Expressive 1-Lipschitz Neural Networks for Robust Multiple Graph Learning against Adversarial Attacks

Xin Zhao [1] **Zeru Zhang** [1] **Zijie Zhang** [1] **Lingfei Wu** [2] **Jiayin Jin** [1] **Yang Zhou** [1] **Ruoming Jin** [3] **Dejing Dou** [4 5] **Da Yan** [6]

Abstract

Recent findings have shown multiple graph learning models, such as graph classification and graph matching, are highly vulnerable to adversarial attacks, i.e. small input perturbations in graph structures and node attributes can cause the model failures. Existing defense techniques often defend specific attacks on particular multiple graph learning tasks. This paper proposes an attack-agnostic graph-adaptive 1-Lipschitz neural network, ERNN, for improving the robustness of deep multiple graph learning while achieving remarkable expressive power. A K_l-Lipschitz Weibull activation function $\bar{f}$ is designed to enforce the gradient norm $\|\nabla \bar{f}(\mathbf{x})\|$ as K_l at layer l. The nearest matrix orthogonalization and polar decomposition techniques are utilized to constraint the weight norm $\|\bar{\mathbf{W}}_l\|$ as $1/K_l$ and make $\bar{\mathbf{W}}_l$ close to the original weight $\mathbf{W}_l$. The theoretical analysis is conducted to derive lower and upper bounds of feasible K_l under the 1-Lipschitz constraint. The combination of norm-constrained $\bar{f}$ and $\bar{\mathbf{W}}_l$ leads to the 1-Lipschitz neural network for expressive and robust multiple graph learning.

1. Introduction

Multiple graph learning aims to automatically extract, manage, infer, and transfer knowledge in multiple graph data. Popular multiple graph learning tasks include graph classification (Rieck et al., 2019; Wu et al., 2019b; Zhao & Wang, 2019; Magelinski et al., 2020; Peng et al., 2020; Ma et al., 2020b; 2021), graph matching (i.e., network alignment) (Zhang & Tong, 2016; Heimann et al., 2018; Li et al., 2019a; Chu et al., 2019; Zhang et al., 2019; Xu et al., 2019a; Du et al., 2019; Huynh et al., 2020; Fey et al., 2020; Yu et al., 2020a;b; Ren et al., 2020; Yan et al., 2020), multi-graph clustering (Ma et al., 2017; Wang et al., 2020b; Fan et al., 2020; Luo et al., 2020; Wang et al., 2020a), multi-view network embedding (Ma et al., 2017; Qu et al., 2017; Liu et al., 2018; Fu et al., 2019; Sun et al., 2019; Fu et al., 2020), and graph kernel (Yanardag & Vishwanathan, 2015; Al-Rfou et al., 2019; Kriege et al., 2019; Togninalli et al., 2019; Oettershagen et al., 2020).

We have witnessed various adversarial defense techniques to improve the robustness of single graph learning tasks against adversarial attacks, such as node classification (Zhu et al., 2019; Miller et al., 2019; Xu et al., 2019b; Tang et al., 2020b; Entezari et al., 2020; Zheng et al., 2020; Zhou & Vorobeychik, 2020; Jin et al., 2020b; Feng et al., 2020; Elinas et al., 2020; Zhang & Zitnik, 2020), network embedding (Dai et al., 2019), graph clustering (Jia et al., 2020), link prediction (Zhou et al., 2019a), and influence maximization (Logins et al., 2020). However, there is still a paucity of robust multiple graph learning methods under adversarial attacks, which is much more difficult to study, since the multiple graph learning tasks need to analyze both intra-graph and inter-graph links of multiple graphs. In addition, the defense strategies for single graph learning models may not work well for multiple graphs with unique characteristics, such as size, density, and degree distribution. Only recently, researchers have started to study how to improve the robustness of deep multiple graph learning methods, including graph classification (Zhang & Lu, 2020; You et al., 2020; Jin et al., 2020a; Gao et al., 2020), graph matching (Yu et al., 2021), and multiple network embedding (Zhou et al., 2020b). However, the above techniques often defend specific attacks on particular learning tasks (e.g., only graph classification or graph matching). Can we design an attack-agnostic graph-adaptive neural architecture for protecting deep multiple graph learning models from adversarial attacks?

Recently, Lipschitz-constrained neural networks are proposed to offer attack-agnostic defense solutions by imposing

[1]Auburn University, USA [2]JD.COM Silicon Valley Research Center, USA [3]Kent State University, USA [4]University of Oregon, USA [5]Baidu Research, China [6]University of Alabama at Birmingham, USA. Correspondence to: Cieua Vvvvv <c.vvvvv@googol.com>, Eee Pppp <ep@eden.co.uk>, Yang Zhou <yangzhou@auburn.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

a Lipschitz constraint on each layer to restrict the diffusion of input perturbations on the neural networks (Cissé et al., 2017; Tsuzuku et al., 2018; Fazlyab et al., 2019). The Lipschitz bound for the entire neural network is the product of the bound on each layer. This allows to constraint the change of its output in proportion to the change in its input. Lipschitz-constrained neural networks are very useful for defending multiple graph learning models since small input perturbations can be propagated within and across graphs, which dramatically amplifies the perturbations in the output space. However, bounding the Lipschitz constant and maintaining the expressive power are often regarded as orthogonal techniques with different optimization goals. Three recent studies of GroupSort (Anil et al., 2019; Cohen et al., 2019) and BCOP (Li et al., 2019b) improve the expressive power of 1-Lipschitz neural networks while enhancing the robustness by enforcing both weight norm and gradient norm as 1. We argue that simply limiting the above two norms to 1 still sacrifices the expressive power, compared with regular neural networks that do not hold the constraints on the weight and gradient. In addition, a 1-Lipschitz neural network with fixed weight and gradient norms may lead to sub-optimal defense when tackling multiple graphs with individual characteristics.

To our best knowledge, this work is the first attack-agnostic graph-adaptive 1-Lipschitz neural network for improving the robustness of deep multiple graph learning while achieving remarkable expressive power, by making the weight and gradient norms adaptive to multiple input graphs and restricting the diffusion of any input perturbations.

Popular 1-Lipschitz activation functions, e.g. ReLU, Sigmoid, and tanh, must trade nonlinear processing for gradient norm preservation, leading to less expressive networks. In statistics, the Weibull distribution can model hazard functions that are monotonically decreasing, increasing, or constant of the proportion of adopters over time, allowing it to describe any phase of an item's lifetime (Weibull, 1951). The major advantage of Weibull analysis is that it is suitable to reliability and failure analysis. In the context of robust deep multiple graph learning, the perturbation diffusion over the layers is similar to a monotonically decreasing hazard function, i.e, the attack failure possibility decreases with the perturbation diffusion. Motivated by this, a K_l-Lipschitz Weibull activation function $\bar{f}$, i.e., $\|\nabla \bar{f}(\mathbf{x})\| = K_l$, is designed to restrict the gradient norm as K_l at layer l of the neural network. In addition, we utilize nearest matrix orthogonalization and polar decomposition techniques (Bjorck & Bowie, 1971; Gander, 1990; Higham et al., 2004) to discover a weight matrix $\bar{\mathbf{W}}_l$ with the norm $1/K_l$ near to the original weight $\mathbf{W}_l$, i.e., $\|\bar{\mathbf{W}}_l\| = 1/K_l$.

By enforcing $\|\nabla \bar{f}(\mathbf{x})\| = K_l$ and $\|\tilde{\mathbf{W}}_l\| = 1/K_l$ at each layer, the composite Lipschitz constant of the entire neural network is constrained to 1. We theoretically derive an important property of our 1-Lipschitz neural network for expressive and robust multiple graph learning: K_l is relevant to and should be adaptive to input graphs and layers. Given an error budget between our 1-Lipschitz neural network and regular neural network without constrained weight and gradient, we validate the existence of feasible K_l under the 1-Lipschitz constraint, i.e., derive lower and upper bounds of feasible K_l for expressive and robust multiple graph learning against adversarial attacks. The theoretical analysis is conducted to demonstrate that our 1-Lipschitz neural network with K_l-Lipschitz Weibull activation function $\bar{f}$ is universal Lipschitz function approximator, i.e., $\bar{f}$ can approximate any linear or nonlinear functions.

Empirical evaluation over graph classification and graph matching demonstrates the superior performance of our ERNN model against state-of-the-art robust graph learning models and Lipschitz-bound neural architectures. We validate that the proposed robust learning strategies are transferable to other popular graph learning tasks in Appendix A.2.

2. Background and Problem Statement

2.1. C-Lipschitz Functions

A function $F : \mathbb{R}^N \mapsto \mathbb{R}^M$ is globally Lipschitz continuous on variable space $\mathcal{X} \subseteq \mathbb{R}^N$ if there exists a nonnegative constant $C \geq 0$ such that for all $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathcal{X}$.

$$\|F(\mathbf{x}_2) - F(\mathbf{x}_1)\| \leq C\|\mathbf{x}_2 - \mathbf{x}_1\|, \ \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \qquad (1)$$

where the smallest such C for which the above inequality holds is the Lipschitz constant of F. If the Lipschitz constant of a function is C, it is called a C-Lipschitz function. If F is everywhere differentiable then its Lipschitz constant is bounded by the operator norm of its Jacobian.

If $\mathbf{x}_2$ is denoted as a perturbation of $\mathbf{x}_1$, i.e., $\mathbf{x}_2 = \mathbf{x}_1 + \delta$, then the Lipschitz constant is the maximum ratio between perturbations $\|F(\mathbf{x}_1 + \delta) - F(\mathbf{x}_1)\|$ in the output space and perturbations $\|(\mathbf{x}_1 + \delta) - \mathbf{x}_1\|$ in the input space. Thus, it is a useful metric to measure the sensitivity of the function F regarding input perturbations.

2.2. Lipschitz-Constrained Neural Networks

Given an input vector $\mathbf{x} \in \mathbb{R}^{N_0}$, a $(L+1)$-layer neural network $\mathbf{y} = F(\mathbf{x})$ is defined as follows: layer 0 takes $\mathbf{h}_0 = \mathbf{x}$ as input, layers $1, \cdots, L-1$ produces the hidden representations $\mathbf{h}_2, \cdots, \mathbf{h}_{L-1}$, and layer L outputs an output variable $\mathbf{y} = \mathbf{z}_L \in \mathbb{R}^{N_L}$.

$$\begin{cases} \mathbf{z}_l = \mathbf{W}_l \mathbf{h}_{l-1} + \mathbf{b}_l, \mathbf{h}_l = f(\mathbf{z}_l), \textbf{if } 1 \leq l \leq L-1, \\ \mathbf{z}_l = \mathbf{W}_l \mathbf{h}_{l-1} + \mathbf{b}_l, \mathbf{y} = \mathbf{z}_l, \textbf{if } l = L. \end{cases} \qquad (2)$$

where N_l is the dimensionality of layer l, $\mathbf{W}_l \in \mathbb{R}^{N_l \times N_{l-1}}$ is the weight matrix between layers $l-1$ and l, and $\mathbf{b}_l \in \mathbb{R}^{N_l}$

is the bias for layer l. $\mathbf{z}_l = [\mathbf{z}_{l1}, \cdots, \mathbf{z}_{lN_l}]$ denotes the pre-activation vector in layer l and $\mathbf{h}_l = [\mathbf{h}_{l1}, \cdots, \mathbf{h}_{lN_l}]$ is the activation vector with $\mathbf{z}_l$. f is the activation function. At layer L, the pre-activation $\mathbf{z}_L$ is used as the final output $\mathbf{y}$.

