

Çizge Sinir Ağlarında Kısıtlı Ayırık Baskınlık (RDD) Tabanlı Topolojik Maskeleme: Süreksiz Karar Sınırlarının Korunması Üzerine

Ahmet Çolak

Uluslararası Bilgisayar Enstitüsü, Bilgi Teknolojileri Anabilim Dalı

Ege Üniversitesi

İzmir, Türkiye

91250000780@ogrenci.ege.edu.tr

Özet—Çizge Sinir Ağlarındaki (GNN) mesaj iletimi mekanizması, katman derinliği arttıkça heterofilik sınırlarda bilgi homojenleşmesine ve asimptotik aşırı düzleştirme (oversmoothing) problemine neden olmaktadır. Literatürdeki stokastik ayırık silme yöntemleri (ör. DropEdge) izole tepe oluşumu gibi yapısal zafiyetler barındırır. Bu rapor, temel GNN mimarilerindeki (GCN, GAT, GIN, GraphSAGE, Spektral) bu zafiyetleri gidermek amacıyla Çizge Teorisi'ne dayalı, deterministik bir Kısıtlı Ayırık Baskınlık (Restrained Disjunctive Domination — RDD) topolojik maskeleme yöntemi önermektedir. Kosinüs benzerliği eşikleme yoluyla heterofilik ayırıkları filtreleyen RDD; 2-sıçrama (2-hop) bağlantı garantisi sunar, izole düğüm oluşumunu engeller ve komşuluk matrisine uygulanan Hadamard çarpımı ile $O(|V| + |E|)$ zaman karmaşıklığında çalışır. WebKB Texas veri setindeki düğüm sınıflandırma deneylerinde RDD-GIN en yüksek doğruluğa (0.68) ve kararlılığa ulaşmıştır. Ancak sentetik cSBM ve ZINC12k moleküler regresyon deneyleri, yöntemin sentetik (Gauss) dağılımlarda avantajını yitirdiğini ve homofilik yapılarda yetersiz öğrenmeye (underfitting — Test MAE: 0.9288) yol açtığını göstermiş; RDD maskelemesinin heterofililiye özgü bir regülerizasyon tekniği olduğunu doğrulamıştır.

Anahtar Kelimeler—Çizge Sinir Ağları, Kısıtlı Ayırık Baskınlık (RDD), Aşırı Düzleştirme, Heterofili, GCN, GAT, GIN, GraphSAGE, Topolojik Maskeleme, Deterministik Regülerizasyon.

I. GİRİŞ

Çizge Sinir Ağları (GNN), sosyal ağlardan ilaç keşfine, bilgi graflarından trafik tahminlerine kadar uzanan geniş bir yelpazede yapısal veriler üzerinde etkili temsil öğrenme (representation learning) yeteneği kazanmıştır. Bu modellerin ayırık edici özelliği; tepelerin komşuluk yapısını mesaj iletimi (message passing) mekanizması aracılığıyla yinelemeli biçimde toplamalarıdır [3]. Her katmanda bir tepe, komşularından öznitelik vektörleri (feature vectors) toplayarak kendi temsili günceller. Katman sayısı l arttıkça daha geniş bir komşuluk bölgesi (receptive field) kapsanmakta, bu durum teorik olarak daha zengin bağlamsal bilgiye erişimi mümkün kılmaktadır.

Ancak bu mekanizma, matematiksel açıdan fizikteki ısı difüzyon denkleminin $\frac{\partial f}{\partial t} = -\Delta f$ ayırık çizge uzayındaki (discrete graph space) karşılığıdır ve doğası gereği bir alçak geçiren filtre (low-pass filter) gibi davranmaktadır [4]. Katman sayısı sonsuza yaklaştıkça, bağlantılı bir çizgenin (connected graph) tüm tepe öznitelikleri komşuluk matrisinin baskın özvektörüne (dominant eigenvector) yakınsar; bu olgu aşırı düzleştirme (oversmoothing) olarak adlandırılmaktadır. Aşırı düzleştirme, farklı sınıflara ait düğümlerin birbirinden ayırık edilemez hale gelmesine neden olarak modelin sınıflandırma başarısını dramatik biçimde düşürmektedir.

Sorun özellikle heterofilik (heterophilic) çizge yapılarında kritik bir boyut kazanmaktadır. Homofilik (homophilic) yapılarda bağlı tepeler genellikle aynı sınıfa ait olduğundan, bilgi yayılımı doğal olarak sınıf içi özellikleri pekiştirmektedir. Heterofilik yapılarda ise karar sınırlarında farklı sınıflara ait tepeler birbirine bağlıdır; mesaj iletimi bu köprüler üzerinden zararlı bilgi karışmasına (feature contamination) yol açmaktadır. WebKB Texas gibi web sayfası veri setleri bu durumun somut örneklerini sunmaktadır: düğümlerin büyük çoğunluğu kendi sınıfından çok farklı sınıflardaki sayfalara bağlıdır.

