



**INSTAGRAM SOSYAL AĞINDAN DÖRTLÜ ÖRÜNTÜ
KEŞFİ VE ETKİ ANALİZİNİ GERÇEK ZAMANLI
GERÇEKLEŞTİRME**

Ahmet Anıl MÜNGEN
Yüksek Lisans Tezi

Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Mehmet KAYA
Haziran-2016

T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

INSTAGRAM SOSYAL AĞINDAN DÖRTLÜ ÖRÜNTÜ KEŞFİ VE ETKİ ANALİZİNİ
GERÇEK ZAMANLI GERÇEKLEŞTİRME

YÜKSEK LİSANS TEZİ

Ahmet Anıl MÜNGEN

(141129112)

Tezin Enstitüye Verildiği Tarih: 16 Mayıs 2016
Tezin Savunulduğu Tarih: 03 Haziran 2016

Tez Danışmanı : Prof. Dr. Mehmet KAYA

Diğer Jüri Üyeleri : Doç. Dr. Muhammed Fatih TALU

:Yrd. Doç. Dr. Taner TUNCER

HAZİRAN 2016

ÖNSÖZ

Sosyal ağ analizi, sosyal ağların gelişmeye başladığı ilk zamanlar olan 2000'li yıllardan beri çalışılan ve bu ağların büyümesine paralel olarak önem kazanan bir araştırma konusudur. Günümüzde sosyal ağlar kullanıcının hem kişisel bilgilerinin hem de ilgi alanları ve beğenilerinin yansıtıldığı platformlar haline dönüşmüştür. Sosyal ağları izleyerek ve takip ederek bir kullanıcının ilgi alanı ve beğenilerini ayrıca kullanıcıların birbirleri arasındaki ilişkilerine dair birçok bilgi elde edilebilir. Sosyal ağ analizi ile bu ilişkiler tespit edilebilir, büyüyen ve karmaşıklaşan sistemde kullanıcı gözü ile görülemeyecek bağlantılar sosyal ağ analizleri ile ortaya çıkartılabilir. Topluluk keşfetme ve topluluk içinde kullanıcı bazında etki değeri belirleme işlemleri yapılabilir.

Lisans eğitimimin başından mezun olduğum güne kadar, çağımızın modern bilgileri ile donanmış, araştırmacı bir kafaya sahip, üretken bir bilgisayar mühendisi olma hayalimde her aşamada kendisini örnek aldığım; keza hayatımın ileriki yıllarını da etkileyecek yüksek lisans tezimin seçimi ve yürütülmesi esnasında ufku açan yol göstericiliği ve değerli katkıları olan sayın hocam Prof. Dr. Mehmet KAYA'ya en derin teşekkürlerimi sunarım.

Ayrıca bu tez kapsamında almış olduğum 2211 - Yurt İçi Lisansüstü Burs Programı (2015) kapsamında "Öncelikli Alanlara Yönelik Teşvik Bursu" nedeniyle TÜBİTAK'a teşekkürlerimi bir borç bilirim (No. 1649B021508870).

Son olarak çalışmalarım sırasında her zaman yanımda olan aileme sonsuz teşekkürlerimi sunarım.

Ahmet Anıl MÜNGEN
ELAZIĞ – 2016

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ	II
İÇİNDEKİLER	III
ŞEKİLLER LİSTESİ	V
TABLolar LİSTESİ	VII
ÖZET	VIII
SUMMARY	IX
1. GİRİŞ	1
1.1 Tezin Amacı	3
1.2. Tezin Organizasyonu.....	3
2. SOSYAL AĞLARDA ÖRÜNTÜ KEŞİFLERİ	5
2.1. Karar Ağaçları ile Örüntü Keşfi	5
2.2. Genetik Algoritmalar ile Örüntü Keşfi.....	7
2.3. Sıklık Desenleri ile Örüntü Keşfi	9
2.4. Segmentasyon Matris İndeksi ile Bağlı Grup Keşfi	11
2.5. Uzaklık Odaklı Örüntü Keşfi.....	13
2.6. Üçlü Kapanış ile Örüntü Keşfi	16
2.6.1. Üçlü Kapanış Kullanım Alanları	16
2.6.2. Üçlü Kapanış Probleminin Tanımlanması.....	17
2.6.3. Veri Toplama ve Gözlem	18
2.6.4. Üçlü Kapanış Faktör Çizgesi.....	20
2.6.4.1. Üçlü Kapanış Faktör Çizgesi Sonuçları.....	21
3. DÖRTLÜ KAPANIŞ İLE ÖRÜNTÜ KEŞFİ	22
3.1. Dörtlü Kapanış için Veri Toplama	22
3.2. Seçilen Öznitelikler ve Gözlemler.....	24
3.2.1. Ağ Topolojileri	24
3.2.2. Kullanıcı Dilleri.....	26
3.2.3. Kullanıcıların Aktif Oldukları Saat Dilimleri.....	27
3.2.4. Yorumlarda Duygu Analizi	29
3.2.5. Sosyal Roller	30

3.2.6. Geçişlilik.....	31
3.2.7. Özet.....	32
3.3. Dörtlü Kapanış Metodu	32
3.4. Dörtlü Kapanış Metodu Deney Sonuçları	34
4. SOSYAL AĞLARDA ETKİ ANALİZİ.....	37
4.1. İstatistiksel Analiz ile Etki Analizi.....	37
4.2. Benzerlik ile Etki Analizi	38
4.3. Sosyal Birleşmeler ve İstatistiki Araçlar ile Etki Analizi.....	38
4.4. Yerel Yakınlık Yayılımı ile Etki Analizi.....	39
4.4.1. Yerel Faktör Çizgesi (TFG).....	40
4.4.2. Test Sonuçları	42
4.5. Geliştirilmiş Yerel Yakınlık Yayılımı (I-TAP)	42
5. SONUÇ	45
KAYNAKLAR.....	46
ÖZGEÇMİŞ	54

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1. Sosyal ağları kullanan Amerikan internet kullanıcıların oranları ve yıllara göre değişimi.....	1
Şekil 1.2. Sosyal ağları kullanan Amerikan internet kullanıcıların yaşlarına ve yıllara göre değişimi	2
Şekil 1.3. Sosyal ağları kullanan Amerikan internet kullanıcıların sosyal ağlara ve yaşlarına göre değişimi	2
Şekil 2.1. Gruptaki arkadaş sayısına(k) göre kullancının gruba dahil olma olasılığı.....	6
Şekil 2.2. Girvan ve Newman'ın çalışmasında modellenen 3 topluluklu ve 21 düğümlü ağ.	8
Şekil 2.3. Girvan ve Newman'ın çalışmasında modellenen ve grup içi bağlantıları gösteren, bağlantılardan üçlü olanların koyulaştırıldığı gösterim şekli	8
Şekil 2.4. Girvan ve Newman'ın çalışmasında modellenen ve birbiri ile yayın yapmış akademisyenleri gösteren tablo	9
Şekil 2.5. Kullanıcılar arası etkileşimin çizge üzerinde gösterimi	11
Şekil 2.6. İki grup arasında segregation matriks indexi	13
Şekil 2.7. DCDD'nin bu set üzerine uygulanmasında çıkan sonuç	16
Şekil 2.8. Açık Üçlü, Üçlü Kapanış ve üçlü kapanış'dan açık üçlü çıkartmak.....	18
Şekil 2.9. Huang ve Arkadaşlarının Faktörlerinden Konum, Cinsiyet ve Popülariteye göre korelasyon değerleri	19
Şekil 2.10. TriadGF modelinin grafikte gösterimi	20
Şekil 2.11. Haung ve arkadaşları yaptıkları Triad-FG modeli Weibo'da uygulanması.....	21
Şekil 3.1. Çalışmada kullanılan 5 farklı topoloji tipi.....	25
Şekil 3.2. En sık kullanılan ve korelasyonu hesaplanan 7 dilin kullanım oranları sunumu	26
Şekil 3.3. Veri havuzumuzda toplanan kullanıcıların dillerine göre Avrupa Kıtasında dağılımı.....	27
Şekil 3.4. Saat dilimine göre aktif kullanıcıların frekansı	28
Şekil 3.5. Kullanıcıların kendi aktif oldukları saat diliminde çeşitli popülerlikteki kullanıcılar ile etkileşimi	28
Şekil 3.6. Kullanıcıların kendi aktif oldukları saat diliminin 8 farklı modda gösterimi.....	29
Şekil 3.7. Yorumların duygu analizi genel sonuçları gösterimi	30

Şekil 3.8. Kullancıların sosyal rollerine göre kendi grupları ile korelasyon değerleri.....	31
Şekil 3.9. Beş aktörlü bir dörtlü kapanış modeli	33
Şekil 3.10. Çalışma sonucunda bulunmuş 3 örnek dörtlü kapanış	36
Şekil 3.11. Altı aktörlü örnek bir etki çizelgesi	36
Şekil 4.1. Temporal Group Algoritması ile bulunan etkileşimleri	39
Şekil 4.2. Stream Modeli ile bulunan etkileşimlerin gösterimi	39
Şekil 4.3. Ortak Yazar ağından sosyal etki ve TAP metodunda sonuçları.....	40
Şekil 4.4. TFG Model Gösterimi	41
Şekil 4.5. Showing how model learning algorithm is finding influence value	43
Şekil 4.6. Altı aktörlü bir etki grafiği örneği	44



TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 2.1. Lars Backstrom'ın kullandığı Karar Ağacı Örneği Öz nitelikleri	6
Tablo 2.2. Bir önceki yıl C konferansında makale yayınlamamış ama yayınlayan akademisyenler ile daha önceki veya sonraki çalışmalarının en az birinde birlikte makale yazmış kişilerin bir sonraki C konferansında makale yayınlama olasılıkları	7
Tablo 2.3. Kullanılan kelime frekanslarına göre kullanıcılar arası sıklık analizi.....	10
Tablo 2.4. DCCD'nin diğer algoritmalara göre başarısı	15
Tablo 2.5. Üçlü Kapanış Tahmin Başarı Performansı.....	21
Tablo 3.1. Dörtlü Kapanış Tahmin Performansı	34
Tablo 3.2. Faktör Etki Analizi	34

ÖZET

Bu tezde sosyal ağ analizi ile topluluk ve örüntü keşfi ve topluluk içinde etki değeri tespiti ele alınmıştır. Sosyal ağlar kullanıcıların aralarında birden fazla yön ve birden fazla şekilde etkileşim oluşturabildikleri platformlardır. Sosyal ağlardaki ilişkileri daha iyi anlayabilmek matematikçilerin, bilgisayar bilimcilerin, ekonomistlerin ve sosyologlar ile birlikte birçok bilim dalının ilgi alanına girer. Örüntü keşfi için grup analizi, karar ağaçları, genetik algoritmalar, sıklık desenleri, Segregasyon Matris indeksi ve uzaklık odaklı grup tespiti yapılan çalışmalardan bazılarıdır. Topluluk içinde etki değeri bulmak için istatistik analiz, benzerlik, sosyal birleşmeler ve istatistiki araçlar ve Yerel Yakınlık Yayılımı (TAP) ise bazı öne çıkan çalışmalardır.

Sosyal ağlarda bulunan veriler ile kullanıcının bilgilerini keşfetmek, kullanıcılar arasında ilişkileri tanımlamak, kullanıcının dahil olduğu grupları keşfetmek ve kullanıcının grup içi etkinliğini ölçmek gibi analizler yapılabilmektedir. Bu tezde öncelikle sosyal ağlarda örüntü keşfi ile ilgili öne çıkan araştırmalar incelenmiş ve karşılaştırmalı olarak sunulmuştur. Daha sonra; Huang ve Tang'ın çok yönlü vektörler ile üçlü ilişki (triadic closure) metodunu temel alarak geliştirilen ve Quad Closure metodu olarak adlandırılan yöntem önerilmiştir. Bununla birlikte; etki analizi konusunda yapılmış çalışmalardan bahsedilmiş ve Tang ve Sun'ın Yerel Yakınlık Yayılımı (TAP) yaklaşımını temel alan ve quad closure method'un sonuçlarında etki analizi yapabilen Geliştirilmiş Yerel Yakınlık Yayılımı metodu önerilmiştir.

Anahtar Kelimeler: Örüntü Keşfi, Etki analizi, Quad Closure, Yerel Yakınlık Yayılımı.

SUMMARY

QUAD PATTERN DISCOVERY AND IMPACT ANALYSE ON THE INSTAGRAM SOCIAL NETWORKS IN REAL TIME

In this study, we focus on pattern discovery on social networks and impact analyze at communities. Social Networks provide a platform that users can interact with each other in many ways. A lot of sciences like computer science, economy and sociology, are interesting in community discovery. Group analysis, decision trees, genetic algorithms, frequency patterns, segregation Matrix Index and Centrality based Community Detection are some approaches for pattern discovery. Statistical analysis, similarity, social cohesion and statistical tools and Topical Affinity Propagation (TAP) are prominent methods for impact analyze in communities.

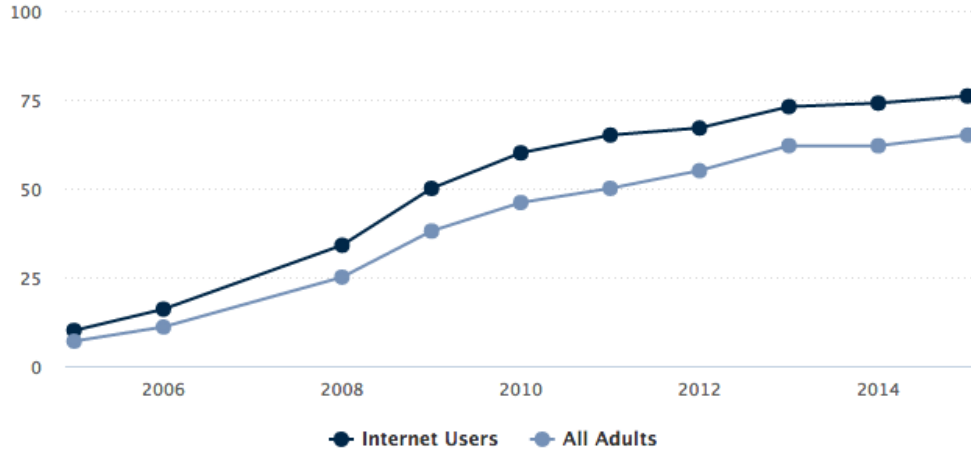
It is possible to define user's information, relationship within users, the groups that the user belongs, user activity in group. We firstly reviewed principal works related with community discovery at social networks in comparison. Then, we proposed a new method named Quad Closure Method which is based on Huang and Tang's triadic closure pattern method. On the other hand, we gave information about different impact analyze methods and we developed Tang and Sun's Topical Affinity Propagation (TAP) method to analyze impact factor on our Quad Closure Method's results.

Keywords: Patern Discovery, Impact Analyze, Quad Closure, Topical Affinity Propagation.

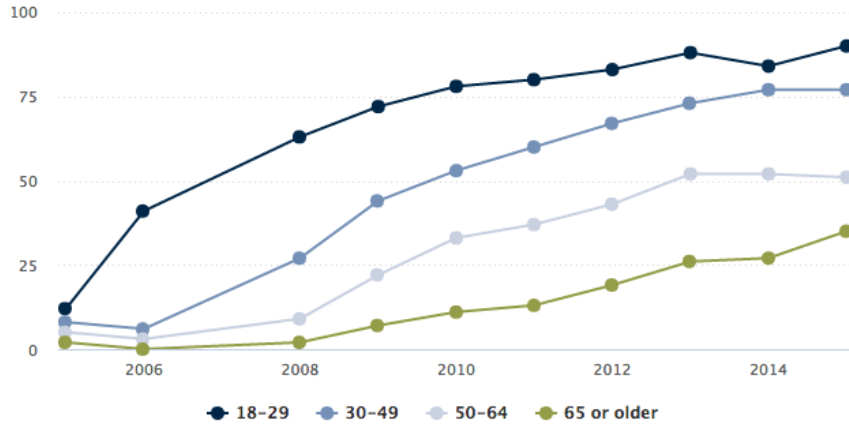
1. GİRİŞ

Sosyal ağlar kullanıcıların genellikle birbirlerine arkadaşlık, takip, beğeni veya yorum yazmak sureti ile etkileşim kurdukları bilgi ağlarıdır. Çalışmanın bu bölümünde sosyal ağları oluşturan kullanıcılar yani düğümler ve düğümler arası ilişkileri gösteren kenarlar açıklandı.

Büyük oranda kullanıcı tarafından veri üretilen sosyal medya sitelerinin kullanım oranı; artan mobil internet bağlantı hizmetleri ve mobil cihaz sayıları ile birlikte her geçen gün artmaktadır. Ocak 2014 de yapılan bir araştırma [1] yetişkinlerin %74'ünün en az bir sosyal medya hesabı olduğu Şekil 1.1 ve Şekil 1.2'de gösterilmektedir. Amazon Alexa [1] verilerine göre dünya çapında en çok ziyaret edilen ilk 500 sitenin yaklaşık üçte biri sosyal medya sitelerinden oluşmaktadır. Kullanıcıların yorum ve beğenileri ile içeriğe katkı sağladığı web siteleri de listeye eklendiğinde bu oran üçte ikiye kadar çıkmaktadır [2].

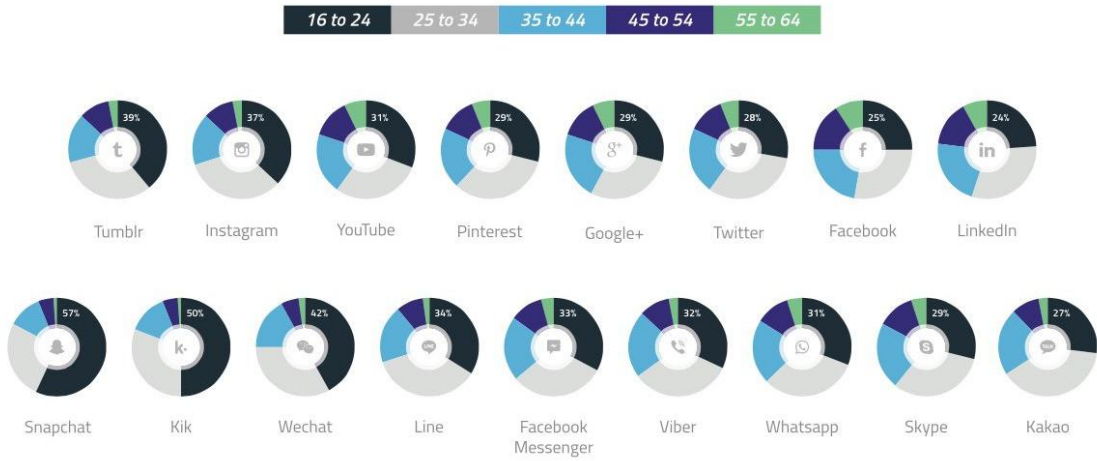


Şekil 1.1 Sosyal ağları kullanan Amerikan internet kullanıcıların oranları ve yıllara göre değişimi[1].



Şekil 1.2 Sosyal ağları kullanan Amerikalı internet kullanıcıların yaşlarına ve yıllara göre değişimi [1]

İnternet üzerinden kullanıcıların hizmetine sunulmuş onlarca farklı sosyal ağ mevcuttur. Bu sosyal ağların çoğu belli bir tematik alana yönelmiştir. Tematik alanlardan önce çıkanlar fotoğraf paylaşımı, video paylaşımı ve mikroblog olarak tarif edilen sınırlı kelime sayısı ile fikir paylaşımı yapan web siteleridir. Bunların dışında birden çok paylaşım türünü kapsayan genel sosyal ağlar da vardır. Dünyadaki en popüler tematik sosyal ağların bir bölümü Şekil 1.3'de sunulmuştur.



Şekil 1.3. Sosyal ağları kullanan Amerikalı internet kullanıcıların sosyal ağlara ve yaşlarına göre değişimi [1]

Sosyal ağ analizi biyologların, matematikçilerin, bilgisayar bilimcilerin, ekonomistlerin, sosyologların ve daha birçok bilim adamların ilgi alanına girmektedir.

Sosyal ağlar kullanıcıları çoğunlukla birbiri ile buluşturan, etkileşim ve iletişim kurmasına izin veren ağlardır. Ağlarda etkileşime giren kullanıcılar zamanla topluluklar ve gruplar oluşturur. Aile, iş arkadaşları veya geziden arkadaşlar sosyal ağlar içinde topluluklara örnektir. Sanal sosyal ağlar diğer bir deyiş ile sosyal medya farklı özellikte insanlardan oluştuğu için gerçek hayattaki ağlar gibi karmaşık yapılardır. Sanal sosyal ağlar da günlük hayatımızdaki ağlar gibi heterojen ve semantik zenginliği olan verilerden oluşurlar. Sanal sosyal ağlarda her birey düğümler ile ifade edilebilir, bireyler arası ilişkiler de kenar olarak tanımlanabilir. Sanal sosyal ağlarda ağı analiz etmek ve toplulukları ve örüntüleri keşfetmek önemlidir. Topluluk ve örüntü keşifleri ile kişiler gruplandırılabilir, gruplara özgü hizmetler verilebilir. Toplulukları keşfetmek, topluluklara yön veren ve en çok etkisi olan insanları keşfetmek için de ilk adım olarak kabul edilebilir. Topluluğa yön veren insanların keşfi ise örüntü keşfi ile ilgili hem sosyoloji, hem bilgisayar bilimleri hem de matematik bilimlerini ilgilendiren bir çok disiplinin ilgisini çeken bir konudur.

1.1 Tezin Amacı

Sosyal ağlar insanlar tarafından her geçen gün daha fazla kullanılmaktadırlar. Sosyal ağlar üzerinden yeni arkadaşlık, müzik grupları hatta iş ortamları kurulmaktadır. Her geçen gün karmaşıklaşan sosyal ağlarda örüntü keşfi yapmak ve keşfedilen örüntü içinde etki analizi yapmak karmaşık bir problemdir. Tezimizde bu problemlere çözüm bulan yaklaşımlar özetlenmiş ve karşılaştırmalı olarak verilmiştir. Bunların dışında hem örüntü keşfi hem de etki analizi için iki yeni yaklaşım sunulmuştur. Böylece karmaşıklaşan sosyal ağlarda daha başarılı örüntü keşfi ve etki analizi yapan yöntemler sunulmaya çalışılmıştır. Bu yöntemler ile topluluklar ve örüntüler keşfedilip topluluklara özel hizmetler veya topluluğa özel tavsiyeler yapılabilecek böylece kullanıcı deneyimi arttırılabilecektir.

1.2 Tezin Organizasyonu

Tezin bundan sonraki bölümleri aşağıdaki şekilde organize edilmiştir. İkinci bölümünde Sosyal Ağlarda Örüntü Keşfi konusunda yapılmış çalışmalar derlenmiştir. Üçüncü Bölümde Dörtlü Kapanış isimli örüntü keşfi metodu önerildi. Dördüncü bölümde Sosyal

Ağlarda Etki Analizi yapan yöntem ve çalışmaları incelendi. Bu bölümün sonunda tarafımızca geliştirilen Geliştirilmiş Yerel Yakınlık Yayılımı (Improvement Topical Affinity Propagation) metodu anlatıldı. Altıncı bölüm de tez çalışmasının sonuçları ve gelecek çalışmaları hakkında öneriler verilir.



