Mathematical AI for Bioscience

Guo-Wei Wei

Mathematics

Michigan State University

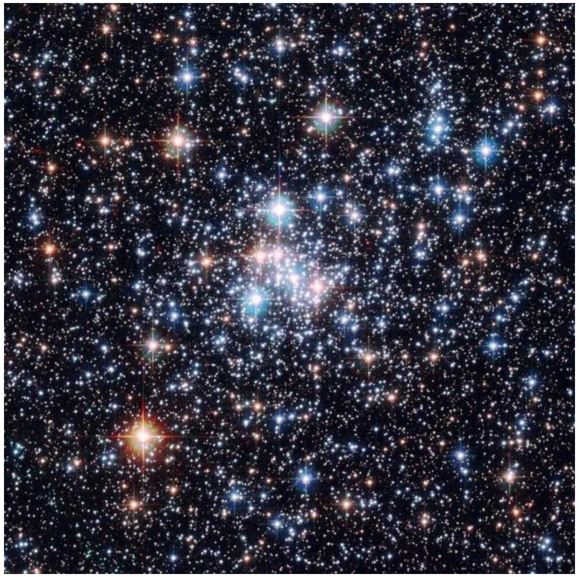
Fostering Cross-Disciplinary Collaboration in Biology, Medicine, and Computational Science

ICERM, Jul 28 - Aug 1, 2025

Grant support: NIH, NSF, DOE, NASA, Pfizer, BMS, MEDC, Georgetown U, COVID-19 HPC Consortium, MSU Foundation, and MSU iCER

Mathematical Foundations of AI and Data
Topological Aspects

Algebraic topology
for point cloud data
(discrete)



Forbes

Differential topology
for data on manifolds
(continuum)

2024 Microsoft

Geometric topology
for 1D curves
(1D in 3D)

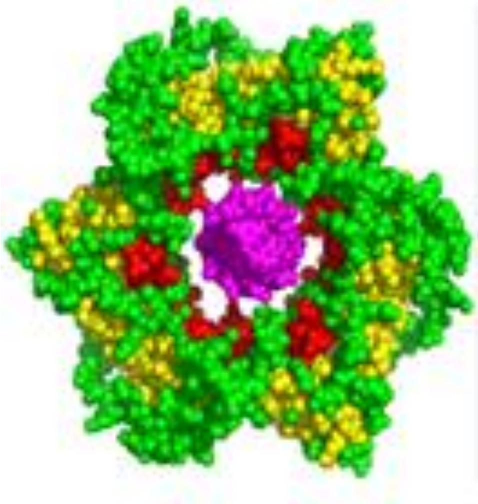
Lookphotos

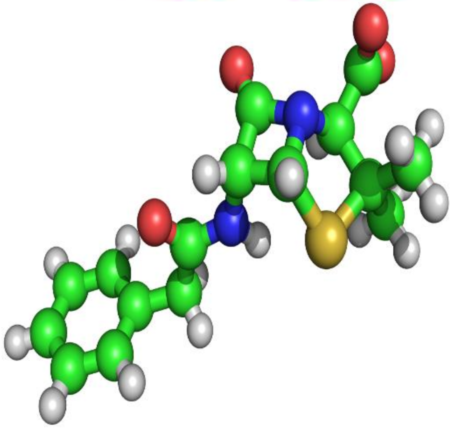
Mathematical foundations of AI and bioscience

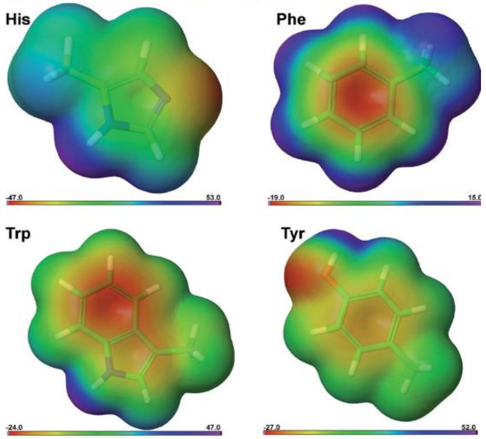
Point cloud

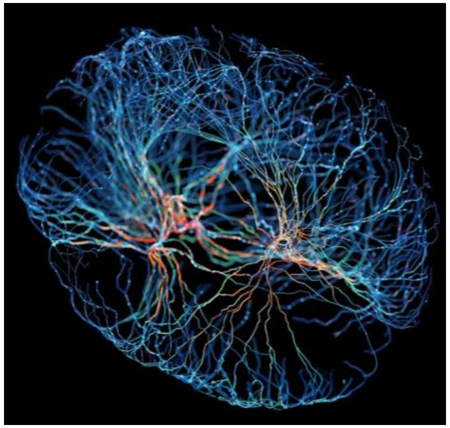
Manifold

curves (1D in 3D)









Algebraic topology/
Commutative algebra

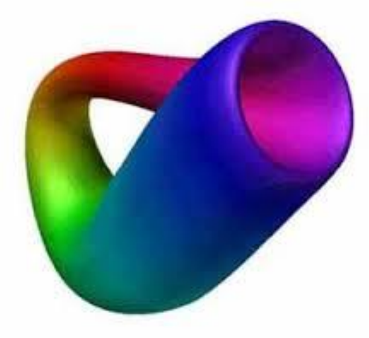
Differential geometry

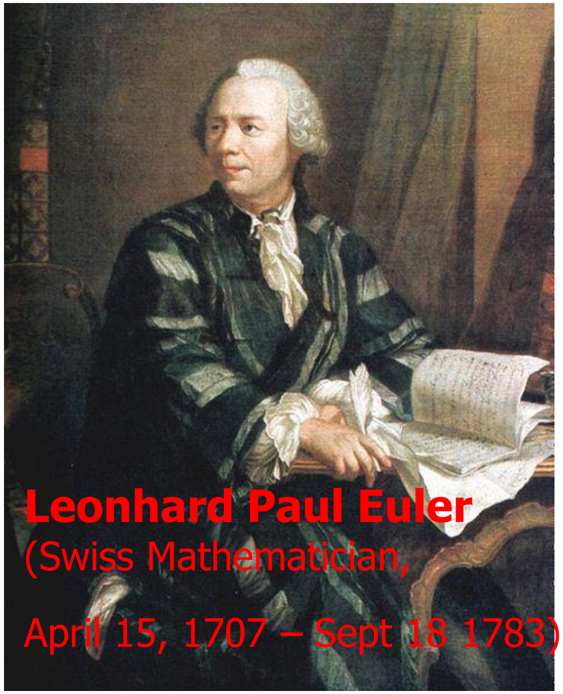
Geometric topology

Topology

Möbius Strips (1858)

Klein Bottle (1882)

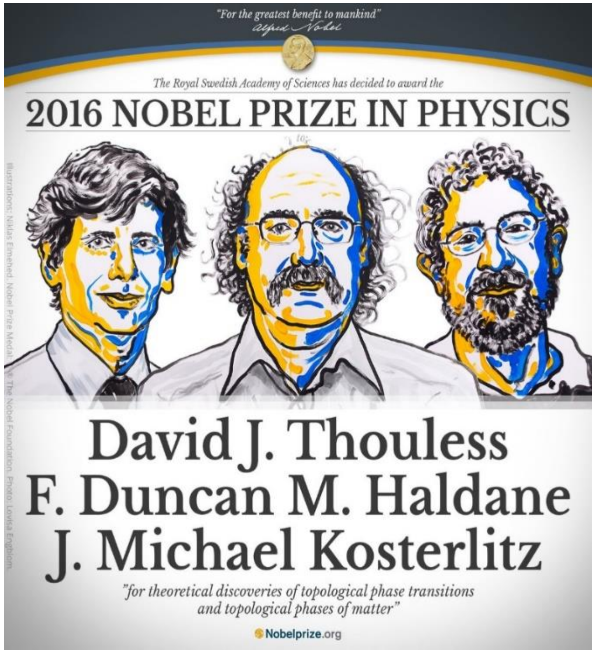




Leonhard Paul Euler
(Swiss Mathematician,
April 15, 1707 – Sept 18 1783)

Torus

Double Torus



"For the greatest benefit to mankind"
Alfred Nobel

The Royal Swedish Academy of Sciences has decided to award the

2016 NOBEL PRIZE IN PHYSICS

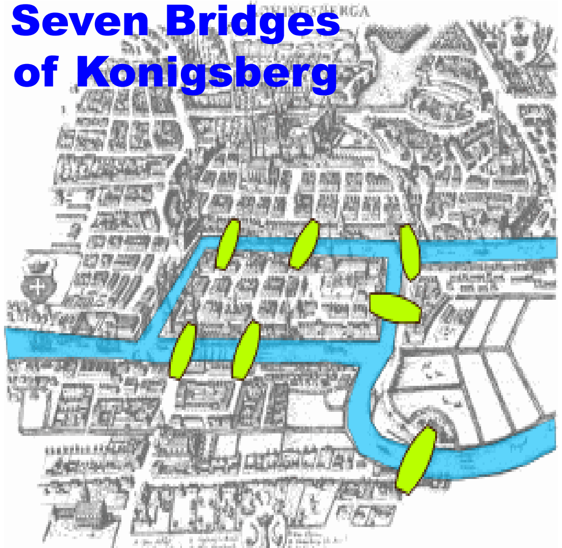
David J. Thouless
F. Duncan M. Haldane
J. Michael Kosterlitz

"for theoretical discoveries of topological phase transitions and topological phases of matter"

Nobelprize.org

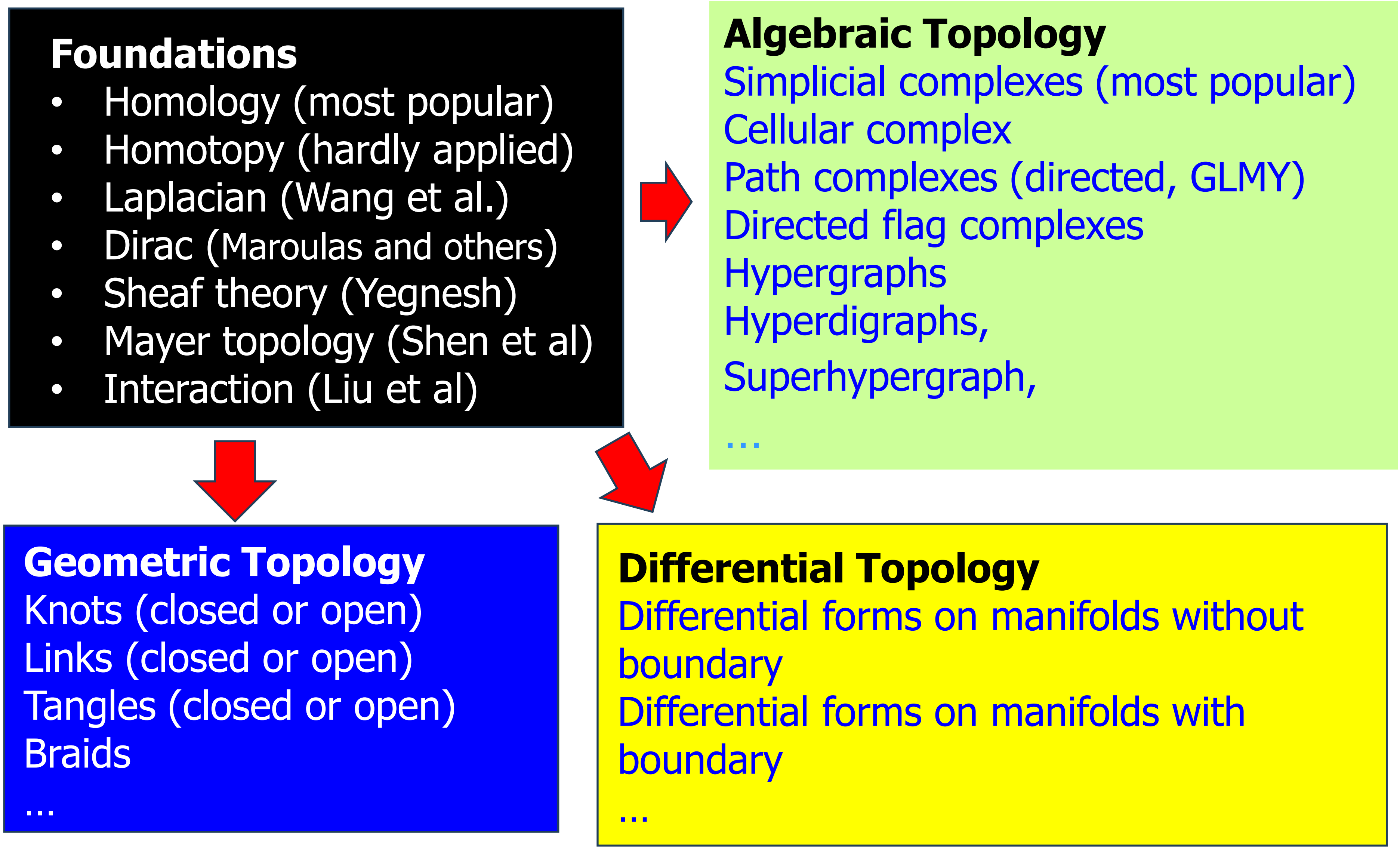
Augustin-Louis Cauchy,
Ludwig Schläfli,
Johann Benedict Listing,
Bernhard Riemann, and
Enrico Betti

Seven Bridges of Konigsberg



Leonhard Euler (1735)

Mathematics of Topological Data Analysis (TDA)



Foundations
- Homology (most popular)
- Homotopy (hardly applied)
- Laplacian (Wang et al.)
- Dirac (Maroulas and others)
- Sheaf theory (Yegnesh)
- Mayer topology (Shen et al)
- Interaction (Liu et al)

Algebraic Topology
Simplicial complexes (most popular)
Cellular complex
Path complexes (directed, GLMY)
Directed flag complexes
Hypergraphs
Hyperdigraphs,
Superhypergraph,
…

Geometric Topology
Knots (closed or open)
Links (closed or open)
Tangles (closed or open)
Braids
…

Differential Topology
Differential forms on manifolds without boundary
Differential forms on manifolds with boundary
…

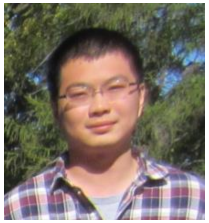
Machine Learning Aspects of TDA

Logistic regression
Support vector machine
Random forest
Ensemble methods
Transfer learning
Active learning
Deep neural network
Convolutional neural network
Nature language processing
Recurrent neural network
Long-short term memory
Graph neural network
Generative AI
ChatGPT
Autoencoder
Transformer
Manifold learning
Graph learning
Geometric learning
PCA
UMAP
t-SNE
Correlated clustering and projection
Topological deep learning
Multiscale Laplacian learning

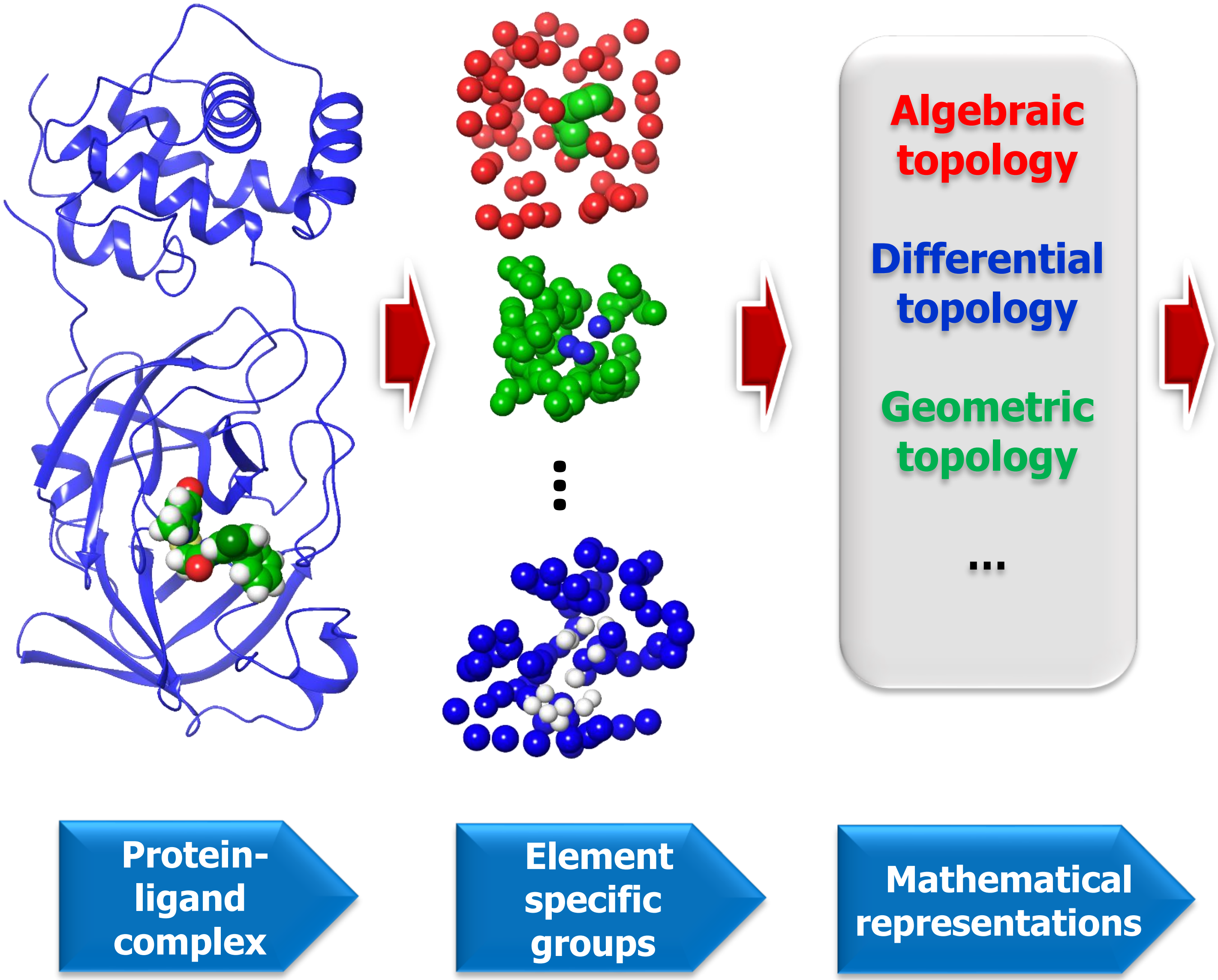
TopologyNet: Cang & Wei, PLOS Comput. Biol. 2017.

Topological Deep Learning (TDL)

The first integration of topology and neural networks



Zixuan Cang



Algebraic topology

Differential topology

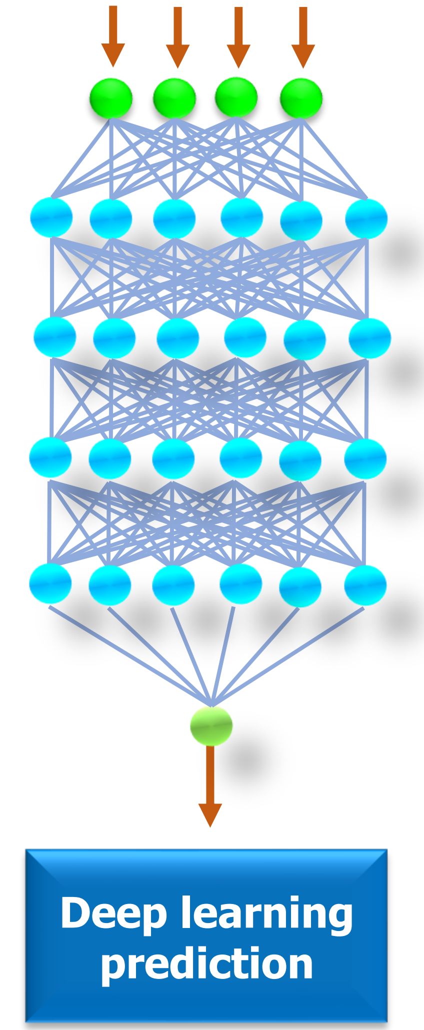
Geometric topology

...