The Lipschitz constant C of neural network is derived below.

$$C = \|\mathbf{W}_L\| \cdot \|\nabla_{\mathbf{z}_{L-1}} f\| \cdot \|\mathbf{W}_{L-1}\| \cdots \|\nabla_{\mathbf{z}_1} f\| \cdot \|\mathbf{W}_1\| \quad (3)$$

Adversarial robustness. When the function F is characterized by a deep neural network, tight bounds on its Lipschitz constant can be extremely useful to improve the robustness of the neural network against adversarial attacks. Concretely, if the Lipschitz constant C of F is limited to a small number, say 1 used in GroupSort (Anil et al., 2019; Cohen et al., 2019) and BCOP (Li et al., 2019b), such that $\|F(\mathbf{x}+\delta) - F(\mathbf{x})\| \leq \|(\mathbf{x}+\delta) - \mathbf{x}\|$ for a 1-Lipschitz neural network, then this can help effectively control the diffusion of input perturbations through the neural networks.

2.3. Multiple Graph Learning

Given a set of S graphs $\mathcal{G} = \{G^1, \cdots, G^S\}$. Each graph is denoted as $G^s = (V^s, E^s)$ $(1 \leq s \leq S)$, where $V^s = \{v_1^s, \cdots, v_{N^s}^s\}$ is the set of N_s nodes and $E^s = \{(v_i^s, v_j^s) : 1 \leq i, j \leq N_s, i \neq j\}$ is the set of edges. Each G^s has an $N_s \times N_s$ binary adjacency matrix $\mathbf{A}^s$, where each entry $\mathbf{A}_{ij}^s = 1$ if there exists an edge $(v_i^s, v_j^s) \in E^s$; otherwise $\mathbf{A}_{ij}^s = 0$. $\mathbf{A}_{i:}^s$ specifies the i^{th} row vector of $\mathbf{A}^s$ and is used to denote the representation of a node v_i^s. In this paper, we focus on enhancing the robustness of two multiple graph learning tasks, but it is straightforward to extend to others.

Graph classification. We associate each graph G^s with a label $y^s \in \mathcal{Y} = \{1, 2, ..., Y\}$, where Y is the number of classes. The training data denotes a set of known graph-label pairs, i.e., $D = \{(G^s, y^s) | G^s \leftrightarrow y^s, G^s \in \mathcal{G}, y^s \in \mathcal{Y}\}$, where $G^s \leftrightarrow y^s$ indicates that G^s and y^s are the corresponding graph-label pair. The goal of graph classification is to employ D as the training data to predict label y^s for graph G^s in the test data. A classifier $F : \mathcal{G} \mapsto \mathcal{Y}$ is optimized to minimize the following loss over all labeled graphs.

$$\mathcal{L} = \frac{1}{|D|} \sum_{s=1}^{|D|} L\big(F(G^s), y^s\big) \quad (4)$$

where L is the cross-entropy loss.

Graph matching. The entire training data consists of a set of training data between pairwise graphs, i.e., $D = \{D^{12}, \cdots, D^{1S}, \cdots, D^{(S-1)S}\}$. Each D^{st} $(1 \leq s < t \leq S)$ specifies a set of pre-aligned node pairs $D^{st} = \{(v_i^s, v_j^t) | v_i^s \leftrightarrow v_j^t, v_i^s \in V^s, v_j^t \in V^t\}$, where $v_i^s \leftrightarrow v_j^t$ represents that two nodes v_i^s and v_j^t are the equivalent ones in two graphs G^s and G^t. The objective of graph matching is to utilize D^{st} as the training data to identify the one-to-one node matchings between nodes v_i^s and v_j^t in the test data. By following the same idea in existing efforts (Man et al., 2016; Zhou et al., 2018a; Yasar & Çatalyürek, 2018; Li et al., 2019a), this paper aims to learn an embedding function F to map the node pairs $(v_i^s, v_j^t) \in D^{st}$ with different features across two graphs into common embedding space, i.e, minimize the distances between projected source nodes $F(v_i^s) \in D^{st}$ and target ones $F(v_j^t) \in D^{st}$. The node pairs $(v_i^s, v_j^t) \in D^{st}$ with the smallest distances in the test data are selected as the matching results.

$$\mathcal{L} = \sum_{s=1}^{S} \sum_{t=s+1}^{S} \mathbb{E}_{(v_i^s, v_j^t) \in D^{st}} \|F(v_i^s) - F(v_j^t)\|_2^2 \quad (5)$$

With the perturbed graphs as input, this paper aims to develop an attack-agnostic graph-adaptive 1-Lipschitz neural network to improve the robustness against adversarial perturbations while achieving remarkable expressive power in the context of multiple graph learning.

3. Expressive and Robust 1-Lipschitz Neural Network for Multiple Graph Learning

Deep learning models have demonstrated their remarkable expressive power by using nonlinear activation functions to stimulate and learn any linear or nonlinear functions representing a question, and provide accurate predictions. Regular neural networks do not hold the constraints on the weight and gradient in order to achieve the superior nonlinearity. GroupSort (Anil et al., 2019; Cohen et al., 2019) and BCOP (Li et al., 2019b) proposed 1-Lipschitz neural networks to achieve the model robustness while improving the expressive power by limiting both weight and gradient norms as 1. Their Lipschitz constant is computed below.

$$\begin{aligned} C &= \|\tilde{\mathbf{W}}_L\| \cdot \|\nabla_{\tilde{\mathbf{z}}_{L-1}} \tilde{f}\| \cdot \|\tilde{\mathbf{W}}_{L-1}\| \cdots \|\nabla_{\tilde{\mathbf{z}}_1} \tilde{f}\| \cdot \|\tilde{\mathbf{W}}_1\| \\ &= 1 \cdot 1 \cdot 1 \cdots 1 \cdot 1 = 1 \end{aligned} \quad (6)$$

The GroupSort activation function $\tilde{f}$ is essentially a permutation operation that sorts and permutes the elements in each $\mathbf{z}_l$ on each layer l. Thus, both $\|\tilde{\mathbf{W}}_l\|$ and $\|\nabla_{\tilde{\mathbf{z}}_l} \tilde{f}\|$ are constrained to 1. This may lead to sub-optimal defense when tackling multiple graphs with individual characteristics.

We propose an attack-agnostic graph-adaptive 1-Lipschitz neural network with a K_l-Lipschitz activation function $\bar{f}$, i.e., $\|\nabla_{\mathbf{z}_{l-1}} \bar{f}\| = K_l$ and a constrained weight matrix $\|\bar{\mathbf{W}}_l\| = 1/K_l$ for achieving better expressive power.

$$\begin{aligned} C &= \|\bar{\mathbf{W}}_L\| \cdot \|\nabla_{\bar{\mathbf{z}}_{L-1}} \bar{f}\| \cdots \|\bar{\mathbf{W}}_2\| \cdot \|\nabla_{\bar{\mathbf{z}}_1} \bar{f}\| \cdot \|\bar{\mathbf{W}}_1\| \\ &= \frac{1}{K_L} \cdot K_L \cdots \frac{1}{K_2} \cdot K_2 \cdot 1 = 1 \end{aligned} \quad (7)$$

where $\|\bar{\mathbf{W}}_l\| = 1/K_l$ if $l > 1$, otherwise $\|\bar{\mathbf{W}}_l\| = 1$. In this paper, we use ∞-norm for both weight and gradient. For ease representation, we use $\|\cdot\|$ to replace $\|\cdot\|_\infty$ in our 1-Lipschitz neural network.

The following theorems validate the existence of feasible K_l under the 1-Lipschitz constraint for robust and expressive multiple graph learning against adversarial attacks. Theorem 2 derives lower bound of feasible K_l and demonstrates

that selecting an appropriate K_l rather than 1 can guarantee that our 1-Lipschitz neural network achieves better expressive power than GroupSort (Anil et al., 2019; Cohen et al., 2019). Theorem 3 derives upper bound of feasible K_l when we are given an error budget between our 1-Lipschitz neural network and regular neural network without constrained weight and gradient. Theorem 3 also exhibits that K_l is relevant to and should be adaptive to input graphs and layers. Definition 1, Lemma1, and Theorem 1 are the preparation of the proof of Theorem 2-3.

Definition 1 *[Finite Partition of an Interval] A partition P of an interval $[a, b]$ on the real line is a sequence of a finite number of subintervals of $[a, b]$*

$$P = \{[x_0, x_1], [x_1, x_2], \cdots, [x_{m-1}, x_m], \cdots, [x_{M-1}, x_M]\} \quad (8)$$

where $a = x_0 < x_1 < x_2 < \cdots < x_{m-1} < x_m < \cdots < x_{M-1} < x_M = b$. The points $x_m, 0 \leq m \leq M$, are called the partition points in P. Each $[x_{m-1}, x_m]$ is referred to as a subinterval of the partition P.

Based on Definition 1, given large enough M, it is always feasible to partition an interval $[a, b]$ into multiple subintervals, such that any continuous nonlinear function f on $[a, b]$ become linear or near-linear on each subinterval.

Lemma 1 *[Lagrange's Mean Value Theorem] For any continuous function f on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there exists a point c in (a, b) such that the tangent at c is parallel to the secant line through the endpoints $(a, f(a))$ and $(b, f(b))$ (Sharma & Vasishtha, 2010).*

$$f'(c) = \frac{f(a) - f(b)}{a - b} \quad (9)$$

When a continuous function f is linear on the interval $[a, b]$, its slope is equal to $f'(c)$. The above observations inspire us to design and transform a nonlinear activation function $\bar{f}$ going through the origin into an approximate piecewise linear function for the proof of the following three theorems and to finally derive the lower and upper bounds of K_l.

