

Büyük Ölçekli Etki Enbüyükleme Problemi İçin Lagrange Gevşetmesi Tabanlı Etkin Bir Çözüm Yöntemi

Evren GÜNEY¹

¹MEF Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Maslak-İstanbul

e-posta: guneye@mef.edu.tr ORCID ID: <http://orcid.org/0000-0001-7572-8627>

Geliş Tarihi: 17.09.2019

Kabul Tarihi: 30.12.2019

Öz

Etki Enbüyükleme Problemi (EEP) büyük bir sosyal ağ içindeki en etkin K tane kişiyi seçen zor bir stokastik kombinatoriyal eniyileme problemidir. Son yıllarda pek çok araştırmacının ilgisini çeken bu problem için çok sayıda etkin yöntem geliştirilmiştir. Sosyal ağdaki bilginin / etkinin yayılımı çeşitli ağ akış modelleri ile tasarlandığında, elde edilen problemin amaç fonksiyonunun alt-birimsel olduğu gözlemlenmiştir. Bu sebeple basit bir açgözlü algoritma ile $(1-1/e)$ en kötü performans garantisine erişilmiştir. Ancak, açgözlü algoritmanın büyük boyutlu problemlerde çok uzun çözüm süreleri gerektirmesi alternatif yöntem arayışlarına neden olmuştur. Son yıllarda geliştirilen yeni yöntemler genelde büyük boyutlu ağlarda kısa sürede iyi çözümler elde ederken $(1-1/e)$ performans garantisini de korumaktadır. Ancak pek az sayıda çalışma problemin sadece en-iyi çözümüne odaklanmıştır. Bu çalışmada Lagrange gevşetmesi tabanlı ve EEP'yi eniyi / eniyeye yakın çözen ve ölçeklenebilen bir yöntem geliştirilmiştir. Bu çerçevede, öncelikle Örneklem Ortalama Yakınsaması ile özgün probleme yakınsayan belirgin bir matematiksel model kurulmuştur. Daha sonra bu model üzerinde düğüm tabanlı Lagrange gevşetmesi tekniği uygulanmıştır. İlgili yöntem bağımsız çağlayan ve doğrusal eşik bilgi yayılım modelleri varsayımı altında çeşitli boyutlardaki sosyal ağ veri setleri (Facebook, Enron, Gnutella, arXiv) üzerinde test edilmiştir. Bütün senaryolarda eniyi / eniyeye yakın çözümlere ulaşılırken yazındaki mevcut yöntemlere göre on kata kadar hızlanma sağlanmıştır.

Anahtar kelimeler

Etki Enbüyükleme;
Sosyal Ağlar; Stokastik
Eniyileme; Lagrange
Gevşetmesi

An Efficient Lagrangean Relaxation Based Solution Method For Large Scale Influence Maximization Problem

Abstract

The Influence Maximization Problem (IMP) is defined as identifying the most influential K individuals in a social network, which is a challenging stochastic combinatorial optimization problem. IMP has attracted great interest among researchers in the last years and many efficient solution methods for this problem has been developed. Under IMP the flow of information through the social network is assumed to be following certain information diffusion models and in certain cases the resulting objective function of the optimization problem is a sub-modular function. Therefore, a naive greedy heuristic can achieve a $(1-1/e)$ worst-case bound. However, the greedy method is not scalable and results in very long computation time for large social networks. Upon this observation, many researchers proposed fast and scalable heuristics that still preserve the $(1-1/e)$ worst-case bound. Contrarily, very few researchers focused on finding methods that provide optimal solutions to IMP. In this work a Lagrangean Relaxation based, fast and scalable method that provides optimal / close-to-optimal solutions is proposed. First, a deterministic optimization problem is constructed by using Sample Average Approximation for the original problem. Then, a node based Lagrangean Relaxation approach is developed to solve the approximation of the original problem. The method is tested with both Independent Cascade and Linear Threshold information diffusion models over various size real-life network instances (Facebook, Enron, Gnutella, arXiv). In all scenarios optimal / close-to-optimal solutions are obtained and our approach outperforms the state-of-the-art approaches up to ten times in terms of solution time.

Keywords

Influence
Maximization; Social
Networks; Stochastic
Optimization;
Lagrangean Relaxation

1. Giriş

Son yıllarda popülerliği ve kullanım sıklığı sürekli artan çevrimiçi sosyal ağlar (ÇSA) ile birlikte bilgi yayılım sistemlerine yönelik ilgide de büyük artış gözlemlenmiştir. Herhangi bir ÇSA'da paylaşılan bir içeriğin pek çok kişi tarafından tekrar tekrar paylaşıldığı durumda tıpkı bir virüs gibi çok hızlı şekilde yayılarak binlerce kişiye kısa sürede ulaştığına çok kere şahit olunmuştur (Li *et al.* 2018). Bu durumun sonucu olarak artık ÇSA'lar günümüzde bilgi paylaşımında ve yayılımında ilk akla gelen ve en sık kullanılan ortamlardır. ÇSA'larda paylaşılan içerikleri gönderen ve alan kişiler genellikle birbirlerini gerçek hayatta da tanır. Bu sebeple bu kanal üzerinden gelen mesajlar/bilgiler alıcı kişi tarafından daha samimi ve inandırıcı bulunur. Dolayısıyla ÇSA'lar üzerinden yürütülen pazarlama kampanyaları geleneksel mecralara göre çoğu zaman daha yüksek performans gösterir (Günneç, 2018). Bu yüzden artık çok daha fazla birey, firma, kar gütmeyen kurum, politikacılar ve buna benzer pek çok kurum ve kuruluş ÇSA'ları kullanarak viral pazarlama ve marka iletişimi faaliyetleri yapmaktadır (Peng *et al.* 2018).

Etki Enbüyükleme Problemi (EEP), bir ÇSA'da yer alan etkileme gücü en yüksek bireyleri tespit etmek şeklinde tanımlanır. EEP'de amaç, herhangi bir bilginin yayılımına tespit edilecek en iyi başlangıç veya çekirdek kümesi ile başladığında, yayılım sona erdiğinde en yüksek sayıda kişiye erişmektir. Sosyal ağlarda her paylaşılan bilgi alıcılar tarafından kabul görmeyebilir. Dolayısıyla bir kişinin diğerini etkilemesi belli bir olasılıkla gerçekleşir. Bu sebeple, EEP stokastik bir ağ üzerinde modellenir ve amaç fonksiyonu beklenen etkinin enbüyüklenmesi şeklinde oluşturulur. Matematiksel olarak EEP $\{max \sigma(S): |S| = k; S \subset V\}$ şeklinde ifade edilir. Burada S çekirdek kümesini, V sosyal ağdaki tüm bireyleri (çekirdek kümesi dahil) temsil eder. Çekirdek kümesinin boyutu k olmalıdır ve yayılıma S kümesi ile başladığında, yayılım tamamlandığında erişilen/etkilenen kişi sayısının beklenen değeri $\sigma(S)$ fonksiyonu ile gösterilir.

