

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



AKADEMİK AĞLARDA BAĞLANTI TAHMİNİ
UYGULAMASI VE ANALİZİ

Yücel BÜRHAN

Yüksek Lisans Tezi
Anabilim Dalı: Yazılım Mühendisliği
Danışman: Doç. Dr. Resul DAŞ

TEMMUZ-2017

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

AKADEMİK AĞLARDA BAĞLANTI TAHMİNİ UYGULAMASI VE ANALİZİ

YÜKSEK LİSANS TEZİ
Yücel BÜRHAN
141137111

Tezin Enstitüye Verildiği Tarih: 17.05.2017

Tezin Savunulduğu Tarih: 09.06.2017

Tez Danışmanı : Doç. Dr. Resul DAŞ

Diğer Jüri Üyeleri : Prof. Dr. Asaf VAROL

: Yrd. Doç. Dr. Özal YILDIRIM

Haziran-2017

TEŞEKKÜR

Çalışmalarım boyunca beni yönlendiren, bana araştırma olanağı sunan,engin bilgi ve geniş hoşgörüsü ile benden hiçbir desteğini esirgemeyen değerli hocam Doç. Dr. Resul DAŞ'a, benden maddi manevi hiçbir desteğini esirgemeyen aileme ve arkadaşlarıma teşekkürü bir borç bilirim.

Yücel BÜRHAN
ELAZIĞ - 2017

İÇİNDEKİLER

	Sayfa No
TEŞEKKÜR	II
İÇİNDEKİLER	III
ÖZET.....	V
SUMMARY	VI
ŞEKİLLER LİSTESİ.....	VII
TABLolar LİSTESİ.....	VIII
KISALTMALAR	IX
1. GİRİŞ.....	1
1.1. Tez Çalışmasının Amacı ve Kapsamı	2
1.2. Literatür Taraması	2
2. AĞLARIN MATEMATİĞİ	5
2.1. Ağlar ve Temsilleri.....	5
2.2. Komşuluk Matrisi.....	5
2.3. Ağırlıklı Ağlar	7
2.4. Yönlü Ağlar	8
2.4.1. Çevrimsiz Yönlü Ağlar	9
2.5. İkili Ağlar	10
2.6. Ağaçlar	11
2.7. Ağlarda Derece.....	12
3. ÖLÇÜMLER VE METRİKLER.....	14
3.1. Derece Merkeziliği	14
3.2. Özvektör Merkeziliği	14
3.3. Katz Merkeziliği.....	15
3.4. Sayfa Derecesi.....	16
3.5. Yakınlık Merkeziliği	17
3.6. Arasındalık Merkeziliği.....	18
3.7. Benzerlik	19
3.7.1. Kosinüs Benzerliği	20
3.7.2. Pearson Korelasyon Katsayıları	21

4.	SOSYAL AĞ ANALİZİ.....	23
4.1.	Sosyal Ağ Analiz Aşamaları	23
4.2.	Sosyal Ağ Analizi Araçları	24
5.	AĞLARDA BAĞLANTI TAHMİNİ YÖNTEMLERİNİN	
	İNCELENMESİ	27
5.1.	Graf Tabanlı Yöntemler	28
5.2.	Olasılıksal Yöntemler.....	29
5.3.	Benzerlik Tabanlı Yöntemler	29
5.3.1.	Komşuluk Tabanlı Ölçütler	30
5.3.2.	Yol Tabanlı Ölçütler.....	32
6.	AKADEMİK AĞLARDA YAZAR-MAKALE BAĞLANTI	
	TAHMİNİ UYGULAMASI.....	34
6.1.	Verilerin Elde Edilmesi.....	36
6.2.	Ön İşlem	37
6.3.	Bağlantı Tahmini Yöntemlerinin Uygulanması:	39
6.4.	Uygulama Sonuçları	44
7.	SONUÇ VE ÖNERİLER.....	54
	KAYNAKLAR.....	56
	ÖZGEÇMİŞ.....	62

ÖZET

Günümüzde, İnternet ve sosyal ağlardaki gelişmeler ile birlikte sosyal ağ analizi ve bağlantı tahmini konusu popülerliğini sürdürmektedir. Bağlantı tahmini probleminin çözümü için, graf yapısıyla beraber web madenciliğinde kullanılan bağlantı tahmini yöntemleri de kullanılmaktadır. Ayrıca gerçek dünya problemlerinin modellenebilmesi için graf yapıları ön plana çıkmaktadır. Sosyal ağ analizi yöntemleri bu modellemelere uygulanmakta, graftaki düğüme karşılık gelen aktörlerin özellikleri ve aktörler arası ilişkilerin durumu ile ilgili olarak çıkarsamalar yapılmaktadır.

Bu tez çalışmasında, sosyal ağ analizi ve bağlantı tahmini için öncelikle kullanılabilecek parametreler ve yöntemler örneklerle açıklanmış, analizi yapılabilecek ağ modelleri matematiksel olarak incelenmiştir. Daha sonra akademik ağlarda bağlantı tahmini üzerine yazılım uygulaması gerçekleştirilmiştir. Uygulamada yazar-makale ağı; ikili ağ olarak modellenmekte, yazarlar ana düğümler olarak nitelendirilmektedir. Çalışmada birlikte yayın yapan yazarlar arasında bağlantılar bulunmakta ve bu bağlantılar ayrıtlar ile temsil edilmektedir. Çalışmada makalelerin anahtar kelimeleri referans alınmaktadır. Yazarlar ile yazdıkları makalelerin anahtar kelimeleri arasında da bağlantılar bulunmaktadır. Yazarlar ile bağlantılı oldukları anahtar kelimeler arasındaki benzerlikler hesaplanarak, yazarlar arasında oluşması muhtemel yeni bağlantılar tahmin edilmeye çalışılmaktadır.

Yapılan çalışma ile yazarların çalışmak istediği konular göz önünde bulundurularak bir yazarın çalışmak istediği alanda birlikte çalışma yapabileceği yazarlar tespit edilmeye çalışılmakta ve elde edilen sonuçlar ile bunun başarılabilirdiği gözlenmektedir. Uygulamanın doğruluğu için yapısı bilinen bir ağ üzerinde çalışılmıştır. Ağda bazı bağlantılar oluşmadan önceki zamana gidilmiş, günümüzde oluşan bağlantıların oluşma ihtimalleri hesaplanmış ve oldukça iyi sonuçlar elde edilmiştir.

Anahtar Kelimeler: Graf Teorisi, Sosyal Ağ Analizi, Bağlantı Tahmini, Sosyal Ağlar, Ağların Matematiği.

SUMMARY

Application and Analysis of Link Prediction in Academic Networks

Nowadays, along with improvements in the Internet and social networks, social network analysis and link prediction are still popular. To solve the link prediction problem, link prediction methods used in web mining are used together with graph structures. In addition, graph structures come into prominence for modeling real-world problems. Social network analysis methods are applied to these models, and inferences are made about properties of the actors corresponding to the node in graph and the situation of inter-actor relations.

In this thesis, the parameters and methods that can be used primarily for social network analysis and link prediction were explained with examples and network models that can be analyzed were mathematically examined. Then, the software on link prediction was performed in the academic networks. Author-article network in software, modeled as a binary network, the authors were considered as the main nodes. There are links between authors who co-publish in the software, and these links are represented by the edges. In the software, the keywords of the articles were taken as reference. There are also links between the authors and the keywords of the articles that they write. By calculating the similarities between the authors and the keywords that they are related to, attempts were made to predict possible new connections between the authors.

By considering the topics that the authors want to study, it is observed that the authors who can work together in an area where an author wants to work are tried to be determined and it is observed that it can be achieved with the obtained results. It has been worked on a known network, for the accuracy of the software. It has been gone to the time before some connections, and the probabilities of the connections occurring today were calculated and very good results were obtained.

Keywords: Graph Teory, Social Network Analysis, Link Prediction, Social Networks, Mathematics of Networks.

ŞEKİLLER LİSTESİ

	Sayfa No
Şekil 2.1. İki küçük ağ.	5
Şekil 2.2. Ağırlıklı ağ.....	7
Şekil 2.3. Yönlü ağ. Kenarları yönleri gösteren oklar ile yönlü bir ağ.	8
Şekil 2.4. Çevrimsiz yönlü ağ.	9
Şekil 2.5. İkili ağın iki tane tek modlu tasarımı.	11
Şekil 2.6. Aynı ağacın iki farklı çizimi.	11
Şekil 3.1. Denklik Örnekleri	19
Şekil 4.1. Sosyal ağ analiz aşamaları	24
Şekil 5.1. Sosyal bir ağda graf yapısı.....	27
Şekil 6.1. Geliştirilen uygulamanın adımları	34
Şekil 6.2. Bağlantı tahmini yöntemleri	35
Şekil 6.3. Geliştirilen uygulamanın veriler eklenmeden önceki ana sayfa görünümü.....	35
Şekil 6.4. IEEE Explore veri tabanından seçilebilecek alanların uygulama arayüzünde görünümü	36
Şekil 6.5. Yazar adı ile arama yapıp istenen verileri çekme	37
Şekil 6.6. Yazar isimleri, yayınlanan makaleler ve yayın yılları	38
Şekil 6.7. Anahtar kelimeler ve yayınlanan makalelerde kullanım sayıları	38
Şekil 6.8. UCINET ile yazar-yazar ilişki ağı	39
Şekil 6.9. PAJEK ile yazar-yazar ilişki ağı.....	40
Şekil 6.10. GEPHI ile yazar-yazar ilişki ağı	40
Şekil 6.11. UCINET ile yazar-konu ilişki ağı örneği.....	41
Şekil 6.12. PAJEK ile yazar-konu ilişki ağı örneği	42
Şekil 6.13. GEPHI ile yazar-konu ilişki ağı örneği	43
Şekil 6.14. Bağlantı tahmini yöntemlerine göre sonuç ekranı	44
Şekil 6.15. UCINET ile yeni yazar-yazar ağı	52
Şekil 6.16. PAJEK ile yeni yazar-yazar ağı.....	52
Şekil 6.17. GEPHI ile yeni yazar-yazar ağı	53

TABLÖLAR LİSTESİ

	Sayfa No
Tablo 3.1. Dört tane özvektör merkeziliği	17
Tablo 4.1. Sosyal Ağ araçlarının Karşılaştırılması.....	26
Tablo 6.1. Jaccard İndeks uygulaması sonuçları	47
Tablo 6.2. Sorensen İndeks Uygulama sonuçları	48
Tablo 6.3. Ortak Komşu uygulama sonuçları.....	49
Tablo 6.4. L. H. Newman İndeks Uygulama Sonuçları	50
Tablo 6.5. Salton İndeks Uygulama Sonuçları.....	51

KISALTMALAR

CEO	: Chief Executive Officer (İcra Kurulu Başkanı)
CFO	: Chief Financial Officer (Finansal Grup Başkanı)
CIO	: Chief Information Officer (Bilgi Sistemleri Grubu Başkanı)
KKO	: Karınca Koloni Optimizasyonu
ISSN	: International Standard Serial Number (Uluslararası Standart Süreli Yayın Numarası)
ISBN	: International Standard Book Number (Uluslararası Standart Kitap Numarası)

1. GİRİŞ

Veri analizi ve verilerin bilgiye dönüştürülebilmesi hayatı kolaylaştırmak açısından büyük önem arz etmektedir. Geçmiş yıllarda bu konuda yapılan çalışmalarda en büyük sorun, verilerin elde edilme zorluğu olarak göze çarpmaktadır. Günümüzde sosyal ağların gelişmesi ile birlikte veriye ulaşmak oldukça kolaylaşmakta ve bu durum sosyal ağ analizi konusunda yapılan çalışmaların artmasına neden olmaktadır.

Sosyal ağlar, ortaya çıktığı andan itibaren son derece yaygın bir şekilde kullanılmaya başlanmıştır. İnsanların sosyal ağlar üzerinden birbirileri ile ilişkiler kurması, kişisel veya sosyal paylaşımlarda oldukça büyük miktarda veriye hiç zorlanmadan ulaşılabilmesini sağlamaktadır. Her geçen gün bu verilerin miktarı muazzam artış göstermektedir.

Sosyal ağlar insanların birbirileri ile ilişki kurdukları; fikir, düşünce, fotoğraf, video, konum gibi verileri paylaştıkları ortamlardır. Sosyal ağlardan elde edilen veriler irdelendiğinde, aktörler ve ilişkileri hakkında çok önemli bilgilere ulaşılabilir. Bu ağların en yaygın olarak kullanılanları facebook, instagram, twitter, linkedin, google+, youtube gibi adreslerdir.

Facebook ilgi alanları, fikir-düşünce, fotoğraf ve video paylaşma, instagram fotoğraf ve video paylaşma; twitter fikir, yorumlar, resim ve video paylaşma; linkedin iş arama, aynı meslek grubundakilerin birbirileri ile iletişime geçebilmesi; google+ birbirini tanıyan veya aynı çevreden olan veya birbirini tanıyan insanların ilişki kurabilmesini sağlama; youtube ise video paylaşma amacı ile kullanılmaktadır.

Sosyal ağlardan kolaylıkla elde edilebilen verilerden gereksiz veriler temizlenip işe yarayacak veriler saklanır, bu veriler işlenir ve kullanıma hazır hale getirilir. Bu veriler bilgiye dönüştükten sonra anlamlı hale gelirler. Sosyal ağlardan elde edilen veriler birçok amaçla işlenir, bilgiler çıkarılır. Veriler, yaygın olarak bağlantı tahmini ve duygu analizi gibi konular açısından incelenmektedir [1].

Son yıllarda sosyal ağ analizi alanında birçok çalışma yapılırken bağlantı tahmini problemi üzerinde de yoğun çalışmalar devam etmektedir. Günümüzde bibliyografik alan [2], moleküler biyoloji [3], adli soruşturmalar [4], tavsiye amaçlı sistemler [5] vb. alanlarda bağlantı tahmini yapısı kullanılmaktadır. Dinamik bir yapıya sahip ağlarda bağlantı tahmininde bulunmak zordur.

1.1. Tez Çalışmasının Amacı ve Kapsamı

Bu tez çalışmasının amacı, yazar-makale ağının irdelenerek yazarlar arasında oluşabilecek bağlantıları tespit eden bir yazılım geliştirmektir. Bu doğrultuda .Net yazılım aracı ile C# dilinde bir otomasyon yazılmış ve bağlantı tahmini yöntemleri kullanılarak ağ analizi yapılmıştır. Ayrıca, UCINET sosyal ağ analiz aracı kullanılarak yazar-yazar ve yazar-anahat kelime ağları çizilmiştir.

1.2. Literatür Taraması

Sosyal ağlarda bağlantı tahmini üzerine literatürde yapılmış birçok çalışma bulunmaktadır. Bağlantı tahmini problemi birçok alanda kendini göstermektedir. Ortak yazar ağı bu alanlar arasında oldukça önemli bir yer tutmaktadır. Ortak yazar ağı üzerinde bağlantı tahmini probleminin bazı örnekleri aşağıda verilmiştir:

Fakhraei ve arkadaşları [6] yaptıkları çalışmada hastalık-ilaç etkileşimini incelemişlerdir. Bu çalışmada ilaçların umulmadık fayda ve zararlarını ortaya çıkarmayı amaçlamışlardır. Elde edilen sonuçlarla diğer olasılıksal yumuşak (soft) mantık kullanan yöntemlerden daha kesin ve etkin bir yöntem önerdiklerini kanıtlamışlardır.

Lakshmi ve Bhavani [7] yaptıkları çalışmayla bağlantı tahmini sorununu eldeki çözümler üzerinde geliştirme yapan bir algoritma önerdiler. Önerilen yöntem heterojen ağlarda bağlantı tahmini için paralel yöntemdir. Bench-mark veri setleri üzerinde bu algoritmaların uygulama sonuçları, çoklu ilişkili bağlantı tahmini performansı için anlamlı bir gelişme kaydettiğini göstermektedir.

Biuk-Aghai [8] yaptığı çalışma ile *wikipediadaki* makaleler ve diğer varlıklar arasındaki karmaşık ilişkileri 3 boyutlu olarak göstermek istemiştir. Bunun için de makalelerin ortak yazarlık ilişkilerinden faydalanmıştır. Bu makalelerin temel katkısı yayınlanan belgelerin benzerliklerini ortaya çıkarmaktır. Bu anlamda yapılan çalışma ile etkili sonuçlar elde edilmiştir.

Krömer ve arkadaşları [9] yaptıkları çalışmada Spektral Bölmeleme ve Karınca Koloni Optimizasyon (KKO) algoritmalarını kullanarak DBLP veri tabanında yazarlar arasındaki ilişkileri belirlemeye çalışmışlardır. Bu makale IT4 Yenilikler Merkezi Mükemmellik Projesi kapsamında hazırlanan projeyi gerçekleştirmede kullanılan veriler ile oluşturulmuştur.

Tsolakidis ve arkadaşları [10] yaptıkları çalışmada akademisyenlerin araştırma faaliyetleri sırasında yaptıkları iş birliğini araştırmak ve yazarlar arasındaki ilişkileri saptayıp görselleştirmek için bir yüksek öğretim kurumunu baz almışlardır. Yaptıkları çalışma sonucunda akademisyenler arasındaki bağın zayıf olduğunu görmüşlerdir.

Meng ve Kennedy [11] çalışmalarında Sidney Teknoloji Üniversitesinde ortak yazarlık ağını grup liderleri ve kümeleme açısından incelemişlerdir. Bu çalışmaya göre sosyal paylaşım ağlarını grup liderleri organize eder. Grup liderlerinin belirlenmesi ile küme sayıları hesaplanabilir. Yaptıkları uygulama diğer küme sayısı hesaplama uygulamaları ile kıyaslandığında oldukça başarılı sonuçlar vermektedir.

Zehnalova ve arkadaşları [12] yaptıkları çalışmada araştırma konuları açısından yazarların evrimi ve ortak yazarlık analizi üzerine çalışmışlardır. Bu çalışmada verilere kolay erişilebilmesi açısından DBLP veri tabanı kullanılmıştır. Ortak yazarlık hakkında genel bir analiz yapmış ve görselleştirmişlerdir.

Ahmedi [13] yaptığı çalışma ile ortak yazarlık ağında yazarların itibarını ölçmeye çalışmıştır. Bu amaçla PageRank ve AuthorRank metriklerini semantik web kural dilinde kullanan bir model önermiştir.

Meng ve Kennedy [14] yaptıkları diğer bir çalışma ile ortak yazarlık ağında verimli araştırma grupları oluşturabilecek ve en verimli bireyleri seçebilecek bir yapı oluşturmayı amaçlamışlardır. Bu amaç doğrultusunda Sidney Teknoloji Üniversitesi akademik işbirliğini incelemişlerdir.

Choobdar ve arkadaşları [15] farklı bilimsel alanların araştırma işbirliğine dayalı olarak karşılaştırılmasını amaçlamaktadır. Araştırma 29 farklı küçük motife dayalı 22 bilim alanında yapılmıştır. Yapılan çalışma ile bazı bilim alanlarının oldukça benzer olduğunu ve kolayca karşılaştırılabileceğini göstermektedir.

Song ve arkadaşları [16] Biyoinformatik dergisi ve BMC Biyoinformatik'in içerik ve ortak yazar şebekesinin benzerliğini analiz etmek için bir çalışma yapmışlardır. Yaptıkları çalışmada Biyoinformatik dergileri arasındaki örtüşmenin gittikçe arttığını ortak yazarlık ağı benzerliğine göre Biyoinformatik araştırmalarına daha fazla araştırma grubunun katıldığını gözlemlemişlerdir.

Krömer ve arkadaşları [17] yaptıkları çalışma ile önemli düğümler (yazarlar) tarafından uyarılan karmaşık ortak yazarlık ağının seçili alt ağlarının ölçeksizlik özelliklerini ve zaman içindeki bu özelliklerinin evrimini analiz etmişlerdir. Önemli yazarlar tarafından indüklenen alt ağlar farklı zamanlarda örneklenir ve güç yasası parametreleri analitik olarak

tahmin edilir. Çalışma, dinamik bir ortak yazarlık ağında güç yasası parametrelerinin ölçek-özgürlüğü ve evrimine ilişkin ampirik bir bakış sağlar.

Bento ve Takeda [18] yaptıkları çalışmada farklı işbirliği alanlarındaki iki veri kümesini kullanarak bilimsel işbirliğini ortaya çıkarabilecek ortak yazarlık ağlarında topluluk bulmak için bir yöntem geliştirmeyi amaçlamışlardır. Çalışmak için Japon araştırma kurumlarında bilgisayar bilimlerinin dünya çapında farklı işbirliği alanlarındaki araştırma konularını ele almışlardır.

Hoang ve arkadaşları [19] yaptıkları çalışmada ortak yazarlık ağında bir yazarın bir konuyu ne kadar çok tercih ettiğini veya tercih etmediğini analiz etmeye çalışmışlardır. Bu çalışma için ILPnet2 veri tabanı kullanılmıştır. Önerdikleri yöntemin doğruluğunu test etmek için de ROC analizi ve Regresyon Tahmini modelini kullanmışlardır. Önerdikleri yöntem ile oldukça başarılı sonuçlar elde etmişlerdir.

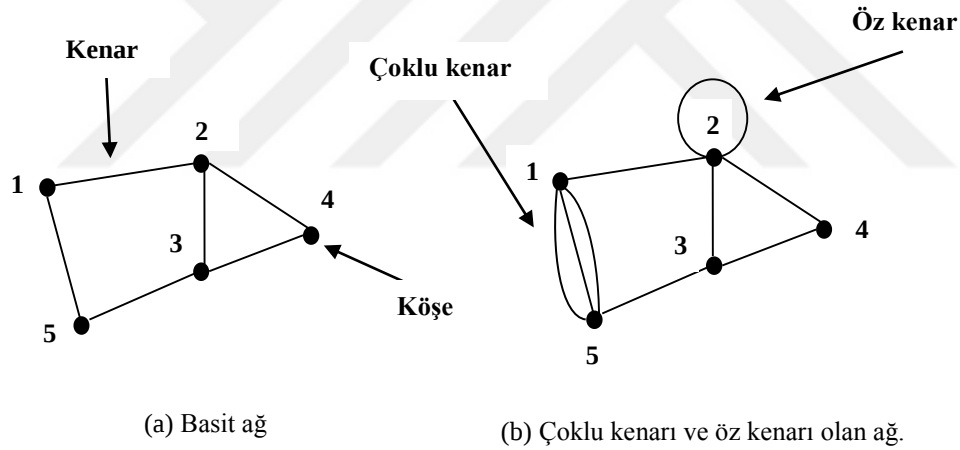
Huang ve arkadaşları [20] yaptıkları çalışmada zamanla değişen bağlantı ağırlık bilgilerinden yararlanan hibrid bir yaklaşım önermektedirler. Bu yaklaşımlarını ortak yazarlık ağında uygulamışlardır. Uygulamada elde ettikleri sonuçlar algoritmalarının oldukça iyi sonuçlar verdiğini göstermektedir.