Theorem 1 *For any K_l-Lipschitz nonlinear activation function $\bar{f} : \mathbb{R}^N \mapsto \mathbb{R}^N$, if $\bar{f}$ is everywhere differentiable and $\bar{f}(\mathbf{x}) = \mathbf{0} \in \mathbb{R}^N$ at $\mathbf{x} = \mathbf{0} \in \mathbb{R}^N$, then there must exist a linear function g such that $\bar{f}(\mathbf{x}) \leq g(\mathbf{x})$ for $\forall \mathbf{x}, \mathbf{x} \in \mathbb{R}^N$.*

Theorem 2 *By following the definition in Eq.(2), we build a $(L+1)$-layer 1-Lipschitz neural network $\bar{F} : \mathbb{R}^{N_0} \mapsto \mathbb{R}^{N_L}$, with a gradient norm preserving activation function $\|\nabla_{\bar{\mathbf{z}}_{l-1}} \bar{f}\| = K_l$ almost everywhere and a norm-constrained weight matrix $\|\bar{\mathbf{W}}_l\| = 1/K_l$ like the definitions in Eq.(7). If $K_l > 1$ for $\forall l, 2 \leq l \leq L$, our 1-Lipschitz neural network $\bar{F}$ achieves better expressive power than the neural network $\tilde{F}$ constructed by the GroupSort model (Anil et al., 2019; Cohen et al., 2019).*

Theorem 3 *Given a regular $(L+1)$-layer fully connected neural network $F : \mathbb{R}^{N_0} \mapsto \mathbb{R}^{N_L}$ with unconstrained ReLU as the activation and unrestricted weight, and our $(L+1)$-layer 1-Lipschitz neural network $\bar{F}$ defined in Theorem 2, if an error budget between each layer of two neural networks is limited to ε, i.e., $\|\bar{\mathbf{z}}_l - \mathbf{z}_l\| \leq \varepsilon$, where $\mathbf{z}_l$ and $\bar{\mathbf{z}}_l$ are the representations at layer l in F and $\bar{F}$ respectively, then $K_l \leq \min\left\{\frac{\|\bar{f}(\bar{\mathbf{z}}_{l-1})\|}{\|\mathbf{W}_l f(\mathbf{z}_{l-1})\| - \varepsilon}, \max\left\{ - \min_j \frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{(l-1)j}}, \max_j \frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{(l-1)j}} \right\}\right\}$.*

Proof. Please refer to Appendix A.1 for detailed proof of the above three theorems.

In this paper, we use a random variable $\mathbf{x}$ denoting the node representation as input of the neural network for multiple graph learning. Since $\mathbf{h}_0 = \mathbf{x}$, $\mathbf{z}_1$ depends on $\mathbf{h}_0$, and other $\mathbf{z}_l$ $(\forall l, 2 \leq l \leq L)$ are related to $\mathbf{z}_1$, the upper bound of K_l is relevant to input graphs and layers.

Based on Theorems 2-3, we need to make $1 < K_l \leq \min\left\{\frac{\|\bar{f}(\bar{\mathbf{z}}_{l-1})\|}{\|\mathbf{W}_l f(\mathbf{z}_{l-1})\| - \varepsilon}, \max\left\{ - \min_j \frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{(l-1)j}}, \max_j \frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{(l-1)j}} \right\}\right\}$, such that our 1-Lipschitz neural network $\bar{F}$ achieves better expressive power than the GroupSort model $\tilde{F}$ and comparable quality to the regular network F at layer l.

3.1. Constraining $\|\bar{\mathbf{W}}_l\| = 1/K_l$

According to Theorem 2 in the GroupSort paper (Anil et al., 2019), when enforcing $\|\bar{\mathbf{W}}_l\| = 1$, it needs to adjust the weight matrix $\mathbf{W}_l$ to have singular values of 1 without sacrificing nonlinear processing capacity. In our case, the equivalent problem is that all singular values of the norm-contained weight matrix $\|K_l \bar{\mathbf{W}}_l\|$ are equal to 1.

Recall that the singular values of a real matrix $\mathbf{A}$ are the eigenvalues of the positive-semidefinite real matrix $\mathbf{A}^T\mathbf{A}$, where $\mathbf{A}^T$ is the transpose of $\mathbf{A}$. The singular values of $\mathbf{A}$ are all 1 iff $\mathbf{A}$ is orthogonal, i.e., $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Thus, the original problem is equivalent to finding the nearest orthonormal matrix of $K_l\mathbf{W}_l$.

For ease of representation, let $\mathbf{A} = K_l\mathbf{W}_l$ and $\mathbf{B} = K_l\bar{\mathbf{W}}_l$. Formally, given an $N_l \times N_{l-1}$ matrix $\mathbf{A}$, we aim to find the nearest $N_l \times N_{l-1}$ matrix $\mathbf{B}$ with N_l orthonormal columns (i.e. $\mathbf{B}^T\mathbf{B} = \mathbf{I}$), i.e., we try to minimize $\|\mathbf{A} - \mathbf{B}\|_F = \sqrt{\mathrm{trace}\left((\mathbf{A} - \mathbf{B})^T(\mathbf{A} - \mathbf{B})\right)}$.

Polar decomposition is an effective technique that finds the nearest orthonormal matrix. However, traditional iterative algorithms have non-trival computational cost based on operation counts, including Björck Orthonormalization (Bjorck & Bowie, 1971) and Newton iteration-based methods (Gander, 1990; Byers & Xu, 2008). In order to improve the cost of the iterative algorithms, a fast hybrid algorithm was proposed to adaptively switches from the matrix inversion based iteration to a matrix multiplication based iteration (Higham & Schreiber, 1990). The hybrid algorithm tends to require

at most 7 iterations for convergence. In addition, if $\mathbf{B}$ is not required to full accuracy then there is no need to iterate to converge–just 1 or 2 iterations may yield a sufficiently accurate approximation to $\mathbf{B}$. In our implementation, we use 3-4 iterations of the fast hybrid algorithm per forward pass to get a good approximation $\|K_l\bar{\mathbf{W}}_l\|$.

Theorem 4 exhibits the uniqueness of the nearest orthonormal matrix $\mathbf{B}$ by using the above fast hybrid algorithm.

Theorem 4 *Given an $N_l \times N_{l-1}$ matrix $\mathbf{A}$, the nearest orthonormal matrix $\mathbf{B}$ of $\mathbf{A}$ is unique. It is equal to $\hat{\mathbf{B}} = \mathbf{A}\mathbf{H}^{-1}$ by using the fast hybrid polar decomposition, where $\mathbf{H} = \sqrt{\mathbf{A}^T\mathbf{A}}$ is positive definite.*

Proof. Please refer to Appendix A.1 for detailed proof.

Practically, we can compute a residual $\mathbf{R} = \mathbf{A}^T\mathbf{A} - \mathbf{I}$ and then use a series to approximate $\hat{\mathbf{B}}$.

$$\begin{aligned}\hat{\mathbf{B}} &= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1/2} = \mathbf{A}(\mathbf{I}+\mathbf{R})^{-1/2} \\ &= \mathbf{A} - \mathbf{A}\mathbf{R}(1/2 - 3\mathbf{R}/8 + 5\mathbf{R}^2/16 - 35\mathbf{R}^4/128 + \cdots)\end{aligned} \tag{10}$$

In order to further improve the computational cost, we can calculate the first few terms without losing much accuracy. If $\|\mathbf{R}\|_F$ comes near to zero, then $\hat{\mathbf{B}}$ can be approximated adequately $\check{\mathbf{B}} = \mathbf{A} - \mathbf{A}\mathbf{R}/2 = \hat{\mathbf{B}}(\mathbf{I} - 3\mathbf{R}^2/8 + \mathbf{R}^3/8 - \cdots)$ as its residual $\check{\mathbf{B}}^T\check{\mathbf{B}} - \mathbf{I} = \mathbf{R}^2(\mathbf{R} - 3\mathbf{I})/4$ is trivial. On the other hand, if $\|\mathbf{R}\|_F << 1/2$, then the columns of $\check{\mathbf{B}}$ will be more nearly orthonormal than those of $\mathbf{A}$; and repeating upon $\check{\mathbf{B}}$ the process performed upon $\mathbf{A}$ can yield an approximation $\hat{\mathbf{B}}(\mathbf{I} - 27\mathbf{R}^4/128 + \cdots)$.

3.2. K_l-Lipschitz Weibull Activation $\|\nabla_{\bar{\mathbf{z}}_l}\bar{f}\| = K_l$

To bound the Lipschitz constant of neural network, the gradient norm must be preserved by each layer in the network during backpropagation. Unfortunately, Theorem 5 exhibits that norm-constrained neural networks with common 1-Lipschitz activation functions (e.g. ReLU, Leaky ReLU, Sigmoid, SoftPlus, or tanh) must trade nonlinear processing for gradient norm preservation, leading to less expressive networks. Namely, such norm-constrained neural networks can only approximate the linear functions with less expressive power. It is straightforward to extend the conclusion of Theorem 5 to our 1-Lipschitz neural network. In addition, for our 1-Lipschitz neural network, most of popular nonlinear activation functions with gradient norm preservation $\|\nabla_{\bar{\mathbf{z}}_{l-1}}f\| = K_l$, can not achieve the feasible K_l, since the maximum values of their derivatives are smaller than the lower bound of the feasible K_l, e.g, 1 for ReLU, Leaky ReLU, and PReLU, 0.25 for Sigmoid, 1 for tanh and Softplus, and 1 for GroupSort (Anil et al., 2019). Therefore, we utilize the Weibull distribution to design an expressive K_l-Lipschitz nonlinear activation function to allow our 1-Lipschitz neural network to approximate any functions.

Theorem 5 *Consider a 1-Lipschitz neural network $\bar{F} : \mathbb{R}^{N_0} \mapsto \mathbb{R}$, built with norm-constrained weights ($\|\bar{\mathbf{W}}_l\| \leq 1$) and 1-Lipschitz, element-wise, monotonic activation functions $\|\nabla_{\bar{\mathbf{z}}_{l-1}}f\| = 1$. If $||\nabla_{\mathbf{x}}\bar{F}(\mathbf{x})|| = 1$ almost everywhere, then $\bar{F}$ is linear (Anil et al., 2019).*

In statistics, the Weibull distribution is a continuous probability distribution (Weibull, 1951). It can model hazard functions that are monotonically decreasing, increasing or constant of the proportion of adopters over time, allowing it to describe any phase of an item's lifetime. Therefore, the major advantage of Weibull analysis is that it is suitable to reliability and failure analysis. In addition, a recent study reports that the non-saturating nonlinear activation functions, such as ReLU and Leaky ReLU, often achieve faster training than the saturating ones, e.g., Sigmoid and tanh (Krizhevsky et al., 2012). In order to achieve the advantage of non-saturating nonlinearity, we combine T Weibull activation functions $\bar{f}_1(z), \cdots, \bar{f}_T(z)$ with different parameters into a composite one, such that the upper bound of $\bar{f}(z)$ is increased to T.

$$\bar{f}(z) = \begin{cases} \sum_{t=1}^{T} \bar{f}_t(z), \text{if } z \geq \mu_t, \\ K_l z, \qquad \text{if } z < \mu_t. \end{cases}, f_t(z) = 1 - e^{-(\frac{z-\mu_t}{\lambda_t})^{\alpha_t}} \tag{11}$$

where $\bar{f}_t$ is the t^{th} Weibull activation function with unique parameters α_t, λ_t, and μ_t. z is an element in $\bar{\mathbf{z}}_l$, $\alpha_t > 0$ is the shape parameter, $\lambda_t > 0$ is the scale parameter, and μ_t is the shift parameter. A value of $\alpha_t < 1$ indicates that the failure rate decreases with time. This happens if there is significant "infant mortality", or few defective parts failing to result in the malfunction of the entire item early and the failure rate decreasing over time as more parts gradually become defective over time. In the context of robust deep multiple graph learning, the perturbation diffusion over the layers is similar to a monotonically decreasing hazard function, i.e, the attack failure possibility decreases with the perturbation diffusion, and the diffusion of any perturbations finally leads to the attack success when enough diffusion is allowed. Thus, the Weibull activation function can effectively model the relationship between the perturbation diffusion and the attack failure.