EEP, yaygın kullanım alanı sebebiyle son yıllarda büyük ilgi görmüştür. Zor bir eniyileme problemi olması nedeniyle de pek çok akademisyen tarafından incelenmiştir. Geniş bir yazını olan EEP'de araştırmacılar problemin değişik yönleri ile ilgilenmiştir (Li *et al.* 2018). EEP'yi bir matematiksel eniyileme problemi olarak ele alan ilk çalışma Kempe ve arkadaşları (Kempe *et al.* 2003) tarafından hazırlanmıştır. Bu çalışmada EEP'nin NP-zor bir problem olduğunu gösteren yazarlar, amaç fonksiyonu $\sigma(S)$ 'in çeşitli yayılım modelleri (bağımsız çağlayan ve doğrusal eşik) durumunda alt-birimsel olduğunu ispatlamışlardır. Dolayısıyla basit bir aç-gözlü algoritma kullanarak EEP'ye $(1-1/e)$ en kötü durum sınır garantisini sağlayan polinom zamanlı bir çözüm yöntemi önermişlerdir. EEP'nin tanımlandığı stokastik ağda $\sigma(S)$ 'in kesin olarak hesaplanması çok zordur, bu sebeple yazarlar Monte Carlo örnekleme yöntemi ile çok sayıda senaryo (örneklem) yaratıp, aç-gözlü algoritmayı bu senaryo kümesi üzerinde koşturmuşlardır. Bu durum aç-gözlü algoritmanın büyük ÇSA'larda yavaş çalışmasına yol açmış ve ölçeklenmesiyle ilgili sıkıntılara neden olmuştur. Bu durumu gözlemleyen pek çok araştırmacı aç-gözlü algoritmaya göre çok daha iyi şekilde ölçeklenebilen farklı algoritmalar geliştirmişlerdir. Leskovec ve arkadaşları CELF (Cost-Effective Lazy Forward) yöntemi ile klasik aç-gözlü algoritmayı 700 kata kadar hızlandırmayı başarmıştır (Leskovec *et al.* 2007). Bu çalışmanın ardından Chen ve ark. EEP için derece indirgemesi yöntemini önererek daha büyük problemleri çözmüşlerdir (Chen *et al.* 2009). Aç-gözlü algoritmayı doğrudan EEP'ye uygulamak yerine, problemi stokastik bir kapsama problemine dönüştürüp "ters erişilebilir küme" yaklaşımını kullanan algoritmalar ise şu anda sezgisel yöntemler içinde en başarılı olanlardır. Bu başlık altında çok sayıda çalışma vardır, ancak özellikle EasySIM (Galhotra *et al.* 2016), IMM (Tang *et al.* 2015), BKRIS (Ko *et al.* 2018) ve SSA (Nguyen *et al.* 2016) bunlar içinde en başarılı olanlardır.

Daha küçük ölçekli ÇSA'lar için yöneylem araştırması ve kombinatoriyal eniyileme yöntemleri kullanarak eniyi çözümü bulmaya odaklanan çalışmalar da mevcuttur. Bu amaçla EEP problemini kesikli karma

tam sayılı eniyileme problemi olarak oluşturup, etkin çözüm yöntemleri öneren çalışmalar yapılmıştır. Güney (2017) EEP'yi zaman indeksli ikili tam sayılı program olarak oluşturmuş ve örneklem ortalama yakınsaması (ÖOY) yöntemi ile çözmüştür. Aynı yazar bir sonraki çalışmasında EEP'nin doğrusal programlama gevşetmesine odaklanarak bu probleme polinom zamanlı $(1-1/e)$ en kötü sınır garantili alternatif bir yöntem geliştirmiştir (Güney, 2019). EEP'ye Bender's ayrıştırması yaklaşımı da oldukça iyi sonuç vermektedir. Wu ve Küçükayvuz EEP'nin Bender's reformülasyonuna ertelenmiş kesi üretimi yöntemi uygulayarak küçük ve orta boyutlu problemlere eniyi çözümler üretmişlerdir (Wu and Küçükayvuz, 2017). Son olarak EEP'ye deterministik şekilde yaklaşarak doğrusal eşik yayılım modeli özelinde sezgisel (Gürsoy ve Günneç, 2017) ve eniyi (Günneç *et al.* 2019) çözümler üreten çalışmalar bulunmaktadır.

Bu çalışmada amacımız EEP'nin eniyi çözümünü yazında bulunan çalışmalara kıyasla daha hızlı bulan ve ölçeklenebilir bir yöntem geliştirmektir. Bu amaçla EEP'nin örnek ortalama yakınsaması formülasyonuna Lagrange gevşetmesi tabanlı bir algoritma önerilmiştir. En iyi performansı elde etmek için klasik Lagrange gevşetmesi yöntemine ek olarak bazı ön-işleme yöntemleri geliştirilmiş, olurlu çözümler üretmek için de yine EEP'ye özgü bir Lagrange sezgiseli oluşturulmuştur. Çeşitli küçük ve orta ölçekli gerçek hayat verisine dayalı sosyal ağ problem örneklerinde uygulanan deneyler sonucunda bilinen-en-iyi yöntem (Wu and Küçükayvuz, 2017) kıyasla 2-10 kat arası hızlanma sağlanmıştır.

2. Materyal ve Metot

Bu bölümde önce EEP'nin formal tanımını yapacak ve varsayımlarımızı ortaya koyacağız. Daha sonra ikili tam-sayı matematiksel modelimizi ve bu modeli Örneklem Ortalama Yakınsaması (ÖOY) yöntemi ile nasıl çözdüğümüzü göstereceğiz. Son olarak da EEP'nin Lagrange gevşetmesini ve olurlu çözüm üretmek için geliştirdiğimiz Lagrange sezgiselini anlatacağız.

2.1 Problem Tanımı ve Varsayımlar

EEP yönlü bir stokastik ağ, $G(V,A,W)$, üzerinde tanımlanır. Burada V ağdaki düğümleri (sosyal ağdaki kişiler), A ağdaki okları ve W da oklar üstündeki ağırlıkları temsil eder. Sosyal ağda iki kişi (i ve j) arasında bir bağlantı veya arkadaşlık ilişkisi olduğu zaman (i,j) oku oluşturulur. i kişinin j kişisini etkileme düzeyi $p_{ij} \in [0,1]$ parametresi ile gösterilir. Eğer ilgilenilen sosyal ağımız yönsüz ise her bir kenar için (i,j) ve (j,i) olacak şekilde iki ok oluşturularak yönlü bir ağ elde edilir.

Kişiler tarafından yapılan paylaşımların veya daha genel bir bakış açısıyla bilginin, sosyal ağ üzerindeki akışı / yayılımı çeşitli yayılım modelleri ile tanımlanır. Bu çalışmada yazında en sık kullanılan iki yayılım modeli üzerinde durulmuştur: Doğrusal Eşik (DE) ve Bağımsız Çağlayan (BÇ).

Her iki modelde de ortak olacak şekilde, yayılımın bir S başlangıç kümesinden (çekirdek kümesi) başladığını ve bu kümedeki bireylerin ağ üzerindeki komşularını etkilemeye / etkinleştirmeye çalıştıkları varsayılır. Yayılımın adım adım ilerlediğini varsayarsak, her adımda etkinleşen bireyler kendi (etkin olmayan) komşularını etkinleştirmeye çalışır. Yayılım, artık ağ üzerinde etkinleştirme denemesi yapılacak düğüm kalmayana kadar devam eder ve durur. Her iki yayılım modelinde de, bir düğüm etkinleştiği zaman tüm yayılım süreci boyunca etkin kalır ve asla edilgen duruma geri dönmez. Bu varsayım EEP yazınında çok yaygın şekilde kullanılır, ancak bazı çalışmalarda bu varsayımın gevşetildiği farklı yayılım modelleri de bulunur (Li *et al.* 2018).

DE ve BÇ yayılım modelleri arasındaki temel fark komşuların nasıl etkinleştirildiği konusunda ortaya çıkar. BÇ yayılım modelinde etkin bir i düğümü yayılımın herhangi bir t adımında henüz etkinleşmemiş bir j komşusunu sadece bir kez deneyecek şekilde p_{ij} olasılığı ile etkilemeye çalışır. Eğer başarılı olursa j düğümü de etkin bir düğüme dönüşür ve $t+1$ adımından itibaren j düğümü de kendi etkin olmayan komşularını etkinleştirmeye çalışır. Eğer i düğümü j düğümünü etkileştirmeyi başaramazsa j düğümünü etkinleştirmek için tekrar deneme hakkı yoktur. Eğer j düğümünü

etkinleştirmek isteyen birden fazla düğüm varsa, düğümler rassal şekilde sıralanır ve hepsi teker teker sıra ile j 'yi etkinleştirmeye çalışırlar. Yayılım süreci başlangıç etkin düğüm kümesi S ile başlar ve başka etkinleştirilebilecek düğüm kalmayana kadar devam eder (Goldenberg *et al.* 2001).

