Şekil 2.1. Görselleştirilmiş sorun işbirliği ağ örneği..	10
Şekil 2.2. Basit graf örneği.....	11
Şekil 2.3. Yönlü graf örneği.....	12
Şekil 2.4. Multipleks graf örneği. ....	12
Şekil 2.5. Ağırlıklı graf örneği. ....	13
Şekil 3.1. Tenis turnuvaları ağı... ..	14
Şekil 5.1. Deneysel çalışmanın uygulama adımları... ..	25
Şekil 5.2. Snooker Dünya Şampiyonası müsabakaları veri seti.....	26
Şekil 5.3. Avusturalya Açık Tenis Şampiyonası müsabakaları ağ görseli.....	27
Şekil 5.4. UEFA avrupa ligi ağları.....	27

## ÇİZELGELER DİZİNİ

### Sayfa

Çizelge 5.1. Dünya şampiyonası ağı için düğümler ve bağlantılar.....	28
Çizelge 5.2. Zaman periyoduna bağlı olarak oluşturulan eğitim ve test ağları.....	29
Çizelge 5.3. Aralığında oluşturulan snooker ağına bağlantı tahmini AUC sonuçları. .....	30
Çizelge 5.4. 2004-2011 Aralığında oluşturulan tenis ağına bağlantı tahmini AUC sonuçları.....	30
Çizelge 5.5. 2004-2011 Aralığında oluşturulan UEFA ağına bağlantı tahmini AUC sonuçları.....	30
Çizelge 5.6. 2007-2014 Aralığında oluşturulan snooker ağına bağlantı tahmini AUC sonuçları.....	32
Çizelge 5.7. 2007-2014 Aralığında oluşturulan tenis ağına bağlantı tahmini AUC sonuçları.....	32
Çizelge 5.8. 2007-2014 Aralığında oluşturulan UEFA ağına bağlantı tahmini AUC sonuçları.....	32
Çizelge 5.9. 2007-2017 Aralığında oluşturulan snooker ağına bağlantı tahmini AUC sonuçları.....	34
Çizelge 5.10. 2007-2017 Aralığında oluşturulan tenis ağına bağlantı tahmini AUC sonuçları.....	34
Çizelge 5.11. 2007-2017 Aralığında oluşturulan UEFA ağına bağlantı tahmini AUC sonuçları.....	35
Çizelge 5.12. 2007-2017 Aralığında oluşturulan tenis ağına bağlantı tahmini AUC sonuçları.....	36
Çizelge 5.13. 2004-2017 Aralığında oluşturulan snooker ağına bağlantı tahmini AUC sonuçları.....	36
Çizelge 5.14. 2004-2017 Aralığında oluşturulan tenis ağına bağlantı tahmini AUC sonuçları.....	37

## SİMGELER VE KISALTMALAR DİZİNİ

### SİMGELER

$G$  : graf

$V$  : düğüm kümesi

$E$  : bağlantı kümesi

$(C_i)^D$ :  $i$  düğümünün derece merkeziliği

$(C_i)^C$ :  $i$  düğümünün yakıncılık merkeziliği

$(C_k)^B$ :  $i$  ve  $j$  düğümleri arasındaki  $n$  düğümünün yakıncılık merkeziliği

$M(v)$  :  $v$  düğümünün bağlı olduğu düğümler kümesi

$\Gamma(x)$  : düğümünün komşularının kümesi

$S_{xy}$  :  $x$  ve  $y$  düğümlerinin benzerlik skoru

$|\Gamma(x)|$  :  $x$  düğümünün komşularının sayısı

### KISALTMALAR

CN : Common Neighbor (Ortak Komşular)

JI : Jaccard Index (Jaccard İndeks)

SI : Sorenson Index (Sorenson İndeks)

PAI : Preferential Attachment Index (Tercihli Bağlantı İndeksi)

AAI : Adamic – Adar Index (Adamic – Adar İndeks)

RAI : Resource Allocation Index (Kaynak Paylaştırma İnde

## BÖLÜM 1

### GİRİŞ

Günümüzde iletişim araçlarının çeşitliliği ve kullanımındaki artış ile insanlar, nesnelere ve olaylar arasındaki etkileşim de artmaktadır. Günlük hayatta gerçekleşen bu etkileşimin içerdiği unsurlar bir ağ yapısını meydana getirmektedir. Ağların yapısından elde edilen bilgilerin analiziyle ilgilenen karmaşık ağ bilimi, son yıllarda popüler bilim dallarından birisi haline gelmiştir. Karmaşık ağ bilimine artan bu yoğun ilginin temelinde çeşitli biyolojik, sosyolojik, teknolojik ve iletişim alanlarında geliştirilen uygulamaların artması ve karmaşık ağ analizine uygun büyük miktarda gerçek verinin bulunması yer almaktadır[1-4]. Ağdaki bağlantı şekillerinin[5-7] ortaya çıkarılması, gelecekte oluşabilecek muhtemel bağlantıların tahmini[8-13] karmaşık ağ analizinde önemli çalışma alanları olmuştur.

Karmaşık ağlarda yapılan bağlantı tahmini, gözlemlenen bağlantı ve düğümlerin özniteliklerinden ve karmaşık ağın yapısal özelliklerinden elde edilen bilgilere dayanarak eksik bağlantıların ortaya çıkarılması veya gelecekte ortaya çıkabilecek yeni bağlantıların tahmin edilmesi işlemlerini kapsamaktadır[14-15]. Karmaşık ağlarda, nitelikleri belirlenmiş düğümlerin özellikleri kullanılarak mevcut olmayan bağlantıların tahmini yapılabilmektedir. Burada önemli olan düğümler arası ilişkilerin ortaya çıkarılması ve düğüm yapılarının ağdaki bağlantıların oluşmasına olan etkisini tespit etmektir. Çünkü ağda gelecekte ortaya çıkabilecek yeni ilişkilerin ihtimali kadar ağa yeni düğümlerin katılması veya mevcut ilişkilerde kopmalar yaşanması da ihtimal dahilindedir. Karmaşık ağların dinamik bir yapıya sahip olması başarılı tahmin işlemlerinin yapılmasını zorlaştırmaktadır[16]. Bağlantı tahmini aslında var olan ağ yapısının analizine dayalıdır. Bu analiz ne kadar iyi yapılır ve düğümler arasındaki ilişkiler doğru ve ayırt edici özelliklere sahip olarak belirlenebilirse tahmin işleminin doğruluğu da o derece artar.

Karmaşık bir ağın bağlantı tahmini açısından analizi yapılırken düğümler arasındaki bağlantının oluşumu, türü ve ağırlığının ortaya çıkarılması kadar yapılan tahminlem işleminin de doğruluğu dikkate alınmalıdır[16]. Önceki yıllarda ağ yapılarına dayanan çok sayıda bağlantı tahmin yöntemi önerilmiştir [8-13].

Bağlantı tahmini yöntemlerinin uygulama alanlarına bakıldığında, dijital ortamda bilginin aşırı yüklenmesiyle birlikte kullanıcıların ihtiyaç duydukları alanlar için tavsiye sistemi gibi çalışan filtreleme sistemlerinde [17-18], online alışveriş sitelerinde e-ticaret önerisi [19], sosyal ağlarda arkadaşlık önerisi [20-21], bilimsel işbirliği ağlarında yazar, konu, hakem önerisi[22-23] gibi öneri sistemlerinde yaygın olarak kullanılmaktadır. Ayrıca sürekli genişlemeye devam eden dinamik ağ modellerini değerlendirmek için de bağlantı tahmini uygulanabilmekte ve tatmin edici sonuçlar elde edilmektedir[24]. Bağlantı tahmini yöntemlerinin, karmaşık ağ bilimi prensiplerine dayalı ve dinamik bir yapıya sahip ağlarda uygulanabilir olması ve başarılı sonuçlar vermesi, bu yöntemlerin uygulama alanlarının genişlemesinde etkili olmaktadır.

Bu çalışmada önceki yıllarda gerçekleştirilen bağlantı tahmini çalışmalarından farklı olarak, değişik branşlardan oluşturulan spor ağlarında geleneksel komşuluk tabanlı bağlantı tahmini yöntemleri uygulanmıştır. Daha önce ağ analizi yapılarak karmaşık ağ prensiplerine uygun yapıda oldukları tespit edilen Dünya Snooker Şampiyonası müsabakaları [25], Avustralya Açık Tenis Turnuvaları [26] ve UEFA Avrupa Ligi karşılaşmaları [27] ağları kullanılarak farklı zaman periyotları için komşuluk tabanlı bağlantı tahmini yöntemleri uygulanmış ve sonuçlar karşılaştırılmıştır. Elde edilen tahmin sonuçlarının başarısını ölçmek için AUC metriği kullanılmıştır[28]. AUC sonuçları, geleneksel komşuluk tabanlı bağlantı tahmini yöntemlerinin spor müsabakalarından oluşturulan ağlarda gelecekte olması muhtemel karşılaşmaları tahmin etmek için kullanılabilir olduğunu göstermektedir.



## 1.1. LİTERATÜR TARAMASI

Bağlantı tahmini konusu ile ilgili sosyal ağlar kapsamında değerlendirilen literatürde birçok çalışma bulunmaktadır. Farklı alanlarda bağlantı tahmini üzerine yapılan çalışmalara rastlanılmaktadır. Bu alanlar arasında ortak yazar ağı önemli bir yer tutmaktadır. Bu konu üzerine yapılmış bazı bağlantı tahmini probleminin örnekleri aşağıdaki şekildedir.

Nasiri ve arkadaşları [29], yedi multipleks ağ üzerinde bağlantı tahminini çözmek için saf rastgele yürüyüşe dayalı yerel rastgele yürüyüşün genişletilmiş bir versiyonunu önermişlerdir. Bir hedef katmanda yeni bir bağlantının ortaya çıkma ihtimalini bulmak için taraflı rastgele bir yürüyüş tanımlamak için bir multipleks ağda katmanlar arası çıkarılan bilgilerden yararlanma yaklaşımları araştırılmıştır. Birçok bağlantı yönteminden daha iyi performans gösterdiği ve tahmin doğruluğunu artırdığı tespit edilmiştir.

Nasiri, Berahmand, Rostami [30], protein-protein etkileşimli tahmin problemini çözmek için özellik seçimine dayalı Deepwalk'un değiştirilmiş bir versiyonunu sunmuştur. Bu şekilde hem ağ yapısı hem de protein özelliklerine fayda sağlanması amaçlanmıştır. Spesifik olarak özellik seçimi iki adımda tamamlanması hedeflenmiştir. Birincisi orijinal özellik kümesinden bir dizi ilgili özellik boyutsal lığı azaltılarak seçilmiş, ikinci olarak ise seçilen öznelikler kümesinde her bir özelliğe önemine göre bir ağırlık verilir böylelikle her bir özelliğin katkısı diğerlerinden ayırt edilmiştir. Deneysel sonuçlar önerilen yaklaşımın diğer bağlantı tahmini yaklaşımlarına göre daha başarılı olduğunu ve tahminin doğruluğunu artırdığını göstermiştir.

Karimi ve arkadaşları [31], multipleks ağlarda bağlantı tahmini, diğer katmanlardan gelen bilgilere dayanarak düğümler arasındaki eksik bağlantıları tespit etme sorununu çözmüşlerdir. Çeşitli multipleks gerçek dünya ağları üzerinde yapılan deneyler, önerilen CLPES yönteminin gelişmiş sonuçlar üretme kabiliyetini ve diğer bağlantı tahmin algoritmalarına karşı üstünlüğünü kanıtlamıştır. Her katmanın dahili

özelliklerinin yeni bir dış benzerlik metriği (ExSim) ile birleştirilmesi, ağda bağlantı oluşumu olasılığı için nihai değerleri elde etmektedir.

Zou ve arkadaşları [32], büyüyen ağlardaki bağlantı tahmini problemlerini araştırmışlardır. Yaygın olarak kullanılmakta olan bağlantı tahmin indekslerine dayalı olarak gelişen ağlar için iki düğüm arasındaki eksik bağlantıların varlığının olasılığını tahmin etmek için zaman dilimine dayalı bir ölçüm önermişlerdir. Önerilen ölçümlerin, özellikle bozulma faktörleri küçük olduğunda, zamanla azalma faktörü olan büyüyen ağlar için mevcut ölçümlerden daha iyi performans gösterdiği tespit edilmiştir. Ayrıca, tahmin verimliliğini ve uygulanabilirliğini geliştirmek için, gerçek dünya ağları için optimal dilim sayısı ve bozulma faktörü için fonksiyon ifadeleri önerilmiştir.

Luo ve arkadaşları [33], düğüm grubu arasındaki farklı ilişki türlerini modelleyen bir multipleks ağdaki farklı katmanların yapısal özelliklerinin bir dereceye kadar birbirleriyle ilişkili olduğunu bulmuşlardır. Önerilen yöntemde her katmanı ağırlıklandırmak için kosinüs benzerliğine dayalı bir katman benzerliği ölçüsü tanımlanmıştır. Önerilen yöntemin performansı kapsamlı deneylerle analiz edilmiştir. Sonuçlar, önerilen yöntemin doğruluk ve çalışma süresi açısından mevcut yöntemlerden daha başarılı olduğunu göstermiştir.

Daud ve arkadaşları [34], sosyal ağların hızlı bir şekilde büyümesi ile birlikte özellikle karmaşık sosyal ağ modellemesinde önemli ilerlemeler olduğunu, bağlantı tahminin ise daha zorlayıcı hale geldiğini tespit etmişlerdir.