2. AĞLARIN MATEMATİĞİ

2.1. Ağlar ve Temsilleri

Graf olarak nitelendirilen ağ yapısı, belirli noktalar kümesi ile bu noktaları birleştiren kenarlar tarafından temsil edilen yapının adıdır. Köşeler ve kenarlar, bilgisayar biliminde düğümler ve bağlantılar; fizikte devre elemanları ve elektrik kabloları; sosyolojide aktörler ve aralarındaki bağlar olarak adlandırılırlar [21].

Bir ağda belirli iki düğüm arasında birden fazla kenar bağlantısı varsa bu kenarların tamamı çoklu kenar olarak isimlendirilirler. Bir düğümden çıkan kenar yine düğümün kendisi ile sonlanıyorsa bu kenara da öz kenar adı verilir. Öz kenarı veya çoklu kenarı olmayan ağa basit ağ veya basit graf denir. Çoklu kenarı olan ağa çoklu graf, öz kenarı olan ağlara da özkenarlı ağ adı verilir [22].



Şekil 2.1. İki küçük ağ.

2.2. Komşuluk Matrisi

Bir ağı matematiksel olarak ifade etmek için farklı yollar vardır. n köşeli yönlendirilmemiş bir ağın köşeleri $1...n$ arası tam sayılarla etiketlenmiş olsun. Eğer (i,j) şeklinde, i ve j köşeleri arasındaki bir kenardan bahsedilirse, tüm kenarların listesini ve n değerini vererek tüm ağ ifade edilebilir. Örneğin Şekil 2.1.a için, $n=6$ olur ayrıca birbirine bağlı olan köşe çiftleri $(1,2)$, $(1,5)$, $(2,3)$, $(2,4)$, $(3,4)$, $(3,5)$ ve $(3,6)$ şeklindedir. Bu ifade şekline kenar listesi adı verilir.

Ağ için bir başka ifade şekli de komşuluk matrisidir [23]. Komşuluk matrisi ile bir ağ daha kolay ve pratik bir şekilde ifade edilebilir. Bir A grafının A_{ij} elemanları ile komşuluk matrisi denklem (2.1) deki gibi yazılabilir:

$$A_{ij} = \begin{cases} 1 & \text{eğer } i \text{ ve } j \text{ köşeleri arasında bir kenar varsa,} \\ 0 & \text{diğer durumlarda,} \end{cases} \quad (2.1)$$

Örneğin, şekil 2.1.a için komşuluk matrisi denklem (2.2) de:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (2.2)$$

şeklinde olur.

Yönsüz ağlarda komşuluk matrisi kullanılarak çoklu kenar ve özkenar da temsil edilebilir. Çoklu kenar için, A_{ij} elemanı 0 veya 1 yerine kenar sayısı değerini alır. Örneğin, i ve j köşeleri arasında 3 tane kenar varsa $A_{ij}=A_{ji}=3$ şeklinde ifade edilir. Öz kenarlarda durum daha farklıdır. i köşesinden kendisine tek bir öz kenar varsa matrisin A_{ii} köşegen elemanı 1 değil 2 değerini alır çünkü yönsüz ağlarda her öz kenarın 2 sonu vardır ve ikisi de aynı köşede son bulmaktadır. Buna göre şekil 2.1.b'nin komşuluk matrisi denklem (2.3) te şöyle gösterilir:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 3 \\ 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (2.3)$$

Bir de çoklu öz kenarlar olabilir (veya çok-öz kenar). Bu tür kenarlar ise komşuluk matrisinin eşdeğer köşegen elemanının kenar sayısının 2 katı ile ifade edilir.

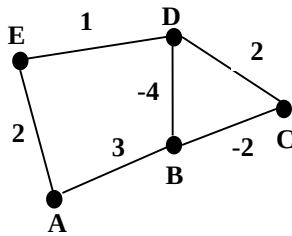
2.3. Ağırlıklı Ağlar

Bazı durumlarda ağların kenarlarını temsil etmek için onların kuvvetleri, ağırlıkları veya değerleri (genellikle gerçel sayılar kullanarak) ile ifade etmek oldukça kullanışlıdır [24]. Böylece İnternette kullanıcılar arasında akan veri miktarı, bir gıda ağında av-avcı etkileşiminde av ile avcı arasındaki toplam enerji akışı veya sosyal ağlarda aktörler (kullanıcılar) arasındaki bağlantı sıklığı ağırlık olarak kullanılabilir. Ağırlıklı ağların komşuluk matrisi oluşturulurken matris elemanları, eşdeğer ağ kenarının ağırlık değeri ile temsil edilir. Ağırlıklı bir ağ için örnek komşuluk matrisi denklem (2.4) te verilmiştir.

$$A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 0.5 \\ 1 & 0.5 & 0 \end{pmatrix} \quad (2.4)$$

Bazı durumlarda ağırlıklı ağlar ile çalışmaktansa çoklu ağlar ile çalışmak daha kullanışlıdır. Ağırlıklı ağın kenar ağırlıkları tamsayı ise bu iki ağ arası dönüşüm yapmak da kolaylaşır. Herhangi iki köşe arasında kenar ağırlığı 0 veya 1 dışında bir değer olursa ağırlıklı tek bir kenar yerine ağırlık değeri kadar kenarı olan çoklu kenar ile temsil etmek de mümkündür. Ters dönüşüm de yapılabilir. Bu dönüşümler özellikle ağın analizi için yararlı olabilir.

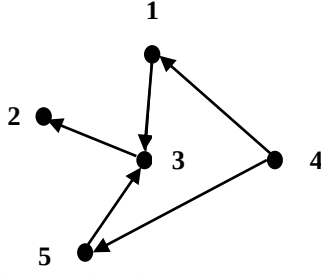
Ağırlıklı ağda ağırlıklar genelde pozitif sayılardır ama negatif olmaması için de hiçbir neden yoktur. Örneğin, sosyal ağ teorisinde insanlar arasındaki sosyal ilişkiler; arkadaşlık ve diğer samimi ilişkileri göstermek için pozitif, düşmanlık için negatif ağırlıklar ile temsil edilir.



Şekil 2.2. Ağırlıklı ağ

2.4. Yönlü Ağlar

Yönlü ağ veya yönlü graf kısaca digraf olarak adlandırılır. Her kenarın bir yöne sahip olduğu ağ türüdür. Bu tür ağların kenarlarına yönlü kenarlar adı verilir. Şekil 2.2’de yönlü bir ağ örneği gösterilmektedir:



Şekil 2.3. Yönlü ağ. Kenarları yönleri gösteren oklar ile yönlü bir ağ.

Yönlü ağlara örnek olarak; bir sayfadan diğerine yönlendirilerek çalıştırılan internet ağı, avdan avcıya gelen enerji ile gıda ağı, bir yayından diğerine alıntı yapılan atıf ağı gösterilebilir.

Yönlendirilmiş bir ağın komşuluk matrisi elemanları denklem (2.5) teki gibi hesaplanır:

$$A_{ij} = \begin{cases} 1 & j' \text{ den } i' \text{ ye bir kenar varsa} \\ 0 & \text{diğer durumlarda} \end{cases} \quad (2.5)$$

Burada kenarın yönüne dikkat edilmelidir. A_{ij} elemanı için kenarın yönü 2. indisten 1. indise doğrudur (j' den i' ye).

Şekil 2.2’deki ağın komşuluk matrisi denklem (2.6) daki gibi olur:

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (2.6)$$

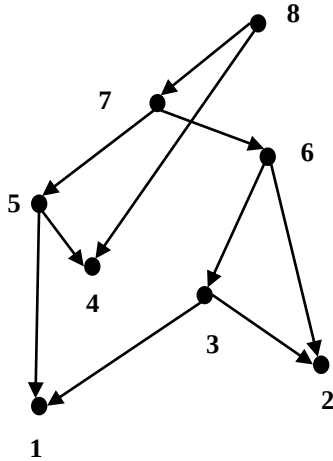
Matrisin simetrik olmadığına dikkat edilmelidir. Genellikle yönlü ağların komşuluk matrisi asimetrik olur.

Gerekli durumlarda yönsüz ağlar yönlü ağlara dönüştürülebilir. Bunu yaparken yönsüz kenar yerine birbirinin tersi yönlü iki tane kenar kullanılır. Böyle bir ağ için komşuluk matrisi orijinal yönsüz ağ ile aynıdır [25].

Yönlü ağlarda da yönsüz ağlar gibi 1'den büyük değerler ile çoklu kenarlar ve sıfır olmayan köşegen değerleri ile öz kenarlar temsil edilebilir. Burada dikkat edilmesi gereken nokta, bir öz kenar temsil edilecekse yönsüz kenarlarda olduğu gibi 2 ile değil 1 ile temsil edilir çünkü yönlü bir kenarın bir tane sonu vardır.

2.4.1. Çevrimsiz Yönlü Ağlar

Yönlü bir ağda çevrim, belli bir köşeden başlanarak geçilen kenarlar sonucunda yine aynı köşeye gelinmesi durumudur. Bazı yönlü ağlarda çevrim bulunmayabilir. Bu ağlara çevrimsiz yönlü ağlar denir [26]. Öz kenarlar da çevrim olarak kabul edilir dolayısıyla çevrimsiz ağlar aynı zamanda hiçbir öz kenarı olmayan ağlardır. Çevrimsiz bir ağ için en klasik örnek makalelerin alıntı ağlarıdır. Bir makale yazılırken, eğer öncesinde yayınlanmış ise başka bir makaleyi alıntı olarak verebiliriz. Ağ grafiksel olarak şekil 2.3'teki gibi tasvir edilebilir. Bu ağda köşeler zaman sıralıdır. Böylece alıntıyı temsil eden tüm kenarlar aşağıya doğru işaret eder.



Şekil 2.4. Çevrimsiz yönlü ağ.

Yönlü ve çevrimsiz bir ağın komşuluk matrisi yukarıda anlatıldığı gibi çizilirse tüm kenarlar bir yöne bakar ve köşeler zaman sıralı olarak eklenirse komşuluk matrisinin

köşegeninin üst kısmı sıfır olmayan elemanlara sahip olur. Buna üst üçgen denir. Örnek bir ağın komşuluk matrisi denklem (2.7) deki gibi olur:

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (2.7)$$

r uzunluğunda bir çevrimin L_r toplam sayısı denklem (2.8) de şöyle hesaplanır:

$$L_r = \sum_{i=1}^n K_i^r \quad (2.8)$$

K_i komşuluk matrisinin öz değeridir. Eğer ağda çevrim varsa $L_r > 0$ olur. Tüm öz değerleri sıfır olan matrislere sıfır kuvvetli matrisler adı verilir.

2.5. İkili Ağlar

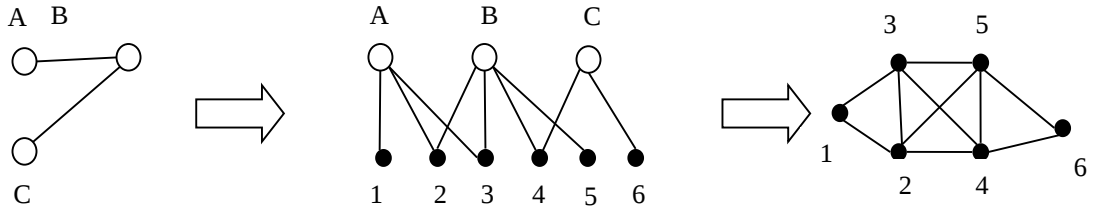
İkili ağlarda iki tür köşe vardır. Biri orijinal köşeleri diğeri ise bu köşelerin ait oldukları grupları temsil eder [27].

İkili ağ için komşuluk matrisi, insidans matrisi olarak adlandırılan dikdörtgen bir matristir. n, ağdaki katılımcıların sayısı ve g grupların sayısını temsil ederse B insidans matrisi g x n boyutlarında B_{ij} elemanlarına sahip bir matris olur ve bu matrisin elemanları denklem (2.9) da şöyle bulunur:

$$B_{ij} = \begin{cases} 1 & \text{eğer } j \text{ köşesi } i \text{ grubuna ait ise} \\ 0 & \text{diğer durumlarda} \end{cases} \quad (2.9)$$

Belirli bir ikili ağda tek tip köşeler arasındaki bağlantıları göstermek için iki modlu formdan tek modlu forma dönüşüm yapmak gerekir. Örnekte verildiği gibi yine film ağında film ve oyuncuların durumları düşünülürse, aynı filmde rol alan oyuncuları göstermek için kenarlar aynı filmde rol almış oyuncular arasında bağlantı kurmalıdır. Yine filmler

arasındaki bağlantı da bir oyuncunun rol aldığı birden fazla film varsa bu filmleri birbirine bağlar. Şekil 2.4, bu durumu temsil eden ikili bir ağı göstermektedir.



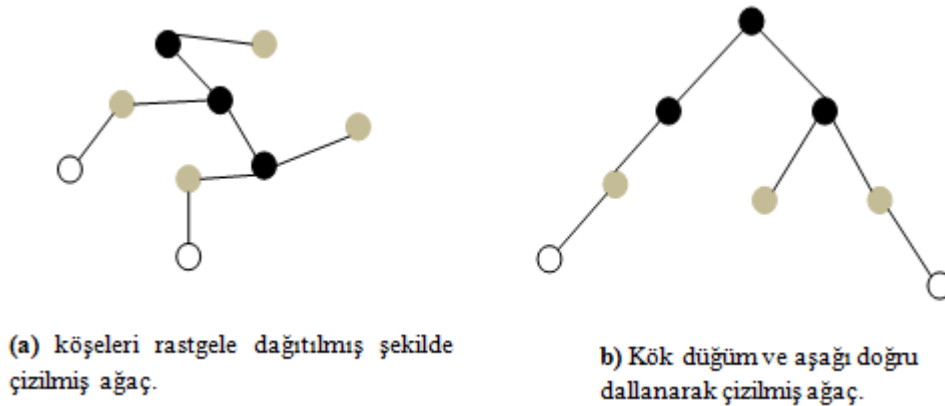
Şekil 2.5. İkili ağın iki tane tek modlu tasarımı.

Şeklin orta kısmı, tek tip üç köşe (içi boş çember ve A-C arası etiketli) ve yedi diğer köşeye (içi dolu çember ve 1-6 arası etiketli) sahip bir ikili ağı göstermektedir. Üst ve alt kısımlar ise köşelerin iki kümesi üzerinde ağın tek modlu yapılarını gösterir.

Tek modlu tasarım çok kullanışlı bir yöntemdir fakat ikili moddan tek moda dönüşüm esnasında ağ ile ilgili birçok bilginin göz ardı edilmesine sebep olur. Yine yukarıdaki örnekten yola çıkılacak olursa film ağı tek moda dönüştürüldüğünde hangi oyuncuların hangi filmlerde rol aldığı bilgisi ikili ağda mevcut iken tek modlu ağda bu bilgiye erişilemez.

2.6. Ağaçlar

Ağaç, kapalı döngüler içermeyen, yönsüz ağlarda bir bağlantıdır. “Bağlantı” ile söylenmek istenen şey ağdaki her köşeye tüm diğerlerinden mutlaka erişilebildiğidir. Bir ağ birbirinden bağımsız iki veya daha fazla parçadan meydana gelebilir. Eğer bu parçaların hiçbir bölümü döngü içermiyorsa ağaç olarak adlandırılır [28].



Şekil 2.6. Aynı ağacın iki farklı çizimi.

Şekil 2.5’te bir ağacın iki farklı çizimi gösterilmiştir. Ağaçlar genellikle Şekil 2.5 b’de görüldüğü gibi köklü olarak çizilirler. En üstte bir kök düğümü ve aşağıya doğru bir dallanma yapısı ile oluşturulurlar. Sadece bir üstteki köşeye bağlı olan, alttaki köşelere yaprak adı verilir. Topolojik olarak bir ağacın özel bir kökü yoktur fakat bazı ağaçlarda belirli bir kök düğüm atanması için özel nedenler vardır.

Ağaçların bilinen en önemli özelliklerinden biri; n köşeli ağacın daima $n-1$ kenarlı olmasıdır. Ağaçlar döngü içermediğinden ve dolayısı ile iki köşe arasında bir kenardan fazla kenar olamayacağından kenar sayısı daima köşe sayısından bir eksiktir.

2.7. Ağlarda Derece

Grafta bir köşenin derecesi ona bağlı olan köşelerin sayısıdır [29]. i köşesinin derecesi k_i olarak gösterilir. n köşeli yönsüz bir grafta derece, komşuluk matrisleri ile denklem (2.10) da şöyle yazılabilir:

$$k_i = \sum_{j=1}^n A_{ij} \quad (2.10)$$

Yönsüz bir grafta her kenarın iki ucu vardır ve eğer toplamda m kenar varsa kenar ucu sayısı $2m$ olur. Kenar ucu sayısı tüm köşelerin dereceleri toplamına eşittir. Dolayısıyla denklem (2.10), denklem (2.11) deki gibi olur:

$$2m = \sum_{i=1}^n k_i \quad (2.11)$$

veya denklem (2.12) deki gibi olur:

$$m = \frac{1}{2} \sum_{i=1}^n k_i = \frac{1}{2} \sum_{i,j} A_{ij} \quad (2.12)$$

Yönsüz bir grafta c köşesinin ortalama derecesi denklem (2.13) te şöyle yazılır:

$$c = \frac{1}{n} \sum_{i=1}^n k_i \quad (2.13)$$

ve denklem(2.11) ile bu sonucu birleştirirsek sonuç denklem (2.14) teki gibi olur:

$$c = \frac{2m}{n} \quad (2.14)$$

Basit bir grafta (çoklu kenarı veya öz kenarı olmayan) kenarların maksimum sayısı $\binom{n}{2} = \frac{1}{2}n(n-1)$ olur. Bir grafın ρ yoğunluğu gerçekten mevcut olan bu kenarın, maksimum olabilecek kenar sayısına bölümüdür. Dolayısıyla sonuç denklem (2.15) teki gibi yazılır:

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1} \quad (2.15)$$

yoğunluk aralığı kesinlikle $0 \leq \rho \leq 1$ dir. Ağların çoğunda yoğunluk $c = \rho / n$ olarak kabul edilir.

Yönlü bir ağda köşe derecesi daha da karmaşıktır. Yönlü bir ağda her köşe iki dereceye sahiptir. Giriş derecesi; bir köşeye bağlı olan, gelen kenarların sayısıdır ve çıkış derecesi, giden kenarların sayısı olarak adlandırılır. Yönlü ağlarda komşuluk matrisi elemanları türünden giriş ve çıkış dereceleri denklem (2.16) daki gibi yazılır:

$$k_i^{in} = \sum_{j=1}^n A_{ij} \quad k_j^{out} = \sum_{i=1}^n A_{ij} \quad (2.16)$$

Yönlü ağda m kenar sayısı tüm köşelere gelen uçların toplam sayılarına eşittir veya giden uçların toplam sayılarına eşittir dolayısıyla bu eşitlik denklem (2.17) de gösterilmiştir:

$$m = \sum_{i=1}^n k_i^{in} = \sum_{j=1}^n k_j^{out} = \sum_{ij} A_{ij} \quad (2.17)$$

Böylece yönlü ağın c_{in} ortalama gelen derecesi ve c_{out} ortalama giden derecesi eşittir. Bu durumda sonuç denklem (2.18) deki gibi olur:

$$c_{in} = \frac{1}{n} \sum_{i=1}^n k_i^{in} = \frac{1}{n} \sum_{j=1}^n k_j^{out} = c_{out} \quad (2.18)$$

Kolaylık olması açısından gelen ve giden derecelerinin ikisini de c ile göstereceğiz. Denklem (2.17) ve denklem (2.18) i birleştirirsek denklem (2.19) u elde ederiz:

$$c = \frac{m}{n} \quad (2.19)$$

3. ÖLÇÜMLER VE METRİKLER

Bir ağın yapısı biliniyorsa ağ topolojisinin belirli özelliklerini yakalayan yararlı niceliklerin çeşitleri veya ölçütleri hesaplanabilir. Tarif edilen yöntemler günümüzde yaygın olarak bilgisayar bilimleri dahil olmak üzere fizik ve biyoloji gibi sosyal bilimlerde dışında kalan alanlarda kullanılmakta ve temel ağ araç kutusunun önemli bir parçasını oluşturmaktadır.

3.1. Derece Merkeziliği

Bir ağda en basit merkezilik ölçütü, köşenin derecesidir, diğer bir deyişle köşeye bağlı kenarların sayısıdır. Derecenin, bazen merkezilik ölçütü olarak kullanıldığının vurgulanması için sosyal ağ literatüründe derece merkeziliği olarak da adlandırılır. Yönlü ağlarda köşeler, hem giriş derecesi hem de çıkış derecesine sahiptir ve ikisi de uygun koşullarda merkezilik ölçütü olarak yararlı olabilir [30].

Derece merkeziliği basit bir merkezilik ölçütü olmasına rağmen çok aydınlatıcı olabilir. Sosyal ağda, diğer birçok kişi ile bağlantısı olan bireylerin daha fazla etkiye sahip olduğunu, bilgiye daha rahat eriştiğini veya daha az bağlantıya sahip olan bireylerden daha prestijli olabileceğini söylemek mümkündür. Sosyal olmayan bir ağ örneği olarak, bilimsel makalelerin değerlendirilmesinde atıf sayılarının kullanımı verilebilir. Bir makalenin, basitçe atıf ağında kendi giriş derecesi olan diğer makalelerden aldığı atıf sayısı, yayının etkili olup olmadığı ve bilimsel araştırmanın etkisini yargılamak için metrik olarak kullanılıp kullanılmadığının ham ölçüsünü verir.

3.2. Özvektör Merkeziliği

Basit derece merkeziliğinin doğal bir uzantısı özvektör merkeziliğidir. Birçok durumda bir ağdaki köşenin önemi, kendileri için önemli olan diğer köşelerle bağlantılarına bağlı olarak artırılır. Bu, özvektör merkeziliğinin arka planıdır. Her komşu için köşeleri bir nokta ile ödüllendirmek yerine özvektör merkeziliği komşularının puanlarının toplamına orantılı olarak her köşeye puan verir [31].

