Deep Learning on Topological Spaces

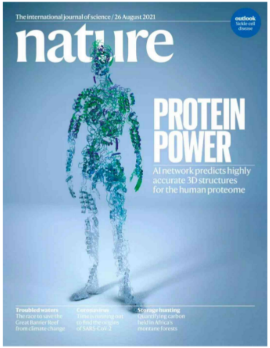
Cristian Bodnar
University of Cambridge

Based on works with F. Frasca, F. di Giovanni, Y.G. Wang, G. Montufar, N. Otter, B. Chamberlain, M. Bronstein, and P. Liò

Deep Exploration of non-Euclidean Data with Geometric and Topological Representation Learning
Kelowna, Canada
July 14, 2022

Geometric Deep Learning

Data often resides on structured domains: molecules, maps, social networks, . . .



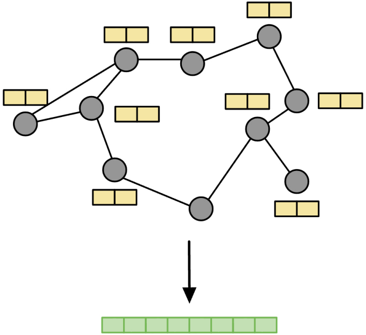
Graph Machine Learning

Data is a graph G with n nodes and a node feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$.

Graph Machine Learning

Data is a graph G with n nodes and a node feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$.

Graph-Level Tasks

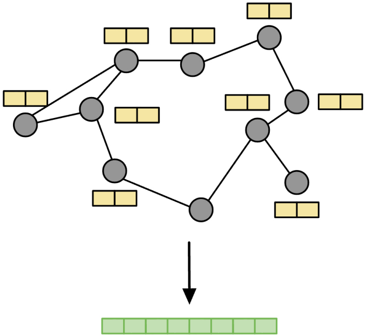


e.g. predicting solubility of molecules

Graph Machine Learning

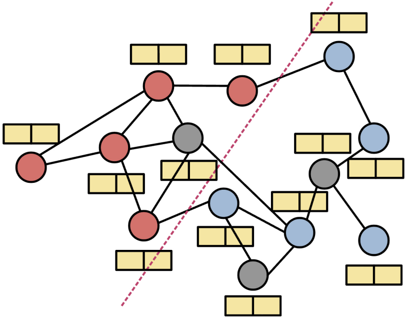
Data is a graph G with n nodes and a node feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$.

Graph-Level Tasks



e.g. predicting solubility of molecules

Node-level Tasks

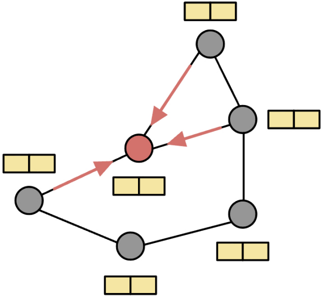


e.g. fraud detection

Message Passing Neural Networks

Most Graph Neural Networks (GNNs) can be understood as message passing:

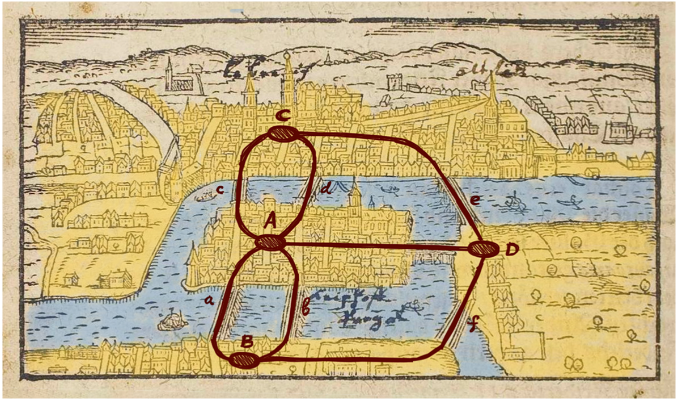
$$m_v^k := \mathrm{AGGREGATE}\left(\left\{h_u^{k-1} \mid u \in \mathcal{N}(v)\right\}\right) \qquad h_v^k := \mathrm{COMBINE}\left(h_v^{k-1}, m_v^k\right)$$



How to design AGGREGATE and COMBINE is a very active area of research.