Protein-ligand complex

Element specific groups

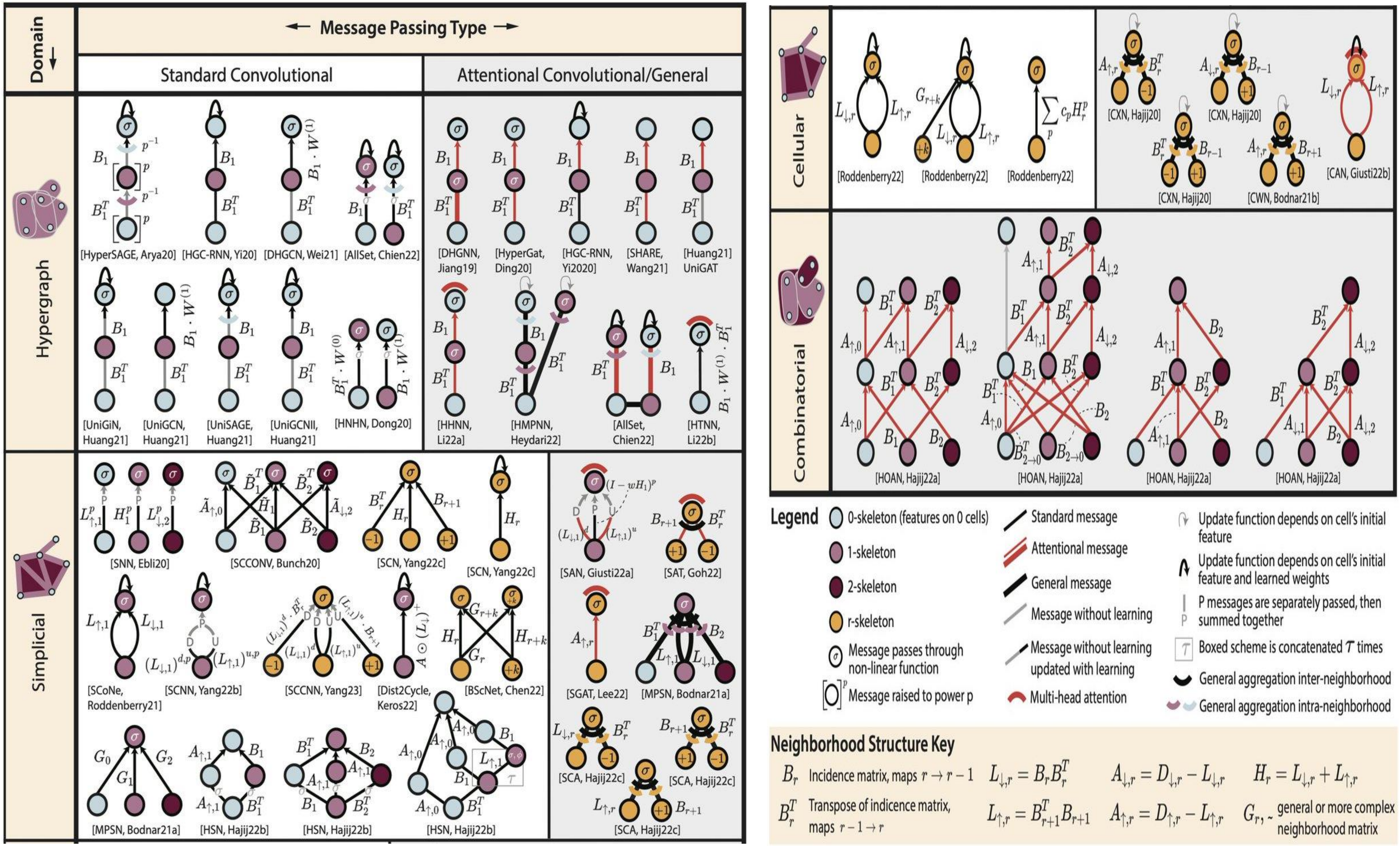
Mathematical representations



Deep learning prediction

Cang & Wei, PLOS Comput. Biol. 2017

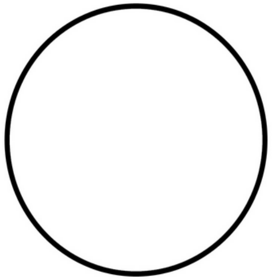
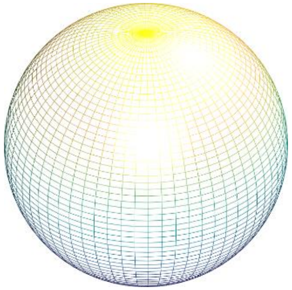
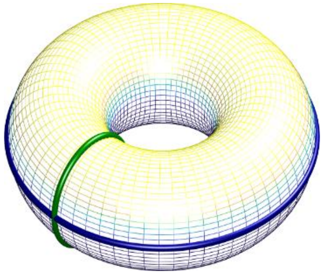
Topological Neural Networks
(Computer Science Aspects of TDL)



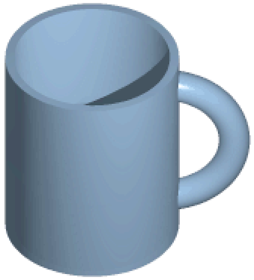
arXiv:2304.10031

Topological invariants: Betti numbers

β_0 is the number of connected components.
β_1 is the number of tunnels or circles.
β_2 is the number of cavities or voids.

Point	**Circle**	**Sphere**	**Torus**
			
$\beta_0 = 1$	$\beta_0 = 1$	$\beta_0 = 1$	$\beta_0 = 1$
$\beta_1 = 0$	$\beta_1 = 1$	$\beta_1 = 0$	$\beta_1 = 2$
$\beta_2 = 0$	$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 1$

Limitation

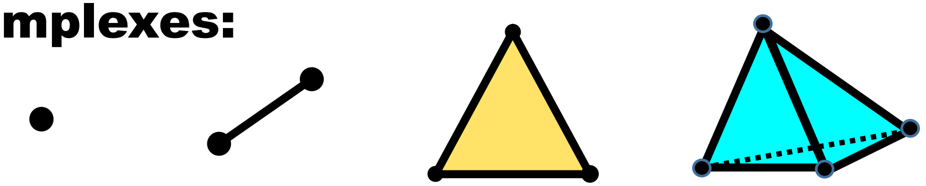


L. Vieira

Crane and Segerman

Algebraic Topology (Persistent homology)

Simplexes:



***0*-simplex** ***1*-simplex** ***2*-simplex** ***3*-simplex**

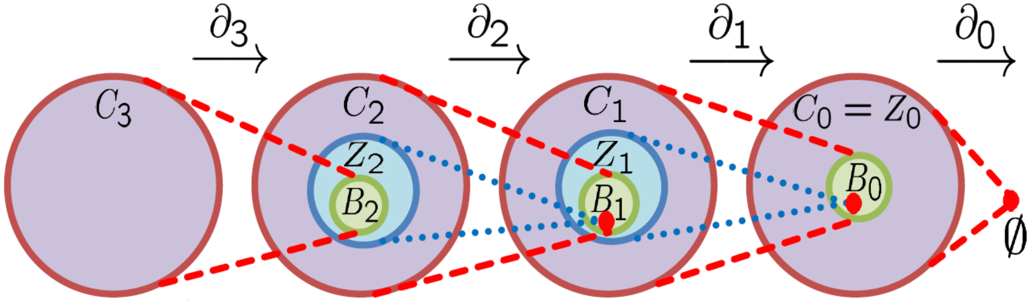
Frosini and Nandi (1999), Robins (1999), Edelsbrunner, Letscher and Zomorodian (2002), Zomorodian and Carlsson (2005), Edelsbrunner and Harer, (2007) Kaczynski, Mischaikow and Mrozek (2004), Ghrist (2008), …

k-chain: $K = \left\{ \sum_j c_j \sigma_j^q \right\}$

Chain group: $C_q(K, \mathbb{Z}_2)$

Boundary operator:

$$\partial_q \sigma^q = \sum_{j=0}^{q} (-1)^j \{v_0, v_1, \dots, \widehat{v_j}, \dots, v_k\}$$



Cycle group: $Z_q = \text{Ker}\, \partial_q$

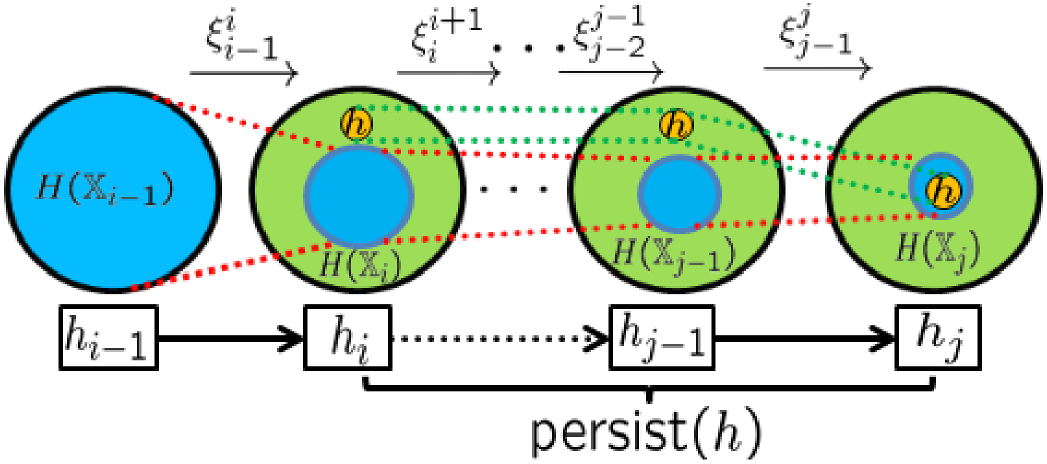
Boundary group: $B_q = \text{Im}\, \partial_{q+1}$

Homology group: $H_q = Z_q / B_q$

Betti number: $\beta_q = \text{Rank}(H_q)$

Xia, Wei, IJNMBE, 2014;
Xia, Feng, Tong, Wei, JCC, 2015

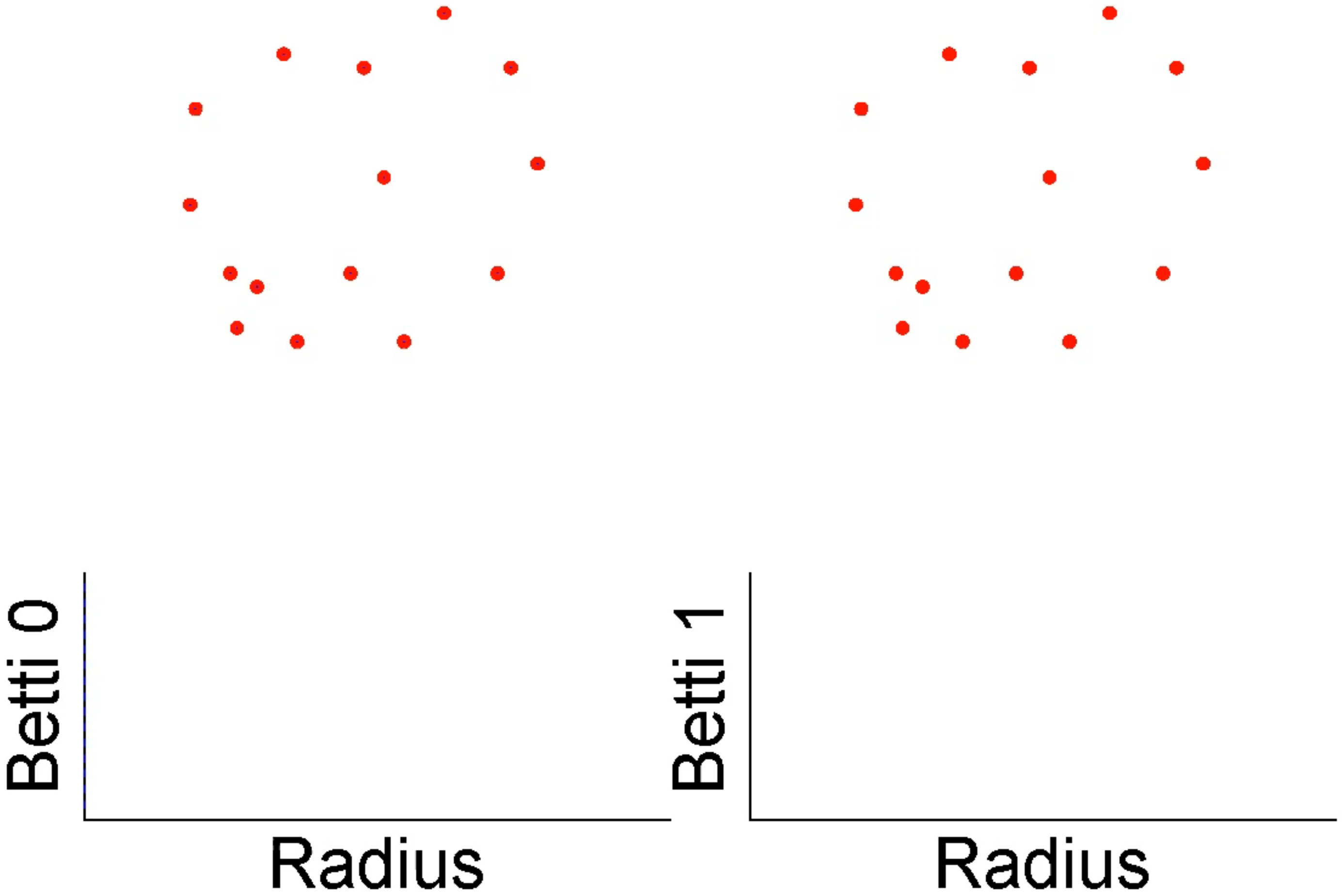
Filtration:



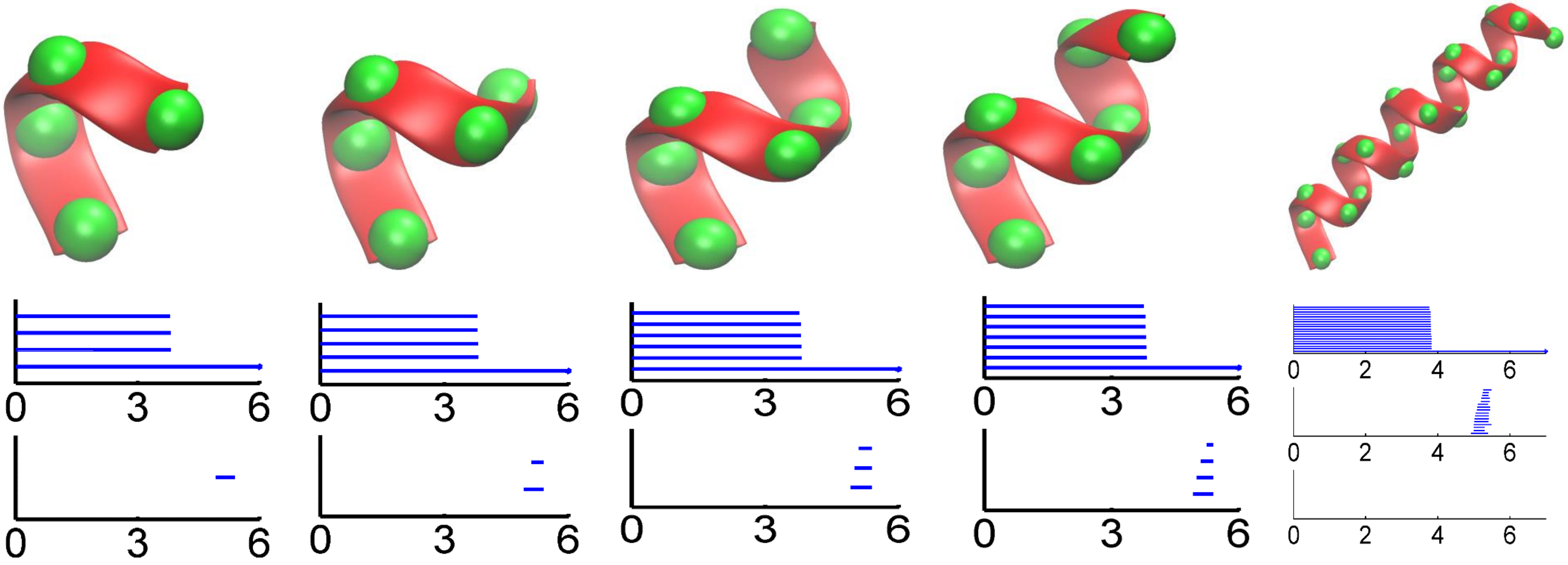
Topological data analysis

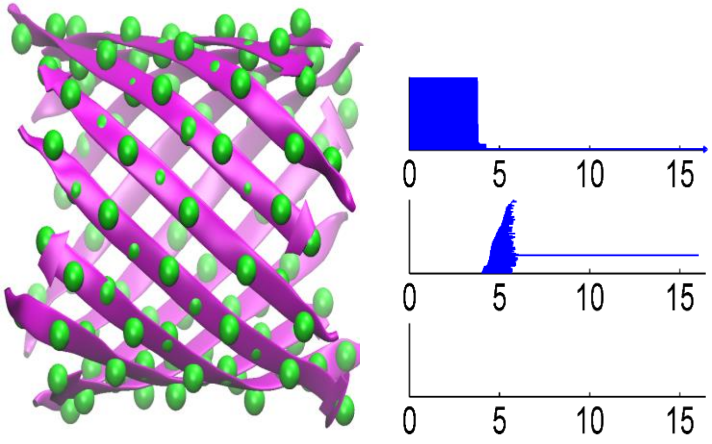
Vietoris-Rips complexes, persistent homology and topological fingerprint

(Xia, Wei, 2014)

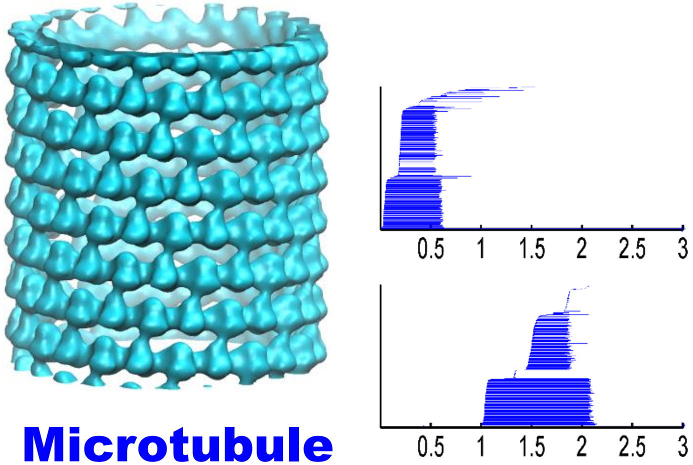


Topological fingerprints of biomolecules





Beta barrel



Microtubule

(Xia & Wei, IJNMBE, 2014, 2015)

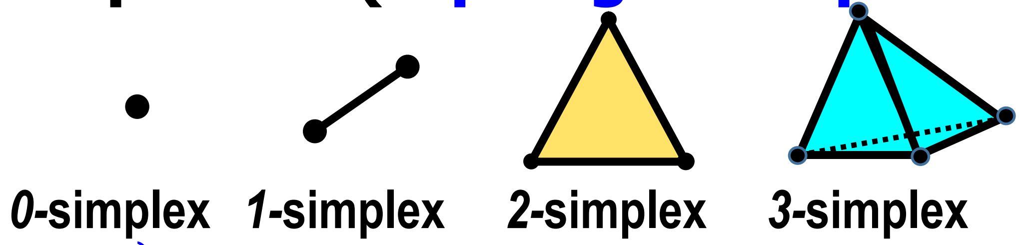
Limitations of Persistent Homology associated with simplicial complex

- It cannot handle heterogeneous information (i.e., different type of objects in the data).
- It is qualitative rather than quantitative (e.g., a 5-member ring is counted the same as a 6-member ring).
- It cannot describe non-topological changes (i.e., homotopic shape evolution over filtration).
- It is incapable of dealing with directed networks and digraphs (polarization, regulation, & control issues).
- It is unable to characterize structured data (e.g., hypergraphs, directed networks).

