Sign Rank Limitations for Inner Product Graph Decoders

Su Hyeong Lee [1] Qingqi Zhang [1] Risi Kondor [1,2]

[1]University of Chicago, CCAM [2]University of Chicago, Dept. of Computer Science

Overview

Goal: Inner product-based decoders are among the most influential frameworks used to extract meaningful data from latent embeddings. However, these decoders have shown limitations in representation capacity particularly for graph reconstruction problems. Can we provide a rigorous theoretical explanation of this phenomenon?

Contributions:

- ✓ To our knowledge, we provide the first theoretical elucidation of why inner product decoders face additional restrictions in graph machine learning using a modified definition of the **sign rank** [1]
- ✓ We derive straightforward modifications to circumvent this issue without deviating from the inner product framework, which is simple to implement
- ✓ We develop a variety of different strategies for proving that a graph cannot be represented by a low-rank embedding
- ✓ We present examples of pedagogical graph structures for which complexifying the latent space provably permits significantly lower dimensional latent encodings to be used
- ✓ We design decoding architectures which drastically expands the representation capacity of inner product decoders that subsume the expressivity of the aforementioned complex GNN

Problem Motivation

Observation: Entry-wise truncation or thresholding allows for drastic matrix rank expansion for graph adjacency matrices.

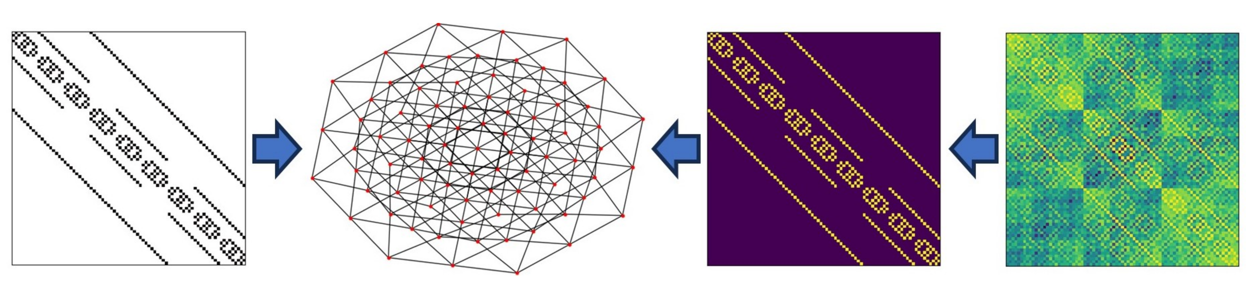


Figure 1. The adjacency $\mathbf{A}$ of the 4-dimensional $3 \times 3 \times 3 \times 3$ grid graph shown on the left has matrix rank 62, whereas the rightmost matrix has matrix rank 6.

Setting for Mechanistic Analysis:

Given a feature matrix $\mathbf{X} \in \mathbb{R}^{N \times d}$ and an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, a *Graph Convolutional Network (GCN)* can be utilized to learn a latent mapping $\mathbf{X} \mapsto \mathbf{Z} \in \mathbb{R}^{N \times f}$, where f is the latent dimension. Form the latent matrix $\mathbf{Z}$ via message passing through the two layer network [2]:

$$\mathrm{GCN}(\mathbf{X}, \mathbf{A}) = \tilde{\mathbf{A}}\,\mathrm{ReLU}(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}_0)\mathbf{W}_1$$

Inference Model

Simple Variational Inference Model:

$$q(\mathbf{Z} \mid \mathbf{X}, \mathbf{A}) = \prod_{i=1}^{N} q(\mathbf{z}_i \mid \mathbf{X}, \mathbf{A})$$

$$q(\mathbf{z}_i \mid \mathbf{X}, \mathbf{A}) = \mathcal{N}\left(\mathbf{z}_i \mid \boldsymbol{\mu}_i, \mathrm{diag}\left(\boldsymbol{\sigma}_i^2\right)\right).$$

