GRAPH EMBEDDING VIA TOPOLOGY AND FUNCTIONAL ANALYSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graphs have been ubiquitous in Machine Learning due to their versatile nature in modelling real world situations. Graph embedding is an important precursor to using graphs in Machine Learning , and much of performance of algorithms developed later depends heavily on this. However very little theoretical work exists in this area , resulting in the proliferation of several benchmarks without any mathematical validation , which is detrimental. In this paper we present an analysis of deterministic graph embedding in general , using tools from Functional Analysis and Topology. We prove several important results pertaining to graph embedding which may have practical importance .One limitation of our work in it's present form is that it's applicable to only deterministic embedding approaches , although we strongly hope to extend it to random graph embedding methods as well in future. We sincerely hope that this work will be beneficial to researchers working in field of graph embedding.

1 INTRODUCTION

Graphs are ubiquitous in Machine learning[10][25].They are versatile in modelling several real world phenomena[6] like drug design to friendship recommendation in social networks[7]. An important challenge is finding a way to represent graphs[14] in a way so that it can be easily exploited for tasks in machine learning[8] , graph embedding is a technique used to perform this. Graph embedding is an important precursor to using graphs in Machine learning[7] [4]. Several approaches have been proposed in this regard[4][5][11][12][13] [15], but no evaluation method was ever proposed.

Analysis in this area has been mostly emperical without much theoretical backing[9][12].In this paper we use tools from functional analysis specifically Hilbert spaces and topology to evaluate all graph embedding approaches in an abstract way .To the author's best knowledge this is the first theoretical work in the area of analyzing several embedding methods under single framework using functional analysis and topology.We consider features to be embedded on only nodes and do not consider features on edges. We also assume that graph is defined by the postion of its Vertices .

Most graph embedding approches proposed till now fall under a sinlge paradigm namely a proximity function which measures the distance between nodes , a linear mapping i.e an encoder which maps this graph to a graph in higher dimensional space , and a proximity function on this higher dimensional graph , finally a loss function which measures discrepancy between the two distance functions[4].

In Section 2 we give necessary mathematical prerequisites like Hilbert spaces[2] , point set topology[1] etc.In Section 3 we translate graph embedding problem into mathematical language namely a 3 tuple space and also prove that the 3 tuple space is a Hilbert space and a closed set .In section 4 we prove some important theorems pertaining to our analysis. Finally we give algorithmic aspects of our approach.Finally we give the conclusions and future work .

As this is the first step , Our paper considers only deterministic embedding approaches[17] and does not consider random walk based aproaches[13][16] ,however we definitely have strong intention to analyze the latter in future.

2 PREREQUISITES

In this section , we introduce mathematical prerequisites namely functional analysis and topology terminology and also some important facts corresponding to them.

2.1 TOPOLOGICAL SPACE

A topological space is an ordered pair (X, τ), where X is a set and τ is a collection of subsets of X, satisfying the following axioms:

The empty set and X itself belong to τ.
Any arbitrary (finite or infinite) union of members of τ still belongs to τ.
The intersection of any finite number of members of τ still belongs to τ.
The elements of τ are called open sets and the collection τ is called a topology on X.
The complement of open sets are called Closed sets

2.2 COMPACT SPACE

Formally, a topological space X is called compact if each of its open covers(open sets which cover the set) has a finite subcover. That is, X is compact if for every collection C of open subsets of X such that

$$X = \bigcup_{x \in C} x$$

In an arbitrary topological space compactness is tough to visualize but in metric spaces it's intuitive which says that compact sets are precisely closed and bounded sets which agrees with our intuition.

2.3 HOMOTOPY AND FUNDAMENTAL GROUP

Homotopy is continuous deformation of functions , more formally it is defined as

H:X×[0,1]→ Y such that
H(x,0) = f(x) and H(x,1) = g(x) for all x.

- Every function is homotopic to itself trivially, by setting H(x,t) = f(x) for all t ∈ [0,1], Hence Homotopy is symmetric .

- Suppose H:X×[0,1]→ Y is a homotopy from f to g then K:X×[0,1]→ Y where K(x,t) = H(x,1-t) for all t ∈ [0,1] now K is a homotopy from g to f,hence Homotopy is reflexive

- Suppose H:X×[0,1]→ Y is a homotopy from f to g and K:X×[0,1]→ Y is a homotopy from g to h

  $$\begin{aligned} L(x,t) &= H(x,t) 0 \leq t \leq 1/2 \\ &= K(x,t) 1/2 \leq t \leq 1 \end{aligned}$$

  So Homotopy is Transitive as well.

