Integrated Defense for Resilient Graph Matching

Jiaxiang Ren [1] **Zijie Zhang** [1] **Jiayin Jin** [1] **Xin Zhao** [1] **Sixing Wu** [2] **Yang Zhou** [1] **Yelong Shen** [3] **Tianshi Che** [1] **Ruoming Jin** [4] **Dejing Dou** [5 6]

Abstract

A recent study has shown that graph matching models are vulnerable to adversarial manipulation of their input which is intended to cause a mismatching. Nevertheless, there is still a lack of a comprehensive solution for further enhancing the robustness of graph matching against adversarial attacks. In this paper, we identify and study two types of unique topology attacks in graph matching: inter-graph dispersion and intra-graph assembly attacks. We propose an integrated defense model, IDRGM, for resilient graph matching with two novel defense techniques to defend against the above two attacks simultaneously. A detection technique of inscribed simplexes in the hyperspheres consisting of multiple matched nodes is proposed to tackle inter-graph dispersion attacks, in which the distances among the matched nodes in multiple graphs are maximized to form regular simplexes. A node separation method based on phase-type distribution and maximum likelihood estimation is developed to estimate the distribution of perturbed graphs and separate the nodes within the same graphs over a wide space, for defending intra-graph assembly attacks, such that the interference from the similar neighbors of the perturbed nodes is significantly reduced. We evaluate the robustness of our IDRGM model on real datasets against state-of-the-art algorithms.

1. Introduction

Graph matching (i.e., network alignment), which aims to identify the same entities (i.e., nodes) across multiple graphs, has been a heated topic in recent years (Chu et al., 2019; Xu et al., 2019a; Wang et al., 2020d; Chen et al., 2020a;b; Zhang & Tong, 2016; Mu et al., 2016; Heimann et al., 2018; Li et al., 2019a; Fey et al., 2020; Qin et al., 2020; Feng et al., 2019; Ren et al., 2020). It has been widely applied to many real-world applications, including protein network alignment in bioinformatics (Liu et al., 2017; Vijayan et al., 2020), user account linking in multiple social networks(Shu et al., 2016; Mu et al., 2016; Feng et al., 2019), object matching in computer vision (Fey et al., 2020; Wang et al., 2020b;e; Yang et al., 2020), knowledge translation in multilingual knowledge bases (Sun et al., 2020; Wu et al., 2020c).

Recently, there has been much interest in developing resilient graph learning techniques to improve the model robustness against adversarial attacks, including node classification (Zhu et al., 2019; Xu et al., 2019b; Tang et al., 2020; 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), graph classification (Jin et al., 2020a), community detection (Jia et al., 2020), network embedding (Dai et al., 2019), link prediction (Zhou et al., 2019a), malware detection (Hou et al., 2019), spammer detection (Dou et al., 2020), fraud detection (Breuer et al., 2020; Zhang et al., 2020a), and influence maximization (Logins et al., 2020). The majority of existing techniques focus on the defenses on single graph learning tasks. Improving the robustness of graph matching against adversarial attacks has not been inadequately investigated yet. Existing techniques for defending single graph learning tasks cannot be directly utilized to improve the robustness of graph matching, as the graph matching has to analyze interactions within and across graphs. To our best knowledge, RGM is the only robust graph matching model (Yu et al., 2021). It enhances the robustness of image matching against visual noise in computer vision, including image deformations, rotations, and outliers, but it fails to defend adversarial attacks on graph topology.

In the context of graph matching, there are two types of topology attacks within and across graphs: (1) **Inter-graph dispersion attacks.** Most of existing graph matching algorithms often aim to minimize the distance or maximize the similarity among the matched nodes in K different graphs in training data by mapping these nodes with different features into common space through either matrix transformation (Zhang & Tong, 2016; Zhang et al., 2019) or network

[1]Auburn University, USA [2]Peking University, China [3]Microsoft Dynamics 365 AI, USA [4]Kent State University, USA [5]University of Oregon, USA [6]Baidu Research, China. Correspondence to: Yang Zhou <yangzhou@auburn.edu>.

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

embedding (Heimann et al., 2018; Chu et al., 2019; Xu et al., 2019a; Fey et al., 2020). The nodes with the smallest distances in K graphs in test data are selected as the matching results. As shown in Figure 1, three matched nodes $\mathbf{v}^1_{i^1}$, $\mathbf{v}^2_{i^2}$, and $\mathbf{v}^3_{i^3}$ in three graphs G^1, G^2, and G^3 are projected into the same space, such that their embeddings $\mathbf{u}^1_{i^1}$, $\mathbf{u}^2_{i^2}$, and $\mathbf{u}^3_{i^3}$ are identical, i.e., $\mathbf{u}^1_{i^1} = \mathbf{u}^2_{i^2} = \mathbf{u}^3_{i^3}$. An inter-graph dispersion attack tries to push the matched nodes in multiple graphs far away from each other for maximizing their distances under an attack budget ϵ. In this case, the attack problem is equivalent to a geometry optimization problem of how to arrange the matched nodes in a hypersphere with radius ϵ such that the distances among them are maximized. Namely, an inscribed regular $(K-1)$-simplex in a hypersphere with radius ϵ is generated by adding/deleting edges to/from the matched nodes, e.g., an inscribed equilateral triangle (i.e., regular 2-simplex) in a circle (1-hypersphere) with radius ϵ in Figure 1. In addition, there is little possibility for the non-matched clean nodes in K graphs to form a regular or near-regular simplex, especially when K is large. Thus, regular or near-regular simplexes within the range of ϵ can be safely treated as the matched nodes under the inter-graph dispersion attacks; and (2) **Intra-graph assembly attacks.** A recent attack solution for graph matching aims to move a node to be attacked to dense region in its graph, such that the distances between its similar neighbors in the same graph and its counterparts in other graphs become smaller than the ones between this perturbed node and its counterparts, and thus to generate a wrong matching result (Zhang et al., 2020b). As shown in Figure 2, two matched nodes $\mathbf{u}^1_{i^1}$ and $\mathbf{u}^2_{i^2}$ are pushed to dense regions in two graphs G^1 and G^2 respectively, such that $\mathbf{u}^1_{i^1}$ is closer to the neighbors of $\mathbf{u}^2_{i^2}$ in G^2, rather than $\mathbf{u}^2_{i^2}$ itself. A wrong matching between $\mathbf{u}^1_{i^1}$ and an neighbor of $\mathbf{u}^2_{i^2}$ will be generated. In addition, since there are many similar neighbors around the perturbed nodes in the dense region, this dramatically increases the possibility of deriving the wrong matching results.

Motivated by the above analysis, we propose an effective simplex detection technique to tackle the inter-graph dispersion attacks. The defense model tries to determine whether the nodes in multiple graphs form inscribed regular simplexes in the hyperspheres with radius ϵ and how regular the simplexes are. The completely regular or near-regular simplexes with the radius $R_K \leq \epsilon$ of their circumscribed hyperspheres are identified as the matching results under the inter-graph dispersion attacks. As shown in Figure 1, the inscribed equilateral triangle consisting of $\mathbf{u}^1_{i^1}$, $\mathbf{u}^2_{i^2}$, and $\mathbf{u}^3_{i^3}$ and its circumscribed circle with radius $R_3 = \epsilon$ are detected as an inter-graph dispersion attack.

Although real clean graphs often follow power-law degree distribution (Kleinberg et al., 1999; Albert et al., 1999; Barabási & Albert, 1999; Aiello et al., 2001; Zügner et al., 2018), most of existing adversarial attack techniques on graph data focus on how to generate imperceptible perturbations within a l_p norm neighborhood but ignore the distribution change from clean graphs to perturbed ones (Bojchevski & Günnemann, 2019; Wang & Gong, 2019; Liu et al., 2019; Chang et al., 2020; Li et al., 2020; Zang et al., 2020). Thus, the perturbed graphs can follow any distributions. The phase-type distribution can be used to approximate any positive-valued distribution (O'Cinneide, 1990). By exploring the phase-type distribution and maximum likelihood estimation (Chakravarthy & Alfa, 1996; Asmussen et al., 1996), we develop a node separation algorithm to handle the intra-graph assembly attacks. We estimate the distribution of perturbed graphs and maximize the distances among the perturbed nodes within the same graphs, for separating the nodes in a narrow space into a wide space, such that the interference from the similar neighbors of the perturbed nodes is significantly reduced. In Figure 2, the nodes in two graphs G^1 and G^2 are separated respectively by maximizing the distances $1/d^1_y$ and $1/d^2_y$ in G^1 and G^2.

