Math-for-industry
Education & Research Hub

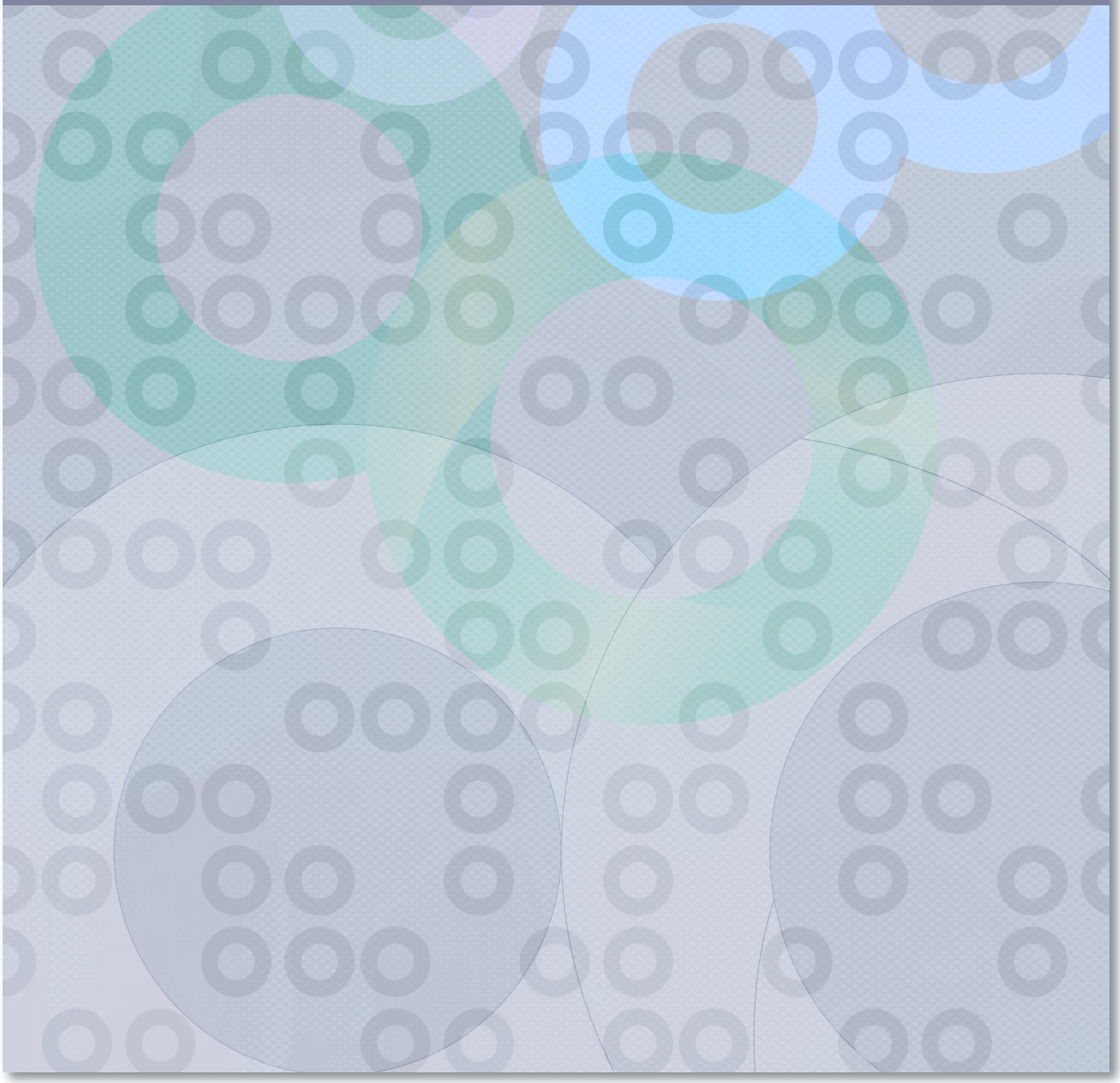
ISSN2188-1200

Young Researchers and Students-Workshop (I)

Topological Data Analysis and Industrial Mathematics

Editor: Keunsu Kim

九州大学マス・フォア・インダストリ研究所



MI Lecture Note Vol.102 : Kyushu University

Young Researchers and Students-Workshop (I)

Topological Data Analysis and Industrial Mathematics

Editor Keunsu Kim

About MI Lecture Note Series

The Math-for-Industry (MI) Lecture Note Series is the successor to the COE Lecture Notes, which were published for the 21st COE Program "Development of Dynamic Mathematics with High Functionality," sponsored by Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) from 2003 to 2007. The MI Lecture Note Series has published the notes of lectures organized under the following two programs: "Training Program for Ph.D. and New Master's Degree in Mathematics as Required by Industry," adopted as a Support Program for Improving Graduate School Education by MEXT from 2007 to 2009; and "Education-and-Research Hub for Mathematics-for-Industry," adopted as a Global COE Program by MEXT from 2008 to 2012.

In accordance with the establishment of the Institute of Mathematics for Industry (IMI) in April 2011 and the authorization of IMI's Joint Research Center for Advanced and Fundamental Mathematics-for-Industry as a MEXT Joint Usage / Research Center in April 2013, hereafter the MI Lecture Notes Series will publish lecture notes and proceedings by worldwide researchers of MI to contribute to the development of MI.

October 2022
Kenji Kajiwara
Director, Institute of Mathematics for Industry

Topological Data Analysis and Industrial Mathematics

MI Lecture Note Vol.102, Institute of Mathematics for Industry, Kyushu University
ISSN 2188-1200
Date of issue: December 22, 2025
Editor: Keunsu Kim
Publisher:
Institute of Mathematics for Industry, Kyushu University
Graduate School of Mathematics, Kyushu University
Motooka 744, Nishi-ku, Fukuoka, 819-0395, JAPAN
Tel +81-(0)92-802-4402, Fax +81-(0)92-802-4405
URL https://www.imi.kyushu-u.ac.jp/

Preface

This lecture note is a collection of slides presented at the workshop “Topological Data Analysis and Industrial Mathematics”, held at Nishijin Plaza, Fukuoka, Japan, from August 5 to 8, 2025. The workshop was organized by the Institute of Mathematics for Industry (IMI), Kyushu University, and supported by JSPS KAKENHI, Grant-in-Aid for Scientific Research (B) 25K00921 and (S) 25H00399, NICT commissioned research No.22301, and the JST Moonshot R&D Program (JPMJMS2021).

The aim of this workshop was to bring together researchers from academia and industry to discuss recent developments and applications of Topological Data Analysis (TDA) in the fields of science, engineering, and industrial mathematics. By bridging topology, geometry, and data-driven modeling, the workshop explored how mathematical frameworks can promote innovation in complex systems, biomedical sciences, and other related areas.

The program was organized around the following three main themes:

- Applications and theoretical extensions of persistent homology
- Applications of TDA to industrial and biomedical sciences
- Meta-level discussions on the role of mathematics for industry

Through these interdisciplinary discussions, this workshop provided a platform for mathematicians, engineers, and data scientists to collaborate and advance both the mathematical foundations and practical applications of topological data analysis.

Organizing Committee Chair: Keunsu Kim

Organizing Committee Members

- Keunsu Kim (Kyushu University)
- Matias de Jong van Lier (Kyushu University)
- Shizuo Kaji (Kyushu University / Kyoto University)
- Jae-Hun Jung (POSTECH/POSTECH MINDS)

2025年度九州大学マス・フォア・インダストリ研究所 共同利用・共同研究 若手・学生研究・研究集会（I） 対面開催

Topological Data Analysis and Industrial Mathematics: Bridging Theory and Applications

Aug. 5 (Tuesday) ~ Aug. 8 (Friday), 2025

Nishijin Plaza, Fukuoka, Japan

URL: https://sites.google.com/view/tda-im-2025

Organizer:

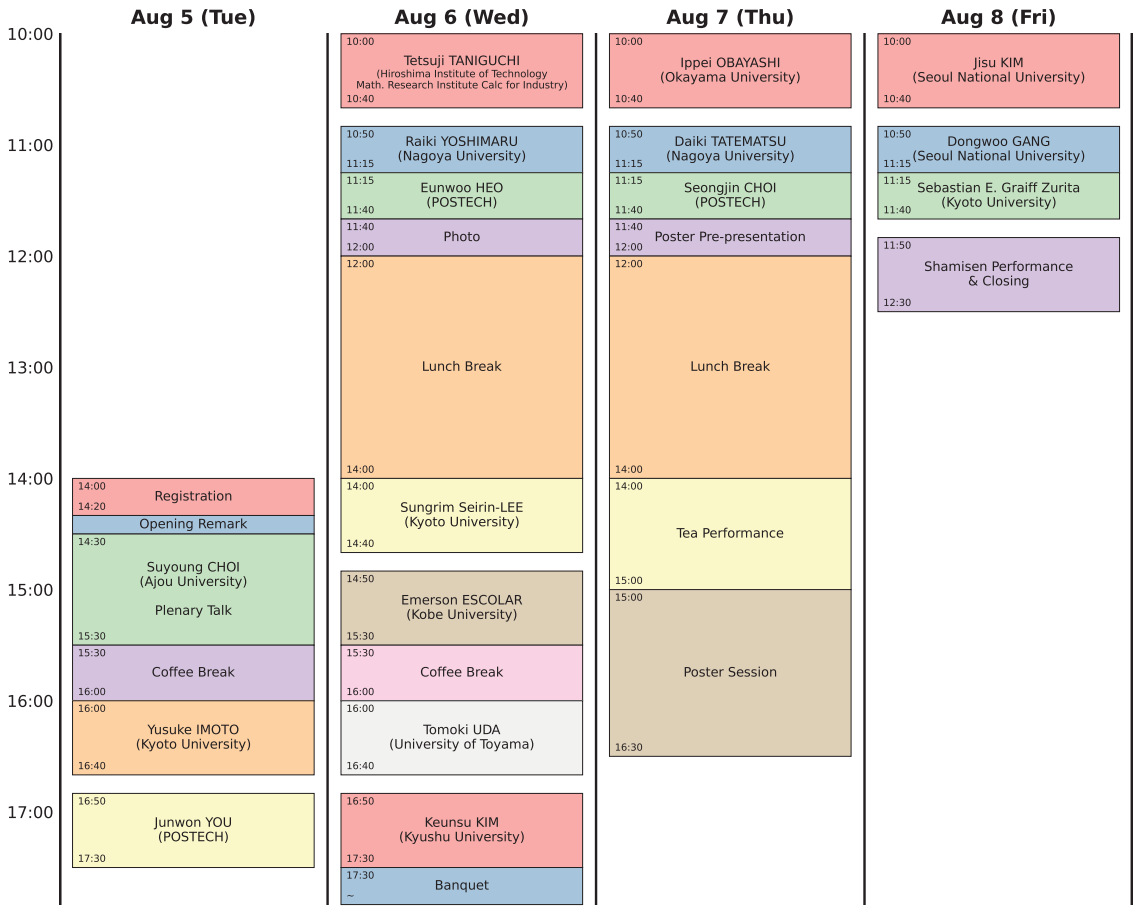
Keunsu Kim (Kyushu University)

Matías de Jong van Lier (Kyushu University)

Shizuo Kaji (Kyushu / Kyoto University)

Jae-Hun Jung (POSTECH / POSTECH MINDS)

問い合わせ先: 共同利用・共同研究拠点事務室　主催：九州大学マス・フォア・インダストリ研究所


Aug 5 (Tue)
Aug 6 (Wed)
Aug 7 (Thu)
Aug 8 (Fri)
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
14:00 Registration 14:20
Opening Remark
14:30 Suyoung CHOI (Ajou University) Plenary Talk 15:30
15:30 Coffee Break 16:00
16:00 Yusuke IMOTO (Kyoto University) 16:40
16:50 Junwon YOU (POSTECH) 17:30
10:00 Tetsuji TANIGUCHI (Hiroshima Institute of Technology Math. Research Institute Calc for Industry) 10:40
10:50 Raiki YOSHIMARU (Nagoya University) 11:15
11:15 Eunwoo HEO (POSTECH) 11:40
11:40 Photo 12:00
12:00 Lunch Break 14:00
14:00 Sungrim Seirin-LEE (Kyoto University) 14:40
14:50 Emerson ESCOLAR (Kobe University) 15:30
15:30 Coffee Break 16:00
16:00 Tomoki UDA (University of Toyama) 16:40
16:50 Keunsu KIM (Kyushu University) 17:30
17:30 Banquet ~
10:00 Ippei OBAYASHI (Okayama University) 10:40
10:50 Daiki TATEMATSU (Nagoya University) 11:15
11:15 Seongjin CHOI (POSTECH) 11:40
11:40 Poster Pre-presentation 12:00
12:00 Lunch Break 14:00
14:00 Tea Performance 15:00
15:00 Poster Session 16:30
10:00 Jisu KIM (Seoul National University) 10:40
10:50 Dongwoo GANG (Seoul National University) 11:15
11:15 Sebastian E. Graiff Zurita (Kyoto University) 11:40
11:50 Shamisen Performance & Closing 12:30

Program

Day 1 (Aug 5, Tue)

- 14:00 – 14:20, Registration
- 14:20 – 14:30, Opening Remark
- 14:30 – 15:30, Suyoung CHOI (Ajou University) – Plenary Talk
- 15:30 – 16:00, Coffee Break
- 16:00 – 16:40, Yusuke IMOTO (Kyoto University)
- 16:50 – 17:30, Junwon YOU (POSTECH)

Day 2 (Aug 6, Wed)

- 10:00 – 10:40, Tetsuji TANIGUCHI (Hiroshima Institute of Technology / Math. Research Institute Calc for Industry)
- 10:50 – 11:15, Raiki YOSHIMURA (Nagoya University)
- 11:15 – 11:40, Eunwoo HEO (POSTECH)
- 11:40 – 12:00, Photo
- 12:00 – 14:00, Lunch Break
- 14:00 – 14:40, Sungrim Seirin-LEE (Kyoto University)
- 14:50 – 15:30, Emerson ESCOLAR (Kobe University)
- 15:30 – 16:00, Coffee Break
- 16:00 – 16:40, Tomoki UDA (University of Toyama)
- 16:50 – 17:30, Keunsu KIM (Kyushu University)
- **17:30 – , Banquet**

  The banquet is supported by POSTECH MINDS (PI: Prof. Jae-Hun Jung).

Day 3 (Aug 7, Thu)

- 10:00 – 10:40, Ippei OBAYASHI (Okayama University)
- 10:50 – 11:15, Daiki TATEMATSU (Nagoya University)
- 11:15 – 11:40, Seongjin CHOI (POSTECH)
- 11:40 – 12:00, Poster pre-persentation
- 12:00 – 14:00, Lunch Break
- 14:00 – 15:00, Tea Performance
- 15:00 – 16:30, Poster Session

Day 4 (Aug 8, Fri)

- 10:00 – 10:40, Jisu KIM (Seoul National University)
- 10:50 – 11:15, Dongwoo GANG (Seoul National University)
- 11:15 – 11:40, Sebastian Elias GRAIFF ZURITA (Kyoto University)
- 11:50 – 12:30, Shamisen Performance and Closing

Contents

POSTER

Topological Data Analysis for Non-Destructive Testing in Civil Engineering

Suyoung Choi

Department of Mathematics, Ajou University, Korea

We explore the potential applications of industrial mathematics in civil engineering, with a focus on non-destructive testing (NDT). In particular, we introduce topological data analysis (TDA) techniques that can enhance model performance when analyzing ground-penetrating radar (GPR) survey images.

References.

[1] Meiyan Kang et al. "A Novel Shape-Aware Topological Representation for GPR Data with DNN Integration". In: *arXiv preprint arXiv:2506.06311* (2025).

[2] Jianwei Lei et al. "GPR detection localization of underground structures based on deep learning and reverse time migration". In: *NDT & E International* 143 (2024), p. 103043.

[3] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[4] Shaoqing Ren et al. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems* 28 (2015).

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

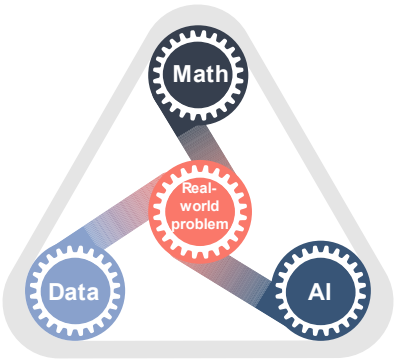
Topological Data Analysis for Non-Destructive Testing in Civil Engineering

2025 Topological Data Analysis and Industrial Math
2025. 8. 5

Suyoung Choi (Ajou Univ.)

Industrial Mathematics

How Do We Solve Industrial Problems Using Math?



AIM: Solve Real world Problem

In the era of the Fourth Industrial Revolution, the integration of **Data**, **Mathematics**, and **Artificial Intelligence (AI)** enables industries to address complex, real-world.

DATA

Collect real-time industrial data from sensors, machines, and digital systems to reflect the physical environment and operational states.

AI

Use artificial intelligence to learn patterns from data and make automated decisions for optimization, fault detection, and predictive maintenance.

Math

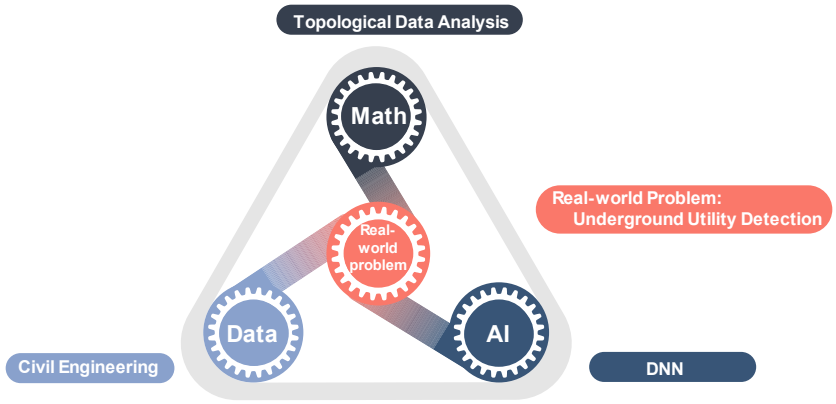
Translate data into interpretable mathematical models that support analysis, simulation, forecasting, and algorithmic reasoning.

Civil Engineering

The planning, design, construction, and maintenance of the built and natural environment including infrastructure such as roads, bridges, canals, dams, airports, pipelines, sewage systems, buildings, and railways.

L. Euler was also one of pioneers of civil engineering.

Industrial Mathematics



Urban problems

- Rapid urbanization increases **underground safety concerns.**
- Many underground facilities are **poorly documented** or **mismatched** with design drawings.
- **Accurate underground information** is essential for safe and efficient urban development.

Urban problems

Industrial Problem – Underground Utility Detection

Underground Utilities: water pipes, gas lines, electrical cables …

Water

Gas

Electrical

Impact our daily lives!

Problems:

Locating and identifying utilities are essential and necessary!

Underground exploration in Korea

- Implementation plan for disaster and safety management technology development 2016

(2016년 재난 및 안전관리기술개발 시행계획)

- Special law on Underground Safety Management 2022.1. 28

(지하안전관리에 관한 특별법)

Developers planning large-scale underground excavation projects, as defined by government regulation, are required to conduct an underground safety assessment.

(대통령령으로 정하는 규모 이상의 지하 굴착공사를 수반하는 사업을 하려는 지하개발사업자는 반드시 지하안전평가를 실시 해야함)

NDT (nondestructive testing)

"A method of inspecting an object without transforming it when you want to know the internal defects or components of the object"

- Does not permanently alter the article being inspected
- Save both money and time in product evaluation, troubleshooting and research
- Can be used to detect flaws in an in-process machine part

NDT (nondestructive testing)

Is it possible to find out the type and characteristics of underground burial without digging the ground?

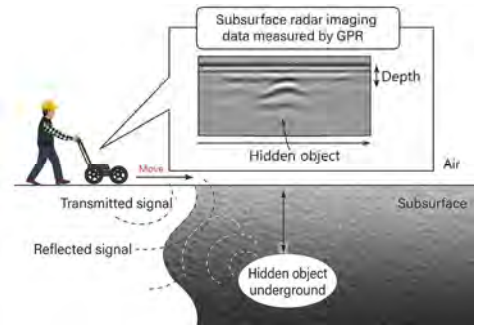
GPR 레이더 탐사
GPRレーダー
GPR radar

전기비저항 탐사
電気探査(比抵抗法)
Electrical Resistivity Surveys

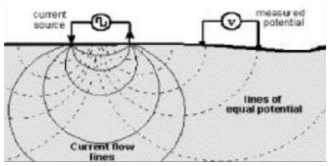
충격반향기법 탐사
衝撃弾性波法
Impact Echo

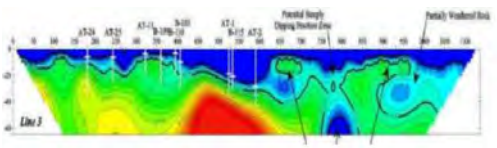
GPR (Ground-penetrating radar)

"A geophysical method that uses radar pulses to image the subsurface"



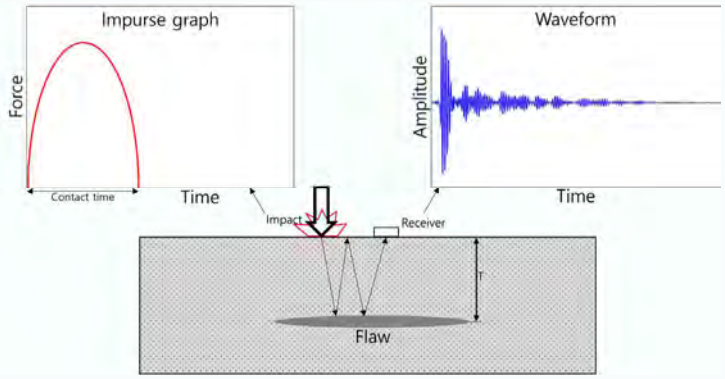
Electrical Resistivity Surveys





geophysical method used to investigate subsurface structures by measuring the soil's resistance to electrical current

Impact-echo method



technique that uses sound waves from small impacts to find cracks or voids inside concrete structures.

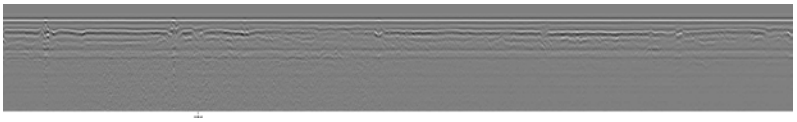
Proj 1 : Pipe detection (GPR)

supported by
KEPCO Research Institute (KEPRI)

GPR (Ground-penetrating radar)

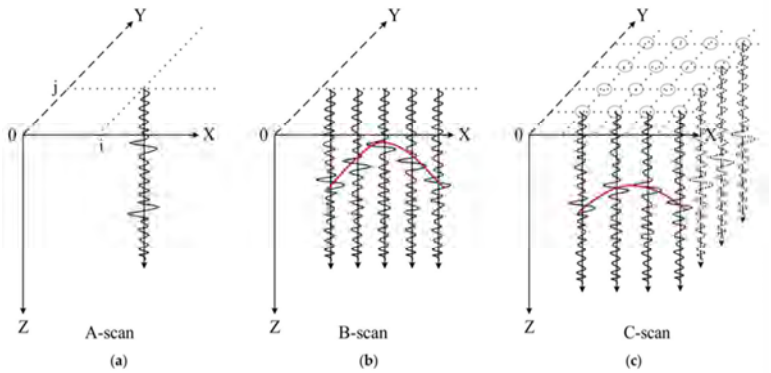
Project Goal:

To detect underground pipes or cavities using GPR data



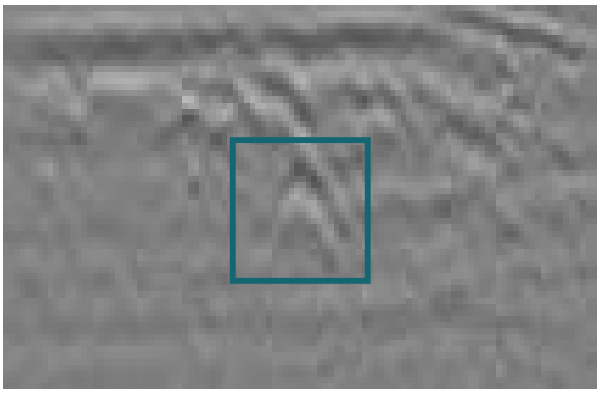
Proj **1 : Pipe detection (GPR)**

GPR (Ground-penetrating radar)



Proj **1 : Pipe detection (GPR)**

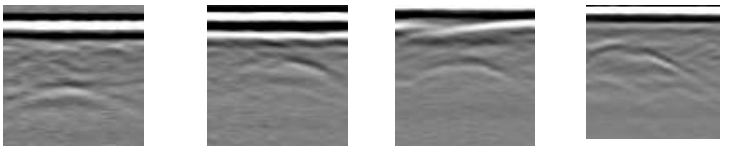
GPR (Ground-penetrating radar) : B-scan image



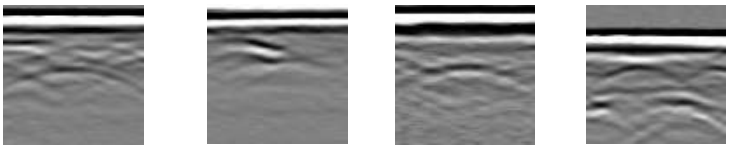
Proj **1 : Pipe detection (GPR)**

Cavity

Pipe

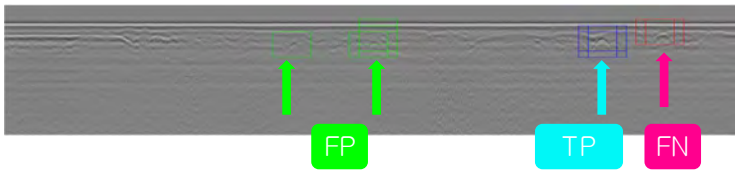


Cavities



Not cavities but similar images

The Risk of False Predictions



- FP (False Positive) : Detecting a pipe **where there is none**
  → Leads to **wasted time, cost, and unnecessary digging**
- FN(False Negative) : **Failing to detect** an actual pipe
  → Can cause **serious safety hazards**, such as accidents or infrastructure damage

20

How to improve?

1. Sim2Real (Simulation to Reality)

- Train models using simulated underground scenarios.
- Transfer knowledge to real GPR data through domain adaptation.
- Reduce data collection cost and improve model robustness.

2. TDA (Topological Data Analysis)

- Extract shape-based features from GPR images.
- Identify persistent topological structures that correspond to underground objects.
- Improve interpretability and detection accuracy.

Why Sim2Real?

1. Needs

- Real GPR data is hard to obtain
- Ground truth is uncertain, even with field data
- On-site surveys are costly and require road closures

(Reference image)

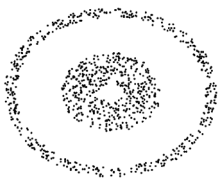
2. Challenges

- Simulated GPR data is often unrealistic
- Simulation is time-consuming to generate

Why TDA?

1. Needs

- Cavity patterns in GPR data are typically homeomorphic to S^1-shaped
- Shapes may distort due to noise, but TDA is stable under such perturbations



2. Challenges

- No straightforward way to apply TDA directly to image data
- Too many 1-cylces appear in persistence diagrams, making interpretation difficult

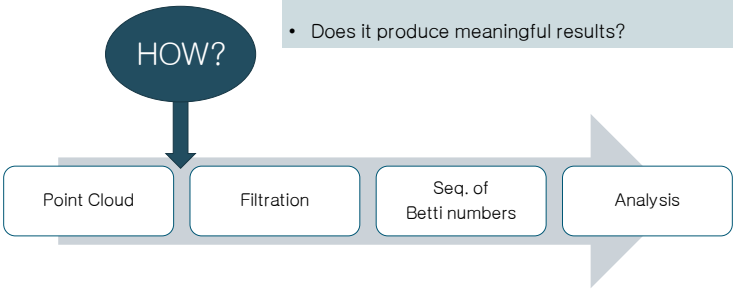
Data Generation and Collection

(A) Simulated B-scan image

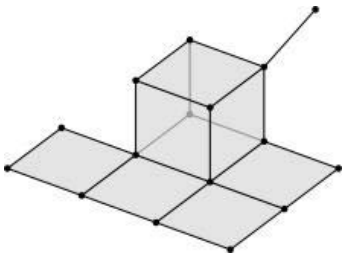
(B) Real-world field B-scan image

Doing TDA

- Does it make sense?
- Is it computable?
- Does it produce meaningful results?



Cubical Complex



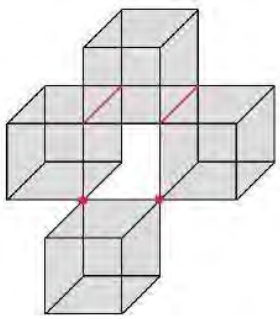
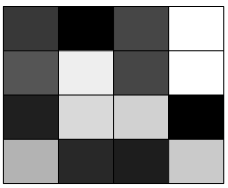
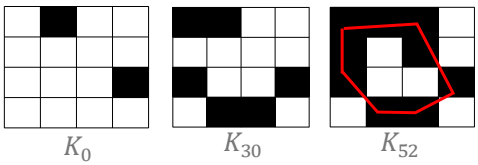


Image to cubical complex

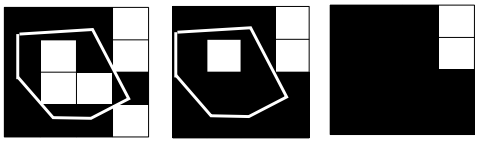


30	0	40	255
52	230	40	255
15	200	190	0
150	20	14	180

Give thresholds



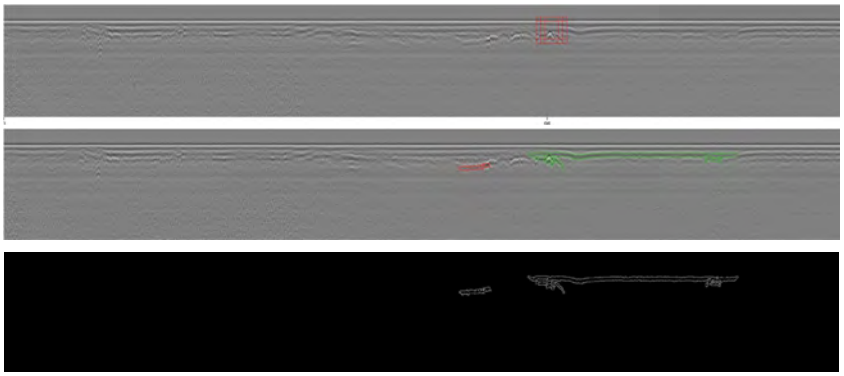
K_0 K_{30} K_{52}



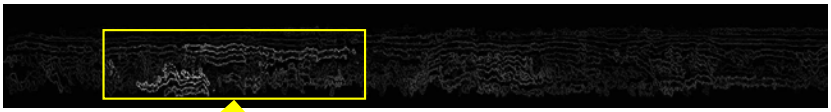
K_{150} K_{200} K_{230}

Proj **1 : Pipe detection (GPR)**

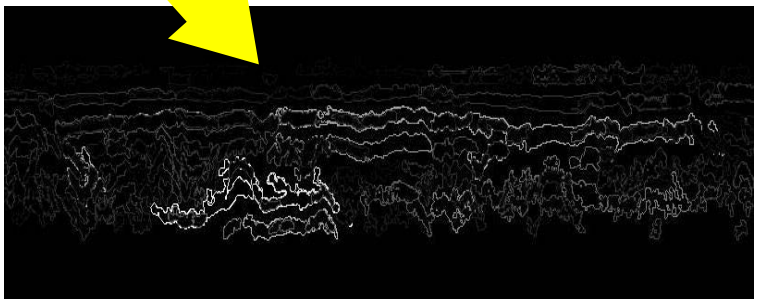
Observation



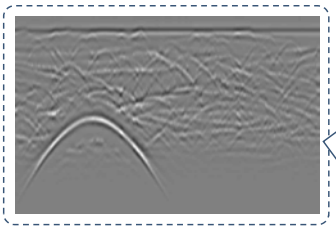
Proj **1 : Pipe detection (GPR)**



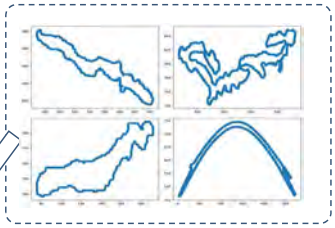
Colored by the life time of each generator of β_1

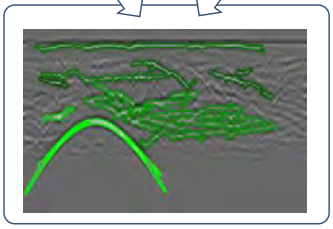


Proj 1 **: Pipe Detection (GPR)**



Threshold: 0.30





Lifetime-weighted topological representation for object detection.

Thank you for your attention!

Suyoung Choi (Ajou Univ.)
schoi@ajou.ac.kr

RNA Landscape Analysis via Combinatorial Hodge Decomposition

Yusuke Imoto

Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Japan

Cell differentiation can be conceptualized as movement on Waddington's epigenetic landscape, yet reconstructing this landscape from high-dimensional single-cell data remains challenging. Here, we propose landscape analysis, a novel framework for single-cell RNA-seq data that reconstructs an RNA landscape, which is Waddington's landscape-like structure, and performs downstream dynamical analysis utilizing this landscape. Single-cell RNA-seq measures transcript levels for approximately 20,000 genes per cell, producing a high-dimensional expression matrix. By applying RNA velocity, we convert these static profiles into vectors that predict each cell's future transcriptional trajectory. We then perform Hodge decomposition on this velocity field to extract the potential that forms the gradient component. The resulting potential surface defines the RNA landscape's height. Finally, by geometrically or statistically analyzing the potential, we derive biologically meaningful insights such as single-cell trajectories, time-resolved differential expression dynamics, and gene functions in cell differentiation. We applied landscape analysis to time-series scRNA-seq data of the PGCLC induction system, identifying differentiation pathways and candidate genes driving induction.

References.

[1] Roberto Buizza and Tim N Palmer. "The singular-vector structure of the atmospheric global circulation". In: *Journal of Atmospheric Sciences* 52.9 (1995), pp. 1434–1456.

[2] Yusuke Imoto and Yasuaki Hiraoka. "V-mapper: Topological data analysis for high-dimensional data with velocity". In: *Nonlinear Theory and Its Applications, IEICE* 14.2 (2023), pp. 92–105.

[3] Martin Leutbecher and Tim N Palmer. "Ensemble forecasting". In: *Journal of computational physics* 227.7 (2008), pp. 3515–3539.

[4] Lek-Heng Lim. "Hodge Laplacians on graphs". In: *Siam Review* 62.3 (2020), pp. 685–715.

[5] Pascal Oettli et al. "Meteorological Landscape of Tropical Cyclone". In: *EGUsphere* 2025 (2025), pp. 1–28.

PHLP: Interpretable Link Prediction via Persistent Homology and Its Extension to Knowledge Graph Completion

Junwon You

Basic Science Research Institute, POSTECH, Korea

We introduce PHLP, a novel and interpretable link prediction framework that utilizes persistent homology to extract topological features from local subgraphs. Unlike conventional GNN-based methods, PHLP offers a transparent feature extraction process that captures topological patterns underlying graph connectivity. PHLP achieves near–state-of-the-art performance across standard benchmarks without relying on GNNs. We also briefly present preliminary results on extending PHLP to knowledge graph completion, demonstrating its potential in capturing relational patterns through topological representations. This research conducted together with Eunwoo Heo and Jae-Hun Jung.

References.

[1] Lada A Adamic and Eytan Adar. "Friends and neighbors on the web". In: *Social networks* 25.3 (2003), pp. 211–230.

[2] Zonghan Wu et al. "A comprehensive survey on graph neural networks". In: *IEEE transactions on neural networks and learning systems* 32.1 (2020), pp. 4–24.

[3] Lin Yao et al. "Link prediction based on common-neighbors for dynamic social network". In: *Procedia Computer Science* 83 (2016), pp. 82–89.

[4] Junwon You, Eunwoo Heo, and Jae-Hun Jung. "Phlp: Sole persistent homology for link prediction-interpretable feature extraction". In: *arXiv preprint arXiv:2404.15225* (2024).

[5] Ziwei Zhang, Peng Cui, and Wenwu Zhu. "Deep learning on graphs: A survey". In: *IEEE Transactions on Knowledge and Data Engineering* 34.1 (2020), pp. 249–270.

PHLP: Interpretable Link Prediction via Persistent Homology and Its Extension to Knowledge Graph Completion

Junwon You
Joint work with Eunwoo Heo and Jae-Hun Jung
TDA+IM 2025

POSTECH MINDS
MATHEMATICAL INSTITUTE FOR DATA SCIENCE

Contents

Preliminaries

Preliminaries: Persistent Homology

- Persistent homology is a method for computing topological features of a data space at different spatial resolutions.
- For given data points, we assume that the given data is a set of samples drawn from some underlying unknown manifold.
- To approximate this manifold with sampled data points, we use a simplicial complex.

Definition 1. Abstract simplicial complex.

An *abstract simplicial complex* is a collection K of non-empty subsets a set $V(K)$ such that, for every element $\sigma \in K$, all non-empty subsets $\sigma' \subseteq \sigma$ are also contained in K.

Preliminaries

Preliminaries: Persistent Homology

Definition 2. Vietoris-Rips complex.

Let (P, d) be a finite metric space. For a real value $r > 0$, the *Vietoris-Rips (Rips) complex* $\mathbb{VR}^r(P)$ is the abstract simplicial complex where a simplex $\sigma \in \mathbb{VR}^r(P)$ if and only if the distance $d(p, q) \leq 2r$ for all pairs of vertices $p, q \in \sigma$.

- The choice of the parameter r is crucial for the approximation of the manifold since it determines the scale at which we construct the simplicial complex.
- Instead of selecting a single r, we use a process called filtration, where we vary r over a range of values and observe how the topological structure evolves.

Preliminaries

Preliminaries: Persistent Homology

Definition 3. Simplicial filtration.

A *filtration* $\mathcal{F} = \mathcal{F}(K)$ of a simplicial complex K is a nested sequence of its subcomplexes:

$$\emptyset = K_0 \hookrightarrow K_1 \hookrightarrow \cdots \hookrightarrow K_n = K.$$

- For a finite metric space (P, d), a Rips filtration $\{\mathbb{VR}^r(P) \hookrightarrow \mathbb{VR}^{r'}(P)\}_{r \leq r'}$ is a sequence of nested Rips complex, each constructed for a different value of the scale parameter r.

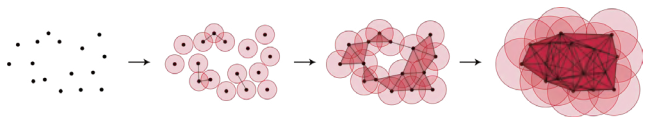


Image from https://christian.bock.ml/posts/persistent_homology/

Preliminaries

Preliminaries: Persistent Homology

Definition 4. Persistent homology groups.

For a simplicial filtration $\mathcal{F}$, the p*-th persistence homology groups* are defined as the images of the homomorphisms $H_p^{i,j} = im(h_p^{i,j}: H_p(K_i) \to H_p(K_j))$ which is induced by the inclusion $K_i \hookrightarrow K_j$ of $\mathcal{F}$, where $0 \leq i \leq j \leq n$ and each $H_p(K_i)$ represents p-th homology group of K_i for $0 \leq i \leq n$. The p*-th persistence Betti numbers* are the dimensions $\beta_p^{i,j} = dim\, H_p^{i,j}$ of the vector space $H_p^{i,j}$.

- Defien $\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j})$ for $0 < i < j \leq n$.
  - $\mu_p^{i,j}$ counts the number of independent classes that are born at K_i and die at K_j.
  - Additionally, we define $\mu_p^{i,n+1}$ (or $\mu_p^{i,\infty}$) to denote the number of independent classes that are born at K_i and never die.
- Note that we can define a function f that induces the filtration $\mathcal{F}$ such that $f(K_i) = a_i$ where $0 \leq i \leq n$ and $a_0 < a_1 < \cdots < a_n$.
  - In the case of the Rips filtration, f is defined as $f(\mathbb{VR}_r(P)) = 2r$.

Preliminaries

Preliminaries: Persistent Homology

Definition 5. Barcode.

The p-*dimensional barcode* $B_p(\mathcal{F}_f)$ of a filtration $\mathcal{F}_f$ is defined as a multi-set of intervals $[a_i, a_j) \in \mathbb{R}^2$ for all non-zero $\mu_p^{i,j}, 0 < i < j \leq n+1$, with multiplicity $\mu_p^{i,j}$, where $\bar{\mathbb{R}} = \mathbb{R} \cup \pm\infty$. The persistence of a class that is born at K_i and dies at K_j is defined as $a_j - a_i$.

Definition 6. Persistence Diagram.

The p-*dimensional persistence diagram* $Dgm_p(\mathcal{F}_f)$ of a filtration $\mathcal{F}_f$ is obtained by plotting each point (a_i, a_j) for all $[a_i, a_j) \in B_p(\mathcal{F}_f)$ on the extended plane $\bar{\mathbb{R}}^2$. Additionally, points on the diagonal $\Delta = (x, x)$ are added with infinite multiplicity.

Preliminaries

Preliminaries: Persistent Homology

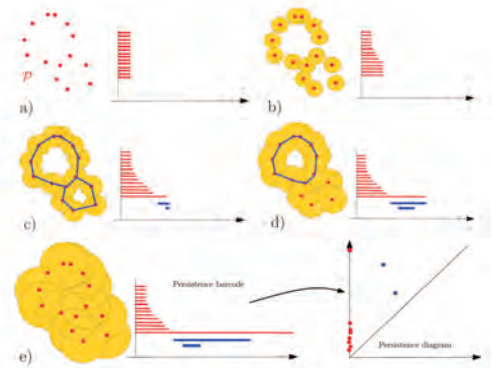


Image from https://www.math.ens.psl.eu/~eaamari/teaching/2022-2023/M2_Jussieu/Lesson%205.pdf

PHLP: Sole Persistent Homology for Link Prediction – Interpretable Feature Extraction

Junwon You[*1], Eunwoo Heo[*1], Jae-Hun Jung[1]
[1]POSTECH

Application1: Link Prediction

Preliminaries: Link Prediction (LP)

- What is Network?

Image from https://www.kbhaskar.com/project/facebook/featured.png

Facebook Social Network

Image from: https://www.sec.gov/Archives/edgar/data/701345/000119312513080344/g486396001.jpg

USAir Airline Network

Application1: Link Prediction

Preliminaries: Link Prediction (LP)

- What is Network?
  - $G = (V, E)$, V: the set of entities, E: the set of links.
  - **Example** V: the set of airports, E: the set of two airport pairs.
  - $V = \{L.A., NewYork, Chicago, \dots\}, E = \{(L.A., NewYork), (L.A., Chicago), \dots\}$.
  - Social network, citation network, biological network.
- What is Link Prediction (LP) problem?
  - **LP** is the problem of predicting the existence of a unseen link between two entities in a network based on observed links.

Application1: Link Prediction

Preliminaries: Link Prediction (LP)

- Problem definition

Problem. Link Prediction.

Consider a network $G = (V, E)$, where V represents the entity nodes in the network, and $E \subseteq |V| \times |V|$ represents the set of "true" links between entities. We are given the set of entities V and a subset $\mathrm{E}' \subset E$ of true links known as observed links. The objective of link prediction is to discover the unobserved true links $E \backslash \mathrm{E}'$.

- In the binary classification formulation of the link prediction task, the potential links L are classified as either true links or false links. Thus, Link prediction model M_b is a function that maps links in L to positive and negative labels i.e. M_b: $L \rightarrow \{0,1\}$.
- In the probability estimation formulation, potential links are associated with existence probabilities. Thus, Link prediction model M_p is a function that maps links in L to a probability i.e. M_p: $L \rightarrow [0,1]$.

Application1: Link Prediction

Preliminaries: Link Prediction (LP)

- Heuristic methods.

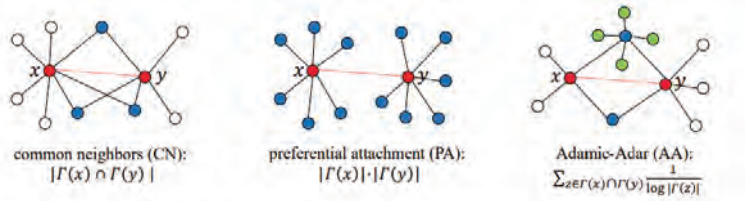


Fig. 10.1: Illustration of three link prediction heuristics: CN, PA and AA.

Image from https://graph-neural-networks.github.io/static/file/chapter10.pdf

Application1: Link Prediction

Introduction to Link Prediction (LP)

- Graph Neural Network (GNN) methods
  - These methods leverage GNNs to learn powerful representations by aggregating node features from neighborhoods.
  - GNN-based models have achieved significant score improvements in capturing intricate relationships within graphs.
  - However, GNN-based methods are comprised of neural networks, making it challenging to understand why a certain prediction was made.

- In this context, we present a novel approach to LP, called PHLP, which uses the topological information of a graph for the prediction.

Application1: Link Prediction

Our methods

- Motivation

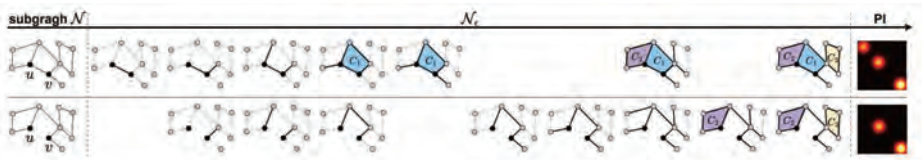


Fig. 2. Topological features in subgraphs with and without a target link (u, v). The diagram illustrates the topological information extraction process for the subgraph $\mathcal{N}$, as described in Section III-D. The presence (top) or absence (bottom) of the target link changes the topological structure of the graph. Top row: When the target link is connected, three features (C_1, C_2, and C_3) are detected shown in the persistence image (PI) in the right column. The PI represents the topological features of the subgraph $\mathcal{N}$ (Section III-E). Bottom row: When the target link is absent, only two features (C_2 and C_3) are detected as depicted in the corresponding PI.

Application1: Link Prediction

Preliminaries: Link Prediction (LP)

- Limitation and Motivation

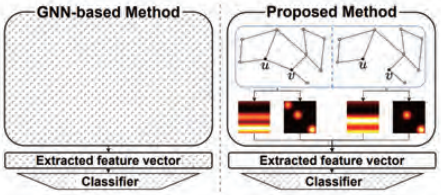


Fig. 1. Difference between the GNN-based and proposed methods. (Left) The GNN-based method extracts feature vectors through optimization (dashed area), making it difficult to interpret what these vectors represent. (Right) The proposed method extracts feature vectors through the designed analysis process, resulting in interpretable vectors.

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

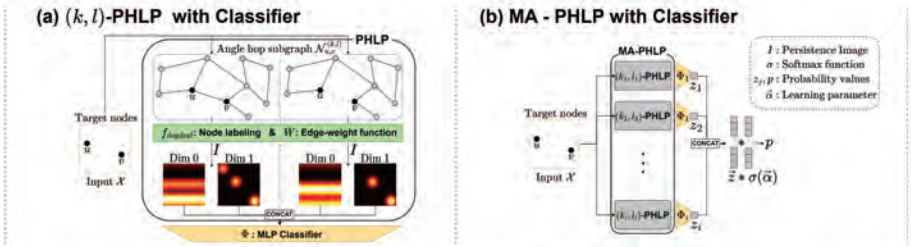


Fig. 3. Overall structure of persistent homology for link prediction (PHLP) and multiangle PHLP (MA-PHLP). (a) PHLP calculates the topological information based on the existence of target links in angle hop subgraphs for each target node. (b) With a classifier, MA-PHLP integrates topological information across various angles to perform LP.

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Extracting angle hop subgraph

- $G = (V, E)$: given graph, $u, v \in V$: given target nodes
- The k-hop enclosing subgraph $\mathcal{N}^k_{u,v} = (V', E')$ for (u, v) is defined as
  - $V' = \{\, z \in V : d(u,z) \le k \text{ or } d(z,v) \le k \,\}$,
  - $E' = \{\, (z,w) \in E : z \in V' \text{ and } w \in V' \,\}$.
- The (k,l)-angle hop enclosing subgraph $\mathcal{N}^{(k,l)}_{u,v} = (V', E')$ for (u, v) is defined as
  - $V' = \{\, z \in V : d(u,z) \le k \text{ or } d(z,v) \le l \,\}$,
  - $E' = \{\, (z,w) \in E : z \in V' \text{ and } w \in V' \,\}$.

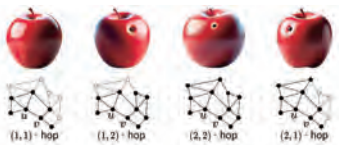


Figure 4: The motivation of (k,l)-angle hop subgraph. Just as viewing photographs of an apple from multiple angles provides a comprehensive understanding. This figure illustrates the capability to extract subgraphs from various perspectives.

Application1: Link Prediction

Ablation Studies: Angles of PHLP

- Table VI presents the performance of PHLP (dim 0) concerning various (k, l)-angle hop subgraphs.
- The best angle varies by datasets, highlighting the importance of varying angles to achieve the best performance.
- Therefore, using MA-PHLP is recommended to maximize performance consistently across datasets.

Table 6
AUC scores for PHLP (dim0) with various (k, l)-angle hops

Dataset	(1,0)	(1,1)
USAir	**96.15 ± 0.83**	95.87 ± 0.83
NS	98.28 ± 0.55	**98.66 ± 0.66**
PB	93.95 ± 0.34	**94.46 ± 0.36**
Yeast	95.52 ± 0.32	**97.31 ± 0.20**
C.ele	86.18 ± 2.12	**87.57 ± 1.20**
Power	73.39 ± 0.99	**77.83 ± 1.44**
Router	92.09 ± 0.57	**93.25 ± 0.47**
E.coli	96.94 ± 0.24	**96.95 ± 0.28**

Dataset	(2,0)	(2,1)	(2,2)
USAir	96.69 ± 0.92	96.74 ± 0.84	**96.85 ± 0.83**
NS	**98.72 ± 0.51**	98.59 ± 0.65	98.56 ± 0.47
PB	94.78 ± 0.30	94.73 ± 0.30	**94.82 ± 0.24**
Yeast	**97.71 ± 0.18**	97.66 ± 0.27	97.58 ± 0.28
C.ele	88.86 ± 1.48	**89.16 ± 1.31**	89.08 ± 1.07
Power	80.27 ± 1.07	83.90 ± 1.29	**86.12 ± 0.86**
Router	95.65 ± 0.44	**95.71 ± 0.39**	94.51 ± 0.69
E.coli	97.26 ± 0.16	97.29 ± 0.24	**97.41 ± 0.21**

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Filtration of the Subgraph

- To apply the Rips filtration, we define an edge-weight function using node labeling that reflects the topology of the given graph.
- DRNL (Double radius node labeling, Zhang et al.)
  - $f_{drnl}^{(a,b)}(w) = 1 + \min\big(d(w,a), d(w,b)\big) + q_w(q_w + r_w - 1),\ where\ d(w,a) + d(w,b) = 2q_w + r_w,\ q_w \in \mathbb{Z},\ r_w \in \{0,1\}.$
- Degree DRNL
  - $f_{degdrnl}^{(a,b)}(w) = f_{drnl}^{(a,b)}(w) + \frac{M - \deg(w)}{M}$, where M denotes the maximum degree of nodes in a subgraph.

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Filtration of the Subgraph

- DRNL (Double radius node labeling, Zhang et al.)
  - $f_{drnl}^{(a,b)}(w) = 1 + \min\big(d(w,a), d(w,b)\big) + q_w(q_w + r_w - 1),\ where\ d(w,a) + d(w,b) = 2q_w + r_w,\ q_w \in \mathbb{Z},\ r_w \in \{0,1\}.$

$(d(w,a), d(w,b))$	(0,-), (-,0)	(1,1)	(1,2), (2,1)	(1,3), (3,1)	(2,2)	(1,4), (4,1)	(2,3), (3,2)	...
$f_{drnl}^{(a,b)}(w)$	1	2	3	4	5	6	7	...

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

- **Filtration of the Subgraph**

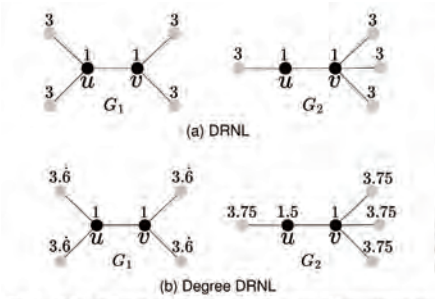


Fig. 4. Node labeling on graphs. (a) Node label values without considering the graph structure cannot distinguish between G_1 and G_2 using DRNL. (b) Applying Degree DRNL allows G_1 and G_2 to be distinguished solely by node label values.

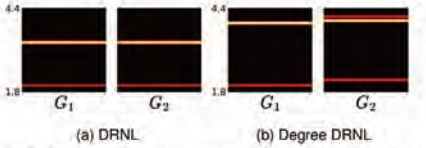


Fig. 5. Persistence images (PIs) for two node labeling methods for the graphs in Fig. 4. (a) DRNL exhibits identical zero-dimensional PIs for G_1 and G_2. (b) Degree DRNL produces distinct outcomes, effectively distinguishing between the two.

Application1: Link Prediction

Ablation Studies: Effects of Degree DRNL

- To assess the proposed Degree DRNL regarding the influence of incorporating degree information, we conducted experiments using DRNL and Degree DRNL and compared the results.
- Across all datasets, MA-PHLP yields higher AUC scores when used with Degree DRNL than with DRNL.
- The substantial improvement observed in the Power dataset is noteworthy, where Degree DRNL yields an increase of over 4 points in the AUC score.

Table 5
AUC scores for MA-PHLP (dim0) by node labeling

Dataset	DRNL	Degree DRNL
USAir	96.73 ± 0.64	$\mathbf{97.10 \pm 0.73}$
NS	98.35 ± 0.58	$\mathbf{98.78 \pm 0.65}$
PB	94.49 ± 0.27	$\mathbf{95.06 \pm 0.28}$
Yeast	97.42 ± 0.27	$\mathbf{97.98 \pm 0.23}$
C.ele	88.97 ± 1.37	$\mathbf{89.88 \pm 1.22}$
Power	88.51 ± 0.81	$\mathbf{92.77 \pm 0.47}$
Router	96.21 ± 0.53	$\mathbf{96.37 \pm 0.43}$
E.coli	97.15 ± 0.18	$\mathbf{97.72 \pm 0.17}$

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Filtration of the Subgraph

- For given subgraph $\mathcal{N} = (V', E')$, $f: V' \to \mathbb{N}$ denotes a node labeling.
- The edge-weight function $W: V' \to \mathbb{R}$ is defined as
  - $W(w,z) = \max\big(f(w), f(z)\big) + \frac{min(f(w),f(z))}{\max(f(w),f(z))}$
  - The min/max term refines values further, enhancing the discriminative power by reducing the occurrence of identical edge weights.

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Predicting the existence of the target link.

- Step1. For the given target nodes (u, v), extract the (k, l)-angle hop subgraph $\mathcal{N}_{u,v}^{(k,l)}$, denoted as $\mathcal{N}^-$, assuming that the target link does not exist during the process.
- Step2. On this subgraph calculates the edge-weights and extracts topological features by calculating PD and its vectorization, persistence image.
- Step3. If $k \neq l$, for symmetry, repeat the same process with the (l, k)-angle hop subgraph and consider the average of the two vectors, denoting this vector as $x_{(u,v)}^-$.
- Step4. To observe the difference in topological features, we consider a subgraph $\mathcal{N}^+$ obtained by connecting the target link. For this graph, $x_{(u,v)}^+$ denotes the vector obtained using this method. Concatenate these vectors as a single vector $x_{(u,v)}$.