Doğrusal Eşik yayılım modelinde ise i düğümünün etkin hale geçebilmesi için en düşük sayıda komşusunun etkin olması gerekir. Bu nedenle DE modelinde her birim düğüm bu en düşük sayıya denk gelecek bir eşik değerini (θ) rassal bir biçimde olarak $[0,1]$ aralığından saptar. Daha sonra i düğümünün komşuluk kümesi N_i' 'de yer alan ve etkin durumda olan her bir j düğümü, i düğümüne p_{ji} seviyesinde sabit bir etki uygular. Böylece i düğümüne gelen toplam etki miktarı $\sum_{j \in N_i'} p_{ji} \leq 1$ ifadesi ile hesaplanır. Toplam etki i düğümünün eşik değerini aştığı zaman, $\sum_{j \in N_i'} p_{ji} \geq \theta_i$, i düğümü etkin hale gelir ve kendi komşularını etkilemeye başlar. Özetle, rassal şekilde belirlenen eşik değerleri bilindiği durumda yayılım S kümesi ile başlar ve kesikli adımlar halinde etkilenen tüm düğümler etkilenene kadar devam eder (Granovetter 1978).

Her iki yayılım yönteminde de süreç çekirdek kümesi S ile başlar ve süreç sona erdiğinde etkin hale geçen tüm düğümler (başlangıç kümesi dâhil) gözlemlenir. Yayılımın durduğu an itibarıyla etkinleşmiş tüm düğümlerin sayısını $I(S)$ fonksiyonu gösterir. Tüm bu yayılım süreci rassallık içerdiği için (oklardaki olasılıklar veya eşik değerlerinin rassallığı) $I(S)$ de bir rassal değişkendir. Dolayısıyla, EEP'de amaç öyle bir eniyi S^* başlangıç kümesi bulunsun ki, yayılım tamamlandığında elde edilen, $\sigma(S) = \mathbf{E}[I(S)]$, yani yayılım sonucu etkinleşen düğümler kümesinin beklenen büyüklüğü, en büyük olsun.

2.2 Etkin Ok ve Tetikleyen Küme Kavramları

EEP için etkin algoritmalar geliştirmede bir başka önemli ölçüt, geliştirilen algoritmanın farklı yayılım modelleri için de geçerli olması ve iyi sonuç vermesidir. Eğer bizden hangi düğümün yayılımın hangi adımında etkinleştirildiği bilgisi istenmiyorsa ve sadece $\mathbf{E}[I(S)]$ 'nin en büyüklenmesi yeterli ise, yayılım modellerinde daha basit şekilde ifade

edilebilen bir eşdeğer gösterim oluşturulabilir. Özgün ağdaki bazı okların silinip, bazılarının silinmediği $G_r \subseteq G$ şeklinde bir ağ alt kümesi tanımlayalım. Yani G_r ağında G' 'de bulunan tüm düğümler aynen yer alırken okların bir kısmının silindiği varsayılır. Bu durumda, G_r 'de yer alan tüm oklara "etkin oklar" denirken geriye kalan oklara da "etkin olmayan oklar" denir. Kempe vd. DE ve BÇ yayılım modellerinin sadece etkin okların kullanılacağı şekilde bir eşdeğer gösterimi olduğunu ortaya koymuşlardır (Kempe *et al.*, 2003). Buna göre BÇ yayılım modeli için şu iki ifade eşdeğerdir:

1. Yayılıma S başlangıç kümesi ile başlayarak her adımda bağlantılı komşuları oklar üzerindeki olasılıklara göre belirleyip, yayılım sonunda beklenen yayılım değeri $\sigma(S)$ 'i hesapla.
2. Ağdaki okları $(1-p_{ij})$ olasılığı ile silerek sadece etkin oklardan oluşan G_r ağını oluştur. Bu yeni ağda S kümesinden erişilebilen (sadece etkin oklardan oluşan patikaları kullanan) tüm düğümleri sayarak $\sigma(S)$ 'i hesapla.

Benzer bir eşdeğerlik DE yayılım modeli için de kurulabilir. Rassal olarak $[0,1]$ aralığından üretilen bir π sayısı eğer $\sum_{j \in N_i'} p_{ji}$ toplamından büyükse i düğümüne giren hiçbir ok etkin değildir. Aksi takdirde öyle bir j^* düğümü vardır ki, $\sum_{j=1}^{j^*} p_{ji} \geq \pi$ ifadesini ilk sağlayan (i 'yi etkinleştiren) düğüm j^* 'dir ve (j^*,i) oku etkin ok olarak kaydedilir, i 'ye gelen diğer tüm oklar etkin olmayan oklar olarak silinir. Burada π sayısı i 'ye bağlı değildir, tamamen rassal üretilir. Aynı şekilde j düğümlerinin sırasıyla ilgili olarak da bir önceliklendirme bulunmamaktadır, rassal şekilde veya düğüm indislerine göre sıralanabilirler ama daha sonra her örneklem oluşturulmasında aynı sıranın kullanılması gereklidir. j düğümlerinin sırası, hangi (j^*,i) okunun etkinleşeceğine doğrudan etki eder ancak, herhangi bir şekilde sıralandıktan sonra sabit bir sıra eldeyen π değerinin rassal seçilmesi yeterlidir ve sıralama yönteminin önemi kalmamaktadır. Özetle, DE yayılım modelinde her bir düğüm için ya gelen oklardan birisi (j^*,i) etkindir, ya da hiçbir gelen ok etkin değildir. Bu çerçevede aşağıdaki iki ifade eşdeğerdir:

1. Yayılıma S başlangıç kümesi ile başlayarak her adımda rassal olarak belirlenen eşik değerlerine göre etkilenen düğümleri saptar

ve yayılım sonunda beklenen yayılım değeri $\sigma(S)$ 'i hesapla.

2. G_r 'yi yukarıda anlatılan şekilde oluştur ve bu ağda S kümesinden erişilebilen (sadece etkin oklardan oluşan patikayı kullanan) tüm düğümleri sayarak $\sigma(S)$ 'i hesapla.

“Tetikleyen küme” kavramı ise bu sözü geçen eşdeğerliklerin bu çerçeveye uyan tüm yayılım modelleri için genelleştirilmiş halidir. Her bir i düğümü için farklı yayılım modellerine göre bu düğümü etkin hale getirebilecek düğümleri rassal şekilde ortaya koyan bir τ_i dağılımı olduğu varsayalım. Bu dağılımdan ortaya çıkan T_i kümesi (örneklemi) bağlı olduğu yayılım modelinin zorunlu kıldığı aktif oklara göre belirlenir. Bu şekilde ortaya çıkan G_r ağı üzerinde S başlangıç kümesinin etkinleştirdiği düğümler ile hesaplanan beklenen etkin düğüm kümesinin büyüklüğü ile yayılımın doğrudan benzetimini yaparak elde edilen $\sigma(S)$ değerleri her zaman birbirine eşittir (Kempe *et al.*, 2003).

2.3 Matematiksel Model

EEP'yi ikili tamsayılı doğrusal programlama problemi olarak modelleyebilmek için öncelikle yayılım yöntemlerinin doğasında olan rassallığın uygun şekilde ifade edilmesi gerekir. Bu amaçla yayılım modeli üzerinden ortaya çıkabilecek bütün olası senaryolar birleştirilerek sonlu (ama çok büyük) bir senaryo kümesi (Ω) elde edilir. BÇ yayılım modelinde her bir ok (i,j) ya p_{ij} olasılıkla etkin, ya da $(1-p_{ij})$ olasılıkla etkin değildir. Dolayısıyla olası tüm senaryoların sayısı $|\Omega|=2^{|A|}$ 'dir. DE yayılım modelinde ise her bir i düğümü için senaryo sayısı, bu düğüme gelen okların sayısından bir fazladır, yani $|N_i|+1$ 'dir. Hatırlanacağı üzere DE yayılım modelinde i düğümünü ya gelen komşularından (sayısı $|N_i|$) biri etkiler ya da toplam etki eşik değerini aşmıyorsa hiçbir komşusu etkileyemez. Bu durumda olası senaryo sayısı $|\Omega|=\prod_{i=1}^{|V|}(|N_i|+1)$ eşitliğiyle hesaplanır. İki durumda da $|\Omega|$ sayısının çok hızlı şekilde arttığını ve küçük ölçekli ağlar için bile çok yüksek boyutlu eniyileme problemlerine neden olduğu kolayca görülür. Bu sebeple tüm senaryoları birleştirmek yerine ÖOY yöntemi

kullanılarak, kısıtlı sayıda senaryo ile ilerlemek daha uygun bir yaklaşımdır. ÖOY yöntemi stokastik eniyileme problemleri için sıklıkla kullanılan ve başlangıçtaki problemin çözümüne yakınsayan çözümlerin başarı ile elde edildiği bir yöntemdir (Homem-de-Mello ve Bayraksan, 2014). Bu yöntemde Monte Carlo örnekleme ile çeşitli sayıda senaryo oluşturulur ve amaç fonksiyonu her bir senaryo için hesaplandıktan sonra ortalaması alınarak başlangıçtaki amaç fonksiyonuna yakınsayan bir sonuca ulaşılır.