Empirical evaluation over real graph datasets demonstrates that the remarkable robustness of IDRGM against state-of-the-art graph matching methods and representative resilient Lipschitz-bound neural architectures. In addition, more experiments, implementation details, and hyperparameter selection and setting are presented in Appendices A.2-A.4.

To our best knowledge, this work is the first to study integrated defense for resilient graph matching against both inter-graph dispersion and intra-graph assembly attacks.

2. Problem Definition

Given a set of K graphs $G^1, \cdots, G^K$ to be matched, each graph is denoted as $G^k = (V^k, E^k)$ $(1 \leq k \leq K)$, where $V^k = \{v^k_1, \cdots, v^k_{N^k}\}$ is the set of N^k nodes and $E^k = \{(v^k_i, v^k_j) : 1 \leq i, j \leq N^k\}$ is the set of edges. Each G^k has an $N^k \times N^k$ binary adjacency matrix $\mathbf{A}^k$, where each entry $\mathbf{A}^k_{ij} = 1$ if there exists an edge $(v^k_i, v^k_j) \in E^k$; otherwise $\mathbf{A}^k_{ij} = 0$. $\mathbf{A}^k_{i:}$ specifies the i^{th} row vector of $\mathbf{A}^k$. In this paper, if there are no specific descriptions, we use $\mathbf{v}^k_i$ to denote a node v^k_i itself and its representation $\mathbf{A}^k_{i:}$, i.e., $\mathbf{v}^k_i = \mathbf{A}^k_{i:}$ and we utilize $\mathbf{v}^k_{ij}$ to specify the j^{th} dimension of $\mathbf{v}^k_i$, i.e., $\mathbf{v}^k_{ij} = \mathbf{A}^k_{ij}$.

The dataset is divided into two disjoint sets: training data D and test data D'. The former denotes a set of known matched nodes across K graphs $D = \{(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}) | \mathbf{v}^1_{i^1} \leftrightarrow \cdots \leftrightarrow \mathbf{v}^K_{i^K}, \mathbf{v}^1_{i^1} \in V^1, \cdots, \mathbf{v}^K_{i^K} \in V^K\}$, where $\mathbf{v}^1_{i^1} \leftrightarrow \cdots \leftrightarrow \mathbf{v}^K_{i^K}$ indicates that K nodes $\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}$ belong to the same entity. The latter, denoted by $D' = \{(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}) | \mathbf{v}^1_{i^1} \leftrightarrow \cdots \leftrightarrow \mathbf{v}^K_{i^K}, \mathbf{v}^1_{i^1} \in V^1, \cdots, \mathbf{v}^K_{i^K} \in V^K\}$, is used to evaluate the graph matching performance, where the nodes (but not their matchings)

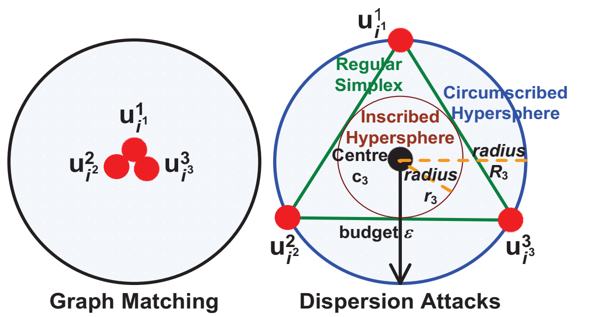


Figure 1: Defenses against Inter-graph Dispersion Attacks

are also observed during training. The goal of graph matching is to use D as the training data to identify the one-to-one matching relationships among nodes $\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}$ belonging to the same entities in the test data D'.

By following the same idea in existing efforts (Zhou et al., 2018a; Yasar & Çatalyürek, 2018; Li et al., 2019a), this paper aims to learn an embedding function M to map the nodes $(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}) \in D$ with different features across K graphs into common embedding space, i.e, minimize the distances among the projected nodes $M(\mathbf{v}^1_{i^1}), \cdots, M(\mathbf{v}^K_{i^K})$. The nodes $(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}) \in D'$ with the smallest distances in the embedding space are selected as the matching results.

$$
\begin{aligned}
&\mathcal{L}_{\mathcal{M}}(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}) = \sum_{k=1, l>k}^{K} 1 - \cos(\mathbf{u}^k_{i^k}, \mathbf{u}^l_{i^l}) \\
&\mathcal{L}_{\mathcal{E}} = \sum_{k=1}^{K} \Big[- \sum_{v^k_{i^k} \in V^k, v^k_{j^k} \in \mathcal{N}(v^k_i)} \max\{0, \cos(\mathbf{u}^k_{i^k}, \mathbf{u}^k_{j^k})\} \\
&\quad + \sum_{j^k=1}^{J} \mathbb{E}_{\mathbf{v}^k_{j^k} \sim p(\mathbf{v}^k_{j^k})} \max\{0, \cos(\mathbf{u}^k_{i^k}, \mathbf{u}^k_{j^k})\} \Big] \\
&\min_M \mathcal{L} = \mathcal{L}_{\mathcal{E}} + \mathbb{E}_{(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}) \in D} \mathcal{L}_{\mathcal{M}}(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K})
\end{aligned}
\tag{1}
$$

where $\mathbf{u}^k_{i^k} = M(\mathbf{v}^k_{i^k})$ denotes an embedding function to map the original representation $\mathbf{v}^k_{i^k}$ of each node $v^k_{i^k}$ in each graph G^k to a low-dimensional representation $\mathbf{u}^k_{i^k}$, i.e., $\mathbf{v}^k_{i^k} : \mathbb{R}^{N^k} \mapsto \mathbf{u}^k_{i^k} : \mathbb{R}^{K-1}$ and $K - 1 << N^k$ $(1 \leq k \leq K)$. cos is the cosine similarity between pairwise node embedding vectors. $\mathcal{N}(v^k_i)$ is the set of neighbors of node v^k_i in graph G^k. $p(\mathbf{v}^k_{j^k})$ denotes the distribution for sampling J negative nodes $v^k_{j^k} \neq v^k_{i^k}$ through the negative sampling method. $\mathcal{L}_{\mathcal{M}}(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K})$ denotes the matching loss among nodes $\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}$ in K graphs, while $\mathcal{L}_{\mathcal{E}}$ is the embedding loss that maximizes/minimizes the similarity between neighbored/disconnected nodes within the same graphs G^k.

With the injected adversarial attacks (including edge insertions and deletions) on K clean graphs $G^1, \cdots, G^K$, leading to perturbed graphs $\hat{G}^1, \cdots, \hat{G}^K$, an adversarial defender is trained to detect or eliminate the perturbations for maintaining the high utility of the matching results by M on $\hat{G}^1, \cdots, \hat{G}^K$.

3. Defenses against Inter-graph Dispersion Attacks

In this section, we propose an effective simplex detection technique to tackle the inter-graph dispersion attacks. In geometry, a simplex is a generalization of the notion of a triangle or tetrahedron to arbitrary dimensions (Elte, 1912). A regular simplex is a simplex that is also a regular polytope. Given K points $\mathbf{u}^1_{i^1}, \cdots, \mathbf{u}^K_{i^K} \in \mathbb{R}^{K-1}$, let $\mathcal{H}_k$ be the hyperplane generated by the points $(\mathbf{u}^l_{i^l})_{l \neq k}$, we can always discover a vector $\mathbf{x}_k \in \mathbb{R}^{K-1}$ and a scalar $z_k \in \mathbb{R}$ such that

$$\mathcal{H}_k = \{\mathbf{y} \in \mathbb{R}^{K-1} : \mathbf{x}_k \cdot \mathbf{y} = z_k\}, \forall k, 1 \leq k \leq K \tag{2}$$

where $\cdot$ represents the inner product between two vectors.

Definition 1 *A $(K-1)$-simplex $\mathcal{S}_{K-1}$, generated by the points $\mathbf{u}^1_{i^1}, \cdots, \mathbf{u}^K_{i^K}$, is defined as the convex hull of these points.*

$$\mathcal{S}_{K-1} = \left\{ \mathbf{s} \middle| \mathbf{s} = \sum_{k=1}^{K} \omega_k \mathbf{u}^k_{i^k}, 0 \leq \omega_k \leq 1, \sum_{k=1}^{K} \omega_k = 1 \right\} \tag{3}$$

The radius r_K of the inscribed hypersphere in $\mathcal{S}_{K-1}$ is given as follows.

$$\frac{1}{r_K} = \sum_{k=1}^{K} \frac{1}{D_k} \tag{4}$$

where $D_k = dist(\mathbf{u}^k_{i^k}, \mathcal{H}_k) = \min_{\mathbf{h} \in \mathcal{H}_k} \|\mathbf{u}^k_{i^k} - \mathbf{h}\|$ represents the distance between $\mathbf{u}^k_{i^k}$ and $\mathcal{H}_k$.