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Predicting the existence of the target link.

- Step5. By above process, we convert any link (u, v) to a single vector $x_{(u,v)}$. Thus we converted LP problem as a binary classification task of given vectors.
- Step6. We employ MLP classifier Φ for prediction. The classifier Φ predicts the existence of a link between two target nodes with the following probability:
  - $z_{uv} = \sigma(\Phi(x_{(u,v)}))$.
- Step7. This classifier Φ is trained with the binary cross-entroy loss function BCE.
  - $\sum_{(u,v)\in\mathcal{X}} BCE(z_{uv}, y_{uv})$, where y_{uv} denotes the label of the target link.

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Multiangle PHLP (MA-PHLP)

- The MA-PHLP maximizes the advantages of PHLP by examining data from various angles through the extraction of subgraphs based on a hyperparameter, the maximum hop (max hop, denoted as H).
- The types of anlges are elements of all combinations of k and l within the set $\{(k, l) \in \mathbb{Z}^2 : 0 \leq l \leq k \leq H, k > 0\}$. Denotes the cardinality of this set as N.
- The MA-PHLP predicts the likelihood of the link existence with the following probability:
  - $p = \sum_{i=1}^{N} \alpha_i z_i$, where $\alpha = (\alpha_1, \cdots, \alpha_N) \in \mathbb{R}^N$ is a trainable parameter, and z_i denotes the prediction probability of a PHLP for each type of angle hop for $i = 1,2, \cdots, N$.

Application1: Link Prediction

Results: MA-PHLP

Table 1
Link prediction performance measured by the AUC on Benchmark datasets (90% observed links)

Dataset	USAir	NS	PB	Yeast	C. ele	Power	Router	E. coli
AA	95.06 ± 1.03	94.45 ± 0.93	92.36 ± 0.34	89.43 ± 0.62	86.95 ± 1.40	58.79 ± 0.88	56.43 ± 0.51	95.36 ± 0.34
Katz	92.88 ± 1.42	94.85 ± 1.10	92.92 ± 0.35	92.24 ± 0.61	86.34 ± 1.89	65.39 ± 1.59	38.62 ± 1.35	93.50 ± 0.44
PR	94.67 ± 1.08	94.89 ± 1.08	93.54 ± 0.41	92.76 ± 0.55	90.32 ± 1.49	66.00 ± 1.59	38.76 ± 1.39	95.57 ± 0.44
WLK	96.63 ± 0.73	98.57 ± 0.51	93.83 ± 0.59	95.86 ± 0.54	89.72 ± 1.67	82.41 ± 3.43	87.42 ± 2.08	96.94 ± 0.29
WLNM	95.95 ± 1.10	98.61 ± 0.49	93.49 ± 0.47	95.62 ± 0.52	86.18 ± 1.72	84.76 ± 0.98	94.41 ± 0.88	97.21 ± 0.27
N2V	91.44 ± 1.78	91.52 ± 1.28	85.79 ± 0.78	93.67 ± 0.46	84.11 ± 1.27	76.22 ± 0.92	65.46 ± 0.86	90.82 ± 1.49
SPC	74.22 ± 3.11	89.94 ± 2.39	83.96 ± 0.86	93.25 ± 0.40	51.90 ± 2.57	91.78 ± 0.61	68.79 ± 2.42	94.92 ± 0.32
MF	94.08 ± 0.80	74.55 ± 4.34	94.30 ± 0.53	90.28 ± 0.69	85.90 ± 1.74	50.63 ± 1.10	78.03 ± 1.63	93.76 ± 0.56
LINE	81.47 ± 10.71	80.63 ± 1.90	76.95 ± 2.76	87.45 ± 3.33	69.21 ± 3.14	55.63 ± 1.47	67.15 ± 2.10	82.38 ± 2.19
SEAL	97.10 ± 0.87	98.25 ± 0.61	95.07 ± 0.39	97.60 ± 0.33	89.54 ± 1.23	86.21 ± 2.89	95.07 ± 1.63	97.57 ± 0.30
WP	98.20 ± 0.57	99.12 ± 0.45	95.42 ± 0.25	98.21 ± 0.17	93.30 ± 0.91	92.11 ± 0.76	97.15 ± 0.29	98.54 ± 0.19
LGLP	97.09 ± 0.13	99.12 ± 0.00	94.70 ± 0.04	97.53 ± 0.13	88.64 ± 0.29	85.63 ± 0.07	95.51 ± 0.07	98.39 ± 0.08
MPLP	97.01 ± 0.54	96.17 ± 0.84	94.06 ± 0.58	94.25 ± 0.43	90.48 ± 0.87	73.71 ± 1.08	91.90 ± 0.50	96.67 ± 0.14
MA-PHLP	97.10 ± 0.69	98.88 ± 0.45	95.10 ± 0.26	97.98 ± 0.22	90.33 ± 1.16	93.05 ± 0.45	96.30 ± 0.43	97.64 ± 0.20
MA-PHLP (dim0)	97.10 ± 0.73	98.78 ± 0.65	95.06 ± 0.28	97.98 ± 0.23	89.88 ± 1.22	93.37 ± 0.41	96.37 ± 0.43	97.72 ± 0.17

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Hybrid Method

- The proposed approach easily integrates with existing subgraph methods. Subgraph methods treat the LP task as a binary classification problem comprising two components: a feature extractor F and classifier P.
- Step1. Subgraph extraction: For the given graph G and target nodes (u, v), k-hop subgraph $\mathcal{N}_{u,v}^{k}$ is extracted.
- Step2. Feature extraction: Existing methods extract features $Z = F(\mathcal{N}_{u,v}^{k})$.
- Step3. Persistence image calculation: Our method extracts topological features and denotes as I. Use MLP Φ to transforms I into a format similar to Z.
- Step4. Concatenate $\alpha_1 Z$ and $\alpha_2 \Phi(I)$ where α_1 and α_2 are trainable parameters. This concatenated vector is classified using the existing method's classifier.

Application1: Link Prediction

Our methods: Persistent Homology for Link Prediction

Hybrid Method

- The proposed a … graph methods treat the LP tas … : a feature extractor F anc …
- Step1. Subgrap … op subgraph $\mathcal{N}_{u,v}^{k}$ is extracted.
- Step2. Feature …
- Step3. Persister … and denotes as I. Use MLP Φ to …
- Step4. Concate… This concatenated vector is classif …

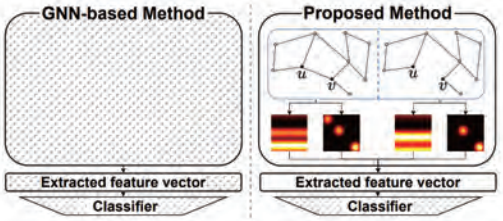


Fig. 1. Difference between the GNN-based and proposed methods. (Left) The GNN-based method extracts feature vectors through optimization (dashed area), making it difficult to interpret what these vectors represent. (Right) The proposed method extracts feature vectors through the designed analysis process, resulting in interpretable vectors.

Application1: Link Prediction

Results: hybrid methods

Table 3
AUC scores for SEAL with and without TDA features

Dataset	SEAL	MA-PHLP + SEAL
USAir	97.10 ± 0.87	$\mathbf{97.41 \pm 0.62}$
NS	98.25 ± 0.61	$\mathbf{98.97 \pm 0.30}$
PB	95.07 ± 0.39	$\mathbf{95.14 \pm 0.39}$
Yeast	97.60 ± 0.33	$\mathbf{97.93 \pm 0.18}$
C.ele	89.54 ± 1.23	$\mathbf{89.61 \pm 1.12}$
Power	86.21 ± 2.89	$\mathbf{95.53 \pm 0.33}$
Router	95.07 ± 1.63	$\mathbf{96.15 \pm 1.26}$
E.coli	97.57 ± 0.30	$\mathbf{97.93 \pm 0.34}$

Table 4
AUC scores for WALKPOOL (WP) with and without TDA features

Dataset	WP	MA-PHLP + WP
USAir	98.20 ± 0.57	$\mathbf{98.27 \pm 0.53}$
NS	99.12 ± 0.45	$\mathbf{99.24 \pm 0.32}$
PB	95.42 ± 0.25	$\mathbf{95.58 \pm 0.32}$
Yeast	98.21 ± 0.17	$\mathbf{98.25 \pm 0.18}$
C.ele	93.30 ± 0.91	$\mathbf{93.32 \pm 0.71}$
Power	92.11 ± 0.76	$\mathbf{96.09 \pm 0.38}$
Router	97.15 ± 0.29	$\mathbf{97.18 \pm 0.24}$
E.coli	98.54 ± 0.19	$\mathbf{98.57 \pm 0.20}$

Application1: Link Prediction

Ablation Studies: Persistence Image Resolution

- We conducted ablation studies to evaluate the sensitivity of MA-PHLP (0-dim) to the resolution of persistence images (PIs).
- As shown in the figure, model performance remains relatively stable as long as the resolution is not too small.
- This suggests that PI resolution is not a highly sensitive hyperparameter.
- Based on this observation, we set $m = 16$ as a balanced choice between computational cost and performance for all experiments.

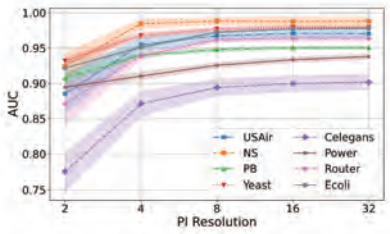


Figure 7: Sensitivity analysis of MA-PHLP (0 dim) to persistence image (PI) resolution size $m \times m$ for $m \in \{2^i \mid i = 1, 2, \ldots, 5\}$. The AUC scores are averaged over 10 runs for each setting. The shaded regions denote the range of one standard deviation around the average.

Application1: Link Prediction

Ablation Studies: Comparison with TLC-GNN

- To demonstrate that the proposed method extracts superior topological information compared to the other method TLC-GNN which use PH, we conducted the experiments. The TLC-GNN was constructed by augmenting the GCN model with PI information. We replaced the PI component of the TLC-GNN model with the PI vector produced by MA-PHLP, resulting in the MA-PHLP-GNN.
- The MA-PHLP-GNN outperformed the TLC-GNN significantly on the CiteSeer and PubMed datasets while achieving similar performance on the Cora dataset.
- The TLC-GNN does not exhibit performance improvement for the PubMed dataset despite adding topological information. However, the proposed MA-PHLP-GNN demonstrates substantial performance enhancement.

TABLE VII
COMPARISON OF AUC SCORES WITH TLC-GNN

Dataset	GCN	TLC-GNN	MA-PHLP-GNN
Cora	92.20 ± 0.83	$\mathbf{93.16 \pm 0.56}$	93.14 ± 0.93
CiteSeer	86.52 ± 1.29	87.38 ± 0.97	$\mathbf{92.08 \pm 0.53}$
PubMed	96.63 ± 0.15	96.30 ± 0.25	$\mathbf{98.07 \pm 0.07}$

Extension to Knowledge Graph Completion

Application2: Knowledge Graph

Preliminaries: Knowledge Graph (KG)

- What is Knowledge Graph?

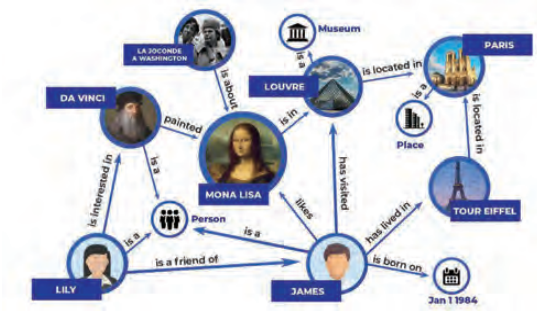


Image from: https://deeppavlov.ai/research/tpost/bn15u1y4v1-improving-knowledge-graph-completion-wit

Application2: Knowledge Graph

Preliminaries: Knowledge Graph (KG)

- Knowledge in graph from:
  - Capture entities, types, and relations
- Nodes are entities
- Nodes are labeled with their types
- Edges between two nodes capture relationships between entities
- KG can store semantic data
- KG can be formulated as $KG = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ where $\mathcal{E}$ is the set of entities, $\mathcal{R}$ is the set of relations and $\mathcal{T} \subset \mathcal{E} \times \mathcal{E} \times \mathcal{R}$ is the set of triples (h, t, r).

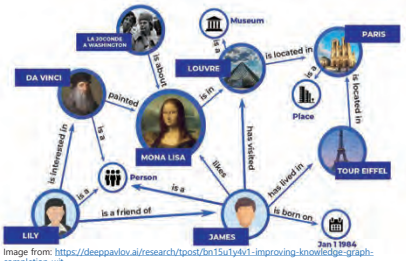


Image from: https://deeppavlov.ai/research/tpost/bn15u1y4v1-improving-knowledge-graph-completion-wit

Application2: Knowledge Graph

Preliminaries: Knowledge Graph (KG)

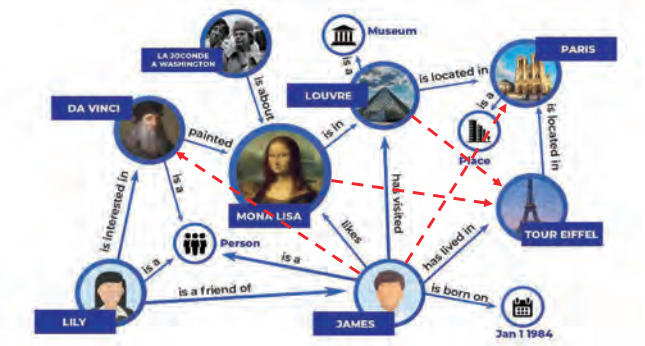
Examples of knowledge graphs

- Google Knowledge Graph Search API
  - powers semantic search results.
- Amazon Product Graph
  - recommend products based on user preferences and item relationships.
- Facebook Graph API
  - captures user connections, interactions, and interests.
- Microsoft Satori Knowledge Graph
  - supports structured information for Bing search and Cortana.

Application2: Knowledge Graph

Preliminaries: Knowledge Graph (KG)

Knowledge Graph Completion (KGC)



(h, t, ?)
(h, ?, r)
(?, t, r)

Application2: Knowledge Graph

Preliminaries: Knowledge Graph (KG)

Knowledge Graph Embedding (KGE)

- Maps entities and relations into a vector space while preserving the structural information of the KG.
- Enables the prediction of triples using embeddings.
  - Define scoring function $f: \mathcal{T} \rightarrow [0,1]$ such that:
  - If $(h,r,t) \in \mathcal{T}$, then $f(h,r,t) \approx 1$
  - Else, $f(h,r,t) \approx 0$.
- One well-known KGE method is TransE (NeurIPS 2013).

Application2: Knowledge Graph

Preliminaries: Knowledge Graph (KG)

TransE (NeurIPS 2013)

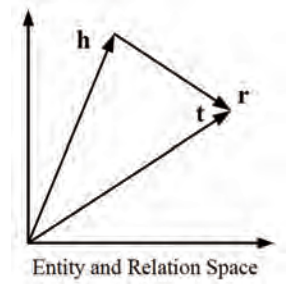


Image from: https://aws-dglke.readthedocs.io/en/latest/kg.html

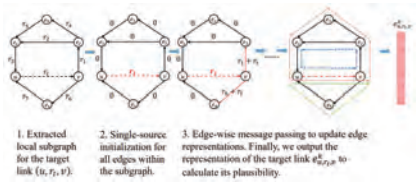
$$h + r \approx t$$

- This method is simple yet effective.
- But it struggles with symmetric relations.
  - Ex. "is a friend of"
  - If A is a friend of B, then:
  - A+r=B and B+r=A, thus, A+2r=A so r=0.
- To address this, many extensions have been proposed:
  - TransR, TorusE, RotatE, DistMult, ComplEx, and others.

Application2: Knowledge Graph

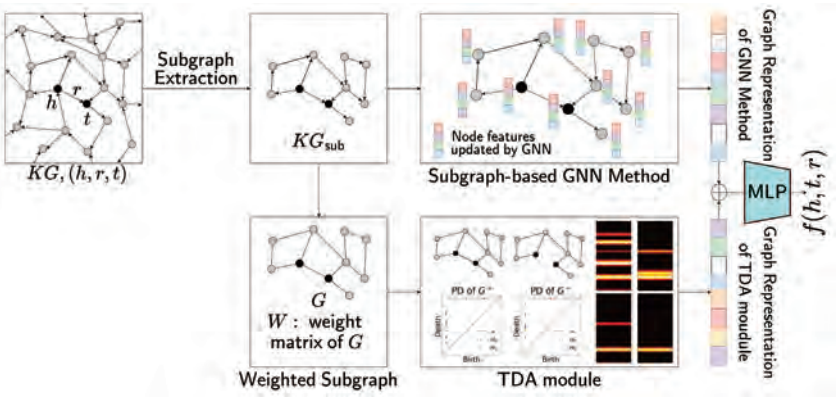
Subgraph-based GNNs for KGC

- Limitation of KGE (e.g., TransE)
  - Rely on manually defined scoring functions.
  - Assume simple relational patterns.
  - Struggle with complex relational structures.
- Subgraph-based GNN Approach
  - For a triple (h, t, r), extract its enclosing subgraph.
  - Then they apply message passing, which uses semantic embeddings of entities and relations to capture the relational context.
  - The resulting embedding reflects the local graph structure and is used for final prediction.
- These subgraph-based GNN methods currently achieve state-of-the-art performance in many KGC benchmarks.
- However, most models still rely heavily on structural and semantic information only.



Application2: Knowledge Graph

Our Method



Application2: Knowledge Graph

Our Method

- **1. Subgraph Extraction KG_{sub}.** Given a target triple (h,t,r), we extract a local enclosing subgraph KG_{sub} from the knowledge graph KG.
- **2. Subgraph-based GNN Encoding.** Let's denote the GNN feature extractor as F(we employed the Rest model [1]). It produces a representation vector $V_F = \mathrm{F}(\mathrm{KG}_{sub})$ based on semantic and structural information.
- **3. Subgraph Conversion to Weighted Simple Graph.** The extracted subgraph KG_{sub} is transformed into a weighted simple graph $G = (V, E)$, $W: E \to \mathbb{R}_{\geq 0}$. Two weighting schemes are considered (next slide).
- **4. Persistent Homology via Graph Filtration.** We construct a filtration on the weighted graph: $K_{\epsilon_1} = \emptyset \hookrightarrow K_{\epsilon_2} \hookrightarrow K_{\epsilon_3} \hookrightarrow \cdots \hookrightarrow K_{\epsilon_m} = \mathbb{X}$ where $K_{\epsilon_i} = \{\tau \in \mathbb{X} \mid if\ |\tau| = 1\ or\ \tau = (u,v)\ s.t. W(u,v) \leq \epsilon_i\}$, $\mathbb{X} = V \cup E$ and $\epsilon_1 \leq \epsilon_3 \leq \epsilon_3 \leq \cdots \epsilon_m$. From this filtration, we compute persistence diagrams (PDs) and convert them to persistence images (PIs). This gives us the topological representation, denoted by V_T.
- **5. Fusion and Prediction.** We concatenate the representations V_F and V_T, and feed the result into a multilayer perceptron (MLP) classifier for final prediction.

[1] Liu, T., Lv, Q., Wang, J., Yang, S., & Chen, H. (2023). Learning rule-induced subgraph representations for inductive relation prediction. *Advances in Neural Information Processing Systems, 36*, 3517-3535.

Application2: Knowledge Graph

Our Method

3. Subgraph Conversion to Weighted Simple Graph.

Given a knowledge graph $KG = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ where $\mathcal{E}$ is the set of entities, $\mathcal{R}$ is the set of relations and $\mathcal{T} \subset \mathcal{E} \times \mathcal{E} \times \mathcal{R}$ is the set of triples (h,t,r), we define a simple graph of KG as $G = (V,E)$ where $V = \mathcal{E}$ and $(u,v) \in E$ if $(u,v,r) \in \mathcal{T}$ or $(v,u,r) \in \mathcal{T}$ for any $r \in \mathcal{R}$.

- **1) Fixed Filtration.** This follows the same edge weighting scheme used in PHLP. We use Degree DRNL, a node labeling method that assigns node labels based on distance and node degrees.
- **2) Filtration Learning.** We introduce a learnable edge weighting function that utilizes the head, tail, and relation features. Let $(h,t,r) \in \mathcal{T}$ be a triple, and let feature vectors derived from a GNN be denoted as x_h, x_t, x_r, respectively. Then the weight function $W(u,v)$ is defined as:

$$W(u,v) = \frac{1}{|\mathcal{T}_{u,v}|} \sum_{(h,t,r) \in \mathcal{T}_{u,v}} \Phi(x_h, x_t, x_r)$$

where Φ is a learnable multilayer perceptron and

$$\mathcal{T}_{u,v} = \{(h,t,r) \in \mathcal{T} \mid (h,t,r) = (u,v,r) \text{ or } (h,t,r) = (v,u,r) \text{ for any } r \in \mathcal{R}\}.$$

This allows the filtration to reflect semantic information from the knowledge graph.

Application2: Knowledge Graph

Results

Table 1: Hits@10 results on the inductive benchmark datasets extracted from WN18RR, FB15k-237 and NELL-995. The results of GraIL, CoMPILE, TACT, SNRI, ConGLR and REST are taken from the paper [1].

	WN18RR				FB15k-237				NELL-995			
	v1	v2	v3	v4	v1	v2	v3	v4	v1	v2	v3	v4
GraIL	82.45	78.68	58.43	73.41	64.11	80.82	83.29	80.29	59.50	93.25	91.41	73.19
CoMPILE	83.60	79.82	60.69	75.49	66.74	84.67	87.44	85.84	58.38	93.78	92.77	75.13
TACT	84.04	81.63	61.31	76.56	68.70	87.97	91.16	89.90	59.65	94.35	94.33	78.49
SNRI	87.23	83.10	67.31	83.32	71.06	89.56	89.59	89.50	66.29	95.37	94.30	82.99
ConGLR	85.64	92.93	70.74	92.90	68.25	88.68	86.91	89.31	81.07	94.92	94.36	81.61
REST	96.28	94.56	79.50	94.19	75.12	91.21	93.06	96.06	88.00	94.96	96.79	92.61
+PH	97.34	94.78	84.29	94.47	80.00	93.72	94.22	96.98	90.00	96.85	97.16	92.75
+PH(FL)	97.34	94.78	83.96	93.91	-ing	-ing	-ing	-ing	90.00	96.00	96.41	-ing

Thank you!

Q&A

Q & A

Preliminaries: Persistent Homology

- Persistence diagrams (barcodes) offer valuable topological insights, but they must be transformed into fixed-dimensional vectors that preserve this information in order to be integrated into deep learning architectures. We use persistence image as a vectorization method.
- First, a barcode B is rotated via the map $T: \mathbb{R}^2 \to \mathbb{R}^2, (b, d) \mapsto (b, d - b)$.
- Next, the persistence surface $\rho_B: \mathbb{R}^2 \to \mathbb{R}$ corresponding to B is defined as

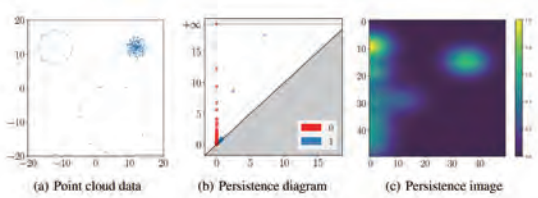
$$\rho_B(z) = \sum_{u \in T(B)} w(u) g_u(z)$$

where w is a weight function and g_u is a Gaussian function.

- The persistence image $\left(I(\rho_B)_{P_i}\right)_i$ is the collection of pixels, where $I(\rho_B)_{P_i} = \iint_{P_i} \rho_B dz$ and $\sqcup_i P_i$ is the partition of a compact subset $A \subseteq \mathbb{R}^2$ (in practice, a rectangular domain divided into $n \times n$ pixels).

Preliminaries

Preliminaries: Persistent Homology



(a) Point cloud data (b) Persistence diagram (c) Persistence image

Figure 6: Examples of persistent homology (PH). (a) Point cloud data (PCD) consisting of two circles and one disk, forming three connected components and two one-dimensional holes. (b) Persistent diagram of the PCD, with three points in 0 dimensions and two points in 1 dimension away from a diagonal line. In (b), the x-axis represents the"birth" times, indicating when specific topological features were born, while the y-axis corresponds to the "death" times, indicating when these features cease to exist. Points far from a diagonal line reflect the global structure of PCD. Points close to the diagonal line reflect the local structure of the PCD and are regarded as noise in this case. (c) A persistence image of the one-dimensional barcode of (b) with a resolution of (50, 50) and a weight function defined by $(death - birth)^2$ to emphasize persistence.

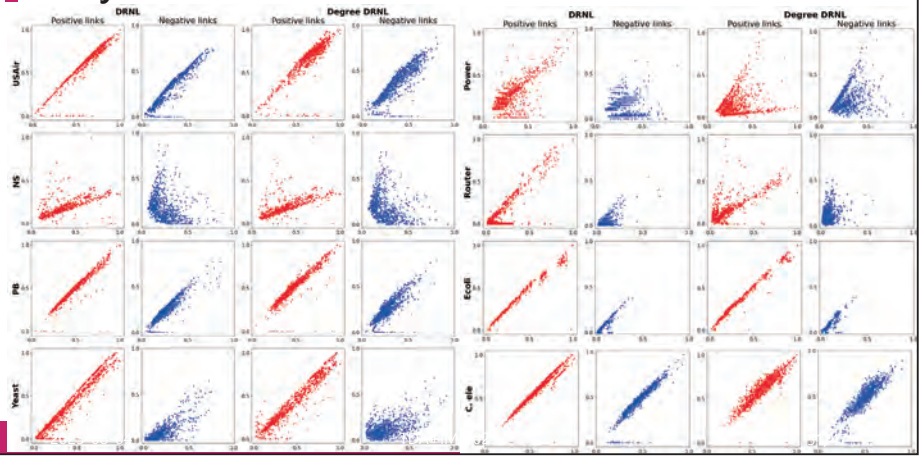
Application1: Link Prediction

Analysis

- Set $\mathcal{Z} \subset \mathbb{R}^{2\times k\times r^2}$, where k is the number of angles, r represents PI resolution and each element $(z_1, z_2) \in \mathcal{Z}$ represents concatenated PIs for cases with (z_1) and without (z_2) a target link.
- Define a function $h: \mathbb{R}^{k\times r^2} \to \mathbb{R}$ as $h(\overrightarrow{v_1}, \cdots, \overrightarrow{v_k}) = \frac{1}{k}\sum_{i=1}^{k} ||\overrightarrow{v_i}||_1$ where $\overrightarrow{v_i}$ are PIs.
- For visualization, transform $\mathcal{Z}$ into the points in $\mathbb{R}^2$ by a function G, defined as $G(z_1, z_2) = \big(h(z_1), h(z_2)\big)$.
- Distribution Analysis:
  - Compared distributions of points for positive and negative links using both DRNL and Degree DRNL.
  - NS and Yeast datasets show significant differences in distributions between positive and negative links, correlating with high performance.
  - C. elegans and Power datasets show more similar distributions with Degree DRNL, aligning with lower performance scores.
  - The difference between DRNL and Degree DRNL on Power dataset supports the highest performance increase.

Application1: Link Prediction

Analysis



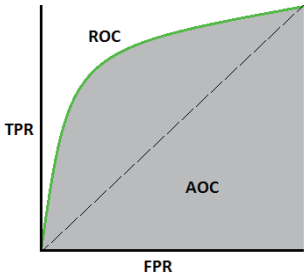
Application1: Link Prediction

AUC score

실제 \ 예측	Positive	Negative
Positive	TP	FN
Negative	FP	TN

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$



https://only-wanna.tistory.com/entry/Classification-Metrics%EB%B6%84%EB%A5%98-%EB%AA%A8%EB%8D%88-%EC%A7%80%ED%91%9C-%EC%95%BC%EC%95%84%EB%B3%84%EA%B8%B0-TPR-FPR%EA%B3%A1%EC-ROC-Curve-%EC%82%AC%EC%9D%84-%EA%B4%80%EA%B3%84-%EB%B0%8F-AUC

Potential of Mathematics for Industry, and the Dilemma in the Midstream

Tetsuji Taniguchi

Hiroshima Institute of Technology / Math. Research Institute Calc for Industry, Japan

In an era where data utilization is essential across all industries, the importance of mathematics for industry has grown dramatically. However, we face the harsh reality that the value of our work is not properly recognized, and our services are often underpriced. In particular, we mathematicians who operate in the "midstream"—translating business challenges into mathematical models—face a serious dilemma: our technical skills are appreciated, but they fail to translate into tangible business outcomes. In this presentation, I will report on my company's real-world experience in confronting this "midstream dilemma." Based on this, I will present to you the fundamental, inherent challenges that we have uncovered within the business model of applying mathematics to industry.

References.

[1] SunMath Calc for Industry, Inc. *SunMath Calc for Industry*. `https://sunmath-calc.co.jp/`. Accessed: 2025-08-25. 2025.

[2] YouTube. *Mathematics for Industry Promotional Video*. `https://www.youtube.com/watch?v=-e64M-gR2H8`. Accessed: 2025-08-25. 2025.

A Data-Driven Framework for Predicting Liver Failure Dynamics and Living Donor Transplant Prognosis

Raiki Yoshimura

Graduate School of Science, Nagoya University, Japan

Acute liver failure (ALF) is one of the most critical hepatic conditions, characterized by rapid progression and a high risk of multi-organ failure and death. Liver transplantation remains the only curative treatment, yet predicting which patients will require it is still a major clinical challenge due to the significant heterogeneity in disease progression. In our first study, we analyzed time-series clinical data from 320 patients with acute liver injury and applied machine learning techniques to identify key prognostic indicators. We found that prothrombin time (PT) serves as a central biomarker for tracking individual disease trajectories. By stratifying patients into six distinct PT dynamic patterns, we were able to quantify the severity of ALF and predict its progression from admission data. Furthermore, we demonstrated the feasibility of modeling future PT dynamics using mathematical approaches, offering a personalized framework for understanding and anticipating ALF progression. While liver transplantation offers a potential cure for end-stage liver disease, outcome prediction remains a critical issue, particularly in living donor liver transplantation (LDLT), which has gained prominence due to shorter wait times and better graft quality. In our second study, we retrospectively analyzed data from 748 LDLT recipients and developed a machine learning model to predict early graft loss (within 180 days postoperatively) with higher accuracy than conventional models. We stratified patients into five groups based on risk and further identified a specific intermediate-risk group (G2) with characteristics similar to those who experienced early graft loss (G1), but with different survival outcomes. Using data available within the first 30 days post- transplant, we constructed a hierarchical model capable of distinguishing these populations, facilitating earlier clinical intervention such as consideration of retransplantation or alternative donor strategies. Together, these studies address the continuum of liver disease—from acute liver injury to post-transplant outcomes—through the lens of time-resolved, individualized prediction. By leveraging machine learning and mathematical modeling, we present a framework that supports more precise and proactive clinical decision-making across the full trajectory of severe liver disease.

References.

[1] William Bernal et al. “Acute liver failure”. In: *The Lancet* 376.9736 (2010), pp. 190–201.

[2] David G Koch et al. “Development of a model to predict transplant-free survival of patients with acute liver failure”. In: *Clinical Gastroenterology and Hepatology* 14.8 (2016), pp. 1199–1206.

[3] Ramesh Kumar et al. “Prospective derivation and validation of early dynamic model for predicting outcome in patients with acute liver failure”. In: *Gut* 61.7 (2012), pp. 1068–1075.

[4] John G O'grady et al. "Early indicators of prognosis in fulminant hepatic failure". In: *Gastroenterology* 97.2 (1989), pp. 439–445.
[5] Raiki Yoshimura et al. "Stratifying and predicting progression to acute liver failure during the early phase of acute liver injury". In: *PNAS nexus* 4.2 (2025), pgaf004.

Topological Data Analysis and
Industrial Mathematics (TDA+IM) 2025
Day2 Aug.6th (Wed.) 10:50 – 11:15

A Data-Driven Framework for Predicting Liver Failure Dynamics and Living Donor Liver Transplant Prognosis

Raiki Yoshimura
Nagoya Univ. Department of Natural Science Interdisciplinary Biology Lab**(iBLab)**

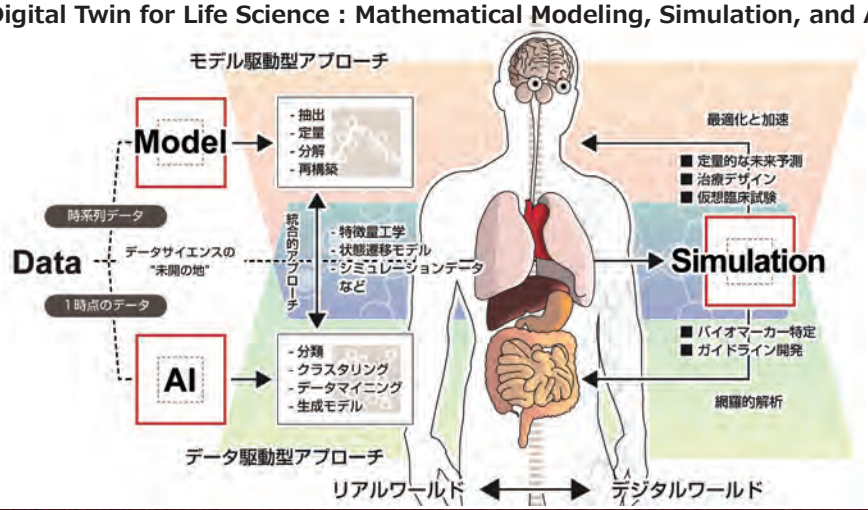
Interdisciplinary Biology Laboratory (iBLab), Nagoya University

PI: Shingo Iwami (Prof.)

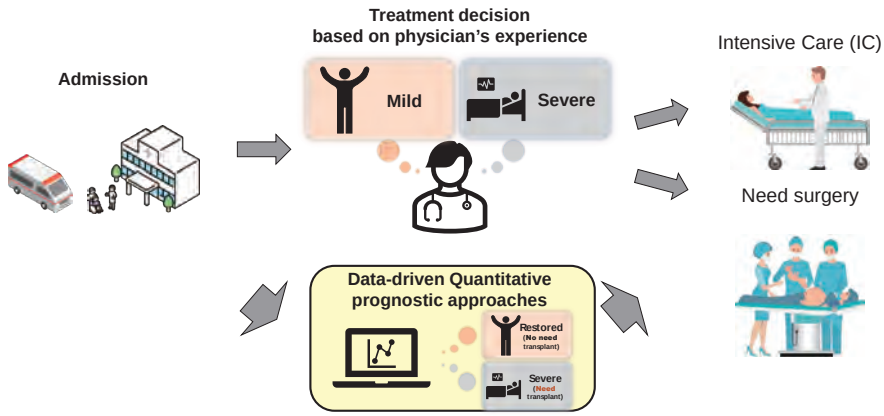
Shoya Iwanami (Lect.)

We study life and health science using mathematical model and AI.

Digital Twin for Life Science : Mathematical Modeling, Simulation, and AI

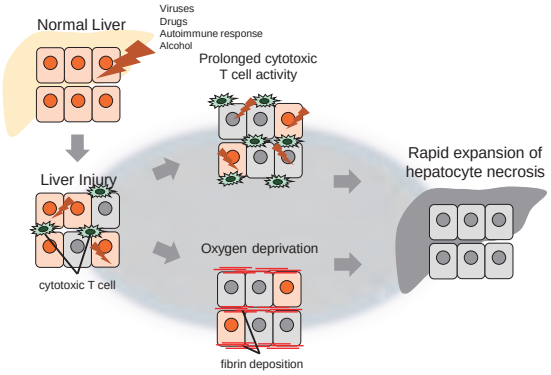


There are no quantitative indicators available to detect aggravation in many diseases



Quantitative prognostication would greatly contribute to the **establishment** **of early treatment.**

Acute liver Injury / Failure (ALI / ALF) as a typical disease



Serious

Life-saving rates without transplantation in Japan

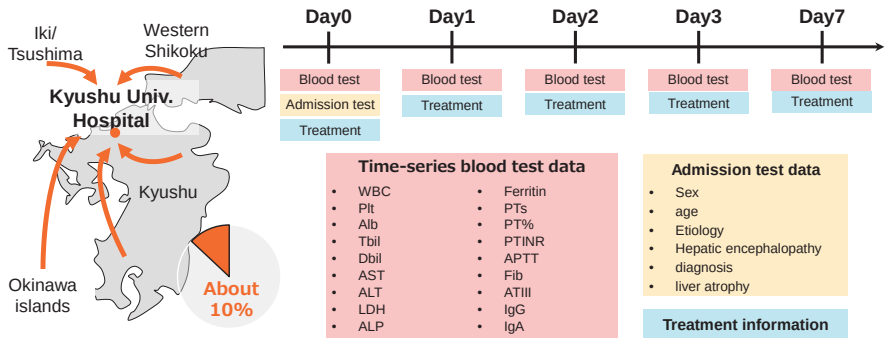
≦50%

Rare

The number of ALF patients in Japan

300 cases/year

Kyushu cohort of 10% of the nation's acute liver failure patients transported



Time-series blood test data

- WBC
- Plt
- Alb
- Tbil
- Dbil
- AST
- ALT
- LDH
- ALP
- gGTP
- Che
- NH3
- BUN
- Cre
- CRP
- Ferritin
- PTs
- PT%
- PTINR
- APTT
- Fib
- ATIII
- IgG
- IgA
- IgM
- AFP
- PIVKA
- sIL2R
- D-bil/T-bil
- MELDscore

Admission test data

- Sex
- age
- Etiology
- Hepatic encephalopathy
- diagnosis
- liver atrophy

Treatment information

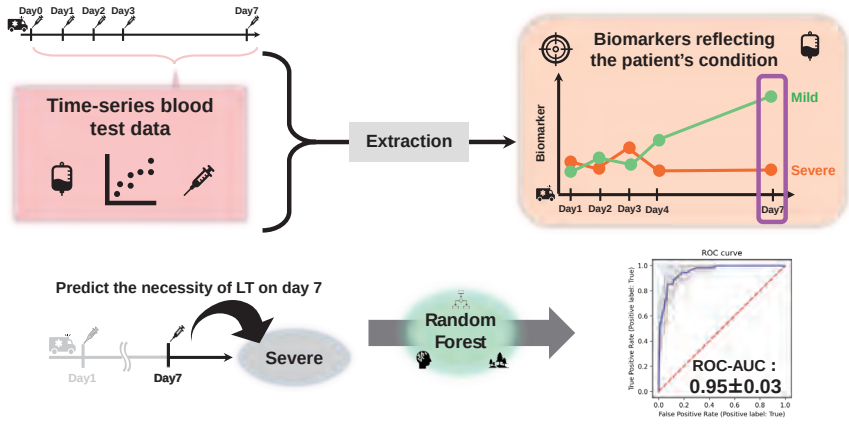
- FFP prescribing (U)
- Anticoagulation
- CHDF
- PE
- Steroid pulse therapy

Used Data : Acute Liver Injury (ALI) patients/ Failure (ALF)

- ALT more than 200 U/l
- T. Bil. more than 4 mg/dl
- PT-INR over 1.2 at admission

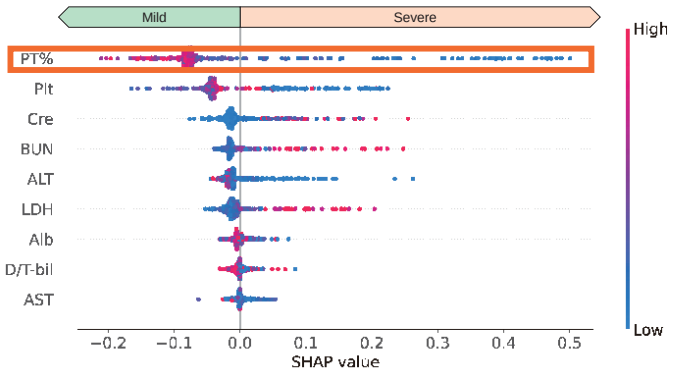
Kyushu cohort data were analyzed in collaboration with Kyushu University Hospital

Does the blood test data contain information on disease progression?



Blood test data may contain information reflecting the progression of the disease.

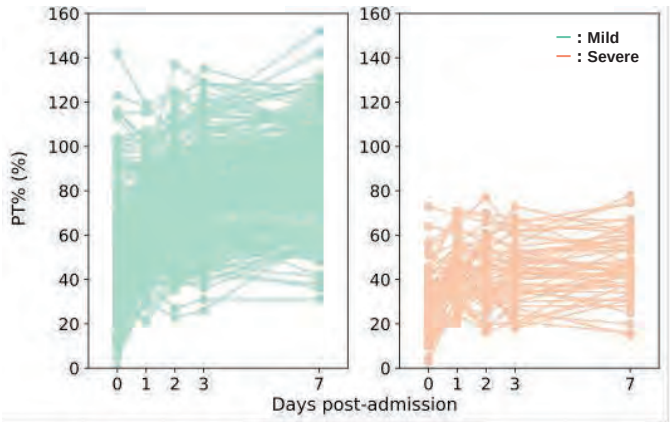
Which factors has important information for the prediction?



∗ **PT%（prothrombin time activity percentage）:**
Hepatic function index using blood clotting time that strongly relies on **clotting factors (CFs)** synthesized mainly in the **liver**

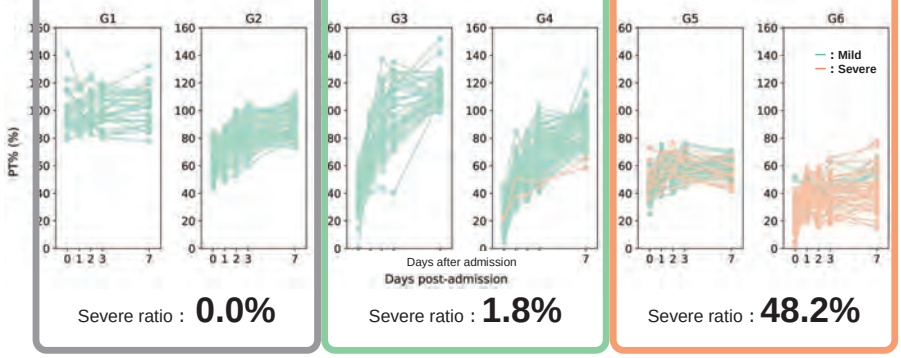
PT% (prothrombin time activity percentage) was identified as an important factor.

Time series dynamics of PT% with large individual differences



Time series dynamics of PT% appear to reflect disease progression

Can patients be stratified by time-series patterns of change in PT%?



PT% itself was suggested to be a **biomarker** reflecting the progression of the patient's condition

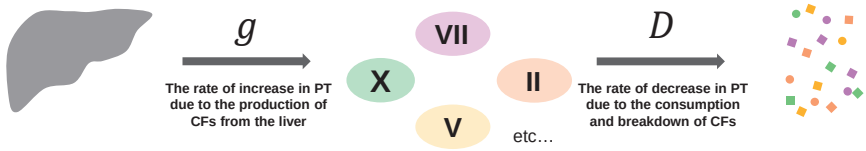
Mathematical model for predicting PT change patterns

∗ **PT% (prothrombin time activity percentage) :**
Hepatic function index using blood clotting time that strongly relies on **clotting factors (CFs)** synthesized mainly in the **liver**

$$\frac{dP(t)}{dt} = g - DP(t)$$

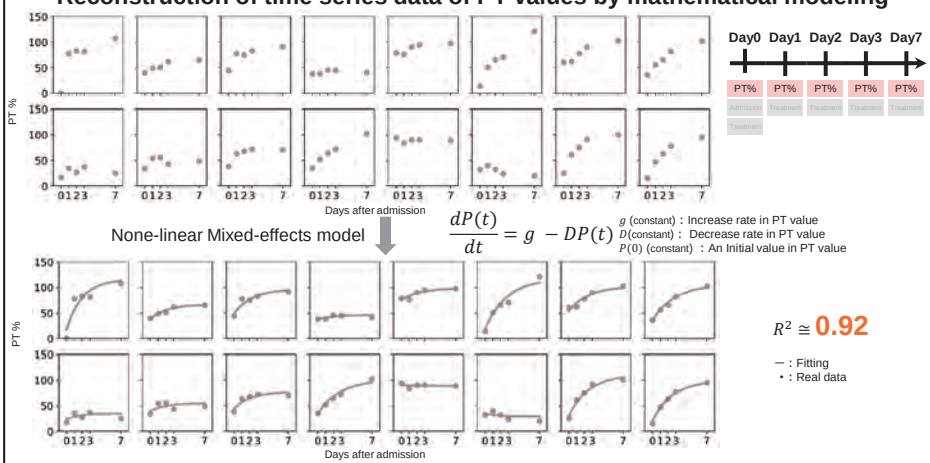
g (constant) : Increase rate in PT value
D(constant) : Decrease rate in PT value
$P(0)$ (constant) : An Initial value in PT value

Coagulation factors (CFs) strongly dependent on PT values



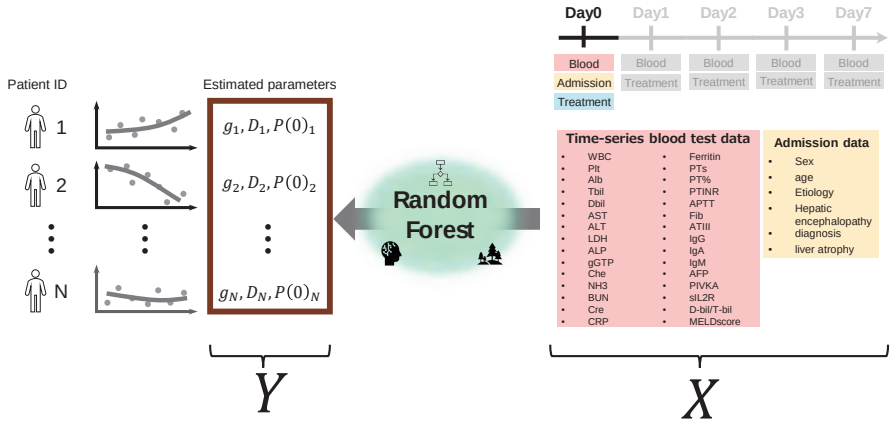
Developed a simple mathematical model based on biological processes

Reconstruction of time series data of PT values by mathematical modeling



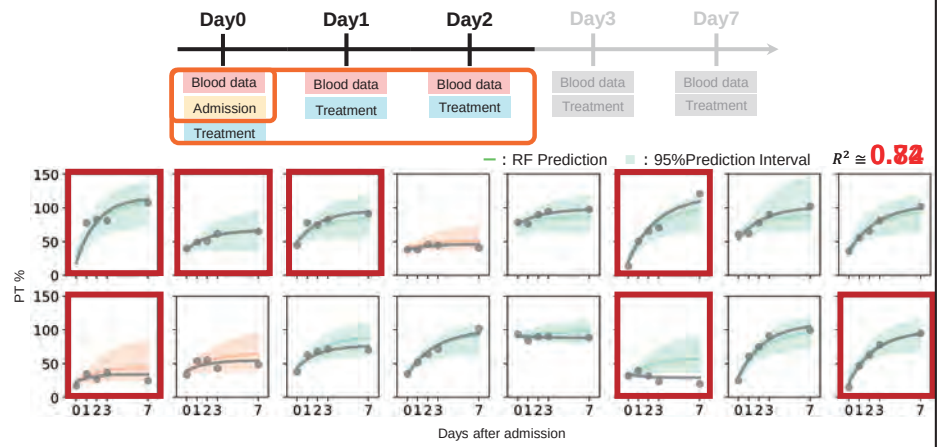
PT changes can be understood in terms of a simple process of biomolecule production and elimination

Estimated parameters as response variables



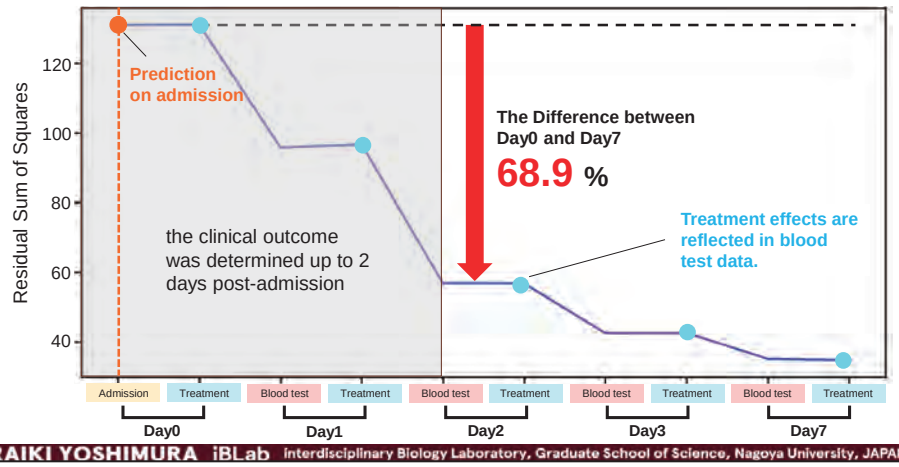
An RF model was trained to predict the parameters as the response variables with the information at the time of admission as the explanatory variable

Prediction performance was changed by adding data from another day



Adding Day2 data, the prediction performance was improved

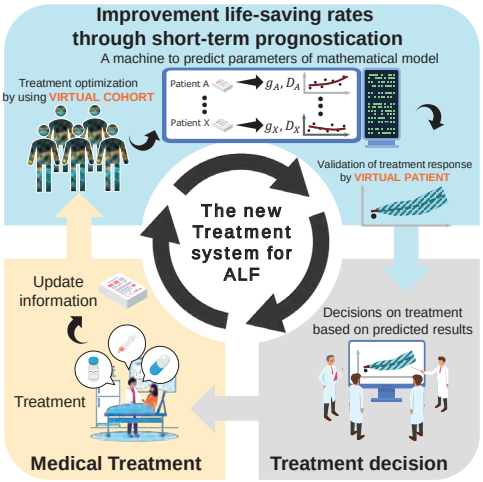
When information from the time of admission can be used to predict with a good prediction performance ?



Summary, Issues and Prospects

- **Exploring biomarker in blood data**
  - PT% are strongly associated with disease progression
- **Quantitative prediction of time-series changes in PT**
  - Development of mathematical models to describe PT dynamics
  - Future prediction of disease progression based on data at the time of admission
- **Evaluation of predictive ability with additional information**
  - Predictive ability improves with additional information up to Day 2
  - Treatment information does not improve predictive ability

Integrate data science and clinical medicine
⇒ **Realize** the new treatment system of ALF



This work has already published!

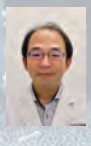
Please check it out if you're interested in

Thank you for listening

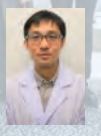
interdisciplinary Biology Laboratory (iBLab) Division of Natural Science,
Graduate School of Science, Nagoya University

Acute Liver Failure study

Professor
Yoshihiro Ogawa
Department of Medicine and Bioregulatory Science,
Kyushu University



Assistant Professor
Masatake Tanaka
Department of Medicine and Bioregulatory Science,
Kyushu University



Assistant Professor
Takeshi Goya
Department of Medicine and Bioregulatory Science,
Kyushu University

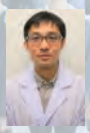
Thank you for listening

interdisciplinary Biology Laboratory (iBLab) Division of Natural Science,
Graduate School of Science, Nagoya University

Acute Liver Failure study

Professor
Yoshihiro Ogawa
Department of Medicine and Bioregulatory Science,
Kyushu University

Assistant Professor
Masatake Tanaka
Department of Medicine and Bioregulatory Science,
Kyushu University



Assistant Professor
Takeshi Goya
Department of Medicine and Bioregulatory Science,
Kyushu University

Liver transplantation study

Professor
Tomoharu Yoshizumi
Faculty of Medical Sciences,
Graduate School of Medical Sciences,
Kyushu University, Japan

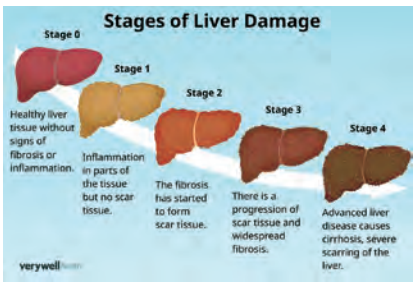
Professor
Takasuke Fukuhara
Faculty of Medical Sciences,
Graduate School of Medical Sciences,
Kyushu University, Japan

An approach to predict how a disease will change and how treatment will work before it happens

- Exploration of biomarkers and predictive models utilizing them
  - Acute Liver Failure
  - Research that combines AI and mathematical models
- **Prediction based on clinical data from early postoperative timing**
  - Liver Transplantation
  - Research that using machine learning

Liver transplantation as a last resort for end-stage liver disease

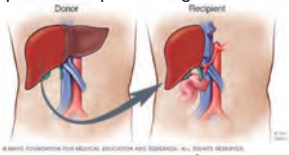
- End-stage disease causes liver dysfunction



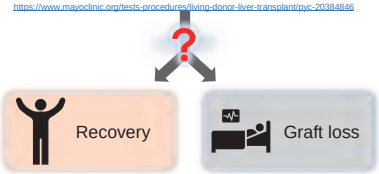
Illustrated by Zoe Hansen in Mark Gurarie, *et.al.*, *verywellhealth*, 2024

HCC　ALF　MASH　CLI
Cirrhosis　MASLD　SLD

- Some patients experience graft loss



Mayo clinic.「Living-donor liver transplant」
https://www.mayoclinic.org/tests-procedures/living-donor-liver-transplant/pyc-20384846



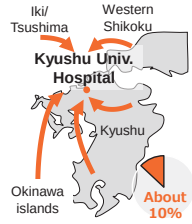
- **Conventional Studies**: simple statistical model

Predictive score, D-MELD, e-GLR： **Limited accuracy**

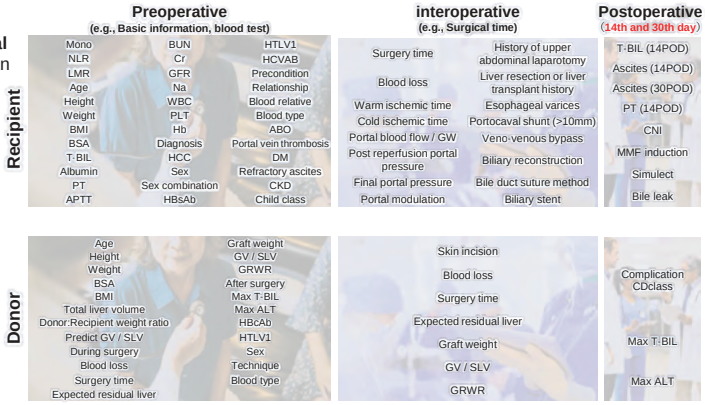
In Western countries, DDLT is the primary method of liver transplantation.