Graphs and Topology

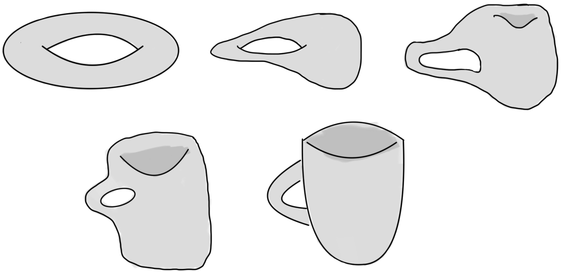
Graph Theory and Topology share a common history: Leonard Euler's paper *geometria situs* ("geometry of location")



Source: Wikipedia

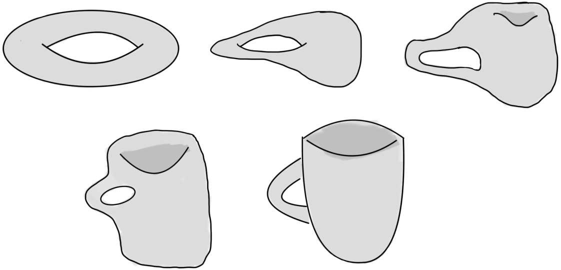
So what is Topology?

It's a branch of mathematics studying objects under continuous deformations. To a topologist, a mug and a donut are the same.



So what is Topology?

It's a branch of mathematics studying objects under continuous deformations. To a topologist, a mug and a donut are the same.

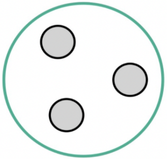


"Topology! The stratosphere of human thought! In the twenty-fourth century, it might possibly be of use to someone." — Aleksandr Solzhenitsyn, In the First Circle (1968)

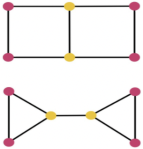
Topological Message Passing

Some problems with graph message passing

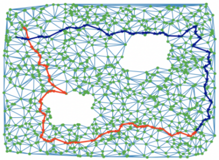
Groupwise interactions



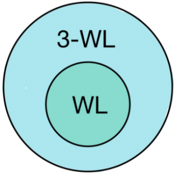
Higher-order structures



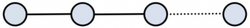
Higher-order signals



Expressive power

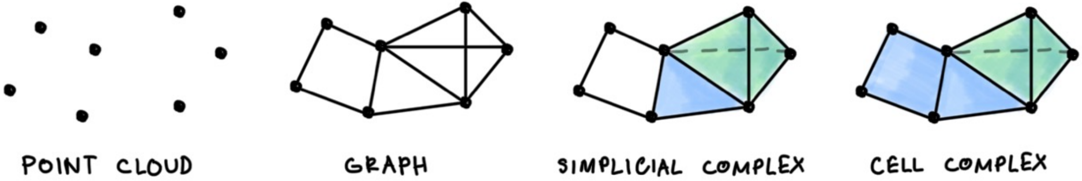


Long-range interactions



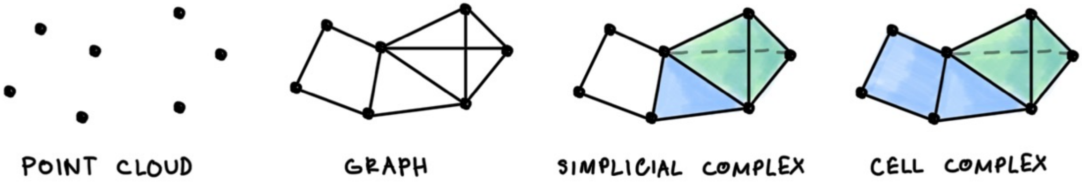
Beyond graphs

Graphs are just a type of topological space with combinatorial structure.



Beyond graphs

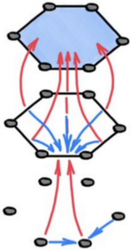
Graphs are just a type of topological space with combinatorial structure.



In this terminology, vertices are 0-cells, edges are 1-cells, 2D surfaces are 2-cells...

Message Passing on Cell Complexes

We can generalise message passing on graphs to cell complexes. We call this *topological message passing*.