The centre $\mathbf{c}_K$ of the inscribed hypersphere in $\mathcal{S}_{K-1}$ is given below.

$$\mathbf{c}_K = r_K \sum_{k=1}^{K} \frac{1}{D_k} \mathbf{u}^k_{i^k} \tag{5}$$

Definition 2 *The centre of gravity $\mathbf{g}_k$ of the k^{th} face of a $(K-1)$-simplex $\mathcal{S}_{K-1}$ is defined as follows.*

$$\mathbf{g}_k = \frac{1}{K-1} \sum_{l=1, l \neq k}^{K} \mathbf{u}^l_{i^l}, 1 \leq k \leq K \tag{6}$$

The k^{th} median of a $(K-1)$-simplex $\mathcal{S}_{K-1}$ is the line segment $[\mathbf{u}^k_{i^k}, \mathbf{g}_k]$.

The centre of gravity $\mathbf{g}_K$ of the $(K-1)$-simplex $\mathcal{S}_{K-1}$ is defined as follows.

$$\mathbf{g}_K = \frac{1}{K} \sum_{k=1}^{K} \mathbf{u}^k_{i^k} \tag{7}$$

Theorems 1-4 demonstrates that for any two different points $\mathbf{u}^k_{i^k}$ and $\mathbf{u}^j_{i^j}$ in a regular simplex $\mathcal{S}_{K-1}$ with centre $\mathbf{c}_K = \mathbf{0}$, the 2-norm of their distance and their inner product depend on only K and the radius r_K but are irrelevant to the coordinates of two points. Therefore, given K and r_K, the inner product between the coordinate vectors of any two points in a standard regular simplex $\mathcal{S}_{K-1}$ with $\mathbf{c}_K = \mathbf{0}$ is a constant. We will utilize this important property to determine whether the nodes in multiple graphs form regular simplexes and how regular the simplexes are.

Theorem 1 *The medians of a $(K-1)$-simplex $\mathcal{S}_{K-1}$ meet at the same point $\mathbf{g}_K$ and they divide each other in the ratio $(K-1):1$.*

In a standard regular simplex $\mathcal{S}_{K-1}$, the centre $\mathbf{c}_K$ of the inscribed hypersphere is equal to $\mathbf{0}$, i.e., $\mathbf{c}_K = \mathbf{0}$, where $\mathbf{0} \in \mathbb{R}^{K-1}$ is an all-zero vector denoting the origin. Based on Eq.(5), we have

$$\sum_{k=1}^{K} \mathbf{u}_{i^k}^k = 0 \tag{8}$$

Without loss of generality, in $\mathcal{S}_{K-1}$ with $\mathbf{c}_K = \mathbf{0}$, there must exist a point with the following coordinate, say $\mathbf{u}_{i^K}^K$.

$$\mathbf{u}_{i^K}^K = [0, \cdots, 0, \mathbf{u}_{i^K(K-1)}^K] \tag{9}$$

where $\mathbf{u}_{i^K(K-1)}^K$ is to be determined.

By symmetry the hyperplane $\mathcal{H}_K$ consisting of $\mathbf{u}_{i^1}^1, \cdots, \mathbf{u}_{i^{K-1}}^{K-1}$ has the Cartesian equation $\mathbf{u}_{i^l(K-1)}^l = -r_K$ for $\forall l, l \neq K$. Namely, the last component of all points $(\mathbf{u}_{i^l}^l)_{l \neq K}$ is $-r_K$.

As the inscribed hypersphere has the radius r_K, based on Theorem 1, then $D_K = dist(\mathbf{u}_{i^K}^K, \mathcal{H}_K) = K r_K$ and $\mathbf{u}_{i^K(K-1)}^K = (K-1) r_K$.

$$\mathbf{u}_{i^K}^K = [0, \cdots, 0, (K-1)r_K] \tag{10}$$

Theorem 2 *By sequentially projecting $\mathcal{S}_{K-1}$, we can generate a series of regular simplexes: $\mathcal{S}_{K-2}$ consisting of $\mathbf{u}_{i^1}^1, \cdots, \mathbf{u}_{i^{K-1}}^{K-1}$ with centre $\mathbf{c}_{K-1} = \mathbf{0}$, $\cdots$, $\mathcal{S}_1$ consisting of $\mathbf{u}_{i^1}^1$ and $\mathbf{u}_{i^2}^2$ with centre $\mathbf{c}_2 = \mathbf{0}$, and $\mathcal{S}_0$ consisting of $\mathbf{u}_{i^1}^1$ with centre $\mathbf{c}_1 = \mathbf{0}$. For radius r_k of $\mathcal{S}_{k-1}$ for any k $(2 \leq k \leq K)$, we have*

$$r_k = \sqrt{\frac{K(K-1)}{k(k-1)}} r_K, 2 \leq k \leq K \tag{11}$$

All points in a standard regular simplex $\mathcal{S}_{K-1}$ with $\mathbf{c}_K = \mathbf{0}$ have the following coordinates.

$$\begin{aligned} \mathbf{u}_{i^K}^K &= [0, \cdots, 0, (K-1)r_K] \\ \mathbf{u}_{i^{K-1}}^{K-1} &= [0, \cdots, 0, (K-2)r_{K-1}, -r_K] \\ \mathbf{u}_{i^{K-2}}^{K-2} &= [0, \cdots, 0, (K-3)r_{K-2}, -r_{K-1}, -r_K] \\ \cdots &= \cdots \end{aligned} \tag{12}$$

Theorem 3 *For any two different points $\mathbf{u}_{i^k}^k$ and $\mathbf{u}_{i^j}^j$ $(1 \leq k, j \leq K, k \neq j)$ in a standard regular simplex $\mathcal{S}_{K-1}$ with $\mathbf{c}_K = \mathbf{0}$, $\|\mathbf{u}_{i^k}^k - \mathbf{u}_{i^j}^j\|_2^2 = 2K(K-1)r_K^2$.*

Theorem 4 *In a standard regular simplex $\mathcal{S}_{K-1}$ with centre $\mathbf{c}_K = \mathbf{0}$, $\|\mathbf{u}_{i^k}^k\|_2^2 = (K-1)^2 r_K^2$ for any k $(1 \leq k \leq K)$. For any two different points $\mathbf{u}_{i^k}^k$ and $\mathbf{u}_{i^j}^j$ $(1 \leq k, j \leq K, k \neq j)$, $\mathbf{u}_{i^k}^k \cdot \mathbf{u}_{i^j}^j = -(K-1)r_K^2$.*

Proof. *Please refer to Appendix A.1 for detailed proof of Theorems 1-4.*

For a regular simplex $\mathcal{S}_{K-1}$, its centre $\mathbf{c}_K$ coincides with the centre of gravity $\mathbf{g}_K$. In addition, $\mathbf{g}_K$ coincides with the centre of the inscribed hypersphere and the circumscribed hypersphere of $\mathcal{S}_{K-1}$. In the context of graph matching, the centre $\mathbf{c}_K$ of a possible regular simplex that consists of K perturbed nodes with the inter-graph dispersion attacks may not be at the origin $\mathbf{0}$ in the embedding space. We calculate $\mathbf{c}_K = \mathbf{g}_K = \frac{1}{K}\sum_{k=1}^{K} \mathbf{u}_{i^k}^k$ and move the simplex by converting the nodes with $\mathbf{w}_{i^k}^k = \mathbf{u}_{i^k}^k - \mathbf{c}_K$ for all k $(1 \leq k \leq K)$, such that the centre is at the origin. Thus, the radius R_K of the circumscribed hypersphere of the simplex is estimated as $R_K = \frac{1}{K}\sum_{k=1}^{K} \|\mathbf{w}_{i^k}^k\|_2$. According to Theorem 1, the radius r_K of the inscribed hypersphere of the simplex is estimated as $r_K = \frac{1}{K-1} R_K$. The attack budget ϵ is estimated with the average $\bar{R}_K$ of the radius R_K of the circumscribed hypersphere of all simplexes, generated by the matched nodes $(\mathbf{u}_{i^1}^1, \cdots, \mathbf{u}_{i^K}^K)$ in the training data D, i.e., $\epsilon = \bar{R}_K$.