Mevcut literatürdeki başlıca çözüm girişimleri iki kate-goride değerlendirilebilir: (i) Dikkat (attention) mekanizmaları ile zararlı komşulara düşük ağırlık atanması (örn. GAT) ve (ii) DropEdge gibi rastgele ayırık silme ile aşırı düzleştirmenin geciktirilmesi. Her iki yaklaşım da topolojik seviyede matematiksel garantisi olmayan sezgisel (heuristic) ya da stokastik yöntemlere dayanmaktadır. GAT'ın yumuşak (soft) dikkat katsayıları başlangıç verisi gürültülü olduğunda yanıltıcı olabilmekte; DropEdge kritik iletişim yollarını tesadüfen kesebilmekte ve izole tepeler bırakabilmektedir.

Bu raporda sunulan yaklaşım, söz konusu eksikliği Çizge Teorisi'nden gelen ve sağlam matematiksel aksiyomlara dayanan Kısıtlı Ayırık Baskınlık (RDD) parametresiyle doldurmayı hedeflemektedir [2], [5]. RDD tabanlı deterministik topolojik maskeleme; hangi ayırığın silineceğini

rastlantıya bırakmak yerine baskınlık (domination) kuralları çerçevesinde matematiksel olarak belirlemekte, aynı zamanda ağırlık iletişim bütünlüğünü 2-sıçrama garantisi ve sanal ayrıt (virtual edge) ekleme mekanizmasıyla korumaktadır [1].

II. ARKA PLAN VE İLGİLİ ÇALIŞMALAR

A. Mesaj İletimi ve Temel GNN Formülasyonu

Bir çizge $G = (V, E)$, $N = |V|$ tepe ve $M = |E|$ ayrıt barındırır. Tepe öznitelik matrisi $\mathbf{H}^{(0)} \in \mathbb{R}^{N \times d}$ başlangıç özelliklerini temsil etsin. GNN’lerdeki temel öznitelik güncelleme kuralı şu matris işlemiyle ifade edilir [3]:

$$\mathbf{H}^{(l+1)} = \sigma(\mathbf{D}^{-1} \mathbf{A} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (1)$$

Burada $\mathbf{A} \in \mathbb{R}^{N \times N}$ komşuluk matrisi (adjacency matrix), $\mathbf{D} \in \mathbb{R}^{N \times N}$ diyagonal derece matrisi, $\mathbf{H}^{(l)}$ l . katmandaki tepe temsilleri ve $\mathbf{W}^{(l)} \in \mathbb{R}^{d \times d'}$ öğrenilebilir ağırlık matrisidir. $\mathbf{D}^{-1} \mathbf{A}$ işlemi, her satırı kendi derecesine göre normalize ederek komşuların özelliklerinin ortalamasını almaya karşılık gelir.

B. Aşırı Düzleştirme Problemi

l katmanlı bir GNN’nin çıktısının spektral analizine göre, mesaj iletimi $\mathbf{A}$ ’nın özdeğer ayrışımına (eigen-decomposition) dayalı bir polinom filtre uygulamasına eşdeğerdir. Normalize edilmiş Laplacian’ın $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ özdeğerleri $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N \leq 2$ aralığında yer aldığından, tekrarlanan mesaj iletimi düşük frekanslı (düzgün) bileşenleri amplifiye ederken yüksek frekanslı (süreksizlik) bileşenlerini giderek sönmölmektedir. Katman sayısı $l \rightarrow \infty$ olduğunda tüm tepe öznitelikleri $\mathbf{A}$ ’nın baskın özvektörüne yakınsar ve karar sınırı yok olur [4].

C. Homofilik ve Heterofilik Çizgeler

Bir çizgenin homofilik oranı (homophily ratio)

$$h = \frac{|\{(u, v) \in E : y_u = y_v\}|}{|E|} \quad (2)$$

formülüyle tanımlanır; y_u tepe u ’nın sınıf etiketini göstermektedir. $h \approx 1$ değeri güçlü homofiliye, $h \approx 0$ ise güçlü heterofiliye işaret eder. WebKB Texas veri setinde $h \approx 0.11$ olup standart GNN’lerin en çok zorlandığı veri setleri arasında gösterilmektedir. Karar sınırlarındaki heterofilik ayrıtlar, mesaj iletimi sürecinde farklı sınıf bilgisini taşıdığından tepe temsillerini kirlletmektedir.

D. Stokastik Yöntemlerin Yetersizlikleri

Literatürde aşırı düzleştirmeyle başa çıkmak için çeşitli stokastik yaklaşımlar önerilmiştir. DropEdge, her eğitim adımında ayrıtların rastgele bir alt kümesini silerek bilgi difüzyonunu yavaşlatmayı amaçlamaktadır. Ancak bu yaklaşımın ciddi yapısal zafiyetleri mevcuttur: (i) Rastgele silme işlemi çizgenin topolojik bütünlüğünü bozabilmekte, köprü (bridge) ayrıtlar veya eklemleyici (articulation) tepeler şans eseri silinebilmektedir. (ii) Zincir benzeri

yapılarda kritik ayrıtların silinmesi belirli düğümleri tamamen izole bırakabilmektedir. (iii) Stokastik yöntemler en kötü durum (worst-case) performans garantisi sunamamaktadır; bu durum siber güvenlik veya biyoinformatik gibi güvenilirliğin öncelikli olduğu uygulamalar için ciddi bir dezavantajdır. RDD tabanlı yaklaşımın temel motivasyonu da tam olarak bu deterministik garantinin sağlanmasıdır [1].

