Oxford ML Summer School 2022

GEOMETRIC DEEP LEARNING
THE ERLANGEN PROGRAMME OF ML

Michael Bronstein

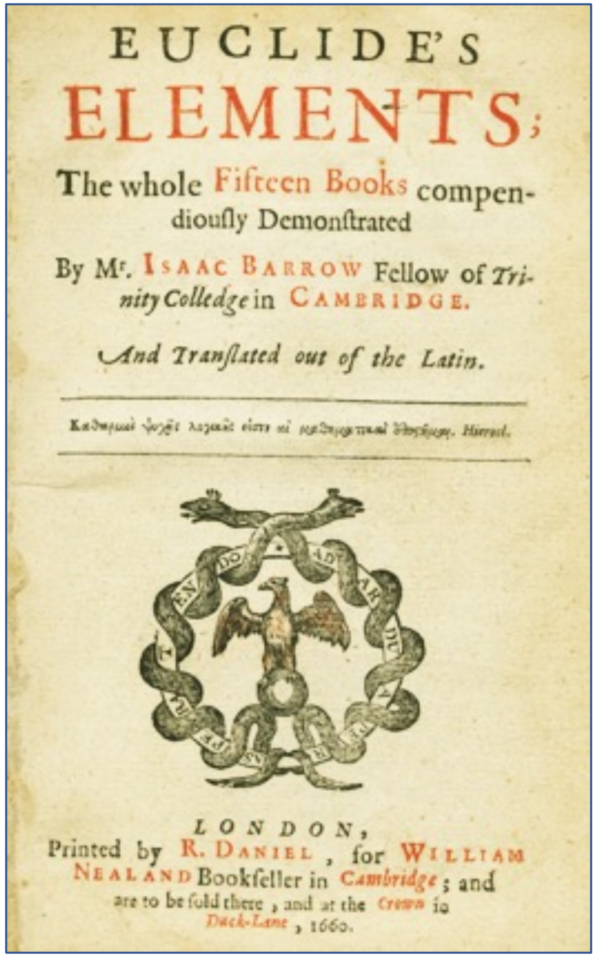
UNIVERSITY OF OXFORD

"Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection"

Hermann Weyl

Weyl 1952

The Origins



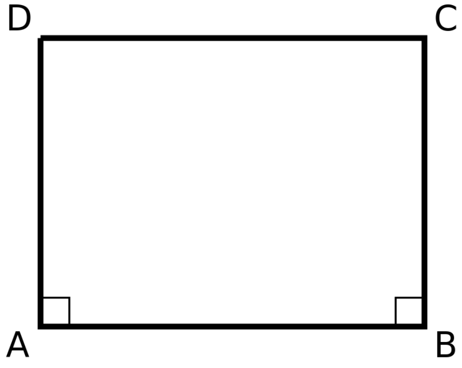
EUCLIDE'S
ELEMENTS;
The whole Fifteen Books compen-
diouſly Demonſtrated
By Mr. ISAAC BARROW Fellow of *Tri-*
nity Colledge in CAMBRIDGE.
And Tranſlated out of the Latin.

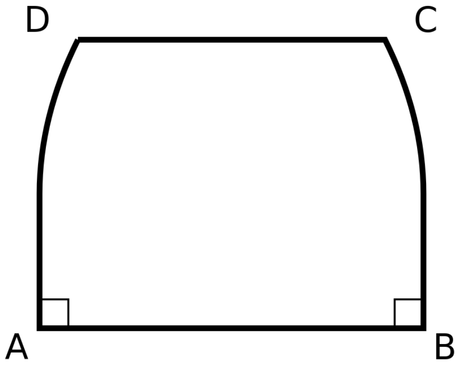
LONDON,
Printed by R. DANIEL, for WILLIAM
NEALAND Bookſeller in *Cambridge*; and
are to be ſold there, and at the *Crown* in
Duck-Lane, 1660.

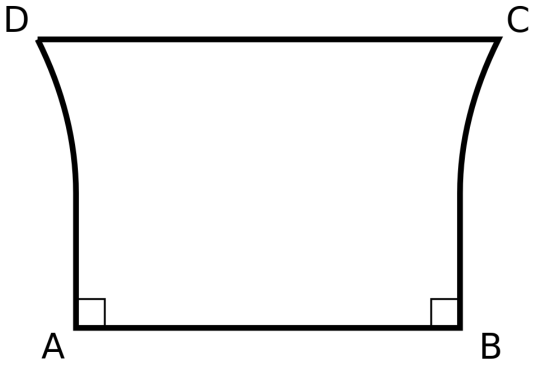
Euclid

~300 BC

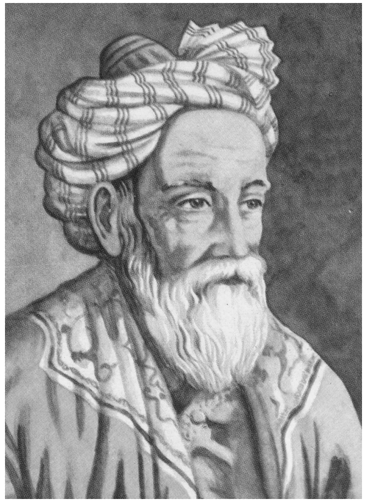
Early attempts







Khayyam-Saccheri quadrilateral



Omar Khayyam

1077

"Three cases of angles in a quadrilateral: Fifth Postulate follows from the right-angle assumption"

Early attempts

Et hujus quidem (poſt multa, ne dicam omnia, conditionatè expenſa) abſolutam falſitatem in XXXIII. tandem oſtendo, quia repugnantis naturæ lineæ rectæ, circa quam multa ibi interſero neceſſaria Lemmata. Tandem verò in præcedente Propoſitione abſolutè demonſtro ſibi ipſi repugnantem hypotheſin anguli acuti .

"repugnant to the nature of straight lines"
— Giovanni Saccheri

EUCLIDES
AB OMNI NÆVO VINDICATUS:
SIVE
CONATUS GEOMETRICUS
QUO STABILIUNTUR
Prima ipſa univerſæ Geometriæ Principia.
AUCTORE
HIERONYMO SACCHERIO
SOCIETATIS JESU
In Ticinenſi Univerſitate Matheſeos Profeſſore.
OPUSCULUM
EX.MO SENATUI
MEDIOLANENSI
Ab Auctore Dicatum.
MEDIOLANI, MDCCXXXIII.
Ex Typographia Pauli Antonii Montani. *Superiorum permiſſu.*

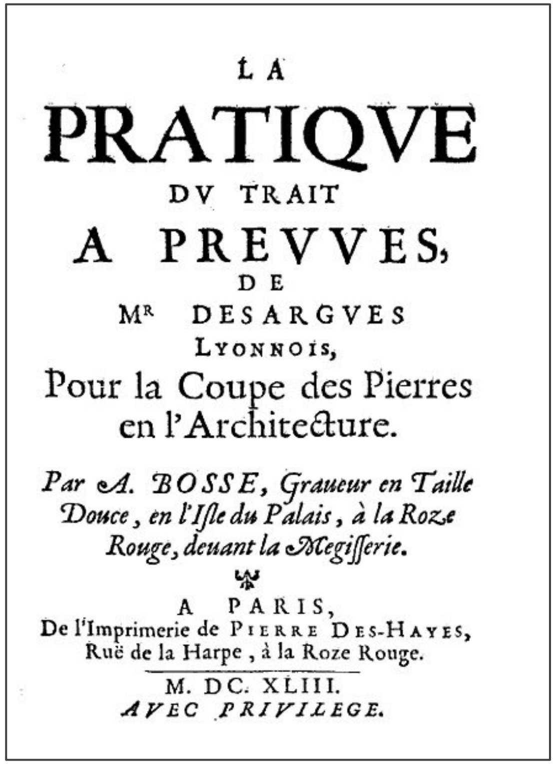
1736


ΕΥΚΛΕΙΔΗΣ

End of Euclid's Monopoly

G. Desargues

1643



LA

PRATIQVE

DV TRAIT

A PREVVES,

DE

MR DESARGVES

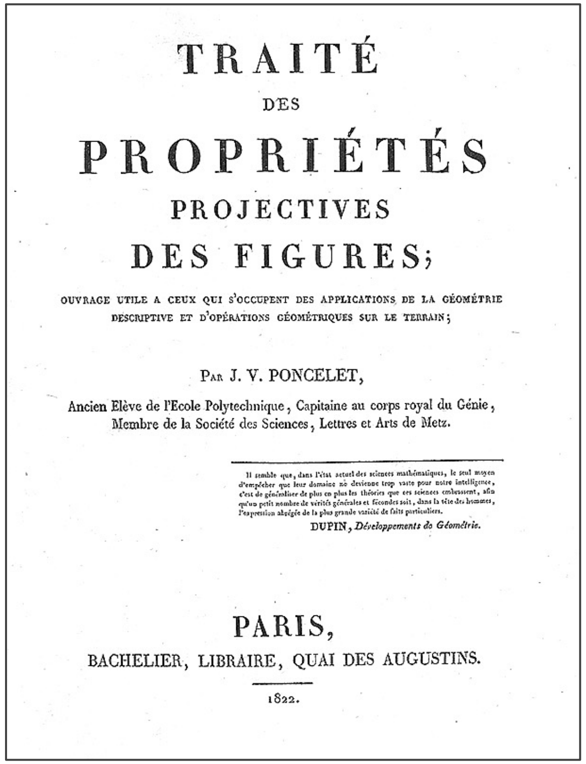
LYONNOIS,

Pour la Coupe des Pierres
en l'Architecture.

Par A. BOSSE, Graueur en Taille Douce, en l'Isle du Palais, à la Roze Rouge, deuant la Megisserie.

A PARIS,
De l'Imprimerie de PIERRE DES-HAYES,
Ruë de la Harpe, à la Roze Rouge.

M. DC. XLIII.
AVEC PRIVILEGE.



TRAITÉ

DES

PROPRIÉTÉS

PROJECTIVES

DES FIGURES;

OUVRAGE UTILE A CEUX QUI S'OCCUPENT DES APPLICATIONS DE LA GÉOMÉTRIE DESCRIPTIVE ET D'OPÉRATIONS GÉOMÉTRIQUES SUR LE TERRAIN;

PAR J. V. PONCELET,

Ancien Elève de l'Ecole Polytechnique, Capitaine au corps royal du Génie,
Membre de la Société des Sciences, Lettres et Arts de Metz.

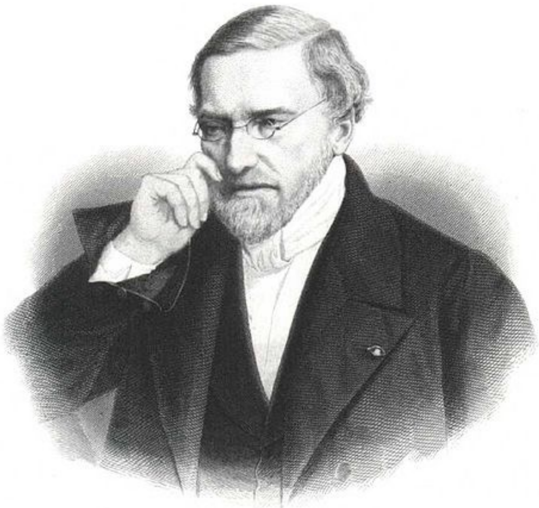
Il semble que, dans l'état actuel des sciences mathématiques, le seul moyen d'empêcher que leur domaine ne devienne trop vaste pour notre intelligence, c'est de généraliser de plus en plus les théories que ces sciences embrassent, afin qu'un petit nombre de vérités générales et fécondes soit, dans la tête des hommes, l'expression abrégée de la plus grande variété de faits particuliers.

DUPIN, *Développements de Géométrie.*

PARIS,
BACHELIER, LIBRAIRE, QUAI DES AUGUSTINS.

1822.

"Projective geometry"



J. V. Poncelet

1822

End of Euclid's Monopoly

C. F. Gauss

"I have discovered such wonderful things that I was amazed...out of nothing I have created a strange new world." — Jánus Bolyai to his father

"To praise it would amount to praising myself. For the entire content of the work...coincides almost exactly with my own meditations [in the] past thirty or thirty-five years."

— Gauss to Farkas Bolyai

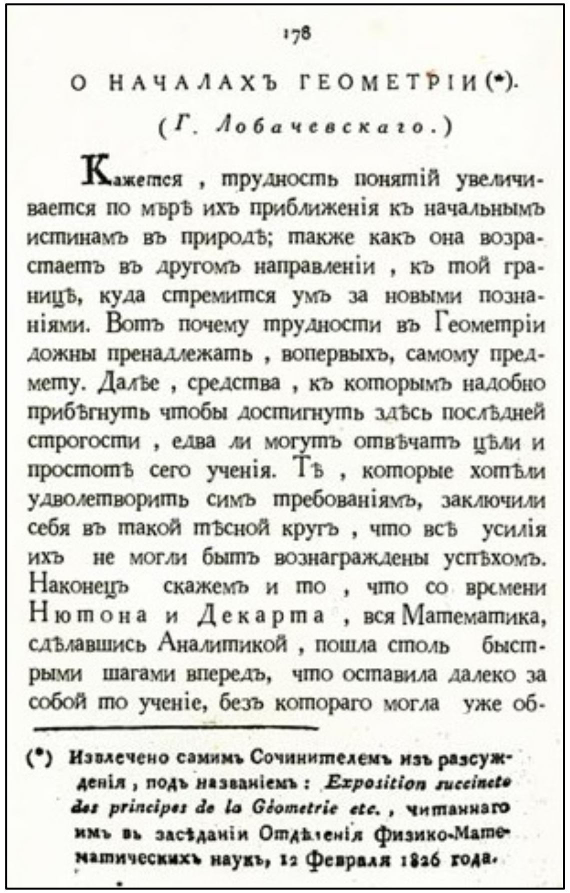
J. Bolyai

1823

Gauss ~1800; Bolyai (1823) 1832

End of Euclid's Monopoly

"In geometry I find certain imperfections which I hold to be the reason why this science […] can as yet make no advance from that state in which it came to us from Euclid. I consider […] the momentous gap in the theory of parallels, to fill which all efforts of mathematicians have so far been in vain."



178

О НАЧАЛАХЪ ГЕОМЕТРІИ (*).

(*Г. Лобачевскаго.*)

Кажется, трудность понятій увеличивается по мѣрѣ ихъ приближенія къ начальнымъ истинамъ въ природѣ; также какъ она возрастаетъ въ другомъ направленіи, къ той границѣ, куда стремится умъ за новыми познаніями. Вотъ почему трудности въ Геометріи дожны пренадлежать, вопервыхъ, самому предмету. Далѣе, средства, къ которымъ надобно прибѣгнуть чтобы достигнуть здѣсь послѣдней строгости, едва ли могутъ отвѣчать цѣли и простотѣ сего ученія. Тѣ, которые хотѣли удволетворить симъ требованіямъ, заключили себя въ такой тѣсной кругъ, что всѣ усилія ихъ не могли бытъ вознаграждены успѣхомъ. Наконецъ скажемъ и то, что со времени Нютона и Декарта, вся Математика, сдѣлавшись Аналитикой, пошла столь быстрыми шагами впередъ, что оставила далеко за собой то ученіе, безъ котораго могла уже об-

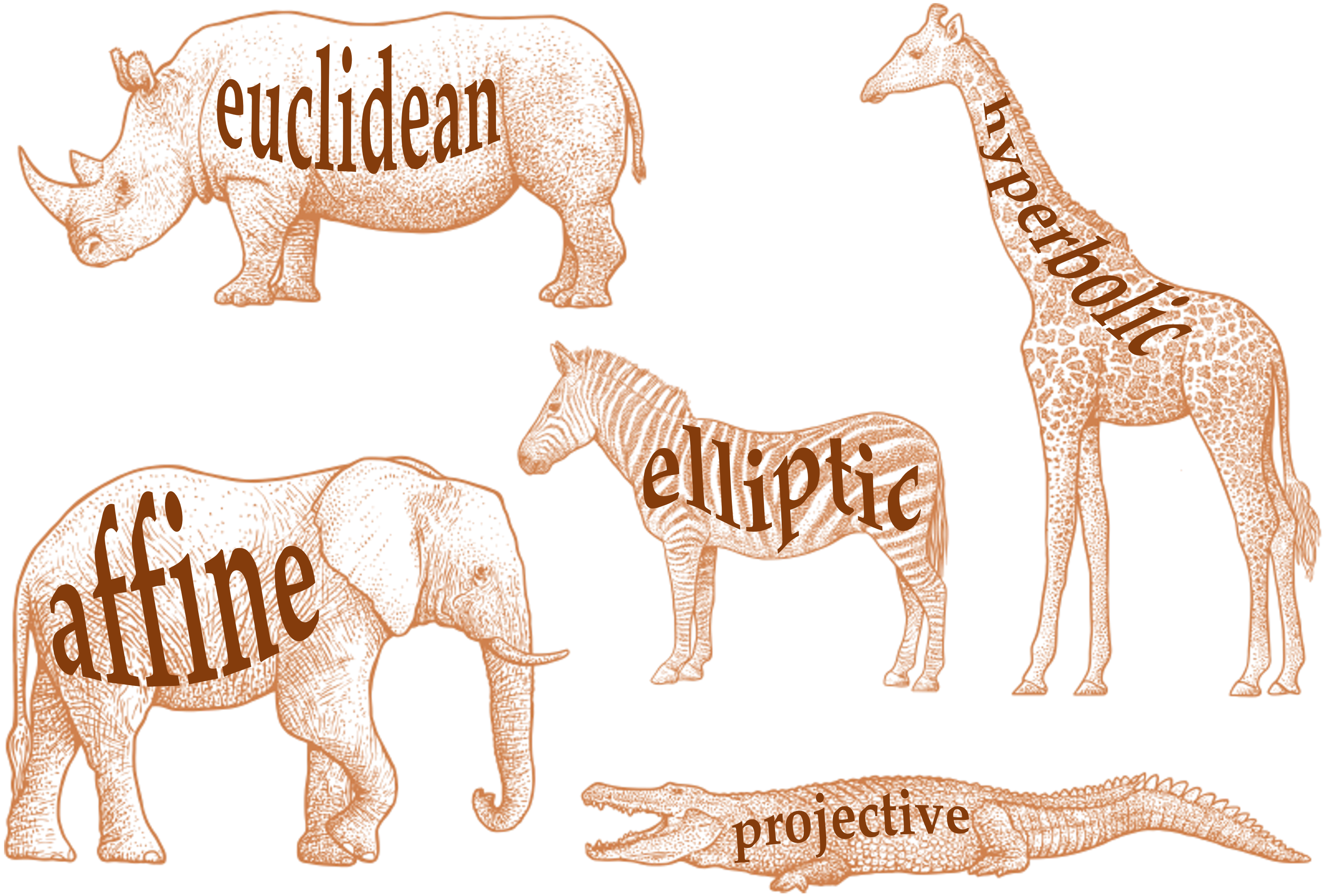
(*) Извлечено самимъ Сочинителемъ изъ разсужденія, подъ названіемъ: *Exposition succincte des principes de la Géometrie etc.*, читаннаго имъ въ засѣданіи Отдѣленія Физико-Математическихъ наукъ, 12 Февраля 1826 года.

N. Lobachevsky

1829

Lobachevsky (1826) 1829

Nineteenth Century Zoo of Geometries



The Erlangen Programme

"Given a [homogeneous] manifold and a transformation group acting [transitively] on it, to investigate those properties of figures on that manifold which are invariant under transformations of that group"

Vergleichende Betrachtungen

über

neuere geometrische Forschungen

von

Dr. Felix Klein,

o. ö. Professor der Mathematik an der Universität Erlangen.

Programm

zum Eintritt in die philosophische Facultät und den Senat der k. Friedrich-Alexanders-Universität zu Erlangen.

Erlangen.

Verlag von Andreas Deichert.

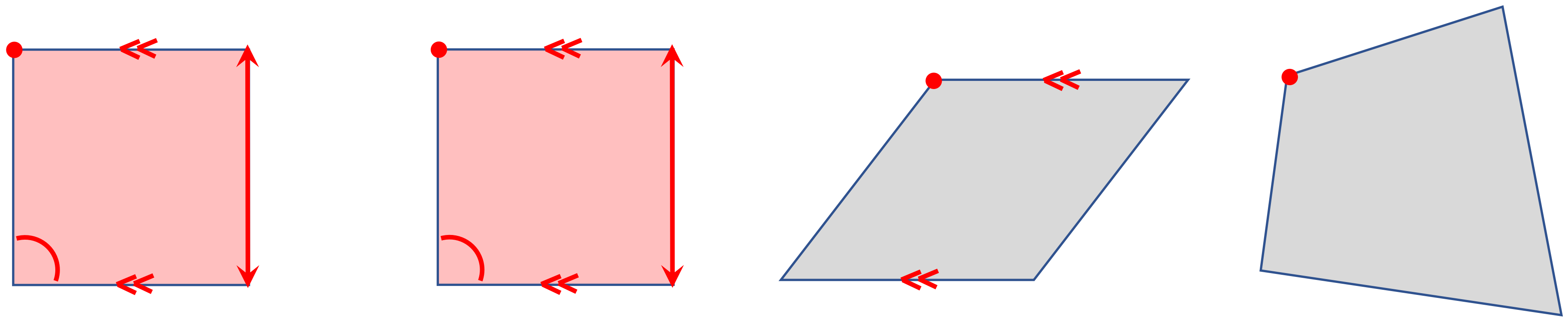
1872.

F. Klein

1872

Klein 1872

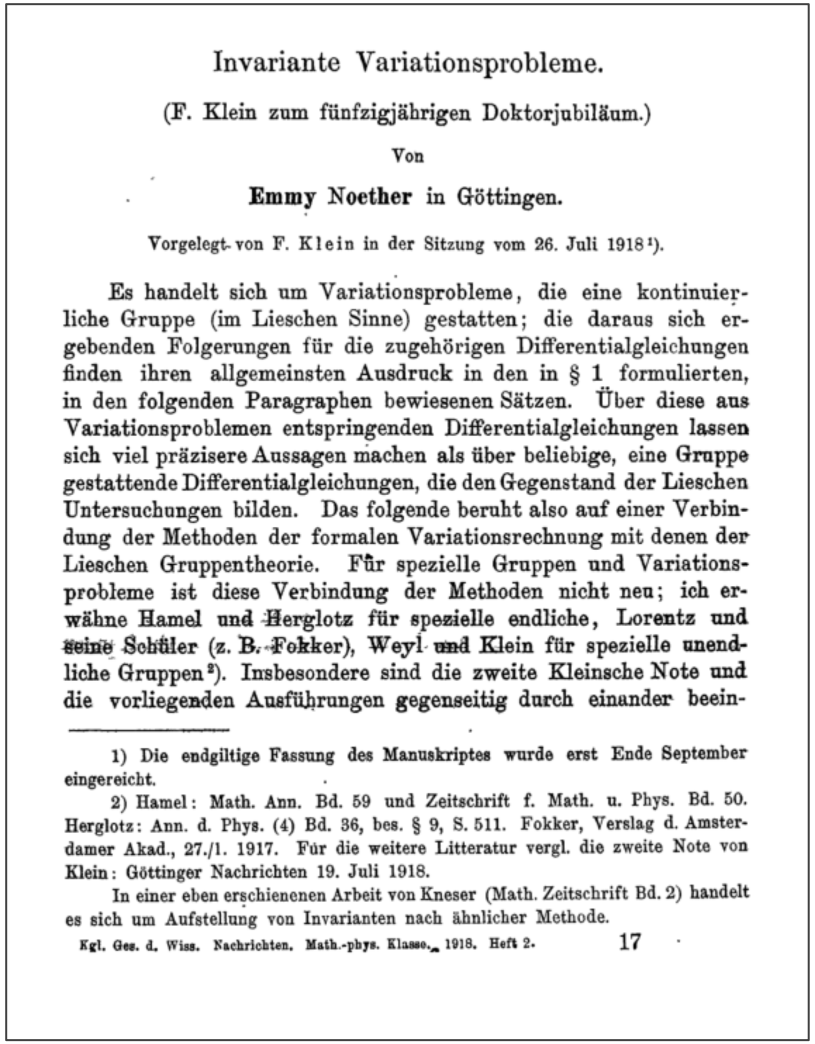
The Erlangen Programme



	Euclidean	**Affine**	**Projective**
angle	+	—	—
distance	+	—	—
area	+	—	—
parallelism	+	+	—
intersection	+	+	+

Noether's Theorem

"Every [differentiable] symmetry of the action of a physical system [with conservative forces] has a corresponding conservation law"



Invariante Variationsprobleme.

(F. Klein zum fünfzigjährigen Doktorjubiläum.)

Von

Emmy Noether in Göttingen.

Vorgelegt von F. Klein in der Sitzung vom 26. Juli 1918 [1]).