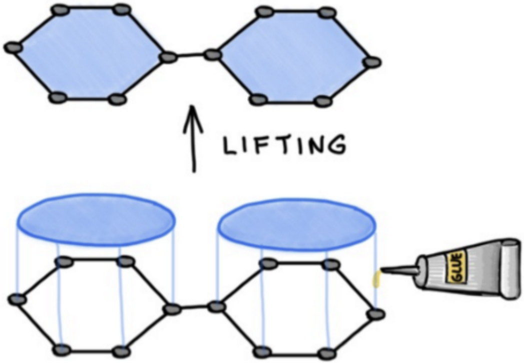


Constructing cell complexes

But how do we obtain a cell complex?

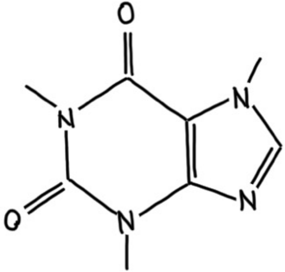
Constructing cell complexes

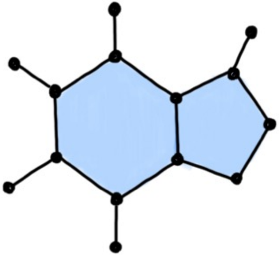
But how do we obtain a cell complex?



Domain alignment

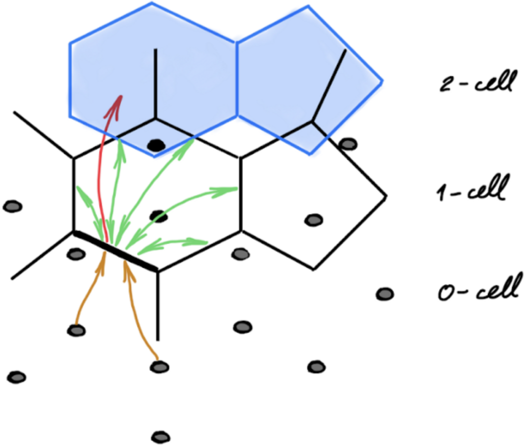
This type of space aligns well with certain applications.





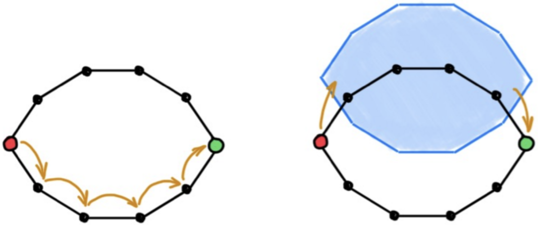
Hierarhical

The message passing becomes hierarchical.



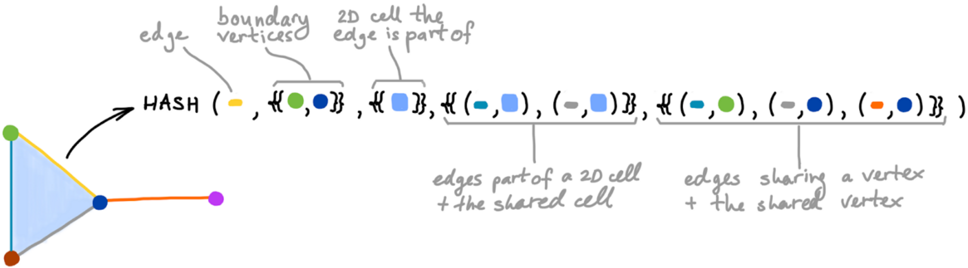
Long-range communication

Fewer layers are needed for long-range communication.



Expressivity

Topological Message Passing can distinguish more pairs of non-isomorphic graphs than message passing GNNs. This can be proven using a cellular version of the Weisfeiler-Lehman Test.



Results

Topological Message Passing can be applied to molecular property prediction.

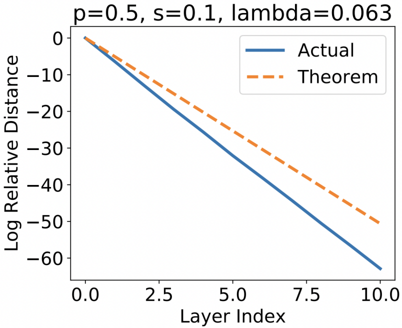
Table 3: ZINC (MAE), ZINC-FULL (MAE) and Mol-HIV (ROC-AUC).