2. SOSYAL AĞLARDA ÖRÜNTÜ KEŞİFLERİ

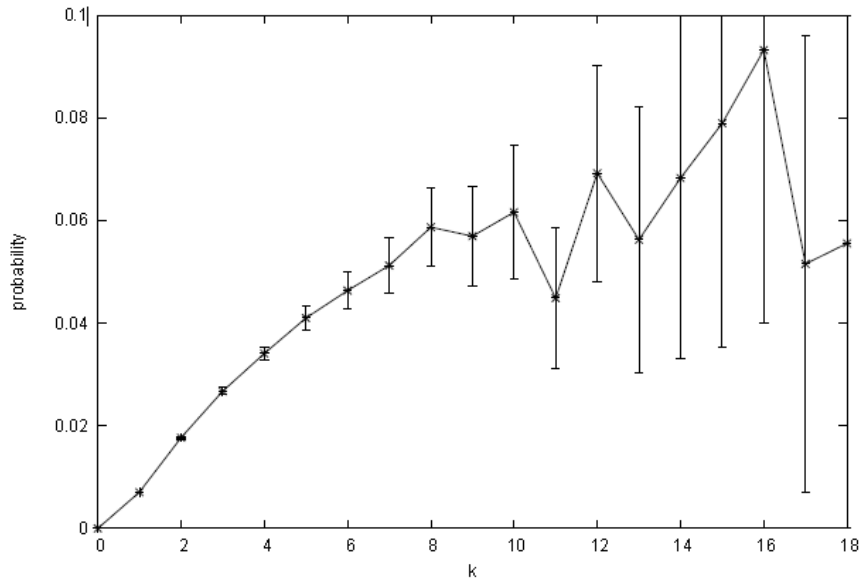
Gerçek hayatta var olan ağlardaki gibi sosyal ağlarda da bireyler keskin bir şekilde gruplandırılmaz ve aralarındaki ilişki net şekilde görülemez. Bireylerin birden fazla topluluk ile etkileşimleri ve ilişkileri farklılık gösterebilir. Gerçek hayattaki gibi sanal ağlarda da bireylerin farklı özellikleri vardır, aynı zamanda bazen bu özellikler değişir ve/veya bireylerin davranışları değişebilir. Her gurubun ve topluluğun da kendi içinde öznitelikleri vardır ve farklı davranışlar gösterir. Bir kullanıcının etkileşimde olduğu kullanıcılar, ailesi, arkadaşları bir topluluğu gösterebilir. Sanal ve Sosyal ağlarda topluluklar çoğu zaman belirtilmemiş / tespit edilmemiştir.

Sanal ve Sosyal ağlarda karmaşık topluluk ve örüntü keşfetmek, ortaya çıkarmak ve modellemek birden fazla disipline hizmet edebilecek araştırmaya değer bir alandır. Tezin bu bölümünde farklı örüntü keşfi metotları ve çalışmaları sunulmuştur.

2.1. Karar Ağaçları ile Örüntü Keşfi

Karar ağaçları birçok alanda olduğu gibi örüntü keşfinde de sık kullanılan bir yöntemdir. Karar ağaçları sensör ağları [2], metin analizi [3] ve kullanıcı deneyimleri içinden istatistiki analiz [4,5] gibi çok geniş bir alanda kullanılırlar. Lars Backstrom, Dan Huttenlocher ve Jon Kleinberg Karar Ağaçları ile örüntü keşfi konusunda yaptıkları çalışmada [6]; kişilerin bir topluluğa üye olma, topluluğun büyümesi için özniteliklerin etkisini araştırma ve DBLP veri tabanını kullanarak karar ağaçları ile kullanıcının bir gruba girme olasılığını çıkarmıştır. Çalışma p değişkenini oran olarak kabul edip k sayıda arkadaşı gruba üye olan birinin gruba üye olma olasılığını bulmayı temel almıştır (Şekil 2.1). Aslında bu çalışma sosyal bilimlerde "diffusion of innovations" olarak adlandırılan üyelerin yayılma mantığını temel alır [7,8].

Karar ağaçlarında uygulanmak üzere bir dizi öznitelik seçilmiştir. Bu özniteliklerden hangisinin gruba girmeyi etkilediğinin tespiti için öznitelikler birbirinden bağımsız seçilmişlerdir.



Şekil 2.1. Gruptaki arkadaş sayısına(k) göre kullanıcının gruba dahil olma olasılığı(p)

Tablo 2.1: Lars Backstrom'ın kullandığı Karar Ağacı Örneği Öz nitelikleri.

Nitelik Seti	Nitelikler
Grup (C) ile ilgili olan özellikler ($EC \subseteq E$.)	• Gruptaki Üye Sayısı (C). • Kullanıcının arkadaşı olan Gruptaki Üye Sayısı (the fringe of C) . • Açık Triad Sayısı: $\{(u, v, w) (u, v) \in EC \wedge (v, w) \in EC \wedge (u, w) \notin EC \wedge u \neq w\}$. • Kapalı Triad Sayısı: $\{(u, v, w) (u, v) \in EC \wedge (v, w) \in EC \wedge (u, w) \in EC\}$. • Kapalı Triad'ın Açık Triad'a oranı • Topluluk üyeleri tarafından atılan bildiriler ve verilen cevaplar
Kullanıcının(U) gruptaki(C) arkadaşları(S) ile ilgili özellikleri	• Gruptaki Arkadaş Sayısı Kullanıcının arkadaşları arasında bitişik çiftlerin sayısı $S(\{(u, v) u, v \in S \wedge (u, v) \in EC\})$. • Ec yolu üzerinde gruptaki arkadaşları ile ilişkili çift sayısı • Ec yolu üzerinde arkadaşları ile ilişki uzaklıkları ortalaması • Arkadaşlarının grupta attığı bildiriler ve verdikleri cevaplar

Sonuçlar gösteriyor ki kişinin bir gruba girme olasılığı, birkaç istisna dışında 'S-Shaped' eğilimine çok yakın şekilde gerçekleşmektedir ve Şekil 4'de gösterilmiştir. Eğer fonksiyon $P(k) = a \log k + b$ olarak çıkartılırsa test veri sonuçları daha uygun formül ile ifade edilmiş olur.

DBLP'den alınan veriler içinde $y-1$ yılı içindeki C konferansında yayın yapmış bireylerin arkadaşlarının y yılında yayın yapmış olma istatistiğini kullanmıştır. Bu bağlamda test 71.618 kişi arasında ve 20 karar ağacı kullanılarak gerçekleştirilmiştir.

Tablo 2.2: Bir önceki yıl C konferansında makale yayınlamamış ama yayınlayan akademisyenler ile daha önceki veya sonraki zamanlarda birlikte en az bir makale yazmış kişilerin bir sonraki C konferansında makale yayınlama olasılıkları

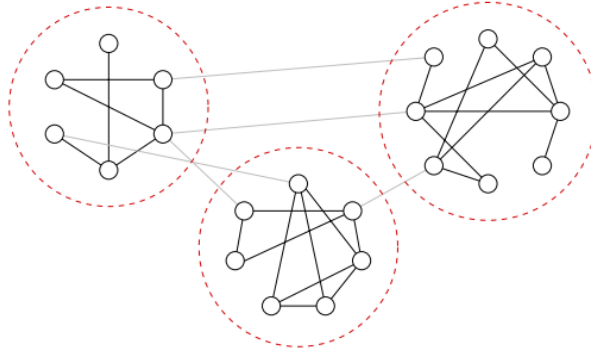
• ROCA: Roc Eğrisi [9]	Kullanılan Özellik	ROCA	APR	CXE
• APR: Average Precision [10]	Arkadaş Sayısı	0.64560	0.01236	0.06123
• CXE: Cross Entropy [11]	Tüm Üyeler	0.74114	0.02562	0.05808

2.2. Genetik Algoritmalar ile Örüntü Keşfi

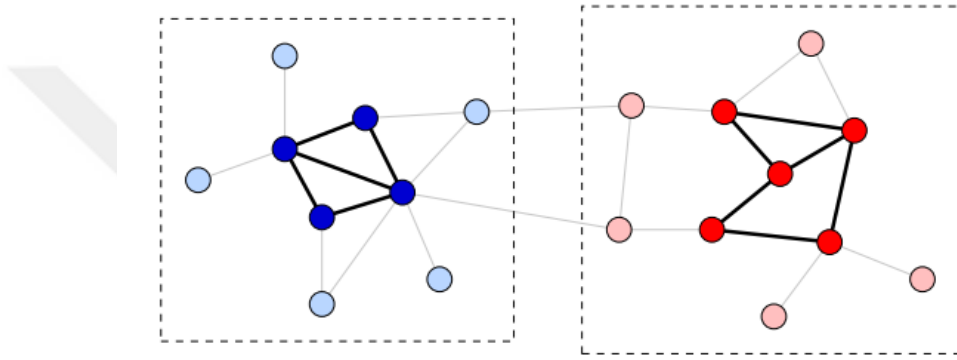
Genetik algoritmalar ilk kez John Holland [12] tarafından ortaya atılmıştır ve bu algoritmayı makine öğrenmesinde kullanan ilk kişilerden biride David Goldberg olmuştur [13]. İlerleyen süreçte Non-Linear Problemlerin çözümü [14], sınıflandırma [15] ve fonksiyon optimizasyonları [16] konularında çalışılmış ve Bulanık Mantık [17] kullanılarak genişletilmiştir. Genetik algoritmaları öne çıkartan özellikler geleneksel optimizasyon metotlarına karşılık parametre kümesi yerine kodlanmış biçimleri kullanmasıdır. Bunun yanında olasılık teoremlerini temel alarak çalışan genetik algoritmalar yalnızca amaç fonksiyonu verilerek de çalışabilirler. Genelde çözüm uzaylarının hepsini aramak zorunda kalmadan belli kısımları tarayarak daha hızlı sürede sonuca ulaşırlar [13]. Genetik algoritmaları öne çıkartan başka bir özellikte tüm uzayı eş zamanlı incelemeleri ve yerel en iyi çözüm noktalarına takılmamalarıdır.

Genetik algoritmalar ile grup tespiti yeni bir konu olmamakla birlikte özellikle biyoloji, sosyal bilimler ve istatistik bilimlerde yapılmış çalışmalar mevcuttur [18–21]. Genetik algoritmaların sosyal ağlarda kullanılması daha yeni bir konudur. Bu alanda yapılan başlıca çalışmalar Wasserman [22] ve Xu'nun [23] 'nun çalışmalarıdır. $Q = \sum_i (e_{ii} - a_i^2)$ olarak tanımlanır ve i indeksi tüm toplulukta çalışır. e_{ii} kenar frekanslarıdır ve 2 düğümü bağlar ve en az 1 noktadan gruba bağlıdır [24]. Son yapılan bazı çalışmalar aynı mantığı temel alarak daha büyük gruplarda grup tespitini Extremal Optimization olan kalite metriği [25-27] kullanarak daha hızlı şekilde yapabilmektedir.

Sosyal ağlarda genetik algoritma kullanımında en çok bilinen örnek Girvan ve Newman'ın çalışmasıdır [27] ve Şekil 2.2, Şekil 2.3 ve Şekil 2.4'de Newman'ın çalışmasında sunulan bazı grafik gösterilmiştir.

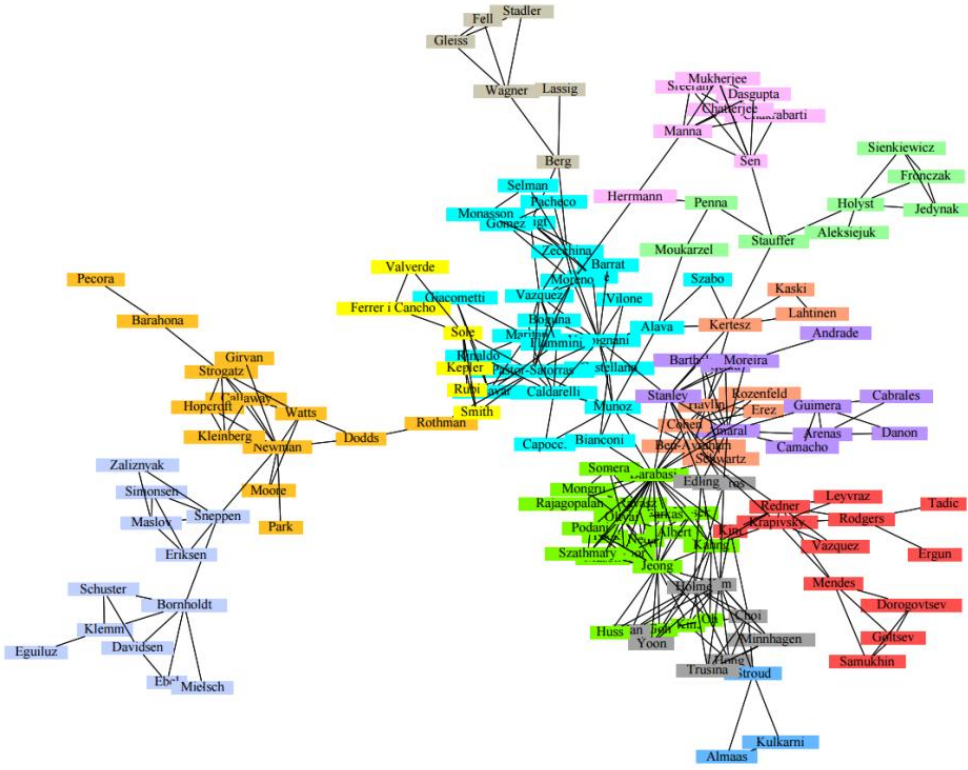


Şekil 2.2. Girvan ve Newman'ın çalışmasında modellenen 3 topluluklu ve 21 düğümlü ağ [27]



Şekil 2.3. Girvan ve Newman'ın çalışmasında modellenen ve grup içi bağlantıları gösteren, bağlantılardan üçlü olanların koyulaştırıldığı gösterim şekli [27]

Pizzuti'nin yaptığı çalışmada [28] fitness fonksiyonunu kullanarak gerçek bir sosyal ağ içindeki grupların tespiti amaçlanmıştır. Pizzuti'nin makalesinde Sosyal Ağ'a SN denilmiş ve bu ağı gösteren çizge $G=(V,E)$ ile göstermiştir. Burada bulunan V düğüm ve kenarların oluşturduğu objelerin hepsi ve E bu objeler arasındaki ilişki ve bağlantıları temsil eder. K boyutlu topluluk keşfinde bir problem K 'nın bilinmemesidir. Pizzuti bu problemi bir kromozomu oluşturan N genlerini $\{1, \dots, N\}$ ifade etmiş ve ininci gen V 'nin i ve j düğümü arasındaki linki olarak tanımlamıştır. Böylece eğer ininci varsa bu i ve j düğümlerinin aynı toplulukta olduğunu göstermektedir. Pizzuti bu çalışması ile tüm grubun sayısı, tüm grup içindeki muhtemel ilişkiler gibi bilgiler bilmeden de grup tespiti yapılabilmektedir.



Şekil 2.4. Girvan ve Newman'ın çalışmasında modellenen ve birbiri ile yayın yapmış akademisyenleri gösteren tablo [27]

Bunların dışında da çizgeleri temel alarak sosyal ağlarda yapılan diğer öncü çalışmalar vardır[29–31]. Bunlardan biri rastgele bir noktadan başlayıp belirgin çizgelerde K-Means algoritmasını kullanarak grup tespiti [30].

2.3. Sıklık Desenleri ile Örüntü Keşfi

Sıklık desenleri analizi, bir grup veya ağ içinde belli özelliklerin sıklık değerlerine bakarak gruplandırma bir istatistik gruplandırma metodolojisi [32,33]. Başlıca sıklık desen analizi/madenciliği; diziler [34], ağaçlar [35] veya çizgeler [36] üzerinde yapılabilir. Başlıca kullanım alanlarından bazıları alışveriş sepeti, internette tıklama analizi ve Genome analizidir [37].

Sıklık desenleri ile varlıklar arası ilişkileri, difüzyon modeli [38] kullanarak, Google arama sonuçlarını [39,40] kullanarak, ve metin bazlı segmentasyon [41] kullanarak tespit etmeye çalışan araştırmalar mevcuttur.

Sıklık Desenleri ile sosyal ağlarda analiz nispeten yeni bir konu olup bu konudaki öncü çalışmalardan biri Adnan ve arkadaşlarının [42] yaptığı çalışmadır. Adnan ve arkadaşları çalışmalarında problemi ve kurulan yapıyı şöyle ifade etmişlerdir: E tüm varlıkların

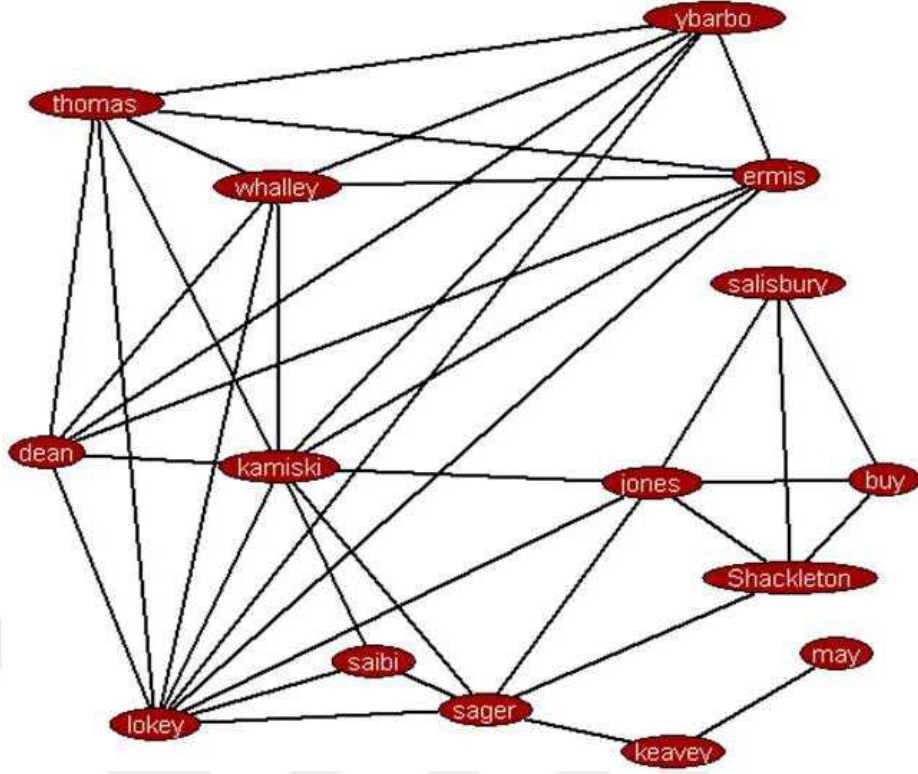
bulunduğu küme olmak üzere her $e1$ varlığı bir $D1$ dataset'i ile ilişkilidir. Bundan dolayı varlık kümesi $V = fv1 = e1; v2 = e2; v3 = e3; :::; vn = eng$ olmak üzere $G(V;A)$ olarak ifade edilebilir. Varlıklar için bir ağ oluştururken $\{e1,e2,...,eN\}$ için bir öznitelik belirlenip bu nitelik Dj ile ilişkilendirilir. Bu nitelik vektörünü yaratmak için her varlık için uygulanır. Öznitelik vektörünü Fj ile ifade edelim, $Fj = (w(f1),w(f2),...,w(fm))$ olur ve $w(fk)$ fk varlığının k 'nıncı özneliğin ağırlığı olarak bulunur. ei ve ej arasındaki benzerlik $(F1.Fj/||Fi||.Fj||)$ olur. Bir kere uzaklık ve benzerlik değerleri çıkarıldığında diğer birçok çalışmada bahsedilen yöntemler ile topluluk keşfedilebilir [42].

Çalışmada kullanılan veri seti 150 Enron şirketi çalışanın 500.000 'den fazla mailini kapsar. Mail kutusunda 1000 E-mail'den fazla e-maili bulunan çalışanlar arasında 15 kullanıcı rastgele seçilerek e-mail'lerinin bir kısmı çalışmaya dahil edilmiştir. Her e-mailin kök kelimeleri çıkartılmış ve bunlar arasında global sıklık analizi yapılmıştır.

Tablo 2.3. Kullanılan kelime frekanslarına göre kullanıcılar arası sıklık analizi (Öklit uzaklığı ile) [42]. Sıklık analizinde IDF veya TF metodlarından biri de kullanılabilir [43].

	buy	dean	ermis	jones	kamiski	keavey	lokey	may	sager	saibi	salisbury	shackleton	thomas	whalley	ybarbo
buy	0.00	0.65	0.57	0.26	0.43	0.41	0.43	0.35	0.32	0.36	0.25	0.22	0.65	0.60	0.59
dean	0.65	0.00	0.13	0.50	0.28	0.50	0.27	0.68	0.40	0.44	0.73	0.64	0.08	0.10	0.13
ermis	0.57	0.13	0.00	0.44	0.22	0.44	0.21	0.61	0.33	0.38	0.65	0.56	0.15	0.14	0.16
jones	0.26	0.50	0.44	0.00	0.27	0.35	0.29	0.38	0.19	0.26	0.36	0.21	0.50	0.47	0.44
kamiski	0.43	0.28	0.22	0.27	0.00	0.31	0.16	0.47	0.17	0.28	0.51	0.39	0.28	0.25	0.25
keavey	0.41	0.50	0.44	0.35	0.31	0.00	0.38	0.25	0.30	0.41	0.45	0.38	0.51	0.47	0.50
lokey	0.43	0.27	0.21	0.29	0.16	0.38	0.00	0.50	0.22	0.25	0.52	0.41	0.27	0.25	0.24
may	0.35	0.68	0.61	0.38	0.47	0.25	0.50	0.00	0.40	0.45	0.35	0.33	0.69	0.65	0.67
sager	0.32	0.40	0.33	0.19	0.17	0.30	0.22	0.40	0.00	0.25	0.44	0.28	0.40	0.36	0.36
saibi	0.36	0.44	0.38	0.26	0.28	0.41	0.25	0.45	0.25	0.00	0.45	0.34	0.43	0.41	0.41
salisbury	0.25	0.73	0.65	0.36	0.51	0.45	0.52	0.35	0.44	0.45	0.00	0.30	0.75	0.70	0.70
shackleton	0.22	0.64	0.56	0.21	0.39	0.38	0.41	0.33	0.28	0.34	0.30	0.00	0.63	0.60	0.59
thomas	0.65	0.08	0.15	0.50	0.28	0.51	0.27	0.69	0.40	0.43	0.75	0.63	0.00	0.09	0.13
whalley	0.60	0.10	0.14	0.47	0.25	0.47	0.25	0.65	0.36	0.41	0.70	0.60	0.09	0.00	0.11
ybarbo	0.59	0.13	0.16	0.44	0.25	0.50	0.24	0.67	0.36	0.41	0.70	0.59	0.13	0.11	0.00

Öklid uzaklığına göre kullanıcıların birbirine ne kadar yakın ve uzak oldukları tespit edilmeye çalışılmıştır. Şekil 2.5'de örnek bir ilişki ağı görülebilir.