Combining these three we see that homotopy is equivalence relation, and its equivalence classes form a group called Fundamental group.

2.4 MULTIVARIATE POLYNOMIALS

A Monomial in 'n' variables $(x_1, x_2, ...x_n)$ is of the form

$$x_1^{a_1} x_2^{a_2} ....x_n^{a_n}$$

Where a_i are indices of the variables and are non-negative integers.

We can simplify the notation for monomials as follows: let $a = (a_1, ..., a_n)$ be an n-tuple of nonnegative integers. Then we set $x^a = x_1^{a_1} x_2^{a_2}$..........$x_n^{a_n}$

Total degree of this Monomial is $a_1 + a_2 +$ .... $+ a_n$

A polynomial f in $x_1, ..., x_n$ with coefficients in 'K' is a finite linear combination (with coefficients in K) of Monomials. We will write a polynomial f in the form

$$f = \sum_\alpha a_\alpha x^\alpha, a_\alpha \in K$$

where the sum is over a finite number of n-tuples $a = (a_1, ..., a_n)$. The set of all polynomials in$x_1, ..., x_n$with coefficients in 'K' is denoted $K[x_1, ..., x_n]$.

2.5 Norm Linear space and Hilbert space

2.5.1 Norm

Given a vector space V over a field 'F' of the Real or Complex numbers , a norm on V is a non-negative valued function p: V $\rightarrow$ R with the following properties:

For all a $\in$ F and all u, v$\in$ V,

- p(u + v) $\leq$ p(u) + p(v) (being sub-a dditive or satisfying the triangle inequality).
- p(av) = | a | p(v) (being absolutely homogeneous or absolutely scalable linearly).
- If p(v) = 0 then v = 0 is the zero vector.

A Vector space V with a norm defined on it is called a norm linear space.A Norm linear space with a complete norm defined on it is called Banach Space.

2.5.2 Inner product space and Hilbert Space

Inner product space is the space satisfying the following properties

- The inner product is conjugate symmetric; that is, the inner product of a pair of elements is equal to the complex conjugate of the inner product of the swapped elements: $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
- Inner product is bilinear mapping i.e $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.
- Inner product is a homogenous mapping $\langle \alpha x, z \rangle = \alpha \langle x, z \rangle$.
- Inner product is a positive mapping i.e $\langle x, x \rangle \geq 0$ for all x .

Given a metric space (X, d), a Sequence $\{x_n\}_{n \geq 1}$ is Cauchy, if for every positive real number $\epsilon \geq 0$ there is a positive integer N such that for all positive integers m, n geq N, the distance $d(x_m, x_n) \leq \epsilon$
A Metric space in which every cauchy sequence converges is called a Complete metric space .
Hilbert space is a Norm linear space which is a complete metric space and also an Inner Product Space

2.6 Connected and Path Connected Spaces

- A topological space X is said to be disconnected if it is the union of two disjoint non-empty open sets. Otherwise, X is said to be connected. A subset of a topological space is said to be connected if it is connected under its subspace topology.
- A path-connected space is a stronger notion of connectedness, requiring the structure of a path. A path from a point x to a point y in a topological space X is a continuous function f from the unit interval to X i.e f : [0,1] $\rightarrow$ X with $f(0) = x$ and $f(1) = y$. A path-component of X is an equivalence class of X under the equivalence relation which makes x equivalent to y if there is a path from x to y. The space X is said to be path-connected (or pathwise connected or 0-connected) if there is exactly one path-component, i.e. if there is a path joining any two points in X

3 ENCODER - DECODER PERSPECTIVE OF GRAPH EMBEDDING

We formulate the problem of problem of graph embedding into the language of mathematics specifically topology and fuctional analysis using encoder decoder formulation , in this paper we closely follow the approach taken by [4] , first we introduce the necessary framework Now we translate our problem into the language of topology and functional analysis.Consider the 3 tuple space $M \times V \times V$

> M = Set of all matrices over the field of real numbers.
> V = Vector space of all multivariate polynomials in prefixed number of variables and degree.