Kyushu University Hospital as a volume center for transplant treatment in Japan

Kyushu University Hospital
10% Transplantation in Japan
(about 500 patients / year)



748 patients
(Graft loss: 177)



Accepts patients from a wide range of locations, with little regional bias and a large number of patients

Definition of three clinical outcomes

Early Graft Loss
Graft loss **within 180 days** (About 20% of all graft loss)
Graft loss caused directly by the effects of the surgery

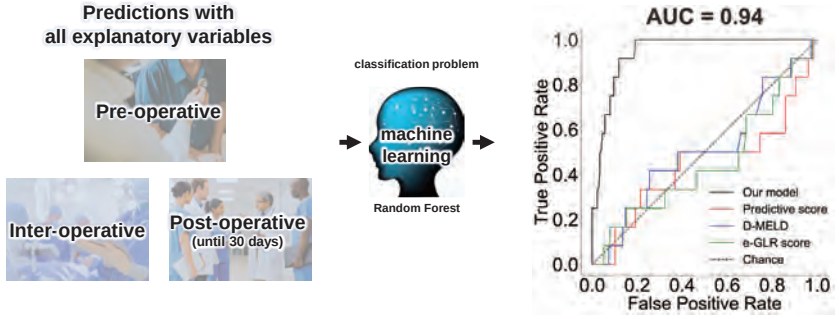
Non-early Graft Loss
Graft loss **after 180 days** (180 ~ 7500 days)
Graft loss not directly related to the effects of the surgery

Definition of three clinical outcomes

Early Graft Loss
Graft loss **within 180 days** (About 20% of all graft loss)
Graft loss caused directly by the effects of the surgery

Non-early Graft Loss
Graft loss **after 180 days** (180 ~ 7500 days)
Graft loss not directly related to the effects of the surgery

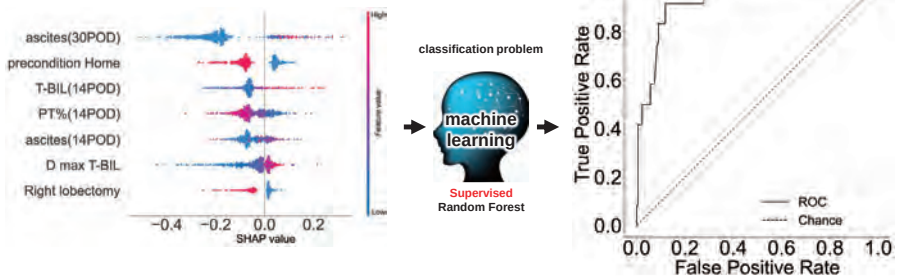
Early graft loss prediction using machine learning



Early graft loss can be predicted with high accuracy by information up to 30 days postoperatively.

Early graft loss prediction using machine learning

Feature selection using SHAP: 7 features



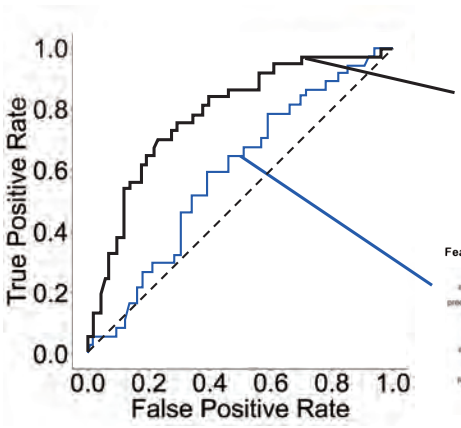
Only **7 factors are** needed to make a prediction with high enough performance

Definition of three clinical outcomes

Early Graft Loss
Graft loss **within 180 days** (About 20% of all graft loss)
Graft loss caused directly by the effects of the surgery

Non-early Graft Loss
Graft loss **after 180 days** (180 ~ 7500 days)
Graft loss not directly related to the effects of the surgery

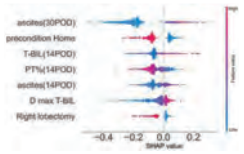
Predicting non-early graft loss using machine learning



Predictions with all explanatory variables

AUC: **0.78**

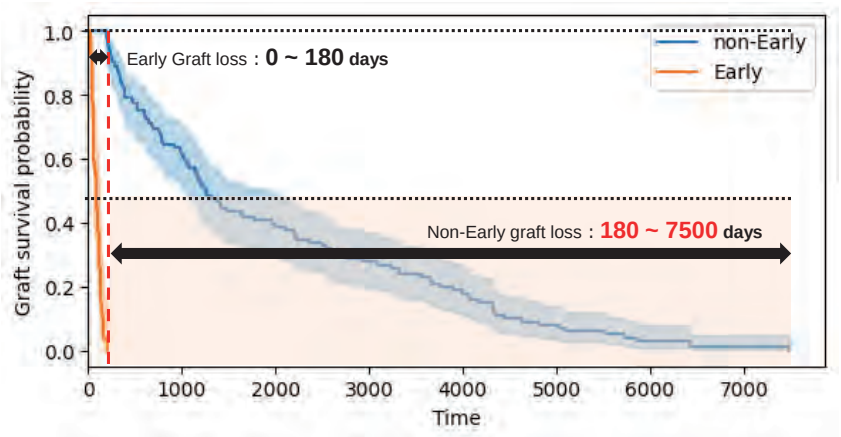
Feature selection using SHAP: 7 features



AUC: **0.59**

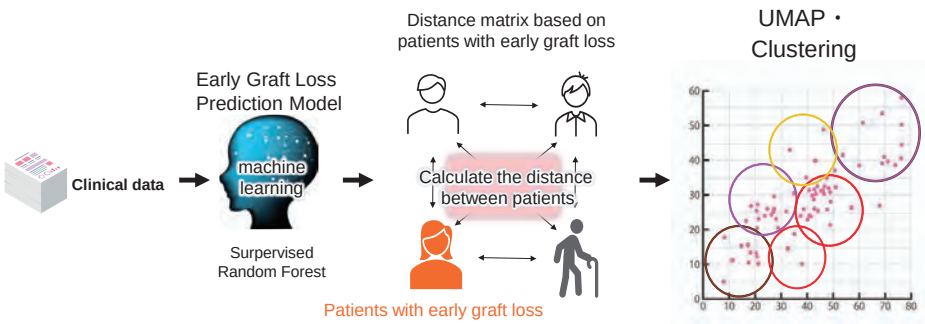
Non-early graft loss is controlled by different factors than early graft loss

Diversity in timing of graft loss



Non-early graft loss patients have high heterogeneity: Needs stratification?

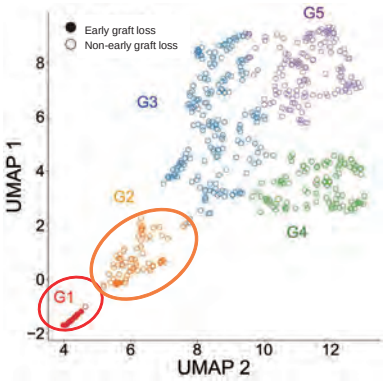
Patient stratification using supervised random forests



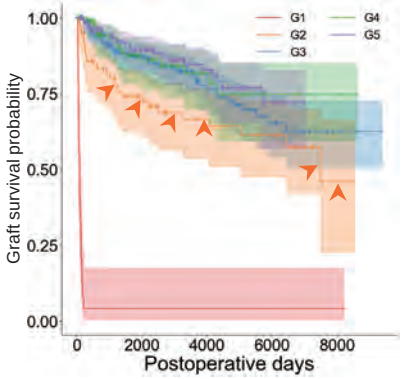
All patients can be discussed within the space based on patients with early graft loss

Patient stratification using supervised random forests

Patient stratification



Survival Time Analysis by Group



Logrank (Bonferroni)

G2-G1 : $p < 0.01$

G2-G3 : $p = 0.055$

G2-G4 : $p = 0.02$

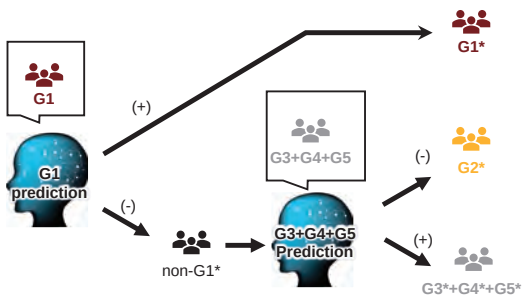
G2-G5 : $p = 0.01$

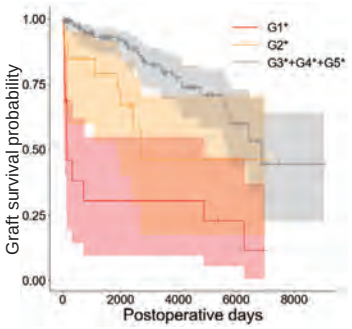
Divided into 3 groups：Early (G1), Intermediate (G2), and Long or No graft loss patients(G3+G4+G5)

RAIKI YOSHIMURA iBLab Interdisciplinary Biology Laboratory, Graduate School of Science, Nagoya University, JAPAN

Hierarchical binary classification model

■ Model detail





Sensitivity : 0.96

Multi-class classification with emphasis on G1 identification by combining two binary classification models

RAIKI YOSHIMURA iBLab Interdisciplinary Biology Laboratory, Graduate School of Science, Nagoya University, JAPAN

Summary and Future directions

■ Summary

- ✓ **Early graft loss can be accurately predicted** using clinical data **up to 30 days** postoperatively
- ✓ This prediction requires **only seven key variables**
- ✓ **Patient stratification**: identification of individuals at risk for **intermediate graft loss (G2)**
- ✓ A **hierarchical prediction** approach enables highly **accurate group-level predictions**

■ Future directions

- ✓ **Establishing the feasibility of outcome prediction** using data obtained over 30 days postoperatively.
- ✓ **Understanding** the dynamics of **pre- and postoperative clinical states**
- ✓ **Investigating the mechanisms** that determine post-transplant outcomes

RAIKI YOSHIMURA iBLab Interdisciplinary Biology Laboratory, Graduate School of Science, Nagoya University, JAPAN

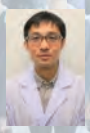
Thank you for listening

interdisciplinary Biology Laboratory (iBLab) Division of Natural Science,
Graduate School of Science, Nagoya University

Acute Liver Failure study

Professor
Yoshihiro Ogawa
Department of Medicine and Bioregulatory Science,
Kyushu University

Assistant Professor
Masatake Tanaka
Department of Medicine and Bioregulatory Science,
Kyushu University



Assistant Professor
Takeshi Goya
Department of Medicine and Bioregulatory Science,
Kyushu University

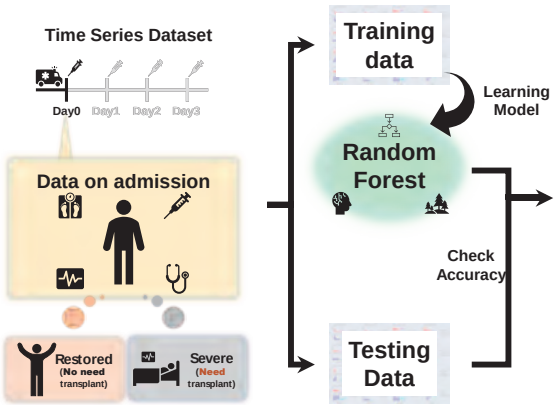
Liver transplantation study

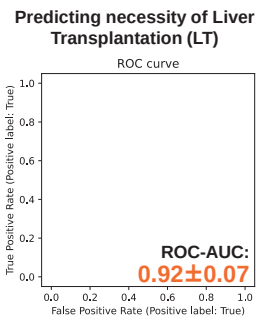
Professor
Tomoharu Yoshizumi
Faculty of Medical Sciences,
Graduate School of Medical Sciences,
Kyushu University, Japan

Professor
Takasuke Fukuhara
Faculty of Medical Sciences,
Graduate School of Medical Sciences,
Kyushu University, Japan

RAIKI YOSHIMURA iBLab interdisciplinary Biology Laboratory, Graduate School of Science, Nagoya University, JAPAN

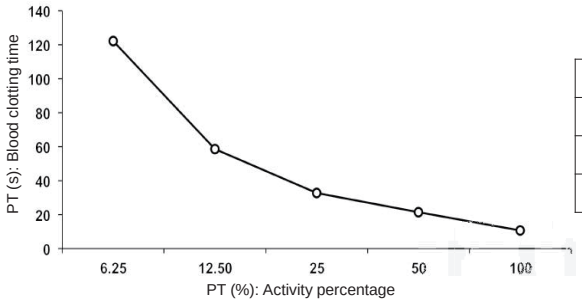
Prediction of the necessity of Liver transplantation on admission





Patients has sufficient information to predict the necessity of LT on admission

RAIKI YOSHIMURA iBLab interdisciplinary Biology Laboratory, Graduate School of Science, Nagoya University, JAPAN

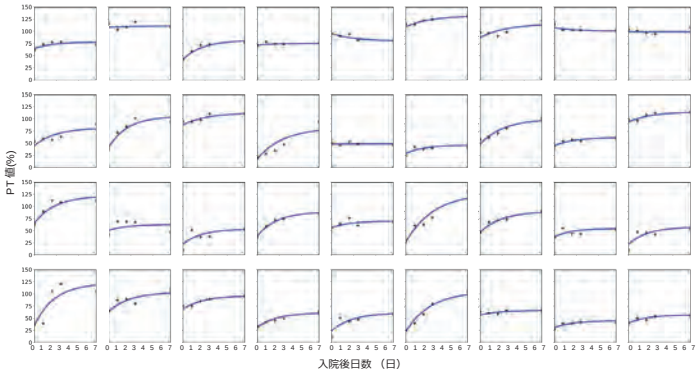


Clotting time	↑	↓
Amount of CFs	↓	↑
PT%	↓	↑

数理モデリングによるPT値の時系列データの再構築

$$\frac{dP(t)}{dt} = g - DP(t)$$

g (定数)：PT値の増加量
D(定数)：PT値の減少率
P_0 (定数) ：PT値の初期値

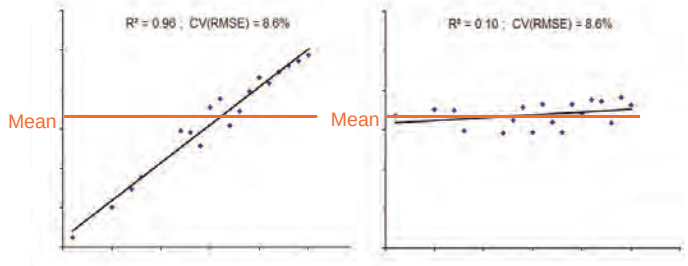


入院時データから三つのパラメータを推定できないか？

R2 VS MSE

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$



Ref: https://www.enmanreg.org/r2/

Quantifying the Topological Structure of Graphs: The Total Persistence Difference

Eunwoo Heo

Department of Mathematics, POSTECH, Korea

Persistent homology (PH) has been widely applied to graph data to extract topological features. However, little attention has been paid to how different distance functions on a graph affect the resulting persistence diagrams and their interpretations. In this paper, we define a class of distances on graphs, called path-representable distances, and investigate structural relationships between their induced persistent homologies. In particular, we identify a nontrivial injection between the 1-dimensional barcodes induced by two commonly used graph distances: the unweighted and weighted shortest-path distances. We formally establish sufficient conditions under which such embeddings arise, focusing on a subclass we call cost-dominated distances. The injection property is shown to hold in 0- and 1-dimensions, while we provide counterexamples for higher-dimensional cases. To make these relationships measurable, we introduce the total persistence difference (TPD), a new topological measure that quantifies changes between filtrations induced by cost-dominated distances on a fixed graph. We prove a stability result for TPD when the distance functions admit a partial order and apply the method to the SNAP EU Research Institution E-Mail dataset. TPD captures both periodic patterns and global trends in the data, and shows stronger alignment with classical graph statistics compared to previously proposed PH-based measures.

References.

[1] David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. "Vines and vineyards by updating persistence in linear time". In: *Proceedings of the twenty-second annual symposium on Computational geometry*. 2006, pp. 119–126.

[2] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[3] Eunwoo Heo, Byeongchan Choi, and Jae-Hun Jung. "Persistent Homology with Path-Representable Distances on Graph Data". In: *arXiv preprint arXiv:2501.03553* (2025).

[4] Mai Lan Tran, Changbom Park, and Jae-Hun Jung. "Topological data analysis of Korean music in Jeongganbo: a cycle structure". In: *Journal of Mathematics and Music* 17.3 (2023), pp. 403–432.

Quantifying the Topological Structure of Graphs

- The Total Persistence Difference.

2025 Topological Data Analysis and Industrial Mathematics
Bridging Theory and Applications

Eunwoo Heo
with Byeongchan Choi, Jae-hun Jung

Department of Mathematics
POSTECH / Ph.D.
hew0920@postech.ac.kr

Persistent Homology (PH)

A mathematical tool in topological data analysis (TDA) that enables the inference of topological information about data.

00 Persistent Homology on graph data

Consider a connected weighted graph $G = (V, E, W_E)$ with its distance $d : V \times V \to \mathbb{R}^{\geq 0}$. Define an abstract simplicial complex $\mathbb{X}$ as the power set of the vertex set V, denoted by $\mathbb{X} := \mathcal{P}(V)$. Consider a real-valued function $f : \mathbb{X} \to \mathbb{R}$ defined as

$$f(\sigma) = \max\{d(v, w) \mid (v, w) \subseteq \sigma \text{ for } v, w \in V\}$$

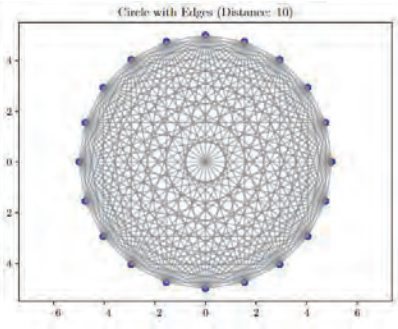
for any k-simplex $\sigma \in \mathbb{X}$ whenever $p \geq 1$, otherwise $f(\sigma) = 0$. For a sequence of real numbers $0 = a_0 \leq a_1 \leq \ldots \leq a_m$, if we denote $\mathbb{X}_{a_l} := f^{-1}((-\infty, a_l])$, then we have a filtration

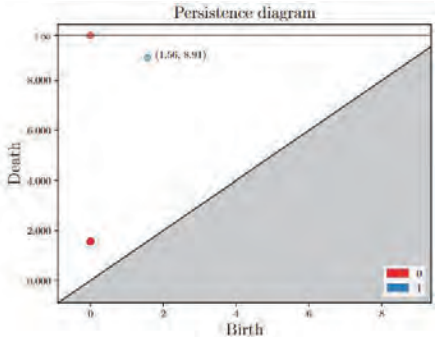
$$V = \mathbb{X}_{a_0} \hookrightarrow \mathbb{X}_{a_1} \hookrightarrow \cdots \hookrightarrow \mathbb{X}_{a_m} = \mathbb{X}.$$

A persistence barcode $\mathbf{bcd}_k(d)$, for each homological dimension k, is the visual representations of homological features as a multiset. They can be represented as follows:

$$\mathbf{bcd}_k(d) = \{[a, b] \mid a, b \in \mathbb{R}, a < b\}.$$

Persistence Diagram (PD)





01

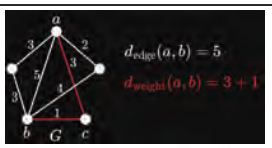
Motivation

However, the result of persistent homology on a given graph $G = (V, E, W_E)$ can vary depending on how the distance d between nodes is defined.

Could there be a hidden relationship between persistent homology results induced by different distances d_1, d_2 on the same graph G?

01

Motivation



Let $G = (V, E, W_E)$ be a connected weighted graph.

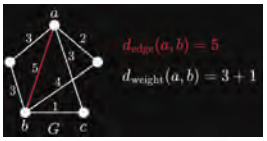
Definition 1. Define the distance $d_{\text{weight}} : V \times V \to \mathbb{R}^{\geq 0}$ such that for any two distinct vertices v and w in V,

$$d_{\text{weight}}(v, w) = \min_{p \in \mathcal{P}(v,w)} \left\{ \sum_{e \in E(p)} W_E(e) \right\},$$

and $d_{\text{weight}}(v, w) = 0$ if $v = w$, where $\mathcal{P}(v, w)$ is the set of all paths connecting v and w, and $E(p)$ denotes the set of edges composing the path p.

01

Motivation



Let $G = (V, E, W_E)$ be a connected weighted graph.

Definition 2. Define the distance $d_{\text{edge}} : V \times V \to \mathbb{R}^{\geq 0}$ such that for any two distinct vertices v and w in V,

$$d_{\text{edge}}(v,w) = \min_{p \in \mathcal{P}_{\min}(v,w)} \left\{ \sum_{e \in E(p)} W_E(e) \right\},$$

and $d_{\text{edge}}(v,w) = 0$ if $v = w$, where $\mathcal{P}_{\min}(v,w)$ is the set of all paths from v to w that attain the minimum number of edges among all such paths and $E(p)$ denotes the set of edges composing the path p.

01

Motivation

Remark. Let d_{edge} and d_{weight} be the distances defined above. Then the following inequality holds:

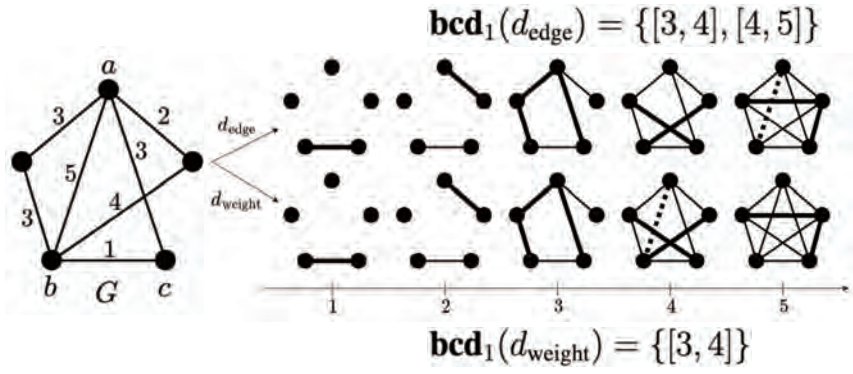
$$d_{\text{weight}}(v,w) \leq d_{\text{edge}}(v,w) \quad \text{for any } v, w \in V,$$

i.e., $d_{\text{weight}} \leq d_{\text{edge}}$.

Proof. Trivial, because $\mathcal{P}_{\min}(v,w) \subseteq \mathcal{P}(v,w)$ for all $v, w \in V$. Since the minimization in d_{weight} is taken over a superset of the one in d_{edge}, the inequality follows. □

01

Motivation



02

Does $d_1 \leq d_2$ implies that $\mathbf{bcd}_1(d_1) \subseteq \mathbf{bcd}_1(d_2)$?

No.

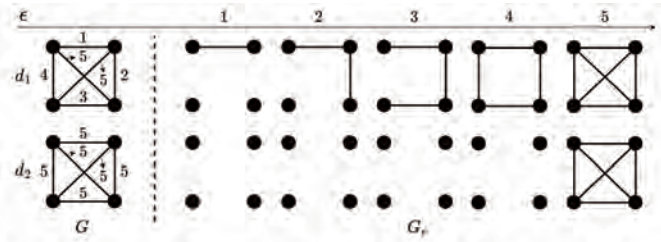


Fig. 2.1: An example where $d_1 \leq d_2$ yet $\mathbf{bcd}_1(d_1) \not\subseteq \mathbf{bcd}_1(d_2)$, with $\mathbf{bcd}_1(d_1) = \{[4,5]\}$ and $\mathbf{bcd}_1(d_2) = \emptyset$. The subgraph G_ϵ of G is defined as $G_\epsilon = (V, E_\epsilon)$, where $E_\epsilon = \{e \in E \mid W_E(e) \leq \epsilon\}$.

02

Does $d_1 \leq d_2$ implies that $\mathbf{bcd}_1(d_1) \subseteq \mathbf{bcd}_1(d_2)$?

No.

We know that $d_{\text{weight}} \leq d_{\text{edge}}$.

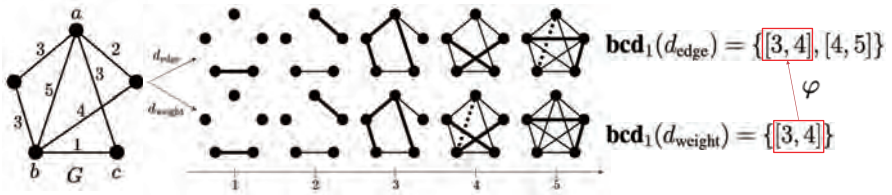
Does this implies that there is an injection from $\mathbf{bcd}_1(d_{\text{weight}})$ to $\mathbf{bcd}_1(d_{\text{edge}})$?

Yes.

Theorem. Let $G = (V, E, W_E)$ be a connected weighted graph. Then there exists an injective function $\varphi : \mathbf{bcd}_1(d_{\text{weight}}) \to \mathbf{bcd}_1(d_{\text{edge}})$ defined as $\varphi([\beta, \delta]) = [\beta, \delta']$ such that $\delta \leq \delta'$ for any $[\beta, \delta] \in \mathbf{bcd}_1(d_{\text{weight}})$.

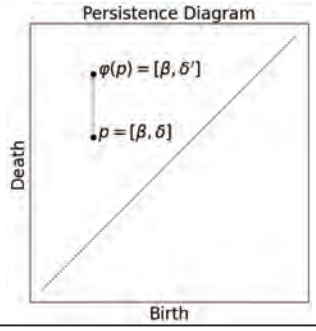
03 Discovery

Theorem. Let $G = (V, E, W_E)$ be a connected weighted graph. Then there exists an injective function $\varphi : \mathbf{bcd}_1(d_{\text{weight}}) \to \mathbf{bcd}_1(d_{\text{edge}})$ defined as $\varphi([\beta, \delta]) = [\beta, \delta']$ such that $\delta \leq \delta'$ for any $[\beta, \delta] \in \mathbf{bcd}_1(d_{\text{weight}})$.



03 Discovery

Theorem. Let $G = (V, E, W_E)$ be a connected weighted graph. Then there exists an injective function $\varphi : \mathbf{bcd}_1(d_{\text{weight}}) \to \mathbf{bcd}_1(d_{\text{edge}})$ defined as $\varphi([\beta, \delta]) = [\beta, \delta']$ such that $\delta \leq \delta'$ for any $[\beta, \delta] \in \mathbf{bcd}_1(d_{\text{weight}})$.



04 When ?

Path-representable distance

Let $G = (V, E, W_E)$ be a connected weighted graph. For any two vertices v, w in V, $P(v, w)$ is the set of paths from v to w that have no repeated vertex, and $\mathcal{P}$ is the union of $P(v, w)$ along all $v, w \in V$. That is,

$$\mathcal{P} = \bigcup_{v,w \in V} P(v, w)$$

A *path choice function* is a function $g : V \times V \to \mathcal{P}$ satisfying $g(a, b) \subseteq g(v, w) \in P(v, w)$ for any vertices v and w in G and any vertices a and b in $g(v, w)$. This property is called *consistency*.

Definition. Consider a connected weighted graph $G = (V, E, W_E)$. A function d on $V \times V$ is called a *path-representable distance* if there is a path choice function g such that

$$d(v, w) = \sum_{e \in g(v,w)} W_E(e) \text{ if } v \neq w$$

and $d(v, w) = 0$ if $v = w$.

04 When ?

Path-representable distance (example)

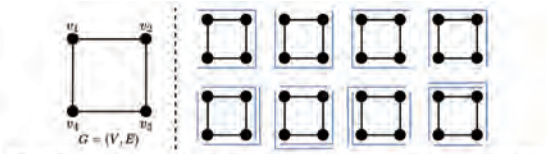


Fig. 3.1: The graph $G = (V, E)$ and all of its possible path choice functions represented by their maximal paths.

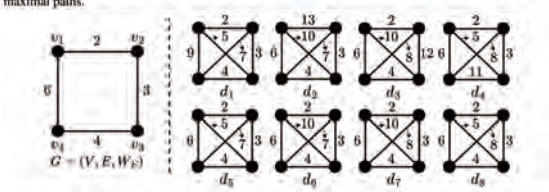


Fig. 3.2: The weighted graph $G = (V, E, W_E)$ and all its possible path-representable distances.

04 When ?

Dominated Path-representable distance

Let $G = (V, E, W_E)$ be a connected weighted graph and g a path choice function on G.

- For an edge $e = (v, w) \in E$:
  - g is *locally weight-dominated* by e if
    $$W_E(e) > W_E(e') \quad \text{for all } e' \in g(v, w)$$
  - g is *locally cost-dominated* by e if
    $$W_E(e) \geq \sum_{e' \in g(v,w)} W_E(e')$$
- We define:
  - g is *weight-dominated* if it is locally weight-dominated for all $e \neq g(v, w)$
  - g is *cost-dominated* if it is locally cost-dominated for all $e \neq g(v, w)$

Let d be a path-representable distance defined by path choice function g on G.

- d is *weight-dominated* if g is weight-dominated
- d is *cost-dominated* if g is cost-dominated

04 When ?

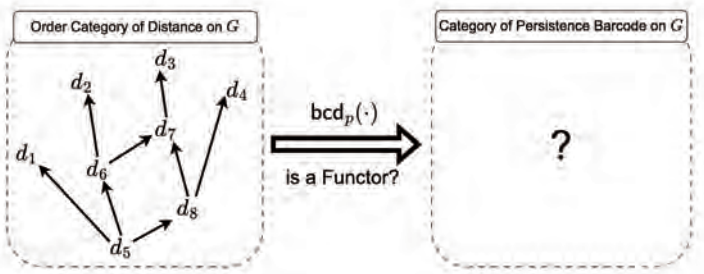
Main Theorem

Theorem. Let $G = (V, E, W_E)$ be a connected weighted graph. Consider two cost-dominated path-representable distances d_i and d_j by path choice functions g_i and g_j on G, respectively. If $d_i(v, w) \leq d_j(v, w)$ for any vertices v and w, then there exists an injective function $\varphi_{i,j} : \mathbf{bcd}_1(d_i) \to \mathbf{bcd}_1(d_j)$ defined as $\varphi_{i,j}([\beta, \delta]) = [\beta, \delta']$ such that $\delta \leq \delta'$ for any $[\beta, \delta] \in \mathbf{bcd}_1(d_i)$.

Proposition. The two distances d_{weight} and d_{edge} are cost-dominated path-representable distances such that $d_{\text{weight}} \leq d_{\text{edge}}$.

04 When ?

Categorical interpretation



04 When ?

Categorical interpretation

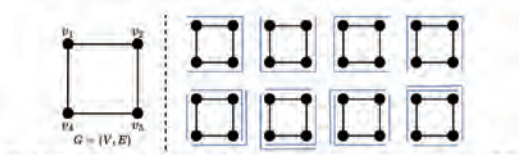


Fig. 3.1: The graph $G = (V, E)$ and all of its possible path choice functions represented by their maximal paths.

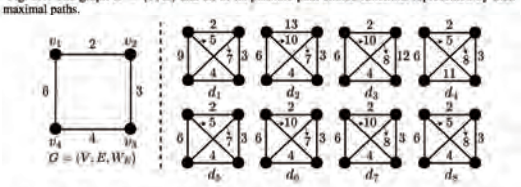
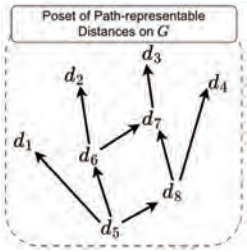
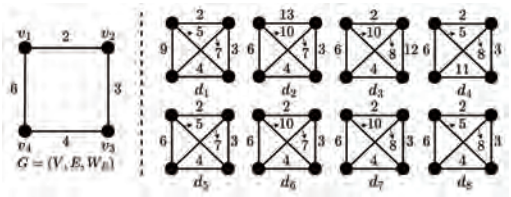


Fig. 3.2: The weighted graph $G = (V, E, W_E)$ and all its possible path-representable distances.

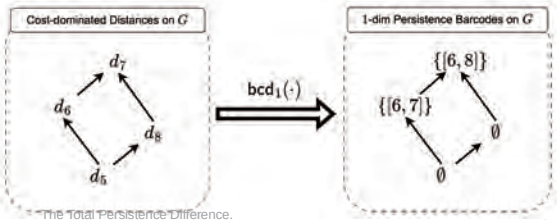


04 When ?

Categorical interpretation



$\mathbf{bcd}_1(\cdot)$ is a well-defined faithful functor.



05 High-dimensional case

STATEMENT 5.1. *Let $G = (V, E, W_E)$ be a connected weighted graph. Consider two cost-dominated path choice functions g_1 and g_2 on G, and its two path-representable distances d_1 and d_2 along g_1 and g_2. For any dimension k if $d_2(v, w) \leq d_1(v, w)$ for any vertices v, w, then there exists an injective function $\varphi : \mathbf{bcd}_k(d_2) \to \mathbf{bcd}_k(d_1)$.*

05 High-dimensional case

STATEMENT 5.1. *Let $G = (V, E, W_E)$ be a connected weighted graph. Consider two cost-dominated path choice functions g_1 and g_2 on G, and its two path-representable distances d_1 and d_2 along g_1 and g_2. For any dimension k if $d_2(v, w) \leq d_1(v, w)$ for any vertices v, w, then there exists an injective function $\varphi : \mathbf{bcd}_k(d_2) \to \mathbf{bcd}_k(d_1)$.*

Table 5.1: True or False for high dimensional cases k for Statement 5.1

Dimension k	0	1	2	3	4	5
(T)rue or (F)alse	T	T	F	F	F	F

Note: This table shows the results of Statement 5.1 for different dimensional cases. The table indicates whether the theorem holds (True) or does not hold (False) for each dimension.

06 Application : the total persistence difference (TPD)

The total persistence difference (TPD)

Definition. Let $G = (V, E, W_E)$ be a connected weighted graph equipped with two cost-dominated path-representable distances d_i and d_j such that $d_i \leq d_j$. Then the total persistence difference between their 1-dimensional persistence diagrams is defined as:

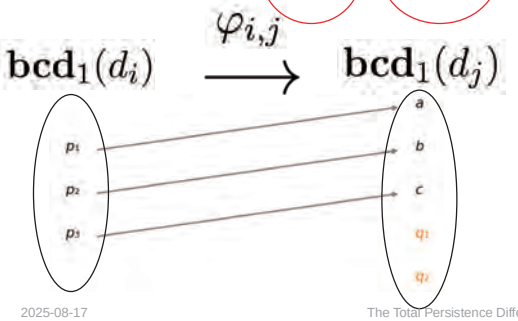
$$\mathrm{Diff}_p(\mathrm{PD}_1(G, d_j),\ \mathrm{PD}_1(G, d_i)) := \left(\sum_{\sigma \in \mathcal{B}_i} (\delta_\sigma^j - \delta_\sigma^i)^p + \sum_{\sigma \in \mathcal{B}_j \setminus \mathcal{B}_i} (\delta_\sigma^j - \beta_\sigma^j)^p \right)^{\frac{1}{p}},$$

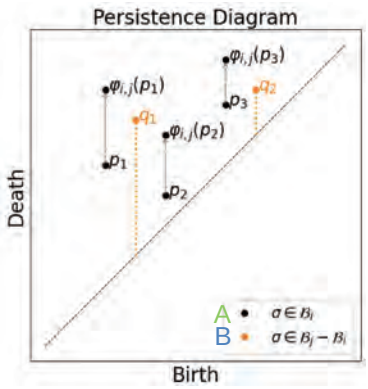
where $\mathcal{B}_i$ and $\mathcal{B}_j$ are the sets of birth simplices (edges) in the persistence diagrams of d_i and d_j, respectively. The values $\beta_\sigma^\bullet$ and $\delta_\sigma^\bullet$ denote the birth and death values of the cycle associated to the birth edge σ of d_i and d_j, respectively.

06 Application : the total persistence difference (TPD)

The total persistence difference (TPD)

$$\mathrm{Diff}_p(\mathrm{PD}_1(G, d_j),\ \mathrm{PD}_1(G, d_i)) := \left(\underbrace{\sum_{\sigma \in \mathcal{B}_i} (\delta_\sigma^j - \delta_\sigma^i)^p}_{\text{A}} + \underbrace{\sum_{\sigma \in \mathcal{B}_j \setminus \mathcal{B}_i} (\delta_\sigma^j - \beta_\sigma^j)^p}_{\text{B}} \right)^{\frac{1}{p}}$$





06 Application : the total persistence difference (TPD)

Stability Theorem for the total persistence difference (TPD)

Theorem. Let $G = (V, E, W_E)$ be a connected weighted graph equipped with two cost-dominated path-representable distances d_i and d_j with $d_i \leq d_j$. Then the total persistence difference between their 1-dimensional persistence diagrams $\mathrm{PD}_1(G, d_j)$ and $\mathrm{PD}_1(G, d_i)$ is bounded above:

$$\mathrm{Diff}_p\big(\mathrm{PD}_1(G, d_j),\ \mathrm{PD}_1(G, d_i)\big) \leq \gamma \, \|d_j - d_i\|_\infty ,$$

where $\|d_j - d_i\|_\infty := \max_{(v,w)\in V\times V}\{d_j(v,w) - d_i(v,w)\}$ and $\gamma = \left(\big||E| - |V| + 1\big|\right)^{\frac{1}{p}}$.

06 Application : the total persistence difference (TPD)

Experiments and Applications (prior work)

Hajij, Mustafa, et al. "Visual detection of structural changes in time-varying graphs using persistent homology." *2018 ieee pacific visualization symposium (pacificvis)*. IEEE, 2018.

- **Dataset:** SNAP EU Email Network (525 days, daily temporal graphs)

To measure topological complexity in a dynamic network, prior work employs an **overlapping sliding window** approach and compares the persistent homology of consecutive subgraphs.

- Given a time-varying graph $G = \{G_0, G_1, \ldots, G_T\}$, they measure topological change over time using the Wasserstein distance between consecutive barcodes:

$$D_t := W_1\big(\mathbf{bcd}_1(G_t),\ \mathbf{bcd}_1(G_{t+1})\big)$$

- The sequence $\{D_t\}_{t=0}^{T-w}$ serves as a **topological complexity signal**, indicating structural changes in the evolving network.

Limitation: Since G_t and G_{t+1} are distinct graphs, the measured topological difference tends to reflect the *degree of change* between them, rather than capturing the *intrinsic complexity* of each individual graph.

06 Application : the total persistence difference (TPD)

Experiments and Applications (prior work)

Hajij, Mustafa, et al. "Visual detection of structural changes in time-varying graphs using persistent homology." *2018 ieee pacific visualization symposium (pacificvis)*. IEEE, 2018.

- **Dataset:** SNAP EU Email Network (525 days, daily temporal graphs)

To measure topological complexity in a dynamic network, prior work employs an **overlapping sliding window** approach and compares the persistent homology of consecutive subgraphs.

- Given a time-varying graph $G = \{G_0, G_1, \ldots, G_T\}$, they measure topological change over time using the Wasserstein distance between consecutive barcodes:

$$D_t := W_1\big(\mathbf{bcd}_1(G_t),\ \mathbf{bcd}_1(G_{t+1})\big)$$

- The sequence $\{D_t\}_{t=0}^{T-w}$ serves as a **topological complexity signal**, indicating structural changes in the evolving network.

Limitation: Since G_t and G_{t+1} are distinct graphs, the measured topological difference tends to reflect the *degree of change* between them, rather than capturing the *intrinsic complexity* of each individual graph.

06 Application : the total persistence difference (TPD)

Experiments and Applications (proposed method)

- **Dataset:** SNAP EU Email Network (525 days, daily temporal graphs)

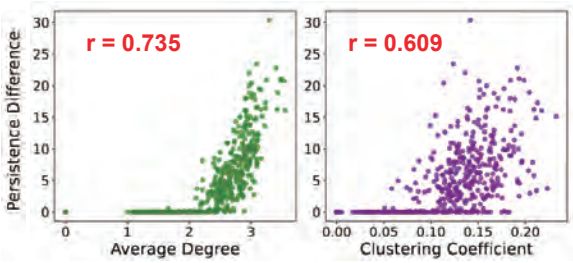
In contrast, our method compares two filtrations on the same graph snapshot G_t, each induced by a different path-representable distance:

$$\mathrm{TPD}_t := \mathrm{Diff}_1\big(\mathbf{bcd}_1(G_t, d_{\mathrm{edge}}),\ \mathbf{bcd}_1(G_t, d_{\mathrm{weight}})\big),$$

thereby quantifying the *intrinsic topological complexity* of the graph itself.

06 Application : the total persistence difference (TPD)

The Pearson correlation for the classical graph statistics



06 Application : the total persistence difference (TPD)

Comparison with Prior Work

Comparison: Our Method vs Prior Work (Hajij et al., 2018)

Aspect	Prior Work	Proposed Method (TPD)
Comparison Basis	Different graphs (overlapping windows)	Same graph with different distance functions
Distance Used	Single (weighted)	d_{edge} and d_{weight}
Weekend Sensitivity	Non-zero values even with minimal structure	**Naturally zero on weekends and holidays**
Interpretability	Hard to isolate cause of change	**Clear attribution to distance-induced differences**
Correlation to Graph Stats	Avg. degree $r = 0.36$	**Stronger:** $r = 0.74$ **(avg. degree)**

Summary

- **Motivation**
  Persistent Homology (PH) is increasingly used in graph data analysis. However, PH results vary significantly depending on the choice of graph distance, and this phenomenon is poorly understood.

- **Main Findings**
  We discovered a nontrivial 1D PH embedding between two commonly used graph distances d_{edge} and d_{weight}. To explain this, we define a broader class of distances — **path-representable distances** — and mathematically characterize when such embeddings occur.

- **Application: Total Persistence Difference (TPD)**
  A new topological measure to quantify the change between filtrations induced by different distances:

$$\text{TPD}_t = \text{Diff}_1\big(\text{PD}_1(G_t, d_{\text{edge}}),\ \text{PD}_1(G_t, d_{\text{weight}})\big)$$

  We also prove a **stability theorem** for TPD when distances are ordered and path-representable.

Thank you

Eunwoo Heo
with Byeongchan Choi, Jae-hun Jung

Department of Mathematics
POSTECH / Ph.D. (conferred Aug 8, 2025)
hew0920@postech.ac.kr

Pathological State Inference System based on Mathematical Model and TDA for Personalized Treatment in Dermatology

Sungrim Seirin-Lee

Kyoto University Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Japan

Skin diseases typically appear as visible information-skin eruptions distributed across the body. However, the biological mechanisms underlying these manifestations are often inferred from fragmented, time-point-specific data such as skin biopsies. The challenge is further compounded for human-specific conditions like urticaria, where animal models are ineffective, leaving researchers to rely heavily on in vitro experiments and sparse clinical observations. In this presentation, I will introduce an innovative methodology that combines mathematical modeling with topological data analysis, allowing for the estimation of patient-specific parameters directly from morphological patterns of skin eruptions. This framework offers a new pathway for personalized analysis and mechanistic insight into complex skin disorders.

References.

[1] Akanksha Maurya et al. "Hybrid topological data analysis and deep learning for basal cell carcinoma diagnosis". In: *Journal of Imaging Informatics in Medicine* 37.1 (2024), pp. 92–106.

[2] John T Nardini et al. "Topological data analysis distinguishes parameter regimes in the Anderson-Chaplain model of angiogenesis". In: *PLOS Computational Biology* 17.6 (2021), e1009094.

[3] Sungrim Seirin-Lee et al. "Mathematical-based morphological classification of skin eruptions corresponding to the pathophysiological state of chronic spontaneous urticaria". In: *Communications Medicine* 3.1 (2023), p. 171.

[4] Thomas Thorne, Paul DW Kirk, and Heather A Harrington. "Topological approximate Bayesian computation for parameter inference of an angiogenesis model". In: *Bioinformatics* 38.9 (2022), pp. 2529–2535.

A Topological Analysis of the Space of Recipes

Emerson Escolar

Graduate School of Human Development and Environment, Kobe University, Japan

In recent years, the use of data-driven methods has provided insights into underlying patterns and principles behind culinary recipes. In this exploratory work, we introduce the use of topological data analysis, especially persistent homology, in order to study the space of culinary recipes. In particular, persistent homology analysis provides a set of recipes surrounding the multiscale "holes" in the space of existing recipes. We then propose a method to generate novel ingredient combinations using combinatorial optimization on this topological information. We made biscuits using the novel ingredient combinations, which were confirmed to be acceptable enough by a sensory evaluation study. Our findings indicate that topological data analysis has the potential for providing new tools and insights in the study of culinary recipes. This talk is based on https://doi.org/10.1016/j.ijgfs.2024.101088

References.

[1] Yong-Yeol Ahn et al. "Flavor network and the principles of food pairing". In: *Scientific reports* 1.1 (2011), p. 196.

[2] Emerson G Escolar, Yuta Shimada, and Masahiro Yuasa. "A topological analysis of the space of recipes". In: *International Journal of Gastronomy and Food Science* 39 (2025), p. 101088.

[3] Juan CS Herrera. "The contribution of network science to the study of food recipes. A review paper". In: *Appetite* 159 (2021), p. 105048.

[4] Donghyeon Park et al. "Kitchenette: Predicting and recommending food ingredient pairings using siamese neural networks". In: *arXiv preprint arXiv:1905.07261* (2019).

A topological analysis of the space of recipes

ESCOLAR, Emerson G.
Kobe University Graduate School of Human Development and Environment

Escolar, E. G., Shimada, Y., & Yuasa, M. (2025). A topological analysis of the space of recipes. *International Journal of Gastronomy and Food Science*, *39*, 101088.

Background for today's talk
Data science & Recipes

- Ahn, Yong-Yeol, et al. "Flavor network and the principles of food pairing." Scientific reports 1.1 (2011): 196.
- Teng, Chun-Yuen, Yu-Ru Lin, and Lada A. Adamic. "Recipe recommendation using ingredient networks." Proceedings of the 4th annual ACM web science conference. 2012.
- H. Lee, Helena, et al. "RecipeGPT: Generative pre-training based cooking recipe generation and evaluation system." Web Conf. WWW 2020. 2020.
- Marın, Javier, et al. "Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.1 (2021): 187-203.
- Goel, Mansi, and Ganesh Bagler. "Computational gastronomy: A data science approach to food." Journal of Biosciences 47.1 (2022): 12.
- etc.
- This research: study the "holes" of recipe space, try to get hints from it to create new combinations of ingredients
  - => Use topological data analysis

Topological Data Analysis

Keyword: the "shape" of data

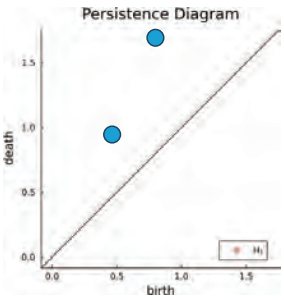
Clusters: Cluster analysis

“Holes” in data

Chazal, Frédéric, and Bertrand Michel. *Frontiers in artificial intelligence* 4 (2021): 667963.

Persistent Homology

A tool in topological data analysis

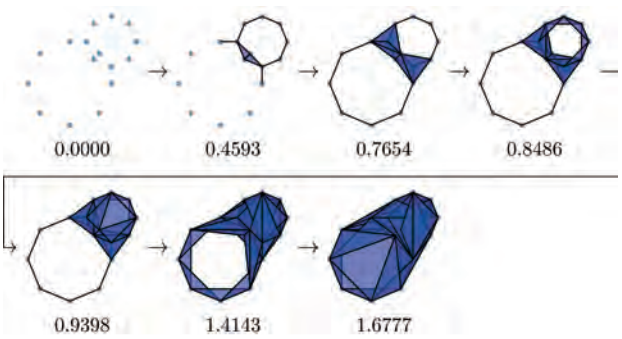


Two holes detected.

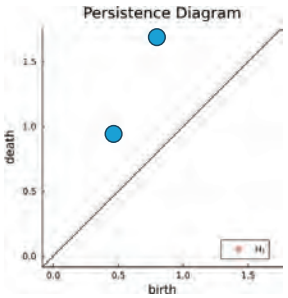
[Edelsbrunner; Letscher; Zomorodian ‘02], [Zomorodian; Carlsson ‘04]

Persistent Homology

A tool in topological data analysis



Vietoris-Rips filtration



Two holes detected
With birth and death:
(0.4593, 0.9398)
(0.7654, 1.6777)

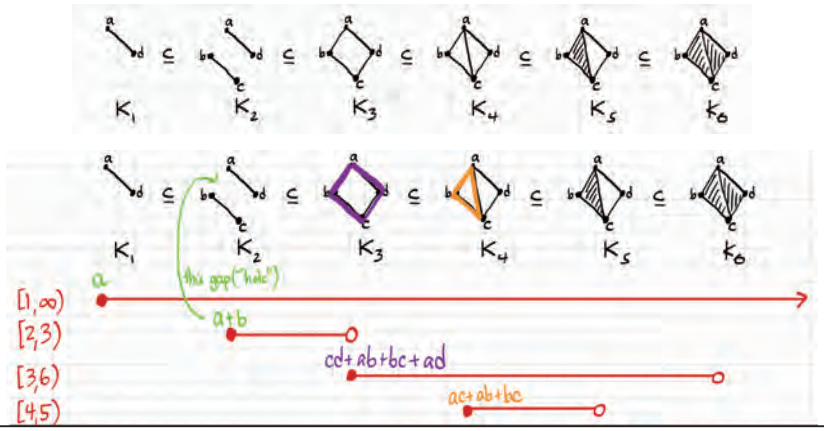
[Edelsbrunner; Letscher; Zomorodian '02], [Zomorodian; Carlsson '04]

In a bit more detail…

- Given a filtration of a finite simplicial complex, there exists a set q-cycles $\{z_i\}_{i=1}^{s}$ and a unique multiset of pairs $\{[b_i, d_i)\}_{i=1}^{s}$ such that the following hold.
  - (for each i) z_i is "born" at b_i in the filtration (birth)
  - (for each i) z_i "dies" at d_i in the filtration (death)
  - For each threshold value t, the homology classes of the z_i "alive" at t (z_i with $b_i \leq t < d_i$) forms a basis for the homology of the simplicial complex at t (basis)

- $\{[b_i, d_i)\}_{i=1}^{s}$ is called the qth persistence diagram
- Each z_i is called a representative cycle – represents some q-dimensional hole (using the interpretation of homology)

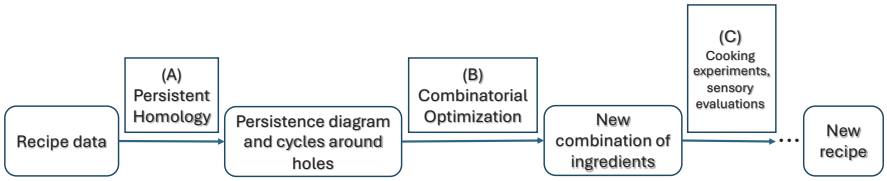
Computed example:



This research: study the “holes” of recipe space, try to get hints from it to create new combinations of ingredients.

The steps of the analysis

- (A) Apply persistent homology to analyze the “space of recipes”
- (B) From the “holes” in the space of recipes, create new combinations of ingredients that can be used as hints for new recipes
- (C) Check using cooking experiments, sensory evaluations



The data

- Ahn, Yong-Yeol, et al. "Flavor network and the principles of food pairing." Scientific reports 1.1 (2011): 196

- Use only list of ingredients
- as 0–1 vectors
- dissimilarity: $d_{\cos}(y,x) = 1 - \frac{\langle y,x \rangle}{\|y\|\|x\|}$
  = 1 – cosine(angle between x and y)

Number of recipes	48,983
Number of ingreds.	381
Ave. num. ingreds. per recipe	8.4936
Std. dev. num. ingreds. per recipe	3.5091

Table 1: Basic Statistics

Ingredient combinations as Vectors

$$\begin{bmatrix} \text{Beef} \\ \text{Potato} \\ \text{Carrot} \\ \text{Onion} \\ \text{Shirataki} \\ \text{Water} \\ \text{Soy sauce} \\ \text{Mirin} \\ \text{Sugar} \\ \text{Hondashi} \\ \text{Sake} \\ \text{Tofu} \\ \vdots \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \quad \cdots$$

Nikujaga
https://cookpad.com/recipe/1519259

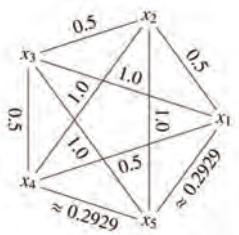
Sukiyaki
https://cookpad.com/recipe/7746800

Example...

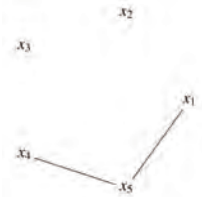
- $x_1 = [1, 0, 0, 1]$
- $x_2 = [0, 0, 1, 1]$
- $x_3 = [0, 1, 1, 0]$
- $x_4 = [1, 1, 0, 0]$
- $x_5 = [1, 0, 0, 0]$

Pairwise dissimilarities using

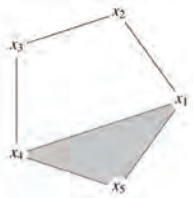
$$d_{\cos}(y, x) = 1 - \frac{\langle y, x \rangle}{\|y\| \|x\|}$$



Vietoris-Rips filtration:

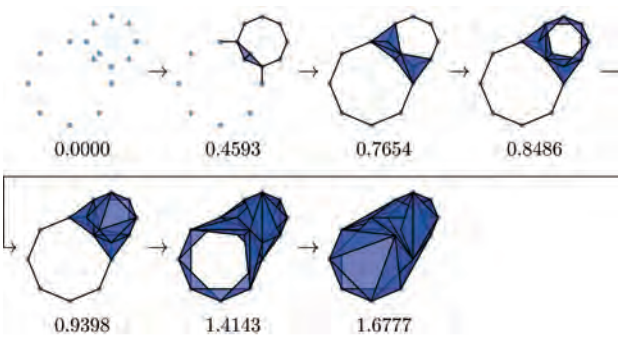


Threshold $t = 1 - \frac{1}{\sqrt{\{2\}}}$

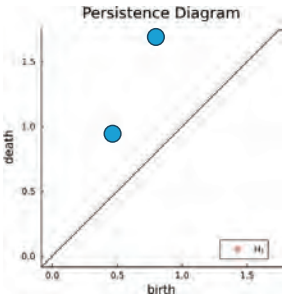


Threshold $t = 0.5$

Recall: Persistent Homology

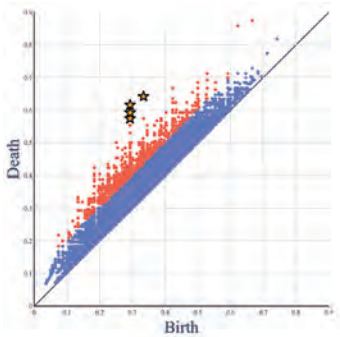


Vietoris-Rips filtration

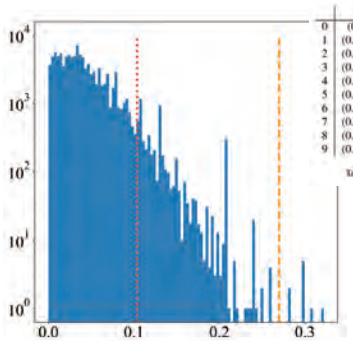


Two holes detected
With birth and death:
(0.4593, 0.9398)
(0.7654, 1.6777)

Persistent homology, applied to recipe data



Degree-1 persistence diagram of the recipe data.



	birth-death pair	lifespan
0	(0.292893, 0.6151)	0.322207
1	(0.333333, 0.641431)	0.308098
2	(0.292893, 0.591752)	0.298859
3	(0.292893, 0.591752)	0.298859
4	(0.292893, 0.591752)	0.298859
5	(0.292893, 0.591752)	0.298859
6	(0.292893, 0.591752)	0.298859
7	(0.292893, 0.573599)	0.280706
8	(0.292893, 0.573599)	0.280706
9	(0.292893, 0.552786)	0.259893

Table A.3: Top 10 largest lifespans

Histogram of the lifespans of the birth-death pairs.