EEP'nin ÖOY formülasyonu için iki değişken kümesi tanımlanır. İlk olarak $y_i, i \in V$ ikili değişkenleri ile hangi düğümlerin S başlangıç kümesinde yer alıp almayacağı belirlenir. Daha sonra her bir $r \in R$ örnekleminde (senaryosunda) hangi düğümlerin etkinleştirildiğini yakalamak için x_{ir} ikili değişkenleri tanımlanır. Bu çerçevede aşağıdaki ikili tam-sayı model ortaya çıkar:

$$\max \frac{1}{|R|} \sum_{r \in R} \sum_{i \in V} x_{ir} \quad (1)$$

$$\text{öyleki } \sum_{i \in V} y_i = k \quad (2)$$

$$x_{ir} \leq \sum_{j \in P_{ir}} y_j \quad \forall i, r \quad (3)$$

$$0 \leq x_{ir} \leq 1, y_i \in \{0,1\} \quad \forall i, r \quad (4)$$

Bu modelde amaç fonksiyonu (1) ağıdaki etkinin tüm senaryolar üzerinden hesaplanan beklenen değerini enbüyükler. İlk kısıt (2) başlangıç kümesinde sadece k tane düğüm olmasını sağlar. (3) numaralı kısıtlara kısaca kapsama kısıtları denir. Burada bir i düğümünün herhangi bir r senaryosunda etkin olabilmesi için, bu düğümün en az bir tane başlangıç kümesinde yer alan düğüm tarafından (etkin oklar üzerinden) erişilebilir olması gerekmektedir. Burada P_{ir} kümesi i düğümünün r senaryosunda hangi başlangıç kümesi düğümleri tarafından erişilebildiğini göstermektedir. Bu kümelerin tespiti için her düğüm-senaryo ikilisi için ilgili senaryodaki etkin oklar üzerinden okların ters yönüne doğru genişlik öncelikli arama yapılır. Bu aramanın adımları sırasında temas edilen tüm düğümler ilgili P_{ir} kümesine eklenir. Modelde son olarak değişkenler üzerindeki kısıtlar ve negatif olamama koşulları (4) tanımlanır. x_{ir} değişkeni ikili değişken olarak

tanımlanmış olmasına rağmen, amaç fonksiyonunun enbüyükleme olması ve eniyi çözümde (3) numaralı kısıtın x_{ir} değerlerini üstten sınırlaması sayesinde bu değişkenler devamlı değişken olarak gevşetilebilir.

2.4 Örneklem Ortalama Yakınsaması Yaklaşımı

ÖOY yaklaşımının az önce tanıtılan modele uygulanmasında dikkat edilmesi gereken çeşitli noktalar vardır. Özellikle kaç örneklem alınacağı ve bunun sonunda oluşan kesikli eniyileme probleminin başlangıç problemine olan yakınsamasının kalitesi önemli başlıklardır. ÖOY ile ilgili çok miktarda çalışma ve geniş bir yazın bulunmaktadır. EEP özelinde anlamlı olan bazı özelliklere burada yer vermekteyiz.

Başlangıç probleminin ve ÖOY'nin eniyi amaç fonksiyon değerleri sırası ile z^* ve z_R^* olsun. Aynı şekilde y^* ve y_R^* bu iki problemin eniyi çözümleri (başlangıç kümeleri) olsun. Son olarak da $(y^*)_\varepsilon$ ve $(y_R^*)_\varepsilon$ bu iki problem için ε -eniyi çözüm kümeleri olsun. Başlangıç ve ÖOY problemlerimizin çözüm kümesi sonlu sayıda eleman içerdikleri için aşağıda yazılan dört koşul sağlanmaktadır:

1. ÖOY'nin amaç fonksiyonunun beklenen değeri, ilk problemin eniyi değerine bir üst sınır teşkil eder, yani, $\mathbf{E}[z_R^*] \geq z^*$,
2. Eğer $R \rightarrow \infty$ sağlanırsa, kesin şekilde $z_R^* \rightarrow z^*$ ve $y_R^* \rightarrow y^*$ sağlanır,
3. Yine, $R \rightarrow \infty$ olduğu durumda, ÖOY'nin eniyi çözümü olan y_R^* 'ın, ilk problemin eniyi çözümü olan y^* 'a yakınsama hızı üsseldir.
4. Herhangi bir, $\rho \in [0, \varepsilon)$ anlamlılık seviyesi verildiğinde ve örneklem sayısının $R \geq \frac{3\sigma_{max}^2}{(\varepsilon-\rho)^2} \log\left(\frac{|X|}{\alpha}\right)$ koşulunu sağlaması durumunda, $P\{(y_R^*)_\varepsilon \subseteq (y^*)_\varepsilon\} \geq 1 - \alpha$ olur, yani ÖOY probleminden elde edilen herhangi bir ρ -eniyi çözüm en az $1 - \alpha$ olasılıkla ilk probleme ε -eniyi bir çözüm olur (Homem-de-Mello ve Bayraksan, 2014).

Burada σ_{max}^2 eniyi çözüm ile diğer çözümler arasındaki varyansı gösterirken $|X|$ ise problemin çözüm kümesinin büyüklüğünü ifade eder. Problemimizde y değişkenleri için olası çözüm sayısı $O(|V|^k)$ mertebesinde olduğu için burada teorik

olarak tespit edilen örneklem sayısı pratik uygulamalar için fazla yüksek çıkabilir. Yine de bu ifadenin logaritması mertebesinde bir örneklem sayısı gerektiği için aslında problem büyüklüğü ile örneklem sayısının ilişkisi $O\left(\frac{|V|^2}{(\varepsilon-\rho)^2} k \log|V|\right)$ seviyesindedir (Homem-de-Mello ve Bayraksan, 2014).

2.5 Lagrange Gevşetmesi

Lagrange Gevşetmesi (LG) büyük ölçekli kombinatoryal eniyileme problemlerinin hızlı çözümünde sık kullanılan yöntemlerden birisidir (Fischer, 1981). Bu yöntemde, başlangıç problemini karmaşıklaştıran ve dolayısıyla çözümünü zorlaştıran bazı kısıtlar gevşetilerek, belirli bir ceza katsayısı ile çarpılmış şekilde amaç fonksiyonuna taşınır. ÖOY formülasyonunda kardinalite kısıtı (2) veya kapsama kısıtları (3) gevşetilebilir. Kardinalite (2) kısıtı gevşetilirse formülasyonun her bir örneklem bazında ayrıştırılabileceği görülür. Stokastik eniyileme yazınında bu yaklaşım ile yapılan çok sayıda ayrıştırma çalışması bulunmaktadır. Kumar vd.'nin çalışmasında önce çekirdek kümesini temsil eden y_i değişkenlerinin her senaryo için bir kopyası oluşturulur (y_{ir}), ardından bu kısıtlar gevşetilerek formülasyonun senaryo bazında ayrışımı sağlanır. Alt-gradyen eniyilemesi yöntemiyle de bu gevşetme eniyi çözüm aranır. Bu çalışmada ise düğüm tabanlı bir gevşetme stratejisi izlemek adına kapsama kısıtlarının gevşetilmesi önerilmiştir. Her bir (i,r) kapsama kısıtı λ_{ir} dual değişkeni (ceza katsayısı) ile çarpılarak amaç fonksiyonuna taşınır ve aşağıdaki LG formülasyonu elde edilir:

$$\max \frac{1}{|R|} \sum_{r \in R} \sum_{i \in V} (1 - \lambda_{ir}) x_{ir} + \sum_{i \in V} \left(\sum_{r \in R} \sum_{j \in F_{ir}} \lambda_{jr} \right) y_i \quad (5)$$

$$\text{öyleki } \sum_{i \in V} y_i = k \quad (6)$$

$$0 \leq x_{ir} \leq 1, y_i \in \{0,1\} \quad \forall i, r \quad (7)$$

Görülebileceği gibi bu formülasyon tüm x_{ir} ve y_i değişkenleri üzerinden ayrıştırılabilir ve basit gözlemlerle çözülebilir. Öyleki, eğer $\lambda_{ir} \geq 0$ ise, $x_{ir} = 1$ olur. Çekirdek kümesini tespit etmek için ise y_i değişkenleri $C_i = \sum_{i=1}^n \left(\sum_{r=1}^R \sum_{j \in F_{ir}} \lambda_{jr} \right)$ kat sayıları üzerinden azalan şekilde sıralanır ve en büyük k

tanesine denk gelen y_i değişkenleri bir yapılıdır. Ayrıca negatif olmayan her λ_{ir} kümesi için LG probleminin amaç fonksiyonu başlangıç problemine bir üst sınır oluşturur. Dolayısıyla artık hedefimiz LG-dual problemini λ_{ir} dual değişkenleri bazında enküçükmektir. Bu formülasyonda F_{ir} kümesi, problemi dualize etmemiz nedeniyle ortaya çıkar ve i düğümünün r senaryosundaki ardıl düğümlerini gösteren kümeyi temsil eder. Bu tanımlar ışığında oluşturulan LG sezgiselinin detaylı anlatımı bir sonraki alt bölümde sunulmuştur.

2.6 Lagrange Sezgiseli

Lagrange sezgiselinin başlangıç adımındaki önemli unsurlardan bir tanesi λ_{ir} dual değişkenlerinin başlangıç değerleridir. Burada tercihimiz $\lambda_{ir} = 1$ şeklinde olmasıdır. Böylece başlangıç çözümünde hem tüm $x_{ir} = 0$ olur hem de her düğüm için C_i değeri o düğümün toplamda (tüm senaryolar bazında) kaç ardılını etkinleştirdiğini gösterir. Böylece başlangıç çözümü, örneklem sonucu ortaya çıkan tüm senaryolarda, bireysel olarak en çok düğümü etkinleştiren k bireye denk gelir. Bu şekilde hesaplanan LG amaç fonksiyonu başlangıç problemi için ilk üst sınırı oluşturur. Ayrıca, bu başlangıç çözümü ile oluşturulan çekirdek kümenin eriştiği tekil düğümlerin sayısı her bir senaryo için hesaplanıp toplandığında başlangıç problemi için olurlu bir çözüm elde edilir. Bu toplam, başlangıç probleminin amaç fonksiyon değeri için bir alt sınırdır.

Lagrange sezgiseli alt-gradyen eniyilemesi ile birlikte göz önüne alındığında iteratif bir yöntemdir. Z_{LG} ve Z_{LG}^* değerleri sırasıyla LG probleminin her bir adımdaki mevcut ve en iyi (enküçük) amaç fonksiyonu değerlerini gösterir. Benzer şekilde Z_{LB} ve Z_{LB}^* değerleri de mevcut ve eniyi (enbüyük) olurlu çözüme denk gelen amaç fonksiyonu değerlerini gösterir. T ile o anki en iyi alt sınır olan Z_{LG}^* değerinin en son değişiminden bu yana geçen alt-gradyen adım sayısı gösterilirken τ ile alt-gradyen eniyilemesinde kullanılan adım uzunluğu ölçek parametresi tanımlanır. Bu açıklamaların ve tanımlamaların ardından Lagrange Sezgiseli'nin adımları aşağıdaki gibidir:

Adım 0 (Sıfırlama). Lagrange parametrelerini sıfırla, öyleki, $\lambda_{ir} = 1, T = 0, \tau = 2, \varepsilon = 0.01$. Her bir düğüm i ve senaryo r için ardıl erişilebilir düğüm kümeleri, F_{ir} 'leri derinlik öncelikli arama ile tespit et. Bu aşamada her bir düğüm için $C_i = |\cup_{r \in R} F_{ir}|$ değerlerini hesapla. C_i değerlerini azalacak şekilde sırala ve ilk k tanesini seçerek başlangıç çözümü y_0 'ı elde et. Bu çözüm üzerinden Z_{LG}^0 ve Z_{LB}^0 amaç fonksiyon değerlerini hesapla. $Z_{LG}^* = Z_{LG}^0$ ve $Z_{LB}^* = Z_{LB}^0$ şeklinde başlangıç alt ve üst sınırları ata. Adım 3'e geç.

Adım 1. LG'yi mevcut λ_{ir} 'leri kullanarak basit gözlem ile çözerek \hat{y} ve \hat{x} anlık çözümlerini ve Z_{LG} değerini elde et. Eğer $Z_{LG} \leq Z_{LG}^*$ ise, $T = 0, Z_{LG}^* = Z_{LG}$ ve $y^* = \hat{y}$ atamalarını yap. Aksi halde $T = T + 1$ atamasını yap ve Adım 2'ye geç.

Adım 2. Yayılım sürecini mevcut \hat{y} başlangıç kümesinden başlayacak şekilde her bir senaryo için koştur ve erişebilenen tekil düğüm sayısını hesaplayarak Z_{LB} değerini elde et. Eğer $Z_{LB} \geq Z_{LB}^*$ ise, $Z_{LB}^* = Z_{LB}$ ve $y^* = \hat{y}$ atamalarını yap.

Adım 3. Eğer $Z_{LG}^* = Z_{LB}$ ise dur, hem ÖÖY hem de LG probleminin en iyi çözümü mevcut y^* değeri ve en iyi amaç fonksiyonu değeri de $Z_{LG}^* = Z_{LB}$ 'dir. Aksi takdirde Adım 4'e geç.

Adım 4. Eğer $T = 30$ ise, tam 30 adımdır alt-gradyen eniyilemesi alt sınır Z_{LB} değerini iyileştiremediğinden adım uzunluğu parametresini $\tau = \tau/2$ şeklinde yarıla ve $T = 0$ yaparak sayacı sıfırla.

Adım 5. Her bir $i \in V, r \in R$ için alt-gradyen değerlerini hesapla, $G_{ir} = x_{ir} - \sum_{j \in P_{ir}} y_j$.

Adım 6. Eğer $\sum_{r \in R} \sum_{i \in V} G_{ir} = 0$ veya $\tau < \varepsilon$ veya zaman limiti aşılsa Adım 8'a git.

Adım 7a. Lagrange katsayılarını güncellemede kullanılacak yeni adım büyüklüğünü hesapla, $\psi = \frac{\tau(1.05 Z_{LG}^* - Z_{LB}^*)}{\sum_{r \in R} \sum_{i \in V} (G_{ir})^2}$.

Adım 7b. Lagrange katsayılarını güncelle. $\lambda_{ir} = \max\{0, \lambda_{ir} + \psi G_{ir}\}$. Eğer λ_{ir} değerinde bir değişiklik olduysa, i düğümü tarafından herhangi bir

r senaryosunda erişilebilen tüm düğümlere ait $C_j, j \in P_{ir}$ değerlerini güncelle. Adım 1'e git.

Adım 8 (Sonuç). y^* kümesini en iyi çekirdek küme, z_{LG}^* değerini ÖOY problemi amaç fonksiyonu için en iyi üst sınır ve z_{LB}^* değerini de en iyi alt sınır olarak kaydet.