Our goal in this work is to encode nodes as low-dimensional vectors that summarize their graph position and the structure of their local graph neighbourhood.These Low dimensional embeddings are encodings of nodes in the original graph, and the geometric relations between nodes in these should reflect the relations in the original graph.

Terminology for Encoder-Decoder perspective :

- Encoder This function maps nodes to the low dimensional space. Mathematically it can be written as
$$ENC : V \rightarrow R^d$$
- Decoder Decoder function as the name suggests decodes the structural informationof the graph.Mathematically it can be written as
$$DEC : R^d \times R^d \rightarrow R^+$$
- Loss function Loss function measures the discrepancy between proximity measures of original and latent space graph ,mathematically it is
$$l : R \times R \rightarrow R$$
- A pairwise proximity function It's a mathematical function which gives measure of proximity of 2 nodes in the original graph, formally it's given by
$$S_G : V \times V \rightarrow R^+$$

Decoder gives the proximity of the nodes in the latent space , our goal should be to construct a decoder which will deviate the least from the original proximity function ,more formally our goal is to minimize

$$\sum_{(v_i,v_j)\in D} l[(DEC(z_i, z_j), S_G((v_i, v_j)]$$

We give a summary about the existing embedding techniques and their corresponding encode,decoder and loss functions in the table1, taken from[4].as we can see all of the involved functions are multivariate fuctions of the coordinates in latent space.Most of the present methods rely on Direct encoding which means the encoder function is a Matrix or a Linear Mapping, so in our analysis we restrict to linear encoders , we give results for optimal embeddings in linear encoders. Now we can easily see that every embedding can be chracterized by the above 3 functions if loss function is fixed to be l_2 norm which we do as it's the most popular choice of loss function in embedding as we can see from the table all functions encoders,proximity measures (although they are discrete spaces they are closed so functions on them can be extended to the whole space)[1] and decoders are all multivariate polynomials in their coordinates in latent space.however although they are polynomials their degree is bounded and number of variables is equal to the the dimension of the latent space.

The 3 tuple space is hence the space of all graph embeddings with linear encoders, polynomial decoders and polynomial proximity measures . Every embedding can be thought as a point in this 3 Tuple space. So we study the topology and analysis of this 3 tuple space to gain further insights into the graph embedding problem, now we prove some basic theorems of our papers.

Table 1: Table summarizing various embedding methods taken from [4]

Method	Decoder	Proximity measure
Laplacian Eigenmaps	$\|z_i - z_j\|_2^2$	General
Graph factorization	$z_i^t z_j$	A_{ij}
GraRep	$z_i^t z_j$	$A_{ij}, A_{ij}^2......A_{ij}^k$
HOPE	$z_i^t z_j$	General

4 Mathematical Results corresponding to Graph embedding

Theorem 1. 1. *Product of Closed sets is Closed.*

2. *Product of Convex sets is Convex.*
3. *Set of all Polynomials in prefixed number of variables and degree is a Vector Space.*
4. *If M is connected , convex and closed space and V is finite dimensional vector space then the Space $M \times V \times V$ is Connected , closed and convex .*

Proof. (i) If A and B are closed sets A^c and B^c are open , because product of open sets is open we know that $A \times B = A^c \times Y \bigcup X \times B^c$ hence $A \times B$ is closed by induction it trivially extends to finitely many products.

(ii)Say A and B are convex sets , consider $A \times B$, (a,b) and (c,d) $\in A \times B$. α(a,b) +(1-α)(c,d) = (αa+ 1- αc) , αb+ 1- αd) $\in A \times B$ because A and B are convex.

(iii)Set of all polynomials with prefixed number of variables and degree is a vector space by compact notation introduced in previous sections $f = \sum_\alpha a_\alpha x^\alpha, a_\alpha \in K$ the proof is the same as that of single variable so if fand g are 2 polynomials f+g is also a polynomial and so is λf forall $\lambda \in F$.

(iv) Suppose M and V are coonected then $M \times V \times V$ is a connected space which is standard proof found in all topology books. However we prove the converse as it's not that common , so we give it here say $M \times V \times V$ is connected now take the projection mapping onto first coordinate π_1

$$\pi_1 : M \times V \times V \to M$$
$$\pi_2 : M \times V \times V \to V$$

π_i are continuous as projection maps are surjective open continuous maps and image of a connected set under continuous mapping is connected , but M and V are connected by the definition of connectedness in the previous sections . ☐

Theorem 2. 1. *There is unique embedding for which loss function takes it's minimal value.*

2. *Encoder matrix can always be UNIQUELY approximated by the matrices of form*
   (a)$\|A\| \leq k$ where k is any real number and$\|.\|$ is any norm on matrices
   (b) Set of all matrices of Operator norm less than 1.(need not be square matrices)
   (c) Set of all Positive Definite matrices .