We address these limitations with new topological methods

Combinatorial Laplacian (topological Laplacian)

- Simplexes (σ^q):



- K-chain: $K = \left\{ \sum_j w_j \sigma_j^q \right\}$

(Eckmann 1944; Goldberg 2002; Horak, Jost, 2013; Serrano, Gomze, 2019,…)

- Chain group: $C_q(K, \mathbb{Z}_2)$
- Boundary operator: $\partial_q : C_q(K) \to C_{q-1}(K)$

$$\partial_q \sigma^q = \sum_{j=0}^{q} (-1)^j \{v_0, v_1, \dots, \widehat{v_j}, \dots, v_q\}$$

- Adjoint boundary operator: $\partial_q^* : C_{q-1}(K) \to C_q(K)$
- q-combinatorial Laplacian operator: $\Delta_q = \partial_{q+1}\partial_{q+1}^* + \partial_q^*\partial_q$
- q-combinatorial Laplacian matrix: $\mathcal{L}_q = \mathcal{B}_{q+1}\mathcal{B}_{q+1}^T + \mathcal{B}_q^T\mathcal{B}_q$
- Betti numbers:

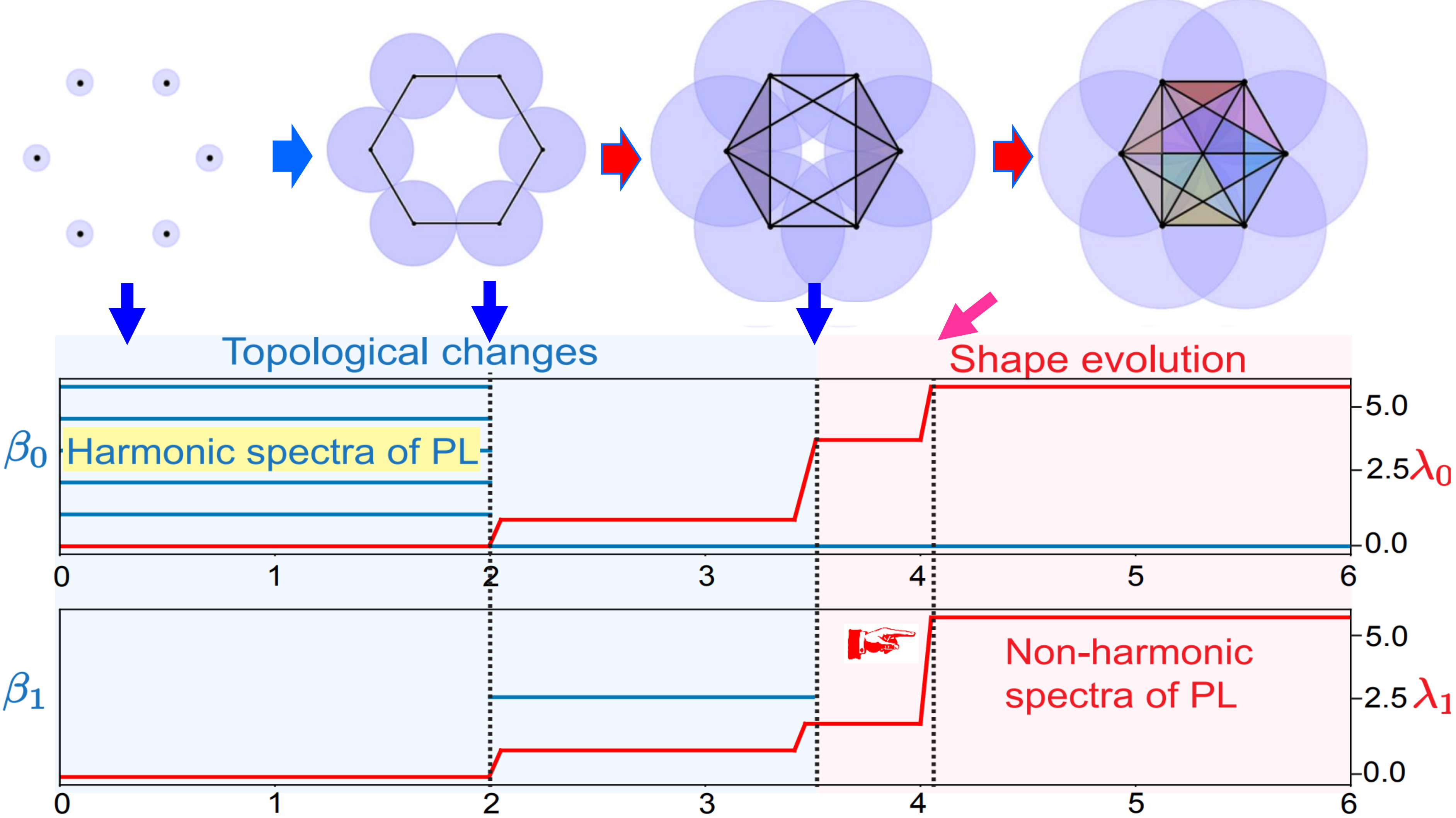
$\beta_q = \dim(\mathcal{L}_q(K)) - \mathrm{rank}\left(\mathcal{L}_q(K)\right)$ = # of zero eigenvalues of $\mathcal{L}_q(K)$

Persistent (Combinatorial) Laplacians

$$\mathcal{L}_q^{t+p} = \mathcal{B}_{q+1}^{t+p}\left(\mathcal{B}_{q+1}^{t+p}\right)^T + \left(\mathcal{B}_q^{t+p}\right)^T\mathcal{B}_q^{t+p}$$

Rui Wang

(Wang, Nguyen, Wei, 2019; Meng et al. 2021; Memoli et al. 2022; Liu and Wu 2023)



Alternative : Persistent Dirac by Maroulas and coworkers, Xia and coworkers

More in our toolbox for TDA

Evolutionary homology

Zixuan Cang, Munch, & Wei, J. Appl. Comput. Topology, 2020

Persistent sheaf Laplacians

Xiaoqi Wei, & Wei, Foundation of Data Science, 2025

Persistent path Laplacians

Rui Wang, & Wei, Foundation of Data Science, 2023

Persistent hyperdigraph Laplacians

Dong Chen, Liu, Wu, & Wei, Foundation of Data Science, 2024

Persistent Mayer Dirac

Suwayyid &Wei, J. Physics: Complexity, 2025

Persistent interaction topology

Jian Liu, Chen, & Wei, FoDS 2025

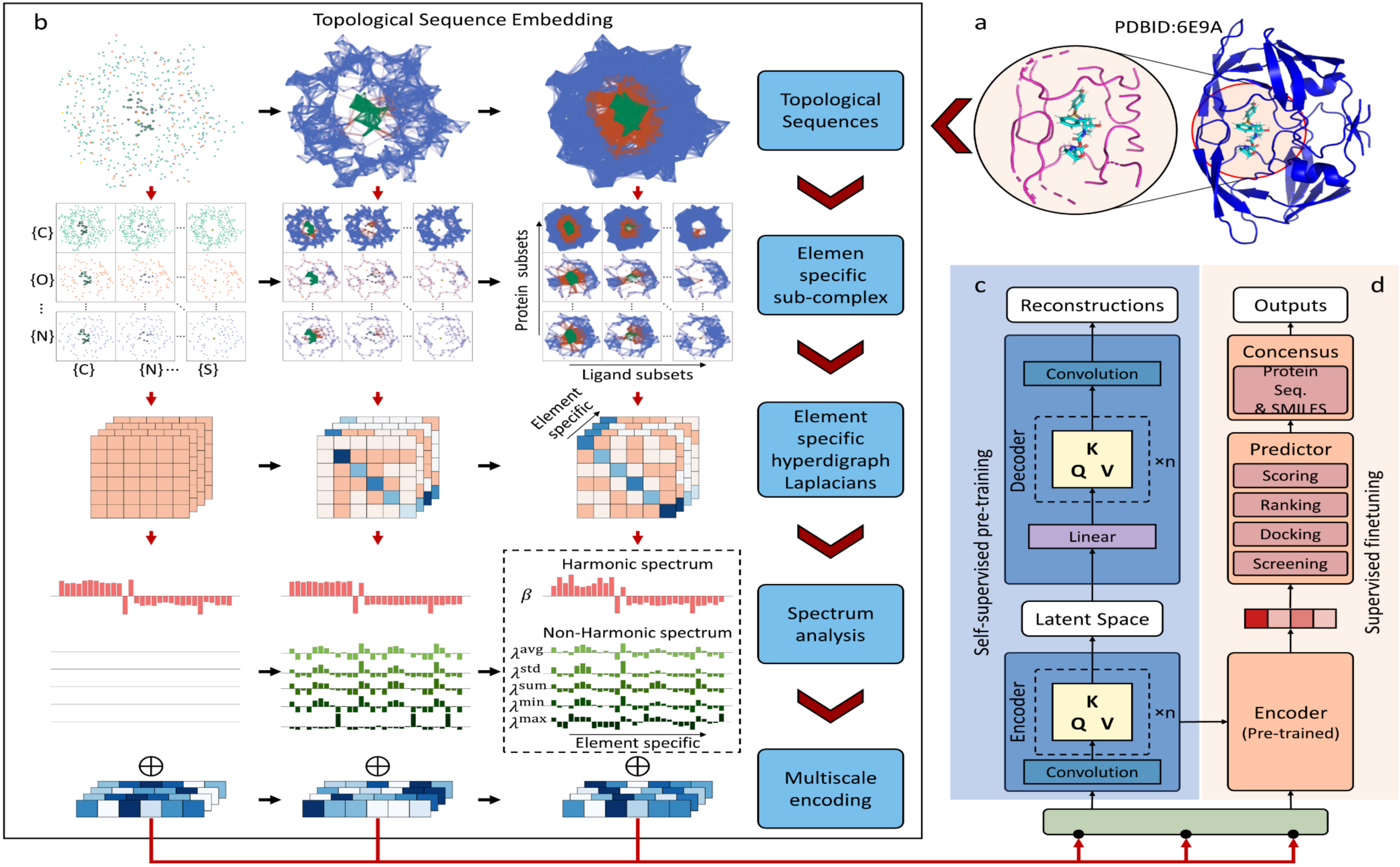
Persistent Mayer topology

Li Shen, Jian Liu, & Wei, Foundation of Data Science, 2024

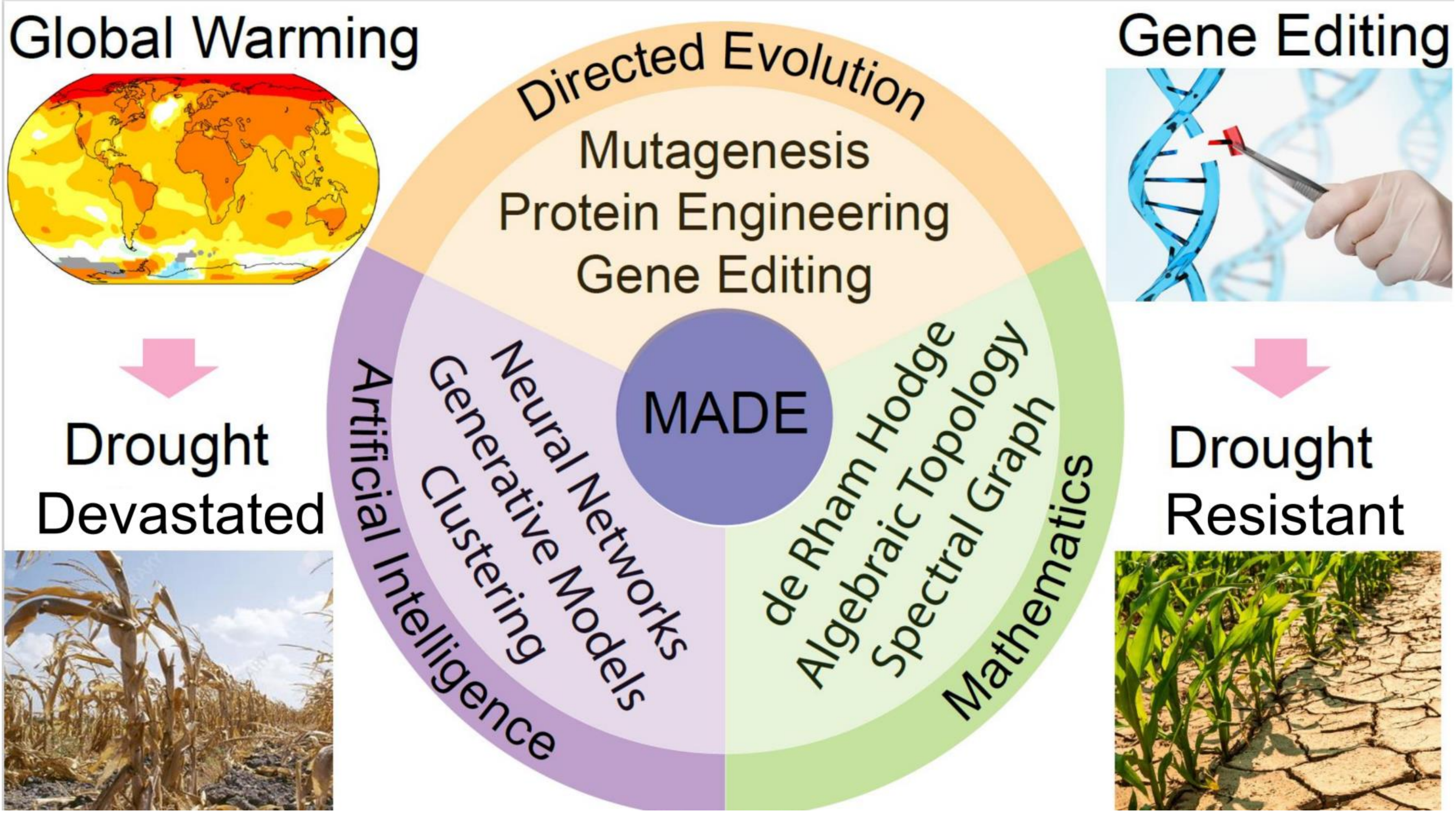
Topology-enabled Transformers (3D to 1D)

Chen, Liu, and Wei, Nature Machine Intelligence, 2024

Dong Chen



Mathematics-Assisted Directed Evolution (MADE)


Global Warming

Drought
Devastated

Directed Evolution

Mutagenesis
Protein Engineering
Gene Editing

MADE

Artificial Intelligence

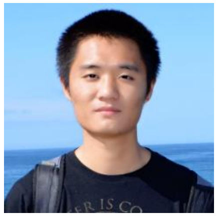
Neural Networks
Generative Models
Clustering

Mathematics

de Rham Hodge
Algebraic Topology
Spectral Graph

Gene Editing

Drought
Resistant


Dr. Yuchi Qiu

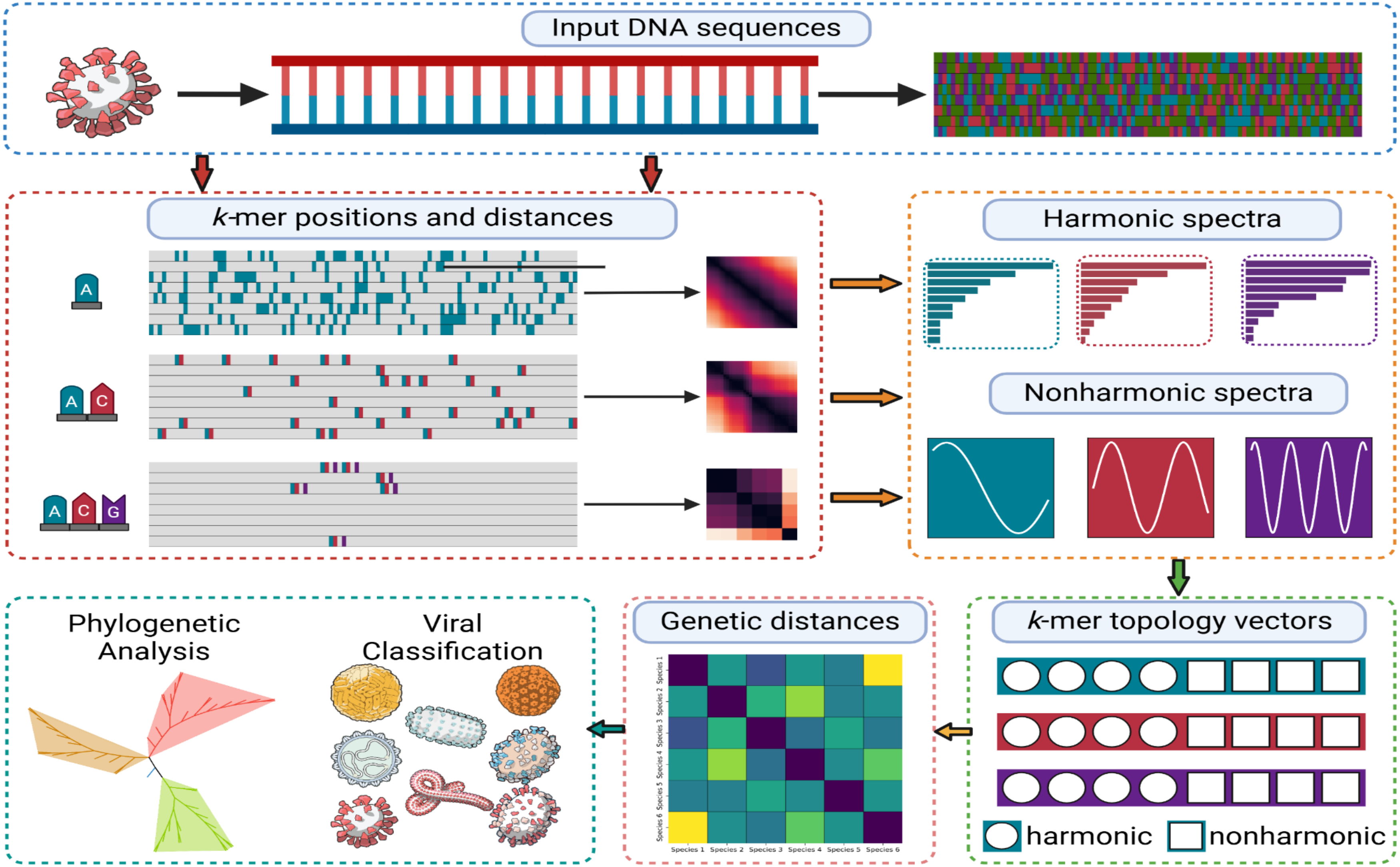
Impact of climate change on food security and global health