Sample Decoder: For i-th *row vector* $\mathbf{w}_i$ of $\mathbf{Z}$, form generative decoding model by taking inner products,

$$p(\mathbf{A} \mid \mathbf{Z}) = \prod_{i=1}^{N}\prod_{j=1}^{N} p(A_{ij} \mid \mathbf{z}_i, \mathbf{z}_j)$$

$$p(A_{ij} = 1 \mid \mathbf{z}_i, \mathbf{z}_j) = \sigma\left(\mathbf{z}_i^\top \mathbf{z}_j\right),$$

where σ denotes the sigmoid.

Sign Rank & Low-Rank Classification

Definition 1: (Sign Rank) Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ and $F = \mathbb{C}$. Then, the *complex sign rank* of $\mathbf{A}$ is the minimal f such that there exists $\mathbf{Z} \in F^{N \times f}$ with columns $\mathbf{z}_i$ satisfying

$$\mathrm{sign}(\mathbf{A}) = \mathrm{sign}\left(\mathfrak{Re}\left(\left[\begin{array}{cccc} | & | & & | \\ \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_f \\ | & | & & | \end{array}\right] \times \left[\begin{array}{c} -\mathbf{z}_1^\top- \\ -\mathbf{z}_2^\top- \\ \vdots \\ -\mathbf{z}_f^\top- \end{array}\right]\right)\right).$$

The *real sign rank* is given by replacing $F = \mathbb{R}$.

Theorem 1

For $k_1, \dots, k_n \in \mathbb{R}$, let

$$\mathbf{Z} = \begin{pmatrix} \cos(k_1 x) & \sin(k_1 x) \\ \cos(k_2 x) & \sin(k_2 x) \\ \vdots & \vdots \\ \cos(k_N x) & \sin(k_N x) \end{pmatrix}, \quad \tilde{\mathbf{A}} = \mathbf{Z}\mathbf{Z}^\top. \tag{1}$$

Then, for any rank 2 adjacency $\mathbf{A}$ representing a connected graph, there exists $(k_1, \dots, k_n)$ such that $\mathrm{sign}(\mathbf{A}) = \mathrm{sign}(\tilde{\mathbf{A}})$.

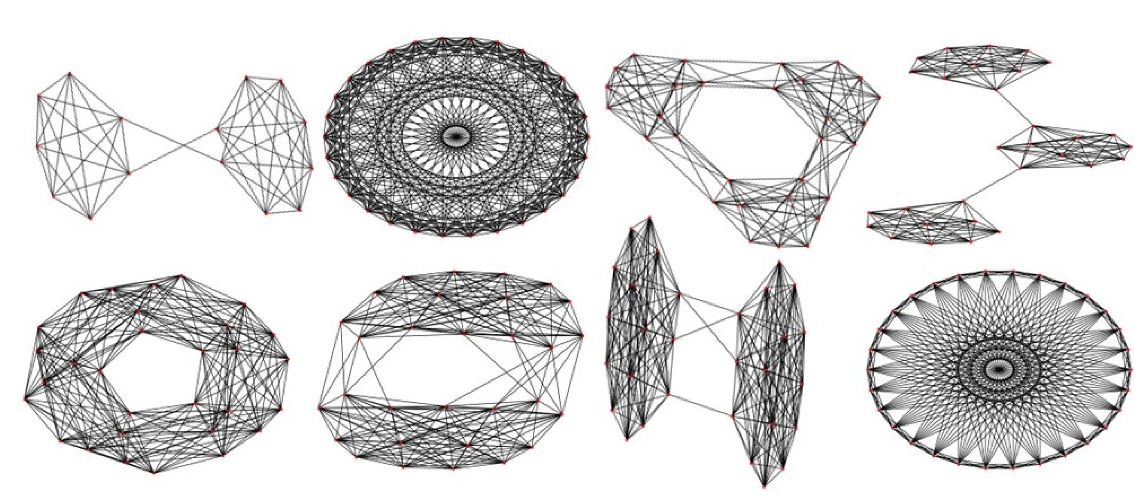


Figure 2. Examples of rank ≤ 2 graphs.