3. *This Optimal embedding can always be attained irrespective of which embedding we start by continuously varying in the 3 tuple space*

Proof. (i) We give the proof in several steps. call $M \times V \times V$ as 'E'.Vector space 'V' is finite dimensional and hence Hilbert space and set of all matrices of order $m \times n$ is a Hilbert space as it can be identified by $R^{m \times n}$ and hence is also a Hilbert space.

If x and y $\in$ E , then (x+y)/2 $\in$ E by convexity taking α to be 2. let δ = inf{$\| x \|$:x$\in$E} by the definition $2\delta \leq \| x+y \|$. we know from the definition of inner roduct and norm that $\|x+y\|^2$ = $\langle x+y, x+y\rangle$ and $\|x-y\|^2$ = $\langle x-y, x-y\rangle$, by adding both the equations we get $\|x+y\|^2$ + $\|x-y\|^2$ = $\langle x+y, x+y\rangle$ + $\langle x-y, x-y\rangle$ = $\langle x, x\rangle$ + $\langle y, y\rangle$ + 2$\langle x, y\rangle$ + $\langle x, x\rangle$ + $\langle y, y\rangle$ - 2$\langle x, y\rangle$, finally we get

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

By the definition of closed set we know that $\exists$ a sequence $\{y_n\} \in$ E such that $\|y_n\| \to \delta$ as n $\to \infty$ now that $\|y_n - y_m\|^2 \leq 2$ ($\|y\|_n^2 + \|y\|_m^2$) - $4\delta^2$
as n $\to \infty$ $\|y_n - y_m\| \to 0$, so the sequence y_n is cauchy.because the space is complete its also convergent so $\exists$ x_0 such that $y_n \to x_0$ and this $x_0 \in$ E as E is closed.so this x_0 is the unique norm element By using δ = inf{$\| x \|$:x$\in$E} and Parallellogram Identity we get

$$\|x-y\|^2 \leq 2 (\|x\|^2 + \|y\|^2) - 4\delta^2$$

If $\|x\|$=$\|y\|$=δ then $\|x-y\|^2 \leq 0 \Rightarrow x = y$ (as norm can't be negative).

so there exists an element of smallest norm which is unique , our only assumption was space is closed and convex so by applying this result to $M \times V \times V$ we get the desired result by inducing the following norm on it . we define norm on it as

$$d(f, g) = \|f - g\|_2.$$

(ii) We use hilbert projection theorem for closed convex sets in Hilbert Spaces , and prove that the sets below are all closed and convex (a) let $\mathcal{A}_k$ = {A: $\|A\| \leq$ k} now clearly $(\infty, k]$ is closed set , also norm is a continuous function and by definition of continuous function we have $\|.\|^{-1}(\infty, k]$ is a closed set too , for convexity we use the below arguement

$$\begin{aligned}
\left\|(\alpha A) + (1-\alpha B)\right\| &\leq \|\alpha\| A + \left\|(1-\alpha)B\right\| && (1)\\
&= \alpha\|A\| + (1-\alpha)\|B\| && (2)\\
&\leq (\alpha)k + (1-\alpha)k && (3)
\end{aligned}$$

so we proved that $\mathcal{A}_k$ is a closed and convex set.

(b) Although this looks like special case of (a), we show that the convex hull here involves orthonomal matrices which enjoy very nice properties and simplify computations , however no explicit characterization can be given in case (a) unlike here , we give a proof when the matrix is square , the proof for general non square called as stiefel manifold is given in [22],
Stiefel manifold (St(p,n))= $\{X \in R^{n \times p} : X^T X = I_p\}$
As 'A' is unitary matrix by its isometry property we know its operator norm is 1,so $\|A\|$ = 1 whenever A*A = I by using triangle inequality we get that

$$\begin{aligned}
\left\|(\alpha A) + (1-\alpha B)\right\| &\leq \|\alpha\| A + \left\|(1-\alpha)B\right\| && (4)\\
&= \alpha\|A\| + (1-\alpha)\|B\| && (5)\\
&\leq (\alpha)1 + (1-\alpha)1 && (6)\\
&\leq 1 && (7)
\end{aligned}$$