Her bir i köşesinin x_i merkeziliği hakkında bazı başlangıç varsayımları yapılırsa, örneğin, tüm i ler için $x_i=1$ ayarlanarak başlanabilir. Açıkçası bu merkezilik ölçütü için kullanışlı değildir fakat daha iyi bir ölçüt olan, i nin merkeziliklerinin toplamı olarak tanımlanan x'_i 'yi hesaplamak için kullanılabilir, böylece denklem (3.1) deki eşitlik elde edilir:

$$x'_i = \sum_j A_{ij}x_j \quad (3.1)$$

elde edilir. A_{ij} komşuluk matrisinin bir elemanıdır. Bu ifade matris gösteriminde $x'=Ax$ olarak yazılır, x ; x_i elemanları ile bir vektördür. Daha iyi tahminler yapmak için bu süreci tekrarlanarak, t adım sonra merkeziliklerin vektörü olan $x(t)$ denklem (3.2) deki gibi elde edilir:

$$x(t) = A^t x(0) \quad (3.2)$$

Özvektör merkeziliği büyük de olabilir çünkü köşenin çok fazla komşusu veya önemli komşuları olabilir. Sosyal ağda bir birey bu ölçüme göre önemli olabilir çünkü birçok insanı (o insanlar onlar için önemli olmasa bile) veya yüksek mevkilerde birkaç kişiyi tanıyor olabilir.

Köşelerin özvektör merkezilikleri negatif olamaz. Bunu görmek için, sadece başlangıç vektörü $x(0)$ 'ın, negatif olmayan elemanlara sahip olduğunda ne olduğuna bakılmalıdır. Hiçbir komşuluk matrisi elemanı da negatif olmadığından A nın negatif elemanları vektör için tanımlanamaz ve denklem (3.2) de $x(t)$ nin hiçbir elemanı negatif olmamalıdır.

Teoride özvektör merkeziliği yönsüz ve yönlü ağlar için hesaplanabilir. Ancak yönsüz ağlarda daha iyi çalışır. Yönlü ağda bazı karışıklıklar ortaya çıkar.

3.3. Katz Merkeziliği

Merkezilik ölçütü ilk olarak Katz tarafından 1953 yılında önerildiğinden Katz merkeziliği olarak adlandırılır [32]. Önceki bölümlerdeki merkeziliklerde karşılaşılan problemler için bir çözüm şudur: her köşeye “karşılıksız” küçük miktarda merkezilik verilir. Diğer bir deyişle, denklem (3.3) teki gibi tanımlanır:

$$x_i = \alpha \sum_j A_{ij} x_j + \beta \quad (3.3)$$

α ve β pozitif sabitlerdir. İlk terim, i 'yi bağlayan köşelerin merkeziliklerini toplayan normal özvektör merkeziliği terimidir ve ikinci terim bağımsız bir parçadır, tüm köşeleri kapsayan sabit ekstra terimdir. İkinci terim eklenerek, sıfır giriş derecesi ile köşeler yine β merkeziliği elde eder ve onlar bir kez sıfır olmayan merkeziliğe sahip olur. Bu, diğer birçok köşeden gelen kenarları olan bir köşenin yüksek merkeziliğe sahip olacağı anlamı taşır [33].

Matris formunda, denklem (3.3), denklem (3.4) tki gibi yazılır:

$$x = \alpha Ax + \beta 1 \quad (3.4)$$

$1, (1, 1, 1, \dots)$ vektörüdür. x yeniden düzenlenerek, $x = \beta(I - \alpha A)^{-1} \cdot 1$ bulunur. Söylendiği gibi, normalde merkezin mutlak büyüklüğü önemsizdir, sadece hangi köşelerin yüksek veya düşük merkezilik değerine sahip olduğu ile ilgilenilir dolayısıyla genel β çarpanı önemsizdir. Kolaylık sağlamak için $\beta=1$ ayarlanır, böylece yeni denklem; denklem (3.5) teki gibi elde edilir:

$$x = (I - \alpha A)^{-1} \cdot 1 \quad (3.5)$$

3.4. Sayfa Derecesi

Katz merkeziliği istenmeyebilecek bir özelliğe sahip olabilir. Yüksek Katz merkeziliği ile bir köşenin diğer köşelere giden kenarlarının olduğu bir durumda diğer köşeler de yüksek merkeziliğe sahip olurlar. Bir milyon köşeye giden kenarı olan yüksek merkezilikli bir köşe, bir milyonun tamamına yüksek merkezilik verir. Bunun her zaman uygun olmadığı öne sürülebilir. Prestijli bir köşeden gelen kenarı alma sayesinde kazanılan merkezilik, diğerleri ile paylaşılacak sureti ile seyreltilir. Mesela, ünlü Yahoo! web dizini herhangi bir web sayfasından bir bağlantı içerebilir fakat milyonlarca diğer sayfadan da bağlantı içerebilir. Yahoo! önemli bir web sitesidir ve herhangi bir mantıklı ölçüm ile yüksek merkeziliğe sahiptir. Yahoo! sitesinin yüksek merkeziliği diğer sitelerle paylaştırıldığında seyreltilmiş olacaktır ve herhangi bir web sayfasının merkeziliğine katkısı çok küçük olacaktır çünkü sayfa, milyonların sadece biridir. Bunun için onların merkeziliklerinin, onların çıkış derecelerine bölünmesi ile orantılı olan ağırlık komşularından türetilen merkezilik olan Katz

merkeziliğinde yeni bir varyasyon tanımlamaya izin verilebilir. Sonra birçok köşeye giden kenarı olan köşeler, diğer köşelerin her birine kendi merkeziliği yüksek olsa bile merkeziliğinin küçük bir miktarını aktarır [34].

Matematiksel olarak bu merkezilik denklem (3.6) daki gibi tanımlanır:

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta \quad (3.6)$$

Bu merkezilik ölçütü, ticaret ünvanı Google web aramaları şirketi tarafından verilen, kendi web sıralama teknolojisinin merkez parçası olarak kullanılan PageRank olarak bilinir [35]. Google web arama motorunun amacı, metin sorgularına yanıt içinde sayfaların önceden monte edilmiş dizinden yararlı web sayfalarının listesini oluşturmaktır. Metin eşleştirme gibi nispeten basit kriterler kullanılarak verilen bir sorgu ile eşleşen sayfalar için ilk dizini (indeks) arayarak bunu yapar ve sonra biri PageRank olan bileşenlerin kombinasyonuna dayanan puanlara göre cevaplar sıralanır.

Tablo 3.1. Dört tane özvektör merkeziliği.

Merkezilik Ölçümleri	Sabit terim ile	Sabit terimsiz
Çıkış derecelerine bölünerek	$x = D(D - \alpha A)^{-1} \cdot 1$ PageRank	$x = AD^{-1}x$ derece merkeziliği
Bölme olmadan	$x = (I - \alpha A)^{-1} \cdot 1$ Katz merkeziliği	$x = K_1^{-1}Ax$ Özvektör merkeziliği

Tablo 3.1’de, farklı matris merkezilik ölçütlerinin bir özeti görülmektedir. Bu ölçütlerin biri kullanılmak istenirse ve şaşırtıcı birçok alternatif bulmak istenirse, özvektör merkeziliği ve PageRank muhtemelen başlangıçta odaklanmak için iki önemli ölçüt olur. Onlar bu türün iki en sık kullanılan ölçütleridir. Katz merkeziliği geçmişte yaygın olarak kullanılmıştır ancak son çalışmalarda daha az tercih edilmiştir, sabit terimsiz PageRank ölçümü yönsüz ağlarda derece merkeziliği ile aynı olurken ve yönlü olanlar için kullanılmaz.

3.5. Yakınlık Merkeziliği

Merkeziliğin tamamen farklı bir ölçütü, bir köşeden diğerine olan uzaklık anlamına gelen yakınlık merkeziliği ile sağlanmaktadır. Örneğin, d_{ij} i den j ye jeodezik yol (en kısa yol) uzunluğudur yani yol boyunca geçilen kenarların sayısı anlamına gelir. Sonra i den j ye

jeodezik mesafenin anlamı, ağda tüm j köşeleri üzerindeki ortalama ve denklem (3.7) deki gibi hesaplanır:

$$l_i = \frac{1}{n} \sum_j d_{ij} \quad (3.7)$$

Bu nicelik ortalamada sadece kısa bir jeodezik mesafe ile diğerlerinden ayrılan köşeler için düşük değerler alır. Bu tür köşeler bilgiye daha iyi erişebilir. Sosyal bir ağda, toplumda diğerlerine daha düşük ortalama mesafesi olan bir kişi diğerlerinin fikirlerine, yüksek ortalama mesafesi olan birinden daha hızlı ulaşır. Yakınlık merkeziliği, merkeziliğin çok doğal bir ölçütüdür bu nedenle sosyal ve diğer ağ çalışmalarında sıklıkla kullanılır [21].

3.6. Arasındalık Merkeziliği

Merkeziliğin çok farklı bir konsepti, diğer köşeler arasındaki yollarda bulunan köşelerin ölçümlerini yapan arasındalık merkeziliğidir. Arasındalık fikri 1977 de Freeman kendini işaret etmemesine rağmen [36] Freeman'a atfedilmiştir. Bu fikir, bağımsız olarak birkaç yıl önce Anthonisse [37] tarafından yayınlanmamış teknik raporda önerilmiştir.

Kenarlar boyunca köşeden köşeye akan bir şey ile bir ağ olduğunu varsayalım. Mesela, bir sosyal ağda mesajlar, haberler, bilgi ya da birinden diğerine akan söylentilere örnek olabilir.

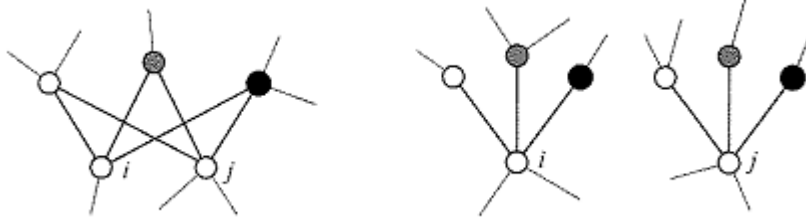
Yüksek arasındalık merkezilikli köşeler, bir ağ içinde diğerleri arasında geçen bilgiler üzerinde kendi kontrolleri sayesinde önemli bir etkiye sahip olabilirler. İletişim senaryomuzda en yüksek arasındalığa sahip köşeler, iletilerin en fazla sayıda geçenleridir ve eğer bu köşeler, onlar geçerken söz konusu iletileri görmek için alırsa veya ileti boyunca geçiş için değer alırsa onlar ağ içinde kendi konumlarından çok fazla güç alırlar. En yüksek arasındalık köşeler, kaldırıldığında diğer köşeler arasındaki iletişimi bozacak olanlardır çünkü onlar iletilerin en çok tercih ettikleri yollardır. Gerçek dünyada tüm köşeler aynı iletişim frekansına sahip değildir ve çoğu durumda iletişim en kısa yoldan yapılmaz. Yine de, arasındalık merkeziliği diğerleri arasındaki bilgi akışı üzerinde köşelerin etkisi için yaklaşık olarak rehberlik edebilir.

3.7. Benzerlik

Sosyal ağ analizinde başka merkezi bir kavram, köşeler arasındaki benzerliktir. Bir ağda köşeler hangi açılardan benzer olabilir ve benzerlik nasıl ölçülebilir? Belirli bir ağda hangi köşeler birbirine en benzerdir? Bu gibi soruların yanıtları sosyal ağlar, bilgi ağları ve diğerlerinde köşelerin türleri ve ilişkilerini ayırmada yardımcı olabilir [38]. Mesela, belirlenen bir sayfaya benzer web sayfalarının listesinin yararlı olduğu düşünülebilir. Aslında, çeşitli web arama motorları zaten böyle bir özellik sağlar: “benzer sayfalar için buraya tıklayın”.

Benzerlik çok farklı şekillerde belirlenebilir ve çoğunun ağlar ile ilgisi yoktur. Örneğin, ticari partner ve çöpçatanlık hizmeti insanların ilgi alanlarını, arka planlarını, sevdikleri veya sevmedikleri şeyleri tanımlamalarını kullanarak insanları başkaları ile eşleştirmeye çalışır. Aslında, bu hizmetler insanların özelliklerine dayalı, insanlar arasındaki benzerlik ölçütlerini hesaplar.

Ağ benzerliğinin ölçütünün inşası için iki temel yaklaşım vardır. Onlar, yapısal denklik [39] ve düzenli denklik [40] olarak adlandırılırlar. İsimler oldukça mantıksız ama temsil ettikleri fikir oldukça basittir. Bir ağda eğer aynı ağ komşularını paylaşıyorlarsa iki köşe yapısal dengittir. Şekil 3.1a da, i ve j köşeleri arasındaki yapısal dengliği gösteren taslak gösteriliyor-bu durumda, her ikisinin de paylaşılmayan komşuları olmasına rağmen aynı üç komşunun iki köşe arasında paylaşıldığı görülür.



(a) Yapısal denklik

(b) Düzenli denklik

Şekil 3.1. Denklik Örnekleri [9]

Düzenli dengliğin çözümü daha zordur. İki düzenli deng köşe mutlaka aynı komşuları paylaşmaz ama kendilerine benzer komşuları vardır. Farklı üniversitelerdeki iki tarih öğrencisi ortak arkadaşına sahip olmayabilir ama her ikisi de diğer tarih öğrencilerinin çoğunu, tarih öğretmenleri ve benzeri bilir, bu anlamda hala benzerdirler. Benzer şekilde, iki farklı

şirkette iki CEO ortak arkadaşlara sahip olmayabilirler ama onlar sırası ile CFO, CIO, yönetim kurulu üyeleri, şirket başkanı ve benzeri profesyonel bağları anlamında benzerdirler. Düzenli denklik Şekil 3.1b’de gösterilmiştir.

3.7.1. Kosinüs Benzerliği

Belki yapısal denkliğin en basit ve açık ölçümü iki köşenin sadece sahip olduğu ortak komşularının sayısını saymak olacaktır. Yönsüz bir ağda i ve j köşelerinin ortak komşularının sayısı n_{ij} denklem (3.8) de şöyle verilir:

$$n_{ij} = \sum_k A_{ik}A_{kj} \quad (3.8)$$

A^2 ’nin ij. elemanı olan değere karşılık gelir.

Ancak, iki köşe için ortak komşuları basitçe saymak çok iyi bir benzerlik ölçümü değildir. Üç ortak komşusu olan iki köşe çok mu az mı benzerdir? Bilindiği kadarıyla köşelerin derecelerini veya diğer köşe çiftlerinin paylaştığı kaç ortak komşuları olduğunu söylemek zordur. Bir strateji, basit bir grafta iki köşenin maksimum ortak komşularının sayısı olduğundan n elemanlı ağda köşelerin toplam sayılarına bölmek olabilir. (teknik olarak maksimum aslında n-2 dir fakat n geniş olduğunda fark küçük olur.) Ancak bu, düşük dereceli köşeleri haksız yere cezalandırmak olur: eğer köşenin derecesi üç ise diğer köşe ile ortak en fazla üç komşusu olabilir ama eğer n böleni çok geniş olursa iki köşe hala bir benzerlik değeri alacak. Daha iyi bir ölçüm köşenin değişen derecelerine izin verir. Böyle bir ölçüm kosinüs benzerliğidir, bazen Salton un kosinüsü olarak da adlandırılır [41].

Geometride, x ve y vektörlerinin iç veya nokta ürünü, $x \cdot y = |x||y|\cos\theta$ olarak verilir. $|x|$, x in büyüklüğüdür ve θ , iki vektör arasındaki açıdır. Yeniden düzenlenirse, açının kosinüsü denklem (3.9) daki gibi yazılabilir:

$$\cos\theta = \frac{x \cdot y}{|x||y|} \quad (3.9)$$

Salton [41], iki vektörün komşuluk matrisinin i. ve j. satırlarını (veya sütun) kabul etmeyi ve benzerlik ölçüsü olarak aralarındaki kosinüs açısını önerdi. İki satırın nokta

ürününün yönsüz ağlar için basitçe $\sum_k A_{ik}A_{kj}$ olduğu kaydedilir. Bu, benzerliği verir ve matematiksel ifadesi denklem (3.10) daki gibidir:

$$\sigma_{ij} = \cos\theta = \frac{\sum_k A_{ik}A_{kj}}{\sqrt{\sum_k A_{ik}^2}\sqrt{\sum_k A_{jk}^2}} \quad (3.10)$$

Eğer bir veya iki köşe sıfır dereceye sahipse Kosinüs benzerliği teknik olarak tanımsızdır fakat genel kabulün bu durumda $\sigma_{ij} = 0$ olduğu söylenir. Kosinüs benzerliği, benzerlik ölçütü için doğal bir ölçek sağlar. Bu değer daima 0-1 aralığındadır. 1 in kosinüs benzerliği, iki köşenin aynı komşulara sahip olduğunu gösterir. Sıfırın kosinüs benzerliği hiç aynı komşuya sahip olmadığını gösterir. Pozitif terimlerin toplamı olarak, kosinüs negatif olsa da Kosinüs benzerliği asla sıfır olamaz [42].

3.7.2. Pearson Korelasyon Katsayıları

Ortak komşularının sayısını normalize etmek için alternatif bir yol, komşularını rastgele seçen köşeleri olan ağdaki sayıyı, beklenen değerle karşılaştırmaktır. Tartışmanın burası Pearson korelasyon katsayısını işaret eder [43].

i ve j köşelerinin sırası ile k_i ve k_j derecelerine sahip oldukları varsayılın. Bu köşeler kaç ortak köşeye sahip olurlar? Bunu hesaplamak eğer onlar, komşularını tamamen rastgele seçerse basittir. i köşesinin komşularını ona açık olan n olasılıktan rastgele seçtiği varsayılın ve j köşesi benzer şekilde komşularını rastgele seçsin. j nin seçtiği ilk komşu için iki seçenektan birini seçme olasılığı k_i/n olur ve sonraki seçimler de benzer şekildedir. (büyük ağ için çok küçük ihtimal olduğundan aynı komşuyu iki kez seçme olasılığı ihmal edilir.) Sonra toplamda iki köşe arasındaki ortak komşuların beklenen sayısı k_j veya $k_i k_j / n$ olacaktır.

İki köşe arasındaki benzerliğin makul ölçüsü olarak, eğer komşularını rastgele seçselerdi, onların sahip oldukları ortak komşularının gerçek sayısının eksi değerinde olması beklenirdi. Bu durumun matematiksel ifadesi denklem (3.11) de şöyle verilir:

$$\begin{aligned} \sum_k A_{ik}A_{jk} - \frac{k_i k_j}{n} &= \sum_k A_{ik}A_{jk} - \frac{1}{n} \sum_k A_{ik} \sum_l A_{jl} \\ &= \sum_k A_{ik}A_{jk} - n \langle A_i \rangle \langle A_j \rangle \end{aligned}$$

$$\begin{aligned}
&= \sum_k [A_{ik}A_{jk} - \langle A_i \rangle \langle A_j \rangle] \\
&= \sum_k (A_{ik} - \langle A_i \rangle) (A_{jk} - \langle A_j \rangle) \quad (3.11)
\end{aligned}$$

$\langle A_i \rangle$, komşuluk matrisinin i. satırının elemanlarının $n^{-1} \sum_k A_{ik}$ ortalamasını ifade eder. Eğer i ve j nin ortak komşularının sayısı tam olarak rastgele şansa dayanarak beklenen değerse, denklem (3.11) sıfır olacaktır. Eğer pozitif ise, i ve j şans ile ikisi arasındaki benzerliğin bir göstergesi olarak beklediğimizden daha çok ortak komşuya sahiptir. Denklem (3.11) negatif de olabilir bu durumda farklılığın olası işareti olarak i ve j beklediğimizden daha az komşulara sahiptir.

Pearson korelasyon katsayısı benzerlik ölçümünde yaygın olarak kullanılmaktadır. Eğer ağdaki bağlantılar rastgele oluşsaydı köşelerden ne beklendiği ile ilgili benzer veya farklı durumları belirtmek için yararlı olabilirdi.

4. SOSYAL AĞ ANALİZİ

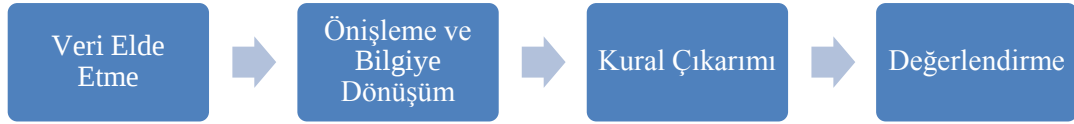
Sosyal ağlar, analiz edilebilmesi bakımından verilerin en rahat elde edilebildiği ve aktörler ile aktörler arası ilişkilerin en rahat modellenenebildiği yapılardır. Sosyal ağlarda aktörler ve aktörler arasındaki ilişkilerin yanı sıra ağın karakteristik yapısının da bilinmesi bu ilişkilerin analizi noktasında büyük kolaylık sağlamaktadır. Sosyoloji, psikoloji, ekonomi, matematik ve daha birçok alanda yoğun bir şekilde kullanılmakta olan sosyal ağ analizinde çoğunlukla aktörler arasındaki bağlantı ve ilişkiler üzerine çalışmalar yapılmaktadır.

Sosyal ağ analizinde üç temel faktör bulunmaktadır. Bunlar: aktör, ilişki ve bağ olarak adlandırılır. Aktörler sosyal ağlarda yer edinmiş kişi kurum veya nesnelerdir. Graf teorisinde bunlar düğüm olarak nitelendirilir. İlişkiler aktörler arasındaki yönlü veya yönsüz münasebetler olarak nitelendirilebilir. Örnek verecek olursak: Facebook, Twitter, Instagram gibi sosyal paylaşım sitelerinde aktörler kullanıcılar ilişkiler ise kullanıcılar arasındaki arkadaşlık veya takip ilişkileri olarak nitelendirilebilir. Bağlar ise güçlü ve zayıf olmak üzere iki şekildedir. Güçlü bağ: yakın ve aktif ilişkileri; zayıf bağ ise seyrek ilişkileri ifade eder [44].

4.1. Sosyal Ağ Analiz Aşamaları

Sadece ilk 5 sırada yer alan sosyal ağ servislerinin verileri göz önüne alındığında analiz için ele alınabilecek verilerin ne kadar büyük olduğu görülebilmektedir. Bu verilerin tamamının kullanılarak analiz yapılması işlemi hem gereksiz verilerle uğraşılması hem de çok uzun sürmesi nedeni ile zorlaşmaktadır. Bu durumun önüne geçebilmek için bu verinin ayıklanması ve veriler arasından sadece analiz için ihtiyaç olan verilerin alınıp geri kalan verilerin yok sayılması hem işlemi hızlandıracak hem de işlem yoğunluğunu azaltarak doğru sonuca ulaşma konusunda yardımcı olacaktır.