Dinamik ağ, ağırlıklı ağ, heterojen ağ ve çapraz ağ gibi birçok farklı alanda bağlantı tahmini analizi ile ilgili çok sayıda uygulama yapılmıştır. Ancak, öneri sistemi, anormallik tespiti, etki analizi ve topluluk tespiti gibi bağlantı tahmini uygulamaları, ağ çeşitliliği, karmaşık ve dinamik ağ bağlamları nedeniyle daha yorucu hale gelmektedir. Mevcut incelemelerin boşluğunu azaltmak için, kapsamlı bir inceleme sağlamak ve sosyal ağlara odaklanarak farklı sosyal ağ bağlamlarında ve analizlerinde bağlantı tahmini uygulamaları yapılmıştır.

Malhotra ve arkadaşları [35], karmaşık ağların, farklı alanlar hakkında bilgi verdiğiğine değinmişlerdir. Denetimli makine öğrenmesi tekniklerini kullanarak tek katmanlı ve çok katlı ağlarda gelecekteki bağlantıları bulmak için bir çözüm önermişlerdir. Makine öğrenimi sınıflandırıcılarını eğitmek için ağın bir dizi topolojik özelliğini ele alınmıştır. Tasarlanan eğitim ve test veri seti oluşturma çerçevesi, önerilen yöntemlerin farklı ağlarda değerlendirilmesine yardımcı olmaktadır.

Berahmand ve arkadaşları [36], yarı yerel yöntemler kategorisinde en iyi bilinen algoritmalarından biri olarak kabul edilen yerel rastgele yürüyüşü araştırmışlardır. Karşılıklı bilgi kullanılarak, düğümlerin asimetrik karşılıklı etkisi kavramını önermişlerdir. Önerilen yöntem ile diğer benzerliğe dayalı yöntemler (yerel, yarı yerel ve küresel) arasında bir karşılaştırma yapılmış ve 11 gerçek dünya ağı için sonuçlar rapor edilmiştir. Önerilen yöntem ile diğer benzerliğe dayalı yöntemler (yerel, yarı yerel ve küresel) arasında bir karşılaştırma yapılmış ve 11 gerçek dünya ağı için sonuçlar rapor edilmiştir. Diğer bağlantı tahmini yaklaşımlarına kıyasla daha yüksek bir tahmin doğruluğuna sahiptir.

Shan ve arkadaşları [37], aynı varlık grubunun farklı etkileşim türlerinde gerçekleştirdikleri çalışmada gerçek karmaşık sistemleri tanımlamak için multipleks ağları yeniden tanımlamışlardır. Multipleks ağlarda bağlantı tahmini uygulamak için denetimli bir yöntem önermişlerdir. Bağlantı tahminini ikili bir sınıflandırma problemi olarak gören yöntemde, bir sınıflandırma modeli, tüm katmanlardan çıkarılan düğüm çiftlerinin ayrıntılı yapısal özellikleri kümesi tarafından beslenir. Önerilen yöntemin etkinliğini analiz etmek için altı ağ üzerinde kapsamlı deneyler yapılmıştır. Sonuçlar, önerilen yöntemin karşılaştırılan yöntemlerden önemli ölçüde daha iyi performans gösterdiğini göstermiştir.

Tang ve arkadaşları [38], çevrimiçi sosyal ağ (OSN) uygulamaları farklı deneyimler sunduğuna gösteren çalışmalarında, örneğin Twitter'da kısa bir metin yayınlamak ve Instagram'da fotoğraf paylaşmak vb, birden çok bir multipleks ağı oluşturulmuştur. Multipleks ağdaki katmanlar arası bağlantı tahmini, farklı OSN'lerdeki hesapların aynı kişiye ait olup olmadığını belirlemeyi amaçlamaktadır. Böylece siber suçlu davranış

modellemesi ve müşteri ilgi analizi gibi uygulamalarda yardımcı olabilmektedir. Çalışmada multipleks ağda katmanlar arası bağlantı tahmini için yinelemeli derece cezası (IDP) algoritması geliştirilmiştir. Öncelikle, daha az bağlantıya sahip ortak bir eşleşen komşuya daha fazla ağırlık veren bir derece cezası ilkesi önerilmiştir. Sonrasında, eşleşmeyen tüm düğüm çiftlerinin eşleşme derecesini verimli bir şekilde elde etmek için düğüm komşuluk matrisi çarpımı uygulanmıştır. Eşleşen derece matrisinden ara katman bağlantı tahmin sonuçlarını elde etmek için onaylanmış maksimum değer yöntemi kullanılmıştır. Tahmin sonuçları, önsel katmanlar arası düğüm çifti kümesine eklenilmiştir. Herhangi bir katmandaki eşleşmeyen tüm düğümler eşleşene veya eşleşmeyen düğüm çiftlerinin tüm eşleşen dereceleri 0'a eşit olana kadar yinelemeli olarak gerçekleştirilmiştir.

Najari ve arkadaşları [39] yaptıkları çalışmada multipleks ağlarda gelecek veya eksik bağlantıları tahmin etmek için yeni bir çerçeve önermektedirler. Bağlantı tahmini çerçevesi, bağlantı tahmininin dikkate alındığı katmandan çıkarılan katmanlar arası benzerlik ve yakınlık tabanlı özelliklere dayanmaktadır. Adamic-Adar ve Jaccard Katsayısı gibi yaygın olarak kullanılan yakınlık tabanlı özellikler göz önünde bulundurulmuştur. Başlangıçta tek katmanlı ağlardaki eksik bağlantıları tahmin etmek için kullanılan özellikler, öğrenme gerektirmez ve bu nedenle hesaplanması kolaydır. Önerilen yöntem, bağlantı tahmini amacıyla katmanlar arası benzerliği hesaba katan sistematik bir yaklaşım sunar. Hem sentetik hem de gerçek multipleks ağlar üzerinde yapılan deneysel sonuçlar, önerilen yöntemin etkinliğini ortaya koymakta ve multipleks ağlarda bağlantı tahmin problemi için önerilen son teknoloji algoritmalarından daha üstün performansını göstermektedir.

Chi ve arkadaşları [40], dinamik ağlarda düğümler arasındaki çekim kuvvetine (DLPA) dayalı yeni bir bağlantı tahmini yaklaşımı önermişlerdir. Önerilen yaklaşımda, kaybolan bağlantıların tespiti ve potansiyel bağlantıların gelecekte gerçek bağlantı haline gelip gelmeyeceğinin tahmin edebilmeyi hedeflemişlerdir. Ayrıca önerilen yaklaşımla hesaplanan potansiyel bağlantıların bağlantı olasılıkları ağın evrimine göre değiştirilebilmiştir. Önerilen yaklaşımın performansını değerlendirmek

için statik ve dinamik gerçek dünya ağları üzerinde deneyler yapılmış ve elde edilen sonuçlar önerilen yaklaşımın tahmin doğruluğu açısından başarılı göstermiştir.

Wang ve arkadaşları [41], geniş ve kapsamlı maksimum bağlantılı alt grafiğe sahip uç noktanın diğer uç noktaları çekme olasılığının daha yüksek olduğu gösteren çalışmalarında hem büyük dereceye hem de H-endeksine sahip olan uç noktanın geniş ve kapsamlı maksimum bağlantılı alt grafiğine sahip olduğunu göstermişlerdir. Çalışmada, derece ve H-endeksinin ağırlıklı sentetik uç nokta etkisine dayalı bir model önerilmiş ve on iki gerçek dünya veri seti üzerinde deneyler gerçekleştirilmiştir. Sonuçlar, ağırlıklı sentetik etkinin bağlantı tahmininde başarılı olduğunu göstermiştir.

Kumar ve arkadaşları [42], ortak komşu yaklaşımları fikrine dayalı olarak, iki kişi arasındaki çok sayıda ortak arkadaş kişileri (düğümler), aralarında daha fazla benzerlik ve daha fazla bağlantı kurma olasılığını ifade eder fikrini savunmuşlar ve çalışmalarını bu doğrultuda gerçekleştirmişlerdir. Kaynak tahsisi sürecinde, iki düğümün ortak komşularının çok sayıda bağlantısı, bilgi akışının da daha hızlı bir süreçte gerçekleştiği temeline dayanmaktadır. Bu yaklaşıma bağlı olarak, yeni bir benzerlik indeksi SHOPI önerilmiştir. Ortak komşuları cezalandırarak bilgi sızıntısını sınırlamaya çalışmışlardır. Ayrıca, daha yüksek dereceli yollar, çekirdek düğüm çifti arasında mevcut olan daha uzun yolları cezalandıran ayırt edici özellikler olarak kullanılır. On iki gerçek dünya ağ veri kümesi (çeşitli alanlardan toplanan) üzerindeki deneysel sonuçlar, SHOPI'nin geleneksel yöntemlerden daha iyi performans gösterdiğini göstermektedir.

Wang ve arkadaşları [43] yaptıkları çalışmada ağ evrimini yönlendiren mekanizmaları ortaya çıkarmanın, karmaşık ağlar alanında süregelen bir zorluk olduğuna değinmişlerdir. Etkili bir araştırma tekniği, bazı mekanizmaları izleyerek kayıp halkaları tahmin etmektir. Bir tahmin edicinin ilkesi, belirli bir ağın mekanizmasıyla tutarlıysa, bu tahmin edicinin daha doğru tahminler sağlaması gerekliliğidir. Bugüne kadar ilgili araştırmaların çoğu sadece yapısal özellikleri dikkate almıştır. Bir düğüm diğer düğümlere bağlantılar oluşturduğunda, etiket bilgisi genellikle önemli bir rol

oyunmaktadır. Çalışmada, verilen bir düğüm  $v$ , aşağıdakilerden oluşan etiket sisteminin homojenliğini ölçmek için entropiyi tanıtılmaktadır.  $v$  etiketleri ve  $v$  'komşuları' etiketlerini ve ardından etkisini yeniden tahmin edin  $v$  diğer düğümlere. Son olarak, etiket sisteminin homojenliğinden yararlanarak ve daha doğru tahminler elde ederek bir bağlantı tahmin algoritması önerilmektedir.

Moradabadi ve arkadaşları [44] yaptıkları çalışmada bir sosyal ağdaki düğümler arasında gelecekteki ilişkilerin ortaya çıkışını tahmin etme problemine değinmişlerdir. Çalışma, ağırlıklı sosyal ağlarda bağlantı tahmini için bir öğrenme otomatı stratejisine odaklanmaktadır. Çalışmada önerilen yöntemde, tahmin edilmesi gereken her test bağlantısı için bir öğrenme otomatı vardır ve her öğrenme otomatı, mevcut ağdaki bağlantıların ağırlığına karşılık gelen bağlantının gerçek ağırlığını öğrenmeye çalışır. Tüm öğrenme otomatları, eylemlerini karşılık gelen bağlantıların ağırlığı olarak yinelemeli olarak seçmektedir. Öğrenme otomata eylemleri seti, eğitim bağlantılarının ağırlığını hesaplamak için kullanılacak ve her öğrenme otomatı, eğitim setinin gerçek ağırlık tahmini üzerindeki etkisine göre ödüllendirilecek veya cezalandırılacaktır. Daha sonra tahmin edilen ağırlıklara dayalı olarak nihai bir tahmin yapılmaktadır. Ortak yazarlık ve e-posta ağları ile yapılan ön bağlantı tahmini deneyleri, ağırlıklar dikkate alındığında tatmin edici sonuçlar vermiştir.

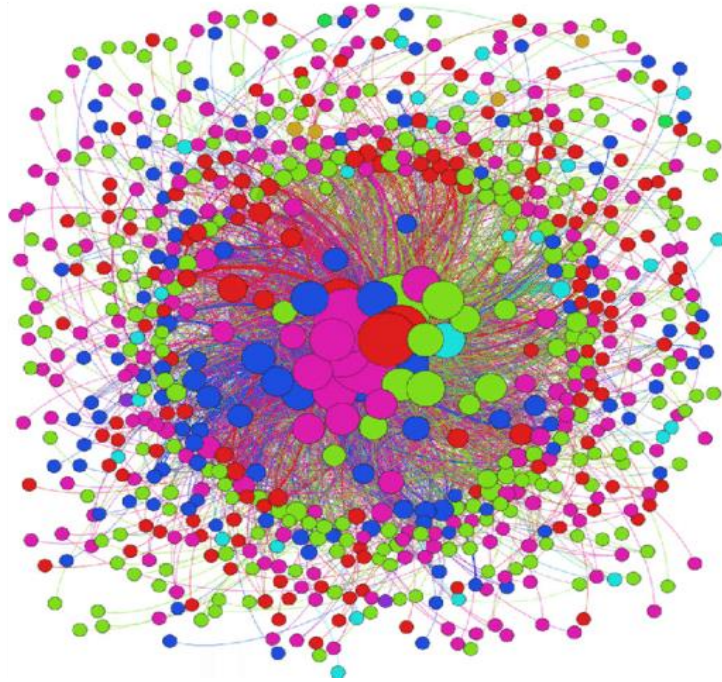
Shakibian ve arkadaşları [45] yaptıkları çalışmalarda heterojen karmaşık ağlardaki bağlantı tahmini ölçümlerinin çoğu, düğümlerin ve yolların önemine daha az dikkat edilirken, düğüm bağlantılarına dayandığına değinmişlerdir. Çalışmada, bağlantı tahmini görevini düzgün bir şekilde gerçekleştirmek için bazı yeni meta-yol tabanlı istatistiksel benzerlik ölçüleri önerilmektedir. Önerilen önlemlerdeki ana fikir, bir meta-yolu izleyen ziyaret edilen düğümler arasında meydana gelen bir dizi birlikte oluşum matrisinde bazı birlikte oluşum olaylarını yönetmektir. Çıkarılan ortak oluşum matrisleri, çeşitli bilgi teorik ölçümlerini belirlemek için enerji, atalet, yerel homojenlik, korelasyon ve korelasyon bilgi ölçüsü açısından analiz edilir. Bağlantı enerjisi, bağlantı ataleti, bağlantı yerel homojenliği, bağlantı korelasyonu ve bağlantı bilgisi korelasyon ölçüsü olarak belirtilen önerilen önlemleri değerlendiririz, standart bir DBLP ağ veri seti kullanarak. AUC puanı ve Kesinlik oranı sonuçları, popüler

meta-yol tabanlı benzerlik ölçütlerine kıyasla önerilen ölçülerin geçerliliğini ve doğruluğunu göstermektedir.