to prove the other direction we use singular value decomposition which says that $A = U\sigma V^*$
$U^*AV = \sigma$
$\|\sigma\| = \|U^*AV\| = \|A\| \leq 1$
so largest singular value of A is 1 so all its singular values are less than one , For convexity we have the following , consider all the points of $\{-1, 1\}^n$ say $s_1, s_2, ...s_{2^n}$, now all elements of σ are

in the convex hull of these 2^n points , the proof is given in [24] .hence σ is in the convex hull of $diag(s_i)$. As $A = U\sigma V^*$, So A is also in the convex hull of orthonormal matrices.

(c) Finally we prove that set of Positive definite matrices is closed and convex , convexity is trivial by the definition of semdefinite matrices , for closedness we use the theorem 3.1 from [23] specifically part(ii) so now we define a mapping from matrix to an 'n' tuple namely $f : A \rightarrow (\delta_1, \delta_2, ...\delta_n)$ This is clearly a continuous map as determinant is continuous map and product of continuous maps is continuous , now the preimage of $(\Delta_1, \Delta_2, ...\Delta_n) \cap \underbrace{(R^+ \times R^+ \times ......R^+)}_{n \text{ times}}$(which is closed) is our set.

Finally when we use Hilbert projection theorem combined with above facts we get the desired result .

((iii) We first prove that the space is path connected and hence the statement follows . To do this we just have to prove that the space $M \times V \times V$ is contractible i.e it has trivial fundamental group , Say E is a convex set, let $x_0 \in E$, we define a function H as

$$H : A \times [0, 1] \rightarrow A \; H(x, t) = tx_0 + (1 - t)x$$

by convexity H(x,t) $\in$ A , H is polynomial in x and t and hence continuous so it's a homotopy by definition also we can check that

$$H(x, 0) = \text{x for all x} \in A$$
$$H(x, 1) = x_0 \text{ for all x} \in A$$

As H is homotopic to Identity map , E is contractible or is homotopic to point space . As fundamental group is homotopy invariant we get that fundamental group of E is trivial. ☐

Remark.

- *In (i) Our theorem we just proved gives important results which say that there is one and only one embedding which is optimal in the sense of minimal loss.Although it might be tempting to say as the set is convex and* $\|.\|^2$ *is a strictly convex function so it's one line proof, however* $\|.\|^2$ *is not a norm (it fails to satisfy the scaling property i.e it scales quadratically not linearly)*

- *Approximating the encoder matrix with certain nice matrix familes will save computational power we empasize when it's possible.The encoder matrix should be non square matrix in most of the cases in practice , because we have to embed it in high low dimensional space , in (ii) our results ii(a),ii(b) hold in the general sense , not necessarily square , hence more practical* ***.Also some families like 'set of matrices of bounded rank' cannot be used for Approximation of encoder , the proof can be found in [24]***

- *In (iii) no matter which embedding we start from we can always reach the optimal embedding by continuously moving from our embedding and traversing through a series of embeddings and whenever integration is involved in 3 tuple space we can safely consider only the straight path without any loss .*
  Also as a byproduct , we also get that as norm is continuous by intermediate value theorem we will have to traverse through all the embeddings whose loss lies between our initial embedding and the optimal embedding .

5 ALGORITHMIC AND COMPUTATIONAL ASPECTS

This section we consider the computational aspects of finding an optimal matrix to the encoder in terms of loss function, this is reduced to the problem of Covex optimization as both the norm and search space are all convex , however our work is better than previous work in the fact that we also know that local minima exists and is unique even when compactness is not assumed , with this the problem of getting stuck in local minima is eliminated.so our problem of finding optimal encoder is equivalent to solving a convex optimization which can be framed as

$$\underset{A \in C}{min} \left\| A - \hat{A} \right\|_2 \; C \text{ is a closed convex set}$$

Problems of the above type are quite banal in Machine Learning , now to finding the optimal embedding in terms of loss function we can frame it as

$$\underset{x \in V \times V}{min} \|x\|_2$$

To analyze the complexity of the algorithm we need to analyze the complexity of search space i.e dimension of the search space suppose that a_i is the maximum power of x_i in any of the multivariable polynomial then the dimension of V is

$$\Pi_{i=1}^{n} a_i$$

and hence the dimension of V× V is $2\Pi_{i=1}^{n} a_i$
Above convex optimization problem is also banal and already well investigated in the ML community , so we are only giving the necessary details necessary to guage the complexity of theproblem , our main goal in this paper is to disclose the relation between two superficially unfamiliar fields and explore the beauty of such connection to further enhance our understanding of both the fields.