Es handelt sich um Variationsprobleme, die eine kontinuierliche Gruppe (im Lieschen Sinne) gestatten; die daraus sich ergebenden Folgerungen für die zugehörigen Differentialgleichungen finden ihren allgemeinsten Ausdruck in den in § 1 formulierten, in den folgenden Paragraphen bewiesenen Sätzen. Über diese aus Variationsproblemen entspringenden Differentialgleichungen lassen sich viel präzisere Aussagen machen als über beliebige, eine Gruppe gestattende Differentialgleichungen, die den Gegenstand der Lieschen Untersuchungen bilden. Das folgende beruht also auf einer Verbindung der Methoden der formalen Variationsrechnung mit denen der Lieschen Gruppentheorie. Für spezielle Gruppen und Variationsprobleme ist diese Verbindung der Methoden nicht neu; ich erwähne Hamel und Herglotz für spezielle endliche, Lorentz und seine Schüler (z. B. Fokker), Weyl und Klein für spezielle unendliche Gruppen [2]). Insbesondere sind die zweite Kleinsche Note und die vorliegenden Ausführungen gegenseitig durch einander beein-

1) Die endgiltige Fassung des Manuskriptes wurde erst Ende September eingereicht.

2) Hamel: Math. Ann. Bd. 59 und Zeitschrift f. Math. u. Phys. Bd. 50. Herglotz: Ann. d. Phys. (4) Bd. 36, bes. § 9, S. 511. Fokker, Verslag d. Amsterdamer Akad., 27./1. 1917. Für die weitere Litteratur vergl. die zweite Note von Klein: Göttinger Nachrichten 19. Juli 1918.

In einer eben erschienenen Arbeit von Kneser (Math. Zeitschrift Bd. 2) handelt es sich um Aufstellung von Invarianten nach ähnlicher Methode.

Kgl. Ges. d. Wiss. Nachrichten. Math.-phys. Klasse. 1918. Heft 2. 17

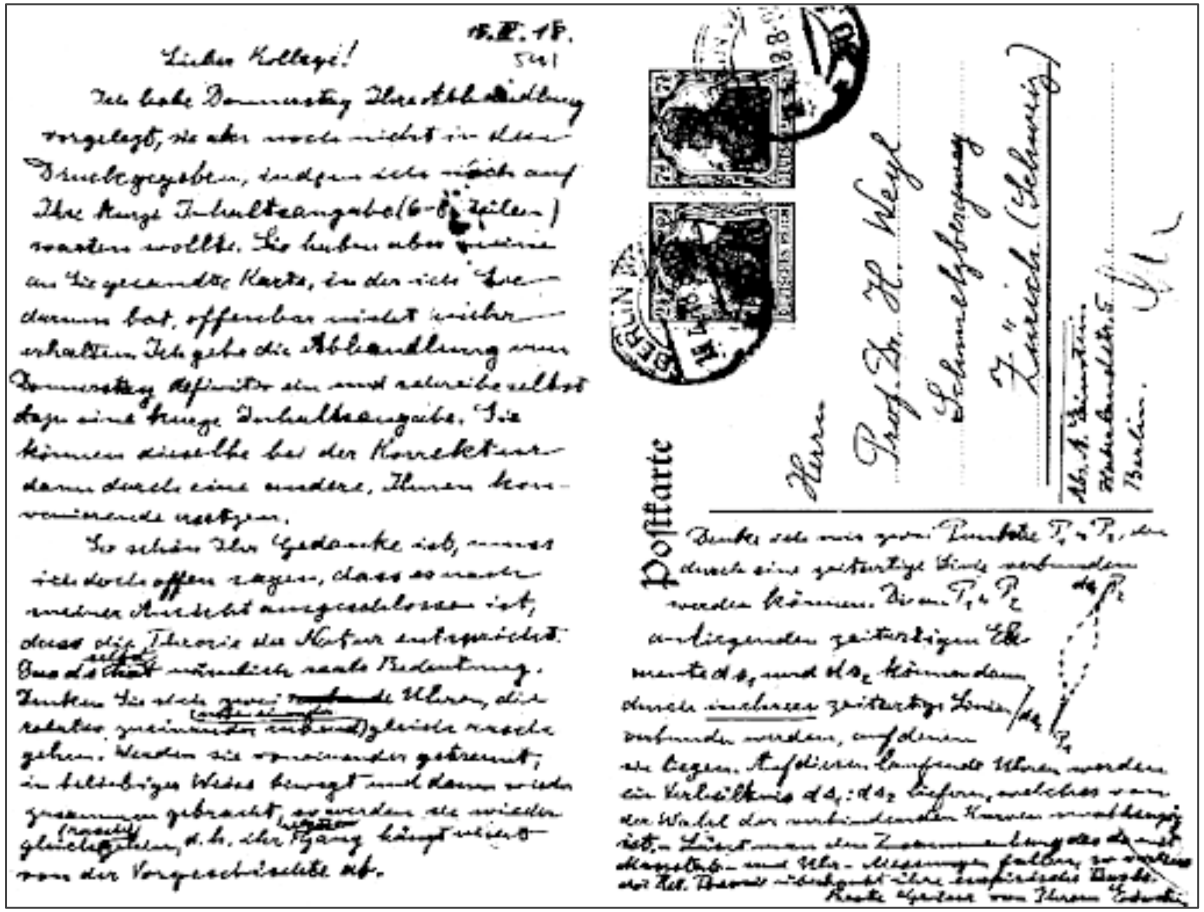
E. Noether

1918

Noether 1918

Gauge invariance

Lieber Kollege! —
Postcard dated 15 April 1918 from Einstein to Weyl arguing with his initially proposed gauge theory



H. Weyl

1929

Weyl 1919; 1929 (see Straumann 1987)

Unification of forces

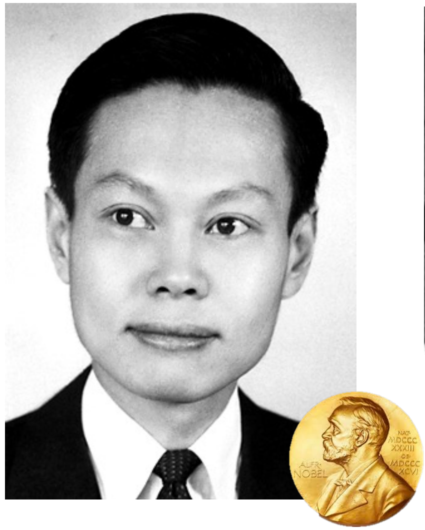
PHYSICAL REVIEW VOLUME 96, NUMBER 1 OCTOBER 1, 1954

Conservation of Isotopic Spin and Isotopic Gauge Invariance*

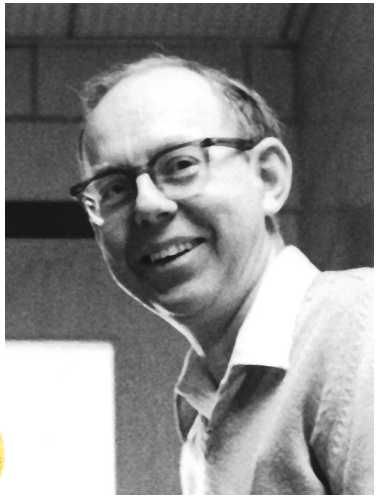
C. N. Yang † and R. L. Mills
Brookhaven National Laboratory, Upton, New York

(Received June 28, 1954)

It is pointed out that the usual principle of invariance under isotopic spin rotation is not consistant with the concept of localized fields. The possibility is explored of having invariance under local isotopic spin rotations. This leads to formulating a principle of isotopic gauge invariance and the existence of a **b** field which has the same relation to the isotopic spin that the electromagnetic field has to the electric charge. The **b** field satisfies nonlinear differential equations. The quanta of the **b** field are particles with spin unity, isotopic spin unity, and electric charge $\pm e$ or zero.



C. N. Yang

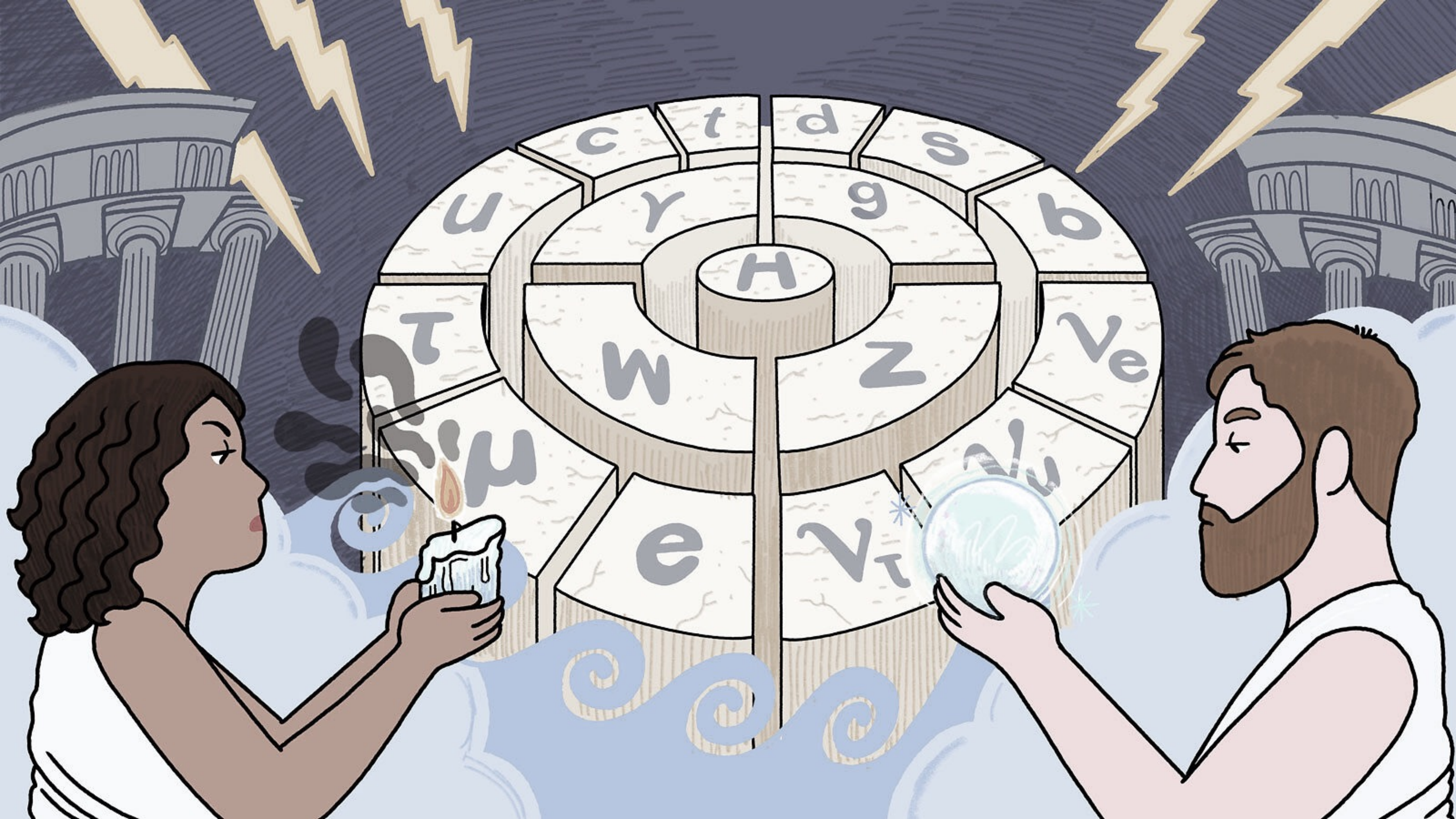


R. L. Mills

Unification of electromagnetic and weak forces (modelled with the groups U(1) × SU(2)) and the strong force (based on the group SU(3))

1954

Yang, Mills 1954



c
t
d
s
u
γ
g
b
H
τ
W
Z
νe
μ
νμ
e
ντ

"It is only slightly overstating the case to say that Physics is the study of symmetry"

P. Anderson

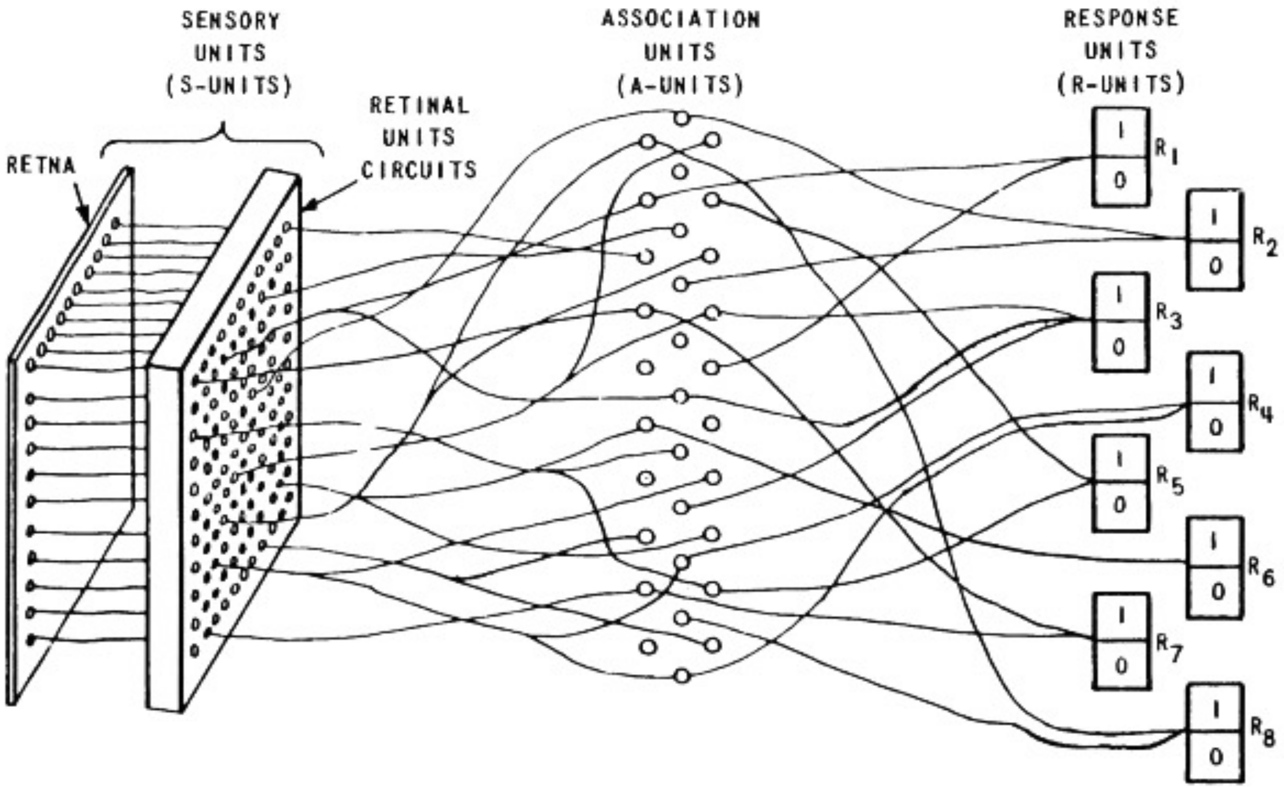
Anderson 1972

?

EARLY NEURAL NETWORKS & THE AI WINTER

Dartmouth AI Conference 1956

Early neural networks



Perceptron, one of the first neural network architectures

F. Rosenblatt

1957

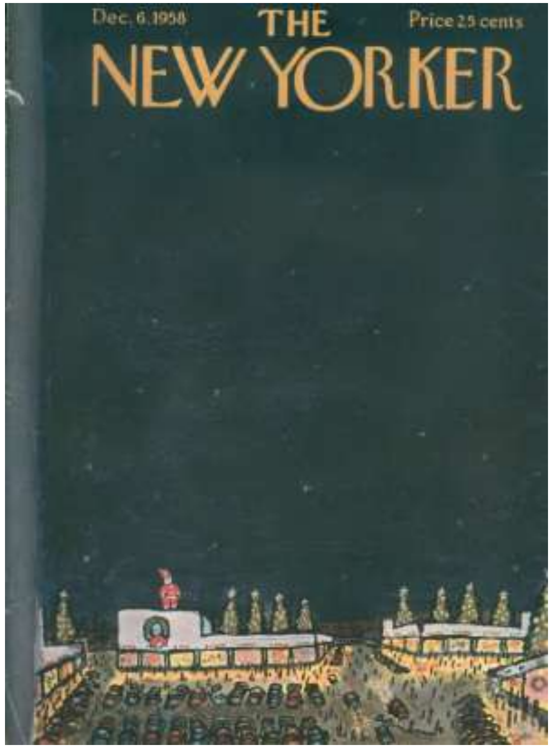
Rosenblatt 1957

Early hype

"First serious rival to the human brain even devised."

"Remarkable machine capable of what amounts to thought"

— The New Yorker



1958

Manson, Stewart, Gill 1958

Early hype

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

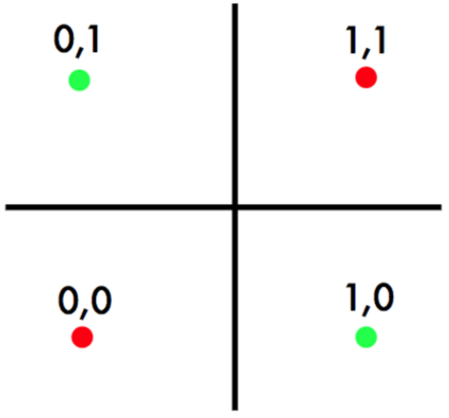
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

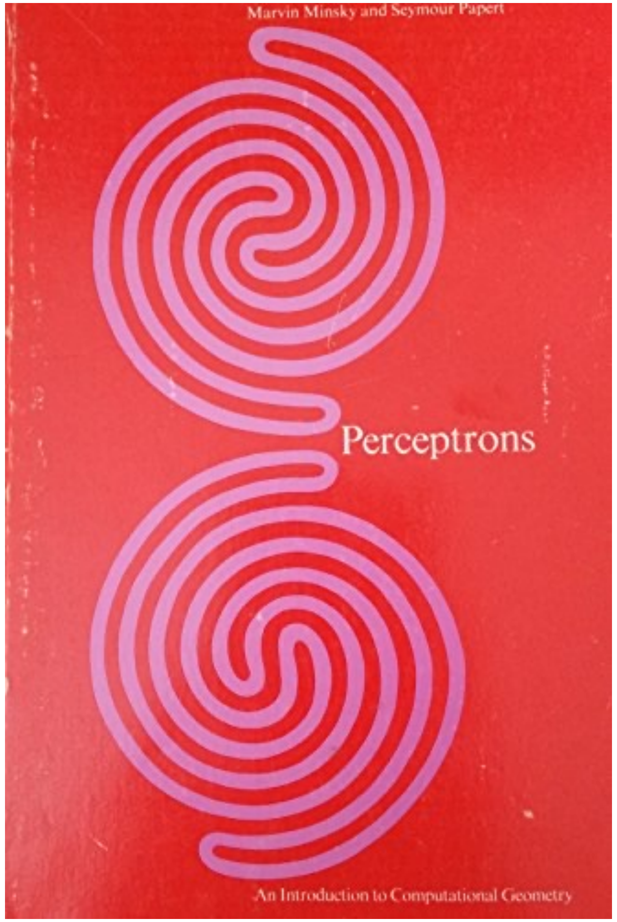
Papert 1966

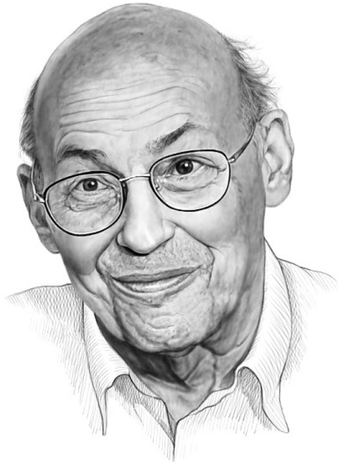
The "XOR Affair"



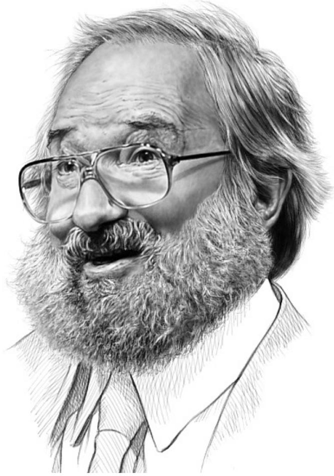
XOR

"[simple] perceptron cannot represent even the XOR function"





M. Minsky



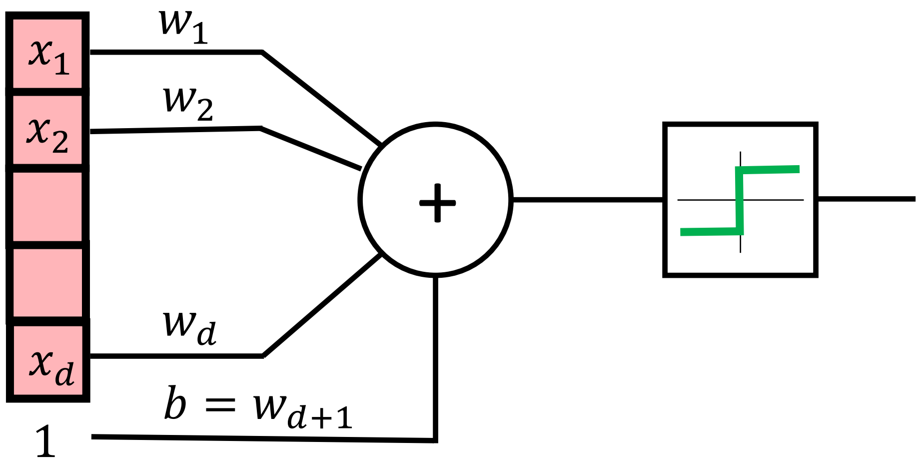
S. Papert

1969

Minsky, Papert 1969


"AI WINTER"

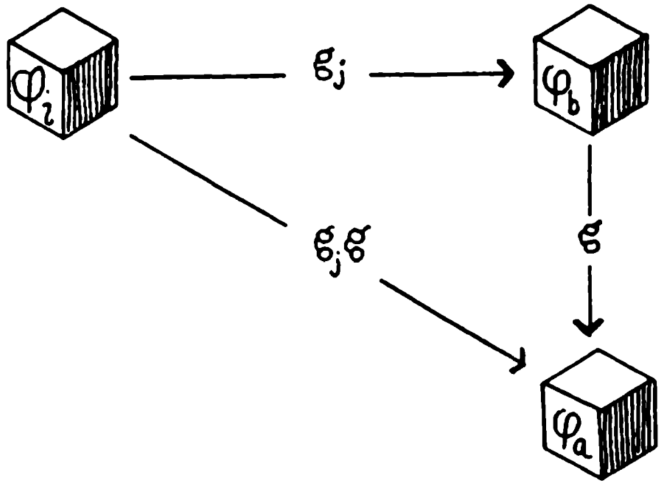
"Simple perceptron"



$$y = \mathrm{sign}(\mathbf{w}^{\mathrm{T}}\mathbf{x})$$

First "geometric" machine learning

Group Invariance Theorem: "if a neural network is invariant to a group, then its output can be expressed as functions of the orbits of the group"