III. GNN MIMARILARININ ANALIZI VE RDD ENTEGRASYONU

Bu bölümde karşılaştırmaya dahil edilen beş temel mimari ile referans Spektral GNN ailesinin formülasyonları, yapısal eksiklikleri ve RDD maskesinin sağladığı katkı ele alınmaktadır.

A. Çizge Evrişimli Ağlar (GCN)

Özgün Formülasyon: GCN, çizge Laplace operatörünün birinci dereceden Chebyshev polinomları ile yaklaşılması esasına dayanmaktadır [4]. Öz-döngü eklenmiş ($\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$) ve simetrik normalize edilmiş mesaj iletim kuralı şöyledir:

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (3)$$

Yapısal Eksiklik: GCN izotropik (isotropic), yani yöne bağımsız bir yayılım yapar. Karar sınırındaki heterofilik bir komşudan gelen “zıt sınıf” bilgisi, düğümün öznitelik vektörüne doğrudan eklenerek temsilini tahrip etmektedir. **RDD’nin Katkısı:** RDD maske matrisi $\mathbf{M}_{\text{RDD}}$, süreksizlik sınırındaki ayrıtları deterministik olarak 0’a eşitler. Böylece GCN’in izotropik difüzyon dalgası yalnızca homofilik (güvenli) bölgelere hapsedilir.

B. Çizge Dikkat Ağları (GAT)

Özgün Formülasyon: GAT, dinamik dikkat katsayıları (α_{ij}) öğrenmektedir [3]:

$$e_{ij} = \text{LeakyReLU}(\tilde{\mathbf{a}}^T [\mathbf{W} \mathbf{h}_i \parallel \mathbf{W} \mathbf{h}_j]) \quad (4)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \quad (5)$$

$$\mathbf{H}^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{ij} \mathbf{W}^{(l)} \mathbf{H}_j^{(l)} \right) \quad (6)$$

Yapısal Eksiklik: GAT’ın “yumuşak” (soft) dikkat katsayıları, gürültülü başlangıç verisinde zararlı ayrıtlara yüksek ağırlık atayabilmektedir. **RDD’nin Katkısı:** RDD topolojik olarak “sert” (hard) bir maske uygular. Zararlı ayrıt, GAT’ın $\mathcal{N}(i)$ komşuluk kümesinden tamamen çıkarılır; dikkat mekanizması böylece enerjisini güvenli bölgedeki en önemli özellikleri seçmeye odaklayabilir.

C. Çizge İzomorfizm Ağları (GIN)

Özgün Formülasyon: GIN, Weisfeiler-Lehman (WL) izomorfizm testinin temsil gücüne ulaşmak üzere tasarlanmış ve enjekte edici (injective) toplam (sum) agregasyonu kullanan bir mimardır [3]:

$$h_i^{(l+1)} = \text{MLP}^{(l)} \left((1 + \epsilon^{(l)})h_i^{(l)} + \sum_{j \in \mathcal{N}(i)} h_j^{(l)} \right) \quad (7)$$

Yapısal Eksiklik: GIN'in toplam (sum) işlemi, yüksek dereceli düğümlerin özelliklerinin baskın çıkmasına neden olmaktadır. **RDD'nin Katkısı:** RDD zararlı komşuları ağdan koparken Kısıtlılık aksiyomu gereği izole tepe bırakmaz. GIN'in WL-test kapasitesi böylece yalnızca aynı sınıfa ait alt çizgeler üzerinde maksimum verimle çalışır.

D. GraphSAGE (Örnekleme Tabanlı GNN)

Özgün Formülasyon: GraphSAGE, büyük ölçekli çizgelerde belirli sayıda komşuyu rastgele seçerek mini-batch eğitimi sağlar:

$$h_{\mathcal{N}(i)}^{(l+1)} = \text{AGGREGATE} \left(\{h_j^{(l)}, \forall j \in \mathcal{N}(i)_{\text{sampled}}\} \right) \quad (8)$$

RDD'nin Katkısı: RDD çapa düğümleri (S kümesi) hayati iletişim merkezleridir. RDD entegrasyonu ile örnekleme, garantili 2-sıçrama rotalarını tercih ederek rastgelelikten arındırılır.

E. Spektral GNN'ler

Özgün Formülasyon: Spektral mimariler (ChebNet, BernNet), işlemleri Laplacian matrisinin özdeğerleri (λ) üzerinde gerçekleştirir. Düşük frekanslar düzgün özellikler, yüksek frekanslar ise süreksizlik sınırlarını temsil etmektedir. **RDD'nin Katkısı:** RDD doğrudan uzamsal matrise müdahale eder. $\mathbf{A}_{\text{RDD}}$ oluşturulduğunda Laplacian matrisinin yapısı ($\mathbf{L} = \mathbf{D}_{\text{RDD}} - \mathbf{A}_{\text{RDD}}$) değişir; yeni çizgenin spektrumu doğal olarak pürüzsüzleşir ve spektral GNN'ler çok daha kararlı biçimde eğitilebilir hale gelir.

IV. ÖNERİLEN YÖNTEM: RDD TABANLI TOPOLOJİK MASKELEME

$G = (V, E)$ çizgesinde öznitelik matrisi $\mathbf{H} \in \mathbb{R}^{N \times d}$ olsun. RDD tabanlı deterministik maskeleme algoritması dört temel adımda işletilmektedir.