Method	ZINC ↓ No Edge Feat.	ZINC ↓ With Edge Feat.	ZINC-FULL ↓ All methods	MOLHIV ↑ All methods
GCN [45]	0.469±0.002	N/A	N/A	76.06±0.97
GAT [67]	0.463±0.002	N/A	N/A	N/A
GatedGCN [10]	0.422±0.006	0.363±0.009	N/A	N/A
GIN [72]	0.408±0.008	0.252±0.014	0.088±0.002	77.07±1.49
PNA [19]	0.320±0.032	0.188±0.004	N/A	79.05±1.32
DGN [5]	0.219±0.010	0.168±0.003	N/A	79.70±0.97
HIMP [26]	N/A	0.151±0.006	0.036±0.002	78.80±0.82
GSN [9]	0.139±0.007	0.108±0.018	N/A	77.99±1.00
CIN-small (Ours)	0.139±0.008	0.094±0.004	0.044±0.003	80.55±1.04
CIN (Ours)	**0.115±0.003**	**0.079±0.006**	**0.022±0.002**	**80.94±0.57**

Neural Sheaf Diffusion

The Oversmoothing Problem

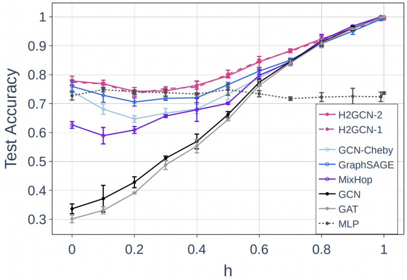
In some GNNs [Oono & Suzuki, 2020] features become progressively smoother with increased depth.



With more layers, GCN approaches a "smooth" subspace where all the node features are constant [Oono & Suzuki, 2020].

The Heterophily Problem

Numerous studies (e.g. [J. Zhu et al., 2020]) remarked that GNNs struggle in heterophilic settings (i.e. graphs where a node tends to be connected to nodes belonging to other classes).



The performance of GNNs is strongly correlated to the homophily level of a graph [J. Zhu et al., 2020].

From Heat Diffusion to Graph Convolutions

Let G be a graph with self loops, degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, normalised Laplacian $\Delta_0 := I - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and node features $\mathbf{X} \in \mathbb{R}^{n \times d}$.

From Heat Diffusion to Graph Convolutions

Let G be a graph with self loops, degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, normalised Laplacian $\Delta_0 := I - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and node features $\mathbf{X} \in \mathbb{R}^{n \times d}$.

We can define a *heat diffusion* process on graphs:

$$\dot{\mathbf{X}}(t) = -\Delta_0 \mathbf{X}(t)$$

From Heat Diffusion to Graph Convolutions

Let G be a graph with self loops, degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, normalised Laplacian $\Delta_0 := I - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and node features $\mathbf{X} \in \mathbb{R}^{n \times d}$.

We can define a *heat diffusion* process on graphs:

$$\dot{\mathbf{X}}(t) = -\Delta_0 \mathbf{X}(t) \rightsquigarrow \mathbf{X}(t+1) = \mathbf{X}(t) - \Delta_0 \mathbf{X}(t)$$

From Heat Diffusion to Graph Convolutions

Let G be a graph with self loops, degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, normalised Laplacian $\Delta_0 := I - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and node features $\mathbf{X} \in \mathbb{R}^{n \times d}$.

We can define a *heat diffusion* process on graphs:

$$\dot{\mathbf{X}}(t) = -\Delta_0 \mathbf{X}(t) \rightsquigarrow \mathbf{X}(t+1) = \mathbf{X}(t) - \Delta_0 \mathbf{X}(t) = \color{red}{(\mathbf{I} - \Delta_0)\mathbf{X}(t)} \tag{1}$$

From Heat Diffusion to Graph Convolutions

Let G be a graph with self loops, degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, normalised Laplacian $\Delta_0 := I - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and node features $\mathbf{X} \in \mathbb{R}^{n \times d}$.