Şekil 2.5. Kullanıcılar arası etkileşimin çizge üzerinde gösterimi [42]

2.4. Segmentasyon Matrisi İndeksi ile Bağlı Grup Keşfi

Segmentasyon analizi ilk olarak 90'lı yıllarda sosyologlar ve matematikçiler tarafından ortaya atılmıştır[44–46]. Segmentasyon analizi ile sosyal topluluklarda grup keşfi konusunda yegane çalışmalardan biri Meir Fershtman'ın çalışmasıdır [47]. Cohesive yani bağlı gruplar iki farklı gruptaki üyelerin aralarındaki kuvvetli bağdan doğan gruplardır.

Bir segmentasyon matrisi indeksini (SMI) ikili bir çizge'de göstermek gerekirse N grup elemanları, A N_a 'lardan oluşmuş bir Grup ve $N_b = N - N_a$ da N 'de olup A 'da olmayan elemanlar olması durumu 2.1'de gösterilmiştir.

$$X_{ij} \begin{cases} \text{Eğer } i, j \text{ seçilirse } 1; \\ \text{Eğer } i, j \text{ seçilmez ise } 0; \end{cases} \quad X_{ii} = 0 \quad (2.1)$$

X_iA ve X_iB 'de elemanın seçebileceği grup sayısı 2.2'de gösterilmiştir.

$$\begin{aligned} X_{iA} &= \sum_{j \in A} X_{ij}; & X_{iB} &= \sum_{j \in B} X_{ij} \\ X_{AA} &= \sum_{i \in A} X_{iA}; & X_{AB} &= \sum_{i \in B} X_{iB} \end{aligned} \quad (2.2)$$

X_{AA} , A grubunun elemanın seçenek sayısı ve X_{AB} , A grubu elemanı olmayanın seçenek sayısı olsun. X_{AA} ve X_{BB} 'yi birçok faktör etkileyebilir. Bundan dolayı genelde göreceli ortalama alınır.

Ortalama seçenek sayısı A'nın elemanı olan ve A'nın elemanı olmayanın sırası ile X_{AA}/N_A ve X_{AB}/N_A olur. A rastgele elemanlardan seçilirse ortalama seçenek sayısı $A(N_A-1)$ ve $A(N_B)$ 'nin içermedikleri olur. Bu ilişki 2.3'de formülize edilmiştir.

$$\frac{X_{AA} / N_A}{X_{AB} / N_A} = \frac{N_A - 1}{N_B} \text{ veya } \frac{X_{AA} / (N_A (N_A - 1))}{X_{AB} / N_A N_B} = 1 \quad (2.3)$$

Seçme yoğunluğu d_{AA} ve d_{AB} olarak ortalama değerden hesaplanırsa sırası ile A'nın elemanı olan ve olmayan seçme yoğunluğu 2.4'de gösterilmiştir..

$$d_{AA} = \frac{X_{AA}}{N_A(N_A-1)} \quad d_{AB} = \frac{X_{AB}}{N_A(N_B)} \quad (2.4)$$

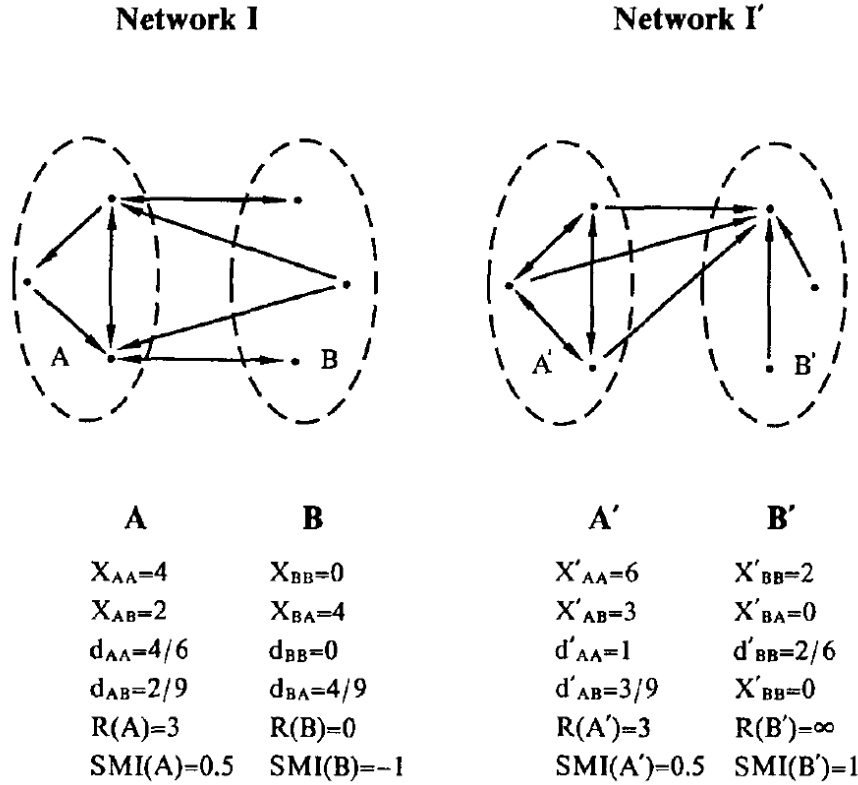
A grubunun segregative davranış eğrisi de 2.5'de gösterildiği gibi ölçülebilir.

$$R(A) = d_{AA} / d_{AB} \quad (2.5)$$

R 0'dan sonsuza giderken segregarion indeksinin formülü de 2.6'da sunulmuştur.

$$SMI(A) = \frac{R(A)-1}{R(A)+1} \quad d_{AB} = \frac{d_{AA} - d_{AB}}{d_{AA} + d_{BB}} \quad (2.6)$$

Şekil 2.6'de A ve A' gruplarının benzer segregarion yüksekliğinde seçme yoğunlukları farklıdır.



Şekil 2.6. İki grup arasında segregation matriks indexi

2.5. Uzaklık Odaklı Örüntü Keşfi

Uzaklık odaklı örüntü keşfi metodu nispeten yeni ortaya atılan metotlardan biridir. Li ve arkadaşları merkezîyet tabanlı örüntü keşfi metodu ile önceden verilmiş topluluk sayısı yani başlangıç değeri olmaksızın her düğümü merkez olarak diğer düğümlere uzaklıklarını bulup düğümler arasında benzerlik ve örüntüleri keşfetmek üzerine bir çalışma önermiştir [48–50]. Birden fazla yakınlık tanımı vardır. Bu çalışmada ise yakınlık tanımı olarak iki düğüm arası uzaklık alınmıştır. Bu uzaklık fiziksel uzaklıktan öte benzerliği anlatır. Yani yakınlık arttıkça iki düğümün birbirine benzerliği artar. Burada yapılan varsayım örnek uzayında benzer düğümlerin birbirine yakın konumlandırılacağıdır.

Merkezi yakınlığı hesaplamak için öncelikli olarak ortalama uzaklıkları hesaplamak gerekir. Bir ağdaki ortalama uzaklıklar 2.7'deki gösterildiği hesaplanır.

$$D_{avg}(V_i) = \frac{1}{n-1} \sum_{j=i}^x g(v_i, v_j) \quad (2.7)$$

Formülde bulunan n ağdaki eleman sayısı $g(V_i, V_j)$ geodesic uzaklıktır. Bundan dolayı V_i 'ye en yakın karşılık $D_{avg}(V_1)$ 2.8'de sunulmuştur.

$$C_c(V_i) = \left[\left[\frac{1}{n-1} \sum_{j=i}^x g(v_i, v_j) \right] \right]^{-1} = \frac{n-1}{\sum_n g(v_i, v_j)} \quad (2.8)$$

Benzerlik, iki düğüm arasındaki öznitelik ve konum benzerliğidir. Benzerliği yüksek olan bireylerin aynı toplulukta olacağı varsayılmıştır. Jaccard benzerliği, yerel bilgiyi temel alarak benzerlik indeksi üretir. İki komşu arasında ortak ve yakın öznitelikler benzerlik oranlarını artırır. Jaccard benzerliğinde $r_{(i)}$ V_i 'in komşusu anlamına gelir. Jaccard benzerliği daha çok hangi komşunun daha benzer olduğunu bulmakta kullanılabilir. Jaccard benzerliği 2.9'da sunulmuştur.

$$S_{Jaccard} = \left| \frac{r(i) \cap r(j)}{r(i) \cup r(j)} \right| \quad (2.9)$$

DCCD Jaccard yakınlığını temel alarak çalışır ve 8 adımı vardır:

1. Tüm düğümler için yakınlıkları hesaplar
2. Yakın düğümleri sıralar ve onları topluluk aday setine alır
3. Jaccard'ı referans alarak yakın komşulara göre bu seti düzenler
4. Merkez düğüme göre yakınlık değişimlerini hesaplar ve merkez düğümü bulur
5. En benzer düğümleri bulur topluluğa bunları ekler
6. Tüm düğümler topluluğa atanmış ise devam eder, atanmamışsa 4. maddeye döner
7. Eğer her topluluğu oluşturan varlık sayısı eşik değerinden düşük ise topluluğu parçalar ve 4. maddeye geri döner
8. Tüm topluluklarda merkez yakınlığını tekrar hesaplar ve en sıkı bağlı toplulukları sıralar.

DVVD ise kesin doğrulama için saflık analizi yapmakla başlar. Burada V tüm ağdaki düğümleri kapsayan olmak üzere $V = \{v_1, v_2, v_3, \dots, v_5\}$ R_i V 'nin bir alt kümesi olmak üzere $V = \{v_1, v_2, v_3, v_4\}$ ve $V = \cup_1^k R_i$ ve $R_1 \cap R_2 = \emptyset (i \neq j)$ olur. Saflık formülü 2.10'da gösterilmiştir:

$$Purity(R, G) = \frac{1}{n} \sum_j \max_i |R \cap G| \quad (2.10)$$

Modülerlik, daha önce bahsedilen Newman [27]'in dağıtık düğümler üzerindeki benzerlik bulma metodudur. Genel olarak A_{ij} yani A uzayının i elemanı ile j elemanı arasındaki bağlantı $A_{ij} = d_i d_j / 2m$ olarak ölçülür ve K topluluklarında Q değeri 2.11'deki gibi olur

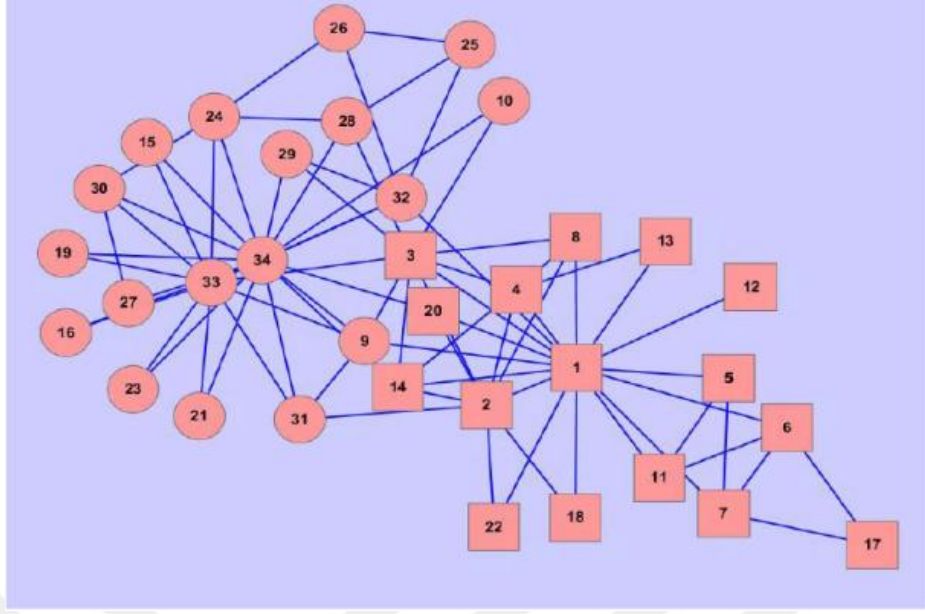
$$\sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m \quad (2.11)$$

$$Q = \frac{1}{2m} \sum_{i=1}^k \sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

Çalışmada sosyal ağ analizi konusunda daha önce kullanılmış [51] ABD'de bir üniversitede karate kulübüne üye olan 34 düğümlü bir veri seti alınmıştır.

Tablo 2.4. DCCD'nin diğer algoritmalara göre başarısı

Algorithm	Safılık	Modülerlik
DCCD	1.00	0.371
TopLeader(2)	1.00	0.371
TopLeader (3)	1.00	0.371
TopLeader (4)	1.00	0.374
Fast Modularity	0.97	0.361
CFinder	0.65	0.380
SCAN	0.764	0.182



Şekil 2.7. DCDD'nin bu set üzerine uygulanmasında çıkan sonuç

Tablo 5 de görüldüğü gibi *DCCD* karate kulübünü 2 ağa ayırmıştır, saflık sonucu 1 ve Q değeri 0.371 olmuştur ve Şekil 2.7'da bir *DCDD* set uygulaması sonucu görüntülenebilir.

2.6. Üçlü Kapanış ile Örüntü Keşfi

Üçlü kapanış (Triadic Closure Pattern), üçlü döngü Alman sosyolog Georg Simmel'in 1908'de yayınlanan kitabında [52] ilk olarak kullanılan bir sosyal ağ terimidir [53]. A , B ve C düğümleri arasındaki ilişkiyi belirtir. Bu ilişki her birinin birbiri ile ilişkili olmasından ziyade A 'nın B ile, B 'nin C ile, C 'nin A ile ilişkili olması durumudur. Üçlü kapanış metodunu tüm ağa uygulamak zor gibi görünse de küçük gruplara uygulanması ve bulunması ağ keşfi açısından önemlidir [54].

2.6.1. Üçlü Kapanış Kullanım Alanları

Üçlü kapanış çok farklı disiplinlerde kullanılan bir metottur. Biomühendislikte çift olan kromozonların [55] bulunması, koordinatlara göre kişileri ilişkilendirme [56] ve çifte 3. kişi ekleme, ağdaki varlıklar arasında kuvvetli ilişkileri olanları bulmak [57] gibi çalışmalar için kullanılmıştır. Coleman [58] ve Uzzi [59] yaptıkları sosyolojik çalışmalarda bireysellikten üçlü kapanış kullanarak gruplama yöntemi ile analiz etmeyi daha başarılı bulmuşlardır. Walker [60] ve Buskens [61] yaptıkları çalışmalarda üçlü kapanış ile tespit edilen grupların karşılıklı güvenlerinin daha yüksek olduğunu tespit etmişlerdir. Reagans

ve McEvily [62] ve Zaheer ve Bell [63] çalışmalarında bu tip grupların daha sıkı ilişki içinde oldukları, daha fazla benzer oldukları ve daha çok bilgi değişimine yatkın olduklarını bulmuştur. Bunlardan dolayı üçlü kapanış sosyal ağlarda varlıklar arası ilişkileri bulma, çiftleri bulma, toplulukları ve örüntüleri keşfetme konusunda en yenilikçi ve başarılı çözümlerden biridir [64].

Üçlü kapanış metodunu Kossinets ve arkadaşları [65] bir üniversitede atılmış mailleri analiz etme, Tiancheng ve arkadaşları [66] ise Twitter verisinde arkadaşının arkadaşını bulmak gibi sosyal ağ analizlerinde kullanmışlardır. Yine Grindroda [67] Twitter verisinin analiz edilmesinde, Huberman [68] yerel ve global ağlardaki farklılıkları karşılaştırmada ve Opsahl [69] sosyal ağların analiz edilmesinde kullanmışlardır. Üçlü kapanış metodu ile sosyal ağlarda örüntü keşfi konusunda en önemli çalışmalardan biri Huang ve arkadaşları tarafından yapılan ve TriadFG'nin sunulduğu çalışmadır [70].

Üçlü kapanış ile sosyal ağ üzerinde çok çeşitli çıkarımlar yapılabilir. Huang ve arkadaşları büyük bir mikro blog servis verisi üzerinde yaptıkları çalışmalarında kişilerin grup olmasına neden olan faktörleri incelemiş, grubu oluşturan kişilerin gruba girmeden önceki girme olasılıklarını tespit etmeye çalışmıştır. Üçlü kapanış metodunu bir adım ilerleterek çizge bazlı TriadFG metodunu öne sürmüşlerdir. Bu metod üçlü kapanış'a dönüşebilecek pattern ve yapıları sosyal ağ'lar üzerinde arayıp modellemeyi sağlamaktadır. Huang ve arkadaşları yaptıkları çalışma ile kişilere yeni arkadaş tavsiye etmeyi ve yeni topluluklar keşfetmeyi çalışmışlardır [70].

2.6.2. Üçlü Kapanış Probleminin Tanımlanması

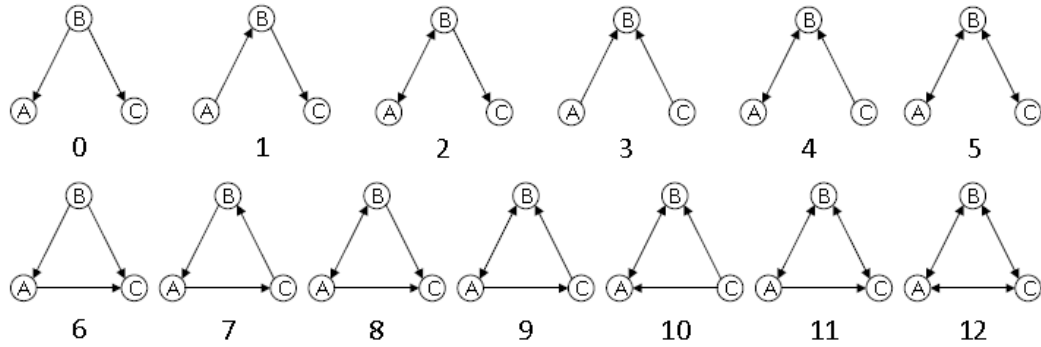
Bu yazının diğer örneklerinde olduğu gibi G sosyal ağını $G=(V,E)$ olarak temsil etmek üzere $E \cup V \times V$ kullanıcılar arasında ilişkiyi gösterir. Uzay ağının çok geniş olmasından ve sınırsız ihtimal bulunmasında dolayı kapalı üçlü kapanış kullanılmıştır. Kapalı üçlü; aynı zaman diliminde 3 kullanıcının (A,B,C) birbiri ile A 'dan B 'ye, B 'den C 'ye ve C 'den A 'ya olmak üzere bağlantılı olmasıdır. Eğer bağlantılardan biri eksik olursa buna da açık üçlü denilebilir. Çalışma açık üçlü olan bir grubun $t+1$ anında kapalı üçlü kapanış olabilme ihtimalini bulmayı amaçlar. y her açık üçlü kapanış adayı için bir değişken olmak üzere $y(t+1)=1$ ve $y(t+1)=0$ olma olasılığı vardır. Bu bağlamda $Y(t+1)$ 2.12'de formülize edilmiştir.

$$f := (\{G^t Y^t\}_{t=1, \dots, T}) \rightarrow Y^{T+1} \quad (2.12)$$

2.6.3. Veri Toplama ve Gözlem

Huang ve arkadaşları çalışmalarında veriyi Weibo isimli Twitter'a benzer bir mikro blog ağından toplamışlardır. Rastgele seçilen 100 kullanıcı alınmış ve toplamda o kullanıcılarla ilişkili olan 695.842 kullanıcı verisi toplanmış ve 423 milyon'a yakın kullanıcı arasındaki ilişkileri gösteren link bulunmuştur. Bu çalışmada kişiler arasında ilişkiyi puanlandıran 5 farklı öznetelik/faktör seçilmiştir. Bunlardan ilki ağ topolojileridir ve Şekil 2.8'de gösterilmiştir. Huang ve arkadaşları [70] her faktör için faktör etki analizi çıkarmışlar ve böylece seçilen özneteliklerin etkisini bağımsız olarak hesaplamışlardır.

Huang ve arkadaşları bu çalışmada ilk önce ağ topolojilerini belirlemeye çalışmışlardır. Şekil 2.8'de gösterilen şekilde 6 açık ve 7 kapalı üçlü bulunmuştur. Topoloji olmaksızın çalışma standart TriadFG'ye göre 0.244 daha başarısız çıkmıştır. Böylece ağ topolojisinin en etkili öznetelik olduğu görülmüştür.



a) 0 ile 5 açık üçlü, 6 ile 12 kapalı üçlülerdir. (A,B ve C kullanıcıları temsil eder)

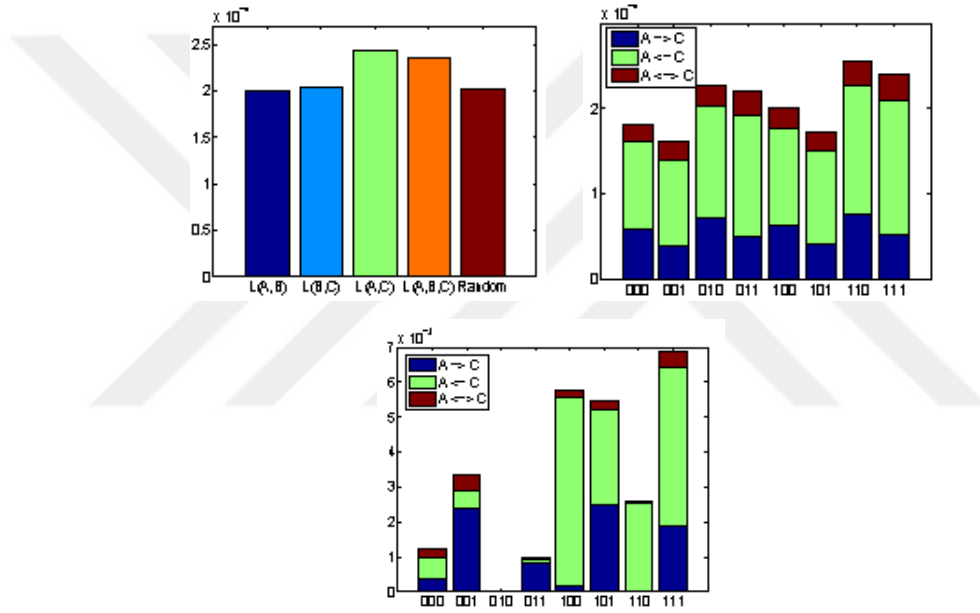
Open Triad $\xrightarrow{A \rightarrow C}$	Triadic Closure $\xrightarrow{A \leftarrow C}$	Open Triad $\xrightarrow{A \leftarrow C}$	Triadic Closure $\xrightarrow{A \leftrightarrow C}$
0 $\xrightarrow{A \rightarrow C}$ 6	0 $\xrightarrow{A \leftarrow C}$ 6	0 $\xrightarrow{A \leftrightarrow C}$ 10	
1 $\xrightarrow{A \rightarrow C}$ 6	1 $\xrightarrow{A \leftarrow C}$ 7	1 $\xrightarrow{A \leftrightarrow C}$ 9	
2 $\xrightarrow{A \rightarrow C}$ 8	2 $\xrightarrow{A \leftarrow C}$ 9	2 $\xrightarrow{A \leftrightarrow C}$ 11	
3 $\xrightarrow{A \rightarrow C}$ 6	3 $\xrightarrow{A \leftarrow C}$ 6	3 $\xrightarrow{A \leftrightarrow C}$ 8	
4 $\xrightarrow{A \rightarrow C}$ 9	4 $\xrightarrow{A \leftarrow C}$ 10	4 $\xrightarrow{A \leftrightarrow C}$ 11	
5 $\xrightarrow{A \rightarrow C}$ 11	5 $\xrightarrow{A \leftarrow C}$ 11	5 $\xrightarrow{A \leftrightarrow C}$ 12	

b) Tablodaki her girdi kullanıcılar arasındaki bağlantıdır. Bu bağlantılara göre açık veya kapalı üçlüler gösterilir.

