



**T.C.
DÜZCE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

SOSYAL AĞLARDA ETKİ MAKSİMİZASYONU

AYBIKE ŞİMŞEK

**DOKTORA TEZİ
ELEKTRİK-ELEKTRONİK VE BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI**

**DANIŞMAN
PROF. DR. RESUL KARA**

DÜZCE, 2018

T.C.
DÜZCE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SOSYAL AĞLARDA ETKİ MAKSİMİZASYONU

Aybike ŞİMŞEK tarafından hazırlanan tez çalışması aşağıdaki jüri tarafından Düzce Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı'nda **DOKTORA TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Prof. Dr. Resul KARA

Düzce Üniversitesi

Jüri Üyeleri

Prof. Dr. Resul KARA

Düzce Üniversitesi

Dr. Öğr. Üyesi Sinan TOKLU

Düzce Üniversitesi

Dr. Öğr. Üyesi Filiz EVRAN ACAR

Düzce Üniversitesi

Dr. Öğr. Üyesi İbrahim Alper DOĞRU

Gazi Üniversitesi

Dr. Öğr. Üyesi Okan ERKAYMAZ

Bülent Ecevit Üniversitesi

Tez Savunma Tarihi: 25/07/2018

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

25 Temmuz 2018

Aybike ŞİMŞEK

TEŐEKKÜR

Doktora öğrenimimde ve bu tezin hazırlanmasında gösterdiği her türlü destek ve yardımdan dolayı çok değerli hocam Prof. Dr. Resul KARA'ya en içten dileklerle teşekkür ederim.

Bu çalışma boyunca yardımlarını ve desteklerini esirgemeyen sevgili eşim Mehmet ŐİMŐEK ve ođlum Davut Mete ŐİMŐEK'e, aileme ve çalışma arkadaşlarıma sonsuz teşekkürlerimi sunarım.

25 Temmuz 2018

Aybike ŐİMŐEK

İÇİNDEKİLER

	<u>Sayfa No</u>
ŞEKİL LİSTESİ	vii
ÇİZELGE LİSTESİ	viii
KISALTMALAR.....	ix
SİMGELER	x
ÖZET	xii
ABSTRACT	xiii
EXTENDED ABSTRACT	xiv
1. GİRİŞ.....	1
1.1. LİTERATÜR TARAMASI.....	1
1.2. TEZ ORGANİZASYONU	9
2. SOSYAL AĞLAR VE ONLINE SOSYAL AĞLAR	11
2.1. SOSYAL AĞIN TEKNİK BAKIŞ AÇISI	11
2.1.1. Ağ Çeşitleri	12
2.1.1.1. Erdős-Renyi Grafi.....	13
2.1.1.2. Ölçek Bağımsız Ağlar	15
2.1.1.3. Küçük Dünya Ağı	16
2.2. ONLINE SOSYAL AĞLARDA ETKİ MAKSİMİZASYONU	16
2.2.1. Ağ Ölçütleri	16
2.2.1.1. Düğüm Ölçütleri	18
2.2.1.2. Kenar Ölçütleri	22
2.2.2. Yayılım Modelleri	23
2.2.2.1. Independent Cascade (IC) Modeli.....	24
2.2.2.2. Linear Threshold (LT) Model	26
2.2.2.3. Polarity Related Independent Cascade (IC-P) Yayılım Modeli.....	28
2.2.2.4. Etki Yayılım Olasılığı Modelleri.....	30
3. ETKİ MAKSİMİZASYONU ve SÜRÜ ZEKÂSI	31

3.1. ETKİ MAKSİMİZASYONU PROBLEMİ	31
3.2. SÜRÜ ZEKÂSI OPTİMİZASYONU.....	33
3.2.1. Bozkurt Optimizasyon Algoritması (Grey Wolf Optimizer, GWO).....	33
3.2.2. Balina Optimizasyon Algoritması (Whale Optimization Algorithm, WOA).....	35
4. ÖNERİLEN YÖNTEM	39
4.1. PROBLEMİN MEVCUT DURUMU.....	39
4.2. SÜRÜ ZEKÂSI ALGORİTMALARINI YEREL OPTİMUMLARDAN KURTARMA	40
4.3. ÖNERİLEN YAKLAŞIM: EM PROBLEMİNDE AMAÇ FONKSİYONUNUN DURUM UZAYINA GENEL BİR EĞİM VERMEK	41
4.3.1. Merkezilik Ölçütlerinden Karma Bir Ölçüt Elde Etme.....	43
4.3.2. İki Çekirdek Düğüm İçin Etki Maksimizasyonu Problemi	45
5. DENEYSEL SONUÇLAR.....	47
5.1. KIYASLAMADA KULLANILAN ALGORİTMALAR.....	48
5.1.1. PageRank.....	48
5.1.2. Kempe'nin Açgözlü Algoritması	49
5.2. FACEBOOK GRAFI ÜZERİNDE YAPILAN GWO VE RASGELE SEÇİM DENEMELERİ	50
5.3. GWO, WOA, PAGERANK VE KEMPE'NİN AÇGÖZLÜ ALGORİTMALARI İÇİN FACEBOOK GRAFI ÜZERİNDE YAPILAN DENEMELER.....	55
5.4. EPINIONS GRAFI ÜZERİNDE YAPILAN GWO VE RASGELE SEÇİM KARŞILAŞTIRMASI	56
6. SONUÇLAR VE DEĞERLENDİRMELER	58
6.1. SONUÇLAR.....	58
6.2. ÇALIŞMANIN GETİRDİĞİ KATKILAR	59
6.3. TARTIŞMA VE ÖNERİLER.....	59
7. KAYNAKLAR.....	61
ÖZGEÇMİŞ.....	67

ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 2.1. Partideki konuklar arasındaki ilişkiler [44].	13
Şekil 2.2. Rasgele Ağlar [44].	14
Şekil 2.3. Yakınlık merkeziliği örneği.	21
Şekil 2.4. Independent Cascade Modeli'nin sözde kodu [48].	25
Şekil 2.5. ICM (Independent Cascade Model) simülasyonu [48].	26
Şekil 2.6. Linear Threshold Model'in sözde kodu [48].	27
Şekil 2.7. LTM simülasyonu [21].	28
Şekil 3.1. Örnek sosyal graf.	33
Şekil 3.2. GWO'nun sözde kodu [7].	35
Şekil 3.3. WOA'nın sözde kodu [8].	38
Şekil 4.1. Erdős-Renyi Grafi üzerinde örnek EM problemi– Normal Graf.	40
Şekil 4.2. Yeniden şekillendirilmiş Erdős-Renyi grafi üzerinde Etki Maksimizasyonu-Çıkış Derecesi Merkeziliği.	42
Şekil 4.3. Yeniden şekillendirilmiş Erdős-Renyi grafi üzerinde Etki Maksimizasyonu – Yakınlık Merkeziliği.	42
Şekil 4.4. Yeniden şekillendirilmiş Erdős-Renyi grafi üzerinde Etki Maksimizasyonu – Birleştirilmiş Ölçüt.	43
Şekil 4.5. Eğim oluşturulan Facebook grafi. (a) Normal Facebook grafi, (b) Çıkış Derecesi Merkeziliği ölçütüne göre yeniden numaralandırılan graf, (c) Yakınlık Merkeziliği ölçütüne göre yeniden numaralandırılan graf, (d) Birleştirilmiş ölçüt ölçütüne göre yeniden numaralandırılmış graf	454
Şekil 5.1. Facebook ağının görüntüsü.	48
Şekil 5.2. PageRank için örnek graf.	49
Şekil 5.3. Farklı sayıdaki çekirdekler için Facebook grafi üzerinde GWO ve rasgele seçim metotlarının deneysel sonuçları: (a) k=1, (b) k=2, (c) k=5, (d) k=10, (e) k=15, (f) k=20, ve (g) k=30.	52
Şekil 5.4. Farklı sayıdaki çekirdekler için Facebook grafi üzerinde GWO, WOA, PageRank ve Kempe ve arkadaşlarının Aç Gözlü Algoritması metotlarının deneysel sonuçları: (a) k=15, (b) k=20, ve (c) k=30.	55
Şekil 5.5. Epinions grafi üzerinde GWO ve rasgele seçim karşılaştırması (k=15, k=20 ve k=30 değerleri için).	57

ÇİZELGE LİSTESİ

	<u>Sayfa No</u>
Çizelge 1.1. İncelenen çalışmaların özeti.	10
Çizelge 5.1. GWO için test istatistikleri.	54
Çizelge 5.2. Rastgele seçim metodu için test istatistikleri.....	54



KISALTMALAR

ABC	Artificial Bee Colony
ALO	Ant Lion Optimizer
BOBYQA	Bound Optimization BY Quadratic Approximation
df	Damping Factor
DP	Dinamik Programlama
EM	Etki Maksimizasyonu
GWO	Grey Wolf Optimizer
ICM	Independent Cascade Model
IC-P	Polarity Related Independent Cascade
IM	Influence Maximization
INT	Influential Node Tracking
LCIP	Least Cost Influence Problem
LIM	Linear Influence Model
LTM	Linear Threshold Model
MLE	Maximum Likelihood Estimation
PPC	Partial Parallel Cascade
PrPaS	Practical Partitioning and Seeding
SASP	Seed Activation Scheduling Problem
SNAP	Stanford Network Analysis Project
UBI	Upper Bound Interchange
US	Uniform Setting
WCS	Weighted Cascade Setting
WOA	Whale Optimization Algorithm
WTSS	Weighted Target Set Selection

SİMGELER

α	Alfa ajanlarının kümesi
β	Beta ajanlarının kümesi
Γ	Sabit
δ	Gama ajanlarının kümesi
θ	LT modelindeki eşik değeri
λ	Sabit
μ	Merkezilik ölçütü
ω	Omega ajanlarının kümesi
A	Komşuluk matrisi
\vec{A}	GWO ve WOA'daki katsayı vektörü
\vec{a}	GWO ve WOA'daki sabit vektörü
b	LT modelinde İki düğüm arasındaki ağırlık
C	Kümeleme katsayısı
C_e	Özvektör merkeziliği
C_{outdeg}	Çıkış derecesi merkeziliği
C_C	Yakınlık merkeziliği
\vec{C}	GWO ve WOA'daki katsayı vektörü
C^{BET}	Aradalık merkeziliği
C^{BON}	Bonacich merkeziliği
C^{DEG}	Derece merkeziliği
C^{KATZ}	Katz merkeziliği
C^{max}	Maksimum dereceye göre normalize edilmiş derece
C^{norm}	Normalize edilmiş derece
C^{toplam}	Derece toplamına göre normalize edilmiş derece
\vec{D}	GWO ve WOA'daki pozisyon vektörü
deg	Bir düğümün derecesi
$deg^{çıkış}$	Bir düğümün çıkış derecesi
$deg^{giriş}$	Bir düğümün giriş derecesi
df	PageRank algoritmasındaki damping factor
E	Kenarlar kümesi
$f_1(\cdot)$	Önerilen yöntemde maksimize edilecek amaç fonksiyonu
$f_2(\cdot)$	Önerilen yöntemde minimize edilecek amaç fonksiyonu
G	Graf
H	Hipotez
k	Etkileyici düğüm olarak seçilen çekirdek düğümlerin sayısı
L	Kenar sayısı
l	Düğümün diğer düğümlere olan ortalama en kısa yol uzunluğu
M	Merkezilik ölçütlerinin kümesi
N	Düğüm sayısı
n	Bir düğümün komşularının kümesi

N_{aktif}^t	t zamanında aktif olan komşu düğümlerin kümesi
n_{Δ}	Ağdaki üçgenlerin sayısı
p	Olasılık
\vec{r}	GWO ve WOA'daki rasgele sayı vektörü
S	Aktif düğümlerin kümesi
$S(\cdot)$	Bağ kuvveti
t	Zaman
u	Keyfi bir düğüm
v	Keyfi bir düğüm
w	Keyfi bir düğüm
\vec{X}	GWO ve WOA'daki pozisyon vektörü
Z	Z skoru



ÖZET

SOSYAL AĞLARDA ETKİ MAKSİMİZASYONU

Aybike ŞİMŞEK

Düzce Üniversitesi

Fen Bilimleri Enstitüsü, Elektrik-Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Prof. Dr. Resul KARA

Temmuz 2018, 66 sayfa

İnsanlar fikirlerini yaymak, yenilikler hakkında bilgi edinebilmek için çevrimiçi sosyal ağları kullanırlar. Bu bağlamda, bilginin sosyal ağlar üzerinde nasıl yayıldığını bilmek önemlidir. Bir sosyal ağ aracılığıyla daha fazla sayıda kişiye bir bilgiyi (örneğin ürün reklamı) yaymak mümkündür. Burada kilit nokta, sosyal ağdaki etkili bireyleri tespit edebilmektir. Bu problem Etki Maksimizasyonu (EM) problemi olarak adlandırılmıştır. EM problemi, belirli bir birey grubunu etkileyen, az sayıdaki çekirdek bireyi bulmaya odaklanır. Literatürde, EM problemi için ağgözlü algoritmalar, stokastik ve evrimsel optimizasyon algoritmaları önerilmiştir. Ancak, bu yöntemler hız veya çözüm kalitesi açısından istenen seviyede değildir. Öte yandan, literatürde hızlı ve optimal çözümler üreten birçok sürü zekâsı algoritması bulunsa da, bu algoritmalar EM problemine doğrudan uygulanamazlar. Bu tezde, sürü zekâsı algoritmalarının EM problemine uygulanabilmesini sağlamak için problemin yapısı değiştirilmiştir. Bir sosyal ağdaki bireyleri düğümler olarak ele alırsak; yapılan işlem, bazı ağ ölçütlerine göre düğümlerin büyükten küçüğe doğru sıralanması ve bu sıraya göre yeniden numaralandırılması olarak tarif edilebilir. Önerilen yaklaşım işaretli ve işaretsiz gerçek graflar ve sentetik graflarla test edilmiştir. Denemelerde, sürü zekâsı algoritmaları olarak Bozkurt Optimizasyonu (Gray Wolf Optimizer, GWO) ve Balina Optimizasyon Algoritması (Whale Optimization Algorithm, WOA); kıyaslama yöntemleri olarak PageRank ve Kempe'nin Ağgözlü Algoritması kullanılmıştır. Deneysel sonuçlar önerilen yaklaşımın iyi çalıştığını göstermiştir.

Anahtar sözcükler: Etki maksimizasyonu, Sosyal ağlar, Sürü zekâsı algoritmaları

ABSTRACT

INFLUENCE MAXIMIZATION IN SOCIAL NETWORKS

Aybike ŞİMŞEK

Düzce University

Graduate School of Natural and Applied Sciences, Department of Electrical-Electronics
and Computer Engineering

Doctoral Thesis

Supervisor: Prof. Dr. Resul KARA

July 2018, 66 pages

People use online social networks to spread ideas, learn about innovations, etc. In this context, it is important to know how information spreads through social networks. It is possible to spread information (e.g., product advertisement) to a larger number of individuals via a social network. The key point is to identify the most influential individuals on the social network. This problem is named as Influence Maximization (IM) problem. The IM problem focuses on finding the small subset of individuals in a social environment who influence a certain group of individuals. In the literature, greedy, stochastic, and evolutionary optimization algorithms have been proposed to solve this problem. However, these methods are not at the desired level in terms of speed or solution quality. On the other hand, although many Swarm Intelligence (SI) algorithms can be found in the literature, these algorithms cannot be directly applied to the IM problem. In this thesis, a change in the structure of the IM problem is suggested in order to tailor it to SI algorithms. If a social network is envisioned as a graph and individuals as nodes, the proposed method means sorting the nodes in descending order according to some graph metrics and renumbering the nodes according to this order. The proposed approach was tested with signed and unsigned real and synthetic graphs. The experiments employed Grey Wolf Optimizer (GWO) and Whale Optimization Algorithm (WOA) SI algorithms and PageRank and Kempe et. al.'s Greedy Algorithm as benchmark methods. Experimental results showed that this approach worked well.

Keywords: Influence Maximization, Social Networks, Swarm Intelligence Algorithms

EXTENDED ABSTRACT

INFLUENCE MAXIMIZATION IN SOCIAL NETWORKS

Aybike ŞİMŞEK

Düzce University

Graduate School of Natural and Applied Sciences, Department of Electrical-Electronics
and Computer Engineering

Doctoral Thesis

Supervisor: Prof. Dr. Resul KARA

July 2018, 66 pages

1. INTRODUCTION

A real social network is of great benefit for information dissemination via communication among its members. This opens the door to applications like viral marketing. However, information dissemination on social networks can also be used for malicious purposes, including the spreading of ‘false news’ or virus-containing messages among people in a short amount of time. Some individuals in a social network are more influential than others for such reasons as social status, charisma, etc. If these individuals can be identified, the spreading of information on a network can be maximized, or the source of diffused content identified. In this respect, detection of influential individuals is the basis of Influence Maximization (IM) problems. In order to solve this problem, various metrics have been used with greedy algorithms. In addition, stochastic and evolutionary methods have been applied to the IM problem. Although, there are many swarm intelligence algorithms in the literature, their application to this problem is very limited. The underlying reason is that swarm intelligence algorithms cannot be directly applied to combinatorial optimization problems. In this thesis, the IM problem was reshaped to enable swarm intelligence algorithms to be applied to these types of problems.

2. MATERIAL AND METHODS

The Facebook and Epinions networks were used in the experiments and were downloaded from the Standard Large Network Dataset Collection

(<http://snap.stanford.edu>). Epinions graph, unlike Facebook graph, also contains information about whether individuals influence each other positively or negatively. Hence, such graphs are referred to as signed graphs and the IM problem is treated as a signed IM problem. In the signed IM problem, it is aimed to maximize the positive effect. The Independent Cascade (IC) model was used to model the influence propagation. In the IC model, there is an influence probability of between 0 and 1, $P(e) = [0,1]$ which is associated with an edge, $e = (u, v)$. Here, u and v are arbitrary neighbor nodes.

Swarm intelligence algorithms assume that the next good solution is around the current state of the state-space and follow the slope of the surface. However, the situation is different in IM problems. What is then needed is to bring the solutions which are similar in quality closer on the surface. In the IM problem, nodes create the state-space surface. Therefore, if the nodes are sorted according to their impacts and renumbered according to this order, a slope is obtained on the state-space surface. We named this process as “reshaping”. More precisely, if a social network is envisioned as a graph and individuals as nodes, reshaping means sorting the nodes in descending order according to some graph metrics and renumbering the nodes according to this order. Thus, the nodes those are close to each other in terms of level of influence become closer to each other in the state-space. This creates a general slope on the state-space surface of the objective function. This simple idea paves the way for applying all swarm intelligence algorithms to this kind of problem.

3. RESULTS AND DISCUSSIONS

Experiments were carried out to see how the reshaping worked. Seed selection was performed both on the original graph and on the modified graphs. In order to determine whether this approach had a significant effect on the reshaped graphs, random seed selection was used as a benchmark method on both the original graph and the modified graphs. In each experiment, the GWO algorithm and the random selection process were run 30 times. There were 50 iterations in each run. Eventually, the best results for the GWO and random selection were found. According to the experimental results, there was no significant difference between the GWO and the random selection for $k = 1$ and $k = 2$. In such a small graph, the problem was relatively simple for a low seed count. In the other experiments, the difference between the GWO and the random selection performance was clear. Especially when the situations for $k = 15$, $k = 20$, $k = 30$ are

examined, it can be seen that the reshaping process worked well. We also compared gives the performances of the GWO, WOA, PageRank and Kempe et. al.'s Greedy Algorithm on the IM problem. Experiments were carried out for $k = 15$, $k = 20$, and $k = 30$ because the performance of the proposed approach had been determined for $k \geq 15$. In the experiments, the reshaped graph was used only according to Closeness Centrality for the evaluation. The GWO and WOA outperformed the PageRank algorithm for all k values, and GWO and WOA gave slightly better results than Kempe et. al.'s Greedy Algorithm for $k = 15$ and $k = 20$. Kempe et. al.'s Greedy Algorithm gave better results for $k = 30$. The performances of the GWO and WOA were close. At this point, we need to specify that GWO and WOA have given competitive result with Kempe et al.'s Greedy Algorithm. Swarm intelligence algorithms are promising for use in the IM problem.

4. CONCLUSION AND OUTLOOK

The main contribution of this study is that it paves the way for applying all swarm intelligence algorithms to IM problems, which are combinatorial optimization problems. In addition, there are some issues on this topic which are open for future research:

It is important to decide the metrics and coefficients for a more healthy reshaping process.

There are many swarm intelligence algorithms in the literature. An important point here is the way different algorithms will yield results for networks where different relationships such as a friendship, a following, or a trust are modeled. Different algorithms may be more appropriate for different social relationships. The basic approach in this study was to adapt the problem to the algorithms; however, the algorithms may need to be modified as well.

Many problems in very different types of networks can be handled as IM problems. For example, detection of the accounts which should be immunized in a network of email contacts in order to minimize the spread of computer viruses. Similarly, detection of the patient who is the origin of the spread of an infection in a contact network of inpatients is an IM problem. Because different network types have their own features, the reshaping process in these networks may vary.

There are many information propagation models. In this context, it is necessary to reconsider the IM problem for different propagation models.

1. GİRİŞ

Gerçek bir sosyal ağda bilginin yayılımı, üyelerin arasındaki iletişim sayesinde çok kolaylıkla yapılabilmektedir. Gerçek sosyal ağları internete aktaran Facebook ve Twitter gibi çevrimiçi sosyal ağlar için de durum aynıdır. Bu çevrimiçi platformların kullanıcıları arasında bilgi çok hızlı bir şekilde yayılabilir. Bu, sözle yapılan pazarlama (viral marketing) gibi uygulamalara kapıları açar. Bununla birlikte, sosyal ağlarda bilgi yayılımı, “yanlış haberlerin” veya virüs içeren mesajların kısa bir süre içinde insanlar arasında yayılması dahil olmak üzere, kötü amaçlı durumlar için de kullanılabilir. Sosyal ağdaki etkili bireyler bilgi yayılmasını artırmak veya dedikodu gibi istenmeyen içeriklerin yayılmasını azaltmak için kullanılabilir [1].

Sosyal bir ağdaki bazı bireyler, sosyal statü, karizma, vb. nedenlerden ötürü diğerlerinden daha etkilidir. Eğer bu bireyler tanımlanabilirse, bilginin bir ağ üzerinde yayılması en üst düzeye çıkarılabilir veya dağıtık içeriklerin kaynağı tespit edilebilir. Bu bağlamda, etkili bireylerin tespiti hem etkileyicinin tespiti hem de etki maksimizasyonu (EM) problemlerinin temelidir. Son yıllarda, EM sorunu yaygın olarak çalışılmış olmasına rağmen etkileyici tespit problemi daha az dikkat çekmiştir. Esasen, EM problemi bir çeşit etkileyici tespit problemidir [2]. EM'de amaç, maksimum sayıda bireyi etkilemek için hangi bireylerin bir tohum olarak seçileceğini belirlemektir. Öte yandan, etkileyici tespiti probleminin amacı, belirli bir aktivasyon durumuna ulaşmak için hangi bireylerin bir tohum olarak seçilmesi gerektiğini belirlemektir[2]. Sonuç olarak, küçük farklılıklar olsa da, her iki problem de etkili bireylerin tespitine dayalı optimizasyon problemleri olarak düşünülebilir. Kempe, “etki maksimizasyon probleminin, maksimum etkiyi gösteren bir k-düğüm seti bulmasını ister” [3]. Bu tanıma dayanarak, etkili bireylerin saptanması problemi, çalışmanın geri kalanında EM olarak adlandırılacaktır.

1.1. LİTERATÜR TARAMASI

EM problemi bir NP-hard kombinatoriyal optimizasyon problemidir [4]. Bu nedenle bu problemi çözebilmek için açgözlü algoritmalar ile beraber çeşitli ölçütler kullanılmıştır.

Bunun yanında EM problemlerine stokastik ve evrimsel metotlar da uygulanmıştır [5]. Literatürde Yapay Arı Kolonisi (Artificial Bee Colony, ABC), Bozkurt Optimizasyonu (Gray Wolf Optimizer, GWO), Balina Optimizasyon Algoritması (Whale Optimization Algorithm, WOA) ve Aslan Karıncası Optimizasyonu (Ant Lion Optimizer, ALO) [6]–[9], gibi birçok sürü zekâsı algoritması olsa da bunların EM problemine uygulanması yeni bir yaklaşımdır [10], [11]. Bu durumun temel nedeni sürü zekâsı algoritmalarının kombinatoriyal optimizasyon problemlerine doğrudan uygulanamamasıdır. EM gibi problemlerde, amaç fonksiyonunun ürettiği durum uzay yüzeyi genel bir eğime sahip değildir. Bu nedenle, sürü zekâsı algoritmaları global yüzeyde optimal bir çözüme yakınsayamazlar. EM gibi kombinatoriyel problemlere sürü zekâsı algoritmalarını uygulamak bu tip algoritmaların avantajlarından faydalanmak anlamına gelecektir. Bu amaçla bu tezde, EM problemi, sürü zekâsı algoritmalarının bu tip problemlere uygulanabilmesi için yeniden şekillendirilmiştir. Yeniden şekillendirmenin anlamı, ele alınan ölçütlere göre düğümlerin azalan sırada yeniden sıralanması ve düğümlerin yeni sıralarına göre yeniden numaralandırılmasıdır. Bu yolla sosyal ağı yeniden şekillendirerek durum uzay yüzeyinde bir eğim oluşturulmuş ve sürü zekâsı algoritmaları EM problemlerine uygulanabilir hale getirilmiştir.

Bu çalışmada öncelikle yeniden şekillendirilmiş ve şekillendirilmemiş graflar üzerinde yaklaşımın nasıl çalıştığını göstermek için GWO ve rasgele seçim metotları karşılaştırılmıştır. Daha sonra, GWO, WOA ve PageRank algoritmaları karşılaştırılmıştır. Orjinal GWO ve WOA metotları sürekli optimizasyon problemleri için çözüm ürettiklerinden bu algoritmalar üzerinde küçük değişiklikler yapılmış ve ayrık (aynı zamanda kombinatoriyal) optimizasyon problemlerine uygulanabilir duruma getirilmişlerdir. EM problemlerinde bu yaklaşımın verimliliği Facebook ve Epinions grafları üzerinde gösterilmiştir [12],[13].

EM problemi yalnızca sosyal ağların bir problemi değildir. Çok farklı tip ağlardaki benzer problemler EM gibi ele alınabilir. Örneğin, bir e-posta ağında bilgisayar virüslerinin yayılımını engellemek için hangi hesapların karantinaya alınması gerektiğinin saptanması veya enfeksiyonun hangi hastadan yayıldığının tespit edilmesi de birer EM problemidir. Bu bakış açısıyla sürü zekâsı algoritmalarının EM problemlerini çözme kapasitesine sahip olması önemlidir.

Sosyal ağ araştırmalarında etki maksimizasyonu çok önemli bir konudur. EM problemi, en etkili bireylerin tespit edilmesine dayanmaktadır. Bu nedenle, literatürdeki çalışmalar

etkili bireyleri işaret eden ölçütlerin geliştirilmesine ve bu ölçütleri kullanan algoritmaların geliştirilmesine odaklanmıştır [2], [14]. Sosyal etki için bazı temel değerlendirme ölçütleri; Derece, Yakınlık, Özvektör, Katz ve Aradalık Merkeziliklerini içerir [14].