We can define a *heat diffusion* process on graphs:

$$\dot{\mathbf{X}}(t) = -\Delta_0 \mathbf{X}(t) \rightsquigarrow \mathbf{X}(t+1) = \mathbf{X}(t) - \Delta_0 \mathbf{X}(t) = (\mathbf{I} - \Delta_0)\mathbf{X}(t) \tag{1}$$

This is strikingly similar to GCN [Kipf & Welling, 2017]:

$$\mathrm{GCN}(\mathbf{X}, \mathbf{A}) := \sigma\left(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{X}\mathbf{W}\right)$$

From Heat Diffusion to Graph Convolutions

Let G be a graph with self loops, degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, normalised Laplacian $\Delta_0 := I - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and node features $\mathbf{X} \in \mathbb{R}^{n \times d}$.

We can define a *heat diffusion* process on graphs:

$$\dot{\mathbf{X}}(t) = -\Delta_0 \mathbf{X}(t) \rightsquigarrow \mathbf{X}(t+1) = \mathbf{X}(t) - \Delta_0 \mathbf{X}(t) = (\mathbf{I} - \Delta_0)\mathbf{X}(t) \tag{1}$$

This is strikingly similar to GCN [Kipf & Welling, 2017]:

$$\mathrm{GCN}(\mathbf{X}, \mathbf{A}) := \sigma\left(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{X}\mathbf{W}\right) = \sigma\left((\mathbf{I} - \Delta_0)\mathbf{X}\mathbf{W}\right) \tag{2}$$

From Heat Diffusion to Graph Convolutions

Let G be a graph with self loops, degree matrix $\mathbf{D}$, adjacency matrix $\mathbf{A}$, normalised Laplacian $\Delta_0 := I - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and node features $\mathbf{X} \in \mathbb{R}^{n\times d}$.

We can define a *heat diffusion* process on graphs:

$$\dot{\mathbf{X}}(t) = -\Delta_0\mathbf{X}(t) \rightsquigarrow \mathbf{X}(t+1) = \mathbf{X}(t) - \Delta_0\mathbf{X}(t) = (\mathbf{I} - \Delta_0)\mathbf{X}(t) \tag{1}$$

This is strikingly similar to GCN [Kipf & Welling, 2017]:

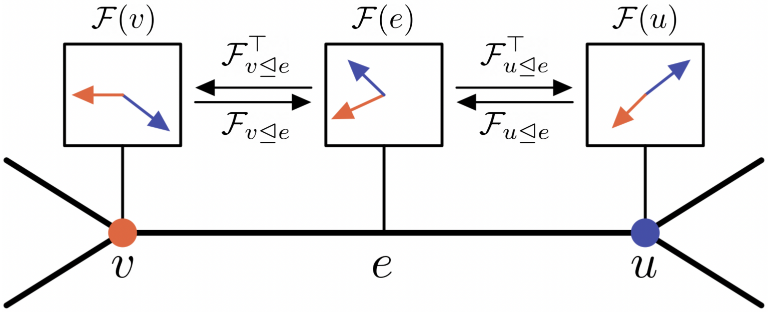
$$\mathrm{GCN}(\mathbf{X}, \mathbf{A}) := \sigma\Big(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{X}\mathbf{W}\Big) = \sigma\Big((\mathbf{I} - \Delta_0)\mathbf{X}\mathbf{W}\Big) \tag{2}$$

Question

How can we make the base diffusion process more powerful?

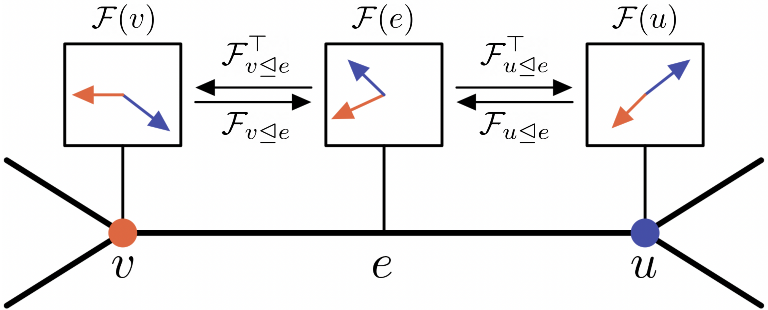
Cellular Sheaves

Cellular sheaves [Curry, 2014] are mathematical objects that attach some additional geometric structure to a graph.