The derivative of $f(z)$ is thus generated as follows.

$$\bar{f}'(z) = \begin{cases} \sum_{t=1}^{T} \frac{\alpha_t}{\lambda_t}\left(\frac{z-\mu_t}{\lambda_t}\right)^{\alpha_t - 1} e^{-(\frac{z-\mu_t}{\lambda_t})^{\alpha_t}}, \text{if } z \geq \mu_t, \\ K_l, \qquad\qquad\qquad\qquad \text{if } z < \mu_t. \end{cases} \tag{12}$$

Theorem 6 *Given T Weibull activation functions with the definition in Eq.(11), there must exist solutions of parameters α_t, λ_t, and μ_t to guarantee $\|\nabla_{\bar{\mathbf{z}}_l}\bar{f}\| = K_l$.*

Proof. Please refer to Appendix A.1 for detailed proof.

In statistics, a scale parameter λ_t is a special kind of numerical parameter of a parametric family of probability distribu-

tions. If it is large, then the distribution will be more spread out; if it is small then it will be more concentrated. Therefore, $\bar{f}(z)$ is more sensitive to λ_t. Notice that $\|\nabla_{\bar{\mathbf{z}}_l}\bar{f}\|_\infty = \bar{f}'(\bar{\mathbf{z}}_{lU}) = \frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{lU}}$, where $\frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{lU}} = \max\{|\frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{l1}}|, \cdots, |\frac{\partial \bar{f}}{\partial \bar{\mathbf{z}}_{lN_l}}|\}$. With selected α_t and μ_t, we utilize the Newton-Raphson method to find an approximate λ_t of the following equation to make $\bar{f}'(\bar{\mathbf{z}}_{lU}) = K_l$ when $z \geq \mu_t$ (Gil et al., 2007).

$$\frac{\alpha_t}{\lambda_t}\Big(\frac{\bar{\mathbf{z}}_{lU}-\mu_t}{\lambda_t}\Big)^{\alpha_t-1} e^{-(\frac{\bar{\mathbf{z}}_{lU}-\mu_t}{\lambda_t})^{\alpha_t}} = K_l \tag{13}$$

3.3. Universal Approximation of K_l-Lipschitz Weibull Activation

The following theorems demonstrate that our 1-Lipschitz neural network architecture with a K_l-Lipschitz activation function $\bar{f}$ is universal Lipschitz function approximator, i.e., $\bar{f}$ can approximate any linear or nonlinear functions.

Definition 2 *We say that a set of functions, $\mathcal{F}$, is a lattice if for any $f, g \in \mathcal{F}$ we have $\max(f, g) \in \mathcal{F}$ and $\min(f, g) \in \mathcal{F}$ (where $\max$ and $\min$ are defined pointwise).*

Lemma 2 *[Restricted Stone-Weierstrass Theorem] Suppose that (X, d_X) is a compact metric space with at least two points and $\mathcal{F}$ is a lattice in $C_{\mathcal{F}}(X, \mathbb{R})$ with the property that for any two distinct elements $x, y \in X$ and any two real numbers a and b such that $|a - b| \leq d_X(x, y)$ there exists a function $f \in \mathcal{F}$ such that $f(x) = a$ and $f(y) = b$. Then $\mathcal{F}$ is dense in $C_{\mathcal{F}}(X, \mathbb{R})$ (Anil et al., 2019).*

Theorem 7 *Let $\mathcal{LN}_\infty^{N_L} : \mathbb{R}^{N_0} \mapsto \mathbb{R}^{N_L}$ denote the class of $(L+1)$-layer 1-Lipschitz neural networks $\bar{F}$ with norm-constrained weight matrices $\|\bar{\mathbf{W}}_l\|_\infty = 1$ $(l = 1)$ and $\|\bar{\mathbf{W}}_l\|_\infty = 1/K_l$ $(l > 1)$, and gradient norm preserving activation function $\|\nabla_{\bar{\mathbf{z}}_{l-1}}\bar{f}\|_\infty = K_l$, by following the definitions in Eqs.(2) and (7). Let input $\mathcal{X}$ be a closed and bounded subset of $\mathbb{R}^{N_0}$ with the L_∞ metric. Then the closure of $\mathcal{LN}_\infty^{N_L}$ is dense in $C_{\mathcal{F}}(\mathcal{X}, \mathbb{R})$.*

Proof. Please refer to Appendix A.1 for detailed proof of the above two theorems.

Theorem 7 demonstrates that the closure of $\mathcal{LN}_\infty^{N_L}$ is dense in $C_{\mathcal{F}}(\mathcal{X}, \mathbb{R})$. Namely, for any input $\mathbf{x} \in \mathbb{R}^{N_0}$, function $\bar{F}(\mathbf{x}) \in \mathcal{LN}_\infty^{N_L}$ can be used to approximate any function $\mathbb{R}^{N_0} \mapsto \mathbb{R}^{N_L}$ in continuous function space $C_{\mathcal{F}}(\mathcal{X}, \mathbb{R})$.

4. Experiments

We perform extensive evaluation on the robustness of our ERNN model for graph classification on three real datasets: BZR, BZR_MD and MUTAG (Jin et al., 2020a; TUD) and for graph matching over three datasets: autonomous systems (AS) (AS), CAIDA relationships datasets (CAI), and DBLP coauthor graphs (DBL), as shown in Table 1.

Graph classification baselines. We compare our ERNN model with one regular graph classification algorithm, two robust node classification models, one general graph denoising method, two state-of-the-art robust graph classification models against adversarial attacks, and two representative Lipschitz-bound neural architectures for restricting the perturbation propagation. **PAN** (Ma et al., 2020c) is a path integral based GNN containing self-consistent convolution and pooling units for producing regular graph classification. **Pro-GNN** (Jin et al., 2020b) jointly learns a clean graph and a robust GNN model for defending node classification. **GRAND** (Feng et al., 2020) is a graph random neural network with random propagation and data augmentation to increase the robustness of node classification. **GCN-SVD** (Entezari et al., 2020) is a general perturbation elimination model irrelevant to specific graph learning architectures. **RoboGraph** (Jin et al., 2020a) is the first certifiably robust graph classification model based on Lagrange dualization and convex envelope. **GraphCL** (You et al., 2020) is a graph contrastive learning framework with data augmentations for GNN pre-training for boosting the robustness of graph classification. **GroupSort** (Anil et al., 2019; Cohen et al., 2019) is a 1-Lipschitz fully-connected neural network that restricts the perturbation propagation by imposing a Lipschitz constraint on each layer. **BCOP** (Li et al., 2019b) is a Lipschitz-constrained convolutional network with expressive parameterization of orthogonal convolution operations. For two node classification models, the average of node labels within the same graphs is output as graph labels.

Graph matching baselines. We compare the ERNN model with six state-of-the-art graph matching algorithms, GroupSort, and BCOP. **FINAL** (Zhang & Tong, 2016) leverages both node and edge attributes to solve the attributed network alignment problem. Its supervised version with prior alignment preference matrix is used for the evaluation. **REGAL** (Heimann et al., 2018) is an unsupervised network alignment framework that infers soft alignments by comparing joint node embeddings across graphs. and by computing pairwise node similarity scores across networks. **MOANA** (Zhang et al., 2019) is a supervised coarsening-alignment-interpolation multilevel network alignment algorithm with the supervision of a prior node similarity matrix. Deep graph matching consensus (**DGMC**) (Fey et al., 2020) is a supervised graph matching method that reaches a data-driven neighborhood consensus between matched node pairs. **CONE-Align** (Chen et al., 2020) models intra-network proximity with node embeddings and uses them to match nodes across networks in an unsupervised manner. **G-CREWE** (Qin et al., 2020) is a rapid unsupervised network alignment method via both graph compression and embedding in different coarsened networks. To our best knowledge, there are no other open-source defense baselines on graph matching available.

Attack models. We validate the robustness with four representative graph attack models. Random attack **(RND)**

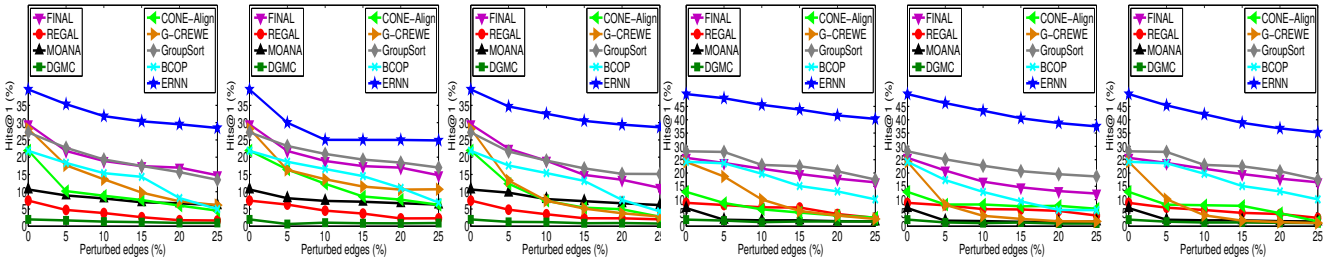
Table 1: Experiment Datasets

Dataset	**AS**		**CAIDA**		**DBLP**	
Graph	G^1	G^1	G^1	G^2	**2013**	**2014**
#Nodes	10,900	11,113	16,493	16,301	28,478	26,455
#Edges	31,180	31,434	33,372	32,955	128,073	114,588
#Matched Nodes	7,943		7,884		4,000	

Dataset	#Graphs	#Avg. Nodes	#Avg. Edges	#Classes
BZR	405	35.75	38.36	2
BZR_MD	306	21.30	225.06	2
MUTAG	188	17.93	19.79	2