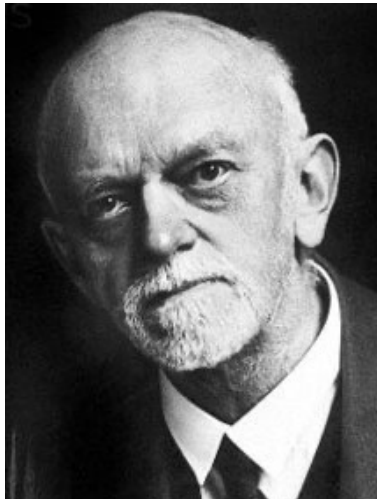


M. Minsky **S. Papert**

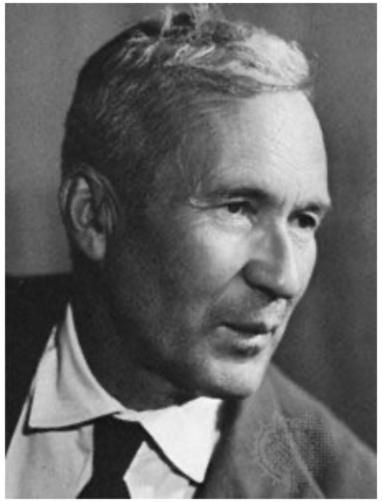
1969

Minsky, Papert 1969

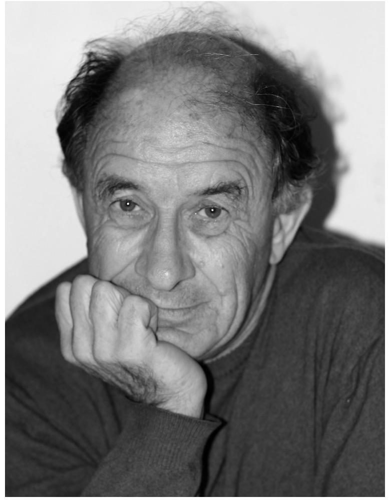
Universal approximation



D. Hilbert



A. Kolmogorov



V. Arnold

G. Cybenko

K. Hornik

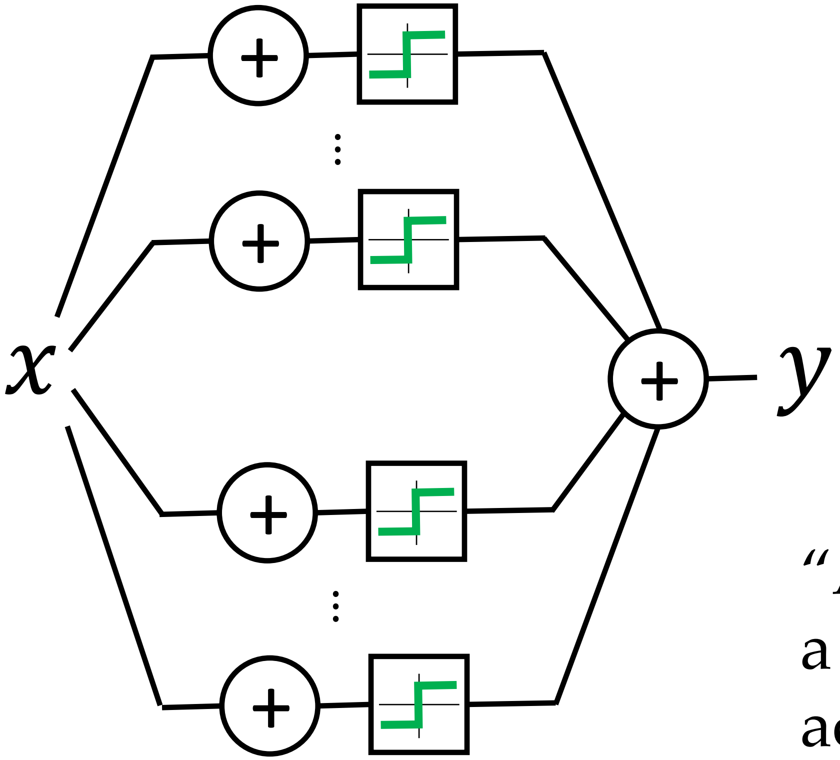
13th Problem

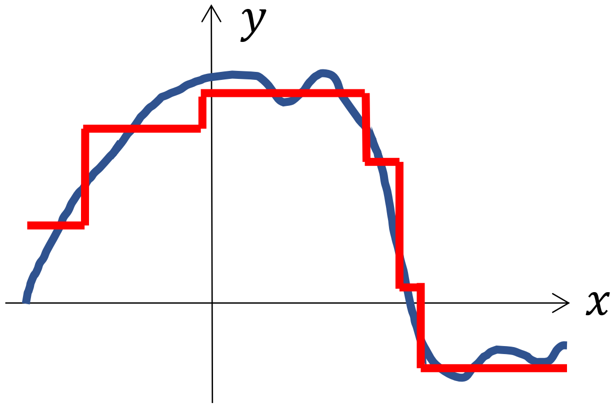
$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$$

Results specific to multilayer neural networks

Hilbert 1900; Arnold 1956; Kolmogorov 1957; Cybenko 1989; Hornik 1991

Universal approximation



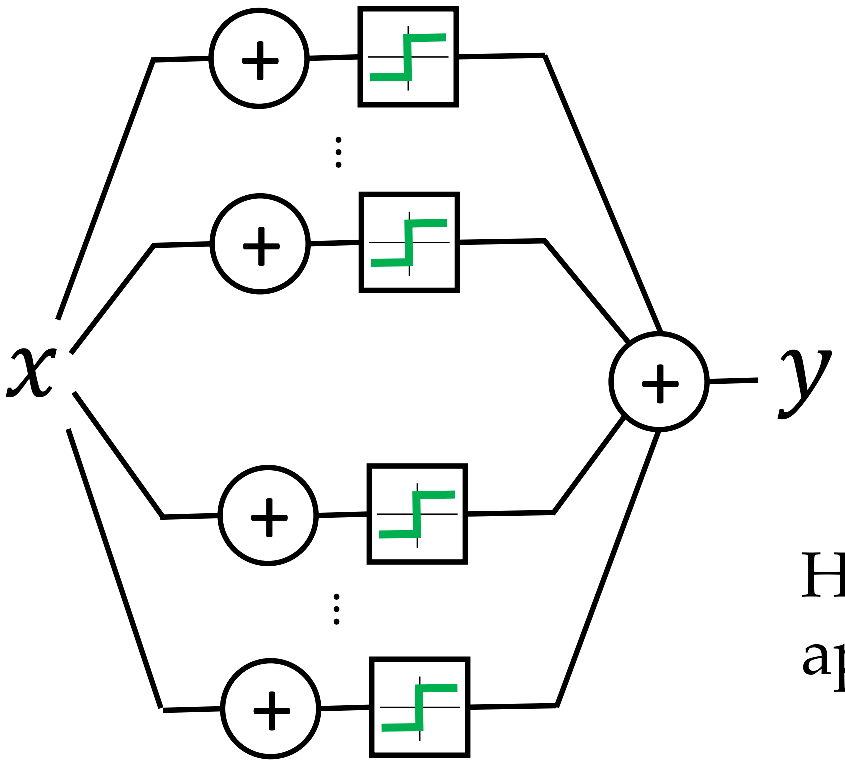


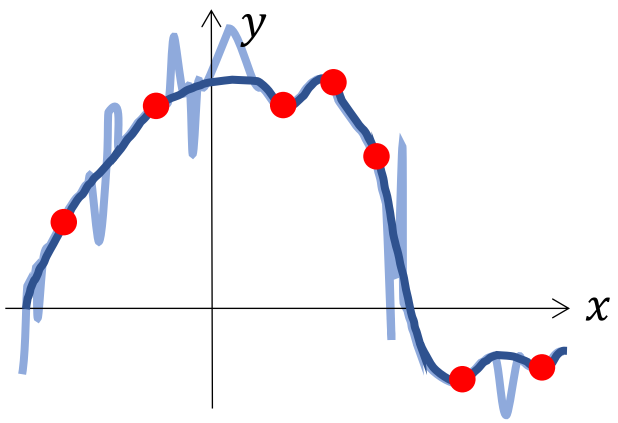
"A 2-layer perceptron can approximate a continuous function to any desired accuracy"

Cybenko 1989; Hornik 1991; Barron 1993; Leshno et al 1993; Maiorov 1999; Pinkus 1999

Deep learning = glorified curve fitting

Universal approximation

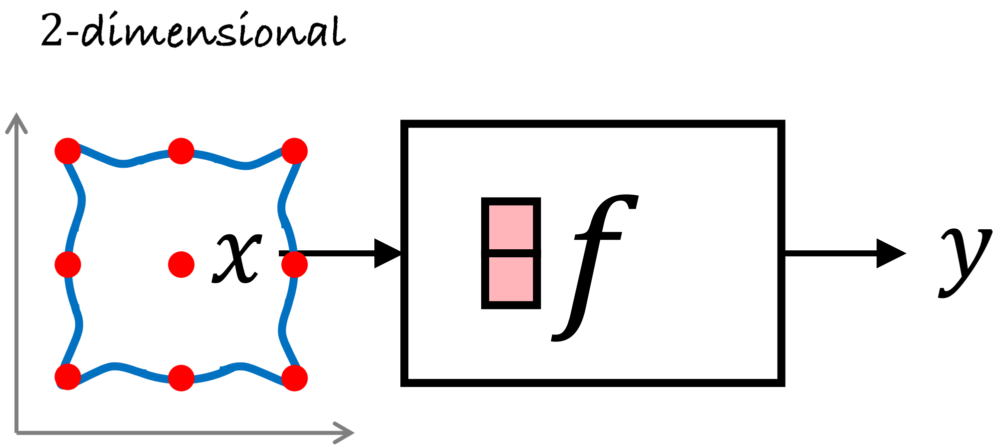




How many samples are needed to approximate to accuracy ε ?

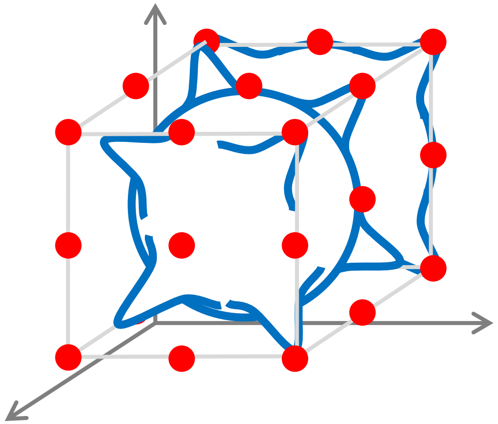
Cybenko 1989; Hornik 1991; Barron 1993; Leshno et al 1993; Maiorov 1999; Pinkus 1999

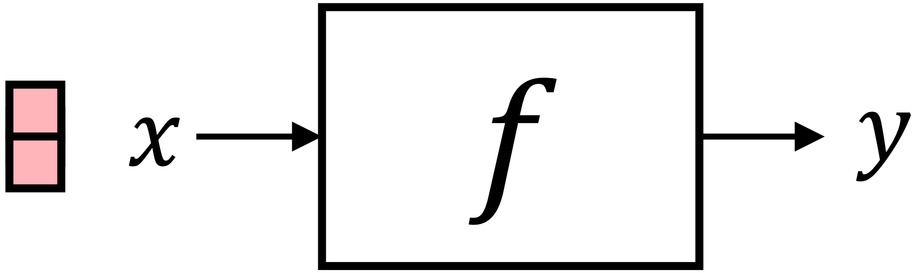
The Curse of Dimensionality



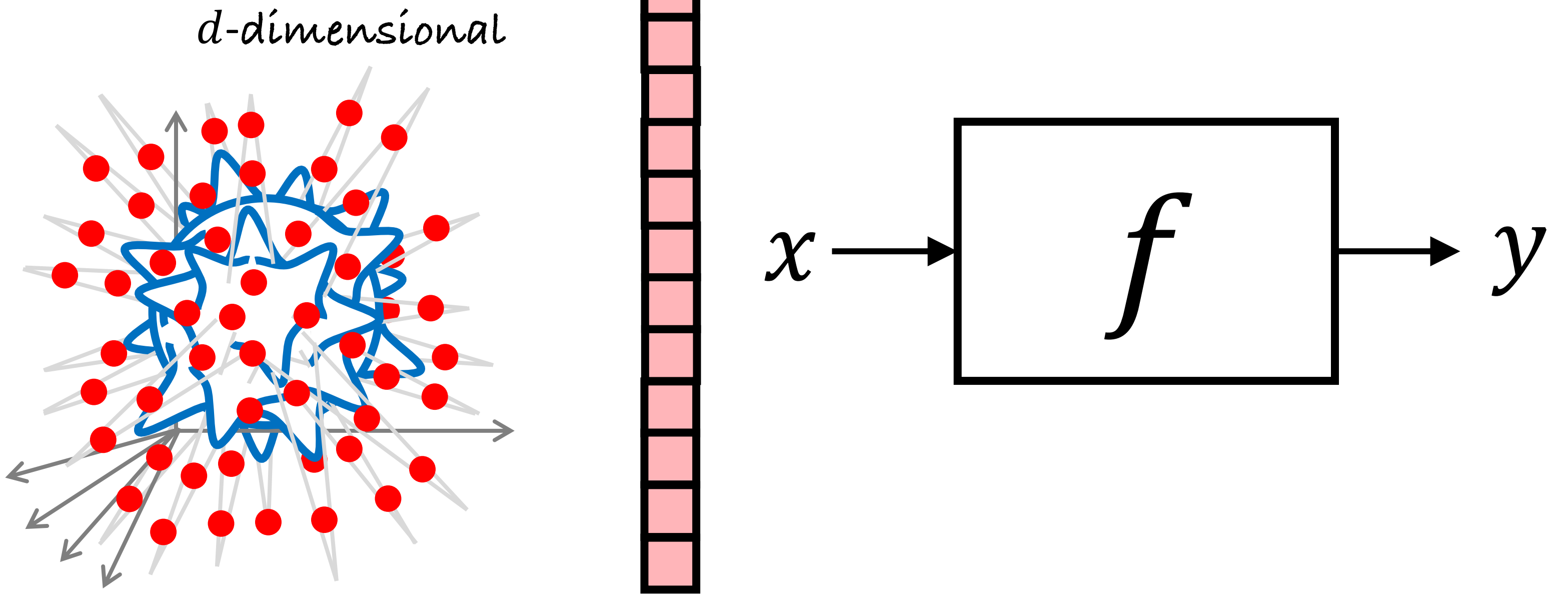
The Curse of Dimensionality

3-dimensional





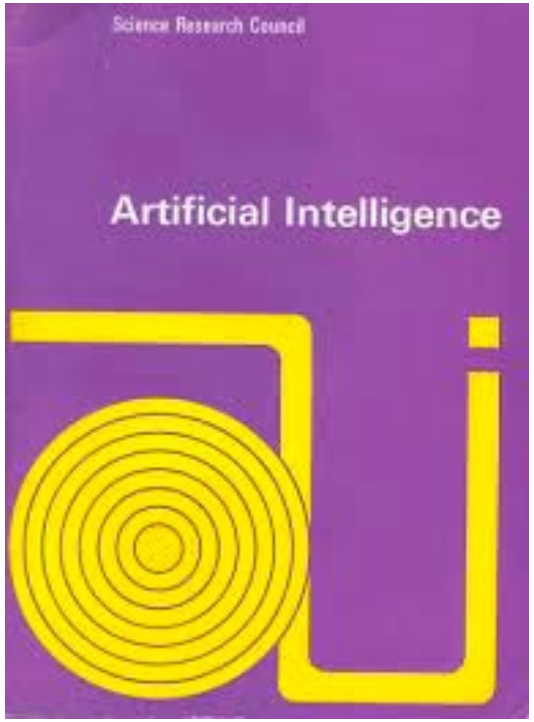
The Curse of Dimensionality



$O(\varepsilon^{-d})$ samples

The Lighthill Report

"Most workers in AI research and in related fields confess to a pronounced feeling of disappointment in what has been achieved in the past twenty-five years. […] In no part of the field have the discoveries made so far produced the major impact that was then promised."



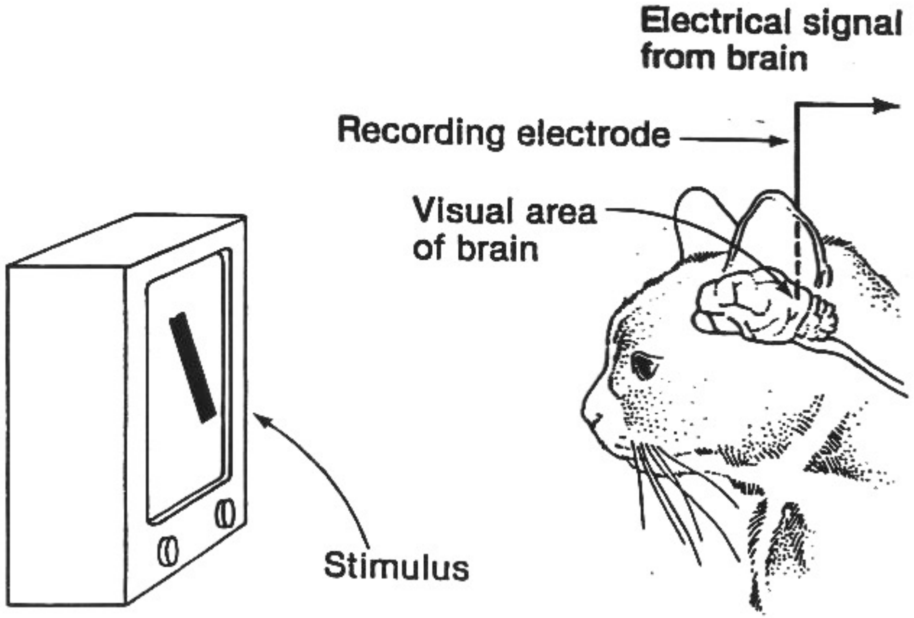
J. Lighthill

1972

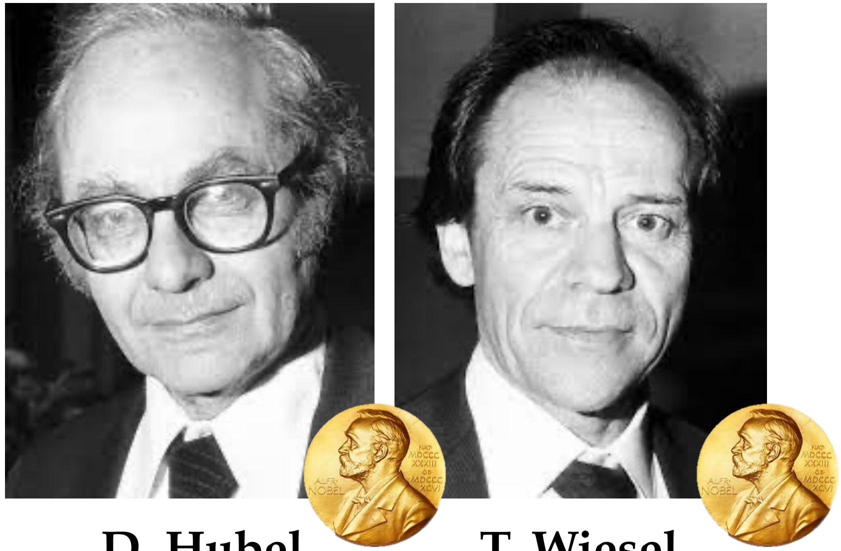
Lighthill 1972

THE EMERGENCE OF GEOMETRIC ARCHITECTURES

Secrets of the visual cortex



Experiments of Hubel and Wiesel that established the structure of the visual cortex