Analiz işleminin en doğru ve en hızlı şekilde gerçekleşebilmesi için analiz işlemi 4 safhada gerçekleştirilmektedir. Bu adımlar veri elde etme, ön işleme ve bilgiye dönüşüm, kural çıkarımı ve değerlendirme [45]. Bu adımlar Şekil 4.1’de gösterilmiştir.



Şekil 4.1. Sosyal ağ analiz aşamaları

a. Veri Elde Etme

Analizin amacına göre sosyal ağlardan toplanan verilerin ihtiyaca göre ayıklanması, aralarından sadece gerekli verilerin alınarak veri ambarlarında saklanması ve gereksiz verilerden arındırılmasıdır. Anket yöntemiyle toplanabilir. Örneklemin nasıl seçileceği ve popülasyonu temsil gücü konusunda güçlü istatistiksel kriterler ve literatür mevcuttur.

b. Önişleme ve Bilgiye Dönüşüm

Verilerin temizlenmesi ve modelleme, sınıflandırma ve filtreleme işlemlerinin yapılmasıdır. Ayrıca veriler burada işlenip kullanılacak hale getirilir.

Bireysel (ilişkisel olmayan) veri setleriyle çalışırken elde edilen veriler bir tabloya dökülür, daha sonra ele alınan olguya ve cevaplamak istenen soruya uygun standart ampirik metotları bu veriye uygulanır.

c. Kural Çıkarımı

Elde edilen bilgilerden anlamlı sonuçlar çıkarılır. Bu sonuçlara bulgu adı verilir. Genel kurallar bu aşamada türetilir.

Yapılan araştırma konusu kapsamında elde edilen verilerden yola çıkarak tanımlayıcı bulgular ifade edilebilir. Ayrıca düğümün kendisi ve çevresi arasındaki mesafe ve ilişkiler irdelenerek yoğunluk ve merkezilik ölçütü bulguları elde edilebilir.

d. Değerlendirme

Elde edilen bulgular ve oluşturulan kurallar çerçevesinde ağ incelenip geleceğe dönük tahminler yapılır ve değerlendirmelerde bulunularak çıkarsamalar yapılır.

4.2. Sosyal Ağ Analizi Araçları

Ucinet: Sosyal ağ analizi için Lin Freeman, Martin Everett ve Steve Borgatti'nin geliştirdiği Windows tabanlı, kapsamlı bir uygulamadır. Metin dosyaları, Excel dosyaları ve

Pajek gibi birkaç sosyal ağ analizi uygulaması ile uyumlu çalışmaktadır. 32.767 düğüme kadar işlem potansiyeli mevcuttur. Tek modlu ve iki modlu ağlar üzerinde işlem yapabilmektedir. Yapılabilecek analiz işlemlerinden bazıları; merkezilik ölçümleri, alt grup tanımlama, rol analizi, graf işlemleri ve permutasyon temelli istatistiksel analiz yöntemleridir. Güçlü matris analiz işlemleri de mevcuttur [46].

Pajek: Boyutu büyük olan ağların analizi ve görselleştirilebilmesi için Vladimir Batagelj ve Andrej Mrvar tarafından geliştirilmiş bu uygulama tüm işletim sistemlerinde çalışabilmektedir. Çok geniş ağları küçük ve çok sayıda alt ağa bölebilir, çok iyi bir görselleştirme sağlar, etkili analiz yöntemleri ile sağlıklı analizler yapabilmektedir. Ucinet ile uyumlu çalışabilmektedir [47].

NetMiner: Geniş boyutlu ağların analizi için kullanılır. Python dilinde geliştirilmiştir. Ağ analizi, görselleştirme, istatistiksel işlemler, veri dönüştürme işlemleri yapabilir, grafik tabloları oluşturabilir. Ucinet, Pajek StOCNET gibi programlarla uyumlu çalışabilmektedir [48].

MultiNet: Geniş ve dağınık ağlarda kullanılan bir analiz programıdır. Matris tabanlı çalışır. Analiz işlemleri için istatistiksel analiz yöntemlerini kullanır. Bağlantıların yönü ve ağırlığı ile düşünüm özellikleri grafiksel olarak gösterilebilmektedir. Ayrıca verilerdeki hatalar tespit edilebilmektedir [49].

Gephi: Gephi, grafik ve ağ analizi için kullanılan açık kaynaklı bir yazılımdır. Büyük ağları gerçek zamanlı olarak görüntülemekte ve bu sosyal ağlarda gerçekleştirilen araştırmaları hızlandırmak için bir 3B oluşturma motoru kullanmaktadır. Esnek ve çok görevli bir mimariye sahip olan Gephi, karmaşık veri setleriyle kolayca çalışmak ve değerli görsel sonuçlar üretmek için yeni olanaklar getirmektedir. Gephi ayrıca ağ verisine kolay ve geniş erişim sağlamak ve filtreleme, navigasyon, manipülasyon ve kümeleme gibi özellikler sağlamaktadır [50].

Igraph: Ağların analizi için kullanılan birçok algoritmayı içerir. 2 ve 3 boyutlu ağları görselleştirebilir. GraphML, GML ve Pajek gibi programları destekler. Igraph, açık kaynak kodlu bir yazılımdır [51].

Statnet: C ve R dili ile gerçekleştirilmiş; analiz, simülasyon ve görselleştirme yapabilen bir yazılımdır. R dilinin komut ara yüzünü kullanır. Network, SNA ve ERGM yazılımlarını içerir [52].

Socnet: Sosyal ağ görüntüleyicisi (SocNetV), sosyal ağ analizi ve görselleştirme için kullanılan çapraz platformlu, kullanıcı dostu ücretsiz bir yazılımdır. SocNetV son

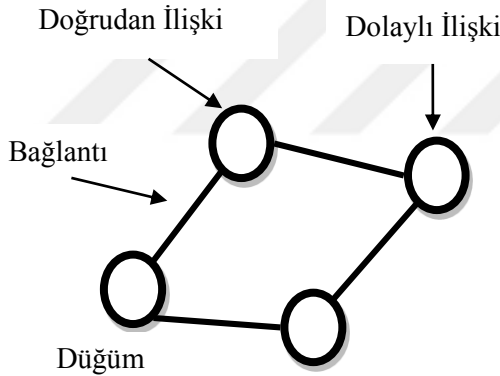
zamanlarda önemli kod yenilemesi, yeni GUI düzeni, performans iyileştirmeleri, Windows 10 uyumluluğu ve çok sayıda hata düzeltme ile yeni bir sürümünü yayımlamıştır. SocNetV, yoğunluk, çap, jeodezi, mesafeler, bağlılık, dışmerkezlilik, kümeleme katsayısı vb. standart grafik ve ağ uyumu ölçümlerini sağlamakta ve bilinen ağ veri takımlarını otomatik olarak oluşturmaktadır. Ayrıca SocNetV, GraphML, Adjacency, Pajek, UCINET, listeler, vb. sosyal ağ biçimlerini desteklemektedir. Birden fazla ilişkiye sahip olan ağları okumakta ve farklı ilişkilerden bağlantıları düzenlemeye ve analiz yapmaya olanak tanımaktadır [53].

Tablo 4.1. Sosyal Ağ araçlarının Karşılaştırılması

Araç	Tip	Son Versiyon/Güncelleme	İşletim Sistemi	Kümeleme	Göstergeler	Görselleştirme
Ucinet	Ücretsiz/Ticari	Ucinet 6.629 - 27 Nisan 2017	Linux, OS X, Windows	✓	✓	✓
NetMiner	Ticari	NetworkMiner 2.1.1 - 19 Ocak 2017	Windows 7/8/8.1/10	✓✓	✓	✓✓
SocNetV	Akademik/Ücretsiz	SocNetV 2.2 - 21 Ocak 2016	Linux, OS X, Windows, Fedora, openSuse	✓✓	✓✓	✓✓
Pajek	Akademik/Ücretsiz	Pajek 5.01 - 1 Şubat 2017	Windows	✓	✓	✓✓
MultiNet	Akademik/Ücretsiz	MultiNet 4.76 - 2007 Versiyon	Windows	✓	✓	✓✓
Gephi	Akademik/Ücretsiz	Gephi 0.9.2 - 14 Ekim 2016	Windows, Mac OS X, Linux	-	✓	✓✓
İgraph	Akademik/Ücretsiz	igraph 1.0.0 - 24 Haziran 2015	Windows, Mac OS X, Linux	-	✓	✓
Statnet	Akademik/Ücretsiz	Statnet 2016.9 - 29 Ağustos 2016	Windows, Mac OS X, Linux	✓✓	✓	✓✓

5. AĞLARDA BAĞLANTI TAHMİNİ YÖNTEMLERİNİN İNCELENMESİ

Bağlantı tahmini yapılabilmesi için ağın yapısının tam olarak bilinmesi gerekmektedir. Sosyal ağlarda bağlantı tahmini için ağ, graf yapısına dönüştürülmelidir. Veriler graflardaki köşeler, ilişkiler ise kenarlar olarak tanımlanmaktadır. Ağ yapısı vektörel olarak da ifade edilebilmektedir. Ağdaki köşe ve kenarların yapıları bilinirse buradan daha oluşmamış bazı bağlantılar tahmin edilebilmektedir. Hatta eklenmesi muhtemel köşelerin de yapıları bilinirse aralarındaki bağlantılar bile tahmin edilebilmektedir. Yine aynı şekilde gelecekte kopacak bağlantılar da tahmin edilebilmekte ve graftan silinmesi muhtemel kenarlar tespit edilebilmektedir. Bu tahminleri bilmek zor problemlerden biridir çünkü ağ dinamik bir yapıya sahiptir. Ağdaki bilgilerin tanımlama şekli de önemli bir problemidir. Mevcut bilgilerin hangilerinin hesaplamalarda kullanılması gerektiği, hangi bilgilerin ne kadar etkin rol oynadığı iyi tespit edilmelidir.



Şekil 5.1. Sosyal bir ağda graf yapısı

Graf olarak modellenen sosyal ağda düğümler (aktörler) ve aralarındaki ilişkiyi temsil eden ayrıtlar Şekil 4.2’de gösterilmektedir. Düğümler arasında direk bağlantı olabileceği gibi dolaylı bağlantılar da olabileceği Şekil 4.2’de görülmektedir.

Bağlantı tahmininin zorlukları üç kısma ayrılabilir: ilki, ağın yapısının yanı sıra düğümlerin ağ içindeki etkinliğini etkileyen özelliklerinin de bilinmesi ağa yeni eklenecek veya ağdan silinecek düğüm ve bağlantıların tahmin edilmesinde önemli bir unsurdur. Örneğin sosyal ağlarda düğümlere karşılık gelen bireylerin sevdikleri veya sevmedikleri şeylerin bilinmesi bağlantı tahmininde temel kriterler olarak göze çarpmaktadır.

İkincisi, ağ hakkında eldeki bilgiler tutarlı olmayabilmektedir. Bu da ağa bağlantı tahmini için uygulanan algoritmaların etkinliğini düşürmektedir.

Üçüncüsü, ağ çok büyük ise yani düğüm ve kenar sayısı çok fazla ise hesaplamalar zorlaşır, yani büyük bir ağda graf modelinin çıkarılması ve tahmin için uygulanan algoritmaların doğru sonuçlar vermesi ağ büyüdükçe zorlaşmaktadır.

Sosyal ağlarda bağlantı tahmini temelde veri madenciliğine dayanmaktadır. Bağlantı tahmini yöntemleri genel olarak graf tabanlı yöntemler, olasılıksal yöntemler ve benzerlik tabanlı yöntemler olmak üzere üç gruba ayrılır. Bu yaklaşımlar içinde yer alan algoritmalar kullanılarak sosyal ağ yapısı ve ağın geleceği hakkında fikir sahibi olunabilir.

5.1. Graf Tabanlı Yöntemler

Graf, ağların yapısını modellemek için kullanılan matematiksel bir modeldir. Ağda yollar, merkezilik ölçümleri, köşe dereceleri, kümeleme katsayısı, gibi özellikler ağın yapısı ve geleceği ile ilgili tahminler yapmamızda çok yararlı olacaktır.

Graf tabanlı yaklaşımlarla yapılacak bağlantı tahmininde, köşeler ve köşeler arasındaki kenarların sayısı önemli bir yer tutar. Bu sayı derece olarak da adlandırılır. Ağların çoğunda düğüm derecelerinin düzensiz olduğu göze çarpmaktadır. Bazı düğümler beklenenden az bazı düğümler beklenenden fazla düğüme sahip olabilir. Yapılan çalışmalarda derece dağılımına bağlı olarak rastgele graf modelleri önerilmektedir [54].

Derece dağılımının hesaplamasının kolaylığına rağmen gerçek dünya problemlerinde ciddi sorunlarla karşılaşmaktadır. Özellikle atıf ağları, world wide web ve bazı sosyal ağlarda bu sorunlar net bir şekilde göze çarpmaktadır. Bu sorunun temel sebebi düğüm dereceleri ile ağırlıklandırılmış ağlarda bağlantı tahmini yaparken yeni eklenecek düğümler ve bu düğümlerle birlikte oluşması muhtemel bağlantıların tespitlerinden kaynaklanmaktadır. Örneğin Barabsi-Albert modelinde [55] her yeni düğümü, mevcut düğümlerin sahip olduğu bağlantıların sayısı ile orantılı bir olasılıkla var olan düğümlere bağlar. Matematiksel olarak yeni bir düğümün i düğüme bağlanma olasılığı denklem (5.1)'de gösterilmektedir:

$$p_i = \frac{k_i}{\sum_j k_j} \quad (5.1)$$

dir. Burada k_i , i 'ninci düğümün derecesi, önceden var olan j düğümlerin derecesine bölünür. Diğer bir deyişle, düşük dereceli düğümlerde yeni bağlantı oluşma olasılığı düşük iken yüksek dereceli düğümlerde bu olasılık çok daha yüksektir.

Graf tabanlı yaklaşımlarda ağlardaki gelişim gözle görülebilecek seviyede iken diğer yaklaşımlarda çok daha büyük ağlar analiz edilebilmiştir.

5.2. Olasılıksal Yöntemler

Olasılıksal yaklaşımlar, ağda olabilecek değişikliklerin olasılığını tahmin etmeye çalışır. Olasılıksal yaklaşımlar muhtemel bağlantıların olasılığını tahmin eden modeller ve bir ağın muhtemel yapılarının olasılığını tahmin eden modeller olmak üzere iki gruba ayrılır. Olasılıksal yöntemler çoğunlukla graf tabanlı yaklaşımlara dayalıdır. Son yıllarda sosyal ağlar için üstel rasgele graf modellerine büyük bir ilgi oluşmuştur. Üstel rasgele graf modelleri bir ağın genel özellikleri, köşeler ve kenarlar kullanarak bütün bir ağ için olasıksal modelleri tahmin etmede kullanılan popüler bir yaklaşımdır. Bu modeller belirli ağ yapılarının oluşup oluşmayacağı hakkında tahminler yapmaya izin veren istatistiksel modellere dayalıdır. Üstel rasgele graflar, sıradan graflardaki sınırlamalarının ortadan kaldırılarak ağlar için makul modeller geliştirmek için kullanılır [56].

5.3. Benzerlik Tabanlı Yöntemler

Benzerlik, bağlantı tahmini probleminde ağın geleceği hakkında yapılabilecek tahminlerin güçlü olmasında önemli bir ölçüttür. Aralarında bağlantı olmayan iki düğüm birbirine ne kadar benzerse gelecekte bu iki düğüm arasında bağlantı oluşma olasılığı o kadar yüksektir. Örneğin, facebook da arkadaş olmayan iki kişinin ortak özellikleri ne kadar çoksa gelecekte arkadaş olma ihtimalleri o kadar yüksektir.

Benzerlik ölçütleri semantik ve topolojik olarak ikiye ayrılır. Semantik ölçütlerde düğümün içeriği benzerlik ölçütü olarak ele alınır. Örneğin yazar işbirliği ağında makalelerin anahtar kelimelerdeki benzerlikle yazarlar arasında gelecekteki etkileşimler tahmin edilebilir [57]. Topolojik ölçütler benzerlik ölçütü olarak ağın yapısını kullanırlar. İki düğüm arasındaki ortak komşuların sayısı topolojik ölçütlere bir örnektir. Topolojik ölçütler literatürde genel olarak komşuluk tabanlı ve yol tabanlı diye kategorize edilmiştir.

5.3.1. Komşuluk Tabanlı Ölçütler

Komşuluk tabanlı ölçütlerde temel fikir x ve y düğümlerinin komşuları $\Gamma(x)$ ve $\Gamma(y)$ nin ne kadar ortak özelliği varsa gelecekte aralarında bağlantı olma ihtimali de o kadar yüksektir. $\Gamma(x)$, x düğümünün ağdaki komşularının kümesini göstermektedir.

Ortak Komşular: x ve y düğümleri için ortak komşuların sayısını ifade etmektedir [58]. Bu ifadenin matematiksel karşılığı denklem (5.2)'de gösterildiği gibidir.

$$OK(x, y) = |\Gamma(x) \cap \Gamma(y)| \quad (5.2)$$

Jaccard Katsayısı: x ya da y den rasgele seçilen bir özelliğin hem x hem de y de birlikte bulunma olasılığıdır. Jaccard, ortak komşuların normalleştirilmiş halidir [58,59]. Matematiksel olarak denklem (5.3)'teki gibi ifade edilir.

$$JK(x, y) = |\Gamma(x) \cap \Gamma(y)| / |\Gamma(x) \cup \Gamma(y)| \quad (5.3)$$

Salton Index: Kosinüs benzerliği de denen bu yöntemde k_x , x düğümünün derecesi k_y de y düğümünün derecesi olsun. Bu durumda formül denklem (5.4)'teki gibi olur [58, 60]:

$$S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{k_x * k_y}} \quad (5.4)$$

Sorensen Index: Bu yöntem ekolojik topluluk verileri için kullanılır [58, 61]. Matematiksel olarak denklem (5.5)'te gösterildiği gibi ifade edilir:

$$S_{xy} = \frac{2|\Gamma(x) \cap \Gamma(y)|}{k_x + k_y} \quad (5.5)$$

Leicht-Holme-Newman Index: Ortak komşusu olan düğümlerin benzerlik değerleri bu yöntemde göre Ortak Komşu İndeks'inden daha yüksek değer alır [58, 62]. Denklem (5.6) L. H. Newman İndeks'in matematiksel karşılığıdır.

$$N_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{k_x * k_y} \quad (5.6)$$

Adamic/Adar Katsayısı: Bu ölçüm iki web sayfasındaki içeriklerin birbirlerine ne kadar yakın olduğunu ölçmektedir. Bunu yapmak için bu sayfaların özellikleri belirlenmelidir [58, 63]. Formülü denklem (5.7)'de verilmiştir:

$$AA(x, y) = \sum_{z: x \text{ ve } y \text{ nin özelliği}} \frac{1}{\log(\text{frekans}(z))} \quad (5.7)$$

Bu nicelikte ortak özelliklerden nadir olanların ağırlık oranları artırılmaktadır. Bağlantı tahmini problemi için bu formül denklem (5.8)'deki gibi güncellenmiştir:

$$AA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(z)} \quad (5.8)$$

Bağlantı tahmini için yapılan çalışmalarda Adamic/Adar diğer ölçütlere oranla daha iyi sonuçlar vermektedir.

Tercihli Bağlılık: Ağda oluşacak yeni bir bağlantının düğümlerinden birinin belirli bir düğüm olma ihtimali, o düğümün komşularının sayısı ile orantılıdır. Yani komşu sayısı çok olan düğümlerin yeni bağlantı oluşturma ihtimali daha yüksektir [59, 3]. Matematiksel formülü denklem (5.9)'daki gibidir:

$$TB(x, y) = |\Gamma(x)| \cdot |\Gamma(y)| \quad (5.9)$$

Kaynak Paylaştırma İndeksi: Karmaşık ağlarda kullanılan bir ölçüttür. Birbiriyle doğrudan bağlantısı olmayan düğümler arasındaki bağlantıları (aktörler arasındaki ilişkileri) ölçer. Düğümler arasında doğrudan bağlantı olmamasına rağmen düğümler komşuları üzerinden bağlantı sağlarlar. Düğümler arasındaki benzerlik aralarındaki veri akışına göre değerlendirilir. Veri akışı yüksek olan düğümler daha benzerdir. Formülde kullanılan $k(z)$, z 'nin derecesidir [58, 64, 65]. Matematiksel ifadesi denklem (5.10) da verilmiştir:

$$KP(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)} \quad (5.10)$$

5.3.2. Yol Tabanlı Ölçütler

Yol tabanlı ölçütler iki düğüm arasındaki en kısa yolların sayısını baz alırlar.

Katz: Katz ölçütü düğümler arasındaki en kısa yolların sayısının toplamını baz alır. Benzerlik hesaplamasında l burada yol uzunluğudur ve uzun yolların hesaplama üzerindeki olumsuz etkisini azaltmak için bir parametreye (β^l) üs olarak dahil edilmiştir [58, 66]. Bu ölçüt denklem (5.11)'de şöyle ifade edilir:

$$Katz(x, y) = \sum_{l=1}^{\infty} \beta^l \cdot |yollar_{x,y}^{(l)}| \quad (5.11)$$

$|yollar_{x,y}^{(l)}|$, x ve y düğümleri arasında l uzunluğundaki yolların sayısıdır. $\beta > 0$ olmalıdır. β değeri ne kadar küçük verilirse ortak komşuların değerine o kadar yaklaşılır.