## BÖLÜM 2

### GRAF TEORİSİ

Graf teorisi, ağ kümesi ve belirli noktalar kümesi olarak nitelendirilen bir matematik biliminin alanıdır. Graf düğümlerden ve mevcut düğümlerin etkileşiminin gösterildiği bir ağ yapısı çeşididir. Graf teorisinin temeli 18.yy'a dayanmaktadır. Graf teorisi günlük hayatta elektrik devreleri, şebeke yapıları, yazılımlar, soy ağaçları, bilgisayar ağları ve biyoloji gibi birçok bilim dalında kullanılmaktadır. İsviçreli matematikçi Leonhard Euler'in şehrinde bir problemin çözümünü bulmasıyla graf teorisi ortaya çıkmıştır. Euler'in bu problemin çözümünde yaratmış olduğu başarı sayesinde, günümüzde hemen hemen her alanda birçok problem çözülmüş ve gelişmelerin temellendirildiği bir bakış açısı yaratmıştır. En önemli olan gelişmelerden biri ise sosyal ağların ortaya çıkışıdır. Sosyal ağların ortaya çıkışıyla ilgili Gephi'de tasarlanmış temsili bir görsel aşağıda gösterilmektedir.



Şekil 2.1. Görselleştirilmiş cinsiyet ağ örneği [65].

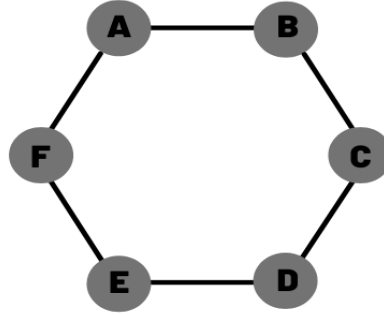


## 2.1.GRAF ÇEŞİTLERİ

Graflar düğümlerin ilişkilerine göre yönlü ve yönsüz graflar olarak isimlendirilebilir. Kenarların ilişkilerine göre ise ağırlıklı veya ağırlıksız olabilirler. İki düğüm arasındaki bağlantı kuvvetini kenar ağırlığı belirlemektedir. Grafları çeşitlendirmek için farklı yollar bulunmaktadır. Bu yollardan bazıları şöyledir; grafların ve kenarların içerikleri dikkate alınarak, ayrıca kenarların sayıları ve yönleri dikkate alınarak gruplandırma yapılabilmektedir.

### 2.1.1. Basit Graf

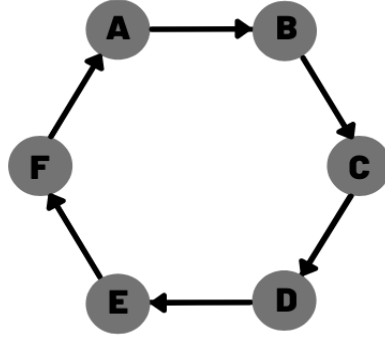
Düğümleri arası ilişkiler yönsüz, döngü içermeyen ve birbirine paralel olmayan graf çeşididir. Basit grafın kenarları simetrik ve herhangi bir yönü yoktur. Mevcut düğümlerin arasında meydana gelen ikili ilişkileri temsil etmektedir.



Şekil 2.2. Basit graf örneği.

### 2.1.2. Yönlü Graf

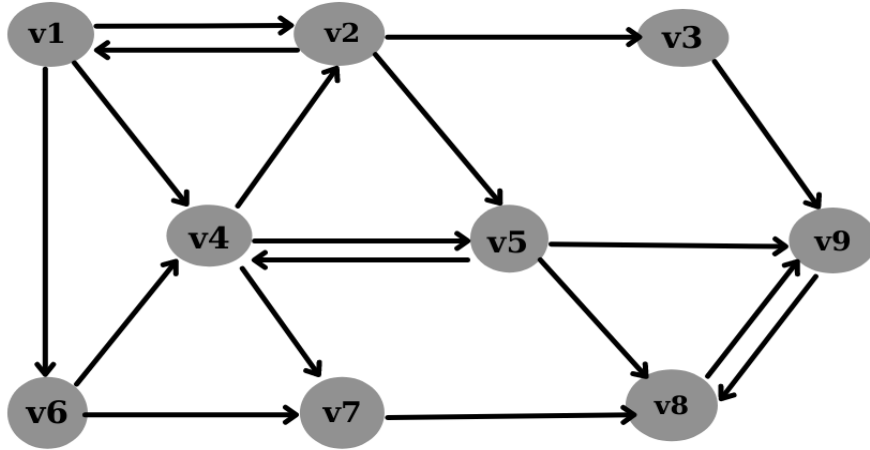
Grafın yönlü olması düğümler arasındaki kenarların yöne sahip olmasıdır. Yönlü graflara birçok örnek verilebilir: sayfalar arası yönlendirme yapan internet ağı, av ile avcı arasındaki ilişkiye dayalı olan gıda ağı, server ve bilgisayarlar arasında kurulan haberleşme ağı şeklinde örnekler çoğaltılabilmektedir.



Şekil 2.3. Yönlü graf örneği.

### 2.1.3.Çoklu(Multipleks) Graflar

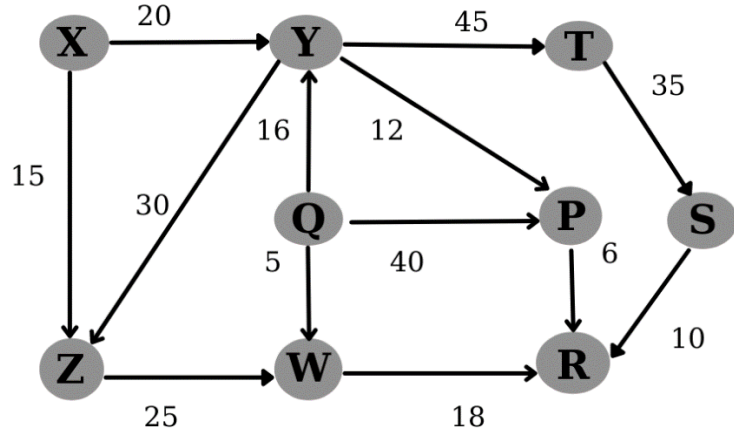
Basit graflar yeterli gelmediğinde kullanılmaktadır. Döngü içermezler, paralel kenarları vardır ve yönsüzdürler. Basit graflar aynı zamanda çoklu graftır ama her çoklu graf bir basit graf değildir.



Şekil 2.4. Multipleks graf örneği.

#### 2.1.4. Ağırlıklı Graflar

Ağırlıklı graflar kenarlarının her biri ağırlıklandırılan, ayrıca numaralandırılan graf türleridir. Her bir kenar ağırlığı düğümlerin arasındaki bağlantının tahmin gücünü belirlemektedir.

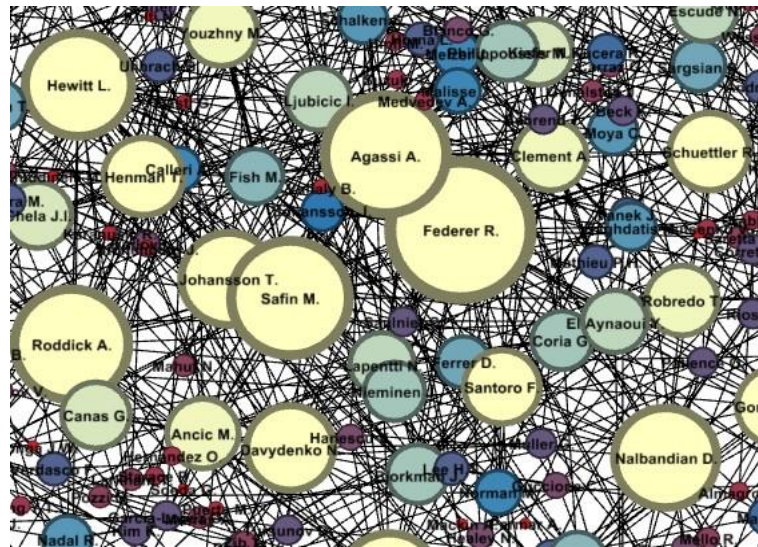


Şekil 2.5. Ağırlıklı graf örneği.

## BÖLÜM 3

### KARMAŞIK AĞLAR

Karmaşık ağlar, herhangi bir sistemin içerisindeki unsurların birbirleriyle olan ilişkilerinin topolojik olarak karmaşık olmasından dolayı karmaşık ağ olarak ifade edilmektedir. Karmaşık sistemler kapsamında kişiler arası sosyal ilişkiler, şirket içerisindeki çalışan ilişkileri, biyolojide hücreler arası etkileşim vb. değerlendirilebilir. Karmaşık ağlardaki düğümler arası ilişkilere örnek olarak; şirket çalışanları arasındaki e-mail kullanımı ve sosyal ağlarda Facebook kullanımı örnek olarak gösterilebilir[61]. Ağlarda karmaşıklığın oluşmasında rol oynayan faktörler; bileşenlerin çeşitlilik göstermesi, büyük ve çok boyutlu olmalarıdır [62]. Ağlarda karmaşıklık seviyesi arttıkça ağ biliminin sunmuş olduğu imkanların ötesinde, yani ağdaki düğümlerin bireysel özelliklerinin dışında ayrı bir ağ doğası oluşmaktadır. Oluşan bu doğa kitle psikolojisi ve birey psikolojisi kadar geniş olabilmektedir. Şekil 2.5’de gösterildiği gibi bir ağı oluşturan unsurlar kenar ve köşeleridir. Varlıkları köşeler temsil edilmekte iken, varlıklar arasındaki ilişkileri ise kenarlar ile temsil edilmektedir.



Şekil 3.1. Tenis turnuvaları ağı [6].

Karmaşık ağlar yapısal olarak incelendiğinde; köşelerin düğümleri, kenarların ise bağlantıları ifade ettiği gözlenebilmektedir. Karmaşık ağların bir başka ifade edilme şekli ise gerçek dünya ağlarının modellenmesidir [46]. İçermiş oldukları köşelerin ve bağlantıların sayısı bakımından bakıldığında karmaşık ağların oldukça büyük yapılar oldukları fark edilmektedir. İletişim teknolojisindeki gelişmeler sayesinde sosyal ağlar yaygınlaşmış böylelikle karmaşık ağların kullanımında artış yaşanmıştır.

### **3.1.SOSYAL AĞLAR**

Belirli bir amaç için oluşturulan internet sitesi içerisinde gündelik hayattaki gibi sembolik işaretler vasıtasıyla insanların birbirleriyle etkileşimde bulunduğu web ortamlarına sosyal ağlar denilmektedir. İlk olarak J.Abundel tarafından 1954 yılında “sosyal ağ” kelimesi kullanılmıştır. Sosyal ağlar kişi ilişkilerini ve video, fotoğraf, ses, konum gibi paylaşımlarını içeren ortamdır. Sosyal ağların günümüzde yaygın olarak kullanılanları Facebook, YouTube, Whatsapp, Facebook Messenger, Instagram, LinkedIn, Twitter şeklinde örneklendirilebilir. Sosyal ağlar sayesinde oldukça fazla miktarda veri elde edilebilir. Sosyal ağlarda her bir kişiyi ya da tüzel kuruluşu birer düğüm temsil etmekte olup, tüm bunlar birbirleriyle ilişkide olan birçok farklı düğümden oluşmaktadırlar.

Sosyal ağ kenar ve düğümler vasıtasıyla birbirine bağlanan kompleks sistemdir. Düğümler arası ilişkileri göstermek amacıyla kenarlar düğümleri birbirine bağlamaktadırlar. Sosyal ağ sistemi doğru oluşturulmuş ise, günlük hayatta karşılaşmış olduğumuz problemleri çözmek için gerekli bilgileri mevcut ağda bulunabilmesi mümkündür.

Sosyal ağlar içerisinde yeni gerçekleşen değişiklikler güncelleme yeteneği sayesinde kaydedilebilmektedir. Meydana gelen değişimler zaman içerisinde, kenar ve düğümler ağa eklenebilmektedirler. Bu nedenle sosyal ağlar dinamik yapılar olarak bilinmektedir.

Kenar ve düğüm sayılarında meydana gelen değişimler, geleceğe yönelik tahmin yapılabilmesine olanak sağlayabilmektedir. Bu nedenle bağlantı tahmin yöntemleri bir ihtiyaç haline gelmiştir. Örneğin sosyal ağlarda bağlantı tahmini, ağdaki mevcut kişilerin aralarında kurdukları ilişkiler baz alınarak, gelecekte ilişkilerinin nasıl olacağını tahmin edilmesi ile temellendirilir. Bağlantı tahmin yöntemleriyle ağdaki mevcut sonuçlara göre yeni sonuçların nasıl olacağını tahmin edilebilir. Örneğin spor ağlarında, müsabakalarda elde edilmiş olan geçmiş sonuçlara göre gelecekteki sonuçlar tahmin edilebilmekte ve doğruluğu test edilebilmektedir [47-48].