Y. Qiu, Hu, Wei, Nature Computational Science, 1(12) 809, 2021.
Y. Qiu, Wei, Nature Computational Science, 3, 149, 2023.

Revealing the Shape of Genome Space via K-mer Topology

Yuta Hozumi and GW Wei   arXiv:2412.20202, 2024

Yuta. Hozumi



Input DNA sequences

k-mer positions and distances

Harmonic spectra

Nonharmonic spectra

k-mer topology vectors

Genetic distances

Phylogenetic Analysis

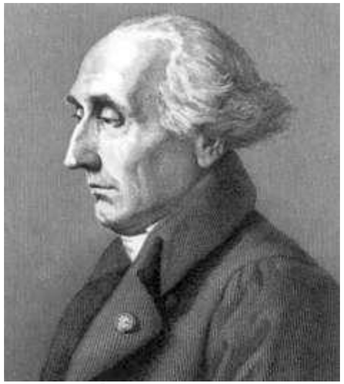
Viral Classification

○ harmonic ☐ nonharmonic

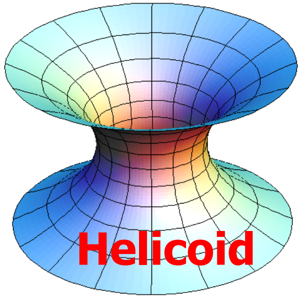
Inspired by Stephen Yau & coworkers

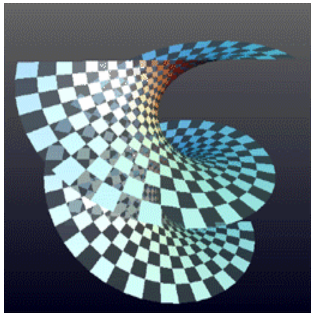
Differential geometry

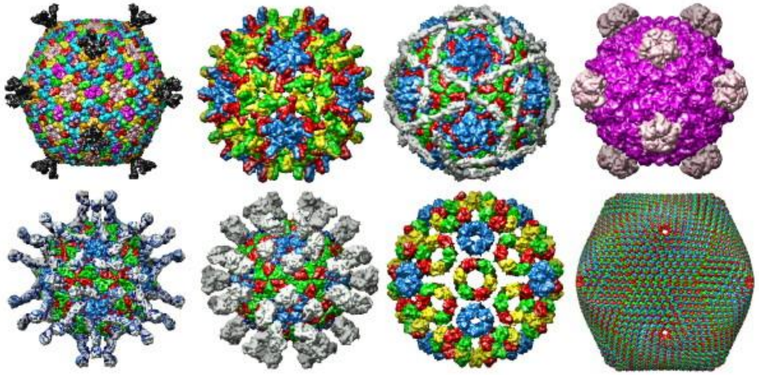
Leonhard P. Euler
(Swiss Mathematician, April 15, 1707 – Sept 18 1783



Joseph L. Lagrange (Italian Mathematician, January 25 1736 – April 10, 1813)



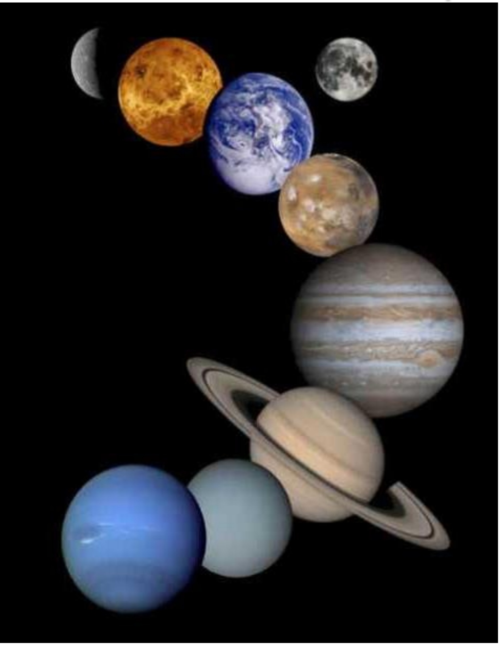


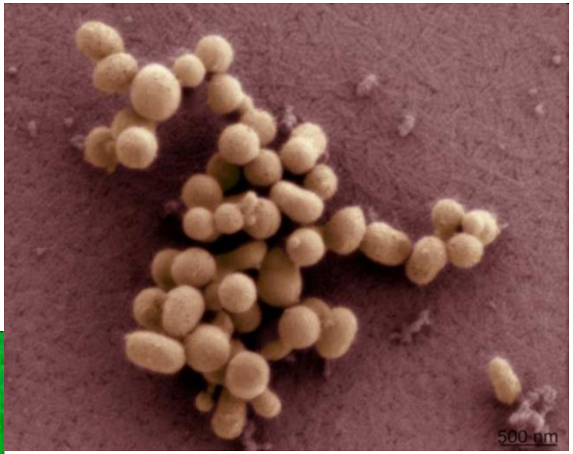


Viral morphology

Minimal Surfaces

A way to minimize energy and maximize stability

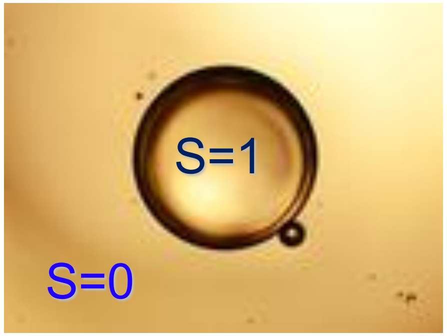




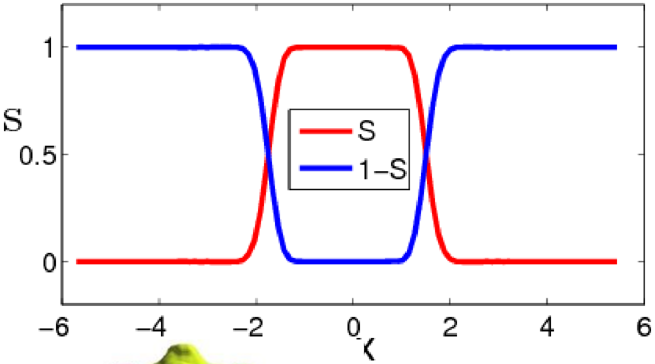
Man-made life, Mycoplasma mycoides

Differential geometry (DG) based multiscale model

$$G = \int \gamma[\text{area}]\, d\mathbf{r} \qquad \text{area} = |\nabla S|$$

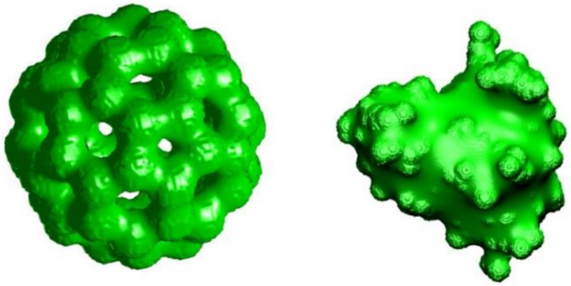


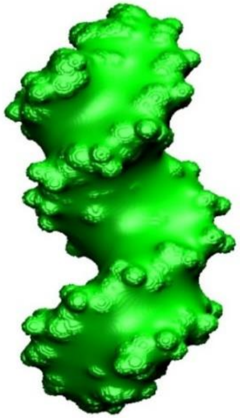
where G is the surface energy, gamma (γ) is the surface tension, and S is a surface characteristic function:

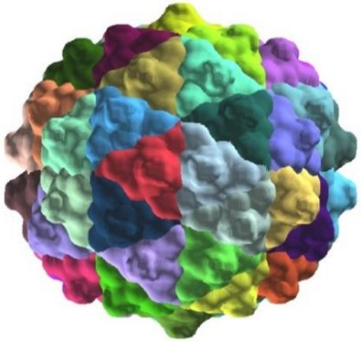


Generalized Laplace-Beltrami flow:

$$\frac{\partial S}{\partial t} = |\nabla S| \underbrace{\left[\nabla \cdot \frac{\gamma \nabla S}{|\nabla S|}\right]}_{\text{Mean curvature}}$$







Shan Zhao

DG-based Poisson-Boltzmann Nernst Planck model

(Bates, Wei, Zhao, 2006; JCC,2008; Wei et al., SIAM Review 2012)

Differential topology

George de Rham 1903-1990

William Hodge 1903-1973

Marston Morse 1892-1977

Heinz Hopf 1894-1971

Henri Poincaré 1854-1912

- De Rham cohomology
- Hodge decomposition
- Morse theory

…

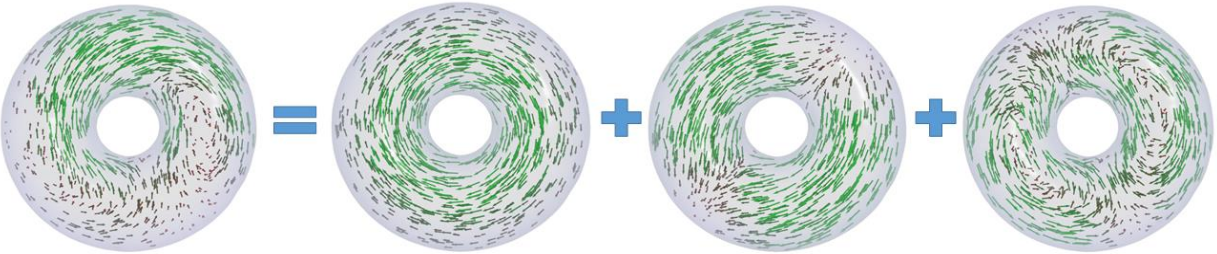
- Discrete exterior calculus
- Finite element exterior calculus

Crane and Segerman

Differential Topology (De Rham-Hodge theory)

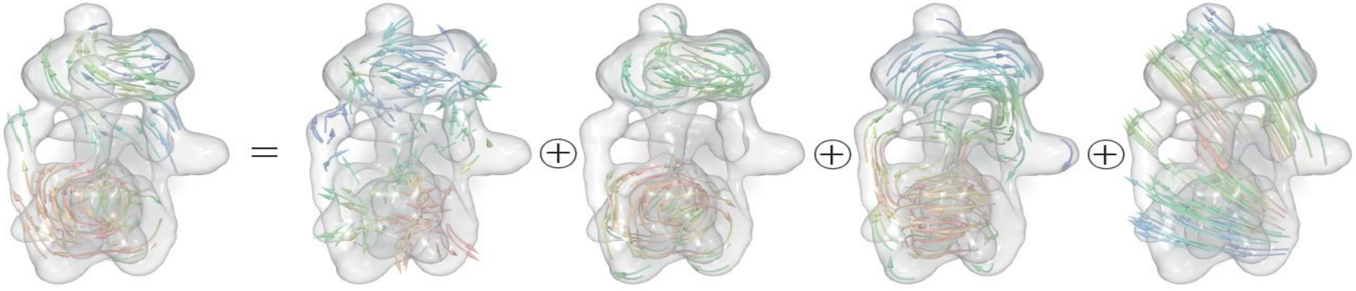
Hodge decomposition:

(Zhao, Wang, Chen, Tong & Wei, BMB, 2020)



A vector field = Harmonic + curl-free + divergent-free

Cryo-EM data:



Input = Normal Gradient ⊕ Tangential Curl ⊕ Tangential Harmonic ⊕ Central Harmonic

---

$$\begin{tikzcd}
\text{0-Form} \arrow[r, "d_0"] \arrow[d, shift right, "\star_0"'] & \text{1-Form} \arrow[r, "d_1"] \arrow[d, shift right, "\star_1"'] & \text{2-Form} \arrow[r, "d_2"] \arrow[d, shift right, "\star_2"'] & \text{3-Form} \arrow[d, shift right, "\star_3"'] \\
\text{Dual 3-Form} \arrow[u, shift right, "\star_0^{-1}"'] & \text{Dual 2-Form} \arrow[l, "d_0^T"] \arrow[u, shift right, "\star_1^{-1}"'] & \text{Dual 1-Form} \arrow[l, "d_1^T"] \arrow[u, shift right, "\star_2^{-1}"'] & \text{Dual 0-Form} \arrow[l, "d_2^T"] \arrow[u, shift right, "\star_3^{-1}"']
\end{tikzcd}$$

$$\begin{tikzcd}
\text{point-based scalar field} \arrow[r, "\nabla"] \arrow[d, leftrightarrow, "\star"] & \text{edge-based vector field} \arrow[r, "\nabla\times"] \arrow[d, leftrightarrow, "\star"] & \text{face-based vector field} \arrow[r, "\nabla\cdot"] \arrow[d, leftrightarrow, "\star"] & \text{cell-based scalar field} \arrow[d, leftrightarrow, "\star"] \\
\text{cell-based scalar field} & \text{face-based vector field} \arrow[l, "\nabla\cdot"'] & \text{edge-based vector field} \arrow[l, "\nabla\times"'] & \text{point-based scalar field} \arrow[l, "\nabla"']
\end{tikzcd}$$

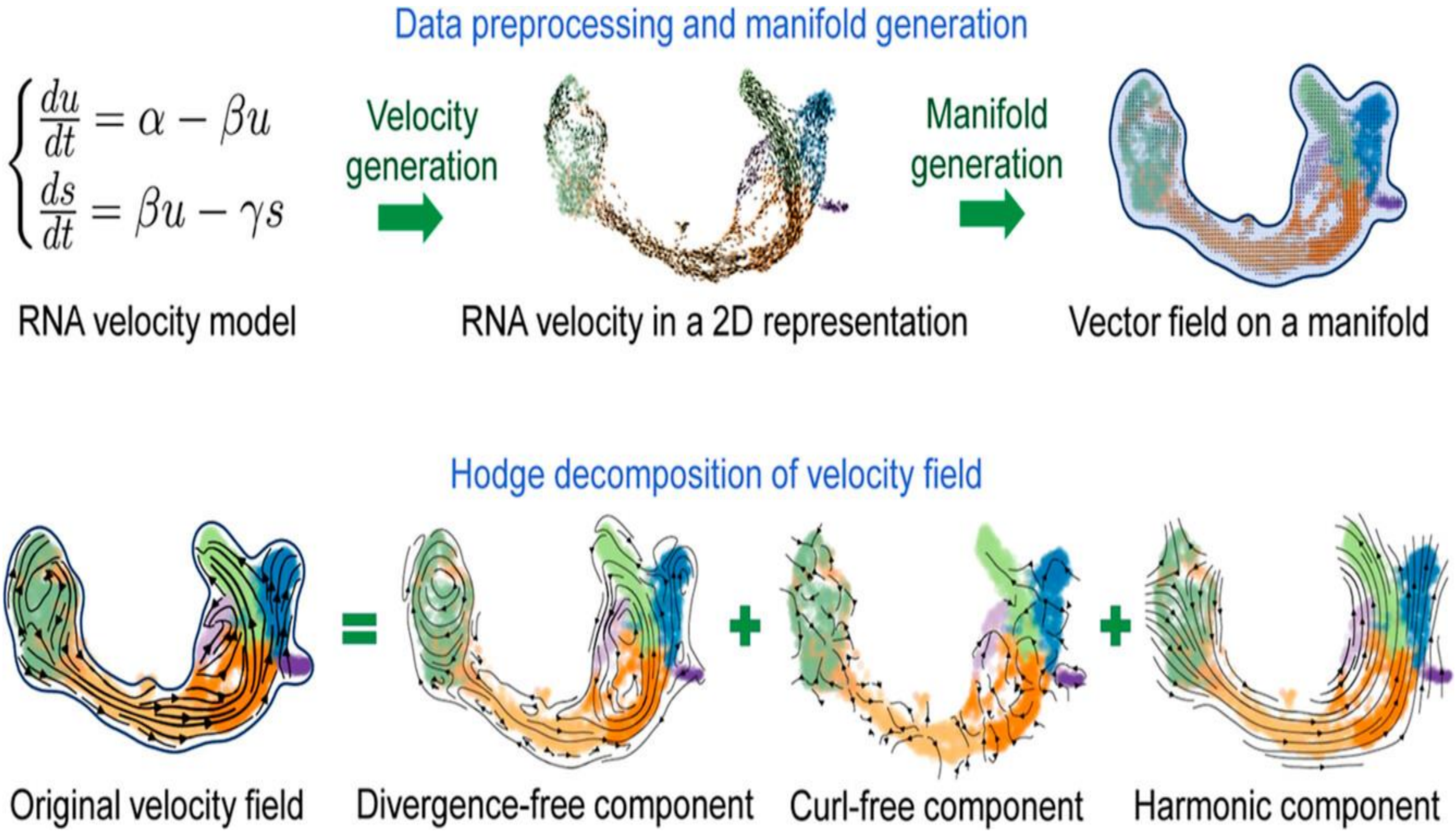
(Douglas Arnold, M Desbrun, AN Hirani, …)

Hodge Decomposition of Single-Cell RNA Velocity

Su, Tong and Wei

J. Chem. Inf. Model. 2024, 64, 8, 3558–3568

Dr. Zhe Su



Data preprocessing and manifold generation

$$\begin{cases} \dfrac{du}{dt} = \alpha - \beta u \\ \dfrac{ds}{dt} = \beta u - \gamma s \end{cases}$$

Velocity generation

Manifold generation

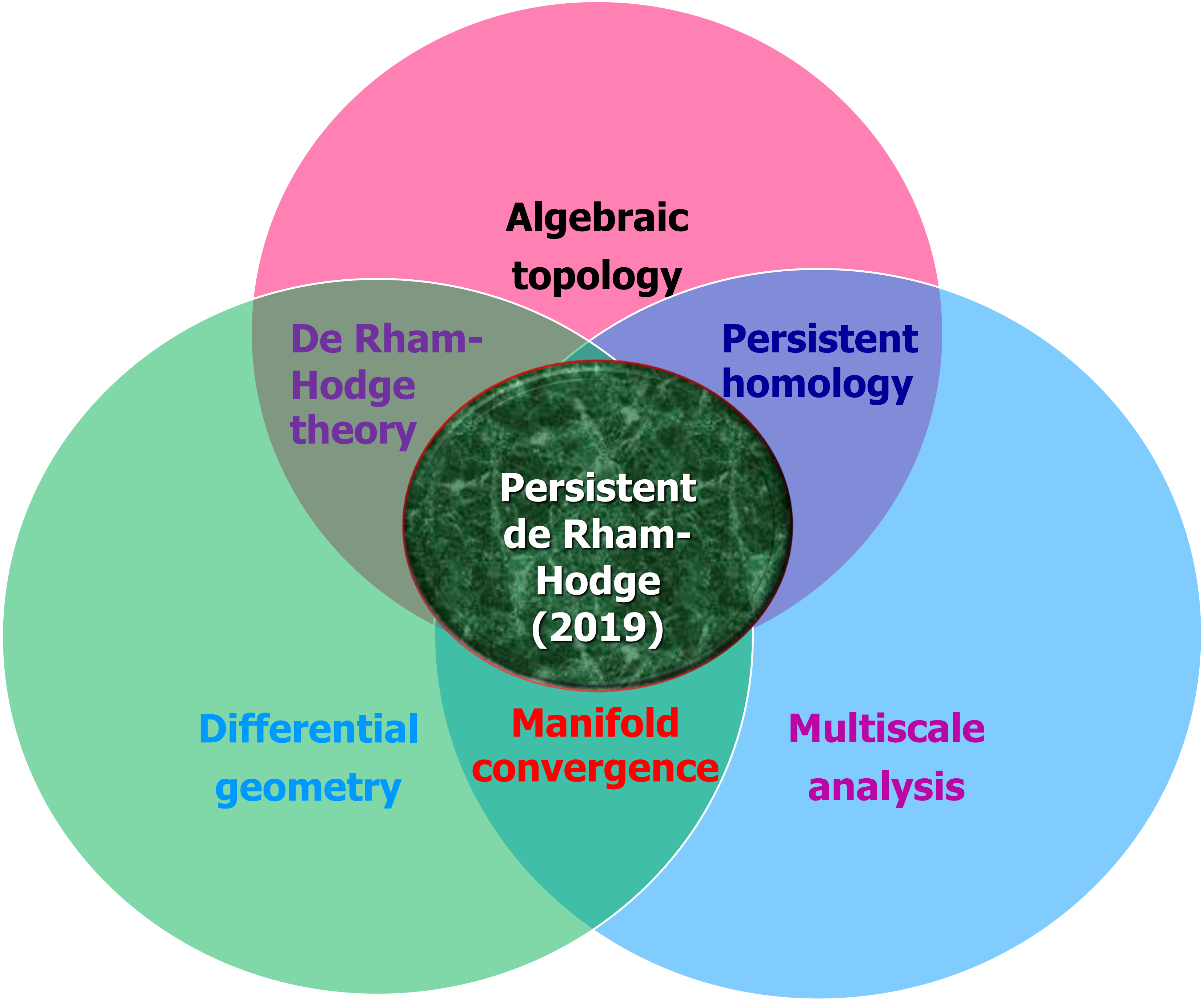
RNA velocity model

RNA velocity in a 2D representation

Vector field on a manifold

Hodge decomposition of velocity field

Original velocity field = Divergence-free component + Curl-free component + Harmonic component