Derece Merkeziliği bir düğümün bitişiğindeki kenarların sayısı olarak hesaplanır [14], [15]. Yönlü graflar iki çeşit dereceye sahiptir: giriş derecesi ve çıkış derecesi. Giriş derecesi, bir düğüme diğer düğümlerden gelen kenarların sayısıdır. Çıkış derecesi, bir düğümden diğer düğümlere giden kenarların sayısıdır. Bir düğümün derecesi ne kadar yüksek olursa, diğer düğümlere o kadar bağlıdır demektir. Bunun anlamı da genel olarak, bu düğümün daha fazla düğümü etkileyebileceğidir.

Yakınlık Merkeziliği, diğer tüm düğümlerden bir düğüme olan ortalama en kısa yol olarak hesaplanır [16], [17]. Yakınlık Merkezinin daha yüksek değerleri, bir düğümün daha yüksek etkiye sahip olduğunu gösterir.

Aradakalmışlık merkeziliği belli bir düğümden geçen iki düğüm arasındaki en kısa yolların yüzdesi olarak hesaplanır [14], [18]. Bilginin en kısa yollarla yayıldığını varsayarsak, daha yüksek aradakalmışlık değerine sahip düğümlerin bilgi yayılımı için köprü görevi gördüğü söylenebilir.

Katz Merkeziliği, belirli bir düğümün komşularının derecelerinin toplamı olarak düşünülebilir [14], [18]. Katz Merkeziliğinin arkasındaki fikir, eğer bir düğümün etkili komşuları varsa, bu düğüm de etkilidir [19].

Özvektör Merkeziliği, temel düşünce ve hesaplama açısından Katz Merkeziliğine benzerdir [14]. Eğer graf güçlü bir şekilde birbirine bağlıysa iyi çalışır. Ancak, fiili yönlü grafların büyük bir bağlı bileşeni yoktur ve bu da pratikte zorluklara yol açmaktadır [19].

Yukarıda açıklanan ölçütlerin dışında, sosyal ağların da belirli metrikleri vardır. Örneğin, bir Twitter kullanıcısının önemini belirlemek için Doğrulanmış Hesap, Üyeyi İçeren Abonelik Listeleri ve Profil Yaşı gibi özellikler kullanılabilir [20]. Bu çalışmada graf tabanlı ölçütler kullanıldığından, kullanıcı tabanlı (nodal) ölçütler dâhil edilmemiştir. Nodal sosyal etki ölçütleri hakkında [14] ve [20]'da detaylı bilgi verilmiştir.

Ancak, EM problemi basitçe belirli bir ölçütün en yüksek değerlerine sahip bireyleri seçilerek çözülemez [21]. Bu nedenle EM problemini çözmek için çeşitli yaklaşımlar

geliştirilmiştir.

Sosyal ağlarda EM problemiyle ilgili yapılan ilk çalışmalardan biri Kempe ve arkadaşlarına aittir [3], [22]. EM probleminin NP-hard, ayrık bir optimizasyon problemi olduğunu göstermişler ve genel bir açgözlü algoritma önermişlerdir. Yaklaşımları bire bir tohumlama stratejisini benimser. Yani, çekirdek adaylarını birer birer değerlendirir ve çekirdek setine ekler. Açgözlü yaklaşımlarını Derece ve Mesafe Merkeziliği sezgisel yöntemleriyle karşılaştırmışlar ve algoritmalarının daha başarılı olduğunu görmüşlerdir.

İlk çalışmalardan bir diğeri Borgatti tarafından gerçekleştirilmiştir [21]. EM problemini anahtar oyuncuları tanımlama problemi olarak tanımladı ve problemi iki perspektiften değerlendirdi: (1) bilginin maksimum yayılmasını sağlamak için anahtar oyuncuların belirlenmesi ve (2) ağı en fazla sayıda alt ağa bölmek için anahtar oyuncuların ağdan çıkarılması. Her iki problem için de basitçe en iyi k oyuncuları seçmek işe yaramadı. Bu nedenle, Borgatti kombinatorial optimizasyon problem olan k -çekirdek dizisini eş zamanlı olarak seçme işlemini ele aldı. Kempe'nin bire bir çekirdekleme yapan açgözlü yaklaşımı ve Borgatti'nin kombinatorial optimizasyon yaklaşımı takip eden çalışmaları büyük ölçüde şekillendirdi. Literatürdeki bazı çalışmalar açgözlü yaklaşımı benimserken, diğerleri problemi kombinatorial optimizasyon problemi olarak ele almıştır.

Raghavan ve Zhang EM problemine farklı bir perspektiften yaklaşmışlar ve bir kombinatorial optimizasyon problemi olan Ağırlıklı Hedef Set Seçimi (Weighted Target Set Selection, WTSS) olarak ele almışlardır [23]. WTSS problemi için ağaçlar üzerinde polinom-zamanda dinamik programlama algoritması geliştirmişlerdir. Daha sonra rasgele grafikler için yaklaşımlarını genelleştirmişlerdir. Benzer bir çalışmada Günneç, Raghavan ve Zhang, En Az Maliyetli Etki Problemini (Least Cost Influence Problem, LCIP) ele almışlardır [24]. Ana motivasyonları, bu tür problemler için matematiksel programlama yaklaşımlarını geliştirmektir. Ağaçlarda LCIP için açgözlü bir algoritma geliştirmişlerdir. Ek olarak, bir ağacı birkaç yıldız grafına ayıran daha iyi bir dinamik programlama (DP) algoritması geliştirmişlerdir.

Lappas ve ark., EM probleminden biraz farklı bir problem ortaya çıkarmış ve k -Effector problemi olarak adlandırmışlardır [4]. K -Effector probleminde, birtakım düğümler bazı tohum düğümleri tarafından başlangıçta aktif hale getirilmiş ve gözlemlenen aktivasyon durumunu en iyi açıklayan efektör düğümlerini bulmayı amaçlamıştır. Rasgele graflar

için problemin optimal çözümünün NP-hard olduğunu ve yaklaşık olarak çözmek için de NP-hard olduğunu gösterdiler. Yaklaşımları iki aşamadan oluşmaktadır: (1) belirli bir ağ ve aktivasyon durumu için bir etki ağacı oluşturma ve (2) ağaç üzerinde en uygun efektörleri seçmek için bir dinamik programlama algoritması kullanma. Algoritmalarının k-Effector problemi için iyi performans sağladığını göstermişlerdir.

Lappas ve ark.'nın k-Effector problemini temel alarak, Tong ve ark., problemi iki perspektiften ele almıştır [2]. İlk olarak, etki mesafesini dikkate almışlar ve etki mesafesi ölçütüne dayalı bir efektör tespit çerçevesi tasarlamışlardır. Daha sonra efektör tespiti için Maksimum Benzerlik Tahminini (Maximum Likelihood Estimation, MLE) kullanmışlardır. Efektör saptama problemini bir optimizasyon problemi olarak formüle etmişler ve 3. Derece yaklaşık çözüm bulan bir algoritma önermişlerdir.

Esasen, sosyal ağlar dinamik yapıya sahiptir ve bu yapı zaman içinde gelişir. Bu bağlamda Song ve ark., etkileyici düğümlerin dinamik bir ağ yapısında izlenmesi gerektiğini iddia etmişler ve geleneksel EM probleminin bir uzantısı olarak Etkili Düğümleri İzleme (Influential Node Tracking, INT) problemini önermişlerdir [25]. EM, statik bir ağda ortak etkiyi en üst düzeye çıkarmak için kullanılsa da, ağ gelişmeye devam ettikçe, INT en yüksek düzeyde etki sahibi olan bir dizi etkileyici düğümünü korumayı amaçlamaktadır. INT problemlerini çözmek için Upper Bound Interchange Greedy (UBI) ve onun varyantı olan UBI + adlı bir açgözlü algoritma geliştirmişlerdir. Başlangıçtan itibaren bir tohum seti oluşturmak yerine, UBI daha önce bulunmuş etkileyici düğümleri kullanır ve toplam etkiyi artırmak için düğümleri değiştirir. Üç gerçek ölçekli dinamik sosyal ağ üzerinde çalıştıklarını, kapsama ve performans açısından daha iyi sonuçlar aldıklarını vurgulamışlardır.

Literatürdeki çalışmaların çoğu, zamandan bağımsız olan IC ve LT yayılım modellerine dayanmaktadır. Yani, bu modellerde, çekirdek seçiminin zamanlaması önemli değildir. Samadi ve ark., birçok problemin zamandan bağımsız olmadığını belirtmişler ve çekirdek düğümlerinin seçimini zamana-bağımlı bir problem olarak ele almışlardır [22], [26]. Uygulanan probleme Çekirdek Aktivasyon Zamanlama Problemi (Seed Activation Scheduling Problem, SASP) denir ve belirli bir bütçe altında çekirdek aktivasyonunun zamanlamasını seçer. Yazarlar, SASP'nin Partial Parallel Cascade (PPC) yayılım modeli altında NP-hard olduğunu göstermişler ve çözümü için karma bir tamsayı programlama yöntemi önermişlerdir. Bu özelliklerinden dolayı SASP, EM problemine yeni bir boyut kazandırmıştır.

Liu ve ark., bir kullanıcıdan diğerine olan etkinin yayılmasında zamanın önemine dikkat çekmiş ve zaman sınırlı etki maksimizasyon problemini ele almışlardır [27]. Geliştirdikleri ağgözlü algoritmanın ölçeklenebilirliğini artırmak için “sosyal ağlarda yayılma yolu” kavramını önermişler ve zaman sınırlı EM problemi için bir dizi yeni algoritma geliştirmişlerdir.

Bazı çalışmalar EM problemine oyun teorisi bağlamında yaklaşmıştır. Bu alandaki temel çalışmalardan ikisi, İrfan ve Ortiz'in ve Molinero ve ark.'nın çalışmalarıdır[28], [29]. İrfan ve Ortiz, ağ üzerindeki “etkileme oyununu”, ağdaki popülasyonun davranışlarını oyun teorisi ile modelleyerek göstermişlerdir. Ayrıca, en etkili bireylerin belirlenmesi problemini ele almışlar Nash-dengesi hesaplamalarını kullanarak yeni bir yaklaşım geliştirmişlerdir. Yazarların çalışmaları önemlidir, çünkü doğrusal-etki oyunlarını oyun teorisindeki önemli modellere bağlar. Molinero ve ark., EM problemini bireylerin işbirlikçi görevleri yerine getirme motivasyonlarının ancak diğer etkili bireyler tarafından bu görevlerin yerine getirilmesine bağlı olduğu bir sistem olarak düşünmüşlerdir. Yani, etkin bireyler bir şeyi yaparlarsa, diğer bireyler de etkili olan bu bireylerden görerek aynı şeyi yapabilirler. Aynı zamanda, bu senaryoyu bir etkileme oyunu olarak modellemişlerdir. Çalışmaları ayrıca, metrikler ve özellikler içeren etki oyunları üzerindeki çeşitli problemlerin hesaplama karmaşıklığını karakterize ettiği için de önemlidir. Oyun teorisi kullanan bir başka çalışma ise Ok ve arkadaşlarının çalışmasıdır [30]. Ok ve ark., EM probleminin bir parçası olarak düşünülebilecek Etki Yayılım Hızının Maksimizasyonu problemini ele almışlardır. Çalışmalarında, oyun teorisini kullanmışlar ve Pratik Paylaştırma ve Çekirdek Seçimi (Practical Partitioning and Seeding , PrPaS) adı verilen bir etkin çekirdek bireylerin seçimi algoritması önermişlerdir. Yazarlar, PrPAS'ın gerçek sosyal ağ topolojisi üzerinden yayılma hızına dayalı diğer algoritmalarından daha üstün olduğu sonucuna varmışlardır.

Diğer ağgözlü yaklaşımlardan farklı olarak, Song ve ark. tarafından böl ve fethet yaklaşımını kullanan bir algoritma geliştirilmiştir [31]. Yazarlar, sosyal ağların en etkili kullanıcılarını bulmanın NP-hard bir problem olduğunu ve ağgözlü algoritmaların yüksek hesaplama maliyetlerine sahip olduğunu belirtmişlerdir. Sonuç olarak, büyük ölçekli mobil sosyal ağlar için paralel hesaplama ile böl ve fethet stratejisini kullanmışlardır. Topluluk Temelli Ağgözlü Algoritması olarak adlandırılan algoritmaları, iki bileşenden oluşur: (1) bilgi yayılımına göre büyük mobil sosyal ağı çeşitli topluluklara bölmek; (2) en etkili düğümleri bulmak için dinamik programlama

yoluyla toplulukları seçmek.

Literatürdeki birçok yaklaşım bir ağdaki tüm düğümleri çekirdek (etkin birey) adayı olarak ele almaktadır. Ancak, bir sosyal çevredeki etkili bireylerin sayısı toplam birey sayısına kıyasla çok küçüktür. Lee ve Chung'a göre EM'nin en zayıf noktası, bazı kullanıcıların diğer kullanıcılardan ayırt edilememesidir [32]. Yazarlara göre, etkin birey seçimi için bütün bireylerin arasından seçim yapmak yerine, belirli bir grup bireye odaklanılmalıdır. Bu yaklaşımla geliştirdikleri algoritma ile mevcut yöntemlere göre çok daha hızlı ve etkin bir sonuç elde etmişlerdir.

Literatürdeki çalışmaların çoğu sadece bir sosyal ağ ile ilgilidir. Ancak, kullanıcılar aynı anda birçok sosyal ağ hakkında bilgi paylaştıklarından, çoklu ağlar üzerinde araştırma yapmak oldukça önemlidir. Zhang ve ark., çoklu sosyal ağlarda En Az Maliyetle Etkileme (Least Cost Influence, LCI) problemini ortaya atmışlardır [33]. Yazarlar, kayıplı ve kayıpsız haritalama şemalarını kullanarak bir ağ kümesini tek bir ağda eşleştirmiştir. Birden fazla ağla çalışmak EM problemini zorlaştırır, çünkü bir bireyin farklı ağlardaki etki düzeyleri farklı olabilir. Bu bağlamda, çoklu ağlarda EM problemi ilginç ve zor bir problemdir.

IC ve LT modelleri literatürde sıklıkla kullanılmaktadır. Ne varki, bu modeller gerçek dünyadaki etki yayılımının yalnızca basit birer modelidir. Bu bağlamda Tong ve ark., yayılım modellerinin gerçekçi olmamalarını eleştirmişlerdir [34]. Yazarlar, gerçek dünyadaki sosyal ağların çok fazla belirsizlik barındırdığını belirtmişler ve dinamik IN modeli olarak adlandırdıkları bir yayılım modeli kullanarak, geliştirdikleri uyarlanabilir çekirdek birey seçimi kavramını kullanmışlardır. Basit bir açgözlü uyarlanabilir çekirdek düğüm seçimi stratejisi önermişler ve daha iyi ölçeklendirme için verimli bir sezgisel algoritma geliştirmişlerdir. Gerçek ve sentetik veri kümeleri üzerinde yaptıkları deneyler sonucunda, önerdikleri stratejinin temel yöntemlerden üstün olduğunu belirtmişlerdir.

Bir başka çalışmada EM sorunu yapısal kontrol edilebilirlik problemi olarak ele alınmış ve problemi çözmek için ağırlıklı bir maksimum eşleştirme algoritması önerilmiştir [35]. Sartor ve ark., çalışmalarında Kempe'nin algoritmasını kıyaslama için referans olarak kullanmışlar ve önerdikleri yöntemin rekabetçi sonuçlar verdiğini göstermişlerdir.

Peng ve ark., mobil sosyal ağlarda sosyal etkiyi ölçmek için bir çerçeve çalışma

hazırlamışlardır [36]. Değerlendirme modellerinin çoğunun çevrimiçi sosyal ağlara odaklandığını, ancak dolaylı etkiyi tanımlayamadığını belirtmişlerdir. Daha sonra, sosyal etkinin yayılma dinamiklerini karakterize etmek için Duyarlı-Enfeksiyöz (SI) salgın modelinden bir yayılım modeli geliştirmişlerdir. Arkadaşlık entropisi ve etkileşim sıklığı entropisini kullanarak doğrudan, dolaylı ve toplam etkiyi ölçmek için bir değerlendirme modeli geliştirmişlerdir. Algoritmaları, düğümleri toplam etkilerine göre azalan düzende sıralar ve ilk k düğümü açgözlü bir şekilde alır. Önerilen yöntemin, rasgele seçim ve düğümlerin derecelerine göre seçim yöntemlerini geride bıraktığını belirtmişlerdir.

Sosyal etki, iki gruba ayrılır: pozitif etki ve negative etki. Bu durumda sosyal etki, işaretli olarak kabul edilir. Li ve ark., bireyler arasında olumsuz ve olumlu ilişkileri içeren işaretli sosyal ağlarda EM problemini ele almışlardır [37]. Amaçları, maksimum pozitif veya maksimum negatif etkiyi yaratacak bir dizi çekirdek bireyi bulmaktır. Yayılım modeli olarak İki Kutuplu (polarity) IC (IC-P) modeli kullanmışlar ve açgözlü bir algoritma geliştirmişlerdir. Benzer bir çalışmada, çekirdek birey seçimi için Tavlama Benzetimi algoritması kullanılmış ve pozitif EM ele alınmıştır. Yazarlar, SNAP (Stanford Network Analysis Project) veri seti koleksiyonundan Epinions, Slashdot ve Wikipedia veri setlerini kullanmışlardır. Yöntemleri, açgözlü algoritmalarından pozitif etki bakımından daha iyi ya da onlarla benzer sonuçlar göstermiştir [38].

Gong ve ark., EM problemi için evrimsel bir algoritma olan ve sosyal ağlarda topluluk temelli etkiyi en üst düzeye çıkarmayı amaçlayan bir memetik algoritma önermişlerdir [39]. En etkili bireyleri bulmak için 2-atlamalı etki yayılımını optimize etmişler ve 3 adet gerçek veri seti ile denemeler yapmışlardır. Yazarlar, önerdikleri yöntemin etkinlik bakımından iyi sonuçlar verdiğini belirtmişlerdir.

Son olarak, Gong ve ark., sosyal ağlarda EM problemi için parçacık sürüsü optimizasyonuna dayalı bir yöntem geliştirmişlerdir [10]. EM problemi için algoritmanın birçok yönünü (parçacıkların güncelleme kuralları, yerel arama stratejileri, vb.) yeniden tanımlamışlardır. Yazarlar algoritmalarının etkili olduğunu ve iyi bir performans sergilediğini bildirmişlerdir.

Literatürde yer alan çalışmalardan görülebileceği gibi, EM problemini çözme girişimleri farklı algoritmik yaklaşımlar ve farklı ölçütler kullanılarak yapılmıştır. Ayrıca, EM problemi ve alt problemleri çeşitli ağ türleri ve çeşitli farklı bilgi yayılımı modelleri için

incelenmiştir. Genel olarak, bir sosyal ağdaki en etkili bireylerin saptanması çoğu durumda NP-hard bir problem olarak görülebilir. Mevcut açgözlü algoritmalar büyük ağlarda yetersizdir. Ayrıca, açgözlü algoritmalar yayılma fonksiyonlarının ağır Monte-Carlo simülasyonlarını gerektirir [40]. Diğer yandan, optimizasyon algoritmaları genellikle daha iyi sonuçlar verir. Dezavantajları ise, daha yavaş çalışmalarıdır. Sürü zekâsı algoritmaları, daha önce belirtilen nedenlerden dolayı EM problemlerinde yaygın olarak kullanılamamaktadır. Daha hızlı sonuçlar veren sürü zekâsı algoritmalarının EM probleminde kullanılması literatürdeki bu boşluğu dolduracaktır.

Çizelge 1.1, incelenen çalışmaların bir özetini sunmaktadır. Çoğu çalışmada açgözlü yaklaşımlar benimsenmiştir. Optimizasyon algoritmaları literatürde daha az kullanılırken, IC modeli en yaygın kullanılan yayılım modelidir. Literatürde, EM problemini yeniden şekillendirerek sürü zekâsı ya da diğer optimizasyon algoritmalarına uygun hale getirmeyi öneren bir çalışma bulunmamaktadır. Yapılan çalışmalarda genellikle ağ ölçütü olarak Derece Merkeziliği ve türevleri kullanılmıştır. Bazı çalışmalar, bu ölçütleri kendi algoritmaları içerisinde kullanırken, bazı çalışmalar yalnızca kıyaslama amacı ile kullanmışlardır. Çoğu çalışmanın, hazır bir veri seti kullandığı; daha az sayıdaki diğer çalışmaların ise kendi veri setlerini, veri toplama, ayıklama vb. işlemler sonucunda hazırladıkları görülmüştür. Hazır veri seti olarak en çok kullanılan veri setleri SNAP veri setleridir.

1.2. TEZ ORGANİZASYONU

Bu tez çalışması aşağıda özetleri verilen bölümler şeklinde organize edilmiştir. 1. Bölümde, gerçek sosyal ağlar ve sosyal ağların online platformlardaki modellerinden bahsedilmiştir. Etki maksimizasyonu problemi kavramı genel olarak tanıtılmış ve literatürde etki maksimizasyonu için yapılan çalışmalara yer verilmiştir. 2. Bölümde, sosyal ağın teknik bakış açısından bahsedilmiş, ağ çeşitlerine yer verilmiştir. Online sosyal ağlarda etki maksimizasyonu kavramı açıklanmış, etkiyi maksimize edebilmek için kullanılan ağ ölçütleri ve yayılım modellerinden bahsedilmiştir. 3. Bölümde, etki maksimizasyonu problemi detaylı olarak ele alınmış ve çalışmadan kullanılan sürü zekâsı algoritmaları açıklanmıştır. 4. Bölümde, etki maksimizasyonu probleminin çözümü için önerilen yöntem verilmiştir. 5. Bölümde, elde edilen deneysel sonuçlara yer verilmiştir. 6. Bölümde, sonuçlar değerlendirilmiş ve çalışmanın literatüre katkılarından bahsedilmiştir. Gelecek çalışmalar için öneriler verilmiştir.

Çizelge 1.1. İncelenen çalışmaların özeti.

Çalışma	Algoritma	Yayımlı Modeli	Ağ Ölçütleri	Veri Seti
[2]	Açgözlü	IC	-	SNAP'ten Facebook ağı.
[25]	Açgözlü	IC	-	Sentetik ağ. SNAP'ten Gnutella ağı.
[30]	Açgözlü	Independent Poisson Clock	Derece Merkeziliği.	SNAP'ten Facebook ağı.
[4]	Dinamik Programlama	IC	Çıkış Derecesi Merkeziliği.	SNAP'ten DBLP ağı.
[41]	Açgözlü	IC	Derece Merkeziliği.	Sentetik ağ. SNAP'ten Email, Gnutella, Slashdot, Amazon ağları.
[27]	Açgözlü	Latency Aware IC	-	SNAP'ten Wiki, Epinions, Slashdot, LiveJournal ağları.
[31]	Açgözlü	IC	Derece Merkeziliği.	China Mobile Firmasından alınan arama ağı.
[32]	Açgözlü	IC	-	SNAP'ten Wiki-Vote, Epinions, Slashdot, Amazon, Pokey ve Gowalla ağları.
[33]	Greedy	LT ve IC	-	MEDLINE, SPIRES, NCSTRL Veritabanlarından elde edilen ortak yazar ağları.
[34]	Açgözlü ve Sezgisel	IC	Derece Merkeziliği.	Sentetik ağ. SNAP'ten Wiki ağı.
[11]	Parçacık Sürü	IC	Derece Merkeziliği, PageRank.	NetGRQC, NetHEPT, NetPHY ağları.
[42]	Açgözlü	IC	Derece ve Yakınlık Merkezilikleri.	NetHEPT ve NetPHY ağları.
[5]	Genetik Algoritma	LT	Derece Merkeziliği.	NetScience, Infectious: Stay Away Ağları. KONECT'ten UCSOCIAL ağı.
[37]	Açgözlü	İki Kutuplu IC	Çıkış Derecesi Merkeziliği.	SNAP'ten Epinions ve Slashdot ağları.
[38]	Tavlama Benzetimi	İki kutuplu IC	Çıkış Derecesi Merkeziliği.	SNAP'ten Epinions, Wikipedia ve Slashdot ağları.
[36]	Açgözlü	SI	Derece Merkeziliği.	Çin'deki bir GSM operatöründen toplanan mesajlaşma ağı.
[22]	Karma bir tamsayı programlama yöntemi	Partial Parallel Cascade (PPC)	Çıkış Derecesi Merkeziliği.	Bir sağlık forumundan elde edilen ağ.

2. SOSYAL AĞLAR VE ONLINE SOSYAL AĞLAR

Gerçek hayatta bireyler diğer bireylerle etkileşim halindedir. Aralarında sevmeme, beğenme, kızma gibi çeşitli ilişkiler mevcuttur. Bireyler arasında bilgi akışı yine bu ilişkiler sayesinde gerçekleştirilir. Gerçek hayattaki bu iletişim ve ilişki modellenmiş ve Facebook, Twitter gibi online sosyal ağlar oluşturulmuştur. Bu tip platformlarda bireyler birbirlerini takip edebilir, paylaşımlarını beğenebilir veya beğenmeyebilir, yorum yapabilir ve hatta aynı paylaşımı kendisi de yeniden paylaşabilir. Gerçek hayatta olduğu gibi bilgi akışı bu platformlar üzerinden de yapılabilir. Buradaki avantaj ve kimi zaman dezavantaj bu akışın günlük hayattakine göre oldukça hızlı gerçekleşmesidir. Paylaşılan bir bilgi birkaç saat içinde milyonlarca kişiye ulaşabilir. Bilginin bu hızlı yayılımı pazarlama şirketlerinin dikkatlerini çekmiş ve ürünlerini bu platformlar üzerinden pazarlamaya, reklamlarını yine bu platformlar üzerinden yapmaya başlamışlardır. Burada önemli olan ağdaki etkili bireylerin tespit edilebilmesidir. Etkili bireyler kavramı diğer kullanıcılar üzerinde normal kullanıcılara göre daha çok etkisi olan kullanıcılar için kullanılmaktadır. Etkili kullanıcılar sosyal statülerinden, karizmalarından, fikir önderi olmalarından veya başka çeşitli nedenlerden dolayı diğer bireyler üzerinde belirli bir etkiye sahiptirler. Onların fikirleri, paylaşımları sosyal ağlar üzerinden daha hızlı yayılabilir ve iyi veya kötü anlamda reklamlar daha hızlı yapılabilir.

Online sosyal ağlarda paylaşılan veriler çok büyük miktarlardadır. Sürekli bir bilgi akışı vardır. Bilginin hangi kullanıcılar tarafından daha etkin şekilde yayıldığı etkili kullanıcıları tespit etmek açısından önemlidir. Binlerce ve hatta milyonlarca kullanıcısı olan bu tip platformlardaki etkili bireylerin tespit edilmesi başlı başına bir problemdir.