Table 2: Graph classification with 5% perturbed edges

Dataset	BZR_MD		MUTAG		BZR	
Metric	**Acc.**	**RMSE**	**Acc.**	**RMSE**	**Acc.**	**RMSE**
PAN	0.541	0.680	0.681	0.570	0.779	0.471
Pro-GNN	0.631	0.610	0.643	0.601	0.738	0.512
GRAND	0.532	0.684	0.633	0.610	0.759	0.492
GCN-SVD	0.648	0.594	0.653	0.593	0.767	0.484
RoboGraph	0.667	0.579	0.644	0.601	0.790	0.458
GraphCL	0.652	0.593	0.545	0.680	0.808	0.439
GroupSort	0.547	0.676	0.606	0.636	0.756	0.494
BCOP	0.582	0.648	0.570	0.656	0.741	0.509
ERNN	**0.691**	**0.540**	**0.788**	**0.461**	**0.820**	**0.425**



(a) RND Attack (b) NEA Attack (c) GMA Attack

Figure 1: Matching on AS with varying perturbed edges

(a) RND Attack (b) NEA Attack (c) GMA Attack

Figure 2: Matching on CAIDA with varying perturbed edges

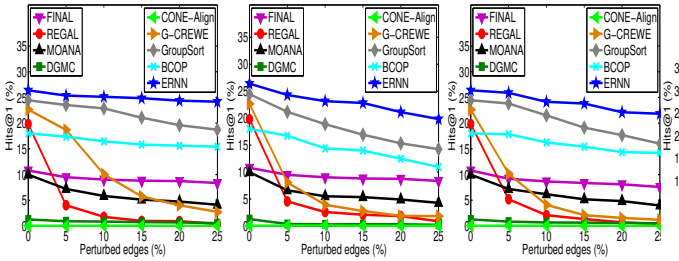
randomly adds and removes edges to generate perturbed graphs. **NEA** (Bojchevski & Günnemann, 2019) is an efficient adversarial attack method that poison the network structure against both network embedding and node classification. **GMA** (Zhang et al., 2020) is the only attack model on graph matching by pushing them to dense regions in two graphs to generate imperceptible and effective attacks. **RL-S2V** (Dai et al., 2018; Zhu et al., 2019) generates adversarial attacks on graph data based on reinforcement learning, which is used to attack both node classification and graph classification models.

Variants of ERNN model. We evaluate four variants to show the strengths of different components. ERNN-1 utilizes a fixed $K_l = 1$ in our 1-Lipschitz neural network. ERNN-R employs the ReLU as the activation. ERNN-N only uses the regular fully-connected neural network. ERNN operates with the full support of graph-adaptive K_l, Weibull activation, and 1-Lipschitz neural network.

Evaluation metrics. We employ two measures to evaluation the quality of graph classification: $Accuracy$ (Jin et al., 2020b; Entezari et al., 2020; Jin et al., 2020a; You et al., 2020) and root-mean-square error ($RMSE$). A higher $Accuracy$ or a smaller $RMSE$ shows a better classification. In addition, we use $Hits@K$ (Yasar & Çatalyürek, 2018; Fey et al., 2020) to verify the quality of graph matching. A larger $Hits@K$ value indicates a better graph matching.

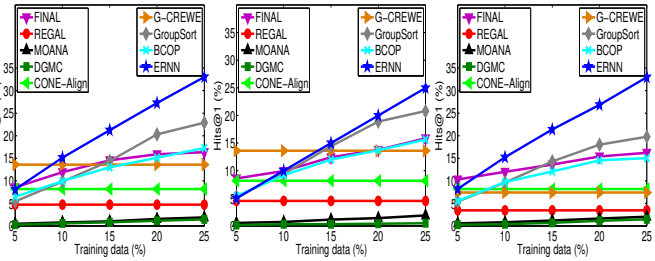
Defense performance on graph classification. Table 2 exhibits the $Accuracy$ and $RMSE$ scores of nine graph classification algorithms under RL-S2V attacks over three groups of datasets. We randomly sample 30% of labeled graphs as training data and the rest as test data. The number of perturbed edges is fixed to 5% in these experiments. It is observed that among nine graph classification methods the ERNN method achieve the highest $Accuracy$ and the smallest $RMSE$ on perturbed graphs in all experiments, showing the robustness of ERNN against adversarial attacks. Compared to the graph classification results by other models, ERNN, on average, achieves 15.2% $Accuracy$ boost and 18.6% $RMSE$ improvement on three groups of datasets. In addition, the promising performance of ERNN over all three datasets implies that ERNN has great potential as a general robust graph classification solution to other datasets, which is desirable in practice.

Defense performance on graph matching with varying perturbation edges. Figures 1-3 present the graph matching quality under three attack models by varying the ratios of perturbed edges from 0% to 25%. We choose 30% of matched node pairs as training data. It is obvious that the quality by each matching algorithm decreases with increasing perturbed edges. This phenomenon indicates that current graph matching methods are sensitive to adversarial attacks. However, ERNN still achieves the highest $Hits@1$ values (> 0.249), which are better than other eight methods in most tests. Especially, when the perturbation ratio is larger than 10%, the $Hits@1$ drop by ERNN becomes slowly.



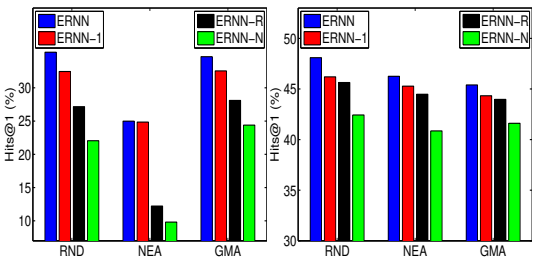
(a) RND Attack (b) NEA Attack (c) GMA Attack

Figure 3: Matching on DBLP with varying perturbed edges



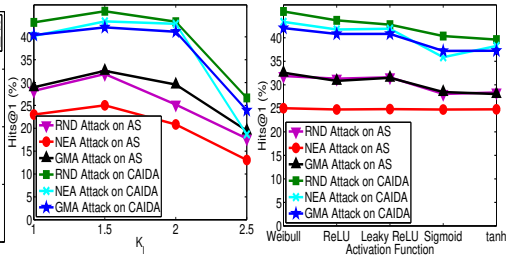
(a) RND Attack (b) NEA Attack (c) GMA Attack

Figure 4: Matching on AS with varying training ratios



(a) AS (b) CAIDA (c) DBLP

Figure 5: $Hits@1$ (%) of ERNN variants with 5% perturbed edges



(a) Fixed K_l (b) Activation Function

Figure 6: $Hits@1$ (%) with varying parameters

Impact of training data ratios. Figure 4 shows the quality of nine graph matching algorithms on AS under three attack models by varying the ratio of training data from 5% to 25%. Here, the number of perturbed edges is fixed to 30%. We make the following observations on the performances by nine graph matching algorithms. (1) The performance curves keep increasing when the training data ratio increases. (2) ERNN outperforms other methods in most experiments with the highest $Hits@1$ scores: $>$ 5.01% . When there are appropriate training data available ($\geq$ 10%), the quality improvement by ERNN is obvious. A reasonable explanation is that more training data makes ERNN be more resilient to poisoning attacks under a small perturbation budget.

Ablation study. Figure 5 presents the $Hits@1$ scores of graph matching on three datasets with four variants of our ERNN model. We observe the complete ERNN achieves the highest $Hits@1$ ($>$ 24.9%) on AS, ($>$ 35.2%) over CAIDA, and ($>$ 17.5%) on DBLP, which are obviously better than other versions. Compared with ERNN-R, ERNN-1 performs well in most experiments. A reasonable explanation is that ReLU must trade nonlinear processing for gradient norm preservation, leading to less expressive neural networks. In addition, ERNN-R achieves the better performance than ERNN-N. A rational guess is that Lipschitz-bounded neural architecture is able to restrict the perturbation propagation on the neural networks, achieving remarkable robustness. These results illustrate all of graph-adaptive K_l, Weibull activation, and Lipschitz-bounded neural network are important in producing robust graph matching.

Impact of K_l. Figure 6 (a) shows the impact of K_l in our ERNN model under three attack methods over three groups of datasets. The performance curves initially raise when K_l increases. As shown in the theoretical analysis, $K_l = 1$ is not the optimal solution and a large K_l can make the 1-Lipschitz neural network more robust. Later on, the performance curves keep relatively stable or even decreasing when K_l continuously increases. A reasonable explanation is that the too large K_l makes the norm-constrained weight matrices very small, such that it may hinder the feedforward of the neural network. Thus, it is important to choose the appropriate K_l for robust training.

Impact of activation function. Figure 6 (b) measures the effect of different activation functions in the ERNN model for the graph matching by using different activation functions. It is observed that among five activation functions, the $Hits@1$ values of our K_l-Lipschitz Weibull activation function outperforms all other competitors. This demonstrates that our Weibull activation is able to better maintain nonlinearity in Lipschitz-bounded neural networks. In addition, ReLU and Leaky ReLU achieve better performance than Sigmoid and tanh. This is consistent with the fact that non-saturating nonlinear activation functions often achieve faster training than saturating ones.

Validation of adversarial robustness on generic learning tasks. We conduct the experiments to validate the adversarial robustness of ERNN on two generic learning tasks: image classification and Wasserstein Distance estimation,

Table 3: Accuracy on image classification

Dataset	MNIST		CIFAR-10	
#Network Depth	3	5	3	5
GroupSort	0.91	0.91	0.48	0.50
BCOP	0.94	0.94	0.45	0.47
ERNN	**0.97**	**0.97**	**0.55**	**0.56**

Table 4: Wasserstein Distance estimation

Dataset	MNIST		CIFAR-10	
#Network Depth	3	5	3	5
GroupSort	2.31	2.55	2.23	2.74
BCOP	5.82	6.04	5.34	6.03
ERNN	**7.19**	**8.03**	**7.26**	**7.88**

Table 5: Lipschitz constant on graph matching

Dataset	AS	CAIDA	DBLP
Unbounded Networks	2880	1530	1260
ERNN	**0.910**	**0.968**	**0.955**

by following similar setting in Table 2 in the GroupSort paper, as shown in Tables 3 and 4. We utilize two standard image datasets: MNIST (Deng, 2012) and CIFAR-10 (Krizhevsky, 2009). Our ERNN model with Weibull activation still achieves the best performance in all tests.

Lipschitz constant estimate. The GroupSort and our ERNN models focus on designing 1-Lipschitz neural networks with enforcing each layer and entire network to be 1-Lipschitz. Table 5 shows the computed Lipschitz constants by ERNN and unbounded neural network on graph matching. The former is close to 1 and much smaller than the latter. This demonstrates that ERNN is able to successfully constrain the Lipschitz constant to 1.