Motivation & Observation

Naturally follows from noting the pairwise periodic regularity of the entries of $\tilde{\mathbf{A}}$ which emerges despite their perpetually imbalanced periods.

Proof Sketch & Generalizations

Two Helpful Lemmas:

Lemma 1

Let $k_i := \sqrt{p_i}$ where $p_1 < p_2 < \dots$ are any sequence of positive integer primes. There exists no $n_1, n_2 \in \mathbb{Z}_{\neq 0}$ such that $n_1 t_1 = n_2 t_2$, where t_1, t_2 are periods of $\tilde{A}_{ij}, \tilde{A}_{i'j'}$ for $\{i, j\} \neq \{i', j'\}$. That is, limiting the indices to the lower triangular portion $i > j$, the periods of $\tilde{A}_{ij}$ can never match.

Lemma 2

Let t_1, t_2 be as in Lemma 1. Then for any $\varepsilon > 0$, there exists $n_1, n_2 \in \mathbb{Z}_{>0}$ such that $|n_1 t_1 - n_2 t_2| < \varepsilon$.

Remark: Any two distinct lower triangular entries of $\tilde{\mathbf{A}}$ may never simultaneously take the value 1 for any $x \neq 0$, but may become arbitrarily close.

Theorem 2

For $k_1, k_1', \dots, k_N, k_N' \in \mathbb{R}$, let

$$\mathbf{Z} = \begin{pmatrix} \cos(k_1 x)\sin(k_1' x) & \cos(k_1 x)\cos(k_1' x) & \sin(k_1 x) \\ \cos(k_2 x)\sin(k_2' x) & \cos(k_2 x)\cos(k_2' x) & \sin(k_2 x) \\ \vdots & \vdots & \vdots \\ \cos(k_N x)\sin(k_N' x) & \cos(k_N x)\cos(k_N' x) & \sin(k_N x) \end{pmatrix}$$

and $\tilde{\mathbf{A}} = \mathbf{Z}\mathbf{Z}^\top$. Then for any connected rank 3 adjacency $\mathbf{A}$, there exists k_i, k_i' such that $\mathrm{sign}(\mathbf{A}) = \mathrm{sign}(\tilde{\mathbf{A}})$.

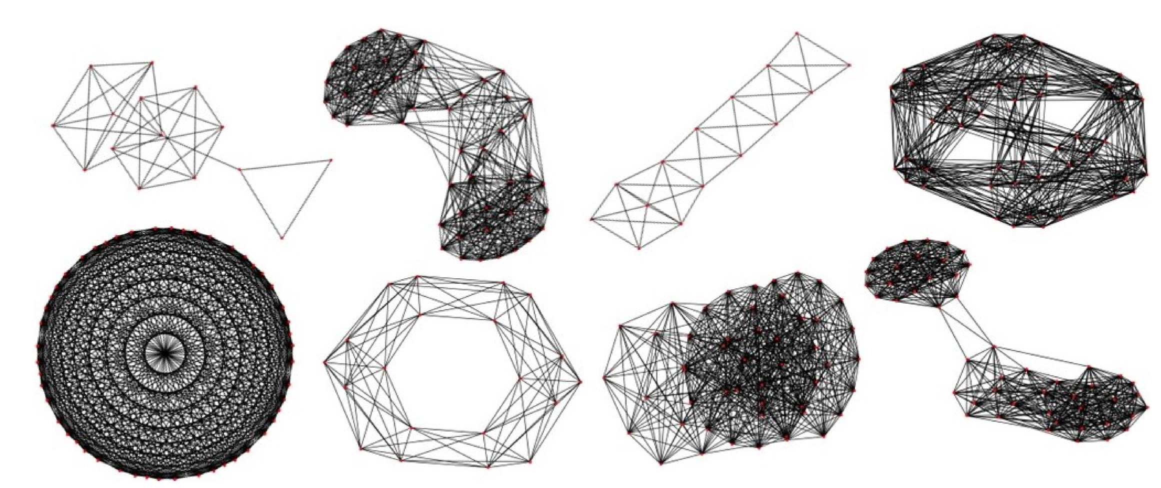


Figure 3. Examples of rank ≤ 3 graphs.