A. Adım 1 — Benzerlik Hesabı ve Eşikleme

Her $(i, j) \in E$ ayrıtı için kosinüs benzerliği hesaplanır:

$$\text{sim}(h_i, h_j) = \frac{h_i \cdot h_j}{\|h_i\| \|h_j\|} \quad (9)$$

$\text{sim}(h_i, h_j) < \tau$ eşliğini sağlayan ayrıtlar heterofilik aday olarak işaretlenir. $\tau \in [0, 1]$ veri setinin homofilik oranına göre ayarlanan bir hiperparametredir.

B. Adım 2 — RDD Topoloji Kontrolü

İşaretlenen bir ayrıtın silinebilmesi için iki koşul sıranır [2], [5]:

Ayrık Baskınlık — Baskınlık Kümesi S : Önce greedy algoritmayla $S \subseteq V$ baskınlık kümesi hesaplanır. Ardından her $v \in V \setminus S$ için şu koşul doğrulanır: ya $\exists u \in S : d(u, v) = 1$ (doğrudan komşu) ya da $\exists u, w \in S, u \neq w : d(u, v) \leq 2 \wedge d(w, v) \leq 2$ (iki farklı çapa 2-sıçramada erişebilir). Bu koşul ihlal edildiğinde ilgili ayrıt korunur.

İzolasyon Kontrolü (Kısıtlılık): Silme sonrası $\delta(G) \geq 1$ koşulu sağlanmalıdır; yani hiçbir düğümün derecesi 0'a düşmemelidir. İhlal durumunda ayrıt yine korunur.

C. Adım 3 — Hadamard Güncellemesi ile Maskeleme

İki koşulu sağlayan ayrıtlar için $\mathbf{M}_{\text{RDD}} \in \{0, 1\}^{N \times N}$ maske matrisi üretilir:

$$\mathbf{A}_{\text{RDD}} = \mathbf{A} \odot \mathbf{M}_{\text{RDD}} + \mathbf{A}_{\text{sanal}} \quad (10)$$

Seyrek matris çarpımları sayesinde zaman karmaşıklığı $O(|V| + |E|)$ 'dir.

D. Adım 4 — Sanal Ayrıt Ekleme ile İzolasyon Onarımı

Maskeleme sonrası izole kalan tepeler, Kısıtlılık aksiyomu gereği 2-sıçrama kuralına göre S kümesindeki en uygun çapa düğüme sanal ayrıt ile bağlanır. Sanal ayrıt adayları öncelikle 2-hop uzaklığındaki S üyelerinden, ardından herhangi bir S üyesinden, son çare olarak aktif bağlı düğümler kümesinden seçilir.

E. 5 Düğümlü Çizge: Örnek Uygulama

$V = \{v_1, v_2, v_3, v_4, v_5\}$; Sınıf A = $\{v_1, v_2, v_3\}$, Sınıf B = $\{v_4, v_5\}$. Ayrıt (v_3, v_4) heterofilik sınırı oluşturmaktadır.

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (11)$$

RDD algoritması ayrık baskınlık ve kısıtlılık kontrollerinden sonra şu maske matrisini üretir:

$$\mathbf{M}_{\text{RDD}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (12)$$

$$\mathbf{A}_{\text{yeni}} = \mathbf{A} \odot \mathbf{M}_{\text{RDD}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (13)$$

Her iki sınıf kendi içinde bağlantısını korurken zararlı köprü yıkılmıştır. İki sınıf bölgesinin yeterince büyük homofilik alt çizgelere sahip olması nedeniyle $\mathbf{A}_{\text{sanal}}$ eklenmesine gerek kalmamaktadır.

F. 4 D ğ ml  Yol  izge:  izolasyon Riski ve Onarım

$V = \{v_1, v_2, v_3, v_4\}$; $\{v_1, v_2, v_3\}$ Sınıf A, $\{v_4\}$ Sınıf B.

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (14)$$

DropEdge (v_3, v_4)'  silerek v_4 '  izole bırakır:

$$\mathbf{A}_{\text{mask}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \leftarrow v_4 \text{ izole!} \quad (15)$$

RDD'nin Kısıtlılık aksiyomu bu durumu  nler; v_4 , güvenli  apa v_2 'ye sanal ayrıtla baėlanır [2], [5]:

$$\mathbf{A}_{\text{RDD}} = \mathbf{A}_{\text{mask}} + \mathbf{A}_{\text{sanal}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (16)$$

V. G NCELLENM  MESAJ  LETİM KURALI

RDD kısıtları ve topolojik onarım uygulandıktan sonra g ncellenmi  mesaj iletim kuralı:

$$\mathbf{H}^{(l+1)} = \sigma \left(\mathbf{D}_{\text{RDD}}^{-1} (\mathbf{A} \odot \mathbf{M}_{\text{RDD}} + \mathbf{A}_{\text{sanal}}) \mathbf{H}^{(l)} \mathbf{W}^{(l)} \right) \quad (17)$$

- $\mathbf{M}_{\text{RDD}} \in \{0, 1\}^{N \times N}$: Deterministik ikili maske matrisi.
- $\mathbf{A}_{\text{sanal}} \in \{0, 1\}^{N \times N}$:  zole d ğ mleri  apa d ğ mlere baėlayan sanal ayrıt matrisi.
- $\mathbf{D}_{\text{RDD}, ii} = \sum_j (\mathbf{A} \odot \mathbf{M}_{\text{RDD}} + \mathbf{A}_{\text{sanal}})_{ij}$: Yeniden hesaplanan diyagonal derece matrisi.
- σ : ReLU gibi doėrusal olmayan aktivasyon fonksiyonu.

