

Rastsal Çizge Modellerinin Yapısal Analizi ile Karmaşık Ağların Sınıflandırılması

Classification of Complex Networks Using Structural Analysis of Random Graph Models

Ali Baran Taşdemir
Bilgisayar Mühendisliği Bölümü
Hacettepe Üniversitesi
Ankara, Türkiye
alibaran@tasdemir.us

Barkın Atasay
Bilgisayar Mühendisliği Bölümü
Hacettepe Üniversitesi
Ankara, Türkiye
barkinatasy96@gmail.com

Lale Özkahya
Bilgisayar Mühendisliği Bölümü
Hacettepe Üniversitesi
Ankara, Türkiye
ozkahya@cs.hacettepe.edu.tr

Özetçe —Karmaşık ağlar, günümüzde, sosyal ilişkiler, beyin aktiviteleri, moleküler yapıları temsil edip, çizge olarak özelliklerinin anlaşılabilmesi ve tahmin yürütülebilmesi için kapsamlı olarak incelenmiştir. Bu çalışmada, gerçek hayattan alınan farklı ağların çizgesel özellikleri kullanılarak, rastsal çizge modellerine göre sınıflandırılması yapılmıştır. Sınıflandırma işleminde, bu iş için en uygun görülen makine öğrenme algoritmaları kullanılmış ve performansları karşılaştırılmıştır. Sınıflandırma işleminin başarı oranını yükseltmek için, rastsal çizge modellerince üretilen sentetik çizgeler de kullanılmıştır. Son olarak, çizge özelliklerinin performansa etkileri istatistiksel yöntemlerle farklı altgruplara ayrılarak incelenmiştir.

Anahtar Kelimeler—sosyal ağların analizi, ağların sınıflandırılması, rastsal çizge modeli.

Abstract—Complex networks representing social interactions, brain activities, molecular structures have been studied widely nowadays to be able to understand and predict their characteristics as graphs. In this study, various real-world networks have been classified according to random graph models by making use of graph features. In the classification process, the most suitable machine learning algorithms are used and their performances are analyzed. Also, synthetic graphs generated by the random graph models are used in order to increase the success rate of the classification. Finally, the graph features are divided into different groups by using statistical tools to study their influence on the performance.

Keywords—Analysis of social networks, network classification, random graph model.

I. GİRİŞ

Karmaşık ya da sosyal ağlar, biyoloji, fizik, bilgisayar mühendisliği gibi bilimin birçok dalında karşımıza çıktığı için, onları nasıl oluştuğunu ve yapısal özelliklerini anlamak temel çalışma konularından biri haline gelmiştir. Gerçek hayat problemlerinden ortaya çıkan bu çizgelere (graph), sosyal ağlar, ortak-yazarlık (co-authorship) ağları, beyin ağları, moleküler ağlar örnek verilebilir. Ağların yapısının incelenmesi, çizge özellikleri hakkında bilgi verecek ve bu karmaşık ağların

da davranışını anlamaya yardım edecektir. Bu tür ağların, derece dağılımı gibi birçok ortak özelliği bulunmaktadır. Örnek olarak, Facebook gibi sosyal ağlarda, aralarında ortak arkadaşı olan iki kişinin de kendi aralarında arkadaş olma sıklığı, bu ağa karşılık gelen çizgede üçgenlerin frekansı ile birebir ilgilidir. Öte yandan, her ağın kendine özgü karakteristikleri bulunmaktadır ve bu sebeple, bu ağların tanımlanması ve davranışının tahmin edilmesine yönelik modeller aranmaktadır.

İleriki bölümlerde ele alınan karmaşık ağların aşağıda listelenen rastsal çizge modellerinden hangisi tarafından daha iyi temsil edildiği incelenecektir. *Köşeler*, ağdaki noktalar (node), *kenarlar* da noktalar arasındaki bağlantılardır. Bir çizge $G = (V, E)$ şeklinde V köşe kümesi ve E kenar kümesi ile tanımlanır. Gerektiğinde, $n = |V|$ ve $m = |E|$ parametreleri bu kümelerin eleman sayılarını temsil edecektir.