In order to determine whether the nodes $\mathbf{w}_{i^1}^1, \cdots, \mathbf{w}_{i^K}^K$ in K graphs form a regular $(K-1)$-simplex, we need to decide whether $\mathbf{w}_{i^k}^k \cdot \mathbf{w}_{i^j}^j = -(K-1)r_K^2$ for all $K(K-1)/2$ pairs of nodes $\mathbf{w}_{i^k}^k$ and $\mathbf{w}_{i^j}^j$ $(1 \leq k < j \leq K)$. However, this operation is non-trivial and practically infeasible. We randomly sample T $(T << K(K-1)/2)$ pairs from $\mathbf{w}_{i^1}^1, \cdots, \mathbf{w}_{i^K}^K$, denoted by $S = \{\mathsf{w}_1^1, \mathsf{w}_2^1\}, \cdots, \{\mathsf{w}_1^T, \mathsf{w}_2^T\}\}$. A function τ is used to define how regular a simplex is.

$$\tau(S) = \frac{1}{T}\sum_{t=1}^{T} g(\mathsf{w}_1^t \cdot \mathsf{w}_2^t + (K-1)r_K^2) \tag{13}$$

where g is the gaussian function with mean $\mu = 0$ and variance $\sigma^2 = 1/(2\pi)$, such that $\tau(S)$ lies between 0 and 1. 1 denotes the simplex is completely regular when $\mathsf{w}_1^t \cdot \mathsf{w}_2^t = -(K-1)r_K^2$ for any two w_1^t and w_2^t. 0 specifies it is least regular if the difference $\mathsf{w}_1^t \cdot \mathsf{w}_2^t$ and $-(K-1)r_K^2$ is large.

By integrating simplex detection for tackling inter-graph dispersion attacks, the overall loss is updated as follows.

$$\begin{aligned} \min_M \mathcal{L} = &\mathcal{L}_{\mathcal{E}} + \mathbb{E}_{(\mathbf{v}_{i^1}^1, \cdots, \mathbf{v}_{i^K}^K) \in D}\Big[\mathcal{L}_{\mathcal{M}}(\mathbf{v}_{i^1}^1, \cdots, \mathbf{v}_{i^K}^K) \times \\ &\big(1 - \tau(S)h(\epsilon + 4 - R_K)\big)\Big] \end{aligned} \tag{14}$$

where h is the sigmoid function. Notice that $h(4) = 0.982 \cdots \approx 1$. Thus, $h(\epsilon + 4 - R_K) \approx 1$ when $R_K \leq \epsilon$, i.e., actual attacks on the matched nodes in all K graphs are observed within the attack budget ϵ. On the other hand, when $R_K > \epsilon$, $h(\epsilon + 4 - R_K) < 1$ and approaches 0. We treat this case as natural outliers or exceptions among the non-matched nodes. Thus, $\tau(S)f(\epsilon + 4 - R_K)$ can be treated as the detection probability of inter-graph dispersion attacks on the matched nodes. It is equal to 1 when the simplex is completely regular and R_K is within ϵ. $\tau(S)h(\epsilon + 4 - R_K)$ keeps decreasing when the simplex becomes less regular and R_K keeps increasing above ϵ. Thus, $\mathcal{L}_{\mathcal{M}}(\mathbf{v}_{i^1}^1, \cdots, \mathbf{v}_{i^K}^K) \times \big(1 - \tau(S)f(\epsilon + 4 - R_K)\big)$ is treated as a matching predictor and a distance function among the matched nodes in both clean and attacked cases.

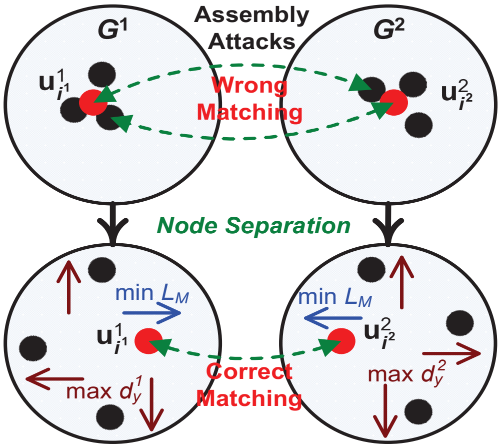


Figure 2: Defense against Intra-graph Assembly Attacks

It is equal to 0 for the attacked case when the simplex is completely regular and $R_K \leq \epsilon$. It is approximately equal to $\mathcal{L}_\mathcal{M}(\mathbf{v}^1_{i^1}, \cdots, \mathbf{v}^K_{i^K}) = 0$ for the clean case, since $\mathrm{w}^t_1 = \mathrm{w}^t_2$ and $r_K = 0$ and thus $\mathrm{w}^t_1 \cdot \mathrm{w}^t_2 + (K-1)r^2_K >> 0$, e.g., $g(x) \leq 0.043\cdots$ when $|x| \geq 1$.

The above discussion is about defenses against inter-graph dispersion attacks on all K graphs. However, the attacker may perturb only some of K graphs. A heuristic strategy is to exclude nodes $\mathbf{u}^k_{i^k}$ if the inner products between they and many other nodes deviate too many from $-(K-1)r^2_K$. We treat $\mathbf{u}^k_{i^k}$ as unattacked nodes and reuse the simplex detection technique to validate the attacks on the rest nodes.

A defender has no idea about which part of the graph is modified or not. A simple margin-based loss will dramatically change the structure of entire graph in the embedding space, especially modify the structure associated with clean nodes. This will result in the matching performance downgrade of clean nodes. Thus, the above simplex detection technique is proposed to detect perturbed nodes and differentiate them from clean nodes. Different defense strategies are adopted for these two types of nodes, which is reflected in Eq.(14).

4. Defense against Intra-graph Assembly Attacks

Theorem 5 demonstrates that the phase-type distribution can be used to approximate any positive-valued distribution. We will utilize the phase-type distribution and maximum likelihood estimation method (Chakravarthy & Alfa, 1996; Asmussen et al., 1996) to estimate the distribution of perturbed graphs and maximize the distances among the perturbed nodes within the same graphs to defense against intra-graph assembly attacks, for separating the nodes in a narrow space into a wide space, such that the interference from the similar neighbors of the perturbed nodes is significantly reduced.

Theorem 5 *The set of phase-type distributions is dense in the field of all positive-valued distributions, namely, it can be used to approximate any positive-valued distribution (O'Cinneide, 1990).*

Definition 3 *Consider a continuous-time Markov process with $m+1$ ($m \geq 1$) states, such that states $1, \cdots, m$ are transient states and state 0 is an absorbing state, a non-negative random variable u has a phase-type distribution if its distribution function is given as follows.*

$$F(x) = P(u \leq x) = 1 - \alpha e^{\mathbf{T}x}\mathbf{1} \equiv 1 - \alpha(\sum_{n=0}^{\infty} \frac{x^n}{n!}\mathbf{T}^n)\mathbf{1}, x \geq 0, \tag{15}$$

where $F(x)$ is the distribution function of u. $\mathbf{1} \in \mathbb{R}^m$ is an all-one column vector. $\alpha \in \mathbb{R}^m$ is a sub-stochastic vector of order m, i.e., α is a row vector with non-negative elements and $\alpha\mathbf{1} \leq 1$. $\mathbf{T}$ is a sub-generator of order m, i.e., $\mathbf{T}$ is an $m \times m$ matrix such that (1) all diagonal elements are negative; (2) all off-diagonal elements are non-negative; (3) all row sums are non-positive; and (4) $\mathbf{T}$ is invertible.

As the embedded nodes $\mathbf{u}^k_1, \cdots, \mathbf{u}^k_{N^k} \in \mathbb{R}^{K-1}$ in each graph G^k ($1 \leq k \leq K-1$) lie in $(K-1)$-dimensional space, we will utilize the multivariate phase-type distribution to estimate their distribution. Without loss of generality, let a $(K-1)$-dimensional random variable $\mathbf{u}$ denote all embedded nodes in G^k.