2.1. SOSYAL AĞIN TEKNİK BAKIŞ AÇISI

Gerçek sosyal ağların modellenmesi sonucu oluşan online sosyal ağlarda gerçek hayattaki durumları modellemek için takip ilişkisi, arkadaşlık ilişkisi gibi ilişki türleri oluşturulmuştur. Twitter gibi ortamlarda bireyler birbirlerini takip ederler ve paylaşımı yeniden paylaşabilirler. Facebook gibi bir sosyal ağda ise arkadaşlık ilişkisi

oluşturulabilir. Paylaşılan gönderiler beğenilebilir, beğenilmeyebilir, yeniden paylaşılabilir, yorum yapılabilir. Bu tip beğenme ve beğenmeme gibi durumlar modellenirken işaretli sosyal ağlar kavramı ortaya çıkmıştır. İşaretli sosyal ağlarda bir kullanıcının diğerinden olumlu etkilenmesi, onun gönderilerini beğenmesi pozitif, beğenmemesi veya olumsuz etkilenmesi negatif durum olarak değerlendirilirken, bireyin diğer bireyi görmezden gelmesi veya hiç etkilenmemesi pasif olarak değerlendirilir.

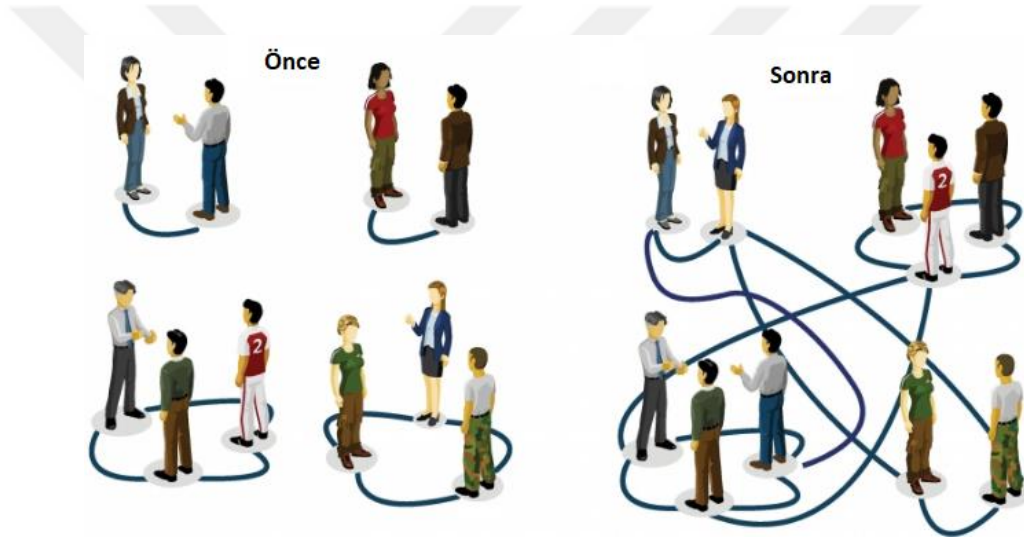
Dünya'daki yaşam birbirine bağlı, çok büyük sayılarda elemanlardan oluşan ağların bir parçası olmak demektir [43]. Geniş ve kompleks ağ yapıları hemen her yerde karşımıza çıkmaktadır. Sosyal ağlar, protein etkileşim ağları, canlıların birbirleri ile olan ilişkilerini tanımlayan ağlar, haberleşme ağları vb. bu ağların en iyi bilinenlerinden birkaçıdır. Bu ağların incelenerek anlamlandırılması ve barındırdıkları bilgilerin elde edilmesi için ilk olarak bilgisayar ortamında modellenmeleri, daha sonra da bu modeller üzerinde çalışacak algoritmaların geliştirilmesi gerekmektedir. Bir sosyal ağ içerisindeki bireylerin etkileşimleri incelenerek, kimin kimle yakın arkadaş olduğu; hangi bireylerin bir sosyal çevre oluşturduğu; belirli bir çevredeki en etkili bireyler bulunabilir. Bu gibi bilgiler, viral pazarlama (marketing) ve anayurt güvenliği gibi birçok alanda kullanılabilir.

Ağ modelleme için graf veri modeli kullanılmakta ve temel graf işlemleri için graf teorisinden faydalanılmaktadır. Her bir alt alanın kendine has problemleri olduğu için o alana özgü algoritmalar geliştirilmekte ve problemler çözülmeye çalışılmaktadır. Graftaki düğümler bireyleri, kenarlar bireyler arasındaki ilişkileri göstermektedir. Kenarlar üzerindeki ağırlıklar bireylerin birbirlerini etkileme olasılıklarını temsil etmektedir.

2.1.1. Ağ Çeşitleri

Bu kısımda EM problemlerini modellemek için kullanılacak rasgele ağlara ve rasgele ağları kullanan graf türlerine değinilmiştir. Şekil 2.1'de verilen örnek ağ modelinde verildiği gibi birbirini tanımayan yüz kişinin çağrılacağı bir parti düzenlendiği varsayalım. İçecek ve yiyecek bir şeyler ikram edildiğinde, kısa sürede bu kişilerin ikili üçlü gruplar oluşturup sohbete başladıkları görülür. Konuklardan birisine, örneğin Ayşe'ye, içecekler hakkında bilgi verilmiş olsun. Eski şişeye sahip içeceğin aslında daha kaliteli olduğu, yalnızca eski şişeye kamufle edildiği, gösterişli şişedekinin

ise daha tatsız olduğu söylenmiş olsun. Ayşe bu bilgiyi öncelikle tanıdıklarına aktaracaktır. Henüz yeni tanışmaya başladığı için bilgi sadece birkaç kişiye aktarılır ve bu aktarım aralarında küçük bir bağlantı oluşturulmuş gibi değerlendirilebilir. Fakat aradan biraz zaman geçince Ayşe'nin bilgi aktardığı Ali başka kişilerle sohbete başlar ve Ayşe'nin şahsen tanımadığı Damla'ya bilgi aktarır. Bu şekilde bilgi aktarımı devam eder ve kısa sürede partideki herkes hangi şişedeki içeceğin daha iyi olduğunu öğrenmiş olur, daha çok kişi arasında konuştuğça bağlantı oluşur. Partideki bu model rasgele ağ modeli denilen bir ağa örnektir. Rasgele ağ modeline göre partideki bilginin herkese aktarılması için Ayşe'nin teker teker herkesle tanışması gerekmez. Ayşe tanıdıklarına, Ayşe'nin tanıdıkları da diğerlerine bilgi aktarımı yapar ve böylelikle bilginin herkese yayılması için görünmez bir ağ oluşur [44].



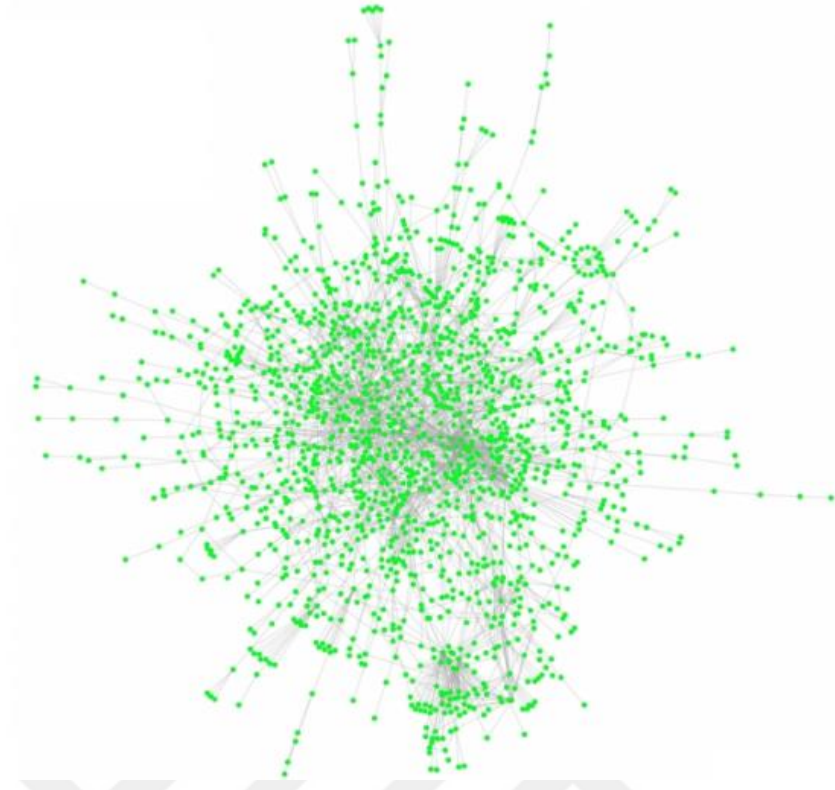
Şekil 2.1. Partideki konuklar arasındaki ilişkiler [44].

Bir kokteyl partisinden rasgele ağlara:

- Bir kokteyl partisinde, rasgele tanışmalar yoluyla bir tanıdık ağı ortaya çıkar.
- Başlangıçta konuklar izole gruplar oluşturur.
- Bireyler karıştığında, gruplar değişir ve hepsini tek bir ağa bağlayan görünmez bir ağ ortaya çıkar [44].

2.1.1.1. Erdős-Renyi Grafi

Ağ bilimi, gerçek ağların özelliklerini yeniden üreten modeller oluşturmayı amaçlamaktadır. Karşılaştığımız birçok ağ, bir kristal kafesin düzenine veya örümcek ağının öngörülebilir radyal mimarisine sahip değildir. Daha ziyade, ilk bakışta sanki rasgele bükülmüş gibidirler (Şekil 2.2). Rasgele ağ teorisi, gerçekten rasgele olan şekilleri yapılandırarak ve karakterize ederek ele alır.



Şekil 2.2. Rasgele Ağlar [44].

Modelleme açısından bakıldığında ağ yalnızca düğümler ve bağlantılardan oluşan nispeten basit bir nesnedir. Bununla birlikte gerçek zorluk, gerçek bir sistemin karmaşıklığını üretebilmek için düğümler arasındaki bağlantıların nereye yerleştirileceğine karar vermektir. Bu bağlamda, rasgele bir ağın arkasındaki felsefe basittir: gerçek ağı modelleyebilmek amacıyla düğümler arasına bağlantılar rasgele yerleştirilir [44].

Rasgele bir ağ, her düğüm çiftinin birbirine p olasılığıyla bağlı olduğu N adet düğümden oluşur. Rasgele ağlar için iki adet tanım mevcuttur:

- $G(N, L)$ Modeli: N etiketli düğümler L rasgele yerleştirilmiş bağlantılarla bağlıdır. Erdős ve Rényi bu tanımlamayı kullanmışlardır [45].
- $G(N, p)$ Modeli: Gilbert tarafından sunulan bu modelde her N etiketli düğüm çifti birbirine p olasılığıyla bağlıdır [45].

$G(N, p)$ Modeli iki düğümün bağlı olma olasılığı p 'yi sabit tutarken, $G(N, L)$ Modeli toplam bağlantı sayısı L 'yi sabit tutar. $G(N, L)$ Modelinde düğümün ortalama derecesi basit olarak Denklem (2.1) ile hesaplanabilirken, $G(N, p)$ Modelinde diğer ağ karakteristikleri kolaylıkla hesaplanabilmektedir.

$$\langle k \rangle = 2L/N \quad (2.1)$$

Rasgele bir ağ oluşturmak için aşağıdaki adımlar izlenir:

- N adet izole düğüm ile başlanır.
- Bir düğüm çifti seçilir ve 0 ile 1 arasında rasgele bir sayı üretilir. Eğer üretilen sayı p olasılığında büyükse seçilen düğümler birbirine bağlanır, aksi takdirde aralarında bağlantı kurulmaz.
- Her $N(N - 1)/2$ düğüm çifti için ikinci adım tekrar edilir.

Yukarıdaki adımlar uygulandıktan sonra elde edilen ağ rasgele graf veya rasgele ağ olarak isimlendirilir. Pál Erdős ve Alfréd Rényi isimindeki iki matematikçi bu ağların özelliklerini anlamamızda önemli bir rol oynamışlardır. Bu yüzden bir rasgele ağ Erdős-Rényi ağı olarak isimlendirilir [44].

2.1.1.2. Ölçek Bağımsız Ağlar

Ölçek bağımsız ağlar yapı olarak rasgele ağlardan çok farklıdır. Ölçek bağımsız ağlarda düğümlerle düğümlerin sahip olduğu derece sayılarının dağılımı kuvvet yasasına (Power Law) göre olmaktadır. Bunun anlamı bir ölçek bağımsız ağdaki az sayıdaki düğümün ağ üzerindeki düğümlerin büyük çoğunluğundan çok daha yüksek dereceye sahip olmasıdır. Ölçek bağımsız ağlar Albert Barabasi ve arkadaşları tarafından literatüre kazandırılmıştır [46]. Gerçek hayattaki birçok ağ ölçek bağımsız özellik göstermektedir. Örneğin, dünya üzerindeki hava alanlarından çok küçük bir kısmının uçuşların büyük bir çoğunluğuna ev sahipliği yapması ve dünya üzerindeki havaalanlarının büyük çoğunluğundan daha yüksek bağlantıya sahip olması gibi.

Ölçek bağımsız ağ aslında k sayıda bağlantı özelliğine sahip $P(k) \sim k^{-\gamma}$ güç yasası dağılımı gösteren birbirine bağlı bir graf veya ağdır. Burada γ , güç yasasının derece eksponentidir.

Ölçeksiz bir ağ, mevcut bir ağa artan bir şekilde düğümler ekleyerek ve mevcut düğümlere bu bağlantıları tanıtarak oluşturulabilir. Bu durumda bir i düğümünün diğerlerine bağlanma olasılığı (linking probability) mevcut k_i bağlantılarının sayısıyla orantılı olacaktır.

$$P(i \text{ düğümüne bağlanma}) \sim \frac{k_i}{\sum_j k_j} \quad (2.2)$$

Ölçek bağımsız ağlar web sayfalarının topolojisini (düğümler web sayfalarını, kenarlar da hiperbağlantıları gösterir) de içeren bilim ve mühendisliğin birçok alanında, Hollywood aktörlerinin ortak ağında (düğümler aktörler, kenarlar ise aynı filmdeki yıldızları gösterir), hakemli literatür ağında (düğümler yayınları, kenarlar ise atıfları gösterir) gibi birçok alanda meydana gelmektedir [46], [47].

2.1.1.3. Küçük Dünya Ağı

Küçük Dünya ağı, düğümlerin hepsinin doğrudan birbirleri ile komşu olmadıkları, hatta ağın tamamına göre çok az düğüm ile komşu oldukları; bununla birlikte, ağ üzerindeki bütün düğümlere az sayıda atlama ile ulaşabildikleri ağ çeşididir. Sosyal ağlar, İnternet üzerindeki web sayfalarının oluşturduğu ağ gibi birçok gerçek ağ, Küçük Dünya ağı özelliği göstermektedir. Bir Küçük Dünya ağındaki düğümlerin birbirlerine olan en kısa yollarının ortalaması, rasgele ağlara göre daha düşüktür. Bununla birlikte, Küçük Dünya ağı özelliği gösteren ağların Kümeleme Katsayıları, rasgele ağlara göre daha yüksektir [44].

2.2. ONLINE SOSYAL AĞLARDA ETKİ MAKSİMİZASYONU

Son yıllarda hızla gelişen Online Sosyal Ağlar sosyal yaşamın dijital ortamda modellenmiş şeklidir. Sosyal yaşamda bazı kişiler daha etkindir ve diğerlerini etkileme yeteneğine sahiptirler. Bu belki sahip oldukları karizmadan, belki sosyal statülerinden veya belki de hitabet yeteneğinden olabilir. Online sosyal ağlarda da bu tip etkili kullanıcılar mevcuttur. Bu kullanıcıların tespit edilmesi etkili kullanıcıların tespiti problemi olarak karşımıza çıkmaktadır. Bu etkinin maksimize edilmesi ise sosyal etki maksimizasyonu olarak bilinmektedir. Bu etki her zaman olumlu yönde olmayabilir, dolayısıyla etkinin de sınıflandırılması anlamlı olmaktadır. Buradaki sınıflar olumlu etki (beğenme), olumsuz etki (beğenmeme) veya etkisiz olarak sayılabilir. Bu durumda problem sosyal ağlarda işaretli etkinin tespiti veya maksimizasyonu olarak isimlendirilir.

2.2.1. Ağ Ölçütleri

2012 yılında Amerikalı basketbol yıldızı Kobe Bryant, Çin mikroblog sitesi Sina Weibo'ya katılmış, birkaç saat içinde 100.000'den fazla takipçisi sayfasına katılmış ve ilk gönderisini yayınlamasını beklemişlerdir. Basın, bu takipçi sayısını Kobe Bryant'ın

Çin’de popüler olması şeklinde yorumlamıştır. Bu örnekte Bryant’ın Çin sosyal medya kullanıcıları arasındaki popülerliğini ölçmek için takipçi sayısı ele alınmıştır [48]. Sosyal etkinin ölçülebilmesi için çeşitli değerlendirme ölçütleri bulunmaktadır.

Graf gösterimi sosyal ortamı analiz ederken sıklıkla kullanılmaktadır. Bu graf arkadaşlık veya sosyal ortam ağındaki kullanıcı etkileşimlerini gösterir. Böyle bir grafta, aşağıdaki soruların cevapları aranır:

- Ağdaki merkezi figürler (etkili bireyler) kimlerdir?
- Arkadaşlar arasında yaygın olan etkileşim desenleri nelerdir?
- Kimler benzer düşünce yapısını sahiptirler ve bu bireyler nasıl bulunabilir?

Bu ve benzeri soruları cevaplayabilmek için öncelikle merkeziliğin niceliğini, etkileşim seviyelerini ve benzerliği belirlemek için ölçütler tanımlanmalıdır. Bu ölçütler yardımıyla ağdaki etkin bireylerin tespiti yapılabilir.

Ölçütler, düğüm ölçütleri ve kenar ölçütleri olarak ayrılmaktadır.

Düğüm ölçütleri:

- Merkezilik
- Derece merkeziliği
- Özvektör merkeziliği
- Katz merkeziliği
- Bonacich merkeziliği
- Aradalık merkeziliği (Betweenness Centrality)
- Yakınlık merkeziliği

Kenar Ölçütleri:

- Bağ kuvveti (Tie Strength)
- Üçlü Yakınlık (Triadic Closure)
- Zayıf Bağ (Weak Tie)
- Kenarın arada bulunma durumu (Edge Betweenness)

2.2.1.1. Dügüm Ölçütleri

Bu kısımda düğüm ölçütlerinin tanımları verilmiştir. Merkezilik, bir düğümün ağda ne kadar önemli olduğunu tanımlar.

Derece Merkeziliği,

En basit ve en popüler merkezilik ölçütüdür. Bir düğümün derecesi, yönlü graflarda giriş derecesi veya çıkış derecesi veya bunların birleşimi şeklinde olabilir. Yönsüz veya bütün kenarları çift yönlü olan graflar için ise sadece derece olarak isimlendirilir ve düğüme bağlı kenar sayısını gösterir. $deg(i)$, i düğümünün derecesi olsun. c_i^{DEG} ifadesi, i düğümünün derece merkeziliği olarak isimlendirilir ve düğümün derecesini verir.

$$c_i^{DEG} = deg(i) \quad (2.3)$$

Derece merkeziliği düğümünden çıkan ve uzunluğu 1 olan yol sayısıdır. Bir düğümünden fazla k adet yol çıkıyorsa buna k -yol merkeziliği denir [49].

Yönlü sosyal ağ graflarında giriş dereceleri kullanıldığında, derece merkeziliği düğümün ne kadar popüler olduğunu ölçer ve ölçüt değeri düğümün şöhretini veya prestijini gösterir. Çıkış derecesi kullanıldığında, derece merkeziliği düğümün topluluğa katılma değerini ölçer. Giriş ve çıkış derecelerini birleştirdiğimizde ise temelde kenarın yönünü göz ardı ediyoruz demektir. Böylelikle kenar yönleri ortadan kalktığında yönsüz graflar için derece merkeziliğini ölçülmüş olunur [48].

$$c_i^{DEG} = deg_i^{giris} \quad (\text{giriş derecesi}) \quad (2.4)$$

$$c_i^{DEG} = deg_i^{cikis} \quad (\text{çıkış derecesi}) \quad (2.5)$$

$$c_i^{DEG} = deg_i^{giris} + deg_i^{cikis} \quad (\text{birleştirilmiş derece}) \quad (2.6)$$

Derece merkeziliği ölçütü ağlar arasında (Facebook ve Twitter gibi) merkezilik değerlerinin karşılaştırılmasına izin vermez. Bu problemin üstesinden gelmek için, derece merkezilik değerleri normalize edilebilir.

Derece merkeziliğini normalize etmek için basit normalizasyon metotları, maksimum olası derece ile normalize etmeyi içerir,

$$C_{deg}^{norm}(i) = \frac{deg_i}{n - 1} \quad (2.7)$$

Denklem 2.7'de, n düğüm sayısını göstermektedir. Aynı zamanda maksimum dereceye

göre de normalize edilebilir,

$$C_{deg}^{max}(i) = \frac{deg_i}{\max_j deg_j} \quad (2.8)$$

Son olarak, derece toplamına göre de normalize edilebilir [48]:

$$C_{deg}^{toplam}(i) = \frac{deg_i}{\sum_j d_j} = \frac{deg_i}{2|E|} = \frac{deg_i}{2m} \quad (2.9)$$

Özvektör Merkeziliği (Eigenvector Merkeziliği),

Derece merkeziliğinde, daha çok bağlantısı olan düğümler daha önemli kabul edilir. Gerçek yaşam senaryolarında çok arkadaşına sahip olmak kişinin önemli olmasını garantilemez. Daha önemli arkadaşlara sahip olmak daha güçlü bir kişi olduğunun göstergesi olabilir.

Özvektör merkeziliği, komşuların (yönlü graflardaki gelen komşuların) önemini de işe katarak derece merkeziliğini genelleştirmeyi dener. Hem yönlü hem de yönsüz graflar için tanımlanmıştır. Komşuların izini tutmak için grafin A komşuluk matrisi kullanılabilir. v_i düğümünün özvektör merkeziliği $C_e(v_i)$ olsun. v_i 'nin merkeziliğinin, onun komşuluklarının merkeziliklerinin bir fonksiyonu olması istenir.

$$c_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^n A_{j,i} C_e(v_j) \quad (2.10)$$

Burada λ , sabit bir katsayıdır. Bütün düğümler için merkezilik vektörleri Denklem (2.11)'deki gibi olsun.

$$C_e = (C_e(v_1), C_e(v_2), \dots, C_e(v_n))^T \quad (2.11)$$

Bu durumda Denklem (2.10) aşağıdaki şekilde yazılabilir [48]:

$$\lambda C_e = A^T C_e \quad (2.12)$$

Katz Merkeziliği,

1953 yılında Leo Katz tarafından tanıtılmıştır. Katz merkeziliği, sosyal ağdaki bir bireyin etkisinin bağlı derecesini ölçmek için kullanılır [50]. Bir birey çifti arasındaki en kısa yolu hesaba katan tipik merkezilik ölçütlerinden farklı olarak Katz merkeziliği, bir birey çifti arasındaki toplam adım sayısını hesaba katarak etkiyi ölçer [49].

Özvektör merkeziliğindeki en büyük problem yönlü graflarda çalışıldığında ortaya çıkar. Merkezilik giden kenarlar olduğunda devam eder ve düğümün yönlü çevrimsiz graflar içinde olması gibi özel durumlarda, düğüme bağlı birçok kenar olmasına rağmen merkezilik sıfır olur. Bu durumda, merkezilik değerine bir bias terimi eklenerek sorun çözülebilir. Bias terimi Γ , ağın topolojisinden bağımsız olarak bütün düğümlere bir merkezilik değeri verilmesini sağlar. Bu değere Katz merkeziliği ismi verilmiştir [48].

Katz merkeziliği bir düğümden başlayan adım sayısını (walk sayısını) hesaplar, uzun adımlara ceza verir. C_i^{KATZ} , i düğümünün Katz merkeziliği olmak üzere,

$$C_i^{KATZ} = e_i^T \left(\sum_{j=1}^{\infty} (\Gamma A)^j \right) \vec{1} \quad (2.13)$$

e_i , i . elemanı 1, diğer bütün elemanları 0 olan sütun vektörü; Γ , 0 ile 1 arasında pozitif ceza sabiti; $\vec{1}$, bütün değerleri 1 olan vektör.

Bonacich Merkeziliği,

Phillip Bonacich tarafından derece merkeziliği modifiye edilerek önerilmiştir [51]. Bonacich merkeziliği aslında Katz'ın bir versiyonudur. Γ ceza sabiti burada da geçerlidir, fakat Katz'dan farklı olarak negatif değerler alabilir. C_i^{BON} ifadesi Bonacich merkeziliğini temsil etmek üzere,

$$C_i^{BON} = e_i^T \left(\frac{1}{\Gamma} \sum_{j=1}^{\infty} (\Gamma A)^j \right) \vec{1} \quad (2.14)$$

Buradaki negatif Γ değeri tek sayıdaki dolaşmaları çift sayıdaki dolaşılardan çıkarmayı sağlar. Bu durum exchange ağlarda anlamlıdır [49].

Aradalık Merkeziliği (Betweenness Centrality),

Kenarlarda olduğu gibi düğümlerde de yüksek aradalık değerine sahip düğüm ağ yapısı içinde kritik pozisyonadadır. Bu nedenle, ağda kritik bir rolü vardır. Sıkı örülmüş gruplardaki düğümlerin taşıdığı büyük miktardaki akış sayesinde düğüm aradalığı tetiklenir. Bu tip düğümler yüksek aradalık değerine sahip kabul edilir.

Aradalık kavramı sosyal ağlarda yapısal boşluklar yayan düğümlerle ilgilidir. Freeman'in aradalık merkeziliği en bilinenidir [52]. Bir düğümün, diğer düğümlerin en kısa yolları üzerinde olması durumunu ölçer.

$$C_i^{BET} = \sum_{j,k} \frac{b_{jik}}{b_{jk}} \quad (2.15)$$

b_{jk} , j düğümünden k düğümüne olan en kısa yolların sayısı, b_{jik} ise j düğümünden k düğümüne i düğümü üzerinden geçerek giden en kısa yolların sayısıdır.

Aradalık değerini hesaplayan en sade algoritma bütün en kısa yol çiftlerini hesaba katar. Bu da $O(n^3)$ zaman ve $O(n^2)$ depolama alanı gerektirir [49].

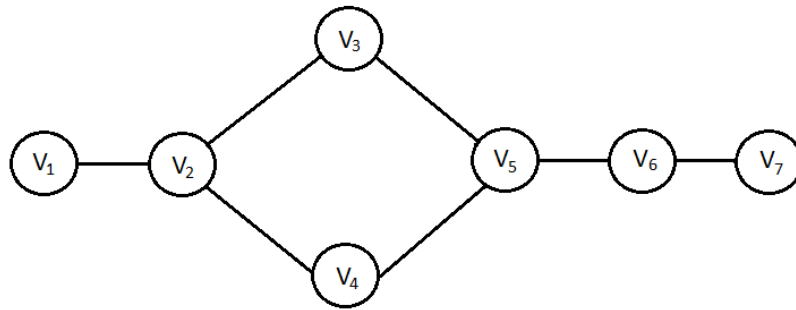
Yakınlık Merkeziliği için ,

En popüler yakınlık merkeziliği Freeman'ın yakınlık merkeziliğidir [53]. Bir düğümün bütün diğer düğümlere olan en kısa uzaklığının ortalamasını hesaplayarak yakınlık merkeziliğini belirler.