Öncelikli Bağlanma (Preferential Attachment) Modeli (PA): Bu modelde her adımda bir köşe ve bu köşeden varolan çizgeye x kenar eklenir. Yeni kenarların her biri, önceden varolan köşelerin dereceleriyle doğru orantılı olarak belirlenir. Burada x değeri aşağıdaki gibi seçildiğinde, $|E|$ 'nin beklenti değeri (expectation) orijinal ağın kenar sayısına eşit olur:

$$\frac{2}{n} + 2x = \frac{2|E|}{n}.$$

Erdős-Rényi (Binomial Rastsal Çizge, $G(n, p)$) Modeli (ER): Köşe ikililerinin sayısı $\binom{n}{2}$ olup, her çift belirlenmiş sabit bir p olasılığıyla kenara dönüşür. Orijinal ağ ile üretilen çizgenin kenar sayısının eşit olması için $p = |E|/\binom{n}{2}$ olarak kabul edilir.

Chung-Lu Modeli (CL): Bu model, ER modelini eşit olmayan kenar olasılıklarıyla genelleştirir. Her bir $\{i, j\}$ köşe çifti birbirine $p_{ij} = (1/C)w_iw_j$ olasılığıyla bağlanır. Burada C bir sabit, w_i , her i köşesi için beklenen (expected) derecedir.

Konfigurasyon (Configuration) Modeli (CFG): Bu modelde, orijinal ağın derece dağılımına sahip çizgeler ailesi içinden bir çizge eşit olasılıkla rastsal seçilmektedir.

Bu çalışmada, gerçek hayattan alınan farklı ağların çizgesel özellikleri kullanılarak, yukarıda bahsedilen rastsal çizge modellerine göre sınıflandırılması yapılmıştır. Sınıflandırma işleminde, bu iş için en uygun görülen makine öğrenme

algoritmaları kullanılmış ve performansları karşılaştırılmıştır. Bonner vd. [1] çizge sınıflandırması için, ağın topoloji öz-niteliklerini derin sinir ağlarını (neural network) kullanarak bir öğrenme yöntemi geliştirmiştir. Rastsal çizge modellerinin çeşitliliği açısından kapsamlı olan çalışmalardan birisi Rossi ve Ahmed [2] tarafından yapılmıştır. Bu çalışmada, birçok farklı kompleks ağ için rastsal çizge modellerine göre bir sınıflandırma modeli oluşturulmuş ve ağların bu rastsal modellere göre sınıflandırılması yapılmıştır. Bu sınıflandırmada, birçok farklı çizge özelliği öz-nitelik olarak kullanılmıştır. Benzeri şekilde, Bonato vd. [3], bu tür bir çalışmayı *karakter ağları* olarak adlandırdıkları, kitap, film gibi eserlerin içinde geçen karakterlerin arasındaki ilişkileri ağ olarak temsil ederek, bu ağlar için rastsal çizge modellerine göre sınıflandırma problemini çalışmışlardır. Bizim çalışmamızda, ağırlıklı olarak keminformatik, biyolojik ve sosyal ağlar çalışılmış, elde edilen sınıflandırma algoritması, keminformatik ağları sınıf olarak Öncelikli Bağlanma (Preferential Attachment) Modeli'ne koyup, sosyal ağları Chung-Lu Modeli (CL)'ne yerleştirmiştir.