Şekil 2.8. Açık Üçlü, Üçlü Kapanış ve üçlü kapanış'dan kapalı üçlü çıkartmak [70].

Şekil 2.8'de üç kullanıcı içinde karşılıklı 2 link olanlar kapalı triad olmaya daha yatkın olduğu örneği gibi her açık triad'ın kapatılma ihtimali gösterilmiştir.

Kullanılan ikinci öznitelik nüfus dağılımıdır ve Şekil 2.9'da gösterilmiştir. Weibo'dan alınan kullanıcının konumu yani yaşadığı bölge/şehir ile kullanıcıların cinsiyeti ve yaş aralıkları da üçüncü öznitelik olarak hesaplamaya dahil edilmiştir ve Şekil 2.9'de gösterilmişlerdir. Nüfus dağılımı en düşük etkili özniteliktir ve çıkarılması durumunda sadece 0.092 oranında algoritmayı olumsuz etkilemiştir.



Şekil 2.9. Huang ve Arkadaşlarının Faktörlerinden Konum, Cinsiyet ve Popüleriteye göre korelasyon değerleri.

Sosyal roller, algoritmada kullanılan dördüncü bir özniteliktir. Kullanıcılar takipçi ve takip eden sayılarına göre üç farklı role bürünmüş ve her rol için kendisi ile ve diğer roller ile korelasyon oranları hesaplanmıştır. Sosyal rollerin algoritma üzerine etkisi ikinci en güçlü öznitelik/faktördür ve çıkarılması durumunda 0.158 oranında olumsuz etkilemektedir. Son öznitelik ise Yapı Deliğidir (Structure Hole).

Geçişlilik [22] birçok sosyal teoride kabul gören bir kavramdır. A,B ve C varlıklar olmak üzere $A \rightarrow B$ ve $B \rightarrow C$ gidiyor ve $A \rightarrow C$ 'ye gidiyor ise geçişli kabul edilir.

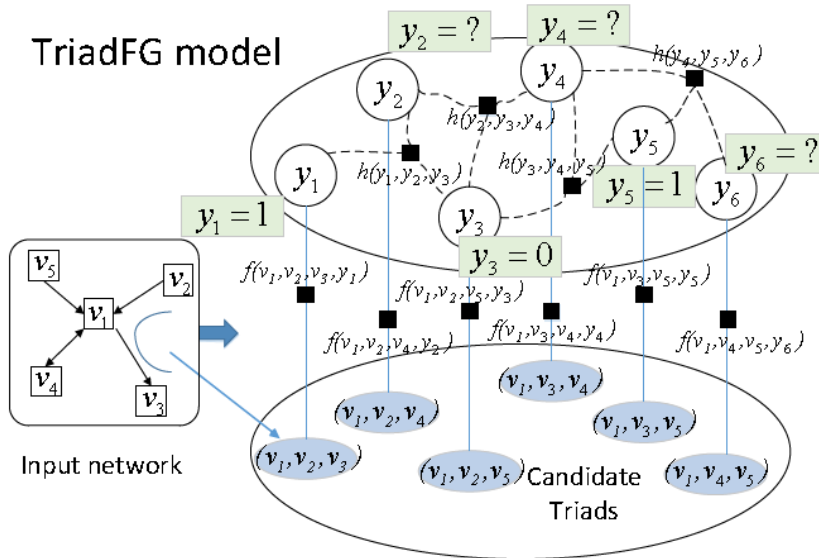
Geçişlilik üçlü kapanış da uygulanan bir faktördür. Weibo'ya uygulanan çalışmaya göre [70] kullanıcıların %72'si bu şekilde diğer kullanıcılar ile etkileşim kurmaktadır.

2.6.4. Üçlü Kapanış Faktör Çizgesi

Üçlü Faktör Çizge'leri bir çeşit Üçlü kapanış modeli olup Huang ve arkadaşları tarafından ortaya atılmıştır [70]. Bu makaleden önce Lou ve arkadaşları da benzer bir model öne sürmüş ve modelde karşılıklı takibi bulmaya çalışmışlardır [66]. Huang ve arkadaşları ise çalışmalarında daha çok açık bir triad'ın hangi koşullarda ve hangi oran ile kapalı üçlü'ye dönüştüğünü ölçmeye çalışmışlardır. Üçlü Kapanış'da olduğu gibi Triad-FG'de de $G = (V, E, X, Y)$ olsun. V düğümlerin bir kümesi $E \cup V \times V$ olma üzere ve Y 'de X 'in $t+1$ anında 1 veya 0 olmak üzere Markov çizge teorimi temel alan formül 2.13'de gösterilmiştir [71,70].

$$P(Y|X, G) = \frac{1}{Z} \exp \left\{ \sum_{i=1}^{|Tr|} \sum_{j=1}^d a_i f_i(x_{ij}, y_i) + \sum_c \sum_k \mu_k h_k(Y_{Tr_c}) \right\} \quad (2.13)$$

Üçlü Faktör Çizgesi modelinin bir örneği Şekil 2.10'da gösterilmiştir. Kapalı üçlü kapanış olmaya aday olan mavi eklipsler aşağıdadır. Beyaz yuvarlaklar gizli değişkenlerdir ve $y_i.f(v_1, v_2, v_3)$ faktor fonksiyonunun özellikleridir. $h(.)$ ise triadlar üzerine fonksiyonun korelasyonudur[70].



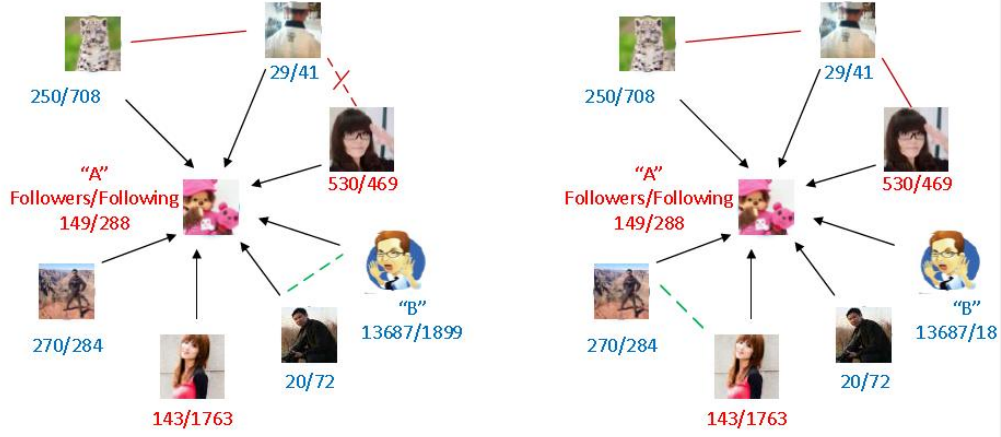
Şekil 2.10. TriadGF modelinin grafikte gösterimi

2.6.4.1. Üçlü Kapanış Faktör Çizgesi Sonuçları

TriadFG, muadili olan SVM ve Logistic algoritmaları ile 4 farklı ölçü birimi ile karşılaştırılmıştır. SVM ve Logistic birer sınıflandırma algoritması olmakla birlikte kullanılan parametreler, veri seti ve girişler neredeyse TriadFG ile aynıdır. Çalışmada TriadFG C++ ile implemente edilmişken SVM ve Logistic için Weka [72] kullanılmıştır. Ölçümler ise Kesinlik Anma, F1-Ölçütü ve Kesinlik ile yapılmıştır. TriadFG algoritması, muadil üçlü kapanış algoritmalarından F1-Ölçütüne göre % 6.99 daha yüksek başarı sağlamıştır. Bir vaka örneği Şekil 2.11'de sunulmuştur.

Tablo 2.5. Üçlü Kapanış Tahmin Başarı Performansı

Algoritma	Kesinlik	Anma	F1-Ölçütü	Doğruluk
SVM	0.890	0.844	0.866	0.882
Logistic	0.882	0.913	0.897	0.85
TriadFG	0.901	0.953	0.926	0.931



Şekil 2.11 Haung ve arkadaşları yaptıkları Triad-FG modeli Weibo'da uygulanması

Şekil 2.11 'de gösterildiği gibi Haung ve arkadaşları yaptıkları Triad-FG modeli Weibo'da uygulandığında A ve B kullanıcılarının üçlü kapanış benzeri arkadaş ilişkilerini modellemektedir [70]. Kırmızı çizgiler bayan, mavi çizgiler erkek olduklarını gösteriyor. Siyah bağlantılar hali hazırda olan bağlantıları işaret ediyor. Kırmızı bağlantılar bir sonraki zaman diliminde kurulacağı düşünülen bağlantıları ifade ediyor. Kırmızı kesikli çizgiler gelecekte kurulan ama tespit edilemeyen çizgiyi ifade ediyor. Yeşil ise tahmin edilen ama bir sonraki zaman diliminde kurulmamış olan bağlantıyı ifade etmektedir.

3. DÖRTLÜ KAPANIŞ İLE ÖRÜNTÜ KEŞFİ

Dörtlü Kapanış (Quad Closure Pattern), Haung [70] ve arkadaşlarının Üçlü Kapanış metodu temel alınarak tarafımızca geliştirilmiş bir örüntü keşfi metodudur. Üçlü Kapanış'dan temel farkı üçlü değil dörtlü ilişkileri ele alarak örüntü keşfi ve bağlantı tahmini yapmasıdır. Çalışmamız ağ topolojilerini ve beş farklı faktörü kapsayan sosyal ağlar için özelleşmiş olasılıklı Dörtlü kapanış metodu oluşturmayı kapsamaktadır.

Bir sosyal ağı çizge ile modellediğimizde bu çizge G ile gösterebiliriz. Problemi ilk önce matematiksel ve çizge olarak gösterilecektir. $G = (V, E)$ olmak üzere G bir çizge V kullanıcı uzayı ve E 'de ilişkiler uzayı olsun. G^{t+1} de $t+1$ anında uzayımızın durumu olsun. Bir dörtlü kapanış dört ilişkili kişinin t anında minimum 3 bağlantısı varken $t+1$ anında bağlantısı olmayan 2 kullanıcının da birbirine bağlantı kurup bir kapalı çizge oluşturup oluşturmadığını bulmaya çalışır. Bir örnek vermek gerekirse üç kullanıcı birbiri ile seri bağlantılı ise bunları $\forall v_i, v_j, v_k \in \Delta$, ile gösterelim ve eğer $e_{ik} \in E$ de varsa biz $\forall$ için kapalı dörtlü kapanış denebilir. Bu anlatımı 3.1de formülize edilmiştir.

$$f := (\{G^t Y^t\}_{t=1, \dots, T}) \rightarrow Y^{T+1} \quad (3.1)$$

Burada gizli bir Y_t vardır ve eğer açık üçlü kapanış kapanırsa $Y_{t+1} = 1$, diğer türlü $y_{t+1} = 0$ olur. Klasik ortak yazar ağlarının aksine, sosyal ağlarda arkadaşlık oluşması yani bağlantının olmasını etkileyen çok çeşitli faktörler vardır ve bu faktörlerin sayısının artması problemi daha karmaşık hale getirir.

3.1. Dörtlü Kapanış için Veri Toplama

Instagram ilk olarak 2011 yılında bir girişimci tarafından kurulmuş ve sadece 18 ay sonra Facebook tarafından 300 milyon dolara satın alınmıştır [73]. Instagram bugün 400 milyondan fazla üyesi olan 40 milyardan fazla fotoğraf paylaşılan, ayda 3.5 milyar beğenin atıldığı günde 80 milyon fotoğraf paylaşılan dünyanın en popüler fotoğraf paylaşma sosyal ağıdır [74]. Dünyadaki tüm internet kullanıcıların 5'inde 1'i Instagram kullanmaktadır 13-20 yaş arası internet kullanan gençlerin yarısından fazlası Instagram üyesidir. Sosyal ağların ve sosyal ağlardaki veri miktarının çok hızlı büyümesi ve değerleri olmasının bir örneği de Instagram'dır [74].

Sosyal Ağlarda özellikle üçüncü parti uygulamaları, kullanıcılara sosyal ağ ile ilgili çeşitli hizmetler sunar. Sosyal ağlar bu uygulamaların kullanması için gelişmiş API

seenekleri sunar. Bylece nc parti uygulamalar izinli veya izne gerek olmaksızın sosyal aęlardan verileri alıp kendi uygulamalarında kullanabilirler.

Instagram da dięer sosyal aęlar gibi ok esnek bir Restful API'ye sahiptir. Bu API ile bir kullanıcının bilgileri (bio), fotoęrafları, fotoęraflara atılan beęeniler ve fotoęraflara yapılan yorumlar kolaylıkla alınabilir. Instagram API'nın Facebook API'na gre başlıca farkı birok profilin public olmasından dolayı kullanıcı iznine ihtiya duymadan veri saęlayabilmesidir.

Bu tezde, Instagram'dan Restful API ile veri toplaması iin bir web rmeęi tasarladı. Bu rmek başlangı noktasını tez yazarının profilinden almak zere veri toplamaya başlamıştır. İlk nce tez yazarının fotoęraflarını ve o fotoęraflara atılan beęeni ve yorumları aldıktan sonra beęeni ve yorumu atanları gezmeye devam etmek sureti ile Instagram iinde ilerlemeye devam etmiştir. Sisteme kaydedilen her veri Instagram'daki ID'si ile kaydedildięi ve bu ID yegane olarak indekslendięi iin mkerrer veri alınmadan kullanıcıları, kullanıcıların fotoęraflarını, fotoęraflara atılan beęeni ve yorumlar toplanmıştır. rmek bu iřleme kontrol edilmemiř ve fotoęrafları alınmamıř son beęeni veya yorum sahibi kalıncaya kadar devam etmektedir.

rmek, sistemin daha dřk sistem zellikleri ile daha verimli alıřması iin fotoęraf veya videoların kendilerini deęil sadece meta datalarını kaydetmekte, yorumların ise hem meta datalarını hem kendilerini kaydetmektedir. Sistemde byk miktarda veriyi tutacaęı ve byk miktarda veride iřlem yapacaęı iin bulut makineler zerinde daęıtık bir yapıda alıřan MongoDB isimli bir NoSQL dokman temelli veritabanının kullanıldı. rmek ise Ubuntu iřletim sistemli bir makinede JAVA dili ile kodlandı.

alıřmamızda kullanılan tm algoritmalar bulut altyapısında ve Map-Reduce programlama modeline uygun olarak yazıldı. alıřmada kullandıęımız veri havuzumuz 20 binden fazla kullanıcı ve yaklaşık 10 milyon baęlantı iermektedir. Bunların dıřında karřılařtırma ve kontrol yapabilmemiz iin veriyi 3 farklı zamanda topladık. İlk grup veriyi Eyll 2015'de topladık ve Ekim 2015'de yeni kullanıcı eklemekten sadece yeni baęlantıları kayıt ettik. Son olarak da Nisan 2015'de veri toplama gerekleřtirdik. İlk ve ikinci veriyi test grubu olarak topladık ikinci ve nc veriyi de kontrol grubu olarak algoritmanın başarı oranlarımızı hesaplamak iin kullandık.

Instagram'dan veri toplarken iki tr kısıtlama mevcuttur. Bunlardan biri, bir insanın dięer bir insana ulařması iin harcayacaęı maksimum baęlantı sayısıdır. Bu sayı Watts [75]'a gre yedi'dir. Yani dnyadaki bir insan dięer bir insana en fazla 7 baęlantı ile

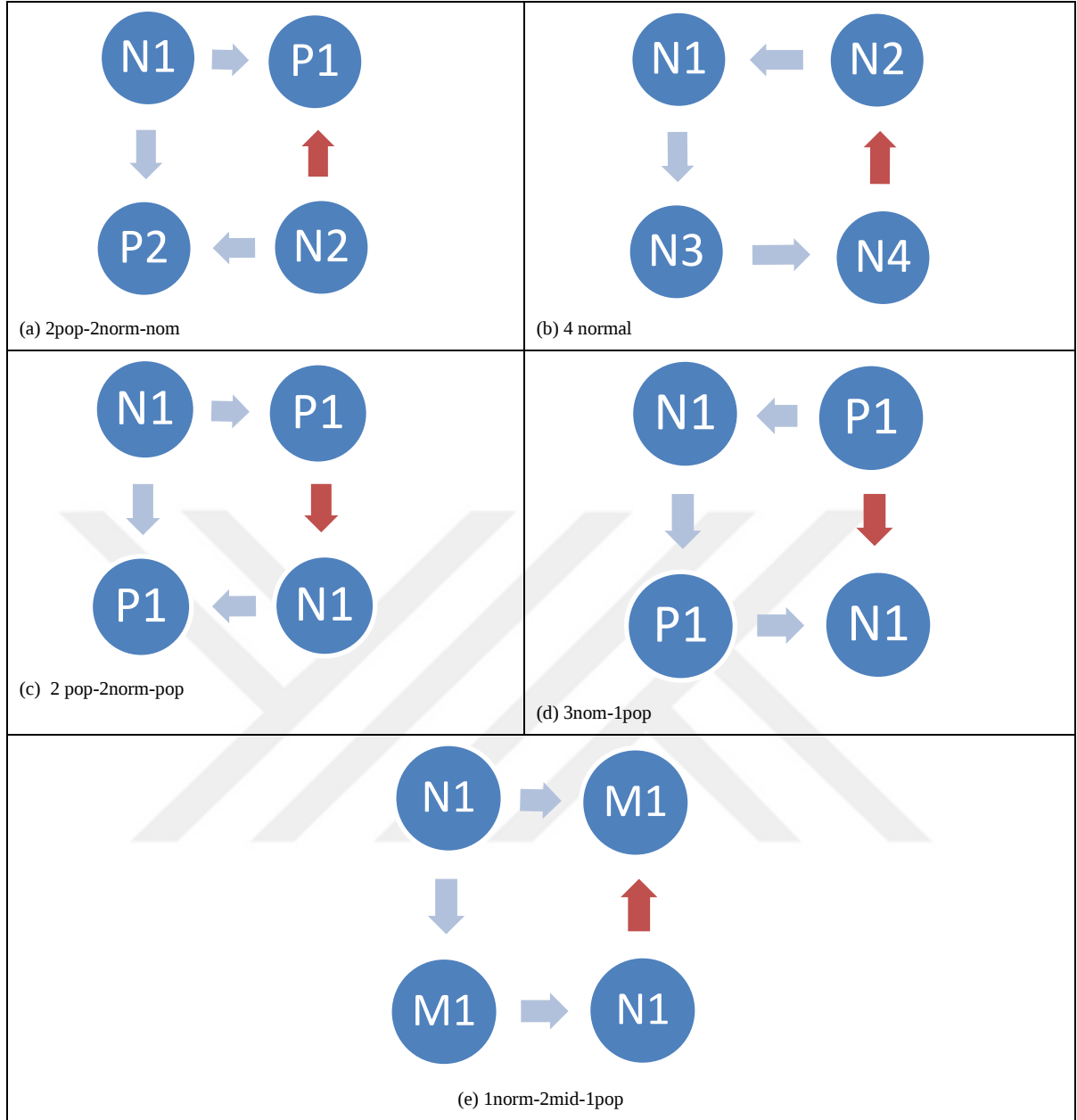
ulařabilir. Dięer limit ise Profil tipidir. Instagram'da profillerini gizli (private) olarak tanımlayan kullanıcıların bilgileri alınamadıęı için bu kullanıcılar sistem dıřı bırakılmıřtır. Bunların dıřında 500'den fazla arkadařı olan ile 10 dan az arkadařı olan kullanıcılar sisteme alınmamıřtır. Bunun nedeni bu tür kullanıcılarda korelasyon sayıları objektif ve tutarlı deęildir. Bir örnek vermek gerekirse 500.000 takipçisi olan bir kullanıcının takipçileri arasında $t+1$ anındaki iliřkiyi tahmin edecek özellik bulmak mümkün deęildir.

3.2 Seçilen Öznitelikler ve Gözlemler

Çalıřmamızda kullandığımız aę topolojilerini seçerken Milo'nun [76] çalıřmasında kullandığı aę topolojilerini temel aldık ve geliřtirdik. Aę topolojisine ek olarak dört farklı faktör ile algoritmanın başarı oranını arttırmaya çalıřtık ve her faktör için ayrı ayrı $h(.)$ korelasyon sayılarını hesapladık. Faktörler; kullanıcının dili, hashtag'in kullanım sıklığı, kullanıcının aktif olduęu zaman dilimleri ve kullanıcı yorumlarının duygu analizidir. Bunların dıřında kullanıcının sosyal rolü de algoritmaya dahil edilmiř dięer bir faktördür.

3.2.1 Aę Topolojileri

Çalıřmamızda bir açık dörtlüyü nasıl kapalı dörtlüye dönüőeceğini belirten beř tip aę topolojisi alt çizge řeklinde gösterilmiřtir. Dört aktörlü bir grupta açık çizgelerin kapalıya dönüřtürölmesi için beřten fazla yol olsa da biz en sık görülen beř aę topolojisini seçtik. Graflardaki tüm yollar tek yönlü ve gri çizgiler halihazırda olan, kırmızı çizgiler ise $t+1$ anında oluřması beklenen baęlantıları temsil eder. Biz topolojilerde 3 farklı kullanıcı tipi kullandık. Birinci tip 'Popüler' veya 'P' ile adlandırılan kullanıcı tipindeki kullanıcılar, onu takip edenlerin takipçi sayısının ortalamasının 4 katından fazla takipçisi olan kullanıcılardır. İkinci tip kullanıcı 'Mid-Popular' olarak adlandırılır 'Popüler' gibi takipçilerinin takipçi sayısının ortalamasının 2 katından fazla olan kullanıcılar bu gruba girer. Son grup ise 'N' ile sembolize edilen 'Normal' dięer bir deyiř ile 'Ordinary' kullanıcıları kapsar bu gruptaki kullanıcıların takipçi ve takip ettikleri sayılar ortalama düzeyindedir. řekil 3.1'de aę topolojileri sunulmuřtur.