### **3.2.SOSYAL AĞ ANALİZİ**

Sosyal ağ yapıları, içerisinde barındırdıkları sosyal ilişkiler hakkında detaylı bilgilere sahiptirler. Sosyal ağlardaki ilişkilerin detaylı olarak analiz edilmesi neticesinde, ağ hakkında tahminler yapılabilmektedir. Bu sayede ağ yapısındaki yönelimler, yorumlar belirlenebilmektedir. Günden güne hız kazanan teknolojik gelişmeler neticesinde sosyal ağların kullanımı hız kazanmıştır. Sosyal ağlarda elde edilen verilerin ölçülebilmesi, anlamlandırılması ve detaylı analizinin gerçekleştirilebilmesi için, ağdaki mevcut ilişkiler sayısallaştırılarak sonuçların çıkarılması maksadıyla bağlantı tahmin yöntemleri kullanılmaktadır. Böylece ağın karakteristik yapısı belirlenmektedir. Sosyal ağ analizi sayesinde anlaşılması zor olan ilişkiler dahi, ağlar modellenerek ve görselleştirilerek gerçekleştirilen detaylı analizler sayesinde anlaşılır kılınabilir.

Sosyal ağ analizi disiplinler arasıdır. Çünkü bilişimden tutun biyoloji, kimya, sosyoloji, ekonomi ve psikoloji gibi birçok bilim dalında yaygın olarak faydalanılmaktadır. Sosyal ağ analizi, ağ içerisindeki öge, ilişki ve bağlantıları baz almaktadır. Mevcut ağdaki temel eleman ögedir. Öğelerden her biri düğüm ile ifade edilebilmektedir. Her bir düğüm arasında ilişkiler mevcuttur, öyle ki bu ilişkiler bağlantılar sayesinde gerçekleşmektedir.

Bağlantılar ise ilişkilerin görsel hale getirilmesidir. İlişkilerde bazı önemli olan parametreler şunlardır: aktiflik, yön, yoğunluk içerik. İki düğüm arasındaki ilişki içerik

sayesinde anlatılabilmektedir. Mevcut iki düğüm arasındaki ilişki yönlü ve yönsüz olarak ikiye ayrılmaktadır. Örneğin Facebook arkadaşlık ilişkileri çift yönlü olması sebebiyle bir yönsüz graftır. Twitter ise öğelerin birbirinin takibini yapabilmesi tek taraflı olduğu için yönlü graflara örnektir [49].

### **3.3.SOSYAL AĞ ANALİZİNDEKİ İŞLEM BASAMAKLARI**

Sosyal ağ analizi için gerekli olan veriler, içerisinde birçok farklı unsur bulunduran geniş ölçekli bilgilerdir. Toplanan veriler kimi zaman devasa boyutlarda olabilmektedir, kimi zaman ise profil bilgilerinden, paylaşımlardan ve oturum hareketlerinden meydana gelebilmektedir. Tüm bu verilerin büyüklüğü ve karmaşıklığı nedeniyle analizlerinin iyi yapılabilmesi için bazı işlem basamakları uygulanmalıdır. Bunlar dört işlem basamağı altında toplanmışlardır [47].

**Analiz için Veri Seçilmesi:** Çevrimiçi ve çevrimdışı ortamlardan analiz için gerekli olan verilerin elde edilmesi aşamasıdır. Ulaşılan veriler veri tabanı yardımıyla saklanır. Verinin elde edilmesi aşamasında gözetilen standart bulunmamaktadır. Bu nedenle verilerin elde edilme şekilleri değişkenlik gösterebilmektedir.

**Analiz için Bilginin Elde Edilmesi ve Ön Değerlendirilmesi:** Ulaşılmış olan veriler ancak analiz edilerek anlamlı hale gelebilmektedir. Bilgiye ulaşılması veri analiziyle meydana gelmektedir. Bilgiye ulaşılmasının ardından anlaşılır hale gelebilmesi amacıyla verinin modellenmesi sağlanır. Analiz yapılabilmesi için ihtiyaç olan filtreler yapılır, analizin hazırlık aşaması böylelikle tamamlanmış olur.

**Analiz için Genelleştirme Yapılması:** Bu işlem basamağında veri analizi sağlanmış olup, gerekli bilgiler çıkarılmıştır. Modelleme ve filtreleme işlemleri gerçekleştirilmiştir. Tümü tamamlandığında spesifik bir model meydana gelmiştir. Meydana gelen modelden genel kurallar elde edilecek şekilde genelleştirilmesi ortak amaçtır.

**Analiz Safhası:** Meydana gelmiş olan genelleştirilmiş model vasıtasıyla toplanan bilgileri anlamlandırabiliriz. Böylece mevcut durum ve olaylardan gelecekte oluşma ihtimali bulunan olasılıkları değerlendirme imkanı bulabilmektedir [34].



## BÖLÜM 4

### BAĞLANTI TAHMİNİ

Günümüzde insanlar arasındaki etkileşimleri modellemenin yaygın bir yolu olarak sosyal ağlar kullanılmaktadır. Ağdaki bir kişinin karşılığı olarak düğüm geldiği gibi, kenarlar arası graflar ise insanlar arasındaki ilişkileri temsil etmektedirler. Graflara sürekli olarak yeni kenar ve düğümler eklenip güncellenebilmesi sebebiyle sosyal ağların dinamik olduğu kanıtlanmaktadır. Herhangi bir sosyal ağı etkileyen birçok parametre bulunması nedeniyle bir sosyal ağın dinamiklerini anlamlandırmak oldukça karmaşıktır.

Herhangi iki düğüm arasındaki ilişkileri anlamlandırabilmek adına birtakım sorulara cevap bulunması gerekmektedir. Bunlardan bazıları şu şekildedir: Zamanla ilişki modelleri ne şekilde değişime uğramaktadır? İki düğüm arasında oluşan ilişkiye ağda mevcut diğer düğümlerin etkisi nedir? Düğümler arası bağlantılara yön veren etmenler nelerdir? Anlamlandırmadaki genel amaç bulunulan durumda mevcut iki durum arasında şu an ilişkinin olmadığını bilip, gelecekteki ilişki olasılığını tahmin edebilmektir. Buna yaygın olarak bağlantı tahmin problemi denilmektedir.

#### 4.1. PROBLEM OLARAK BAĞLANTI TAHMİNİ

Karmaşık bir ağ yapısı içerisinde  $t_0$  zamanındaki ağ yapısı örnek alınıp,  $t_0$  anından  $t_1$  anına kadarki zaman zarfında ağda oluşabilecek olası yeni bağlantıların tahin edilmesi işlemine bağlantı tahmini denilmektedir[50]. Yaklaşım olarak bağlantı tahmini problem olarak ele alındığında, ağdaki genel yapının analizi ve zamana bağlı bağlantı değişimlerinin analizine dayalı olarak yani, mevcut durumdan ağda oluşabilecek yeni bağlantılarının tahmin edilmesi problemine çözüm sağlamaktadır. Bağlantı tahmin sürecindeki başarının artırılabilmesi için ise mevcut ağdan ne kadar çok bilgi elde edilirse o kadar çok bağlantı tahmin neticeleri doğru olmaktadır.

Bir karmaşık ağda  $t_0$  zamanında anlık olarak, bağlantılar  $G = (V, E)$  formül yapısıyla ifade edilirken,  $(E)$  ise  $t_0$  ve  $t_1$  anları arasında bağlantı kümesinde meydana gelebilecek olası değişimleri ifade etmek için kullanılmaktadır.  $t_0$  ve  $t_1$  anları arasında aralarında bağlantı bulunmayan düğümler birbiriyle bağlantı kurabilirler. Fakat düğümler arasında bağlantı kurulmama ihtimalide bulunmaktadır. Yaklaşımsal olarak bağlantı tahmin problemi tam olarak bu noktadaki belirsizlik nedeniyle çözüm konusunda zorlanmaktadır. Bağlantı tahmin probleminin çözüme kavuşabilmesi için  $t_0$  anının çok iyi analiz edilmesi gerekmektedir. Yani mevcut ağ hakkında ne kadar çok bilgi elde edilirse o kadar iyi olmaktadır. Böylece ağdaki düğümler birbirinden ayırt edilebilmekte ve bağlantı tahmin başarısı da bir o kadar artmaktadır.

## **4.2. BAĞLANTI TAHMİN YÖNTEMLERİ**

Bağlantı tahmin problemini çözmek adına günümüzde birçok yöntem geliştirilmiştir. Tüm bu yöntemler karmaşık ağların analizinde çok önemli yer tutmaktadır. Geçmişten günümüze geliştirilmiş olan bağlantı tahmin yöntemleri incelendiğinde, geliştirilmiş olan ilk yöntemleri veri madenciliği yöntemine dayandığı belirlenmektedir [51]. Sonrasında geliştirilmiş olan yöntemlere bakıldığında ise mevcut ağ yapısındaki bilgiler kullanılarak daha çok graf tabanlı modellere rastlanmıştır [52]. Literatürde yerel ağ yapısıyla yapılmış olan çalışmalara da rastlanılmaktadır [53]. Ancak geliştirilen tüm yöntemlere bakıldığında daha çok, mevcut ağın genel özellikleriyle ve düğümlerden elde edilmiş olan öznitelik bilgilerinin kullanılmasıyla oluşan başarılı sonuçlar veren analiz yöntemleri geniş ölçekte kabul görmektedir [54]. Bağlantı tahmin yöntemleri, düğüm öznitelikleri ve ağın yapısal özelliklerinin çeşidine göre bölümlere ayrılıp değerlendirilmektedir.

### **4.2.1 Benzerlik Tabanlı Bağlantı Tahmin Yöntemleri**

Geliştirilmiş olan yöntemlerin en eskilerinden olan benzerlik tabanlı bağlantı tahmin yöntemleri, mevcut ağın yapısını ve özelliklerini göz önünde bulundurarak en sağlıklı yaklaşımların, analizlerin oluşmasını sağlayan en basit yöntemdir. Bu yöntemlerde bağlantı tahmini gerçekleştirilirken öncelikli olarak mevcut ağın yapısal özellikleri ve

öznitelikleri göz önünde bulundurularak var olan düğümler arasındaki benzerlik oranı hesaplanarak, gelecekte oluşabilecek yeni olası bağlantıların belirlenmesi için çalışmalarda bulunulur.

Bu yöntemler yol tabanlı yöntemler, yerel benzerlik tabanlı yöntemler ve yarı yerel benzerlik tabanlı yöntemler olarak kısımlara ayrılırlar. Benzerlik tabanlı tahmin yöntemlerinde bazı temel ölçütler aşağıdaki formüller sayesinde ifade edilebilmektedirler:

- Herhangi bir karmaşık ağ yapısında  $x$  düğümünün komşularının kümesi  $\Gamma(x)$  ile ifade edilmektedir.
- Herhangi bir karmaşık ağ yapısında  $x$  ve  $y$  düğümlerinin benzerliklerinin rakamsal olarak karşılığı ise  $S_{xy}$  ile ifade edilmektedir.
- Herhangi bir karmaşık ağ yapısında  $x$  düğümünün komşularının sayısı yani derecesi  $|\Gamma(x)|$  ile ifade edilebilmektedir.

#### **4.2.1.1. Yerel Benzerlik Tabanlı Bağlantı Tahmin Yöntemleri**

Yerel benzerlik tabanlı bağlantı tahmin yöntemleri oluşturulurken, mevcut karmaşık ağ yapısındaki düğüm çiftlerinin ortak komşularından yola çıkılarak gelecekte olası bağlantı durumlarını tahmin etmeye dayanmaktadır.

Daha kıvıllı sonuçlar alabilmek için mevcut ağ yapısının özelliklerini tümünden ele almak yerine, sadece ortak komşulardan yola çıkılarak elde edilen bilgiler doğrultusunda bağlantı tahmin çalışması gerçekleştirilebilmektedir. Sadece komşuluk ilişkilerine dayalı bilgilerin kullanımı analiz başarısına negatif etkide bulunmaktadır.

#### **Ortak Komşular Yöntemi (Common Neighbor)**

Ortak komşular(CN) bağlantı tahmini amacıyla geliştirilen yöntemlerin en basitlerinden biri olmasına karşı tahmin başarısı oldukça yüksektir. Bu nedenle en çok kullanılan yöntemlerden bir tanesidir. Ortak komşular yönteminin bağlantı

tahminindeki bakış açısı şöyledir; iki düğümün ortak komşuları ne kadar fazla ise bağlantı başarısı o kadar yüksek olmaktadır [54].

$$S_{xy} = |\Gamma(x) \cap \Gamma(y)| \quad (4.1)$$

Ortak komşu yöntemi ağırlıklı ağlar için geliştirilmiş bağlantı tahmin yöntemine göre iki düğümün birbirleri ile etkileşime girme olasılığı, ortak komşularının düğümlerle aralarında ağırlıklarının toplanılmasına bağlı olarak hesabı yapılmaktadır. Ortak komşu yönteminin bakış açısı şöyledir; düğümlerin ağırlıkları toplamda ne kadar çok ise gelecekteki düğümlerin birbirleriyle etkileşime girme olasılıkları o kadar fazla olmaktadır [55].

$$S_{xy}^{WCN} = \sum_{z \in |\Gamma(x) \cap \Gamma(y)|} w(x, z) + w(y, z) \quad (4.2)$$

### Jaccard İndeksi (Jaccard Index)

Daha çok veri madenciliğinde kullanılan jaccard indeksi, benzerlik tabanlı bağlantı tahmininde uygulanan başarı oranı yüksek yöntemlerinden biridir.