Yakınlık merkeziliğinde, düğümler ne kadar merkezilerse diğer düğümlere o kadar çabuk ulaşabilirler. Bu düğümler diğerlerinden daha küçük ortalama en kısa yol uzunluğuna sahip olmalıdırlar.

$$C_c(v_i) = \frac{1}{\bar{l}_{v_i}} \quad (2.16)$$

Burada $\bar{l}_{v_i} = \frac{1}{n-1} \sum_{v_j \neq v_i} l_{i,j}$ değeri v_i 'nin diğer düğümlere olan ortalama en kısa yol uzunluğudur. Ortalama en kısa yol uzunluğu ne kadar düşük olursa düğümün merkeziliği de o kadar yüksek olur [48]. Şekil 2.3, örnek bir grafi göstermektedir.



Şekil 2.3. Yakınlık merkeziliği örneği

Bu graftaki düğümler için yakınlık merkeziliği değerleri aşağıdaki gibi hesaplanır:

$$C_c(v_1) = 1/((1 + 2 + 2 + 3 + 4 + 5)/6) = 0,35 \quad (2.17)$$

$$C_c(v_2) = 1/((1 + 1 + 1 + 2 + 3 + 4)/6) = 0,5 \quad (2.18)$$

$$C_c(v_3) = C_c(v_4) = 1/((2 + 1 + 2 + 1 + 2 + 3)/6) = 0,55 \quad (2.19)$$

$$C_c(v_5) = 1/((3 + 2 + 1 + 1 + 1 + 2)/6) = 0,6 \quad (2.20)$$

$$C_c(v_6) = 1/((4 + 3 + 2 + 2 + 1 + 1)/6) = 0,46 \quad (2.21)$$

$$C_c(v_7) = 1/((5 + 4 + 3 + 3 + 2 + 1)/6) = 0,33 \quad (2.22)$$

Hesaplanan değerlere göre v_5 düğümü en yüksek yakınlık merkeziliğine sahiptir.

2.2.1.2. Kenar Ölçütleri

Bağ kuvveti (Tie Strength),

İki düğüm arasındaki bağ kuvveti ne kadar çok ortak komşuları olduğuna bağlıdır. A ve B düğümü ne kadar çok ortak komşuya sahiplerse aralarındaki bağ kuvveti de o kadar güçlü olur. A ve B düğümünün komşuluklarının örtüşmesi büyükse A ve B düğümleri birbirlerine güçlü bağlıdırlar denir. Aksi durumda da zayıf bir bağa sahiptirler denir. A ve B düğümü arasındaki bağ kuvvetini $S(A, B)$ ile gösterirsek,

$$S(A, B) = \frac{|n_A \cap n_B|}{|n_A \cup n_B|} \quad (2.23)$$

olur. Burada n_A ve n_B sırasıyla A ve B düğümlerinin komşularını göstermektedir. İki düğüm birbirlerine güçlü bir bağ kuvvetiyle bağlılarsa birbirlerine güvenmeleri daha kolaydır, çünkü dürüst olmayan bir davranışı ortak arkadaşları vasıtasıyla fark edebilirler. Bağ kuvvetinin sıfır olması iki düğümün hiçbir ortak komşusunun olmaması anlamına gelir. Bu nedenle bu tip iki düğümün birbirine güvenmesi risklidir, çünkü birbirlerinin davranışlarını doğrulayacak ortak arkadaşları yoktur [49].

Üçlü Yakınlık (Triadic Closure),

Bağ kuvvetinin bir sonucu olarak ortaya çıkan bir hipotezdir. A, B ve C gibi üç adet düğüm arasındaki bağların doğasıyla ilgilidir. Eğer A ve B arasında ve A ile C arasında güçlü bir bağ varsa o zaman B ile C arasında da güçlü bir bağ vardır denir. Benzer şekilde eğer A-B ve A-C aralarında zayıf bağ kuvvetine sahiplerse B ve C de zayıf bağa sahiptirler. Üçlü yakınlık ölçütü ağdaki katsayılar kümelenecek ölçülür. A düğümünün kümeleme katsayısı, A düğümünün rasgele seçilen iki arkadaşının birbiriyle arkadaş olma olasılığı olarak tanımlanır. Yani, A'nın birbirine bağlı arkadaş çiftlerinin kesiridir. Aslında bu da ağdaki üçgenleri sayma problemi ile ilgilidir. n_A ağdaki üçgenlerin sayısı ve $|E|$ de kenar sayısı olsun. Kümeleme katsayısı aşağıdaki gibi hesaplanır:

$$C = \frac{6n_{\Delta}}{|E|} \quad (2.24)$$

n_{Δ} üçgenlerin sayısını hesaplama işlemi temelinde pahalı bir işlemdir. n_{Δ} arasındaki bağlantı ve ağı özdeğerleri Tsourakakis tarafından bulunmuştur. Bu çalışmaya göre n_{Δ} , özdeğerlerin üçüncü adımına eşit olmaktadır ($\sum \lambda_i^3$, λ_i değeri i. özdeğer olmak üzere) [49].

Zayıf Bağ (Weak Tie),

A ve B düğümlerinin komşuluklarının kesişimi küçük olduğunda A-B arasındaki bağlantı zayıf bağ olarak kabul edilir. Hiçbir ortak komşu yoksa A-B bağlantısı yerel köprü olarak isimlendirilir. A-B bağlantısı yerel köprü ise, istisnai durumda, A-B düğümünü ağdan kaldırmak A ve B'nin bağlı olduğu bileşenlerin ağ ile bağlantılarının kopması şeklinde sonuçlanabilir. Böylesi bir durumda, A-B bağlantısı global köprü olarak kabul edilir. Ağda, global köprüler yerel köprülere göre daha az rastlanır fakat ikisinin de ağa etkisi neredeyse aynıdır [49].

Kenarın arada bulunma durumu (Edge Betweenness),

Önemli ölçütlerden bir tanesi de kenarın üzerinden geçen toplam akış miktarını ölçen kenarın arada bulunma durumudur. A ve B arasındaki bilgi akışının, A ve B arasındaki en kısa yollar aracılığıyla gerçekleştiği kabul edilir. Kenarın arada bulunma durumuna graf bölümlendirme örnek verilebilir. Graf bölümlendirme, ağı birbirinden kopmuş bileşenlere çevirmek için kademeli olarak yüksek arada bulunma durumu skoruna sahip kenarların kaldırılması işlemidir. Bu birbirinden kopan bileşenler ağdaki düğüm kümelerini oluşturur [49].

2.2.2. Yayılım Modelleri

İnsanlar çeşitli kanallar aracılığıyla birbirleriyle bilgi alış verişinde bulunurlar. Doğrudan arkadaşlarıyla bilgi paylaşabildikleri gibi yerel medya (TV, gazete vs.) gibi araçlarla da bilgi paylaşımında bulunabilirler. Bu bilgi akışına farklı araştırma alanları farklı bakış açılarıyla bakmaktadır. Sosyal ağlar açısından bilgi yayılımı, bir miktar bilginin yayılması ve bireylere ulaşması anlamına gelir. Bilgi yayılımının üç temel parçası vardır:

1. Gönderici(ler): Bilgi yayılımı işlemi başlatan bir veya birkaç kullanıcı.
2. Alıcı(lar): Yayılan bilgiyi alan bir veya birkaç kullanıcı. Genellikle, alıcı sayısı

gönderici sayısından fazladır ve göndericileri kapsayabilir.

3. Ortam: Yayılımın gerçekleştiği ortamdır. Örneğin, bir dedikodu yayılımındaki ortam bireyler arasındaki kişisel iletişim olabilir [48].

Online sosyal ağlar bireylerin diğer bireylerle etkileşim halinde buldukları ağlardır. Bu ağlar bazen gözlemlenemez veya kullanılamaz olsalar da bilgi yayılımı işlemi gerçekleşir. Bireyler bilginin akmasına karar vererek bilginin yayılımını kolaylaştırır. Örneğin, bir söylenti yayılırken, bireyler bu söylentiye komşularına yayma ya da yaymama kararını verirler. Bu kararı bağımsız olarak veya bağımlı (başkalarından aldıkları bilgiye bağlı) bir şekilde verebilirler. Bağımlı kararlar verdiklerinde, bireylerin birbirleri üzerindeki bağımlılık seviyelerini ölçmek önemlidir. Bu bağımlılık bireyin kararlarının kendi komşularına(arkadaşlarına) bağlı olduğu durumda yerel bağımlılık, karar vermeden önce ağdaki bütün bireyler gözlemleniyorsa global bağımlılıktır.

Literatürde bilgi yayılımını modellemek için kullanılan çeşitli teknikler mevcuttur. Aşağıda, bu modeller sırası ile incelenmiştir.

2.2.2.1. *Independent Cascade (IC) Modeli*

Etkileşimli parçacık sistemleri modelini ve market araştırmalarını temel alan IC modeli ilk olarak Kempe ve arkadaşları tarafından tanıtılmıştır [54]. Aynı zamanda epidemik modellerle de ilişkilidir. Modelin anahtar özelliği sosyal ağlardaki her kenar üzerindeki yayılım olayının karşılıklı olarak bağımsız olmasıdır.

Bu modelin varsayımları aşağıdaki şekilde özetlenebilir:

- Ağ yönlü bir graf ile gösterilir. Dğümler aktörleri ve kenarlar da bunlar arasındaki iletişim kanallarını gösterir. Her düğüm yalnızca bağlı olduğu düğümleri etkileyebilir.
- Kararlar binary'dir. Yani nodlar ya aktif ya da pasif olabilirler. Dğümün aktif olması demek davranışı, yeniliği veya kararı benimsediği anlamına gelir.
- Dğüm bir kere aktif olduktan sonra komşu düğümlerini aktive edebilir.
- Aktivasyon işlemi progresif bir işlemdir, yani düğümler pasiften aktif duruma geçebilirler fakat tersi söz konusu değildir.

Aktif düğümler gönderenler, aktive edilen düğümler de alıcılar olarak kabul edilirse, IC

modelde göndericiler alıcıları aktive etmektedirler. Bu nedenle, IC model gönderici-merkezli model olarak kabul edilir. Bu modelde t anında aktif olan düğümün bir sonraki $t + 1$ anında kendi komşularını aktive edebilmek için bir şansı vardır. t anında v düğümü aktif olsun. Herhangi bir w komşusu için, $t + 1$ anında w düğümünün aktif olabilme olasılığı $p_{v,w}$ 'dir. t anında aktif olan v düğümünün w komşu düğümünü aktif edebilmek için bir şansı vardır ve bu aktivasyon yalnızca $t + 1$ anında yapılabilir. Bir dizi aktif düğüm ile işleme başlanır ve yeni bir aktivasyon olmayana kadar devam edilir. Şekil 2.4'te IC modeli sözde kod olarak verilmiştir.

```

Gereksinim:  $G(V, E)$  yayılım grafi,  $A_0$  başlangıçta aktif olan düğümlerin dizisi,  $p_{v,w}$  aktivasyon olasılıkları
1: return Aktive olan düğümlerin final dizisi  $A_\infty$ 
2:  $i = 0$ ;
3: while  $A_i \neq \{\}$  do
4:  $i = i + 1$ ;
5:  $A_i = \{\}$ ;
6: for all  $v \in A_{i-1}$  do
7:   for all  $v$ 'nin  $w$  komşuları,  $w \notin \cup_{j=0}^i A_j$  do
8:      $rand = [0,1]$  aralığında rasgele bir sayı üret;
9:     if  $rand < p_{v,w}$  then
10:       $w$ 'yi aktive et
11:       $A_i = A_i \cup \{w\}$ ;
12:     end if;
13:   end for;
14: end for;
15: end while;
16:  $A_\infty = \cup_{j=0}^i A_j$ ;
17: Return  $A_\infty$ ;

```

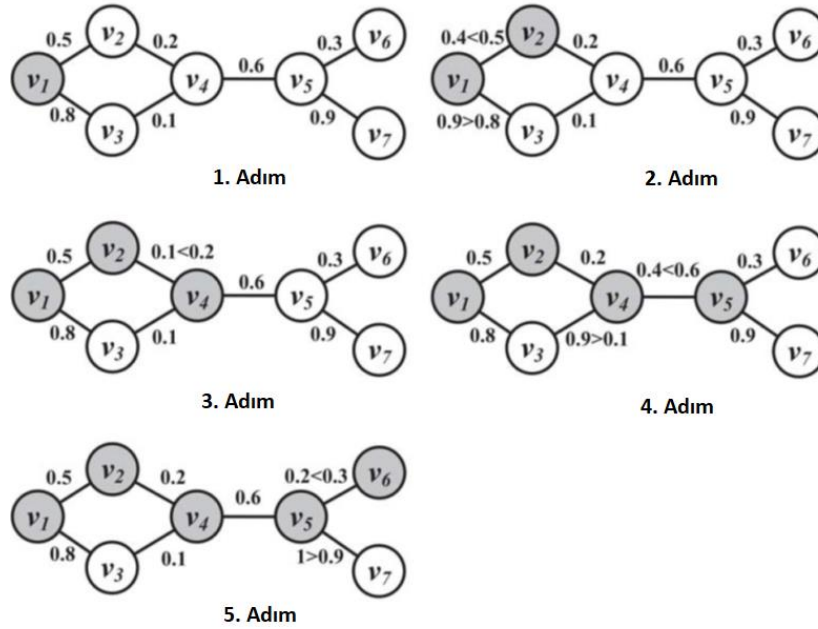
Şekil 2.4. Independent Cascade Modeli'nin sözde kodu [48].

Örnek olarak Şekil 2.5 ele alınsın. Ağ yönsüzdür, bu nedenle $p_{v,w} = p_{w,v}$ olur. Yönsüz olduğu için bir kenarla birbirine bağlı herhangi iki düğümün birbirlerini aktive etme şansları eşittir. Ağdaki 1. Adımı ele alalım. Kenarlar üzerindeki değerler $p_{v,w}$ 'ları göstermektedir. IC modeli prosedürü aktif bir dizi düğüm ile başlar. Örnekte başlangıçtaki aktif düğüm v_1 'dir. Aktive edilen her düğüm komşularını aktive etmek için yalnızca bir şansa sahiptir. Aktif hale gelen düğüm her komşusu için rasgele bir sayı üretir. Eğer bu üretilen rasgele sayı komşusunun $p_{v,w}$ 'sundan küçükse komşu aktive edilir. Şekilde soldaki rasgele üretilen sayı ve sağdaki ise $p_{v,w}$ 'yu temsil etmektedir. Bu prosedür izlenerek beş adım geçtikten sonra altı adet düğüm etkilenmiş ve IC modeli prosedürü sonlanmıştır.

Kenarlar üzerindeki sayılar $p_{v,w}$ ağırlıklarını göstermektedir. Eşitsizlik olması

durumunda aktivasyon şartı kontrol edilir. Soldaki sayı rasgele üretilen sayıyı, sağdaki ise $p_{v,w}$ ağırlığını göstermektedir.

IC modeli bilgi yayılımı işlemini karakterize etmektedir. Gönderici merkezlidir ve düğüm bir kere aktif olduktan sonra bütün komşu düğümlerini aktif etme amacındadır. IC modelinde düğüm aktivasyonu olasılıksal bir işlemdir. Bu nedenle farklı denemelerde farklı sonuçlar elde edilebilir. IC modeli ile ilgilenirken sorulabilecek önemli sorulardan biri verilen bir ağda sonuçta maksimum sayıda etkilenmiş düğüm elde edebilmek için başlangıçta hangi düğüm setini seçmemiz gerektiğidir.



Şekil 2.5. ICM (Independent Cascade Model) simülasyonu [48].

2.2.2.2. Linear Threshold (LT) Model

Etkiyi modellemede amaç bireylerin birbirlerini nasıl etkilediklerini açıklayan modeller tasarlamaktır. Sosyal medyanın doğası gereği, etkinin birbirine bağlı kullanıcılar arasında olduğunu varsaymak doğru olur. Bazı zamanlar ağ dışardan gözlemlenebilir (açık ağlar) ve bazen de gözlemlenemez (kapalı ağlar, implicit). Örneğin, sosyal medya üzerindeki bir online servise katılım için o ağa bağlı bir kişinin referansı gerekiyorsa bu tip ağlar gözlemlenebilir ağlardır. Hangi kullanıcının kim tarafından ağa katıldığı yani etkilendiği bilinmektedir. Tam tersi olan gözlemlenemez ağlarda insanlar ürünleri satın almak için etkilenebilirler ve çoğu durumda da satıcı alıcının kimden etkilendiğini bilmez, fakat zaman içinde ürün satışındaki artışı değerlendirir. Gözlemlenebilir ağlarda (açık ağlar, explicit) etkiyi modellemek için LT modeli gibi eşik modellerine

başvurulur. Kapalı ağlarda farklı zamanlarda etkilenen bireylerin sayısını giriş olarak alan (haftadaki müşteri sayısı gibi) LIM (Linear Influence Model) gibi metotlar çalıştırılabilir.

Eşik modelleri açık ağlarda etkinin modellenmesi için basit fakat etkilidir. Bu modellerde, düğümler komşularının(veya yönlü grafta gelen komşularının) sayısına veya fraksiyonuna (eşik) göre karar verirler. Eşik modelleri Granovetter ve Schelling'in çalışmalarında 1970'lerin başlarında kullanılmıştır [55], [56]. Schelling, eşik modelini kullanarak, aynı renge ve aynı tercihlere sahip olan küçük grupların varlığının, global ağda bir ırk ayırımı olduğunu göstermiştir [48]. LT modeli eşik modellerine bir örnektir. Bu modelde, v düğümü herhangi bir w komşusundan $b_{v,w}$ ağırlığı ile etkilenir ($\sum_{v' \text{nin komşusu } w} b_{v,w} \leq 1$). Her v düğümü $[0,1]$ aralığında rasgele bir değer olacak şekilde bir θ_v eşik değeri seçer. Eşik değeri v düğümünün aktif olabilmesi için, aktif durumda olması gereken komşularının ağırlığını gösterir. Rasgele seçilmiş eşik değerleri ve A_0 aktif düğümlerinin başlangıç seti (bütün diğer düğümler pasif) eşliğinde yayılım ayrık adımlarla deterministik olarak devam eder. t anında; $t - 1$ anında aktif olan bütün düğümler aktif olarak kalırlar ve herhangi bir v düğümünün aktive edilebilmesi için, v düğümünün bütün aktif komşularının ağırlıkları toplamının θ_v eşik değerine eşit veya büyük olması gerekir [3].

$$\sum_{v' \text{nin aktif komşusu } w} b_{v,w} \geq \theta_v \quad (2.25)$$

LT modelinin sözde kodu Şekil 2.6'da verilmiştir.

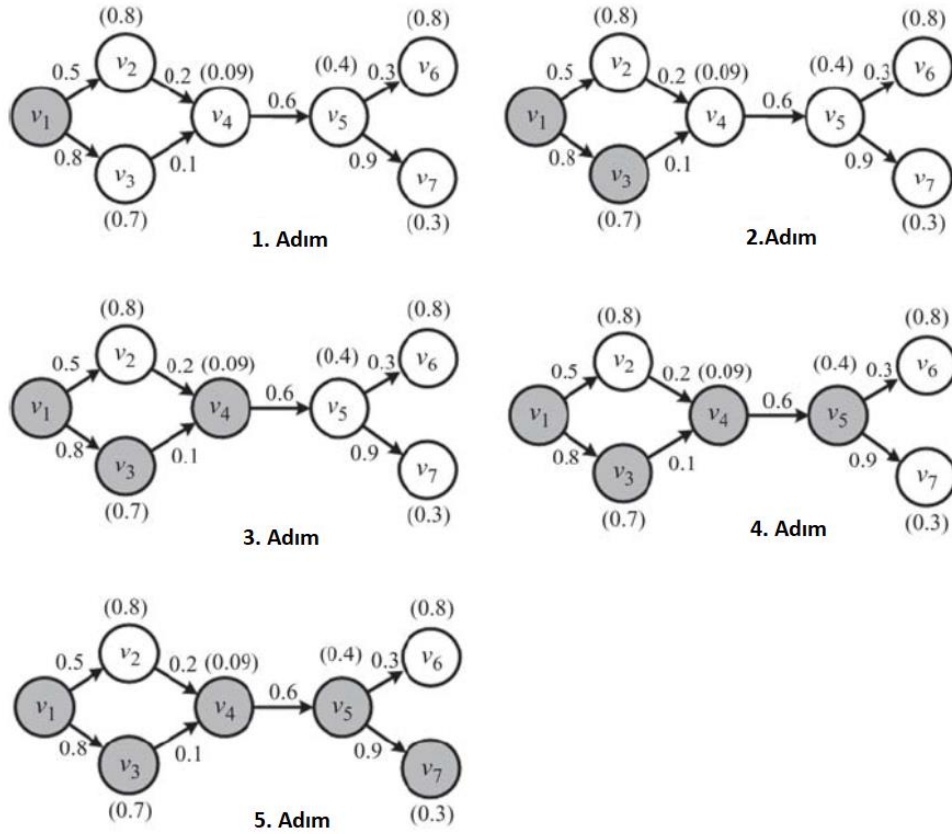
```

Gereksinim:  $G(V, E)$  grafi,  $A_0$  başlangıçta aktif olan düğümlerin dizisi
1: return Aktive olan düğümlerin final dizisi  $A_\infty$ 
2:  $i = 0$ ;
3:  $[0,1]$  aralığındaki rasgele  $\theta_v$  eşik değerlerini düzgün bir şekilde (uniformly) ata
4: while  $i = 0$  veya  $(A_{i-1} \neq A_i, i \geq 1)$  do
5:  $A_{i+1} = A_i$ 
6: aktifOlmayanlar =  $V - A_i$ ;
7: for all  $v \in$  aktifOlmayanlar do
8:   if  $\sum_{v' \text{ye bağılı } j' \text{ler, } j \in A_i} w_{j,v} \geq \theta_v$  then
9:      $v$ 'yi aktive et;
10:     $A_{i+1} = A_{i+1} \cup \{v\}$ ;
11:   end if
12: end for
13:  $i=i+1$ ;
14: end while
15:  $A_\infty = A_i$ ;
16: Return  $A_\infty$ ;

```

Şekil 2.6. Linear Threshold Model'in sözde kodu [48].

Şekil 2.7’de kenarlar üzerindeki değerler ağırlıkları, düğümler üzerindeki değerler ise eşik değerlerini göstermektedir. v_1 düğümü başlangıçta aktif düğüm olarak seçilmiştir. Düğümü etkileyen aktif komşuların ağırlıkları toplamı etkilenecek düğümün eşik değerinden büyük veya eşit ise düğüm etkilenmiş ve aktive olmuş aksi takdirde pasif olarak durumunu korumuştur. Şekilde $t = 0$ anında v_1 düğümü aktive edilmiştir. $t + 1$ anında hem v_2 hem de v_3 , v_1 düğümünden etki alır. v_2 düğümü $0,5 < 0,8$ olduğundan aktive olmazken, v_3 düğümü $0,8 > 0,7$ olduğundan aktive olur. Algoritma benzer şekilde devam eder ve sonunda beş adet düğüm aktive olur.



Şekil 2.7. LTM simülasyonu [21].

Düğümlerin üzerindeki etiketler θ_i eşik değerlerini; kenarlar üzerindeki değerler ise $w(i,j)$ ağırlıklarını göstermektedir [48].

2.2.2.3. Polarity Related Independent Cascade (IC-P) Yayılım Modeli

Li ve arkadaşları, IC modeli işaretli sosyal ağlarda kullanılmak üzere genişletmişler ve Polarity-related Independent Cascade (IC-P) yayılım modelini geliştirmişlerdir [37].

İşaretsiz sosyal ağlarda kullanılan IC modelinde bir düğüm aktif veya pasif olmak üzere iki durumdan birinde olmak zorundadır. Herhangi bir u düğümünün aktif olması, bu

düğümün ağda yayılan bir bilgiyi benimsemesi anlamına gelir. Pasif olması ise bilgiyi benimsememesi anlamındadır. Yayılım işlemi başlangıç aktif düğüm seti S ile başlar ve ayrık adımlarla devam eder. t adımında, $t - 1$ adımında aktif olan herhangi bir düğüm, pasif komşuları v 'yi etkileyebilmek için bir şansa sahiptir. Her düğüm diğerlerini etkileyebilmek için yalnızca bir şansa sahiptir ve $A_{u,v}$ olasılıkla u düğümü v düğümünü etkiler. u başarılı olursa $t + 1$ adımında v aktif olacaktır. u başarılı olsa da olmasa da, v 'yi etkileyebilmek için yeni girişimlerde bulunamaz. Yeni bir aktivasyon mümkün olmayıncaya kadar işlem devam eder. Eğer v düğümü bir zaman adımında birden fazla düğümü aktive ederse, bunların aktivasyon girişimleri keyfi bir sıralamada yapılır. IC-P yayılım modeli “düşmanımın dostu benim düşmanımdır” ve “düşmanımın düşmanı benim dostumdur” sosyal prensibine dayanır. İşaretli sosyal ağlardaki kullanıcılar arasındaki ilişkinin polaritesini birleştirir. IC modelinin işaretli sosyal ağlarda kullanılmak üzere genişletilmiş şeklidir. IC-P modelinde düğümlerin aktif durumları pozitif ve negatif durum olmak üzere ikiye ayrılmıştır. Bu nedenle, IC-P modelindeki her düğüm pozitif, negatif veya pasif olmak üzere üç duruma sahiptir. Bir u düğümü için pozitif durum, ilgili kullanıcının ağda yayılan bilgiyi kullanması ve sonra desteklemesi veya güvenmesi anlamına gelir. u düğümünün negatif durumunun anlamı, ilgili kullanıcının yayılan bilgiyi kullanması fakat ona karşı çıkması veya güvenmemesi anlamına gelir. u 'nun pasif durumda olması yayılan bilgiyi kullanmaması, onunla ilgilenmemesi anlamına gelir. U düğümünün durumu $S(u)$ ile gösterilsin. $1, -1, 0$ değerleri ise sırasıyla pozitif, negatif ve pasif durumu göstermek için kullanılır [37].

IC-P modelinde yayılım işlemi bir dizi aktif düğüm seti S ile başlar. S hem pozitif hem de negatif düğümleri içerir. S içinde olmayan diğer bütün düğümler sosyal ağ grafında pasif durumdadırlar. İşlem ayrık adımlarda gerçekleşir. $t - 1$ zaman adımında aktive edilen bir u düğümü, t zaman adımında pozitif veya negatif durumda olacaktır. Daha sonra bu u düğümü mevcut pasif v komşularını t zaman adımında aktive etmek için sadece bir şansa sahiptir. t anında pozitif veya negatif olan v düğümünün komşu seti $N_{aktif}^t(v)$ olsun. t anında, $u \in N_{aktif}^t$ olmak üzere her düğüm keyfi bir sıralamada v düğümünü $A_{v,u}$ olasılığıyla aktive eder. v düğümü N_{aktif}^t içindeki bir düğüm tarafından bir kere aktive edildikten sonra, $N_{aktif}^t(v)$ içindeki diğer düğümler artık bu düğümü aktive edemezler. IC modelinden farklı olarak v düğümü bir zaman anında yalnızca bir kere aktive edilebilir.