Definition 4 *For a $(K-1)$-dimensional random variable $\mathbf{u} = [\mathbf{u}_1, \cdots, \mathbf{u}_{K-1}]$ and $0 \leq x_1 \leq \cdots \leq x_{K-1}$, a multivariate phase-type distribution is defined as follows.*

$$\begin{aligned}\bar{F}(x_1, \cdots, x_{K-1}) &= P(\mathbf{u}_1 > x_1, \cdots, \mathbf{u}_{K-1} > x_{K-1}) \\ &= \alpha e^{\mathbf{T}x_1}\mathbf{D}_1 e^{\mathbf{T}(x_2-x_1)}\mathbf{D}_2 \cdots e^{\mathbf{T}(x_{K-1}-x_{K-2})}\mathbf{D}_{K-1}\mathbf{1}\end{aligned} \tag{16}$$

where $\bar{F}(x_1, \cdots, x_{K-1})$ is the survival function of $\mathbf{u}$. $\mathbf{D}_k$ ($1 \leq k \leq K-1$) is a diagonal matrix with the diagonal elements of 0 or 1. The absolutely continuous component of the joint distribution F has the following density.

$$\begin{aligned}f(x_1, \cdots, x_{K-1}) = (-1)^{K-1}\alpha e^{\mathbf{T}x_1}(\mathbf{T}\mathbf{D}_1 - \mathbf{D}_1\mathbf{T}) \\ e^{\mathbf{T}(x_2-x_1)}(\mathbf{T}\mathbf{D}_2 - \mathbf{D}_2\mathbf{T}) \cdots e^{\mathbf{T}(x_{K-1}-x_{K-2})}\mathbf{T}\mathbf{D}_{K-1}\mathbf{1}\end{aligned} \tag{17}$$

Assuming that $\mathbf{u}$ has the same boundary on all $K-1$ dimensions, i.e., $0 \leq x_1 = \cdots = x_{K-1} = x$, we have

$$\bar{F}(x_1, \cdots, x_{K-1}) = \alpha e^{\mathbf{T}x}\mathbf{D}\mathbf{1} \tag{18}$$

where $\mathbf{D} = \prod_{k=1}^{K-1} \mathbf{D}_k$ is still a diagonal matrix with the diagonal elements of 0 or 1. Now, we utilize maximum likelihood estimation (MLE) (Chakravarthy & Alfa, 1996; Asmussen et al., 1996) to estimate parameters α, $\mathbf{T}$, and $\mathbf{D}$.

$$L(\alpha, \mathbf{T}, \mathbf{D}|x) = P(x)\log Q(x) + (1-P(x))\log(1-Q(x)) \tag{19}$$

where $P(x)$ denotes the distribution of actual data and $Q(x) = 1 - \bar{F}(x_1, \cdots, x_{K-1})$ specifies the estimated phase-type distribution. The partial derivatives w.r.t. the parameters are computed below.

Algorithm 1 Expressive Parameter Estimation

Input: graph $G^k = (V^k, E^k)$, node embeddings $\mathbf{u}_1^k, \cdots, \mathbf{u}_{N^k}^k$, boundary parameter X, initial parameters α_0, $\mathbf{T}_0$, $\mathbf{D}_0$, and $\mathcal{D}_0$, and number of iterations I.
Output: Estimated distribution $Q_{\mathcal{D}}(x)$.

```
Initialize α, T, D, and 𝒟 with α_0, T_0, D_0, and 𝒟_0;
Normalize u_1^k, ··· , u_{N^k}^k into a bounded range [0, X].
for i = 1 to I
    x = i/I * X;
    Compute P(x) = #(u ≤ [x, ··· , x]) / N^k  for ∀u ∈ {u_1^k, ··· , u_{N^k}^k};
    Calculate Q_𝒟(x) with α, T, D, and 𝒟;
    Utilize MLE to optimize L(α, T, D, 𝒟|x);
    Update α, T, D, and 𝒟;
Return Q_𝒟(x).
```

$$
\begin{aligned}
\frac{\partial L}{\partial \alpha} &= \frac{P(x)e^{\mathbf{T}x}\mathbf{D}\mathbf{1}}{\alpha e^{\mathbf{T}x}\mathbf{D}\mathbf{1}-1} - \frac{P(x)-1}{\alpha} = 0 \\
\frac{\partial L}{\partial \mathbf{T}} &= \frac{P(x)\alpha x e^{\mathbf{T}x}\mathbf{D}\mathbf{1}}{\alpha e^{\mathbf{T}x}\mathbf{D}\mathbf{1}-1} - x(P(x)-1) = 0 \\
\frac{\partial L}{\partial \mathbf{D}} &= \frac{-P(x)\alpha e^{\mathbf{T}x}\mathbf{1}}{1-\alpha e^{\mathbf{T}x}\mathbf{D}\mathbf{1}} - \frac{(P(x)-1)\alpha e^{\mathbf{T}x}\mathbf{1}}{\alpha e^{\mathbf{T}x}\mathbf{D}\mathbf{1}} = 0
\end{aligned} \tag{20}
$$

We solve the above equations and get

$$
\begin{aligned}
\alpha &= \mathbf{1}^{-1}\mathbf{D}^{-1}e^{-\mathbf{T}x}(1-P(x)) \\
\mathbf{T} &= \frac{\log(\alpha^{-1}(1-P(x))\mathbf{1}^{-1}\mathbf{D}^{-1})}{x} \\
\mathbf{D} &= e^{-\mathbf{T}x}\alpha^{-1}(1-P(x))\mathbf{1}^{-1}
\end{aligned} \tag{21}
$$

where matrix inverse operator is used to represent vectors α^{-1} and $\mathbf{1}^{-1}$ such that $\mathbf{1}^{-1} \times \mathbf{1} = 1$ and $\alpha \times \alpha^{-1} = 1$, although the vectors do not have the inverse.

Fitting phase-type distributions often face the dilemma of unexpressive estimation, due to the restrict of binary diagonal elements of 0 or 1 in $\mathbf{D}_k$ (Chakravarthy & Alfa, 1996; Asmussen et al., 1996). We propose an expressive parameter estimation method for multivariate phase-type distribution by introducing one additional parameter $\mathcal{D}$.

Theorem 6 *The estimation with $\bar{F}_{\mathcal{D}}(x_1, \cdots, x_{K-1}) = \alpha e^{\mathbf{T}x}\mathbf{D}\mathcal{D}\mathbf{1}$ is more expressive than the one with $\bar{F}(x_1, \cdots, x_{K-1}) = \alpha e^{\mathbf{T}x}\mathbf{D}\mathbf{1}$, where $\mathcal{D} = \operatorname{diag}(h(d_1), \cdots, h(d_m))$ is a diagonal matrix to be estimated and h is the sigmoid function.*

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

Based on newly introduced expressive factor $\mathcal{D}$, we have corresponding survival function $\bar{F}_{\mathcal{D}}(x_1, \cdots, x_{K-1})$, distribution function $F_{\mathcal{D}}(x_1, \cdots, x_{K-1}) = 1 - \bar{F}_{\mathcal{D}}(x_1, \cdots, x_{K-1})$ (i.e., $Q_{\mathcal{D}}(x)$), and likelihood function $L(\alpha, \mathbf{T}, \mathbf{D}, \mathcal{D}|x)$. The expressive parameter estimation of multivariate phase-type distribution is presented in Algorithm 1.

In terms of the estimated distribution of the perturbed node embeddings in each graph G^k $(1 \leq k \leq K-1)$, we utilize the random-restart hill-climbing method (Russell & Norvig, 1995) to find Y nodes $\mathsf{u}_1^k, \cdots, \mathsf{u}_Y^k$ with local maximal densities. For each u_y^k $(1 \leq y \leq Y)$, we sample Z nearest neighbors in terms of the embedding features to form a group and calculate the average distance d_y^k between pairwise node embeddings within the group.

By combining node separation for handling intra-graph assembly attacks, the overall loss function is updated below.

$$
\begin{aligned}
\min_M \mathcal{L} = &\mathcal{L}_{\mathcal{E}} + \mathbb{E}_{(\mathbf{v}_{i^1}^1, \cdots, \mathbf{v}_{i^K}^K) \in D}\Big[\mathcal{L}_{\mathcal{M}}(\mathbf{v}_{i^1}^1, \cdots, \mathbf{v}_{i^K}^K) \times \\
&\big(1 - \tau(S)h(\epsilon + 4 - R_K)\big)\Big] + \sum_{k=1}^{K-1}\sum_{y=1}^{Y}\frac{1}{d_y^k}
\end{aligned} \tag{22}
$$

The combination of minimizing $\mathcal{L}_{\mathcal{M}}$ and $1/d_y^k$ by training M offers a defense solution against intra-graph assembly attacks. On one hand, minimizing $\mathcal{L}_{\mathcal{M}}$ can pull the matched nodes across graphs close to each other. On the other hand, minimizing $1/d_y^k$ is like a bombing operation at the densest locations and push the nodes within graphs far away from each other, such that the interference from the similar neighbors of the perturbed node is significantly reduced.