Jaccard İndeksi(JI) yöntemi benzerlik oranını hesaplarken, ortak komşuların sayısını kullanmanın yanı sıra tüm komşuların toplam sayısını da göz önünde bulundurarak iki düğümün gelecekte etkileşime girebilme olasılığını hesaplamaktadır [56].

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (4.3)$$

Jaccard İndeksin ağırlıklı ağ yapısı için geliştirilen yöntem ile iki düğümün gelecekte etkileşime girebilme olasılığı, ortak komşularının düğümler ile aralarındaki ağırlıkları toplamının bütün komşularının ağırlıkları toplamına bölünmesiyle elde edilmektedir [57].

$$S_{xy} = \sum_{z \in |\Gamma(x) \cap \Gamma(y)|} \frac{w(x, z) + w(y, z)}{\sum_{a \in \Gamma(x)} w(a, x) + \sum_{a \in \Gamma(y)} w(b, y)} \quad (4.4)$$

### Sorenson İndeksi (Sorenson Index)

Sorenson İndeksi yöntemi benzerlik oranını hesaplarken, ortak komşuların sayısını kullanmanın yanı sıra tüm komşuların derecelerini de göz önünde bulundurarak iki düğümün gelecekte etkileşime girebilme olasılığını hesaplamaktadır. Dereceleri düşük olan düğümlerin gelecekte oluşabilecek bağlantılarının yüksek olduğunu göstermeyi hedeflemektedir.

$$S_{xy} = \frac{2|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x)| + |\Gamma(y)|} \quad (4.5)$$

### Tercihli Bağlantı İndeksi (Preferential Attachment)

Ölçüsüz olarak büyüyen ağlarda tercih edilmekte olan tercihli bağlantı indeksi, mevcut ağa yeni katılan herhangi bir düğümün yüksek dereceli düğümler ile etkileşime girme olasılığının daha fazla olduğunu öne sürmektedir. Yani yaklaşıma göre mevcut düğüm derecesi ne kadar çok olursa gelecekteki bağlantıları artırma olasılığı o kadar fazla olduğu görüşünü savunulmaktadır. Tercihli bağlanma indeksine (PAI) göre mevcut ağda düğümlerin komşu sayısı ne kadar fazla ise herhangi iki düğüm arasındaki bağlantı oluşma olasılığı da o kadar çok yüksektir [57].

$$S_{xy} = \Gamma(x) * \Gamma(y) \quad (4.7)$$

Tercihli bağlanma indeksinin ağırlıklı ağlarda düğümlerin komşuların ağırlıklı toplamları ve komşu sayıları işleme tabi tutulmaktadır. Mevcut düğümlerin komşularıyla aralarındaki ağırlıklarının kuvveti ve komşu sayıları başka düğümlere bağlanma olasılığını etkilemektedir.

$$S_{xy} = \sum_{a \in \Gamma(x)} w(a, x) * \sum_{b \in \Gamma(x)} w(b, y) \quad (4.8)$$

### Adamic Adar İndeksi (Adamic-Adar Index )

Adamic adar indeksinin(AAI) bakış açısı, mevcut ağda bağlantı sayısı düşük olan komşulara önem verilerek bağlantı olasılığının hesaplanmasında dayanmaktadır. Örneğin ortak komşuları  $\Gamma(x)$  ve  $\Gamma(y)$  olan  $z$  düğümünün kaç düğüm ile bağlantısının olduğunu bulmak için  $z$  düğümünün derecesine, bağlantı olasılığı hesabı yapılırken önem verilip işleme alınmalıdır [58].

$$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(|\Gamma(z)|)} \quad (4.9)$$

Ağırlıklı ağlarda ise Adamic Adar ölçütüyle düğümlerin birbirine bağlanma ihtimali aşağıdaki formül yardımıyla hesaplanmaktadır.

$$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{w(x,z) + w(y,z)}{\log \log(1 + \sum_{c \in \Gamma(z)} w(z,c))} \quad (4.10)$$

### Kaynak Paylaştırma İndeksi (Resource Allocation Index )

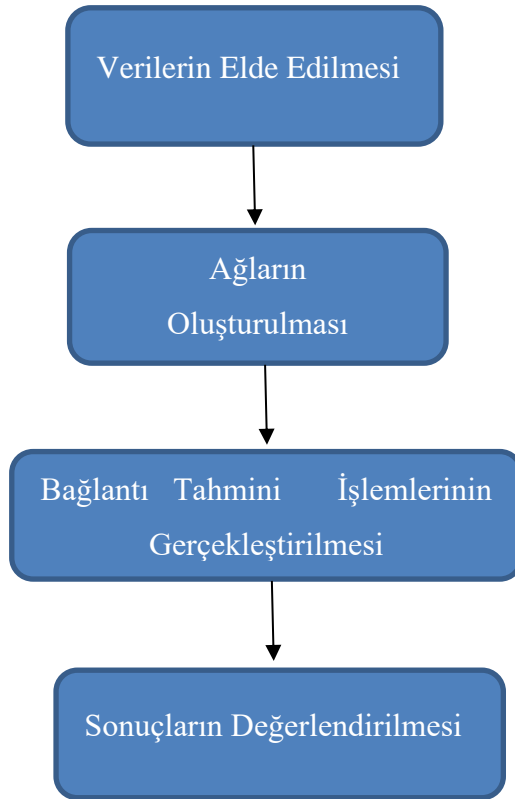
Kaynak paylaşırma (RAI) indeksi birbirleriyle direkt olarak bağlantısı olmayan düğüm çiftlerinin aralarındaki bağlantının ne kadar güçlü olduğunu ölçmekte olup karmaşık ağlarda yaygın olarak kullanılmaktadır. Bir başka deyişle doğrudan birbirleriyle bağlantıları olmadığı halde ortak komşular üzerinden iletişimde olan düğüm çiftlerinin benzerliklerini hesaplamak maksadıyla kullanılmaktadır. Bu benzerlik oranının hesabı ise birbirlerinden almış oldukları kaynaklar doğrultusunda yapılmaktadır [59]. Aşağıda bulunan formülde  $k_z$ ,  $\Gamma(x)$  ve  $\Gamma(y)$  'nin ortak komşusu olan  $z$ 'nin komşularının sayısını ifade etmektedir.

$$S_{xy} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z} \quad (4.11)$$

## BÖLÜM 5

### DENEYSEL ÇALIŞMA

Bağlantı tahmini uygulamalarının genellikle sosyal ağlar ve etkileşim ağları üzerinde yoğunluk kazanması bu yöntemlerin farklı dinamik ağlardaki performansının yeterince değerlendirilmesinin önüne geçmektedir. Bu nedenle bağlantı tahmini yöntemlerinin spor ağı[5,6] gibi farklı branşlardan oluşturulmuş ağlardaki başarısını değerlendirmek bu çalışmanın temelini oluşturmaktadır. Ayrıca spor müsabakalarında sporcuların ya da takımların gelecekte birbiriyle karşılaşabilme olasılıklarının hesaplanmasındaki zorluklara karşı komşuluk tabanlı bağlantı tahmini yöntemlerinin kullanılabilirliğinin değerlendirilmesi çalışmanın bir başka amacı olmuştur. Deneysel çalışmanın genel olarak adımları Şekil 1’de gösterilmektedir.



Şekil 5.1. Deneysel çalışmanın uygulama adımları.

## 5.1.VERİ SETİ

Gerçekleştirilen deneysel çalışmada, Dünya Snooker Şampiyonası[25], Avustralya Açık Tenis Turnuvaları[26] ve UEFA Avrupa Ligi[27] karşılaşmalarından ağlar oluşturulmuştur. Çalışmada bu veri setlerinden oluşturulan ağların kullanılma nedeni ağ teorisine uygun olarak karmaşık ağ özelliklerini gösteren gerçek dünya ağları olmalarıdır.

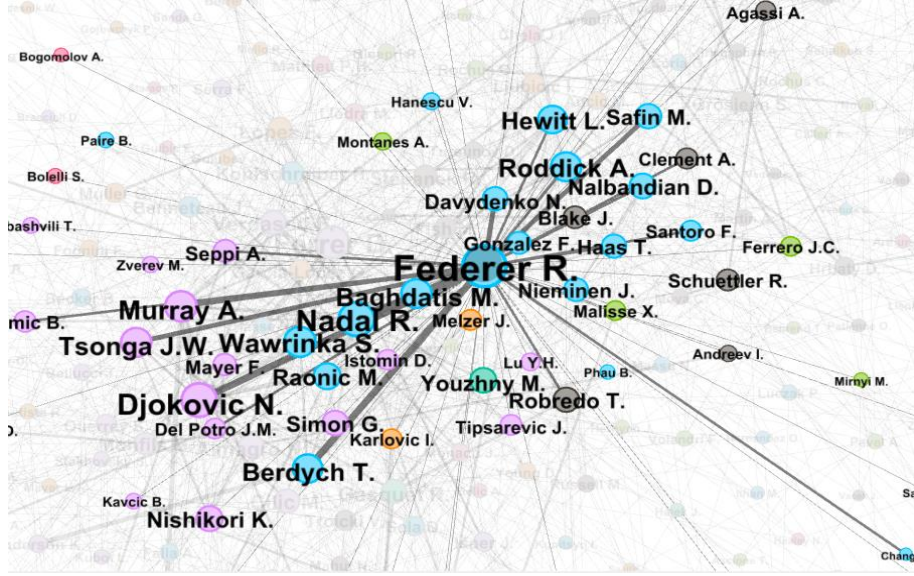
Dünya Snooker Şampiyonası veri seti [60] Şekil 5.2 gösterilmiştir.

tournament_id	stage	player1_name	player2_name	score1	score2
772	Final	Alex Higgins	Ray Reardon	18	15
772	Semi-final	Alex Higgins	Jimmy White	16	15
772	Semi-final	Ray Reardon	Eddie Charlton	16	11
772	Quarter-final	Eddie Charlton	Tony Knowles	13	11
772	Quarter-final	Alex Higgins	Willie Thorne	13	10
772	Quarter-final	Ray Reardon	Silvino Francisco	13	8
772	Quarter-final	Jimmy White	Kirk Stevens	13	9
772	Last 16	Eddie Charlton	Bill Werbeniuk	13	5
772	Last 16	Silvino Francisco	Dean Reynolds	13	8
772	Last 16	Alex Higgins	Doug Mountjoy	13	12
772	Last 16	Tony Knowles	Graham Miles	13	7
772	Last 16	Ray Reardon	John Virgo	13	8
772	Last 16	Kirk Stevens	Patsy Fagan	13	7
772	Last 16	Willie Thorne	John Spencer	13	5
772	Last 16	Jimmy White	Perrie Mans	13	6
772	Last 32	Eddie Charlton	Cliff Wilson	10	5
772	Last 32	Patsy Fagan	David Taylor	10	9
772	Last 32	Silvino Francisco	Dennis Taylor	10	7
772	Last 32	Alex Higgins	Jim Meadowcroft	10	5
772	Last 32	Tony Knowles	Steve Davis	10	1
772	Last 32	Perrie Mans	Tony Meo	10	8
772	Last 32	Graham Miles	Dave Martin	10	5
772	Last 32	Doug Mountjoy	Rex Williams	10	3

Şekil 5.2. Snooker Dünya şampiyonası müsabakaları veri seti.

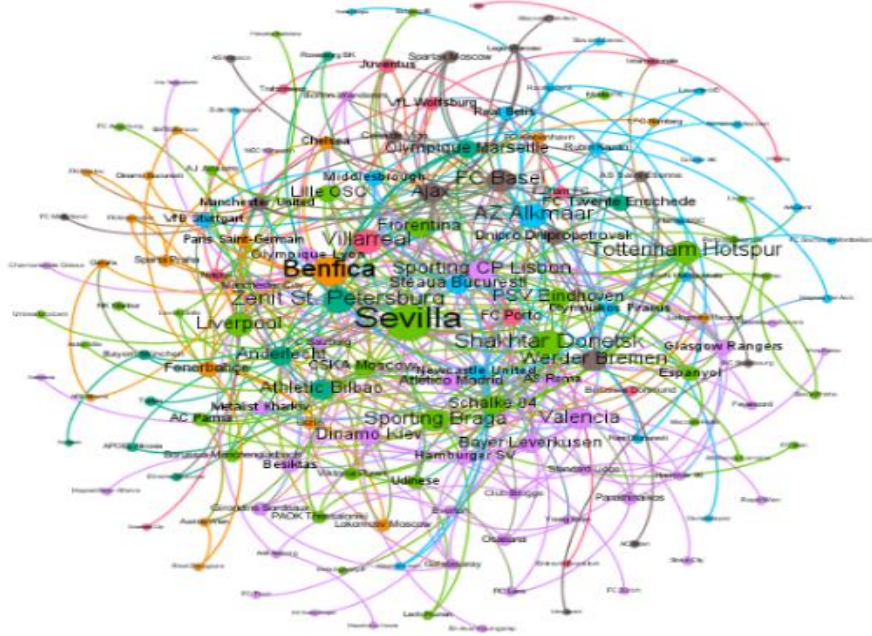
Avusturalya Açık Tenis Şampiyonası müsabakaları ağı [63] müsabakalarından oluşan karmaşık ağ Şekil 5.3 gösterilmiştir.





Şekil 5.3. Avusturalya Açık Tenis Şampiyonası müsabakaları ağı [63].

UEFA Avrupa Ligi Ağları [64] müsabakalarından oluşan karmaşık ağ Şekil 5.4 gösterilmiştir.



Şekil 5.4. UEFA Avrupa Ligi ağları [64].