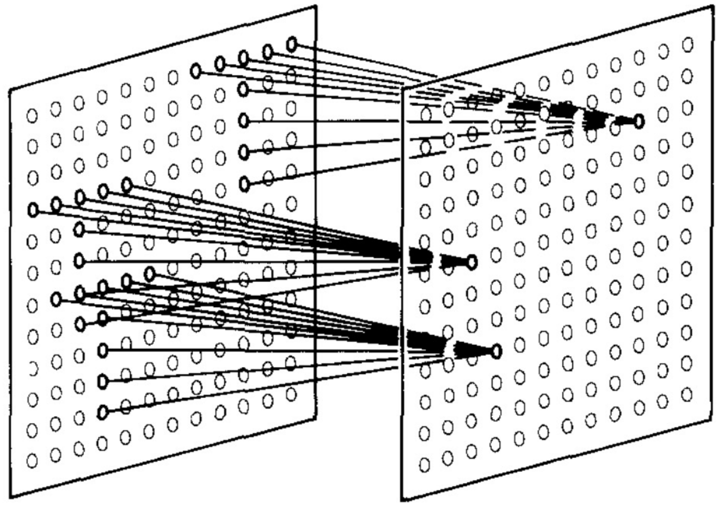


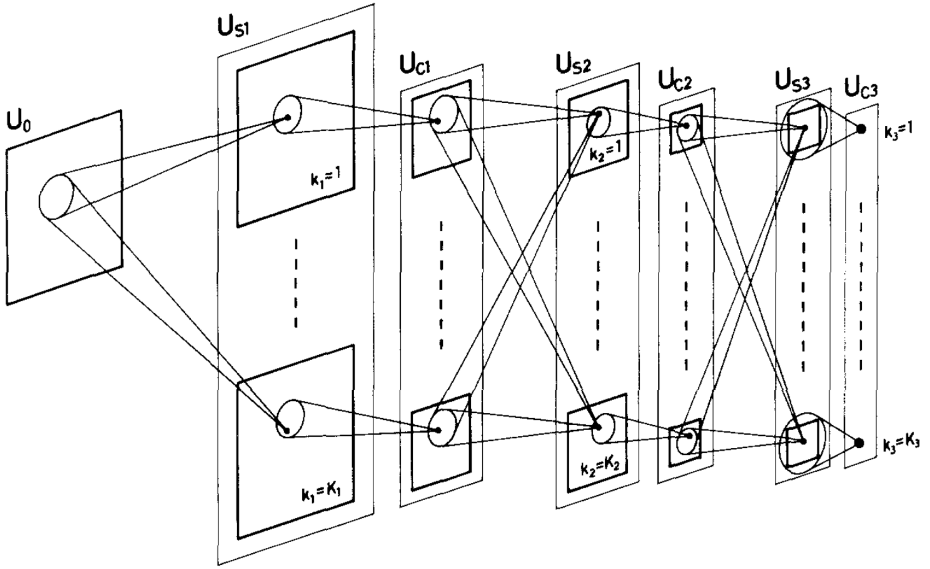
D. Hubel **T. Wiesel**

1959

Hubel, Wiesel 1959; 1962

Neocognitron



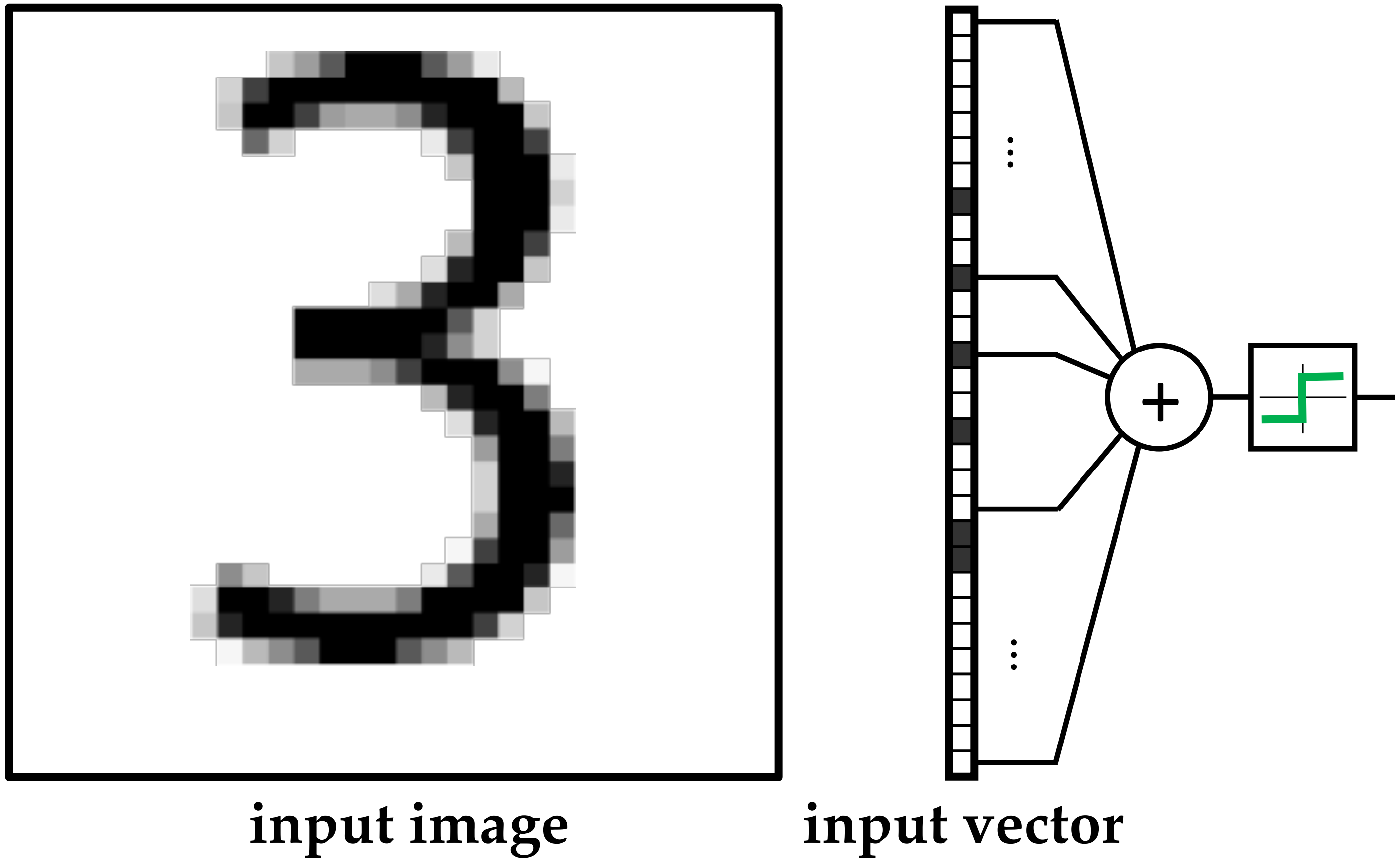


Neocognitron, an early geometric neural network

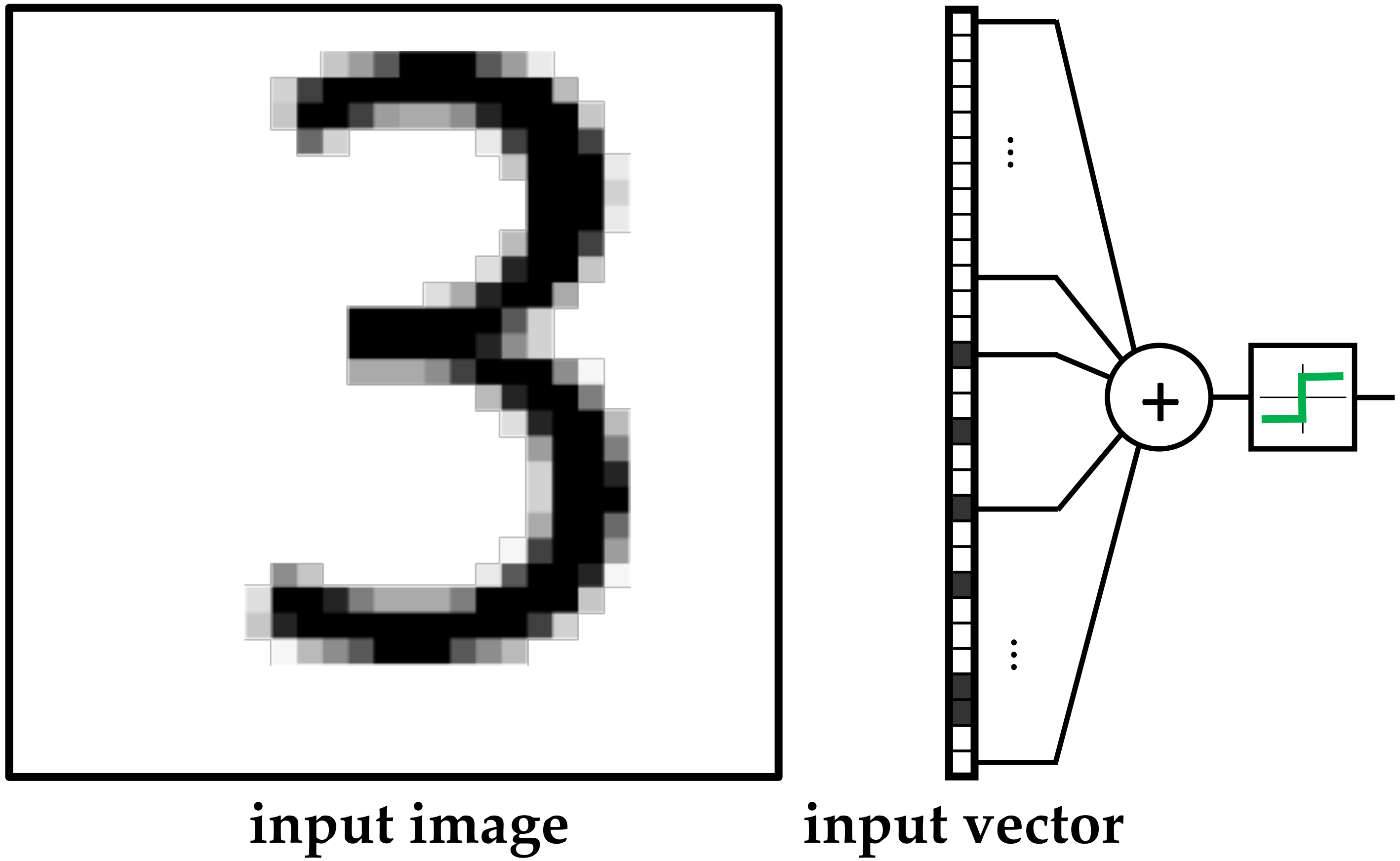
K. Fukushima

1980

Fukushima 1980

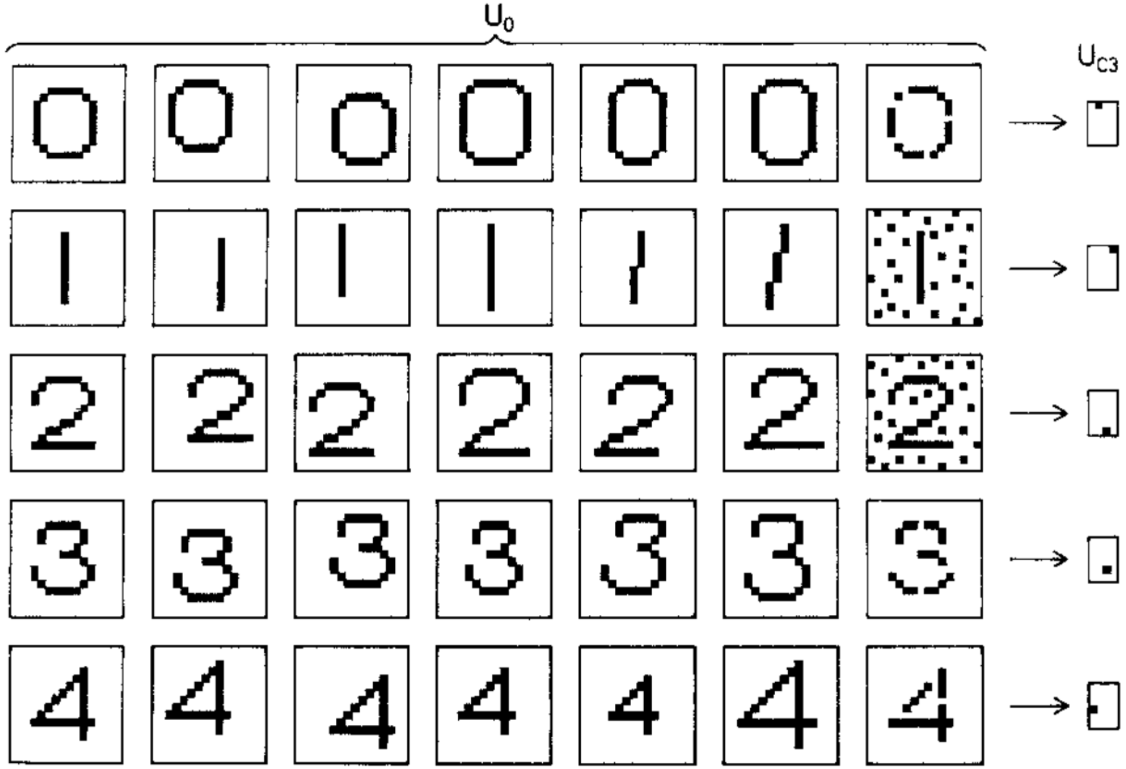


input image
input vector



"The response of [Perceptrons] was severely affected by the shift in position [...] of the input patterns. Hence, their ability for pattern recognition was not so high."— Fukushima

Neocognitron



Experimental evaluation of the Neocognitron

K. Fukushima

1980

Fukushima 1980

Neocognitron

- Deep neural network (7 layers tested)
- Local connectivity ("receptive fields")
- Nonlinear filters with shared weights (S-layers)
- Average pooling (C-layers)
- ReLU activation function
- "Self-organised" (unsupervised) **– no backprop yet!**

K. Fukushima

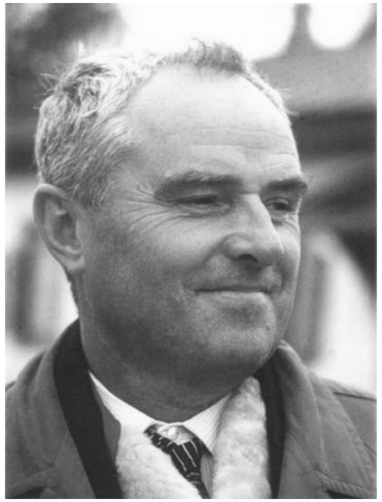
1980

Fukushima 1980

How to train your neural network?

F. Rosenblatt

Perceptron
learning rule
(1 layer)



A. Ivakhnenko

Group method of
data handling

S. Linnainmaa

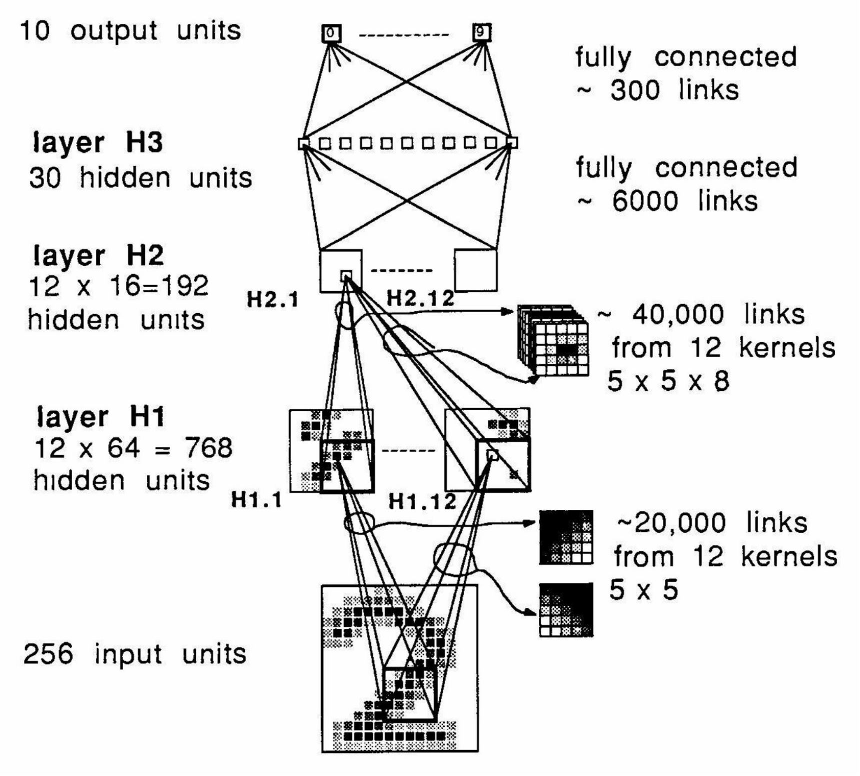
P. Werbos

Backpropagation

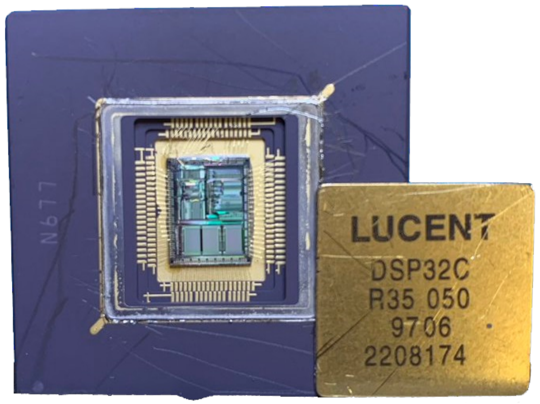
D. Rumelhart

Rosenblatt 1957; Ivakhnenko, Lapa 1966; Linnainmaa 1970; Werbos 1982; Rumelhart et al. 1986

Convolutional neural networks



First version of a CNN

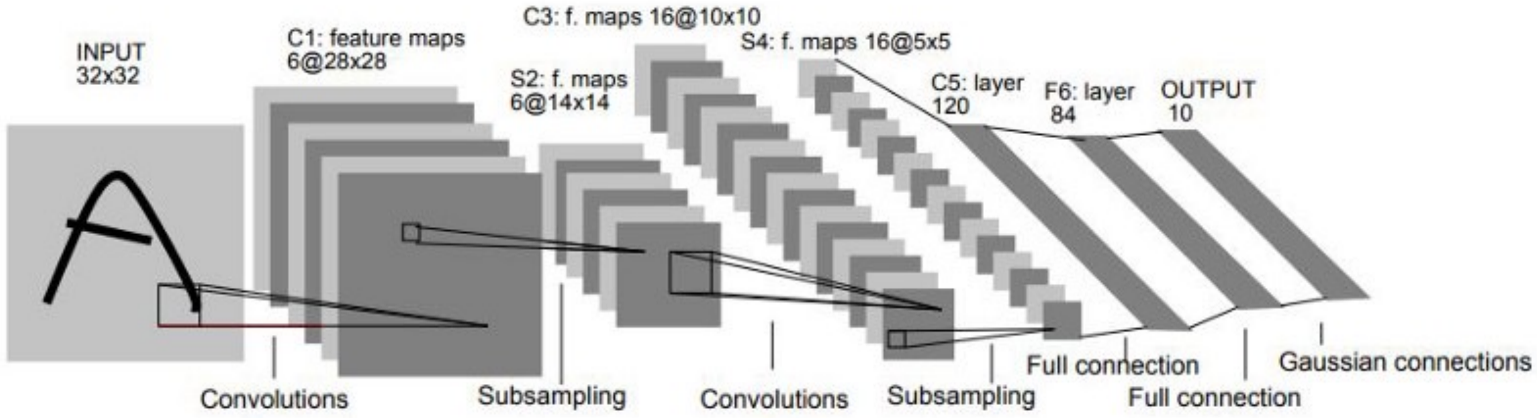


AT&T DSP-32C
capable of 125m floating
point multiply-accumulate
operations/sec

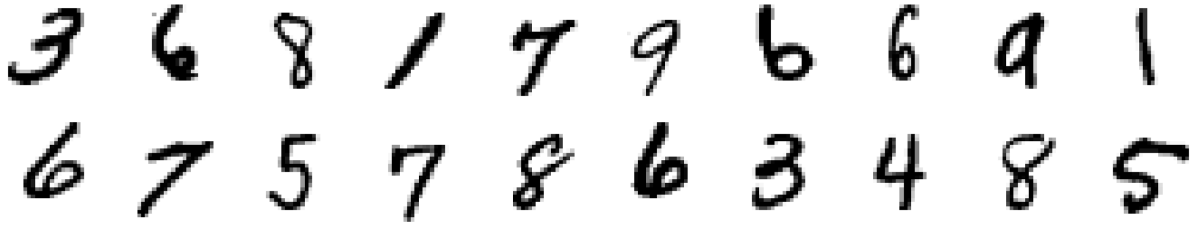
Y. LeCun

LeCun et al. 1989

LeNet-5



LeNet-5 classical CNN architecture

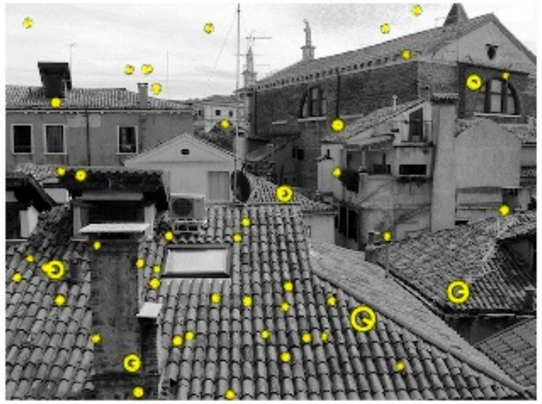


MNIST digits dataset

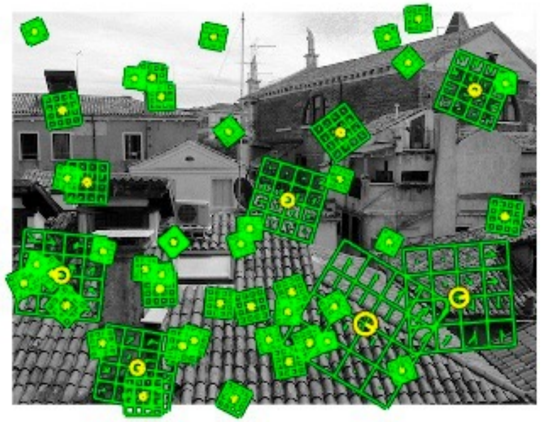
Y. LeCun

LeCun et al. 1998

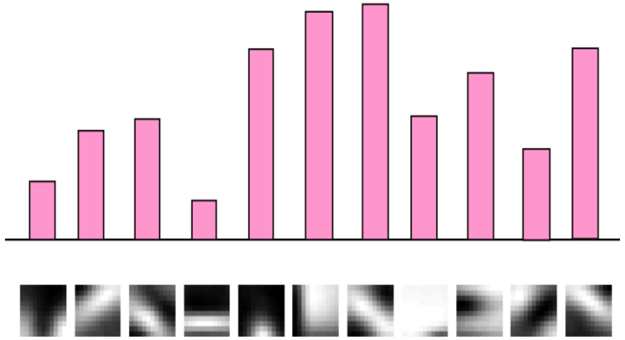
Computer vision in the 2000s



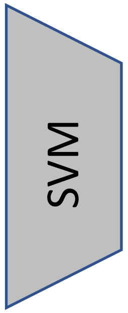
Feature detection



Feature description



Feature aggregation

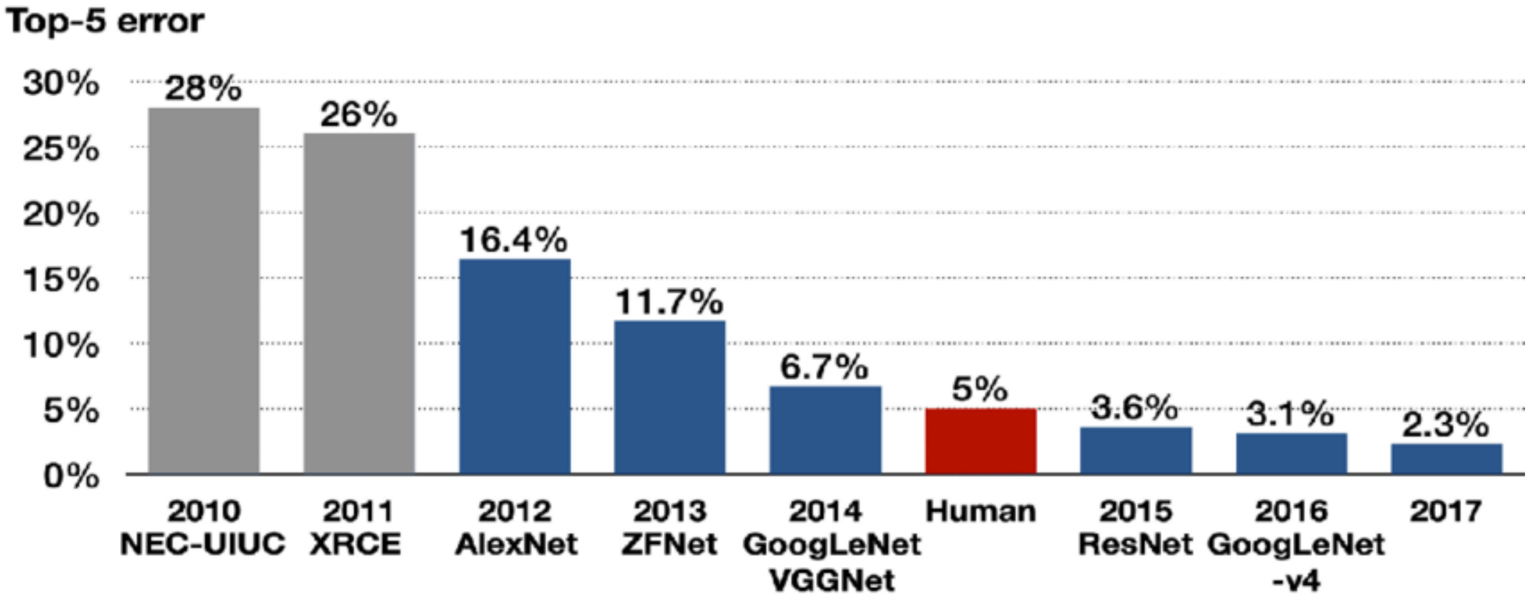


Classification

A typical image classification pipeline from the 2000s

Sivic, Zisserman 2003

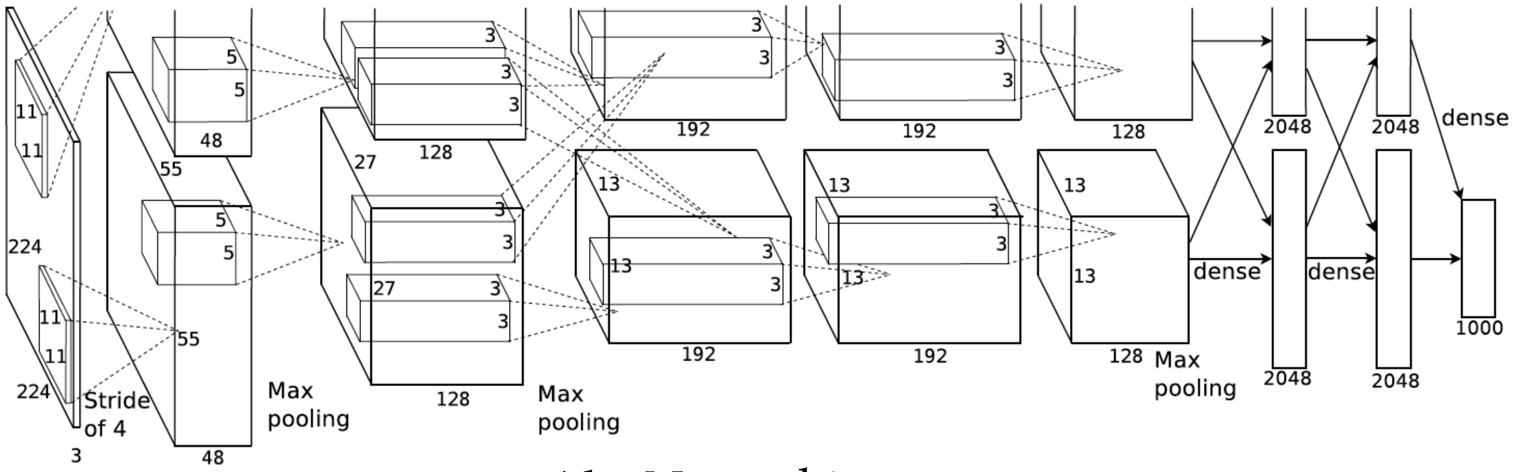
ImageNet



L. Fei-Fei

AlexNet beating all "handcrafted" approaches on ImageNet benchmark—the moment of truth for computer vision