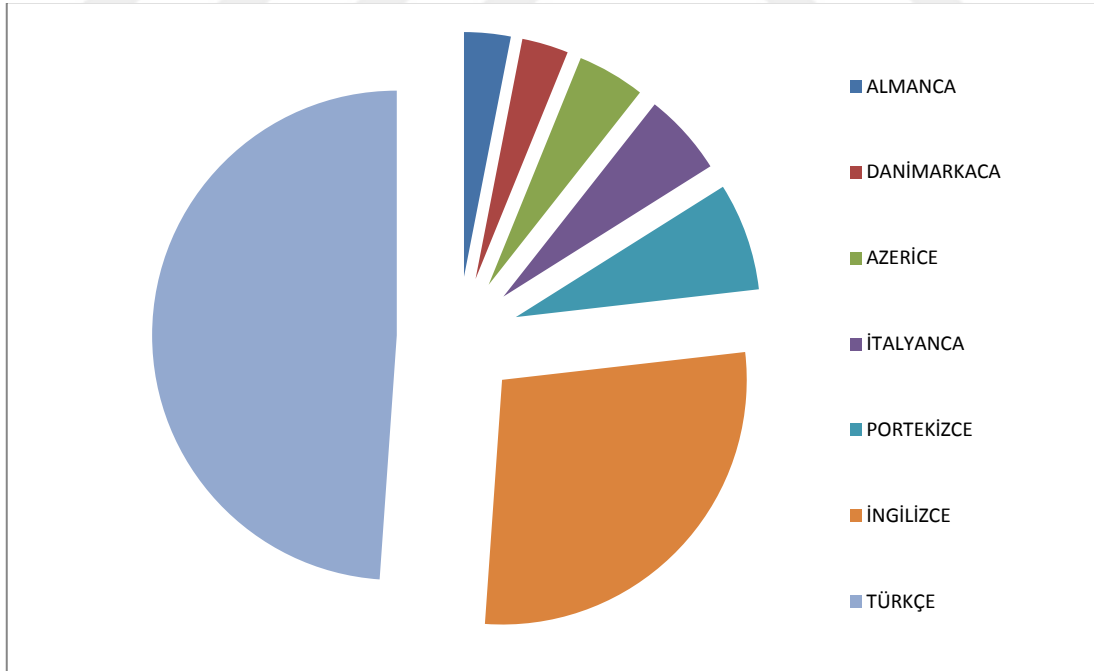


Şekil 3.1. Çalışmada kullanılan 5 farklı topoloji tipi

Şekil 3.1'de gösterilen topolojilerden (a) çizgesindeki $N1$, $P1$ ve $P2$ 'nin fotoğrafını beğenmiştir ve $N2$ 'de $P2$ 'nin fotoğrafını beğenmiştir. Bununla birlikte $P1$ ve $P2$ 'nin popüler kullanıcılar olduğunu biliyoruz. Biz $N2$ 'nin $t+1$ anında $P2$ 'ye beğeni atmak veya yorum yapmak sureti ile arkadaş olup olmayacağını tahmin etmeye çalışmaktadır.

3.2.2 Kullanıcı Dilleri

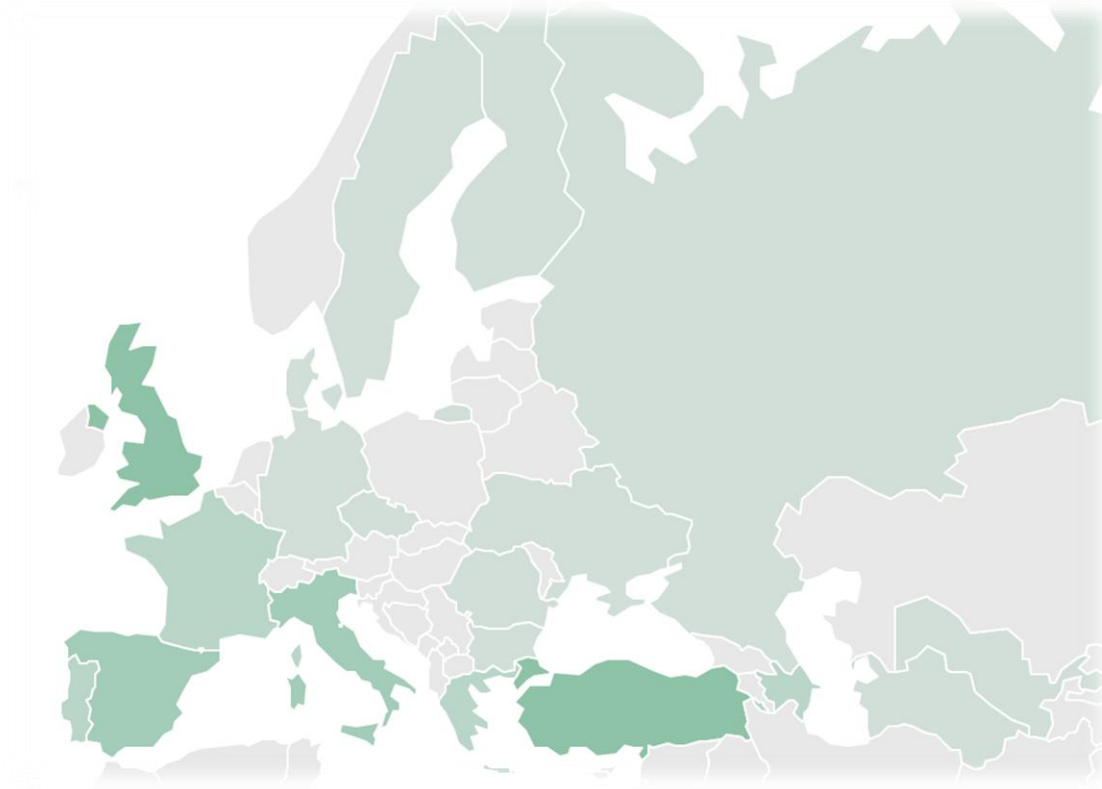
Dil başka bir insan ile iletişim kurmanın yegane yöntemidir. Benzer dili kullanan kullanıcılar diğer kullanıcılara göre nispeten daha fazla etkileşime girer ve yeni arkadaşlıklar/bağlantılar oluşturur bundan dolayı dil bağlantı tahmininde önemli bir faktördür. Instagram kullanıcıların kullandıkları dil ile ilgili bilgi tutmamakta ve/veya bu veriyi API ile sunmamaktadır. Çalışmamızda Google Translation API ile kullanıcıların yorumları analiz edilip kullanıcının dili tespit edilmeye çalışılmıştır. Kullanıcıların kullandıkları tüm diller veri havuzundaki kullanım sıklığına göre 7 popüler dil ve diğerleri olarak sınıflandırılmıştır. Sistem sadece Türkçe-Türkçe gibi aynı dili kullanan kullanıcıların korelasyonunu değil aynı zamanda Türkçe - İngilizce gibi farklı dilleri kullanan kullanıcıların da korelasyonunu hesaplamaktadır. Eğer kullanıcı diğer olarak sınıflandırılan dillerden birini kullanıyorsa dil korelasyonu sıfır olarak sisteme geçmektedir. Unutulmamalıdır ki başlangıç noktası tezin yazarı olduğundan örümceğin topladığı kullanıcıların büyük bir bölümü Türkçe konuşan kullanıcılardır. Sistemde bulunan kullanıcıların en çok konuştuğu diller Şekil 3.2'de ve en sık kullanılan dillerin dünya sıcaklık haritasında gösterimi Şekil 3.3.'de verilmiştir.



Şekil 3.2. En sık kullanılan ve korelasyonu hesaplanan 7 dilin kullanım oranları sunumu

Google Translation API'nin bir parçası olan Google Detect Language servisi hali hazırda birçok çalışmada kullanılmıştır ve Bergsma [77] 'a göre Twitter'daki twit'lerin dil tespitini yaklaşık 85% doğruluk oranında yapmaktadır.

Şekil 3.3'de, kullanıcının konuştuğu tespit edilen dilin konuşulduğu ana ülke'de yaşadığı varsayılarak ülke dağılımı yapılmıştır. Bir örnek vermek gerekirse tüm İngilizce yorum yapan kullanıcıların İngiltere de yaşadığı varsayılmıştır.

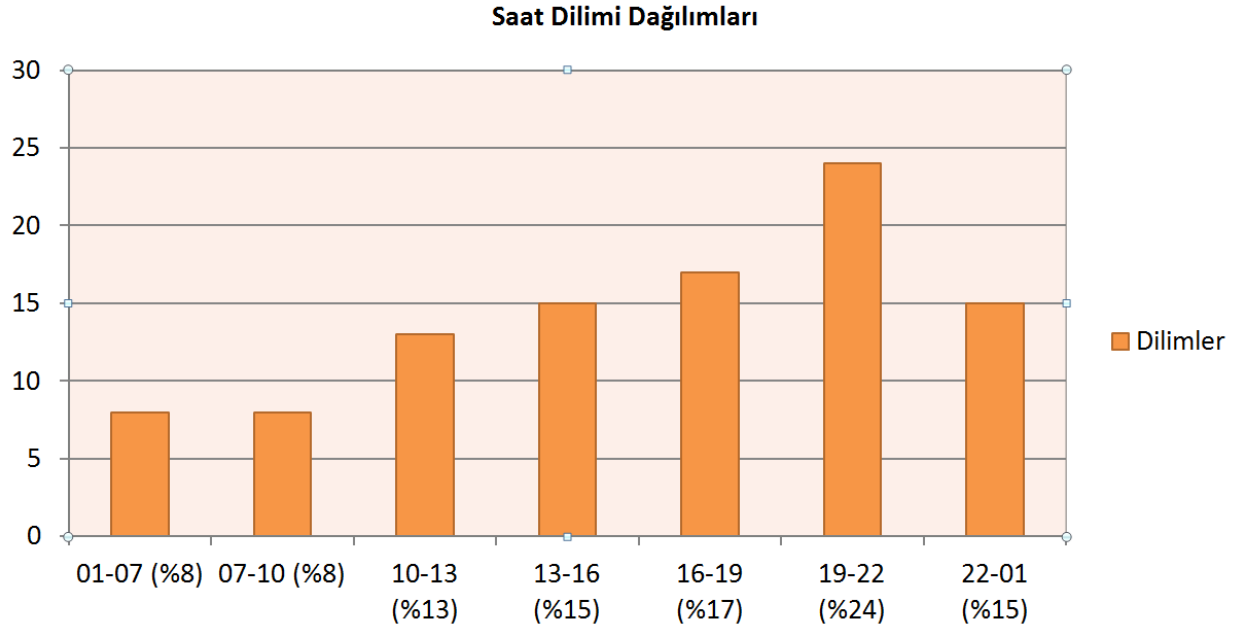


Şekil 3.3. Veri havuzumuzda toplanan kullanıcıların dillerine göre Avrupa Kıtasında dağılımı.

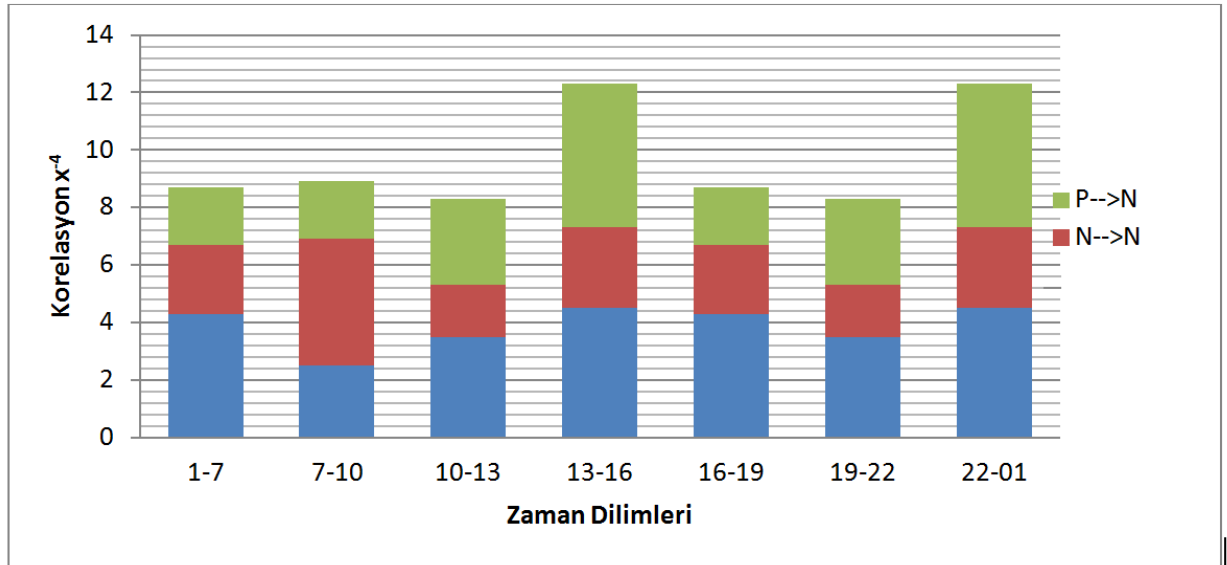
3.2.3. Kullanıcıların Aktif Oldukları Saat Dilimleri

Benzer saatlerde aktif olan kullanıcıların birbiri ile etkileşime girmeleri farklı saatlerde aktif olan kullanıcılara göre daha yüksek ihtimaldir. Saat dilimlerine göre etkili iletişimi gösteren bir çalışma Mucha [78] tarafından yapılmıştır. Bundan dolayı çalışmamızda 24 saatlik periyodu 7 farklı zaman dilimine böldük. Kullanıcıların fotoğraf paylaşma saatlerini ve beğeni atma saatlerine göre en aktif saatini bulduk. Saat dilimlerini gösterirken üç basamaklı ikili kodlama kullandık. Buna göre $X=0$ veya $X=1$ olmak üzere XXX kullanıcının aktif olduğu dilimi gösteriyor. 000 günün ilk vakitlerini göstermekteyken, 111 günün son kısmı demek oluyor. Kullanıcıların tüm işlem vakitleri sunucu saatine göre

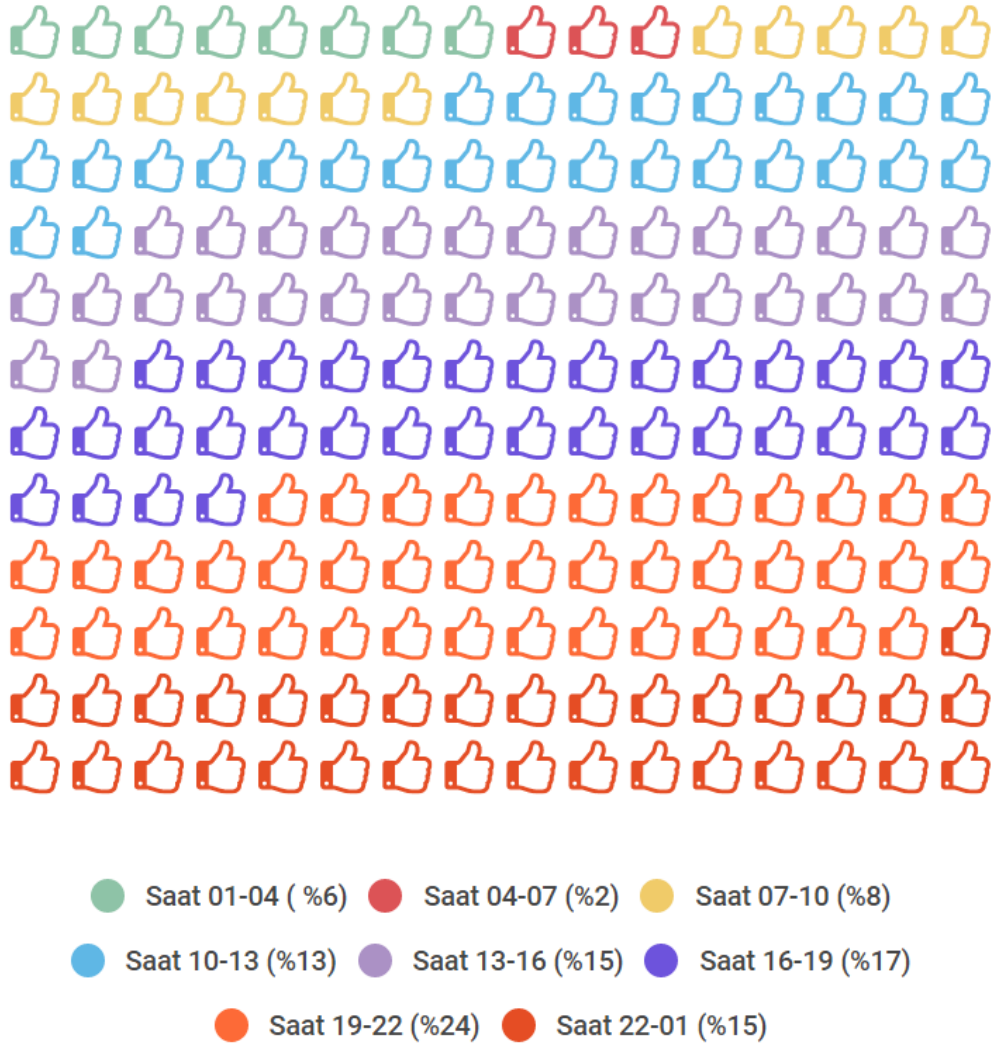
kaydedildiği için veri kaynağından sunucu saatine göre alabildik. Saatlerin dağılım grafiği Şekil 3.4'de sunulmuştur. Saat dilimlerine ve popüleriteye göre korelasyon oranları Şekil 3.5'de gösterilmektedir. Şekil 3.6'da ise 8 farklı dilimde dağılım gösterilmiştir.



Şekil 3.4. Saat dilimine göre aktif kullanıcıların frekansı. En çok kullanıcı 19-22 saatleri arasında aktiftir.



Şekil 3.5. Kullanıcıların kendi aktif oldukları saat diliminde çeşitli popülerlikteki kullanıcılar ile etkileşimi. (Mavi N->P 'yi temsil eder).

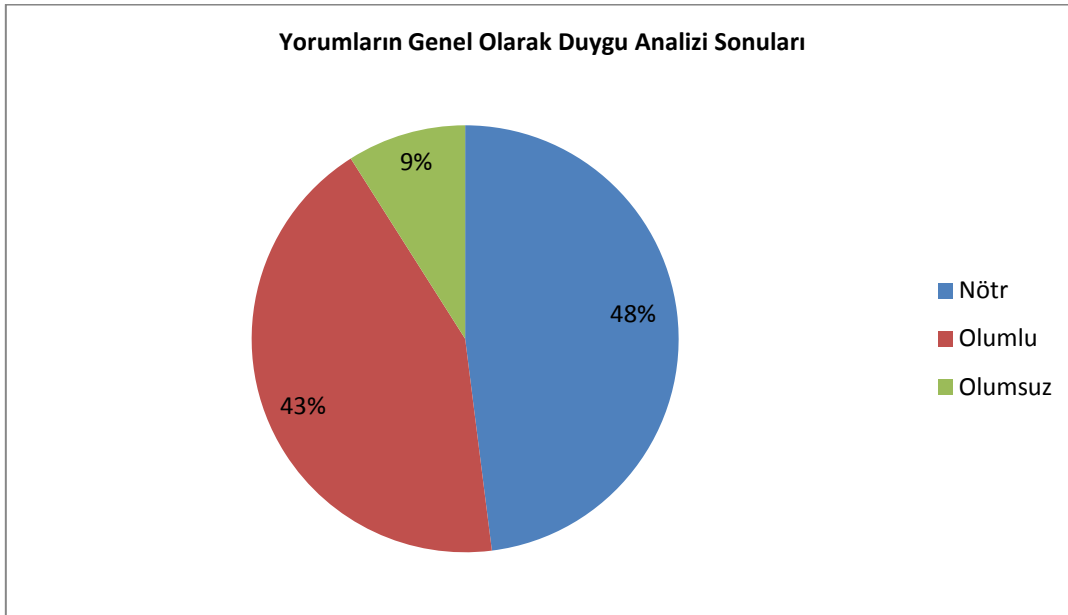


Şekil 3.6. Kullanıcıların kendi aktif oldukları saat diliminin 8 farklı modda gösterimi

3.2.4. Yorumlarda Duygu Analizi

Sosyal ağlarda duygu analizi ile ilgili birçok çalışma yapılmıştır [79,80]. Sosyal medya'daki duygu halinin kullanıcının gerçek psikolojik durumunu yansıtmadığına inanan çalışmalar da [81,82] yapılmış olmasına rağmen uzmanlar kullanıcının psikolojisinin yeni arkadaşlıklar oluşturma sürecinde önemli olduğuna inanmaktadır. Bundan dolayı, tüm yorumlarda duygu analizi yaptık ve kullanıcının duygusunu tespit etmeye çalıştık. Duygu analizi yapılan yorumların %48'i nötr çıkmıştır. Bu oranın çok yüksek çıkmasının bir nedeni yorumların genelde duygu analizi yapılamayacak kadar kısa olmasıdır. Duygu analizi yapan birçok çalışma ve kütüphane vardır. Biz çalışmamızda web servis olarak hizmet sunan 'Sentigem' duygu analizi aracını seçtik. Cieliebak ve arkadaşlarına [83] göre Sentigem yaklaşık 0.61 doğruluk oranına sahip olan bir duygu analizi aracıdır. Yorumların

genelinin duygu analizi sonuçları Şekil 3.7’de sunulmuştur. Bunların dışında hashtagleri de bir faktör olarak alınıp kullanılmak istenmiştir. Buna karşın hashtag’e bağlı herhangi bir korelasyon bulunamadığı için yani hashtag arkadaşlık oluşturmak için ortak özellik olmaktan çok uzak olduğu için alınmadı. Hashtag’i almamamızın bir diğer nedeni de birçok hashtag’in yaşam ömrünün değerlendirme yapılamayacak kadar kısa olmasıdır. Sosyal ağlar’da kullanılan birçok hashtag sadece birkaç saatlik ve birkaç gün aktif olarak kullanılmaktadır.