Buradaki çalışmanın temel katkısı, benzerlerinden farklı olarak ilk defa 5 köşeli motiflerin frekanslarından sınıflandırma ve öğrenme başarısını yükseltenlerini analiz etmesi, hangi motiflerin öz-nitelik olarak öneminin daha belirgin olduğunu sunması ve benzeri çalışmalarla karşılaştırıldığında çok az sayıda öz-nitelik kullanarak aynı seviyede başarı yüzdesi sağlamasıdır. Verilen gerçek bir ağın öz-nitelik vektörü elde edilip, rastsal çizge modelleri yardımıyla üretilen çizgeler, sınıflandırma işleminin eğitim (training) aşamasında kullanılmış, bunun sonucunda hangi modelin orijinal ağa en benzer çizgeleri üretebildiği bulunmuştur.

II. METOT

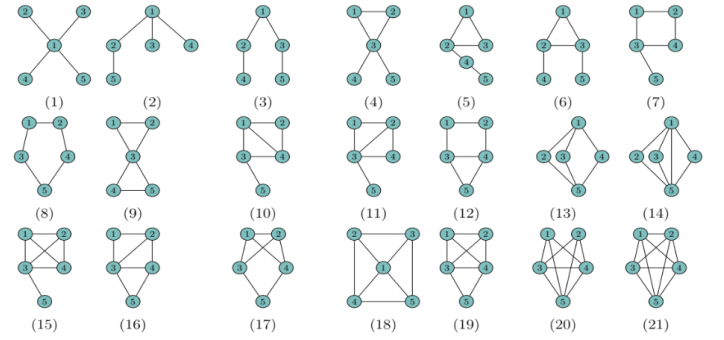
A. Çizge Öz-nitelikleri ve Modelleri

Kullanılan öz-nitelikler, Tablo I'de de görülen yoğunluk, maksimum derece, ortalama derece, maksimum k -çekirdek, ortalama kümelenme katsayısı, ortalama yol uzunluğu, toplam üçgen sayısı, ortalama özvektör merkezliyeti ve bunların yanında 4-profil ve 5-profilin içerdiği bağlantılı (connected) motiflerin frekanslarıdır. Aşağıda, bu çizge öz-niteliklerinin tanımları verilmektedir. Her köşenin bağlantısı olan komşu köşelerin sayısı, o köşenin *derecesi* olup, maksimum, minimum ve ortalama (n üzerinden) değerleri çizgeler üzerine tanımlanan renklendirme, eşleştirme problemleri gibi alanlarda önemli rol oynayan niteliklerdendir. Algoritmik çalışma süresini etkileyen en önemli çizge parametrelerinden biri olan *çizge yoğunluğunun* değeri $m/\binom{n}{2}$ şeklinde hesaplanır. Bu bağlamda, bir çizgenin n^2 seviyesinde kenar içermesi yoğun, aksi halde de seyrek, olduğu anlamına gelir.

Ortalama yol uzunluğu, ağdaki köşelerin bağlılığına (connectivity) dair bilgi veren bir özellik olup, olası tüm köşe çiftleri arasındaki en kısa yolların uzunluklarının ortalamasıdır. Bir çizgenin kümelenme katsayısı (clustering coefficient), çizgedeki köşelerin kümelenme eğilimini ölçer [4]. *Yerel (local) kümelenme katsayısı*, bir i köşesi için $C_i = T_i/W_i$ olarak tanımlanır. Burada, T_i ve W_i , sırasıyla, i köşesini içeren üçgen sayısı ve i 'yi ortada bulunduran "açık üçgen" de denen ikili kenarların (wedge) sayısını belirtir. *Ortalama kümelenme katsayısı*, C_i değerinin tüm köşeler üzerinden ortalamasıdır. *Özvektör merkezliyeti* bir köşenin diğer köşeler üzerindeki

etkisini ölçer. Bir köşenin özvektör merkezliyeti, komşuluk matrisinin en büyük özvektöründeki bu köşeye karşılık gelen hane değeri. Eğer bir köşenin derecesi yüksekse, bu köşenin özvektör merkezliyeti yüksek olacaktır.