Yerel Rastgele Gezinti: Yerel rastgele gezinti $H_{(x,y)}$, x düğümünden başlayan rastgele gezintinin y düğümüne ulaşıldığında elde edilen adım sayısıdır. Yerel rastgele gezinti ne kadar düşük olursa x ve y düğümleri birbirine o kadar benzerdir ve aralarında bağlantı olma ihtimali o kadar yüksektir. Düşük olması x ve y nin benzer olduğunu aralarında bağlantı olabileceğini göstermektedir. Yönlü ağlarda bu ölçüt simetrik değildir. O yüzden bunun yerine gidiş/dönüş zamanı (commute time), $C_{(x,y)} = H_{(x,y)} + H_{(y,x)}$, kullanılmalıdır. y düğümün çok geniş bir dağılım olasılığına sahip olduğunda $H_{(x,y)}$ çok küçük olacaktır. Bunu dengelemek için ölçüt normalleştirilebilir [58, 67]. Ulaşma zamanı matematiksel olarak denklem (5.12)'de verilmiştir:

$$UZ(x, y) = -(H_{(x,y)} \cdot \pi_y + H_{(y,x)} \cdot \pi_x) \quad (5.12)$$

Köklü PageRank: Ulaşma zamanında x ve y düğümleri arasındaki yollar çok kısa olsa da rastgele yürüyüşlerle x düğümünden y düğümüne ulaşmak için çok fazla düğüm geçilmesi gerekebilir. Bu durumun önüne geçebilmek için rasgele yürüyüş her adımda β parametresindeki olasılık değeri ile yeniden başa döndürülebilir. Böylece ağda rastgele yürüyüşler mümkün olan en kısa yollardan yapılabilir. Rasgele yürüyüşün belli bir olasılıkla yeniden başlatılması web sayfalarındaki PageRank ölçütünün temelidir. Rasgele yürüyüş β [0,1] olasılığı ile başa döner, $1-\beta$ olasılığı ile o an bulunan düğümün komşularından rasgele birine gider. Bu işlem her adımda uygulanır. i düğümünün tüm koşulları için diagonal derece

matrisinde $D[i,j] = \sum_j A[i,j]$ dir. $N = D^{-1}A$, komşuluk matrisinin satırlarının 1'e normalize edilmesidir [58, 68]. Formülü denklem (5.13)'te verilmiştir:

$$KPR = (1 - \beta)(I - \beta N)^{-1} \quad (5.13)$$

SimRank: Bu ölçüt, iki düğüm benzer düğümler ile bağlantılı ise bu iki düğüm benzerdir esasına dayanır [58, 69]. Matematiksel ifadesi denklem (5.14)'te şöyle ifade edilir:

$$Benzerlik(x,y) := \gamma \cdot \frac{\sum_{a \in \Gamma(x)} \sum_{b \in \Gamma(y)} Benzerlik(a,b)}{|\Gamma(x)| \cdot |\Gamma(y)|} \quad (5.14)$$

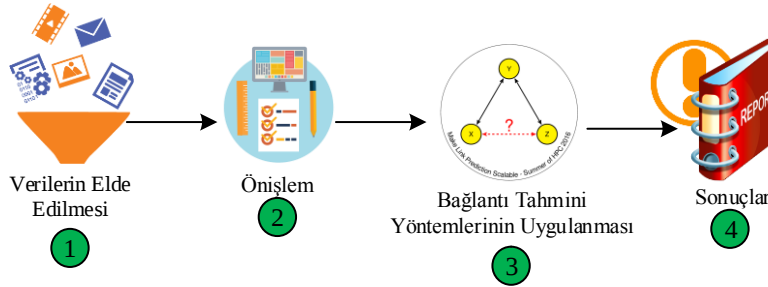
Bu yöntemler dışında pek çok yöntem ve yaklaşım mevcuttur. Bunlardan bazıları sık örüntü madenciliği, rastgele yürüyüş ve yayma yöntemleridir.

6. AKADEMİK AĞLARDA YAZAR-MAKALE BAĞLANTI TAHMİNİ UYGULAMASI

Akademik alanda yapılan çalışmalarda en dikkat çeken zorluk, bir akademisyenin çalışmak istediği bir konuda kendine kaynak veya birlikte çalışacağı başka bir akademisyeni bulabilmesidir. Bu çalışmada seçilmiş olan yazarların çalışma konuları ele alınarak, yazarlar arasında çalışma konuları bazında oluşan benzerlik hesaplanmıştır. Böylece yazarların çalışma alanlarında kendileri ile benzer konuları çalışan yazarların tespit edilmesi amaçlanmıştır.

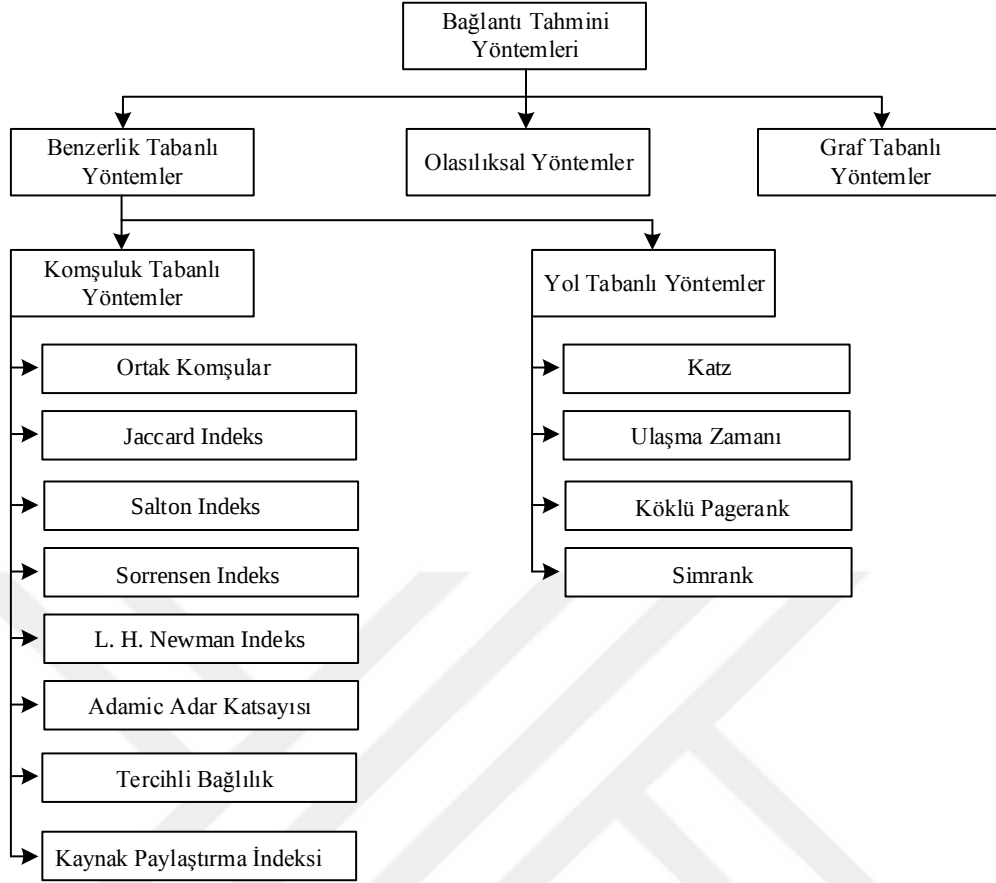
Geliştirilen bu araç ile gerçekleştirilmek istenen amaç, yazarların çalışma alanları arasındaki benzerliği hesaplayarak, yazarlar arasındaki ilişkileri analiz etmektir. Bu amaç doğrultusunda, yazarların çalıştıkları alanlara göre benzer konularda çalışan ilgili yazarları tespit etmek, ortak çalışma yapan yazar gruplarını ortaya koymak ve ortaya çıkarılan bu bilgiler kullanılarak elde edilecek sonuçlara göre birlikte yayın yapabilecek yazarları tahmin edecek bir yapının temellerini oluşturmasını sağlamaktır.

Gerçekleştirilen uygulamanın genel adımları Şekil 6.1’de gösterilmektedir.



Şekil 6.1. Geliştirilen uygulamanın adımları

Gerçekleştirilen uygulamada kullanılan bağlantı tahmini yöntemleri ile birlikte diğer bağlantı tahmini yöntemleri Şekil 6.2’de gösterilmektedir.



Şekil 6.2. Bağlantı tahmini yöntemleri

Geliştirilen uygulamanın ana ekran görüntüsü şekil 6.3'te gösterilmektedir.



Şekil 6.3. Geliştirilen uygulamanın veriler eklenmeden önceki ana sayfa görünümü

Şekil 6.3'te uygulamanın ana form ekranı görülmektedir. Uygulama; Microsoft Visual Studio .Net ile, C# programlama dilinde ve SQL Server veri tabanı kullanılarak gerçekleştirilmiştir. Uygulamada, veri tabanına eklenen yazarlar sol tarafta listelenir. Seçilen yazarın makale isimleri ve yayınlanma yılları makale sekmesinde liste halinde gösterilir. yine seçilen yazarın yayınladığı makalelerin anahtar kelimeleri ve her bir anahtar kelimenin kullanım sayısı anahtar kelimeler sekmesinde listelenir. Sonra yazarlar ve yayınlanan makalelerdeki anahtar kelimeler arasındaki benzerlik değerleri 5 farklı yöntem ile benzerlik sekmesinde hesaplanmaktadır.

6.1. Verilerin Elde Edilmesi

Makale Yazar ilişkisinin tespitinde *IEEE Explore*, *AIP*, *IET*, *AVS*, *IBM*, *Semantic Scholar*, *CiteSeerX*, *BibSonomy*, *MathSciNet*, *PubMed*, *RePEc*, *zbMATH* gibi birçok farklı veri tabanı kullanılabilir. Bu veri tabanlarında makaleye ilişkin özet, makale adı, anahtar kelimeler, yayınlanma tarihi, DOI, ISSN ve ISBN bilgileri, kontrol ve index terimleri gibi birçok veri elde edilebilir. Ayrıca makale seçimleri, konferans, dergi, kitap, eğitim kursu gibi yayınlanma türüne göre de ayrılabilir. Bu veri kümeleri kullanım amaçlı olarak herkese açıktır. Şekil 6.4'te IEEE Explore veri tabanında yazar ve makaleler ile ilgili seçilebilecek alanlar görülmektedir.

Şekil 6.4. IEEE Explore veri tabanından seçilebilecek alanların uygulama arayüzünde görünümü

Bu çalışmamızda IEEE Explore veri tabanından rastgele belirlenen 20 adet yazara ait;

- Makale isimleri,
- Makalelerin yayınlanma yılları,
- Makalelerin anahtar kelimeleri

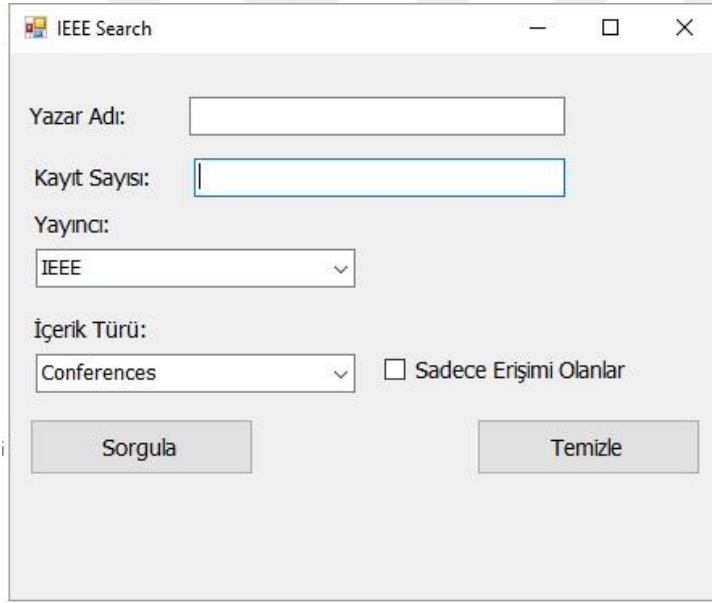
- Anahtar kelimelerin kaç kez kullanıldığı

verileri Microsoft Visual Studio. Net programlama aracı yardımıyla elde edilmiştir.

Alınan veriler MS SQL Server veri tabanına aktarılmıştır.

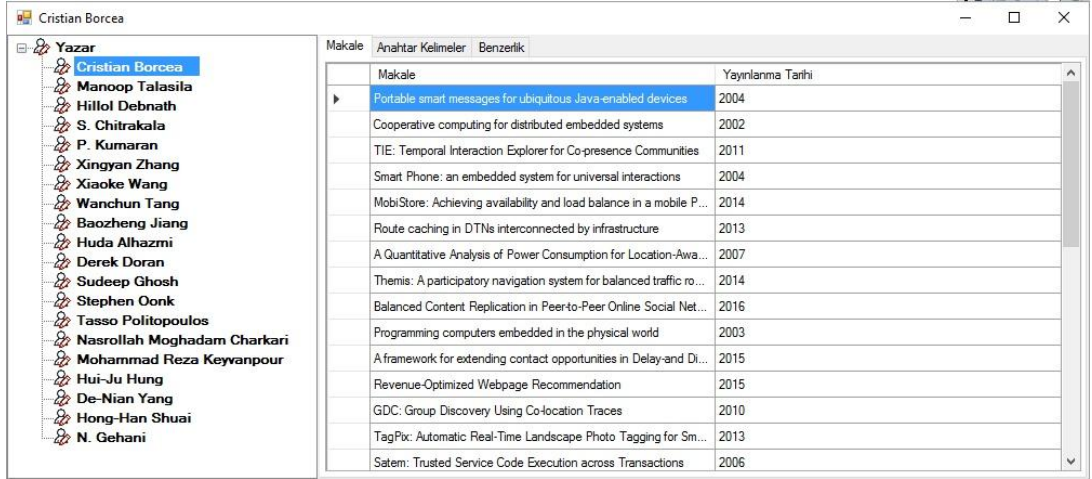
6.2. Ön İşlem

Yazarlara ilişkin veri tabanından alınan tüm veriler incelenerek, gerçekleştirilen uygulama için uygun parametreler belirlenmiştir. IEEE Explore sisteminden alınabilecek parametreler; yazar adı, yayın yılı, kurum, yayın başlığı, anahtar kelimeler, kontrol terimleri, özet, ISSN, ISBN'dir. Bu parametrelerden yazar adı ve yayınlanan makalelerin anahtar kelimeleri uygulama için yeterli görüldüğünden filtrelenerek alınmıştır. Özet, ISSN ve ISBN benzerlik için hiçbir anlam ifade etmemektedir.



Şekil 6.5. Yazar adı ile arama yapıp istenen verileri çekme

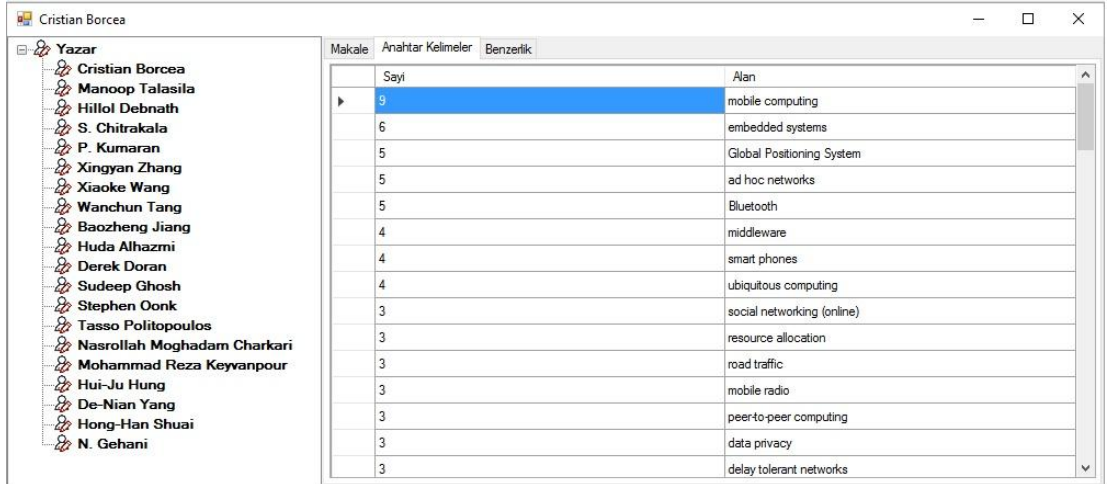
Şekil 6.5'te yazar isimleri ile arama yapıp, istenen verilerin (yazar adı-soyadı, makale başlığı, makalelerin anahtar kelimeleri ve yayınlanma yılları) veri tabanına kayıt edildiği form ekranı görülmektedir.



Yazar	Makale	Yayınlanma Tarihi
Cristian Borcea	Portable smart messages for ubiquitous Java-enabled devices	2004
Manoop Talasila	Cooperative computing for distributed embedded systems	2002
Hillol Debnath	TIE: Temporal Interaction Explorer for Co-presence Communities	2011
S. Chitrakala	Smart Phone: an embedded system for universal interactions	2004
P. Kumaran	MobiStore: Achieving availability and load balance in a mobile P...	2014
Xingyan Zhang	Route caching in DTNs interconnected by infrastructure	2013
Xiaoke Wang	A Quantitative Analysis of Power Consumption for Location-Awa...	2007
Wanchun Tang	Themis: A participatory navigation system for balanced traffic ro...	2014
Baozheng Jiang	Balanced Content Replication in Peer-to-Peer Online Social Net...	2016
Huda Alhazmi	Programming computers embedded in the physical world	2003
Derek Doran	A framework for extending contact opportunities in Delay-and Di...	2015
Sudeep Ghosh	Revenue-Optimized Webpage Recommendation	2015
Stephen Oonk	GDC: Group Discovery Using Co-location Traces	2010
Tasso Politopoulos	TagPic: Automatic Real-Time Landscape Photo Tagging for Sm...	2013
Nasrollah Moghadam Charkari	Satim: Trusted Service Code Execution across Transactions	2006
Mohammad Reza Keyvanpour		
Hui-Ju Hung		
De-Nian Yang		
Hong-Han Shuai		
N. Gehani		

Şekil 6.6. Yazar isimleri, yayınlanan makaleler ve yayın yılları

Şekil 6.6’da veri tabanına kayıt edilmiş olan yazarlar yazarların IEEE Explore sistemindeki konferans makaleleri ve yayınlanma yılları listelenmektedir.



Yazar	Makale	Anahtar Kelimeler	Benzerlik
Cristian Borcea	9	mobile computing	
Manoop Talasila	6	embedded systems	
Hillol Debnath	5	Global Positioning System	
S. Chitrakala	5	ad hoc networks	
P. Kumaran	5	Bluetooth	
Xingyan Zhang	4	middleware	
Xiaoke Wang	4	smart phones	
Wanchun Tang	4	ubiquitous computing	
Baozheng Jiang	3	social networking (online)	
Huda Alhazmi	3	resource allocation	
Derek Doran	3	road traffic	
Sudeep Ghosh	3	mobile radio	
Stephen Oonk	3	peer-to-peer computing	
Tasso Politopoulos	3	data privacy	
Nasrollah Moghadam Charkari	3	delay tolerant networks	
Mohammad Reza Keyvanpour			
Hui-Ju Hung			
De-Nian Yang			
Hong-Han Shuai			
N. Gehani			

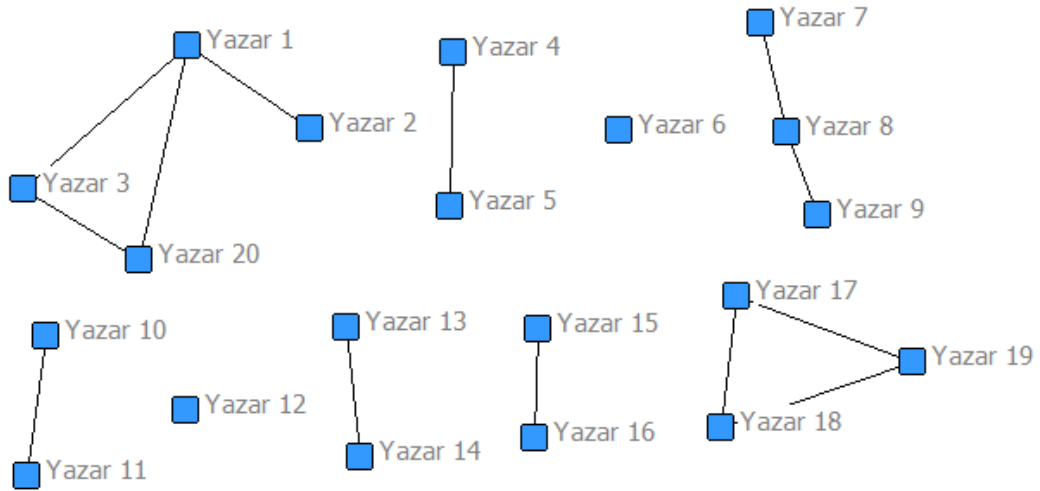
Şekil 6.7. Anahtar kelimeler ve yayınlanan makalelerde kullanım sayıları

Şekil 6.7’de veri tabanına kaydedilmiş olan yazarlar, yazarların makalelerindeki anahtar kelimeler ve anahtar kelimelerin kaç defa kullanıldığı listelenmektedir.

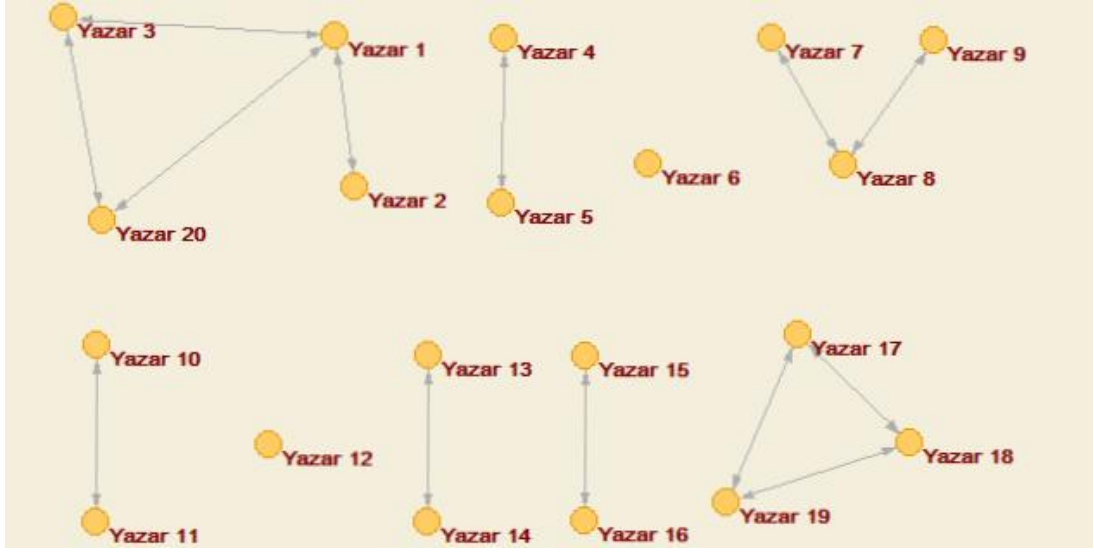
6.3. Bağlantı Tahmini Yöntemlerinin Uygulanması:

Belirlenen veri kümesine *Jaccard Index*, *Sorensen Index*, *Ortak Komşu Index*, *L. H. Newman Index* ve *Salton Index* komşuluk tabanlı bağlantı tahmini yöntemleri uygulanmıştır. Bu yöntemlerin seçilmesindeki en önemli etmenler, yöntemlerin uygulanma kolaylığı ve eldeki verilerin formülasyondaki temsil biçimleri açısından bu yöntemlere uygun olmasıdır [59]. Bu uygulamada geliştirilen yazılım için, Microsoft Visual Studio .Net ortamında C# programlama dili kullanılmıştır.