AlexNet

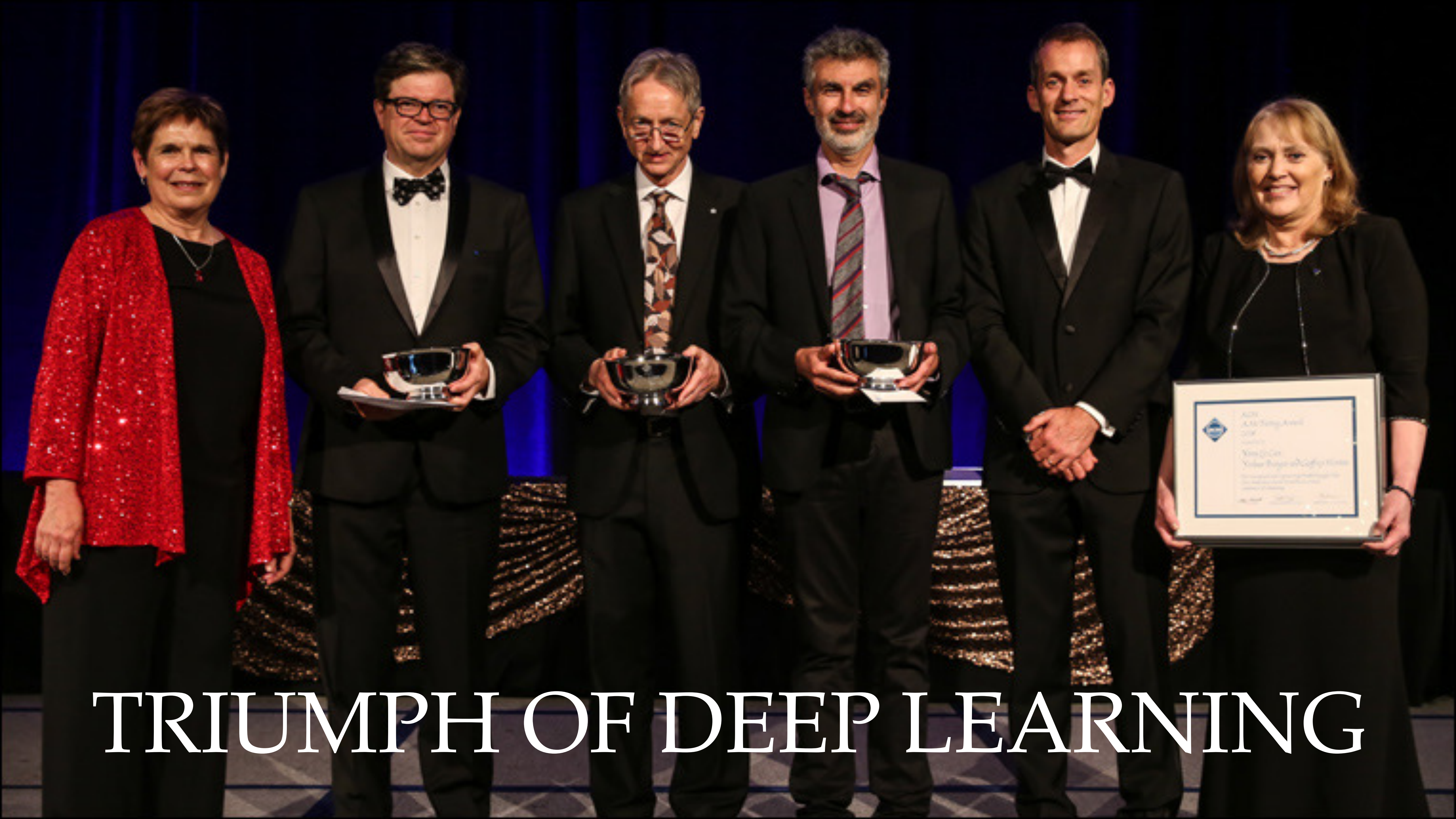


AlexNet architecture

Nvdia GTX 580 GPU capable of
~200G FLOP/sec

A. Krizhevsky

Krizhevsky et al. 2012


TRIUMPH OF DEEP LEARNING

GRAPH NEURAL NETWORKS & THEIR CHEMICAL PRECURSORS

Early chemoinformatics

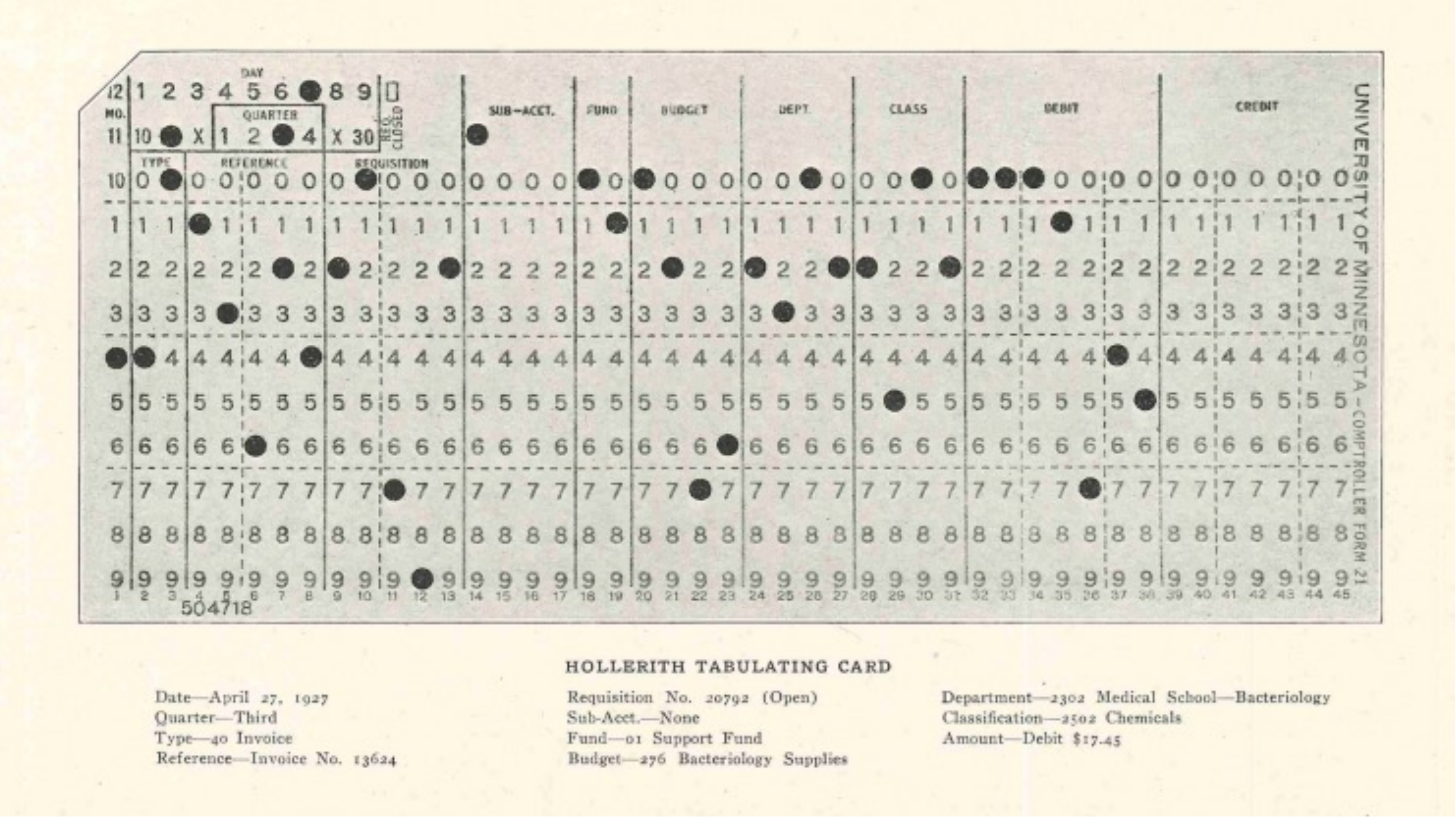
First chemical abstracts journal
Chemisches Zentralblatt 1830–1969

Beilstein Handbusch
~500 volumes, 500k pages

CAS
A division of the
American Chemical Society

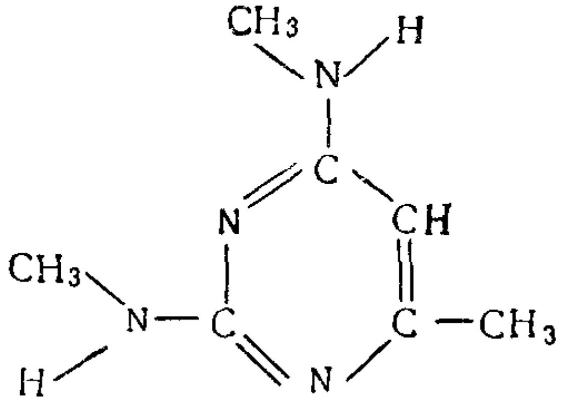
Chemical Abstracts Service
as of today ~200m compounds

Early chemoinformatics



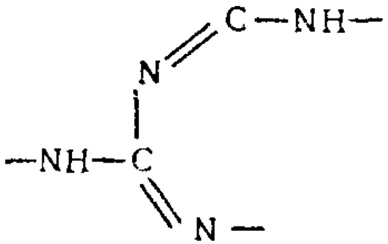
Punch card for early computer

Structural similarity of molecules



II

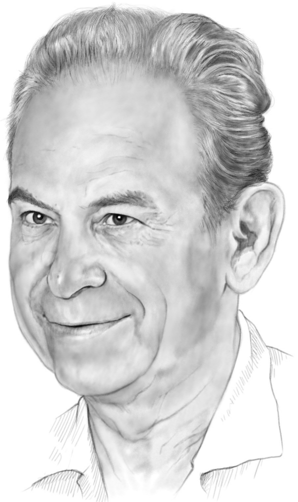
$H^1N^3J^1C^3NC^3J^1MC^3NX3N^3H^1J^1$



a)

$H^1N^3R^1C^3R^1NC^3NR^1N^3H^1R^1$

Early “chemical ciphers” used for molecule representations
fail to capture structural similarity

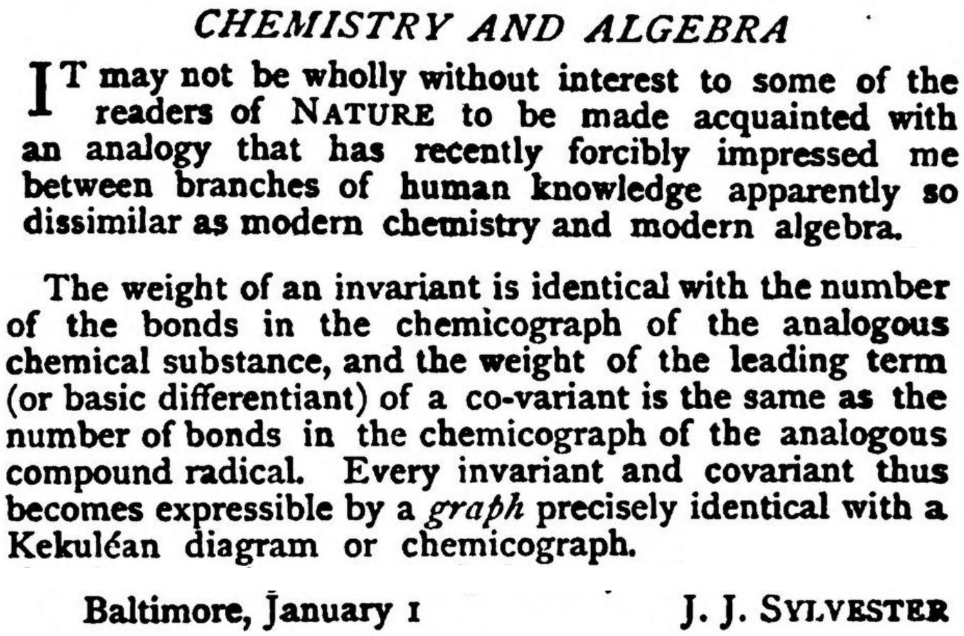


G. Vlăduţ

1959

Vlăduţ et al. 1959; Portrait: Ihor Gorskiy (from a photo courtesy of Serge Vlăduţ)

Graph theory & Chemistry

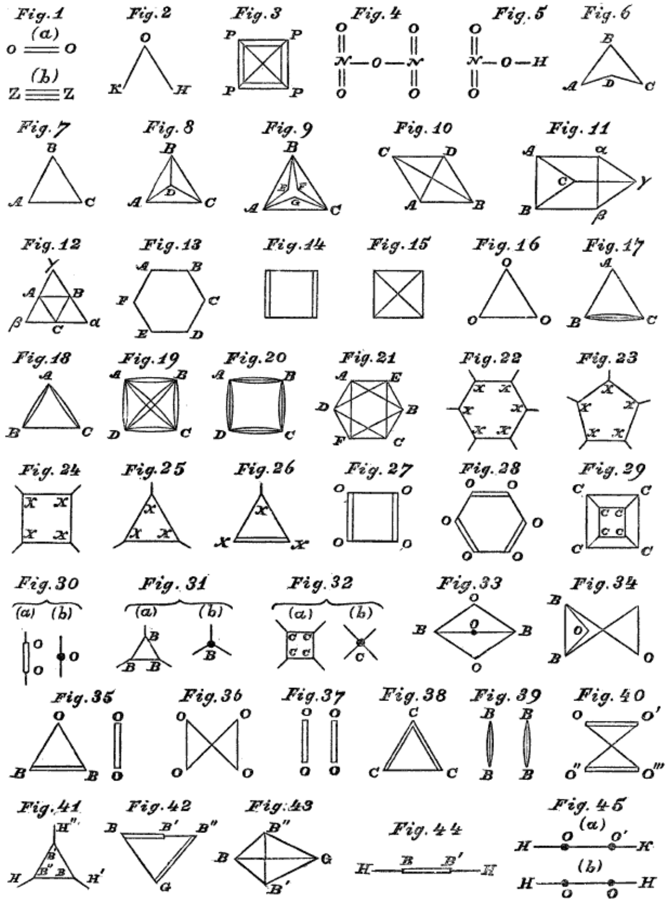


CHEMISTRY AND ALGEBRA

IT may not be wholly without interest to some of the readers of NATURE to be made acquainted with an analogy that has recently forcibly impressed me between branches of human knowledge apparently so dissimilar as modern chemistry and modern algebra.

The weight of an invariant is identical with the number of the bonds in the chemicograph of the analogous chemical substance, and the weight of the leading term (or basic differentiant) of a co-variant is the same as the number of bonds in the chemicograph of the analogous compound radical. Every invariant and covariant thus becomes expressible by a *graph* precisely identical with a Kekuléan diagram or chemicograph.

Baltimore, January 1 J. J. SYLVESTER



J. Sylvester

1878

The term "graph" appeared first in the chemical context

Weisfeiler-Lehman test

УДК 519.1

ПРИВЕДЕНИЕ ГРАФА К КАНОНИЧЕСКОМУ ВИДУ И ВОЗНИКАЮЩАЯ ПРИ ЭТОМ АЛГЕБРА

Б. Ю. ВЕЙСФЕЙЛЕР, А. А. ЛЕМАН

Рассматривается алгоритм приведения заданного конечного мультиграфа Г к каноническому виду. В процессе такого приведения возникает новый инвариант графа— алгебра $\mathfrak{A}$ (Г). Изучение свойств алгебры $\mathfrak{A}$ (Г) оказывается полезным при решении некоторых задач теории графов.

Выдвигаются и обсуждаются некоторые предположения относительно связи между свойствами алгебры $\mathfrak{A}$ (Г) и группой автоморфизмов графа Aut (Г). Построен пример неориентированного графа Г, алгебра $\mathfrak{A}$ (Г) которого совпадает с групповой алгеброй некоторой некоммутативной группы.

An algorithm is considered, reducing the specified finite multigraph Г to canonical form. In the course of this reduction, a new invariant of the graph is generated — algebra $\mathfrak{A}$ (Г). Study of the properties of the algebra $\mathfrak{A}$ (Г) proves helpful in solving a nu mber of graph-theoretic problems. Some propositions concerning the relationships between the properties of the algebra $\mathfrak{A}$ (Г) and the graph's automorphism group *Aut* (Г) are discussed. An example of non—oriented graph Г is constructed whose algebra $\mathfrak{A}$ (Г) coincides with the group algebra of a non—commutative group.

1. Рассмотрим произвольный конечный граф Г и его матрицу смежности $A(\Gamma)=\{a_{ij}\}$; здесь a_{ij}—число ребер, ведущих из i-й вершины графа в j-ую; $i, j=1, 2, \ldots, n$. В случае неориентированного графа полагаем $a_{ij}=a_{ji}$. Каноническим видом графа мы будем называть его матрицу смежности при канонической нумерации вершин, т. е. при таком частичном упорядочении множества вершин, при котором из того, что a и b несравнимы следует, что существует автоморфизм графа, переводящий вершину a в b и сохраняющий отношение смежности.

В п. 6, 7 описан процесс приведения графа к каноническому виду, состоящий в поэтапном переупорядочении строк и столбцов матрицы $A(\Gamma)$, который, грубо говоря, сводится к следующему.

Рассмотрим для простоты неориентированный граф без кратных связей. Сначала каждой вершине графа сопоставим характеристический вектор, единственная компонента которого равна числу соседей данной вершины. Затем разобьем вершины на классы, так чтобы вершины с одинаковым характеристическим вектором принадлежали одному классу; классы при этом упорядочим в соответствии с естественным порядком в множестве характеристических векторов. Далее, каждой вершине сопоставим характеристический вектор $v_i=\{l, v_{i1}, v_{i2}, \ldots\}$, где v_{ik}—число соседей k-го класса у i-ой вершины, l—номер класса, к которому принадлежит i-ая вершина. Теперь снова разобьем вершины на классы в соответствии с новыми характеристическими векторами, упорядоченными лексикографически, и т. д. Заметим, что если вершины a и b на некотором шаге принадлежали разным классам и было выполнено условие $a<b$, то и в дальнейшем это условие всегда будет выполнено. Отсюда следует, что описанный процесс обрывается не позже, чем через n шагов—и либо все вершины относятся к разным классам (т. е. построена каноническая нумерация), либо дальнейшего разбиения на классы не происходит.

В случае, если Г—ориентированный мультиграф, возьмем в качестве характеристического вектора v_i упорядоченную i-ую строку матрицы $A(\Gamma)$ (считая при этом, что диагональный элемент предшествует всем остальным). Вместо различных элементов a_{ij} введем различные независимые переменные $x_1, x_2, \ldots$, упорядочив их в соответствии с порядком среди a_{ij}. Полученную таким способом матрицу обозначим $X=X(\Gamma)$. При очередном разбиении вершин на классы к одному классу, как и прежде, отнесем вершины с одинаковым характеристическим вектором; при этом k-ая компонента вектора v_i есть по определению сумма элементов i-ой строки матрицы $X(\Gamma)$, соответствующих вершинам k-го класса (предыдущего разбиения). Матрица $X(\Gamma)$ разбивается, таким образом, на блоки, в каждом из которых мы можем ввести свои независимые переменные и т. д. (точное определение этих операций см. п. 6, операции α_1, β_1).

Заметим, что описанная до сих пор процедура аналогична методам, изложенным в [1] и [2].

Для дальнейшего разбиения вершин на классы рассмотрим элемент u_{ij} матрицы $U=X\cdot X'$, где X'—матрица, полученная из X заменой переменных $x_1, x_2, \ldots$ переменными $x'_1, x'_2, \ldots$, причем все переменные $x_1, x_2, \ldots, x'_1, x'_2, \ldots$ независимы. Элемент u_{ij} является многочленом второй степени от $x_1, x_2, \ldots, x'_1, x'_2, \ldots$ Если теперь обозначить различные многочлены различными новыми независимыми переменными, то к полученной матрице снова можно будет применять все описанные выше операции и т. д., до тех пор, пока и этот процесс не оборвется (см. п. 6, операция α_2, β_2, β_3).

2. Геометрически введение независимых переменных и матрицы $U=X\cdot X'$ означает следующее. Если наряду с ребрами данного графа Г рассматривать ребра дополнительного графа $\overline{\Gamma}$, то на первом же шаге нашей процедуры этим ребрам будут соответствовать наверняка разные переменные— т. е. произойдет раскраска ребер данного и дополнительного графов в разные цвета. В дальнейшем каждое новое введение дополнительных переменных вводит новую раскраску ребер, причем 1) ребра, окрашенные на каком-то шаге по разному, в дальнейшем также будут окрашены по-разному; 2) разбиение вершин на классы производится в соответствии с количеством ребер каждого цвета, исходящих из вершины.

Известно, далее, что элемент $a_{ij}\in A^2$, $A=A(\Gamma)$, где Г—неориентированный граф без кратных связей, равен числу путей длины 2, ведущих из i-й вершины в j-ую. Аналогично, коэффициент при $x_k x'_l$ в многочлене $u_{ij}\in U=X\cdot X'$ равен числу путей, ведущих из i-й вершины в j-ую по ребрам сначала k-го, а затем l-го цвета.

3). Дальнейшая процедура приведения графа к каноническому виду использует применение описанных выше операций к матрице, полученной из $X(\Gamma)$ вычеркиванием некоторого ее столбца и соответствующей строки. Если для матриц порядка $k\leqslant n-1$ приведение к каноническому виду определено, то возникает новая возможность для разбиения вершин на классы: старшей считается та вершина, после вычеркивания которой получается лексикографически старший канонический вид оставшегося графа (см. операции α_3, α_4, β_4, Ы п. 7). Очевидно, что и такое разбиение на некотором шаге оборвется. Доказано, что любые две вершины a и b, отнесенные к одному классу при последнем разбиении, эквивалентны, т. е. существует автоморфизм графа Г, переводящий a в b и сохраняющий отношения смежности.

Рассмотрим снова такую матрицу $X=X(\Gamma)$, что в матрице $U=X\cdot X'$ на месте одинаковых переменных матрицы X стоят одинаковые многочлены. Матрица $X(\Gamma)$ является тогда общей точкой некоторой матричной алгебры $\mathfrak{A}(\Gamma)$, т. е. если задано кольцо K (например, кольцо целых чисел $\mathbb{Z}$ или поле рациональных чисел Q), то множество матриц, полученных подстановкой в X вместо ее переменных элементов кольца K образует алгебру $\mathfrak{A}_K(\Gamma)=\mathfrak{A}(\Gamma)\otimes K$. Алгебра $\mathfrak{A}(\Gamma)$ является, очевидно, инвариантом графа. Некоторые соотношения между

A. Lehman

B. Weisfeiler

1968

Weisfeiler, Lehman 1968; Portraits: Ihor Gorskiy

First Graph Neural Networks

A. Sperduti

C. Goller

A. Küchler

M. Gori

F. Scarselli

A. Micheli

Labeling RAAM

1994

Backprop through structure

1996

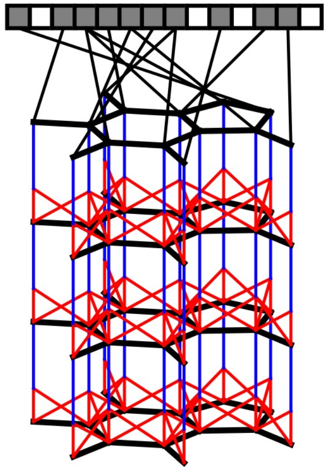
"Graph Neural Networks"

2005, 2008

"NN4G"

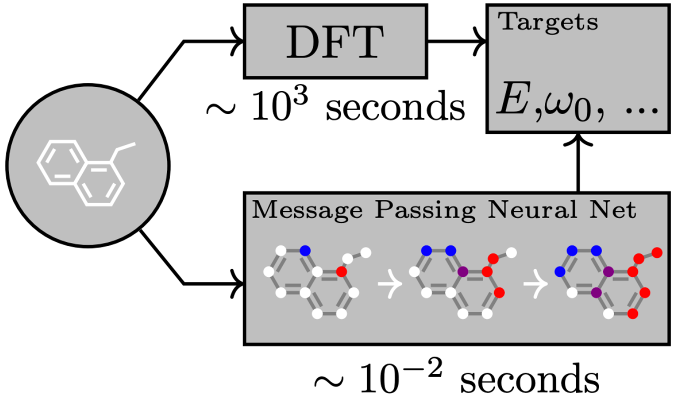
2009

Back to the chemical roots



GNN-based
molecular fingerprints

D. Duvenaud

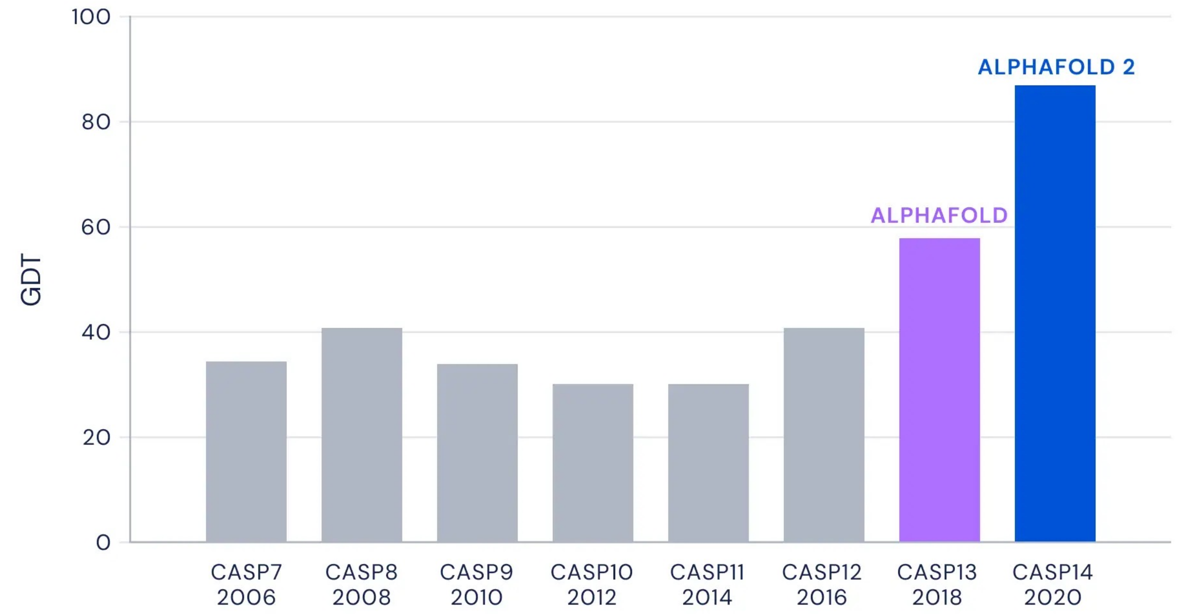


Chemical property prediction
using message passing GNNs

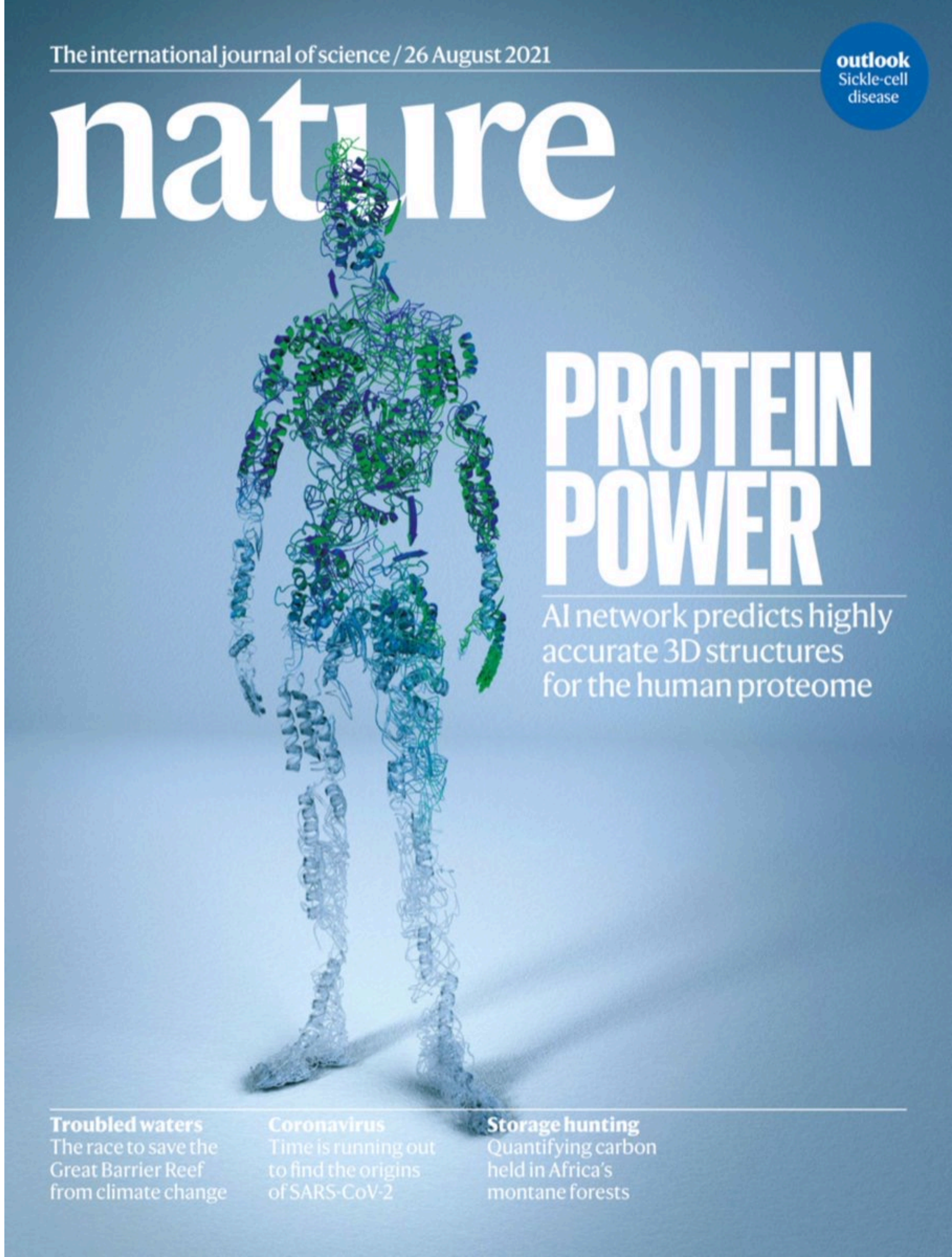
J. Gilmer

Duvenaud et al. 2015; Gilmer et al. 2017

Back to the chemical roots



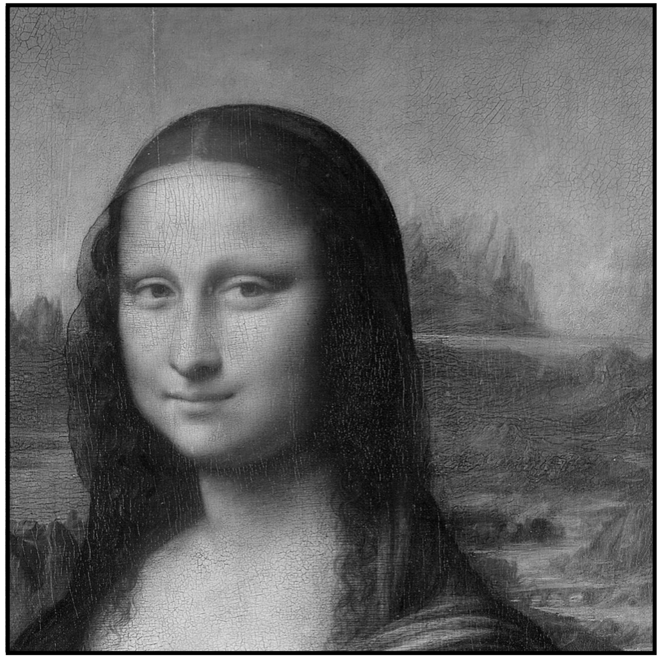
An "ImageNet" moment of structural biology