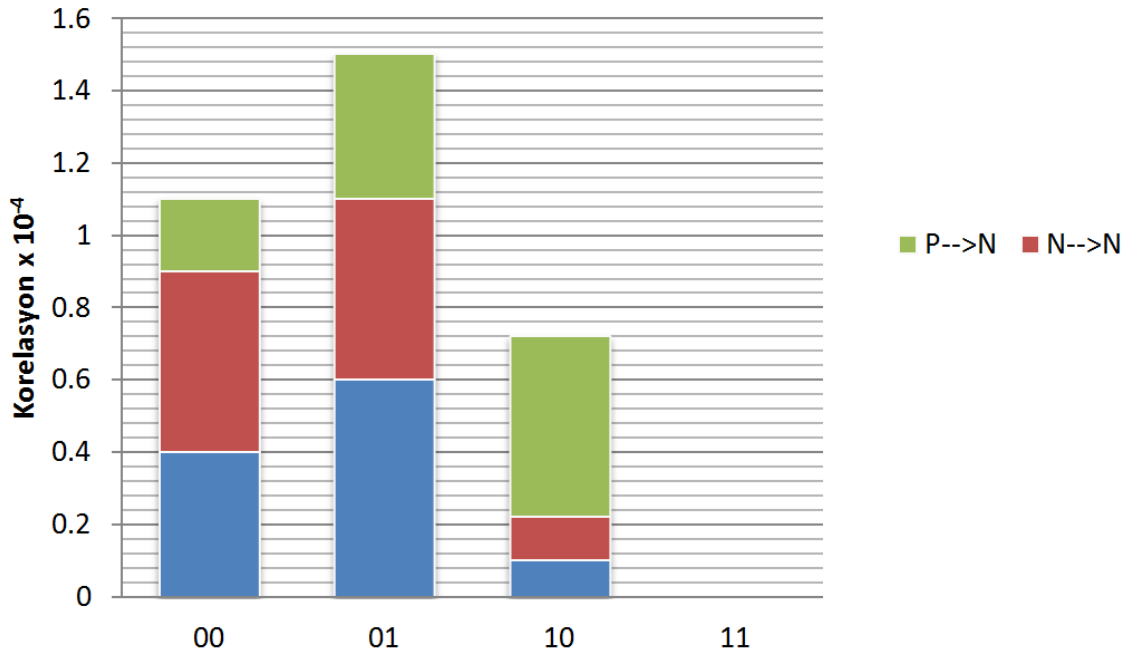


Şekil 3.7. Yorumların duygu analizi genel sonuçları gösterimi

3.2.5. Sosyal Roller

Sosyal ağlardaki sosyal roller hem topolojilerde hem de kullanıcı korelasyonlarında kullanılmıştır. Çalışmada kullanıcıları 3 grupta sınıflandırdık. İlk grup P ile ifade edilen Popüler isimli gruptur ve bu grup kendi takipçilerinin takipçi sayısının ortalamasının 4 katı olan kullanıcıları kapsar. Popüler grubuna genel olarak ünlü yazarlar, TV oyuncular, başkanlar ve futbolcular gibi gerçek hayatlarında ünlü olan kullanıcılar girer. Popüler’e benzer olarak Mid-Popüler grubu M ile gösterilir ve kendi takipçilerinin takipçi sayılarının ortalamasının 2 katı ile 4 katı arasında olan gruptur. Bu grup genelde yerel veya bir alanda ünlü kişileri kapsar. Son grup N ile gösterilen standart kullanıcı grubudur. Aynı zamanda toplam kullanıcı sayısından en büyük paya sahip gruptur. Bu gruptaki kullanıcılar, takipçi sayısı ile takip edilen sayısı ortalama olan kullanıcılardır. Çalışmamızda sosyal rolleri ikili

kodlama sisteminde iki karakter ile gösterdik. 00 standart kullanıcıları temsil ederken 10 popüler kullanıcıları temsil eder. 11 kodu Instagram'da 1 milyondan fazla takipçisi olan kullanıcıları ifade eder ve 264 kullanıcıyı kapsar. Bu kullanıcılar algoritmada değerlendirmeye alınmazlar çünkü bu kullanıcıları takip edenler arasında herhangi bir ilişki ve ortak özellik olması beklenmemektedir. Şekil 2.18.'da kullanıcıların kendi sosyal gruplarına ve diğer sosyal gruplara göre korelasyon oranları sunulmuştur.



Şekil 3.8. Kullanıcıların sosyal rollerine göre kendi grupları ile korelasyon değerleri (Mavi N-->P'yi Temsil Eder).

3.2.6. Geçişlilik

Sosyal ilişki teoreminde Geçişlilik [22] etkili bir kavramdır. Örneğin A , B ve C birbirinden bağımsız düğümler olsun ve arasında $A \rightarrow B$ ve $B \rightarrow C$ bağlantıları olsun $A \rightarrow C$ bağlantısının oluşup bu düğümlerin bir üçlü kapanış olma olasılığı diğer hiç bir faktör ve topolojiye bakılmaksızın yaklaşık %72'dir. Ancak, bu hesaplama gerçekleştirilirken popüler kullanıcılar başarı oranını çok düşürmemesi için işleme katılmamıştır.

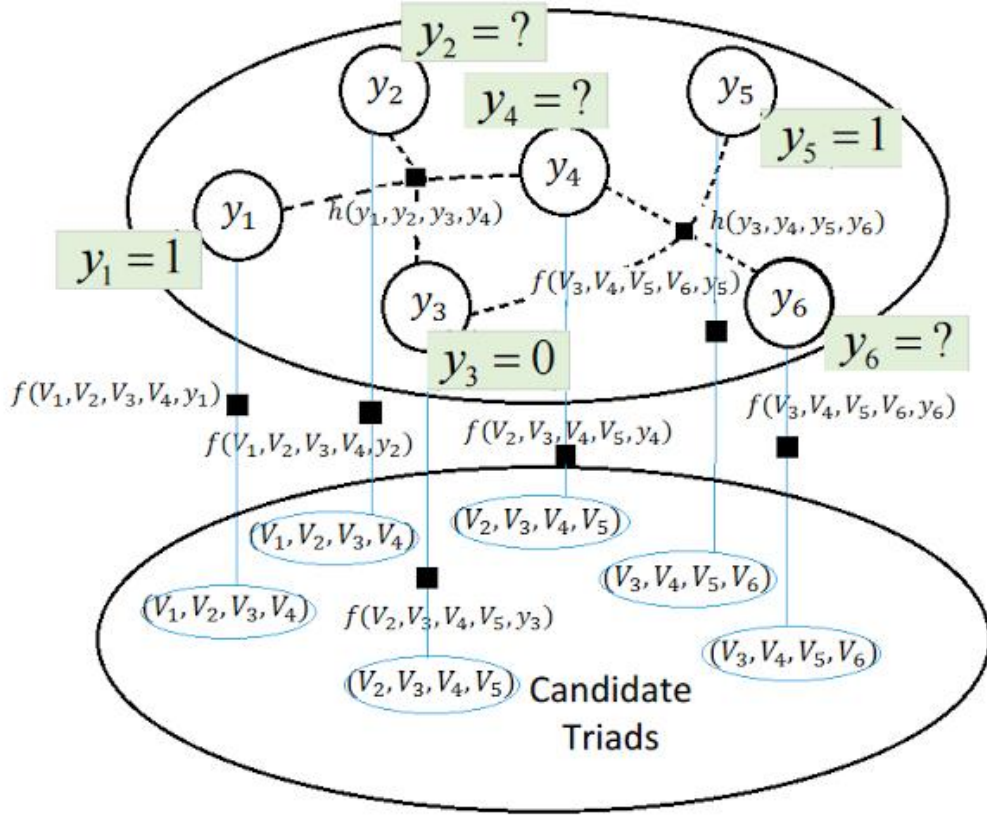
3.2.7. Özet

Çalışmada kullandığımız faktörlere genel olarak baktığımızda duygu analizinin korelasyon değerlerinin çok düşük olduğunu ve ayırt edici bir faktör olmadığını görebiliyoruz. Bununla birlikte hashtag'e bağlı korelasyon değerlerinin de, hashtag'lerin çok geniş kitlelere dağılması ve çok kısa ömürlü olmasından dolayı iyi sonuç vermediği aşikardır. Bunlardan dolayı, duygu analizi ve hashtag'in yeni arkadaşlıklar oluşturulması yani link tahmini yapmada önemli faktörler olmadığını söyleyebiliriz.

Diğer yandan, kullanıcının aktif olduğu saat dilimleri bağlantı tahmini için önemlidir. Çünkü aynı saat diliminde Instagram'ı kullanan kullanıcıların büyük oranı aynı ülkede yaşayan veya aynı yaşama stiline sahip olan insanlardır ve bundan dolayı aralarındaki korelasyon yüksektir. En önemli faktör ise dildir. Birbiri ile etkileşime en açık olanlar aynı dili konuşanlardır ve bunlar korelasyon oranı en yüksek olan kullanıcılarıdır.

3.3. Dörtlü Kapanış Metodu

Topoloji, bağlantı tahmini yapmak için yeterli değildir. Bundan dolayı topolojiye ek olarak birkaç faktör sisteme entegre edilmelidir. Çalışmada; ağ topolojileri, kullanıcı biyografik bilgileri ve faktörlerin etkileri dahil edilerek bir açık dörtlü kapanışının kapalı olabilme ihtimalini bulan bir model ortaya çıkartılmıştır. Yöntem, TriadFG'yi [70] temel alır ancak TriadFG'nin aksine önerilen modelde aktör sayısı 3 yerine 4 olduğu ve 4 kişinin arasındaki ilişkileri takip ettiğimiz için daha karmaşıktır. Bu nedenle TriadFG den farklı faktörler ve farklı topolojiler kullanılmaktadır. Şekil 3.9'da Mavi elipsler açık olup kapalı triad olmaya aday triadları gösterir, $f(v1, v2, v3)$ faktör fonksiyonunu gösterir ve $h(.)$ korelasyon fonksiyonunu ifade eder.



Şekil 3.9. Beş aktörlü bir Dörtlü Kapanış modeli

T anında bir $G^t = \{V, E, X, Y\}$ ağımız olsun. V düğümlerin listesi ve $E \subset V \times V$ bağlantılar olmak üzere $\textcircled{R}$ daha önce bahsettiğimiz faktörler ve Y açık dörtlünün kapanıp kapanmadığını gösterir değişken olur. Örneğin Şekil 3.9'da basit bir dörtlü kapanış sunulmuştur. Burada açık dörtlüler $(\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4)$ dir. Tahmin sürecinde 6 aday mavi elips ile çizilmiştir. F fonksiyonu kapalı dörtlü olma durumunu gösterir. Bunların dışında, $h(\cdot)$ ise f fonksiyonu ise korelasyon değerini gösterir. Beş aktörlü bir dörtlü kapanış örneği Şekil 3.10'de sunulmuştur.

$G^t = \{V^t, E^t, X^t\}$ t anında grafın durumu ve y şimdilik bilinmeyen değişkendir. Bizim amacımız $t+1$ anında y 'nin değerini bulmaya çalışmaktır. Biz Markov Dağılım ve Hammersley-Clifford [84] teoremi ile $P(Y|X, G)$ 'nin sonlu olasılığını 3.2'de gösterildiği gibi hesaplanır.

$$P(Y|X, G) = \frac{1}{Z} \exp \left\{ \sum_{i=1}^{|Tr|} \sum_{j=1}^d a_i f_i(x_{ij}, y_i) + \sum_c \sum_k \mu_k h_k(Y_{Tr_c}) \right\} \quad (3.2)$$

$|Tr|$ açık üçlü kapanış adaylarını gösterir, d ise hesaplamaya giren faktör sayısıdır. x_{ij} ve j^{th} dörtlünün değeri ve c korelasyonu temsil eder. Örneğin, $h(y_1, y_2, y_3, y_4)$ bu

düğümün korelasyon fonksiyonunu gösterir ve a_j ile μ_k bu korelasyon fonksiyonun parametreleridir. Son olarak Z normalleştirme faktörüdür.

Çalışmamızda, korelasyon faktörü değişkenleri işlem yaptıkça geçmiş işlem bilgisine göre güncellenir. Bundan dolayı bu işlemi hesaplayan bir formülümüz daha vardır. $\theta = (\{a_j\}, \{\mu_k\})$ ve $\theta = \arg \max O(\theta)$ olmak üzere *gradient descent* metodunu öğrenmek sürecinde uyguluyoruz. Bu metodun ilk adımı, değişkenlerin başlangıç değerlerini rastgele atamaktır ve sistem işledikçe de dönüp bu değerleri revize ederek Lou ve arkadaşlarının [66] çalışmasına benzer şekilde gerçekleştirmektedir.

3.4 Dörtlü Kapanış Metodu Deney Sonuçları

Çalışmayı yaparken 3 farklı zamanda verinin kaydını aldık. İlk kayıt ile ikinci kayıt öğrenme süreci için, ikinci kayıt ile üçüncü kayıt ise yaptığımız tahminleri test etmek için kullandık. Başarı kriterimizi ölçmek için 3 farklı metot kullandık. Bunlar Kesinlik, Anma ve F-Ölçütü. Tüm algoritmalar Java dili ile yazılmış ve Intel® Core(TM) i5-4300 CPU @ 1.90GHz 8 GB hafızalı bir Windows 7 bilgisayarda çalıştırılmıştır.

Tablo 3.1. Dörtlü Kapanış Tahmin Performansı

Algoritma	Kesinlik	Anma	F-Ölçütü
SVM	0.786	0.798	0.788
TriadFG	0.798	0.815	0.812
Dörtlü	0.812	0.850	0.825

Faktör Etki Analizi: Tüm faktörlerin korelasyon etkisi ayrı ayrı hesaplanmıştır. L, TZ, H ve S sırası ile dil, zaman dilimi, Hashtag kullanımı ve duygu analizini temsil eder.

Tablo 3.2. Faktör Etki Analizi

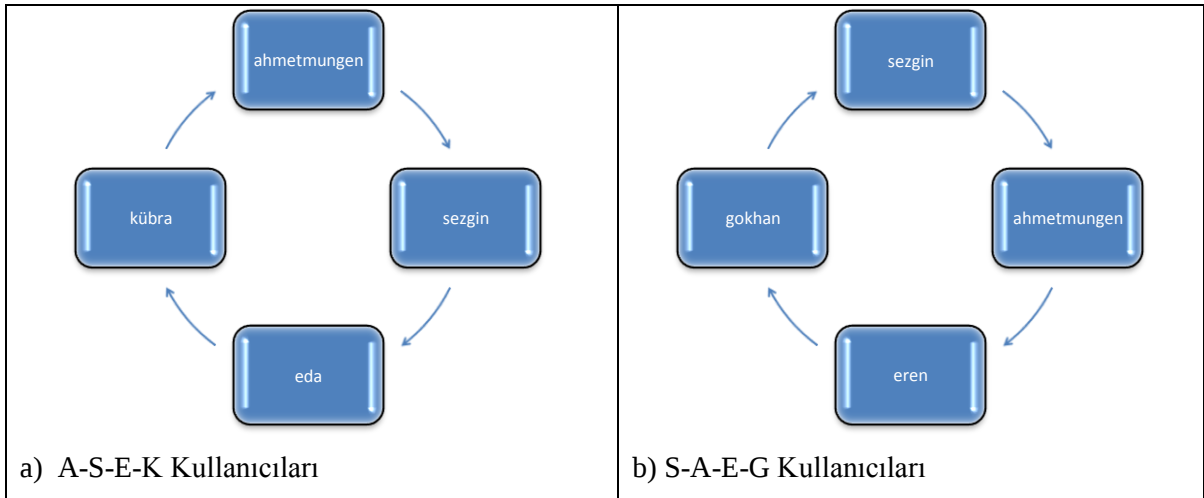
Method	Dörtlü(Q)	Q-L	Q-TZ	Q-H	Q-S
F-Measure	0.812	0.584	0.665	0.799	0.807

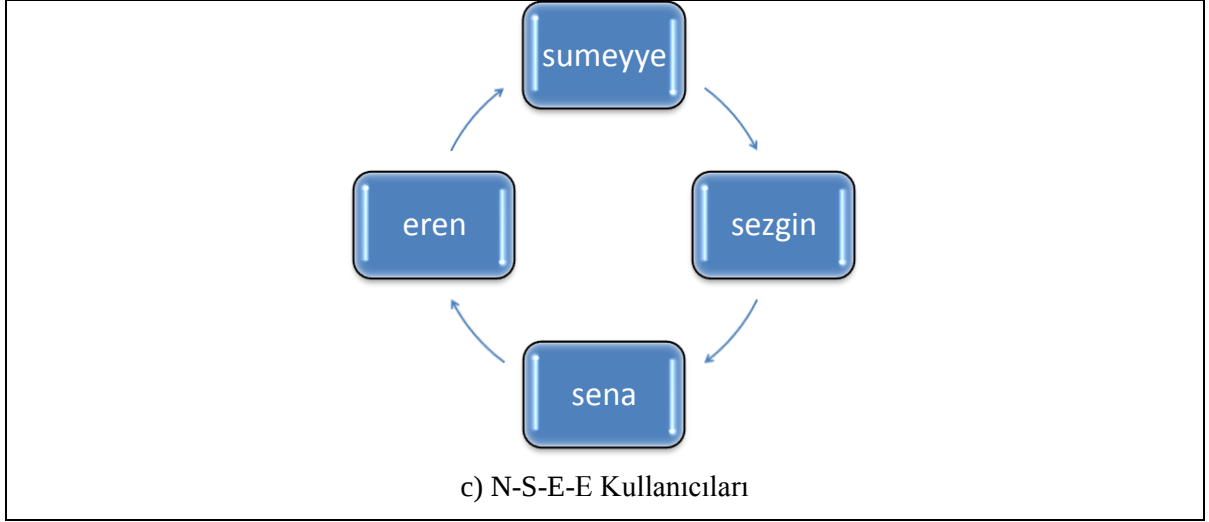
Q-L Dörtlü formülünün dil korelasyonu olmadan başarı oranını, Q-TZ ise dörtlü kapanış fonksiyonun zaman dilimleri olmadan başarı oranını gösterir. Tablo 3.2'de görüleceği gibi duygu analizi en az etkili ve dil ile zaman dilimleri en çok etkili

faktörlerdir. SVM ve TriadFG'ye göre Dörtlü metodu (+2.3%, $p \ll 001$) daha başarılı olmuştur.

Vaka Örneği 1 : Diğer algoritmaların bulamadığı ve bizim çalışmamızın bulduğu bir örnek Şekil 3.10'da görülebilir. Başlangıç olarak *sezgin_zerdelci* isimli kullanıcı ele alınmıştır. Kullanıcının 1. seviye arkadaş sayısı: 5, 2. seviye arkadaş sayısı yani arkadaşlarının arkadaş sayısı 77 ve 3. seviye arkadaş sayısı 4105'dir. Bu kullanıcı ile ilgili faktörler katılmadan sadece topolojik olarak aday döngülerin bulunması için 3521 ihtimal çıkartılmıştır. Bu ihtimallerden 3'ü Şekil 3.10'de sunulmuştur.

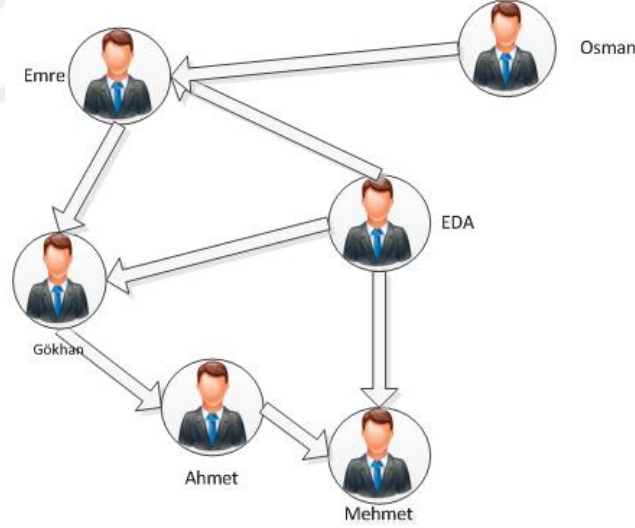
Şekil 3.10'deki *a* figüründe; sistemde *t* anında *ahmetmungen* kullanıcısının *sezgin* kullanıcısına, *sezgin* kullanıcısının *eda* kullanıcısına, *eda* kullanıcısının *kübra* kullanıcısına bağlantısı vardır. *T+1* anında *kübra* kullanıcısının *ahmetmungen* kullanıcısına beğeni atmak veya yorum yapmak sureti ile bağlantısı gerçekleşmiştir. *B* figüründe sistemde *t* anında *sezgin* kullanıcısının *ahmetmungen* kullanıcısına, *ahmetmungen* kullanıcısının *eren* kullanıcısına, *eren* kullanıcısının *gokhan* kullanıcısına bağlantısı vardır. *T+1* anında *gokhan* kullanıcısının *sezgin* kullanıcısına beğeni atmak veya yorum yapmak sureti ile bağlantısı gerçekleşmiştir. *C* figüründe sistemde *t* anında *sumeyye* kullanıcısının *sezgin* kullanıcısına, *sezgin* kullanıcısının *sena* kullanıcısına, *sena* kullanıcısının *eren* kullanıcısına bağlantısı vardır. *T+1* anında *eren* kullanıcısının *sumeyye* kullanıcısına beğeni atmak veya yorum yapmak sureti ile bağlantısı gerçekleşmiştir.





Şekil 3.10. Çalışma sonucunda bulunmuş 3 örnek Dörtlü Kapanış.

Vaka Örneği 2: Diğer algoritmaların bulamadığı ve bizim çalışmamızın bulduğu bir diğer örnek Şekil 3.11'de sunulmuştur. Bu örnek çalışmanın etki analizi bölümünde de kullanılmıştır.



Şekil 3.11. Altı aktörlü örnek bir etki çizelgesi

Şekil 4.11'de *Emre Gökhan* ve *Eda* bir üçlü kapanış içindedir. Tüm algoritmalar bunu tespit edebiliyor. Buna karşın *Eda*, *Gökhan*, *Ahmet* ve *Mehmet*'in içinde bulunduğu dörtlü dörtlü kapanış sadece bizim çalışmamızdaki algoritma ile tespit edilebilmiştir. *T* anında *Ahmet* ve *Mehmet* arasında bir bağlantı yoktur. Algoritma *t* anında böyle bir bağlantı olabileceğini gösterdi. Bağlantı *t+1* anında *Ahmet*'den *Mehmet*'e beğeni atılması ile oluşmuştur.

4. SOSYAL AĞLARDA ETKİ ANALİZİ

Günlük hayatımızda arkadaşlarımız ailelerimiz kararlarımızda etkili rol oynar ve bizde başkalarını verdiğimiz tavsiye, yaptığımız davranışlar ve paylaştığımız bilgiler ile kararlarını etkileriz. Karmaşık ağlarda grup bulmak kadar bulunan grup içinde kullanıcıyı etkileyen diğer kişileri bulmak ve gruba en çok etki edeni bulmak hem sosyoloji hem matematik bilimi hem de bilgisayar bilimi açısından önemli bir araştırma alanıdır. Sosyal ağlarda arkadaşının ya da takip ettiği kişinin yaptığını yapma ve onun ilgilendiği alana ilgi duyma bir yönlendirme ve etki şeklidir [85]. Sanal sosyal ağlarda yani diğer bir deyişle sosyal medyada birçok zaman salgın şekline dönüşen trend'ler ortaya çıkmaktadır. Bu trendlerin yayılması da o trendlere ilk ilgi duyan kişilerin etki alanları ile ilgilidir bu tür etkilere sosyal korelasyon denir [86]. Sosyal ağlarda görülen bir paylaşım veya fotoğrafa birini etiketlemek etkiyi arttıran yöntemlerden biridir ve kullanıcılar paylaşımlarını genelde en iyi anlaştıkları başka bir deyiş ile en iyi etkiledikleri veya etkilendikleri kişiyi etiketleyerek yaparlar [87]. Anagnostopoulos ve arkadaşlarına göre [85] etki bir kullanıcının diğerini bir konuda aksiyon alması için tetiklemesidir. Çalışmanın bu bölümünde kullanıcıların birbiri üzerinde yaptıkları etkileri ölçen farklı metotlardan bahsedip karşılaştırmalı olarak sunulacaktır.