Çizelge 1'de gösterildiği gibi Snooker ve tenis ağlarında düğümler sporculardan, düğümler arasındaki bağlantılar ise sporcular arasındaki müsabakalardan ve UEFA

Avrupa Ligi ağlarında ise düğümler takımlardan, düğümler arasındaki bağlantılar ise takımlar arasındaki karşılaşmalardan oluşturulmuştur.

Çizelge 5.1. Dünya şampiyonası ağı için düğümler ve bağlantılar.

	<b>Snooker Dünya Şampiyonası</b>	<b>Avustralya Açık Tenis Şampiyonası</b>	<b>Uefa Avrupa Ligi</b>
<b>Nodes</b>	615	553	154
<b>Edges</b>	3033	2156	388

## **5.2.ZAMAN PERİYODUNA BAĞLI OLARAK AĞLARIN OLUŞTURULMASI**

Deneysel çalışmada, elde edilen veri setlerinden 4 farklı zaman periyodunda ağırlıksız ve yönsüz ağlar oluşturulmuştur.

Çizelge 2’de gösterildiği gibi oluşturulan ağlarda eğitim için zaman periyotları farklı aralıklarda seçilerek zaman periyodunun eğitim ağındaki önemi araştırılmıştır. Bununla birlikte eğitim ağlarının başarısını ölçmek için eğitim ağını oluşturulan zaman periyodunun devamındaki 4 yıllık periyot test ağı olarak kullanılmıştır.

Eğitim ve test ağları oluşturulduktan sonra Bölüm 4’de anlatılan geleneksel komşuluk tabanlı bağlantı tahmini yöntemlerinden Ortak Komşular, Jaccard İndeks, Tercihli Atanma İndeksi, Adamic-Adar İndeks, Kaynak Paylaştırma İndeksi ve Sorenson İndeks yöntemleri uygulanmıştır.

Çizelge 5.2. Zaman periyoduna bağlı olarak oluşturulan eğitim ve test ağları.

<b>Eğitim Ağı Zaman Periyodu</b>	<b>Test Ağı Zaman Periyodu</b>
2004-2007	2008,2009,2010,2011
2007-2010	2011,2012,2013,2014
2007-2013	2014,2015,2016,2017
2004-2013	2014,2015,2016,2017

### **5.3.DENEY SONUÇLARI**

#### **5.3.1. 2004-2011 Arası Oluşturulan Ağlar**

2004-2011 yılları arasında veriler için 2004-2007 arası zaman periyodundaki ağ eğitim için kullanılırken, 2008, 2009, 2010 ve 2011 yılları ise test ağları olarak kullanılmıştır. Çizelge 3-5' e bakıldığında AUC metriğinden elde edilen sonuçlar, bağlantı tahmini yöntemlerinin spor ağlarının tamamı için istenilen düzeyde başarılı olmadığı görülmektedir.

Bunun temel nedeni olarak eğitim ağı için kullanılan zaman periyodunun kısa olmasından kaynaklandığı görülmektedir. Ancak, eğitim ağından kaynaklanan negatif etkiye rağmen eğitim için oluşturulan ağların zaman periyodu arttıkça bağlantı tahmini yöntemlerinin başarısı artmıştır.

Çizelge 5.3. 2004-2011 aralığında oluşturulan snooker ağında bağlantı tahmini AUC sonuçları.

	2008	2009	2010	2011
<b>Ortak Komşu (CN)</b>	0,542	0,780	0,881	0,899
<b>Jaccard Index</b>	0,697	0,947	0,968	0,987
<b>Adamic Adar (AA)</b>	0,560	0,876	0,946	0,965
<b>Tercihli Bağlantı İndeksi</b>	0,776	0,838	0,854	0,871
<b>Sorenson Index</b>	0,465	0,696	0,774	0,790
<b>Kaynak Tahsisli Index(RA)</b>	0,576	0,614	0,771	0,787

Çizelge 5.4. 2004-2011 aralığında oluşturulan tenis ağında bağlantı tahmini AUC sonuçları.

	2008	2009	2010	2011
<b>Ortak Komşu (CN)</b>	0,516	0,742	0,839	0,856
<b>Jaccard Index</b>	0,664	0,902	0,922	0,940
<b>Adamic Adar (AA)</b>	0,533	0,834	0,901	0,919
<b>Tercihli Bağlantı İndeksi</b>	0,739	0,799	0,814	0,830
<b>Sorenson index</b>	0,443	0,663	0,737	0,753
<b>Kaynak Tahsisli Index(RA)</b>	0,549	0,584	0,734	0,749

Çizelge 5.5. 2004-2011 aralığında oluşturulan UEFA ağında bağlantı tahmini AUC sonuçları.

	2008	2009	2010	2011
<b>Ortak Komşu (CN)</b>	0,304	0,437	0,494	0,504
<b>Jaccard Index</b>	0,391	0,531	0,615	0,627
<b>Adamic Adar (AA)</b>	0,314	0,491	0,53	0,541
<b>Tercihli Bağlantı İndeksi</b>	0,435	0,47	0,603	0,615
<b>Sorenson index</b>	0,261	0,39	0,434	0,443
<b>Kaynak Tahsisli Index(RA)</b>	0,323	0,344	0,432	0,441

Bağlantı tahmin yöntemlerinin uygulandığı ağlarda branş düzeyindeki tahmin başarılarına bakıldığında, Çizelge 3’de görüldüğü gibi Snooker müsabakalarından meydana gelen ağların diğer spor ağlarından daha başarılı tahmin sonuçlar elde edildiği görülmektedir. En düşük tahmin başarısı ise UEFA Avrupa Ligi karşılaşmalarından oluşturulan ağlarda olduğu görülmektedir.

Bunun nedeni ise eğitim ve test ağlarındaki düğüm ve bağlantı sayılarının tahmin başarısına olan etkisinden kaynaklanmaktadır. Eğitim ve test ağı için kullanılan düğüm ve bağlantı sayısının fazlalığı yani ağ için için yapılacak analizde kullanılan verilerin niteliği bağlantı tahminini doğru oranda arttırmaktadır.

Bağlantı tahmini yöntemlerinin 2004-2011 yılları arasında oluşturulan ağlardaki başarıları karşılaştırıldığında, Çizelge 3'deki Snooker ağlarında 2008 yılı tahmininde Tercihli Bağlantı İndeksi, 2009, 2010 ve 2011 yılı tahminlerinde Jaccard İndeks, Çizelge 4'teki Tenis ağlarında 2008 yılı tahmininde Tercihli Bağlantı İndeksi, 2009, 2010 ve 2011 yılı tahminlerinde Jaccard İndeks, Çizelge 5'deki UEFA Avrupa Ligi tahminlerinde 2008 yılı tahmininde Tercihli Bağlantı İndeksi, 2009, 2010 ve 2011 yılları bağlantı tahminlerinde Jaccard İndeks yöntemlerinin başarılı olduğu görülmektedir.

Tüm ağlar için genel olarak bağlantı tahmin yöntemlerinin başarılarına bakıldığında 2004-2011 yılları arasındaki ağlarda Tercihli Bağlantı İndeksi ve Jaccard İndeks yöntemlerinin diğer yöntemlere göre başarılı oldukları görülmektedir. Tercihli Bağlantı İndeksi yönteminin test ağının dar zaman periyodundan oluşturulduğu ağlarda diğer yöntemlerden daha başarılı olduğu, test ağının zaman periyodu arttıkça Jaccard İndeks yönteminin diğer yöntemlerden daha başarılı olduğu görülmektedir. Bağlantı tahmininde başarıların birbirinden farklılık göstermesi, uygulanan bağlantı tahmini yöntemlerinin ağlardaki düğümlerin birbirine bağlanma şekli, düğüm sayısı, bağlantı sayısı ve hub düğümlerin ağdaki güçlerini farklı şekilde analiz etmelerinden kaynaklanmaktadır.

### **5.3.2. 2007-2014 Arası Oluşturulan Ağlar**

2007-2014 yılları arasında veriler için 2007-2010 arası zaman periyodundaki ağ eğitim için kullanılırken, 2011, 2012, 2013 ve 2014 yılları ise test ağları olarak kullanılmıştır. Çizelge 6-8' de AUC metriğinden elde edilen sonuçlara bakıldığında, Çizelge 3-5'te görüldüğü gibi 2004-2011 yılları arasında oluşturulan ağlardaki sonuçlara benzer sonuçların elde edildiği görülmektedir. Bunun nedeni ise 2004-2011 ve 2007-2014

yıllarındaki kullanılan veri setlerindeki verilerin benzer özellikler taşımasıdır. Bu durum eğitim ve test ağları için oluşturulan ağlarda zaman periyodunun seçiminin doğru belirlenmesinin önemini göstermektedir Eğitim ve test ağlarının oluşturulduğu zaman periyotlarının benzer uzunlukta olması bağlantı tahmini yöntemlerinin benzer sonuçlar üreteceği sonucu görülmektedir.

Çizelge 5.6. 2007-2014 aralığında oluşturulan snooker ağında bağlantı tahmini AUC sonuçları.

	2011	2012	2013	2014
<b>Ortak Komşu (CN)</b>	0,543	0,779	0,881	0,899
<b>Jaccard Index</b>	0,597	0,811	0,939	0,958
<b>Adamic Adar (AA)</b>	0,527	0,852	0,921	0,939
<b>Tercihli Bağlantı İndeksi</b>	0,663	0,717	0,922	0,940
<b>Sorenson index</b>	0,547	0,820	0,911	0,930
<b>Kaynak Tahsisli Index(RA)</b>	0,646	0,722	0,907	0,926

Çizelge 5.7. 2007-2014 aralığında oluşturulan tenis ağında bağlantı tahmini AUC sonuçları.

	2011	2012	2013	2014
<b>Ortak Komşu (CN)</b>	0,540	0,775	0,877	0,894
<b>Jaccard Index</b>	0,574	0,780	0,903	0,921
<b>Adamic Adar (AA)</b>	0,507	0,819	0,885	0,903
<b>Tercihli Bağlantı İndeksi</b>	0,638	0,690	0,887	0,903
<b>Sorenson index</b>	0,526	0,788	0,876	0,894
<b>Kaynak Tahsisli Index(RA)</b>	0,622	0,694	0,872	0,891

Çizelge 5.8. 2007-2014 aralığında oluşturulan UEFA ağında bağlantı tahmini AUC sonuçları.

	2011	2012	2013	2014
<b>Ortak Komşu (CN)</b>	0,373	0,535	0,605	0,617
<b>Jaccard Index</b>	0,479	0,650	0,753	0,768
<b>Adamic Adar (AA)</b>	0,376	0,607	0,656	0,669
<b>Tercihli Bağlantı İndeksi</b>	0,532	0,575	0,738	0,753
<b>Sorenson index</b>	0,319	0,478	0,531	0,542
<b>Kaynak Tahsisli Index(RA)</b>	0,377	0,421	0,529	0,540

Bağlantı tahmini yöntemlerinin 2007-2014 yılları arasında oluşturulan ağlardaki başarıları karşılaştırıldığında, Çizelge 6'deki Snooker ağlarında 2011 yılı tahmininde Tercihli Bağlantı İndeksi, 2012 tahmininde Adamic Adar, 2013 ve 2014 yılı tahminlerinde Jaccard İndeks, Çizelge 7'teki Tennis ağlarında 2011 yılı tahmininde Tercihli Bağlantı İndeksi, 2012 yılı tahmininde Adamic Adar, 2013 ve 2014 yılı tahminlerinde Jaccard İndeks, Çizelge 8'deki UEFA Avrupa Ligi tahminlerinde 2011 yılı tahmininde Tercihli Bağlantı İndeksi, 2012 yılı tahmininde Adamic Adar, 2013 ve 2014 yılları bağlantı tahminlerinde Jaccard İndeks yöntemlerinin başarılı olduğu görülmektedir.

Tüm ağlar için genel olarak bağlantı tahmin yöntemlerinin başarılarına bakıldığında 2007-2014 yılları arasındaki ağlarda Tercihli Bağlantı İndeksi, Adamic Adar ve Jaccard İndeks ve yöntemlerinin diğer yöntemlere göre başarılı oldukları görülmektedir. Bu zaman periyodunda Adamic Adar yönteminin 2012 ağı için tüm spor ağlarında diğer yöntemlerden başarılı sonuçlar verdiği görülmektedir. Bunun nedeni ise Adamic Adar'ın düğüm sayısı fazla bağlantı sayısı zayıf ağlarda bağlantı tahmin gücünün daha yüksek olmasından kaynaklanmaktadır. Bununla birlikte, Tercihli Bağlantı İndeksi ve Jaccard İndeks yöntemlerinin 2004-2011 aralığında ağlarda olduğu gibi benzer şekilde başarılı olduğu görülmektedir. Ayrıca uygulanan tüm yöntemlerin genel olarak başarıların 2004-2011 aralığındaki tahminlerden daha başarılı olmaları, 2007-2014 aralığında oluşturulan ağlardaki düğüm ve kenar sayılarının daha fazla olmasından ve hub düğümlerin davranışlarındaki değişimlerden kaynaklanmaktadır.

### **5.3.3. 2007-2017 Arası Oluşturulan Ağlar**

2007-2017 yılları arasında veriler için 2007-2013 arası zaman periyodundaki ağ eğitim için kullanılırken, 2014, 2015, 2016 ve 2017 yılları ise test ağları olarak kullanılmıştır.