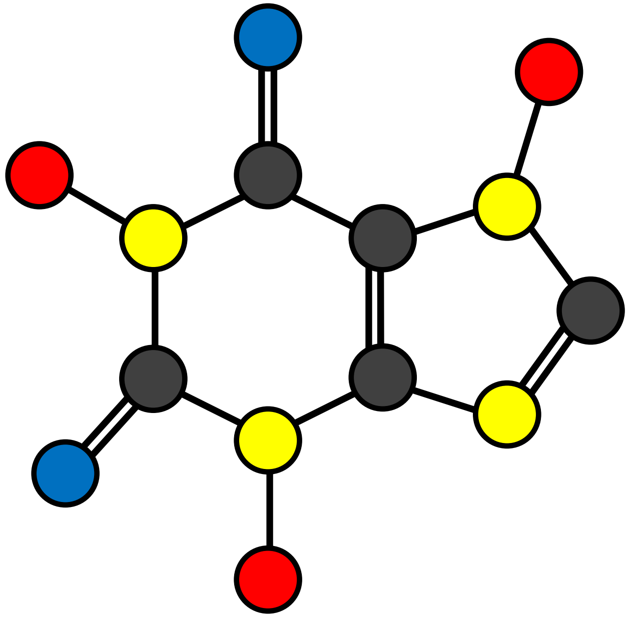
Jumper et al. 2021

THE BLUEPRINT

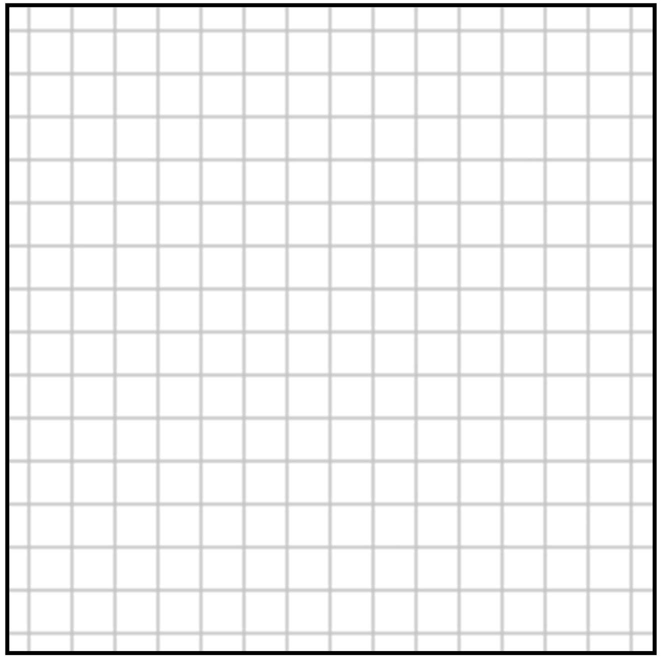
Convolutional Neural Network



Graph Neural Network

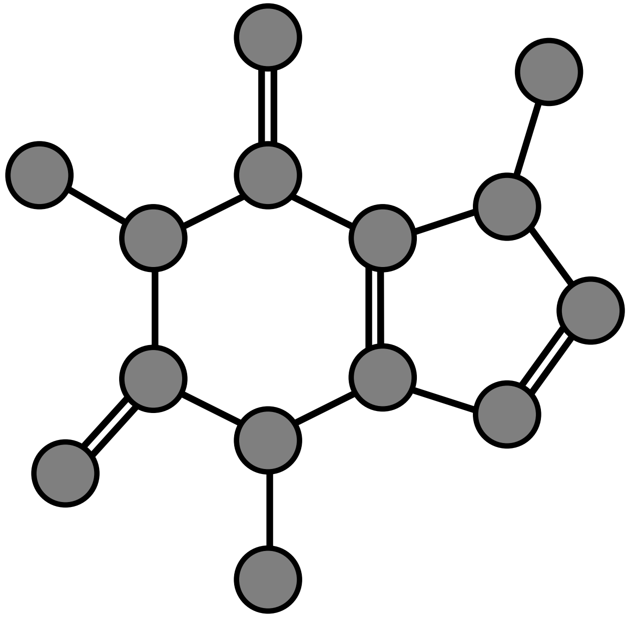


Convolutional Neural Network



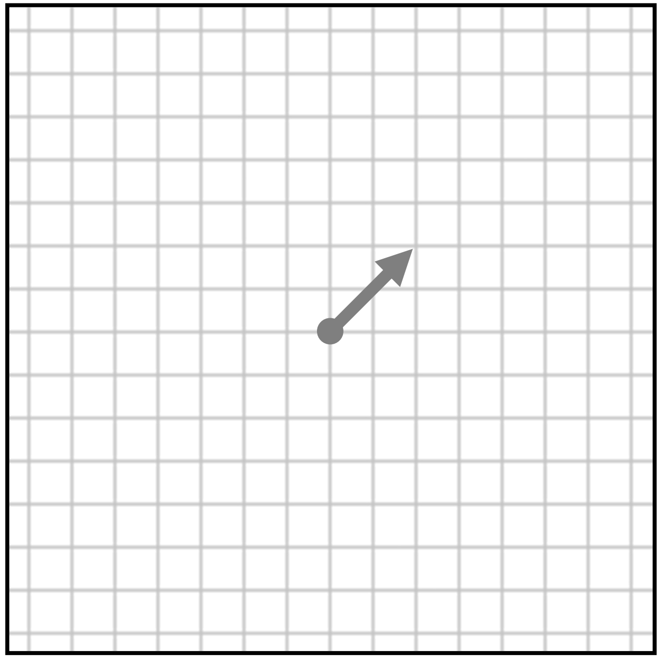
Underlying domain:
grid

Graph Neural Network



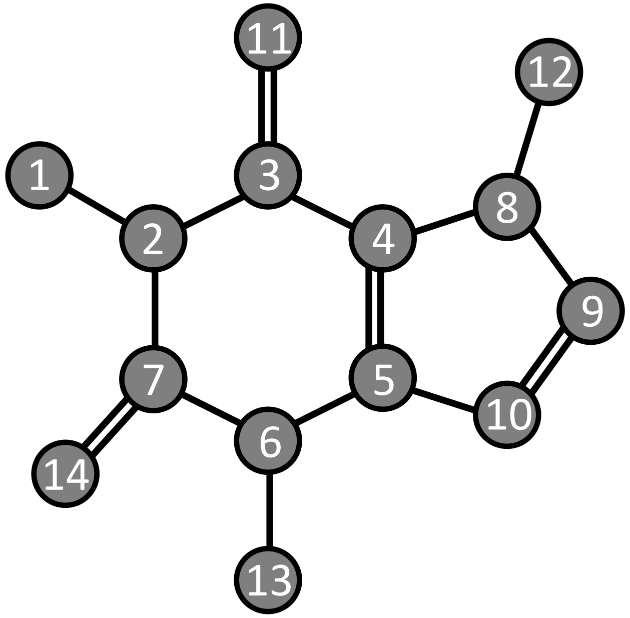
Underlying domain:
graph

Convolutional Neural Network



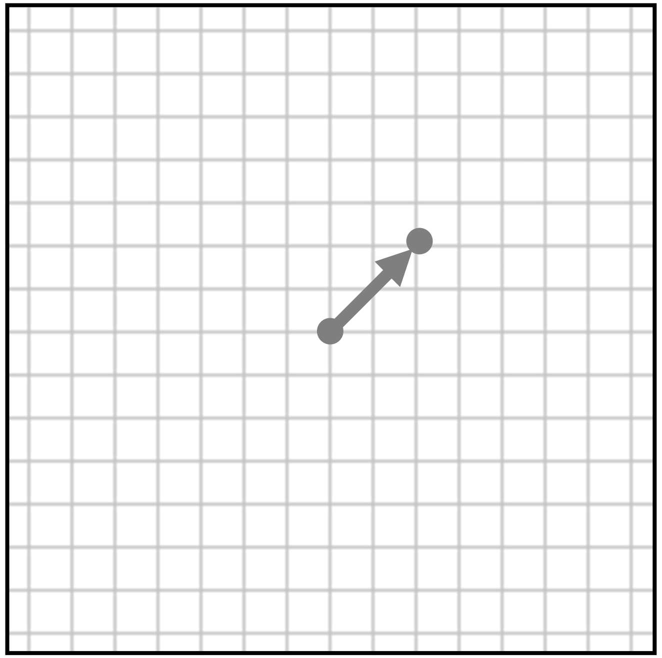
Symmetry:
Translation

Graph Neural Network



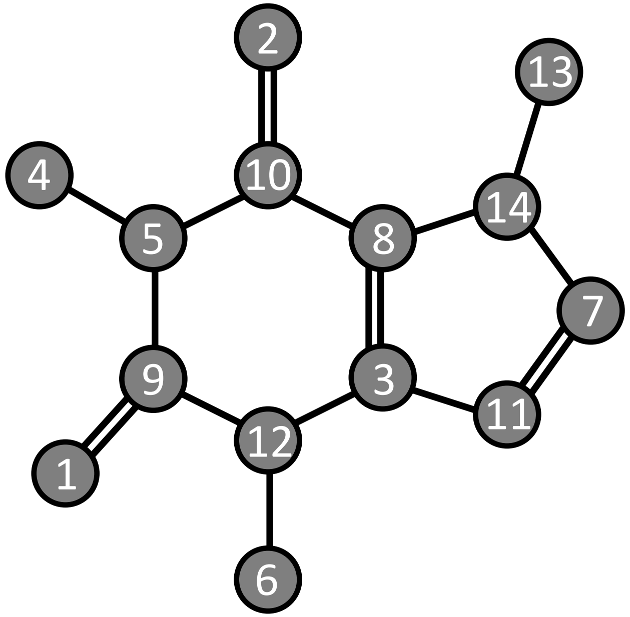
Symmetry:
Permutation

Convolutional Neural Network



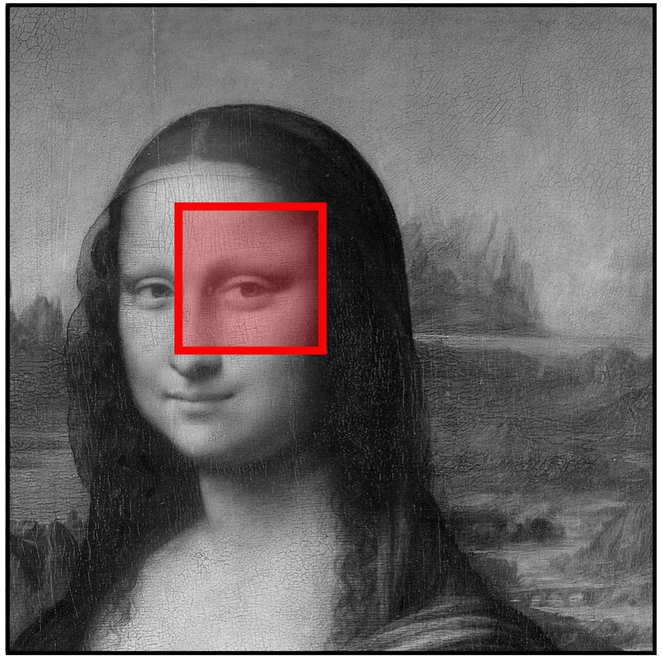
Symmetry:
Translation

Graph Neural Network



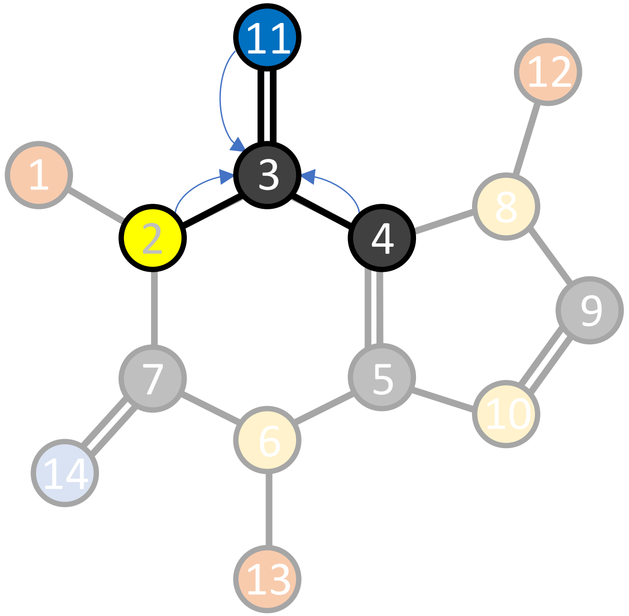
Symmetry:
Permutation

Convolutional Neural Network



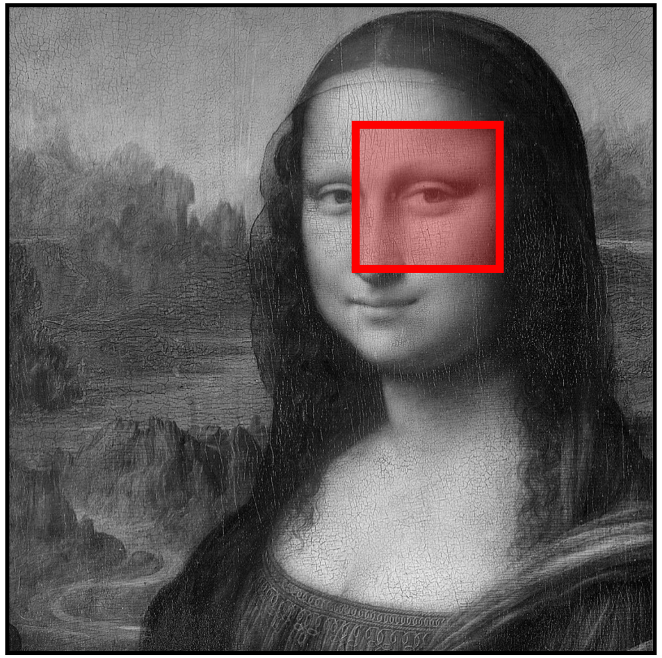
Convolution:
translation equivariant

Graph Neural Network



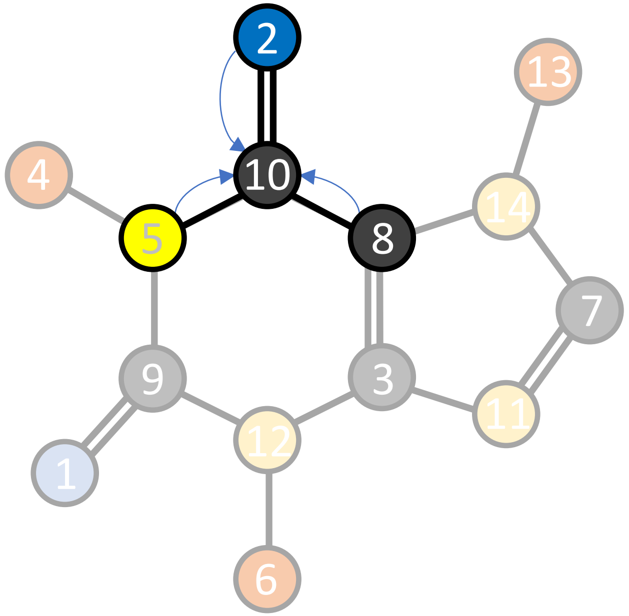
Message passing:
permutation equivariant

Convolutional Neural Network

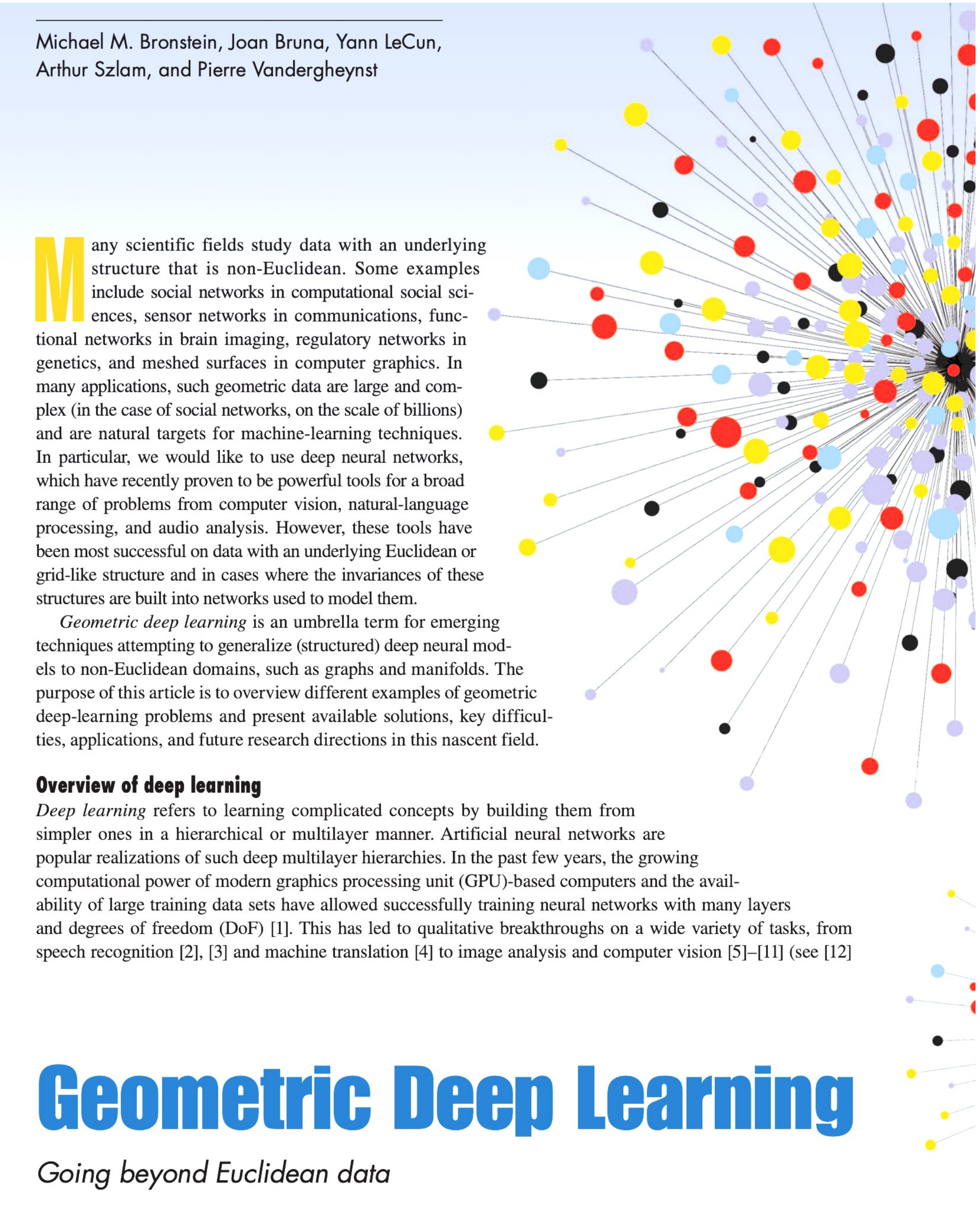


Convolution:
translation equivariant

Graph Neural Network



Message passing:
permutation equivariant



Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst

Many scientific fields study data with an underlying structure that is non-Euclidean. Some examples include social networks in computational social sciences, sensor networks in communications, functional networks in brain imaging, regulatory networks in genetics, and meshed surfaces in computer graphics. In many applications, such geometric data are large and complex (in the case of social networks, on the scale of billions) and are natural targets for machine-learning techniques. In particular, we would like to use deep neural networks, which have recently proven to be powerful tools for a broad range of problems from computer vision, natural-language processing, and audio analysis. However, these tools have been most successful on data with an underlying Euclidean or grid-like structure and in cases where the invariances of these structures are built into networks used to model them.

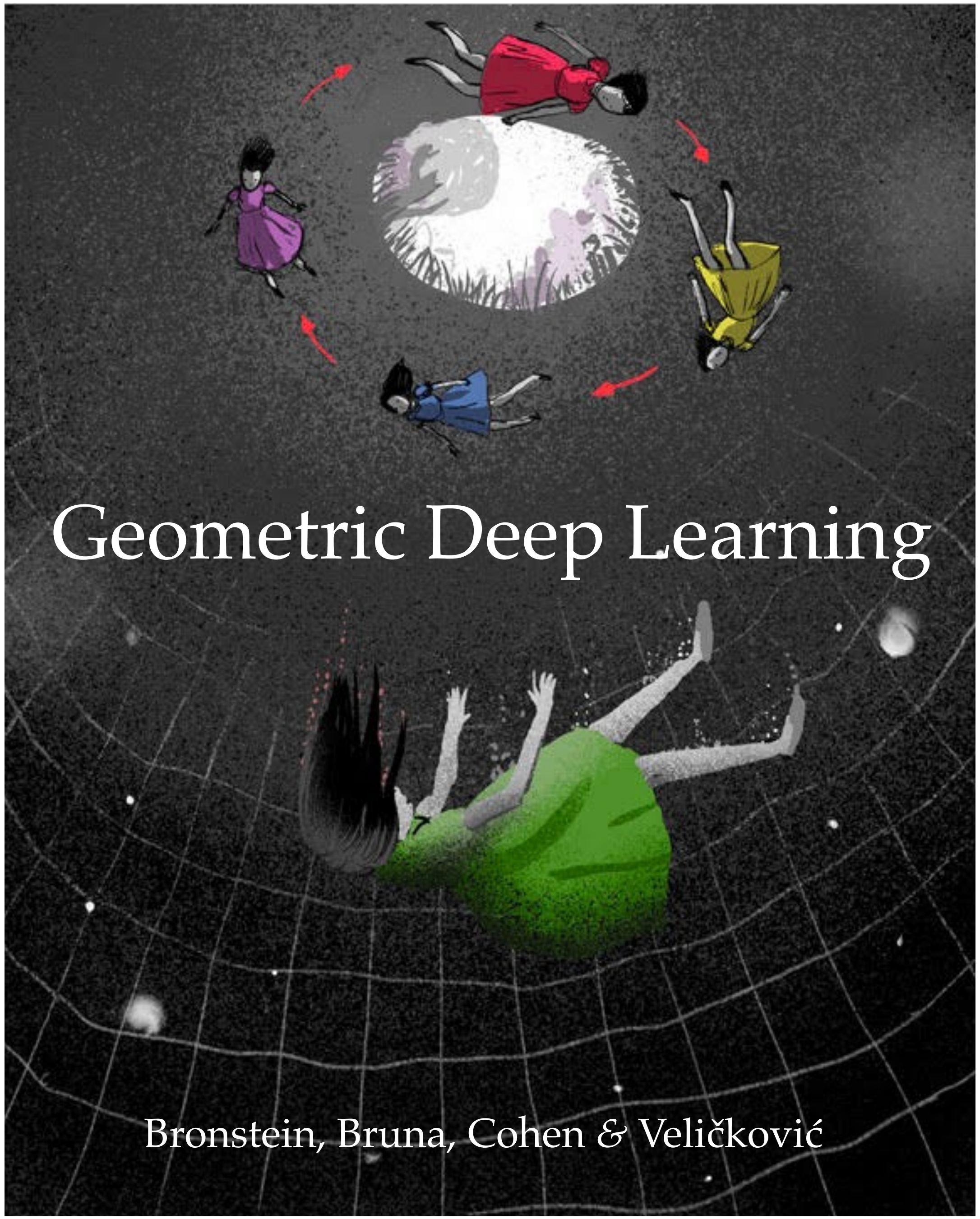
Geometric deep learning is an umbrella term for emerging techniques attempting to generalize (structured) deep neural models to non-Euclidean domains, such as graphs and manifolds. The purpose of this article is to overview different examples of geometric deep-learning problems and present available solutions, key difficulties, applications, and future research directions in this nascent field.