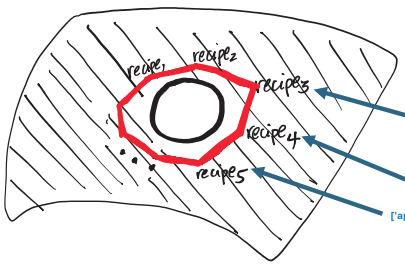
[Edelsbrunner; Letscher; Zomorodian '02], [Zomorodian; Carlsson '04]

In a bit more detail…

- Given a filtration of a finite simplicial complex, there exists a set q-cycles $\{z_i\}_{i=1}^{s}$ and a unique multiset of pairs $\{[b_i, d_i)\}_{i=1}^{s}$ such that the following hold.
  - (for each i) z_i is "born" at b_i in the filtration (birth)
  - (for each i) z_i "dies" at d_i in the filtration (death)
  - For each threshold value t, the homology classes of the z_i "alive" at t (z_i with $b_i \leq t < d_i$) forms a basis for the homology of the simplicial complex at t (basis)

- $\{[b_i, d_i)\}_{i=1}^{s}$ is called the qth persistence diagram
- Each z_i is called a representative cycle – represents some q-dimensional hole (using the interpretation of homology)

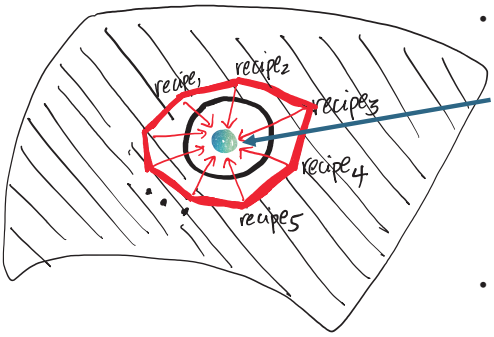
Representative cycle z_i: List of recipes surrounding a hole



Just an "image", do not take literally

['butter', 'cinnamon', 'cream', 'egg', 'milk', 'nutmeg', 'vanilla', 'wheat']
['butter', 'cane_molasses', 'milk', 'rye_flour', 'wheat', 'yeast']
['cherry', 'cream', 'egg', 'gelatin', 'wheat']
['butter', 'olive', 'olive_oil', 'wheat']
['butter', 'cane_molasses', 'egg', 'oat', 'raisin', 'vanilla', 'walnut', 'wheat']
['gin', 'lemon']
['butter', 'cane_molasses', 'cinnamon', 'cream', 'egg', 'milk', 'vanilla', 'wheat']
['apple', 'butter', 'cane_molasses', 'cinnamon', 'egg', 'milk', 'vanilla', 'walnut', 'wheat']
['lemon', 'lime', 'orange_juice']
['butter', 'wheat', 'yeast']
['cherry', 'lemon', 'lime', 'orange', 'orange_juice', 'pineapple']
['butter', 'cane_molasses', 'milk', 'wheat', 'whole_grain_wheat_flour', 'yeast']
['butter', 'cream', 'milk', 'potato', 'wheat']
['butter', 'cane_molasses', 'egg', 'milk', 'rye_flour', 'wheat', 'yeast']
['butter', 'cane_molasses', 'cinnamon', 'egg', 'milk', 'pecan', 'vanilla', 'vegetable_oil', 'wheat']
['cherry', 'cream', 'gelatin', 'wheat']
['gin']
['butter', 'cane_molasses', 'cinnamon', 'egg', 'lard', 'milk', 'oat', 'vanilla', 'walnut', 'wheat']
['butter', 'milk', 'wheat', 'yeast']
['butter', 'cane_molasses', 'cocoa', 'coconut', 'egg', 'milk', 'oat', 'vanilla', 'walnut', 'wheat']
['butter', 'honey', 'milk', 'wheat', 'whole_grain_wheat_flour', 'yeast']
['butter', 'lard', 'milk', 'wheat']
['butter', 'cane_molasses', 'cinnamon', 'egg', 'lard', 'milk', 'oat', 'vanilla', 'wheat']
['butter', 'cherry', 'cream', 'gelatin', 'wheat']
['butter', 'cane_molasses', 'cinnamon', 'egg', 'lard', 'oat', 'raisin', 'vanilla', 'walnut', 'wheat']
['apple', 'butter', 'cane_molasses', 'cinnamon', 'egg', 'milk', 'pecan', 'vanilla', 'vegetable_oil', 'walnut', 'wheat']
['butter', 'cane_molasses', 'cinnamon', 'cocoa', 'egg', 'milk', 'vanilla', 'wheat', 'yeast']
['apple', 'butter', 'cane_molasses', 'cinnamon', 'egg', 'milk', 'pecan', 'vanilla', 'vegetable_oil', 'wheat']
['butter', 'egg', 'milk', 'nutmeg', 'raisin', 'wheat']
['gin', 'lemon', 'orange_juice']
['butter', 'cane_molasses', 'cocoa', 'egg', 'milk', 'vanilla', 'walnut', 'wheat']
['butter', 'cane_molasses', 'coconut', 'egg', 'oat', 'raisin', 'vanilla', 'walnut', 'wheat']
['butter', 'cream', 'wheat']
['cherry', 'lemon', 'lime', 'orange', 'pineapple']
['gin', 'lemon', 'lime', 'orange_juice', 'tea']
. . .

Can we use it to make a new recipe? Center?



- Average number of ingredients in original data: 8.49±3.51
- Ingredients in this cycle:
  ['almond', 'bell_pepper', 'black_pepper', 'butter', 'cane_molasses', 'champagne_wine', 'cheddar_cheese', 'cherry', 'chicken', 'cinnamon', 'cocoa', 'cranberry', 'cream', 'cream_cheese', 'egg', 'garlic', 'gelatin', 'lard', 'lemon', 'lime', 'milk', 'oat', 'onion', 'orange', 'orange_juice', 'pepper', 'pineapple', 'potato', 'raspberry', 'rum', 'seed', 'vanilla', 'walnut', 'wheat', 'yeast']
  35 ingredients
- Too many ingredients!

19

Finding new combinations using optimization

- Let S be the set of all ingredients used in the recipes surrounding a hole. Think of these as candidate ingredients
- Consider the following problem:

$$y_* = \underset{y \subset S,\, |y|=t}{\operatorname{argmax}}\, d_{\cos}(y, X) = \underset{y \subset S,\, |y|=t}{\operatorname{argmin}}\, \max_{x \in X} \frac{\langle y, x \rangle}{\|y\| \|x\|}$$

- Find the combination y that is the most dissimilar to the set of existing recipes X
- Under the condition that y is exactly t ingredients from the candidate set S

Finding new combinations using optimization

Problem formulation (original):

$$y_* = \underset{y \subset S,\, |y|=t}{\operatorname{argmax}}\, d_{\cos}(y, X) = \underset{y \subset S,\, |y|=t}{\operatorname{argmin}}\, \max_{x \in X} \frac{\langle y, x \rangle}{\|y\| \|x\|}$$

- Using the "epigraph trick", we turn min max problem into a minimization problem:
  The $\max_{x \in X} \frac{\langle y,x \rangle}{\|y\| \|x\|}$ is equal to $\min \left\{ \lambda \middle| \lambda \geq \frac{\langle y,x \rangle}{\|y\| \|x\|} \;\; \forall x \in X \right\}$
- We also apply a projection to S to decrease the dimensionality of the problem

Finding new combinations using optimization
(Mixed Integer Linear Programming Problem)

- Thus, equivalent to the following Problem:

$$\begin{cases} \text{minimize} & \lambda \\ \text{subject to} & v^T z - \lambda \leq 0 \qquad \forall\, v \in \left\{ \dfrac{\pi_S(x)}{\sqrt{t}\|x\|} \,\middle|\, x \in X \right\} \\ & 1^T z = t \\ & z \in \{0,1\}^s, \qquad \lambda \in \mathbb{R} \end{cases}$$

- Use software (e.g. GLPK; coin-or/Cbc; IBM® ILOG® CPLEX®) to solve

Some example optimal solutions (with $t = 5$)

- ...
  ('cranberry', 'cream cheese', 'gin', 'olive oil', 'raisin')
  ('cranberry', 'cream cheese', 'gin', 'raisin', 'starch')
  ('cranberry', 'cream cheese', 'gin', 'raisin', 'whole grain wheat flour')
  ('cranberry', 'cream cheese', 'gin', 'starch', 'whole grain wheat flour')
  ('cranberry', 'cream cheese', 'raisin', 'starch', 'whole grain wheat flour')
  ...

- many optimal solutions with the same dissimilarity to existing recipes!

Biscuits preparation

- Four solutions (NCS, NR, NG, NCB) with $t = 5$ ingredients were selected.
- The ingredients suggested biscuits.

Table: Compositions of the biscuits

	Control	No corn starch (NCS)	No raisin (NR)	No gin (NG)	No cranberry (NCB)
Sugar*	30	30	30	30	30
Whole grain wheat flour	45	90	45	45	45
Starch (cornstarch)	45	-	45	45	45
Cranberry (dried cranberry)	10	10	20	10	-
Raisin	10	10	-	10	20
Gin	10	10	10	-	10
Water	-	-	-	10	-
Cream cheese	80	80	80	80	80

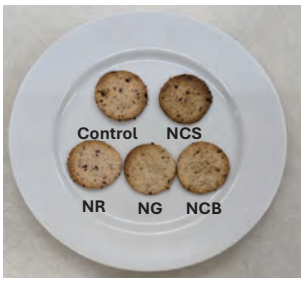
*Sugar was added because the recipe data does not contain seasonings (e.g. salt and sugar) as ingredients

24

Sensory evaluation of biscuits

Comparison sample

Control, NCS, NR, NG, NCB



Panel

19 untrained non-expert Japanese male and female students (21.3 ± 2.6 years old)

Factors

Scoring method: -3 ~ +3

Intensity of color, texture, sweetness, sourness

Preferences for aroma, color, texture, taste, palatability, overall judgment

Results of the sensory evaluation

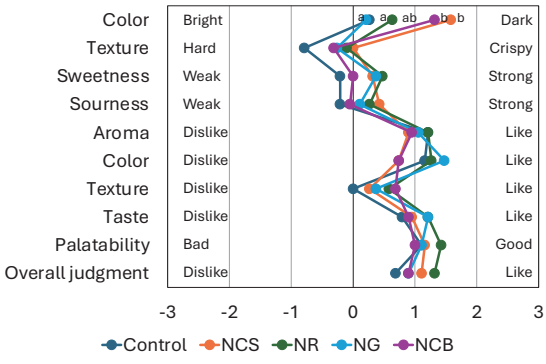


Figure 1: Sensory properties of biscuits

Mean (n=19).
Only "color" exhibits significant difference ($p < 0.05$, Tukey's HSD test)

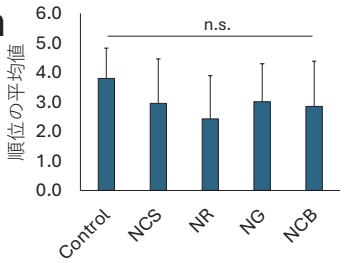


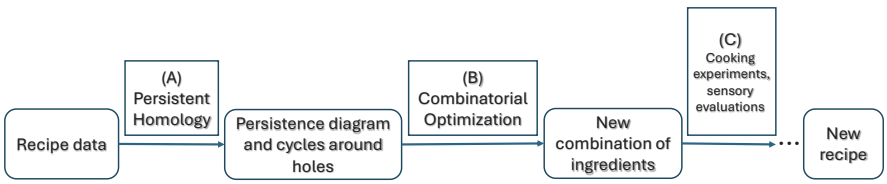
Figure 2: Ranking order

Mean (n=19). No significant difference in ranking order of biscuits, by Newell and MacFarlane test

Scores of palatability and overall judgment of all biscuits were about 1.0, indicating that the suggested ingredient combinations are potentially viable for recipes of biscuits

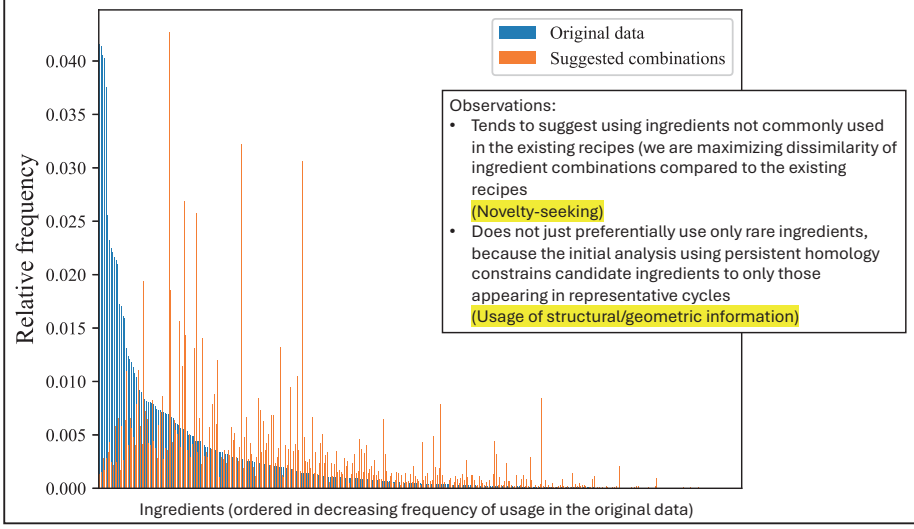
Recall: The steps of the analysis

- (A) Apply persistent homology to analyze the "space of recipes"
- (B) From the "holes" in the space of recipes, create new combinations of ingredients that can be used as hints for new recipes
- (C) Check using cooking experiments, sensory evaluations



Testing novelty (simple check)

- Generate 31,478 suggestions from our method (up to 20 solutions for the representative cycles of the top 5% birth–death pairs)
- Result:
  - ➢Equal to existing recipes: 61 (0.19%)
  - ➢Strict subcombinations of existing recipes: 506 (1.6%)
- Point: most of the suggestions from our method do not fall into either of these cases
- Next slide: look at the big picture of ingredient usage patterns



Summary

- Analysis of recipes
  - Persistent homology to explore geometric structure
  - Combinatorial optimization to find new combinations
- TDA (+optimization) has the potential for providing new insights in the study of culinary recipes

Thank you for your attention!

Thanks to Funding:

Casio Science Promotion Foundation, Japan (Grant number 40-29)

Research Grant of Graduate School of Human Development and Environment, Kobe University

JSPS Grant-in-Aid for Transformative Research Areas (A) (22H05105)
"Establishing data descriptive science and its cross-disciplinary applications"

JSPS Grant-in-Aid for Scientific Research (C) (24K06846)
"Development of new methods for data analysis of processes using topological ideas"

Ellipse Cloud: Anisotropy-Aware Persistent Homology

Tomoki Uda

Faculty of Science, University of Toyama, Japan

Persistent homology is a widely used tool in topological data analysis, yet standard filtration methods often fail to capture anisotropic structures inherent in real-world data. We propose Ellipse Cloud, a preprocessing-based approach that enhances anisotropic features in persistent homology. Our method constructs a Vietoris–Rips (VR) filtration using ellipse tangency times instead of pairwise Euclidean distances, extending the standard VR filtration to an anisotropic setting. This formulation allows anisotropy to be incorporated into persistent homology while remaining compatible with standard computational frameworks. A key computational challenge in this framework involves determining critical time points at which expanding ellipses first interact, which we address through an efficient numerical algorithm.

To evaluate the effectiveness of our approach, we apply it to a toy problem involving a highly noisy two-dimensional point cloud with multiple ring structures. While standard persistent homology struggles to capture the underlying rings due to excessive noise, our anisotropic filtration successfully identifies optimal 1-cycles that preserve the original structures to a greater extent. More generally, our proposed preprocessing technique tends to increase the lifetime of significant persistence pairs, lowering birth values and raising death values compared to standard VR filtrations. These results suggest that incorporating anisotropic filtrations can provide more informative topological summaries of geometrical structures in data. Potential applications include sensor coverage problems, where sensors often exhibit directional sensitivity rather than isotropic coverage.

In the talk, we will also introduce the Python library 'EllPHi' for anisotropic persistent homology analysis. 'EllPHi' provides the fast and accurate ellipse-tangency solver. The related source codes are available in [GitHub (https://github.com/t-uda/ellphi)](https://github.com/t-uda/ellphi).

References.

[1] Vincent Peter Grande and Michael T Schaub. "Non-isotropic persistent homology: Leveraging the metric dependency of ph". In: *Learning on Graphs Conference*. PMLR. 2024, pp. 17–1.

[2] t-uda. *EllPHi: A Fast Ellipse-Tangency Solver for Anisotropic Persistent Homology*. `https://github.com/t-uda/ellphi`. Open-source software, MIT License. 2025.

Ellipse Cloud:

Anisotropy-Aware Persistent Homology

Tomoki Uda
Faculty of Science
University of Toyama
（富山大学学術研究部理学系数理情報学プログラム）

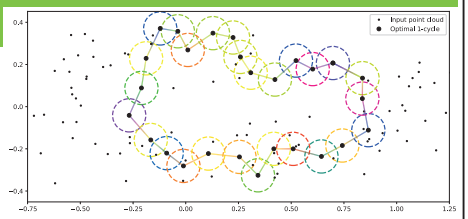
2025-08-06 (Wed) TDA+IM @ Fukuoka

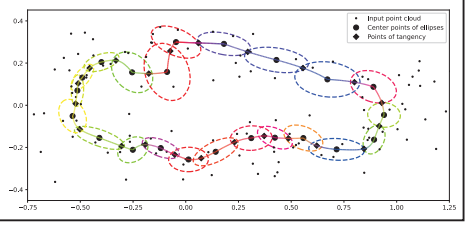
Persistent Homology (PH)

Standard PH
w/ circular disks

?

Anisotropic PH
w/ ellipses



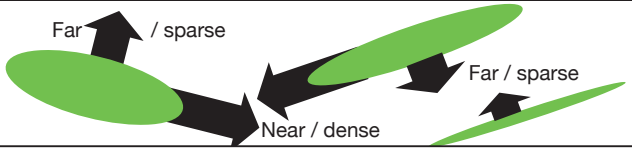


Definition. Anisotropy?

Given spatial data with a specific metric, it is said to exhibit **anisotropy** if its local properties differ depending on the direction of measurement.

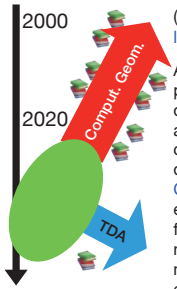
Typical example:
Sensors or radars have directivity.

We hereafter focus on anisotropic distance structure.



Preceding Studies

TL;DR: The existing TDA methods have not well explored anisotropy yet. Much research should still be done in this direction.



While computational geometry has explored anisotropic approaches (e.g., anisotropic Voronoi diagrams), such ideas have been less extensively developed within persistent homology frameworks.

A TDA-based approach, Non-isotropic Persistent Homology (NIPH), proposed by Grande et al. (2023), incorporates anisotropy through directional scaling. However, NIPH inherently focuses on anisotropy along a single, specified direction, potentially overlooking more complex or multi-directional anisotropic structures in real-world data. Kališnik et al. (arXiv 2024) recently introduced the Ellipsoid Complex, which infers local anisotropy by fitting data-driven ellipsoids before computing persistent homology. While their framework pursues a goal similar to ours, its reliance on iterative numerical heuristics brings additional computational overhead and may limit robustness in high-dimensional, noisy settings, leaving efficiency and accuracy as open challenges.

Q. How to find the point of tangency (or intersection) of two ellipses?

Intersection of quadratic curves

Does anybody know by whom and when it was invented?

To be honest, I do not know the first appearance of the calculation method for conics' intersections.
Any relevant literature would be quite helpful! Thanks!

- Mathematicians studied quadratic curves (a.k.a. conic sections) from B.C.
- Apollonius wrote "Conics" around the 2nd century B.C.
- In history, conics have a connection to astronomy. (J. Kepler 16th century)
- Wikipedia - [[Conic sections]] - § Intersecting two conics:
  — one can locate the intersection by identifying the degenerate pencil.
- There is an implementation note and existing software for computing and judging ellipse intersections.
  davideberly/GeometricTools

Other (hopefully) related information already provided:
- "On Heuristic Detection of Ellipsoids Collision" — by engineers
- "Origami Math Book states the pencil method" — by a school teacher
- "Projective Geometry Book is a citation source" — Wikipedia

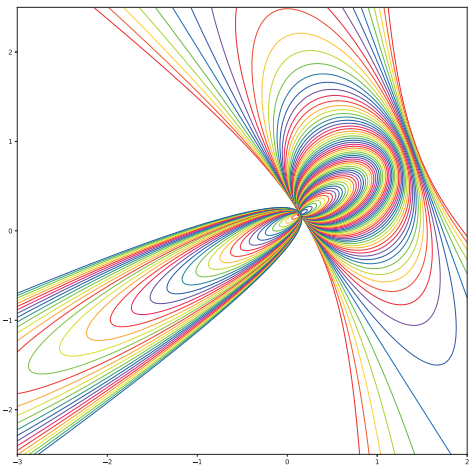
"Well-known" method [citation needed]

Find the degenerate pencil

1. Given two conics Q_0 and Q_1 (expressed by quadratic polynomials)
2. Consider their pencil Q_μ, a linear combination of Q_0 and Q_1
3. Find μ for which Q_μ degenerates to two lines
   (The degeneracy condition is a cubic equation in μ)
4. Obtain the intersection point (x, y)
   by combining the line equation with either Q_0 or Q_1

Key Fact: Q_μ go through the intersections of Q_0, Q_1

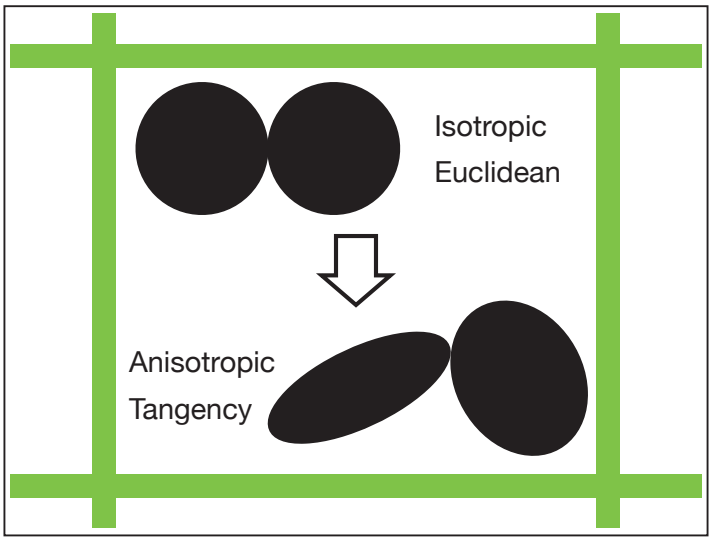
Pencil of conics



Yet another way: Bézoutian

- Fact: Two (quadratic) polynomials share some roots
  $\iff$ the Bézoutian, determinant of the Bézout matrix, is 0.
- davideberly/GeometricTools takes this approach.
  1. Collect the ellipse polynomial terms in x
  2. Compute Bézout's determinant → quartic(= degree 4) in y
  3. Solve the quartic equation in y
  4. Back-substitute y into the ellipse equation → find x
  5. Reject **non-intersecting falsy solutions** (x, y)

Bad from computational viewpoint (wasting many computations!)



Ellipse Cloud

Simultaneous Expansion Model

Given ellipses $Q_i(t,\mathbf{x}) = {}^{\top}\mathbf{x}\, A_i\, \mathbf{x} + 2\mathbf{b}_i \cdot \mathbf{x} + c_i = 0$, decompose it as

$$Q_i(t,\mathbf{x}) = {}^{\top}\left(\mathbf{x}-\bar{\mathbf{x}}_i\right) {}^{\top}P_i \begin{pmatrix} r_i & 0 \\ 0 & r_i' \end{pmatrix}^{-2} P_i \left(\mathbf{x}-\bar{\mathbf{x}}_i\right) - t^2 .$$

- $\bar{\mathbf{x}}_i$ i-th center point
- P_i axes (orthogonal)
- $r_i \geq r_i' > 0$ reference radii
- t time, expansion parameter

given

For each pair $(\mathrm{Q}_i, \mathrm{Q}_j)$, find the tangency time t.

Because the model includes a parameter t, the intersection problem becomes complicated.

Another smart way

Lagrange multiplier method

$$\underset{\mathbf{x}\in\mathbf{R}^2,\ t\geq 0}{\text{minimize}}\ t^2 \quad \text{subject to} \quad \begin{cases} Q_0(t,\mathbf{x}) = 0, \\ Q_1(t,\mathbf{x}) = 0. \end{cases}$$

↓

Find $(\mathbf{x}, t, \mu)$ such that $\nabla F = 0$, where
$F(\mathbf{x}, t, \mu) = t^2 + \mu_0 Q_0(t,\mathbf{x}) + \mu_1 Q_1(t,\mathbf{x})$.

Constrained minimization

Find a stationary point of the Lagrangian

By $\nabla_{\mathbf{x}} F = 0$, we have $\mathbf{x} = \bar{\mathbf{x}}(\mu) = -A_\mu^{-1}\mathbf{b}_\mu$, where

$$A_\mu = \mu_0 A_0 + \mu_1 A_1, \quad \mathbf{b}_\mu = \mu_0 \mathbf{b}_0 + \mu_1 \mathbf{b}_1.$$

(Impose normal vectors to be linearly dependent.)

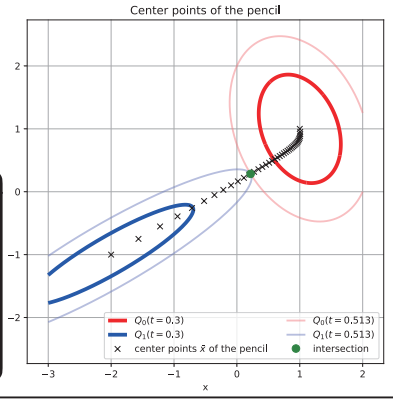
What is $\bar{\mathbf{x}}(\mu)$?

$\bar{\mathbf{x}}(\mu)$ = the center point of the pencil

Completing the square of $Q_\mu = \mu_0 Q_0 + \mu_1 Q_1$ yields the center point of Q_μ, which coincides with $\bar{\mathbf{x}}(\mu)$.

Key points:

- $\bar{\mathbf{x}}(\mu)$ does NOT depend on t. (The center is always located even when Q_μ is an imaginary ellipse.)
- Find a stationary point only on the center curve $\bar{\mathbf{x}}(\mu)$.



Note: Lagrange multiplier method

- davideberly/GeometricTools also uses the Lagrange multiplier to determine *whether* two given ellipses intersect or not.
  - It implements a Boolean query of the relative position of them:
    A. If two ellipses intersect.
    B. If one includes another strictly.
    C. Neither; if two ellipses are separated.

- It is straightforward to solve the tangency problem in the same way, but it has not been implemented in this software yet.
- Thus, our proposed method and software are still new.
- Especially, we can prove **the unique existence of the solution**.

$$\mu \mapsto Q_\mu(t, \bar{\mathbf{x}}(\mu))$$

Takes the maximum value

at $\mu = \bar{\mu}$.

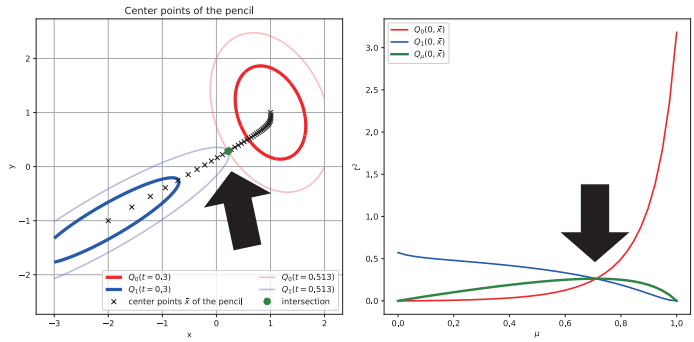
Futhermore, at time

$$\bar{t} = \sqrt{Q_{\bar{\mu}}(0, \bar{\mathbf{x}}(\bar{\mu}))},$$

$\bar{\mathbf{x}}(\bar{\mu})$ is the point of tangency.

Sketch of proof and the plot of Q_μ

Show convexity. Calculate the 2nd derivative of $Q_\mu(t, \bar{\mathbf{x}}(\mu))$ w.r.t. μ. Proved.



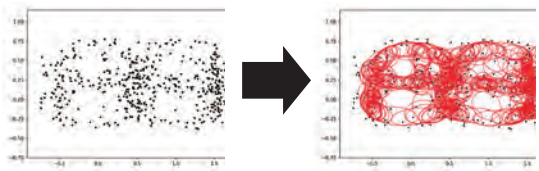
Numerical performance

TABLE 1. Performance evaluation results ($N = 1000$ ellipses). The first column `method` indicates the solver specified by the `scipy.optimize.root_scalar` function in Python: (1) bisection, (2) Brent, and (3) Newton. In (4), we adopt a hybrid approach that runs Brent for eight iterations to obtain an initial estimate and then applies Newton for three iterations. The second column shows the execution time per call, while the subsequent columns show relative errors. The relative error for Q_i is defined as $|Q_i - Q_\mu|/Q_\mu$ $(i = 0, 1)$, and the listed values are averages. The adjacent column shows the median of the corresponding relative error.

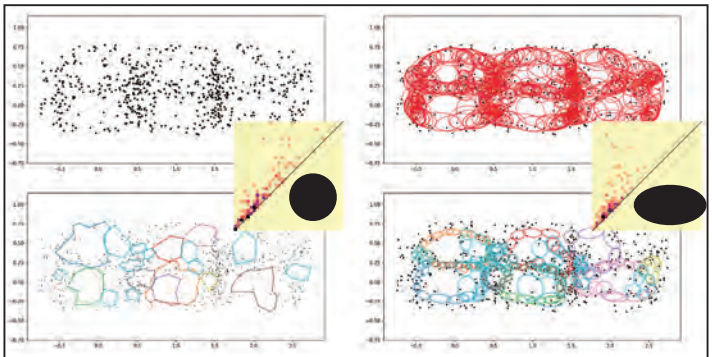
method	Execution Time	Rel. Error Q_0	(Median)	Rel. Error Q_1	(Median)
(1) `bisect`	0.320 ms	$3.80 \cdot 10^{-12}$	$3.21 \cdot 10^{-12}$	$3.83 \cdot 10^{-12}$	$3.23 \cdot 10^{-12}$
(2) `brentq`	0.089 ms	$1.84 \cdot 10^{-13}$	$2.65 \cdot 10^{-15}$	$1.85 \cdot 10^{-13}$	$2.68 \cdot 10^{-15}$
(3) `newton`	0.133 ms	$2.58 \cdot 10^{-02}$	$4.85 \cdot 10^{-16}$	$3.11 \cdot 10^{-02}$	$4.78 \cdot 10^{-16}$
(4) (2)→(3)	0.118 ms	$9.59 \cdot 10^{-15}$	$4.26 \cdot 10^{-16}$	$1.14 \cdot 10^{-14}$	$4.20 \cdot 10^{-16}$

Solve the algebraic equation $Q_0(0, \bar{\mathbf{x}}(\mu)) = Q_1(0, \bar{\mathbf{x}}(\mu))$ to obtain $\bar{\mu}$

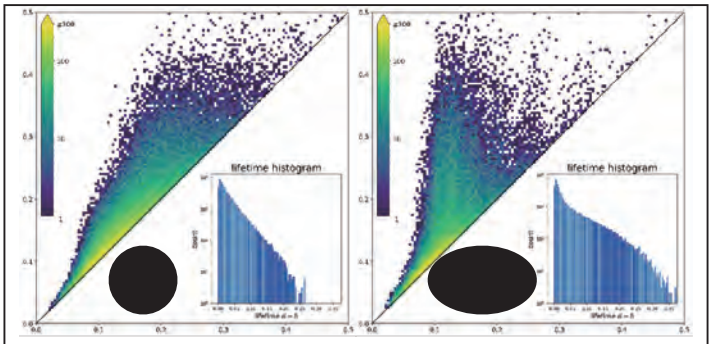
k-NN covariance construction



1. For each point x_i, take the k-nearest neighbour
2. Compute the mean $\bar{x}_i$ and the covariance matrix, and eigendecompose it
3. Place ellipse $Q_i(t, x) = {}^\top(x - \bar{x}_i)^\top P_i \operatorname{diag}(r_1, r_2)\, P_i(x - \bar{x}_i) - t^2 = 0$
4. Compute a quasi-metric d_{ij} by ellipse tangency
5. Build VR filtration from (d_{ij}) and compute PH (PD)



Ellipse PH captures anisotropic ring structures in greater extent. In this toy experiment, we found many optimal 1-cycles whose corresponding persistence are much longer.



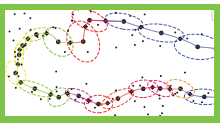
Cumulative histograms of 1000 PDs in the same experiment configuration.

Compared to the standard PH, the anisotropic PH exhibits:

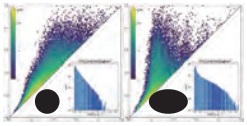
- Birth time is earlier and death time is later.
- Fewer birth-death pairs have shorter lifetimes.
- More birth-death pairs have longer lifetimes.

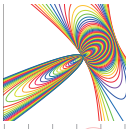
ellphi

Summary

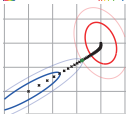


- Seeking detailed references on conics' intersections 📚🙏; Suggestions welcome!
- Proposed method
  - Ellipse Cloud construction
  - Fast and accurate computation of ellipse tangency
  - Software implementation





ellphi



- Looking for applications to real-world data
- Suggestions of anisotropic problems are welcome too!

Q & A

- Inverse analysis? → 🆗 possible as in the standard VR filtration
- PD optimization? → 🚧 possible but not implemented yet [WIP]
  - ✅ The gradient of the tangency problem is computable.
  - 🚧 Future work: ellipses' axes optimization? What about DoF?
- Shapes other than ellipses? → 🚧 possible in principle but hard
  - ✅ The Lagrange multiplier method is applicable in a similar way.
  - ❓ No closed-form solutions nor assurance, in general, though.
- Higher dimension (ellipsoids)? → 🆗 extensible in the same way
- Other filtrations like alpha complex?
  - ✅ Ellipsoid Complex proposed by Kališnik et al. (arXiv 2024)
  - ❓ Computational cost issues? Intersection of ≥ 2 ellipses?

Nonnegative Matrix Factorization with Topological Regularization

Keunsu Kim

Institute of Mathematics for Industry, Kyushu University, Japan

In this study, we propose Top-NMF, a novel model that incorporates topological regularization into Nonnegative Matrix Factorization (NMF), a widely used dimensionality reduction technique. While conventional regularization methods focus on preserving relationships between data points to guide low-dimensional representations, Top-NMF explicitly controls the topological structure of the support of each basis vector. We interpret each data point as a real-valued function defined over a structured domain (such as a grid or a graph), and treat each basis vector in the same way. Our focus is on the support of each basis vector, and we introduce quantitative topological descriptors derived from persistent homology as regularization terms. These descriptors encourage the support to exhibit desirable properties such as connectedness and modularity. These regularization terms can be applied across diverse domains including time series, images, and graphs and guide the model to learn basis vectors that reflect meaningful structures. We provide a theoretical formulation, describe the optimization scheme, and demonstrate through experiments that Top-NMF achieves structurally faithful and interpretable decompositions.

References.

[1] Mathieu Carriere et al. "Optimizing persistent homology based functions". In: *International conference on machine learning*. PMLR. 2021, pp. 1294–1303.

[2] Patrik O Hoyer. "Non-negative matrix factorization with sparseness constraints". In: *Journal of machine learning research* 5.Nov (2004), pp. 1457–1469.

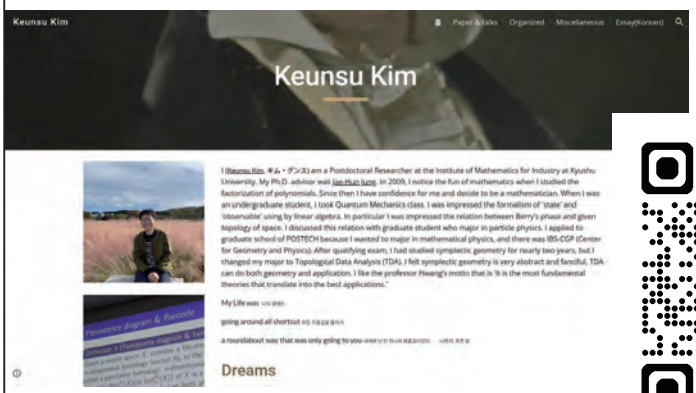
[3] Daniel D Lee and H Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization". In: *nature* 401.6755 (1999), pp. 788–791.

Nonnegative matrix factorization with topological regularization

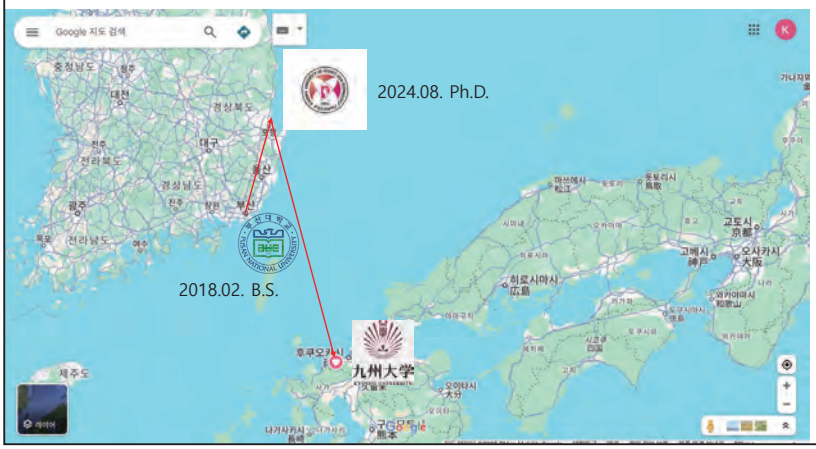
Keunsu Kim (キム・グンス)
Assistant professor (Non-tenure track)
Institute of Mathematics for Industry
at Kyushu University
2025.08.06
(with Matias de Jong van Lier and Shizuo Kaji)

Workshop for Topological Data Analysis and Industrial Mathematics

https://sites.google.com/view/keunsukim/



Background



Linear Dimensionality Reduction

Linear Dimensionality Reduction

Given n d-dimensional data points $\{\mathbb{x}_1, \cdots, \mathbb{x}_n\}$ and $\mathbb{x}_i \in \mathbb{R}^d$.

Want to reduce the dimensionality to $r < d$.

i.e., find basis vectors $\mathbb{v}_j \in \mathbb{R}^d$ and corresponding coefficients w_{ij}.

$W = [w_{ij}] \in \mathbb{R}^{n \times r}$
Coefficient matrix

$V = \begin{bmatrix} \mathbb{v}_1 \\ \vdots \\ \mathbb{v}_r \end{bmatrix} \in \mathbb{R}^{r \times d}$
Basis matrix

$$\mathbb{x}_i \approx w_{i1}\mathbb{v}_1 + \cdots + w_{ir}\mathbb{v}_r \quad \Leftrightarrow \quad X \approx WV$$

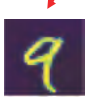
Criterion is to minimize the Loss function L.

Ex) $L_{apx}(\mathbb{w}, \mathbb{v}) = \sum_{i=1}^{n} \left\| \mathbb{x}_i - \sum_{j=1}^{r} w_{ij}\mathbb{v}_j \right\|^2$

$\mathbb{w} := \{w_{ij}\}_{\substack{i=1,\cdots,n \\ j=1,\cdots,r}}$

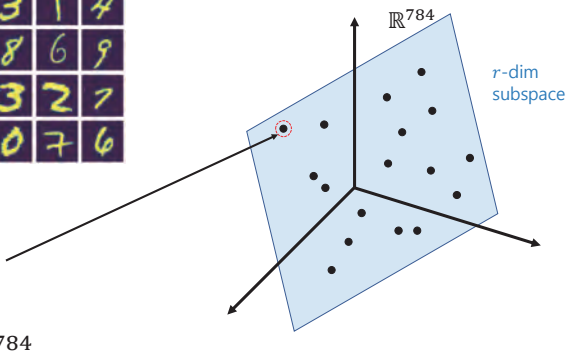
$\mathbb{v} := \{\mathbb{v}_j\}_{j=1,\cdots,r}$

More clarify data points $\{\mathbb{x}_i\}_{i=1}^{40}$ and each $\mathbb{x}_i \in \mathbb{R}^{28\times 28}$

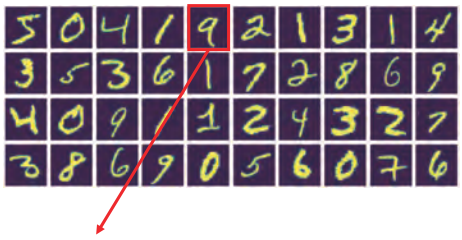


unfold

$\mathbb{x}_5 \in \mathbb{R}^{28\times 28}$ $\mathbb{x}_5 \in \mathbb{R}^{784}$

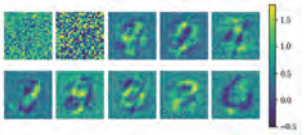


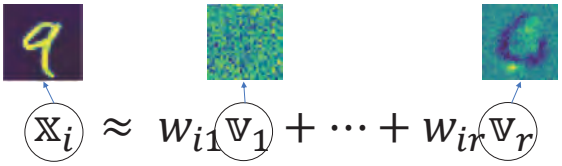
Principal Component Analysis(PCA)



minimize $L_{apx}(\mathbb{w}, \mathbb{v}) = \sum_{i=1}^{n} \left\| \mathbb{x}_i - \sum_{j=1}^{r} w_{ij} \mathbb{v}_j \right\|^2$

Ex: PCA basis, $r = 10$.





$\mathbb{x}_i \approx w_{i1}\mathbb{v}_1 + \cdots + w_{ir}\mathbb{v}_r$

$i = 5$

(Daniel D. Lee, and H. Sebastian Seung) Learning the parts of objects by non-negative matrix factorization, *nature* (1999):

Nonnegative Matrix Factorization(NMF)

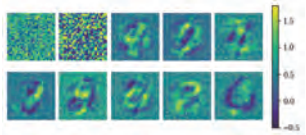
Given dataset $\{\mathbb{x}_1, \cdots, \mathbb{x}_n\}$ and $\mathbb{x}_i \in \mathbb{R}^d_{\geq 0}$,

minimize

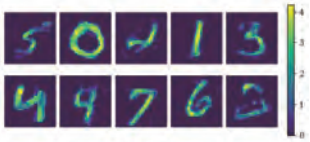
$$L_{apx}(\mathbb{w}, \mathbb{v}) = \sum_{i=1}^{n} \left\| \mathbb{x}_i - \sum_{j=1}^{r} w_{ij} \mathbb{v}_j \right\|^2$$

subject to

$w_{ij}, \mathbb{v}_j \geq 0$, nonnegative constraint



PCA basis



NMF basis

NMF: Nonnegative Matrix Factorization

In this study, given a $\{\mathbb{x}_1, \cdots, \mathbb{x}_n\}$ and $\mathbb{x}_i \in \mathbb{R}^d_{\geq 0}$,

Goal: Obtain **interpretable basis vectors** $\mathbb{v}_j \in \mathbb{R}^d_{\geq 0}$.

Method:

1. Quantify **desirable** properties of $\mathbb{v}_j$ (e.g., sparseness, **connectedness**) into a scalar $L_{reg}(\mathbb{v})$

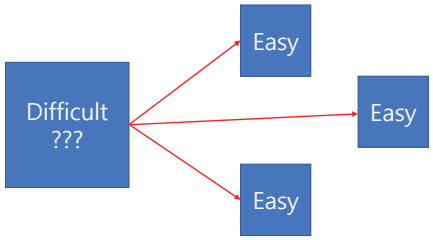
2. Encourage the desirable properties by using them as a regularization term.

   $L(\mathbb{w}, \mathbb{v}) = L_{apx}(\mathbb{w}, \mathbb{v}) + \lambda_{reg} \cdot L_{reg}(\mathbb{v})$

3. Optimize w_{ij} and $\mathbb{v}_j$ jointly under nonnegativity constraints.

Philosophy

To divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.
- René Descartes



Quantify desirable properties of $\mathbb{v}_j$

Quick Review of **sparsity** of $\mathbb{v}_j$

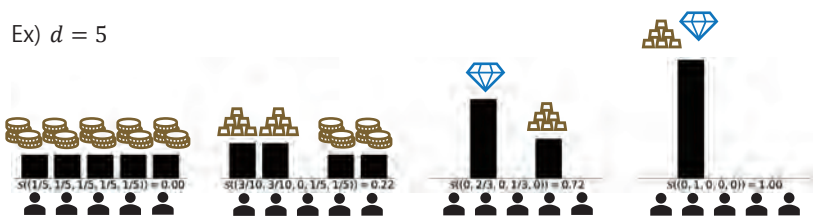
(Patrik O. Hoyer) Non-negative matrix factorization with sparseness constraints, Journal of machine learning research (2004)

Sparsity score of $\mathbb{v}_j \in \mathbb{R}^d$

Cf. $\|\mathbb{v}_j\|_2 \le \|\mathbb{v}_j\|_1 \le \sqrt{d}\|\mathbb{v}_j\|_2$

$$\mathcal{S}(\mathbb{v}_j) = \frac{1}{\sqrt{d}-1} \cdot \left(\sqrt{d} - \frac{\|\mathbb{v}_j\|_1}{\|\mathbb{v}_j\|_2} \right) \in [0,1]$$

Ex) $d = 5$



Nonnegative Matrix Factorization with sparsity regularization (S-NMF)

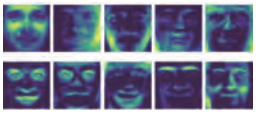
$$L_{apx}(\mathbb{w},\mathbb{v}) = \sum_{i=1}^{n}\left\|\mathbb{x}_i - \sum_{j=1}^{r} w_{ij}\mathbb{v}_j\right\|^2$$

+

$$L_{spa}\left(v;\{a_j\}_{j=1,\cdots,r}\right) = \sum_{j=1}^{r}\|S(v_j) - a_j\|^2$$

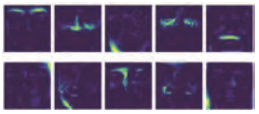
subject to
$w_{ij}, \mathbb{v}_j \geq 0$, nonnegative constraint
$*a_j$ is desirable sparsity score

Fetch olivetti faces dataset



NMF basis

Each basis vector is localized.



s-NMF basis $a_j = 0.7$

S-NMF: Nonnegative Matrix Factorization with sparsity regularization

There is no guarantee of acquiring a localized basis.

Top-NMF

In this study, given a $\{\mathbb{x}_1, \cdots, \mathbb{x}_n\}$ and $\mathbb{x}_i \in \mathbb{R}^d_{\geq 0}$,

Recall

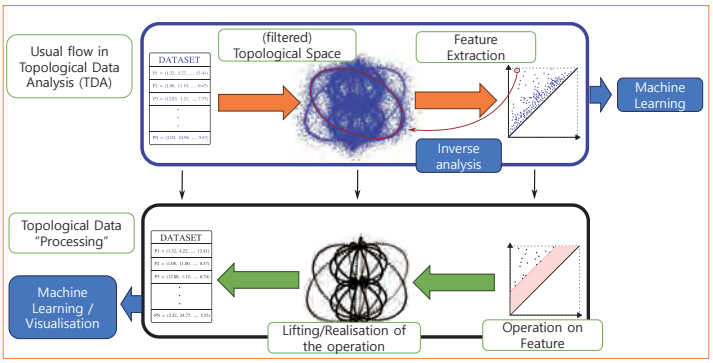
Goal: Obtain **interpretable basis vectors** $\mathbb{v}_j \in \mathbb{R}^d$.

Method:

1. **Quantify desirable topological** properties of $\mathbb{v}_j$ (e.g., sparseness, **connectedness**) into a scalar $L_{top}(\mathbb{v})$
2. Encourage desirable properties by using $L_{top}(\mathbb{v})$ as a regularization term.

   $L(\mathbb{w},\mathbb{v}) = L_{apx}(\mathbb{w},\mathbb{v}) + \lambda_{top} \cdot L_{top}(\mathbb{v})$ subject to $w_{ij}, \mathbb{v}_j \geq 0$.
3. Optimize w_{ij} and $\mathbb{v}_j$ jointly under nonnegativity constraints.

From Topological Data "Analysis" to Data "Processing"



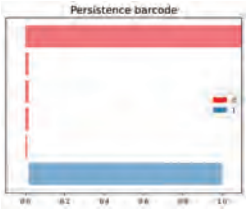
Made by Matias de Jong van Lier

0-dim homology: # components
1-dim homology: # holes

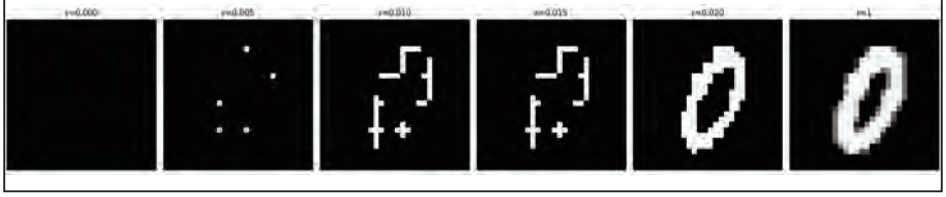
Filtration and barcode

x:  $\in [0,1]^{28\times28}$ $PH(x)$:

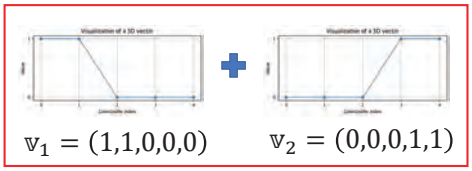
Persistent Homology



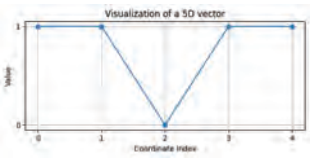
Superlevel set filtration on a cubical complex: pixel intensity $\geq 1-\varepsilon$



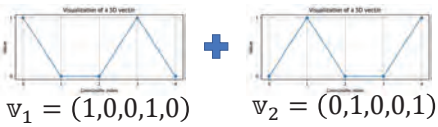
Want



$\mathbb{v}_1 = (1,1,0,0,0)$ $\mathbb{v}_2 = (0,0,0,1,1)$



$\mathbb{x}_1 = (1,1,0,1,1)$



$\mathbb{v}_1 = (1,0,0,1,0)$ $\mathbb{v}_2 = (0,1,0,0,1)$

Each basis time series has a sparsity score ≈ 0.66.

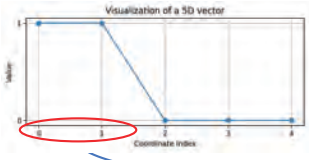
Sparsity score cannot distinguish between the two decompositions

Superlevel set filtration on a cubical complex: pixel intensity $\geq 1-\varepsilon$

Observe the variation of

$\mathbb{v}_1 = (1,1,0,0,0)$

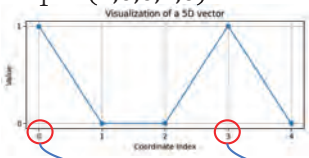
$supp_{\geq 1-\varepsilon}(\mathbb{v}_1) = \{i \in \{0,1,2,3,4\} \mid \mathbb{v}_1[i] \geq 1-\varepsilon\}$



$0 < \varepsilon < 1$

$\varepsilon < 0$

$\mathbb{v}_1 = (1,0,0,1,0)$

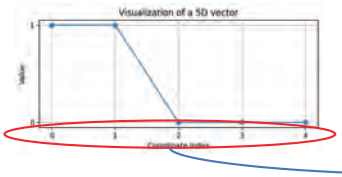


Superlevel set filtration on a cubical complex: pixel intensity $\geq 1-\varepsilon$

Observe the variation of sublevel set

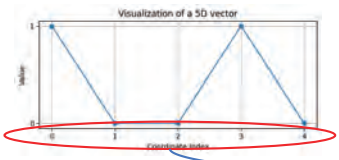
$v : \{0,1,2,3,4\} \to \mathbb{R}_{\geq 0}$

$supp_{\geq 1-\varepsilon}(\mathbb{v}_1) = \{i \in \{0,1,2,3,4\} \mid \mathbb{v}_1[i] \geq 1-\varepsilon\}$



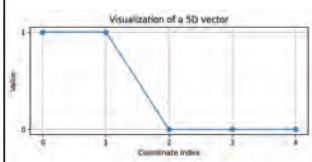
$0 < \varepsilon < 1$

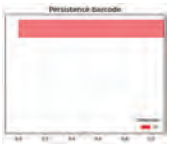
$\varepsilon < 0$



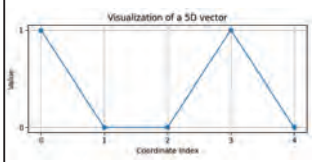
superlevel set filtration on a cubical complex

$\mathrm{PE}^{(k)}(\mathbb{v}_j) := \sum_{(b,d)\in \mathrm{PH}_k^{\mathrm{fin}}(v_j)} (d-b)^2$





$\mathrm{PE}^{(0)}(v) = 0$



$\mathrm{PE}^{(0)}(v) = 1$

We can distinguish between the two decompositions using the persistent homology.

Top-NMF on image dataset

$\mathbb{v}_j$

Superlevel cubical filtration

$\mathrm{PH}_k^{\mathrm{fin}}(\mathbb{v}_j)$

$$\mathrm{PE}^{(k)}(\mathbb{v}_j) := \sum_{(b,d)\in \mathrm{PH}_k^{\mathrm{fin}}(v_j)} (d-b)^2 \quad \text{for } k = 0,1.$$

$$L_{top}(\mathbb{v}) = \sum_{j=1}^{r} \mathrm{PE}^{(0)}(\mathbb{v}_j)$$

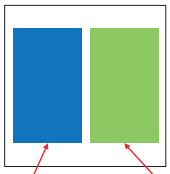
Find a solution $w_{ij}, v_j \geq 0$, which minimizes

$$\mathrm{L}(\mathbb{w}, \mathbb{v}) = L_{apx}(\mathbb{w}, \mathbb{v}) + \lambda_{top} \cdot L_{top}(\mathbb{v})$$

i.e. Top-NMF yields basis vectors with connected supports.

Korean character example

$\mathbb{x}_1$ $\mathbb{x}_2$ $\mathbb{x}_3$ $\mathbb{x}_4$ $\mathbb{x}_5$ $\mathbb{x}_6$



ㅏ, ㅑ

ㄱ,ㄷ,ㄹ

母音(vowel): あいうえお

子音(consonant): except for あいうえお

We desire the basis as follows:

$\mathbb{v}_1$ $\mathbb{v}_2$ $\mathbb{v}_3$ $\mathbb{v}_4$ $\mathbb{v}_5$

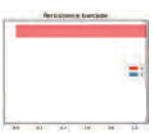
$$\mathbb{x}_i, \mathbb{v}_j \in \mathbb{R}^{64\times 64} = \mathbb{R}^{4096}$$

$\mathbb{x}_1 = \mathbb{v}_2 + \mathbb{v}_1$

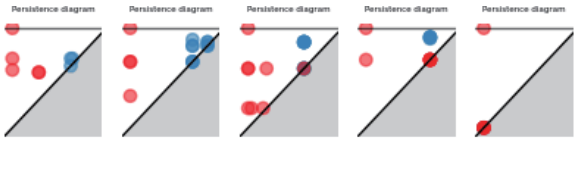
Observation

NMF basis

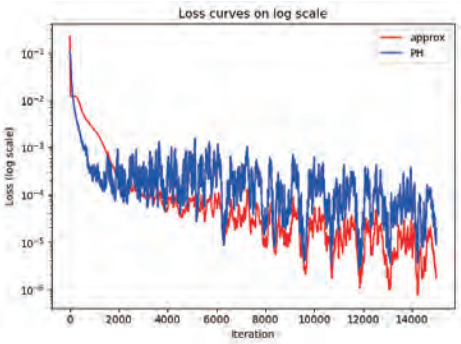
Top-NMF basis



Top-NMF learning process

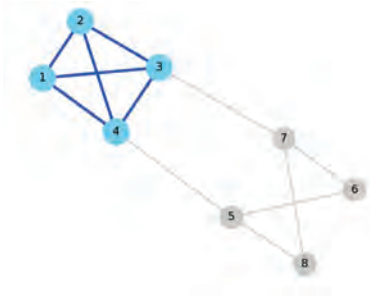


Top-NMF learning process

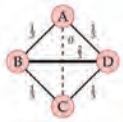


Clique Deviation Metric in a Graph

Given an undirected graph,
clique is a fully connected subgraph.



$G = (\mathcal{V}, \mathcal{E})$: fully connected weighted graph, with edge weights in the range [0,1] (not vertex weights).