Cellular Sheaves

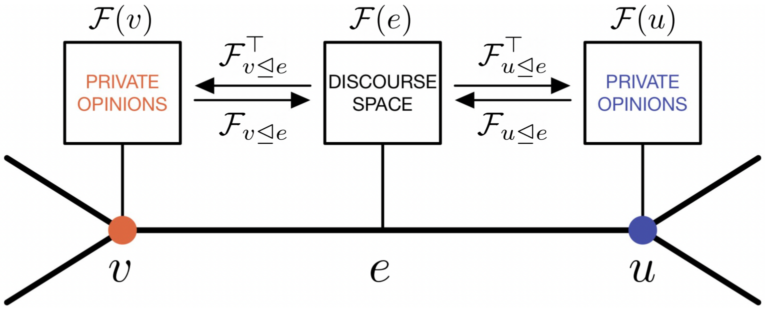
Cellular sheaves [Curry, 2014] are mathematical objects that attach some additional geometric structure to a graph.



Sheaves were invented by the French mathematician Jean Leray during WWII.

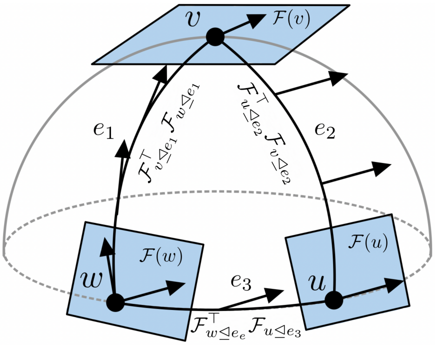
Opinion Dynamics

Opinion dynamics [Hansen & Ghrist, 2020] provides a nice mental picture of cellular sheaves.



Discrete Vector Bundles

The sheaves with orthogonal restriction maps are *discrete $O(d)$-bundles*.



Analogy between parallel transport on a sphere and transport on a discrete vector bundle. A tangent vector is moved from $\mathcal{F}(w) \to \mathcal{F}(v) \to \mathcal{F}(u)$ and back.

Expressive Power of Sheaf Diffusion

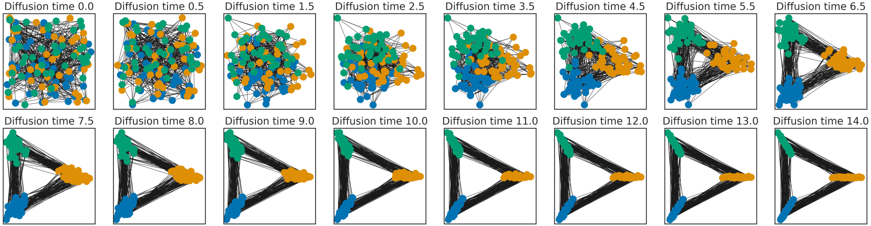
We are analysing the expressive power of the following *sheaf diffusion* PDE:

$$\mathbf{X}(0) = \mathbf{X}, \quad \dot{\mathbf{X}}(t) = -\Delta_{\mathcal{F}}\mathbf{X}(t)$$

Expressive Power of Sheaf Diffusion

We are analysing the expressive power of the following *sheaf diffusion* PDE:

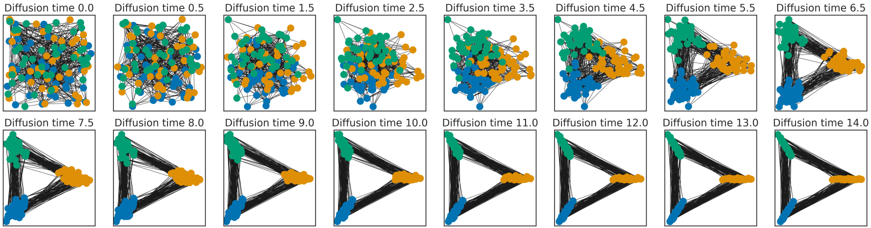
$$\mathbf{X}(0) = \mathbf{X}, \quad \dot{\mathbf{X}}(t) = -\Delta_{\mathcal{F}}\mathbf{X}(t)$$