Overview of deep learning

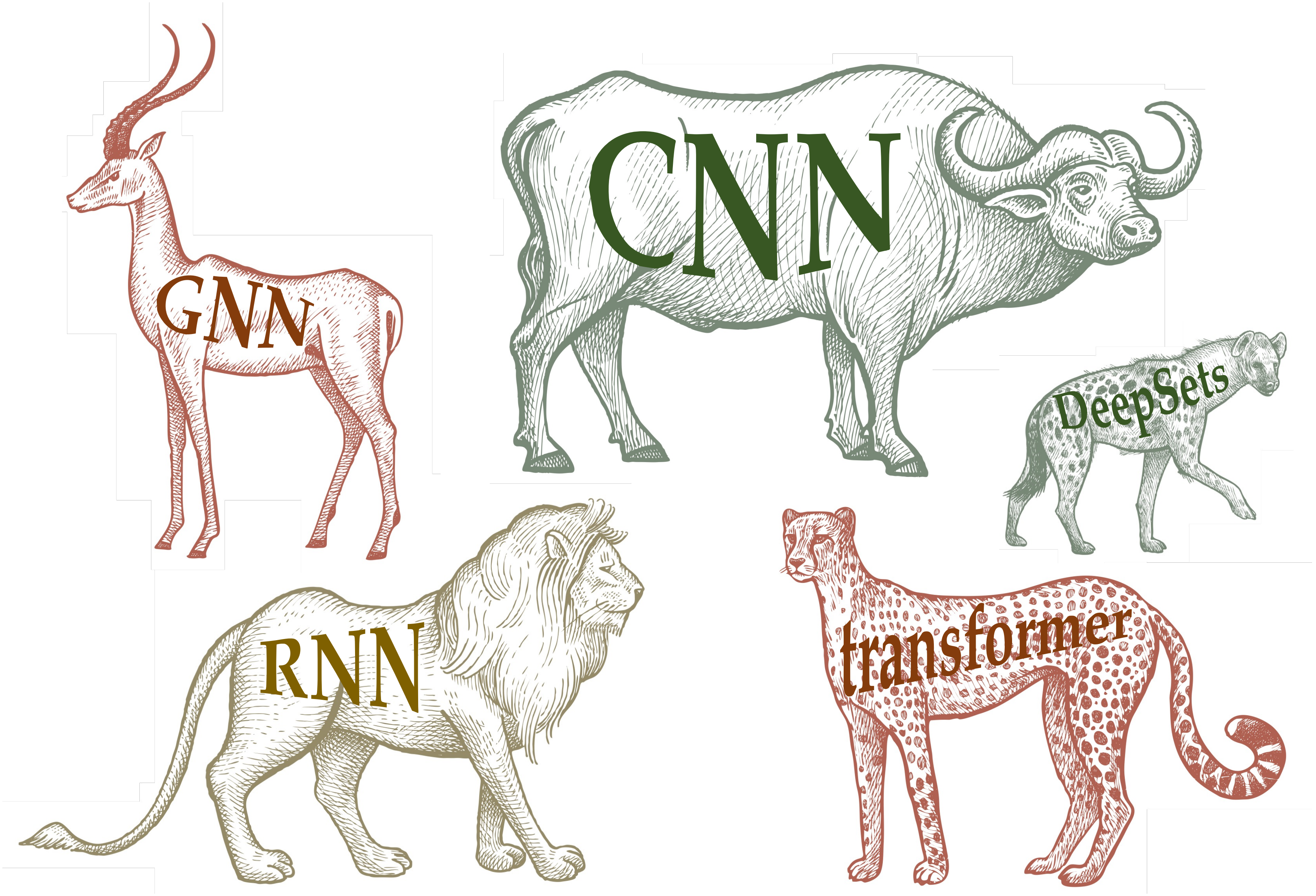
Deep learning refers to learning complicated concepts by building them from simpler ones in a hierarchical or multilayer manner. Artificial neural networks are popular realizations of such deep multilayer hierarchies. In the past few years, the growing computational power of modern graphics processing unit (GPU)-based computers and the availability of large training data sets have allowed successfully training neural networks with many layers and degrees of freedom (DoF) [1]. This has led to qualitative breakthroughs on a wide variety of tasks, from speech recognition [2], [3] and machine translation [4] to image analysis and computer vision [5]–[11] (see [12]

Geometric Deep Learning

Going beyond Euclidean data



Twentieth Century Zoo of Neural Network Architectures

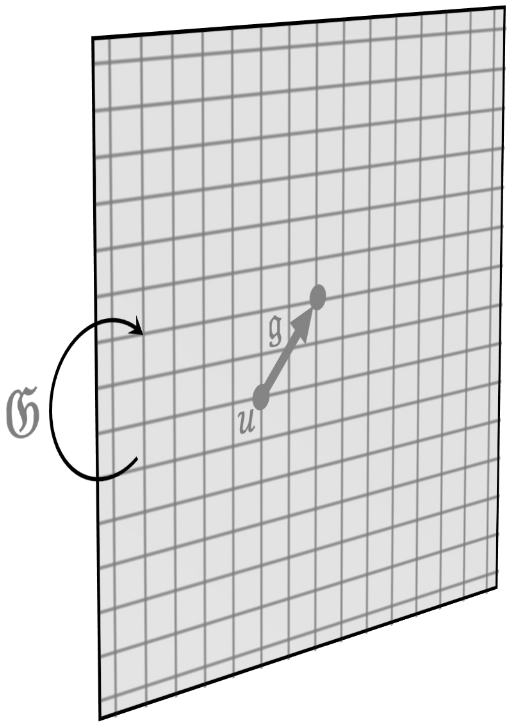


The Erlangen Programme of ML

Geometric Deep Learning

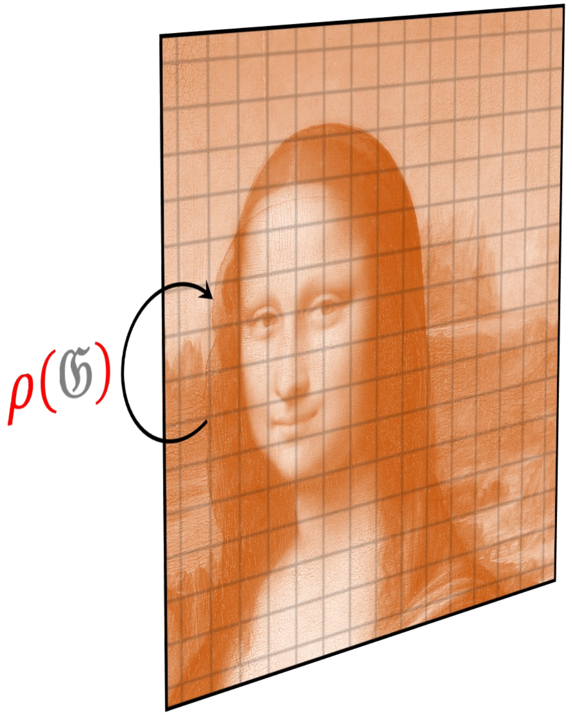
Geometric Deep Learning Blueprint

domain Ω



symmetry group $\mathfrak{G}$

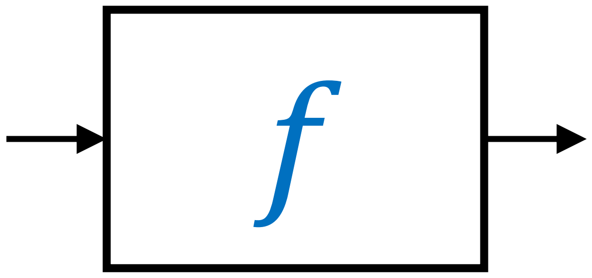
signals $\mathcal{X}(\Omega)$



group representation $\rho(\mathfrak{G})$

$$\rho(\mathfrak{g})x(u) = x(\mathfrak{g}^{-1}u)$$

functions $\mathcal{F}(\mathcal{X}(\Omega))$



equivariance

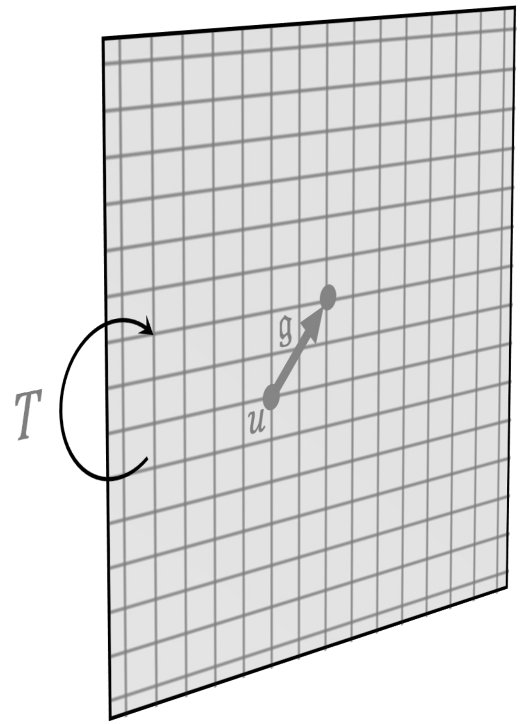
$$f(\rho(\mathfrak{g})x) = \rho(\mathfrak{g})f(x)$$

invariance

$$f(\rho(\mathfrak{g})x) = f(x)$$

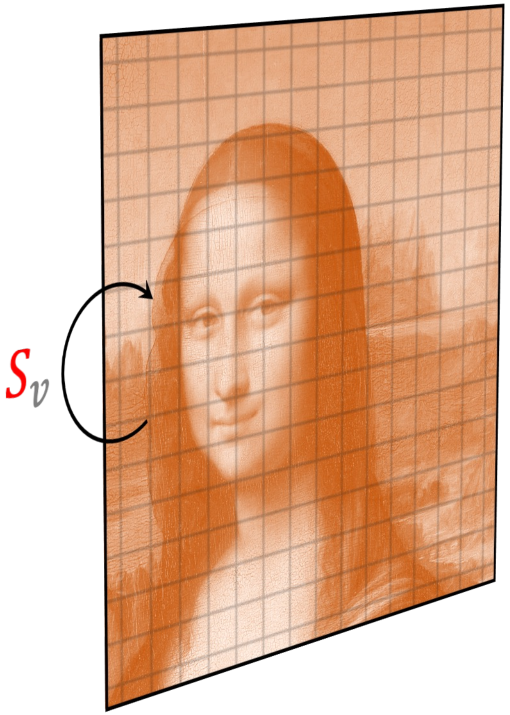
Example: Convolutional Neural Networks

Plane $\mathbb{R}^2$



Translation group $T(2)$

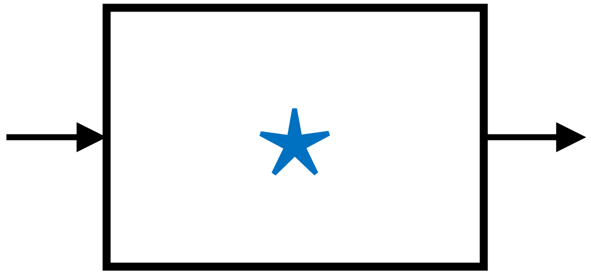
images $\mathcal{X}(\mathbb{R}^2)$



Shift operator S

$S_v x(u) = x(u - v)$

functions $\mathcal{F}(\mathcal{X}(\Omega))$

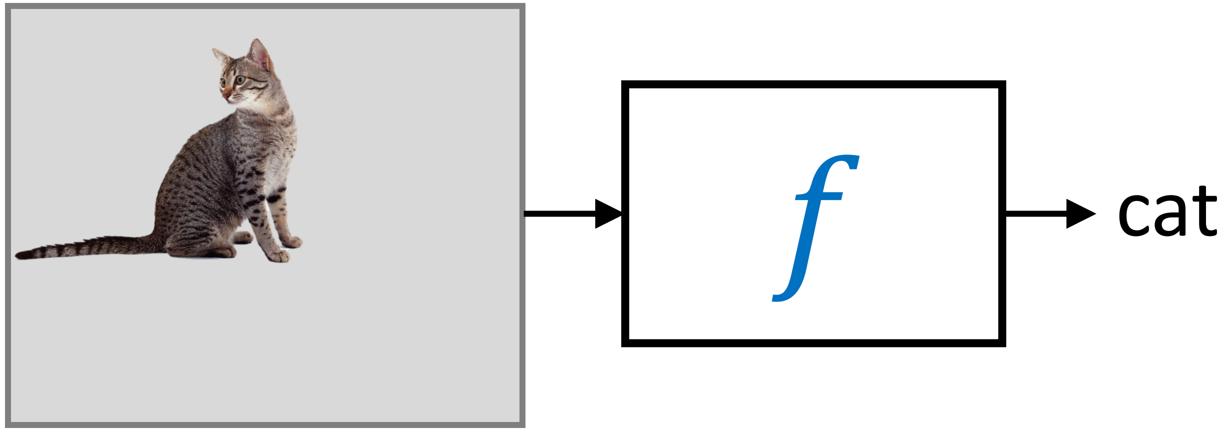


Convolutional layer

$(Sx \star y) = S(x \star y)$

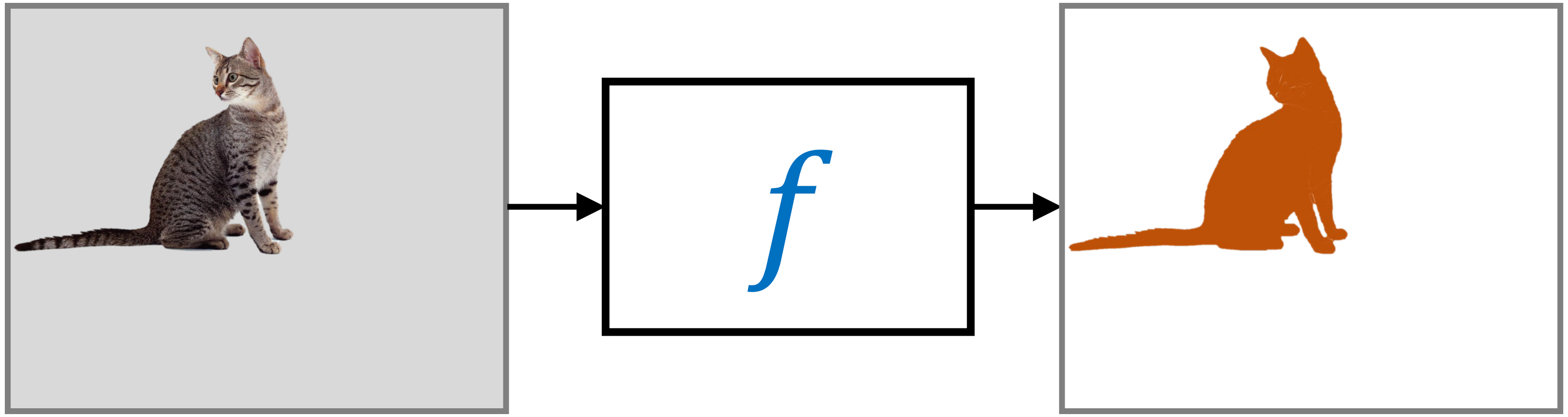
Invariant functions: Image Classification

$\mathfrak{G}$-invariance $\quad f(\rho(\mathfrak{g})x) = f(x)$



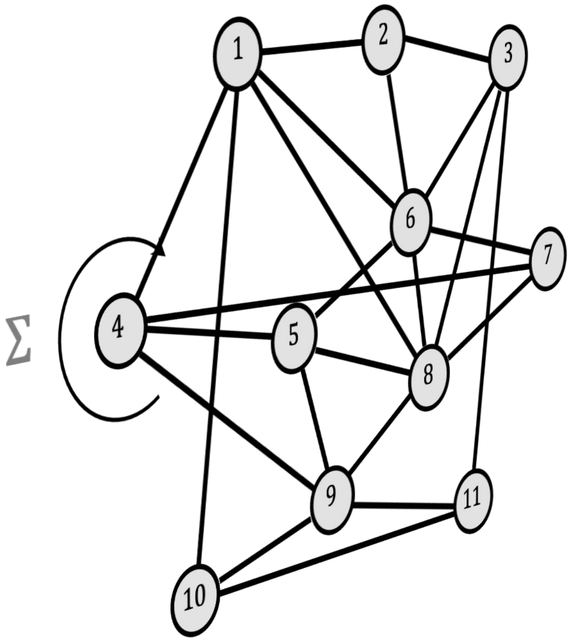
Equivariant functions: Image Segmentation

$\mathfrak{G}$-equivariance $\quad f(\rho(\mathfrak{g})x) = \rho(\mathfrak{g})f(x)$



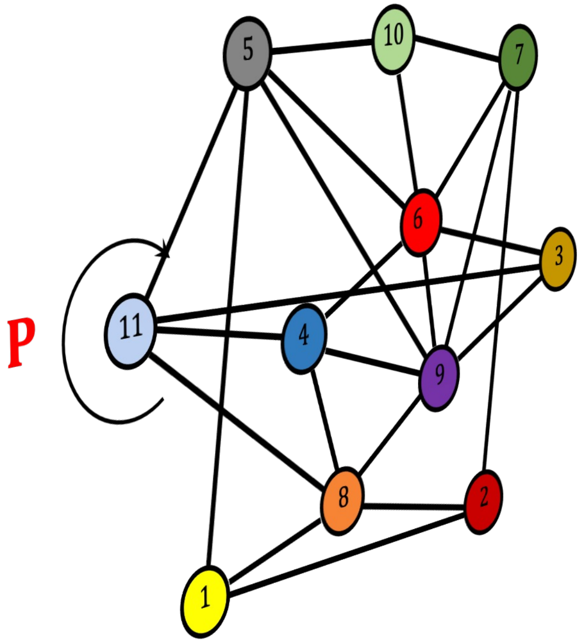
Example: Graph Neural Networks

Graph $G = (V, E)$



Permutation group Σ_n

Node features $\mathcal{X}(G)$



Permutation matrix P

$\mathbf{PX} = \left(x_{\pi^{-1}(i),j}\right)$

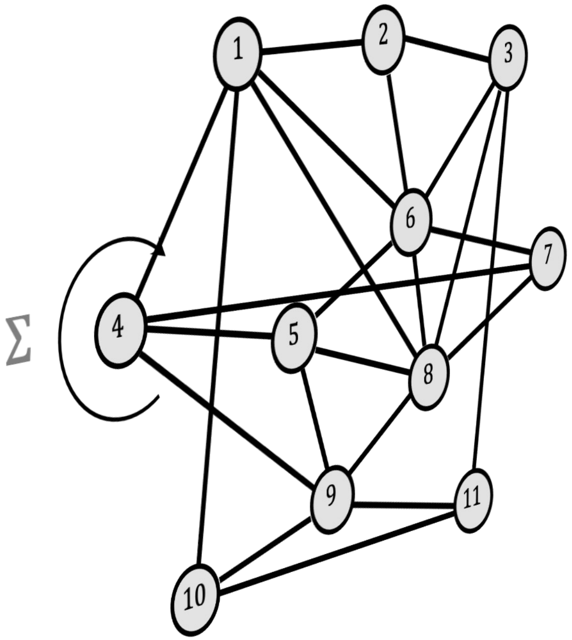
functions $\mathcal{F}(\mathcal{X}(\Omega))$

Message passing

$\mathbf{F}\left(\mathbf{PX}, \mathbf{PAP}^{\top}\right) = \mathbf{PF}(\mathbf{X}, \mathbf{A})$

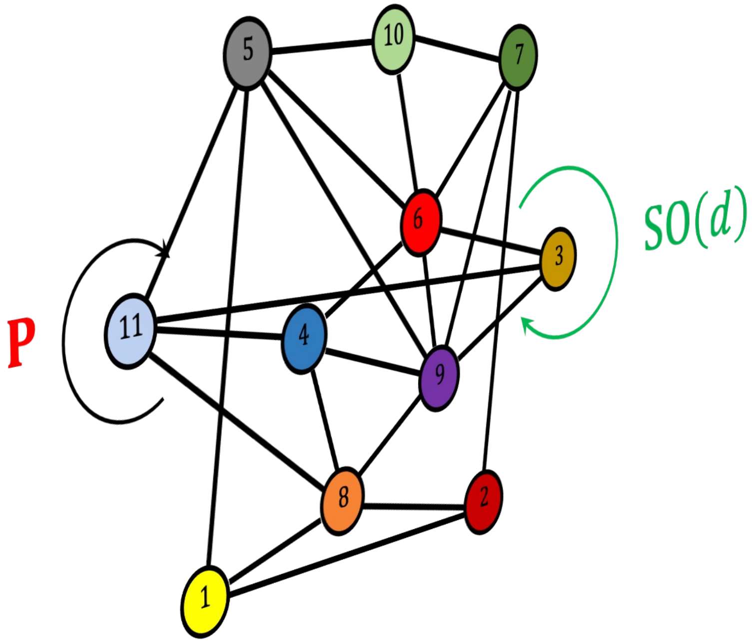
Example: Equivariant Graph Neural Networks

Graph $G = (V, E)$



Permutation group Σ_n

Node features $\mathcal{X}(G)$



Permutation matrix P

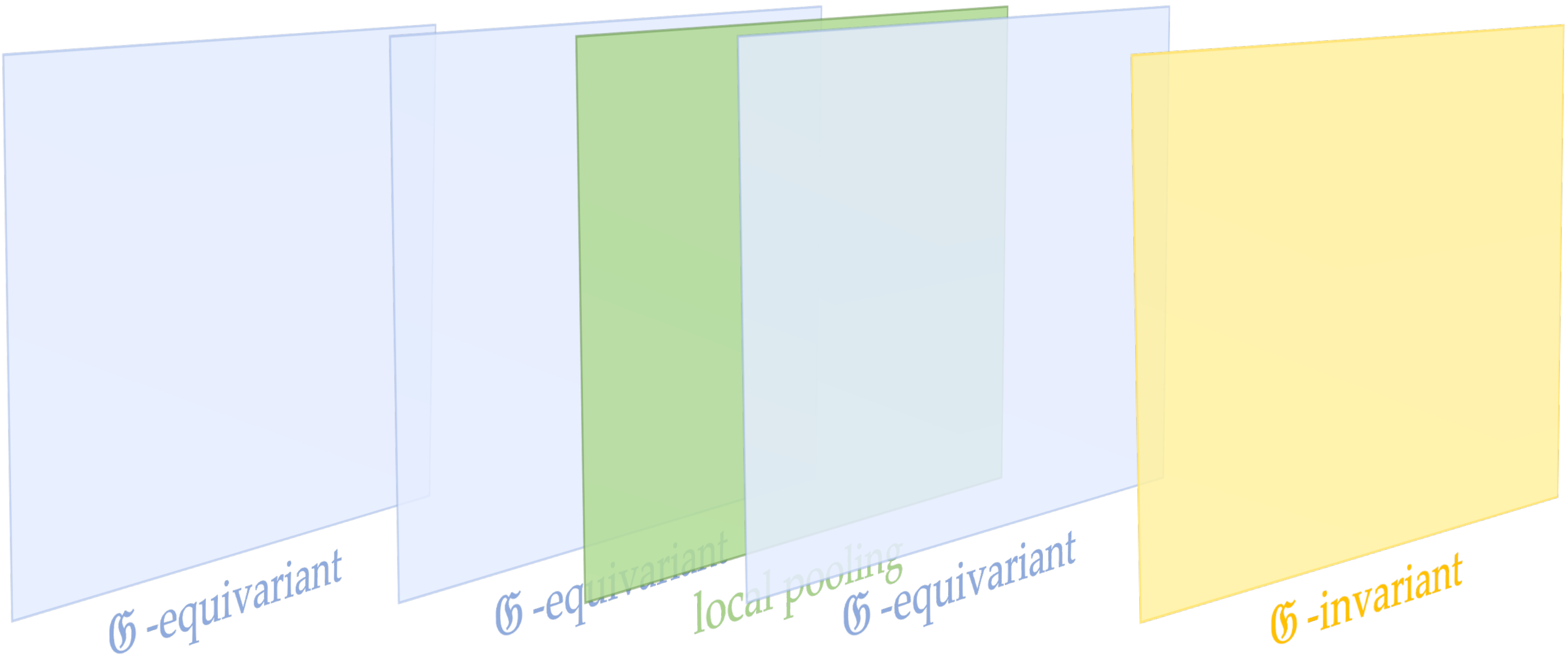
Rotation R

functions $\mathcal{F}(\mathcal{X}(\Omega))$

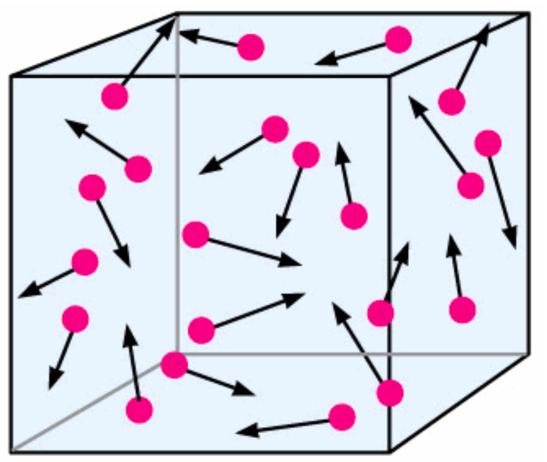
Equivariant message passing

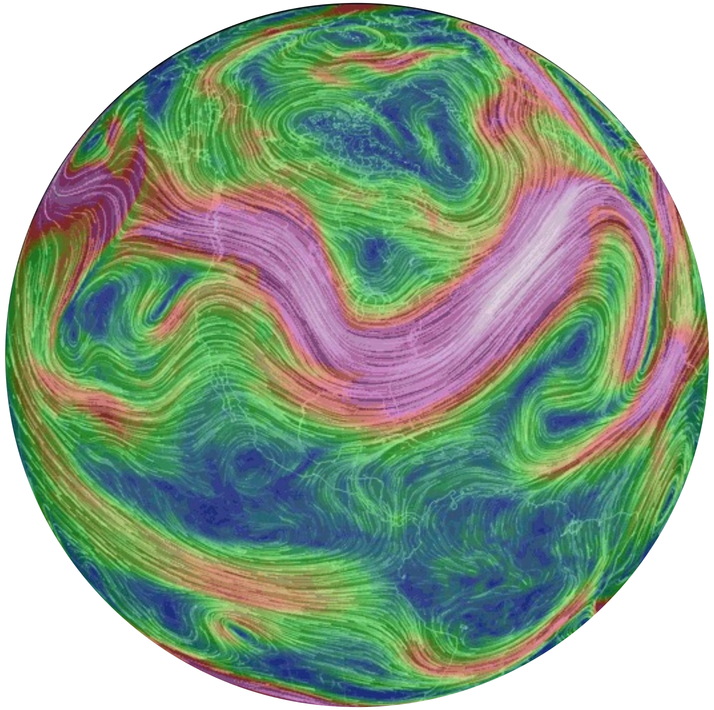
$$\mathbf{F}(\mathbf{PXR}, \mathbf{PAP}^\top) = \mathbf{PF}(\mathbf{X}, \mathbf{A})\mathbf{R}$$