Algebraic topology

De Rham-Hodge theory

Persistent homology

Persistent de Rham-Hodge (2019)

Differential geometry

Manifold convergence

Multiscale analysis

Evolutionary de Rham-Hodge (Differential Topology)

Zenghouyi Chime Bells (433 BC in China)

Ribando-Gros, et al. SIAM Review 2024

Chen, Zhao, Tong & Wei, DCDS-B, 2020

Manifold evolution

$$M_0 \xrightarrow{\mathfrak{I}_{0,1}} M_1 \xrightarrow{\mathfrak{I}_{1,2}} M_2 \xrightarrow{\mathfrak{I}_{2,3}} \cdots \xrightarrow{\mathfrak{I}_{n-1,n}} M_n \xrightarrow{\mathfrak{I}_{n,n+1}} M$$

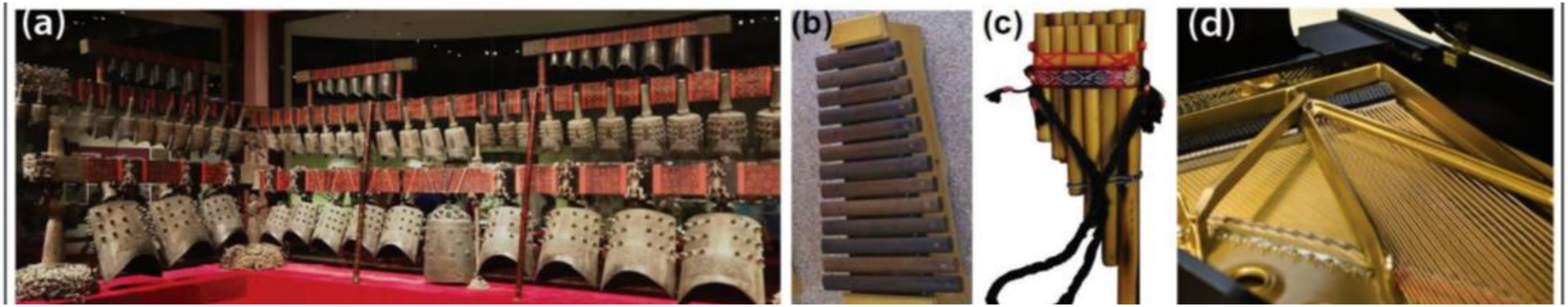
Filtration-induced de Rham complexes:

$$\begin{tikzcd}
\Omega_n^0(M_0) \arrow[r, "d^0"] \arrow[d, "\mathfrak{E}_{0,}"] & \Omega_n^1(M_0) \arrow[r, "d^1"] & \Omega_n^2(M_0) \arrow[r, "d^2"] & \Omega_n^3(M_0) \arrow[d, "\mathfrak{E}_{0,1}"] \\
\Omega_n^0(M_1 \arrow[d, "\mathfrak{E}_{1,}"] & & & \Omega_n^3(M_1) \arrow[d, "\mathfrak{E}_{1,1}"] \\
\Omega_n^0(M_2 \arrow[d, "\mathfrak{E}_{2,1}"] & \arrow[d, "\mathfrak{E}_{2,1}"] & \arrow[d, "\mathfrak{E}_{2,1}"] & \Omega_n^3(M_2) \arrow[d, "\mathfrak{E}_{2,1}"] \\
\cdots & \cdots & \cdots & \cdots
\end{tikzcd}$$

Persistent Hodge Laplacians:

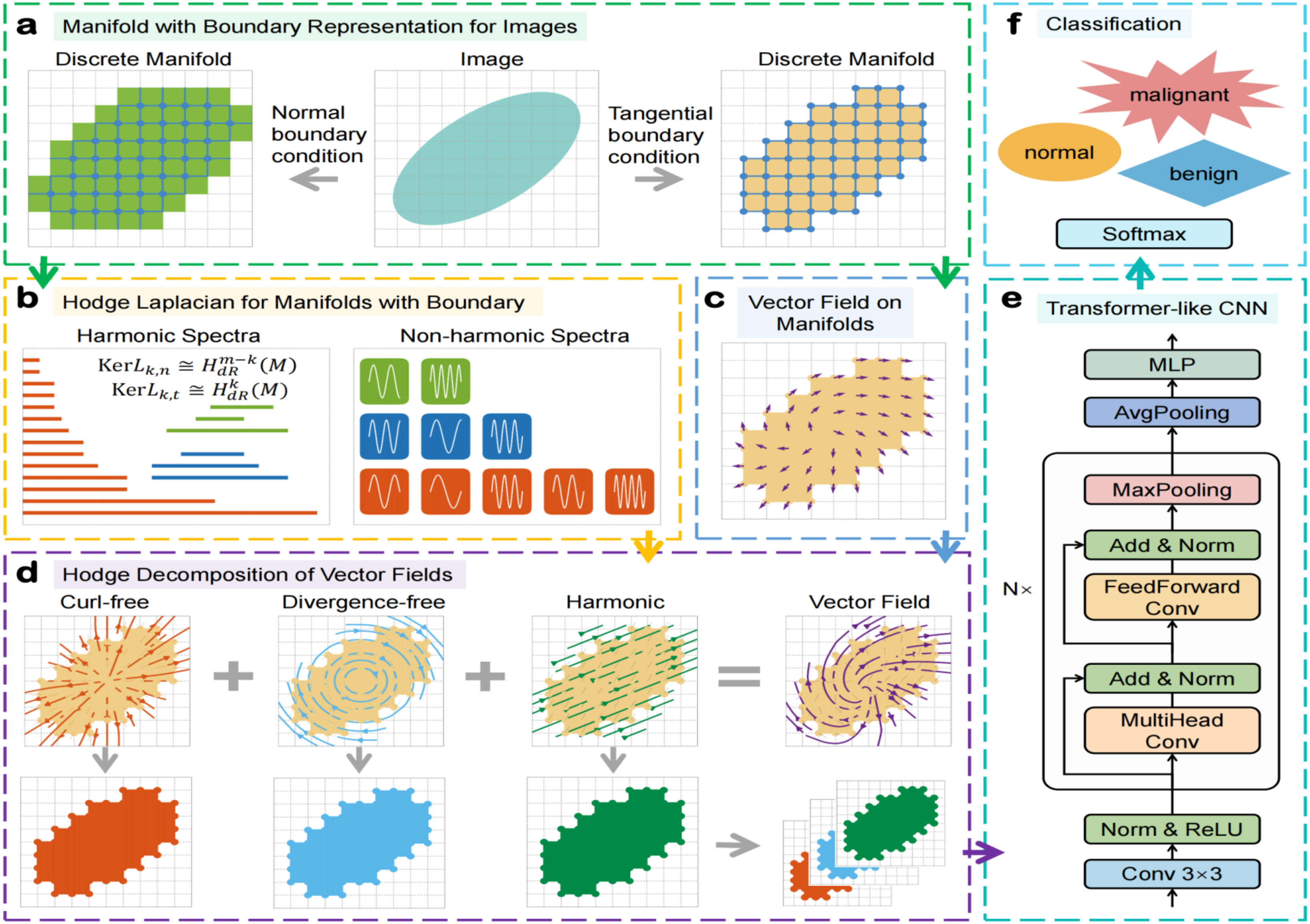
$$\Delta_k^{l,p} = \partial_{k+1}^{l} d_k^{l} + d_{k-1}^{l+p} \partial_k^{l+p}$$

Su, Tong & Wei, Mathematics, 2024



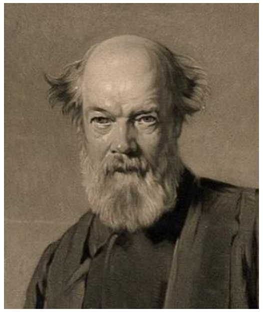
Medical image classification using differential topology

Liu, Wang, Song, Su, & Wei, Nat Comm., Under rev.



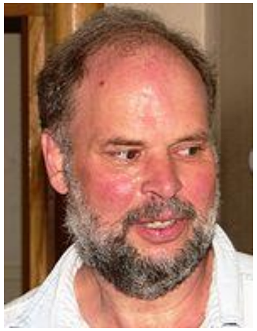
Xiang Liu

Geometric topology



Peter Tait
1831-1901

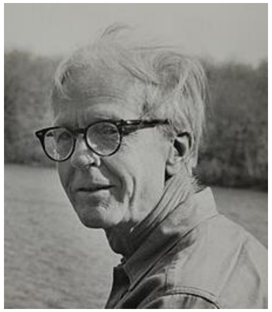
Kurt
Reidemeister
1893-1971



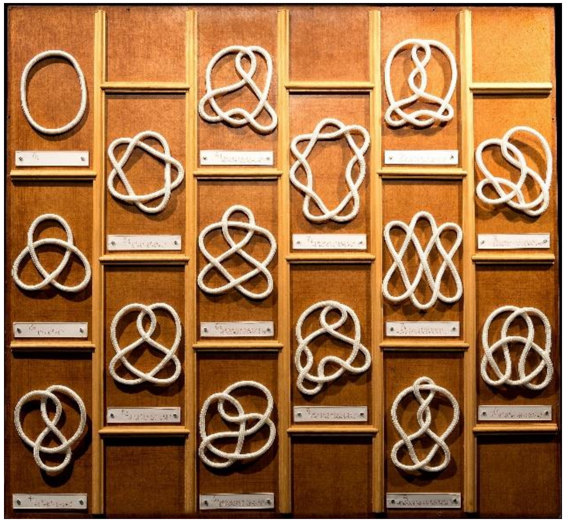
Vaughan Jones
1952-2020

Mikhail
Khovanov
1972-

Henri
Poincaré
1854-1912



Hassler
Whitney
1907-1989

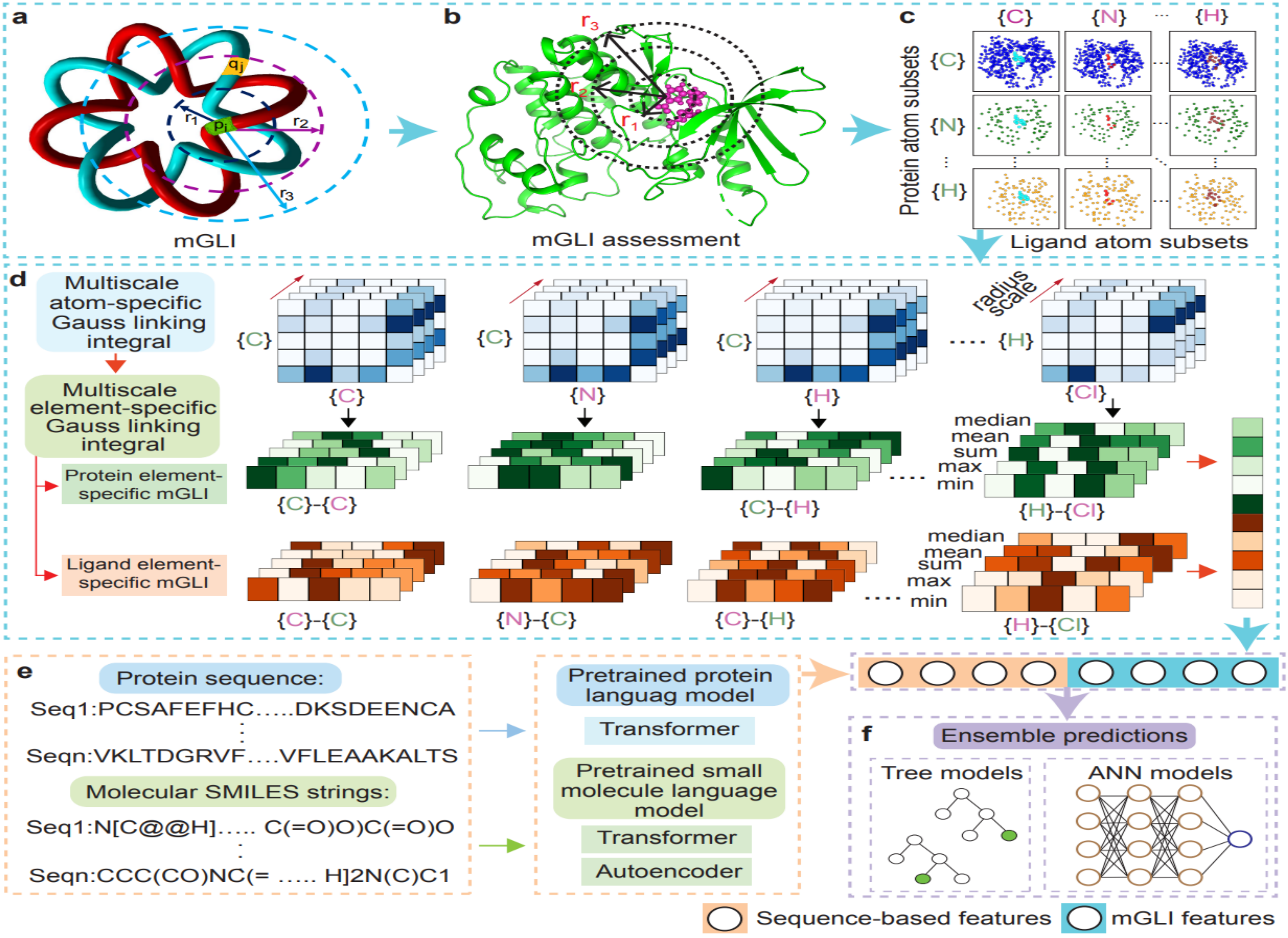


Matemateca

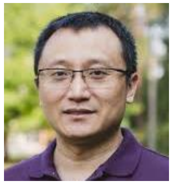
Book of Kells

Multiscale Gauss link integral

Shen, Feng, Li, Lei, Wu, & Wei, PNAS, 2024



Li Shen

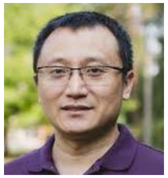


Hongsong Feng

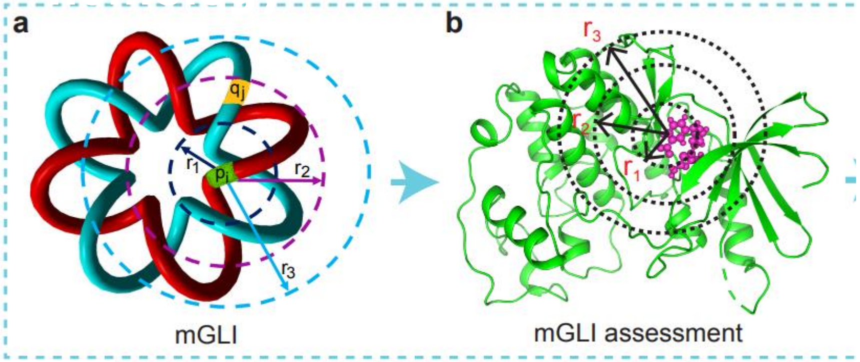
E Panagiotou, LH Kauffman,…

Geometric Topology (1D curve in 3-space)


Li Shen

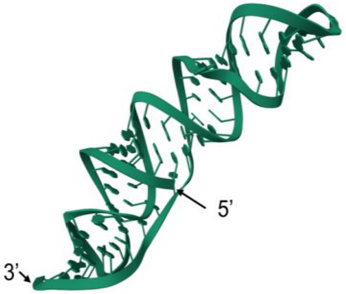

H Feng

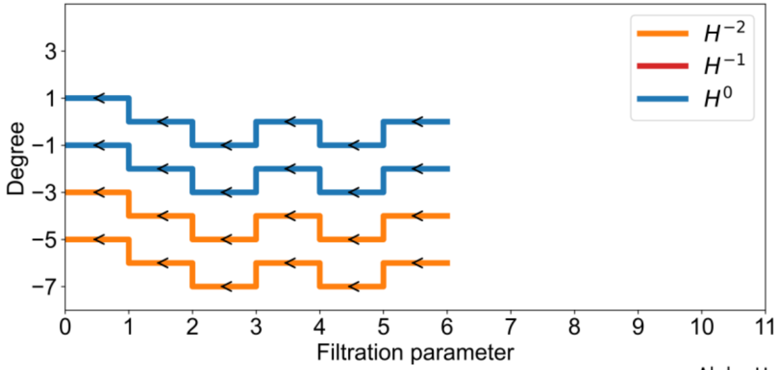
Multiscale Gauss link integral



Shen, Feng, Li, Lei, Wu, & Wei, PNAS, 2024

Persistent Khovanov homology





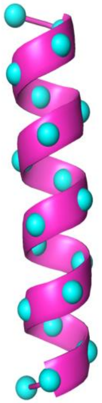

Jian Liu

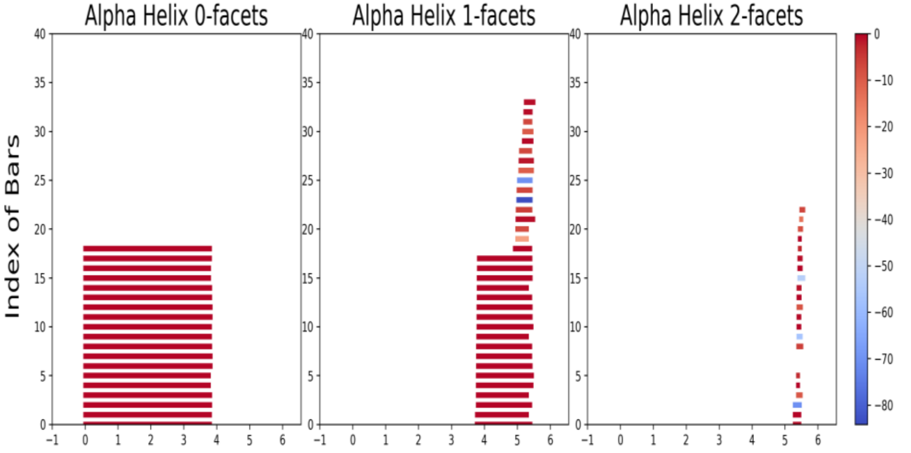

Li Shen

Shen et al. AIMS Mathematics, 2024

Persistent Jones Polynomial

Ruzhi Song, Fengling Li, Jie Wu, Fengchun Lei, and GW Wei, Mathematics 2025



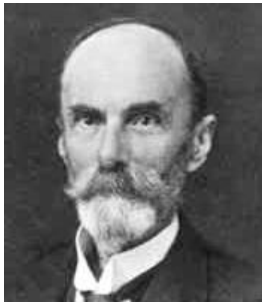


Commutative Algebra

Wolfgang Krull
(1899 – 1971)

Oscar Zariski
(1899 – 1986)

Melvin Hochster
(1943 --)



Francis Macaulay
(1862 –1937)

Richard Stanley
(1944-)

Persistent Commutative Algebra (Suwayyid and Wei, FoDS, 2025)

- Introduce multiscale algebraic invariants to data analysis
- Integrate commutative algebra, algebraic topology, algebraic geometry, and combinatorics with machine learning.
- Offer a new rational learning platform.