Expressive Power of Sheaf Diffusion

We are analysing the expressive power of the following *sheaf diffusion* PDE:

$$\mathbf{X}(0) = \mathbf{X}, \quad \dot{\mathbf{X}}(t) = -\Delta_{\mathcal{F}}\mathbf{X}(t)$$



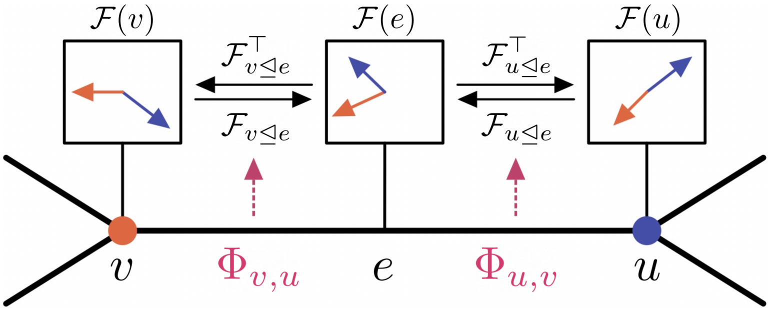
Theorem (Informal)

For any node classification task on a connected graph, there exists a sheaf that can linearly separate the classes by performing diffusion infinitely long.

Learning sheaves

Each $d \times d$ matrix $\mathcal{F}_{v \trianglelefteq e}$ is learned via a parametric function $\Phi : \mathbb{R}^{d \times 2} \to \mathbb{R}^{d \times d}$:

$$\mathcal{F}_{v \trianglelefteq e := (v,u)} = \Phi(\mathbf{x}_v, \mathbf{x}_u) \tag{3}$$



The restriction maps are learned from data.

Real-World Evaluation

We evaluate on multiple node-classifications tasks with various degrees of homophily [Rozemberczki et al, 2019, Pei et al, 2019].

	Texas	Wisconsin	Film	Squirrel	Chameleon	Cornell	Citeseer	Pubmed	Cora
Hom level	**0.11**	**0.21**	**0.22**	**0.22**	**0.23**	**0.30**	**0.74**	**0.80**	**0.81**
#Nodes	183	251	7,600	5,201	2,277	183	3,327	18,717	2,708
#Edges	295	466	26,752	198,493	31,421	280	4,676	44,327	5,278
#Classes	5	5	5	5	5	5	7	3	6
Diag-NSD	**85.67**±6.95	**88.63**±2.75	**37.79**±1.01	**54.78**±1.81	**68.68**±1.73	**86.49**±7.35	**77.14**±1.85	89.42±0.43	87.14±1.06
O(d)-NSD	**85.95**±5.51	**89.41**±4.74	**37.81**±1.15	**56.34**±1.32	**68.04**±1.58	**84.86**±4.71	76.70±1.57	**89.49**±0.40	86.90±1.13
Gen-NSD	82.97±5.13	**89.21**±3.84	**37.80**±1.22	53.17±1.31	67.93±1.58	**85.68**±6.51	76.32±1.65	89.33±0.35	87.30±1.15
GGCN	**84.86**±4.55	86.86±3.29	37.54±1.56	**55.17**±1.58	**71.14**±1.84	**85.68**±6.63	**77.14**±1.45	89.15±0.37	**87.95**±1.05
H2GCN	**84.86**±7.23	87.65±4.98	35.70±1.00	36.48±1.86	60.11±2.15	82.70±5.28	77.11±1.57	**89.49**±0.38	**87.87**±1.20
GPRGNN	78.38±4.36	82.94±4.21	34.63±1.22	31.61±1.24	46.58±1.71	80.27±8.11	77.13±1.67	87.54±0.38	**87.95**±1.18
FAGCN	82.43±6.89	82.94±7.95	34.87±1.25	42.59±0.79	55.22±3.19	79.19±9.79	N/A	N/A	N/A
MixHop	77.84±7.73	75.88±4.90	32.22±2.34	43.80±1.48	60.50±2.53	73.51±6.34	76.26±1.33	85.31±0.61	87.61±0.85
GCNII	77.57±3.83	80.39±3.40	37.44±1.30	38.47±1.58	63.86±3.04	77.86±3.79	**77.33**±1.48	**90.15**±0.43	**88.37**±1.25
Geom-GCN	66.76±2.72	64.51±3.66	31.59±1.15	38.15±0.92	60.00±2.81	60.54±3.67	**78.02**±1.15	**89.95**±0.47	85.35±1.57
PairNorm	60.27±4.34	48.43±6.14	27.40±1.24	50.44±2.04	62.74±2.82	58.92±3.15	73.59±1.47	87.53±0.44	85.79±1.01
GraphSAGE	82.43±6.14	81.18±5.56	34.23±0.99	41.61±0.74	58.73±1.68	75.95±5.01	76.04±1.30	88.45±0.50	86.90±1.04
GCN	55.14±5.16	51.76±3.06	27.32±1.10	53.43±2.01	64.82±2.24	60.54±5.30	76.50±1.36	88.42±0.50	86.98±1.27
GAT	52.16±6.63	49.41±4.09	27.44±0.89	40.72±1.55	60.26±2.50	61.89±5.05	76.55±1.23	87.30±1.10	86.33±0.48
MLP	80.81±4.75	85.29±3.31	36.53±0.70	28.77±1.56	46.21±2.99	81.89±6.40	74.02±1.90	75.69±2.00	87.16±0.37