5. Experimental Evaluation

We will show the robustness of our IDRGM model for resilient graph matching over three datasets: autonomous systems (AS) (AS), CAIDA relationships datasets (CAI), and DBLP coauthor graphs (DBL), as shown in Table 1.

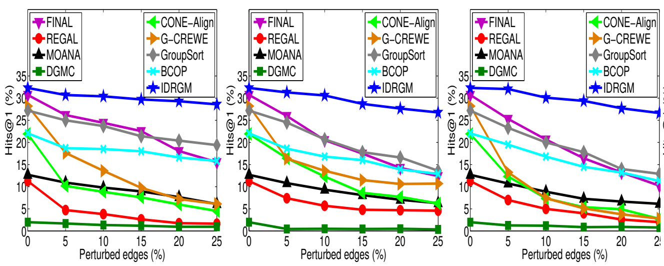
Graph matching baselines. We compare the IDRGM model with six state-of-the-art graph matching algorithms and two representative representative resilient Lipschitz-bound neural architectures. **FINAL** (Zhang & Tong, 2016) leverages both node and edge attributes to solve the attributed network alignment 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., 2020b) 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. **Group-Sort** (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 orthogonal convolution operations. To our best knowledge, there are no other

Table 1: Statistics of the Datasets

Dataset	AS			CAIDA		
Graph	G^1	G^2	G^3	G^1	G^2	G^3
#Nodes	10,900	11,113	11,019	16,655	16,493	16,301
#Edges	31,180	31,434	31,761	33,340	33,372	32,955
#MatchedNodes	7,943			7,884		
Dataset	**DBLP**					
Graph	2013	2014	2015			
#Nodes	28,478	26,455	27,543			
#Edges	128,073	114,588	133,414			
#MatchedNodes	4,000					

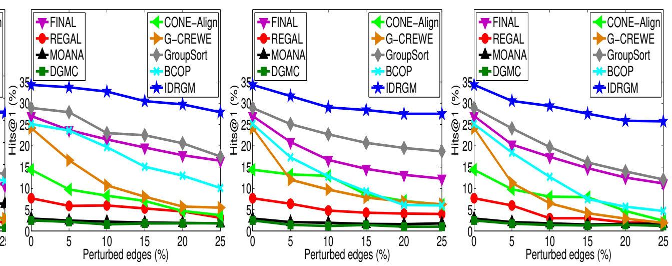
Table 2: $Hits@1$ (%) with 5% perturbed edges

Dataset	AS			CAIDA			DBLP		
Attack Model	RND	NEA	GMA	RND	NEA	GMA	RND	NEA	GMA
FINAL	25.2	19.7	21.3	23.7	20.9	20.3	12.4	9.5	9.2
REGAL	4.7	7.4	7.0	5.9	6.4	6.0	9.6	4.5	5.8
MOANA	2.8	2.5	2.6	2.5	2.1	2.1	3.8	3.1	3.1
DGMC	1.7	0.5	1.3	2.1	1.5	1.7	0.9	0.4	0.9
CONE-Align	10.2	9.4	12.3	7.8	8.3	6.8	3.2	5.3	4.6
G-CREWE	17.6	16.3	13.3	16.6	12.1	11.3	18.7	8.1	10.2
GroupSort	25.0	24.5	23.3	27.9	25.1	24.1	21.7	18.0	20.8
BCOP	18.7	18.6	19.5	23.6	17.3	18.4	15.6	14.7	15.9
IDRGM	**30.7**	**31.3**	**32.1**	**33.8**	**31.7**	**30.5**	**24.3**	**22.7**	**23.4**



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

Figure 3: AS with varying perturbed edges



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

Figure 4: CAIDA with varying perturbed edges

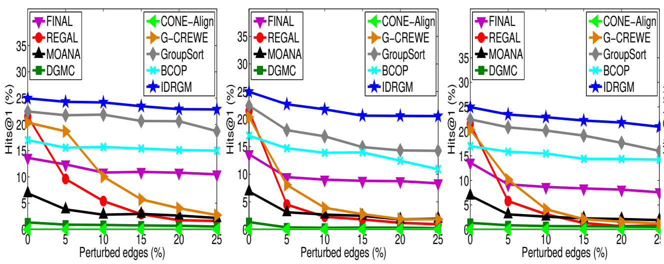
open-source defense baselines on graph matching available. This work is the first to study integrated defense for robust graph matching against adversarial attacks.

Attack models. We evaluate the model resilience with three representative graph attack methods. Random attack (**RND**) randomly adds and removes edges to generate perturbed graphs. **NEA** (Bojchevski & Günnemann, 2019) is an efficient adversarial attack method that poison the network structure and have a negative effect on the quality of network embedding and node classification. **GMA** (Zhang et al., 2020b) is the only attack model on graph matching by estimating and maximizing the densities of nodes to be attacked, for pushing them to dense regions in two graphs to generate imperceptible and effective attacks.

Versions of IDRGM model. We compare four versions of IDRGM to validate the strengths of different defense components. IDRGM-D only utilize the simplex detection to tackle inter-graph dispersion attacks. IDRGM-A only employs the node separation for addressing intra-graph assembly attacks. IDRGM-N uses the basic graph matching model without any defense techniques. IDRGM operates with the full support of both defense techniques.

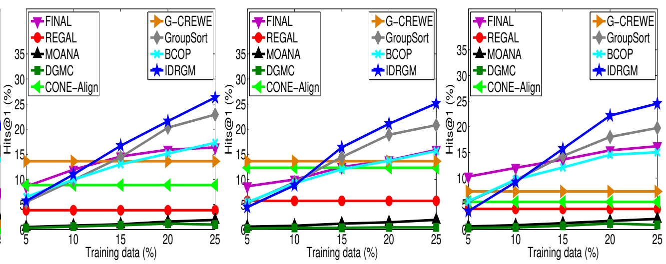
Defense performance under different attack models. We report $Hits@K$ (Yasar & Çatalyürek, 2018; Fey et al., 2020) to evaluate and compare our model to previous lines of work, where $Hits@K$ measures the proportion of correctly matched nodes ranked in the top-K list. A larger $Hits@K$ value demonstrates a better graph matching result. Table 2 exhibits the $Hits@K$ of nine graph matching algorithms on test data by three attack models over three groups of datasets. We randomly sample 30% of known matched node pairs as training data and the rest as test data. We repeat the selection process of matched node pairs five times and report the average scores. For all attack models, the number of perturbed edges is fixed to 5% in these experiments. For random attacks, we randomly add and remove edges with the half perturbation ratio (i.e., 2.5%) to three groups of datasets respectively. We use the default parameter settings for other attack models in the authors' implementation. We have observed that among nine graph matching methods, no matter how strong the attacks are, the IDRGM method achieve the highest $Hits@K$ scores on perturbed graphs in all experiments, showing the resilience of IDRGM to the adversarial attacks. Compared to the best graph matching results by other methods, IDRGM, on average, achieves 21.4%, 24.6%, and 16.8% improvement of $Hits@K$ on AS, CAIDA, and DBLP respectively. In addition, the promising performance of IDRGM under different attack models implies that IDRGM has great potential as a general defense solution to other graph matching methods.

Defense performance with varying perturbation edges. Figures 3-5 present the graph matching quality under three attack models by varying the ratios of perturbed edges from 0% to 25%. We perform the defense test for all night algorithms on the modified graphs with different perturbation ratios. It is obvious the quality by each matching mthod decreases with increasing perturbed edges. This phenomenon



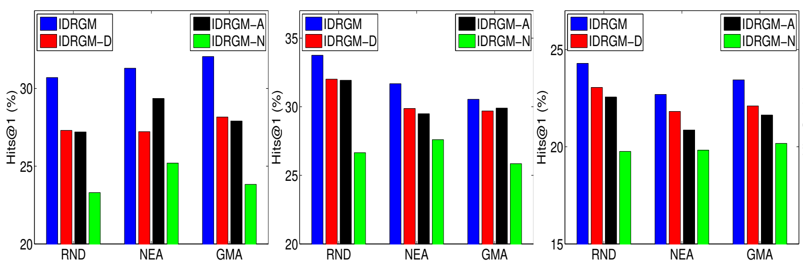
(a) RND Attack (b) NEA Attack (c) GMA Attack

Figure 5: DBLP with varying perturbed edges



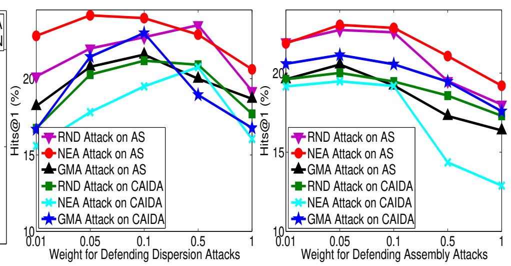
(a) RND Attack (b) NEA Attack (c) GMA Attack

Figure 6: AS with varying training ratios



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

Figure 7: $Hits@1$ (%) of IDRGM variants with 5% perturbed edges



(a) Strong Attack Weight (b) Weak Attack Weight

Figure 8: $Hits@1$ (%) with varying parameters

indicates that current graph matching methods are sensitive to adversarial attacks. However, IDRGM still achieves the highest $Hits@1$ values (> 0.206), which are better than other eight methods in most tests. In addition, the $Hits@1$ drop by our IDRGM model is slower than other methods.