Çizelge 9-11' de AUC metriğinden elde edilen sonuçlara bakıldığında, Çizelge 3-8 gibi 2004-2011 ve 2007-2014 yılları arasında oluşturulan ağlardaki tahminlere göre daha başarılı tahminler yapıldığı görülmektedir. Bunun nedeni ise eğitim ağı için

zaman periyodunun aralığındaki artış test ağının zaman periyodu kısa olsa da daha başarılı tahminlerin yapılmasında etkili olmuştur. Test ağındaki zaman periyodunun geniş alınmasıyla birlikte özellikle 2016 ve 2017 yıllarında kullanılan tüm bağlantı yöntemlerinin yüksek bir tahmin başarısı gösterdiği görülmektedir. Bu durum hem eğitim hem de test için oluşturulan ağların zaman periyodundaki artışın tahmin sonuçlarını da olumlu yönde etkilediğini sonucunu göstermektedir.

Komşuluk tabanlı bağlantı tahmini yöntemlerinin genel yaklaşımı düğümlerin topolojik olarak komşularıyla olan durumlarını analiz ederek düğümler arasındaki benzerliği hesaplaması şeklinde olmasından dolayı, ağdaki topolojik bilgilerin fazla olması daha başarılı tahminler yapabilmesine doğrudan etki etmektedir.

Çizelge 5.9. 2007-2017 aralığında oluşturulan snooker ağında bağlantı tahmini AUC sonuçları.

	2014	2015	2016	2017
<b>Ortak Komşu (CN)</b>	0,547	0,787	0,889	0,907
<b>Jaccard Index</b>	0,704	0,956	0,972	0,992
<b>Adamic Adar (AA)</b>	0,552	0,893	0,963	0,983
<b>Tercihli Bağlantı İndeksi</b>	0,782	0,845	0,855	0,873
<b>Sorenson index</b>	0,469	0,702	0,782	0,798
<b>Kaynak Tahsisli Index(RA)</b>	0,547	0,619	0,778	0,794

Çizelge 5.10. 2007-2017 aralığında oluşturulan tenis ağında bağlantı tahmini AUC sonuçları.

	2014	2015	2016	2017
<b>Ortak Komşu (CN)</b>	0,511	0,735	0,831	0,848
<b>Jaccard Index</b>	0,658	0,894	0,909	0,928
<b>Adamic Adar (AA)</b>	0,516	0,834	0,900	0,919
<b>Tercihli Bağlantı İndeksi</b>	0,730	0,790	0,799	0,816
<b>Sorenson index</b>	0,438	0,656	0,730	0,745
<b>Kaynak Tahsisli Index(RA)</b>	0,511	0,579	0,727	0,742



Çizelge 5.11. 2007-2017 aralığında oluşturulan UEFA ağında bağlantı tahmini AUC sonuçları.

	2014	2015	2016	2017
<b>Ortak Komşu (CN)</b>	0,310	0,446	0,504	0,514
<b>Jaccard Index</b>	0,399	0,542	0,627	0,64
<b>Adamic Adar (AA)</b>	0,313	0,506	0,546	0,557
<b>Tercihli Bağlantı İndeksi</b>	0,443	0,479	0,615	0,628
<b>Sorenson index</b>	0,266	0,398	0,443	0,452
<b>Kaynak Tahsisli Index(RA)</b>	0,320	0,351	0,441	0,450

Bağlantı tahmini yöntemlerinin 2007-2017 yılları arasında oluşturulan ağlardaki başarıları karşılaştırıldığında, Çizelge 9'daki Snooker ağlarında 2014 yılı tahmininde Tercihli Bağlantı İndeksi, 2015, 2016 ve 2017 yılı tahminlerinde Jaccard İndeks, Çizelge 10'teki Tenis ağlarında 2014 yılı tahmininde Tercihli Bağlantı İndeksi, 2015, 2016 ve 2017 yılı tahminlerinde Jaccard İndeks, Çizelge 11'deki UEFA Avrupa Ligi tahminlerinde 2014 yılı tahmininde Tercihli Bağlantı İndeksi, 2015, 2016 ve 2017 yılları bağlantı tahminlerinde Jaccard İndeks yöntemlerinin başarılı olduğu görülmektedir.

Tüm ağlar için genel olarak bağlantı tahmin yöntemlerinin başarılarına bakıldığında 2004-2011 yılları arasındaki ağlarda Tercihli Bağlantı İndeksi, Adamic Adar ve Jaccard İndeks ve yöntemlerinin diğer yöntemlere göre başarılı oldukları görülmektedir. Bu zaman periyodunda Jaccard İndeks yönteminin 2014 yılı hariç tüm yıllarda diğer yöntemlerden daha başarılı tahminlerde bulunduğu anlaşılmaktadır. 2012 yılında ise önceli zaman aralıklarında olduğu gibi Tercihli Bağlantı İndeksi yönteminin zaman aralığının kısa olduğu test ağında daha başarılı tahmin yaptığı sonucu doğrulanmış olmaktadır.

#### **5.3.4. 2004-2017 Arası Oluşturulan Ağlar**

2004-2017 yılları arasında veriler için 2004-2013 arası zaman periyodundaki ağ eğitim için kullanılırken, 2014, 2015, 2016 ve 2017 yılları ise test ağları olarak kullanılmıştır. Çizelge 12-14' de AUC metriğinden elde edilen sonuçlara bakıldığında, Çizelge 9-

11’de olduğu gibi 2007-2017 yılları arasında oluşturulan ağlardaki tahminlere benzer şekilde tahminler yapıldığı görülmektedir. 2007-2017 zaman aralığındaki tahminlerden farklı olarak Çizelge 12-14’e bakıldığında eğitim ağında kullanılan kısa zaman aralıklarındaki tahminlerin önceki zaman periyotlarındaki tahminlerden daha başarılı olduğu anlaşılmaktadır. Bunun nedeni ise eğitim için oluşturulan ağın uzun zaman periyodundan oluşmasıdır. Çünkü, eğitim için oluşturulan ağın zaman aralığı uzun olması tahmin için daha fazla topolojik bilgilerin kullanılmasını sağlamakta ve bu da kısa zaman aralığında tahmin başarısını artırıcı bir etken olmaktadır. Bununla birlikte test için kullanılan ağın zaman aralığının uzun olması önceli zaman periyotlarında olduğu gibi benzer derecede yüksek bir başarı elde edilmesini sağlamaktadır.

Çizelge 5.12. 2004-2017 aralığında oluşturulan snooker ağında bağlantı tahmini AUC sonuçları.

	2014	2015	2016	2017
<b>Ortak Komşu (CN)</b>	0,627	0,7159	0,815	0,930
<b>Jaccard Index</b>	0,724	0,790	0,861	0,938
<b>Adamic Adar (AA)</b>	0,634	0,717	0,809	0,914
<b>Tercihli Bağlantı İndeksi</b>	0,804	0,861	0,912	0,940
<b>Sorenson index</b>	0,524	0,597	0,687	0,796
<b>Kaynak Tahsisli Index(RA)</b>	0,610	0,696	0,801	0,841

Çizelge 5.13. 2004-2017 aralığında oluşturulan tenis ağında bağlantı tahmini AUC sonuçları.

	2014	2015	2016	2017
<b>Ortak Komşu (CN)</b>	0,615	0,701	0,799	0,912
<b>Jaccard Index</b>	0,710	0,774	0,844	0,920
<b>Adamic Adar (AA)</b>	0,622	0,702	0,793	0,896
<b>Tercihli Bağlantı İndeksi</b>	0,788	0,844	0,894	0,922
<b>Sorenson index</b>	0,514	0,585	0,673	0,780
<b>Kaynak Tahsisli Index(RA)</b>	0,598	0,682	0,785	0,825

Çizelge 5.14. 2004-2017 aralığında oluşturulan UEFA ağında bağlantı tahmini AUC sonuçları.

	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>
<b>Ortak Komşu (CN)</b>	0,609	0,695	0,792	0,903
<b>Jaccard Index</b>	0,703	0,767	0,836	0,911
<b>Adamic Adar (AA)</b>	0,616	0,696	0,786	0,888
<b>Tercihli Bağlantı İndeksi</b>	0,781	0,836	0,886	0,913
<b>Sorenson index</b>	0,509	0,58	0,667	0,773
<b>Kaynak Tahsisli Index(RA)</b>	0,593	0,676	0,778	0,817

Bağlantı tahmini yöntemlerinin 2004-2017 yılları arasında oluşturulan ağlardaki başarıları karşılaştırıldığında, Çizelge12-14'de görüldüğü gibi Tercihli Bağlantı İndeksi, eğitim için oluşturulan bütün zaman aralıklarında ve bütün spor branşlarında diğer yöntemlere göre daha başarılı tahmin yapmıştır. Bunun nedeni ise, Tercihli Bağlantı İndeksi yönteminin eğitim için kullanılan ağlarda zaman periyodunun uzun olmasından kaynaklı olarak ağdaki topolojik bilgileri kullanarak daha fazla benzerlik oranı hesaplaması tahminlerdeki başarısını arttırmıştır. Bununla birlikte 2004-2017 yılları arasında oluşturulan eğitim ve test ağlarında kullanılan komşuluk tabanlı bağlantı tahmini yöntemlerinin başarılı olduklarını söylemek mümkündür.

## BÖLÜM 6

### SONUÇLAR

Bu tez çalışmasında, karmaşık ağlarda gelecekte muhtemel bağlantıları tahmin etmek için kullanılan bağlantı tahmini yöntemlerinden komşuluk tabanlı bağlantı tahmini yöntemlerinin spor ağlarındaki başarısı araştırılmıştır. Çalışmada 2004-2017 yılları arasında, Snooker Dünya Şampiyonası, Avustralya Açık Tenis Turnuvası ve UEFA Avrupa Ligi verileri kullanılarak 4 farklı zaman periyodunda ağlar oluşturulmuştur. Oluşturulan ağlarda geleneksel komşuluk tabanlı bağlantı tahmini yöntemleri kullanılarak bağlantı tahmini işlemleri gerçekleştirilmiştir. Yapılan bağlantı tahmini işlemlerinin başarısı AUC metriği kullanılarak çıkarılmıştır. Gerçekleştirilen deneysel çalışmalar sonucunda geleneksel komşuluk tabanlı bağlantı tahmini yöntemlerinin spor ağlarında bağlantı tahmini işlemleri için kullanılabilir olduğu tespit edilmiştir. Deneysel çalışmalardan elde edilen önemli sonuçlardan birisi de oluşturan ağların zaman aralıklarının önemi olmuştur. Özellikle eğitim seti için oluşturulan ağların zaman aralığı kısa olması durumunda tahmin başarılarının düşük olduğu görülmektedir. Zaman aralığı kısa olan test ağlarında en başarılı yöntem Tercihli Bağlantı İndeksi olurken, test ağında zaman aralığı arttıkça Jaccard İndeks yöntemi daha başarılı sonuçlar vermiştir. Bununla birlikte eğitim ağının uzun olduğu 2004-2017 zaman aralığındaki tahmin sonuçlarına bakıldığında Çizelge 12-14'e bakıldığında Tercihli Bağlantı İndeksi yönteminin diğer yöntemlerden daha başarılı olduğu görülmektedir. Bunun sebebi ise Tercihli Bağlantı İndeksi yönteminin topolojik verilerin fazla olması durumunda zaman aralığı kısa ağlarda bile tahmin gücünün yüksek olmasından kaynaklanmaktadır. Geleneksel komşuluk tabanlı bağlantı tahmini yöntemlerinin uygulandığı ağların bağlantı tahmini sonuçları karşılaştırıldığında en başarılı tahminlerin Snooker ağlarında olduğu görülmektedir. Diğerlerine göre en başarısız tahmin sonuçları ise UEFA Avrupa Ligi ağlarında olmuştur. Bunun sebebi ise tahmin işlemlerinde kullanılan verilerdeki düğüm ve

baęlantı sayılarının etkili olmasıdır. Dügüm ve baęlantı sayısı fazla olan aęlarda komşuluk tabanlı baęlantı tahmini yöntemleri daha başarılı olmaktadır.

Karmaşık aęlarda baęlantı tahmini için kullanılan topolojik ve topolojik olmayan dięer yöntemlerin spor aęlarındaki başarılarının araştırılması bu tez çalışmasından çıkan sonuçlar dahilinde önerilmektedir.

## KAYNAKLAR

1. Newman, M. E. J.: *Siam Rev.* 45, 167 (2003).
2. Dorogovtsev ,S. N.,Mendes ,J. F. F.: In Evolution Of Networks, **Oxford University Press**, Oxford, (2003).
3. Dodds ,P. S., Muhamad ,R., Watts ,D. J.: *Science* 301, 827 (2003).
4. Watts ,D. J., S. Strogatz,H.: *Nature* (London) 393, 440 (1998).
5. Watts Dj, Strogatz Sh: Collective Dynamics Of 'Small-World' Networks. *Nature*. 393 (6684): 440–2, (1998)..
6. Erdos, P., Rényi, A, On Random Graphs. I(Pdf). *Publicationes Mathematicae*. 6: 290–297, (1959).
7. Barabasi ,A. -L., Albert ,R.: Emergence Of Scaling In Random Networks, *Science*, 286, 509 (1999).
8. Newman M.E.J.: Clustering And Preferential Attachment In Growing Networks, *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* 64 (2001).
9. Paul: Etude De La Distribution Florale Dans Une Portion Des Alpes Et Du Jura, *Bull. La Soc. Vaudoise Des Sci. Nat.* 37, 547–579,(1901)
10. Barabasi Albert-Laszlo, Albert Reka, Emergence Of Scaling In Random Networks, *Science* (80-. ). 286 (1999) 509–512.
11. Barabasi ,A. -L., Albert ,R.: Statistical Mechanics Of Complex Networks, *Reviews Of Modern Physics*. 74 (1): 47–97,(2002)
12. Zhou,T., Lü, L., Zhang, Y.C.: Predicting Missing Links Via Local Information, *Eur. Phys. J. B.* 71 623–630, (2009)
13. Sorensen T.: A Method Of Establishing Groups Of Equal Amplitude In Plant Socio-logy Based On Similarity Of Species Content., Det. Kong. Danske Vidensk, *Selesk Biol. Skr.* 5,1–34, (1948) .
14. Liben-Nowell, D., Kleinberg,J.: The Link-Prediction Problem For Social Networks, *J. Am. Soc. Inf. Sci. Technol.* 58, 1019–1031, (2007).