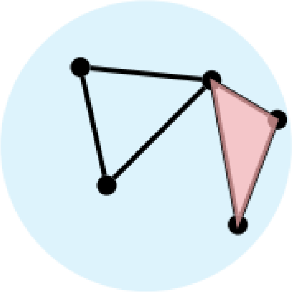
Stanley-Reisner rings and ideals

Faisal Suwayyid & Wei, Foundations of Data Science, 2025.

- Given a field *k*, define the **polynomial ring:**
  $S = k[x_1, x_2, \dots, x_n]$ on the vertex set V.

- A simplicial complex Δ^t induces the **Stanley–Reisner ideal (**i.e., **non-faces** of Δ^t**)**:
  $$I^t(\Delta^t) := \langle x_{i_1}, \dots, x_{i_r} | \{x_{i_1}, \dots, x_{i_r}\} \notin \Delta^t \rangle$$

- This ideal can also be expressed by **minimal facet ideals** P_σ :
  $$I^t = \bigcap_{\sigma \in \mathcal{F}(\Delta^t)} \mathrm{P}_\sigma , \qquad \text{where,} \quad \mathrm{P}_\sigma := (x_i | x_i \notin \sigma)$$
  with $\mathcal{F}(\Delta^t)$ denoting the facets (maximal faces) of Δ^t.



$$\rightarrow \quad I^t = \cap \ \mathrm{P}_\sigma$$

Persistent Facet Ideals

Faisal Suwayyid & Wei, Foundations of Data Science, 2025.

- For a fixed dimension i, define the set

$$\mathcal{P}_i^t = \{\mathrm{P}_\sigma \mid \dim(\sigma) = i\}.$$

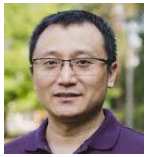
- The **persistent prime ideals (or persistent facet ideals)** of dimension i between two-time steps t and t' are given by the intersection:

$$\mathcal{P}_i^{t,t'} := \mathcal{P}_i^t \cap \mathcal{P}_i^{t'}$$

This captures the common i-dimensional combinatorial features that persist across the filtration levels.

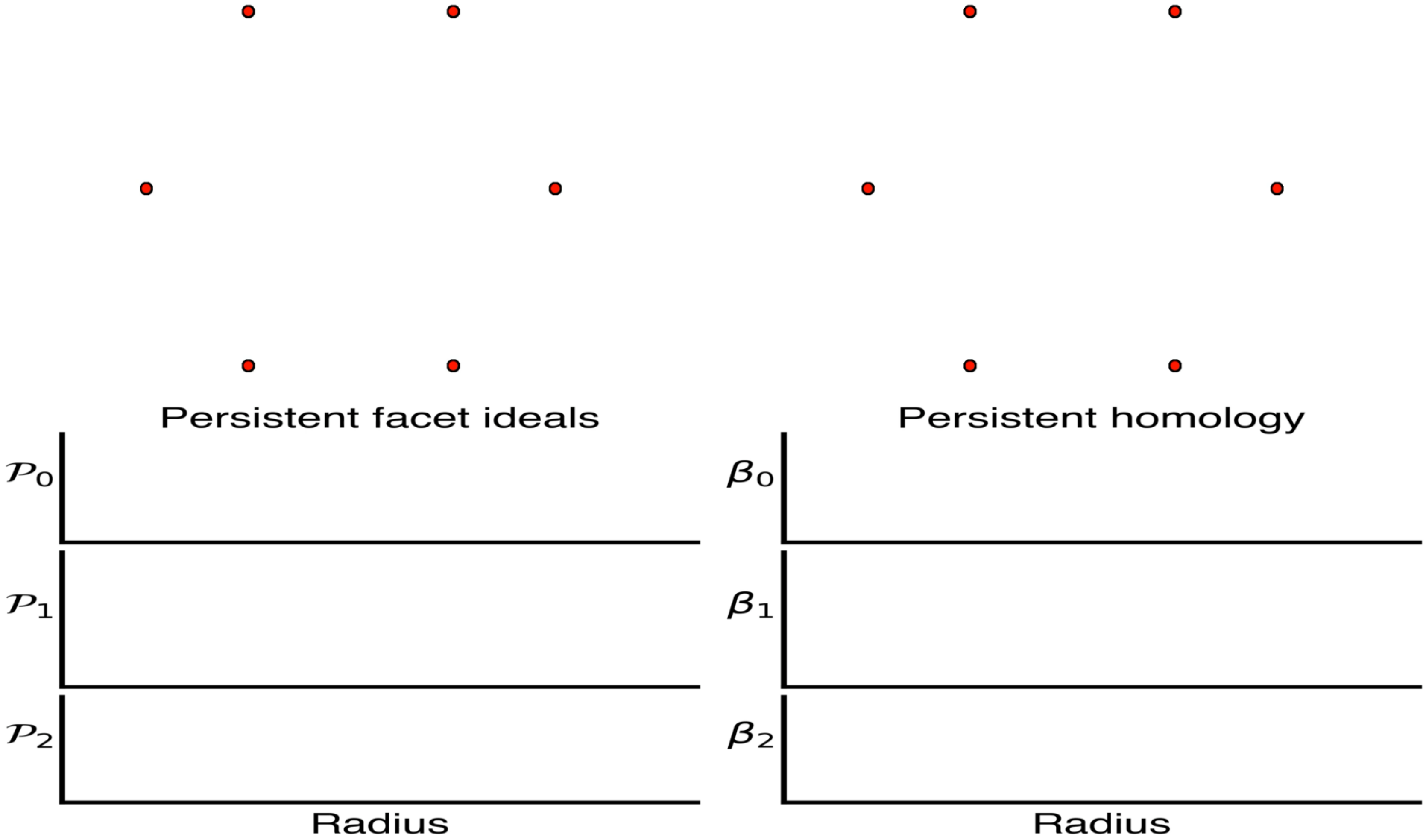
- The **facet persistence Betti number** $\beta_i^{t,t'} := |\mathcal{P}_i^t \cap \mathcal{P}_i^{t'}|$ counts the number of such persistent prime ideals, offering an algebraic measure of the topological stability in dimension i.

Comparison between Persistent Facet Ideals and Persistent Homology



Feng et al., JCIM, 2025

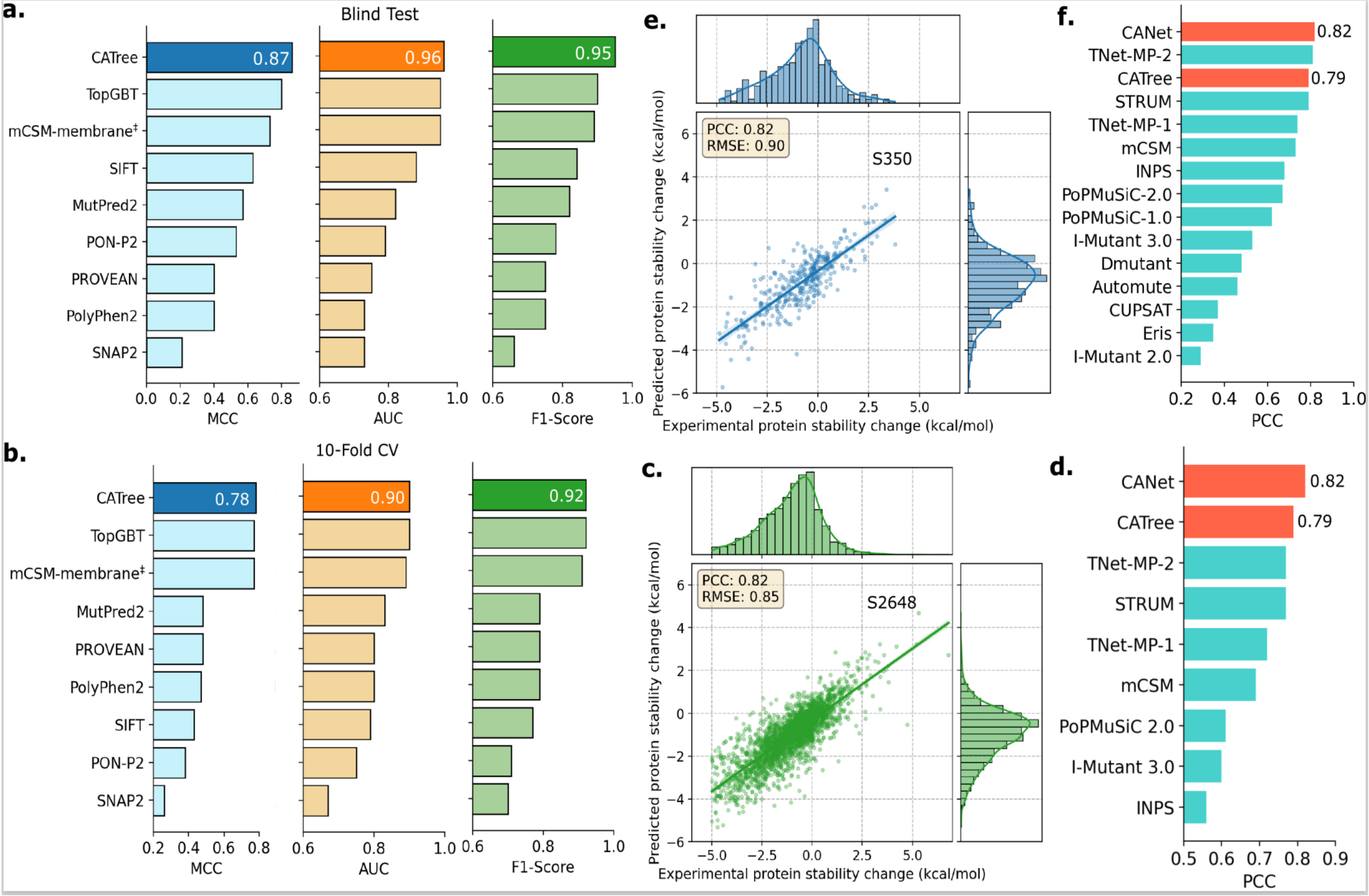
Hongsong Feng



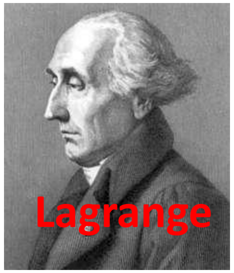
Commutative Algebra Tree/Net

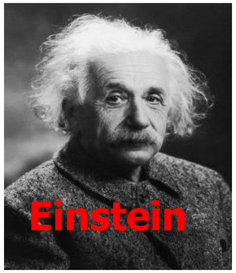
Protein mutations

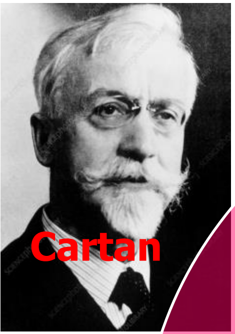
JunJie Wee



Our Strategy





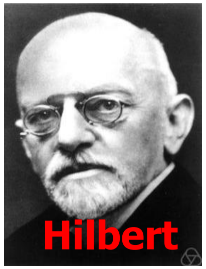


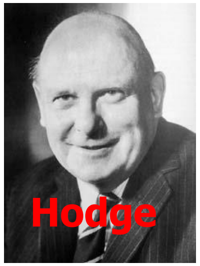
Algebraic topology
Geometric topology
Differential geometry
Nonlinear algebra
A.C.G.S.T. graphs
Variational PDEs

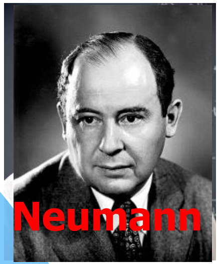
Sequence data
Structure data
Omics data
Biophysics
Bioinformatics
Systems biology
Systems pharmacology

Data-driven biological discovery

Machine learning
Deep learning
Manifold learning
Transformer
Autoencoder
Generative AI
Large language model







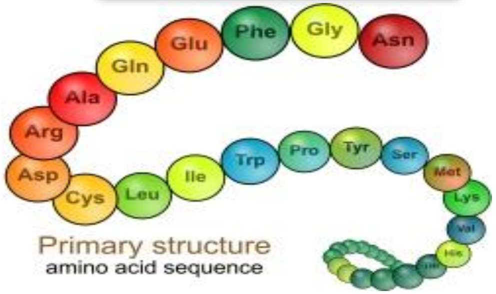
Drug Design Data Resource (D3R) Grand Challenges

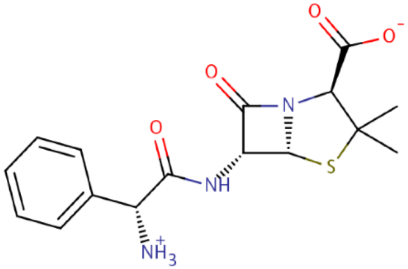
- Funded in part by National Institute of General Medical Sciences
- Hosted at the University of California, San Diego
- Annually 2015-2019

GC4 (2018 -2019): 55 research groups

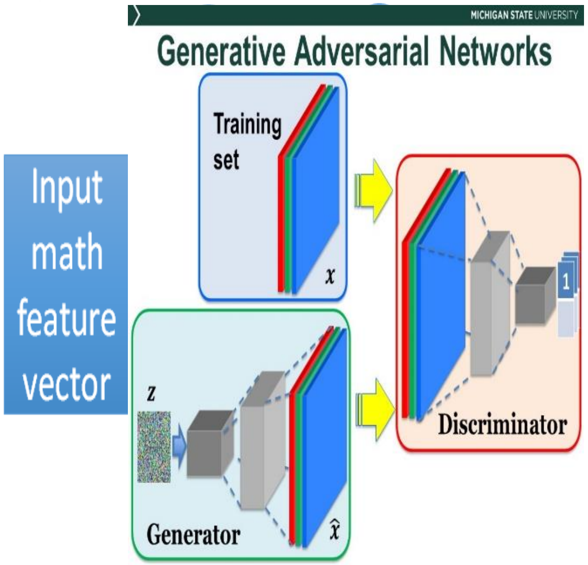
Drug Design Data Resource (D3R) Grand Challenge

Given data

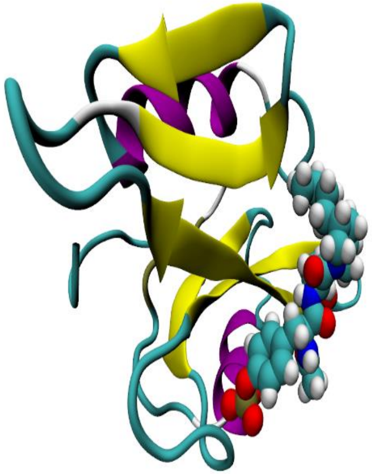


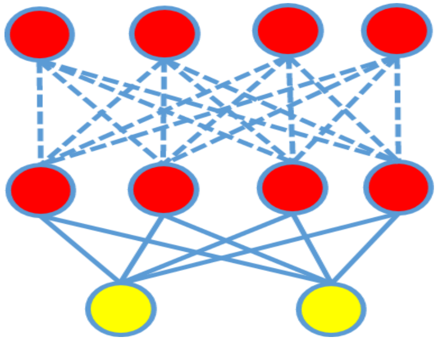


Math based GAN

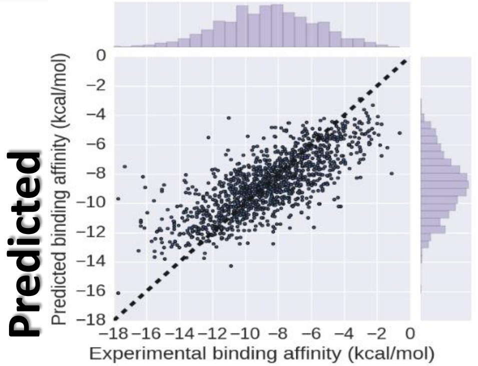


Predicted complex

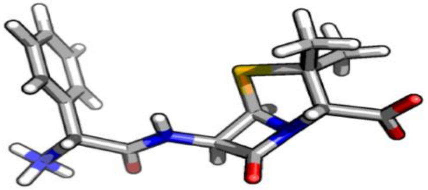




Final predictions to be compared with experiments



Experimental



Drug pose

(Nguyen et al, JCAMD, 2018)

Our performance in D3R Grand Challenges, worldwide competitions in computer-aided drug design, 2016-2019.

D3R Grand Challenge 4 (2018-2019)

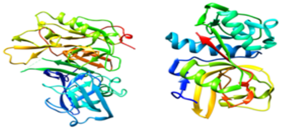
Pose Predictions

BACE Stage 1A
Pose Predictions (Partials) 🥇 2/3 🥈 2/3

BACE Stage 1B
Pose Prediction (Partials) 🥈 2/2 🥉 1/2

Affinity Predictions

Cathepsin Stage 1
Combined Ligand and Structure Based Scoring 🥇 2/5 🥈 2/3 🥉 2/4
Ligand Based Scoring (No participation)
Structure Based Scoring 🥇 2/4 🥈 3/3 🥉 3/3
Free Energy Set 🥇 1/7 🥈 1/7 🥉 2/5



BACE Stage 1
Combined Ligand and Structure (No participation)
Ligand Based Scoring(Partials) (No participation)
Structure Based Scoring(Partials)(No participation)
Free Energy Set (No participation)

BACE Stage 2
Combined Ligand and Structure
Ligand Based Scoring(No participation)
Structure Based Scoring (Partials)
Free Energy Set 🥈 3/4 🥉 1/4

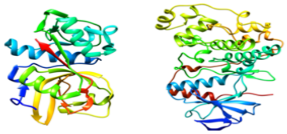
D3R Grand Challenge 3 (2017-2018)

(Nguyen et al, JCAMD, 2018)

Pose Prediction

Cathepsin Stage 1A
Pose Predictions (partials)

Cathepsin Stage 1B
Pose Prediction



Affinity Rankings excluding Kds > 10 μM

Cathepsin Stage 1
Scoring (partials)
Free Energy Set

Cathepsin Stage 2
Scoring (partials)
Free Energy Set

VEGFR2
Scoring (partials)

JAK2 SC3
Scoring
Free Energy Set 🥇 4/4 🥈 4/4 🥉 4/4

JAK2 SC2
Scoring (partials)

TIE2
Scoring 🥇 3/3 🥈 1/2
Free Energy Set 2 🥇 4/4 🥈 5/5 🥉 4/4

p38-α
Scoring

ABL1
Scoring (partials) 🥇 2/4 🥈 4/5 🥉 2/3

Active / Inactive Classification

VEGFR2
Scoring (partials)

JAK2 SC3
Scoring
Free Energy Set 🥇 1/5 🥈 1/4 🥉 1/2

JAK2 SC2
Scoring (partials)

TIE2
Scoring (partials) 🥇 1/5 🥈 1/4 🥉 1/2
Free Energy Set 1 🥇 1/1 🥈 4/5 🥉 1/1

p38-α
Scoring (partials)

ABL1
Scoring (partials)

Affinity Rankings for Cocrystalized Ligands

Cathepsin Stage 1
Scoring (partials)
Free Energy Set 🥇 3/17 🥉 9/17

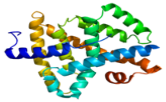
Cathepsin Stage 2 🥇 2/2 🥈 2/2
Scoring (partials)
Free Energy Set 🥇 19/44 🥈 3/20 🥉 1/4

D3R Grand Challenge 2 (2016-2017)

Given: Farnesoid X receptor (FXR) and 102 ligands
Tasks: Dock 102 ligands to FXR, and predict their poses, binding free energies and energy ranking

Stage 1
Pose Predictions (partials)
Scoring (partials)
Free Energy Set 1 (partials)
Free Energy Set 2 (partials)