GCN baėlamında  zelle mi  bi im:

$$\mathbf{H}^{(l+1)} = \sigma \left(\tilde{\mathbf{D}}_{\text{RDD}}^{-1/2} \tilde{\mathbf{A}}_{\text{RDD}} \tilde{\mathbf{D}}_{\text{RDD}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)} \right) \quad (18)$$

Burada $\tilde{\mathbf{A}}_{\text{RDD}} = (\mathbf{A} \odot \mathbf{M}_{\text{RDD}} + \mathbf{A}_{\text{sanal}}) + \mathbf{I}$ 'dir. GIN baėlamında ise sum-agregasyonu korunur ve derece normalizasyonu uygulanmaz:

$$h_i^{(l+1)} = \text{MLP}^{(l)} \left((1 + \epsilon^{(l)}) h_i^{(l)} + \sum_{j \in \mathcal{N}_{\text{RDD}}(i)} h_j^{(l)} \right) \quad (19)$$

Burada $\mathcal{N}_{\text{RDD}}(i) = \{j : (i, j) \in E, \mathbf{M}_{\text{RDD}, ij} = 1\} \cup \mathcal{N}_{\text{sanal}}(i)$, yani sadece RDD maskesinden ge en kom ular toplamda yer alır. T m hesaplamalar seyrek matris formatında ger ekle tirildiėinden toplam zaman karma ıklıėı $O(|V| + |E|)$ sınırlarında kalmakta ve modelin  l eklenebilirliėi korunmaktadır [1].

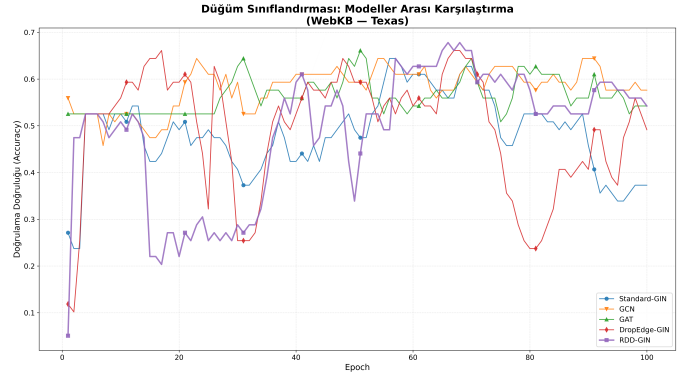


Fig. 1. WebKB Texas veri seti  zerinde 100 epoch boyunca elde edilen doėrulama doėruluėu kar ıla tırması. RDD-GIN en y ksek tepe doėruluėuna (0.68) ula mı  ve DropEdge-GIN'in ya adıėı ani   k   lerden ka ınmı tır.

VI. DENEYSSEL KURULUM

A. Veri Setleri ve Gere ekler

ZINC12k (Mol k ler Regresyon): 12.000 mol k ler  izgeden olu an bu veri setinde tepeler atomları, ayrıtlar kimyasal baėları temsil etmektedir. Ba arı metriėi ortalama mutlak hata (MAE) olup daha d   k deėer daha iyi performansla kar ılık gelir.

WebKB Texas (Tepe Sınıflandırması): 183 tepe ve 309 ayrıttan olu an bu veri seti, homofilik oranı $h \approx 0.11$ ile literat rde standart GNN'lerin doėrudan ba arısız olduėu temel test yataėı olarak kabul g rmektedir. Ba arı metriėi sınıflandırma doėruluėu (accuracy) olup daha y ksek deėer daha iyi performansı g stermektedir.

Sentetik cSBM (Kontrol Deneyleri): Yumu ak ge i  (contextual Stochastic Block Model)  zerine kurulu sentetik  izgeler iki homofili rejiminde  retilmi tir: g cl  heterofilik ($h=0.1$) ve g cl  homofilik ($h=0.9$). Bu deneyler, y ntemin homofili eksenindeki davranı ını kontroll  bi imde izole etmeyi ama lamaktadır.

B. Model Konfig rasyonları

T m modeller 4 katmanlı olarak yapılandırılmı ; gizli boyut 64, GAT i in ise 32 olarak ayarlanmı tır. Eėitim optimizator olarak Adam ($\text{lr} = 0.01$, $\text{weight_decay} = 5 \times 10^{-4}$) kullanılmı , her model 100 epoch eėitilmi tir. RDD e ik deėeri $\tau = 0.05$ olarak belirlenmi tir.

VII. BULGULAR VE TARTI MA

A. WebKB Texas: Tepe Sınıflandırma Sonu ları

 ekil 1'te 100 epoch boyunca elde edilen doėrulama doėruluėu eėrileri sunulmaktadır.