Graph data 1

AB AC AD BC BD CD

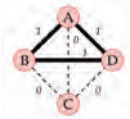
$\mathbb{x}_1 = \left(\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}\right)$

Graph data 2

AB AC AD BC BD CD

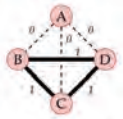
$\mathbb{x}_2 = \left(\frac{1}{3}, 0, \frac{1}{3}, \frac{2}{3}, 1, \frac{2}{3}\right)$

Want decompose the structure via clique graphs.



Basis graph 1

$\mathbb{v}_1 = (1,0,1,0,1,0)$



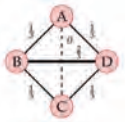
Basis graph 2

$\mathbb{v}_2 = (0,0,0,1,1,1)$

Ex)

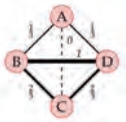
$\left(\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}\right) = \frac{1}{3} \cdot (1,0,1,0,1,0) + \frac{1}{3} \cdot (0,0,0,1,1,1)$

$$\mathbb{x}_1 = \frac{1}{3} \cdot \mathbb{v}_1 + \frac{1}{3} \cdot \mathbb{v}_2$$



Graph data 1

$\mathbb{x}_1 = \left(\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}\right)$

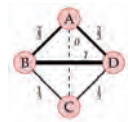


Graph data 2

$\mathbb{x}_2 = \left(\frac{1}{3}, 0, \frac{1}{3}, \frac{2}{3}, 1, \frac{2}{3}\right)$

NMF basis

Basis graph 1



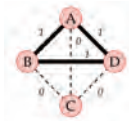
Basis graph 2

$\mathbb{v}_1 = \left(\frac{4}{13}, 0, \frac{4}{13}, \frac{9}{13}, 1, \frac{9}{13}\right)$ $\mathbb{v}_2 = \left(\frac{2}{3}, 0, \frac{2}{3}, \frac{1}{3}, 1, \frac{1}{3}\right)$

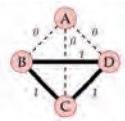
$$\mathbb{x}_1 = \frac{13}{42} \cdot \mathbb{v}_1 + \frac{5}{14} \cdot \mathbb{v}_2$$

$$\mathbb{x}_2 = \frac{13}{14} \cdot \mathbb{v}_1 + \frac{1}{14} \cdot \mathbb{v}_2$$

Desirable basis



Basis graph 1



Basis graph 2

$\mathbb{v}_1 = (1,0,1,0,1,0)$ $\mathbb{v}_2 = (0,0,0,1,1,1)$

$$\mathbb{x}_1 = \frac{1}{3} \cdot \mathbb{v}_1 + \frac{1}{3} \cdot \mathbb{v}_2$$

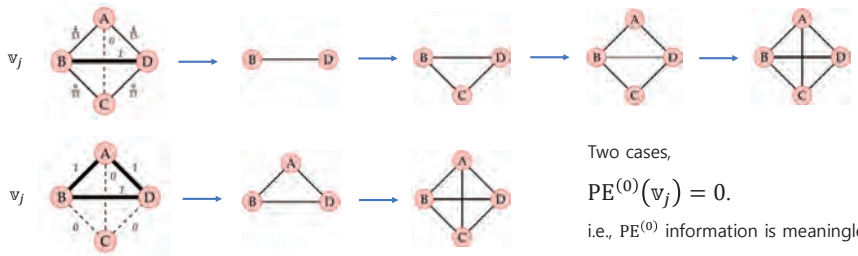
$$\mathbb{x}_2 = \frac{1}{3} \cdot \mathbb{v}_1 + \frac{2}{3} \cdot \mathbb{v}_2$$

To distinguish two decomposition,
we introduce the clique deviation metric

$$\mathrm{PE}^{(k)}(\mathbb{v}_j) := \sum_{(b,d)\in \mathrm{PH}_k^{\mathrm{fin}}(v_j)} (d-b)^2$$

Key point: We assign a fixed weight of 1 to each vertex.

To justify the key point, assume each vertex is assigned the maximum edge weight among its incident edges.



Two cases,

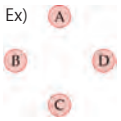
$\mathrm{PE}^{(0)}(\mathbb{v}_j) = 0.$

i.e., $\mathrm{PE}^{(0)}$ information is meaningless.

We assign a fixed weight of 1 to each vertex.

$$CDM(\mathbb{v}_j) := -\mathrm{PE}^{(0)}(\mathbb{v}_j) - \alpha \cdot \mathrm{PE}^{(1)}(\mathbb{v}_j) \quad \text{for} \quad \alpha > 0.$$

Ex)

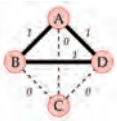


$\mathrm{PE}^{(0)} = 4$

Interpretation of Low CDM Value

A small value of $CDM(\mathbb{v}_j)$ implies high values of $\mathrm{PE}^{(0)}(\mathbb{v}_j)$ and $\mathrm{PE}^{(1)}(\mathbb{v}_j)$,
By using this as the L_{top} , we encourage each basis vector to:

- $\mathrm{PE}^{(0)}$: connect only the necessary vertices
- $\mathrm{PE}^{(1)}$: create as many holes as possible among the connected vertices.



Theorem:
v is a local minimizer of CDM if and only if it takes the value 1 on a disjoint union of cliques.
*Clique: #node ≥ 3

Topological quantifiers and L_{top}

$\mathrm{PE}^{(k)}(v) := \sum_{(b,d)\in \mathrm{PH}_k^{\mathrm{fin}}(v)} (d-b)^2$	Persistence Energy
$CDM(v) := -\mathrm{PE}^{(0)}(v) - \mathrm{PE}^{(1)}(v)$	Clique Deviation Metric
$PerScore(v) := \frac{\max_{(b,d)\in \mathrm{PH}_1^{\mathrm{fin}}(v)} (d-b)}{\sqrt{3}}$	Periodic Score
$\mathcal{W}\mathrm{PE}^{(k)}(v) := \sum_{(b,d)\in \mathrm{PH}_k^{\mathrm{fin}}(v)} (1-d)\cdot(d-b)^2$	Weighted Persistence Energy

Top-NMF

Find a solution $w_{ij}, \mathbb{v}_j \geq 0$, which minimizes

$$\mathrm{L}(\mathbb{w}, \mathbb{v}) = L_{apx}(\mathbb{w}, \mathbb{v}) + \lambda_{top} \cdot L_{top}(\mathbb{v})$$

Theorem:

Gradient descent algorithm converges.

(Damek Davis, et al.) Stochastic subgradient method converges on tame functions, *Foundations of computational mathematics* (2020).

(Mathieu Carrière, et al.) Optimizing persistent homology based functions, *International conference on machine learning*. PMLR (2021).

We solve this problem using gradient descent.

$$\mathbb{w} \leftarrow \mathbb{w} - \frac{\partial L}{\partial \mathbb{w}}$$

$$\mathbb{v} \leftarrow \mathbb{v} - \frac{\partial L}{\partial \mathbb{v}}$$

In this study, given a $\{\mathbb{x}_1, \cdots, \mathbb{x}_n\}$ and $\mathbb{x}_i \in \mathbb{R}^d_{\geq 0}$,

Recall

Goal: Obtain **interpretable basis vectors** $\mathbb{v}_j \in \mathbb{R}^d$.

Method:

1. **Quantify desirable topological properties** of $\mathbb{v}_j$ (e.g., sparseness, connectedness) into a scalar $L_{top}(\mathbb{v})$
2. Encourage desirable properties by using the quantifiers as a regularization term.

   $L(\mathbb{w}, \mathbb{v}) = L_{apx}(\mathbb{w}, \mathbb{v}) + \lambda_{top} \cdot L_{top}(\mathbb{v})$ subject to $w_{ij}, \mathbb{v}_j \geq 0$.

3. Optimize w_{ij} and $\mathbb{v}_j$ jointly under nonnegativity constraints.

Applications of Persistent Homology to Materials Science, and Persistent Homology Software HomCloud

Ippei Obayashi

Center for Artificial Intelligence & Mathematical Data Science, Okayama University, Japan

In this presentation, I will discuss persistent homology, a mathematical tool that characterizes the shape of data using topology. Mathematical foundations and applications to materials science will be presented. Our persistent homology-based software, HomCloud, will also be introduced.

References.

[1] Yasuaki Hiraoka et al. "Refinement of interval approximations for fully commutative quivers". In: *arXiv preprint arXiv:2310.03649* (2023).

[2] Emi Minamitani et al. "Topological descriptor of thermal conductivity in amorphous Si". In: *The Journal of Chemical Physics* 156.24 (2022).

[3] Ippei Obayashi, Yasuaki Hiraoka, and Masao Kimura. "Persistence diagrams with linear machine learning models". In: *Journal of Applied and Computational Topology* 1.3 (2018), pp. 421–449.

[4] Ippei Obayashi, Takenobu Nakamura, and Yasuaki Hiraoka. "Persistent homology analysis for materials research and persistent homology software: HomCloud". In: *journal of the physical society of japan* 91.9 (2022), p. 091013.

Applications of Persistent homology to materials science, and persistent homology software HomCloud

Ippei Obayashi

Okayama University

Aug. 7, 2025

岡山大学
OKAYAMA UNIVERSITY

Introduction to persistent homology

Persistent homology

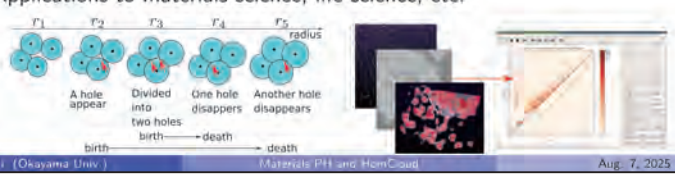
- *Topological Data Analysis* (TDA)
  - Data analysis using the mathematical idea of topology
  - Quantitatively characterize the shape of data
    - ★ Connected components, rings, cavities,

Persistent homology (PH): Mail tool of TDA

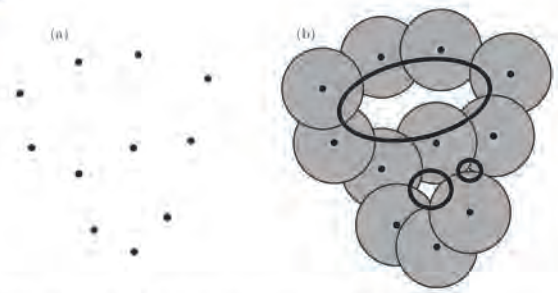
- Using the idea of homology
- PH gives good descriptors for the shape of data (called persistence diagram)

Developed in the 21st century

- Mathematical theory and algorithm
- Software
- Applications to materials science, life science, etc.



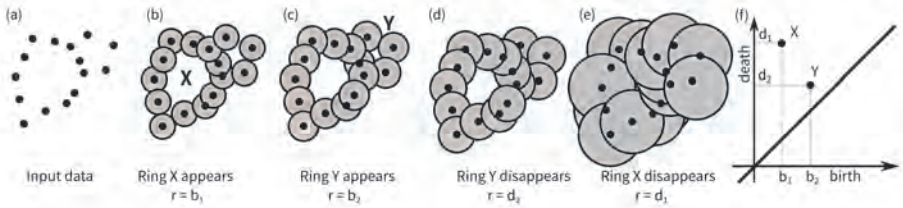
r-Ball model



- The input data is a set of points (a pointcloud)
  The pointcloud has no holes/rings, but it looks like some holes
  Put discs with radii r
  Three holes
  We can characterize the pointcloud by these holes

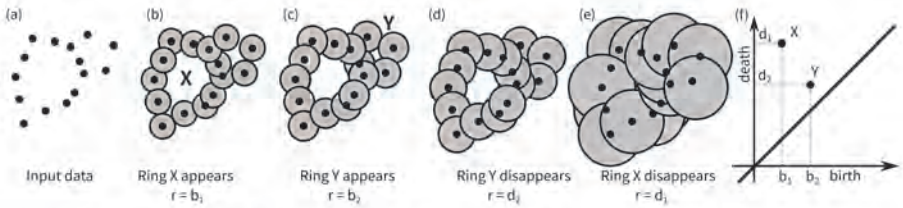
Filtration

The radius r becomes larger, and holes appear and disappear. The theory of persistent homology makes pairs of the appearance and disappearance of holes



Persistence diagram

These pairs are called *birth-death pairs*, and the collection of birth-death pairs is called a *persistence diagram* (PD). The scatter plot (or 2D histogram plot) of the pairs is used to visualize PDs.



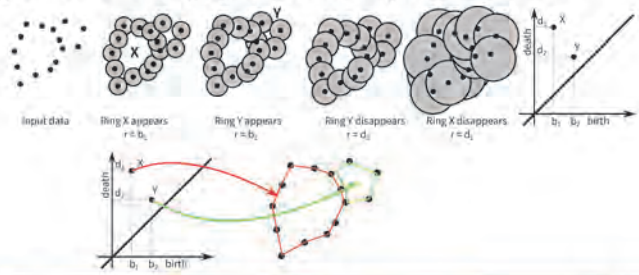
The persistence diagram gives a geometric summary of the data.

Inverse analysis

- In a PD, each birth-death pair corresponds to a ring or cavity

  It is very useful to identify such a ring/cavity

  Since there are many candidates, the best ring or cavity is selected using mathematical optimization.
  - Tightness of a homology generator is usually used as a criterion



Applications to materials research

Joint work with E. Minamitani, T. Shiga, and M. Kashiwagi [1]

Outline of the research

- Aim: Understanding the relationship between nanostructures of amorphous silicon and thermal conductivity of the material

  Method: The thermal conductivity and persistence diagrams are calculated from a lot of atomic configurations of amorphous silicon by molecular dynamical simulations, and two statistical methods (ridge regression and principal component analysis) and an inverse analysis is applied to the diagrams to identify local structures that correlate with the thermal conductivity

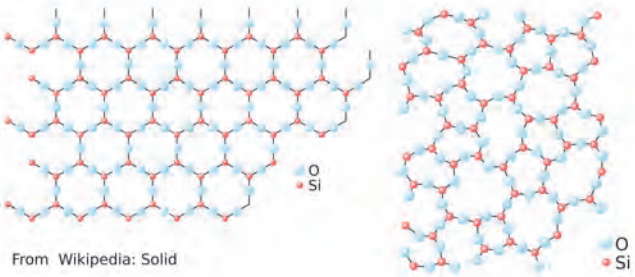
  Results: Thermal conductivity changes when pentagons in the atomic configurations are deformed. We also successfully explained why the deformation changes the thermal conductivity from the viewpoint of physics, using vibrational modes on local structures

Background

Two types of solid materials: crystal and amorphous

- Crystal: periodic atomic structures

  Amorphous: No periodic structures



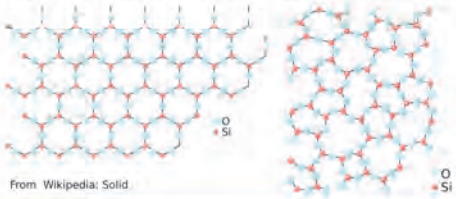
From Wikipedia: Solid

Short-range order, medium-range order, and long-range order

- Short-range order (SRO): Order on neighbor atoms
  - Numbers of adjacent atoms, bond lengths, and bond angles
  - Common for crystals, liquids, and amorphous

  Long-range order (LRO): Order on periodic (crystalline) structures
  - Only crystals

  Medium-range order (MRO): Order larger than short-range order
  - Common for crystals and amorphous, and liquids do not have this type of order
  - Clarifying medium-range order is an important problem on amorphous study



From Wikipedia: Solid

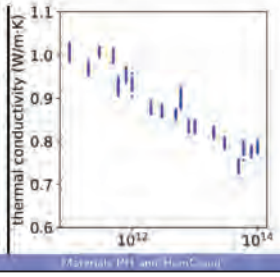
Thermal conductivity

Heat flows from a hot environment to a cold one. Heat flux, a flow of energy per unit area and per unit time, is described by Fourier's low:

$$-k\nabla T \tag{1}$$

where ∇T is the temperature gradient, and k is a constant determined by the material. k is called *thermal conductivity*.

- Usually, crystals have higher thermal conductivity than amorphous materials
  Amorphous materials have different thermal conductivity depending on their structures
  - Amorphous structures are calculated in molecular dynamical simulations by rapid cooling of hot materials
  - The cooling rate and thermal conductivity are correlated (the figure)
  - The cooling rate and medium-range order are also considered to be correlated
  - Therefore, we have a *hypothesis* that medium-range order determines thermal conductivity



Problem

How do we characterize medium-range orders quantitatively?

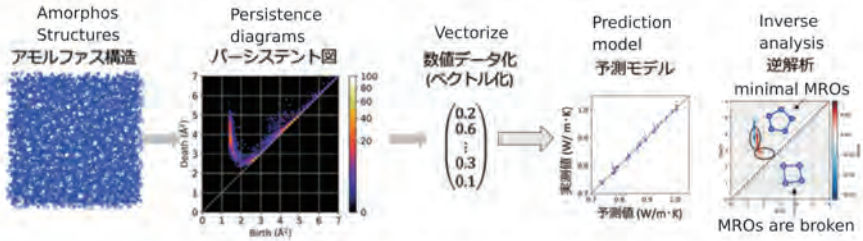
- Descriptors suitable for our purpose

  Recently persistent homology has successfully described amorphous structures like silica glass and metallic glass

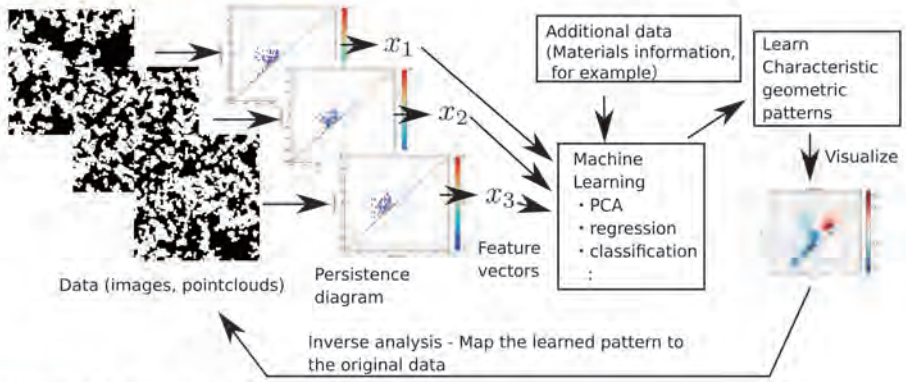
We tried to apply PH for our purpose.

Method

The thermal conductivity and persistence diagrams are calculated from a lot of atomic configurations of amorphous silicon by molecular dynamical simulations, and two statistical methods and the inverse analysis is applied to the 1st diagrams to identify local structures that correlate with the thermal conductivity.



Machine learning framework



Obayashi et al. (2018)[2]

Machine learning concepts

- PH can summarize the *geometric* structure of the data quantitatively

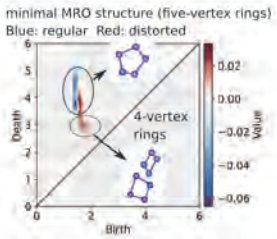
  Machine learning can find *characteristic patterns* from the data

  The combination of these two concepts, we can extract *characteristic geometric patterns* from the data

  By combining *linear machine learning model*, *histogram-based vectorization* of persistence diagrams, and *inverse analysis*, we realize *explainable* PH machine learning!

Results

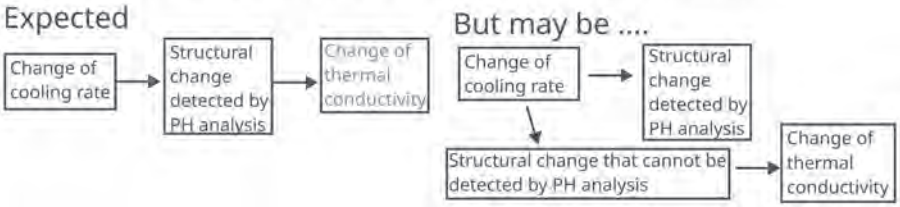
Thermal conductivity changes when pentagons in the atomic configurations are deformed.



- The blue area positively contributes to thermal conductivity, and the red area negatively contributes to thermal conductivity

  The deformation of five-vertex rings are correlated with thermal conductivity

In the PH analysis, we only found the correlation between thermal conductivity and 5-vertex / 4-vertex rings, and both of the following two graphs are possible:



We need to identify the cause of the relationship through:

- theoretical analysis
  further experiments

We successfully explained why the deformation changes the thermal conductivity from the viewpoint of physics, using vibrational modes on local structures

- We estimated the influence of local structures on thermal conductivity using vibrational modes from Allen–Feldman theory in physics
  We found that the deformation of five-vertex rings makes the thermal conductivity lower

Software

Which software do we use? You can use any software you like, but our software, HomCloud, is a good option!

- HomCloud has powerful inverse analysis tools, such as optimal and stable volumes, PH trees, and optimal 1-cycles
  HomCloud has vectorization functionality by persistence image and persistence codebook
  HomCloud has a useful Python interface, and you can use HomCloud with Python's rich machine-learning ecosystem
  All necessary components for our PH+ML framework are available and you can combine these components flexibly
  HomCloud is available on Linux, Apple Silicon Mac, Windows, and Google Colaboratory

`https://homcloud.dev`

Development

For better software, we are implementing the following practices:

- Dogfooding
  - The developer always uses HomCloud for PH data analysis
  - Software quality can be monitored at all times
  - Usability is always tested by the developers
- Continuous integration
  - The software is automatically built and tested on Linux, Windows, and Mac every time the code is changed

How to learn HomCloud

HomCloud has the Python interface and CLI, but the Python interface is recommended. If you want to try HomCloud, it is best to start with Google Colaboratory.

- Only a web browser and a Google account
- Lightweight installation process

HomCloud provides tutorials at `https://homcloud.dev/tutorials.en.html`, starting with "3D Point Cloud Data Analysis". After completing that tutorial, you should proceed to other tutorials that interest you.

HomCloud Documentation

HomCloud provides two types of documentation on the website.

- Tutorials
  - Each tutorial provides the complete set of introductory knowledge about persistent homology data analysis
  - You can choose tutorials on topics that interest you.
    - 3D point cloud analysis, binary or grayscale image analysis
    - More advanced topics: atomic configurations, machine learning
- Python API documentation
  - Detailed descriptions of each class and method
  - If the API document is insufficient, you can use the source code of HomCloud itself

How to cite HomCloud

You can use HomCloud freely since it is open source software.

Please cite the following review paper [3] about persistent homology and HomCloud. You may also refer to the URL of this website, `https://homcloud.dev/index.en.html`, with HomCloud's version.

I. Obayashi, T. Nakamura, and Y. Hiraoka. "Persistent Homology Analysis for Materials Research and Persistent Homology Software: HomCloud". In: J. Phys. Soc. Jpn. 91.9 (2022), p. 091013. doi: 10.7566/JPSJ.91.091013.

Future of HomCloud

Now we plan:

- Continuous code refinement

  Updating documents
  - Recently, the tutorials have been reviewed and refreshed
  - The next step is to prepare code snippets

  Publicly publishing the source code repository
  - Now the source code of HomCloud is available by downloading the .tar.gz file, but the Git repository is not available.

More ambitious ideas for developing HomCloud:

- Parallel computation for optimal or stable volumes
  - Since they require significantly more cost than computing PDs.

  Non-rectangular unit cell for periodic boundary condition for 3d alpha filtrations
  - CGAL supports a rectangular unit cell, and HomCloud uses the feature
  - CGAL does not support a non-rectangular parallelepiped unit cell
  - It is a desirable feature, but difficult to implement

  Integration with connected PD [4] software, one of extensions of PDs using multi-parameter persistent homology

Acknowledgments

Collaborators:
Y. Hiraoka, M. Kimura, E. Minamitani, T. Shiga, M. Kashiwagi

Funding Support:
JSPS Kakenhi Grant number 16K17638, 19H00834, 19KK0068, 20H05884, 22H05106, JST PRESTO Grant Number JPMJPR1923, JST CREST Grant Number JPMJCR15D3.

Wrap up

- Applications of persistent homology to material research
- Our software, HomCloud, is available for PH data analysis

References

[1] Emi Minamitani, Takuma Shiga, Makoto Kashiwagi, and Ippei Obayashi. Topological descriptor of thermal conductivity in amorphous Si. Journal of Chemical Physics **156**, 244502 (2022)
[2] Ippei Obayashi, Yasuaki Hiraoka, and Masao Kimura. Persistence diagrams with linear machine learning models. Journal on Applied and Computational Topology **1**, 3–4, 421–449, (2018).
[3] I. Obayashi, T. Nakamura, and Y. Hiraoka. Persistent Homology Analysis for Materials Research and Persistent Homology Software: HomCloud. J. Phys. Soc. Jpn. **91**, 091013 (2022)
[4] Y. Hiraoka, K. Nakashima, I. Obayashi, and C. Xu. Refinement of Interval Approximations for Fully Commutative Quivers. Arxiv preprint (2023). `https://arxiv.org/abs/2310.03649`

Understanding Depression during the COVID-19 Pandemic as Topographical Maps

Daiki Tatematsu

Graduate School of Science, Nagoya University, Japan

The COVID-19 pandemic changed our lifestyles. It is expected that the changes in mental health also occurred because of these changes. In this study, we used the questionnaire responses that asked high school students in Tokyo about their depression before and during the COVID-19 pandemic and analyzed the group characteristics of changes in depression as topographical maps using energy landscape analysis (ELA), a method of multidimensional time-series data analysis. As a result, we visualized how the topographical maps of the depression changed in the COVID-19 pandemic and found that the COVID-19 pandemic has made the students less likely to become depressed. These results suggest that ELA is useful for the analysis of psychiatric questionnaires.

References.

[1] Takahiro Ezaki et al. "Energy landscape analysis of neuroimaging data". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 375.2096 (2017), p. 20160287.

[2] Ronald C Kessler et al. "Screening for serious mental illness in the general population". In: *Archives of general psychiatry* 60.2 (2003), pp. 184–189.

[3] Ingrid A van de Leemput et al. "Critical slowing down as early warning for the onset and termination of depression". In: *Proceedings of the National Academy of Sciences* 111.1 (2014), pp. 87–92.

[4] Takamitsu Watanabe et al. "Energy landscape and dynamics of brain activity during human bistable perception". In: *Nature communications* 5.1 (2014), p. 4765.

Topological Data Analysis and Industrial Mathematics 2025 (TDA+IM 2025)
Thursday, Aug 7th 2025, 10:50AM-11:15AM(JST)

Understanding depression during the COVID-19 pandemic as topographical maps

Daiki Tatematsu[1], Shingo Iwami[1]
[1]interdisciplinary Biology Laboratory (iBLab), Dept. Biological Science, Grad. Sch. Science, Nagoya University, Japan

iBLab

DAIKI TATEMATSU iBLab interdisciplinary Biology Laboratory, Graduate School of Science, Nagoya University, JAPAN

Depression in the COVID-19 Pandemic

Research the Relationship between the **COVID-19 Pandemic** and **Depression**
for the Next Pandemic

DAIKI TATEMATSU iBLab interdisciplinary Biology Laboratory, Graduate School of Science, Nagoya University, JAPAN

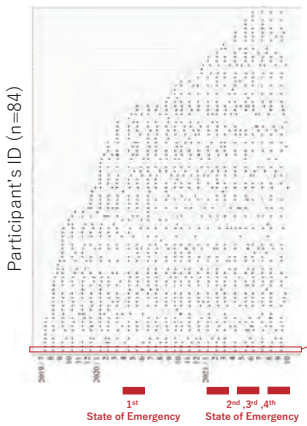
Survey for High School Students in Tokyo

【青春期の健康・発達調査】
TEEN COHORT

Analyze the data collected from
the survey for **84 high school students**
who participate in the **Tokyo Teen Cohort**

S.Ando *et al.*, ***Int. J. Epidemiol.***, 2019
N. Okada *et al.*, ***Psychiatry Clin. Neurosci.***, 2019

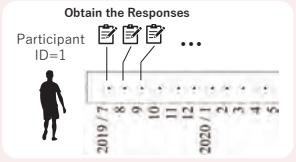
Monthly Questionnaire for Approximately 2 years, including the COVID-19 Pandemic



Questionnaire about Depression (K6)

Item	During the last 30 days,	None of the time	A little of the time	Some of the time	Most of the time	All of the time
K6-1	how often did you feel nervous?	0	1	2	3	4
K6-2	how often did you feel hopeless?	0	1	2	3	4
K6-3	how often did you feel restless or fidgety?	0	1	2	3	4
K6-4	how often did you feel so depressed that nothing could cheer you up?	0	1	2	3	4
K6-5	how often did you feel that everything was an effort?	0	1	2	3	4
K6-6	how often did you feel worthless?	0	1	2	3	4

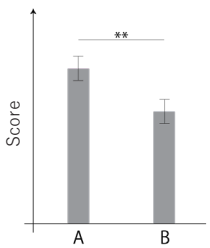
Kessler R. C. *et al.*, ***Arch. Gen. Psychiatry***, 2003



Understand the Mental Health as Topographical Maps

Psychiatric questionnaires
such as K6 are commonly analyzed
with **significance tests after scoring**

The mental state should be understood
simply and clearly as **topographical maps**



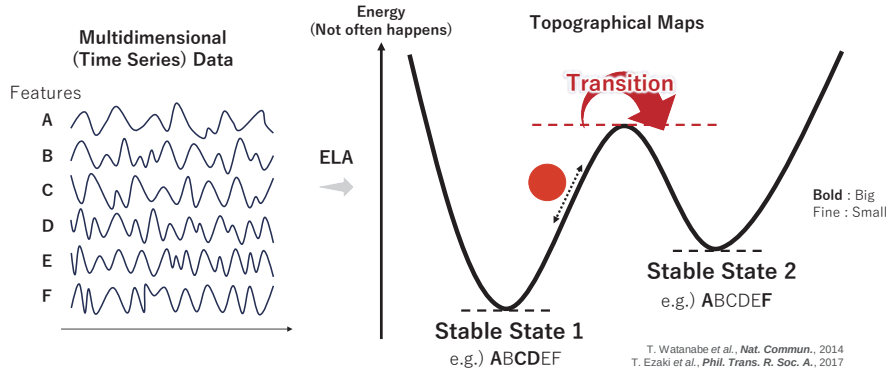
A **deep** valley is formed.
→ **Small** Variance & **Small** Autocorrelation

A **shallow** valley is formed.
→ **Large** Variance & **Small** Autocorrelation

I. A. Van de Leemput *et al.*, ***PNAS***, 2014
M. Scheffer *et al.*, ***JAMA Psychiatry***, 2024

These previous studies used **time series analysis**
and **the topographical map of mental Health is only used as a concept**

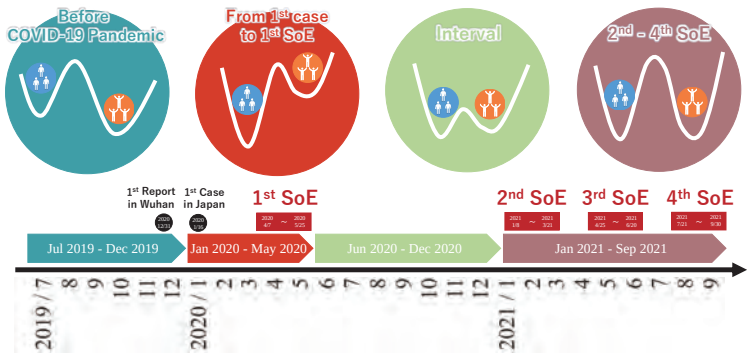
Energy Landscape Analysis (ELA)



T. Watanabe *et al.*, ***Nat. Commun.***, 2014
T. Ezaki *et al.*, ***Phil. Trans. R. Soc. A.***, 2017

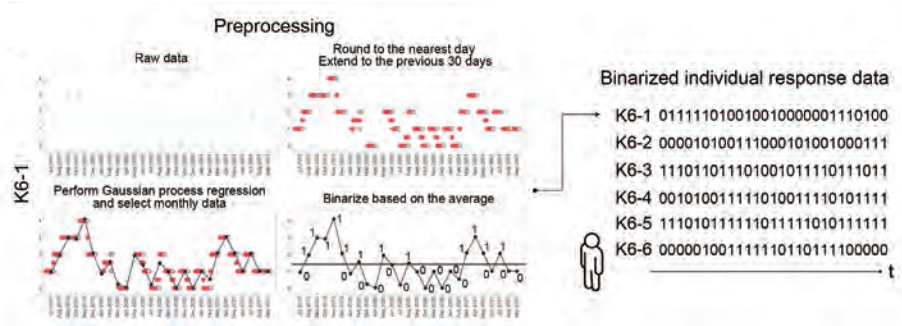
Understand **Multidimensional (Time Series) Data** as **Topographical Maps**

Topographical Maps of Depression during COVID-19 Pandemic



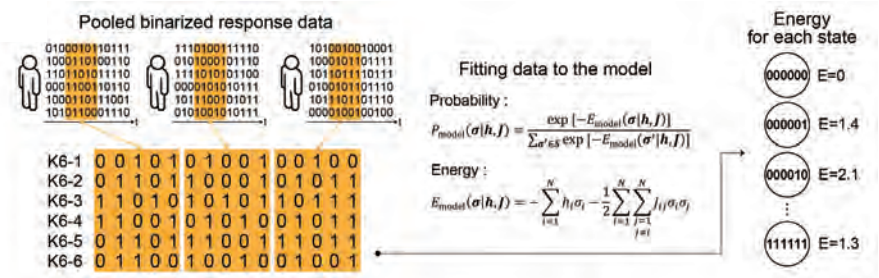
Analyze **the Changes in Topographical Maps of Depression**
during **the COVID-19 Pandemic**

Flow Chart of the Analysis 1 : Preprocessing



D. Tatematsu, N. Nakamura *et al.*, ***in revision***

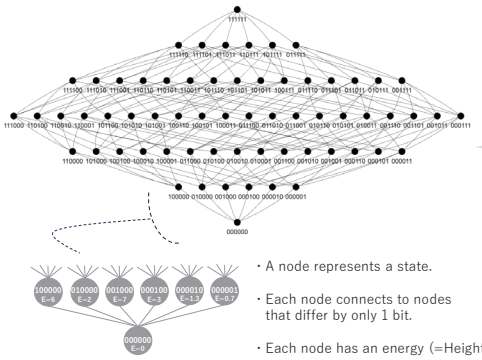
Flow Chart of the Analysis 2 : ELA (Maximum Entropy Model)



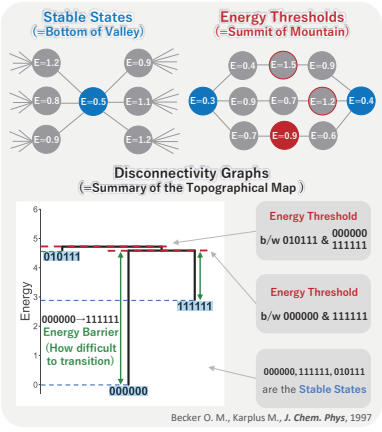
Calculate **the Energies (=Heights)** of **All Possible States (2^6=64 states)**

Flow Chart of the Analysis 3 : ELA (Disconnectivity Graphs)

The 6 Features Network
Expressing the **Transition** in This Research
(=Topographical Map of Depression)

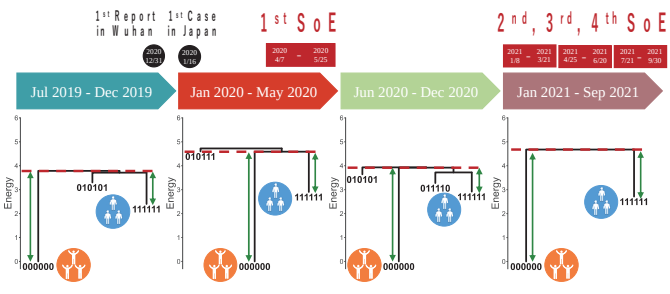


- A node represents a state.
- Each node connects to nodes that differ by only 1 bit.
- Each node has an energy (=Height).



Becker O. M., Karplus M., *J. Chem. Phys*, 1997

Topographical Maps of Depression in 4 Periods



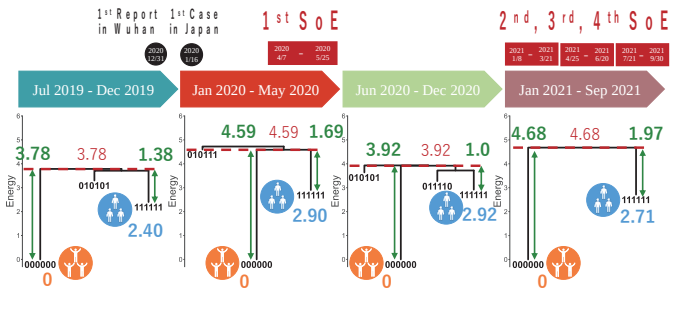
e.g.)

010111

0 ··· Below average
1 ··· Above average within individual

Item	During the last 30 days,
K6-1	how often did you feel nervous?
K6-2	how often did you feel hopeless?
K6-3	how often did you feel restless or fidgety?
K6-4	how often did you feel so depressed that nothing could cheer you up?
K6-5	how often did you feel that everything was an effort?
K6-6	how often did you feel worthless?

Topographical Maps of Depression in 4 Periods



e.g.)

010111

0 ··· Below average
1 ··· Above average within individual

Item	During the last 30 days,
K6-1	how often did you feel nervous?
K6-2	how often did you feel hopeless?
K6-3	how often did you feel restless or fidgety?
K6-4	how often did you feel so depressed that nothing could cheer you up?
K6-5	how often did you feel that everything was an effort?
K6-6	how often did you feel worthless?

Less Likely to Become Depressed during **the COVID-19 Pandemic**
More Likely to Remain in the Previous State during **the SoEs**

D. Tatematsu, N. Nakamura *et al.*, ***in revision***

Comparison with Previous Studies

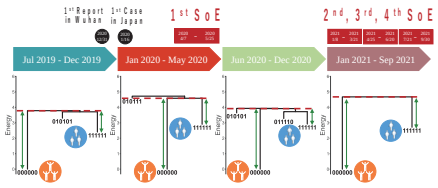
Chinese junior high school students (13-15 y/o)
significantly decreased depression scores
during the online class period under the lockdown.

→ **Scored** and **2 point Significance Test**

Japanese high school students (16-18 y/o)
are **less likely to become depressed**
during the COVID-19 pandemic
(and **more likely to remain in the previous state** during the SoEs).

→ **Topographical Maps of Depression**

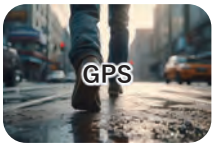
Qu M., Yang K., Cao Y. *et al.*, ***JAMA Netw. Open***, 2022



D. Tatematsu, N. Nakamura *et al.*, ***in revision***

As overall population trends, **the COVID-19 lifestyle**
seems to have **a positive impact** on the student's depression

Next Step



Other Types of Data Associated with the Subjects

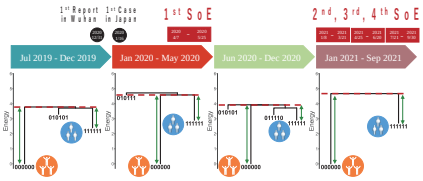
Summary

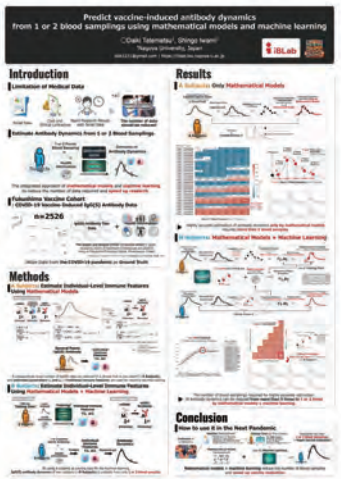
- **Questionnaire** for Approximately 2 years, including **the COVID-19 Pandemic**, for **High School Students in Tokyo**

  → We analyzed the relationship between **the COVID-19 pandemic** and **depression** using **Energy Landscape Analysis (ELA)**.

- **Topographical Maps of Depression** during **the COVID-19 Pandemic**

  → **Less Likely to Become Depressed** during **the COVID-19 Pandemic.**
  More Likely to Remain in the Previous State during **the SoEs.**





Poster Session

Predict vaccine-induced antibody dynamics from 1 or 2 blood samplings using mathematical models and machine learning

While biomedical data is highly accurate, the amount of data is limited, and there is a need to develop analytical methods that effectively utilize a small amount of data. In this study, we used data collected from approximately 2,500 individuals in the Fukushima vaccine cohort, Japan's largest and longest cohort for the COVID-19 vaccine. By applying an integrated approach of mathematical models and machine learning, we estimated IgG(S) antibody titer dynamics from 1 or 2 IgG(S) antibody titer data, age, and sex. This means that IgG(S) antibody titer data at any given time can be predicted from 1 or 2 blood samples. This approach can be applied to speeding up vaccine evaluation.

Acknowledgments

Data Acquisition, Psychiatry

Shinsuke Koike
University of Tokyo Institute for Diversity & Adaptation of Human Mind (UTIDAHM), Tokyo, Japan

Kiyoto Kasai
Department of Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

Analysis Methods (ELA)

Tetsuo Ishikawa
Department of Extended Intelligence for Medicine, The Ishii-Ishibashi Laboratory, Keio University School of Medicine, Tokyo, Japan

Masato S Abe
Faculty of Culture and Information Science, Doshisha University, Kyoto, Japan

Takahiro Ezaki
Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

Naoki Masuda
Department of Mathematics, State University of New York at Buffalo, Buffalo, NY, USA

Eiryo Kawakami
Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan

Symmetric Simplicial Lifting for Hypergraph Learning

Seongjin Choi

Department of Mathematics, POSTECH, Korea

The formulation of higher-degree sheaf Laplacians on hypergraphs is hindered by the fundamental challenges of sparsity and orientational ambiguity. To address this, we propose a foundational methodology: symmetric simplicial lifting. This technique embeds a hypergraph into a richer structure, allowing for the systematic construction of Laplacians of arbitrary degree. We validate our framework with the Hypergraph Neural Sheaf Diffusion (HNSD) model, which leverages a degree-zero sheaf Laplacian to learn diffusion processes. The model achieves strong performance on key hypergraph benchmarks, demonstrating that our approach offers a principled pathway for higher-order signal analysis. This work is joint with Gahee Kim and Yong-Geun Oh.

References.

[1] Seongjin Choi, Gahee Kim, and Yong-Geun Oh. "Hypergraph Neural Sheaf Diffusion: A Symmetric Simplicial Set Framework for Higher-Order Learning". In: *IEEE Access* (2025).

[2] Iulia Duta et al. "Sheaf hypergraph networks". In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 12087–12099.

[3] Yifan Feng et al. "Hypergraph neural networks". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 3558–3565.

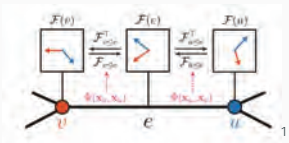
[4] Jakob Hansen and Thomas Gebhart. "Sheaf neural networks". In: *arXiv preprint arXiv:2012.06333* (2020).

Symmetric Simplicial Lifting for Hypergraph Learning

Seongjin Choi
POSTECH
joint work with Gahee Kim(KAIST AI), Yong-Geun Oh(POSTECH)

TDA + IM 2025
7th August 2025

Background: Sheaf Laplacian on graphs



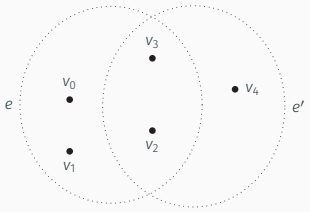
[1]

Figure 1: Node feature x_v is diffused to u by $-\mathcal{F}^T_{u \trianglelefteq e}\mathcal{F}_{v \trianglelefteq e}(x_v)$.

- Cellular sheaf transfers node feature to its incident edge.
- Graph Laplacian $\Longrightarrow$ Sheaf Laplacian: How the node features are diffused to its adjacent nodes in complicated ways.

[1] Image from Bodnar, Cristian, et al. "Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns." Advances in Neural Information Processing Systems 35 (2022): 18527-18541.

Adjacency in hypergraphs



- Nodes v, v' are adjacent if $\exists e \ni v, v'$ with $|e| = 2$.
- Hyperedges e, e' are adjacent if $|e| = |e'|$ and $\exists e''$ such that
  - $e, e' \subset e'', |e''| - |e| = 1$: upper adjacent OR
  - $e, e' \supset e'', |e| - |e''| = 1$: lower adjacent
- Degree n ($n > 1$) sheaf Laplacian describes how features of hyperedge of cardinality $n + 1$ are diffused to its adjacent hyperedges.

Problem: Sparsity of adjacent hyperedges

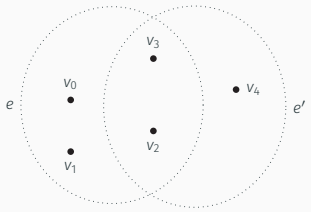
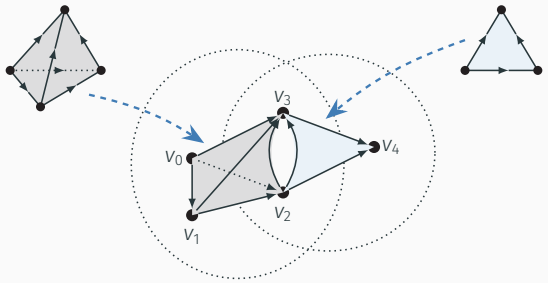


Figure 2: Degree n sheaf Laplacian vanishes for $\forall n$.

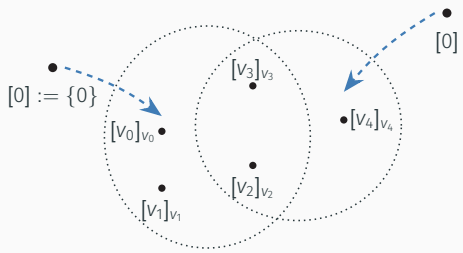
- The existence of a hyperedge does not imply non-vanishing Laplacians.
- No hyperedge e'' with $|e''| = 2 \implies L^0 = 0$.
- No adjacent hyperedge pair $\implies L^n = 0$ for $n > 0$.

Idea: Embed hypergraph into its configuration space



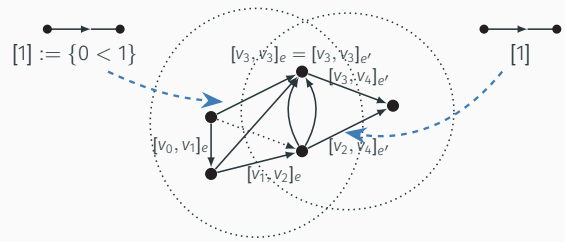
- hyperedge e, e': observed relations
- All possible oriented subrelations from the observed relations.
- Ex. (v_0, v_1, v_2, v_3) in e, (v_2, v_3, v_4) in e'.

0-dim'l objects in configuration space



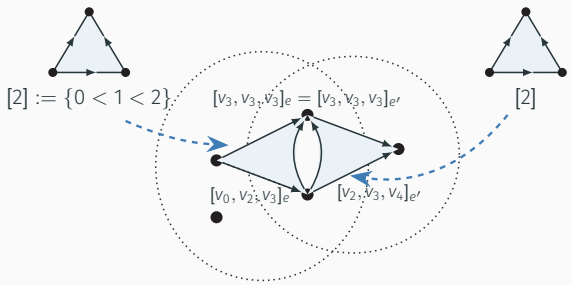
- $\{f : [0] \to x \mid x \in E \bigsqcup V\}$: f is represented by $(v)_x$ for $v \in x$
- $(v_2)_e \sim (v_2)_{e'} \sim (v_2)_v$: Represent the shared node v_2
- $\Delta(H)_0 := \{f : [0] \to x \mid x \in E \bigsqcup V\} / \sim$

1-dim'l objects in configuration space



- $\{f : [1] \to x \mid x \in E \bigsqcup V\}$: f is represented by $(v, w)_x$ for $v, x \in x$
- $(v_2, v_2)_e \sim (v_2, v_2)_{e'} \sim (v_2, v_2)_{v_2}$: Represent the shared node v_2
- $\Delta(H)_1 := \{f : [1] \to x \mid x \in E \bigsqcup V\} / \sim$

2-dim'l objects in configuration space



- $\{f : [2] \to x \mid x \in E \bigsqcup V\}$: f is represented by $(u, v, w)_x$ for $u, v, w \in x$
- $(v_2, v_2, v_2)_e \sim (v_2, v_2, v_2)_{e'} \sim (v_2, v_2, v_2)_{v_2}$: Represent the shared node v_2
- $\Delta(H)_2 := \{f : [2] \to x \mid x \in E \bigsqcup V\} / \sim$

Mathematical structure of $\Delta(H)$

We will show that $\Delta(H) := \{\Delta(H)_n\}_{n \in \mathbb{N}}$ is a symmetric simplicial set.

Intuitions of symmetric simplicial set

- A collection of oriented n-dimensional objects (n-simplices) represented by tuples.
- Given any n-simplex $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_n)$, any new tuple formed from the vertex set $\{\sigma_0, \ldots, \sigma_n\}$ must also be a simplex (e.g., $(\sigma_{\mu(0)}, \cdots, \sigma_{\mu(m)})$ for $\mu : [m] \to [n]$).
- Think of it as a 'tuple' version of a simplicial complex.

Symmetric simplicial set

Defn. Symmetric simplicial set

A collection $X = \{X_n\}_{n\in\mathbb{N}}$ is called a *symmetric simplicial set* if it is equipped with a family of functions

$$\{X(\mu) : X_n \to X_m\}_{\{\mu:[m]\to[n]\}} \tag{1}$$

satisfying if $\mu : [m] \to [n], \nu : [n] \to [p]$ then

$$X(\nu \circ \mu) = X(\mu) \circ X(\nu). \tag{2}$$

An element of X_n is called an n-simplex in X.

Equation (1) says $X(\mu) : (\sigma_0, \cdots, \sigma_n) \to (\sigma_{\mu(0)}, \cdots, \sigma_{\mu(n)})$.
Equation (2) says

$$X(\nu\circ\mu)(\sigma_0, \cdots, \sigma_p) = (X(\mu) \circ X(\nu))\,(\sigma_0, \cdots, \sigma_p) = (\sigma_{\nu\circ\mu(0)}, \cdots, \sigma_{\nu\circ\mu(p)}).$$

9

Tuple representation of n-simplex

Let $(i)_{[n]} : [0] \to [n]$ is a function with $(i)_{[n]}(0) = i$ for $i \in [n]$.
Equation (1) induces a map

$$X((i)_{[n]}) : X_n \to X_0.$$

Hence $\sigma_i := X((i)_{[n]})(\sigma) \in X_0$ for $i \in [n]$.

Defn. Tuple representation of σ

$(\sigma_0, \cdots, \sigma_n)$ is called the *tuple representation* of σ.

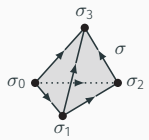


Figure 3: Tuple representation of 3-simplex σ

10

Facet of n-simplex

Let $\delta_i^n : [n-1] \to [n]$ is the unique order preserving injection from $\{0 < \cdots < n-1\}$ to $\{0 < \cdots < \hat{i} < \cdots < n\}$. Equation (1) induces a map

$$X(\delta_i^n) : X_n \to X_{n-1}.$$

Hence $d_i^n(\sigma) := X(\delta_i^n)(\sigma) \in X_{n-1}$ for $i \in [n]$.

Defn. ith facet of σ

The *ith facet* of σ, denoted by $d_i^n(\sigma)$, is defined as

$$X\left(\delta_i^n\right)(\sigma) \in X_{n-1}.$$

- We denote $\sigma \prec \tau$ and say σ is a facet of τ (or τ is a cofacet of σ) if σ is ith facet of τ for some $i \in [n]$.
- For $\sigma \prec \tau$, $[\sigma : \tau] := (-1)^i$ where $\sigma = d_i^n(\tau)$.

11

Illustration of facets

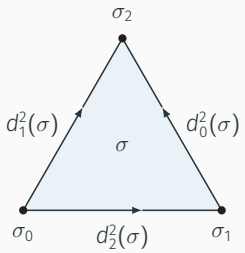


Figure 4: 3 Facets of 2-simplex σ.

Rmk.

The concept of an ith facet does not require an orientation on X.

12

Adjacency of two n-simplices

Since the 'facet' notion is defined, we can discuss adjacency.

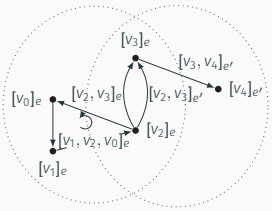
Defn. Upper/lower adjacency

Let σ, σ' be two n-simplices. σ, σ' are said to be

1. *upper adjacent* if $\exists \tau \in X_{n+1}$ such that
$$\sigma \prec \tau, \sigma' \prec \tau.$$
2. *lower adjacent* if $\exists \mu \in X_{n-1}$ such that
$$\mu \prec \sigma, \mu \prec \sigma'.$$

13

Example



Ex. Symmetric simplicial lifting $\Delta(H)$

- $\Delta(H)(\mu)\left([v_{j_0}, \cdots, v_{j_n}]_e\right) := [v_{j_{\mu(0)}, \cdots, v_{j_{\mu(m)}}}]$ for each $\mu : [m] \to [n]$.
- Tuple representation of $[v_1, v_2, v_0]_e$ is (v_1, v_2, v_0).
- 0th facet of $[v_1, v_2, v_0]_e$ is $[\widehat{v_1}, v_2, v_0]_e = [v_2, v_0]_e$.
- $[v_1, v_2]_e, [v_2, v_3]_e$ are lower adjacent since $[v_2]_e$ is common facet.
- $[v_2, v_3]_{e'}, [v_3, v_4]_{e'}$ are upper adjacent since $[v_2, v_3, v_4]_{e'}$ is common cofacet.

14

Cellular sheaf on symmetric simplicial set

Defn. Cellular sheaf of degree m

A cellular sheaf $(X, \mathcal{F})$ of degree m consists of the following data:

- For $n \in [m]$, n-simplex $\sigma \in X_n$, a $\mathbb{R}$-vector space $\mathcal{F}(\sigma)$, called the *stalk at* σ.
- For $n \in [m]$, n-simplex $\sigma \in X_n$ and facet $d_i^n(\sigma)$, a linear map

$$\mathcal{F}\left(d_i^n(\sigma) \prec \sigma\right) : \mathcal{F}\left(d_i^n(\sigma)\right) \to \mathcal{F}(\sigma)$$

satisfying the following compatibility conditions:

$$\begin{aligned} &\mathcal{F}\left(d_j^{n-1}(d_i^n(\sigma)) \prec d_i^n(\sigma)\right) \circ \mathcal{F}\left(d_i^n(\sigma) \prec \sigma\right) \\ &\quad = \mathcal{F}\left(d_{i-1}^{n-1}(d_j^n(\sigma)) \prec d_j^n(\sigma)\right) \circ \mathcal{F}\left(d_j^n(\sigma) \prec \sigma\right) \end{aligned} \tag{3}$$

for every $n \in [m]$, $i \in [n]$, $j < i$, and $\sigma \in X_n$.

Elements of $\mathcal{F}(\sigma)$ are called *features at simplex* σ.

Compatibility conditions

- We can easily check that

$$\delta_i^n \circ \delta_j^{n-1} = \delta_j^n \circ \delta_{i-1}^{n-1} : [n-2] \to [n].$$

- Equation (2) implies

$$d_j^{n-1} \circ d_i^n(\sigma) = d_{i-1}^{n-1} \circ d_j^n(\sigma). \tag{4}$$

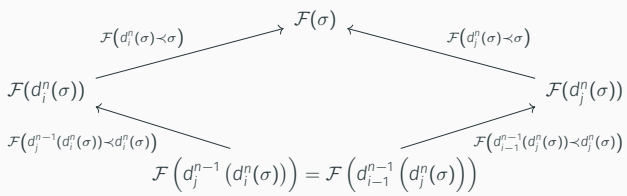


Figure 5: Condition (3) respects the identity (4).