5. Related Work

Lipschitz-bounded neural networks. Enforcing Lipschitz constraints in the training of neural networks is useful for ensuring adversarial robustness against adversarial attacks. Existing research activities can be classified into three broad categories: (1) Regularization techniques penalize or bound the Jacobian of the neural network, constraining the Lipschitz constant locally (Drucker & LeCun, 1992; Sokolic et al., 2017; Gulrajani et al., 2017; Krishnan et al., 2020). It is easy to train networks under the gradient penalties, but these methods cannot enforce the Lipschitz constraint globally; (2) Architecture constraint-based methods constrain the operator norm of each layer's weights, such as the matrix spectral norm (Cissé et al., 2017; Yoshida & Miyato, 2017; Miyato et al., 2018). These approaches enforce the Lipschitz constraint but come at a cost in expressive power; and (3) Gradient norm preserving architectures enforce both weight norm and gradient norm as 1 to constraint 1-Lipschitz neural networks globally, which improves the expressive power to a certain degree (Anil et al., 2019; Li et al., 2019b; Cohen et al., 2019). However, simply limiting the above two norms to 1 still sacrifices the expressive power, in comparison with regular neural networks without constrained weight and gradient.

Adversarial defenses on multiple graph learning. Graph data analysis have attracted active research in the last decade (Cheng et al., 2009; Zhou et al., 2009; 2010; Cheng et al., 2011; Zhou & Liu, 2011; Cheng et al., 2012; Lee et al., 2013; Su et al., 2013; Zhou et al., 2013; Zhou & Liu, 2013; Palanisamy et al., 2014; Zhou et al., 2014; Zhou & Liu, 2014; Su et al., 2015; Zhou et al., 2015b; Bao et al., 2015; Zhou et al., 2015d; Zhou & Liu, 2015; Zhou et al., 2015a;c; Lee et al., 2015; Zhou et al., 2016; Zhou, 2017; Palanisamy et al., 2018; Zhou et al., 2018c;b; Ren et al., 2019; Zhou et al., 2019c;b;d; Zhou & Liu, 2019; Goswami et al., 2020; Wu et al., 2020a; 2021a; Zhou et al., 2020a;b; Zhang et al., 2020; Zhou et al., 2020c; 2021; Jin et al., 2021; Wu et al., 2021b; Zhang et al., 2021). The majority of existing techniques focus on tackling vulnerability and improving robustness on single graph learning tasks under adversarial attacks. Recently, researchers have demonstrated that multiple graph learning models, especially deep learning-based models, are highly sensitive to adversarial attacks, including graph classification (Dai et al., 2018; Tang et al., 2020a; Xi et al., 2020) and graph matching (Zhang et al., 2020). Several adversarial defense models have been developed to improve the robustness of multiple graph learning models in graph classification (Zhang & Lu, 2020; You et al., 2020; Jin et al., 2020a; Gao et al., 2020), graph matching (Yu et al., 2021), and multiple network embedding (Zhou et al., 2020b). RGM is a robust graph matching model against visual noise, including image deformations, rotations, and outliers for image matching, but it fails to defend adversarial attacks on graph topology (Yu et al., 2021). A common characteristics of the above techniques is that they often defend specific attacks on particular learning tasks, rather than attack-agnostic defense models.

6. Conclusions

In this work, we proposed an expressive 1-Lipschitz neural network to improve the robustness of multiple graph learning. First, the theoretical analysis is conducted to derive lower and upper bounds of feasible K_l under the 1-Lipschitz constraint. Second, a K_l-Lipschitz nonlinear activation function is designed to enforce the gradient norm as K_l at each layer. Finally, the nearest matrix orthogonalization and polar decomposition techniques are utilized to constraint the weight norm as $1/K_l$.

Acknowledgements

This research was partially supported by the grants of NSF IIS-2041065 and ALDOT 931-055.

References

https://snap.stanford.edu/data/Oregon-2.html.

https://snap.stanford.edu/data/as-Caida.html.

http://dblp.uni-trier.de/xml/.

https://chrsmrrs.github.io/datasets/docs/datasets/.

Al-Rfou, R., Perozzi, B., and Zelle, D. DDGK: learning graph representations for deep divergence graph kernels. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pp. 37–48, 2019.

Anil, C., Lucas, J., and Grosse, R. B. Sorting out lipschitz function approximation. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 291–301, 2019.

Bao, X., Liu, L., Xiao, N., Zhou, Y., and Zhang, Q. Policy-driven autonomic configuration management for nosql. In *Proceedings of the 2015 IEEE International Conference on Cloud Computing (CLOUD'15)*, pp. 245–252, New York, NY, June 27-July 2 2015.

Bjorck, A. and Bowie, C. An iterative algorithm for computing the best estimate of an orthogonal matrix. *SIAM J. Numer. Anal.*, 8:358–364, 1971.

Bojchevski, A. and Günnemann, S. Adversarial attacks on node embeddings via graph poisoning. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 695–704, 2019.

Byers, R. and Xu, H. A new scaling for newton's iteration for the polar decomposition and its backward stability. *SIAM J. Matrix Anal. Appl.*, 30(2):822–843, 2008.

Chang, H., Rong, Y., Xu, T., Huang, W., Zhang, H., Cui, P., Zhu, W., and Huang, J. A restricted black-box adversarial framework towards attacking graph embedding models. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, New Yrok, NY, USA, February 7 - 12, 2020*, 2020.

Chen, X., Heimann, M., Vahedian, F., and Koutra, D. Cone-align: Consistent network alignment with proximity-preserving node embedding. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pp. 1985–1988, 2020.

Cheng, H., Lo, D., Zhou, Y., Wang, X., and Yan, X. Identifying bug signatures using discriminative graph mining. In *Proceedings of the 18th International Symposium on Software Testing and Analysis (ISSTA'09)*, pp. 141–152, Chicago, IL, July 19-23 2009.

Cheng, H., Zhou, Y., and Yu, J. X. Clustering large attributed graphs: A balance between structural and attribute similarities. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(2):1–33, 2011.

Cheng, H., Zhou, Y., Huang, X., and Yu, J. X. Clustering large attributed information networks: An efficient incremental computing approach. *Data Mining and Knowledge Discovery (DMKD)*, 25(3):450–477, 2012.

Chu, X., Fan, X., Yao, D., Zhu, Z., Huang, J., and Bi, J. Cross-network embedding for multi-network alignment. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pp. 273–284, 2019.

Cissé, M., Bojanowski, P., Grave, E., Dauphin, Y. N., and Usunier, N. Parseval networks: Improving robustness to adversarial examples. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pp. 854–863, 2017.

Cohen, J. E. J., Huster, T., and Cohen, R. Universal lipschitz approximation in bounded depth neural networks. *CoRR*, abs/1904.04861, 2019.

Dai, H., Li, H., Tian, T., Huang, X., Wang, L., Zhu, J., and Song, L. Adversarial attack on graph structured data. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pp. 1123–1132, 2018.

Dai, Q., Shen, X., Zhang, L., Li, Q., and Wang, D. Adversarial training methods for network embedding. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pp. 329–339, 2019.

Deng, L. The MNIST database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Process. Mag.*, 29(6):141–142, 2012.

Drucker, H. and LeCun, Y. Improving generalization performance using double backpropagation. *IEEE Trans. Neural Networks*, 3(6):991–997, 1992.

Du, X., Yan, J., and Zha, H. Joint link prediction and network alignment via cross-graph embedding. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pp. 2251–2257, 2019.

Elinas, P., Bonilla, E. V., and Tiao, L. Variational inference for graph convolutional networks in the absence of graph data and adversarial settings. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, Online*, December 6-12 2020.

Entezari, N., Al-Sayouri, S., Darvishzadeh, A., and Papalexakis, E. All you need is low (rank): Defending against adversarial attacks on graphs. In *Proceedings of the 13th ACM International Conference on Web Search and Data Mining, WSDM 2020, Houston, TX, February 3-7, 2020*, 2020.

Fan, S., Wang, X., Shi, C., Lu, E., Lin, K., and Wang, B. One2multi graph autoencoder for multi-view graph clustering. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pp. 3070–3076, 2020.

Fazlyab, M., Robey, A., Hassani, H., Morari, M., and Pappas, G. J. Efficient and accurate estimation of lipschitz constants for deep neural networks. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pp. 11423–11434, 2019.

Feng, W., Zhang, J., Dong, Y., Han, Y., Luan, H., Xu, Q., Yang, Q., Kharlamov, E., and Tang, J. Graph random neural networks for semi-supervised learning on graphs. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, Online*, December 6-12 2020.

Fey, M., Lenssen, J. E., Morris, C., Masci, J., and Kriege, N. M. Deep graph matching consensus. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020.

Fu, D., Xu, Z., Li, B., Tong, H., and He, J. A view-adversarial framework for multi-view network embedding. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pp. 2025–2028, 2020.

Fu, Y., Wang, P., Du, J., Wu, L., and Li, X. Efficient region embedding with multi-view spatial networks: A perspective of locality-constrained spatial autocorrelations. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 906–913, 2019.

Gander, W. Algorithms for the polar decomposition. *SIAM J. Scientific Computing*, 11(6):1102–1115, 1990.

Gao, Z., Hu, R., and Gong, Y. Certified robustness of graph classification against topology attack with randomized smoothing. In *2020 IEEE Global Communications Conference, GLOBECOM 2020, Taipei, Taiwan, December 8-10, 2020*, 2020.

Gil, A., Segura, J., and Temme, N. M. *Numerical methods for special functions*. SIAM, 2007.

Goswami, S., Pokhrel, A., Lee, K., Liu, L., Zhang, Q., and Zhou, Y. Graphmap: Scalable iterative graph processing using nosql. *The Journal of Supercomputing (TJSC)*, 76 (9):6619–6647, 2020.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5767–5777, 2017.

Hasanzadeh, A., Hajiramezanali, E., Narayanan, K. R., Duffield, N., Zhou, M., and Qian, X. Semi-implicit graph variational auto-encoders. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 10711–10722, 2019.

Heimann, M., Shen, H., Safavi, T., and Koutra, D. REGAL: representation learning-based graph alignment. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, pp. 117–126, 2018.

Higham, N. J. and Schreiber, R. S. Fast polar decomposition of an arbitrary matrix. *SIAM J. Sci. Comput.*, 11(4):648–655, 1990.

Higham, N. J., Mackey, D. S., Mackey, N., and Tisseur, F. Computing the polar decomposition and the matrix sign decomposition in matrix groups. *SIAM J. Matrix Anal. Appl.*, 25(4):1178–1192, 2004.

Huynh, T. T., Tong, V. V., Nguyen, T. T., Yin, H., Weidlich, M., and Hung, N. Q. V. Adaptive network alignment with unsupervised and multi-order convolutional networks. In *36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA, April 20-24, 2020*, pp. 85–96, 2020.

Jia, J., Wang, B., Cao, X., and Gong, N. Z. Certified robustness of community detection against adversarial structural perturbation via randomized smoothing. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pp. 2718–2724, 2020.