Bu çalışmada, *k-profil*, sadece k 'nin değerinin 4 ve 5 olduğu durumlar için kullanılacak olup, k tane köşe üzerinde oluşabilecek tüm altçizgelerin (motif) görülme sıklığı bilgisini içerir. Şekil 1'de, 5-profilde yer alan motifler örnek olarak verilmiştir. Yine motif frekansları gibi, diğer bir önemli ağ parametresi de çizgenin k -çekirdeğidir [5], [6]. Bir çizgenin k -çekirdeği, içerisindeki her köşenin en az k komşusunu da içeren olabilecek en büyük altçizgedir. Motiflerin frekansı



Şekil 1: 5-profil motifleri [7]

çizgelerin yapısı hakkında hala yeterli bilgi vermiyor olsa da, birçok ağın yerel (local) yapısı hakkında oldukça önemli bilgi sağladığı doğrulanmıştır [8]–[10]. Bu sayede, ağ yapısının karakterize edilmesi için yeterli derecede istatistiksel bilgi sağlandığı da görülmüştür [11]. Motif frekanslarının hesaplayan algoritmaların daha hızlı çalışması, bu makalenin konusu olmamakla beraber, oldukça aktif bir araştırma konusudur. Bu çalışmada motif frekanslarının hesaplanması, güncel olarak bilinen en hızlı ve kesin sonuçları üretmiş Pinar, Seshadhri ve Jha'nın [7]'nin algoritmasına dayandırılmaktadır.

B. Öz-nitelik Seçimi

Temel olarak kullanılan makine öğrenimi algoritmaları Destek Vektör Makinesi (SVM: support vector machine), Lojistik Regresyon, Karar Ağacı (decision tree), Rastsal Orman (random forest), Adaboost ve En-Yakın- k -Komşu, k NN, (k Nearest Neighbors) olarak sayılabilir. Tüm öz-nitelikler, normalize edilerek kullanılmıştır. İlk aşamalarda, makine öğrenmesi modelleri, yapay ağ modellerini ayırt edebilmesi için bu yapay ağ modelleri ile üretilmiş ağlar ile eğitilmiştir. Daha sonra, bu makine öğrenmesi modellerinin, tamamen farklı ağlar olan gerçek dünya ağları için tahmin gerçekleştirmesi amaçlanmıştır. Bu tahminlerde herhangi bir doğru veya yanlış olmamakla beraber, değişik makine öğrenmesi modellerinin yaptığı tahminleri birarada görmek, sınıflandırma sonuçlarını güvenilirliğini sağlamak açısından önemlidir.

Sınıflandırma modeli, 4 farklı rastsal çizge modeli kullanılarak üretilen, her biri için 100'er olmak üzere 400 sentetik ağ ile eğitilmiştir. Eğitim sürecinde sınıflandırma modelinin hiper parametreleri 5-katlamalı çapraz doğrulama (5-fold cross validation) ile seçilmiştir. Çapraz doğrulama, ana veri kaynağının bir alt kümesindeki makine öğrenme algoritmalarını

TABLO I: Seçilen ağların öznetelikleri

Ağ	Düğüm Sayısı	Kenar Sayısı	Yoğunluk	Maks. Derece	Ort. Derece	Maks. k-Çekirdek	Ort. Kümelene Katsayısı	Diameter	Ort. Yol Uzunluğu	Üçgen Sayısı	Ort. Özvektör Merkeziyeti
celegans-dir	453	2025	0.0198	237	8.94	10	0.64646	7	2.6638	3284	0.0311
macaque-rhesus_brain_2	91	582	0.1421	87	12.79	11	0.86005	3	1.8681	1902	0.0848
mouse_brain_1	213	16089	0.7126	205	151.07	111	0.75826	2	1.2874	622414	0.0677
ENZYMES_g300	49	93	0.0791	7	3.80	3	0.13780	16	5.8061	17	0.0854
ENZYMES_g540	49	92	0.0782	6	3.76	3	0.27551	17	6.1607	23	0.0880
soc-dolphins	62	159	0.0841	12	5.13	4	0.25896	8	3.3570	95	0.0908

değerlendirmek için yeniden örnekleme prosedürüdür. Veri kümesi 5 eşit parçaya ayrılarak, her bir yinelemede, bir parça test veri kümesi, diğer parça da eğitim veri kümesi olarak tanımlanmıştır.