Diffusion of features

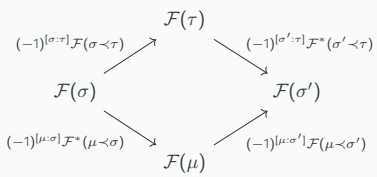


Figure 6: adjacent σ, σ' with one common cofacet τ, one common facet μ.

Feature x_σ at simplex σ is diffused to its adjacent simplex σ' by

$$\begin{aligned} &(-1)^{[\sigma:\tau]+[\sigma':\tau]}\mathcal{F}^*(\sigma' \prec \tau)\,\mathcal{F}(\sigma \prec \tau)(x_\sigma) \\ &+ (-1)^{[\mu:\sigma]+[\mu:\sigma']}\mathcal{F}(\mu \prec \sigma')\,\mathcal{F}^*(\mu \prec \sigma)(x_\sigma). \end{aligned}$$

Aggregating all such possible feature diffusions results in the construction of the *sheaf Laplacian*.

Degree n sheaf Laplacian

Defn. Degree n sheaf Laplacian

Let $\mathcal{F}$ be a cellular sheaf on X and $k \in \mathbb{N}$.

1. A *k-cochain*, denoted by $\mathbf{x} = (x_\sigma)_{\sigma \in X_k}$, is an element of the direct sum of stalks over all k-simplices in X.
2. For a k-cochain $\mathbf{x}$, the σ-component of the *degree k sheaf Laplacian* $L^k_{\mathcal{F}}(\mathbf{x})_\sigma$ is defined by

$$\sum_{\sigma',\tau} (-1)^{[\sigma:\tau]+[\sigma':\tau]} \mathcal{F}^*(\sigma \prec \tau)\mathcal{F}(\sigma' \prec \tau)(x_{\sigma'}) + \sum_{\sigma'',\mu} (-1)^{[\mu:\sigma]+[\mu:\sigma'']} \mathcal{F}(\mu \prec \sigma)\mathcal{F}^*(\mu \prec \sigma'')(x_{\sigma''})$$

for all possible σ', σ'', τ, and μ.

18

Normalized degree n sheaf Laplacian

Defn. Normalized degree n sheaf Laplacian

1. The σ-component of the diagonal blocks, denoted by $D^k_{\mathcal{F}}(\mathbf{x})_\sigma$, is defined as

$$\sum_{\{\tau \mid \sigma \prec \tau\}} \mathcal{F}^*(\sigma \prec \tau)\mathcal{F}(\sigma \prec \tau)(x_\sigma) + \sum_{\{\mu \mid \mu \prec \sigma\}} \mathcal{F}(\mu \prec \sigma)\mathcal{F}^*(\mu \prec \sigma)(x_\sigma).$$

2. For a k-cochain $\mathbf{x}$, the σ-component of the *normalized degree k sheaf Laplacian*, denoted by $\mathcal{L}^k_{\mathcal{F}}(\mathbf{x})_\sigma$, is defined by

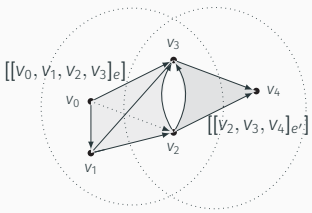
$$\left((D^k_{\mathcal{F}})^{-\frac{1}{2}} L^k_{\mathcal{F}} (D^k_{\mathcal{F}})^{-\frac{1}{2}}\right)(\mathbf{x})_\sigma.$$

19

Embedding property of Δ

Thm. Δ is an embedding.

$$\Delta(H) \cong \Delta(H') \iff H \cong H'.$$



Idea:

- $e \iff \{[v_0, v_1, v_2, v_3]_e, [v_1, v_0, v_2, v_3]_e, \cdots\}$
- $e' \iff \{[v_2, v_3, v_4]_{e'}, [v_3, v_2, v_4]_{e'}, \cdots,\}$

20

Degree 0 sheaf Laplacian of $\Delta(H)$

Thm. Formula for degree 0 sheaf Laplacian

For a 0-cochain $\mathbf{x}$, $L^0_{\mathcal{F}}(\mathbf{x})_{[v]_v}$ is

$$\sum_{\substack{\{w|w,v\in e \text{ for } e\}\\ \{e|v,w\in e\}}} \mathcal{F}^*([v]_v \prec [v,w]_e)\mathcal{F}([v]_v \prec [v,w]_e)(x_v)$$
$$+ \sum_{\substack{\{w|w,v\in e \text{ for } e\}\\ \{e|v,w\in e\}}} \mathcal{F}^*([v]_v \prec [w,v]_e)\mathcal{F}([v]_v \prec [w,v]_e)(x_v)$$
$$- \sum_{\substack{\{w|w,v\in e \text{ for } e\}\\ \{e|v,w\in e\}}} \mathcal{F}^*([v]_v \prec [v,w]_e)\mathcal{F}([w]_w \prec [v,w]_e)(x_w)$$
$$- \sum_{\substack{\{w|w,v\in e \text{ for } e\}\\ \{e|v,w\in e\}}} \mathcal{F}^*([v]_v \prec [w,v]_e)\mathcal{F}([w]_w \prec [w,v]_e)(x_w).$$

21

Generalization of sheaf Laplacian on graphs

Thm.

Suppose G is a graph and $\mathcal{F}$ is a cellular sheaf on G. Then $\mathcal{F}$ induces a cellular sheaf $\widehat{\mathcal{F}}$ on $\Delta(G)$ satisfying

$$\mathcal{L}^0_{\widehat{\mathcal{F}}} = \mathcal{L}_{\mathcal{F}}.$$

Idea: Define a cellular sheaf $\widehat{\mathcal{F}}$ on $\Delta(G)$ by

$$\widehat{\mathcal{F}}([v_{i_0},\cdots,v_{i_n}]_e) \coloneqq \begin{cases} \mathcal{F}(v) & \text{if } \{v_{i_0},\cdots,v_{i_n}\} = \{v\} \\ \mathcal{F}(e) & \text{otherwise.} \end{cases}$$

and linear map $\widehat{\mathcal{F}}([v_{i_0},\cdots,\widehat{v_{i_l}},\cdots,v_{i_n}]_e \prec [v_{i_0},\cdots,v_{i_n}]_e)$ by

$$\begin{cases} \mathcal{F}(v\in e) & \text{if } \{v_{i_0},\cdots,\widehat{v_{i_l}},\cdots,v_{i_n}\} \neq \{v_{i_0},\cdots,v_{i_n}\} \\ \mathrm{Id} & \text{otherwise.} \end{cases}$$

22

Cora dataset

Cora is a citation network of scientific publications.

- Node: a single scientific publication, a total of 2,708 papers.
- Hyperedge: connect papers that are co-cited by another publication.
- Node features: The presence (1) or absence (0) of each word from a dictionary of 1,433 unique words is recorded in this vector.
- Node labels: {Case Based, Genetic Algorithms, Neural Networks, Probabilistic Methods, Reinforcement Learning, Rule Learning, Theory}

23

Hypergraph Neural Sheaf Diffusion (HNSD)

Given: Hypergraph $H = (V, E)$ with 0-cochain $\mathbf{y}$ (node features)
Task: Predict a label of each node in H

1. Preprocess 0-cochain $\mathbf{y}$ with MLP so that $\mathbf{x} \in \mathbb{R}^d$. We fix $d = 8$.
2. $\mathbf{X}_0 \in \mathbb{R}^{|V|d \times f_0}(f_0 = 1)$ is a matrix with the first column $\mathbf{x}$
3. A general layer of HNSD on H is defined by

$$\mathbf{X}_{t+1} = \alpha\left((\mathrm{Id} - \mathcal{L}^0_{\mathcal{F}(t)})(\mathrm{Id} \otimes \mathbf{W}^t_1)\mathbf{X}_t\mathbf{W}^t_2\right) \in \mathbb{R}^{|V|d \times f_{t+1}}$$

where

- $t \in \{0, \cdots, L\}$ is layer with f_L= the number of labels, $1 \leq L \leq 8$
- α is nonlinear function
- $\mathbf{W}^1_t \in \mathbb{R}^{d \times d}, \mathbf{W}^2_t \in \mathbb{R}^{f_t \times f_{t+1}}$ are learnable matrices at layer t
- $\mathcal{F}(t)$ is cellular sheaf on $\Delta(H)$ at layer t

Hypergraph Neural Sheaf Diffusion (HNSD) (Continued)

We learn the cellular sheaf $\mathcal{F}$ on $\Delta(H)$ of degree 1 via

$$\mathcal{F}([v]_v \prec [v, w]_e) = \mathrm{MLP}(x_v \,\|\, x_{[v,w]_e}) \in \mathbb{R}^{d^2}$$

where

$$x_{[v,w]_e} = \alpha'\left(M\alpha\left(W^T\begin{bmatrix} x_v \\ 1 \end{bmatrix} \odot \begin{bmatrix} x_w \\ 1 \end{bmatrix}\right)\right)$$

with $\alpha = \mathrm{ReLU}$, $\alpha' = \tanh$.

Results: Node classification accuracy

	Cora	Citeseer	Cora-CA	DBLP-CA	Senate	AVG.
CEGCN	75.32 ± 1.69	71.43 ± 1.34	76.68 ± 1.30	87.19 ± 0.30	48.17 ± 3.68	71.74
HNHN	76.36 ± 1.92	72.64 ± 1.57	77.19 ± 1.49	86.78 ± 0.29	50.85 ± 3.35	72.76
LEGCN	72.23 ± 1.60	71.84 ± 1.17	72.23 ± 1.60	84.26 ± 0.40	73.24 ± 10.29	74.76
HCoN	51.77 ± 2.23	43.48 ± 1.12	72.37 ± 1.08	89.98 ± 0.26	46.28 ± 4.66	60.78
HyperGCN	74.19 ± 1.41	69.42 ± 3.49	70.00 ± 3.74	86.78 ± 2.39	53.66 ± 6.35	70.78
AllDeepSets	76.88 ± 1.80	70.83 ± 1.63	81.97 ± 1.50	**91.27** ± 0.27	48.17 ± 5.67	73.82
SheafHyperGNN	77.80 ± 2.24	73.93 ± 1.06	81.65 ± 1.50	88.93 ± 0.66	74.65 ± 5.90	79.39
HNSD(OURS)	**79.28** ± 0.82	**74.40** ± 1.47	**82.58** ± 1.15	89.85 ± 0.44	**78.45** ± 5.87	**80.91**

Summary

- We embed hypergraph into symmetric simplicial set, 'tuple' version of simplicial complex.
- Symmetric simplicial lifting solves problems of adjacency sparsity and lack of orientation. This enables the definition of higher-degree sheaf Laplacians.
- HNSD shows competitive performance across established benchmarks.

Thank you!

Seongjin Choi
Ph.D. Candidate

jincs10@postech.ac.kr

https://sites.google.com/view/seongjinchoi/home

Topological Data Analysis for Feature Extraction and Model Evaluation

Jisu Kim

Department of Statistics, Seoul National University, Korea

Topological Data Analysis (TDA) generally refers to utilizing topological features from data. A central topic in TDA is persistent homology, which observes data at various resolutions and summarizes topological features that persistently appear. TDA has been proven valuable in enhancing machine learning applications. This work explores how TDA can enhance machine learning workflows, focusing on two areas: feature extraction and model evaluation. Persistent homology, while rich in structural information, is often challenging to integrate directly into statistical and machine learning frameworks. To address this, various featurization techniques map persistence-based information into Euclidean or functional spaces, enabling its incorporation into neural networks and other learning algorithms. I will examine different approaches that efficiently transform topological summaries into differentiable layers and leverage geometric representations for visualization and dimensionality reduction. In addition to feature extraction, TDA has recently been applied to evaluate data quality and model performance. By quantifying the topological structure of generated or transformed data, TDA-based methods provide robust evaluation metrics that improve the reliability of model assessment. I will present how this is done in particular in generative modeling scenarios.

References.

[1] Kwangho Kim et al. "Pllay: Efficient topological layer based on persistent landscapes". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 15965–15977.

[2] Hajin Lee, Jisu Kim, and Kwangho Kim. "ECLayr: Fast and Robust Topological Layer based on Differentiable Euler Characteristic Curve". In: ().

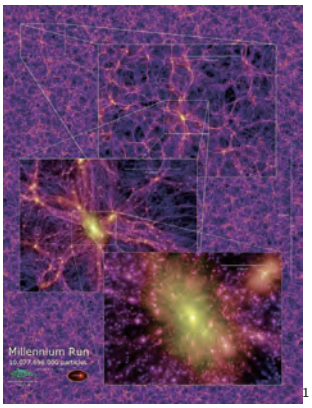
[3] Hengrui Luo et al. "Generalized penalty for circular coordinate representation". In: *arXiv preprint arXiv:2006.02554* (2020).

Topological Data Analysis for Feature Extraction and Model Evaluation

Jisu KIM

Topological Data Analysis and Industrial Mathematics
2025-08-08

Topological structures in the data provide information.



[1]

[1] http://www.mpa-garching.mpg.de/galform/virgo/millennium/poster_half.jpg

Persistent Homology: observe topological structure with multi resolutions.

4 / 65

Persistent Homology: observe topological structure with multi resolutions.

5 / 65

Persistent Homology: observe topological structure with multi resolutions.

6 / 65

Persistent Homology: observe topological structure with multi resolutions.

- Georges Seurat, A Sunday afternoon on the island of La Grande Jatte (Un dimanche après-midi à l'Île de la Grande Jatte)

A (very) rough introduction to Machine Learning

- For given problem and data, machine learning / deep learning learns a parametrized model.
  - Given data $\mathcal{X}$,
  - Parametrized model f_θ,
  - Loss funciton $\mathcal{L}$ adapted to a problem,
  - Machine Learning computes a solution that minimizes the loss function: $\arg\min_\theta \mathcal{L}(f_\theta, \mathcal{X})$.
- For many cases, computing an explicit formula for the minimizer is impossible or too expensive (e.g. inverting a large matrix). So, we often use gradient descent using $\nabla_\theta \mathcal{L}(f_\theta, \mathcal{X})$:

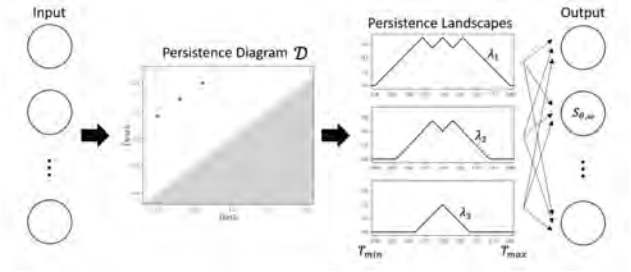
$$\theta_{n+1} = \theta_n - \lambda \nabla_\theta \mathcal{L}(f_\theta, \mathcal{X}).$$

Topological Data Analysis is applied to Machine Learning.

- A Survey of Topological Machine Learning Methods (Hensel, Moor, Rieck, 2021)
- Roughly, there are two directions applying Topological Data Analysis (TDA) to Machine Learning:
  - Make features from TDA to add topological features to data $\mathcal{X}$: more common
    - PLLay: Efficient Topological Layer based on Persistence Landscapes (Kim, Kim, Zaheer, Kim, Chazal, Wasserman, 2020)
    - Generalized penalty for circular coordinate representation (Luo, Patania, Kim, Vejdemo-Johansson, 2021)
    - ECLayr: Fast and Robust Topological Layer based on Differentiable Euler Characteristic Curve (Lee, Kim, Kim, 2025?)
  - Evaluate quality of data $\mathcal{X}$ or model f_θ using TDA: recently of interest
    - TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models (Kim, Jang, Kim, Yoo, 2024)

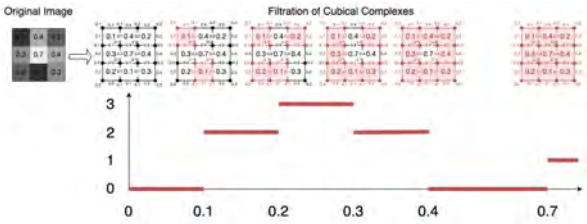
Topological structure is featurized as persistence landscape to be further applied in machine learning framework.

- Featurization using Persistence Landscape
  - PLLay: Efficient Topological Layer based on Persistent Landscapes (Kim, Kim, Zaheer, Kim, Chazal, Wasserman, 2020)



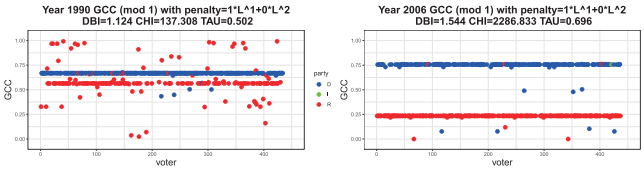
Topological structure is featurized as euler characteristic curve to be further applied in machine learning framework.

- Featurization using Euler Characteristic Curve
  - ECLayr: Fast and Robust Topological Layer based on Differentiable Euler Characteristic Curve (Lee, Kim, Kim, 2025?)



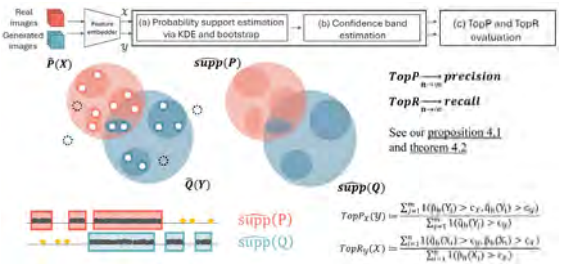
Topological structure is featurized as circular coordinates to be further applied in machine learning framework.

- Featurization using Circular Coordinates
  - Generalized penalty for circular coordinate representation (Luo, Patania, Kim, Vejdemo-Johansson, 2021)

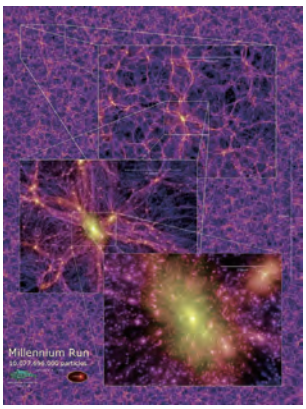


Data or Model is evaluated using Topological Data Analysis.

- Evaluation using Confidence of Topological Data Analysis
  - TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models (Kim, Jang, Kim, Yoo, 2024)



Topological holes in the data provide information.



The number of holes is used to summarize geometrical features.

- Geometrical objects :
  - ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ
  - A, あ, い, う, え, お
  - 福, 岡, 九, 州, 大, 学, 西, 新, プ, ラ, ザ
- The number of holes of different dimensions is considered.
  1. β_0 =# of connected components 
  2. β_1 =# of loops (holes inside 1-dim sphere) 
  3. β_2 =# of voids (holes inside 2-dim sphere) : if $dim \geq 3$ 

Example : Objects are classified by homologies.

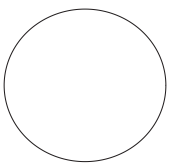
1. β_0 =# of connected components 
2. β_1 =# of loops 

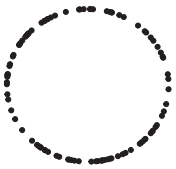
$\beta_0 \setminus \beta_1$	0	1	2	5
1	ㄱ, ㄴ, ㄷ, ㄹ, ㅅ, ㅈ, ㅋ, ㅌ, 九, 大	ㅁ, ㅇ, ㅂ, ㅍ, A	あ,西	
2	ㅊ, い, う, え, ラ	お, 新, プ		
3	岡, ザ	ㅎ		
4	学			
6	州			
7				福

Homology of finite sample is different from homology of underlying ma6nifold, hence it cannot be directly used for the inference.

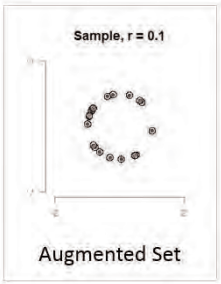
- When analyzing data, we prefer robust features where features of the underlying manifold can be inferred from features of finite samples.
- Homology is not robust:

Underlying circle: $\beta_0 = 1$, $\beta_1 = 1$ 100 samples: $\beta_0 = 100$, $\beta_1 = 0$



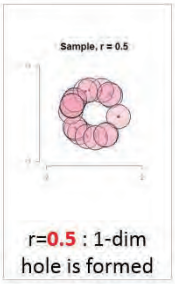


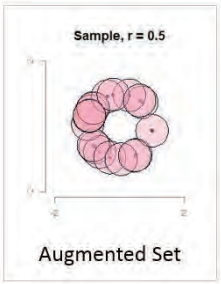
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.



19 / 65

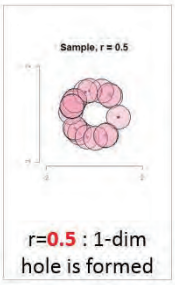
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

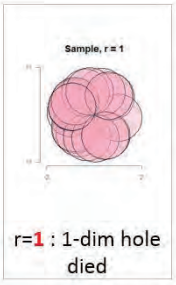


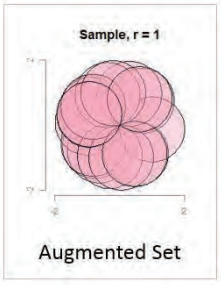


20 / 65

Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

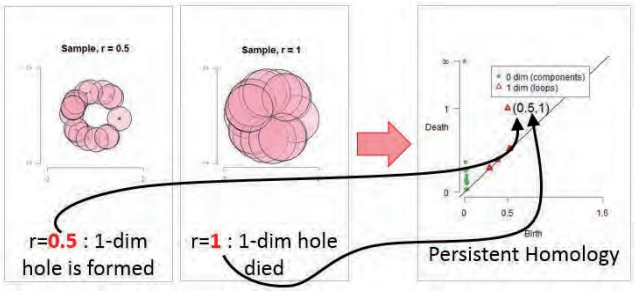






21 / 65

Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

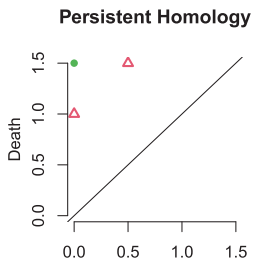


22 / 65

23 / 65

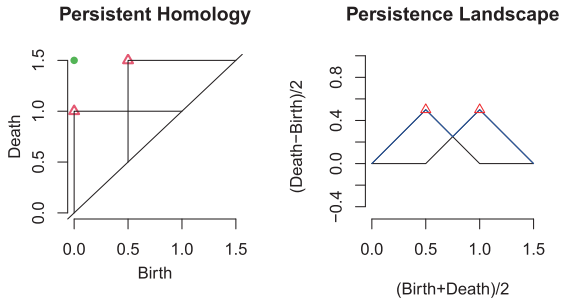
Persistent homology is further summarized and embedded into a Euclidean space or a functional space.

- The space of the persistent homology is complex, so directly applying in machine learning is difficult.
- If the persistent homology is further summarized and embedded into a Euclidean space or a functional space, then applying in machine learning becomes much more convenient.
  - e.g., Persistence Landscape, Persistence Silhouette, Persistence Image



24 / 65

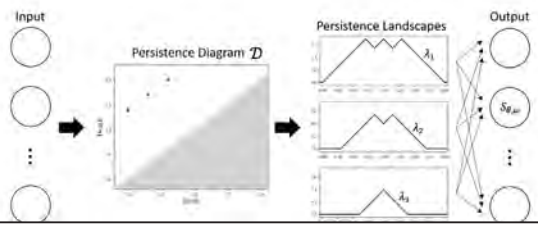
Persistence Landscape is a functional summary of the persistent homology.



PLLay: Build topological layer using Persistence Landscape

- PLLay: Efficient Topological Layer based on Persistence Landscapes (Kim, Kim, Zaheer, Kim, Chazal, Wasserman, 2020)

1. From data X, choose an appropriate simplicial complex K and a function f to compute the Persistece diagram $\mathcal{D}$.
2. From the persistence diagram $\mathcal{D}$, compute the persistence landscape $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$.
3. Compute the weighted average function $\bar{\lambda}_\omega(t) := \sum_{k=1}^{K_{\max}} \omega_k \lambda_k(t)$, and vectorize to get $\bar{\Lambda}_\omega \in \mathbb{R}^m$.
4. For a parametrized differentiable map $g_\theta : \mathbb{R}^m \to \mathbb{R}$, compute $S_{\theta,\omega}(\mathcal{D}) := g_\theta(\bar{\Lambda}_\omega)$.



PLLay is differentiable.

- A deep learning model learns its parameters by back propagation, which is to apply gradient descent layer-wise.
- For a deep learning layer to be learnable, it should be differentiable:

Theorem (Theorem 3.1 in Kim et al. [2020])

The PLLay function $S_{\theta,\omega}$ is differentiable with respect to the input data X.

PLLay is stable.

- PLLay is stable with respect to changes in persistence diagrams:

Theorem (Theorem 4.1 in Kim et al. [2020])

For two persistence diagrams $\mathcal{D}, \mathcal{D}'$,

$$|S_{\theta,\omega}(\mathcal{D}) - S_{\theta,\omega}(\mathcal{D}')| = O(d_B(\mathcal{D}, \mathcal{D}')),$$

where d_B is the bottleneck distance.

PLLay is stable.

- PLLay is stable with respect to perturbations in input X:

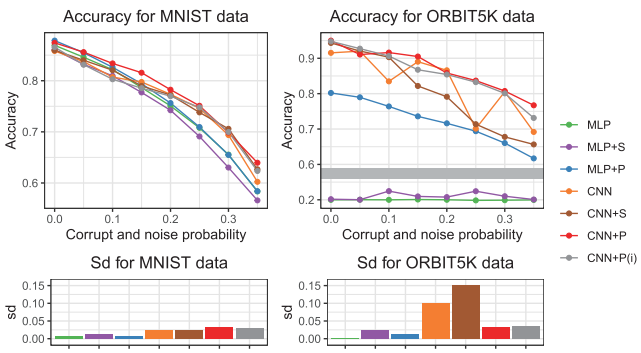
Theorem (Theorem 4.2 in Kim et al. [2020])

Let $X \sim P$ and P_n be the empirical distribution. Further, let $\mathcal{D}_P, \mathcal{D}_X$ be the persistence diagrams of P, X, respectively. Then

$$|S_{\theta,\omega}(\mathcal{D}_X) - S_{\theta,\omega}(\mathcal{D}_P)| = O(W_2(P_n, P)),$$

where W_2 is 2-Wasserstein distance.

Experiments

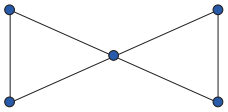


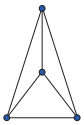
Euler Characteristic is computationally efficient.

- Euler Characteristic of a simplex or cubical complex is an alternating sum of betti numbers: for a simplex / cubical complex K,

$$\chi(K) = \sum_{k=0}^{\infty}(-1)^k|K^k| = \sum_{k=0}^{\infty}(-1)^k\beta_k,$$

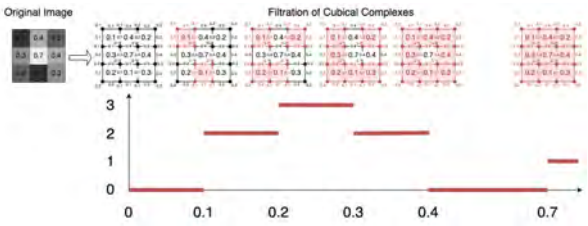
  where K^k is the set of k-dimensional simplices in K, and β_k is the k-th Betti number of K.

-  $\chi(K) = 5 - 6 = 1 - 2 = -1$

-  $\chi(K) = 4 - 6 = 1 - 3 = -2$

Euler Characteristic Curve is computationally efficient compared to Persistent Homology.

- Euler Characteristic Curve (ECC) $\mathcal{C} : \mathbb{R} \to \mathbb{R}$ computes the Euler characteristic along a filtration.
- ECC does not involve computing persistent homology, hence more computationally efficient compard to persistent homology.



EClayr: Build topological layer using Euler Characteristic Curves

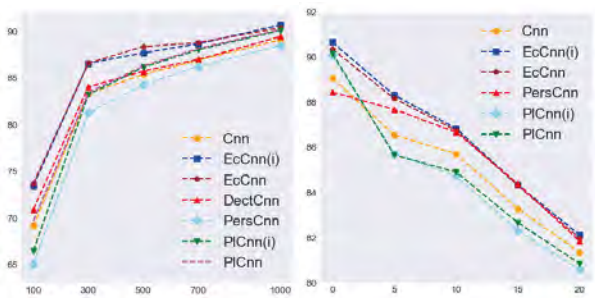
- ECLayr: Fast and Robust Topological Layer based on Differentiable Euler Characteristic Curve (Lee, Kim, Kim, 2025?)

1. From data X, choose an appropriate simplicial complex K and a function f to build a filtration.
2. From the filtration, compute the Euler Characteristic Curve $\mathcal{C} : \mathbb{R} \to \mathbb{R}$, and vectorize to get $\mathcal{E} \in \mathbb{R}^v$.
3. For a parametrized differentiable map $g_\theta : \mathbb{R}^m \to \mathbb{R}$, compute $\mathcal{O}_\theta := g_\theta(\mathcal{E})$.

Computation Time

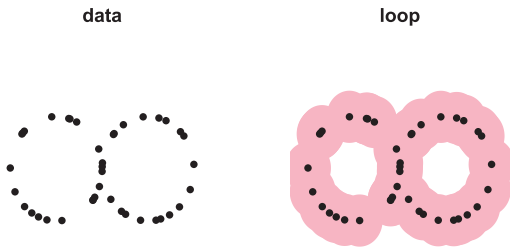
Model	Data		
	MNIST	Br35H	Synthetic
ECC	3.129 sec	0.458 sec	2.17 sec
PH	33.700 sec	11.033 sec	59.288 sec

Experiments



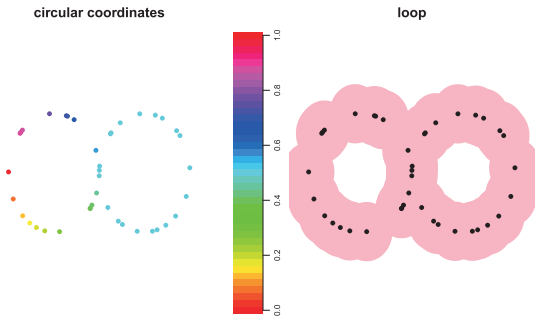
Circular coordinates provide topological representations of reduced dimension.

- Persistent cohomology and circular coordinates (de Silva, Morozov, Vejdemo-Johansson, 2011)
- Topological Learning for Motion Data via Mixed Coordinates (Vejdemo-Johansson, Pokorny, Skraba, Kragic, 2015)



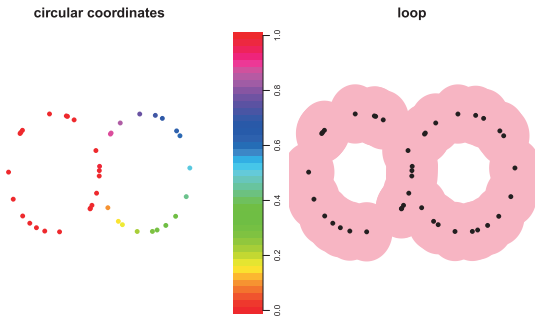
Circular coordinates provide topological representations of reduced dimension.

- circuiar coordinate is a function that maps from data points X to circle S^1.



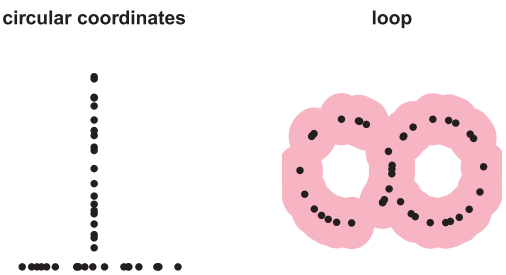
Circular coordinates provide topological representations of reduced dimension.

- circuiar coordinate is a function that maps from data points X to circle S^1.



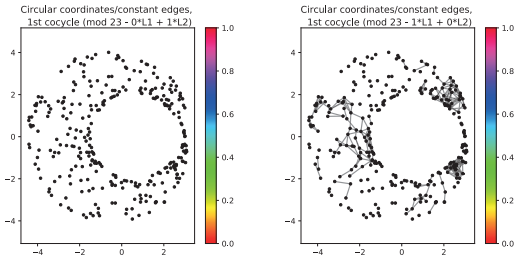
Circular coordinates provide topological representations of reduced dimension.

- circuiar coordinate is a function that maps from data points X to torus $\mathbb{T}^k = (S^1)^k$.



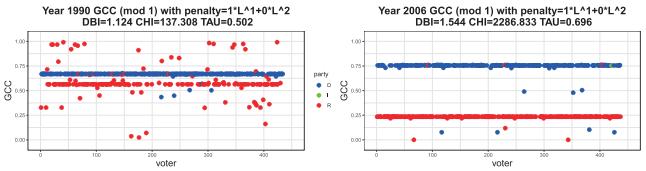
Circular coordinates with generalized penalty better visualizes topological information from data.

- Generalized penalty for circular coordinate representation (Luo, Patania, Kim, Vejdemo-Johansson, 2021)
- When computing circular coordinates, we solve an optimization problem.
- We switch L_2 loss by L_1 loss for circuiar coordinate values to change more abructly: better visualizes topological information from data.



Circular coordinates with generalized penalty better visualizes topological information from data.

- Generalized penalty for circular coordinate representation (Luo, Patania, Kim, Vejdemo-Johansson, 2021)
- Voting data in 2006 is more bipolarized than voting data in 1990.

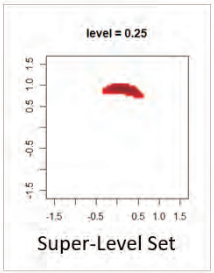


We rely on the kernel density estimator to extract topological information of the underlying distribution.

- The kernel density estimator is

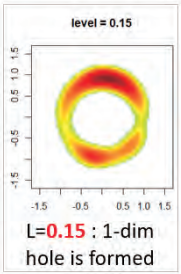
$$\hat{p}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

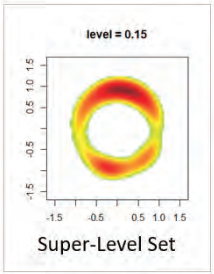
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.



46 / 65

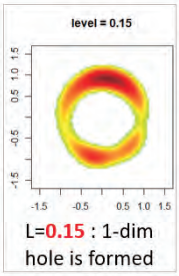
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

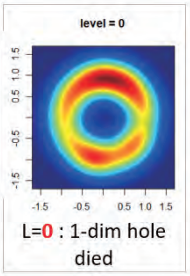


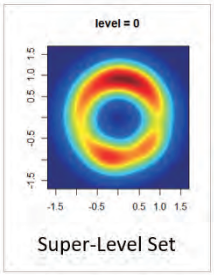


47 / 65

Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

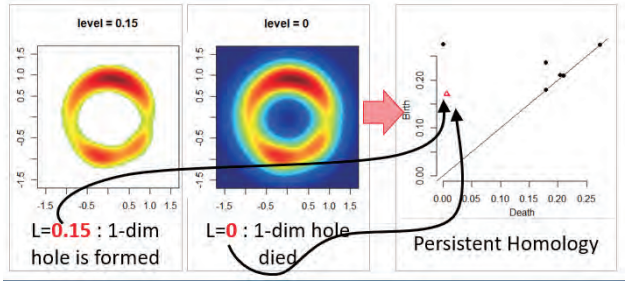






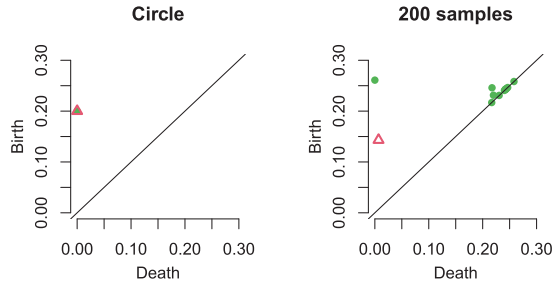
48 / 65

Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.



49 / 65

Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.

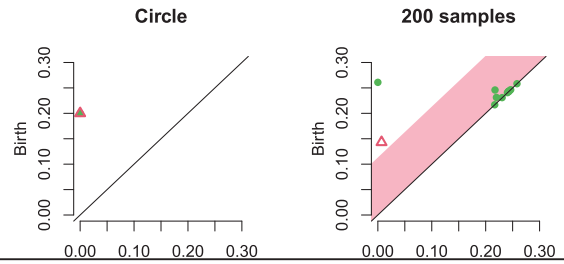


50 / 65

Confidence band for persistent homology separates homological signal from homological noise.

Let M be a compact manifold, and $X = \{X_1, \cdots, X_n\}$ be n samples. Let f_M and f_X be corresponding functions whose persistent homology is of interest. Given the significance level $\alpha \in (0, 1)$, $(1 - \alpha)$ confidence band $c_n = c_n(X)$ is a random variable satisfying

$$\mathbb{P}\left(d_B(Dgm(f_M), Dgm(f_X)) \leq c_n\right) \geq 1 - \alpha.$$



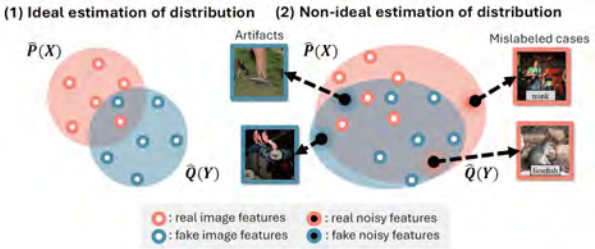
51 / 65

Confidence band for the persistent homology can be computed using the bootstrap algorithm.

1. Given a sample $X = \{x_1, \ldots, x_n\}$, compute the kernel density estimator $\hat{p}_h$.
2. Draw $X^* = \{x_1^*, \ldots, x_n^*\}$ from $X = \{x_1, \ldots, x_n\}$ (with replacement), and compute $\theta^* = \sqrt{nh^d}||\hat{p}_h^*(x) - \hat{p}_h(x)||_\infty$, where $\hat{p}_h^*$ is the density estimator computed using X^*.
3. Repeat the previous step B times to obtain $\theta_1^*, \ldots, \theta_B^*$
4. Compute $\hat{z}_\alpha = \inf\left\{q : \frac{1}{B}\sum_{j=1}^{B} I(\theta_j^* \geq q) \leq \alpha\right\}$
5. The $(1-\alpha)$ confidence band for $\mathbb{E}[p_h]$ is $\left[\hat{p}_h - \frac{\hat{z}_\alpha}{\sqrt{nh^d}}, \hat{p}_h + \frac{\hat{z}_\alpha}{\sqrt{nh^d}}\right]$.

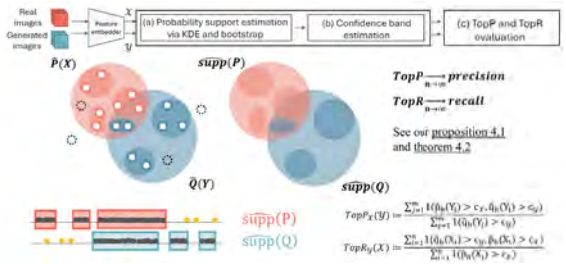
Existing evaluation metrics for generative models are vulnerable to noise.

- TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models (Kim, Jang, Kim, Yoo, 2024)
- To evaluate generative models, metrics compare the support of real image distributions and fake image distributions.
- Existing evaluation metrics tend to overestimate the support of the data distribution: vulnerable to noise



TopP&R robustly evaluates generative models by retaining only topologically and statistically significant features with confidence.

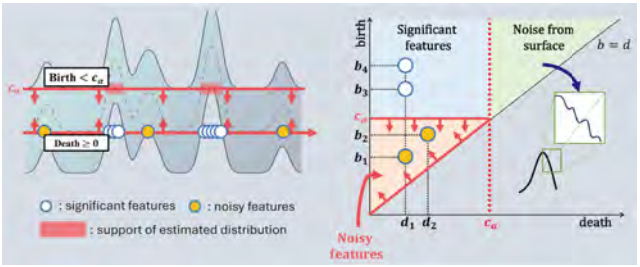
- TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models (Kim, Jang, Kim, Yoo, 2024)



55 / 65

We find threshold c_α that selects statistically and topologically significant features.

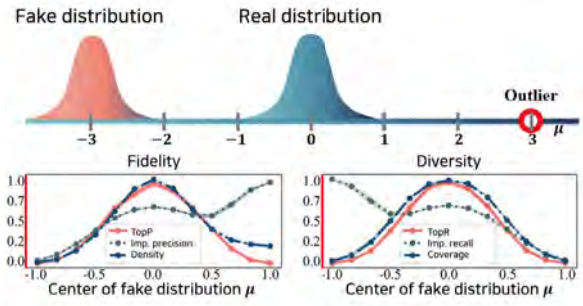
- TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models (Kim, Jang, Kim, Yoo, 2024)



56 / 65

Experiments

- TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models (Kim, Jang, Kim, Yoo, 2024)



57 / 65

There are many programs for Topological Data Analysis.

- There are many programs for Topological Data Analysis: e.g., Dionysus, DIPHA, GUDHI, javaPlex, Perseus, PHAT, Ripser, TDA, TDAstats

R Package TDA provides an R interface for C++ libraries for Topological Data Analysis.

- website: `https://cran.r-project.org/web/packages/TDA/index.html`
- Author: Brittany Terese Fasy, Jisu Kim, Fabrizio Lecci, Clément Maria, David Milman, and Vincent Rouvreau.
- R is a programming language for statistical computing and graphics.
- R has short development time, while C/C++ has short execution time.
- R package TDA provides an R interface for C++ library GUDHI/Dionysus/PHAT, which are for Topological Data Analysis.

References I

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *arXiv preprint arXiv:1207.6437*, 2012.

Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. *arXiv preprint arXiv:1207.3674*, 2012.

Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Annual Symposium on Computational Geometry*, pages 474–483. ACM, 2014.

Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. Persistent cohomology and circular coordinates. *Discrete & Computational Geometry*, 45(4):737–759, 2011.

H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. Applied mathematics. American Mathematical Society, 2010. ISBN 9780821849255. URL `http://books.google.com/books?id=MDXa6gFRZuIC`.

References II

Felix Hensel, Michael Moor, and Bastian Rieck. A survey of topological machine learning methods. *Frontiers Artif. Intell.*, 4:681108, 2021. doi: 10.3389/frai.2021.681108. URL `https://doi.org/10.3389/frai.2021.681108`.

Kwangho Kim, Jisu Kim, Manzil Zaheer, Joon Sik Kim, Frédéric Chazal, and Larry Wasserman. PLLay: Efficient Topological Layer based on Persistent Landscapes. *arXiv e-prints*, art. arXiv:2002.02778, February 2020.

Pum Jun Kim, Yoojin Jang, Jisu Kim, and Jaejun Yoo. TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models. *arXiv e-prints*, art. arXiv:2306.08013, June 2024. doi: 10.48550/arXiv.2306.08013.

Hengrui Luo, Alice Patania, Jisu Kim, and Mikael Vejdemo-Johansson. Generalized penalty for circular coordinate representation. *Foundations of Data Science*, 3(4):729–767, 2021.

References III

Mikael Vejdemo-Johansson, Florian T Pokorny, Primoz Skraba, and Danica Kragic. Cohomological learning of periodic motion. *Applicable algebra in engineering, communication and computing*, 26(1):5–26, 2015.

Thank you!

Statistical Inference for Persistent Homology

Featurization using Persistent Homology

R Package TDA: Statistical Tools for Topological Data Analysis
- Sample on manifolds, Distance Functions, and Density Estimators
- Persistent Homology and Persistence Landscape
- Statistical Inference on Persistence Homology and Persistence Landscape

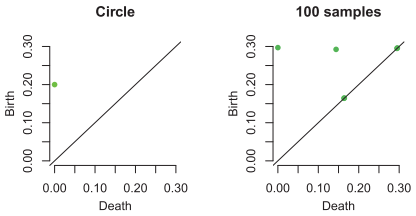
Bottleneck distance gives a metric on the space of persistent homology.

Definition

Let D_1, D_2 be multiset of points. Bottleneck distance is defined as

$$d_B(D_1, D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.



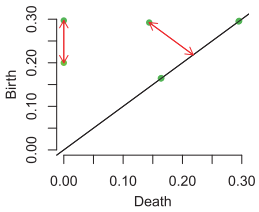
Bottleneck distance gives a metric on the space of persistent homology.

Definition

Let D_1, D_2 be multiset of points. Bottleneck distance is defined as

$$d_B(D_1, D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.



$$\sup_{x \in D_1} \|x - \gamma_1(x)\|_\infty = 0.1$$

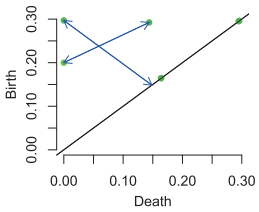
Bottleneck distance gives a metric on the space of persistent homology.

Definition

Let D_1, D_2 be multiset of points. Bottleneck distance is defined as

$$d_B(D_1, D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.



$$\sup_{x \in D_1} \|x - \gamma_2(x)\|_\infty = 0.15$$

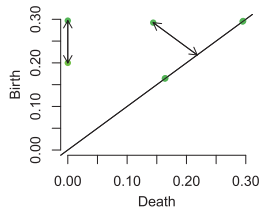
Bottleneck distance gives a metric on the space of persistent homology.

Definition

Let D_1, D_2 be multiset of points. Bottleneck distance is defined as

$$d_B(D_1, D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.



$$\inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty = 0.1$$

Bottleneck distance can be controlled by the corresponding distance on functions: Stability Theorem.

Theorem

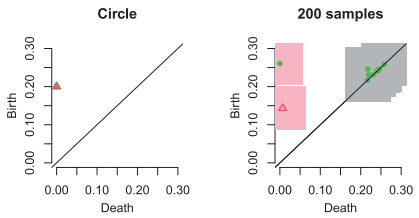
[Edelsbrunner and Harer, 2010][Chazal, de Silva, Glisse, and Oudot, 2012] Let $\mathbb{X}$ be finitely triangulable space and f, $g : \mathbb{X} \to \mathbb{R}$ be two continuous functions. Let $Dgm(f)$ and $Dgm(g)$ be corresponding persistence diagrams. Then

$$d_B(Dgm(f), Dgm(g)) \leq \|f - g\|_\infty.$$

Confidence band for the persistent homology is a random quantity containing the persistent homology with high probability.

Let M be a compact manifold, and $X = \{X_1, \cdots, X_n\}$ be n samples. Let f_M and f_X be corresponding functions whose persistent homology is of interest. Given the significance level $\alpha \in (0, 1)$, $(1 - \alpha)$ confidence band $c_n = c_n(X)$ is a random variable satisfying

$$\mathbb{P}\left(Dgm(f_M) \in \{\mathcal{D} : d_B(\mathcal{D}, Dgm(f_X)) \leq c_n\}\right) \geq 1 - \alpha.$$



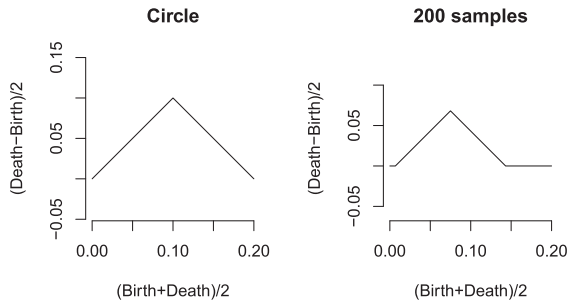
Confidence band for the persistent homology can be obtained by the corresponding confidence band for functions.

From Stability Theorem, $\mathbb{P}(||f_M - f_X|| \leq c_n) \geq 1 - \alpha$ implies

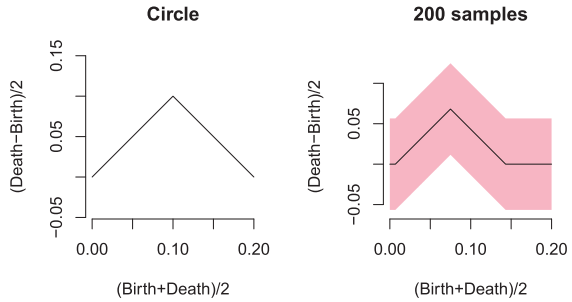
$$\mathbb{P}(d_B(Dgm(f_M), Dgm(f_X)) \leq c_n) \geq \mathbb{P}(||f_M - f_X||_\infty \leq c_n) \geq 1 - \alpha,$$

so the confidence band of corresponding functions f_M can be used for confidene band of persistent homologies $Dgm(f_M)$.

Persistence Landscape of the underlying manifold can be inferred from Persistence Landscape of finite samples.



Confidence band for persistent homology quantifies the randomness of the persistence landscape.

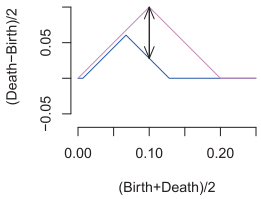


∞-landscape distance gives a metric on the space of persistence landscapes.

Definition

[Bubenik, 2012] Let D_1, D_2 be multiset of points, and λ_1 , λ_2 be corresponding persistence landscapes. ∞-landscape distance is defined as

$$\Lambda_\infty(D_1, D_2) = \|\lambda_1 - \lambda_2\|_\infty.$$



∞-landscape distance can be controlled by the corresponding distance on functions: Stability Theorem.

Theorem

Let $f, g : \mathbb{X} \to \mathbb{R}$ be two functions, and let $Dgm(f)$ and $Dgm(g)$ be corresponding persistent homologies. Then

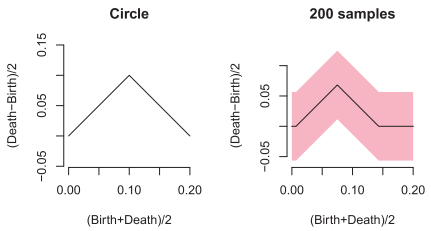
$$\Lambda_\infty(Dgm(f), Dgm(g)) \leq \|f - g\|_\infty.$$

Confidence band for the persistence landscape can be computed using the bootstrap algorithm.

- Let λ_M and λ_X be persistence landscapes of the manifold M and samples X. From Stability Theorem, $\mathbb{P}(||f_M - f_X|| \leq c_n) \geq 1 - \alpha$ implies

  $$\mathbb{P}(\lambda_X(t) - c_n \leq \lambda_M(t) \leq \lambda_X(t) + c_n \, \forall t) \geq \mathbb{P}(||f_M - f_X|| \leq c_n) \geq 1 - \alpha,$$

  so the confidence band of corresponding functions f_M can be used for confidene band of the persistence landscape λ_M.



Confidence band for the persistence landscape can be computed using the bootstrap algorithm.

- Confidence band for the persistence landscape can be also computed using multiplier bootstrap; see [Chazal, Fasy, Lecci, Rinaldo, and Wasserman, 2014].

Statistical Inference for Persistent Homology

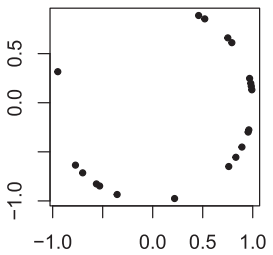
Featurization using Persistent Homology

R Package TDA: Statistical Tools for Topological Data Analysis
Sample on manifolds, Distance Functions, and Density Estimators
Persistent Homology and Persistence Landscape
Statistical Inference on Persistence Homology and Persistence Landscape

R Package TDA provides a function to sample on a circle.

The function circleUnif() generates n sample from the uniform distribution on the circle in $\mathbb{R}^2$ with radius r.

```
circleSample <- circleUnif(n = 20, r = 1)
plot(circleSample, xlab = "", ylab = "", pch = 20)
```



R Package TDA provides distance functions and density functions over a grid.

Suppose $n = 400$ points are generated from the unit circle, and grid of points are generated.

```
X <- circleUnif(n = 400, r = 1)

lim <- c(-1.7, 1.7)
by <- 0.05
margin <- seq(from = lim[1], to = lim[2], by = by)
Grid <- expand.grid(margin, margin)
```

R Package TDA provides KDE function over a grid.

The Gaussian Kernel Density Estimator (KDE) $\hat{p}_h : \mathbb{R}^d \to [0, \infty)$ is defined as

$$\hat{p}_h(y) = \frac{1}{n(\sqrt{2\pi}h)^d} \sum_{i=1}^{n} \exp\left(\frac{-\|y - x_i\|_2^2}{2h^2}\right),$$

where h is a smoothing parameter.
The function kde() computes the KDE function $\hat{p}_h$ on a grid of points.

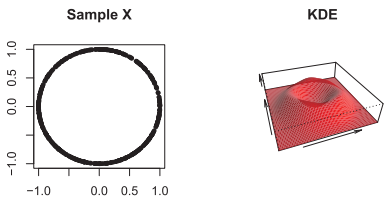
```
h <- 0.3
KDE <- kde(X = X, Grid = Grid, h = h)

par(mfrow = c(1,2))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
persp(x = margin, y = margin,
  z = matrix(KDE, nrow = length(margin), ncol = length(margin)),
  xlab = "", ylab = "", zlab = "", theta = -20, phi = 35, scale = FALSE,
  expand = 3, col = "red", border = NA, ltheta = 50, shade = 0.5,
  main = "KDE")
```


R Package TDA provides KDE function over a grid.

The Gaussian Kernel Density Estimator (KDE) $\hat{p}_h : \mathbb{R}^d \to [0, \infty)$ is defined as

$$\hat{p}_h(y) = \frac{1}{n(\sqrt{2\pi}h)^d} \sum_{i=1}^{n} \exp\left(\frac{-\|y - x_i\|_2^2}{2h^2}\right),$$

where h is a smoothing parameter.
The function kde() computes the KDE function $\hat{p}_h$ on a grid of points.



R Package TDA computes Persistent Homology over a grid.

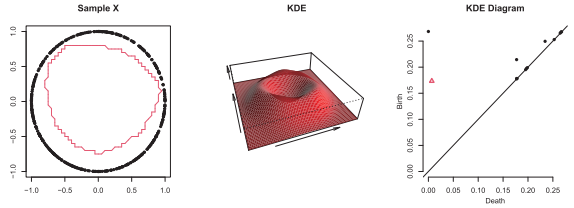
- The function gridDiag() computes the persistence diagram of sublevel (and superlevel) sets of the input function.
  - gridDiag() evaluates the real valued input function over a grid.
  - gridDiag() constructs a filtration of simplices using the values of the input function.
  - gridDiag() computes the persistent homology of the filtration.
- The user can choose to compute persistent homology using either C++ library GUDHI, Dionysus, or PHAT.

R Package TDA computes Persistent Homology over a grid.

```
DiagGrid <- gridDiag(X = X, FUN = kde, lim = c(lim, lim), by = by,
    sublevel = FALSE, library = "Dionysus", location = TRUE,
    printProgress = FALSE, h = h)

par(mfrow = c(1,3))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
one <- which(DiagGrid[["diagram"]][, 1] == 1)
for (i in seq(along = one)) {
  for (j in seq_len(dim(DiagGrid[["cycleLocation"]][[one[i]]])[1])) {
    lines(DiagGrid[["cycleLocation"]][[one[i]]][j, , ], pch = 19, cex = 1,
        col = i + 1)
  }
}
persp(x = margin, y = margin,
  z = matrix(KDE, nrow = length(margin), ncol = length(margin)),
  xlab = "", ylab = "", zlab = "", theta = -20, phi = 35, scale = FALSE,
  expand = 3, col = "red", border = NA, ltheta = 50, shade = 0.9,
  main = "KDE")
plot(x = DiagGrid[["diagram"]], main = "KDE Diagram")
```


R Package TDA computes Persistent Homology over a grid.

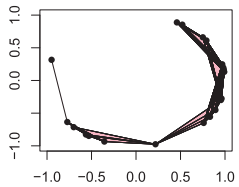
- The function gridDiag() computes the persistent homology of sublevel (and superlevel) sets of the input function.
  - gridDiag() evaluates the real valued input function over a grid.
  - gridDiag() constructs a filtration of simplices using the values of the input function.
  - gridDiag() computes the persistent homology of the filtration.
- The user can choose to compute persistent homology using either GUDHI, Dionysus, or PHAT.