Impact of training data ratios. Figure 6 shows the quality of nine graph matching algorithms on AS by varying the ratio of training data from 5% to 25%. The number of perturbed edges is fixed to 5%. We make the observations on the quality by nine matching methods. (1) The performance curves keep increasing when the training data ratio increases. (2) IDRGM outperforms other methods in most experiments with the highest $Hits@1$ scores ($> 5.12\%$). When there are many training data available ($\geq 15\%$), the quality improvement by IDRGM is obvious. A reasonable explanation is more training data makes IDRGM be more resilient to poisoning attacks under small perturbation budget.

Ablation study. Figure 7 presents the $Hits@1$ scores of graph matching on three datasets with four variants of our IDRGM model. We observe the complete IDRGM achieves the highest $Hits@1$ ($> 30.7\%$) on AS, ($> 30.5\%$) over CAIDA, and ($> 22.7\%$) on DBLP, which are obviously better than other versions. Compared with IDRGM-A, IDRGM-D performs better in most experiments. A reasonable explanation is that they focus on different types of adversarial attacks. IDRGM-D utilize the simplex detection technique to tackle inter-graph dispersion attacks. IDRGM-A employ the node separation method to defend intra-graph assembly attacks. However, the prediction of graph matching mainly depends on inter-graph links. Thus, addressing inter-graph dispersion attacks is more critical to maintaining the robustness of graph matching. However, IDRGM-A achieves the better performance than IDRGM-N. These results illustrate that both defense techniques for defending two types of adversarial attacks are important in producing robust graph matching results.

Impact of weight for defending inter-graph dispersion attacks. We assign a weighting factor to $\tau(S)h(\epsilon+4-R_K)$ in the overall loss function in Eq.(22. Figure 8 (a) measures the performance effect of weight for the graph matching by varying ϵ from 0.01 to 1. It is observed that when increasing ϵ, the $Precision$ of the IDRGM model initially increases and finally decreases. This demonstrates that there must exist an optimal weighting factor for for defending inter-graph dispersion attacks. A too large weight may reduce the ratio of defending intra-graph assembly attacks, although addressing inter-graph dispersion attacks is more critical to maintaining the robustness of graph matching. Thus we suggest well handling inter-graph dispersion attacks for the graph matching task with weight between 0.05 and 0.5.

Impact of weight for defending intra-graph assembly attacks. In addition, we assign a weighting factor to $\sum_{k=1}^{K-1}\sum_{y=1}^{Y}\frac{1}{d_y^k}$ in the overall loss function in Eq.(22. Figure 8 (b) shows the impact of weight in our IDRGM model over two groups of datasets. The performance curves initially raise when the weight increases. Intuitively, the IDRGM with small weight can help defend intra-graph assembly attacks. Later on, the performance curves decrease quickly when the weight continuously increases. A reasonable explanation is that the too large weight is able to reduce the ratio of defending inter-graph dispersion attacks, as addressing inter-graph dispersion attacks is more important to achieve a graph matching result.

Time complexity analysis. Let K be #graphs, N^k be #nodes and M^k be #edges in graph G^k, $|D|$ be the size of training data, T be #sampled node pairs in Eq.(13), Z be #sampled nearest neighbors in Page 6, I be #iteration of Algorithm 1, m be #states in phase-type distribution, the cost of simplex detection that is dominated by the computation of $\tau(S)$ in Eq.(13) is $O(T(K-1))$. For each graph G^k, the cost of embedding M of all nodes is $O((N^k)^2(K-1))$, the cost of random-restart hill-climbing is bounded with $O(N^k)$, the cost of the average distance d_y^k within each of Y groups is $O(Z^2(K-1))$, the cost of distribution estimation in Algorithm 1 is approximately equal to $O(Im^3)$. The costs of computing $\mathcal{L}_{\mathcal{M}}$ and $\mathcal{L}_{\mathcal{E}}$ are $O(K(K-1)^2/2)$ and $O(K(M^k+N^kJ)(K-1))$. The cost of computing overall loss in Eq.(22) is thus $O(K(M^k+N^kJ)(K-1)+|D|(K(K-1)^2/2+T(K-1))+Z^2(K-1)^2Y)$. As $M^k, N^k >>$ all other variables, the total cost is approximately equal to $O(M^k+N^kJ)$ by ignoring other variables.

6. Related Work

Recent defense techniques on graph learning models against adversarial attacks can be broadly classified into two categories: adversarial defense and certifiable robustness. We have witnessed various effective adversarial defense models to improve the robustness of graph mining models against adversarial attacks in node classification (Zhu et al., 2019; Xu et al., 2019b; Tang et al., 2020; 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; Luo et al., 2021), network embedding (Dai et al., 2019; Entezari et al., 2020; Wu et al., 2020b), link prediction (Zhou et al., 2019a), malware detection (Hou et al., 2019), spammer detection (Dou et al., 2020), fraud detection (Breuer et al., 2020; Zhang et al., 2020a), graph classification (Zhang & Lu, 2020; You et al., 2020), graph matching (Yu et al., 2021), and influence maximization (Logins et al., 2020). Certifiable robustness techniques aim to design robustness certificates to measure the safety of individual nodes under adversarial perturbation. Training learning models jointly with these certificates can lead to a safety guarantee of more nodes in various tasks, include node classification (Zügner & Günnemann, 2019; Bojchevski & Günnemann, 2019; Bojchevski et al., 2020; Zügner & Günnemann, 2020; Wang et al., 2020f; Schuchardt et al., 2021), graph classification (Jin et al., 2020a; Gao et al., 2020), and community detection (Jia et al., 2020).

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., 2020c;d; Zhang et al., 2020b; Zhou et al., 2020e; 2021; Jin et al., 2021; Wu et al., 2021b; Zhang et al., 2021). Graph matching, also well known as network alignment, has been a heated topic in recent years (Chu et al., 2019; Xu et al., 2019a; Wang et al., 2020d; Chen et al., 2020a;b; Zhang & Tong, 2016; Mu et al., 2016; Heimann et al., 2018; Li et al., 2019a; Fey et al., 2020; Qin et al., 2020; Feng et al., 2019; Ren et al., 2020). Research activities can be classified into three broad categories. (1) Topological structure-based techniques, which rely on only the structural information of nodes to match two or multiple graphs, including CrossMNA (Chu et al., 2019), MOANA (Zhang et al., 2019), GWL (Xu et al., 2019a), DPMC (Wang et al., 2020d), MGCN (Chen et al., 2020a), GraphSim (Bai et al., 2020), ZAC (Wang et al., 2020b), GRAMPA (Fan et al., 2020), CONE-Align (Chen et al., 2020b), and DeepMatching (Wang et al., 2020a); (2) Structure and/or attribute-based approaches, which utilize highly discriminative structure and attribute features for ensuring the matching effectiveness, such as FINAL (Zhang & Tong, 2016), ULink (Mu et al., 2016), gsaNA (Yasar & Çatalyürek, 2018), REGAL (Heimann et al., 2018), SNNA (Li et al., 2019a), CENALP (Du et al., 2019), GAlign (Huynh et al., 2020), Deep Graph Matching Consensus (Fey et al., 2020), CIE (Yu et al., 2020), RE (Zhou et al., 2020b), MetaNA (Zhou et al., 2020a), G-CREWE (Qin et al., 2020), and GA-MGM (Wang et al., 2020c); (3) Heterogeneous methods employ heterogeneous structural, content, spatial, and temporal features to further improve the matching performance, including HEP (Zheng et al., 2018), DPLink (Feng et al., 2019), BANANA (Ren et al., 2020). Several papers review key achievements of graph matching across online information networks including state-of-the-art algorithms, evaluation metrics, representative datasets, and empirical analysis (Shu et al., 2016; Yan et al., 2020).