4 farklı teknik kombine edilmiş ve bu sonuçlarla 3 farklı öznetelik vektörü seçilmiştir. Bu aşamada, Ki-Kare (Chi-Square) testi, Lojistik Regresyon, Özyinelemeli Öznetelik Eleme (RFE: recursive feature elimination) ve Rastsal Orman sınıflayıcıları kullanılmıştır. İki olayın bağımsızlığını hesaplamak için kullanılan istatistiksel bir yöntem olan Ki-kare testi ile özneteliklerin birbirinden ne kadar bağımsız olduğu ölçülmektedir. RFE [12], en önemli öznetelikleri seçmek için özyinelemeli bir öznetelik eleme yöntemidir. Bu algoritma verilerle çalışır ve her adım için en zayıf özelliği düşer ve bir sonraki adımı başlatır. Bu işlem, algoritma verilen belirli sayıda öznetelik kalana kadar devam eder. Algoritma, en zayıf özneteliği katsayılarla tanımlar. En küçük katsayı değerine sahip öznetelik, en zayıf öznetelik olarak tanımlanmaktadır. Lojistik regresyon ve rastsal orman test aşamasında kullanılıp, her bir özneteliğin katsayısına karar vermek için kullanılmıştır.

Yukarıda bahsedilen analiz sonucu seçilen öznetelik vektörü bileşenleriyle beraber sıralanmıştır. Vektör-4 herhangi bir indirgemeye tabi tutulmamış ve bir önceki bölümde bahsedilen tüm öznetelikleri içeren vektördür. Önceki paragrafta bahsedilen indirgemeler sonucu öznetelik açısından Vektör-3 \subseteq Vektör-2 \subseteq Vektör-1 \subseteq Vektör-4 olmak üzere üç farklı vektör ortaya çıkmıştır. İlk indirgeme sonucu elde edilen *Vektör-1*, Vektör-4'teki tüm öznetelikleri içerir, ama 5-profil motiflerinden hepsini değil de 4, 9, 13, 14'ün (Şekil 1) dışındakilerin frekanslarını öznetelik olarak içermektedir. *Vektör-2*, Vektör-1'den yoğunluk, ve 5-profil motiflerinden 5, 8, 10, 15'in çıkarılmasıyla oluşturulmuştur. Son olarak, *Vektör-3*, Vektör-2'den ortalama derece, maksimum k-çekirdek, ortalama yol uzunluğu, toplam üçgen sayısı, ve 5-profil motiflerinden 1, 3, 7, 11, 12, 16, 18, 19'un çıkarılmasıyla oluşturulmuştur. *Vektör-5* ise Vektör-4'ten 5-profile yönelik tüm özneteliklerin çıkarılmasıyla elde edilmiştir. Sonuçlar kısmında, Vektör-3'ün, sadece 8 öznetelik kullanarak en kapsamlı öznetelik kümesini içeren Vektör-4'ün elde ettiği sonuçlara çok yakın bir başarı yüzdesi sağladığı gösterilmektedir. Bu öznetelikler: maksimum derece, ortalama kümelene katsayısı, ortalama özvektör merkeziyeti ve 5-profil motiflerinden 2, 6, 17, 20 ve 21'dir.