R Package TDA computes Vietoris-Rips Persistent Homology.

- Vietoris-Rips complex consists of simplices whose pairwise distances of vertices are at most $2r$ apart, i.e.

$$\text{Rips}(\mathcal{X}, r) = \{\{x_1, \ldots, x_k\} \subset \mathcal{X} : d(x_i, x_j) < 2r, \text{ for all } 1 \leq i, j \leq k\}.$$



- Rips filtration is formed by Rips complices with gradually increasing r.

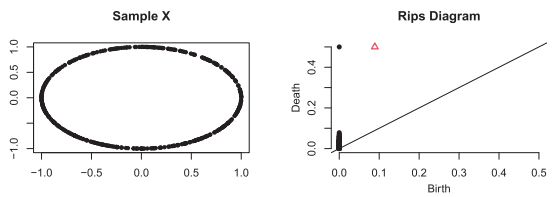
R Package TDA computes Vietoris-Rips Persistent Homology.

- The function ripsDiag() computes the persistence diagram of the Rips filtration built on top of a point cloud.
  - ripsDiag() constructs the Vietoris-Rips filtration using the data points.
  - ripsDiag() computes the persistent homology of the Vietoris-Rips filtration.
- The user can choose to compute persistent homology using either C++ library GUDHI, Dionysus, or PHAT.

```
DiagRips <- ripsDiag(X = X, maxdimension = 1, maxscale = 0.5,
    library = c("GUDHI", "Dionysus"), location = TRUE)

par(mfrow = c(1,2))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
plot(x = DiagRips[["diagram"]], main = "Rips Diagram")
```


R Package TDA computes Vietoris-Rips Persistent Homology.

- The function ripsDiag() computes the persistence diagram of the Rips filtration built on top of a point cloud.
  - ripsDiag() constructs the Vietoris-Rips filtration using the data points.
  - ripsDiag() computes the persistent homology of the Vietoris-Rips filtration.
- The user can choose to compute persistent homology using either C++ library GUDHI, Dionysus, or PHAT.



R Package TDA computes Persistence Landscape.

- Let Λ_p be created by tenting each point $p = (x, y) = \left(\frac{b+d}{2}, \frac{d-b}{2}\right)$ representing a birth-death pair (b, d) in the persistence diagram D.
- The persistence landscape of D is the collection of functions

$$\lambda_k(t) = k \max_p \Lambda_p(t), \quad t \in [0, T], k \in \mathbb{N},$$

  where k max is the kth largest value in the set.
- The function landscape() evaluates the persistence landscape function $\lambda_k(t)$.

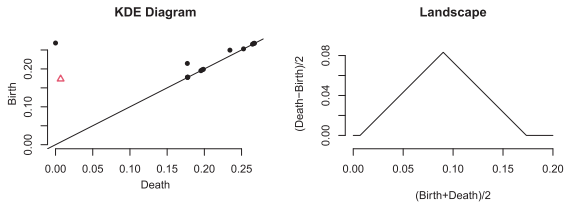
```
tseq <- seq(0, 0.2, length = 1000)
Land <- landscape(DiagGrid[["diagram"]], dimension = 1, KK = 1, tseq = tseq)

par(mfrow = c(1,2))
plot(x = DiagGrid[["diagram"]], main = "KDE Diagram")
plot(tseq, Land, type = "l", xlab = "(Birth+Death)/2",
     ylab = "(Death-Birth)/2", asp = 1, axes = FALSE, main = "Landscape")
axis(1); axis(2)
```


R Package TDA computes Persistence Landscape.

- Let Λ_p be created by tenting each point $p = (x, y) = \left(\frac{b+d}{2}, \frac{d-b}{2}\right)$ representing a birth-death pair (b, d) in the persistence diagram D.
- The persistence landscape of D is the collection of functions

$$\lambda_k(t) = k \max_p \Lambda_p(t), \quad t \in [0, T], k \in \mathbb{N},$$

  where k max is the kth largest value in the set.
- The function landscape() evaluates the persistence landscape function $\lambda_k(t)$.



R Package TDA computes the bootstrap confidence band for a function.

The function bootstrapBand() computes $(1-\alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{\rho}_h]$.

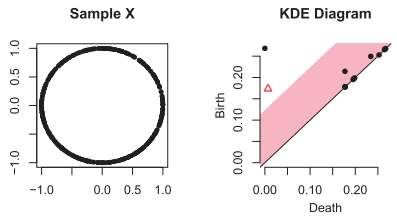
```
bandKDE <- bootstrapBand(X = X, FUN = kde, Grid = Grid, B = 20,
    parallel = FALSE, alpha = 0.1, h = h)
print(bandKDE[["width"]])
```

```
##        90%
## 0.06189347
```


R Package TDA computes the bootstrap confidence band for the persistent homology.

The $(1-\alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{\rho}_h]$ is used as the confidence band for the persistent homology.

```
par(mfrow = c(1,2))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
plot(x = DiagGrid[["diagram"]], band = 2 * bandKDE[["width"]],
     main = "KDE Diagram")
```



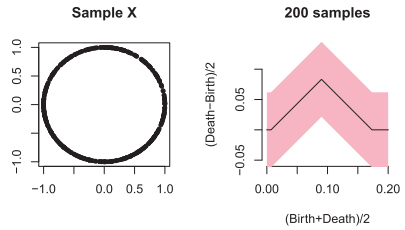
R Package TDA computes the bootstrap confidence band for the persistence landscape.

The $(1-\alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{\rho}_h]$ is used as the confidence band for the persistence landscape.

```
par(mfrow = c(1,2))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
plot(tseq, Land, type = "l", xlab = "(Birth+Death)/2",
     ylab = "(Death-Birth)/2", asp = 1, axes = FALSE, main = "200 samples")
axis(1); axis(2)
polygon(c(tseq, rev(tseq)), c(Land - bandKDE[["width"]],
        rev(Land + bandKDE[["width"]])), col = "pink", lwd = 1.5,
        border = NA)
lines(tseq, Land)
```

R Package TDA computes the bootstrap confidence band for the persistence landscape.

The $(1-\alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{p}_h]$ is used as the confidence band for the persistence landscape.



Persistent Vector Bundles and Stiefel–Whitney Classes in Data Analysis

Dongwoo Gang
Department of Mathematical Sciences, Seoul National University, Korea

In this talk, we introduce a new method to estimate Stiefel–Whitney classes—topological invariants that detect features like orientability and embedding obstructions—directly from point cloud data. We first extend classical vector bundle theory to persistent vector bundles in the setting of topological data analysis. By applying cohomology operations to persistent cohomology, we compute these classes in a persistent setting. We demonstrate the method with applications in image analysis, molecular conformation, and high-dimensional data.

References.

[1] Marco Contessoto et al. "Persistent cup-length". In: *arXiv preprint arXiv:2107.01553* (2021).
[2] Dongwoo Gang. "Persistent Stiefel-Whitney Classes of Tangent Bundles". In: *arXiv preprint arXiv:2503.15854* (2025).
[3] Umberto Lupo, Anibal M Medina-Mardones, and Guillaume Tauzin. "Persistence steenrod modules". In: *Journal of Applied and Computational Topology* 6.4 (2022), pp. 475–502.
[4] John Willard Milnor and James D Stasheff. *Characteristic classes*. 76. Princeton university press, 1974.

Stiefel–Whitney Classes of Tangent Bundles in Data Analysis

Gang Dongwoo

Department of Mathematical Sciences, Seoul National University

August 8, 2025

Table of Contents

Invariants of dataset

- Suppose a point cloud $X \subset \mathbb{R}^D$ lies in smooth closed d-dimensional manifold $M \subset \mathbb{R}^D$.

  One of the major goals in topological data analysis(TDA) is to extract the **geometric/topological invariants** of the underlying manifold M from the sampled dataset X.

  Common examples include (persistent) homology and (filtered) Euler characteristic.

Tangent bundle of the manifold

- One important structure of a smooth manifold M is its **tangent bundle** TM, which is a union of local linear approximations(called *tangent spaces*) across all points on the manifold.

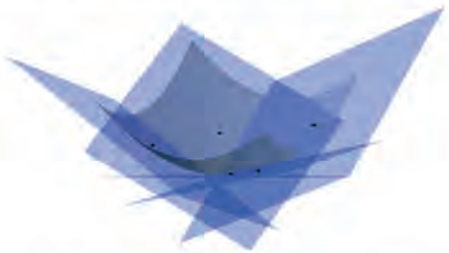


Figure: Some tangent planes (blue) on the manifold M (gray). Points of tangency are marked with bullets.

Orientability

- A smooth manifold M is called **orientable** if it admits an atlas where all transition functions have the positive determinant.

  For example, S^2 is orientable, whereas $\mathbb{RP}^2$ is not.

  The **orientability** of a smooth manifold is a property of its **tangent bundle**, indicating whether the structure group can be reduced to the special orthogonal group $SO(n)$.

  Despite of it, the top-dimensional homology class of M determines its orientability.

Torus and Klein bottle

Figure: Torus and Klein bottle projected into $\mathbb{R}^3$.

Torus and Klein bottle

- The $\mathbb{Z}/2$ (co)homology group, commonly used for the persistent cohomology, cannot distinguish them.

  However, they are not homeomorphic, as the torus is orientable while the Klein bottle is not. Therefore, the tangent bundle for a smooth manifold provides more refined invariant of the base space.

Space	$H^0(-;\mathbb{Z}/2)$	$H^1(-;\mathbb{Z}/2)$	$H^2(-;\mathbb{Z}/2)$
T^2	$\mathbb{Z}/2$	$\mathbb{Z}/2\oplus\mathbb{Z}/2$	$\mathbb{Z}/2$
K	$\mathbb{Z}/2$	$\mathbb{Z}/2\oplus\mathbb{Z}/2$	$\mathbb{Z}/2$

Table: Cohomology groups $H^\bullet(X;\mathbb{Z}/2)$ over $\mathbb{Z}/2$ for the torus T^2 and the Klein bottle K.

Axiomatic Definition of Stiefel–Whitney Class

Let ξ be a real vector bundle of rank n over a paracompact base space B. For non-negative integer i, the **i-th Stiefel–Whitney class** $w_i(\xi) \in H^i(B;\mathbb{Z}/2)$ is the unique element of $H^i(B;\mathbb{Z}/2)$ satisfying the following axioms:

- **Normalization:** For the Möbius line bundle γ_1^1 over $\mathbb{RP}^1$, $w_1(\gamma_1^1) \neq 0$.

  Naturality: If the map between base spaces $f : B(\xi) \to B(\eta)$ is covered by a bundle map $\xi \to \eta$, then $w_i(\xi) = f^*(w_i(\eta))$.

  Whitney Product Formula: For bundles ξ and η over the same base,

$$w_k(\xi \oplus \eta) = \sum_{i=0}^{k} w_i(\xi) \smile w_{k-i}(\eta),$$

  Rank: $w_0(\xi) = 1$, and for all $i > n$, $w_i(\xi) = 0$.

Stiefel–Whitney Class

- Intuitively, the Stiefel-Whitney classes of a vector bundle measures how much twisted the bundle is. If the vector bundle ξ is trivial *iff* $w_i(\xi) = 0$ for every non-negative integer i.

Figure: Comparison of the trivial line bundle TS^1 (left) and the Möbius line bundle γ_1^1 (right) over the same base S^1.

From the axioms, we find that $w_1(TS^1) = 0$ but $w_1(\gamma_1^1) \neq 0$.

A smooth manifold M is orientable *iff* $w_1(TM) = 0$. Therefore, the circle S^1 is orientable.

Naturality

- Since the SW class satisfies the naturality axiom, it is reasonable to expect a persistent version of this concept.

  To formalize such type of theory, one needs the notion of a **persistent vector bundle**(or a **vector bundle filtration**), modeled as a the functor from $(\mathbb{R}, \leqslant)^{\mathrm{op}}$ to the category of rank-d real vector bundles. Due to time constraints, we will not discuss about the persistent vector bundle in detail here.

Cohomology operations

- For a paracompact space X, two important cohomology operations on $H^\bullet(X, \mathbb{Z}/2)$ are the **cup product** and **Steenrod Squares**. Under Poincaré duality over $\mathbb{Z}/2$, for two cocycles α and β in $H^\bullet(X; \mathbb{Z}/2)$, their cup product $\alpha \smile \beta$ represents their **transversal intersection**, whereas the i-th Steenrod Square $\mathrm{Sq}^i(\alpha)$ of α captures the ($\deg \alpha - i$ perverse) **self-intersection** of the cocycle α.

Figure: The torus (gray) with its 1-cocycles (blue and red).

Both operations are *functorial* and can be extended to barcode-level operations in persistent cohomology over $\mathbb{Z}/2$.

Wu Formula

- Let M be a smooth, closed n-dimensional manifold. The **i-th Wu class** $v_i \in H^i(M; \mathbb{Z}/2)$ is the unique element satisfying:

$$\mathrm{Sq}^i(x) = v_i \smile x \quad \text{for all } x \in H^{n-i}(M; \mathbb{Z}/2).$$

  Then, the k-th SW class $w_k(TM)$ is given by the **Wu formula**:

$$w_k(TM) = \sum_{i+j=k} \mathrm{Sq}^i(v_j).$$

  In other words, the SW classes of the tangent bundle of the smooth manifold M are completely determined by the cohomology operations of M, which are purely topological in nature.

Persistent Wu formula

Algorithm (Persistent Wu formula)

*Let $\mathbb{X} : ([s,t], \leqslant) \to$ **Top** be a filtered space that is homotopy equivalent to a smooth, closed n-dimensional manifold M. Then for each $0 \leqslant i \leqslant n$, there is a unique persistent cohomology class $v_i \in H^i(\mathbb{X}; \mathbb{Z}/2)$ such that:*

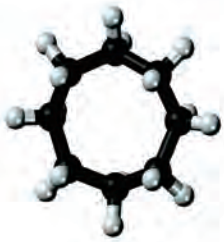
$$\mathrm{Sq}^i(x) = v_i \smile x \quad \text{for all } x \in H^{n-i}(\mathbb{X}; \mathbb{Z}/2).$$

Moreover, the persistent cohomology class:

$$w_k = \sum_{i+j=k} \mathrm{Sq}^i(v_j) \in H^k(\mathbb{X}; \mathbb{Z}/2)$$

corresponds to $w_k(TM)$ via the homotopy equivalences between $\mathbb{X}$ and M.

Application : Conformation space of C_8H_{16}



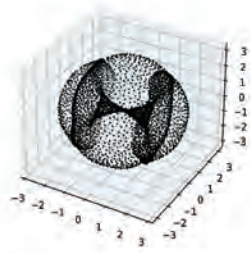


Figure: Molecular structure of cyclooctane (C_8H_{16}) and a point cloud sampled from its conformation space, projected into $\mathbb{R}^3$.

Application : Conformation space of C_8H_{16}

- After suitable preprocessing, we observe that the conformation space of C_8H_{16} forms a singular manifold $A \coprod_{S^1 \sqcup S^1} B$, modeled as the adjunction(pushout) of two closed surfaces A and B intersecting along two disjoint circles $S^1 \sqcup S^1$.

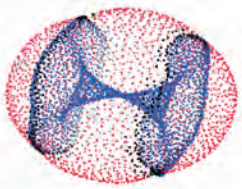


Figure: Point cloud X sampled from the singular manifold, where A (red) and B (blue) intersect along two circles (black).

Application : Conformation space of C_8H_{16}

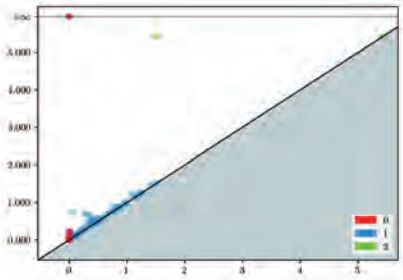


Figure: Persistence diagram of the point cloud from the surface A over $\mathbb{Z}/2$. The diagram suggests that surface A is homeomorphic to a 2-sphere S^2.

Application : Conformation space of C_8H_{16}

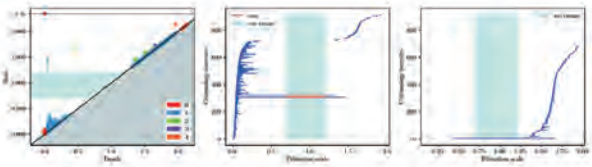


Figure: persistence diagram of point cloud the from surface B over $\mathbb{Z}/2$, along the cohomology barcodes together with the estimated first and second SW classes of the tangent bundle TB. The observed topological invariants match those of the Klein bottle, both in terms of cohomology over $\mathbb{Z}/2$ and the Stiefel–Whitney classes of its tangent bundle.

References

- Gang, D. (2025). Persistent Stiefel-Whitney Classes of Tangent Bundles. arXiv preprint arXiv:2503.15854.
- Milnor, J. W., Stasheff, J. D. (1974). Characteristic classes (No. 76). Princeton university press.
- Lupo, U., Medina-Mardones, A. M., Tauzin, G. (2022). Persistence steenrod modules. Journal of Applied and Computational Topology, 6(4), 475-502.
- Contessoto, M., Mémoli, F., Stefanou, A., Zhou, L. (2021). Persistent cup-length. arXiv preprint arXiv:2107.01553.


Thank you.
Gang Dongwoo (Department of Mathematica
Stiefel-Whitney Classes of Tangent Bundles i
August 8, 2025
19 / 19

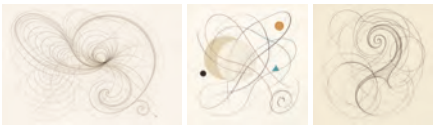
Geometric properties of curves in ensemble forecasting

Sebastian Elias Graiff Zurita

Kyoto University Institute for Advanced Study, Kyoto University, Japan

In this talk, we analyze ensemble forecasting trajectories from a geometric point of view. We focus on the oriented turning angles to cluster and distinguish different weather scenarios. Ensemble forecasting is a method used in weather prediction, which consists of running multiple forecast simulations, each with slightly varied initial conditions or model parameters, to capture the inherent uncertainty in weather forecasting. Collectively, these outputs map the range and likelihood of future states, and it is crucial to identify and label them to take the proper preventive actions in each situation. We quantify the shape of the trajectory with the Frenet frame, which is a coordinate system attached to a moving point along a curve. In two dimensions, the curvature at each point of the curve defines the frame; in three dimensions, the torsion is additionally included; and analogous quantities extend to higher dimensions. As a first step in this research, we emphasize the oriented turning angle (the cumulative signed change of direction) as a feature for grouping ensemble forecasting data. We apply this to heavy-rain datasets and show that the turning angle helps distinguish different meteorological scenarios. We think that studying these techniques further will improve interpretation of ensemble information and uncertainty assessment.

Geometric Properties of Curves in Ensemble Forecasting



Sebastián Elías Graiff Zurita

Geometry of curves

Let us start with the following concepts in 2D, 3D, and general dimension:

- Smooth curves
  - Arc-length parameterization
  - Frenet frame
  - Orientation
  - Turning angle
- Discrete analogues

Smooth curves (2D)

The principal object of study is the parametric **curve**:

$$\gamma(t) : [t_0, t_1] \to \mathbb{R}^2$$

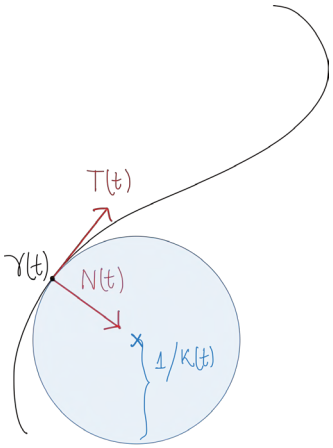
For convenience, we reparameterize it by the **arc-length**

$$s(t) = \int_{t_0}^{t} \|\gamma'(\tau)\| \, \mathrm{d}\tau$$

giving the **arc-length parameterization** of the curve:

$$\gamma(s) : [0, L] \to \mathbb{R}^2$$

$$\|\gamma'(s)\| = 1, \forall s$$



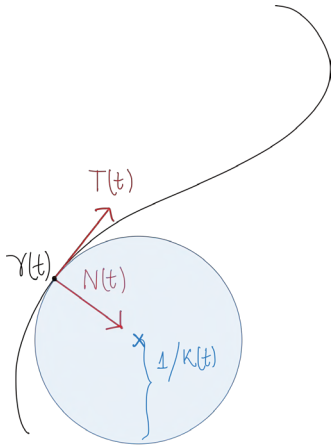
Smooth curves (2D)

Given a arc-length parameterized curve, consider:

$$\gamma(s) : [0, L] \to \mathbb{R}^2 \qquad \|\gamma'(s)\| = 1, \forall s$$

$$T(s) := \gamma'(s) \qquad \text{(Tangent)}$$
$$N(s) := \mathrm{R}_{\pi/2} T \qquad \text{(Normal)}$$

$$\kappa(s) := \langle T'(s), N(s) \rangle \qquad \text{(Curvature)}$$



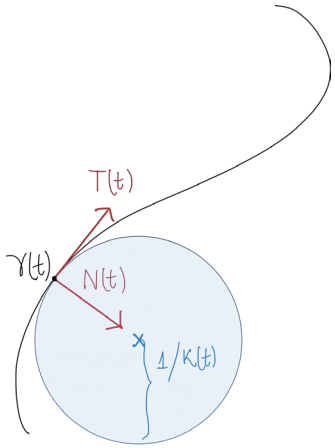
Frenet frame (2D)

The **Frenet frame** describes the geometric properties of a particle moving along a differentiable curve.

$$F(s) = [T(s), N(s)] \in \mathbb{R}^{2\times 2} \qquad \text{(Frenet frame)}$$

The Frenet frame satisfies the **Frenet formula:**

$$F'(s) = F(s)K(s)$$
$$K(s) := \begin{pmatrix} 0 & -\kappa(s) \\ \kappa(s) & 0 \end{pmatrix}$$



Smooth curves (3D)

Similarly, we consider an arc-length parameterized curve,

$$\gamma(s) : [0, L] \to \mathbb{R}^3 \qquad \|\gamma'(s)\| = 1, \forall s$$

$$T(s) := \gamma'(s) \qquad \text{(Tangent)}$$
$$N(s) := \frac{\gamma''(s)}{\|\gamma''(s)\|} \qquad \text{(Normal)}$$
$$B(s) := T(s) \times N(s) \qquad \text{(Binormal)}$$

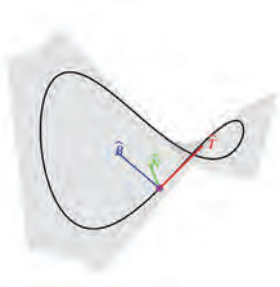


Image: Commons.Wikimedia/Bevin Maultsby

Smooth curves (3D)

$$T(s) := \gamma'(s) \qquad \text{(Tangent)}$$

$$N(s) := \frac{\gamma''(s)}{\|\gamma''(s)\|} \qquad \text{(Normal)}$$

$$B(s) := T(s) \times N(s) \qquad \text{(Binormal)}$$

Comparing with a two dimensional space, now we have an extra dimension where the curve can *curl*, giving raise to the torsion:

$$\kappa(s) := \langle T'(s), N(s) \rangle \qquad \text{(Curvature)}$$

$$\tau(s) := \langle N'(s), B(s) \rangle \qquad \text{(Torsion)}$$

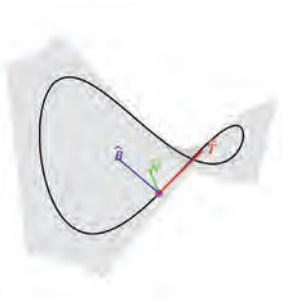


Image: Commons.Wikimedia/Bevin Maultsby

Frenet frame (3D)

In three dimensions, the Frenet frame is given by the normal, tangent and binormal vectors,

$$F(s) = [T(s), N(s), B(s)] \in \mathbb{R}^{3\times 3}$$

Which satisfies the **Frenet-Serret formulas:**

$$F'(s) = F(s)K(s)$$

$$K(s) := \begin{pmatrix} 0 & -\kappa(s) & 0 \\ \kappa(s) & 0 & -\tau(s) \\ 0 & \tau(s) & 0 \end{pmatrix}$$

Smooth curves (ND)

All the previous concepts are generalized for any dimension. The generalized tangents, are obtained recursively.

$$\gamma(s) : [0, L] \to \mathbb{R}^n \qquad \|\gamma'(s)\| = 1, \forall s$$

$$u_1(s) = \gamma'(s), \qquad E_1(s) := \frac{u_1(s)}{\|u_1(s)\|},$$

$$u_2(s) = \gamma''(s) - \langle E_1(s), \gamma''(s)\rangle E_1(s), \qquad E_2(s) := \frac{u_2(s)}{\|u_2(s)\|},$$

$$\vdots$$

$$u_n(s) = \gamma^{(n)}(s) - \sum_{k=1}^{n-1} \left\langle E_k(s), \gamma^{(n)}(s) \right\rangle E_k(s), \qquad E_n(s) := \frac{u_n(s)}{\|u_n(s)\|}.$$

Smooth curves (ND)

All the previous concepts are generalized for any dimension. The generalized tangents, are obtained recursively.

$\gamma(s) : [0, L] \to \mathbb{R}^n \qquad \|\gamma'(s)\| = 1, \forall s$

$$[E_1(s), \ldots, E_n(s)] = \text{Gram–Schmidt}[\gamma'(s), \gamma''(s), \ldots, \gamma^{(n)}(s)]$$

Frenet frame (ND)

The Frenet frame is generalized by a system of orthonormal vectors that aligns with higher derivatives of a curve.

$F(s) = [E_1(s), \ldots, E_n(s)] \in \mathbb{R}^{n \times n}$

$$F'(s) = F(s)K(s)$$

$$K(s) := \begin{pmatrix} 0 & -\kappa_1(s) & 0 & 0 \\ \kappa_1(s) & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & -\kappa_{n-1}(s) \\ 0 & 0 & \kappa_{n-1}(s) & 0 \end{pmatrix}$$

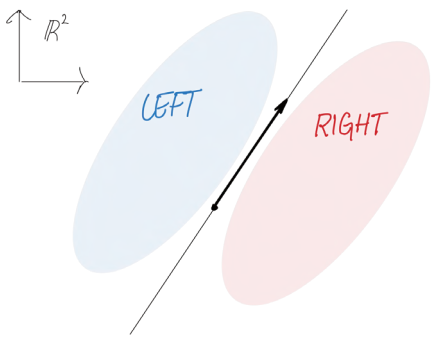
$$\kappa_k(s) := \left\langle E_k'(s), E_{k+1}(s) \right\rangle, \quad k = 1, 2, \ldots, n-1$$

Orientation (ND)

A Frame, seeing it as an ordered basis, has an *orientation depending on the sign of the determinant* being positive or negative.

The orientation distinguishes between two possible “handedness” of coordinate systems:

- In 2D, this corresponds to **clockwise** vs **counter-clockwise.**
- In 3D, **right-handed** vs **left-handed**.

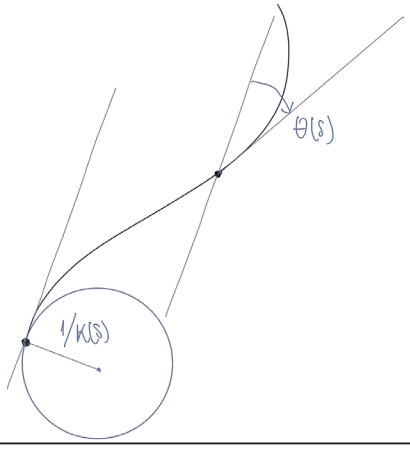


Turning angle (ND)

In two dimensions, the turning angle has a natural sign depending on the orientation of the Frenet frame.

Let us consider its extension to high dimensions by simply taking

$$\psi(s) := \int_0^s \operatorname{sign}\det(F(\bar{s}))\left\langle E_1'(\bar{s}), E_2(\bar{s})\right\rangle \mathrm{d}\bar{s}$$



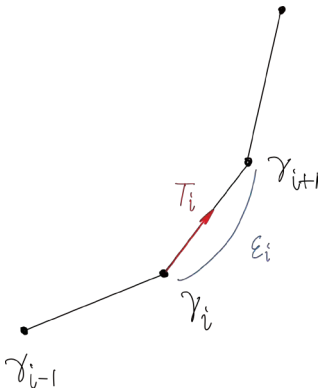
Discrete analogues

A discrete curve is simply a sequence of points,

$$\gamma_i \in \mathbb{R}^n$$

were we consider

$$v_i := \Delta\gamma_i \quad \text{(velocity)}$$
$$\varepsilon_i := \|\Delta\gamma_i\| \quad \text{(step size)}$$
$$T_i := \frac{v_i}{\|v_i\|} \quad \text{(tangent)}$$



Discrete analogues

The k-finite difference,

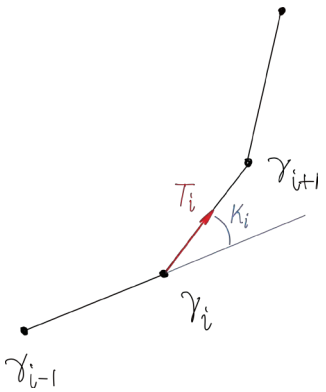
$$\begin{cases} \Delta^k\gamma_i := \Delta(\Delta^{k-1}\gamma_i) \\ \Delta\gamma_i := \gamma_{i+1} - \gamma_i \end{cases} \quad \text{(k-difference)}$$

is used to define the **discrete Frenet frame**,

$$F_i := \text{Gram–Schmidt}[\Delta^1\gamma_i, \Delta^2\gamma_i, \ldots, \Delta^n\gamma_i]$$

and the ***signed*** **deflection** and **turning angle:**

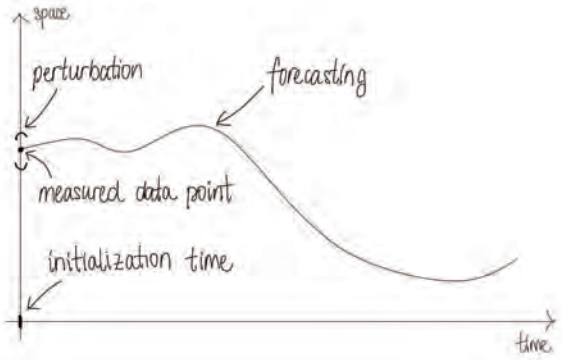
$$K_i := \operatorname{sign}\det(F_i)\arccos(\langle T_i, T_{i-1}\rangle) \quad \text{(deflection)}$$
$$\psi_i := \sum_{j\le i} K_j \quad \text{(turning angle)}$$



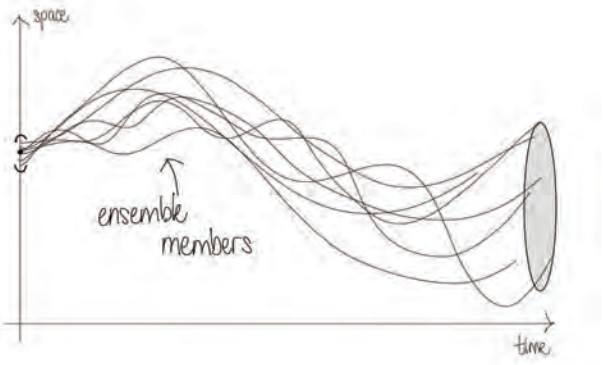
Ensemble Forecasting

Ensemble forecasting is a method used in weather prediction that involves running multiple forecast simulations, each with slightly varied initial conditions or model parameters, to capture the uncertainty of the atmosphere and improve forecast reliability.

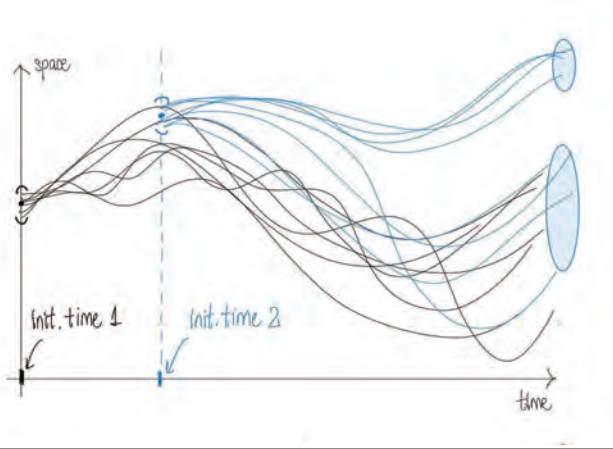
Ensemble Forecasting



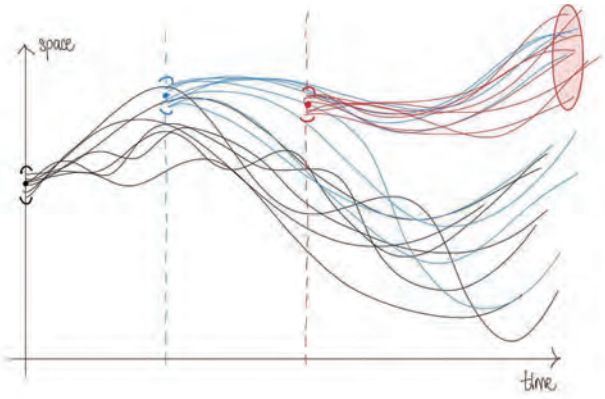
Ensemble Forecasting



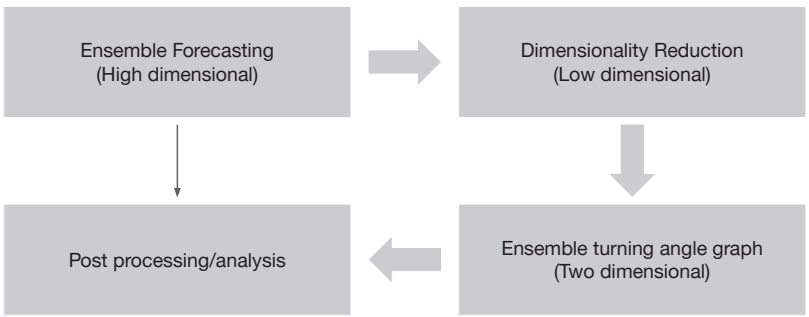
Ensemble Forecasting



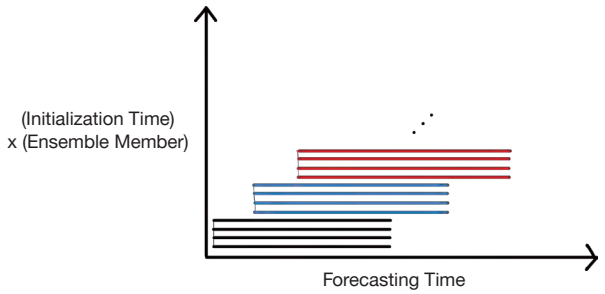
Ensemble Forecasting



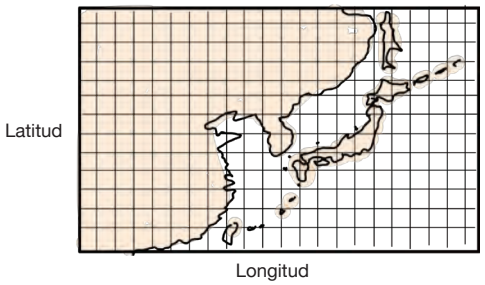
Data analysis workflow



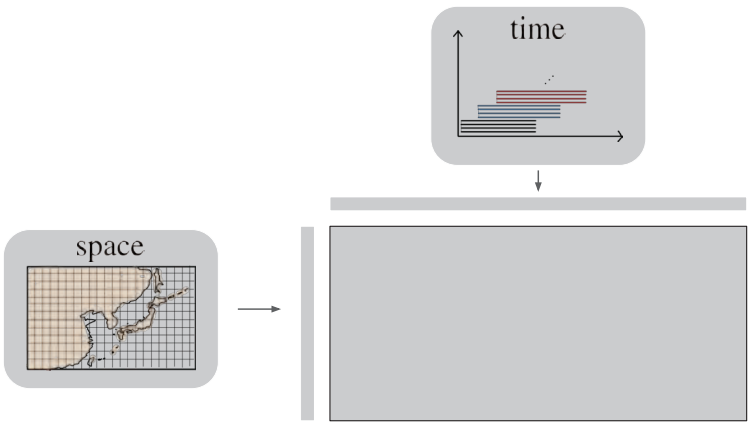
Input Data (time axis)



Input Data (space axis)



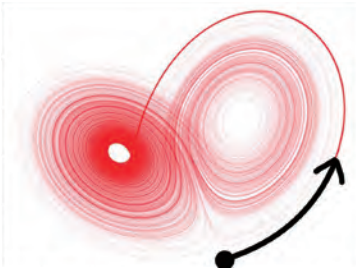
Input Data (full matrix)



Example: Lorenz ‘63 model

The Lorenz system (1963) is a classical example of a system that can exhibit chaotic behavior.

$$\begin{cases} x' = \sigma(y - x), \\ y' = x(\rho - z) - y, \\ z' = xy - \beta z. \end{cases}$$



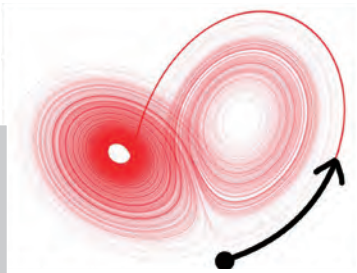
Example: Lorenz ‘63 model

$$\begin{cases} x' = \sigma(y - x), \\ y' = x(\rho - z) - y, \\ z' = xy - \beta z. \end{cases}$$

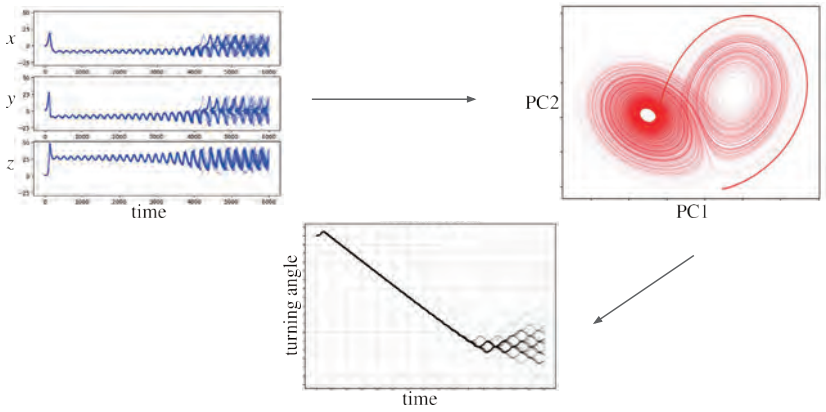
$$\left(\rho = 28, \sigma = 10, \beta = \frac{8}{3}\right)$$

$$\{X^j : [0, T] \to \mathbb{R}^3\}_{j=1}^{50} \quad \text{(50 ensembles)}$$

$$X^j(0) \sim \mathcal{N}(X_0, r) \qquad \begin{cases} X_0 = (0, 1, 1.05) \\ r = 0.15 \end{cases}$$



Example: Lorenz ‘63 model



Example: Lorenz ‘96 model

The Lorenz ‘96 model is a dynamical system defined in N dimensions. It is commonly used as a model problem in geosciences:

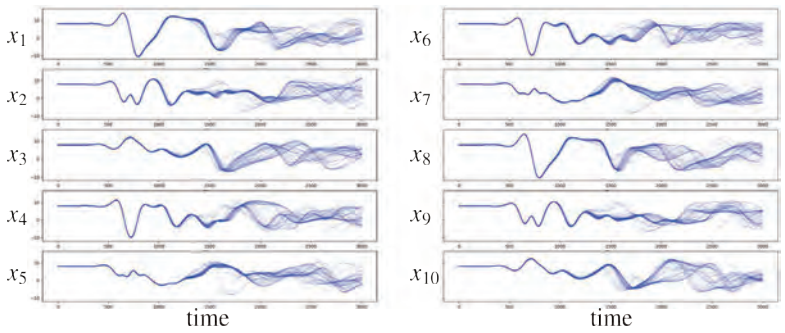
$$x_i' = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \quad i \in (1, 2, \ldots, N)_{\text{cyclic}}$$

$$(N = 10, F = 8)$$

$$\{x^j : [0, T] \to \mathbb{R}^{10}\}_{j=1}^{50} \quad (50 \text{ ensembles})$$

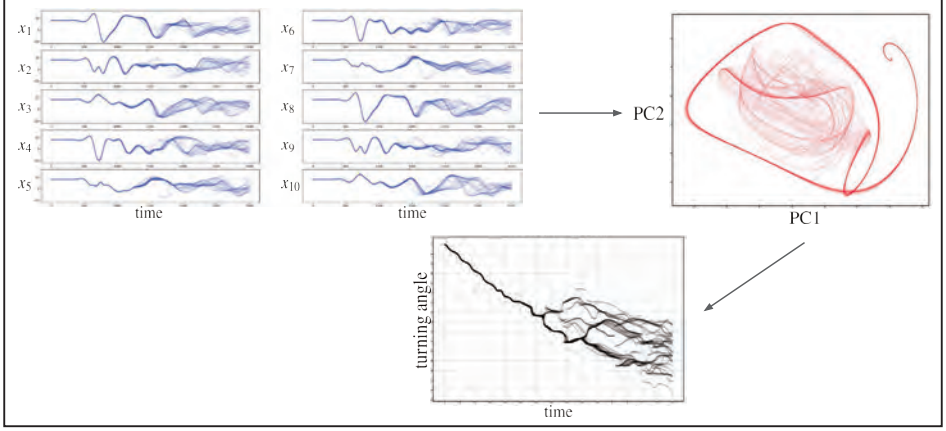
$$x^j(0) \sim \mathcal{N}(x_0, r) \qquad \begin{cases} x_0 &= (8.01, 8, 8, \ldots, 8) \\ r &= 5 \times 10^{-5} \end{cases}$$

Example: Lorenz ‘96 model



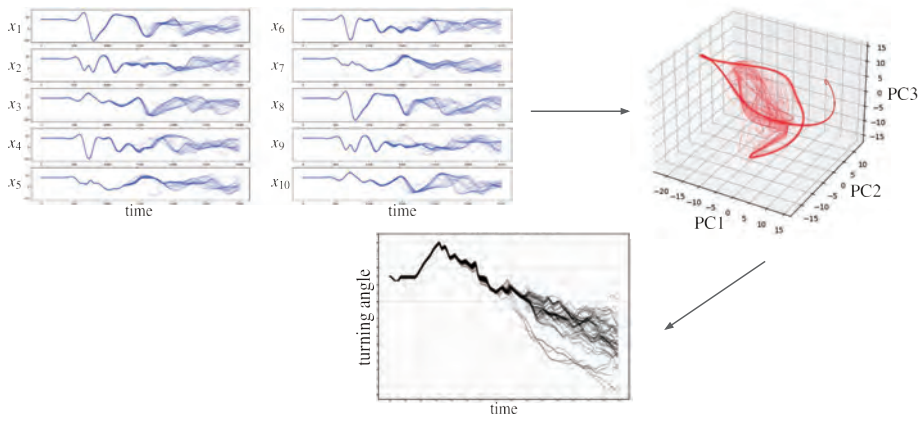
Example: Lorenz ‘96 model

(**2D** PCA projection)



Example: Lorenz ‘96 model

(**3D** PCA projection)



TIGGE Dataset (2018)

The **THORPEX Interactive Grand Global Ensemble** (TIGGE) is an implementation of ensemble forecasting for global weather forecasting established in 2006 by the World Meteorological Organization.

Initialization Time:
Every 6 hours.
Forecasting Time:
Every 6 hours, for up to 2 weeks.
Ensemble Members:
20 perturbed, 1 control point.
Data:
Geopotential height at 500 hPa.

TIGGE Dataset (2018)

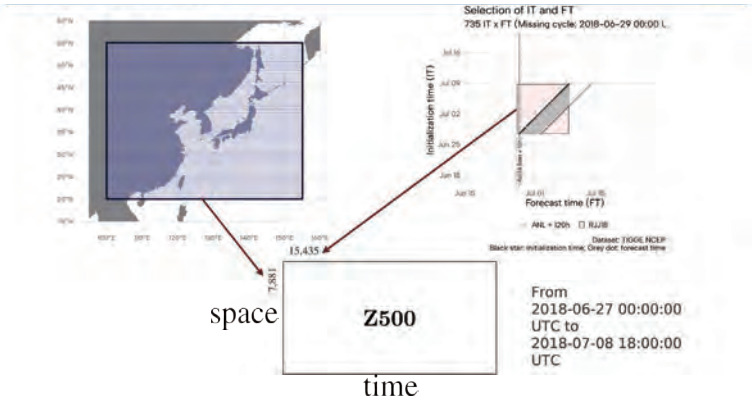
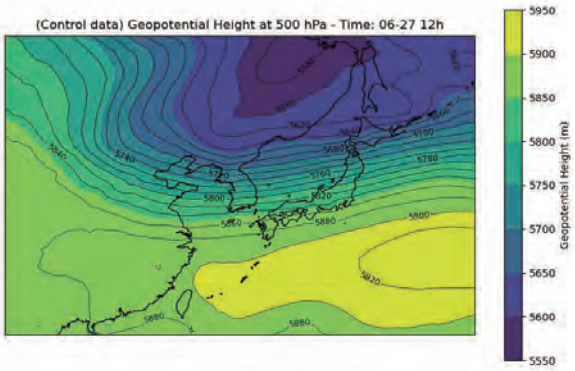


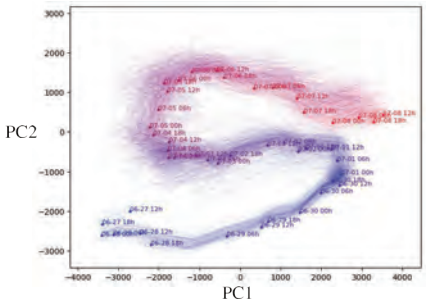
Image: Pascal OETTLI, Center for Environmental Remote Sensing, Chiba University

TIGGE Dataset (2018)

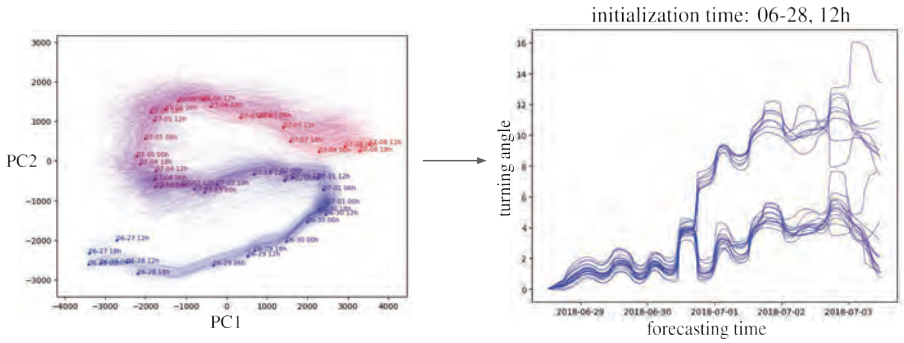


TIGGE Dataset (2018)

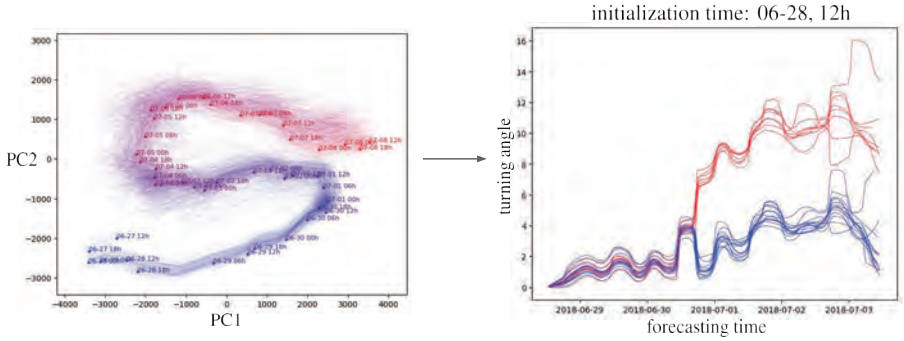
(**2D** PCA projection)



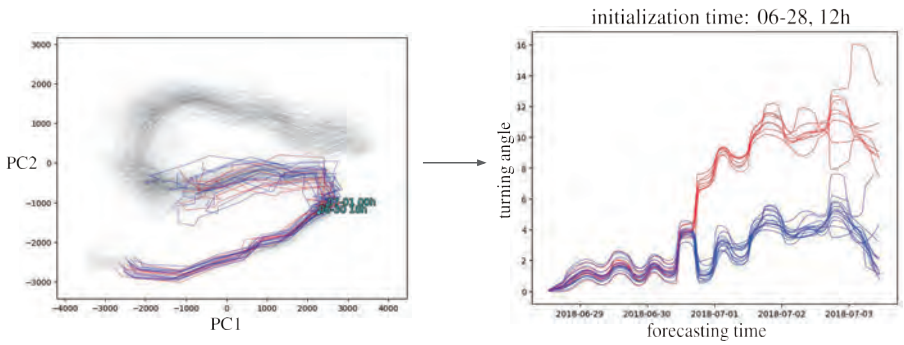
TIGGE Dataset (2018)



TIGGE Dataset (2018)

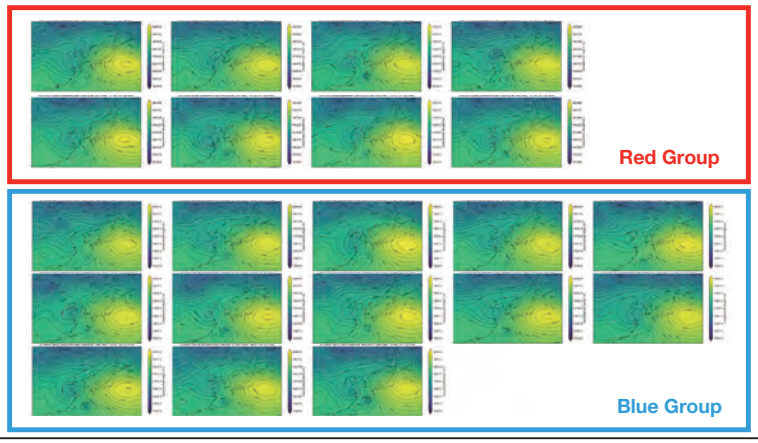


TIGGE Dataset (2018)



TIGGE Dataset (2018)

initialization time: 06-28, 12h



TIGGE Dataset (2018)

(Statistical Analysis)

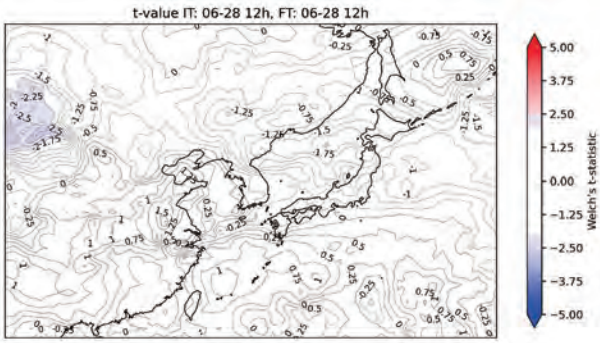
Welch's t-test

$$\text{t-value} = \frac{Z_{\text{Red}} - Z_{\text{Blue}}}{\sqrt{\frac{s^2_{\text{Red}}}{N_{\text{Red}}} + \frac{s^2_{\text{Blue}}}{N_{\text{Blue}}}}}$$

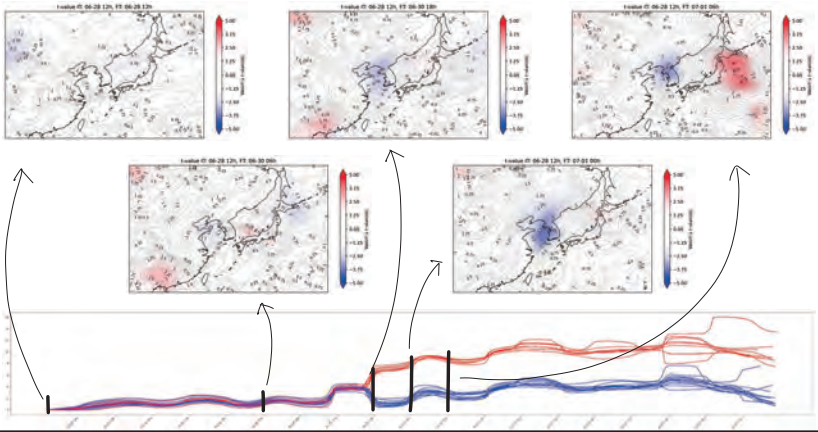
Z_i : group mean

s_i^2 : group variance

N_i : group count



TIGGE Dataset (2018)



Thanks!

2025 Topological Data Analysis and Industrial Mathematics

A Topology and Distribution-Based Method for Pipe Localization in Ground Penetrating Radar Data

MEIYAN KANG
Ph.D. Candidate
miyeon@ajou.ac.kr

SUYOUNG CHOI
Advisor. Professor
schoi@ajou.ac.kr

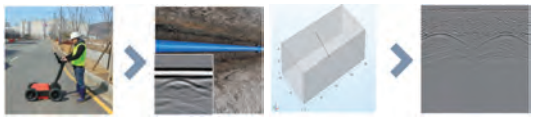
亞洲大学校
AJOU UNIVERSITY

Introduction

Ground Penetrating Radar (GPR) is a widely used non-destructive testing (NDT) technique for subsurface exploration, particularly in infrastructure inspection and maintenance. However, it is often limited structural awareness. In this study, we propose a novel mathematical framework that combines Topological Data Analysis (TDA) with Kernel Density Estimation (KDE). TDA is used to extract robust topological features from B-scan images, and KDE models the distributions of these features across space. By computing distances between the resulting distributions, we can localize buried objects.

Data Description

A typical GPR survey setup is illustrated in Fig. 1(a) illustrates in the field, included here for visual context. While this study does not use real-world data, the same principles are replicated in simulation. We generated 341 synthetic B-scan images using the open-source FDTD simulator gprMax. The simulation domain included a single buried pipe and multiple soil layers. A representative simulation setup and its synthetic B-scan output are shown in Fig. 1(b).



(a) Field data collection (b) gprMax synthetic B-scan generated.

Fig. 1: Field Data Collection and Simulated Data Generation

Toplogical Feature Extraction

We applied topological data analysis (TDA) to extract structural patterns from simulated GPR images as illustrated in Fig.2. H_1 loops are extracted, some sample are shown in Fig. 3, which indicate circular or arc-like structures that potentially correspond to buried pipes.

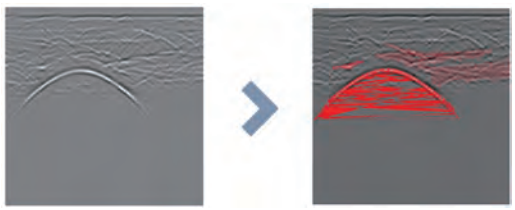


Fig. 2: From raw B-scan to extracted H_1 loops for shape analysis.

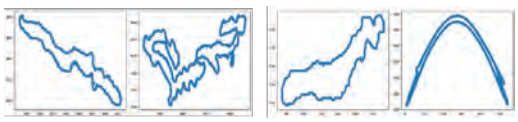


Fig. 3: Detected H_1 features (in red) highlighting arc-like pipe structures.

Distribution-Based Detection

To identify pipe-like patterns, we focus on classifying the detected H_1 loops based on their spatial distributions.
Fig. 4 shows these procedures.

Step 1. Select a representative parabolic shape as a reference.

Step 2. Using Kernel Density Estimation (KDE) get the spatial distribution of each H_1 features.

Step 3. The similarity between distributions is quantified using the Wasserstein distance.

Step 4. Features exhibiting similar KDE profiles are classified as pipe-related, while dissimilar ones are discarded.