4.1. İstatistiksel Analiz ile Etki Analizi

Anagnostopoulos ve arkadaşları [85] "Influence and correlation in Social Networks" isimli makalelerinde kullanıcılar arasındaki sosyal korelasyonu bulmak için istatistiki ihtimal analizi yapmaktadır. Çalışmaya göre sosyal medyada her kullanıcı arkadaşının paylaştığı yazıyı gördüğü anda yazıyı beğenme, paylaşma veya yorum yapma şansı vardır. Anagnostopoulos ve arkadaşları $p(a)$ aksiyon alma ihtimali olmak üzere alfa ve beta da coefficient olmak üzere 4.1'deki formülleri kullanmaktadırlar.

$$p(a) = \frac{e^{\alpha \ln(a+1) + \beta}}{1 + e^{\alpha \ln(a+1) + \beta}} \quad (4.1)$$
$$\ln\left(\frac{p(a)}{1 - p(a)}\right) = \alpha \ln(a + 1) + \beta$$

A sosyal korelasyon katsayısı olmak üzere a 'nın büyüklüğü korelasyonun büyüklüğünü göstermektedir ve α ile β 'yi kullanarak *logistic regression*'u maksimum olmasına göre

işlem yapılır. Daha net olmak gerekirse Y_a, t anında a 'nın aktif arkadaşları olsun. N_a, t de t anından aktif olmayan arkadaşları olsun. Sonuç olarak formülümüz 4.2'de sunulmuştur.

$$\prod_a p(a)^{Y_a} (1 - p(a))^{N_a} \quad (4.2)$$

4.2. Benzerlik ile Etki Analizi

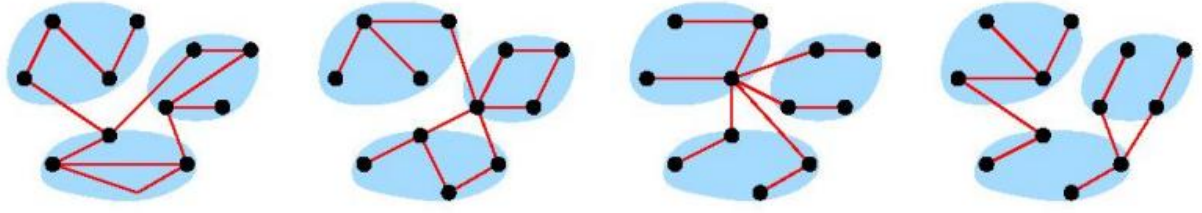
Crandall ve arkadaşları [88] benzerliğin sosyal etkisi konusunda yaptıkları çalışmada kullanıcıların benzerliklerinin sosyal ağlarda da benzer tepkiler vermeleri ve benzerliği yüksek olan diğer kullanıcıların yaptıklarını daha çok yapmaya eğilimli olduklarını söylemektedir. Buradan sosyal etkiyi arttıran nedenlerden birinin benzerlik olduğunu göstermeye çalışmıştır. Word-of-Mount [7,89] yani söyleyişle geçen salgın sosyal ağlarda hızla artan bir etki biçimidir. Sanal ağlarda bu salgın söyleyiş yerini yazı ve paylaşımlara bırakmıştır. Crandall ve arkadaşlarının çalışmasında sosyal etkiyi test etmek için Wikipedia verisini kullanarak kullanıcıların benzerliklerinin ilgi alanları ve hareketlerini etkileyip etkilemediği ortaya çıkarılmaya çalışılmıştır. Diğer modellerde olduğu gibi Crandall'da zamanı en önemli değişken olarak düşünüp zaman bazlı bir vektör ile kullanıcının aktivitelerini temsil etmiştir. v kullanıcı t zaman ve i de aktivite olmak üzere cosine metriğine göre formül 4.3'de gösterilmiştir.

$$\left[\text{Cosine}(\vec{u}, \vec{v}) = \cos \vec{u} - \vec{v} = \frac{\vec{u} \cdot \vec{v}}{||\vec{u}||_2 ||\vec{v}||_2} \right] \quad (4.3)$$

4.3. Sosyal Birleşmeler ve İstatistiki Araçlar ile Etki Analizi

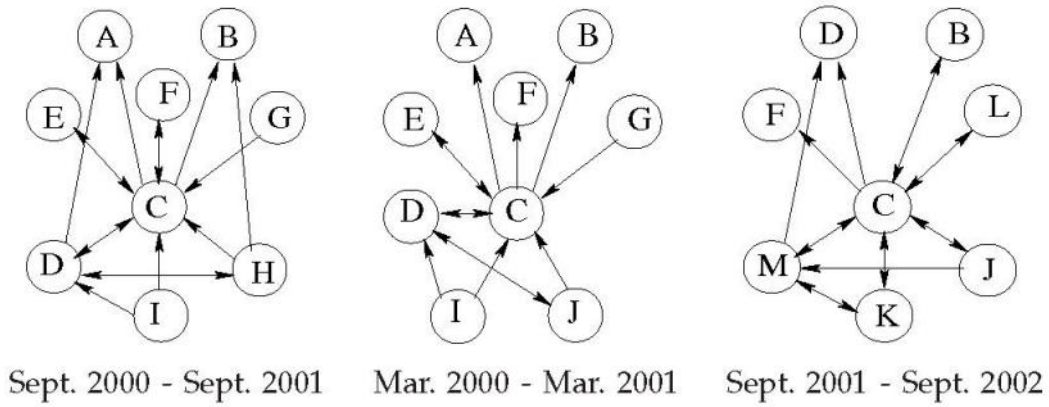
Baumes ve arkadaşları [90] tarafından ortaya çıkarılan SIGHTS bir algoritmik modelden öte verileri toplayıp çizgeler aracılığı ile modellemeye yardımcı olan, önceden tanımlanmış fonksiyonların bu verilere uygulanmasına izin veren bir çatı programıdır. SIGHTS, blog, sosyal medya, forum siteleri gibi internet servislerinden verileri toplayıp iyi yapılandırılmış XML'e dönüştürüp kaydeder. Bu verileri kullanıcının belirleyeceği filtrelerle göre parçalar, çizgelere dönüştürür ve örüntü keşfi yapmaya çalışır [91,92].

SLIGHT'ın kapsadığı algoritmalarından biri Baumes ve arkadaşlarınca yapılan akan iletişimde grup bulmayı amaçlayan *Temporal Group* Algoritmasıdır. Diğer bir algoritma Döngü Modeli algoritmasıdır ve kullanıcılar arasında etkileşimlerle birlikte oluşan döngüleri bulmayı amaçlar. *Temporal Group* Algoritmalarının 4 örneği Şekil 4.1 'de sunulmuştur.



Şekil 4.1. Temporal Group Algoritması ile bulunan etkileşimleri

SLIGHT'ın kapsadığı bir diğer algoritma Stream modelidir. Bu algoritma kümeleri bulmayı amaçlar. Şekil 4.2 de görülen grafik ENRON mail grubu içinden çıkartılmış kullanıcılar arası ilişkiyi ifade eder.



Şekil 4.2. Stream modeli ile bulunan etkileşimlerin gösterimi

4.4. Yerel Yakınlık Yayılımı ile Etki Analizi

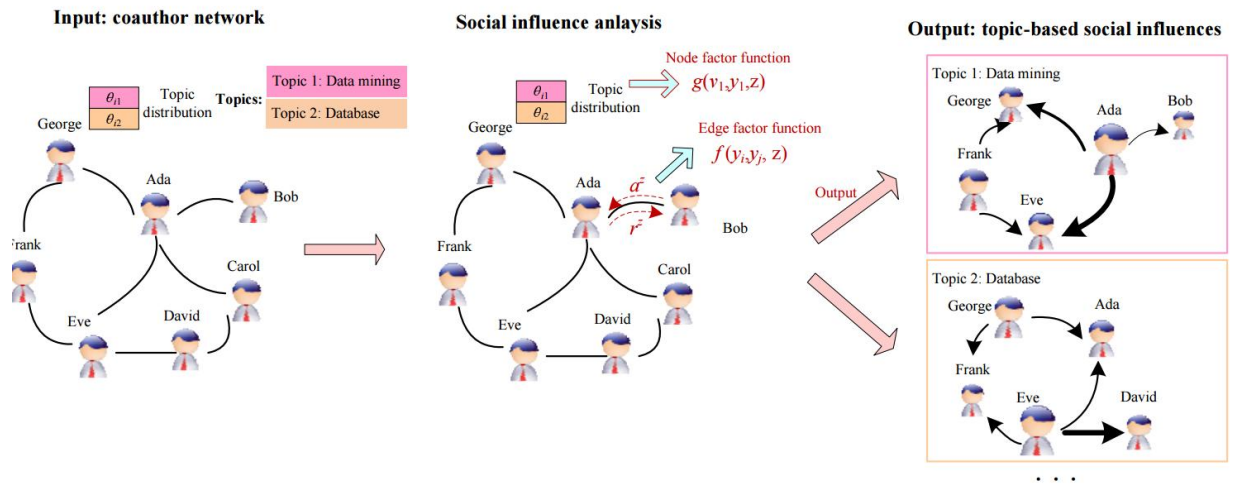
İnsanların sosyal ağlarda birbirini etkilediklerini çalışmamız içinde birçok defa bahsettik. Akademik çalışma yapanların da birbirlerinin çalışmalarından etkilendikleri ve arkadaşlarının çalışmaları veya çalıştıkları alana ilgi duymaları bilinen bir davranıştır. Akademisyenler gibi üretken meslek gruplarının birbirinden ne oranda etkilendikleri, bundan da önemlisi bu etkinin büyük bir sosyal ağı uygulandığında nasıl sonuç vereceği bilinmemektedir. Tang ve arkadaşları makalelerinde [93] önerdikleri Yerel Yakınlık Yayılımı(Topical Affinity Propagation(TAP)) metodu ile bu sorunun cevabını bulmaya çalışmışlardır.

Birçok sosyal ağ analizi yöntemi derece dağılımları, diameter, kümeleme katsayısı [94–97] gibi makro-seviyede modeller sunmayı amaçlar. Sosyal etki üzerine yapılan çalışmaların çoğunluğu da nitel bir değer aramaktadır[98, 99]. Tang ve arkadaşları çalışmalarında belli bir konuda ve/veya belli bir grup içinde en etkili bireyi aramaktalar ve

topluluğu en çok etkileyen diğer bir deyiş ile en fazla bilgiye sahip uzmanı, dinamik sosyal ağlarda hızlı bir şekilde aramayı sağlayan model sunmaktadırlar.

Tang ve arkadaşları modellerini oluştururken modellerinin üç ana özelliği içermesine dikkat etmişlerdir. Bu özellikler çok yönlülük, düğüm spesifik, ölçeklenebilirliktir. Çok yönlülük A'nın B'yi güçlü etkilemesinin yanında B'ninde A'yı zayıf da olsa etkileyebileceğini hesaba katmaktır. Düğüm spesifik: sosyal etki global olarak değil düğüm bazında yani kullanıcı bazında veya az sayıda kullanıcı bazında ölçülebileceğidir. Ölçeklenebilirlik ise modelin gerçek bir sosyal medyada çok büyük miktarda veri setinde çalışabilir olmasıdır.

Şekil 4.3.'deki ilk ağ ortak yazı yayınlayan yedi yazarın birbiri ile ilişkisini göstermektedir. Ortadaki grafikte aynı anahtar kelimeleri kullanan yazarlar arasında ilişki gösterilmiştir. Sağdaki sonuç ise TAP yöntemi ile analiz edilmiş sonucu vermektedir. Ada isimli kullanıcı veri madenciliği konusunda anahtar isimdir Eve ise veritabanları konusunda anahtar isimdir.



Şekil 4.3. Ortak Yazar ağından sosyal etki ve TAP metodunda sonuçları [93]

4.4.1. Yerel Faktör Çizgesi (TFG)

Bir TFG de değişkenler $\{v_i=1\}$, gizli vektörler $\{y_i\}$ ve düğümlerden (N) oluşan bir ağıdır. Gizli vektör $y_i \in \{1, \dots, N\}$ T konu bazlı olarak ifade edilirse her y, z i elemanı $\{1, \dots, N\}$ arasında bir değer alır ve z konusu üzerinde v_i 'nin maksimum etkisini gösterir.

- $g(v_i, y_i, z)$ v_i düğümünün z konusu üzerinde etkinliğini gösterir.

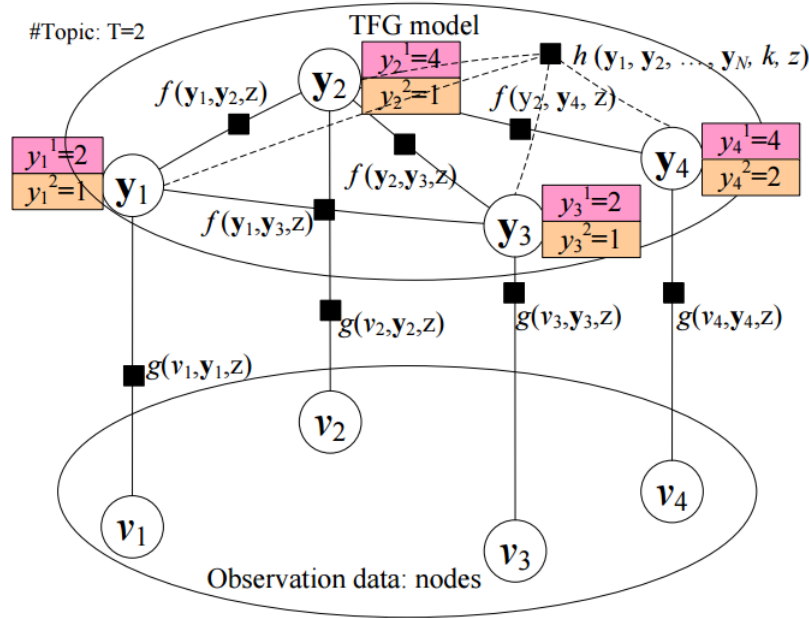
- $f(y_i, y_j, z)$ z konusu üzerinde y_i 'yi gösteren düğümünün etkinliğini gösteren bağlantıdır
- $h(y_1, \dots, y_N, k, z)$ tüm düğümlerin z konusu üzerinde değerlerini gösterir.
- Buradan v_i düğümünün z konusu üzerinde etkinliğin genel formülü 4.4'de sunulmuştur.

$$g(v_i, \mathbf{y}_i, z) = \begin{cases} \frac{w_{iy_i}^z}{\sum_{j \in NB(i)} (w_{ij}^z + w_{ji}^z)} & y_i^z \neq i \\ \frac{\sum_{j \in NB(i)} w_{ji}^z}{\sum_{j \in NB(i)} (w_{ij}^z + w_{ji}^z)} & y_i^z = i \end{cases} \quad (4.4)$$

Ortak Dağılım için faktör değer modelleri, bu formüle göre uygulanması 4.5'de formülize edilmiştir.

$$\begin{aligned} P(v, Y) &= \frac{1}{Z} \prod_{k=1}^N \prod_{z=1}^T h(y_1, \dots, y_N, k, z) \\ &\prod_{i=1}^N \prod_{z=1}^T g(v_i, y_i, z_i) P(v, Y) \\ &= \frac{1}{Z} \prod_{k=1}^N \prod_{z=1}^T h(y_1, \dots, y_N, k, z) \end{aligned} \quad (4.5)$$

TFG'nin 4 aktörlü Çizge üzerindeki görünümünü Şekil 4.4'de sunulmuştur.



Şekil 4.4. TFG Gösterimi [93]

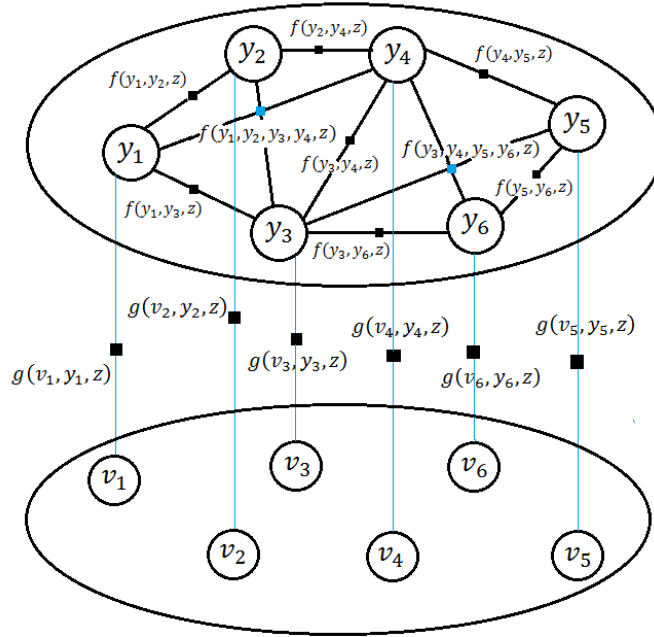
4.4.2. Test Sonuçları

Tang ve arkadaşları [93] önerdikleri bu yeni metodun testini biri heterojen ikisi homojen olmak üzere akademik makaleler verisinde yapmışlardır. Toplamda 640 bin'den fazla yazar ve bu yazarlarla ilgili 1.5 milyon'dan fazla ilgili yazar ve 2 milyon'dan fazla makale ile 12 milyon'dan fazla atıf ele alınmıştır. Çalışmada Author-Conference-Topic tarafından seçilmiş 200 konu ele alınmıştır. Test edilen ve bizim tez çalışmamız ile birebir ilgili olan diğer veri de *wikipedia* verisidir. 18 bin film, 10 bin aktör ve 9 bin yazar ile yapılan testde film-yönetmen, film-aktör ve film-yazar arasındaki ilişkiler ve üçlü kapanışlar bulunmaya çalışılmıştır. Akademik makaleler arasında olduğu gibi burada da 10 kategori seçilmiştir.

Modelin başarısını ölçmek için CPU zaman, örnek olay ve uygulama geliştirme ele alınmıştır. CPU zime kriterinde bakılan veri miktarına göre modelin çalışma süresidir. Örnek olaya göre ise konu bazlı olarak gerçek hayattan çekilmiş verilerdeki başarı ölçülmüştür. Uygulama geliştirme ise yapılan model içinde uzman bulma işleminin başarı oranı ele alınmıştır. Tang ve arkadaşları Matlab 2007b kullanarak Windows 2003 işletim sistemi üzerinde Dual-Core Intel Xeon işlemci ve 8 GB Ram kullanmışlardır. Map-Reduce programlaması için Hadoop kullanılmış ve 6 bilgisayarlı 48GB'li bir cluster oluşturulmuştur.

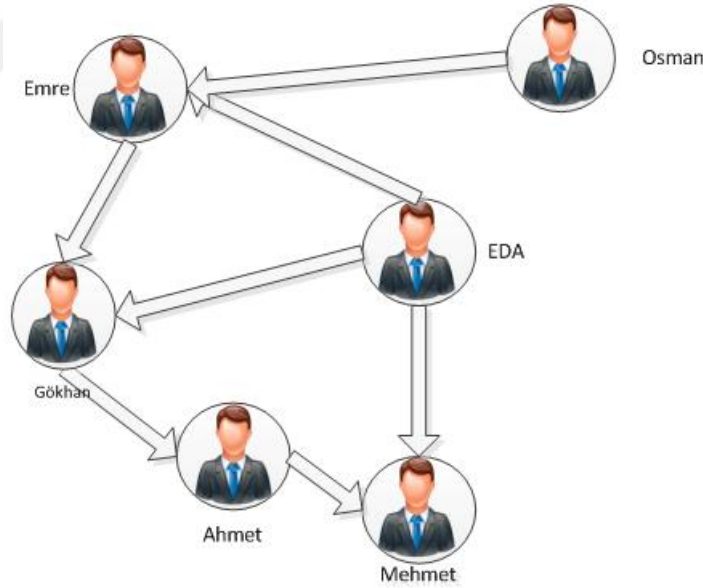
4.5. Geliştirilmiş Yerel Yakınlık Yayılımı (I-TAP)

Önerilen Yerel Yakınlık Yayılımı (Topical Affinity Propagation - I-TAP) Modeli, Tang ve Sun'ın "Topical Affinity Propagation (TAP)" [11] yaklaşımı temel alınarak gruptaki en etkili kişiyi bulmaya yönelik olarak geliştirilmiştir. I-TAP ile üçlü kapanış gruplarda değil dört aktörlü dörtlü kapanış gruplarında etki değeri ölçümü yapılmıştır. Şekil 4.5'de I-TAP'in 5 aktörlü basit bir gösterimi sunulmuştur.



Şekil 4.5. Geliştirilmiş Yerel Yakınlık Yayılımı için 6 aktörlü bir örnek

I-TAP'da TAP'a benzer olarak gruptaki yaprak düğüm başlangıç olarak kabul edilmiştir. Döngülü çizgelerde yaprak düğüm olmadığı için gruptaki biyografik bilgiye göre en düşük puanlı birey yaprak olarak kabul edilir. Başlangıç iç puanı bireyin yayınladığı fotoğraf sayısının aldığı beğeni sayısına bölümü ile bulunur. Eğer döngü varsa en düşük birey yaprak kabul edilir. Döngülü ise yaprak kullanıcı, değilse gruptaki kimseden beğeni almayan birey'den başlanır ve beğeni attığı kullanıcının puanına beğeni atan kullanıcının puanı kadar ekleme yapılır. Ekleme yapılan kullanıcı da başka bir kullanıcıya beğeni atarsa hem kendi puanını hem de yaprak kullanıcısından gelen puanı beğeni attığı kişiye puan olarak iletir. Sistem tüm düğümleri en az 3 kere gezer. Bunun başlıca nedeni aslında güçlü bir kullanıcıdan beğeni almış ama kendi iç puanı düşük olan kullanıcıların da puanının doğru hesaplanmasını sağlamaktır. Hem döngülü hem uç düğümün olduğu bir örnek Şekil 4.6'da sunulmuştur.



Şekil 4.6. Altı aktörlü bir etki grafiği örneği

Şekil 4.6'da *Osman* kendi öz seviyesinden *Emre*'ye puan vermiştir. *Emre* hem *Osman*'dan hem *Eda*'dan puan almıştır. *Gökhan*, *Emre* ve *Eda*'dan puan almıştır. *Gökhan*'ın *Emre*'den aldığı puanın bir bölümü *Osman*'dan gelmiştir. *Ahmet*, *Gökhan*'dan puanı almış bu puan en az 3 kullanıcının puanları toplamıdır. Aslında *Mehmet* sadece 2 beğeni almasına karşılık onu beğenen *Ahmet* en az 3 kişiden puan aldığı için bunun yanında *Eda* tarafından da beğeni aldığı için en az beğeni aldığı halde en yüksek puanlı kullanıcı olmuştur. Herkesin iç puanını 1 kabul edersek ve her ok sadece 1 beğeni gösteriyor ise *Osman*'ın puanı 1 olacak *Emre*'nin puanı ise 3 olacaktır. *Gökhan*, *Eda*'dan 1 , *Emre*'den 3 puan alacak bir puanda kendi iç değerinden geldiği için 4 puan olacaktır. *Mehmet*, *Ahmet*'ten 5 puan alacak *Eda*'dan gelen 1 puan ile birlikte 6 puan ile en yüksek puanlı birey olacaktır. *Mehmet*'den sonra *Ahmet* sadece 1 beğeni aldığı halde en yüksek ikinci puanlı kişidir.