Lagrange sezgiselinde kullanılan parametrelerin ilk değerleri genel olarak deneysel gözleme dayanmaktadır ve probleme özgü şekilde algoritmanın farklı performans göstermesine neden olabilir. λ_{ir} değerlerinin neden bire eşitlendiği alt bölüm başında açıklanmıştır. Adım uzunluğu parametresi τ için ise $\tau = 2$ değeri yazında en çok önerilen değer olması sebebiyle tercih edilmiştir (Fischer, 1981). Lagrange sezgiseli olurlu bir başlangıç çözümü ile başlar ve her adımda alt ve üst sınırları birbirine yaklaştırmaya çalışır. Eğer zaman ve adım uzunluğu sınırlarına gelmeden bu sağlanırsa ÖOY problemi için eniyi çözüm kısa sürede elde edilir. Ancak bazı durumlarda alt ve üst sınır arasındaki farkın azalma hızı çok düşebilir. Bu da algoritmanın çok uzun süre çalışmasına (özellikle büyük boyutlu ağlarda) neden olabilir. Bu sebepten ötürü Adım 3'te $z_{LG}^* = z_{LB}^*$ eşitliğini talep etmek yerine $z_{LG}^* - z_{LB}^* < \epsilon'$ şeklinde bir yaklaşık eşitlik yakalamak da yeterli görülebilir.

3. Bulgular

Bu bölümde, Lagrange Sezgiseli'nin eniyi / eniyeye çok yakın çözümleri, var olan eniyileme yöntemlerine göre çok daha kısa sürede bulup bulamadığını sınamak için yapılan kapsamlı deneysel çözümler sunulmuştur.

3.1 Deneysel Ortam

Yapılan bütün deneyler Windows 10 işletim sistemi ile çalışan 64GB RAM ve 2.6GHz Xeon işlemcili bir iş istasyonunda gerçekleştirilmiştir. Algoritmalar MS Visual Studio 2017 Community arayüzünden C# dili ile .NET Core Konsol Uygulaması olarak kodlanmıştır. Deneylerde Lagrange Sezgiseli (LS), Wu ve Küçükyavuz tarafından geliştirilmiş olan ve halihazırda en iyi performansı gösteren Ertelenmiş Kesi Oluşturma (EKO) yöntemi ile kıyaslanmıştır. Bir

önceki bölümde ayrıntıları verilmiş olan LS'de tüm alt problemler basit gözlem ile çözülebildiği için herhangi bir ticari çözücüye ihtiyaç duyulmamıştır. EKO yönteminde ise Benders ana-problemi çözümünde ticari çözücü gerekmektedir. Bu amaçla Gurobi 8.0 çözücüsü (Gurobi, 2019) standart ayarları ile kullanılmıştır.

Bu çerçevede (Wu ve Küçükyavuz, 2017) çalışmasında kullanılan dört gerçek yaşam tabanlı sosyal ağ verileri kullanılmıştır. Bu sosyal ağ verilerine Stanford Büyük Ağ Verileri (SNAP) web sayfasından (İnt. Kyn. 1) ulaşılabilir. İlgili sosyal ağların özelliklerini içeren özet bilgiler Çizelge 1'de sunulmuştur.

UCI (University of California Irvine), California Üniversitesi öğrencilerinin kendi aralarında kullandıkları bir mesajlaşma platformudur. Kişiler arasında gönderilen mesajlar üzerinden düğüm ve kenarlar oluşturulmuştur. GNU veri seti Gnutella dosya paylaşım uygulamasını kullanan kişilerin oluşturduğu ağdan elde edilmiştir. HEP veri seti ise açık kaynak akademik paylaşım ortamı olan arXiv platformundaki Yüksek Enerji Fizik bilim dalı ile ilgili yayınlardaki atıflar üzerinden oluşturulmuştur. Burada yazarlar düğüm olarak tanımlanırken, atıf yapılan yayındaki yazarlardan atıf yapan yayındaki yazarlara doğru oklar oluşturulur. Bu deneysel seti EEP yazınındaki pek çok çalışmada da kullanılmıştır (Kempe et al. 2003, Chen et al. 2009, Tang et al. 2015, Güney 2017). Son olarak ENRON veri setinde, Enron firması çalışanlarının kendi aralarında yaptıkları e-posta yazışmaları üzerinden bir sosyal ağ oluşturulmuştur.

Çizelge 1. Deneysel Analizlerde Kullanılan Gerçek Sosyal Ağlar ile ilgili Özet Bilgiler

Sosyal Ağ	Kategori	Düğüm Sayısı	Kenar Sayısı	Ağ Tipi
UCI	Mesajlaşma	1,899	59,835	Yönlü
GNU	P2P	10,876	39,994	Yönsüz
HEP	Atıf	37,153	231,568	Yönsüz
ENRON	E-Posta	36,692	367,662	Yönlü

Görüleceği üzere UCI ve ENRON veri setlerindeki kenarlar yönlü iken, GNU ve HEP'teki kenarlar

yönsüzdür. EEP’de yönlü oklara ihtiyaç duyduğumuz için, yazında çok uygulanan, yönsüz kenarları iki tersdüz okla değiştirme yöntemi uygulanarak yönlü ağlar elde edilmiştir.

Deneyler hem Bağımsız Çağlayan hem de Doğrusal Eşik yayılım modelleri için yapılmıştır. Bağımsız Çağlayan yayılım modeli için oklardaki etkileme ağırlıkları $p=0.1$ olarak her ok için eşit tutulmuştur. Doğrusal Eşik modelinde ise her bir düğüm için önce içe gelen ok sayısını temsil eden $N_i, i \in V$ değeri hesaplanmıştır. Daha sonra da o düğüme gelen her bir okun ağırlığı için $1/N_i$ değeri alınmıştır. Deneylerde başlangıç küme büyüklüğü için $k=2,3,4,5$ değerleri kullanılırken, senaryo sayısı için $R=100,200,300,500$ değerleri kullanılmıştır. Bu çalışmada kullanılan ok ağırlık değeri ($p=0.1$) ve parametre değerleri (k,R) Wu ve Küçükyavuz (2017) çalışmasında da birebir kullanılan değerlerdir. Özetle, 4 deney seti, 2 yayılım modeli, 4 başlangıç kümesi boyutu ve 4 de senaryo sayısı değeri olmak üzere, toplam 128 ayrı vaka elde edilmiş ve hepsi 10’ar kez LS ve EKO yöntemleri ile çözümlenerek ortalama değerler (çözüm ve çözüm süresi) hesaplanmıştır.

3.2 Yöntemlerin Performans Kıyaslamaları

Önceki bölümde tanıtılan 4 veri seti kullanılarak, 2 ayrı yayılım modeli ve 16 farklı başlangıç kümesi ve senaryo sayısı kombinasyonu için elde edilen sonuçlar Çizelge-2,3,4,5’te sunulmuştur. Bu çalışmada amaç, eniyi / eniyiye yakın çözümleri en kısa sürede elde etmek olduğu için ilk kıyaslama LS ve EKO’nun çözüm süreleri üzerinden yapılmıştır. İki yöntemde de 3600 saniye süre limiti ve %1 eniyilik aralığı koşulu (eniyi çözüme olan uzaklığın %1 veya daha az olması) varsayılmıştır.

Çizelge 2. UCI ve GNU Veri Setleri İçin Koşma Süresi Kıyaslamaları (Bağımsız Çağlayan Yayılım Modeli)

		UCI		GNU	
k	R	t_{LS}	t_{EKO}	t_{LS}	t_{EKO}
2	100	3.3	53.8	19.3	392.4
3	100	3.6	60.5	40.2	399.5
4	100	4.1	54.2	10.0	465.9
5	100	4.9	57.2	13.5	433.0
2	200	6.1	67.5	29.3	398.0
3	200	7.2	68.3	34.7	430.0
4	200	9.2	63.0	28.5	434.4

5	200	10.8	66.1	73.1	466.8
2	300	9.5	203.2	52.0	194.5
3	300	10.3	197.9	45.9	198.2
4	300	11.9	220.1	65.6	228.6
5	300	16.1	203.2	36.8	233.6
2	500	15.5	153.9	111.8	389.5
3	500	17.3	165.2	87.5	399.4
4	500	19.8	155.2	74.7	536.1
5	500	28.3	158.4	114.2	482.7
Ortalama		11.1	121.7	52.3	380.2

Çizelge 2.’de UCI ve GNU veri setleri ile ilgili Bağımsız Çağlayan yayılım modeline göre elde edilen sonuçlar listelenmiştir. Tüm senaryolarda LS yöntemi EKO yönteminden iyi performans göstermiş ve ortalama hızlanma UCI’da 10 kat, GNU’da 7 kattan fazla olmuştur.