7. Conclusions

In this work, we have proposed an integrated defense model for resilient graph matching. First, we identify and analyze two types of unique topology attacks in graph matching: inter-graph dispersion and intra-graph assembly attacks. Second, a simplex detection technique is proposed to tackle inter-graph dispersion attacks. Finally, a node separation method is developed to defend intra-graph assembly attacks.

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/.

Aiello, W., Graham, F. C., and Lu, L. A random graph model for power law graphs. *Exp. Math.*, 10(1):53–66, 2001.

Albert, R., Jeong, H., and Barabási, A. Diameter of the world-wide web. *Nature*, 401:130–131, 1999.

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.

Asmussen, S., Nerman, O., and Olsson, M. Fitting phase-type distributions via the em algorithm. *Journal of Statistics*, 23(4):419–441, 1996.

Bai, Y., Ding, H., Gu, K., Sun, Y., and Wang, W. Learning-based efficient graph similarity computation via multi-scale convolutional set matching. 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. 3219–3226, 2020.

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.

Barabási, A. and Albert, R. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

Bojchevski, A. and Günnemann, S. Certifiable robustness to graph perturbations. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2019, 8-14 December 2019, Vancouver, Canada*, pp. 8317–8328, 2019.

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.

Bojchevski, A., Klicpera, J., and Günnemann, S. Efficient robustness certificates for discrete data: Sparsity-aware randomized smoothing for graphs, images and more. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 12-18 July 2020*, 2020.

Breuer, A., Eilat, R., and Weinsberg, U. Friend or faux: Graph-based early detection of fake accounts on social networks. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pp. 1287–1297, 2020.

Chakravarthy, S. and Alfa, A. S. *Matrix-Analytic Methods in Stochastic Models*. CRC Press, 1996.

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, H., Yin, H., Sun, X., Chen, T., Gabrys, B., and Musial, K. Multi-level graph convolutional networks for cross-platform anchor link prediction. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pp. 1503–1511, 2020a.

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, 2020b.

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.

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

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.

Dou, Y., Ma, G., Yu, P. S., and Xie, S. Robust spammer detection by nash reinforcement learning. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pp. 924–933, 2020.

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.

Elte, E. L. Iv. five dimensional semiregular polytope. In *The Semiregular Polytopes of the Hyperspaces*. Simon & Schuster, 1912.

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, Z., Mao, C., Wu, Y., and Xu, J. Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, pp. 2985–2995, 2020.

Feng, J., Zhang, M., Wang, H., Yang, Z., Zhang, C., Li, Y., and Jin, D. Dplink: User identity linkage via deep neural network from heterogeneous mobility data. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pp. 459–469, 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.

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.

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.

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.

Hou, S., Fan, Y., Zhang, Y., Ye, Y., Lei, J., Wan, W., Wang, J., Xiong, Q., and Shao, F. α*Cyber*: Enhancing robustness of android malware detection system against adversarial attacks on heterogeneous graph based model. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, pp. 609–618, 2019.

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.

Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. The web as a graph: Measurements, models, and methods. In *Computing and Combinatorics, 5th Annual International Conference, COCOON '99, Tokyo, Japan, July 26-28, 1999, Proceedings*, pp. 1–17, 1999.

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, J., Zhang, H., Han, Z., Rong, Y., Cheng, H., and Huang, J. Adversarial attack on community detection by hiding individuals. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pp. 917–927, 2020.

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, X., Si, S., Zhu, X., Li, Y., and Hsieh, C. A unified framework for data poisoning attack to graph-based semi-supervised learning. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2019, 8-14 December 2019, Vancouver, Canada*, pp. 9777–9787, 2019.

Liu, Y., Ding, H., Chen, D., and Xu, J. Novel geometric approach for global alignment of PPI networks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI 2017, February 4-9, 2017, San Francisco, California, USA.*, pp. 31–37, 2017.

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., Cheng, W., Yu, W., Zong, B., Ni, J., Chen, H., and Zhang, X. Learning to drop: Robust graph neural network via topological denoising. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM 2021, Online, March 8-12, 2021*, 2021.

Mu, X., Zhu, F., Lim, E., Xiao, J., Wang, J., and Zhou, Z. User identity linkage by latent user space modelling. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pp. 1775–1784, 2016.

O'Cinneide, C. A. Characterization of phase-type distributions. *Communications in Statistics: Stochastic Models*, 6:1–57, 1990.

O'Cinneide, C. A. Phase-type distribution: open problems and a few properties. *Communications in Statistics: Stochastic Models*, 15:731–757, 1999.

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.

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.

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.

Russell, S. and Norvig, P. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, 1995.

Schuchardt, J., Bojchevski, A., Klicpera, J., and Günnemann, S. Collective robustness certificates. In *9th International Conference on Learning Representations, ICLR 2021, Online, May 4-7, 2021, Conference Track Proceedings*, 2021.

Shu, K., Wang, S., Tang, J., Zafarani, R., and Liu, H. User identity linkage across online social networks: A review. *SIGKDD Explorations*, 18(2):5–17, 2016.

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, Z., Wang, C., Hu, W., Chen, M., Dai, J., Zhang, W., and Qu, Y. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. 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. 222–229, 2020.

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*, 2020.

Vijayan, V., Gu, S., Krebs, E. T., Meng, L., and Milenkovic, T. Pairwise versus multiple global network alignment. *IEEE Access*, 8:41961–41974, 2020.

Wang, B. and Gong, N. Z. Attacking graph-based classification via manipulating the graph structure. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019*, pp. 2023–2040, 2019.

Wang, C., Wang, Y., Zhao, Z., Qin, D., Luo, X., and Qin, T. Credible seed identification for large-scale structural network alignment. *Data Min. Knowl. Discov.*, 34(6): 1744–1776, 2020a.

Wang, F., Xue, N., Yu, J., and Xia, G. Zero-assignment constraint for graph matching with outliers. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pp. 3030–3039, 2020b.

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*, 2020c.

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, 2020d.

Wang, T., Liu, H., Li, Y., Jin, Y., Hou, X., and Ling, H. Learning combinatorial solver for graph matching. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pp. 7565–7574, 2020e.

Wang, Y., Liu, S., Yoon, M., Lamba, H., Wang, W., Faloutsos, C., and Hooi, B. Provably robust node classification via low-pass message passing. In *Proceedings of the 20th IEEE International Conference on Data Mining (ICDM'20)*, Online, November 17-20 2020f.

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.

Wu, Y., Liu, X., Feng, Y., Wang, Z., and Zhao, D. Neighborhood matching network for entity alignment. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pp. 6477–6487, 2020c.

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.

Yang, X., Liu, Z., and Qiaoxu, H. Incorporating discrete constraints into random walk-based graph matching. *IEEE Trans. Syst. Man Cybern. Syst.*, 50(4):1406–1416, 2020.

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.

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*, 2020.

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.

Zang, X., Xie, Y., Chen, J., and Yuan, B. Graph universal adversarial attacks: A few bad actors ruin graph learning models. *CoRR*, abs/2002.04784, 2020.

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, S., Yin, H., Chen, T., Hung, Q. V. N., Huang, Z., and Cui, L. Gcn-based user representation learning for unifying robust recommendation and fraudster detection. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, pp. 689–698, 2020a.

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 2020b.

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.

Zheng, V. W., Sha, M., Li, Y., Yang, H., Fang, Y., Zhang, Z., Tan, K., and Chang, K. C. Heterogeneous embedding propagation for large-scale e-commerce user alignment. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, pp. 1434–1439, 2018.

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, F., Cao, C., Trajcevski, G., Zhang, K., Zhong, T., and Geng, J. Fast network alignment via graph meta-learning. In *39th IEEE Conference on Computer Communications, INFOCOM 2020, Toronto, ON, Canada, July 6-9, 2020*, pp. 686–695, 2020a.

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, T., Lim, E., Lee, R. K., Zhu, F., and Cao, J. Retrofitting embeddings for unsupervised user identity linkage. In *Advances in Knowledge Discovery and Data Mining - 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11-14, 2020, Proceedings, Part I*, pp. 385–397, 2020b.

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, 2020c.

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 2020d.

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 2020e.

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. and Günnemann, S. Certifiable robustness and robust training for graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pp. 246–256, 2019.

Zügner, D. and Günnemann, S. Certifiable robustness of graph convolutional networks under structure perturbations. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2020, August 23-27, 2020*, 2020.

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.