III. DENEYSEL SONUÇLAR

Deneyler için sosyal, biyoloji ve kimyabilimi alanlarındaki ağlar üzerinde çalışılmıştır [13]. Tüm deneyler Python 3.7.5 yardımıyla yapılmıştır. Tüm çizgeler yönsüz (undirected), ağırlıksız (unweighted) ve bağlantılıdır (connected). Ağların özellikleri Tablo-I'de gösterilmiştir. Bu tabloda, sınır sistemlerinin topolojisi, beyin ağları (Macaque-rhesus-brain-2 ve mouse-brain-1), kimyasal yapılar arasındaki ilişkiler keminformatik ağlar (ENZYMES-G300 and ENZYMES-G540), biyolojik

sistemler biyolojik ağlar (celegans-dir), canlıların arasındaki iletişim bilgileri sosyal canlı ağları (soc-dolphins) ile örneklenmiştir.

Tablo-II'de, 5 farklı öznetelik vektörünün makine öğrenmesi eğitimi sonucu F1 skorları gösterilmektedir. *F1-skoru*, kesinlik (precision) ve hassasiyet (recall) değerlerinin harmonik ortalaması alınarak hesaplanan istatistiksel bir ölçüttür. Bu sonuçlarda bütün vektörlerin, F1 skorlarının % 95'ten yüksek olduğu, bunun yanında, hassasiyet (recall) ve kesinlik (precision) değerlerinin de sırasıyla % 95 ve %96'nın üzerinde olduğu gözlemlenmiştir. Başka bir önemli sonuç ise, Vektör-2 ve 3 ile diğerlerine göre daha az öznetelik kullanılarak yüksek skorlar elde edilmesidir. Vektör-3'ün, sadece 8 öznetelik kullanarak en kapsamlı öznetelik kümesini içeren Vektör-4'ün elde ettiği sonuçlara çok yakın bir başarı yüzdesi sağladığı gösterilmektedir. Bu öznetelikler: maksimum derece, ortalama kümelene katsayısı, ortalama özvektör merkeziyeti ve 5-profil motiflerinden 2, 6, 17, 20 ve 21'dir. Vektör-5'in ve Vektör-4'ün ortalama F1 skorlarına bakıldığında, 5-profil motiflerinin sınıflandırma için önemli bir öznetelik olduğu görülmektedir. Tablo-II'de gözlemlenen, Vektör-3'ün tatmin edici seviyede başarılı sonuçlar verdiğidir. Bu sebeple, sınıflandırma işleminin test aşamasında Vektör-3, öznetelik vektörü olarak kullanılacaktır.

TABLO II: Öznetelik vektörlerinin eğitim skorları

	F1 Skorları				
	Vektör-1	Vektör-2	Vektör-3	Vektör-4	Vektör-5
CL	0.9686	0.9372	0.9217	0.9930	0.9802
CNFG	0.9849	0.9678	0.9634	0.9833	0.9695
GNP	0.9720	0.9358	0.9239	0.9783	0.9563
PA	0.9986	0.9988	0.9939	0.9976	0.9986
Ortalama	0.9810	0.9599	0.9507	0.9881	0.9762

Bir sonraki aşama, yapay ağlarla eğitilmiş makine öğrenmesi modellerinin gerçek dünya ağları üzerine tahminleridir. Bunun için, her bir makine öğrenmesi modeli için farklı skor sistemleri belirlenmiştir. Örneğin, SVM modelleri için bu skor hiperdüzleme(hyperplane) olan uzaklıktır. Skor sistemlerinin tümü için skor ne kadar büyük olursa, tahmin edilen sınıfa olan güvenin de o kadar yüksek olduğu anlamına gelmektedir. Tablo III'te, Vektör-3 ile eğitilmiş olan makine öğrenmesi modelinin, ENZYMES g300 ve g540 ağları için tahmin sonuçları görülmektedir. Bu ağların ikisi de keminformatik ağlardır. Sonuçlarda görüldüğü gibi, her iki ağ da sınıf olarak Öncelikli Bağlanma (Preferential Attachment) Modeli'ne yerleştirmiştir. Bu beklenen bir sonuçtur, çünkü moleküler ağlarda da genelde PA modelinde olduğu gibi daha fazla komşuya sahip köşeler ileride de komşu sayısının artacağı en kuvvetli adaylardır [14].