5. SONUÇ

Bu çalışmada Sosyal ağlarda grup tespiti konusunda yapılan çalışmalar gözden geçirilmiş, grup tespiti için Haung [70] ve arkadaşlarının Üçlü kapanış metodunu temel olarak geliştirilen dörtlü kapanış önerilmiştir. Dörtlü kapanış metodu grup tespiti yapan benzeri diğer metotlara ve özellikle temel aldığı üçlü kapanış'a göre (%2.3, $p \ll 001$) daha başarılı tahminler yapmıştır. Ancak bulduğu örnek sayısı üçlü kapanış'a göre daha düşüktür. Faktör değer analizine göre ağ topolojileri grup tespitinde en önemli faktör olurken kullanıcının aktif olduğu zaman dilimleri ikinci en önemli faktör olduğu tespit edilmiştir. Sosyal ağlarda etki değeri hesaplayan birçok yaklaşım analiz edilmiş ve Tang ve arkadaşlarının [93] önerdikleri Yerel Yakınlık Yayılımı (TAP) metodunun sosyal ağlar için uygulanabilirliği ve kapsamı açısından muadillerinden bir adım önde olduğu görülmüştür. Bölüm 2'de geliştirdiğimiz dörtlü kapanış metodunda bulduğumuz gruplarda etki analizi yapabilmesi için TAP'daki ağ topolojisi geliştirilmiş ve dörtlü kapanış için uygun hale getirilerek "Geliştirilmiş Yerel Yakınlık Yayılımı " olarak isimlendirilerek sunulmuştur. Çalışma sonucunda bu yaklaşımlar Instagram sosyal ağından toplanan veriler ile dağıtık bir mimari üzerinde test edilmiş, çıkan sonuçlar çalışmanın ilgili bölümünde detaylı olarak sunulmuştur.

Çalışmada kullanılan ağ topolojileri çalışmanın ana omurgasını oluşturup ve en yüksek faktör değerine sahip olsa da, sosyal ağlar gibi çok dinamik yapılarda kullanıcıları belli topolojilere sıkıştırmaya çalışmak uzun vadede başarı oranını düşürecektir. Topolojiye daha az bağımlı veya dinamik topolojiler ile çalışmanın desteklenmesi başarı oranlarını arttıracaktır. Çalışmamız ve benzer çalışmalarda zaman çizgileri bir veya iki tanedir. Buna karşın zaman çizgilerini çoğaltıp kullanıcıları daha yakından takip etmek özellikle sosyal ağı çok aktif kullanan kullanıcılar için önemlidir. İnsanların ilgi alanları hayatlarındaki dönüm noktalarına göre değişmektedir. Bir örnek vermek gerekirse, evlenen veya çocuğu olan bir birey takip ettiği yurtdışı seyahatleri veya yeni araç modelleri yerine artık çocuk bakımı ile ilgili konulara ilgi gösterecektir. Bir süre devam eden bu ilgi alanı çocuğun büyümesine paralel olarak çocuklu sosyal aktiviteler bulmaya kayacaktır. Ne bizim önerdiğimiz model ne de diğer modeller bireyin de ilgi alanının değişebileceğini varsayıp zaman çizgisini ona göre özelleştirmemektedir. Hatta bu tür değişiklikler birçok algoritmanın başarı oranını için olumsuz etki yapmaktadır.

KAYNAKLAR

- [1] **D. Chaffey**, 2015. "Global Social Media Research Summary 2015," *SmartInsights*. [Online]. Available: <http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>.
- [2] **M. a. Friedl ve C. E. Brodley**, 1997. "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ.*, vol. 61, no. 3, pp. 399–409,
- [3] **D. M. Magerman**, 1995. "Statistical Decision-Tree Models for Parsing," *Wall Str. J.*, pp. 276–283,
- [4] **L. Rokach ve O. Maimon**, 2005. "Decision Tree," *Data Min. Knowl. Discov. Handb.*, p. pp 165–192,
- [5] **T. M. Mitchell**, "Decision Tree Learning," *Machine Learning*. pp. 52–80, 1997.
- [6] **L. Backstrom**, 2006. "Group formation in large social networks: membership, growth, and evolution," in *In KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 44–54.
- [7] **E. M. Rogers**, 1995. *Diffusion of innovations*.
- [8] **D. Strang ve S. a. Soule**, 1998. "Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills," *Annu. Rev. Sociol.*, vol. 24, pp. 265–290,
- [9] **A. P. Bradley**, 1997. "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159,
- [10] **Y. Yue, T. Finley, F. Radlinski, ve T. Joachims**, 2007. "A support vector method for optimizing average precision," *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '07*, vol. Amsterdam, p. 271,
- [11] **D. P. Kroese**, 2010. "Cross-Entropy Method," in *Wiley Encyclopedia of Operations Research and Management Science*, pp. 1–12.
- [12] **J. H. Holland**, 1975. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, vol. 1, no. 1.
- [13] **D. E. Goldberg**, 1989. *Genetic Algorithms in Search, Optimization, and Machine*

Learning, vol. Addison-We..

- [14] **L. Costa ve P. Oliveira**, 2001. “Evolutionary algorithms approach to the solution of mixed integer non-linear programming problems,” *Comput. Chem. Eng.*, vol. 25, no. 2–3, pp. 257–266,
- [15] **L. B. Booker, D. E. Goldberg, ve J. H. Holland**, 1989. “Classifier systems and genetic algorithms,” *Artificial Intelligence*, vol. 40, no. 1–3. pp. 235 282,
- [16] **C. R. Houck ve M. G. Kay**, 2008. “A Genetic Algorithm for Function Optimization : A Matlab Implementation,” *Ncsuie Tr*, vol. 95, no. 919, pp. 1–14,
- [17] **M. Kaya ve R. Alhajj**, 2005. “Genetic algorithm based framework for mining fuzzy association rules,” *Fuzzy Sets Syst.*, vol. 152, pp. 587–601,
- [18] **a. Arenas, L. Danon, a. Díaz-Guilera, P. M. Gleiser, ve R. Guimerà**, 2004. “Community analysis in social networks,” *Eur. Phys. J. B*, vol. 38, no. 2, pp. 373–380,
- [19] **P. Holme, M. Huss, ve H. Jeong**, 2003. “Subnetwork hierarchies of biochemical pathways,” *Bioinformatics*, vol. 19, no. 4, pp. 532–538,
- [20] **K. A. Eriksen, I. Simonsen, S. Maslov, ve K. Sneppen**, 2003. “Modularity and extreme edges of the internet,” *Phys. Rev. Lett.*, vol. 90, no. 14, p. 148701,
- [21] **J.-P. Eckmann ve E. Moses**, 2002. “Curvature of co-links uncovers hidden thematic layers in the World Wide Web,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 9, pp. 5825–9,
- [22] **S. Wasserman ve K. Faust**, 1994. *Social network analysis: methods and applications*, vol. 24.
- [23] **R. Xu ve D. Wunsch**, 2005. “Survey of clustering algorithms,” *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645–678,
- [24] **M. Tasgin ve a Herdagdelen**, 2007. “Community detection in complex networks using genetic algorithms,” *Arxiv Prepr. arXiv0711.0491*, pp. 1–6,
- [25] **A. Clauset, M. E. J. Newman, ve C. Moore**, 2004. “Finding community structure in very large networks,” *Phys. Rev. E. Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics*, vol. 70, no. 6 Pt 2, p. 066111,
- [26] **J. Duch ve A. Arenas**, 2005. “Community detection in complex networks using extremal optimization,” *Phys. Rev. E*, vol. 72, no. 2, p. 027104,
- [27] **M. Newman ve M. Girvan**, 2004. “Finding and evaluating community structure in

- networks,” *Phys. Rev. E*, vol. 69, no. 2, pp. 1–16,
- [28] **C. Pizzuti**, 2008. “Community detection in social networks with genetic algorithms,” in *Proceedings of the 10th annual conference on Genetic and evolutionary computation - GECCO '08*, 2008, p. 1137.
 - [29] **G. Palla, I. Derényi, I. Farkas, ve T. Vicsek**, 2005. “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, no. 7043, pp. 814–8,
 - [30] **J. B. MacQueen**, 1967. “Kmeans Some Methods for classification and Analysis of Multivariate Observations,” *5th Berkeley Symp. Math. Stat. Probab.* 1967, vol. 1, no. 233, pp. 281–297,
 - [31] **B. W. Kernighan ve S. Lin**, 1970. “An Efficient Heuristic Procedure for Partitioning Graphs,” *Bell System Technical Journal*, vol. 49. pp. 291–307,
 - [32] **J. Pei, J. Han, Q. Chen, B. Mortazavi-Asl, M.-C. Hsu, Jianyong Wang, H. Pinto, ve U. Dayal**, 2004. “Mining sequential patterns by pattern- growth: the PrefixSpan approach,” *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1424–1440, 2004.
 - [33] **J. Han, H. Cheng, D. Xin, ve X. Yan**, 2007. “Frequent pattern mining: current status and future directions,” *Data Min. Knowl. Discov.*, vol. 15, pp. 55–86, 2007.
 - [34] **J. Han ve J. Pei**, 2000. “FreeSpan: frequent pattern-projected sequential pattern mining,” 200AD, pp. 355–359.
 - [35] **J. Han, J. Pei, Y. Yin, ve R. Mao**, 2004. “Mining Frequent Patterns without CandidateGeneration: A Frequent-Pattern Tree Approach,” in *Data Mining and Knowledge Discovery*, vol. 8, no. 1, 2004, pp. 53–87.
 - [36] **X. Yan ve J. Han**, 2002 . “gSpan: Graph-Based Substructure Pattern Mining,” in *Data Mining, 2002. ICDM 2003. Proceedings. ...*, 2002, pp. 721–724.
 - [37] **C. Borgelt**, 2012. “Frequent item set mining,” *Wiley Interdiscip. Rev. Data Min. Knowl Discov.*, vol. 2, no. 6, pp. 437–456, 2012.
 - [38] **R. Agrawal, S. Rajagopalan, R. Srikant, ve Y. Xu**, 2003. “Mining newsgroups using networks arising from social behavior,” *Proc. twelfth Int. Conf. World Wide Web - WWW '03*, p. 529, 2003.
 - [39] **S. Staab, P. Domingos, P. Mika, J. Golbeck, Li Ding, T. Finin, a. Joshi, a.**

- Nowak, ve R.R. Vallacher**, 2005. "Social Networks Applied," *IEEE Intell. Syst.*, vol. 20, pp. 80–93, 2005.
- [40] **P. Mika**, 2004. "Bootstrapping the FOAF-Web: An Experiment in Social Network Mining," *Friend a Friend Soc. Netw. Semant.*, pp. 1–10,
- [41] **W. Khan, F; Fisher, T; Shuler, L; Wu, T; Pottenger**, 2002. "Mining Chat-room Conversations for Social and Semantic Interactions," in *ech. Rep. LUCSE-02-011, Lehigh University*,
- [42] **M. Adnan, R. Alhajj, ve J. Rokne**, 2009. "Identifying Social Communities by Frequent Pattern Mining," *2009 13th Int. Conf. Inf. Vis.*, pp. 413–418,
- [43] **W. Zhang, T. Yoshida, ve X. Tang**, 2011. "A comparative study of TF*IDF, LSI and multi-words for text classification," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765,
- [44] **F. Meir ve C. Michael**, 1993. "The Segregation Matrix: A new sociometric index," *Henrietta Szold Inst.*, vol. 4, pp. 563–581,
- [45] **S. P. Borgatti, M. G. Everett, ve P. R. Shirey**, 1990. "LS sets, lambda sets and other cohesive subsets," *Soc. Networks*, vol. 12, no. 4, pp. 337–357, 1990.
- [46] **R. D. Alba**, 1973. "A graph- theoretic definition of a sociometric clique†," *J. Math. Sociol.*, vol. 3, no. 1, pp. 113–126.
- [47] **M. Fershtman**, 1997. "Cohesive group detection in a social network by the segregation matrix index," *Social Networks*, vol. 19, pp. 193–207,
- [48] **J. Li, L. Huang, T. Bai, ve Z. Wang**, 2012. "A novel high-quality community detection algorithm based on modified K-means clustering," *Int. J. Adv. Comput. Technol.*, vol. 4, no. 11, pp. 248–256,
- [49] **K. R.R, O. J Chen, ve Z. R**, 2010. "Top Leaders Community Detection Approach in Information Networks," 2010, pp. 1–9.
- [50] **X.-F. Wang ve Y.-B. Liu**, 2009. "Overview of algorithms for detecting community structure in complex networks," *J. Univ. Electron. Sci. Technol. China*, vol. 38, no. 5, pp. 537–543, 2009.
- [51] **W. W. Zachary**, 1977. "An Information Flow Model for Conflict and Fission in Small Groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, 1977.
- [52] **G. Simmel**, 1908. *Sociology: investigations on the forms of sociation*. Berlin Germany: Duncker & Humblot, 1908.
- [53] **A. L. J. Ter Wal**, 2014. "The dynamics of the inventor network in German

- biotechnology: geographic proximity versus triadic closure,” *J. Econ. Geogr.*, vol. 14, no. 3, pp. 589–620, 2014.
- [54] **T. Clark ve D. Iacobucci**, 2004. “Six Degrees: The Science of a Connected Age (Book),” *J. Mark.*, vol. 68, no. 1, pp. 166–167, 2004.
- [55] **J. a. Davis**, 1970. “Clustering ve Hierarchy in Interpersonal Relations: Testing Two Graph Theoretical Models on 742 Sociomatrices,” *Am. Sociol. Rev.*, vol. 35, no. 5, p. 843,
- [56] **D. Obstfeld**, 2005. “Social Networks, the Tertius Iungens Orientation, ve Involvement in Innovation,” *Adm. Sci. Q.*, vol. 50, no. 1, pp. 100–130,
- [57] **J. Skvoretz**, 1991. “Theoretical and methodological models of networks and relations,” *Soc. Networks*, vol. 13, no. 3, pp. 275–300,
- [58] **J. S. Coleman**, 1988. “Social Capital in the Creation of Human Capital,” *Am. J. Sociol.*, vol. 94, p. S95, 1988.
- [59] **B. Uzzi**, 1997. “Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness,” *Adm. Sci. Q.*, vol. 42, pp. 35–67,
- [60] **G. Walker, B. Kogut, ve W. Shan**, “Social Capital, Structural Holes and the Formation of an Industry Network,” *Organ. Sci.*, vol. 8, no. 2, pp. 109–125, 1997.
- [61] **V. Buskens**, 2002. *Social networks and trust*, vol. 2.
- [62] **R. Reagans ve B. McEvily**, 2003. “Network Structure and Knowledge Transfer: The Effects of Cohesion and Range,” *Adm. Sci. Q.*, vol. 48, no. 2, pp. 240–267,
- [63] **A. Zaheer ve G. G. Bell**, 2005. “Benefiting from network position: Firm capabilities, structural holes, and performance,” *Strateg. Manag. J.*, vol. 26, no. 9, pp. 809–825,
- [64] **L. Fleming, C. King, ve a. I. Juda**, 2007. “Small Worlds and Regional Innovation,” *Organ Sci.*, vol. 18, no. 6, pp. 938–954,
- [65] **G. Kossinets ve D. J. Watts**, 2006. “Empirical analysis of an evolving social network,” *Science*, vol. 311, no. 5757, pp. 88–90,
- [66] **T. Lou, J. I. E. Tang, J. Hopcroft, Z. Fang, ve X. Ding**, 2013. “Learning to predict reciprocity and triadic closure in social networks,” *ACM Trans. Knowl. Discov. Data*, vol. 7, no. 2, pp. 1–25,
- [67] **M. C. Grindroda, Peter Highamb, Desmond J Parsons**, 2012. “Bistability

- through Triadic Closure,” *Internet Math.*, vol. 8, no. 14, pp. 402–423,
- [68] **B. A. Huberman, D. M. Romero, ve F. Wu**, 2008. “Social networks that matter: Twitter under the microscope,” *Proc. Intl. Conf. Comput. Coop. Work*, pp. 107–114,
 - [69] **T. Opsahl**, 2013. “Triadic closure in two-mode networks: Redefining the global and local clustering coefficients,” *Soc. Networks*, vol. 35, no. 2, pp. 159–167,
 - [70] **H. Huang, J. Tang, S. Wu, L. Liu, ve X. Fu**, 2014 . “Mining Triadic Closure Patterns in Social Networks,” in *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, 2014, pp. 499–504.
 - [71] **J. . Hammersley ve P. Clifford**, 1971. “Markov field on finite hraps and lattices,” *Unpubl. Manuscr.*,
 - [72] **M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, ve I. H. Witten**, 2009. “The WEKA data mining software: an update,” *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18,
 - [73] **Wikipedia**, 2016. “Wikipedia - Instagram,” *Wikipedia*, 2016. [Online]. Available: <https://en.wikipedia.org/wiki/Instagram>. [Accessed: 10-May-2016].
 - [74] **Instagram**, 2016. “Instagram Stats.” [Online]. Available: <https://instagram.com/press/>.
 - [75] **D. J. Watts ve S. H. Strogatz**, 1998. “Collective dynamics of ‘small-world’ networks.,” *Nature*, vol. 393, no. 6684, pp. 440–2,
 - [76] **R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, ve U. Alon**, 2002. “Network motifs: simple building blocks of complex networks.,” *Science (80-.)*, vol. 298, no. 5594, pp. 824–827,
 - [77] **S. Bergsma, P. McNamee, M. Bagdouri, C. Fink, ve T. Wilson**, 2012. “Language Identification for Creating Language-Specific Twitter Collections,” “*Proceedings 2nd Work. Lang. Soc. Media NAACL-HLT’12*,” no. Lsm 2012, pp. 65–74,
 - [78] **P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, ve J.-P. Onnela**, 2010. “Community Structure in Time-Dependent, Multiscale, and Multiplex Networks,” *Science (80-.)*, vol. 328, no. 5980, pp. 876–878,
 - [79] **S. Asur ve B. A. Huberman**, 2010. “Predicting the Future with Social Media,” *Arxiv Prepr. arXiv10035699*, vol. abs/1003.5, pp. 492–499,

- [80] **B. Pang ve L. Lee**, 2006. “Opinion Mining and Sentiment Analysis,” *Found. Trends® Informatio*Pang, B., Lee, L. (2006). *Opin. Min. Sentim. Anal. Found. Trends® Inf. Retrieval*, 1(2), 91–231. doi10.1561/1500000001n Retr., vol. 1, no. 2, pp. 91–231,
- [81] **R. Kraut, M. Patterson, V. Lundmark, S. Kiesler, T. Mukopadhyay, ve W. Scherlis**, 1998.“Internet paradox: A Social Technology That Reduces Social Involvement and Psychological Well-Being?,” *Am. Psychol.*, vol. 53, no. 9, pp. 1017–1031,
- [82] **I. Brissette, M. F. Scheier, ve C. S. Carver**, 2002. “The role of optimism in social network development, coping, and psychological adjustment during a life transition,” *J. Pers. Soc. Psychol.*, vol. 82, no. 1, pp. 102–111,
- [83] **M. Cieliebak, O. Dierr, ve F. Uzdilli**, 2013 . “Potential and limitations of commercial sentiment detection tools,” in *CEUR Workshop Proceedings*, 2013, vol. 1096, pp. 47–58.
- [84] **H. J ve P. Clifford**, 1971. “Markov field on finite graphs and lattices,” *Unpubl. Manuscr.*,
- [85] **A. Anagnostopoulos, R. Kumar, and M. Mahdian**, 2008. “Influence and correlation in social networks,” *Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. KDD 08*, vol. 10, no. 1, p. 7,
- [86] **L. Backstrom, D. Huttenlocher, J. Kleinberg, ve X. Lan**, 2006. “Group formation in large social networks,” in *the 12th ACM SIGKDD international conference*, 2006, pp. 44–54.
- [87] **C. Marlow, M. Naaman, D. Boyd, ve M. Davis**, 2006. “HT06, tagging paper, taxonomy, Flickr, academic article, to read,” *Proc. seventeenth Conf. Hypertext hypermedia- HYPERTEXT '06*, p. 31,
- [88] **D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, ve S. Suri**, 2008. “Feedback effects between similarity and social influence in online communities,” *Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD08*, p. 160,
- [89] **C. D. Manning ve P. Raghavan**, 2009. “An Introduction to Information Retrieval,”
- [90] **J. Baumes, M. Goldberg, M. Hayvanovych, S. Kelley, M. Magdon-Ismail, K. Mertsalov, and W. Wallace**, 2007 . “SIGHTS: A Software System for Finding Coalitions and Leaders in a Social Network,” in *Intelligence and*

Security Informatics, 2007 IEEE, 2007, pp. 193–199.

- [91] **J. Baumes, M. Goldberg, M. Magdon-Ismail, ve W. Wallace**, 2005. “On hidden groups in communication networks,” *Trans.*, pp. 1–30, 2005.
- [92] **J. Baumes, M. Goldberg, M. Hayvanovych, M. Magdon-Ismail, W. Wallace, ve M. Zaki**, 2006. “Finding hidden group structure in a stream of communications,” *Lecture Notes in Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3975 LNCS. pp. 201–212,
- [93] **J. Tang, J. Sun, C. Wang, ve Z. Yang**, 2009. “Social influence analysis in large-scale networks,” *Proc. SIGKDD’09*, pp. 807–816,
- [94] **S. H. Strogatz**, 2001. “Exploring complex networks.,” *Nature*, vol. 410, no. 6825, pp. 268–276,
- [95] **M. E. J. Newman**, 2003. “The Structure ve Function of Complex Networks,” *SIAM Review*, vol. 45, no. 2. pp. 167–256,
- [96] **M. Faloutsos, P. Faloutsos, ve C. Faloutsos**, 1999. “On Power-Law Relationships of the Internet Topology,” in *In SIGCOMM*, 1999, pp. 251–262.
- [97] **R. Albert ve A. L. Barabasi**, 2002. “Statistical mechanics of complex networks,” *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.
- [98] **D. Krackhardt**, 1992. “The strength of strong ties: The importance of Philos in organizations,” *Networks Organ. Struct. Form, Action*, vol. 216, pp. 216–239,
- [99] **M. S. Granovetter**, 1973. “The Strength of Weak Ties,” *American Journal of Sociology*, vol. 78, no. 6. pp. 1360–1380,

ÖZGEÇMİŞ

Ahmet Anıl MÜNGEN

Elazığ - Türkiye

amungen@gmail.com

Kişisel Bilgiler:

Doğum Tarihi : 31.10.1990

Doğum Yeri : Elazığ

Eğitim

2010 - 2014 : Fırat Üniversitesi Bilgisayar Mühendisliği, Elazığ

2004 - 2007 : Yavuz Selim Anadolu Lisesi, Elazığ

İş Tecrübeleri

Kurum Adı : CNS AR-GE Bilişim, Türkiye (2015-Halen çalışıyorum)

Kurum Adı : Ernst & Young Türkiye (2014-2015)

TÜBİTAK Destekleri ve Bursları

2209/A Üniversite Öğrencileri Yurt İçi Araştırma Projeleri Destek Programı (2012)

2241/A Sanayi Odaklı Lisans Bitirme Tezi Destekleme Programı (2013)

2210 Yurt İçi Yüksek Lisans Burs Programı (2015)