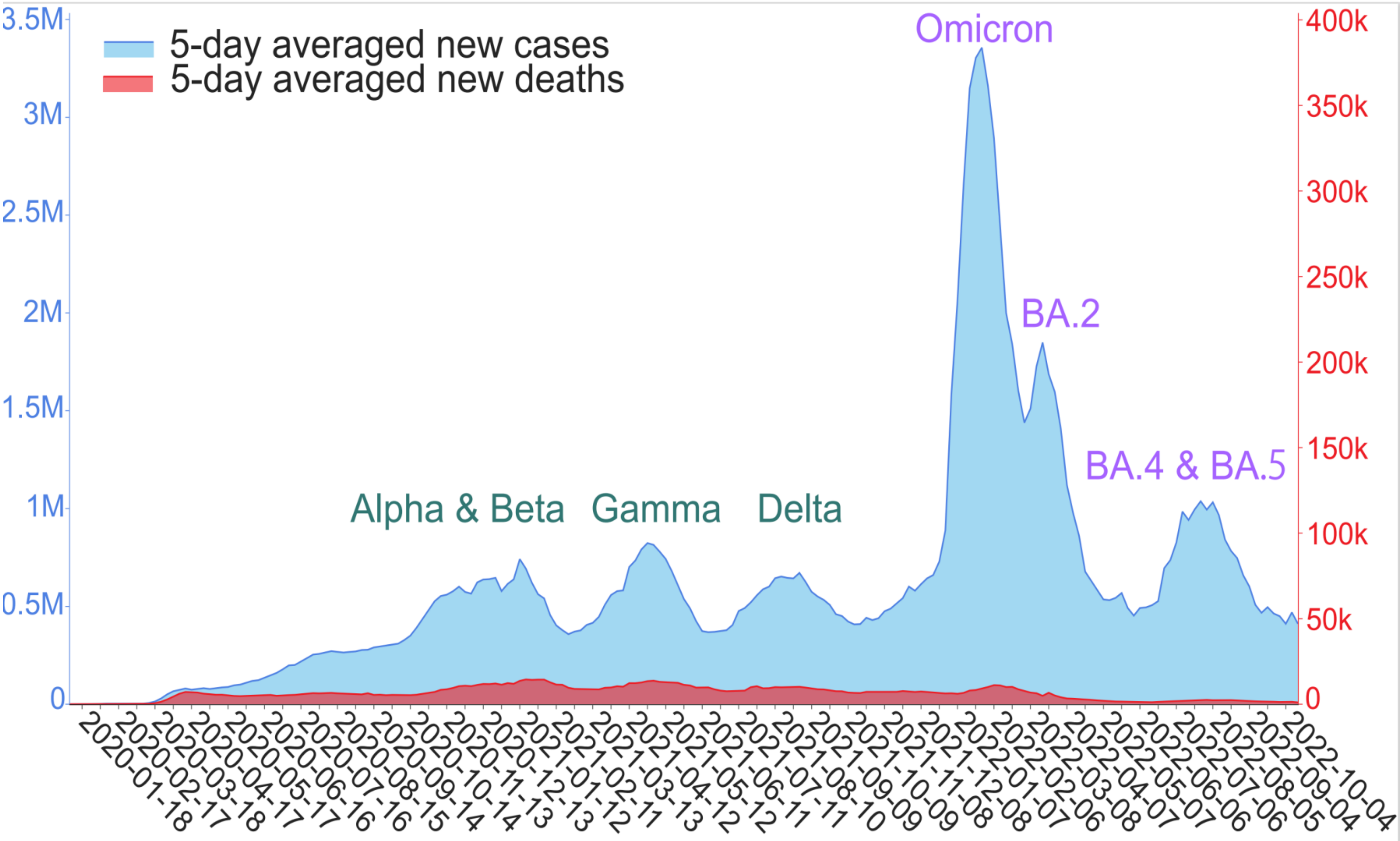
Stage 2
Scoring (partials)
Free Energy Set 1 (partials) 🥇 1/3 🥈 2/2 🥉 2/2
Free Energy Set 2 (partials)



We made many people upset at AI/ML/DL as noticed by Pedro Ballester (ICL).

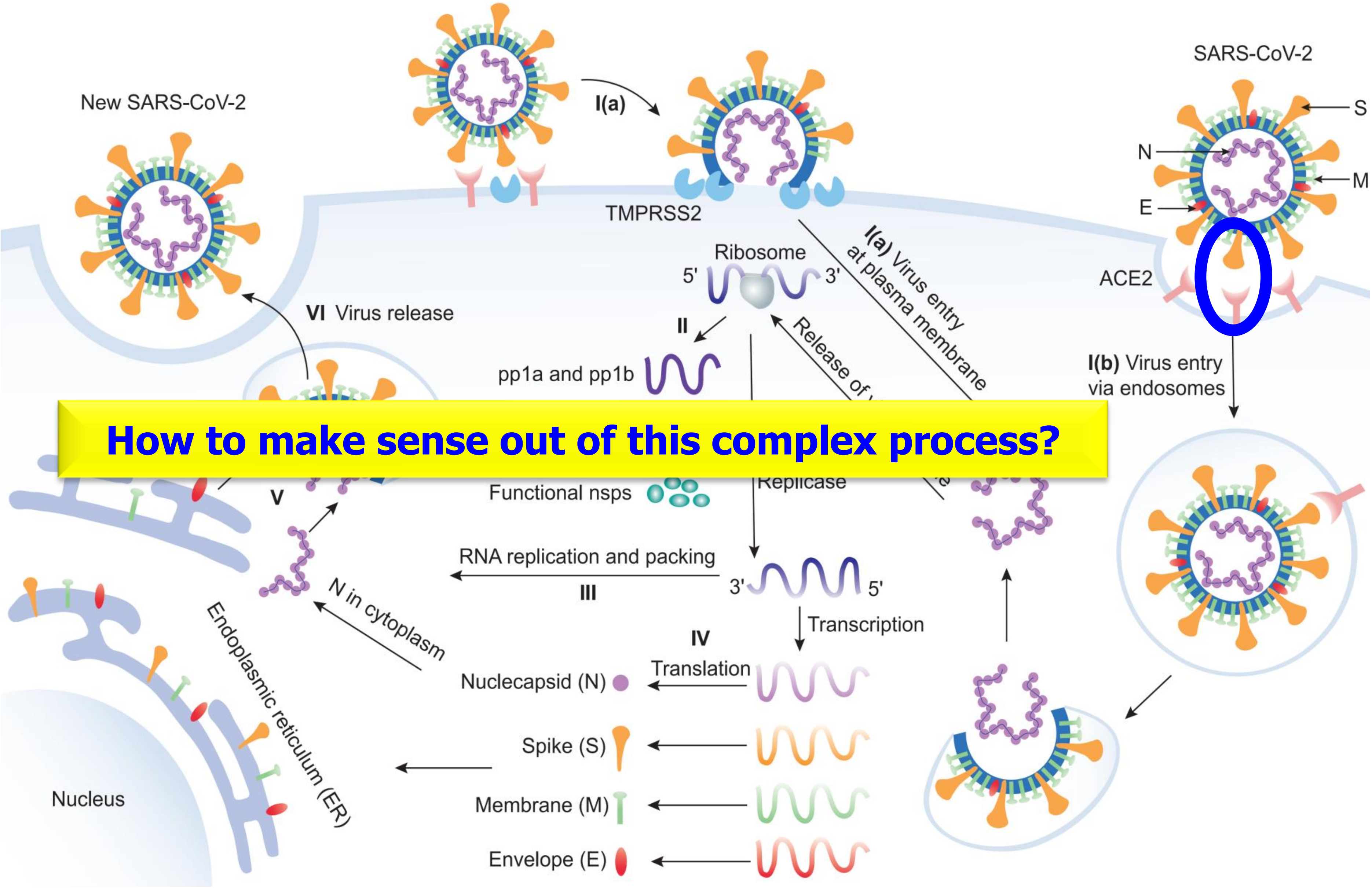
AlphaFold started to change people mind about AI/ML/DL in 2018.

Evolution of SARS-CoV-2 variants

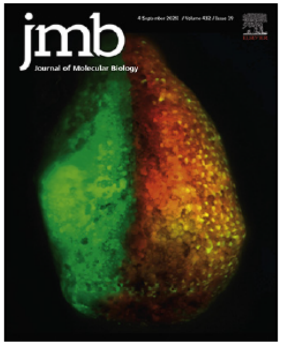


What are the evolutionary mechanisms?

Life cycle of SARS-CoV-2 in a host cell



How to make sense out of this complex process?

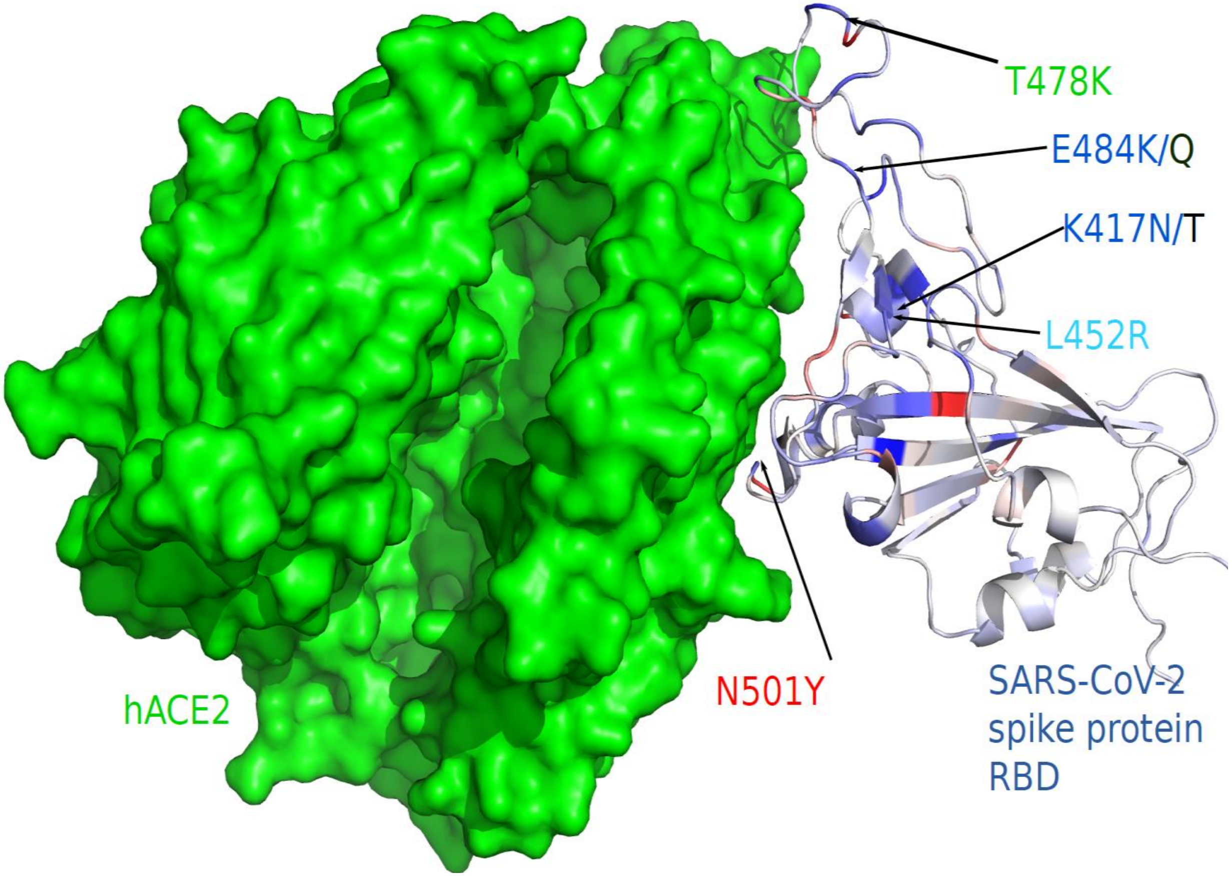


Mutations Strengthened SARS-CoV-2 Infectivity

Dr Jiahui Chen

We predicted prevailing SARS-CoV-2 variants to occur at residues 452 and 501 using T-deep mutational scanning

Jiahui Chen[1], Rui Wang[1], Menglun Wang[1] and Guo-Wei Wei[1,2,3]



Alpha: N501Y

Beta: K417N, E484K, N501Y

Gamma: K417T, E484K, N501Y

Delta: L452R, T478K

Epsilon: L452R

Theta: E484K, N501Y

Kappa: L452R, E484Q

Lambda: L452Q, F490S

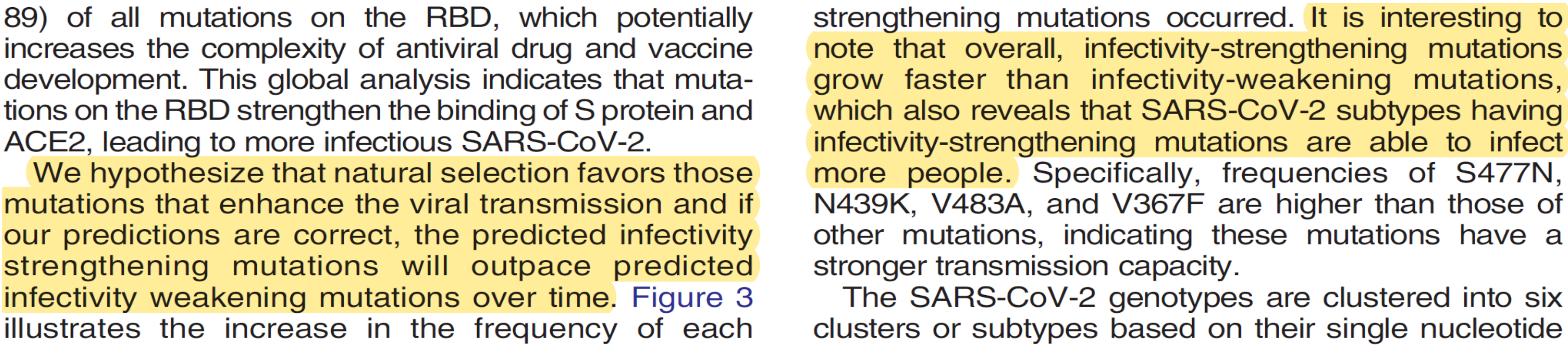
Mu: R346K,E484K,N501Y

Omicron: G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H;

BA.2.12.1: Omicron + L452Q;

BA.4/BA.5: Omicron + L452R

We discovered the mechanism of viral transmission and evolution



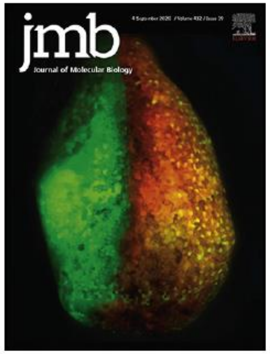
89) of all mutations on the RBD, which potentially increases the complexity of antiviral drug and vaccine development. This global analysis indicates that mutations on the RBD strengthen the binding of S protein and ACE2, leading to more infectious SARS-CoV-2.

We hypothesize that natural selection favors those mutations that enhance the viral transmission and if our predictions are correct, the predicted infectivity strengthening mutations will outpace predicted infectivity weakening mutations over time. Figure 3 illustrates the increase in the frequency of each strengthening mutations occurred. It is interesting to note that overall, infectivity-strengthening mutations grow faster than infectivity-weakening mutations, which also reveals that SARS-CoV-2 subtypes having infectivity-strengthening mutations are able to infect more people. Specifically, frequencies of S477N, N439K, V483A, and V367F are higher than those of other mutations, indicating these mutations have a stronger transmission capacity.

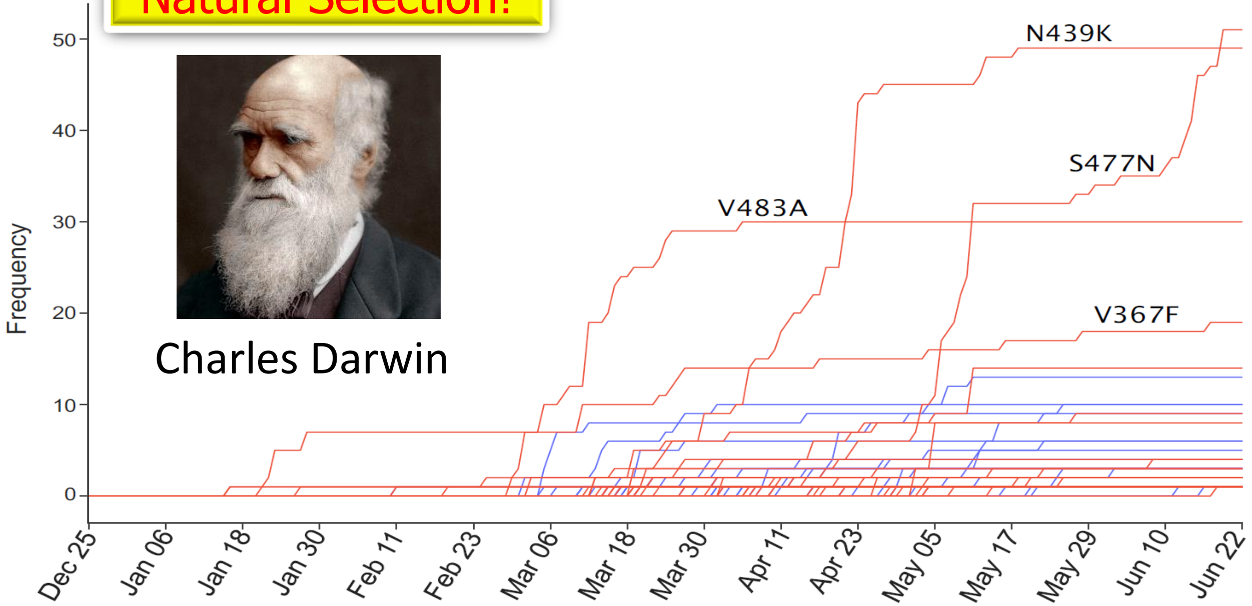
The SARS-CoV-2 genotypes are clustered into six clusters or subtypes based on their single nucleotide

Natural Selection!