Jin, H., Shi, Z., Peruri, A., and Zhang, X. Certified robustness of graph convolution networks for graph classification under topological attacks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020 (NeurIPS'20)*, Online, December 6-12 2020a.

Jin, R., Li, D., Gao, J., Liu, Z., Chen, L., and Zhou, Y. Towards a better understanding of linear models for recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'21)*, Virtual Event, August 14-18 2021.

Jin, W., Ma, Y., Liu, X., Tang, X., Wang, S., and Tang, J. Graph structure learning for robust graph neural networks. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pp. 66–74, 2020b.

Kriege, N. M., Neumann, M., Morris, C., Kersting, K., and Mutzel, P. A unifying view of explicit and implicit feature maps of graph kernels. *Data Min. Knowl. Discov.*, 33(6): 1505–1547, 2019.

Krishnan, V., Makdah, A. A. A., and Pasqualetti, F. Lipschitz bounds and provably robust training by laplacian smoothing. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Krizhevsky, A. Learning multiple layers of features from tiny images. *Technical Report*, 2009.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pp. 1106–1114, 2012.

Lee, K., Liu, L., Tang, Y., Zhang, Q., and Zhou, Y. Efficient and customizable data partitioning framework for distributed big rdf data processing in the cloud. In *Proceedings of the 2013 IEEE International Conference on Cloud Computing (CLOUD'13)*, pp. 327–334, Santa Clara, CA, June 27-July 2 2013.

Lee, K., Liu, L., Schwan, K., Pu, C., Zhang, Q., Zhou, Y., Yigitoglu, E., and Yuan, P. Scaling iterative graph computations with graphmap. In *Proceedings of the 27th IEEE international conference for High Performance Computing, Networking, Storage and Analysis (SC'15)*, pp. 57:1–57:12, Austin, TX, November 15-20 2015.

Li, C., Wang, S., Wang, Y., Yu, P. S., Liang, Y., Liu, Y., and Li, Z. Adversarial learning for weakly-supervised social network alignment. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 996–1003, 2019a.

Li, Q., Haque, S., Anil, C., Lucas, J., Grosse, R. B., and Jacobsen, J. Preventing gradient attenuation in lipschitz constrained convolutional networks. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pp. 15364–15376, 2019b.

Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J. Understanding of internal clustering validation measures. In *Proc. 2010 Int. Conf. on Data Mining (ICDM'10)*, pp. 911–916, Sydney, Australia, Dec. 2010.

Liu, Y., He, L., Cao, B., Yu, P. S., Ragin, A. B., and Leow, A. D. Multi-view multi-graph embedding for brain network clustering analysis. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 117–124, 2018.

Lloyd, S. P. Least squares quantization in pcm. *IEEE Trans. Information Theory*, 28:128–137, 1982.

Logins, A., Li, Y., and Karras, P. On the robustness of cascade diffusion under node attacks. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pp. 2711–2717, 2020.

Luo, D., Bian, Y., Yan, Y., Liu, X., Huan, J., and Zhang, X. Local community detection in multiple networks. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pp. 266–274, 2020.

Ma, G., He, L., Lu, C.-T., Shao, W., Yu, P. S., Leow, A. D., and Ragin, A. B. Multi-view clustering with graph embedding for connectome analysis. In *Proc. 2017 Int. Conf. Information and Knowledge Management (CIKM'17)*, pp. 127–136, Singapore, November 6-10 2017.

Ma, J., Ding, S., and Mei, Q. Towards more practical adversarial attacks on graph neural networks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, Online*, December 6-12 2020a.

Ma, N., Bu, J., Yang, J., Zhang, Z., Yao, C., Yu, Z., Zhou, S., and Yan, X. Adaptive-step graph meta-learner for few-shot graph classification. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pp. 1055–1064, 2020b.

Ma, T., Wang, H., Zhang, L., Tian, Y., and Al-Nabhan, N. Graph classification based on structural features of significant nodes and spatial convolutional neural networks. *Neurocomputing*, 423:639–650, 2021.

Ma, Z., Xuan, J., Wang, Y. G., Li, M., and Liò, P. Path integral based convolution and pooling for graph neural networks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020c.

Magelinski, T., Beskow, D. M., and Carley, K. M. Graph-hist: Graph classification from latent feature histograms with application to bot detection. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 5134–5141, 2020.

Man, T., Shen, H., Liu, S., Jin, X., and Cheng, X. Predict anchor links across social networks via an embedding approach. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pp. 1823–1829, 2016.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pp. 3111–3119, 2013.

Miller, B. A., Camurcu, M., Gomez, A. J., Chan, K., and Eliassi-Rad, T. Improving robustness to attacks against vertex classification. In *Proceedings of the 15th International Workshop on Mining and Learning with Graphs colocated with 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, MLG@KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, 2019.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

Oettershagen, L., Kriege, N. M., Morris, C., and Mutzel, P. Temporal graph kernels for classifying dissemination processes. In *Proceedings of the 2020 SIAM International Conference on Data Mining, SDM 2020, Cincinnati, Ohio, USA, May 7-9, 2020*, pp. 496–504, 2020.

Palanisamy, B., Liu, L., Lee, K., Meng, S., Tang, Y., and Zhou, Y. Anonymizing continuous queries with delay-tolerant mix-zones over road networks. *Distributed and Parallel Databases (DAPD)*, 32(1):91–118, 2014.

Palanisamy, B., Liu, L., Zhou, Y., and Wang, Q. Privacy-preserving publishing of multilevel utility-controlled graph datasets. *ACM Transactions on Internet Technology (TOIT)*, 18(2):24:1–24:21, 2018.

Peng, H., Li, J., Gong, Q., Ning, Y., Wang, S., and He, L. Motif-matching based subgraph-level attentional convolutional network for graph classification. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 5387–5394, 2020.

Perozzi, B., Al-Rfou, R., and Skiena, S. Deepwalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'14)*, pp. 701–710, New York, NY, August 24-27 2014.

Qin, K. K., Salim, F. D., Ren, Y., Shao, W., Heimann, M., and Koutra, D. G-CREWE: graph compression with embedding for network alignment. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pp. 1255–1264, 2020.

Qu, M., Tang, J., Shang, J., Ren, X., Zhang, M., and Han, J. An attention-based collaboration framework for multi-view network representation learning. In *Proc. 2017 Int. Conf. Information and Knowledge Management (CIKM'17)*, pp. 1767–1776, Singapore, November 6-10 2017.

Ren, F., Zhang, Z., Zhang, J., Su, S., Sun, L., Zhu, G., and Guo, C. BANANA: when behavior analysis meets social network alignment. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pp. 1438–1444, 2020.

Ren, J., Zhou, Y., Jin, R., Zhang, Z., Dou, D., and Wang, P. Dual adversarial learning based network alignment. In *Proceedings of the 19th IEEE International Conference on Data Mining (ICDM'19)*, pp. 1288–1293, Beijing, China, November 8-11 2019.

Rieck, B., Bock, C., and Borgwardt, K. M. A persistent weisfeiler-lehman procedure for graph classification. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 5448–5458, 2019.

Sen, P., Namata, G., Bilgic, M., Getoor, L., Gallagher, B., and Eliassi-Rad, T. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

Sharma, J. and Vasishtha, A. *Kirshna's Real Analysis: (General), Thirty Eighth Edition*. Krishna Prakashan Media, 2010.

Sokolic, J., Giryes, R., Sapiro, G., and Rodrigues, M. R. D. Robust large margin deep neural networks. *IEEE Trans. Signal Process.*, 65(16):4265–4280, 2017.

Su, Z., Liu, L., Li, M., Fan, X., and Zhou, Y. Servicetrust: Trust management in service provision networks. In *Proceedings of the 10th IEEE International Conference on Services Computing (SCC'13)*, pp. 272–279, Santa Clara, CA, June 27-July 2 2013.

Su, Z., Liu, L., Li, M., Fan, X., and Zhou, Y. Reliable and resilient trust management in distributed service provision networks. *ACM Transactions on the Web (TWEB)*, 9(3): 1–37, 2015.

Sun, Y., Wang, S., Hsieh, T., Tang, X., and Honavar, V. G. MEGAN: A generative adversarial network for multi-view network embedding. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pp. 3527–3533, 2019.

Tang, H., Ma, G., Chen, Y., Guo, L., Wang, W., Zeng, B., and Zhan, L. Adversarial attack on hierarchical graph pooling neural networks. *CoRR*, abs/2005.11560, 2020a.

Tang, X., Li, Y., Sun, Y., Yao, H., Mitra, P., and Wang, S. Transferring robustness for graph neural network against poisoning attacks. In *Proceedings of the 13th ACM International Conference on Web Search and Data Mining, WSDM 2020, Houston, TX, February 3-7, 2020*, 2020b.

Togninalli, M., Ghisu, M. E., Llinares-López, F., Rieck, B., and Borgwardt, K. M. Wasserstein weisfeiler-lehman graph kernels. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 6436–6446, 2019.

Tsuzuku, Y., Sato, I., and Sugiyama, M. Lipschitz-margin training: Scalable certification of perturbation invariance for deep neural networks. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada*, pp. 6542–6551, 2018.

Wang, D., Cui, P., and Zhu, W. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'16)*, pp. 1225–1234, San Francisco, CA, August 13-17 2016.

Wang, R., Yan, J., and Yang, X. Graduated assignment for joint multi-graph matching and clustering with application to unsupervised graph matching network learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020a.

Wang, T., Jiang, Z., and Yan, J. Multiple graph matching and clustering via decayed pairwise matching composition. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pp. 1660–1667, 2020b.

Weibull, W. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18:293–297, 1951.

Wu, H., Wang, C., Tyshetskiy, Y., Docherty, A., Lu, K., and Zhu, L. Adversarial examples for graph data: Deep insights into attack and defense. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pp. 4816–4823, 2019a.

Wu, J., He, J., and Xu, J. Demo-net: Degree-specific graph neural networks for node and graph classification. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pp. 406–415, 2019b.

Wu, S., Li, Y., Zhang, D., Zhou, Y., and Wu, Z. Diverse and informative dialogue generation with context-specific commonsense knowledge awareness. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, (ACL'20)*, pp. 5811–5820, Online, July 5-10 2020a.

Wu, S., Li, Y., Zhang, D., Zhou, Y., and Wu, Z. Topicka: Generating commonsense knowledge-aware dialogue responses towards the recommended topic fact. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, (IJCAI'20)*, pp. 3766–3772, Online, January 7-15 2021a.

Wu, S., Wang, M., Zhang, D., Zhou, Y., Li, Y., and Wu, Z. Knowledge-aware dialogue generation via hierarchical

infobox accessing and infobox-dialogue interaction graph network. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence, (IJCAI'21)*, Online, August 21-26 2021b.