Çizelge 3. HEP ve ENRON Veri Setleri İçin Koşma Süresi Kıyaslamaları (Bağımsız Çağlayan Yayılım Modeli)

		HEP		ENRON	
k	R	t_{LS}	t_{EKO}	t_{LS}	t_{EKO}
2	100	370.4	477.9	852.4	1487.7
3	100	369.0	587.1	815.5	1316.9
4	100	368.8	823.8	815.7	1217.6
5	100	368.9	284.2	893.5	1251.7
2	200	373.8	204.0	1680.1	3223.4
3	200	374.5	358.3	1575.6	2695.1
4	200	370.7	1515.4	1497.6	2767.5
5	200	372.2	573.8	1655.6	2818.3
2	300	375.7	391.3	2699.3	3896.4
3	300	378.6	421.1	2593.0	3792.5
4	300	377.9	653.9	2608.9	4768.6
5	300	377.4	572.9	2564.2	3926.8
2	500	403.7	480.5	2017.8	6507.6
3	500	394.5	1013.3	4275.5	7304.7
4	500	418.7	1102.9	4479.5	7369.3
5	500	408.1	1841.1	3971.1	6965.7
Ortalama		381.4	706.4	2187.2	3831.9

Çizelge 3’te ise daha büyük boyutlu olan HEP ve ENRON veri setleri yine Bağımsız Çağlayan yayılım modeli varsayımı altında test edilmiştir. Burada LS’nin performansı çoğu veride EKO’dan iyi olmasına rağmen fark bir önceki çizelgedeki sonuçlara göre azalmıştır. Veri boyutunun büyümesi her iki algoritmayı da oldukça olumsuz etkilemektedir.

Çizelge 4. UCI ve GNU Veri Setleri İçin Koşma Süresi Kıyaslamaları (Doğrusal Eşik Yayılım Modeli)

		UCI		GNU	
k	R	t_{LS}	t_{EKO}	t_{LS}	t_{EKO}
2	100	9.4	36.0	1.0	20.1
3	100	10.7	61.2	10.4	26.1
4	100	11.6	90.2	47.7	21.2

5	100	12.3	68.3	47.0	26.7
2	200	17.6	73.3	13.0	171.1
3	200	21.8	69.7	87.4	207.7
4	200	23.0	137.8	89.1	183.6
5	200	24.2	288.8	82.6	269.6
2	300	27.5	169.0	2.2	460.9
3	300	34.4	130.4	123.3	514.0
4	300	34.4	235.7	138.3	743.8
5	300	37.8	358.2	141.8	1578.1
2	500	45.9	76.9	217.4	525.8
3	500	55.9	138.7	228.6	426.9
4	500	62.5	127.5	254.8	475.6
5	500	69.3	423.1	249.3	959.4
Ortalama		31.2	155.3	108.4	413.2

Çizelge 5. HEP ve ENRON Veri Setleri İçin Koşma Süresi Kıyaslamaları (Doğrusal Eşik Yayılım Modeli)

k	R	HEP		ENRON	
		t _{LS}	t _{EKO}	t _{LS}	t _{EKO}
2	100	27.8	87.1	158.5	169.2
3	100	97.6	130.2	269.4	275.6
4	100	125.7	135.4	248.8	373.4
5	100	159.6	291.4	231.3	48.4
2	200	76.4	551.0	450.4	438.2
3	200	136.7	198.7	460.5	422.0
4	200	337.2	282.5	468.0	530.7
5	200	224.1	369.8	418.7	599.4
2	300	146.3	262.0	632.5	340.9
3	300	152.8	346.8	677.3	439.7
4	300	238.0	335.3	677.6	834.1
5	300	646.4	795.8	644.3	630.3
2	500	12.7	588.2	1046.8	1213.2
3	500	525.6	557.7	1068.3	1271.2
4	500	550.8	824.6	1218.4	1596.3
5	500	302.6	942.3	1265.1	1785.3
Ortalama		235.0	418.7	621.0	685.5

Benzer analizler bu sefer Doğrusal Eşik yayılım modeli varsayımı ile yapılmış ve bulgular Çizelge 4 ve Çizelge 5'te gösterilmiştir. Benzer desenler bu tablolarda da gözükmemektedir. LS daha küçük boyutlu olan UCI ve GNU veri setlerinde hemen hemen bütün vakalarda daha iyi sonuç vermiştir ve ortalama 4-5 kat hızlanma sağlamıştır. Daha büyük boyutlu veri setlerinde ise hızlanma 2 kat civarındadır.

Başka bir gözlem ise k değeri büyüdükçe çözüm sürelerinin birkaç istisna dışında uzamasıdır. Bunun temel nedeni k değeri büyüdükçe V düğüm kümesi içinden yapılabilecek k seçim sayısının, yani çözüm uzayı büyüklüğünün kombinatoryal şekilde artmasıdır.

3.3 Lagrange Sezgiseli ile İlgili Diğer Özellikler

Bu bölümde Lagrange Sezgiseli ile ilgili iki önemli özellik hakkında deneysel veriler sunulmuştur. Bir önceki bölümden hatırlanacağı üzere iki yöntem için de %1 eniyilik aralığı varsayımı kullanılmıştır. LS'de bu aralık algoritma koşumu sırasında tespit edilen en iyi alt ve üst sınırlar kullanılarak hesaplanmaktadır. Dolayısıyla LS'nin bulunduğu eniyilik aralığı aslında gerçek problemin eniyi çözümü ile LS'nin en iyi çözümü (alt sınırı) üzerinden hesaplanacak olan gerçek eniyilik aralığı için de bir üst sınırdır. Ayrıca LS sezgiseli ile ilgili önemli bir bilgi de kaç adım sonunda algoritmanın durduğudur. Bazı durumlarda ve bazı problem tiplerinde LS sezgiselinde alt ve üst sınırdaki iyileşmeler son adımlarda azalarak algoritmanın çok uzun süreler çalışmasına neden olabilmektedir. Çizelge 6 ve Çizelge 7'de iki farklı yayılım modeli için dört farklı veri setinden elde edilen sonuçlar verilmiştir. “%” işareti LS algoritması durduğunda (adım büyüklüğü parametresi yeterince küçülürse veya alt ve üst sınır arasındaki fark %1'in altına inerse) alt ve üst sınır arasındaki yüzdesel farkı göstermektedir. “#” işareti ise LS'nin kaç adım sonunda durduğunu göstermektedir. Bağımsız Çağlayan yayılım modeli sonuçlarına bakıldığında HEP dışında diğer üç veri setinde de düşük adım sayısı ve %1'den küçük eniyilik aralığı görülmüştür. HEP veri setinde ise açıklıklar %1'den büyüktür, bu durumda algoritma, ancak adım parametresinin yeterince küçüldüğü durumda durmuştur.