Tablo-IV'te "dolphins"(yunuslar) ağı için yapılan tahminlerin sonuçları bulunmaktadır. Bu ağ, bir grup yunusun birbirleri ile olan sosyal etkileşimlerini temsil etmektedir. Bu ağın

TABLO III: Keminformatik ağlar sonuçları (Vektör-3).

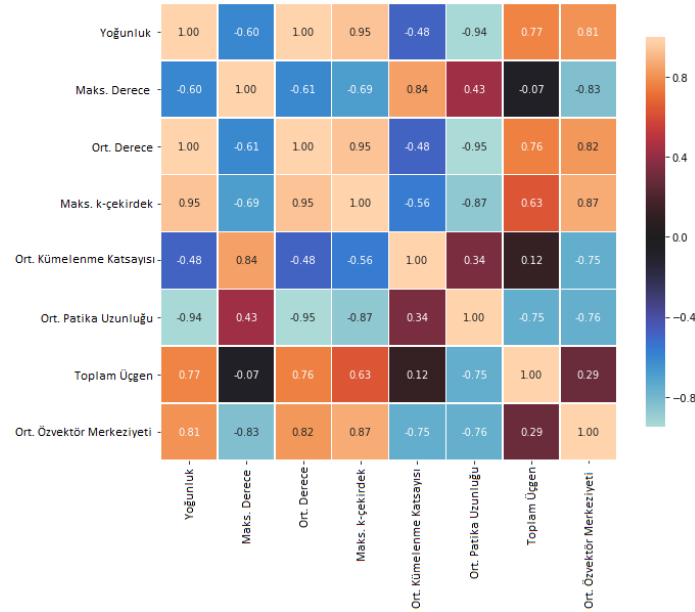
Ağ	Sınıflandırıcı	CL	CNFG	GNP	PA
ENZYMES G-300	SVM L1	0.50	-48.38	-34.35	8.95
	SVM L2	1.98	-32.19	-34.69	13.68
	Lojistik Regresyon	4.10	-21.32	-107.61	24.18
	Adaboost	17.94	-19.46	-21.18	22.71
	Rastal Orman	0.15	0.15	0.18	0.52
	kNN	0.0	0.0	0.0	1.0
ENZYMES G-540	SVM L1	-13.26	-13.97	-33.80	7.16
	SVM L2	-15.73	-17.43	-34.29	17.53
	Lojistik Regresyon	-1.66	-22.58	-107.30	30.52
	Adaboost	3.31	-7.89	-5.37	9.95
	Rastal Orman	0.0	0.0	0.0	1.0
	kNN	0.0	0.0	0.0	1.0

“Chung-Lu (CL) Modeli” olarak sınıflandırılması, bu modelin derece dağılımı gibi yapısal özelliklerinin, pek çok sosyal ağda görülen, kuvvet kuralına (power law) uygun olmasındandır [15]. Ek olarak, Şekil 2’de, başarı performansı en yük-

TABLO IV: Sosyal ağ tahmin sonuçları (Vektör-3)

Ağ	Sınıflandırıcı	CL	CNFG	GNP	PA
soc-dolphins	SVM L1	21.45	-9.97	-7.91	-6.64
	SVM L2	16.75	-5.10	-5.48	-2.10
	Lojistik Regresyon	41.59	-14.27	-16.17	-5.95
	Adaboost	8.50	3.24	-7.35	-4.39
	rastal orman	0.56	0.26	0.04	0.14
	kNN	1.0	0.0	0.0	0.0

sek olarak gözlemlenen öznelitliklerin arasındaki korelasyonun, çizgelerin yapısı açısından da anlamlı olduğu görülmüştür. Örnek olarak, üçgen sayısı ve yoğunluk arasında, ortalama kümelenme katsayısı ile maksimum derece arasında yüksek korelasyon bulunmuştur.