Wu, T., Ren, H., Li, P., and Leskovec, J. Graph information bottleneck. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020 (NeurIPS'20)*, Online, December 6-12 2020b.

Xi, Z., Pang, R., Ji, S., and Wang, T. Graph backdoor. *CoRR*, abs/2006.11890, 2020.

Xu, H., Luo, D., Zha, H., and Carin, L. Gromov-wasserstein learning for graph matching and node embedding. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 6932–6941, 2019a.

Xu, K., Chen, H., Liu, S., Chen, P., Weng, T., Hong, M., and Lin, X. Topology attack and defense for graph neural networks: An optimization perspective. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pp. 3961–3967, 2019b.

Yan, J., Yang, S., and Hancock, E. R. Learning for graph matching and related combinatorial optimization problems. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pp. 4988–4996, 2020.

Yanardag, P. and Vishwanathan, S. V. N. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015*, pp. 1365–1374, 2015.

Yasar, A. and Çatalyürek, Ü. V. An iterative global structure-assisted labeled network aligner. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pp. 2614–2623, 2018.

Yoshida, Y. and Miyato, T. Spectral norm regularization for improving the generalizability of deep learning. *CoRR*, abs/1705.10941, 2017.

You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., and Shen, Y. Graph contrastive learning with augmentations. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020 (NeurIPS'20)*, Online, December 6-12 2020.

Yu, T., Wang, R., Yan, J., and Li, B. Learning deep graph matching with channel-independent embedding and hungarian attention. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020a.

Yu, T., Yan, J., and Li, B. Determinant regularization for gradient-efficient graph matching. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pp. 7121–7130, 2020b.

Yu, Y., Xu, G., Jiang, M., Zhu, H., Dai, D., and Yan, H. Joint transformation learning via the $l_{2,1}$-norm metric for robust graph matching. *IEEE Trans. Cybern.*, 51(2): 521–533, 2021.

Zhang, G., Zhou, Y., Wu, S., Zhang, Z., and Dou, D. Cross-lingual entity alignment with adversarial kernel embedding and adversarial knowledge translation. *CoRR*, abs/2104.07837, 2021.

Zhang, L. and Lu, H. A feature-importance-aware and robust aggregator for GCN. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pp. 1813–1822, 2020.

Zhang, S. and Tong, H. FINAL: fast attributed network alignment. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016, San Francisco, CA, USA, August 13-17, 2016*, pp. 1345–1354, 2016.

Zhang, S., Tong, H., Maciejewski, R., and Eliassi-Rad, T. Multilevel network alignment. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pp. 2344–2354, 2019.

Zhang, X. and Zitnik, M. Gnnguard: Defending graph neural networks against adversarial attacks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, Online*, December 6-12 2020.

Zhang, Z., Zhang, Z., Zhou, Y., Shen, Y., Jin, R., and Dou, D. Adversarial attacks on deep graph matching. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020 (NeurIPS'20)*, Virtual, December 6-12 2020.

Zhao, Q. and Wang, Y. Learning metrics for persistence-based summaries and applications for graph classification. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 9855–9866, 2019.

Zheng, C., Zong, B., Cheng, W., Song, D., Ni, J., Yu, W., Chen, H., and Wang, W. Robust graph representation learning via neural sparsification. In *Proceedings of*

the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2019, Online, 2020.

Zhou, F., Liu, L., Zhang, K., Trajcevski, G., Wu, J., and Zhong, T. Deeplink: A deep learning approach for user identity linkage. In *2018 IEEE Conference on Computer Communications, INFOCOM 2018, Honolulu, HI, USA, April 16-19, 2018*, pp. 1313–1321, 2018a.

Zhou, K. and Vorobeychik, Y. Robust collective classification against structural attacks. In *Proceedings of the Thirty-Sixth Conference on Uncertainty in Artificial Intelligence, UAI 2020, virtual online, August 3-6, 2020*, pp. 119, 2020.

Zhou, K., Michalak, T. P., and Vorobeychik, Y. Adversarial robustness of similarity-based link prediction. In *IEEE International Conference on Data Mining, ICDM 2019, Beijing, China, November 8-11, 2019*, 2019a.

Zhou, Y. *Innovative Mining, Processing, and Application of Big Graphs*. PhD thesis, Georgia Institute of Technology, Atlanta, GA, USA, 2017.

Zhou, Y. and Liu, L. Clustering analysis in large graphs with rich attributes. In Holmes, D. E. and Jain, L. C. (eds.), *Data Mining: Foundations and Intelligent Paradigms: Volume 1: Clustering, Association and Classification*. Springer, 2011.

Zhou, Y. and Liu, L. Social influence based clustering of heterogeneous information networks. In *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'13)*, pp. 338–346, Chicago, IL, August 11-14 2013.

Zhou, Y. and Liu, L. Activity-edge centric multi-label classification for mining heterogeneous information networks. In *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'14)*, pp. 1276–1285, New York, NY, August 24-27 2014.

Zhou, Y. and Liu, L. Social influence based clustering and optimization over heterogeneous information networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(1):1–53, 2015.

Zhou, Y. and Liu, L. Approximate deep network embedding for mining large-scale graphs. In *Proceedings of the 2019 IEEE International Conference on Cognitive Machine Intelligence (CogMI'19)*, pp. 53–60, Los Angeles, CA, December 12-14 2019.

Zhou, Y., Cheng, H., and Yu, J. X. Graph clustering based on structural/attribute similarities. *Proceedings of the VLDB Endowment (PVLDB)*, 2(1):718–729, 2009.

Zhou, Y., Cheng, H., and Yu, J. X. Clustering large attributed graphs: An efficient incremental approach. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM'10)*, pp. 689–698, Sydney, Australia, December 14-17 2010.

Zhou, Y., Liu, L., Perng, C.-S., Sailer, A., Silva-Lepe, I., and Su, Z. Ranking services by service network structure and service attributes. In *Proceedings of the 20th International Conference on Web Service (ICWS'13)*, pp. 26–33, Santa Clara, CA, June 27-July 2 2013.

Zhou, Y., Seshadri, S., Chiu, L., and Liu, L. Graphlens: Mining enterprise storage workloads using graph analytics. In *Proceedings of the 2014 IEEE International Congress on Big Data (BigData'14)*, pp. 1–8, Anchorage, AK, June 27-July 2 2014.

Zhou, Y., Liu, L., and Buttler, D. Integrating vertex-centric clustering with edge-centric clustering for meta path graph analysis. In *Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'15)*, pp. 1563–1572, Sydney, Australia, August 10-13 2015a.

Zhou, Y., Liu, L., Lee, K., Pu, C., and Zhang, Q. Fast iterative graph computation with resource aware graph parallel abstractions. In *Proceedings of the 24th ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC'15)*, pp. 179–190, Portland, OR, June 15-19 2015b.

Zhou, Y., Liu, L., Lee, K., and Zhang, Q. Graphtwist: Fast iterative graph computation with two-tier optimizations. *Proceedings of the VLDB Endowment (PVLDB)*, 8(11): 1262–1273, 2015c.

Zhou, Y., Liu, L., Pu, C., Bao, X., Lee, K., Palanisamy, B., Yigitoglu, E., and Zhang, Q. Clustering service networks with entity, attribute and link heterogeneity. In *Proceedings of the 22nd International Conference on Web Service (ICWS'15)*, pp. 257–264, New York, NY, June 27-July 2 2015d.

Zhou, Y., Liu, L., Seshadri, S., and Chiu, L. Analyzing enterprise storage workloads with graph modeling and clustering. *IEEE Journal on Selected Areas in Communications (JSAC)*, 34(3):551–574, 2016.

Zhou, Y., Amimeur, A., Jiang, C., Dou, D., Jin, R., and Wang, P. Density-aware local siamese autoencoder network embedding with autoencoder graph clustering. In *Proceedings of the 2018 IEEE International Conference on Big Data (BigData'18)*, pp. 1162–1167, Seattle, WA, December 10-13 2018b.

Zhou, Y., Wu, S., Jiang, C., Zhang, Z., Dou, D., Jin, R., and Wang, P. Density-adaptive local edge representation learning with generative adversarial network multi-label edge classification. In *Proceedings of the 18th IEEE International Conference on Data Mining (ICDM'18)*, pp. 1464–1469, Singapore, November 17-20 2018c.

Zhou, Y., Jiang, C., Zhang, Z., Dou, D., Jin, R., and Wang, P. Integrating local vertex/edge embedding via deep matrix fusion and siamese multi-label classification. In *Proceedings of the 2019 IEEE International Conference on Big Data (BigData'19)*, pp. 1018–1027, Los Angeles, CA, December 9-12 2019b.

Zhou, Y., Ling Liu, Qi Zhang, K. L., and Palanisamy, B. Enhancing collaborative filtering with multi-label classification. In *Proceedings of the 2019 International Conference on Computational Data and Social Networks (CSoNet'19)*, pp. 323–338, Ho Chi Minh City, Vietnam, November 18-20 2019c.

Zhou, Y., Ren, J., Wu, S., Dou, D., Jin, R., Zhang, Z., and Wang, P. Semi-supervised classification-based local vertex ranking via dual generative adversarial nets. In *Proceedings of the 2019 IEEE International Conference on Big Data (BigData'19)*, pp. 1267–1273, Los Angeles, CA, December 9-12 2019d.

Zhou, Y., Liu, L., Lee, K., Palanisamy, B., and Zhang, Q. Improving collaborative filtering with social influence over heterogeneous information networks. *ACM Transactions on Internet Technology (TOIT)*, 20(4):36:1–36:29, 2020a.

Zhou, Y., Ren, J., Dou, D., Jin, R., Zheng, J., and Lee, K. Robust meta network embedding against adversarial attacks. In *Proceedings of the 20th IEEE International Conference on Data Mining (ICDM'20)*, pp. 1448–1453, Sorrento, Italy, November 17-20 2020b.

Zhou, Y., Ren, J., Jin, R., Zhang, Z., Dou, D., and Yan, D. Unsupervised multiple network alignment with multinominal gan and variational inference. In *Proceedings of the 2020 IEEE International Conference on Big Data (BigData'20)*, pp. 868–877, Atlanta, GA, December 10-13 2020c.

Zhou, Y., Zhang, Z., Wu, S., Sheng, V., Han, X., Zhang, Z., and Jin, R. Robust network alignment via attack signal scaling and adversarial perturbation elimination. In *Proceedings of the 30th Web Conference (WWW'21)*, pp. 3884–3895, Virtual Event / Ljubljana, Slovenia, April 19-23 2021.

Zhu, D., Zhang, Z., Cui, P., and Zhu, W. Robust graph convolutional networks against adversarial attacks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pp. 1399–1407, 2019.

Zügner, D., Akbarnejad, A., and Günnemann, S. Adversarial attacks on neural networks for graph data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pp. 2847–2856, 2018.