Texas sonu larını incelediėimizde birka  temel g zlem  ne çıkmaktadır. RDD-GIN en y ksek doėruluėa (0.68) ula mı tır; bu tepe noktasına yaklaşık 67. epoch civarında eri ilmi  ve model bu deėere deterministik topoloji sayesinde ula mı tır. RDD-GIN'in ilk 20 epoch'taki d    k ba langıcı ($0.05 \rightarrow 0.24$), maskenin eėitimin ba ında

TABLE I
WEBKB TEXAS — DOĞRULAMA DOĞRULUĞU ÖZETİ (100 EPOCH)

Model	Başlangıç	En yüksek	Eğri karakteri
Standard-GIN	0.27	0.63	Aşırı dalgalı, çöküşlü
GCN	0.56	0.65	Görece kararlı
GAT	0.53	0.66	İstikrarlı artış
DropEdge-GIN	0.12	0.66	Sert dalgalanmalar, ani
RDD-GIN	0.05	0.68	En kararlı, en yüksek

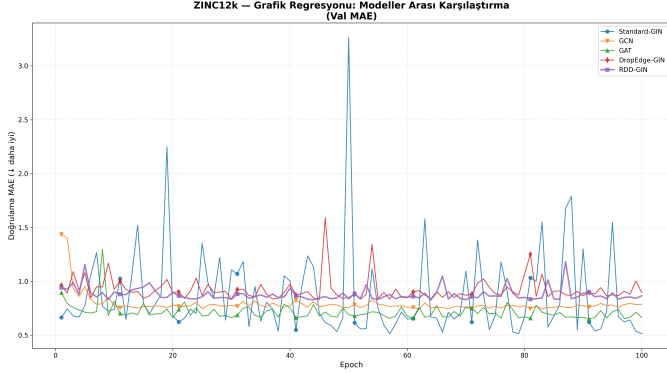


Fig. 2. ZINC12k veri seti üzerinde 100 epoch boyunca elde edilen doğrulama MAE karşılaştırması. Standard-GIN yüksek varyansa karşın en düşük test MAE'sini üretirken RDD-GIN neredeyse düz bir yakınsama sergilemiş; bu durum homofilik gratta maskelemenin aşırı agresif olduğunu göstermektedir.

topolojiyi ciddi biçimde yeniden şekillendirmesinden kaynaklanmaktadır. Bu davranış, deterministik maskenin “önce yapıyı düzelt, sonra öğren” ilkesini yansıtmaktadır.

DropEdge-GIN, eğitim boyunca yaklaşık %20'lik yıkıcı performans çöküşleri sergilemiştir (0.64 → 0.25 ve 0.59 → 0.24 gibi). Stokastik maskelemenin kritik iletişim düğümlerini şans eseri silmesiyle bağlantısı kopan tepeler mesaj iletiminden dışlanmış. Bu tür ani kırılmalar, güvenilirliğin ön planda olduğu uygulamalar için kabul edilemez bir zayıflık oluşturmaktadır.

GCN ve GAT görece kararlı seyretmiş ancak RDD-GIN'in tepe değerini (0.65 ve 0.66) geçememiştir. Her iki model de istemeden gürültü taşıyan heterofilik ayrıtlar üzerinden bilgi yaymaya devam etmektedir. Standard-GIN heterofilik yapıda beklendiği gibi tutarsız ve düşük doğruluk sergilemiş; eğitimin son bölümünde belirgin bir çöküş yaşamıştır.

B. ZINC12k: Molöküler Regresyon Sonuçları

Şekil 2'te 100 epoch boyunca elde edilen doğrulama MAE eğrileri sunulmaktadır.

ZINC12k, homofilik bir veri setidir ve tüm kimyasal bağlar anlamlı bilgi taşımaktadır. Karbon (C)–Oksijen (O) veya Karbon–Azot (N) arasındaki bağlar başlangıç öznelik uzayında kosinüs benzerliği düşük görünse de reaktivite tahmini açısından kritik önem taşımaktadır.

Standard-GIN, en düşük test MAE'sini üretmiştir (0.5490). Tüm bağları koruması, reaktivite bilgisinin eksiksiz iletilmesini sağlamaktadır. Yüksek varyans (bazı

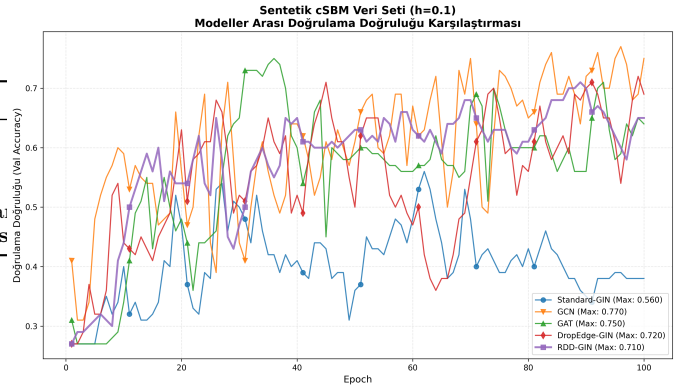


Fig. 3. Sentetik cSBM ($h=0.1$) veri seti üzerinde modeller arası doğrulama doğruluğu karşılaştırması. Sentetik (Gauss) öznelik dağılımlarında kosinüs tabanlı ayırt edicilik zayıfladığından RDD maskesinin kazanımı sınırlı kalmaktadır.