Yeni aktive edilen v düğümünün $S(v)$ durumu, v düğümünü aktive eden u düğümünün durumuyla ve u ile v düğümü arasındaki ilişkinin polaritesiyle ilişkilidir ve $S(v) = S(u) \times P_{u,v}$ şeklinde ifade edilir. Bu nedenle,

- Eğer u düğümü pozitifse ve u ile v arasındaki ilişki de pozitifse, v düğümü pozitif olacaktır.
- Eğer u düğümü negatifse ve u ile v arasındaki ilişki de pozitifse, v düğümü negatif olacaktır.
- Eğer u düğümü pozitifse ve u ile v arasındaki ilişki de negatifse, v düğümü negatif olacaktır.
- Eğer u düğümü negatifse ve u ile v arasındaki ilişki de negatifse, v düğümü pozitif olacaktır [37].

Bir düğüm pozitif veya negatif olduktan sonra durumunu bir daha deęiřtirmez. Yeni aktive edilen düğüm kalmayıncaya kadar iřlem devam eder.

2.2.2.4. Etki Yayılım Olasılıęı Modelleri

Tekdüze Ayarlama (Uniform Setting, US),

Aę hakkında çok fazla bilgi sahibi olunmadıęı durumlarda tekdüze ayarlama kullanılır. Her kenar için etkileme olasılıęı $P(u, v) = 0.01$ olarak tanımlanır.

Aęırlıklandırılmıř Kademeli Ayarlama (Weighted Cascade Setting, WCS),

$\text{deg}(v)$, v düğümüne baęlı düğümlerin sayısı olmak üzere u düğümünün v düğümünü etkileme olasılıęı $P(u, v) = 1/\text{deg}(v)$ olarak hesaplanır. Tekdüze ayarlama yüksek dereceye sahip düğümler çok etkili olmaktadır. Aęırlıklandırılmıř kademeli ayarlama ise derece ile yayılım olasılıęını dengeler [2].

3. ETKİ MAKSİMİZASYONU VE SÜRÜ ZEKÂSI

Bu başlık altında etki maksimizasyonu problemi tanımlanacaktır. Etki maksimizasyonunda kullanılan sürü zekâsı optimizasyon yöntemlerinin detayları verilecektir.

3.1. ETKİ MAKSİMİZASYONU PROBLEMİ

Etki Maksimizasyonu problemi, bir etkinin bir ağda bulunan elemanlar arasında yayılımının eniyilenmesi olarak tanımlanabilir. Sosyal ağlar açısından düşünülecek olunursa problem şu şekilde tarif edilebilir: İstenen bir etkinin bir sosyal ağ üzerindeki en fazla sayıdaki bireye ulaştırılması için en etkili k adet bireyin seçilmesi [3]. Eğer, seçilmiş olan k adet birey aktif edilirse, istenen etki bu k adet bireyden başlayarak ağ üzerinde yayılacaktır. Benzer şekilde, yayılması istenmeyen bir etki için eğer bu k adet bireye “bağışıklık kazandırılırsa” (istenmeyen etkiyi yaymaları engellenirse), geniş çaplı yanlış etki yayılımının önüne geçilebilir [57].

Etki Maksimizasyonu probleminin ele alınabilmesi için ilk olarak bireylerin birbirlerini etkileme olasılıklarının belirlenmesi ve etki yayılımının modellenmesi gerekmektedir. Bir sosyal ağ, yönlü ve ağırlıklı bir graf olarak modellenir. Graftaki düğümler bireyleri, kenarlar bireyler arasındaki etkileşimi ve etkileşimin yönünü, ağırlıklar ise bireylerin komşularından etkilenme olasılıklarını gösterir.

Bir bireyin, komşu bireyden etkilenme olasılığını belirlemek için kabul görmüş temel iki yaklaşım Weighted Cascade Setting (WCS) ve Uniform Setting (US) [54]. WCS’de bir düğümün komşusundan etkilenme olasılığı, o düğümün giriş derecesi ile ters orantılıdır. Bir düğümün giriş derecesi, bu düğüme gelen yönlü kenarların sayısıdır. Şöyle ki; u ve v düğümleri komşu düğümler olsun. u düğümünden v düğümüne yönlü bir kenar olması şartı ile; v düğümünün giriş derecesi $deg_{in}(v)$ olarak ifade edilirse, u düğümünün v düğümünü etkileme olasılığı $1/deg_{in}(v)$ olarak belirlenir. US’te ise bir düğümün diğer düğümü etkileme olasılığı 0.01 gibi sabit bir değer olarak belirlenir.

Etki yayılımının modellenmesi için literatürde kabul görmüş temel iki model Independent Cascade (IC) model ve Linear Threshold (LT) modeldir [54]. IC modeli şu şekilde çalışmaktadır: u düğümünün v düğümünü etkileyip etkilemediğini belirlemek için rasgele bir sayı üretilir. Eğer bu sayı, u düğümünün v düğümünü etkileme olasılığından küçükse, v düğümü etkilenir ve aktif hale gelir. Aktif hale gelen v düğümü, kendi komşularını etkileyebilir. LT modeli ise şu şekilde çalışmaktadır: bir v düğümünün etkilenebilmesi için belirli bir eşik değer vardır. Eğer v düğümünün aktif olan (etkilenmiş olan) komşularının v düğümünü etkileme olasılıklarının toplamı bu eşik değerden yüksekse v düğümü etkilenir ve aktif hale gelir. Aktif hale gelen v düğümü, kendi komşularının etkilenmesine katkıda bulunur.

Bireylerin birbirlerini etkileme olasılıklarının belirlenmesi ve etki yayılımının modellenmesinden sonra, en etkili k adet bireyin seçimi yapılabilir. $k=1$ durumu için düşünürsek, en etkin birey aday olarak bir düğümün seçilmesi/aktif kabul edilmesi ve etkisinin modellenerek kaç bireyi etkilediğinin tespit edilmesi işlemleri yapılır. Etkinin yayılımı olasılıksal olduğundan, aday olarak seçilen bireyin etkisinin modellenmesi defalarca (en az birkaç bin kez) yapılır ve bütün modellemelerde etkilenen birey sayısının ortalaması alınarak, seçilen adayın etkisi hesaplanır [38]. Problem, $k>1$ durumları için kombinatoriyal eniyileme problemidir ve kaba kuvvet yaklaşımı ile çözümü çok uzun süreler almaktadır.

Bu çalışmada yayılım modeli olarak literatürde sıklıkla kullanılan IC model kullanılmıştır. IC modelde, 0 ve 1 arasında değişen bir etkileme olasılığı vardır, $P(e) = [0,1]$. Bu olasılık bir e kenarıyla ilişkilendirilir, $e = (u, v)$. Burada, u ve v keyfi iki komşu düğümdür. Sosyal bir ağ yönlü bir graf ile gösterilir, $G = (V, E)$. Burada $V = \{i_1 \dots i_n\}$ bireyler kümesi, $E = \{e_1 \dots e_m\}$ ise bireyler arasındaki ilişkiler yani kenarlar kümesidir. n toplam birey yani düğüm sayısı iken, m graftaki toplam ilişki yani kenar sayısıdır. Örnek bir sosyal graf Şekil 3.1’de verilmiştir.

EM probleminde amaç en fazla sayıda bireyi etkileyecek k adet çekirdek bireyi tespit etmektir. $S = \{s_1 \dots s_k\}$ kümesi çekirdek düğümleri içeren küme ve $A = \{a_1 \dots a_m\}$ kümesi de k adet çekirdek düğümden etkilenen bireylerin kümesi olsun. Dolayısıyla burada $S \subseteq A$ olur. Sonuç olarak, problem basitçe şu şekilde ifade edilebilir: Denklem (3.1)’deki amaç fonksiyonunu maksimize edebilmek için hangi k bireyler çekirdek düğüm olarak seçilmelidir? Bu problem kombinatoriyal bir optimizasyon problemidir ve büyük n and m değerleri için makul bir süre içerisinde kaba kuvvet yoluyla çözülemez.

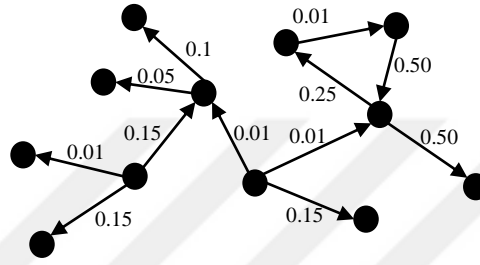
Bu nedenle sürü zekâsı algoritmalarını kullanmak problemin çözümüne katkı sağlayacaktır.

$$f_1(S) = |A \setminus S| \quad (3.1)$$

Problem minimizasyon problemi olarak açıklanacak olunursa amaç fonksiyonu aşağıdaki gibi yazılabilir:

$$f_2(S) = |V| - |A \setminus S| \quad (3.2)$$

Denklem (3.2)'nin amacı ağdaki etkilenmeyen birey sayısını minimize etmektir.



Şekil 3.1. Örnek sosyal graf.

3.2. SÜRÜ ZEKÂSİ OPTİMİZASYONU

Sürü zekâsı algoritmaları gerçek yaşamdaki sürü davranışlarını modeller. Sürünün avını arama, izleme ve yakalamadaki becerisini optimizasyon problemlerinde kullanmayı hedefler. Bu çalışmada Bozkurt ve Balina Optimizasyon algoritmaları kullanılmıştır.

3.2.1. Bozkurt Optimizasyon Algoritması (Grey Wolf Optimizer, GWO)

GWO algoritması, yeni geliştirilen sürü zekâsı algoritmalarından olup, birçok mühendislik problemine başarı ile uygulanmıştır [58]–[66]. GWO algoritması bozkurtların avlanma davranışları temel alınarak geliştirilmiştir [7]. Bozkurtlarda sıkı bir hiyerarşi vardır. Bu hiyerarşide 4 tip kurt vardır: alfa, beta, delta, omega. Hiyerarşideki en üst seviyede Alfa kurtları yer alırlar. Bunlar bir dişi bir de erkek kurttan oluşur ve gruptaki baskın kurtlardır. Alfa avlanma, uyunacak yer, yürüme zamanı ile ilgili kararları verir. Bazı durumlarda diğer kurtları izleyebilir. Diğer kurtlar Alfa'yı izlemek durumundadır. Beta kurtları ikinci seviyede yer alırlar. Alfa kurtlarına karar vermelerinde yardımcı olurlar. Dişi veya erkek olabilirler ve alfa kurdu öldüğünde veya çok yaşlandığında yerine geçmeye aday en iyi kurtlardır. Alfa kurtlarının emirlerini diğer kurtlara iletirler ve alfa'ya geri bildirimde bulunurlar. Üçüncü seviyede

delta kurtları bulunur. Alfa ve Beta kurtlarına boyun eğerler fakat omega kurtlarına karşı baskındırlar. Delta kendi içinde izci, nöbetçi, yaşlı, avcı ve bakıcı olmak üzere gruplara ayrılır. İzci kurtların görevleri mıntıkanın sınırlarını izlemek ve tehlike durumunda sürüyü uyardır. Nöbetçi kurtlar sürüyü korumakla görevlendirilmişlerdir. Yaşlı kurtlar daha önceden alfa veya beta olmuş olan kurtlardır. Avcılar alfa ve beta kurtlarına avlanmada yardım etmekle görevlidir. Bakıcı kurtlar ise sürüdeki hasta, zayıf ve yaralı kurtlara bakmakla görevlidirler. Son seviyede bulunan omega kurtları diğer bütün kurtlara boyun eğmek zorundadırlar. Sürüdeki en önemsiz kurtlar olarak görülürler ve avın yenme sırasında sonda yer alırlar. Omega kurtları olmadığında bütün diğer kurtlar birbiriyle kavga etmektedirler [7].

Her aday çözüm, birer avdır. Alfa, beta, delta ve omega ajanları popülasyonun çözüm uzayını ararlar. En iyi çözüm alfa olarak ele alınır. İkinci ve üçüncü en iyi çözüm ise sırasıyla beta ve delta olarak kabul edilir. Kalan bütün diğer çözümler omegadır. GWO algoritmasında ajan alfa, beta ve deltayı temsilen α , β ve δ olarak gösterilir. Omega ise geriye kalan tüm çözümlerdir ve ω olarak gösterilir. Her iterasyonda kurtlar avı çevrelerler. Sonuçta, kurtlar ava doğru harekete geçerler ve en iyi aday, çözüm olarak seçilir.

Avı çevreleme davranışının matematiksel modeli aşağıdaki gibidir [7]:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad (3.3)$$

$$\vec{X}(t + 1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (3.4)$$

Burada t mevcut iterasyonu, \vec{A} ve \vec{C} katsayı vektörlerini, \vec{X}_p avın pozisyon vektörünü ve \vec{X} bozkurtun pozisyon vektörünü gösterir. \vec{A} ve \vec{C} vektörleri aşağıdaki şekilde hesaplanır [7]:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 \vec{a} \quad (3.5)$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad (3.6)$$

\vec{a} 'nın bileşenleri iterasyon boyunca 2'den 0'a doğru doğrusal olarak azalır. \vec{r}_1 ve \vec{r}_2 vektörleri [0,1] aralığında rasgele vektörlerdir.

Avlanma işlemi genelde alfa tarafından yönlendirilir. Beta ve delta da avlanmaya zamanla katılabilir. Matematiksel modelde alfa, beta ve deltanın potansiyel en iyi

çözüm hakkında daha çok bilgisi olduğu kabul edilmiştir. En iyi üç çözüm saklanır ve diğer adımlarda ajanlar en iyi çözümlere göre pozisyonlarını ayarlarlar [7].

Alfa hareket etmediğinde avlanma işlemi biter. A vektörü $[-2a, 2a]$ aralığında rasgele bir değerdir. $|A| < 1$ olduğunda kurtlar ava doğru saldırırlar. Buna sömürme işlemi denir. $|A| > 1$ olduğunda kurtlar daha iyi bir av bulmak üzere avdan uzaklaşırlar. GWO'nun sözde kodu Şekil 3.2'de verilmiştir [7].

```
Gereksinim:  $X_i, (i = 1, 2, \dots, n)$  bozkurt popülasyonunu,  $a, A$  ve  $C$ 'yi başlat  
1: Her arama ajanı için amaç fonksiyonunu hesapla  
2:  $X_\alpha = \text{en iyi arama ajanı}$   
3:  $X_\beta = \text{en iyi ikinci arama ajanı}$   
4:  $X_\delta = \text{en iyi üçüncü arama ajanı}$   
5: while( $t < \text{maksimum iterasyon sayısı}$ ) do  
6:   for all arama ajanı  
7:     mevcut arama ajanının pozisyonunu Denklem (3.4)'ü kullanarak güncelle  
8:   end for  
9:    $a, A$  ve  $C$ 'yi güncelle  
10:  Bütün ajanlar için amaç fonksiyonunu hesapla  
11:   $X_\alpha, X_\beta$  ve  $X_\delta$ 'yi güncelle  
12:   $t = t + 1$ ;  
13: end while  
14: Return  $X_\alpha$ 
```

Şekil 3.2. GWO'nun sözde kodu [7].

GWO hızlı ve yeni bir algoritmadır. Orijinal GWO, sürekli optimizasyon problemleri için tasarlanmıştır. Yani durum uzayından ürettiği aday çözümler reel sayılardan oluşur. Ancak, etkin bireylerin tespiti problemindeki durum uzayı tam sayılardan oluşur. Algoritmanın seçmesi gereken şeyler, tamsayı olarak numaralandırılmış düğümlerdir. Bu nedenle GWO tamsayı çözümler üretecek şekilde değiştirilmiştir. Böylece algoritma her aşamada tamsayı çözümler üretir. GWO üzerinde bu değişikliklerin dışında başka bir değişiklik yapılmamıştır.

3.2.2. Balina Optimizasyon Algoritması (Whale Optimization Algorithm, WOA)

WOA da, GWO gibi yeni geliştirilmiş olan sürü zekâsı algoritmalarındandır [8]. Çok parametrelili optimizasyon problemlerine başarı ile uygulanmaktadır [67]–[74]. Kambur balinalar avı arar, çevreler ve avın etrafında sarmal şekilde hava kabarcıkları oluştururlar. Sonra yüzeye doğru yüzerler, avlarını yüzeye çıkmaya zorlarlar. WOA kambur balinaların bu davranışını taklit eder. GWO gibi WOA da optimizasyon problemleri için tasarlanmıştır. WOA algoritması bir dizi rasgele çözümle başlar. Her bir yinelemede, arama ajanları rasgele seçilen bir arama ajanına veya şimdiye kadar elde

edilen en iyi çözümlere göre konumlarını günceller [8].

Kambur balinalar avın konumunu belirleyip onu çevreleyebilirler. Optimal çözümün konumun arama yüzeyinde nerede bulunduğu önceden bilinmediğinden dolayı, WO algoritması mevcut en iyi çözüm adayını hedef av olarak ya da ona yakın bir optimal çözüm olarak kabul eder. En iyi arama ajanı belirlendikten sonra, diğer ajanlar kendi konumlarını bu en iyi ajana doğru gidecek şekilde güncellerler. Bu davranış, aşağıdaki formüller ile ifade edilebilir [8].

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (3.7)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (3.8)$$

Burada t , mevcut iterasyonu; \vec{A} ve \vec{D} , katsayı vektörlerini; X^* , şimdiye kadar elde edilmiş en iyi çözümün pozisyon vektörünü; \vec{X} , pozisyon vektörünü; $|\cdot|$, mutlak değeri; \cdot ise vektörlerin eleman eleman çarpımını ifade etmektedir.

\vec{A} ve \vec{C} vektörleri şu şekilde hesaplanmaktadır [8]:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (3.9)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (3.10)$$

Burada \vec{a} , 2'den 0'a doğrusal olarak her bir iterasyonda azalan bir vektör; \vec{r} , [0,1] arasında rasgele sayılardan oluşan vektördür.

Kambur balinaların kabarcık ağı oluşturma davranışlarını modellemek için 2 aşamalı bir yaklaşım kullanılabilir [8].

1. Küçülen sarma mekanizması

Bu davranış, Denklem (3.9)'daki \vec{a} değerinin düşürülmesi ile elde edilir. \vec{A} vektörünün değer değişim aralığının \vec{a} kullanılarak azaltıldığına dikkat etmek gerekir. Bir başka deyişle \vec{A} vektörü, $[-a, a]$ aralığında rasgele değerlerden oluşur. Burada a , her bir iterasyonda 2'den 0'a doğru azalan bir sayıdır.

2. Sarmal pozisyon güncelleme

İlk olarak (X, Y) pozisyonunda bulunan balina ile (X^*, Y^*) pozisyonunda bulunan av arasındaki mesafe hesaplanır. Daha sonra, kambur balinaların sarmal şeklindeki hareketini modellemek için balina ile avın pozisyonları arasındaki mesafe aşağıdaki gibi bir

denklem oluşturulur [8]:

$$\vec{X}(t + 1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (3.11)$$

Burada $\vec{D}' = |\vec{X}^*(t) - \vec{X}(t)|$ olarak yazılır ve i. balinanın ava (şimdiye kadarki en iyi çözüme) uzaklığını ifade eder. b, logaritmik sarmalın şeklini tanımlamak için kullanılan bir sabit; l, $[-1,1]$ aralığında rasgele bir sayı; \cdot ise vektörlerin eleman eleman çarpımını ifade etmektedir.

Kambur balinaların avları etrafında daralan çemberlerden oluşan sarmal şeklinde hareket ettiklerine dikkat etmek gerekir. Optimizasyon esnasında bu eş zamanlı davranışları modellemek için, küçülen sarma mekanizması ile sarmal pozisyon güncelleme mekanizmasının gerçekleşme olasılıkları %50 olarak alınır. Bunun matematiksel modeli aşağıdaki gibidir [8]:

$$\vec{X}(t + 1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0,5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0,5 \end{cases} \quad (3.12)$$

Burada p, $[0,1]$ aralığında rasgele bir sayıdır.

\vec{A} vektörünü değiştirmek için kullanılan mekanizmanın bir benzeri, av arama (keşif) için de kullanılabilir. Gerçekte, kambur balinalar diğer kambur balinaların pozisyonlarına göre rasgele arama yaparlar. Bundan dolayı, atılım aşamasının aksine, keşif aşamasında ajanların pozisyonları o ana kadarki en iyi çözümün pozisyonuna göre değil rasgele olarak güncellenir. Bu mekanizma ile WOA'nın global arama yapabilmesi sağlanır. Bu durumun matematiksel modeli şöyledir [8]:

$$\vec{D} = |\vec{C} \cdot \vec{X}_{\text{rand}} - \vec{X}| \quad (3.13)$$

$$\vec{X}(t + 1) = \vec{X}_{\text{rand}} - \vec{A} \cdot \vec{D} \quad (3.14)$$

Burada \vec{X}_{rand} , mevcut popülasyondan seçilen bir rasgele pozisyon vektörüdür (rasgele bir balina). WOA'nın sözde kodu Şekil 3.3'de verilmiştir.

Orijinal WOA, sürekli optimizasyon problemleri için tasarlanmıştır. Yani durum uzayından ürettiği aday çözümler reel sayılardan oluşur. Ancak, etkin bireylerin tespiti problemindeki durum uzayı tam sayılardan oluşur. Algoritmanın seçmesi gereken şeyler, tamsayı olarak numaralandırılmış düğümlerdir. Bu nedenle WOA tamsayı

çözümler üretecek şekilde değiştirildi. Böylece algoritma her aşamada tamsayı çözümler üretti. WOA üzerinde bu değişikliklerin dışında başka bir değişiklik yapılmamıştır.

```
Gereksinim:  $X_i, (i = 1, 2, \dots, n)$  balina popülasyonunu başlat,  
1: Her arama ajanı için amaç fonksiyonunu hesapla  
2:  $X^* = \text{en iyi arama ajanı}$   
3: while ( $t < \text{maksimum iterasyon sayısı}$ ) do  
4:   for all arama ajanları  
5:      $a, A, C, l$  ve  $p$ 'yi güncelle  
6:     if1 ( $p < 0.5$ )  
7:       if2 ( $|A| < 1$ )  
8:         Denklem (3.7)'ye göre mevcut arama ajanının pozisyonunu güncelle  
9:       else if2 ( $|A| \geq 1$ )  
10:        Rasgele arama ajanı ( $X_{rand}$ )'ı seç  
11:        Denklem (3.14)'e göre mevcut arama ajanının pozisyonunu güncelle  
12:       end if2  
13:     else if1 ( $p \geq 0,5$ )  
14:       Denklem (3.11)'e göre mevcut arama ajanının pozisyonunu güncelle  
15:     end if1  
16:   end for  
17: Herhangi bir arama ajanının arama alanının ötesine geçip geçmediğini kontrol et ve düzelt  
18: Her arama ajanı için amaç fonksiyonunu hesapla  
19: Eğer daha iyi bir çözüm varsa  $X^*$ 'yi güncelle  
20:  $t = t + 1$ ;  
21:end while  
22: Return  $X^*$ 
```

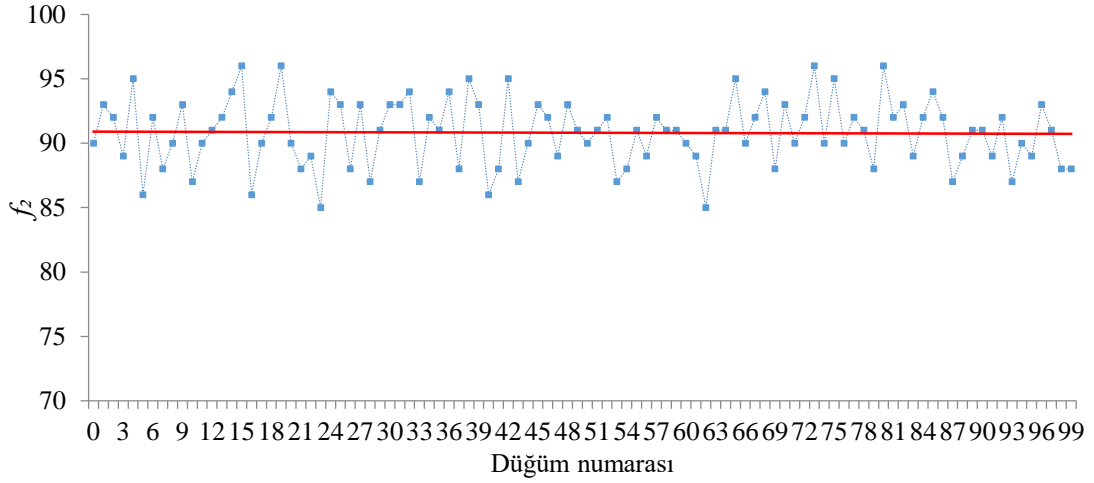
Şekil 3.3. WOA'nın sözde kodu [8].

4. ÖNERİLEN YÖNTEM

4.1. PROBLEMİN MEVCUT DURUMU

Öncelikle $k = 1$ için küçük bir graf üzerinde EM problemi değerlendirilmiştir. Şekil 4.1'de, $k = 1$ için EM problemi 100 düğümlü bir Erdős-Renyi grafı için modellenmiştir. Dikey eksen düğümleri göstermektedir. Yatay eksen ilgili düğümleri çekirdek olarak seçtiğimizde elde edilen f_2 değerlerini göstermektedir. Modellemede IC kullanılmıştır. Bireylerin birbirlerini etkileme olasılıkları (etkinin yayılma olasılığı, graftaki kenarların ağırlıkları) uniform ayara göre $P(e) = 0.01$ olarak alınır. Eğer, eldeki sosyal ağ ile ilgili yeteri kadar bilgiye sahip değilsek bu ayar kullanılır [2]. Erdős-Renyi grafı rasgele bir graf olduğu için bu tip graflar ile birlikte uniform ayar kullanılabilir. Daha önce de bahsedildiği gibi EM problemi ayrık kombinatoriyal optimizasyon problemidir. Şekil 4.1'de gösterilen değerler ayrık değerlerdir. Bir an için bu değerlerin oluşturduğu eğri bir yüzey gibi düşünülürse, bu yüzey üzerinde genel bir eğim olmadığı görülür. Örneğin, 63 numaralı düğüm etkin birey olarak iyi bir seçimken, 62 numaralı birey çok kötü bir seçim olacaktır. Genel olarak bir sürü zekâsı algoritması, bir durum uzayı meydana getiren değişkenlerin farklı değerleri için bir amaç fonksiyonunu maksimize ya da minimize etmeye çalışır. Bunu yaparken, bir sonraki iyi çözümün mevcut iyi çözümün yakınında olduğunu varsayar. Ancak, mevcut problemde bir sürü zekâsı algoritması komşu durumları kontrol ederek global minimum (ya da en azından yakınına) yakınsayamaz. Sonuç olarak, bu tip problemler için ABC, GWO gibi sürü zekâsı algoritmalarının, durum uzayı üzerindeki dolaşmaları, eğer algoritmaların yeni çözüm üretme mekanizmalarına dışarıdan bir müdahale yapılmamışsa rasgele salınımdan öteye geçmeyecektir.