6 Conclusion and Future work

In this paper we to the best of our knowledge propose a first ever theoretical model for analyzing graph embedding methods and give a formal abstract method for evaluating various embedding methods at once and also prove certain results correspondingly. We prove certain theoretical results which we think might have practical significance like can there be multiple extrema of graph embedding ,also can we approximate encoder matrix with matrix families having certain nice properties . We din't consider questions like , Can there be embeddings with non linear encoders and can achieve better performance?, If yes , under what conditions ?.Also our approach works only for deterministic embedding methods but we also need to analyze random walk based approach for graph embedding as well.Our future research work would be focusing on this aspect.

7 References

[1] James R munkres- Topology - fourth edition.

[2]Walter Rudin - Real and Complex analysis.

[3] Wlter Rudin - Functional Analysis .

[4]William L. Hamilton,rex ying, Jure Leskovec .Inductive Representation Learning on Graphs: Methods and Applications *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* .

[5]Shuai Zhang,Yi Tay,Qi Liu.Quaternion Knowledge Graph Embeddings*Advances in neural information processing systems* 2019.

[6]Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, pages 2787–2795, 2013.

[7]Seyed Mehran Kazemi and David Poole. Simple embedding for link prediction in knowledge graphs. *Advances in Neural Information Processing Systems*, pages 4289–4300, 2018.

[8]Takuma Ebisu and Ryutaro Ichise. Toruse: Knowledge graph embedding on a lie group. *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[9]Belkin, Mikhail, Niyogi, Partha, and Sindhwani, Vikas. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *JMLR,* 7:2399–2434, 2006.

[10]Zhilin Yang ,William W. Cohen , Ruslan Salakhutdinov.Revisiting Semi-Supervised Learning with Graph Embeddings*33 rd International Conference on Machine Learning*2016.

[11]Weston, Jason, Ratle, Frederic, Mobahi, Hossein, and Collobert, Ronan. Deep learning via semi-supervised embedding. *Neural Networks: Tricks of the Trade* pp.639–655. Springer, 2012

[12]Dijun Luo,Chris Ding,Feiping Nie,Heng Huang.Cauchy Graph Embedding*Proceedings of the 28 th International Conference on Machine Learning*

[13]B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. *KDD*, 2014

[14] Bryan Perozzi. Local Modeling of Attributed Graphs: Algorithms and Applications. *PhD thesis, Stony Brook University*, 2016

[15]S. Chang, W. Han, J. Tang, G. Qi, C.C. Aggarwal, and T.S. Huang. Heterogeneous network embedding via deep architectures. *KDD* 2015

[16]L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. *WSDM*, 2011

[17]H. Chen, B. Perozzi, Y. Hu, and S. Skiena. Harp: Hierarchical representation learning for networks. *arXiv preprint arXiv:1706.07845*, 2017

[18]David A. Cox, John Little, Donal O'Shea- Ideals, Varieties, and Algorithms : An Introduction to Computational Algebraic Geometry and Commutative Algebra*Springer, Undergraduate Texts in Mathematics , 4 edition*

[19] Allen Hatcher - Algebraic topology *Cambridge University Press*

[20]L. van der Maaten and G. Hinton. Visualizing data using t-sne. JMLR, 9:2579–2605, 2008

[21] Walter Rudin - Principles of Mathematical Analysis [22] Note on the convex hull of the Stiefel manifold - Kyle A Gallivan,P.A.Absil [23] Hiriart-Urruty, J., Malick, J. A Fresh Variational-Analysis Look at the Positive Semidefinite Matrices World. J Optim Theory Appl 153, 551–577 (2012). https://doi.org/10.1007/s10957-011-9980-6 [24]Hiriart-Urruty, J., Le, H.Y. Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function. Optim Lett 6, 841–849 (2012). https://doi.org/10.1007/s11590-011-0304-4. [25]F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model.*IEEE Transactions on Neural Networks,* 20(1):61–80, 2009.

A APPENDIX

You may include other additional sections here.