15. Linyuan, L.L., Zhou, T.: Link Prediction In Complex Networks: A Survey, *Phys. A Stat. Mech. Its Appl.* 390, 1150–1170, (2011)
16. Huang, Z.: Link Prediction Based On Graph Topology: The Predictive Value Of Gene-ralized Clustering Coefficient, *Ssrn.* (2010).
17. Resnick, P., Varian, H.R.: Recommender Systems Mmende Tems, *Commun. Acn.* **40**, 56–58, (1997).
18. Lü, L., Medo, M., Yeung, C.H., Zhang, Y.-C., Zhang, Z.-K., Zhou, T.: Recommender Systems, *Phys. Rep.* 519, 1–49, (2012).
19. Huang, Z., Li, X., Chen, H.: Link Prediction Approach To Collaborative Filtering, *Proc. 5th Acn/Ieee-Cs Jt. Conf. Digit. Libr. - Jcdl '05.* (2005)
20. Kleinberg, J.: Analysis Of Large-Scale Social And Information Networks Subject Areas , *Author For Correspondence* , (2013).
21. Zhang, Q.M., Lü, L., Wang, W.Q., Zhu, Y.X., Zhou, T.: Potential Theory For Directed Networks, *Plos One.* 8 (2013)
22. Bürhan, Y., Daş, R.: Akademik Veritabanlarından Yazar-Makale Bağlantı Tahmini, *Politeknik Dergisi Journal Of Polytechnic*, 787-800, Ankara(2017).
23. Hanley, J.A., Mcneil, B.J.: The Meaning And Use Of The Area Under A Receiver Operating Characteristic (Roc) Curve., *Radiology.* 143, 29–36 (1982).
24. Wang, W.Q., Zhang, Q.M., Zhou, T., Evaluating Network Models: A *Likelihood Analysis*, Epl. 98, 1–6, (2012)
25. Joseph D O'Brien, James P Gleeson, A complex networks approach to ranking professional Snooker players, Journal of Complex Networks, Volume 8, *Issue* 6, 1 December 2020
26. Fındık ,O., Özkaynak ,E.: Complex Network Analysis Of Players In Tennis Tourna-ments, International Conference On Advanced Technologies, *Computer Engineering And Science (Icatces'18)*, 383-388, Karabük (2018)
27. Sulak,E.E., Yılmaz,H. Özkaynak,E.: Complex Network Analysis Of Uefa Europe Lea-gue Competitions, International Conference On Advanced Technologies, *Computer Engineering And Science (Icatces'18)*, 389-393, Karabük (2018)
28. Hanley, J. A. and McNeil, B. J., "The meaning and use of the area under a receiver operating characteristic (ROC) curve.", *Radiology*, 143 (1): 29–36 (1982).

29. Nasiri, E., Berahmand, K., & Li, Y. (2021). A new link prediction in multiplex networks using topologically biased random walks. *Chaos, Solitons & Fractals*, *151*, 111230. doi: 10.1016/J.CHAOS.2021.111230
30. Berahmand, K., Nasiri, E., Forouzandeh, S., & Li, Y. (2021). A preference random walk algorithm for link prediction through mutual influence nodes in complex networks. *Journal of King Saud University - Computer and Information Sciences*. doi: 10.1016/J.JKSUCI.2021.05.006
31. Karimi, F., Lotfi, S., & Izadkhah, H. (2021). Community-guided link prediction in multiplex networks. *Journal of Informetrics*, *15*(4), 101178. doi: 10.1016/J.JOI.2021.101178
32. Zou, L., Wang, C., Zeng, A., Fan, Y., & Di, Z. (2021). Link prediction in growing networks with aging. *Social Networks*, *65*, 1–7. doi: 10.1016/J.SOCNET.2020.11.001
33. Luo, H., Li, L., Zhang, Y., Fang, S., & Chen, X. (2021). Link prediction in multiplex networks using a novel multiple-attribute decision-making approach. *Knowledge-Based Systems*, *219*, 106904. doi: 10.1016/J.KNOSYS.2021.106904
34. Daud, N. N., Ab Hamid, S. H., Saadoon, M., Sahran, F., & Anuar, N. B. (2020). Applications of link prediction in social networks: A review. *Journal of Network and Computer Applications*, *166*, 102716. doi: 10.1016/J.JNCA.2020.102716
35. Malhotra, D., & Goyal, R. (2021). Supervised-learning link prediction in single layer and multiplex networks. *Machine Learning with Applications*, *6*, 100086. doi: 10.1016/J.MLWA.2021.100086
36. Berahmand, K., Nasiri, E., Forouzandeh, S., & Li, Y. (2021). A preference random walk algorithm for link prediction through mutual influence nodes in complex networks. *Journal of King Saud University - Computer and Information Sciences*. doi: 10.1016/J.JKSUCI.2021.05.006
37. Singh, S. P., & Mishra, B. N. (2016). Major histocompatibility complex linked databases and prediction tools for designing vaccines. *Human Immunology*, *77*(3), 295–306. doi: 10.1016/J.HUMIMM.2015.11.012
38. Wang, J., Zhang, Q. M., & Zhou, T. (2019). Tag-aware link prediction algorithm in complex networks. *Physica A: Statistical Mechanics and Its Applications*, *523*, 105–111. doi: 10.1016/J.PHYSA.2019.02.028
39. Najari, S., Salehi, M., Ranjbar, V., & Jalili, M. (2019). Link prediction in multiplex networks based on interlayer similarity. *Physica A: Statistical*



*Mechanics and Its Applications*, 536, 120978. doi: 10.1016/J.PHYSA.2019.04.214

40. Sulak, E.E., Yılmaz, H. Özkaynak, E.: Complex network analysis of UEFA Europe league competitions. In: International Conference on Advanced Technologies, *Computer Engineering and Science (ICATCES 2018)*, Karabük, pp. 389–393 (2018)
41. Chi, K., Yin, G., Dong, Y., & Dong, H. (2019). Link prediction in dynamic networks based on the attraction force between nodes. *Knowledge-Based Systems*, 181, 104792. doi: 10.1016/J.KNOSYS.2019.05.035
42. Yao, L., Wang, L., Pan, L., & Yao, K. (2016). Link Prediction Based on Common-Neighbors for Dynamic Social Network. *Procedia Computer Science*, 83, 82–89. doi: 10.1016/J.PROCS.2016.04.102
43. Kumar, A., Mishra, S., Singh, S. S., Singh, K., & Biswas, B. (2020). Link prediction in complex networks based on Significance of Higher-Order Path Index (SHOPI). *Physica A: Statistical Mechanics and Its Applications*, 545, 123790. doi: 10.1016/J.PHYSA.2019.123790
44. Wu, J. (2018). A generalized tree augmented naive Bayes link prediction model. *Journal of Computational Science*, 27, 206–217. doi: 10.1016/J.JOCS.2018.04.006
45. Moradabadi, B., & Meybodi, M. R. (2018). Link prediction in weighted social networks using learning automata. *Engineering Applications of Artificial Intelligence*, 70, 16–24. doi: 10.1016/J.ENGAPPAL.2017.12.006
46. Barabasi, A.-L., Albert, R.: Emergence of scaling in random networks. *Science* 286, 509 (1999)
47. Fındık ,O., Özkaynak ,E.: Complex Network Analysis Of Players In Tennis Tourna-ments, International Conference On Advanced Technologies, *Computer Engineering And Science (Icatces'18)*, 383-388, Karabük (2018)
48. Sulak,E.E., Yılmaz,H. Özkaynak,E.: Complex Network Analysis Of Uefa Europe Lea-gue Competitions, International Conference On Advanced Technologies, *Computer Engineering And Science (Icatces'18)*, 389-393, Karabük (2018)
49. Ruhnau, B.: *Eigenvector centrality: a node-centrality?* *Soc. Netw.* 22(4), 357–365 (2000)
50. Wang, P., Xu, B., Wu, Y., and Zhou, X., "Link prediction in social networks: the state-of-the-art", *Science China Information Sciences*, 58 (1): 1–38 (2015).
51. Sarukkai, R. R., "Link prediction and path analysis using Markov

- chains", *Computer Networks*, 33 (1–6): 377–386 (2000).
52. Guo, F., Yang, Z., and Zhou, T., "Predicting link directions via a recursive subgraph-based ranking", *Physica A: Statistical Mechanics And Its Applications*, 392 (16): 3402–3408 (2013).
53. Zhou, T., Lü, L., Zhang, Y.C.: Predicting missing links via local information. *Eur. Phys. J. B* 71, 623–630 (2009)
54. Newman M.E.J.: Clustering and Preferential Attachment in Growing Networks, *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* 64 (2001).
55. Murata, T. and Moriyasu, S., "Link Prediction of Social Networks Based on Weighted Proximity Measures", *IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)*, Fremont, 85–88 (2007).
56. Jaccard, P., "Etude de la distribution florale dans une portion des Alpes et du Jura", *Bulletin De La Societe Vaudoise Des Sciences Naturelles*, 37: 547–579 (1901).
57. de Sa, H. R. and Prudencio, R. B. C., "Supervised link prediction in weighted networks", *The 2011 International Joint Conference on Neural Networks*, San Jose, 2281–2288 (2011).
58. Adamic, L. A. and Adar, E., "Friends and neighbors on the Web", *Social Networks*, 25 (3): 211–230 (2003).
59. Zhou, T., Lü, L., Zhang, Y.C.: Predicting missing links via local information. *Eur. Phys. J. B* 71, 623–630 (2009)
60. [https://figshare.com/articles/dataset/Snooker\\_datasets/13604480/1](https://figshare.com/articles/dataset/Snooker_datasets/13604480/1)
61. Karadeniz, E., Temel, M. M., & Karci, A. (2019). Prediction of Collaboration Between Universities of Turkey. *2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018*. doi: 10.1109/IDAP.2018.8620900
62. Bürhan, Y., & Daş, R. (2017). Akademik Veritabanlarından Yazar-Makale Bağlantı Tahmini. *Journal of Polytechnic*. doi: 10.2339/politeknik.368989
63. İnternet: IEEE Xplore, "2000-2017 Arası Yayınlanmış Makaleler", <http://www.tennis-data.co.uk/alldata.php>
64. <https://Kassiesa.Ho-me.Xs4all.Nl/Bert/Uefa/Data/Index.Html> (2018).

65. Ortu, M., Destefanis, G., Counsell, S., Swift, S., Tonelli, R., & Marchesi, M. (2017). How diverse is your team? Investigating gender and nationality diversity in GitHub teams. *Journal of Software Engineering Research and Development*, 5(1). doi: 10.1186/s40411-017-0044-y

## ÖZGEÇMİŞ

Mine KELEŞ; ilk ve orta öğrenimini aynı şehirde tamamladı. Çarşamba Anadolu Lisesi Fen Bilimleri Bölümü'nden mezun oldu. Ardından Karabük Üniversitesi Bilgisayar Mühendisliği Bölümü'nde öğrenime başlayıp 2017 yılında mezun oldu. Ardından 2017 yılında Karabük Üniversitesi Bilgisayar Mühendisliği Bölümünde yüksek lisans eğitimine başladı. 2018-2019 teknopark bünyesinde Yazılım Destek Personeli olarak çalıştı. 2019-2022 yılları arası eğitimlik yaptı. Spor ağlarında bağlantı tahmini konulu yüksek lisans tezini 2022 yılında tamalamıştır. Lisans ve lisansüstü eğitimleri sırasında staj bilgileri ve kullandığı bireysel program projeleri aşağıda verilmiştir:

**Bilgisayar Bilgileri:** Word, Excel, Notepad, Visual Studio, Adobe Dreamweaver, Matlab ArcGis, Android Stdio, Eclips, SQL Server

**Yetenekler:** C, PHP, HTML, CSS, XML, JavaScript, C ++, MATLAB, Python, MySQL, Java

**Stajlar:** Samsun Büyükşehir Belediyesi (C # Kütüphane ve C # Otel Otomasyonu)

**Lisans Projesi:** ArcGis ile Perakende Pazarlarının Yeri için GIS Uygulaması

**Lisans Bireysel Projeler:**

- Php MVC yapısı ve Codegnighter Framework ile Site Tasarımı
- Eclips ile Java Soket Programlama Mini Sohbet Uygulaması
- Diyabet Tespiti için Mobil Uygulama(Android Studio)
- Tarif Kitabı için Mobil Uygulama(Android Studio)
- Çeviri için oluşturulan Mobil Uygulama(Android Studio)

**Hacettepe Teknopark Projesi:** Canlı Dershanecilik Eğitim Uygulaması(PHP-FrontEnd ve Backend Çalışmaları)

**Yüksek Lisans Bildiri Çalışmaları:**

- Cuda ile Acile Gelen Hastaların Toplam Acil Süresinin Ortalama Hesaplanması(2017)

- Karmaşık Spor Ağlarında Bağlantı Tahmini(2021)

**Yüksek Lisans Projesi:** Spor Ağlarında Bağlantı Tahmini Yöntemlerinin Karşılaştırılması(2022)