Bu durumda sürü zekâsı algoritmaları global olarak yakınsayamazlar ve yerel optimumlarda dolaşırlar.



Şekil 4.1. Erdős-Rényi Grafi üzerinde örnek EM problemi– Normal Graf.

4.2. SÜRÜ ZEKÂSI ALGORİTMALARINI YEREL OPTİMUMLARDAN KURTARMA

Sürü zekâsı algoritmaları çok parametrelili optimizasyon problemlerine başarı ile uygulanabilirler. Ne var ki, sürü zekâsı algoritmalarının erken yakınsama ve zayıf yerel arama kabiliyetleri gibi bazı dezavantajları vardır [75]. Son yıllarda sürü zekâsı algoritmalarının bu dezavantajlarını ortadan kaldırmak için bazı hibrit yaklaşımlar geliştirilmiştir. Bu yaklaşımlardan bazıları Nelder-Mead ve BOBYQA (Bound Optimization BY Quadratic Approximation) gibi yerel arama yöntemlerinin sürü zekâsı algoritmaları ile birlikte kullanılmasını önermektedir [76], [77]. Nelder-Mead yöntemi yerel aramada iyidir, ancak bu algoritmanın yakınsaması başlanılan noktaya son derece hassas biçimde bağımlıdır. Sürü zekâsı algoritmaları global aramada iyidirler. Ne var ki, bu algoritmalarda yüksek hesaplama gücü istemektedirler [78]. Benzer şekilde BOBYQA yerel arama için sıklıkla kullanılan bir başka algoritmadır. Bahsedilen bu hibrit tekniklerin amacı global seviyedeki yakınsama için sürü zekâsı algoritmalarının kullanılması ve yerel aramalar için de Nelder-Mead ve BOBYQA gibi algoritmaları kullanmaktır. Ne var ki, bu hibrit yöntemler de sürü zekâsı algoritmalarının EM problemi için global yakınsamasını iyileştirmez. Çünkü sürü zekâsı algoritmaları global arama yeteneklerini kullanabilmek için problemin durum uzayı yüzeyi üzerinde genel bir eğimin bulunmasına ihtiyaç duyarlar.

4.3. ÖNERİLEN YAKLAŞIM: EM PROBLEMİNDE AMAÇ FONKSİYONUNUN DURUM UZAYINA GENEL BİR EĞİM VERMEK

Daha önce bahsedildiği gibi sürü zekâsı algoritmaları mevcut durumdan bir sonraki iyi duruma geçebilmek için yüzey eğimini takip ederler. Bu algoritmalar bir sonraki iyi çözümün durum uzayı üzerindeki mevcut durumun yakınında olduğunu varsayar. Ne varki EM probleminde durum farklıdır. EM problemini sürü zekâsı algoritmaları için uygun hale getirmek amacıyla kalite açısından benzer olan çözümleri durum uzayı üzerinde birbirlerine yaklaştırmak gerekir. EM probleminde durum uzayı yüzeyini düğümler oluşturur. Böylece eğer düğümleri etki seviyelerine göre sıralayıp bu sıraya göre düğümler yeniden numaralandırılırsa EM probleminin durum uzayı yüzeyinde genel bir eğim elde edilir.

EM probleminde bir düğümün etki seviyesini tahmin etmek için kullanılan bazı ölçütler vardır. Eğer düğümleri bu ölçütlerden biri ya da birkaçına göre küçükten büyüğe (ya da büyükten küçüğe) doğru sıralanıp, düğümler buna göre yeniden numaralandırılırsa yukarıda bahsedilen eğim verme işlemi gerçekleştirilmiş olur.

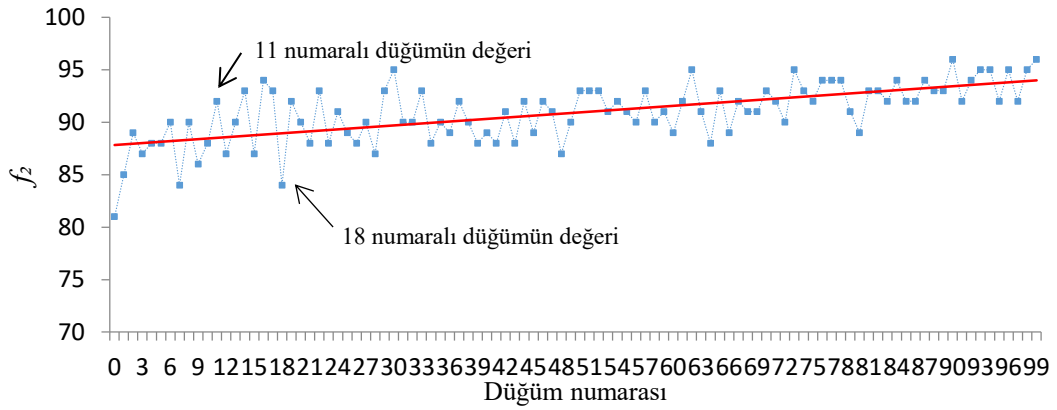
Bu ölçütlerden sık kullanılan iki tanesi çıkış derecesi merkeziliği ve yakınlık merkeziliğidir. Çıkış derecesi merkeziliği yönlü bir graftaki bir düğümden çıkan kenarların sayısıdır. Çıkış derecesi merkeziliği C_{outdeg} , bir yönlü graftaki u_j düğümü için $C_{outdeg}(u_j) = d_j^{out}$ olarak gösterilir. Yakınlık merkeziliği bir graf üzerindeki bir düğümün diğer bütün düğümlere olan en kısa mesafelerinin ortalaması olarak tanımlanabilir. Biçimsel olarak Denklem (4.1)'deki gibi ifade edilebilir.

$$C_c(u_j) = \frac{1}{\bar{l}_{u_j}} \quad (4.1)$$

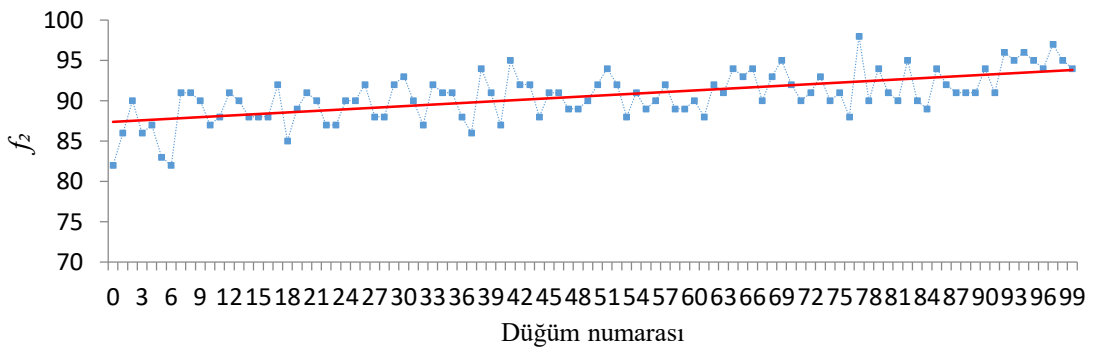
$$\bar{l}_{u_j} = \frac{1}{n-1} \sum_{u_j \neq u_m} l_{j,m} \quad (4.2)$$

Denklem (4.2)'de bulunan l ifadesi, u_j ve u_m arasındaki en kısa yoldur [79]. Daha büyük C_{outdeg} veya C_c değeri bir düğümün diğerlerinden daha etkin olduğunu gösteren bir işarettir. Bu yolla kullanılan Erdős-Renyi grafindaki düğümler çıkış derecelerine göre büyükten küçüğe doğru sıralanıp yayılım modeli tekrarlanırsa Şekil 4.2'deki gibi bir eğim elde edilir. Benzer şekilde bu graftaki düğümler yakınlık merkeziliklerine göre büyükten küçüğe göre sıralanırsa Şekil 4.3 elde edilir. Bu yolla elde edilen grafların

yapıları orijinal graflar ile tamamen aynıdır; yalnızca düğümler yeniden numaralandırılmıştır. Şimdi sürü zekâsı algoritmaları bu grafların düğümlerinin oluşturduğu durum uzayı yüzeyinde hareket edebilir. Burada şunu belirtmekte fayda vardır, en etkili k adet bireyin seçilmesi yeniden şekillendirilmiş graftaki ilk k adet bireyin seçilmesi gibi basit bir işlem değildir. Şekil 4.2 ve Şekil 4.3'den görülebileceği gibi düğümleri yeniden numaralandırılmış graf üzerindeki ilk 10 birey en etkili 10 birey değildir. Elbette hesaplanan ölçütler ve düğümlerin etkileri arasında bir ilişki vardır ancak bu ilişki doğrusal değildir. Örneğin Şekil 4.2'de $C_{\text{outdeg}}(11) > C_{\text{outdeg}}(18)$ (11 numaralı düğüm 18 numaralı düğümden daha büyük çıkış derecesine sahiptir.) olmasına rağmen $f_2(11) > f_2(18)$. Bu problem bir minimizasyon problemi olarak ele alınmıştır, dolayısıyla daha düşük f_2 değerleri daha iyidir. Ek olarak bir düğüm tarafından etkilenen bireylerin sayısı yalnızca bu düğümün çekirdek olarak seçildiği durumda, bu düğümün başka bireylerle birlikte çekirdek olarak seçilmesi durumlarında farklı olabilir. Bu durum alt modülerlik olarak tarif edilmektedir [3].



Şekil 4.2. Yeniden şekillendirilmiş Erdős-Renyi grafi üzerinde EM-Çıkış Derecesi Merkeziliği.



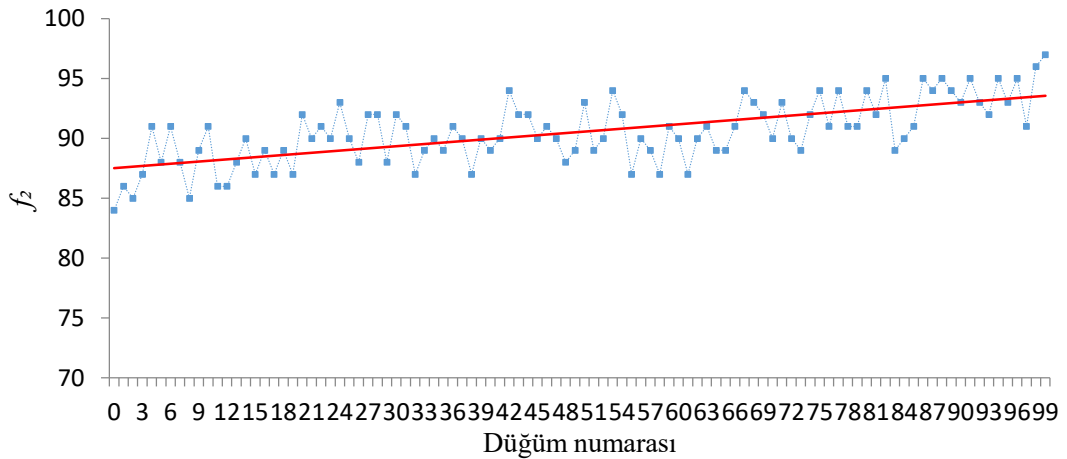
Şekil 4.3. Yeniden şekillendirilmiş Erdős-Renyi grafi üzerinde EM - Yakınlık Merkeziliği.

4.3.1. Merkezilik Ölçütlerinden Karma Bir Ölçüt Elde Etme

Literatürde merkezilik ölçütleri genelde tek başlarına kullanılırlar. Ancak bir düğümün etkinlik seviyesini yalnızca bir ölçüte göre belirlemek çok sağlıklı bir durum değildir. Birden fazla ölçüt tek bir ölçüt çatısı altında birleştirilerek daha iyi sonuçlar elde edilebilir. Bu durum önerilen yönteme uyarlanacak olunursa, düğümler yalnızca bir ölçüt yerine birden fazla ölçütten oluşmuş karma bir ölçüte göre de sıralanabilir. Birden fazla ölçütü birleştirmek için yapılması gereken ilk iş hangi ölçütün ne kadar önemli olduğunu tespit etmektir. Karmaşıklıktan uzak olması için birlikte kullanılan ölçütlerin eşit derecede önemli olduğu varsayılmıştır. $M = \{\mu_0 \dots, \mu_N\}$ ölçüt kümesi olsun. Herbir ölçüt için belirlediğimiz ağırlıklar (önem) $(0,1)^N$ 'dan oluşan $W = \{w_0 \dots, w_N\}$ olarak gösterilsin. Bu şekilde oluşturulan bileşik ölçüt aşağıdaki gibi gösterilebilir:

$$\mu_{\text{combined}} = W \cdot M \quad (4.3)$$

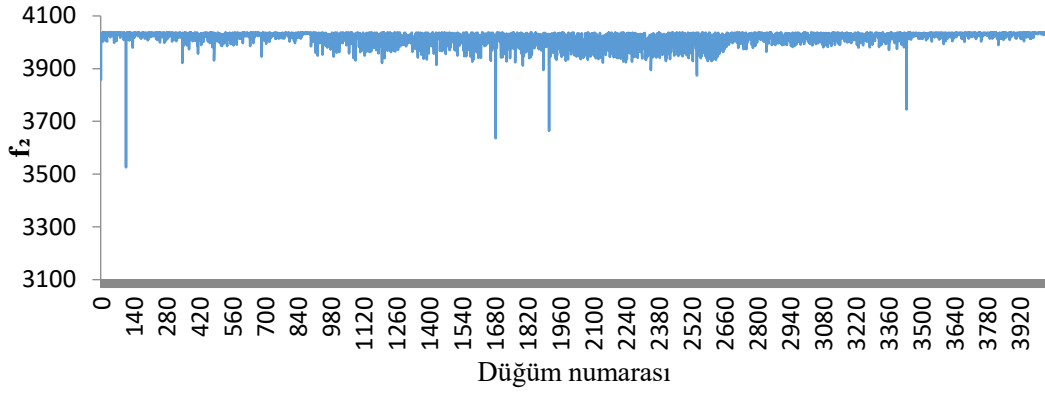
Eğer bir ölçütün diğerinden daha önemli olup olmadığını bilmiyorsak bütün ölçütler için aynı ağırlık kullanılabilir. Örneğin iki ölçüt $\mu_0 = C_{\text{outdeg}}$ ve $\mu_1 = C_c$; bu ölçütler için belirlenen ağırlıklar $W = \{0.5, 0.5\}$ olsun. Elde edilen sonuç $\mu_{\text{combined1}} = 0.5 \times C_{\text{outdeg}} + 0.5 \times C_c$ olur. Kullanılan Erdős-Renyi grafindaki düğümler $\mu_{\text{combined1}}$ ölçütüne göre yeniden numaralandırılırsa Şekil 4.4 elde edilir. Şekilden de görülebileceği gibi yine genel bir eğim elde edilmiş olur.



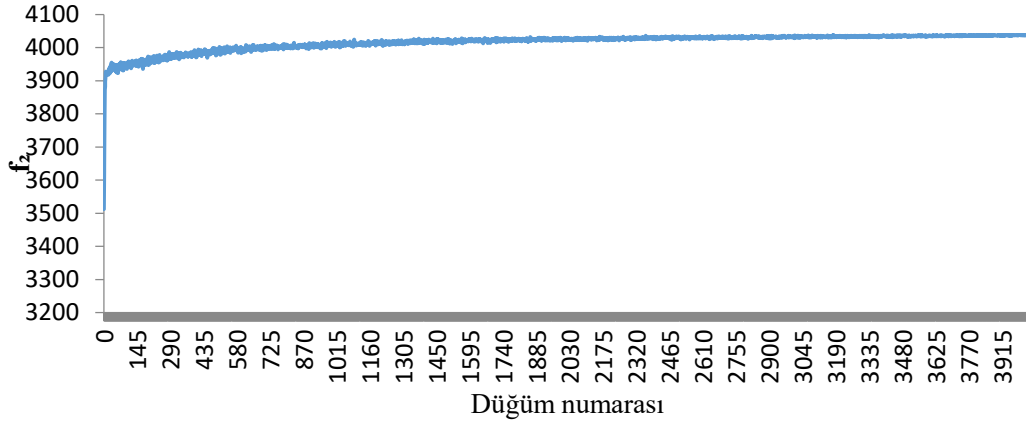
Şekil 4.4. Yeniden şekillendirilmiş Erdős-Renyi grafi üzerinde EM – Birleştirilmiş Ölçüt.

Gerçek veriseti olarak Stanford Üniversitesi Geniş Ağ VeriSeti koleksiyonundan (<http://snap.stanford.edu>) Facebook sosyal ağ grafi kullanılmıştır [12],[13]. Facebook

grafı 4.039 düğüm ve 88.234 yönsüz kenardan oluşmaktadır. İlk olarak yönsüz kenarların herbiri iki yönlü kenarla değiştirilmiş ve böylece toplamda 176.468 yönlü kenar elde edilmiştir. Facebook grafiındaki düğümler C_{outdeg} , C_c ve $\mu_{combined}$ ölçütlerine göre sıralanarak 3 yeni graf elde edilmiştir. Erdős-Renyi grafında kullanılan $k = 1$ için EM problemi modellemesi Facebook grafı için de gerçekleştirilmiştir. Buradaki amaç, gerçek bir graf üzerinde beklenildiği gibi bir eğim oluşup oluşmadığını görmektir. Şekil 4.5’de görüldüğü gibi Facebook grafı üzerinde de bir eğim oluşmuştur.

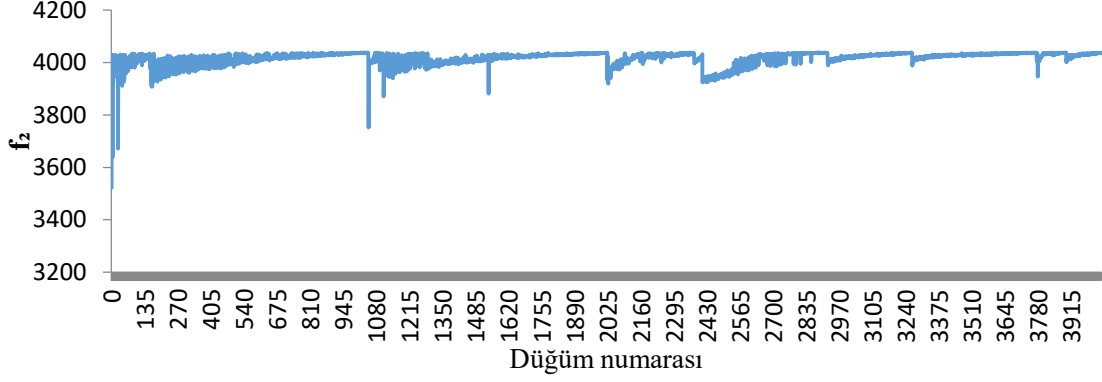


(a)

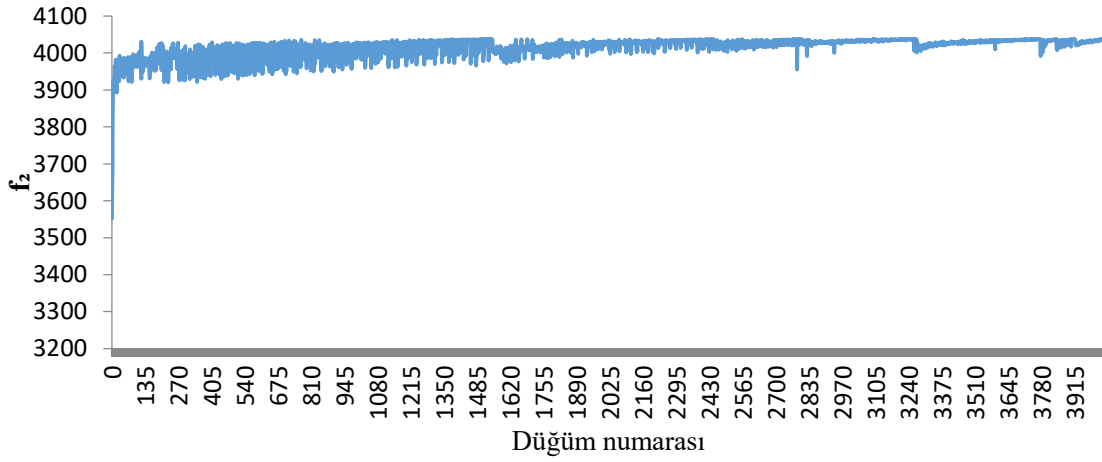


(b)

Şekil 4.5. Eğim oluşturulan Facebook grafı. (a) Normal Facebook grafı, (b) Çıkış Derecesi Merkeziliği ölçütüne göre yeniden numaralandırılan graf, (c) Yakınlık Merkeziliği ölçütüne göre yeniden numaralandırılan graf, (d) Birleştirilmiş ölçüt ölçütüne göre yeniden numaralandırılmış graf.



(c)



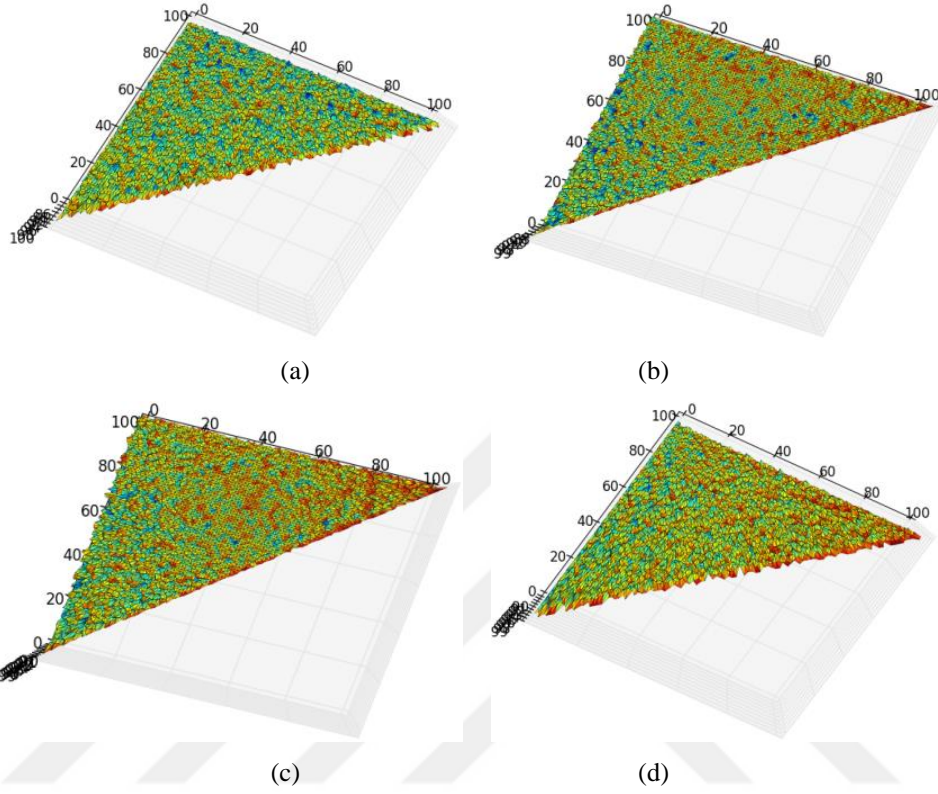
(d)

Şekil 4.5. (devam). Eğitim oluşturulan Facebook grafi. (a) Normal Facebook grafi, (b) Çıkış Derecesi Merkeziliği ölçütüne göre yeniden numaralandırılan graf, (c) Yakınlık Merkeziliği ölçütüne göre yeniden numaralandırılan graf, (d) Birleştirilmiş ölçüt ölçütüne göre yeniden numaralandırılmış graf.

4.3.2. İki Çekirdek Düğüm İçin Etki Maksimizasyonu Problemi

Problemin durum uzayını 3 boyutlu olarak görmek için aynı Erdős-Renyi grafi üzerinde problem $k=2$ için ele alınmıştır. Orijinal graf normal graf olarak isimlendirilmiştir. Normal graf için ele edilen grafik Şekil 4.6-a'da, çıkış derecesine göre yeniden şekillendirilmiş graf için elde edilen grafik Şekil 4.6-b'de ve yakınlık merkeziliğine göre yeniden şekillendirilmiş graf için elde edilen grafik Şekil 4.6-c'de gösterilmektedir. Şekil 4.6'da mavi renk çukurları (minimumları) ve kahverengi renk tepeleri (maksimumları) göstermektedir. Şeklin x ve y ekseni düğüm numaraları, z ekseni ise f_2 değerleridir. Şekillerden de görülebileceği gibi minimum noktaları normal graf üzerinde rasgele bir dağılım göstermektedir. Buna karşın minimum noktaları yeniden

şekillendirilmiş graflar üzerinde belirli bölgelerde toplanmıştır. Bu, amaç fonksiyonunun durum uzayı yüzeyi üzerinde belirli bir eğim elde edildiğini gösterir. Böylece bir sürü zekâsı algoritması bu yüzey üzerinde hareket edebilir.



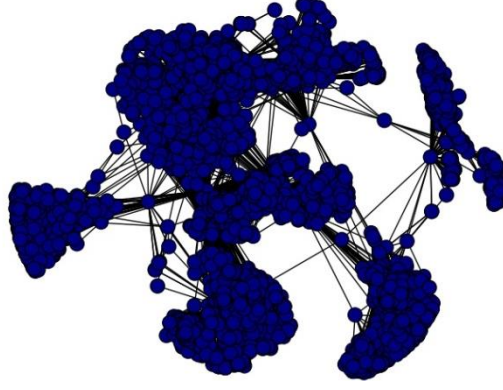
Şekil 4.6. Erdős-Rényi Grafı üzerinde EM: (a) normal graf, (b) çıkış derecesi merkeziliğine göre yeniden şekillendirilmiş graf ve (c) yakınlık merkeziliği ölçütüne göre yeniden şekillendirilmiş graf, (d) birleştirilmiş merkeziliğe göre yeniden şekillendirilmiş graf.

5. DENEYSEL SONUÇLAR

Önerilen yöntemin geçerliliğini göstermek için SNAP VeriSeti koleksiyonundan Facebook ve Epinions sosyal ağ grafları üzerinde denemeler gerçekleştirilmiştir [12],[13]. Facebook grafi 4.039 düğüm ve 88.234 yönsüz kenardan oluşmaktadır. İlk olarak yönsüz kenarların herbiri iki yönlü kenarla değiştirilmiş ve böylece toplamda 176.468 yönlü kenar elde edilmiştir. Facebook grafinin bir görüntüsü Şekil 5.1'de verilmiştir. Facebook grafi, yalnızca bireylerin arkadaşlık ilişkilerini göstermektedir. Facebook Grafi bağlı bir graftır. Bütün düğümlerden diğer bütün düğümlere ulaşılabilir. Ortalama Kümeleme Katsayısı (clustering coefficient) 0,6055 ve üçgenlerin sayısı 1.612.010'dir. Facebook grafinin çapı (en uzun en kısa yol) 8'dir.