Çizelge 6. Lagrange Sezgiseli Eniyilik Aralığı ve Adım Sayısı (Bağımsız Çağlayan Yayılım Modeli)

k	R	UCI		GNU		HEP		ENRON	
		%	#	%	#	%	#	%	#
2	100	0.22	2	0.72	8	1.33	75	0.26	13
3	100	0.24	2	0.70	7	2.34	273	0.21	14
4	100	0.22	2	0.74	10	2.34	277	0.23	13
5	100	0.21	2	0.79	12	2.34	275	0.24	14
2	200	0.31	2	0.87	21	2.28	67	0.40	12
3	200	0.31	2	0.82	8	3.37	269	0.33	21
4	200	0.33	2	0.70	8	3.37	279	0.34	19
5	200	0.33	2	0.60	10	3.57	276	0.36	14
2	300	0.37	2	0.75	4	3.12	130	0.47	18
3	300	0.45	2	0.64	8	4.22	263	0.44	15
4	300	0.41	2	0.87	14	4.61	277	0.47	14
5	300	0.44	2	0.71	7	4.63	276	0.48	12
2	500	0.30	2	0.41	8	3.92	77	0.58	11
3	500	0.30	6	0.31	39	4.90	256	0.62	12
4	500	0.28	3	0.61	7	5.43	270	0.56	20
5	500	0.30	5	0.72	24	5.34	285	0.56	13
Ortalama		0.31	2.4	0.68	12.0	3.57	226.6	0.41	14.7

Çizelge 7. Lagrange Sezgiseli Eniyilik Aralığı ve Adım Sayısı (Doğrusal Eşik Yayılım Modeli)

k	R	UCI		GNU		HEP		ENRON	
		%	#	%	#	%	#	%	#
2	100	1.41	11	0.12	6	0.66	23	0.89	150
3	100	3.25	16	0.09	67	0.86	101	3.56	261
4	100	5.94	20	1.07	282	0.83	137	5.89	277
5	100	6.52	20	2.08	294	0.84	160	3.98	273
2	200	1.64	48	0.24	41	0.16	46	1.76	258
3	200	2.71	50	1.32	287	0.89	82	2.17	260
4	200	6.22	36	2.00	287	0.73	203	4.75	281
5	200	6.10	46	1.73	284	0.70	154	6.73	275
2	300	2.66	42	0.44	5	0.73	63	2.42	250
3	300	3.31	24	0.57	283	0.68	66	3.18	269
4	300	4.98	60	1.04	280	0.71	102	5.05	277
5	300	7.52	39	1.18	298	1.51	278	5.13	276
2	500	2.49	42	0.93	281	0.42	11	1.53	260
3	500	3.99	48	0.91	285	0.90	127	1.73	261
4	500	4.96	56	1.76	293	0.79	133	4.65	272
5	500	7.57	64	1.17	293	0.87	74	6.70	282
Ortalama		4.46	38.9	1.04	222.9	0.77	109.4	3.76	261.4

Doğrusal Eşik yayılım modeli durumundaki sonuçlarda ise UCI ve ENRON veri setlerinde eniyilik açıklıklarının yüksek olduğu görülmüştür. Adım sayısı ise kararsız davranışlar sergilemiştir. Bu bilgiler ışığında LS'nin hızlı sonuç bulma yetkinliğine karşın veriye olan hassasiyeti nedeniyle bazı durumlarda kötü performans gösterme riski bulunmaktadır. Bu çerçevede LS'nin dengelenmesine yönelik yazında yer alan (kutu yöntemi gibi) çeşitli yöntemlerden faydalanılabilir.

4. Tartışma ve Sonuç

Bu çalışmada son yıllarda birçok araştırmacının ilgisini çeken ve çok sayıda çalışma yapılan çevrimiçi sosyal ağlardaki en etkin kişileri bulan Etki Enbüyükleme Problemi üzerinde durulmuştur. Yazındaki çoğu çalışmanın aksine problemin eniyi çözümünden çok fazla ödün vermeden ölçeklenebilir ve hızlı bir yöntem geliştirme amacıyla Lagrange gevşetmesi tabanlı bir sezgisel yöntem geliştirilmiştir. EEP, kesin çözümü çok zor olan stokastik bir eniyileme problemi olduğu için aynı problemin Örneklem Ortalama Yakınsaması kullanılarak Lagrange sezgiseli, daha küçük boyutlu bu problem üzerinde koşturulmuştur. LS'nin performansı yazındaki mevcut en iyi yöntem olan Ertelenmiş Kesi Oluşturma (EKO) yöntemi ile kıyaslanmıştır. EEP yazınında çok sık kullanılan dört gerçek hayat sosyal ağ veri seti üzerinde yapılan

deneyler sonucunda LS'nin ortalama 2 ila 10 kat arasında bir hızlanma sağladığı görülmüştür. Ayrıca çoğu vakada azalan çözüm süresine rağmen eniyilik aralığının %1'in altında kaldığı gözlemlenmiştir. Deneysel sonuçlar, LS'nin iyi performansına rağmen zaman zaman kararsız bir davranış izleyerek yüksek oynaklık gösterdiğini tespit etmiştir. Gelecek çalışmalarda bu oynaklığı azaltacak yöntemler üzerinde durulabilir. Ayrıca LS benzeri farklı ayrıştırma yöntemleri de denenerek çözüm süresini hızlandıracak seçenekler geliştirilebilir.

5. Kaynaklar

- Chen, W., Wang, Y. and Yang, S., 2009. Efficient influence maximization in social networks. *Proceedings of the 15th ACM SIGKDD International conference on Knowledge Discovery and Data Mining, ACM*, 199-208.
- Fischer, M., 1981. The lagrangian relaxation method for solving integer programming problems. *Management Science*, **27**, 1-18.
- Galhotra, S., Arora, A. and Roy, S., 2016. Holistic influence maximization: Combining scalability and efficiency with opinion-aware models. *Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16, ACM*, 743-758.
- Goldenberg, J., Libai, B. and Muller, E., 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 211-223.
- Granovetter, M., 1978. Threshold models of collective behaviour. *The American Journal of Sociology*, **83**, 1420-1443.
- Gurobi Optimization, 2019, Gurobi 8.0 User's Manual.
- Güney, E., 2019. On the optimal solution of budgeted influence maximization problem in social networks. *Operational Research*, **19**, 817-831.
- Güney, E., 2019. An efficient linear programming based method for the influence maximization problem in social networks. *Information Sciences*, **503**, 589-605.

- Günneç, D., 2018. Etki Enbüyükleme Problemi için Ajan-bazlı Modelleme Yaklaşımı (An Agent-based Modeling Approach for the Influence Maximization Problem). *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, **18**, 701-709.
- Günneç, D., Raghavan, S., Zhang, R. 2019. Least Cost Influence Maximization on Social Networks. *Inform's Journal on Computing*, (Online First), <https://doi.org/10.1287/ijoc.2019.0886>, 1-15.
- Gürsoy, F. and Günneç, D., 2018. Influence Maximization in Social Networks under Deterministic Linear Threshold Model. *Knowledge-Based Systems*, **161**, 111-123.
- Homem-de-Mello, T. and Bayraksan, G., 2014. Monte carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, **19**, 56-85.
- Kempe, D., Kleinberg, J. and Tardos, E., 2003. Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, ACM, 137-146.
- Ko, Y.Y., Cho, K.J. and Kim, S.W., 2018. Efficient and effective influence maximization in social networks: A hybrid-approach. *Information Sciences*, **465**, 144-161.
- Kumar, A., Wu, X. and Zilberstein, S., 2012. Lagrangian relaxation techniques for scalable spatial conservation planning. *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, **439**, 309-315.
- Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. and Glance, N., 2007. Cost effective outbreak detection in networks. *Proceedings of the 13th ACM SIGKDD International conference on Knowledge Discovery and Data Mining*, ACM, 420-429.
- Li, Y., Fan, J., Wang, Y. and Tan, K. L., 2018. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, **1**, 1852-1872.
- Nguyen, H. T., Thai, M. T. and Dinh, T. N., 2016. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. *Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16*, ACM, 695-710.
- Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J. and Jia, W., 2018. Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, **106**, 17-32.
- Tang, Y., Shi, Y., and Xiao, X., 2015. Influence maximization in near-linear time: A martingale approach. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, SIGMOD15*, ACM, 1539-1554.
- Wu, H. H. and Küçükyavuz, S., 2017. A two stage stochastic programming approach for influence maximization in social networks. *Computational Optimization and Applications*, **69**, 1-33.

İnternet kaynakları

1-<http://snap.stanford.edu/data>, (09.06.2019)