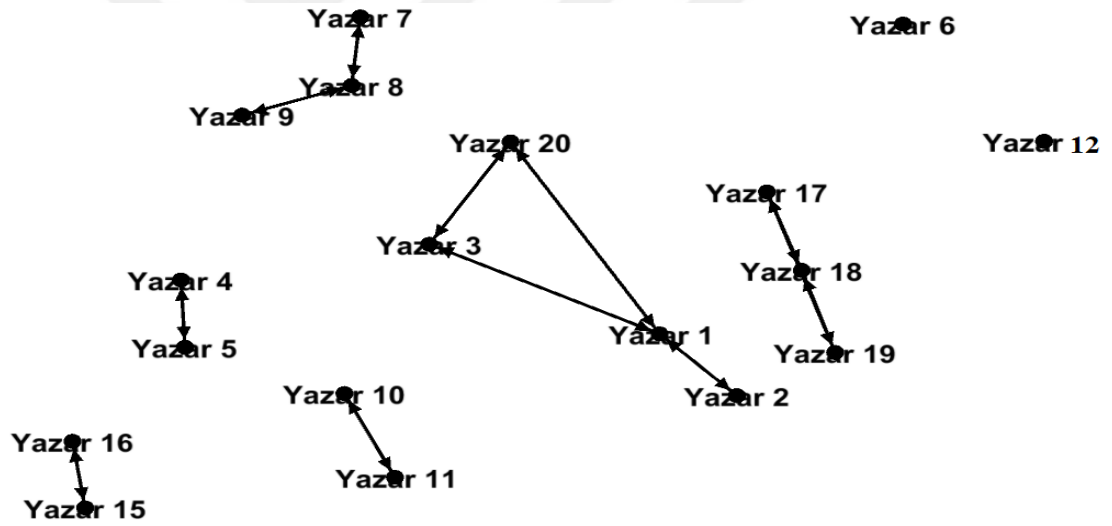
Şekil 6.8’de UCINET sosyal ağ analizi yazılımı, Şekil 6.9’de PAJEK sosyal ağ analizi yazılımı, Şekil 6.10’da GEPHI sosyal ağ analizi yazılımı ile rastgele seçilmiş 20 yazar arasında mevcut olan bağlantılar görsel olarak sunulmuştur. Yazarlar arasındaki bağlantı, yazarların birbiri ile yayın yaptığı anlamını taşımaktadır. Örneğin; yazar 17 ile yazar 18, yazar 17 ile yazar 19 ve yazar 18 ile yazar 19 birlikte yayın yapmıştır. Dolayısı ile Şekil 6.8’de aralarında bağlantı bulunmaktadır.



Şekil 6.8. UCINET ile yazar-yazar ilişki ağı

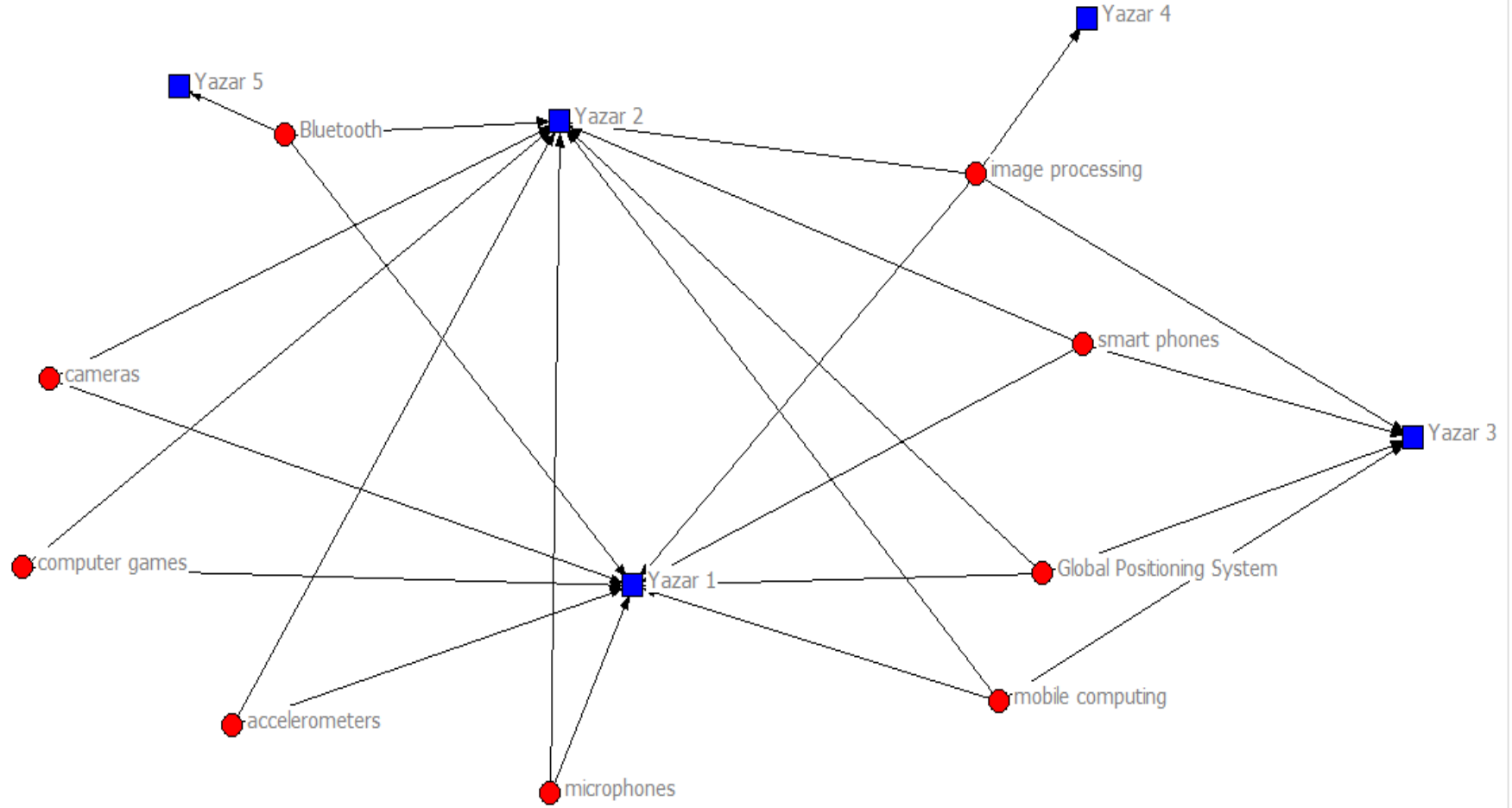


Şekil 6.9. PAJEK ile yazar-yazar ilişki ağı

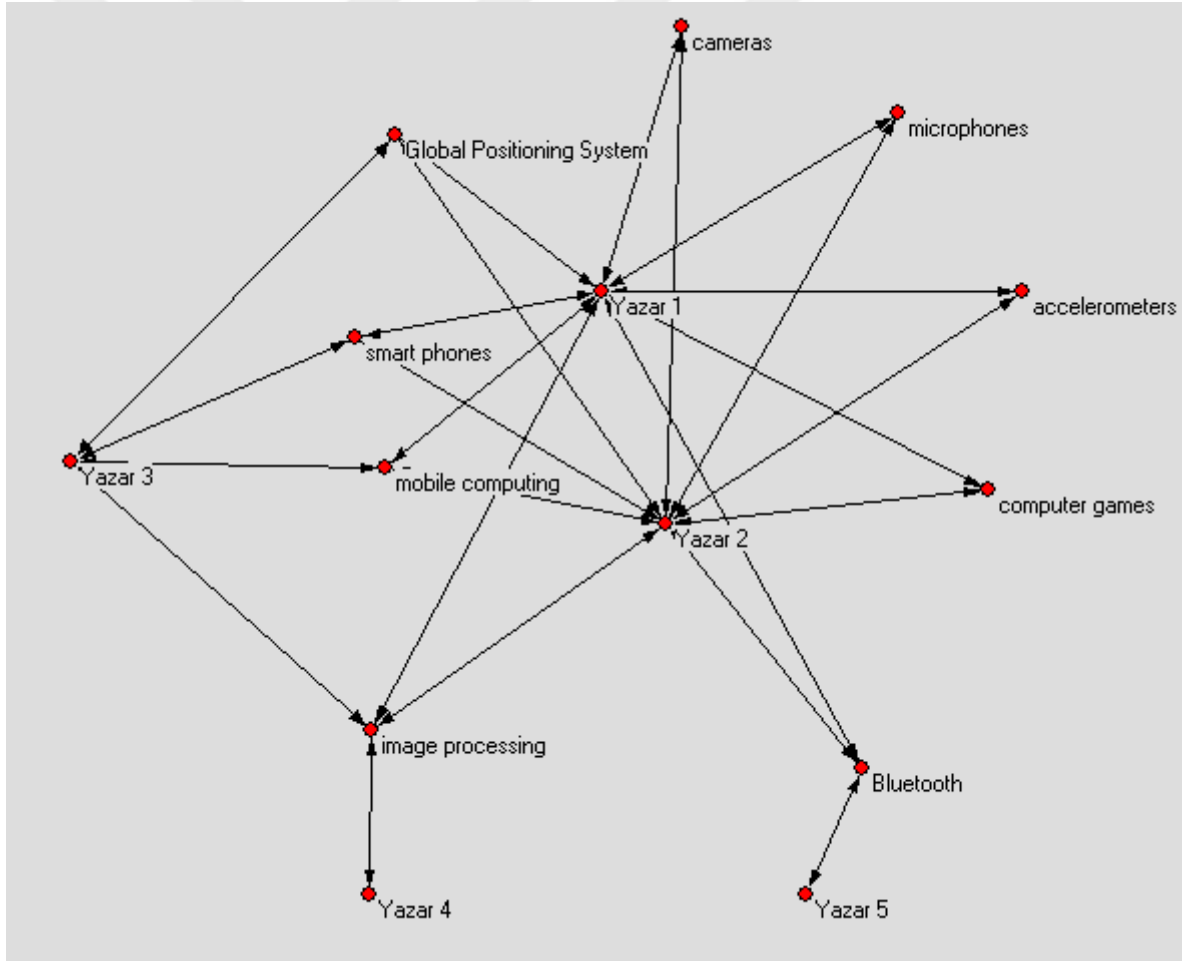


Şekil 6.10. GEPHI ile yazar-yazar ilişki ağı

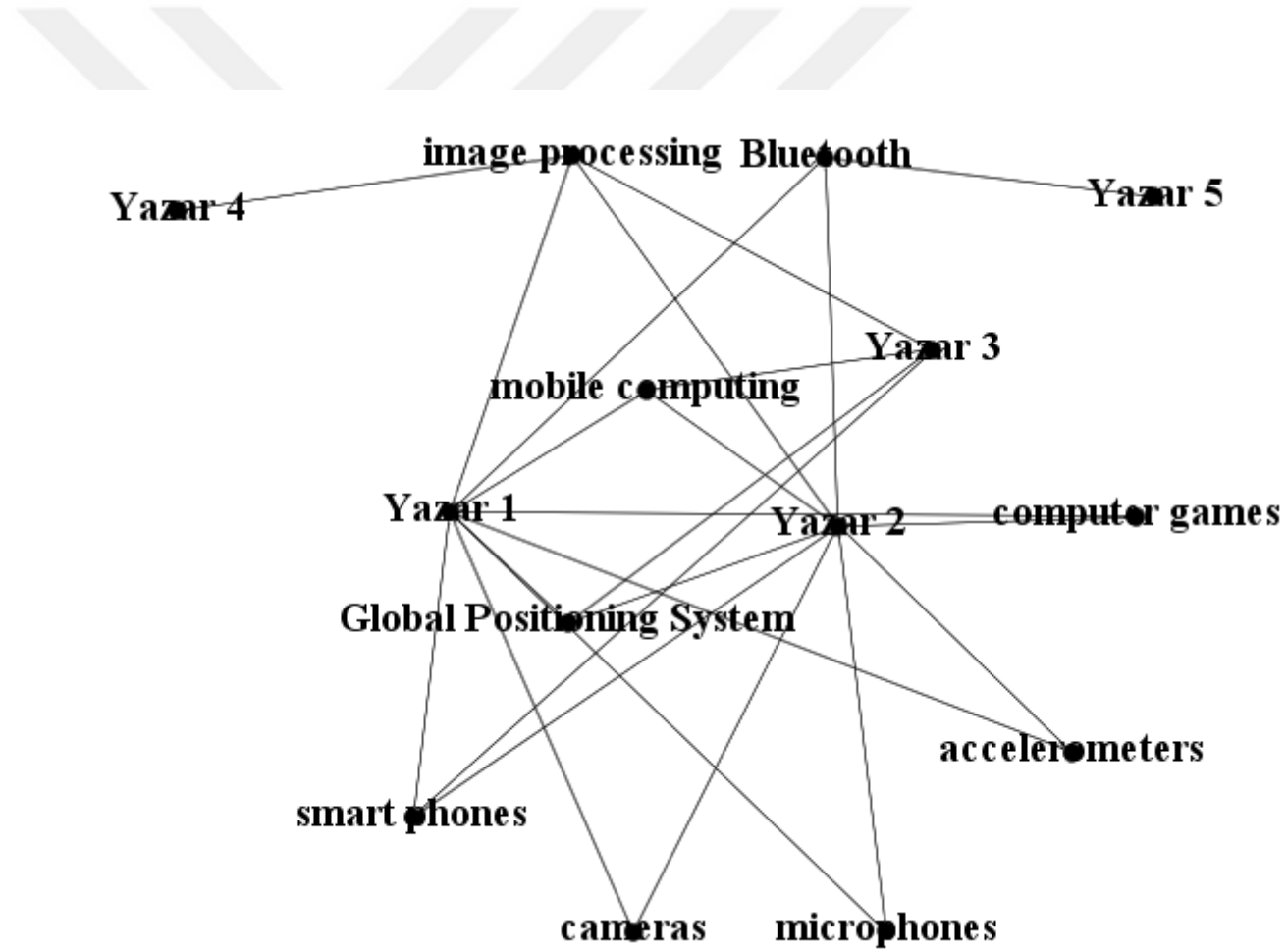
Şekil 6.11 UCINET sosyal ağ analizi yazılımı, Şekil 6.12 PAJEK sosyal ağ analizi yazılımı ve Şekil 6.13 de GEPHI sosyal ağ analizi yazılımı ile; yazarlar ve yayınladıkları makalelerin anahtar kelimeleri arasındaki ilişkileri görsel olarak ifade etmektedir. Yani yazarlar ile kullandıkları anahtar kelimeler arasında bağlantı vardır.



Şekil 6.11. UCINET ile yazar-konu ilişki ağı örneği



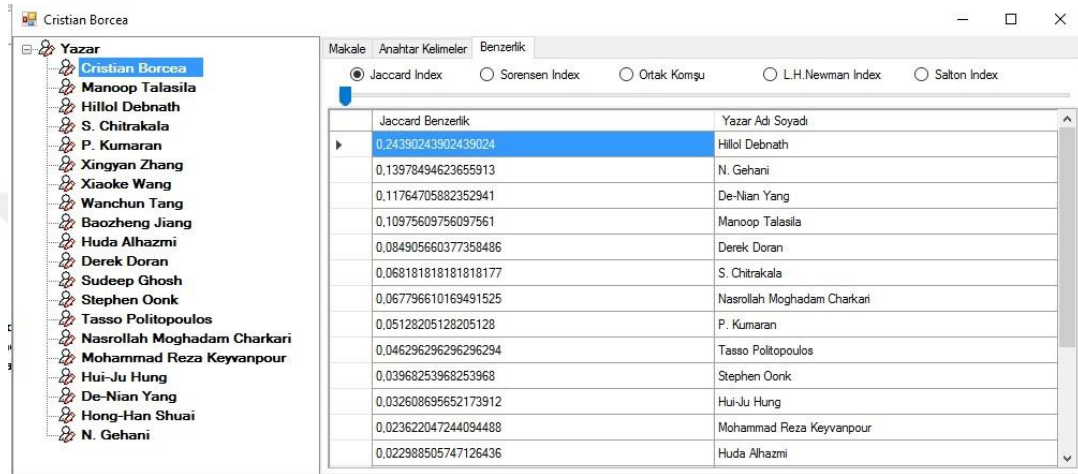
Şekil 6.12. PAJEK ile yazar-konu ilişki ağı örneği



Şekil 6.13. Gephi ile yazar-konu ilişkisi ağı örneği

6.4. Uygulama Sonuçları

Geliştirilen uygulama ile yirmi adet yazarın çeşitli konferanslarda yayınladığı makalelerin anahtar kelimeleri arasındaki benzerlikler beş farklı matematiksel model uygulanarak analiz edilmiş ve sonuçları aşağıda beş farklı tabloda verilmiştir. Yazarlar rastgele seçilmiş, IEEE Explore sisteminde yayınlanmış olan konferans makaleleri veri seti hazırlamak için tercih edilmiştir.



Jaccard Benzerlik	Yazar Adı Soyadı
0.24390243902439024	Hillol Debnath
0.13978494623655913	N. Gehani
0.11764705882352941	De-Nian Yang
0.10975609756097561	Manoop Talasila
0.084905660377358486	Derek Doran
0.068181818181818177	S. Chitrakala
0.067796610169491525	Nasrollah Moghadam Charkari
0.05128205128205128	P. Kumaran
0.046296296296296294	Tasso Politopoulos
0.03968253968253968	Stephen Oonik
0.032608695652173912	Hui-Ju Hung
0.023622047244094488	Mohammad Reza Keyvanpour
0.022988505747126436	Huda Alhazmi

Şekil 6.14. Bağlantı tahmini yöntemlerine göre sonuç ekranı

Şekil 6.14'te veri tabanına kaydedilmiş yazarlar ve yazarların yazdıkları makalelerde kullandıkları anahtar kelimelerden yola çıkarak elde edilen benzerlik değerleri 5 farklı yöntemle listelenmektedir.

Yazar-Yazar ağı incelendiğinde bazı yazarların birlikte yayın yaptıkları görülmektedir. Yapılan bu analiz çalışmasında yazarlar arasındaki benzerlikler irdelendiğinde ortak yayın yapmamış bazı yazar çiftlerinin benzerlik değerlerinin ortak yayın yapmış olan yazarlar arası benzerlik değerinden daha yüksek olduğu görülür. Bu da bu yazarlar arasında da bir ilişki kurulabileceği (örneğin ortak yayın yapabilecekleri, yeni ilişki kurulan yazarların çalışmaları muhtemel konuların tespiti) anlamına gelir.

Yapılan uygulamada Tablo 6.1, yazarlar arasındaki benzerliğin Jaccard İndex yöntemi ile hesaplanan sonuçları içermektedir. Tablo 6.2, Sorensen İndex yöntemi ile; Tablo 6.3, Ortak Komşu İndex yöntemi ile; Tablo 6.4, L. H. Newman İndex yöntemi ile, Tablo 6.5 ise Salton İndex yöntemi ile hesaplanan sonuçları içermektedir. Şekil 6.8, Şekil 6.9 ve Şekil 6.10'da yazarlar arasındaki mevcut bağlantılar gösterilmektedir. Aralarında bağlantı olan

yazarlar birlikte yayın yapmış olan yazarlardır. Yazarlar arasında oluşması muhtemel yeni bağlantıları tahmin etmek için Şekil 6.8’de gösterildiği gibi birbiri ile bağlantılı olan yazar çiftlerinden her tablo için bir yöntem kullanılarak benzerlik değerleri hesaplanmış ve her tabloda bu yazarlar arasındaki en büyük değer eşik değer olarak belirlenmiştir.

Tablo 6.1- Tablo 6.5 dikkate alındığında yazar 1 ile yazar 2, yazar 1 ile yazar 3 ve yazar 1 ile yazar 20 arasında oluşan benzerlik değerleri incelendiğinde, yazar 1 ile yazar 3’ten elde edilen değerler en büyük değerler olduğundan eşik değerleri olarak belirlenmiştir. Tablolar ayrı ayrı analiz edildiğinde beş tabloda da belirlenen eşik değerlerinden büyük tek değer yazar 1 ile yazar 18 arasında oluşan benzerlik değeri olduğu görülmektedir. Uygulanan beş yöntemde de bu değerlerin eşik değerlerinden yüksek çıkması yazar 1 ile yazar 18 arasında yeni bir bağlantı oluşabileceği anlamına gelmektedir.

Aynı şekilde Şekil 6.8’den yola çıkarak, yazar 2 sadece yazar 1 ile bağlantılı olduğundan eşik değerleri beş yöntemde de yazar 2 ile yazar 1 arasında oluşan benzerlik değerleri olur. Tüm tablolar ayrı ayrı incelendiğinde Tablo 6.1, Tablo 6.2, Tablo 6.4 ve Tablo 6.5’te yazar 2 ile yazar 3 arasındaki benzerlik değerlerinin eşik değerlerden büyük olduğu görülmektedir. Bu durumda Jaccard İndex, Sorensen İndex, L. H. Newman İndex ve Salton İndex yöntemlerine göre yazar 2 ile yazar 3 arasında yeni bir bağlantı oluşabilmektedir. Sadece Tablo 3’te Ortak Komşu yöntemine göre bu iki yazar arasında bağlantı oluşmamaktadır.

Yazar 4 ile sadece yazar 5 arasında bağlantı olduğundan eşik değerleri yazar 4 ile yazar 5 arasında hesaplanan değerlerdir. Tablolar incelendiğinde yazar 4 ile yazar 15, yazar 4 ile yazar 16 ve yazar 4 ile yazar 19 arasındaki benzerlik değerlerinin tüm tablolarda eşik değerlerinden büyük olduğu görülmektedir. Bu durumda Jaccard İndex, Sorensen İndex, Ortak Komşu İndex, L. H. Newman İndex ve Salton İndex yöntemlerine göre yazar 4 ile yazar 15, yazar 4 ile yazar 16 ve yazar 4 ile yazar 19 arasında da bağlantılar oluşabileceği söylenebilir.

Yazar 10 ile sadece yazar 11 bağlantılı olduğundan eşik değerleri bu iki yazar arasında hesaplanan değerler olur. Tablolar incelendiğinde yazar 10 ile yazar 19 arasındaki benzerlik değerlerinin Tablo 6.1, Tablo 6.2, Tablo 6.4 ve Tablo 6.5’te eşik değerlerinden büyük olduğu görülür. Sadece Ortak Komşu İndex yönteminde bu değer sağlanmamaktadır. Bu durumda yazar 10 ile yazar 19 arasında da bir bağlantı olabileceği söylenebilir.

Yazar 15 ile sadece yazar 16’nın bağlantısı olduğundan burada da eşik değerleri yazar 15 ile yazar 16 arasında hesaplanan değerlerdir. Tablolar analiz edildiğinde yazar 15 ile

yazar 4 arasındaki benzerlik değerlerinin beş yöntemin tamamında eşik değerlerinden yüksek olduğu gözlenmektedir. Bu durumda yazar 15 ile yazar 4 arasında da bağlantı oluşabileceği söylenebilir.