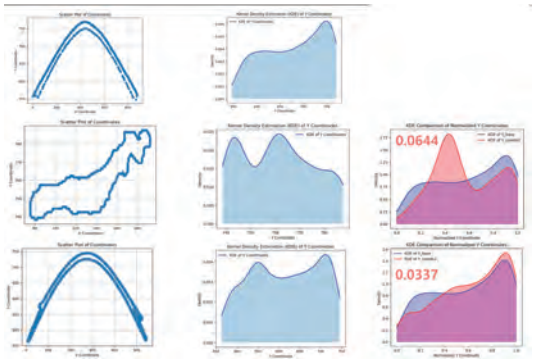


Fig. 4: Classification of KDE-based features using Wasserstein distance.

Results

The framework achieved an F1-score of 0.86, with balanced precision and recall (both 0.86), as summarized in Fig. 5. Although the overall accuracy was 0.76, the high F1-score highlights the model's robustness in handling potential class imbalance. The confusion matrix shows a high number of true positives and relatively few false detections, confirming the method's reliability.

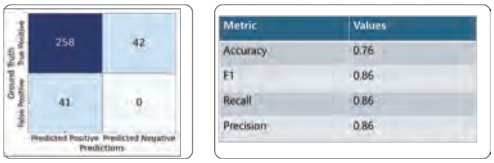


Fig. 5: Evaluation metrics showing F1-score, precision, recall, and accuracy.

Conclusion

This study explores a topology- and distribution-driven approach, which shows initial promise. However, its performance can be further improved, and real-world validation remains necessary.

Support Estimation with Topological guarantee

Hyeongyu Kim
Seoul National University

1. Introduction

A support of a probability distribution P is the minimal closed set whose P-probability is 1. Support estimation is one of the main interests of statistics, machine Learning, and industry due to its wide range of applications like anomaly detection (Devroy and Wise, 1980) and clustering (Cuevas et al., 2000). Support estimation is closely related to density estimation since support of a P whose probability density function is p is often defined as $\overline{\{x : p(x) > 0\}}$. Cuevas and Fraiman (1997) proposed a plug-in method for support estimation using the upper-level set of a kernel density estimator as a support estimator and showed its measure-based metric and Hausdorff metric convergence rates.
Although the plug-in estimator offers useful geometric inference of the support, it does not provide any guarantees regarding accurate inference of the topological structure of the support. One approach to performing topological inference on the support is to construct a set which is homotopy equivalent to the support.
In this poster, I investigate the conditions under which the plug-in estimator (Cuevas and Fraiman, 1997) is homotopy equivalent to the support for both fixed and random data settings.

2. Background

2.1. Notation

- Distance between a set $\mathbb{X}$ and a point y is defined as

$$d(y, \mathbb{X}) = \inf_{x \in \mathbb{X}} \|y - x\|_2$$

where $\|\cdot\|_2$ is usual Euclidean L_2 norm.

- For a subset $\mathbb{X} \subset \mathbb{R}^d$ and $r > 0$, its r-offset $\mathbb{X}^r$ is defined as

$$\mathbb{X}^r = \left\{ y \in \mathbb{R}^d : d(y, \mathbb{X}) < r \right\} = \bigcup_{x \in \mathbb{X}} B(x, r)$$

and 0-offset $\mathbb{X}^0$ be just $\mathbb{X}$ itself.

2.2. Homotopy equivalence

- A **homotopy** between two continuous maps $f_0, f_1 : \mathbb{X} \to \mathbb{Y}$ is a continuous map $F : \mathbb{X} \times [0, 1] \to \mathbb{Y}$ such that for all $x \in \mathbb{X}$, $F(x, 0) = f_0(x)$, and $F(x, 1) = f_1(x)$. Two maps f_0 and f_1 are called **homotopic** if such homotopy F exists, and denote $f_0 \simeq f_1$.

Using homotopy between maps, we define the homotopy equivalence between two topological spaces as follows.

- A continuous map $f : \mathbb{X} \to \mathbb{Y}$ is called a **homotopy equivalence** if there exists a map $g : \mathbb{Y} \to \mathbb{X}$ such that $g \circ f \simeq \mathrm{id}_{\mathbb{X}}$ and $f \circ g \simeq \mathrm{id}_{\mathbb{Y}}$. $\mathbb{X}$ and $\mathbb{Y}$ are said to be **homotopy equivalent** or have the same homotopy type if such homotopy equivalence exists, and denote $\mathbb{X} \simeq \mathbb{Y}$.

It is worth knowing that being homotopy equivalent is an equivalence relation.
A special case of homotopy equivalence called a deformation retract is a widely used deformation map from a space to its subspace.

- A **deformation retract** of a space $\mathbb{X}$ onto a subspace $A \subset \mathbb{X}$ is a continuous map $F : \mathbb{X} \times [0, 1] \to \mathbb{X}$ such that

$$F(x, 0) = x, \ F(x, 1) \in A \ \text{ for } \ \forall x \in \mathbb{X},$$
$$F(a, t) = a \ \text{ for } \ \forall a \in A \ \text{ and } \ \forall t \in [0, 1]$$

The subspace $A \subset \mathbb{X}$ is called a **deformation retract** of $\mathbb{X}$ if such a deformation retract exists. And of course, $A \simeq \mathbb{X}$ if $A \subset \mathbb{X}$ is a deformation retract of $\mathbb{X}$.

2.3. Kernel Density Estimation

Let P be a Borel probability distribution in $\mathbb{R}^d$ with probability density p. A kernel density estimation is a nonparametric density estimation that estimates p using a kernel function $K : \mathbb{R}^d \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}^d} K(x)dx = 1, \ \int_{\mathbb{R}^d} xK(x)dx = 0, \ \text{and } 0 < \int_{\mathbb{R}^d} x^2 K(x)dx < \infty.$$

Here we define a kernel density estimator(KDE) $\hat{p}_h : \mathbb{R}_d \to \mathbb{R}$ as

$$\hat{p_h}(x) := \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

for given data $\{X_1, \cdots, X_n\} \subset \mathbb{R}^d$ and bandwidth $0 < h < \infty$.
An expected value of KDE $E_P[\hat{p}_h(x)] := p_h(x)$ is often called a smoothed density. A smoothed density can be viewed as a convolution of p and $K_h(\cdot) = K(\cdot/h)/h^d$, and it is useful for geometrical inference.

- (Gine and Guillou, 2002)

$$\|\hat{p}_h - p_h\|_\infty = O_P\left(\sqrt{\frac{\log 1/h}{nh^d}}\right)$$

3. Topological Consistency of plug-in Estimator

Cuevas and Fraiman (1997) defines the plug-in based support estimator $\hat{p}_h^{-1}[\lambda_n, \infty)$ where $\{\lambda_n\}_{n \in \mathbb{N}}$ is a decreasing sequence of positive numbers, converging to 0. A main result of this paper is to investigate the conditions when this plug-in based estimator is homotopy equivalent to the true support $\mathrm{supp}(P)$ of P. Directly targeting $\mathrm{supp}(P)$ requires $h \to 0$, however this requires the existence of a density function p, which is often violated due to a manifold assumption. It also incurs the curse of dimensionality in the convergence rate of $\hat{p}_h$. Instead, I would like to fix $h > 0$. Then plug-in estimator will naturally converge to $\mathrm{supp}(p_h)$, and mild geometric conditions on the kernel K and $\mathrm{supp}(P)$ ensure that $\mathrm{supp}(p_h)$ and $\mathrm{supp}(P)$ are homotopy equivalent.

Lemma 1.

Let P be a probability distribution in $\mathbb{R}^d$ and K be the kernel function. If the kernel K and $\mathrm{supp}(P)$ satisfy mild geoemtric conditions, then $\mathrm{supp}(p_h)$ deformation retracts to $\mathrm{supp}(P)$. i.e., $\mathrm{supp}(p_h) \simeq \mathrm{supp}(P)$, and $\mathrm{supp}(p_h) \simeq p_h^{-1}[\xi, \infty)$ for small enough $\xi > 0$.

Now the problem becomes whether the plug-in estimator $\hat{p}_h^{-1}[\lambda_n, \infty)$ constructed from a finite sample is homotopy equivalent to $p_h^{-1}[\xi, \infty)$. I will examine the conditions under which this equivalence holds for both fixed and random data.

1. Luc Devroy and Gary L. Wise. (1980) Detection of abnormal behavior via nonparametric estimation of the support. SIAM J. Appl. Math., 38(3):480-488, 1980.
2. Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. (2000) Estimating the number of clusters. Canadian Journal of Statistics, 28(2):367-382, 2000.
3. Antonio Cuevas and Ricardo Fraiman. (1997). A plug-in approach to support estimation. Ann. Statist., 25(6):2330-2321, 1997.
4. Evarist Gine and Armelle Guillou. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. Annales de l'Institut Henri Poincare (B) Probability and Statistics, 38(6):907-921, 2002
5. Jisu Kim, Jaehyeok Shin, Alessandro Rinaldo, and Larry A. Wasserman. (2019). Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension. Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019

3.1. Fixed Data

Let f and g be two real-valued functions. To see whether the level sets of these functions deformation retract to each other, we first need the level sets to be interleaved.

Lemma 2.

Let $f, g : \mathcal{U} \subset \mathbb{R}^d \to \mathbb{R}$ be two real valued functions. Let $\xi_1, \xi_2, \lambda \in \mathbb{R}$ be satisfying $\xi_1 < \lambda < \xi_2$ and $\sup_{x \in f^{-1}[\xi_1, \xi_2]} |f(x) - g(x)| \leq \min\{\lambda - \xi_1, \xi_2 - \lambda\}$. Then the following holds:

$$f^{-1}[\xi_2, \infty) \subset g^{-1}[\lambda, \infty) \subset f^{-1}[\xi_1, \infty)$$

Whether there are deformation retracts between level sets is a well-known question in Morse Theory and is solved by flows. Let $\mathbb{Y} := f^{-1}[\xi, \infty) \subset g^{-1}[\lambda, \infty) := \mathbb{X}$. When $f = g$, a flow integrating the gradient ∇f would give a deformation retract of $\mathbb{X}$ to $\mathbb{Y}$, where a flow $\psi : \mathbb{D} \subset (\mathbb{X} \backslash \mathbb{Y}) \times [0, \infty) \to \mathbb{X} \backslash \mathbb{Y}$ satisfies

$$\frac{d}{ds}\psi(x, s) = \nabla f(\psi(x, s))$$

Hence we can hope that when ∇f and ∇g are close, we can build a vector field W that is "close" to both ∇f and ∇g to make a deformation retract.

Lemma 3.

Let $f, g : \mathbb{R}^d \to \mathbb{R}$ be two real valued smooth functions. Let $\xi_1 < \xi_2$ be such that f has no critical points in $D := f^{-1}[\xi_1, \xi_2]$. Suppose ∇f and ∇g satisfy

$$\|\nabla f(x) - \nabla g(x)\|_2 \leq \frac{1}{2} \min_{x \in D} \|\nabla f(x)\|_2 \quad \text{for all } \ x \in D.$$

Then there exists a smooth vector field $W : D \to \mathbb{R}^d$ that approximates both $\nabla f(x)$ and $\nabla g(x)$

With Lemma 3, we can finally build a deformation retraction from $\mathbb{X}$ to $\mathbb{Y}$

Lemma 4.

Let $f, g : \mathbb{R}^d \to \mathbb{R}$ be two nonnegative real valued smooth functions. Suppose $\mathbb{X} := f^{-1}[\xi, \infty)$ and $\mathbb{Y} := g^{-1}[\lambda, \infty)$ satisfies that $\mathbb{Y} \subset \mathbb{X}$. Suppose $W : \mathbb{X} \backslash \mathbb{Y} \to \mathbb{R}^d$ is a vector field approximating both $\nabla f(x)$ and $\nabla g(x)$. Then $\mathbb{X}$ deformation retracts to $\mathbb{Y}$. In particular, $\mathbb{X} \simeq \mathbb{Y}$.

Now combining all the above lemmas, we can say the homotopy equivalence between level sets of plug-in estimator and smoothed density.

Theorem 5.

Suppose the kernel function K is smooth. Let $0 < \xi_1 < \lambda_n < \xi_2$ satisfying that p_h has no critical points in $D := p_h^{-1}[\xi_1, \xi_2]$. Now, suppose $\hat{p}_h$ satisfies the followings:

$$\sup_{x \in D} |\hat{p}_h(x) - p_h(x)| \leq \min\{\lambda_n - \xi_1, \xi_2 - \lambda_n\}, \tag{1}$$

$$\sup_{x \in D} \|\nabla \hat{p}_h(x) - \nabla p_h(x)\|_2 \leq \frac{1}{2} \min_{x \in D} \|\nabla p_h(x)\|_2. \tag{2}$$

Then $p_h^{-1}[\xi_1, \infty)$ deformation retracts to $\hat{p}_h^{-1}[\lambda_n, \infty)$, and $\hat{p}_h^{-1}[\lambda_n, \infty)$ deformation retracts to $p_h^{-1}[\xi_2, \infty)$. In particular, $p_h^{-1}[\xi_1, \infty) \simeq \hat{p}_h^{-1}[\lambda_n, \infty) \simeq p_h^{-1}[\xi_2, \infty)$.

The following corollary summarizes the topological consistency of the plug-in estimator under mild conditions.

Corollary 6.

Let $0 < \xi_{1,n} < \lambda_n < \xi_{2,n} < \xi_0, \forall n$, where ξ_0 is minimum nonzero homological critical value of p_h, and $\xi_{2,n} \searrow 0$. Suppose for each n, Theorem 5 is satisfied for $\xi_{1,n}, \xi_{2,n}$. Then under mild geometric assumption, $p_h^{-1}[\lambda_n, \infty) \simeq supp(p_h) \simeq supp(P)$ for all large n and small enough h.

3.2. Random Data

To control the probability of geometric conditions (1) and (2), the function space of kernel should not be too complex. One common approach is to assume that the function space is a uniformly bounded VC-class. Hence we need the following additional assumption.

- **uniformly bounded VC class.** Let $K : \mathbb{R}^d \to \mathbb{R}$ be a kernel function with $\|K\|_\infty, \|K\|_2 < \infty$. I assume that,

$$\mathcal{F}_h := \{K_{x,h} : x \in \mathcal{U}\},$$

is a uniformly bounded VC class with dimension ν. That is, there exists $A, \nu > 0$, for every probability measure Q on $\mathbb{R}^d$ and for every $\epsilon \in (0, \|K\|_\infty)$, the covering numbers $\mathcal{N}(\mathcal{F}_h, L_2(Q), \epsilon)$ satisfies

$$\mathcal{N}(\mathcal{F}_h, L_2(Q), \epsilon) \leq \left(\frac{A\|K\|_\infty}{\epsilon}\right)^\nu.$$

Under Assumption 3, we have the desired high probability bound, which is Corollary 13 and 21 in Kim et al. (2019).

Proposition 7.

(Corollary 13 and 21 in Kim et al. (2019)) Let P be a probability distribution in $\mathbb{R}^d$, and $X_1, \cdots, X_n$ be i.i.d. from P. Let K be a kernel function having compact support and of a uniformly bounded VC class. Then, for any $\epsilon > 0$, with probability at least

$$1 - \exp\left(C\left(\log(h^{-1} \wedge 1) - nh^{2d+2}\epsilon^2\right)\right),$$

we have

$$\max\left\{\sup_{x \in \mathcal{U}} |\hat{p}_h(x) - p_h(x)|, \sup_{x \in \mathcal{U}} \|\nabla \hat{p}_h(x) - \nabla p_h(x)\|_2\right\} \leq \epsilon,$$

where C is a constant depending only on P and K.

In our setting, let $\epsilon := \min\{\lambda_n - \xi_1, \xi_2 - \lambda_n, \frac{1}{2}\min_{x \in D} \|\nabla p_h(x)\|_2\}$ gives the high probability bound for Theorem 5, and let $\epsilon_n := \min\{\lambda_n - \xi_{1,n}, \xi_{2,n} - \lambda_n, \frac{1}{2}\min_{x \in D_n} \|\nabla p_h(x)\|_2\}$ gives the high probability bound for Corollary 6.

4. Conclusion

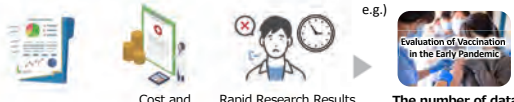
Compared to existing methods such as the union of balls or simplicial complexes, the plug-in estimator is more robust to outliers, enabling more reliable inference of the distribution's support. Moreover, since it is an application of a statistically well-studied object-KDE-probabilistic analysis can be conducted more easily. Fixing the bandwidth $h > 0$ of the kernel function helps alleviate the curse of dimensionality, and much faster convergence rates of both the kernel density estimator and the plug-in estimator can be achieved.

Predict vaccine-induced antibody dynamics from 1 or 2 blood samplings using mathematical models and machine learning

○Daiki Tatematsu[1], Shingo Iwami[1]
[1]Nagoya University, Japan
dikt1231@gmail.com | https://iblab.bio.nagoya-u.ac.jp

Introduction

Limitation of Medical Data



Estimate Antibody Dynamics from 1 or 2 Blood Samplings

The Integrated Approach of **Mathematical Models** and **Machine Learning** to Reduce the Number of Data Required and **Speed Up Research**

Fukushima Vaccine Cohort : COVID-19 Vaccine-Induced IgG(S) Antibody Data

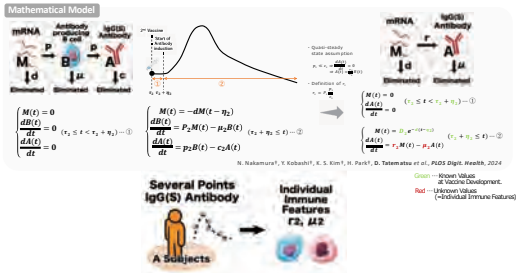
The largest and longest COVID-19 vaccine cohort in Japan, consisting mainly of healthcare professionals and patients living in Fukushima Prefecture (Soma, Minamisoma, Hirata)

N. Nakamura[†], Y. Kobashi[†], K. S. Kim[†], H. Park[†], **D. Tatematsu** *et al.*, ***PLOS Digit. Health***, 2024

Utilize Data from **the COVID-19 pandemic** as **Ground Truth**

Methods

A Subjects: Estimate Individual-Level Immune Features Using Mathematical Models



N. Nakamura[†], Y. Kobashi[†], K. S. Kim[†], H. Park[†], **D. Tatematsu** *et al.*, ***PLOS Digit. Health***, 2024

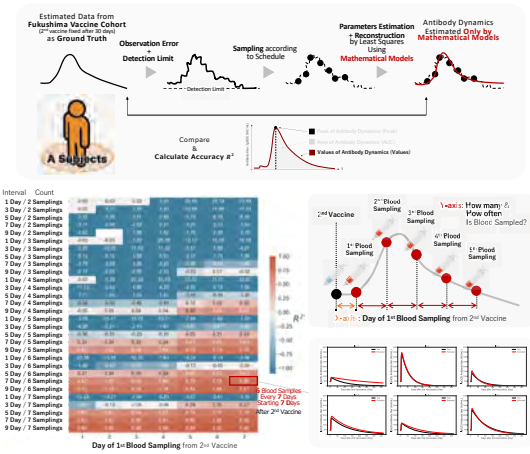
A comparatively large number of IgG(S) data are collected in a clinical trial or pre-cohort (=**A Subjects**), and **estimated parameters r_2 and μ_2** (=**individual immune features**) are used for machine learning training.

B Subjects: Estimate Individual-Level Immune Features Using Mathematical Models + Machine Learning

By using A subjects as training data for the machine learning, **IgG(S) antibody dynamics** of new subjects (=**B Subjects**) is available from only **1 or 2 blood samples**.

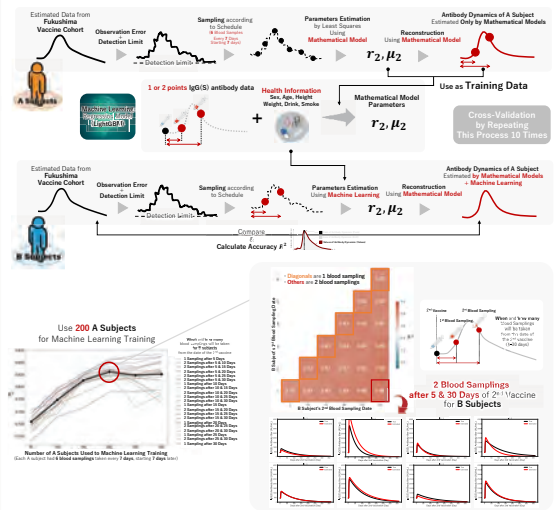
Results

A Subjects: Only Mathematical Models



Highly accurate estimation of antibody dynamics **only by mathematical models** requires **more than 5 blood samples**.

B Subjects: Mathematical Models + Machine Learning



The number of blood samplings required for highly accurate estimation of antibody dynamics can be reduced **from more than 5 times to 1 or 2 times by mathematical models + machine learning**.

Conclusion

How to use it in the Next Pandemic

Mathematical models + machine learning reduce the number of blood sampling and **speed up vaccine evaluation**.

Limit Theorems for Verbose Persistence Diagrams

Jeong-hwi Joe* — KAIST — jhjoe@kaist.ac.kr

Woojin Kim — KAIST — woojin.kim@kaist.ac.kr

Cheolwoo Park — KAIST — parkcw2021@kaist.ac.kr

Introduction

- At the intersection of random topology and TDA, the asymptotic behavior of persistence diagrams has been widely studied.
- The verbose diagram [UZ16] refines the persistence diagram by including extra points along the diagonal.
- In this work, we initiate the study of the asymptotic behavior of verbose diagrams. In particular, we extend all of the main theorems in [HST18] and [SS22] to the setting of verbose diagram.

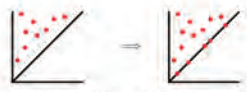
Preliminaries

Vague Convergence and Marked Point Processes

- $\mathscr{R}(S)$: the set of **Radon measures** on a second-countable locally compact Hausdorff space S, i.e., the set of measures on S which are finite on compact sets.
- A sequence $\{\mu_n\}_n$ of measures in $\mathscr{R}(S)$ is said to **converge vaguely** to a measure $\mu \in \mathscr{R}(S)$, and we write $\mu_n \xrightarrow{v} \mu$, if
$$\int_S f \, d\mu_n \to \int_S f \, d\mu \quad \text{as } n \to \infty$$
for every continuous function $f : S \to \mathbb{R}$ with compact support.
- Let $\mathbb{M}$ be the set of **marks**, which is also second-countable locally compact Hausdorff. A **marked point process** on S with marks in $\mathbb{M}$ is a random variable that outputs a (multi-)subset of $S \times \mathbb{M}$ such that (i) the projection on S is always locally finite and (ii) almost surely, no two elements have the same first coordinate.
- A **stationary** marked point process on $\mathbb{R}^N$ is a marked point process whose distribution does not change after arbitrary translation on the first coordinate.

Verbose Diagrams

- To get the **verbose diagram** of a (simplicial) filtration, start with an ordinary persistence diagram and **additionally record the cycles that are created and filled in simultaneously** on the **diagonal line**.



Persistence Diagram (PD) — Verbose diagram (VD)

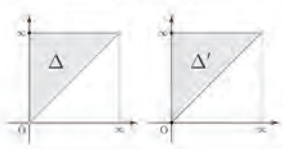
- **Example.** In the filtration below, a 1-cycle is filled in as soon as it is created at $t = 3$. This is captured by the point $(3, 3)$ in the degree-1 verbose diagram.

Why should we utilize VDs instead of PDs?

- The VDs are **easy to calculate** as provide a **stronger stability** than the PDs [ZM24]. are **finer invariants** than
- The use of VD makes the *persistent homology transform* [TMB14] **more efficient**, requiring fewer directional diagrams to ensure faithfulness [FMS24].

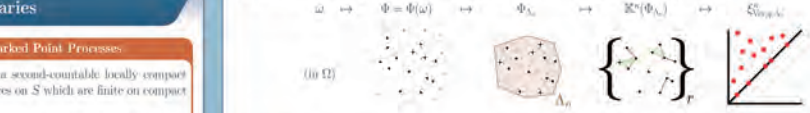
Diagrams as Measures

- Let $(X, \mathcal{A})$ be a measurable space. A multiset A of points in X is identified with the discrete measure
$$\xi = \sum_{a \in A} \delta_a,$$
where δ_a is the Delta measure at a.
- Thus, persistence diagrams and verbose diagrams can be represented as corresponding discrete measures on Δ and Δ', respectively.



Procedure of Generating a Random Diagram

- $(\Omega, \mathcal{F}, \mathbb{P})$: a probability space
- Φ: a stationary marked point process on $\mathbb{R}^N$ (with marks in $\mathbb{M}$)
- $\{\Lambda_n\}_n$: an increasing sequence of bounded Borel convex subsets of $\mathbb{R}^N$ that eventually covers every point of $\mathbb{R}^N$
- Φ_{Λ_n}: restriction of Φ to Λ_n.
- $\mathbb{K}^\kappa(\Phi_{\Lambda_n})$: a filtration built on Φ_{Λ_n}. It can be Čech or Vietoris–Rips filtration, or more general filtrations whose rule is specified with the filtration function κ.
- $\xi_{\mathrm{Vrb},q}(\mathcal{C}(\mathbb{K}^\kappa(\Phi_{\Lambda_n}))) = \xi^\kappa_{\mathrm{Vrb},q,\Lambda_n}$: the q-th verbose diagram of $\mathbb{K}^\kappa(\Phi_{\Lambda_n})$, which is represented as a discrete measure on Δ'.
- An illustration of the procedure of generating a random verbose diagram (marks oppressed):



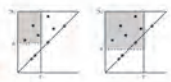
- $\xi_q(\mathbb{K}^\kappa(\Phi_{\Lambda_n})) = \xi^\kappa_{q,\Lambda_n}$: the q-th (ordinary) persistence diagram of $\mathbb{K}^\kappa(\Phi_{\Lambda_n})$, which is represented as a discrete measure on Δ.

Main Results

More notation:

- $R_q(\kappa)$ (resp. $R'_q(\kappa)$): the subset of Δ (resp. Δ') that consists of all locations where persistence (resp. verbose) diagrams can place points.
- $\beta^{\kappa,r,s}_q$: the q-th persistent Betti number of $\mathbb{K}^\kappa(\Phi_{\Lambda_n})$ from r to s.

For any filtration, we extend the notion of persistent Betti number β^κ_q from $r, s \in [0, \infty)$ to also allow the case $r, s = \infty$. For any $r, s \in [0, \infty]$, $\beta^{\kappa,r,s}_q$ equals the number of points of the verbose diagram that is in the region $[0, s] \times (r, \infty]$. See on the right.



The table below is the summary of our main theorems. The second column contains previous results and the third column contains their extensions to the verbose diagrams. The third row is our original result.

	PD	VD (this work)
Strong law of large numbers	[HST18, Theorem 1.5], [SS22, Theorem 2.6] Assumptions omitted for brevity $\frac{1}{\mathrm{vol}(\Lambda_n)} \mathbb{E}[\xi^\kappa_{q,\Lambda_n}] \xrightarrow{v} \nu_q \quad as\ n \to \infty.$ *Moreover, if Φ is ergodic, then almost surely,* $\frac{1}{\mathrm{vol}(\Lambda_n)} \xi^\kappa_{q,\Lambda_n} \xrightarrow{v} \nu_q \quad as\ n \to \infty.$	**Theorem 1.** Assumptions omitted for brevity $\frac{1}{\mathrm{vol}(\Lambda_n)} \mathbb{E}[\xi^\kappa_{\mathrm{Vrb},q,\Lambda_n}] \xrightarrow{v} \nu_q \quad as\ n \to \infty.$ *Moreover, if Φ is ergodic, then almost surely,* $\frac{1}{\mathrm{vol}(\Lambda_n)} \xi^\kappa_{\mathrm{Vrb},q,\Lambda_n} \xrightarrow{v} \nu_q \quad as\ n \to \infty.$
Total mass of the limiting measure		**Theorem 2.** *Suppose that the intensity λ of Φ, defined by $\lambda = \mathbb{E}[\Phi([0,1)^N)]$, is positive, and that $\kappa(\sigma) < \infty$ for all $\sigma \in \mathscr{F}(\mathbb{R}^N)$. Then* $\nu'_q(\Delta') = \begin{cases} \lambda, & \text{if } q = 0, \\ \infty & \text{otherwise.} \end{cases}$
Support of the limiting measure	[HST18, Theorem 1.9] Assumptions omitted for brevity. Then $\mathrm{supp}(\nu_q) = \overline{R_q(\kappa)}$ *for every $q \geq 0$, where the closure is taken in Δ.*	**Theorem 3.** Assumptions omitted for brevity. Then $\mathrm{supp}(\nu_q) = \overline{R_q(\kappa)}$ *for every $q \geq 0$, where the closure is taken in Δ.*
CLT for the persistent Betti number	[HST18, Theorem 1.12] Assumptions omitted for brevity. Then, for any $0 \leq r \leq s < \infty$ and $q \geq 0$, there exists a constant $\sigma^2_{r,s} = \sigma^2_{r,s}(\kappa, q)$ such that $\frac{\beta^{\kappa,r,s}_q - \mathbb{E}[\beta^{\kappa,r,s}_q]}{L^{N/2}} \xrightarrow{d} \mathcal{N}(0, \sigma^2_{r,s})$ *as $L \to \infty$.*	**Theorem 4.** Assumptions omitted for brevity. Then, for any $r, s \in [0, \infty]$ and $q \geq 0$, there exists a constant $\sigma^2_{r,s} = \sigma^2_{r,s}(\kappa, q)$ such that $\frac{\beta^{\kappa,r,s}_q - \mathbb{E}[\beta^{\kappa,r,s}_q]}{L^{N/2}} \xrightarrow{d} \mathcal{N}(0, \sigma^2_{r,s})$ *as $L \to \infty$.*

References

[FMS24] Brittany Terese Fasy, David L Millman, and Anna Schenfisch. "How Small Can Faithful Sets Be? Ordering Topological Descriptors". In: *arXiv preprint arXiv:2402.13622* (2024).

[HST18] Yasuaki Hiraoka, Tomoyuki Shirai, and Khanh Duy Trinh. "Limit theorems for persistence diagrams". In: *The Annals of Applied Probability* 28.5 (2018), pp. 2740–2780. DOI: 10.1214/17-AAP1371. URL: https://doi.org/10.1214/17-AAP1371.

[SS22] Tomoyuki Shirai and Kiyotaka Suzuki. "A limit theorem for persistence diagrams of random Čhrevoi complexes built over marked point processes". In: *Modern Stochastics: Theory and Applications* 10.1 (2022), pp. 1–18.

[TMB14] Katharine Turner, Sayan Mukherjee, and Doug M Boyer. "Persistent homology transform for modeling shapes and surfaces". In: *Information and Inference: A Journal of the IMA* 3.4 (2014), pp. 310–344.

[UZ16] Michael Usher and Jun Zhang. "Persistent homology and Floer–Novikov theory". In: *Geometry & Topology* 20.6 (Dec. 2016), pp. 3333–3430. ISSN: 1465-3060. DOI: 10.2140/gt.2016.20.3333. URL: http://dx.doi.org/10.2140/gt.2016.20.3333.

[ZM24] Ling Zhou and Facundo Mémoli. "Ephemeral persistence features and the stability of filtered chain complexes". In: *Journal of Computational Geometry* 15.2 (2024), pp. 258–328.

Partially supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-00515946).

Partially supported supported by the National Research Foundation of Korea (NRF-2021R1A2C1092925).

MI レクチャーノートシリーズ刊行にあたり

本レクチャーノートシリーズは、文部科学省 21 世紀 COE プログラム「機能数理学の構築と展開」(H15-19 年度) において作成した COE Lecture Notes の続刊であり、文部科学省大学院教育改革支援プログラム「産業界が求める数学博士と新修士養成」(H19-21 年度) および、同グローバル COE プログラム「マス・フォア・インダストリ教育研究拠点」(H20-24 年度) において行われた講義の講義録として出版されてきた。平成 23 年 4 月のマス・フォア・インダストリ研究所 (IMI) 設立と平成 25 年 4 月の IMI の文部科学省共同利用・共同研究拠点として「産業数学の先進的･基礎的共同研究拠点」の認定を受け、今後、レクチャーノートは、マス・フォア・インダストリに関わる国内外の研究者による講義の講義録、会議録等として出版し、マス・フォア・インダストリの本格的な展開に資するものとする。

2022 年 10 月
マス・フォア・インダストリ研究所
所長　梶原 健司

Young Researchers and Students-Workshop (I)

Topological Data Analysis and Industrial Mathematics

発　行　2025年 12 月 22 日
編　集　Keunsu Kim
発　行　九州大学マス・フォア・インダストリ研究所
九州大学大学院数理学府
〒819-0395 福岡市西区元岡744
九州大学数理・IMI 事務室
TEL 092-802-4402　FAX 092-802-4405
URL https://www.imi.kyushu-u.ac.jp/

印　刷　城島印刷株式会社
〒810-0012 福岡市中央区白金 2 丁目 9 番 6 号
TEL 092-531-7102　FAX 092-524-4411

シリーズ既刊

Issue	Author／Editor	Title	Published
COE Lecture Note	Mitsuhiro T. NAKAO Kazuhiro YOKOYAMA	Computer Assisted Proofs - Numeric and Symbolic Approaches - 199pages	August 22, 2006
COE Lecture Note	M.J.Shai HARAN	Arithmetical Investigations - Representation theory, Orthogonal polynomials and Quantum interpolations- 174pages	August 22, 2006
COE Lecture Note Vol.3	Michal BENES Masato KIMURA Tatsuyuki NAKAKI	Proceedings of Czech-Japanese Seminar in Applied Mathematics 2005 155pages	October 13, 2006
COE Lecture Note Vol.4	宮田　健治	辺要素有限要素法による磁界解析 - 機能数理学特別講義 21pages	May 15, 2007
COE Lecture Note Vol.5	Francois APERY	Univariate Elimination Subresultants - Bezout formula, Laurent series and vanishing conditions - 89pages	September 25, 2007
COE Lecture Note Vol.6	Michal BENES Masato KIMURA Tatsuyuki NAKAKI	Proceedings of Czech-Japanese Seminar in Applied Mathematics 2006 209pages	October 12, 2007
COE Lecture Note Vol.7	若山　正人 中尾　充宏	九州大学産業技術数理研究センター キックオフミーティング 138pages	October 15, 2007
COE Lecture Note Vol.8	Alberto PARMEGGIANI	Introduction to the Spectral Theory of Non-Commutative Harmonic Oscillators 233pages	January 31, 2008
COE Lecture Note Vol.9	Michael I.TRIBELSKY	Introduction to Mathematical modeling 23pages	February 15, 2008
COE Lecture Note Vol.10	Jacques FARAUT	Infinite Dimensional Spherical Analysis 74pages	March 14, 2008
COE Lecture Note Vol.11	Gerrit van DIJK	Gelfand Pairs And Beyond 60pages	August 25, 2008
COE Lecture Note Vol.12	Faculty of Mathematics, Kyushu University	Consortium “MATH for INDUSTRY” First Forum 87pages	September 16, 2008
COE Lecture Note Vol.13	九州大学大学院 数理学研究院	プロシーディング「損保数理に現れる確率モデル」 ― 日新火災・九州大学 共同研究2008年11月 研究会 ― 82pages	February 6, 2009

シリーズ既刊

Issue	Author／Editor	Title	Published
COE Lecture Note Vol.14	Michal Beneš, Tohru Tsujikawa Shigetoshi Yazaki	Proceedings of Czech-Japanese Seminar in Applied Mathematics 2008 77pages	February 12, 2009
COE Lecture Note Vol.15	Faculty of Mathematics, Kyushu University	International Workshop on Verified Computations and Related Topics 129pages	February 23, 2009
COE Lecture Note Vol.16	Alexander Samokhin	Volume Integral Equation Method in Problems of Mathematical Physics 50pages	February 24, 2009
COE Lecture Note Vol.17	矢嶋　　徹 及川　正行 梶原　健司 辻　　英一 福本　康秀	非線形波動の数理と物理 66pages	February 27, 2009
COE Lecture Note Vol.18	Tim Hoffmann	Discrete Differential Geometry of Curves and Surfaces 75pages	April 21, 2009
COE Lecture Note Vol.19	Ichiro Suzuki	The Pattern Formation Problem for Autonomous Mobile Robots —Special Lecture in Functional Mathematics— 23pages	April 30, 2009
COE Lecture Note Vol.20	Yasuhide Fukumoto Yasunori Maekawa	Math-for-Industry Tutorial: Spectral theories of non-Hermitian operators and their application 184pages	June 19, 2009
COE Lecture Note Vol.21	Faculty of Mathematics, Kyushu University	Forum “Math-for-Industry” Casimir Force, Casimir Operators and the Riemann Hypothesis 95pages	November 9, 2009
COE Lecture Note Vol.22	Masakazu Suzuki Hoon Hong Hirokazu Anai Chee Yap Yousuke Sato Hiroshi Yoshida	The Joint Conference of ASCM 2009 and MACIS 2009: Asian Symposium on Computer Mathematics Mathematical Aspects of Computer and Information Sciences 436pages	December 14, 2009
COE Lecture Note Vol.23	荒川　恒男 金子　昌信	多重ゼータ値入門 111pages	February 15, 2010
COE Lecture Note Vol.24	Fulton B.Gonzalez	Notes on Integral Geometry and Harmonic Analysis 125pages	March 12, 2010
COE Lecture Note Vol.25	Wayne Rossman	Discrete Constant Mean Curvature Surfaces via Conserved Quantities 130pages	May 31, 2010
COE Lecture Note Vol.26	Mihai Ciucu	Perfect Matchings and Applications 66pages	July 2, 2010

シリーズ既刊

Issue	Author／Editor	Title	Published
COE Lecture Note Vol.27	九州大学大学院 数理学研究院	Forum “Math-for-Industry” and Study Group Workshop Information security, visualization, and inverse problems, on the basis of optimization techniques 100pages	October 21, 2010
COE Lecture Note Vol.28	ANDREAS LANGER	MODULAR FORMS, ELLIPTIC AND MODULAR CURVES LECTURES AT KYUSHU UNIVERSITY 2010 62pages	November 26, 2010
COE Lecture Note Vol.29	木田　雅成 原田　昌晃 横山　俊一	Magma で広がる数学の世界 157pages	December 27, 2010
COE Lecture Note Vol.30	原　　隆 松井　　卓 廣島　文生	Mathematical Quantum Field Theory and Renormalization Theory 201pages	January 31, 2011
COE Lecture Note Vol.31	若山　正人 福本　康秀 高木　　剛 山本　昌宏	Study Group Workshop 2010 Lecture & Report 128pages	February 8, 2011
COE Lecture Note Vol.32	Institute of Mathematics for Industry, Kyushu University	Forum “Math-for-Industry” 2011 “TSUNAMI-Mathematical Modelling” Using Mathematics for Natural Disaster Prediction, Recovery and Provision for the Future 90pages	September 30, 2011
COE Lecture Note Vol.33	若山　正人 福本　康秀 高木　　剛 山本　昌宏	Study Group Workshop 2011 Lecture & Report 140pages	October 27, 2011
COE Lecture Note Vol.34	Adrian Muntean Vladimír Chalupecký	Homogenization Method and Multiscale Modeling 72pages	October 28, 2011
COE Lecture Note Vol.35	横山　俊一 夫　　紀恵 林　　卓也	計算機代数システムの進展 210pages	November 30, 2011
COE Lecture Note Vol.36	Michal Beneš Masato Kimura Shigetoshi Yazaki	Proceedings of Czech-Japanese Seminar in Applied Mathematics 2010 107pages	January 27, 2012
COE Lecture Note Vol.37	若山　正人 高木　　剛 Kirill Morozov 平岡　裕章 木村　正人 白井　朋之 西井　龍映 栄　伸一郎 穴井　宏和 福本　康秀	平成23年度 数学・数理科学と諸科学・産業との連携研究ワークショップ 拡がっていく数学 ～期待される “見えない力” ～ 154pages	February 20, 2012

シリーズ既刊

Issue	Author／Editor	Title	Published
COE Lecture Note Vol.38	Fumio Hiroshima Itaru Sasaki Herbert Spohn Akito Suzuki	Enhanced Binding in Quantum Field Theory 204pages	March 12, 2012
COE Lecture Note Vol.39	Institute of Mathematics for Industry, Kyushu University	Multiscale Mathematics: Hierarchy of collective phenomena and interrelations between hierarchical structures 180pages	March 13, 2012
COE Lecture Note Vol.40	井ノ口順一 太田　泰広 筧　　三郎 梶原　健司 松浦　　望	離散可積分系・離散微分幾何チュートリアル2012 152pages	March 15, 2012
COE Lecture Note Vol.41	Institute of Mathematics for Industry, Kyushu University	Forum “Math-for-Industry” 2012 “Information Recovery and Discovery” 91pages	October 22, 2012
COE Lecture Note Vol.42	佐伯　　修 若山　正人 山本　昌宏	Study Group Workshop 2012 Abstract, Lecture & Report 178pages	November 19, 2012
COE Lecture Note Vol.43	Institute of Mathematics for Industry, Kyushu University	Combinatorics and Numerical Analysis Joint Workshop 103pages	December 27, 2012
COE Lecture Note Vol.44	萩原　　学	モダン符号理論からポストモダン符号理論への展望 107pages	January 30, 2013
COE Lecture Note Vol.45	金山　　寛	Joint Research Workshop of Institute of Mathematics for Industry (IMI), Kyushu University “Propagation of Ultra-large-scale Computation by the Domain-decomposition-method for Industrial Problems (PUCDIP 2012)” 121pages	February 19, 2013
COE Lecture Note Vol.46	西井　龍映 栄　伸一郎 岡田　勘三 落合　啓之 小磯　深幸 斎藤　新悟 白井　朋之	科学・技術の研究課題への数学アプローチ ―数学モデリングの基礎と展開― 325pages	February 28, 2013
COE Lecture Note Vol.47	SOO TECK LEE	BRANCHING RULES AND BRANCHING ALGEBRAS FOR THE COMPLEX CLASSICAL GROUPS 40pages	March 8, 2013
COE Lecture Note Vol.48	溝口　佳寛 脇　　隼人 平坂　　貢 谷口　哲至 島袋　　修	博多ワークショップ「組み合わせとその応用」 124pages	March 28, 2013

シリーズ既刊

Issue	Author／Editor	Title	Published
COE Lecture Note Vol.49	照井　章 小原　功任 濱田　龍義 横山　俊一 穴井　宏和 横田　博史	マス・フォア・インダストリ研究所　共同利用研究集会 II 数式処理研究と産学連携の新たな発展　137pages	August 9, 2013
MI Lecture Note Vol.50	Ken Anjyo Hiroyuki Ochiai Yoshinori Dobashi Yoshihiro Mizoguchi Shizuo Kaji	Symposium MEIS2013: Mathematical Progress in Expressive Image Synthesis　154pages	October 21, 2013
MI Lecture Note Vol.51	Institute of Mathematics for Industry, Kyushu University	Forum “Math-for-Industry” 2013 “The Impact of Applications on Mathematics”　97pages	October 30, 2013
MI Lecture Note Vol.52	佐伯　修 岡田　勘三 髙木　剛 若山　正人 山本　昌宏	Study Group Workshop 2013 Abstract, Lecture & Report　142pages	November 15, 2013
MI Lecture Note Vol.53	四方　義啓 櫻井　幸一 安田　貴徳 Xavier Dahan	平成25年度　九州大学マス・フォア・インダストリ研究所 共同利用研究集会　安全・安心社会基盤構築のための代数構造 ～サイバー社会の信頼性確保のための数理学～　158pages	December 26, 2013
MI Lecture Note Vol.54	Takashi Takiguchi Hiroshi Fujiwara	Inverse problems for practice, the present and the future　93pages	January 30, 2014
MI Lecture Note Vol.55	栄　伸一郎 溝口　佳寛 脇　隼人 渋田　敬史	Study Group Workshop 2013 数学協働プログラム Lecture & Report 98pages	February 10, 2014
MI Lecture Note Vol.56	Yoshihiro Mizoguchi Hayato Waki Takafumi Shibuta Tetsuji Taniguchi Osamu Shimabukuro Makoto Tagami Hirotake Kurihara Shuya Chiba	Hakata Workshop 2014 ~ Discrete Mathematics and its Applications ~　141pages	March 28, 2014
MI Lecture Note Vol.57	Institute of Mathematics for Industry, Kyushu University	Forum “Math-for-Industry” 2014: “Applications + Practical Conceptualization + Mathematics = fruitful Innovation”　93pages	October 23, 2014
MI Lecture Note Vol.58	安生健一 落合啓之	Symposium MEIS2014: Mathematical Progress in Expressive Image Synthesis　135pages	November 12, 2014

シリーズ既刊

Issue	Author／Editor	Title	Published
MI Lecture Note Vol.59	西井　龍映 岡田　勘三 梶原　健司 髙木　　剛 若山　正人 脇　　隼人 山本　昌宏	Study Group Workshop 2014 数学協働プログラム Abstract, Lecture & Report　196pages	November 14, 2014
MI Lecture Note Vol.60	西浦　　博	平成26年度九州大学 IMI 共同利用研究・研究集会（I） 感染症数理モデルの実用化と産業及び政策での活用のための新たな展開　120pages	November 28, 2014
MI Lecture Note Vol.61	溝口　佳寛 Jacques Garrigue 萩原　　学 Reynald Affeldt	研究集会 高信頼な理論と実装のための定理証明および定理証明器 Theorem proving and provers for reliable theory and implementations (TPP2014)　138pages	February 26, 2015
MI Lecture Note Vol.62	白井　朋之	Workshop on “β-transformation and related topics”　59pages	March 10, 2015
MI Lecture Note Vol.63	白井　朋之	Workshop on “Probabilistic models with determinantal structure” 107pages	August 20, 2015
MI Lecture Note Vol.64	落合　啓之 土橋　宜典	Symposium MEIS2015: Mathematical Progress in Expressive Image Synthesis　124pages	September 18, 2015
MI Lecture Note Vol.65	Institute of Mathematics for Industry, Kyushu University	Forum “Math-for-Industry” 2015 “The Role and Importance of Mathematics in Innovation”　74pages	October 23, 2015
MI Lecture Note Vol.66	岡田　勘三 藤澤　克己 白井　朋之 若山　正人 脇　　隼人 Philip Broadbridge 山本　昌宏	Study Group Workshop 2015 Abstract, Lecture & Report 156pages	November 5, 2015
MI Lecture Note Vol.67	Institute of Mathematics for Industry, Kyushu University	IMI-La Trobe Joint Conference “Mathematics for Materials Science and Processing” 66pages	February 5, 2016
MI Lecture Note Vol.68	古庄　英和 小谷　久寿 新甫　洋史	結び目と Grothendieck-Teichmüller 群 116pages	February 22, 2016
MI Lecture Note Vol.69	土橋　宜典 鍛治　静雄	Symposium MEIS2016: Mathematical Progress in Expressive Image Synthesis　82pages	October 24, 2016
MI Lecture Note Vol.70	Institute of Mathematics for Industry, Kyushu University	Forum “Math-for-Industry” 2016 “Agriculture as a metaphor for creativity in all human endeavors” 98pages	November 2, 2016
MI Lecture Note Vol.71	小磯　深幸 二宮　嘉行 山本　昌宏	Study Group Workshop 2016 Abstract, Lecture & Report　143pages	November 21, 2016

シリーズ既刊

Issue	Author／Editor	Title	Published
MI Lecture Note Vol.72	新井　朝雄 小嶋　　泉 廣島　文生	Mathematical quantum field theory and related topics　133pages	January 27, 2017
MI Lecture Note Vol.73	穴田　啓晃 Kirill Morozov 須賀　祐治 奥村　伸也 櫻井　幸一	Secret Sharing for Dependability, Usability and Security of Network Storage and Its Mathematical Modeling　211pages	March 15, 2017
MI Lecture Note Vol.74	QUISPEL, G. Reinout W. BADER, Philipp MCLAREN, David I. TAGAMI, Daisuke	IMI-La Trobe Joint Conference Geometric Numerical Integration and its Applications　71pages	March 31, 2017
MI Lecture Note Vol.75	手塚　　集 田上　大助 山本　昌宏	Study Group Workshop 2017 Abstract, Lecture & Report　118pages	October 20, 2017
MI Lecture Note Vol.76	宇田川誠一	Tzitzéica 方程式の有限間隙解に付随した極小曲面の構成理論 ―Tzitzéica 方程式の楕円関数解を出発点として―　68pages	August 4, 2017
MI Lecture Note Vol.77	松谷　茂樹 佐伯　　修 中川　淳一 田上　大助 上坂　正晃 Pierluigi Cesana 濵田　裕康	平成29年度　九州大学マス・フォア・インダストリ研究所 共同利用研究集会（I） 結晶の界面，転位，構造の数理　148pages	December 20, 2017
MI Lecture Note Vol.78	瀧澤　重志 小林　和博 佐藤憲一郎 斎藤　　努 清水　正明 間瀬　正啓 藤澤　克樹 神山　直之	平成29年度　九州大学マス・フォア・インダストリ研究所 プロジェクト研究　研究集会（I） 防災・避難計画の数理モデルの高度化と社会実装へ向けて 136pages	February 26, 2018
MI Lecture Note Vol.79	神山　直之 畔上　秀幸	平成29年度　AIMaP チュートリアル 最適化理論の基礎と応用　96pages	February 28, 2018
MI Lecture Note Vol.80	Kirill Morozov Hiroaki Anada Yuji Suga	IMI Workshop of the Joint Research Projects Cryptographic Technologies for Securing Network Storage and Their Mathematical Modeling　116pages	March 30, 2018
MI Lecture Note Vol.81	Tsuyoshi Takagi Masato Wakayama Keisuke Tanaka Noboru Kunihiro Kazufumi Kimoto Yasuhiko Ikematsu	IMI Workshop of the Joint Research Projects International Symposium on Mathematics, Quantum Theory, and Cryptography　246pages	September 25, 2019
MI Lecture Note Vol.82	池森　俊文	令和2年度　AIMaP チュートリアル 新型コロナウイルス感染症にかかわる諸問題の数理 145pages	March 22, 2021

シリーズ既刊

Issue	Author／Editor	Title	Published
MI Lecture Note Vol.83	早川健太郎 軸丸　芳揮 横須賀洋平 可香谷　隆 林　　和希 堺　　雄亮	シェル理論・膜理論への微分幾何学からのアプローチと その建築曲面設計への応用　49pages	July 28, 2021
MI Lecture Note Vol.84	Taketoshi Kawabe Yoshihiro Mizoguchi Junichi Kako Masakazu Mukai Yuji Yasui	SICE-JSAE-AIMaP Tutorial Advanced Automotive Control and Mathematics　110pages	December 27, 2021
MI Lecture Note Vol.85	Hiroaki Anada Yasuhiko Ikematsu Koji Nuida Satsuya Ohata Yuntao Wang	IMI Workshop of the Joint Usage Research Projects Exploring Mathematical and Practical Principles of Secure Computation and Secret Sharing　114pages	February 9, 2022
MI Lecture Note Vol.86	濱田　直希 穴井　宏和 梅田　裕平 千葉　一永 佐藤　寛之 能島　裕介 加葉田雄太朗 一木　俊助 早野　健太 佐伯　　修	2020年度採択分　九州大学マス・フォア・インダストリ研究所 共同利用研究集会 進化計算の数理　135pages	February 22, 2022
MI Lecture Note Vol.87	Osamu Saeki, Ho Tu Bao, Shizuo Kaji, Kenji Kajiwara, Nguyen Ha Nam, Ta Hai Tung, Melanie Roberts, Masato Wakayama, Le Minh Ha, Philip Broadbridge	Proceedings of Forum “Math-for-Industry” 2021 -Mathematics for Digital Economy-　122pages	March 28, 2022
MI Lecture Note Vol.88	Daniel PACKWOOD Pierluigi CESANA, Shigenori FUJIKAWA, Yasuhide FUKUMOTO, Petros SOFRONIS, Alex STAYKOV	Perspectives on Artificial Intelligence and Machine Learning in Materials Science, February 4-6, 2022　74pages	November 8, 2022

シリーズ既刊

Issue	Author／Editor	Title	Published
MI Lecture Note Vol.89	松谷　茂樹 落合　啓之 井上　和俊 小磯　深幸 佐伯　　修 白井　朋之 垂水　竜一 内藤　久資 中川　淳一 濵田　裕康 松江　　要 加葉田雄太朗	2022年度採択分 九州大学マス・フォア・インダストリ研究所 共同利用研究集会 材料科学における幾何と代数 III　356pages	December 7, 2022
MI Lecture Note Vol.90	中山　尚子 谷川　拓司 品野　勇治 近藤　正章 石原　　亨 鍛冶　静雄 藤澤　克樹	2022年度採択分 九州大学マス・フォア・インダストリ研究所 共同利用研究集会 データ格付けサービス実現のための数理基盤の構築　58pages	December 12, 2022
MI Lecture Note Vol.91	Katsuki Fujisawa Shizuo Kaji Toru Ishihara Masaaki Kondo Yuji Shinano Takuji Tanigawa Naoko Nakayama	IMI Workshop of the Joint Usage Research Projects Construction of Mathematical Basis for Realizing Data Rating Service 610pages	December 27, 2022
MI Lecture Note Vol.92	丹田　　聡 三宮　　俊 廣島　文生	2022年度採択分 九州大学マス・フォア・インダストリ研究所 共同利用研究集会 時間・量子測定・準古典近似の理論と実験 ～古典論と量子論の境界～　150pages	Janualy 6, 2023
MI Lecture Note Vol.93	Philip Broadbridge Luke Bennetts Melanie Roberts Kenji Kajiwara	Proceedings of Forum “Math-for-Industry” 2022 -Mathematics of Public Health and Sustainability-　170pages	June 19, 2023
MI Lecture Note Vol.94	國廣　　昇 池松　泰彦 伊豆　哲也 穴田　啓晃 縫田　光司	2023年度採択分 九州大学マス・フォア・インダストリ研究所 共同利用研究集会 現代暗号に対する安全性解析・攻撃の数理　260pages	Janualy 11, 2024
MI Lecture Note Vol.96	澤田　茉伊	2023年度採択分 九州大学マス・フォア・インダストリ研究所 共同利用研究集会 デジタル化時代に求められる斜面防災の思考法　70pages	March 18, 2024

シリーズ既刊

Issue	Author／Editor	Title	Published
MI Lecture Note Vol.97	Shariffah Suhaila Syed Jamaludin Zaiton Mat Isa Nur Arina Bazilah Aziz Taufiq Khairi Ahmad Khairuddin Shaymaa M.H.Darwish Ahmad Razin Zainal Abidin Norhaiza Ahmad Zainal Abdul Aziz Hang See Pheng Mohd Ali Khameini Ahmad	International Project Research-Workshop (I) Proceedings of 4th Malaysia Mathematics in Industry Study Group (MMISG2023)　172pages	March 28, 2024
MI Lecture Note Vol.98	中澤　　嵩	2024 年度採択分 九州大学マス・フォア・インダストリ研究所 共同利用研究集会 自動車性能の飛躍的向上を目指す Data-Driven 設計　92pages	January 30, 2025
MI Lecture Note Vol.99	Jacques Garrigue	2024 年度採択分 九州大学マス・フォア・インダストリ研究所 共同利用研究集会 コンピュータによる定理証明支援とその応用　308pages	March 17, 2025
MI Lecture Note Vol.100	Yutaka Jitsumatsu Masayoshi Ohashi Akio Hasegawa Katsutoshi Shinohara Shintaro Mori	IMI Workshop of the Joint Usage Research Projects Mathematics for Innovation in Information and Communication Technology 274pages	March 19, 2025
MI Lecture Note Vol.101	Makoto Ohsaki Yoshiki Jikumaru	IMI Workshop of the Joint Usage Research Projects Evolving Design and Discrete Differential Geometry:towards Mathematics Aided Geometric Design 528pages	October 1st, 2025

九州大学マス・フォア・インダストリ研究所
九州大学大学院 数理学府
〒819-0395 福岡市西区元岡744　TEL 092-802-4402　FAX 092-802-4405
URL https://www.imi.kyushu-u.ac.jp/