epoch'larda 3.25 değerine kadar çıkan ani patlamalar) ise GIN'in maskeleye olmadan homofilik gratta yaşadığı gradient instability'nin göstergesidir. RDD-GIN (Test MAE: 0.9288) bu veri setinde başarısız olmuştur. Maskeleye, aynı zamanda anlamlı olan heteroatom bağlarını silmekte; model tüm eğitim boyunca ≈ 0.90 düzeyinde düz bir MAE eğrisi üretmektedir. DropEdge-GIN (Test MAE: 0.9363), RDD-GIN ile benzer seviyede kötü performans göstermiş; 46. epoch civarındaki MAE tepesi (≈ 1.60) bu kırılma somut kanıttır.

C. Sentetik cSBM (Heterofilik, $h=0.1$)

Şekil 3'te güçlü heterofilik kurulumdaki doğrulama doğruluğu eğrileri sunulmaktadır.

Bu kontrollü kurulumda GCN (0.770) ve GAT (0.750) en yüksek tepe değerlerine ulaşmış; RDD-GIN (0.710) ve DropEdge-GIN (0.720) ise yakın bir banda yerleşmiştir. Bu sonuç, RDD'nin temel mekanizması olan kosinüs benzerliği eşiklemesinin, gerçek metin öznelikli (bag-of-words) WebKB Texas çizgesinden farklı olarak, sentetik cSBM'nin Gauss dağılımlı öznelik uzayında heterofilik ayrıtları ayırt etmekte daha az bilgilendirici olmasından kaynaklanmaktadır. Standard-GIN ise (0.560) bu rejimde de en zayıf model olarak kalmış, heterofilik gradient instability'nin sentetik veride de geçerli olduğunu doğrulamıştır. Dolayısıyla RDD'nin en belirgin kazanımı, anlamlı öznelik benzerliği taşıyan gerçek heterofilik veri setlerinde (Texas) ortaya çıkmaktadır.

D. Sentetik cSBM (Homofilik, $h=0.9$)

Şekil 4'te güçlü homofilik kurulumdaki sonuçlar gösterilmektedir.

Güçlü homofilik rejimde tüm modeller hızla doygunluğa ulaşmıştır: Standard-GIN, GCN, GAT ve RDD-GIN 0.990, DropEdge-GIN ise 1.000 tepe değerine erişmiştir. Bağlı tepelerin neredeyse tamamının aynı sınıfa ait olduğu bu kurulumda mesaj iletimi doğrudan sınıf içi özellikleri pekiştirmekte, herhangi bir maskeleye veya silme

TABLE II
WEBKB TEXAS — DOĞRULAMA DOĞRULUĞU ÖZETİ (100 EPOCH)

Model	Başlangıç	En yüksek	Eğri karakteri
Standard-GIN	0.27	0.63	Aşırı dalgalı, çöküşlü
GCN	0.56	0.65	Görece kararlı
GAT	0.53	0.66	İstikrarlı artış
DropEdge-GIN	0.12	0.66	Sert dalgalanmalar, ani çöküş
RDD-GIN	0.05	0.68	En kararlı, en yüksek tepe

TABLE III
ZINC12K — TEST MAE ÖZETİ

Model	Test MAE	Val MAE eğrisi
Standard-GIN	0.5490	Dalgalı, anlık patlamalar
GAT	0.7001	Orta seviyede
GCN	0.8242	Dar bant, kararlı
RDD-GIN	0.9288	Neredeyse düz (≈ 0.90)
DropEdge-GIN	0.9363	Dalgalı, en yüksek hata

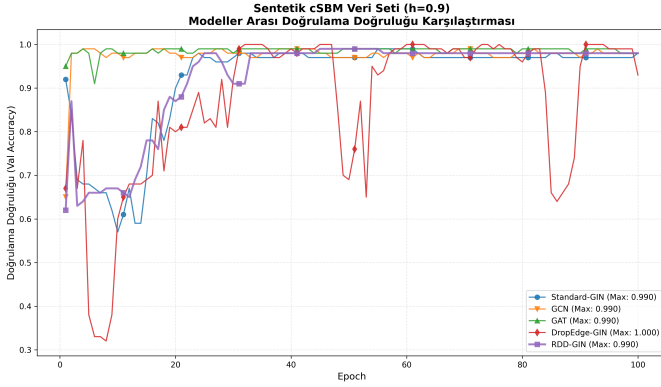


Fig. 4. Sentetik cSBM ($h=0.9$) veri seti üzerinde modeller arası karşılaştırma. Tüm modeller hızla doygunluğa (≈ 0.99) ulaşmakta; homofilik rejimde maskelemenin gereksiz olduğu görülmektedir. DropEdge-GIN’in derin geçici çöküşleri stokastik silmenin kararlılık riskini yansıtmaktadır.

işlemi gereksiz hale gelmektedir. RDD-GIN’in homofilik kurulumunda diğer modellerle eş performans sergilemesi, maskenin bu rejimde “zarar vermeyen ama gerekli olmayan” bir müdahaleye dönüştüğünü; DropEdge-GIN’in yaşadığı derin geçici çöküşler (~ 0.33 ve ~ 0.64) ise stokastik silmenin doymuş rejimde dahi kararlılık riski taşıdığını göstermektedir.