Lower Bounds for Graph Rank

Theorem 3

The real rank of an $(N-1)$-star graph is lower bounded by $(N+1)/2$. Moreover, any star graph is complex rank 1.

Idea : For graphs with node connectivity imbalance, complex latent space is far more economical. We can synthesize examples for which compression gap is arbitrarily large!

Corollary 1

For graph G, let H be the largest induced $(N_H - 1)$-star subgraph. Then any faithful real latent encoding of G must possess at least $\lceil (N_H + 1)/2 \rceil$ dimensions.

Architectures

Generalization:

$$\mathrm{sign}(\mathbf{A}) = \mathrm{sign}(\mathbf{Z}_1\mathbf{C}_1\mathbf{Z}_1^\top - \mathbf{Z}_2\mathbf{C}_2\mathbf{Z}_2^\top)$$

DGAE:

$$\mathbf{A}\,\mathrm{ReLU}\left(\mathbf{A}\mathbf{X}\mathbf{W}_0\right)\mathbf{W}_1 = \mathbf{Z}_0, \quad \hat{\mathbf{A}} = \mathrm{sign}\left(\mathbf{Z}_0\mathbf{C}_0\mathbf{Z}_0^\top\right)$$

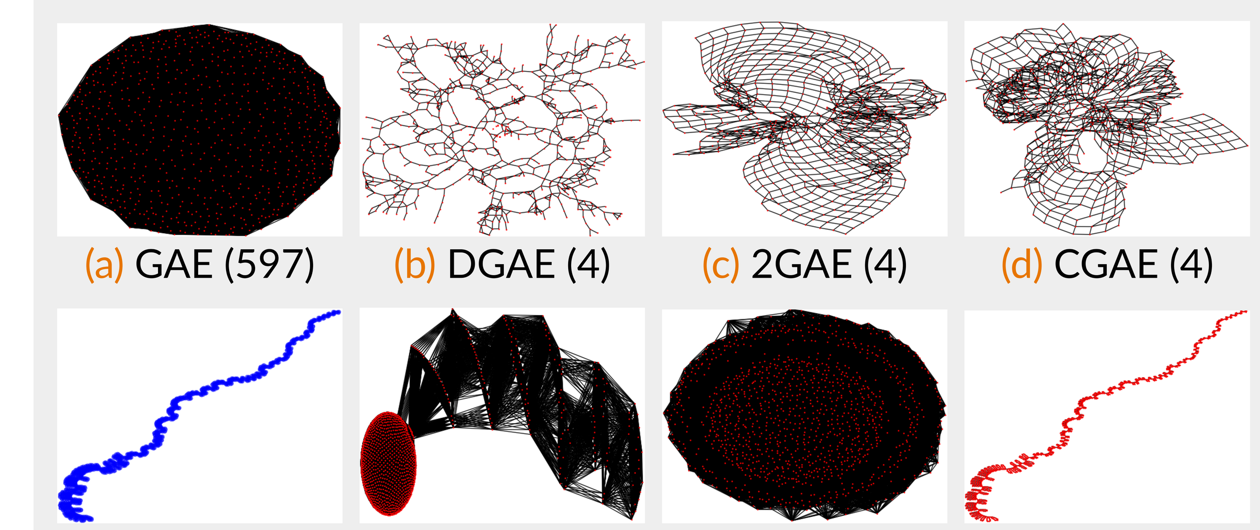
m-GAE:

$$\mathbf{A}\,\mathrm{ReLU}\left(\mathbf{A}\mathbf{X}\mathbf{W}_{2n+1}\right)\mathbf{W}_{2n+2} = \mathbf{Z}_n$$

$$\hat{\mathbf{A}} = \mathrm{sign}\left(\sum_{n=0}^{k}(-1)^n \mathbf{C}_{4n}\mathbf{Z}_n\mathbf{C}_{4n+1}\mathbf{C}_{4n+2}\mathbf{Z}_n^\top\mathbf{C}_{4n+3}\right)$$

Sample Experiments

Performance Enhancement after Intervention



(a) GAE (597) (b) DGAE (4) (c) 2GAE (4) (d) CGAE (4)

(e) Original (f) GAE (8) (g) GAE (597) (h) 2GAE (8)

Minimal Rank Embeddings of Real-World Data

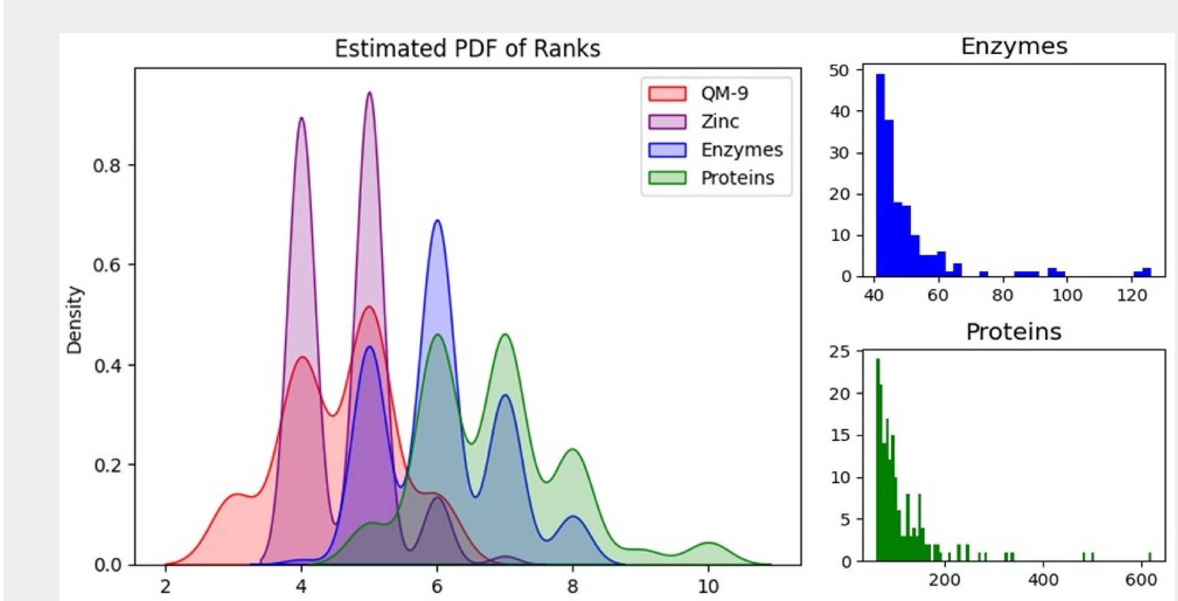


Figure 5. All molecules with more than 27, 36, 40, 60 nodes from QM-9, Zinc, TU-Enzymes, TU-Protein datasets are treated via GAE and DGAE, totaling 35, 118, 162, 176 graphs. GAE mostly fails for up to 80 latent dimensions, but DGAE embeds all in ≤ 10.

References

[1] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus vapnik-chervonenkis dimension. Sbornik: Mathematics, 208(12), 2017.

[2] Thomas N. Kipf and Max Welling. Variational graph auto-encoders. NeurIPS Workshop on Bayesian Deep Learning, 2016.