Şekil 2: ENZYME G300 ağı için korelasyon matrisi

IV. SONUÇLAR

Her bir gerçek dünya ağı (real-world network) için rastsal çizge modelleri ile örnek ağlar üretilmiştir. Bu örneklerle 6

farklı makine öğrenmesi modeli eğitilmiş ve bu modellerden çizge modellerini ayırt etmesi beklenilmiştir. Eğitimin sonunda, yüksek bir başarıyla yapay ağ modellerinin ayırt edilebildiği görülmüştür. İstatistiksel yöntemlerle, eğitim aşamasında kullanılan çizge öznelitlikleri üzerinde indirgemeler yapılmış ve başarılı bir sınıflandırma için gereken öznelitlik miktarı azaltılmıştır. Böylece, büyük ağlar için hesaplama zamanı açısından çok daha pahalı olacak bu işlemler daha kısa zamanda yapılabilecektir. Gelecekte, bu yöntemin büyük ölçekli ağları da kapsayan çok daha geniş bir ağ kümesine uygulanması planlanmaktadır.

BİLGİLENDİRME

Bu çalışma, kısmen Bilim Akademisi tarafından sağlanan BAGEP Ödülü ve TÜBİTAK 118E283 no’lu proje kapsamında desteklenmiştir.

KAYNAKLAR

- [1] S. Bonner, J. Brennan, G. Theodoropoulos, I. Kureshi, and A. S. McGough, “Deep topology classification: A new approach for massive graph classification,” in *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016, pp. 3290–3297.
- [2] R. A. Rossi and N. K. Ahmed, “Complex networks are structurally distinguishable by domain,” *Social Network Analysis and Mining*, vol. 9, no. 1, p. 51, 2019.
- [3] A. Bonato, D. R. D’Angelo, E. R. Elenberg, D. F. Gleich, and Y. Hou, “Mining and modeling character networks,” in *International workshop on algorithms and models for the web-graph*. Springer, 2016, pp. 100–114.
- [4] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, p. 440, 1998.
- [5] A. E. Sarıyüce, C. Seshadhri, and A. Pinar, “Parallel local algorithms for core, truss, and nucleus decompositions,” *arXiv.org e-Print archive*, <https://arxiv.org/abs/1704.00386>, 2017.
- [6] P. Liu and A. E. Sarıyüce, “Analysis of core and truss decompositions on real-world networks,” *MLG workshop at Knowledge Discovery and Data Mining Conference, Anchorage, AK, USA*, 2019.
- [7] A. Pinar, C. Seshadhri, and V. Vishal, “Escape: Efficiently counting all 5-vertex subgraphs,” in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 1431–1440.
- [8] P. W. Holland and S. Leinhardt, “Local structure in social networks,” *Sociological methodology*, vol. 7, pp. 1–45, 1976.
- [9] K. Faust, “A puzzle concerning triads in social networks: Graph constraints and the triad census,” *Social Networks*, vol. 32, no. 3, pp. 221–233, 2010.
- [10] O. Frank, “Triad count statistics,” in *Annals of Discrete Mathematics*. Elsevier, 1988, vol. 38, pp. 141–149.
- [11] N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt, “Efficient graphlet kernels for large graph comparison,” in *Artificial Intelligence and Statistics*, 2009, pp. 488–495.
- [12] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.
- [13] R. A. Rossi and N. K. Ahmed, “The network data repository with interactive graph analytics and visualization,” in *AAAI*, 2015. [Online]. Available: <http://networkrepository.com>
- [14] S. Light, P. Kraulis, and A. Elofsson, “Preferential attachment in the evolution of metabolic networks,” *Bmc Genomics*, vol. 6, no. 1, p. 159, 2005.
- [15] F. Chung and L. Lu, “The average distances in random graphs with given expected degrees,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 25, pp. 15 879–15 882, 2002.