Kullanılan ikinci graf olan Epinions grafi 131.828 düğüm ve 841.372 yönlü kenardan oluşmaktadır. Epinions Grafi bağlı bir graf değildir. Bağlantısız alt grafları vardır. Bağlı en büyük alt grafindaki düğüm sayısı 119.130'dur. Ortalama Kümeleme Katsayısı (clustering coefficient) 0,1279 ve üçgenlerin sayısı 4.910.076'dır. grafinin çapı (en uzun en kısa yol) 14'dür. Epinions grafi, Facebook grafindan farklı olarak, bireylerin birbirlerini olumlu ya da olumsuz etkilediklerine ilişkin bilgi de barındırmaktadır. Bundan dolayı, bu tip graflar işaretli graf olarak anılmakta ve EM problemi, işaretli EM problemi olarak ele alınmaktadır. İşaretli EM probleminde, genel olarak olumlu etkinin maksimize edilmesi amaçlanır. Bu bölümde ilk olarak Facebook grafi ile gerçekleştirilen denemeler verilmiş, ardından Epinions grafi ile gerçekleştirilen işaretli EM denemeleri ele alınmıştır.

Tez kapsamında geliştirilen bütün uygulamalar Python dili ile kodlanmış ve graf fonksiyonları için Networkx aracı kullanılmıştır [80].



Şekil 5.1. Facebook ağının görüntüsü.

Bir sosyal ağ üzerinde bireylerin birbirlerini etkileme olasılıklarını modellemek amacıyla iki popüler yaklaşım kullanılır. Bunlardan ilki Uniform Setting (US) diğeri Weighted Cascade Setting (WCS)'dir. Daha önce bahsedildiği gibi US'de $P(e) = 0.01$ kullanılır. WCS'de v düğümünün u düğümü tarafından etkilenme olasılığı $P((u,v)) = 1/\text{deg}(v)$ 'dir. Burada $\text{deg}(v)$, v düğümünün derecesidir. Daha gerçekçi sonuçlar elde edebilmek için Facebook ve Epinions grafi üzerinde yapılan denemelerde WCS kullanılmıştır [2], [3].

Facebook grafi için yapılan denemeler iki aşamada ele alınmıştır. İlk aşamada GWO ve rasgele seçim yöntemi kıyaslanmıştır. Böylece graf üzerinde yapılan yeniden şekillendirme işleminin iyi sonuç verdiği gösterilmiştir. İkinci aşamada literatürde karşılaştırma amacıyla sık kullanılan Kempe'nin açgözlü algoritması ve PageRank algoritmaları ile GWO ve bir diğeri sürü zekâsı algoritması olan WOA kıyaslanmıştır. Kıyaslanan sürü zekâsı algoritmaları ve diğeri algoritmalarından aşağıda kısaca bahsedilmiştir.

5.1. KIYASLAMADA KULLANILAN ALGORİTMALAR

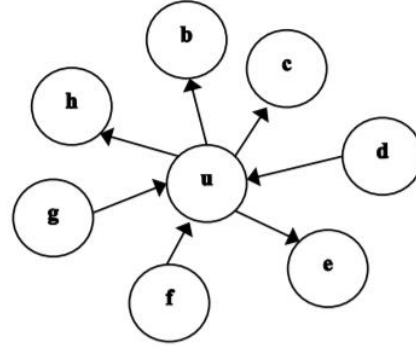
Daha önce bahsedilen sürü zekâsı algoritmaları ile kıyaslama yapmak için literatürde sıklıkla kullanılan PageRank ve Kempe'nin açgözlü algoritması kullanılmıştır.

5.1.1. PageRank

PageRank, Google şirketinin kurucuları olan Larry Page ve Sergey Brin ile arkadaşları tarafından geliştirilmiş olan bir algoritmadır[81]. Esasen, web sayfalarının ne kadar önemli olduklarını göstermek için geliştirilmiştir. Bunun yanı sıra, genel olarak bir ağ üzerindeki bir düğümün ne kadar önemli olduğunu tespit etmek için de kullanılabilir.

Güncel çalışmalarda, sosyal ağlardaki etkin bireylerin tespiti için kullanılmaktadır [82].

A_u , u düğümüne yönelen kenarların kaynak düğümlerinin kümesi, B_u , u 'dan diğer düğümlere yönelen kenarların hedef düğümlerinin kümesi olsun. Ör: Şekil 5.2'deki graf için $A_u = \{d, f, g\}$ ve $B_u = \{b, c, e, h\}$ olur.



Şekil 5.2. PageRank için örnek graf.

Bir düğümün PageRank değeri şu şekilde hesaplanır [81]:

- Başlangıçta bütün düğümlerin PageRank değeri “1/toplam düğüm sayısı” olarak alınır (örnek için 1/8).
- Bütün düğümler için aşağıdaki formül ile PageRank değerleri hesaplanır. Bu işlem 40 kez (daha hassas sonuçlar için daha fazla da olabilir) tekrar edilir.
- Sonuçta elde edilen PageRank değerleri kullanılır.
- Formüldeki df (damping factor) varsayılan olarak 0,85 alınır.

$$PageRank(u) = (1 - df) \times 0,1 + df \times \left(\sum_{v \in A_u} \left(\frac{PageRank(v)}{|B_v|} \right) \right) \quad (5.1)$$

5.1.2. Kempe'nin Açgözlü Algoritması

EM probleminde çok önemli bir algoritmadır, kilometre taşı denilebilir. Çekirdek düğümler seçilirken kullanılır.

Çalışma mantığı [3]:

- Daha önce seçilmemiş olan düğümlerden yayılımı maksimize edeni seç
- Bir önceki adımı istenen çekirdek sayısına ulaşana kadar tekrarla

Dezavantajı her adımda bütün düğümler için yayılım modellenmek zorundadır.

5.2. FACEBOOK GRAFI ÜZERİNDE YAPILAN GWO VE RASGELE SEÇİM DENEMELERİ

Denemeler için GWO'nun <http://www.alimirjalili.com/GWO.html> sitesinden indirilen Python kodları kullanılmıştır. Kodlar üzerinde 3.2.1 ve 3.2.2'de bahsedilen değişiklikler gerçekleştirildikten sonra amaç fonksiyonu olarak $f_2(S)$ kullanılmıştır.

Bu denemelerin amacı yeniden şekillendirme işleminin işe yarayıp yaramadığını görmektir. Çekirdek seçme işlemi hem orijinal graf üzerinde hem de yeniden şekillendirilmiş graflar üzerinde yapılmıştır. Önerilen yöntemin yeniden şekillendirilmiş graflar üzerinde anlamlı bir fark yaratıp yaratmadığını görmek için kıyas yöntemi olarak rasgele çekirdek seçimi yapılmıştır. Her bir denemede GWO ve rasgele seçim işlemleri 30'ar defa çalıştırılmıştır. Her bir çalıştırma 50 iterasyon içermektedir. Sonuçta GWO ve rasgele seçim yönteminin verdiği en iyi sonuçlar kıyaslama için kullanılmıştır. Yöntemler $k = 1$, $k = 2$, $k = 5$, $k = 10$, $k = 15$, $k = 20$, ve $k = 30$ çekirdek sayıları için denenmiştir. Her bir denemede elde edilen sonuçlar Şekil 5.3'de sırasıyla verilmektedir.

Problemi yeniden şekillendirme yaklaşımının temel motivasyonu problemin durum uzayı üzerinde bir eğim elde etmek ve böylece sürü zekâsı algoritmalarının bu yüzey üzerinde hareket edebilmelerini sağlamaktır. Graflar yeniden şekillendirildiğinde kalite bakımından birbirine yakın olan çözümler durum uzayı yüzeyi üzerinde de birbirine yakınlaştırılmış olur. Böyle bir ilişki problemin orjinal halindeki çözümler arasında yoktur. Dahası, bir sürü zekâsı algoritmasındaki ajanlar bir sonraki en iyi çözüm için mevcut en iyi çözümün etrafını ararlar. Sürü zekâsı algoritmalarından beklenen şey ajanların komşu çözümler üzerinde hareket ederek optimal çözüme ulaşmalarıdır. Açıklanan bu nedenlerden dolayı eğer yeniden şekillendirme işlemi işe yararsa sürü zekâsı algoritmalarının yeniden şekillendirilmiş graflar üzerinde normal graflara göre daha iyi sonuçlar vermesi gerekir.

DeneySEL sonuçlara göre $k = 1$ ve $k = 2$ durumları için GWO ve rasgele seçim yöntemleri arasında anlamlı bir fark yoktur. Böyle bir graf üzerinde düşük çekirdek sayılarında problem nispeten basit sayılabilir. Daha yüksek çekirdek sayıları için yapılan denemelerde GWO ve rasgele seçim yönteminin performans farkları daha bariz olarak görülmektedir. Özellikle $k = 15$, $k = 20$, $k = 30$ durumları incelendiğinde (Şekil 5.3 e-f-g), yeniden şekillendirme işleminin işe yaradığı açıkça görülmektedir.

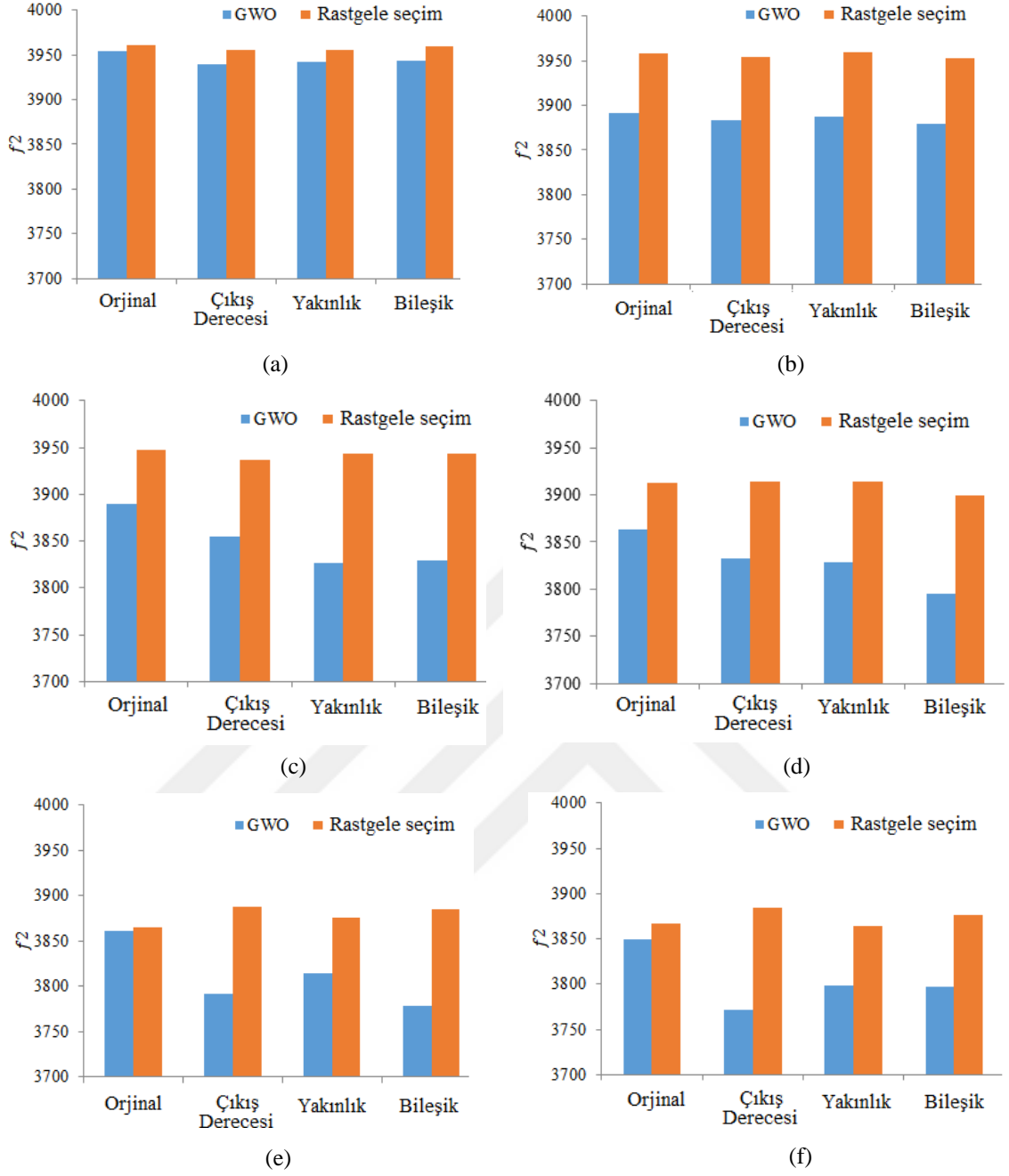
Şekil 5.3 e-f-g’de normal graf üzerinde GWO ve rasgele seçim yöntemleri çok yakın sonuçlar vermiştir. Diğer taraftan çıkış derecesi, yakınlık merkeziliği ve birleştirilmiş merkezilik ölçütlerine göre yeniden şekillendirilmiş graflar üzerinde GWO çok daha iyi sonuçlar vermiştir.

Seçilecek çekirdek sayısı arttıkça problem daha zor hale gelir. Daha zor durumlar için GWO’nun performansındaki anlamlı artış yeniden şekillendirme işleminin işe yaradığının bir göstergesidir. Sonuç olarak GWO algoritması yeniden şekillendirilmiş problem üzerinde iyi sonuçlar verirken orjinal problem üzerindeki performansı rasgele seçim yönteminden pek de farklı olmamıştır.

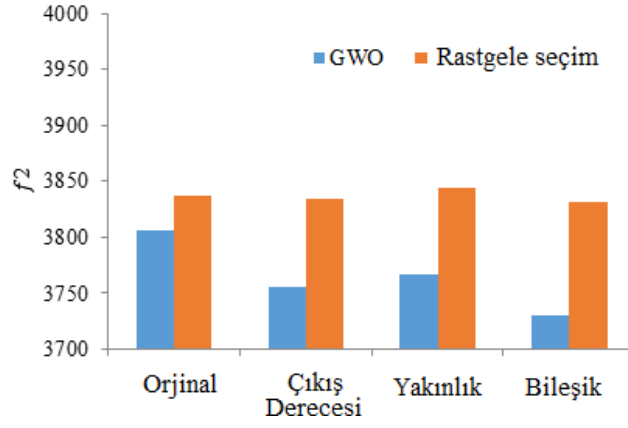
Önerilen ek bir yaklaşım olan karma ölçüt üretme ve grafi buna göre yeniden şekillendirme işleminin işe yaradığı söylenebilir. Şekil 5.3 incelendiğinde, GWO’nun karma ölçüte göre yeniden şekillendirilmiş graf üzerindeki performansının çıkış derecesi merkeziliği ve yakınlık merkeziliği derecelerine göre ayrı ayrı yeniden şekillendirilmiş graflar üzerindeki performansından daha iyi olduğu görülebilir. Buradan, eğer birden fazla ölçüt uygun bir şekilde birleştirilerek bir karma ölçüt elde edilirse bu ölçütün bir bireyin etkinlik düzeyini karma ölçütü oluşturan ölçütlere göre daha iyi temsil ettiği söylenebilir. Şekil 5.3’de görüldüğü gibi önerilen yaklaşım sürü zekâsı algoritmalarının EM problemlerinde kullanılmasını ve başarılı sonuçlar elde edilmesini sağlamıştır. Aynı zamanda, sonuçların istatistiksel olarak anlamlı olup olmadığını görmek için bazı testler uygulanmıştır. Eğer önerilen yeniden şekillendirme işlemi başarılıysa GWO’nun yeniden şekillendirilmiş graflar üzerindeki performansı normal graf üzerindeki performansından anlamlı düzeyde farklı olmalıdır. $P_{reshaped}^{GWO}$ GWO’nun yeniden şekillendirilmiş graflar üzerindeki performansı; $P_{regular}^{GWO}$ GWO’nun normal graflar üzerindeki performansı olsun. Hipotez şu şekilde kurulur ve anlamlılık derecesi belirlenir.

$$H_0: \left| P_{yenidenŞekillenmişGraf}^{GWO} - P_{normalGraf}^{GWO} \right| \text{ sıfırdır,}$$

$$H_1: \left| P_{yenidenŞekillenmişGraf}^{GWO} - P_{normalGraf}^{GWO} \right| \text{ pozitiftir, } \alpha = 0.05$$



Şekil 5.3. Farklı sayıdaki çekirdekler için Facebook grafi üzerinde GWO ve rasgele seçim metotlarının deneysel sonuçları: (a) k=1, (b) k=2, (c) k=5, (d) k=10, (e) k=15, (f) k=20, ve (g) k=30.



(g)

Şekil 5.3. (devam). Farklı sayıdaki çekirdekler için Facebook grafi üzerinde GWO ve rasgele seçim metotlarının deneysel sonuçları: (a) $k=1$, (b) $k=2$, (c) $k=5$, (d) $k=10$, (e) $k=15$, (f) $k=20$, and (g) $k=30$.

Ön test (yeniden şekillendirme yapmadan önce) ile son test(yeniden şekillendirme yaptıktan sonra) arasındaki farkı gözlemlemek için örneklem büyüklüğünün <30 olduğu durumlarda kullanılan parametrik olmayan Wilcoxon İşaretli Sıra Testi kullanılmıştır. Farklı çekirdek sayıları $k = 1, k = 2, k = 5, k = 10, k = 15, k = 20, ve k = 30$ hesaba katıldığında yedi adet örneklem olmaktadır. GWO için elde edilen test istatistikleri Çizelge 5.1’de verilmiştir. Çizelgede, GWO2 yeniden şekillendirilmiş graf üzerindeki GWO performansını, GWO1 ise normal graf üzerindeki GWO performansını göstermektedir. 10, 15, 20 ve 30 için Wilcoxon İşaretli Sıra Testi analiz sonuçlarına göre yeniden şekillendirilmiş graf üzerindeki GWO’nun performansı normal graftakine göre belirgin şekilde daha iyidir ($p = .018, p < \alpha$) ve sonuç olarak H_0 hipotezi reddedilir.

Diğer taraftan, yeniden şekillendirilmiş graflar ve normal graf üzerindeki rasgele seçim metodunun performansı neredeyse aynı olmalıdır. $P_{reshaped}^R$ rasgele seçimin yeniden şekillendirilmiş graflar üzerindeki performansı ve $P_{regular}^R$ rasgele seçimin normal graflar üzerindeki performansı olsun. Hipotez şu şekilde kurulur ve anlamlılık derecesi belirlenir.

$$H_0: |P_{yenidenŞekillenmişGraf}^R - P_{normalGraf}^R| \text{ sıfırdır,}$$

$$H_1: |P_{yenidenŞekillenmişGraf}^R - P_{normalGraf}^R| \text{ pozitiftir, } \alpha = 0.05$$

Rasgele seçim ile elde edilen test istatistikleri Çizelge 5.2’de gösterilmiştir. Çizelgede, Random2 yeniden şekillendirilmiş graflar üzerindeki rasgele seçim performansını ve

Random1 normal graf üzerindeki rasgele seçim performansını göstermektedir. Hiçbir çekirdek sayısı için rasgele seçim yönteminin yeniden şekillendirilmiş graflar ve normal graf üzerindeki performansı arasında istatistiksel olarak anlamlı bir fark gözlemlenmemiştir ($p = .735$, $p > .05$), sonuç olarak H_0 hipotezi kabul edilir.

Çizelge 5.1. GWO için test istatistikleri.

	GWO2 - GWO1
Z	-2.366
Asymp. Sig. (2-tailed)	.018

Çizelge 5.2. Rasgele seçim metodu için test istatistikleri.

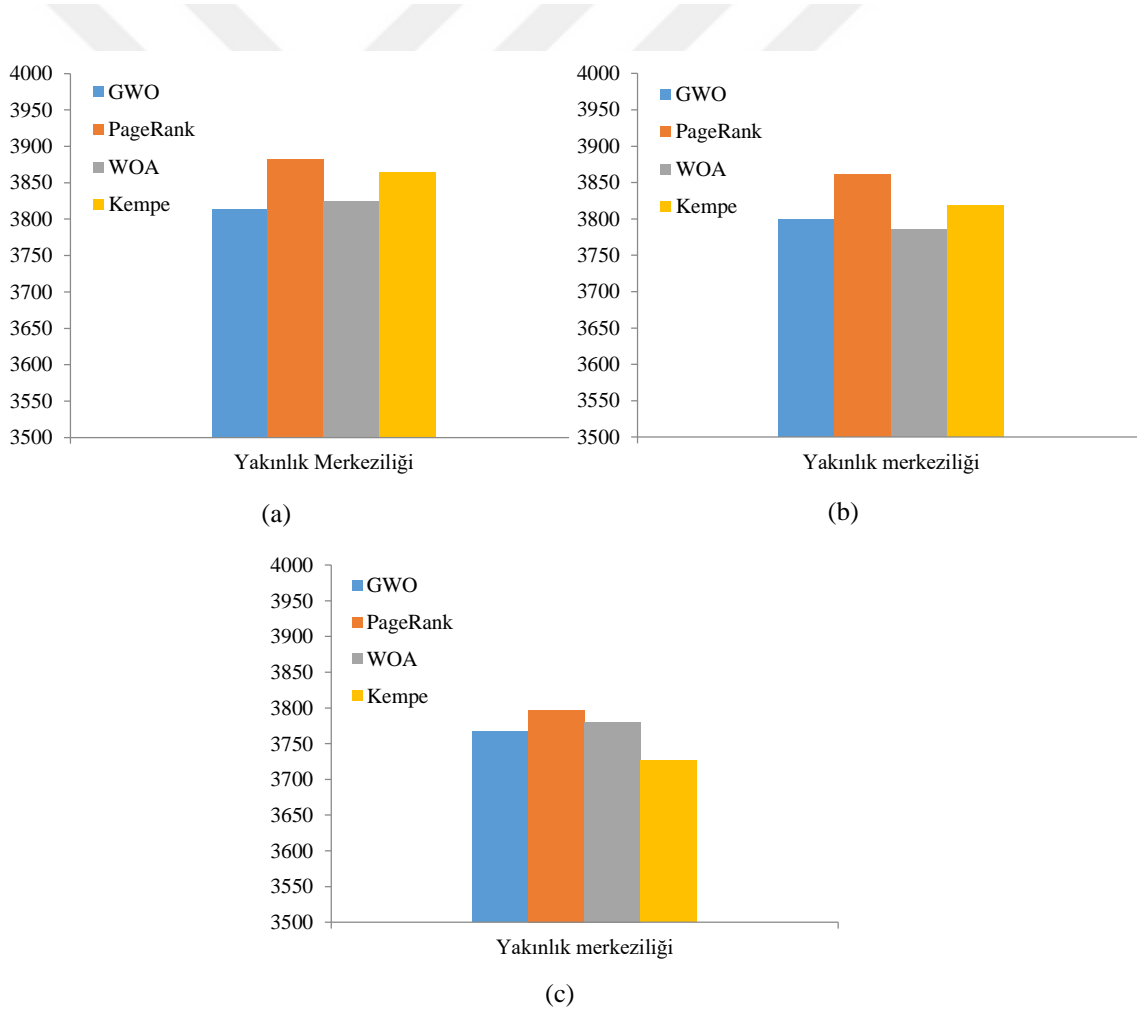
	Random2 - Random1
Z	-.339
Asymp. Sig. (2-tailed)	.735

Önerilen Yöntemin algoritmik karmaşıklığı da hesaplanmıştır. Yeniden şekillendirme işleminin ilk adımı kullanılacak ölçütlerin hesaplanmasıdır. Yeniden şekillendirme işleminin algoritmik karmaşıklığı hesaplanacak ölçütlerin algoritmik karmaşıklığına büyük ölçüde bağlıdır. Örneğin, çıkış derecesi merkeziliği ölçütü hesabının karmaşıklığı $O(|V|^2)$ 'dir [4]. Yakınlık merkeziliği ölçütünün karmaşıklığını hesaplamak için ise ilk olarak Johnson'un bütün çiftler arasındaki en kısa yol hesabı algoritması kullanılabilir. Johnson'un algoritmasının karmaşıklığı $O(|V|^2 \log|V| + |V||E|)$ 'dir. Bütün düğümler arasındaki en kısa yollar hesaplandıktan sonra, bütün düğümler için ortalama en kısa yol değerleri hesaplanır. Bunun için iç içe 2 döngüye ihtiyaç vardır. Bu da, karmaşıklığa $O(|V|^2)$ ekler. Sonuç olarak, yakınlık merkeziliği ölçütü hesabının algoritmik karmaşıklığı $O(|V|^2 \log|V| + |V|^2 + |V||E|)$ olarak elde edilir. Yeniden şekillendirme işleminin bir sonraki aşaması, hesaplanan ölçüt ya da ölçütlere göre düğümlerin sıralanmasıdır. Eğer, Birleşmeli Sıralama gibi etkin bir algoritma seçilirse bu işlemin algoritmik karmaşıklığı $O(|V| \log|V|)$ olacaktır.

5.3. GWO, WOA, PAGERANK VE KEMPE'NİN AÇGÖZLÜ ALGORİTMALARI İÇİN FACEBOOK GRAFI ÜZERİNDE YAPILAN DENEMELER

Bir önceki bölümde yeniden şekillendirme işleminin işe yaradığı gösterilmişti. Bu bölümde GWO, WOA, PageRank ve Kempe'nin Aç Gözlü Algoritmalarının EM problemi üzerindeki performansları karşılaştırılmıştır. GWO'nunkine benzer olarak WOA'nın Python kodları <http://www.alimirjalili.com/WOA.html> adresinden indirilmiştir ve amaç fonksiyonu olarak Denklem (3.2)'de verilen $f_2(S)$ kullanılmıştır.

Önerilen yöntemin faydası daha yüksek çekirdek sayıları için daha belirgin olduğundan denemeler $k = 15$, $k = 20$, ve $k = 30$ için yapılmıştır. Herbir denemenin sonucu Şekil 5.4'te sırasıyla gösterilmektedir.



Şekil 5.4. Farklı sayıdaki çekirdekler için Facebook grafi üzerinde GWO, WOA, PageRank ve Kempe ve arkadaşlarının Aç Gözlü Algoritması metotlarının deneysel sonuçları: (a) $k=15$, (b) $k=20$, ve (c) $k=30$.

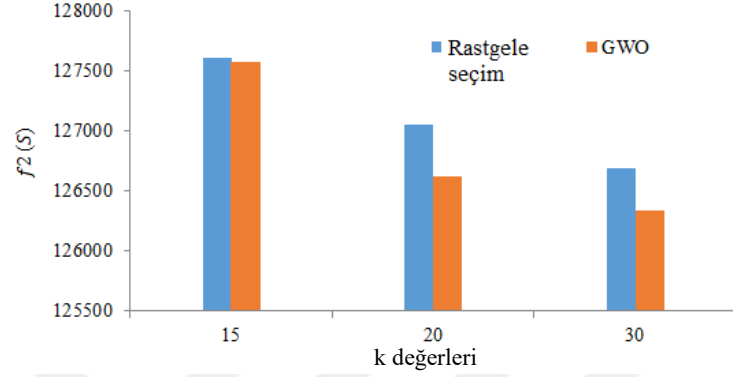
Denemelerde yalnızca yakınlık merkeziliğine göre yeniden şekillendirilen Facebook grafi kullanılmıştır. GWO ve WOA bütün denemelerde PageRank algoritmasından daha iyi sonuç vermiştir ve Kempe'nin algoritmasından $k = 15$ ve $k = 20$ durumları için daha iyi sonuç vermiştir. $k = 30$ durumu için Kempe'nin açgözlü algoritması en iyi sonucu üretmiştir. Yine $k = 30$ için GWO ve WOA'nın performansları birbirine yakındır.