Based on 15,140 viral genomes



Chen, Wang, Wang, Wei, JMB, 432, 5212, July 2020



Charles Darwin

Dr Jiahui Chen

Rui Wang

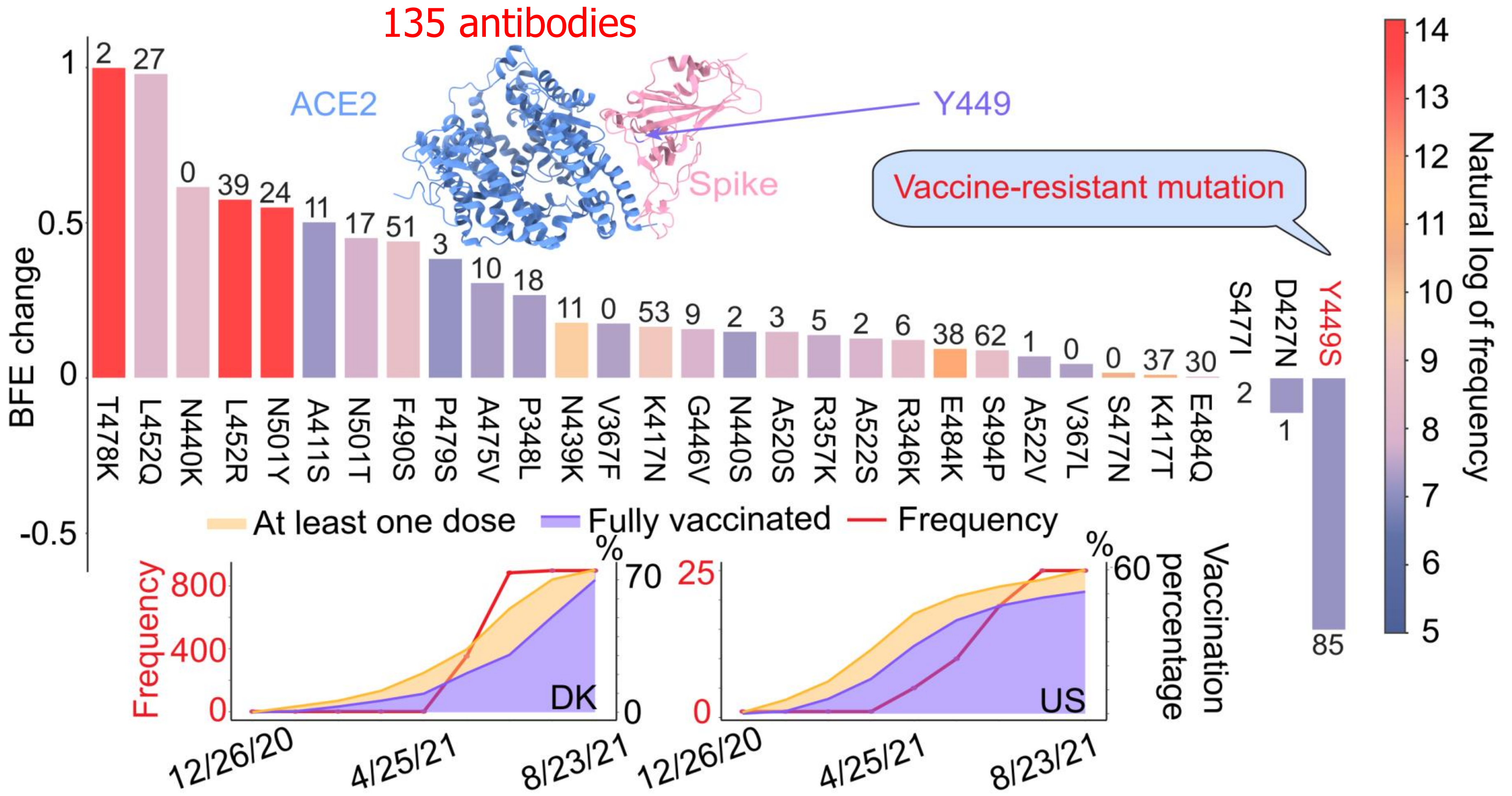
Figure 3. The time evolution of 89 SARS-CoV-2 S protein RBD mutations. The red lines represent the mutations that strengthen the infectivity of SARS-CoV-2 (i.e., $\Delta\Delta G$ is positive), and the blue lines represent the mutations that weaken the infectivity of SARS-CoV-2 (i.e., $\Delta\Delta G$ is negative). Many mutations overlap their trajectories. Here, the collection date of each genome sequence that deposited in GISAID is applied.

We discovered vaccine-breakthrough mechanism

By genotyping 2,298,349 viral genomes isolated from patients

Wang, Chen, and Wei, J. Phys. Chem. Letter, 12. 11850-11857 2021

Rui Wang



Evolution mechanisms --- Natural selection via two complementary transmission pathways: Infectivity strengthening and vaccine breakthrough

Omicron BA.2 (B.1.1.529.2): high potential to becoming the next dominating variant

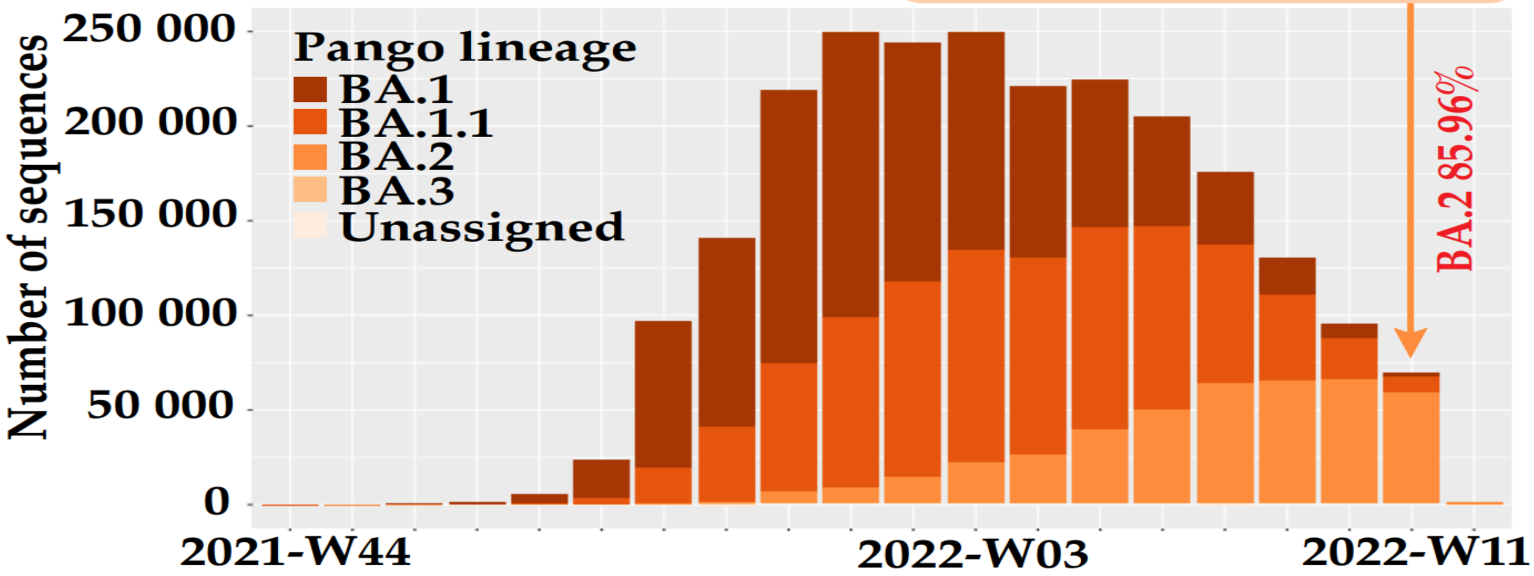
Jiahui Chen[1] and Guo-Wei Wei[1,3,4*]
[1] Department of Mathematics,
Michigan State University, MI 48824, USA.

Dr Jiahui Chen

World Health Organization

COVID-19 Weekly Epidemiological Update
Edition 84, published 22 March 2022

On 2/10/2022, we predicted that BA.2 will become the dominant variant. This became the reality in later March according to WHO



This was confirmed by WHO on March 22, 2022!

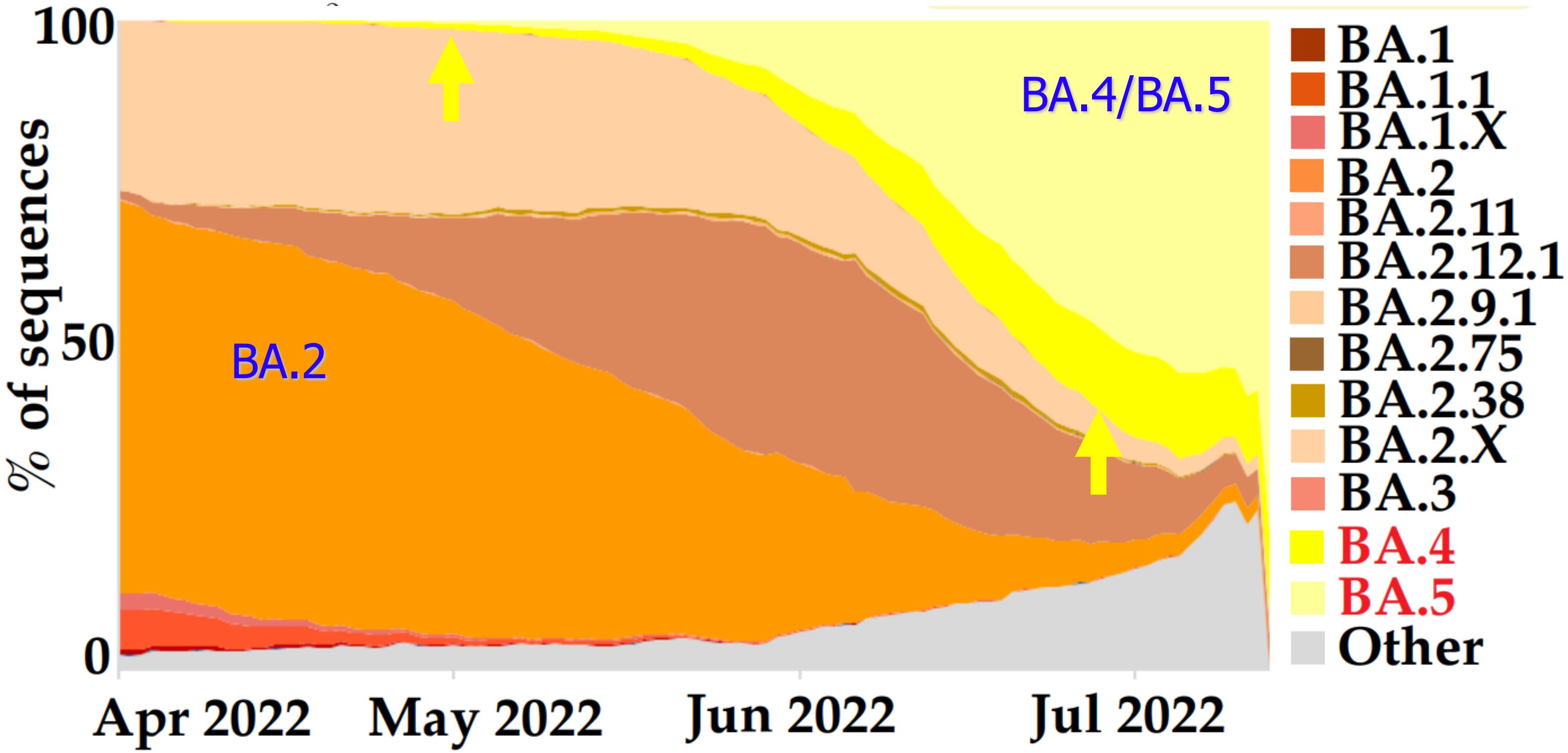
All other predictions were confirmed within 50 days

Chen, Wei, J. Phys. Chem. Lett., 13, 2840-3849, 2022.

Persistent Laplacian projected Omicron BA.4 and BA.5 to become new dominating variants

Dr Jiahui Chen

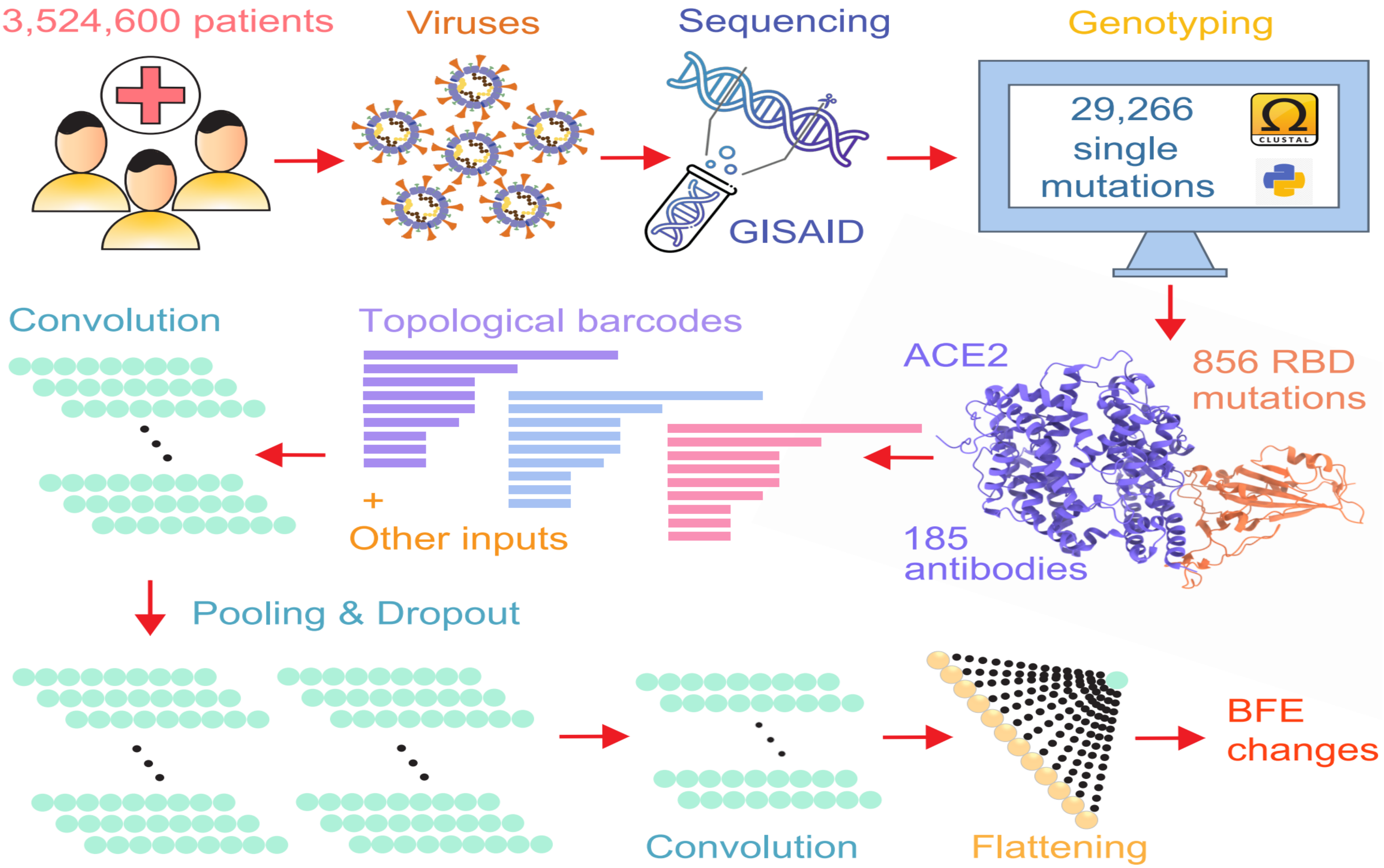
Jiahui Chen[1], Yuchi Qiu[1], Rui Wang[1], and Guo-Wei Wei[1,2,3]*
[1] Department of Mathematics,
Michigan State University, MI 48824, USA.
East Lansing, MI 48823 USA.



This was confirmed by WHO in early July (WHO weekly update release number 101)

Jiahui Chen, Qiu, Wang, and Wei, Computers in Biology and Medicine, 151, 106262 (2023)

Genome-Math-AI-experimental data modeling of protein-protein binding affinity changes following mutations



3,524,600 patients

Viruses

Sequencing

GISAID

Genotyping

29,266 single mutations

ACE2

856 RBD mutations

185 antibodies

Topological barcodes

+ Other inputs

Convolution

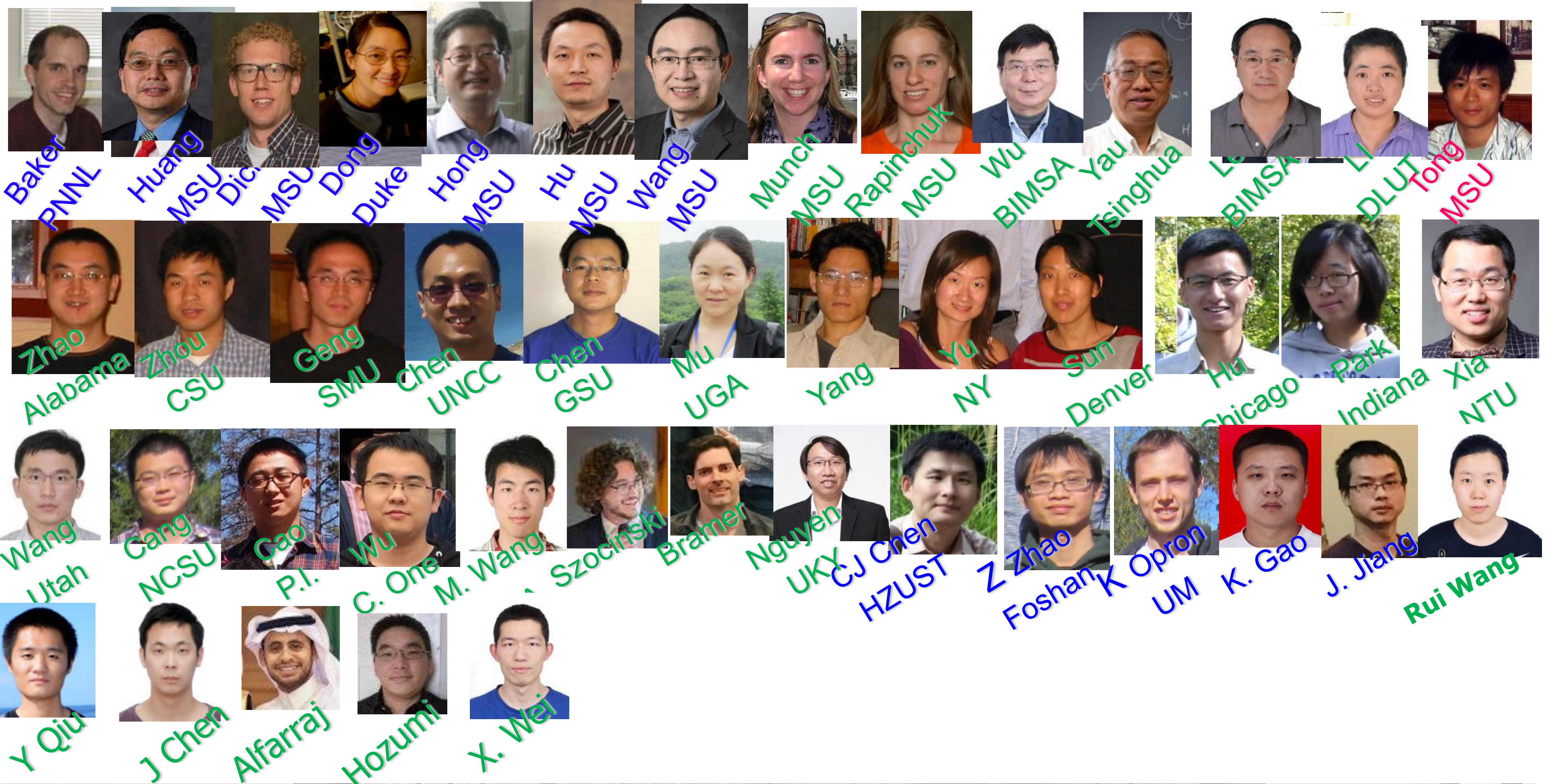
Pooling & Dropout

Convolution

Flattening

BFE changes

(Wei, Nature Machine Intell. (2019); Wang et al. Nature Machine Intell. 2020; Chen et al. Nature, Commun., 2021; Qiu et al. Nature Comput. Sci. 2021)


Baker
PNNL
Huang
MSU
Dic
MSU
Dong
Duke
Hong
MSU
Hu
MSU
Wang
MSU
Munch
MSU
Rapinchuk
MSU
Wu
BIMSA
Yau
Tsinghua
Le
BIMSA
Li
DLUT
Tong
MSU
Zhao
Alabama
Zhou
CSU
Geng
SMU
Chen
UNCC
Chen
GSU
Mu
UGA
Yang
Yu
NY
Sun
Denver
Hu
Chicago
Park
Indiana
Xia
NTU
Wang
Utah
Cang
NCSU
Cao
P.I.
Wu
C. One
M. Wang
Szocinski
Bramer
Nguyen
UKY
CJ Chen
HZUST
Z Zhao
Foshan
K Opron
UM
K. Gao
J. Jiang
Rui Wang
Y Qiu
J Chen
Alfarraj
Hozumi
X. Wei