Yazar 19 ile yazar 17 ve yazar 18 arasında bağlantı bulunmaktadır. Tablolar analiz edildiğinde yazar 19 ile yazar 17 arasında hesaplanan değerlerin daha büyük olduğu gözlenir. Yani eşik değerleri bu iki yazar arasında hesaplanan değerlerdir. Tablolar incelenince yazar 19 ile yazar 10 arasındaki benzerlik değerlerinin Tablo 6.1, Tablo 6.2, Tablo 6.4 ve Tablo 6.5'te; yazar 19 ile yazar 4 ve yazar 19 ile yazar 11 arasındaki benzerlik değerlerinin ise yöntemlerin tamamında eşik değerlerinden büyük olduğu gözlenmektedir. Bu durumda yazar 19 ile yazar 4, yazar 19 ile yazar 10 ve yazar 19 ile yazar 11 arasında yeni bağlantı oluşabileceği söylenebilir.

Yazarlar arasında oluşması muhtemel bağlantılar da yazar-yazar ağına eklenmiş ve sonuçta yeni bir ağ elde edilmiştir. Yeni Yazar-Yazar ağı Şekil 6.15'te UCINET sosyal ağ analizi yazılımı, Şekil 6.16'te PAJEK sosyal ağ analizi yazılımı ve Şekil 6.17'da GEPHI sosyal ağ analizi yazılımı ile gösterilmiştir, yeni oluşan bağlantılar kırmızı renktedir.

Tablo 6.1. Jaccard İndeks uygulaması sonuçları

Jaccard	Yazar 1	Yazar 2	Yazar 3	Yazar 4	Yazar 5	Yazar 6	Yazar 7	Yazar 8	Yazar 9	Yazar 10	Yazar 11	Yazar 12	Yazar 13	Yazar 14	Yazar 15	Yazar 16	Yazar 17	Yazar 18	Yazar 19	Yazar 20
Yazar 1	-	0,10976	0,24390	0,06818	0,05128	0,01852	0,01869	0,01685	-	0,02299	0,08491	0,02198	0,03968	0,04630	0,06780	0,02362	0,03261	0,25765	0,01031	0,13978
Yazar 2	0,10976	-	0,16000	0,01493	0,02041	-	-	-	-	-	0,02439	-	0,01754	0,02564	0,01370	-	-	0,01739	-	0,03125
Yazar 3	0,24390	0,16000	-	0,02597	0,03390	0,02128	0,02174	-	-	0,03846	0,06000	0,03333	0,02985	0,04082	0,04959	-	0,06452	0,05785	0,02857	0,33333
Yazar 4	0,06818	0,01493	0,02597	-	0,08696	0,01163	0,02381	-	-	0,06452	0,06977	-	0,02857	0,03448	0,22059	0,11458	0,04348	0,07742	0,11940	-
Yazar 5	0,05128	0,02041	0,03390	0,08696	-	0,02985	0,01493	-	-	0,02128	0,02778	0,01961	-	-	0,01370	-	0,03846	0,04930	0,01786	0,01563
Yazar 6	0,01852	-	0,02128	0,01163	0,02985	-	0,03774	-	-	0,02941	0,05172	-	0,02667	-	0,03053	-	0,05128	0,03817	0,02326	-
Yazar 7	0,01869	-	0,02174	0,02381	0,01493	0,03774	-	0,13514	0,07895	0,03030	0,05263	-	-	-	0,01515	-	0,02564	0,01504	0,02381	-
Yazar 8	0,01685	-	-	-	-	-	0,13514	-	0,14141	-	-	0,00917	-	-	-	-	-	0,01471	-	-
Yazar 9	-	-	-	-	-	-	0,07895	0,14141	-	-	-	0,04167	-	-	-	-	-	-	-	-
Yazar 10	0,02299	-	0,03846	0,06452	0,02128	0,02941	0,03030	-	-	-	0,14286	-	-	-	0,03636	0,07843	0,11111	0,02679	0,15000	-
Yazar 11	0,08491	0,02439	0,06000	0,06977	0,02778	0,05172	0,05263	-	-	0,14286	-	0,02326	0,03797	0,03226	0,06061	0,06579	0,04545	0,09302	0,11364	0,05556
Yazar 12	0,02198	-	0,03333	-	0,01961	-	-	0,00917	0,04167	-	0,02326	-	0,01695	-	0,00855	0,01724	-	-	-	0,06061
Yazar 13	0,03968	0,01754	0,02985	0,02857	-	0,02667	-	-	-	-	0,03797	0,01695	-	0,50943	0,05405	0,05435	-	0,01948	0,04839	0,02817
Yazar 14	0,04630	0,02564	0,04082	0,03448	-	-	-	-	-	-	0,03226	-	0,50943	-	0,03759	0,03947	-	0,00725	0,02174	0,03774
Yazar 15	0,06780	0,00870	0,04959	0,22059	0,01370	0,03053	0,01515	-	-	0,03636	0,06061	0,00855	0,05405	0,03759	-	0,15672	0,03448	0,08040	0,06034	0,01550
Yazar 16	0,02362	-	-	0,11458	-	-	-	-	-	0,07843	0,06579	0,01724	0,05435	0,03947	0,15672	-	0,01667	0,03311	0,06667	-
Yazar 17	0,03261	-	0,06452	0,04348	0,03846	0,05128	0,02564	-	-	0,11111	0,04545	-	-	-	0,03448	0,01667	-	0,12037	0,16000	0,02778
Yazar 18	0,11765	0,01739	0,05785	0,07742	0,04930	0,03817	0,01504	0,01471	-	0,02679	0,09302	-	0,01948	0,00725	0,08040	0,03311	0,12037	-	0,10714	0,03937
Yazar 19	0,01031	-	0,02857	0,11940	0,01786	0,02326	0,02381	-	-	0,15000	0,11364	-	0,04839	0,02174	0,06034	0,06667	0,16000	0,10714	-	-
Yazar 20	0,13978	0,03125	0,33333	-	0,01563	-	-	-	-	-	0,05556	0,06061	0,02817	0,03774	0,01550	-	0,02778	0,03937	-	-

Tablo 6.2. Sorensen İndeks Uygulama sonuçları

Sorensen	Yazar 1	Yazar 2	Yazar 3	Yazar 4	Yazar 5	Yazar 6	Yazar 7	Yazar 8	Yazar 9	Yazar 10	Yazar 11	Yazar 12	Yazar 13	Yazar 14	Yazar 15	Yazar 16	Yazar 17	Yazar 18	Yazar 19	Yazar 20
Yazar 1	-	0,19780	0,39216	0,12766	0,09756	0,03636	0,03670	0,03315	-	0,04494	0,15652	0,04301	0,07634	0,08850	0,12698	0,04615	0,06316	0,39353	0,02041	0,24528
Yazar 2	0,19780	-	0,27586	0,02941	0,04000	-	-	-	-	-	0,04762	-	0,03448	0,05000	0,01724	-	-	0,03419	-	0,06061
Yazar 3	0,39216	0,27586	-	0,05063	0,06557	0,04167	0,04255	-	-	0,07407	0,11321	0,06452	0,05797	0,07843	0,09449	-	0,12121	0,10938	0,05556	0,50000
Yazar 4	0,12766	0,02941	0,05063	-	0,16000	0,02299	0,04651	-	-	0,12121	0,13043	-	0,05556	0,06667	0,36145	0,20561	0,08333	0,14371	0,21333	-
Yazar 5	0,09756	0,04000	0,06557	0,16000	-	0,05797	0,02941	-	-	0,04167	0,05405	0,03846	-	-	0,02703	-	0,07407	0,09396	0,03509	0,03077
Yazar 6	0,03636	-	0,04167	0,02299	0,05797	-	0,07273	-	-	0,05714	0,09836	-	0,05195	-	0,05926	-	0,09756	0,07353	0,04545	-
Yazar 7	0,03670	-	0,04255	0,04651	0,02941	0,07273	-	0,23810	0,14634	0,05882	0,10000	-	-	-	0,02985	-	0,05000	0,02963	0,04651	-
Yazar 8	0,03315	-	-	-	-	-	0,23810	-	0,24779	-	-	0,01818	-	-	-	-	-	0,02899	-	-
Yazar 9	-	-	-	-	-	-	0,14634	0,24779	-	-	-	0,08000	-	-	-	-	-	-	-	-
Yazar 10	0,04494	-	0,07407	0,12121	0,04167	0,05714	0,05882	-	-	-	0,25000	-	-	-	0,07018	0,14545	0,20000	0,05217	0,26087	-
Yazar 11	0,15652	0,04762	0,11321	0,13043	0,05405	0,09836	0,10000	-	-	0,25000	-	0,04545	0,07317	0,06250	0,11429	0,12346	0,08696	0,17021	0,20408	0,10526
Yazar 12	0,04301	-	0,06452	-	0,03846	-	-	0,01818	0,08000	-	0,04545	-	0,03333	-	0,01695	0,03390	-	-	-	0,11429
Yazar 13	0,07634	0,03448	0,05797	0,05556	-	0,05195	-	-	-	-	0,07317	0,03333	-	0,67500	0,10256	0,10309	-	0,03822	0,09231	0,05479
Yazar 14	0,08850	0,05000	0,07843	0,06667	-	-	-	-	-	-	0,06250	-	0,67500	-	0,07246	0,07595	-	0,01439	0,04255	0,07273
Yazar 15	0,12698	0,01724	0,09449	0,36145	0,02703	0,05926	0,02985	-	-	0,07018	0,11429	0,01695	0,10256	0,07246	-	0,27097	0,06667	0,14884	0,11382	0,03053
Yazar 16	0,04615	-	-	0,20561	-	-	-	-	-	0,14545	0,12346	0,03390	0,10309	0,07595	0,27097	-	0,03279	0,06410	0,12500	-
Yazar 17	0,06316	-	0,12121	0,08333	0,07407	0,09756	0,05000	-	-	0,20000	0,08696	-	-	-	0,06667	0,03279	-	0,21488	0,27586	0,05405
Yazar 18	0,21053	0,03419	0,10938	0,14371	0,09396	0,07353	0,02963	0,02899	-	0,05217	0,17021	-	0,03822	0,01439	0,14884	0,06410	0,21488	-	0,19355	0,07576
Yazar 19	0,02041	-	0,05556	0,21333	0,03509	0,04545	0,04651	-	-	0,26087	0,20408	-	0,09231	0,04255	0,11382	0,12500	0,27586	0,19355	-	-
Yazar 20	0,24528	0,06061	0,50000	-	0,03077	-	-	-	-	-	0,10526	0,11429	0,05479	0,07273	0,03053	-	0,05405	0,07576	-	-

Tablo 6.3. Ortak Komşu uygulama sonuçları

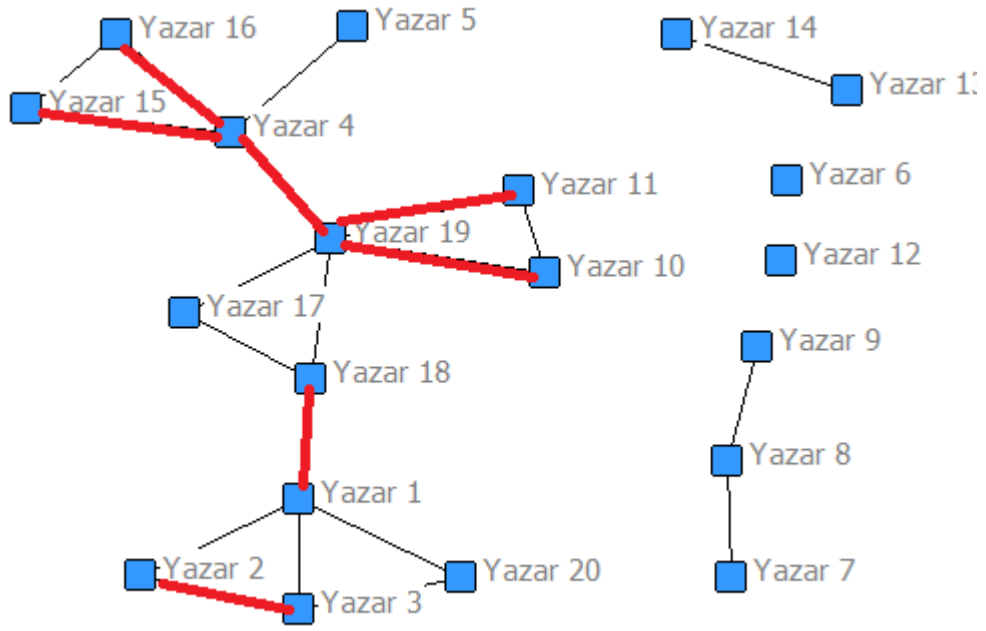
Ortak Komşu	Yazar 1	Yazar 2	Yazar 3	Yazar 4	Yazar 5	Yazar 6	Yazar 7	Yazar 8	Yazar 9	Yazar 10	Yazar 11	Yazar 12	Yazar 13	Yazar 14	Yazar 15	Yazar 16	Yazar 17	Yazar 18	Yazar 19	Yazar 20
Yazar 1	-	9	20	9	6	2	2	3	-	2	9	2	5	5	12	3	3	21	1	13
Yazar 2	9	-	4	1	1	-	-	-	-	-	1	-	1	1	1	-	-	2	-	1
Yazar 3	20	4	-	2	2	1	1	-	-	1	3	1	2	2	6	-	2	7	1	11
Yazar 4	9	1	2	-	8	1	2	-	-	4	6	-	3	3	30	11	3	12	8	-
Yazar 5	6	1	2	8	-	2	1	-	-	1	2	1	-	-	2	-	2	7	1	1
Yazar 6	2	-	1	1	2	-	2	-	-	1	3	-	2	-	4	-	2	5	1	-
Yazar 7	2	-	1	2	1	2	-	15	3	1	3	-	-	-	2	-	1	2	1	-
Yazar 8	3	-	-	-	-	-	15	-	14	-	-	1	-	-	-	-	-	3	-	-
Yazar 9	-	-	-	-	-	-	3	14	-	-	-	1	-	-	-	-	-	-	-	-
Yazar 10	2	-	1	4	1	1	1	-	-	-	5	-	-	-	4	4	2	3	3	-
Yazar 11	9	1	3	6	2	3	3	-	-	5	-	1	3	2	8	5	2	12	5	3
Yazar 12	2	-	1	-	1	-	-	1	1	-	1	-	1	-	1	1	-	-	-	2
Yazar 13	5	1	2	3	-	2	-	-	-	-	3	1	-	27	8	5	-	3	3	2
Yazar 14	5	1	2	3	-	-	-	-	-	-	2	-	27	-	5	3	-	1	1	2
Yazar 15	12	1	6	30	2	4	2	-	-	4	8	1	8	5	-	21	4	16	7	2
Yazar 16	3	-	-	11	-	-	-	-	-	4	5	1	5	3	21	-	1	5	4	-
Yazar 17	3	-	2	3	2	2	1	-	-	2	2	-	-	-	4	1	-	13	4	1
Yazar 18	20	2	7	12	7	5	2	3	-	3	12	-	3	1	16	5	13	-	12	5
Yazar 19	1	-	1	8	1	1	1	-	-	3	5	-	3	1	7	4	4	12	-	-
Yazar 20	13	1	11	-	1	-	-	-	-	-	3	2	2	2	2	-	1	5	-	-

Tablo 6.4. L. H. Newman İndeks Uygulama Sonuçları

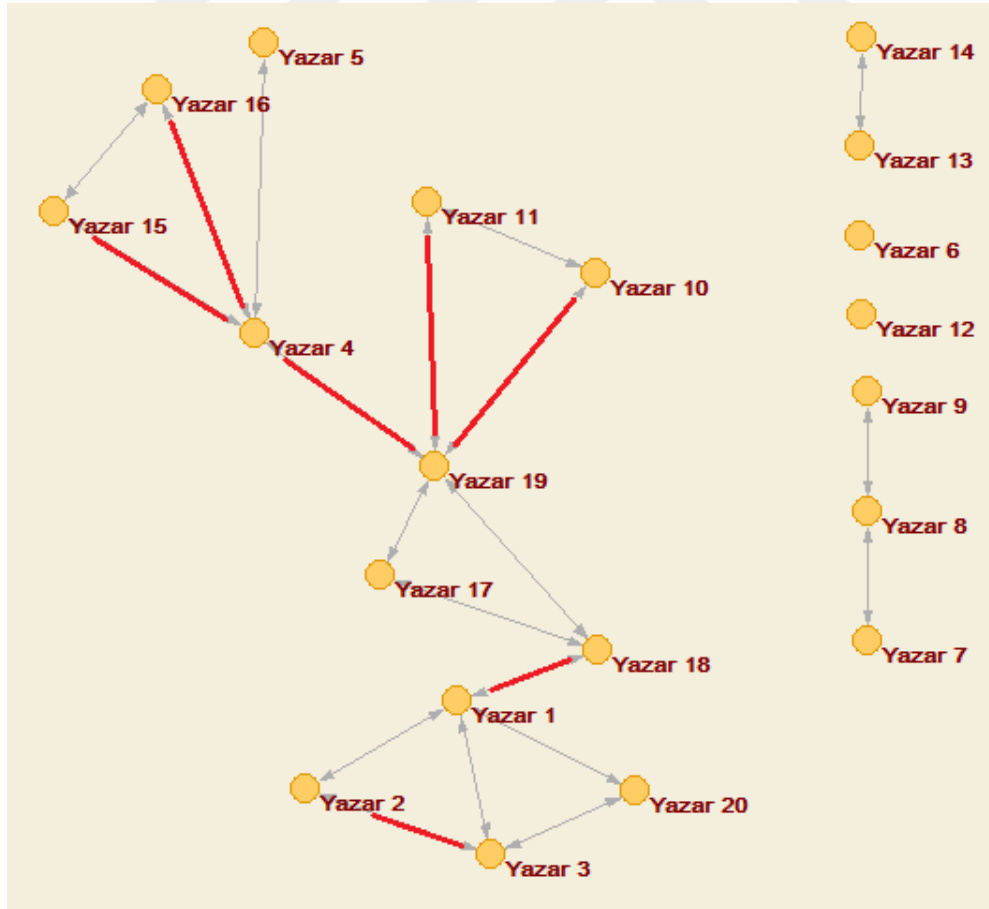
L. H. Newman	Yazar 1	Yazar 2	Yazar 3	Yazar 4	Yazar 5	Yazar 6	Yazar 7	Yazar 8	Yazar 9	Yazar 10	Yazar 11	Yazar 12	Yazar 13	Yazar 14	Yazar 15	Yazar 16	Yazar 17	Yazar 18	Yazar 19	Yazar 20
Yazar 1	-	0,02439	0,02439	0,00372	0,00357	0,00174	0,00181	0,00074	-	0,00697	0,00665	0,00443	0,00249	0,00393	0,00274	0,00152	0,00563	0,02552	0,00152	0,01321
Yazar 2	0,02439	-	0,04444	0,00377	0,00542	-	-	-	-	-	0,00673	-	0,00454	0,00717	0,00208	-	-	0,00412	-	0,00926
Yazar 3	0,02439	0,04444	-	0,00339	0,00488	0,00357	0,00370	-	-	0,01429	0,00909	0,00909	0,00408	0,00645	0,00561	-	0,01538	0,00648	0,00625	0,04583
Yazar 4	0,00372	0,00377	0,00339	-	0,00661	0,00121	0,00251	-	-	0,01937	0,00616	-	0,00208	0,00328	0,00950	0,00777	0,00782	0,00377	0,01695	-
Yazar 5	0,00357	0,00542	0,00488	0,00661	-	0,00348	0,00181	-	-	0,00697	0,00296	0,00443	-	-	0,00091	-	0,00750	0,00316	0,00305	0,00203
Yazar 6	0,00174	-	0,00357	0,00121	0,00348	-	0,00529	-	-	0,01020	0,00649	-	0,00292	-	0,00267	-	0,01099	0,00331	0,00446	-
Yazar 7	0,00181	-	0,00370	0,00251	0,00181	0,00529	-	0,01122	0,01587	0,01058	0,00673	-	-	-	0,00138	-	0,00570	0,00137	0,00463	-
Yazar 8	0,00074	-	-	-	-	0,01122	-	0,02020	-	-	0,00184	-	-	-	-	-	-	0,00056	-	-
Yazar 9	-	-	-	-	-	0,01587	0,02020	-	-	-	0,01299	-	-	-	-	-	-	-	-	-
Yazar 10	0,00697	-	0,01429	0,01937	0,00697	0,01020	0,01058	-	-	-	0,04329	-	-	-	0,01068	0,02381	0,04396	0,00794	0,05357	-
Yazar 11	0,00665	0,00673	0,00909	0,00616	0,00296	0,00649	0,00673	-	-	0,04329	-	0,00551	0,00371	0,00391	0,00453	0,00631	0,00932	0,00673	0,01894	0,00758
Yazar 12	0,00443	-	0,00909	-	0,00443	-	-	0,00184	0,01299	-	0,00551	-	0,00371	-	0,00170	0,00379	-	-	-	0,01515
Yazar 13	0,00249	0,00454	0,00408	0,00208	-	0,00292	-	-	-	-	0,00371	0,00371	-	0,03555	0,00305	0,00425	-	0,00113	0,00765	0,00340
Yazar 14	0,00393	0,00717	0,00645	0,00328	-	-	-	-	-	-	0,00391	-	0,03555	-	0,00301	0,00403	-	0,00060	0,00403	0,00538
Yazar 15	0,00274	0,00208	0,00561	0,00950	0,00091	0,00267	0,00138	-	-	0,01068	0,00453	0,00170	0,00305	0,00301	-	0,00818	0,00575	0,00277	0,00818	0,00156
Yazar 16	0,00152	-	-	0,00777	-	-	-	-	-	0,02381	0,00631	0,00379	0,00425	0,00403	0,00818	-	0,00321	0,00193	0,01042	-
Yazar 17	0,00563	-	0,00625	0,00782	0,00750	0,01099	0,00570	-	-	0,04396	0,00932	-	-	-	0,00575	0,00321	-	0,01852	0,03846	0,00641
Yazar 18	0,00452	0,00412	0,00648	0,00377	0,00316	0,00331	0,00137	0,00056	-	0,00794	0,00673	-	ssss	0,00060	0,00277	0,00193	0,01852	-	0,01389	0,00386
Yazar 19	0,00152	-	-	0,01695	0,00305	0,00446	0,00463	-	-	0,05357	0,01894	-	0,00765	0,00403	0,00818	0,01042	0,03846	0,01389	-	-
Yazar 20	0,01321	0,00926	0,04583	-	0,00203	-	-	-	-	-	0,00758	0,01515	0,00340	0,00538	0,00156	-	0,00641	0,00386	-	-