5.4. EPINIONS GRAFI ÜZERİNDE YAPILAN GWO VE RASGELE SEÇİM KARŞILAŞTIRMASI

Önerilen yeniden şekillendirme yönteminin daha büyük sosyal ağlarda ve işaretli etki yayılımı probleminde başarılı sonuçlar verip vermediğini görmek için Epinions grafi üzerinde GWO ve rasgele seçim algoritmaları karşılaştırılmıştır. Burada Kempe'nin algoritmasını karşılaştırma için kullanmamamızın nedeni, Kempe'nin algoritmasının her bir adımda etkiyi maksimize eden düğümü çekirdek düğüm olarak seçmesi mekanizmasıdır. İşaretli sosyal ağlarda bir düğüm çekirdek düğüm olarak seçildiği zaman belirli düğümleri olumlu etkilemesinin yanı sıra bazı düğümleri de olumsuz etkileyecektir. Bu durum, gerçekte hangi düğümün o andaki düğümler içerisinde en yüksek etkiye sahip olduğunu tespit etmek için Kempe'nin orijinal algoritmasından farklı bir algoritma gerektirir.

GWO ve rasgele seçim algoritmaları denemeleri 15, 20 ve 30 çekirdek sayısı için, yakınlık merkeziliğine göre yeniden şekillendirilmiş graf üzerinde gerçekleştirilmiştir. Karşılaştırma sonuçları Şekil 5.5'te gösterilmektedir. 15 çekirdek sayısı için GWO ve rasgele seçim arasında anlamlı bir fark gözlemlenmemiş ancak daha yüksek sayılarda GWO algoritması rasgele seçime göre daha iyi sonuçlar vermiştir. Epinions grafi Facebook grafindan yaklaşık 33 kat daha büyüktür. Bu durumda 15 çekirdek sayısı nispeten az bir değer olarak düşünülebilir. Bundan dolayı GWO ve rasgele seçim arasında çok anlamlı bir fark oluşmaması olağan karşılanabilir. Burada PageRank ve Kempe'nin açgözlü algoritması kıyaslama için kullanılmamıştır. PageRank bir düğümün graf üzerindeki konumu ile ilgili bir ölçüt olup bu düğümün komşularıyla olan olumlu/olumsuz etkileme ilişkisini göz önünde bulundurmaz. Örneğin, PageRank değeri daha yüksek olan bir düğüm komşularının çoğunu olumsuz etkiliyorken, PageRank değeri daha düşük olan bir düğüm, komşularının çoğunu olumlu etkileyebilir. Bu durumda PageRank değeri yüksek olan düğümün çekirdek düğüm olarak seçilmesi

hatalı olacaktır. Benzer şekilde Kempe'nin açgözlü algoritması da bir çekirdek düğümü seçerken bu düğümün kaç komşusunu etkilediğine bakar. Bunu yaparken olumlu/olumsuz etkiyi göz ardı eder. Dolayısıyla, PageRank ve Kempe'nin açgözlü algoritmalarının işaretli ağlara uygulanabilmeleri için bu ağların özellikleri dikkate alınarak algoritmalar üzerinde değişiklikler yapılması gereklidir.



Şekil 5.5. Epinions grafi üzerinde GWO ve rasgele seçim karşılaştırması ($k=15$, $k=20$ ve $k=30$ değerleri için).

6. SONUÇLAR VE DEĞERLENDİRMELER

6.1. SONUÇLAR

EM gibi problemlerde, amaç fonksiyonunun ürettiği durum uzay yüzeyi genel bir eğime sahip değildir. Bu nedenle, sürü zekâsı algoritmaları global yüzeyde optimal bir çözüme yakınsayamazlar. EM gibi kombinatoriyel problemlere sürü zekâsı algoritmalarını uygulamak bu algoritmaların avantajlarından faydalanmak demektir. Bu tezde, EM problemi, sürü zekâsı algoritmalarının bu tip problemlere uygulanabilmesi için yeniden şekillendirilmiştir. Yeniden şekillendirmenin anlamı, ele alınan ölçütlere göre düğümlerin azalan sırada yeniden sıralanması ve düğümlerin yeni sıralarına göre yeniden numaralandırılmasıdır. Bu yöntem kullanılarak sosyal ağı yeniden şekillendirerek durum uzay yüzeyinde bir eğim oluşturulmuş ve sürü zekâsı algoritmaları EM problemlerine uygulanabilir hale getirilmiştir.

Tez kapsamında gerçekleştirilen deneysel çalışmaların sonuçları şu şekilde özetlenebilir. Facebook grafi üzerinde gerçekleştirilen GWO ve Rasgele Seçim yöntemlerinin performansları kıyaslandığında, GWO algoritmasının performansının Rasgele Seçim yönteminin performansına göre ortalama %48 daha iyi olduğu görülmektedir. Bu performans kıyası, her iki yöntem uygulandığı zaman ağ üzerindeki etkilenen düğüm sayılarının birbirlerine oranı ile bulunmuştur. Benzer şekilde, Facebook grafi üzerinde gerçekleştirilen GWO, WOA, ve PageRank algoritmalarının performansı kıyaslandığında; GWO ve WOA'nın performanslarının çok yakın olduğu ve bu algoritmalarının, PageRank algoritmasına göre ortalama %28 daha iyi sonuç verdiği söylenebilir. Aynı denemelerdeki Kempe'nin algoritmasının performansı ile GWO ve WOA'nın performansları yakındır.

Epinions grafi üzerinde gerçekleştirilen GWO ve Rasgele Seçim yöntemlerinin performansları kıyaslandığında, GWO algoritmasının performansının Rasgele Seçim yönteminin performansına göre ortalama %8 daha iyi olduğu görülmektedir. Burada şunu belirtmek de gerekir: GWO algoritmasının, işaretli sosyal ağlar üzerinde daha iyi sonuç vermesi için bu algoritmaya hiçbir müdahalede bulunulmamış, yalnızca problem yeniden şekillendirilmiştir.

6.2. ÇALIŞMANIN GETİRDİĞİ KATKILAR

Literatürde sürü zekâsı algoritmalarının EM probleminde kullanımı son derece kısıtlıdır. Çoğu sürü zekâsı algoritması kombinatoriyal optimizasyon problemlerine doğrudan uygulanamazlar. Sürü zekâsı algoritmalarının bu tip problemlere uygulanabilmesi için ya algoritma ya da problem amaca uygun olarak modifiye edilmelidir. Bu tez çalışmasının en temel katkısı sürü zekâsı algoritmalarının kombinatoriyal bir optimizasyon problemi olan EM problemine uygulanmasını sağlamasıdır. Bu tez çalışması sürü zekâsı algoritmalarının EM problemine sosyal ağ grafının yapısını değiştirmeden yeniden şekillendirilerek uygulanabileceğini göstermiştir. Elde edilen deneysel sonuçlar ve yapılan istatistiksel analizler bunu desteklemektedir. GWO ve WOA'nın sürü zekâsı algoritmaları olarak seçilmiş olmalarına rağmen farklı sürü zekâsı algoritmaları da EM problemlerine uygulanabilir.

Bu çalışmanın literatüre katkıları şu şekilde sıralanabilir:

- Sürü zekâsı algoritmalarının EM problemlerine uygulanmasının yolu açılmıştır.
- İki adet sürü zekâsı algoritmasının (GWO and WOA) başarılı bir şekilde EM problemlerine uygulandığı gösterilmiştir.
- Merkezilik ölçütlerinin başarılı bir şekilde birleştirildiği ve uygulandığı gösterilmiştir.

6.3. TARTIŞMA VE ÖNERİLER

Gelecek çalışmalar için bu konu üzerinde bazı açık sorunlar bulunmaktadır.

Ölçütlere ve Katsayılarına Karar Vermek

Daha önce de bahsedildiği gibi merkezilik ölçütleri tek başlarına kullanılabilir gibi bileşik olarak da kullanılabilirler. Bu çalışmada çıkış derecesi merkeziliği ve yakınlık merkeziliğini kullanarak bir bileşik metrik oluşturulmuştur. Bunların yanı sıra eigenvector merkeziliği ve Katz merkeziliği gibi farklı ölçütler de vardır. Bir araştırma konusu, bir ölçütün farklı durumlarda bir bireyin etkinliğini ne kadar gösterdiğidir. Örneğin, güven ilişkisi, takip etme ilişkisi, arkadaş ilişkisi gibi farklı ilişkilerin modellendiği ağlar için farklı ölçütler değerlendirilebilir. Bir başka konu da bileşik ölçüt oluştururken kullanılan katsayıların belirlenmesidir. Bu çalışmada bileşik ölçüt

oluştururken her iki ölçütün de eşit öneme sahip oldukları varsayılarak 0,5 katsayısı kullanıldı ancak farklı ağ tipleri, farklı sosyal ilişkiler, farklı yayılım modelleri gibi farklı durumlar için katsayı seçimi farklı şekillerde yapılabilir.

Problem için Sürü Zekâsı Algoritmasının Seçimi

Bu çalışmanın amacı sürü zekâsı algoritmalarının EM problemine uygulanabileceğini göstermektir. Bu amaçla GWO ve WOA algoritmaları kullanıldı. Ancak literatürde birçok farklı sürü zekâsı algoritması bulunmaktadır. Yukarıda bahsedilen farklı durumlar için değişik sürü zekâsı algoritmalarının kullanılması denenebilir.

Algoritmayı Probleme Göre Modifiye Etmek

Bu çalışmadaki temel yöntem problemi sürü zekâsı algoritmalarına uyarlamaktır. Bununla birlikte algoritmaların da probleme uygun hale getirilmesi mümkündür. Bütün sürü zekâsı algoritmaları problemin durum uzayı yüzeyi üzerinde hareket etmek için, yerel minimum veya yerel maksimumlardan kurtulmak için kendi stratejilerine sahiptir. Algoritmaların bu stratejileri EM probleminin doğasına göre yeniden uyarlanarak daha başarılı sonuçlar elde edilebilir.

Farklı Ağ Tipleri İçin Problemi Yeniden Şekillendirme

Giriş kısmında da bahsedildiği gibi EM problemi yalnızca sosyal ağlar üzerindeki bir problem değildir. Çok farklı ağ tiplerindeki benzer problemler EM problemi olarak ele alınabilir. Örneğin bir e-posta haberleşme ağında bir virüs yayılımını minimize etmek için hangi hesapların özellikle korunması gerektiğinin tespiti de bir EM problemidir. Değişik ağ tipleri kendilerine has özelliklere sahip oldukları için yeniden şekillendirme işlemi bir ağdan diğer ağa değişebilir.

EM Problemini Farklı Yayılım Modelleri ve Sürekli-Zaman Yayılımı Altında Ele Alma

IC modelinden daha karmaşık ve gerçekçi bilgi yayılımı modelleri vardır. Bu modeller daha yüksek işlemci gücüne ihtiyaç duyarlar. Ek olarak IC ve LT modelleri ayrık zamanlı çalışan modellerdir. Ne varki, gerçek dünyada bilgi ayrık zamanlı olarak yayılmaz. Bu nedenden dolayı daha gerçekçi modellemeler için sürekli-zamanlı yayılım modelleri kullanılabilir [22].

7. KAYNAKLAR

- [1] T. Johansson, “Gossip spread in social network Models,” *Physica A: Statistical Mechanics and its Applications*, vol. 471, pp. 126–134, 2017.
- [2] G. A. Tong, S. Li, W. Wu, and D.-Z. Du, “Effector Detection in Social Networks,” *IEEE Transactions on Computational Social Systems*, vol. 3, no. 4, pp. 151–163, 2016.
- [3] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” *ACM SIGKDD International Conference On Knowledge Discovery And Data Mining (KDD)*, Washington, USA, 2003, pp. 137-146.
- [4] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila, “Finding effectors in social networks,” *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, Washington, USA, 2010, pp. 1059-1068.
- [5] K. Zhang, H. Du, and M. W. Feldman, “Maximizing influence in a social network: Improved results using a genetic algorithm,” *Physica A: Statistical Mechanics and its Applications*, vol. 478, pp. 20–30, 2017.
- [6] D. Karaboga, B. Akay, and C. Ozturk, “Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks,” *International Conference on Modeling Decisions for Artificial Intelligence (MDAI)*, Kitakyushu, Japan, 2007, pp. 318–329.
- [7] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey Wolf Optimizer,” *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [8] S. Mirjalili and A. Lewis, “The Whale Optimization Algorithm,” *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.
- [9] S. Mirjalili, “The Ant Lion Optimizer,” *Advances in Engineering Software*, vol. 83, pp. 80–98, 2015.
- [10] M. Gong, J. Yan, B. Shen, L. Ma, and Q. Cai, “Influence maximization in social networks based on discrete particle swarm optimization,” *Information Sciences*, vol. 367–368, pp. 600–614, 2016.
- [11] Qixiang Wang, M. Gong, Chao Song, and Shanfeng Wang, “Discrete particle swarm optimization based influence maximization in complex networks,” *IEEE Congress on Evolutionary Computation (CEC)*, San Sebastian, Spain, 2017, pp. 488–494.
- [12] J. McAuley and J. Leskovec, “Learning to discover social circles in ego networks,” *International Conference on Neural Information Processing Systems (NIPS)*, Lake Tahoe, USA, 2012, pp. 539–547.
- [13] J. Leskovec and A. Krevl. (2014, Jun 1). [Online]. Available: <https://snap.stanford.edu/data>
- [14] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, and W. Jia, “Influence analysis in social

- networks: A survey,” *Journal of Network and Computer Applications*, vol. 106, pp. 17–32, 2018.
- [15] S. P. Borgatti, “Centrality and network flow,” *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.
- [16] E. Bergamini, M. Borassi, P. Crescenzi, A. Marino, and H. Meyerhenke, “Computing Top- k Closeness Centrality Faster in Unweighted Graphs,” *Algorithm Engineering & Experiments (ALENEX)*, Arlington, USA, 2016, pp. 68–80.
- [17] S. P. Borgatti, K. M. Carley, and D. Krackhardt, “On the robustness of centrality measures under conditions of imperfect data,” *Social Networks*, vol. 28, no. 2, pp. 124–136, 2006.
- [18] R. Rabade, N. Mishra, and S. Sharma, “Survey of Influential User Identification Techniques in Online Social Networks,” *International Symposium on Intelligent Informatics (ISI)*, Mysore, India, 2014, pp. 359–370.
- [19] A. Temizsoy, G. Iori, and G. Montes-Rojas, “Network centrality and funding rates in the e-MID interbank market,” *Journal of Financial Stability*, vol. 33, pp. 346–365, 2017.
- [20] J.V. Cossu, V. Labatut, and N. Dugué, “A review of features for the discrimination of twitter users: application to the prediction of offline influence,” *Social Network Analysis and Mining*, vol. 6, no. 1, p. 25, 2016.
- [21] S. P. Borgatti, “Identifying sets of key players in a social network,” *Computational and Mathematical Organization Theory*, vol. 12, no. 1, pp. 21–34, 2006.
- [22] M. Samadi, R. Nagi, A. Semenov, and A. Nikolaev, “Seed activation scheduling for influence maximization in social networks,” *Omega*, vol. 77, pp. 96–114, 2018.
- [23] S. Raghavan and R. Zhang, “Weighted Target Set Selection on Social Networks,” The Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, Maryland, USA, Tech. Rep.
- [24] D. Günneç, S. Raghavan, and R. Zhang, “Tailored Incentives and Least Cost Influence Maximization on Social Networks,” The Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, Maryland, USA, Tech. Rep.
- [25] G. Song, Y. Li, X. Chen, X. He, and J. Tang, “Influential Node Tracking on Dynamic Social Network: An Interchange Greedy Approach,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 359–372, 2017.
- [26] M. Samadi, A. Nikolaev, and R. Nagi, “The temporal aspects of the evidence-based influence maximization on social networks,” *Optimization Methods and Software*, vol. 32, no. 2, pp. 290–311, 2017.
- [27] B. Liu, G. Cong, Y. Zeng, D. Xu, and Y. M. Chee, “Influence spreading path and its application to the time constrained social influence maximization problem and beyond,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1904–1917, 2014.

- [28] M. T. Irfan and L. E. Ortiz, "On influence, stable behavior, and the most influential individuals in networks: A game-theoretic approach," *Artificial Intelligence*, vol. 215, pp. 79–119, 2014.
- [29] X. Molinero, F. Riquelme, and M. Serna, "Cooperation through social influence," *European Journal of Operational Research*, vol. 242, no. 3, pp. 960–974, 2015.
- [30] J. Ok, Y. Jin, J. Shin, and Y. Yi, "On Maximizing Diffusion Speed Over Social Networks With Strategic Users," *IEEE/ACM Transactions on Networking*, vol. 24, no. 6, pp. 3798–3811, 2016.
- [31] G. Song, X. Zhou, Y. Wang, and K. Xie, "Influence maximization on large-scale mobile social network: a divide-and-conquer method," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1379–1392, 2015.
- [32] J.-R. Lee and C.-W. Chung, "A query approach for influence maximization on specific users in social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 340–353, 2015.
- [33] H. Zhang, D. T. Nguyen, H. Zhang, and M. T. Thai, "Least cost influence maximization across multiple social networks," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 929–939, 2016.
- [34] G. Tong, W. Wu, S. Tang, and D.-Z. Du, "Adaptive influence maximization in dynamic social networks," *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 112–125, 2017.
- [35] G. Sartor, Y. K. Chia, L. Wynter, and J. Ruths, "A Weighted Maximum Matching Algorithm for Influence Maximization and Structural Controllability," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14447–14453, 2017.
- [36] S. Peng, A. Yang, L. Cao, S. Yu, and D. Xie, "Social influence modeling using information theory in mobile social networks," *Information Sciences*, vol. 379, pp. 146–159, 2017.
- [37] D. Li, Z.-M. Xu, N. Chakraborty, A. Gupta, K. Sycara, and S. Li, "Polarity related influence maximization in signed social networks," *PLoS One*, vol. 9, no. 7, pp. 1–12, 2014.
- [38] D. Li, C. Wang, S. Zhang, G. Zhou, D. Chu, and C. Wu, "Positive influence maximization in signed social networks based on simulated annealing," *Neurocomputing*, vol. 260, pp. 69–78, 2017.
- [39] M. Gong, C. Song, C. Duan, L. Ma, and B. Shen, "An efficient memetic algorithm for influence maximization in social networks," *IEEE Computational Intelligence Magazine*, vol. 11, no. 3, pp. 22–33, 2016.
- [40] C. Zhou, P. Zhang, W. Zang, and L. Guo, "On the upper bounds of spread for greedy algorithms in social network influence maximization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2770–2783, 2015.
- [41] H. Nguyen and R. Zheng, "On budgeted influence maximization in social networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1084–1094, 2013.
- [42] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," *ACM SIGKDD International Conference On Knowledge Discovery And Data Mining (KDD)*, Paris, France, 2009, pp. 199–208.

- [43] J. Gao, B. Barzel, and A.-L. Barabási, “Universal resilience patterns in complex networks,” *Nature*, vol. 530, no. 7590, pp. 307–312, 2016.
- [44] A.L. Barabási, *Network Science*, Cambridge, United Kingdom: Cambridge University Press, 2016.
- [45] E. N. Gilbert, “Random Graphs,” *Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.
- [46] A. L. Barabási, R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [47] A.-L. Barabási, R. Albert, and H. Jeong, “Mean-field theory for scale-free random networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1–2, pp. 173–187, 1999.
- [48] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining*. New York, USA: Cambridge University Press, 2014.
- [49] J. Sun and J. Tang, “A survey of models and algorithms for social influence analysis,” in *Social Network Data Analytics*, C. C. Aggrawal, Eds. Boston, MA: Springer, 2011, pp. 177–214.
- [50] L. Katz, “A new status index derived from sociometric analysis,” *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.
- [51] P. Bonacich, “Power and centrality: a family of measures,” *American Journal of Sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.
- [52] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, vol. 40, no. 1, p. 35, 1977.
- [53] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.
- [54] W. Chen, L. V. S. Lakshmanan, and C. Castillo, *Information and Influence Propagation in Social Networks*, vol. 5, no. 4. 2013.
- [55] M. S. Granovetter, “The strength of weak ties,” *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.
- [56] T. C. Schelling, “Dynamic models of segregation,” *The Journal of Mathematical Sociology*, vol. 1, no. 2, pp. 143–186, 1971.
- [57] F. Morone and H. A. Makse, “Influence maximization in complex networks through optimal percolation,” *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.
- [58] A. K. Tripathi, K. Sharma, and M. Bala, “A novel clustering method using enhanced grey wolf optimizer and mapreduce,” *Big Data Research*, In press.
- [59] P. K. Gujarathi, V. A. Shah, and M. M. Lokhande, “Grey wolf algorithm for multidimensional engine optimization of converted plug-in hybrid electric vehicle,” *Transportation Research Part D: Transport and Environment*, vol. 63, pp. 632–648, 2018.
- [60] A. K. M. Khairuzzaman and S. Chaudhury, “Multilevel thresholding using grey wolf optimizer for image segmentation,” *Expert Systems with Applications*, vol. 86, pp. 64–76, 2017.
- [61] M. Fahad *et al.*, “Grey wolf optimization based clustering algorithm for vehicular

- ad-hoc networks,” *Computers & Electrical Engineering*, 2018.
- [62] L. K. Panwar, S. Reddy K, A. Verma, B. K. Panigrahi, and R. Kumar, “Binary Grey Wolf Optimizer for large scale unit commitment problem,” *Swarm and Evolutionary Computation*, vol. 38, pp. 251–266, 2018.
- [63] X.-Q. Bian, L. Zhang, Z.-M. Du, J. Chen, and J.-Y. Zhang, “Prediction of sulfur solubility in supercritical sour gases using grey wolf optimizer-based support vector machine,” *Journal of Molecular Liquids*, vol. 261, pp. 431–438, 2018.
- [64] M. H. Qais, H. M. Hasanien, and S. Alghuwainem, “Augmented grey wolf optimizer for grid-connected PMSG-based wind energy conversion systems,” *Applied Soft Computing*, vol. 69, pp. 504–515, 2018.
- [65] W. Long, J. Jiao, X. Liang, and M. Tang, “An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization,” *Engineering Applications of Artificial Intelligence*, vol. 68, pp. 63–80, 2018.
- [66] W. Long, J. Jiao, X. Liang, and M. Tang, “Inspired grey wolf optimizer for solving large-scale function optimization problems,” *Applied Mathematical Model*, vol. 60, pp. 112–126, 2018.
- [67] S. Raj and B. Bhattacharyya, “Optimal placement of TCSC and SVC for reactive power planning using Whale optimization algorithm,” *Swarm and Evolutionary Computation*, vol. 40, pp. 131–143, 2018.
- [68] H. H. Mehne and S. Mirjalili, “A parallel numerical method for solving optimal control problems based on whale optimization algorithm,” *Knowledge-Based Syst.*, vol. 151, pp. 114–123, 2018.
- [69] M. Mafarja and S. Mirjalili, “Whale optimization approaches for wrapper feature selection,” *Applied Soft Computing*, vol. 62, pp. 441–453, 2018.
- [70] M. Abd Elaziz and D. Oliva, “Parameter estimation of solar cells diode models by an improved opposition-based whale optimization algorithm,” *Energy Conversion and Management*, vol. 171, pp. 1843–1859, 2018.
- [71] M. A. El Aziz, A. A. Ewees, and A. E. Hassanien, “Whale Optimization Algorithm and Moth-Flame Optimization for multilevel thresholding image segmentation,” *Expert Systems with Applications*, vol. 83, pp. 242–256, 2017.
- [72] A. M. Al-Zoubi, H. Faris, J. Alqatawna, and M. A. Hassonah, “Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts,” *Knowledge-Based Systems*, vol. 153, pp. 91–104, 2018.
- [73] D. B. Prakash and C. Lakshminarayana, “Optimal siting of capacitors in radial distribution network using Whale Optimization Algorithm,” *Alexandria Engineering Journal*, vol. 56, no. 4, pp. 499–509, 2017.
- [74] A. Tharwat, Y. S. Moemen, and A. E. Hassanien, “Classification of toxicity effects of biotransformed hepatic drugs using whale optimized support vector machines,” *Journal of Biomedical Informatics*, vol. 68, pp. 132–149, 2017.
- [75] C. K. Monson and K. D. Seppi, “Adaptive diversity in PSO,” *Annual Conference On Genetic And Evolutionary Computation (GECCO)*, Seattle, USA, 2006, pp. 59–66.
- [76] S. Cheng, H. Zhan, and Z. Shu, “An innovative hybrid multi-objective particle

- swarm optimization with or without constraints handling,” *Applied Soft Computing*, vol. 47, pp. 370–388, 2016.
- [77] E. Zahara and Y.-T. Kao, “Hybrid Nelder–Mead simplex search and particle swarm optimization for constrained engineering design problems,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3880–3886, 2009.
- [78] S.-K. S. Fan and E. Zahara, “A hybrid simplex search and particle swarm optimization for unconstrained optimization,” *European Journal of Operational Research*, vol. 181, no. 2, pp. 527–548, 2007.
- [79] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*. New York, USA: Cambridge University Press, 2014.
- [80] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using NetworkX,” *Python in Science Conference (SciPy)*, Pasadena, USA, 2008, pp. 11–15.
- [81] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” InfoLab, Stanford University, San Francisco, USA, Tech. Rep.
- [82] Z. Zengin Alp and Ş. Gündüz Öğüdücü, “Identifying topical influencers on twitter based on user behavior and network topology,” *Knowledge-Based Systems*, vol. 141, pp. 211–221, 2018.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı :Aybike ŞİMŞEK
Doğum Tarihi ve Yeri :08.04.1982 / Horasan
Yabancı Dili :İngilizce
E-posta :aybikesimsek@duzce.edu.tr

ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Doktora	Elektrik- Elektronik ve Bilgisayar Mühendisliği	Düzce Üniversitesi	2018
Y. Lisans	Bilgisayar Mühendisliği	Gazi Üniversitesi	2010
Lisans	Bilgisayar Mühendisliği	Selçuk Üniversitesi	2003
Lise		Kenan Evren Anadolu Lisesi	1999

TEZDEN ÇIKAN YAYINLAR

Simsek A, Kara R., “Using Swarm Intelligence Algorithms to Detect Influential Individuals for Influence Maximization in Social Networks”, Expert Systems with Applications, Temmuz, 2018, doi: 10.1016/j.eswa.2018.07.038.

Simsek A, Kara R., “Influence Maximization in Social Networks : Defining Starting Point Criteria” , III International Conference on Engineering and Natural Science (ICENS 2017), Budapeşte/Macaristan, Mayıs 3-7, 2017.

Simsek A, Kara R., “A Simple Approach for Influence Maximization in Social Networks” , 5th International Conference on Advanced Technology & Sciences (ICAT'17), İstanbul/Türkiye, Mayıs 9-12, 2017.