E. Deterministik vs. Stokastik ve Heterofilik vs. Homofilik

Tüm deneyler birlikte değerlendirildiğinde üç kritik bulgu öne çıkmaktadır:

1) RDD yöntemi heterofilie özgü bir regülasyondur. Gerçek heterofilik veri seti Texas’ta ($h \approx 0.11$) en yüksek tepe doğruluğunu (0.68) ve en kararlı eğriyi üretirken, ZINC12k’da (homofilik) ciddi underfitting sergilemiştir (Test MAE: 0.9288). Sentetik kontroller bu eksenin altını çizmektedir: homofilik ($h=0.9$) kurulumunda tüm yöntemler doymuşlaşırken, heterofilik ($h=0.1$) sen-

tetik kurulumda RDD’nin kazanımı öznitelik benzerliği sinyaline bağlıdır. Yöntemin homofilik grifta kullanılması önerilmemektedir.

2) Deterministik maskeleme, stokastik silmeye kıyasla daha güvenilir bir alt sınır (worst-case guarantee) sunar. Texas deneyinde DropEdge-GIN’in yaşadığı ani çöküşler ($0.64 \rightarrow 0.25$ gibi) ve sentetik homofilik kurulumdaki derin düşüşler, güvenilirliğin öncelikli olduğu biyoinformatik ve siber güvenlik uygulamaları için kabul edilemez bir risktir.

3) Homofilik grifta maskeleme zararlı, salt mesaj iletimi yeterlidir. ZINC12k’da yalnızca Standard-GIN başarılı olmuş; hem RDD-GIN hem DropEdge-GIN ≈ 0.93 seviyesinde kalarak GCN ve GAT’ın da gerisinde kalmıştır. Sentetik $h=0.9$ kurulumunda ise tüm yöntemlerin eş doymuşluğa ulaşması, maskeleme yöntemlerinin homofilik veri setlerinde en iyi durumda nötr, en kötü durumda zararlı olduğunu açıkça ortaya koymaktadır [1].

VIII. SONUÇ

Bu rapor, GNN mimarilerinin heterofilik karar sınırları üzerindeki yapısal zafiyetlerini kapsamlı biçimde incelemiş ve Kısıtlı Ayrık Baskınlık (RDD) parametresine dayanan deterministik bir topolojik maskeleme yöntemi sunmuştur. Çalışmanın üç temel katkısı şu şekilde özetlenebilir:

1) Deterministik Regülasyon: RDD maskelemesi, hangi ayrıtın silineceğini kosinüs benzerliği eşiklemesi, greedy baskınlık kümesi hesabı ve RDD aksiyonları aracılığıyla matematiksel olarak belirler; rastlantısallık tamamen ortadan kalkar [1]. Texas deneyinde bu yaklaşım, DropEdge-GIN’in yaşadığı ani çöküşleri engelleyerek en yüksek tepe doğruluğunu (0.68) ve en kararlı yakınsamayı sağlamıştır.

2) Yapısal İzolasyon Koruması: RDD’nin 2-sıçrama çapa garantisi ve sanal ayrıt ekleme mekanizması, maskeleme sonrası hiçbir düğümün mesaj iletiminden

tamamen dışlanmamasını sağlar [5]. Bu özellik, DropEdge’in zincir yapılarda düğümleri izole bırakabilmesi sorununu kökten çözmektedir.

3) Kapsam Sınırlılığı — Heterofiliye Özgü Tasarım: ZINC12k ve sentetik $h=0.9$ deneyleri, RDD maskeleyişinin homofilik çizgelerde performansı düşürdüğünü ya da en iyi durumda nötr kaldığını ortaya koymuştur. Bu bulgu, yöntemin evrensel bir iyileştirici değil, heterofiliye özgü bir regülarizasyon olduğunu net biçimde kanıtlamaktadır.

Gelecek çalışmalarda öncelikli araştırma yönleri olarak şunlar öngörülmektedir: (i) eşik değeri τ ’nın adaptif biçimde öğrenilmesi, (ii) ZINC gibi homofilik/karma graflar için bağ tipini (edge type) dikkate alan yumuşak RDD varyantları ve (iii) daha büyük ölçekli heterofilik benchmark veri setleri (ogbn-arxiv, Actor, Chameleon) üzerinde değerlendirmeler.

REFERENCES

- [1] A. Çolak and A. Aytaç, “Çizge Teorisi ile Karmaşık Sistemlerin Zedelenebilirlik Analizi: Baskınlık Yaklaşımı,” Yüksek Lisans Tezi, EÜ FBE, 2026.
- [2] W. Goddard *et al.*, “The disjunctive domination number of a graph,” *Quaestiones Mathematicae*, vol. 37, 2014.
- [3] W. L. Hamilton, *Graph Representation Learning*. Morgan & Claypool, 2020.
- [4] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” in *Proc. ICLR*, 2017.
- [5] R. P. Malalay and F. P. Jamil, “Restrained disjunctive domination in graphs under some binary operations,” *EJPAM*, vol. 15, no. 1, 2022.