Tablo 6.5. Salton İndeks Uygulama Sonuçları

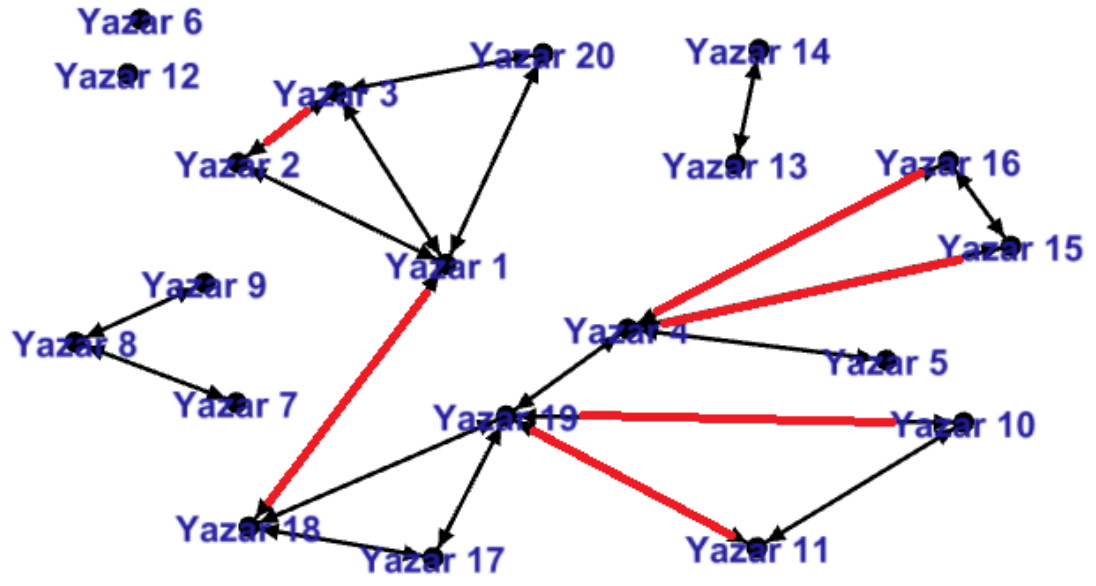
Salton	Yazar 1	Yazar 2	Yazar 3	Yazar 4	Yazar 5	Yazar 6	Yazar 7	Yazar 8	Yazar 9	Yazar 10	Yazar 11	Yazar 12	Yazar 13	Yazar 14	Yazar 15	Yazar 16	Yazar 17	Yazar 18	Yazar 19	Yazar 20
Yazar 1	-	0,94346	1,98030	0,75794	0,54100	0,19069	0,19157	0,22299	-	0,21200	0,83925	0,20739	0,43685	0,47036	0,87287	0,26312	0,30779	1,99095	0,10102	1,26267
Yazar 2	0,94346	-	0,74278	0,12127	0,14142	-	-	-	-	-	0,15430	-	0,13131	0,15811	0,09285	-	-	0,18490	-	0,17408
Yazar 3	1,98030	0,74278	-	0,22502	0,25607	0,14434	0,14586	-	-	0,19245	0,41208	0,17961	0,24077	0,28006	0,53241	-	0,34816	0,61872	0,16667	1,65831
Yazar 4	0,75794	0,12127	0,22502	-	0,80000	0,10721	0,21567	-	-	0,49237	0,62554	-	0,28868	0,31623	2,32845	1,06341	0,35355	0,92859	0,92376	-
Yazar 5	0,54100	0,14142	0,25607	0,80000	-	0,24077	0,12127	-	-	0,14434	0,23250	0,13868	-	-	0,16440	-	0,27217	0,57346	0,13245	0,12403
Yazar 6	0,19069	-	0,14434	0,10721	0,24077	-	0,26968	-	-	0,16903	0,38411	-	0,22792	-	0,34427	-	0,31235	0,42875	0,15076	-
Yazar 7	0,19157	-	0,14586	0,21567	0,12127	0,26968	-	1,33631	0,46852	0,17150	0,38730	-	-	-	0,17277	-	0,15811	0,17213	0,15250	-
Yazar 8	0,22299	-	-	-	-	-	1,33631	-	1,31701	-	-	0,09535	-	-	-	-	-	0,20851	-	-
Yazar 9	-	-	-	-	-	-	0,46852	1,31701	-	-	-	0,20000	-	-	-	-	-	-	-	-
Yazar 10	0,21200	-	0,19245	0,49237	0,14434	0,16903	0,17150	-	-	-	0,79057	-	-	-	0,37463	0,53936	0,44721	0,27975	0,62554	-
Yazar 11	0,83925	0,15430	0,41208	0,62554	0,23250	0,38411	0,38730	-	-	0,79057	-	0,15076	0,33129	0,25000	0,67612	0,55556	0,29488	1,01058	0,71429	0,39736
Yazar 12	0,20739	-	0,17961	-	0,13868	-	-	0,09535	0,20000	-	0,15076	-	0,12910	-	0,09206	0,13019	-	-	-	0,33806
Yazar 13	0,43685	0,13131	0,24077	0,28868	-	0,22792	-	-	-	-	0,33129	0,12910	-	3,01869	0,64051	0,50767	-	0,23943	0,37210	0,23408
Yazar 14	0,47036	0,15811	0,28006	0,31623	-	-	-	-	-	-	0,25000	-	3,01869	-	0,42563	0,33753	-	0,08482	0,14586	0,26968
Yazar 15	0,87287	0,09285	0,53241	2,32845	0,16440	0,34427	0,17277	-	-	0,37463	0,67612	0,09206	0,64051	0,42563	-	1,68676	0,36515	1,09119	0,63117	0,17474
Yazar 16	0,26312	-	-	1,06341	-	-	-	-	-	0,53936	0,55556	0,13019	0,50767	0,33753	1,68676	-	0,12804	0,40032	0,50000	-
Yazar 17	0,30779	-	0,34816	0,35355	0,27217	0,31235	0,15811	-	-	0,44721	0,29488	-	-	-	0,36515	0,12804	-	1,18182	0,74278	0,16440
Yazar 18	1,45095	0,18490	0,61872	0,92859	0,57346	0,42875	0,17213	0,20851	-	0,27975	1,01058	-	0,23943	0,08482	1,09119	0,40032	1,18182	-	1,07763	0,43519
Yazar 19	0,10102	-	0,16667	0,92376	0,13245	0,15076	0,15250	-	-	0,62554	0,71429	-	-	0,14586	0,63117	0,50000	0,74278	1,07763	-	-
Yazar 20	1,26267	0,17408	1,65831	-	0,12403	-	-	-	-	-	0,39736	0,33806	0,23408	0,26968	0,17474	-	0,16440	0,43519	-	-



Şekil 6.15. UCINET ile yeni yazar-yazar ağı



Şekil 6.16. PAJEK ile yeni yazar-yazar ağı



Şekil 6.17. GEPHI ile yeni yazar-yazar ağı

7. SONUÇ VE ÖNERİLER

Özellikle sosyal ağların gelişmesi ile birlikte bağlantı tahmini problemi popüler bir hal almıştır. Günümüzde birçok alanda yaygın olarak kullanılmaktadır. Bazıları; sosyal ağlarda kullanıcıların arkadaşlık ilişkilerini tahmin etme, yazar-makale ağlarında yazarların gelecekte yapacakları yayın konularını tahmin etme ve yazarların ortak yayın çıkarma olasılığını belirleme, hastalık-ilaç ağlarında hastalıkları tedavi etmede kullanılabilecek ilaçların belirlenmesi alanlarıdır.

Bağlantı tahmininin başarımlarının yüksek olabilmesi için seçilecek kriterlerin ve parametrelerin doğru belirlenmesi gerekmektedir. Ayrıca bağlantı tahmini problemini çözme yöntemleri graf yapıları üzerinde uygulandığından, graf yapılarının da gayet iyi bilinmesi gerekmektedir.

Bu çalışmada bağlantı tahmini yöntemleri akademik alanda yazar ve yayınladıkları makalelerin anahtar kelimeleri arasındaki ilişkiler üzerine uygulanmış ve çalışma konuları arasındaki benzerlikler ortaya konmuştur. Özellikle yerel indeksleme yöntemlerinden Jaccard Index, Sorensen Index, Ortak Komşu, L. H. Newman Index, Salton Index yöntemlerinin analiz işlemleri için matematiksel model olarak uygun olması sebebi ile bu yöntemler seçilmiştir.

Yapılan analizler sonucunda veri setindeki yazarların çalışma alanları arasında benzerlikler tespit edilmiştir. Özellikle bu benzerliklerden yola çıkarak hangi yazarların birbiri ile birlikte çalışma yapabileceği hakkında tahminler yapılmıştır. Tahmin için yapılan hesaplamalarda görülmüştür ki, ortak yayın veya ortak anahtar kelime sayısı (ortak komşu indeks yönteminde net olarak görülmektedir) tek başına doğru sonuçlar vermemektedir.

Yapılan çalışmada; yazar adı, yayın yılı, kurum, yayın başlığı, anahtar kelimeler, kontrol terimleri, özet, ISSN (Uluslararası Standart Süreli Yayın Numarası), ISBN (Uluslararası Standart Kitap Numarası) parametreleri IEEE Explore veritabanından çekilebilmektedir. Bu parametreler arasından yazar adı ve anahtar kelimeler benzerlik ölçümü için en baskın terimler olarak tespit edilmiştir. Özet, yayın başlığı, ISSN ve ISBN'nin benzerlik ölçümü için yararlı özellikler olmadığı ve yayın yılı, kurum, kontrol terimlerinin ise yazar adı ve anahtar kelimeler ile birlikte kullanılarak benzerlik ölçümünün daha hassas bir şekilde tahmin yapılması sağlanabilir. Bu parametrelerin eklenmesinin

dezavantajı ise işlem yükünün artması ve eklenecek parametrelerin (düğümlerin) ağırlıklarının hesaplanma zorluğudur.

Yapılan uygulamanın doğruluğunun tespiti için, çalışma tüm özellikleri bilinen bir yazar-makale ağında gerçekleştirilmiştir. Ağda günümüzde var olan bazı bağlantıların olmadığı zamana gidilerek ağın günümüzdeki hali tahmin edilmeye çalışılmış ve başarılı sonuçlar elde edilmiştir. Gelecekteki çalışmalarda bu parametreler eklenerek daha detaylı çalışmalar yapılması amaçlanmaktadır.



KAYNAKLAR

- [1] Sen, P., Namata, G., Bilgic, M., Getoor, L., Gallagher B., Eliassi-Rad T., 2008. Collective Classification in Network Data. *AI Magazine*, 29 (3), 93-106.
- [2] Hwang, S.Y., Wylie, B.N., Wei, C.P. ve Liao, Y.F., 2010. Coauthorship Networks And Academic Literature Recommendation, *Electronic Commerce Research and Applications*, 9, 323–334.
- [3] Barabási, A. L. ve Albert, R., 1999. Emergence Of Scaling in Random Networks, *Science* 286, 509–512.
- [4] Hasan, M.A., Chaoji, V., Salem, S. ve Zaki, M.J., 2006. Link Prediction Using Supervised Learning, (SIAM 2006) Workshop on Link Analysis, Counterterrorism and Security, Bethesda, MD.
- [5] Folino, F. ve Pizzuti, C., 2010. A Comorbidity-Based Recommendation Engine For Disease Prediction, *Proc of 23rd IEEE International Symposium on Computer Based Medical Systems (CBMS 2010)*, 6-12.
- [6] Fakhraei, S., Huang, B., Raschid, L., Getoor, L., 2014. Network-Based Drug-Target Interaction Prediction With Probabilistic Soft Logic, *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 11(5), 775-787.
- [7] Lakshmi, T. J., Bhavani, S. D., 2014, Heterogeneous Link Prediction Based On Multi Relational Community Information. In *Communication Systems and Networks (COMSNETS)*, Sixth International Conference on, 1-4, IEEE.
- [8] Biuk-Aghai, R. P., 2006. Visualizing co-authorship networks in online Wikipedia. In *Communications and Information Technologies, 2006. ISCIT'06. International Symposium on* (pp. 737-742). IEEE.
- [9] Krömer, P., Snášel, V., Platoš, J., Kudělka, M., and Horák, Z. 2012. An ACO inspired weighting approach for the spectral partitioning of co-authorship networks. In *Evolutionary Computation (CEC), 2012 IEEE Congress on* (pp. 1-7). IEEE.
- [10] Tsolakidis, A., Sgouropoulou, C., Papageorgiou, E., Terraz, O., and Miaoulis, G. 2012. Co-authorship networks in academic research communities: the role of network strength. In *Informatics (PCI), 2012 16th Panhellenic Conference on* (pp. 150-155). IEEE.

- [11] Meng, Q., and Kennedy, P. J. 2012. Determining the number of clusters in co-authorship networks using social network theory. In *Cloud and Green Computing (CGC), 2012 Second International Conference on* (pp. 337-343). IEEE.
- [12] Zehnalova, S., Horak, Z., Kudelka, M., and Snasel, V. 2012. Evolution of Author's Topic in Authorship Network. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on* (pp. 1207-1210). IEEE.
- [13] Ahmedi, L. 2012. AuthorRank+ FOAF: ranking for co-authorship networks on the web. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)* (pp. 315-321). IEEE Computer Society.
- [14] Meng, Q., and Kennedy, P. J. 2012. Using field of research codes to discover research groups from co-authorship networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on* (pp. 289-293). IEEE.
- [15] Choobdar, S., Ribeiro, P., Bugla, S., and Silva, F. 2012. Comparison of co-authorship networks across scientific fields using motifs. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on* (pp. 147-152). IEEE.
- [16] Song, M., Yang, C. C., Tang, X., and Han, W. S. 2012. Mapping the field of Bioinformatics with a content and co-authorship analysis. In *Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on* (pp. 774-781). IEEE.
- [17] Krömer, P., Kudelka, M., Horák, Z., and Snáel, V. 2013. Evolution of Scale-Freeness in a Co-authorship Network. In *Intelligent Networking and Collaborative Systems (INCoS), 2013 5th International Conference on* (pp. 668-672). IEEE.
- [18] Bento, C., and Takeda, H. 2013. Finding Research Communities and their Relationships by Analyzing the Co-authorship Network. In *Information Visualisation (IV), 2013 17th International Conference* (pp. 132-140). IEEE.
- [19] Le Hoang, N., Khoa, P. V. D., and Phuc, D. 2013. Predicting preferred topics of authors based on co-authorship network. In *Computing and Communication*

- Technologies, Research, Innovation, and Vision for the Future (RIVF)*, 2013 *IEEE RIVF International Conference on* (pp. 70-75). IEEE.
- [20] Huang, S., Tang, Y., Tang, F., and Li, J. 2014. Link prediction based on time-varied weight in co-authorship network. In *Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on* (pp. 706-709). IEEE.
 - [21] Newman, M., 2010. *Networks: an introduction*, Oxford University Press.
 - [22] Voigt, K., 2011. *Structural Graph-Based Metamodel Matching*, Doctoral dissertation, Technische Universität Dresden, Berlin, Germany
 - [23] Ying, X., Wu, X., 2008. Randomizing Social Networks: a Spectrum Preserving Approach, In *SDM*, 8, 739-750.
 - [24] Barrat, A., Barthélemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The Architecture Of Complex Weighted Networks, *Proceedings of the National Academy of Sciences of the United States of America*, 101(11), 3747-3752.
 - [25] Lin, P., Jia, Y., Li, L., 2008. Distributed Robust H_∞ Consensus Control In Directed Networks Of Agents With Time-Delay, *Systems and Control Letters*, 57(8), 643-653.
 - [26] Abdulla, P. A., Atig, M. F., Rezine, O., 2013. Verification Of Directed Acyclic Ad Hoc Networks, In *Formal Techniques for Distributed Systems*, 193-208, Springer Berlin Heidelberg.
 - [27] Dormann, C. F., Strauss, R., 2014. A Method For Detecting Modules In Quantitative Bipartite Networks, *Methods in Ecology and Evolution*, 5(1), 90-98.
 - [28] Deo, N., 2016. *Graph theory with applications to engineering and computer science*, Courier Dover Publications.
 - [29] Bollobás, B., 2013. *Modern graph theory* (Vol. 184), Springer Science and Business Media.
 - [30] Saxena, A., Malik, V., Iyengar, S. R. S., 2015. Estimating the Degree Centrality Ranking of a Node, *arXiv preprint arXiv:1511.05732*.
 - [31] https://en.wikipedia.org/wiki/Centrality#Eigenvector_centrality 8 Aralık 2016
 - [32] Katz, L., 1953. A New Status Index Derived From Sociometric Analysis, *Psychometrika*, 18, 39-43.

- [33] Aprahamian, M., Higham, D. J., Higham, N. J., 2015. Matching Exponential-Based And Resolvent-Based Centrality Measures, *Journal of Complex Networks*, cnv016.
- [34] Devi, P., Gupta, A., Dixit, A., 2014. Comparative Study of HITS and PageRank Link Based Ranking Algorithms, *International Journal of Advanced Research in Computer and Communication Engineering*, 3(2).
- [35] Brin, S., Page, L., 2012. Reprint of: The Anatomy Of A Large-Scale Hypertextual Web Search Engine, *Computer networks*, 56(18), 3825-3833.
- [36] Freeman, L. C., 2004. *The Development of Social Network Analysis*, Empirical Press, Vancouver.
- [37] Anthonisse, J. M., 1971. The Rush In A Directed Graph, Technical Report BN 9/71, StichtingMathematisch Centrum, Amsterdam.
- [38] Valari, E., Papadopoulos, A. N., 2013. Continuous Similarity Computation over Streaming Graphs, In *Machine Learning and Knowledge Discovery in Databases*, 638-653, Springer Berlin Heidelberg.
- [39] Fujimoto, K., Valente, T. W., 2012. Social Network Influences On Adolescent Substance Use: Disentangling Structural Equivalence From Cohesion, *Social Science and Medicine*, 74(12), 1952-1960.
- [40] Batagelj, V., Doreian, P., and Ferligoj, A., 1992. An optimizational approach to regular equivalence. *Social Networks*, 14(1-2), 121-135.
- [41] Salton, G., 1989, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, Reading, MA.
- [42] Hamers, L., Hemeryck, Y., Herweyers, G., Janssen, M., Keters, H., Rousseau, R., Vanhoutte, A., 1989, Similarity Measures In Scientometric Research: The Jaccard Index Versus Salton's Cosine Formula. *Information Processing and Management*, 25(3), 315-318.
- [43] Monedero, I., Biscarri, F., León, C., Guerrero, J. I., Biscarri, J., Millán, R., 2012, Detection of Frauds and Other Non-Technical Losses in A Power Utility Using Pearson Coefficient, Bayesian Networks and Decision Trees. *International Journal of Electrical Power and Energy Systems*, 34(1), 90-98.
- [44] Chang, W., Lin, T., 2010. A Cluster-Based Approach for Automatic Social Network Construction, *IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust*, 601-606.

- [45] Ting, I.H., 2008. Web Mining Techniques for On-line Social Network Analysis, IEEE, 696-700.
- [46] <https://sites.google.com/site/ucinetsoftware/home> UCINET'in özellikleri 10 Nisan 2017
- [47] <http://vlado.fmf.uni-lj.si/pub/networks/pajek> PAJEK'in özellikleri 10 Nisan 2017
- [48] <http://www.netminer.com> NETMİNER'in özellikleri 10 Nisan 2017
- [49] Huisman, M., ve van Duijn, M.A.J., 2011. The SAGE Handbook of Social Network Analysis, s. 578-600
- [50] <https://gephi.org> GEPHI'nin özellikleri 10 Nisan 2017
- [51] <http://igraph.org> IGRAPH'in özellikleri 11 Nisan 2017
- [52] <http://www.statnet.org> STATNET'in özellikleri 11 Nisan 2017
- [53] socnetv.org SOCNETV'nin özellikleri 11 Nisan 2017
- [54] Lo, S., and Lin, C., 2006. WMR--A graph-based algorithm for friend recommendation. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 121-128). IEEE Computer Society.
- [55] Barabási, A. L., 2009. From networks to human behavior, In *Proceedings of Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 435.
- [56] Wang, C., Satuluri V. and Parthasarathy, S., 2007, Local Probabilistic Models For Link Prediction. In *Proceedings of 7th IEEE International Conference on Data Mining*, 322–331.
- [57] Xiang, E. W., 2008. A survey on link prediction models for social network data. *Science and Technology*.
- [58] Lü, L., and Zhou, T., 2011. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6), 1150-1170.
- [59] P. Jaccard., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura, *Bull. Soc. Vaud. Sci. Nat.* 37-547
- [60] G. Salton, M.J. McGill, 1983. *Introduction to Modern Information Retrieval*, McGraw-Hill, Auckland.

- [61] T. Sørensen, 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons, *Biol. Skr.* 5-1.
- [62] Leicht, E. A., Holme, P., and Newman, M. E., 2006. Vertex similarity in Networks, *Physical Review E*, 73(2), 026120.
- [63] Adamic, L. A., and Adar, E., 2003. Friends and neighbors on the web, *Social networks*, 25(3), 211-230.
- [64] Zhou, T., Lü, L., and Zhang, Y. C., 2009. Predicting missing links via local information, *The European Physical Journal B*, 71(4), 623-630.
- [65] Q. Ou, Y.-D. Jin, T. Zhou, B.-H. Wang, B.-Q. Yin, 2007. Power-law strength-degree correlation from resource-allocation dynamics on weighted networks, *Phys. Rev. E* 75, 021102.
- [66] L. Katz, 1953. A new status index derived from sociometric analysis, *Psychometrika* 18-39.
- [67] W. Liu, L. Lü, 2010. Link prediction based on local random walk, *Europhys. Lett.* 89, 58007
- [68] S. Brin, L. Page, 1998. The anatomy of a large-scale hypertextual web search engine, *Comput. Netw. ISDN Syst.* 30-107
- [69] G. Jeh, J. Widom, 2002. SimRank: a measure of structural-context similarity, in: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, pp. 271–279

ÖZGEÇMİŞ

Yücel BÜRHAN

Kişisel Bilgiler

Doğum Tarihi : 19.01.1985

Doğum Yeri : Bingöl

Medeni Durumu : Bekar

Eğitim:

Lise : 1999-2003 Elazığ Anadolu Lisesi

Lisans : 2004-2009 Fırat Üniversitesi Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü

Çalıştığı Kurumlar:

2011- Devam ediyor Munzur Üniversitesi Tunceli MYO

2009-20011 Servus Bilgisayar A.Ş.