References

Cristian Bodnar*, Fabrizio Frasca*, Yu Guang Wang*, Nina Otter, Guido Montúfar*, Pietro Liò and Michael Bronstein (2021)
Weisfeiler and Lehman Go Topological: Message Passing Simplicial Networks
ICML 2021

Cristian Bodnar*, Fabrizio Frasca*, Nina Otter, Yu Guang Wang, Pietro Liò, Guido Montúfar and Michael Bronstein (2021)
Weisfeiler and Lehman Go Cellular: CW Networks
NeurIPS 2021

Vidit Nanda
Computational Algebraic Topology
University of Oxford Course, 2020-2021 - `https://people.maths.ox.ac.uk/nanda/cat/`

Keenan Crane
Discrete Differential Geometry
CMU Course, Spring 2021 - `https://brickisland.net/DDGSpring2021/`

References

Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Liò, Michael M. Bronstein (2022)
Neural Sheaf Diffusion: A Topological Perspective on Heterophily and Oversmoothing in GNNs
Preprint 2022

Thomas N. Kipf, Max Welling
Semi-Supervised Classification with Graph Convolutional Networks
ICLR 2017

Kenta Oono, Taiji Suzuki (2020)
Graph Neural Networks Exponentially Lose Expressive Power for Node Classification
ICLR 2020

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, Danai Koutra
Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs
NeurIPS 2020

References

Justin Curry
Sheaves, cosheaves and applications
PhD Dissertation, University of Pennsylvania, 2014

Jakob Hansen, Robert Ghrist
Opinion Dynamics on Discourse Sheaves
SIAM J. Appl. Math., 81(5), 2033–2060

Jakob Hansen, Robert Ghrist
Toward a Spectral Theory of Cellular Sheaves
Journal of Applied and Computational Topology volume 3, pages 315–358 (2019)

Yujun Yan, Milad Hashemi, Kevin Swersky, Yaoqing Yang, Danai Koutra
Two Sides of the Same Coin: Heterophily and Oversmoothing in Graph Convolutional Neural Networks
Preprint (2021)

References

Eli Chien, Jianhao Peng, Pan Li, Olgica Milenkovic
Adaptive Universal Generalized PageRank Graph Neural Network
ICLR 2021

Chen Cai, Yusu Wang
A Note on Over-Smoothing for Graph Neural Networks
ICML 2020 Graph Representation Learning Workshop

Jakob Hansen, Thomas Gebhart
Sheaf Neural Networks
NeurIPS 2020 Workshop on TDA and Beyond

Benedek Rozemberczki, Carl Allen, Rik Sarkar
Multi-scale Attributed Node Embedding
Journal of Complex Networks

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, Bo Yang
Geom-GCN: Geometric Graph Convolutional Networks
ICLR 2020

Thank you for your attention!

Email: `cb2015@cam.ac.uk`
Twitter: `@crisbodnar`