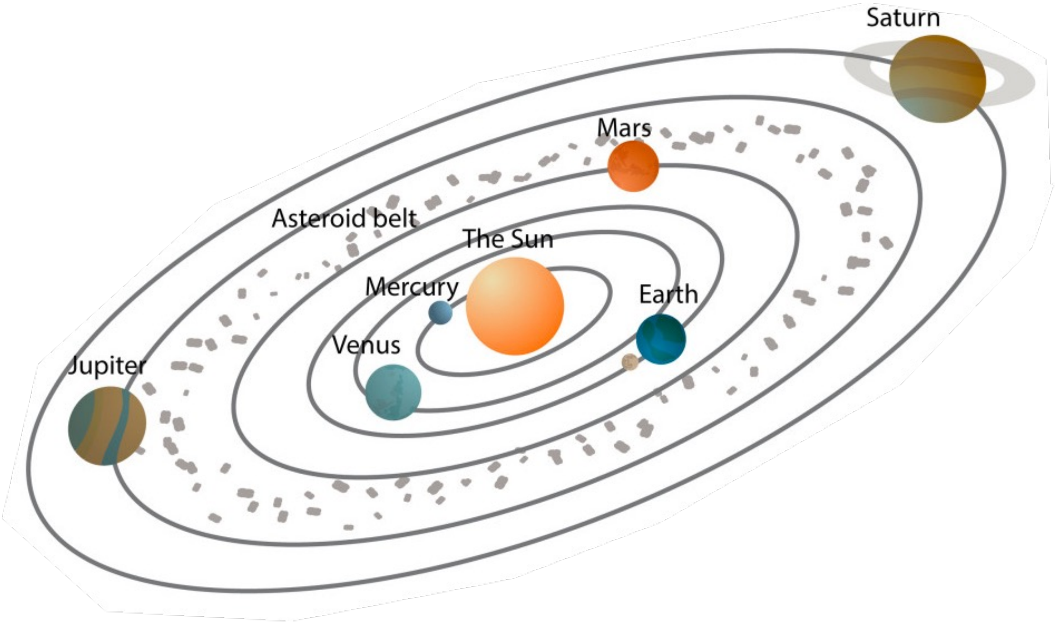
Geometric Deep Learning Blueprint



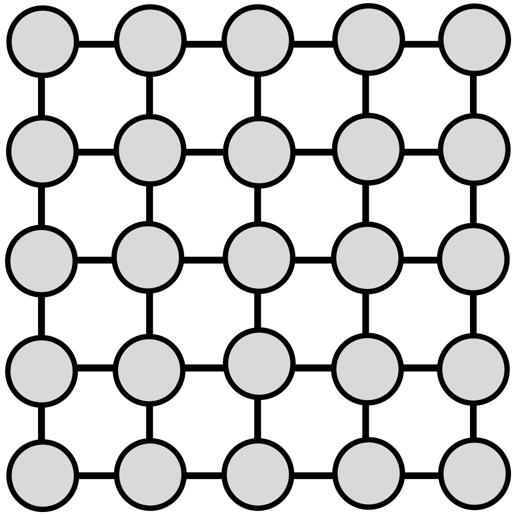
Scale Separation in Physics





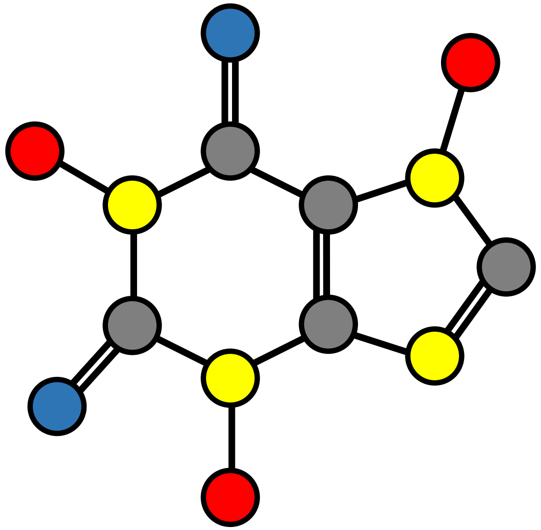


The "5G" of Geometric Deep Learning

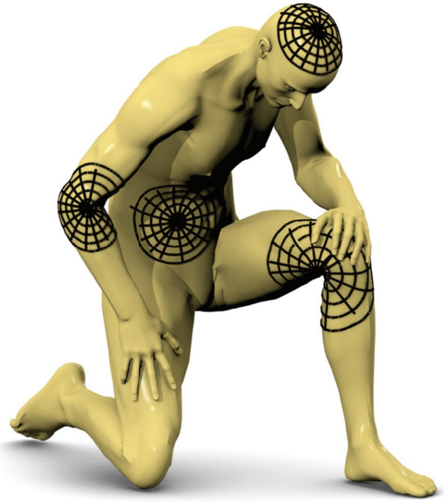


Images & Sequences

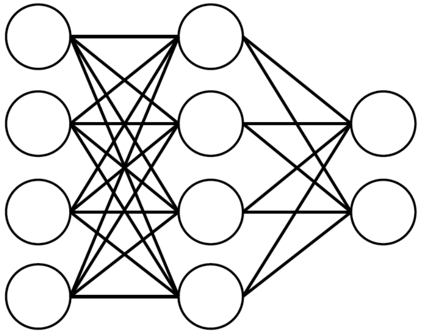
Homogeneous spaces



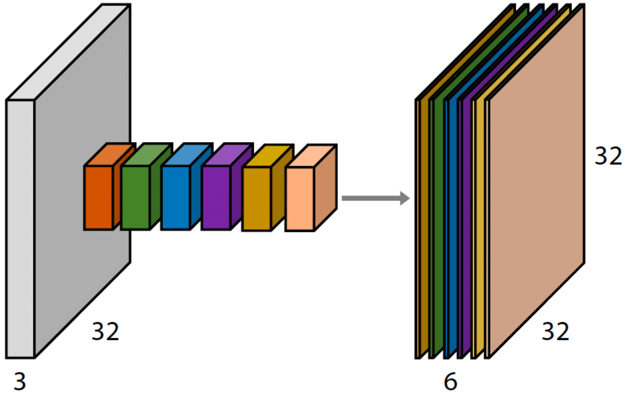
Graphs & Sets



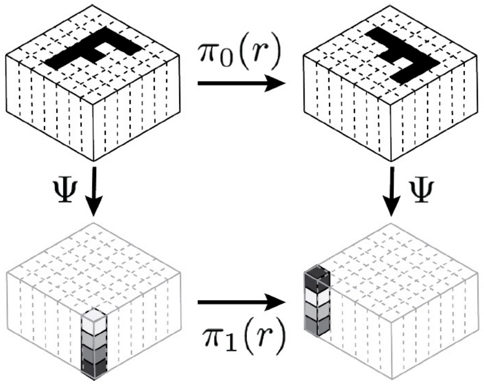
Manifolds, Meshes & Geometric graphs



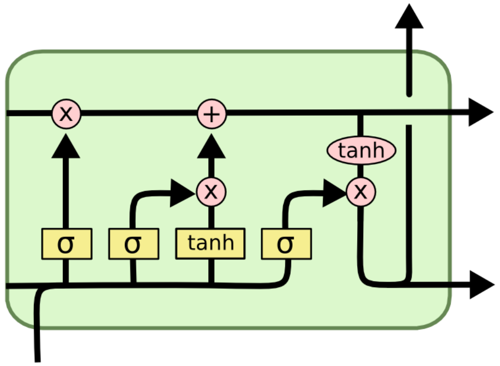
Perceptrons
Function regularity



CNNs
Translation

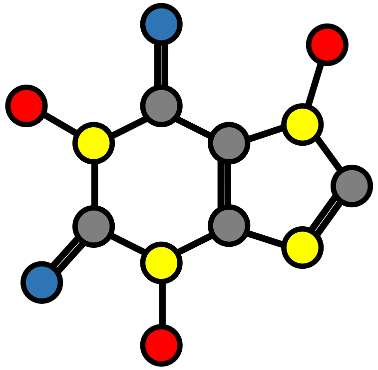


Group-CNNs
Translation+Rotation,
Global groups

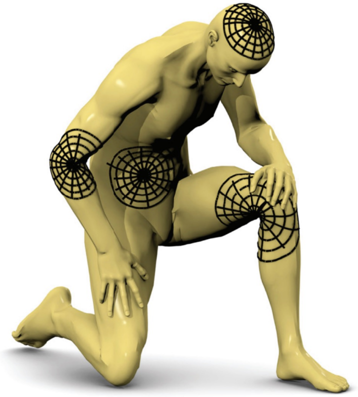


LSTMs
Time warping

DeepSets / Transformers
Permutation

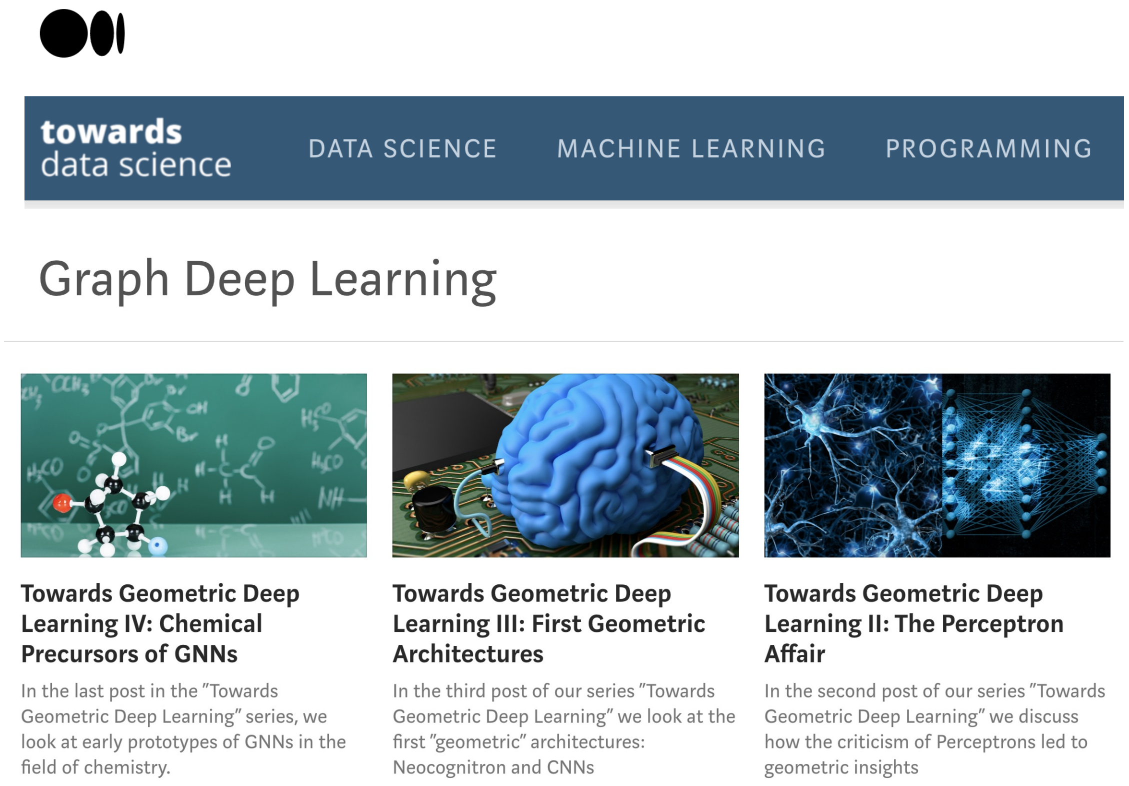


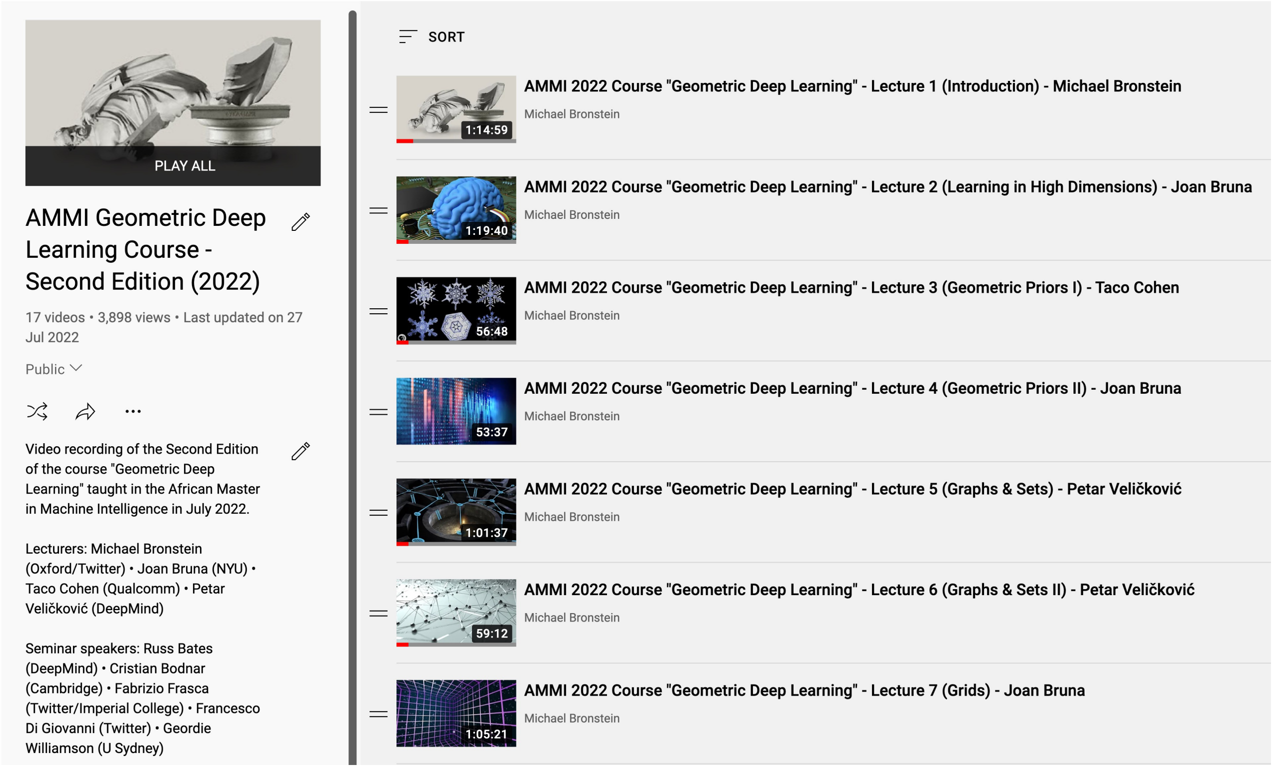
GNNs
Permutation



Intrinsic CNNs
Isometry / Gauge choice

geometricdeeplearning.com



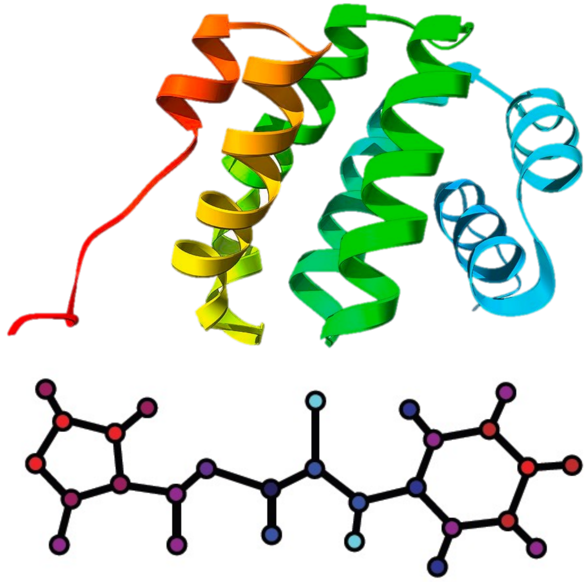


Medium Blog

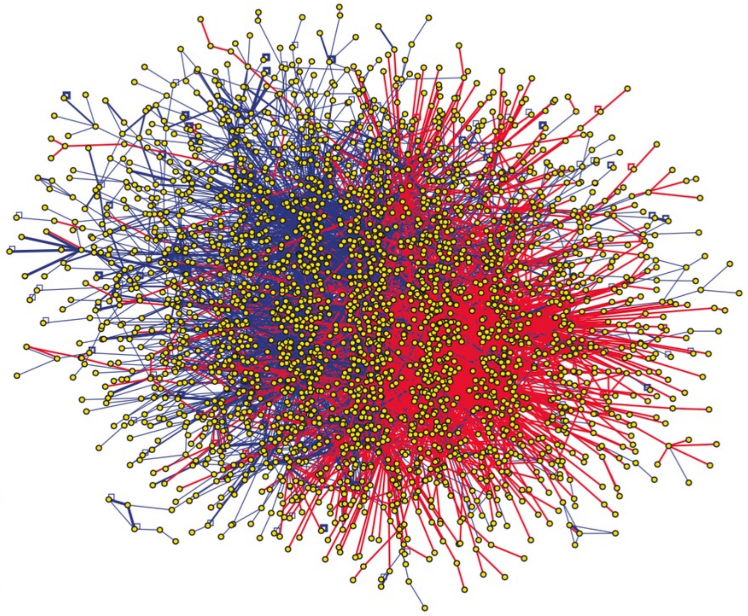
YouTube Channel

GRAPHS

Graphs = Systems of Relations and Interactions



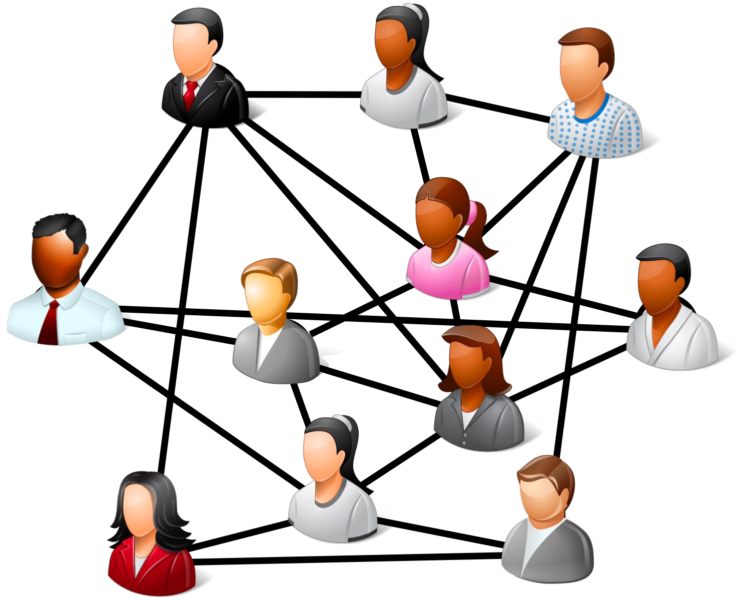
Molecules



Interactomes

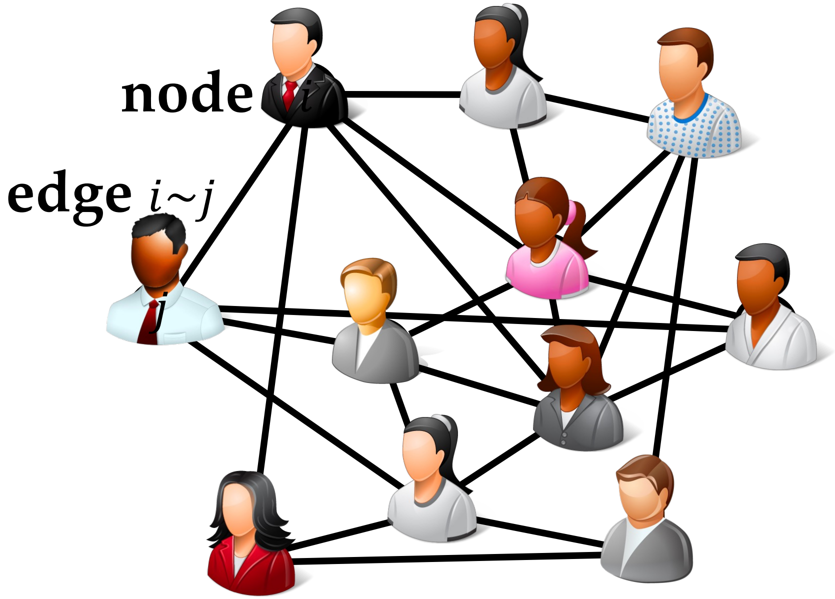
Social networks

Graphs



social network

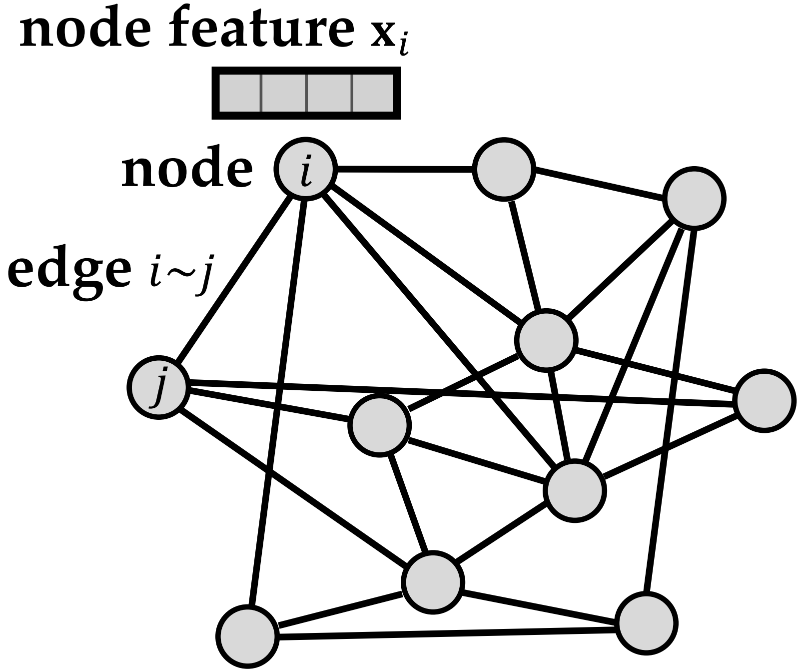
Graphs



social network

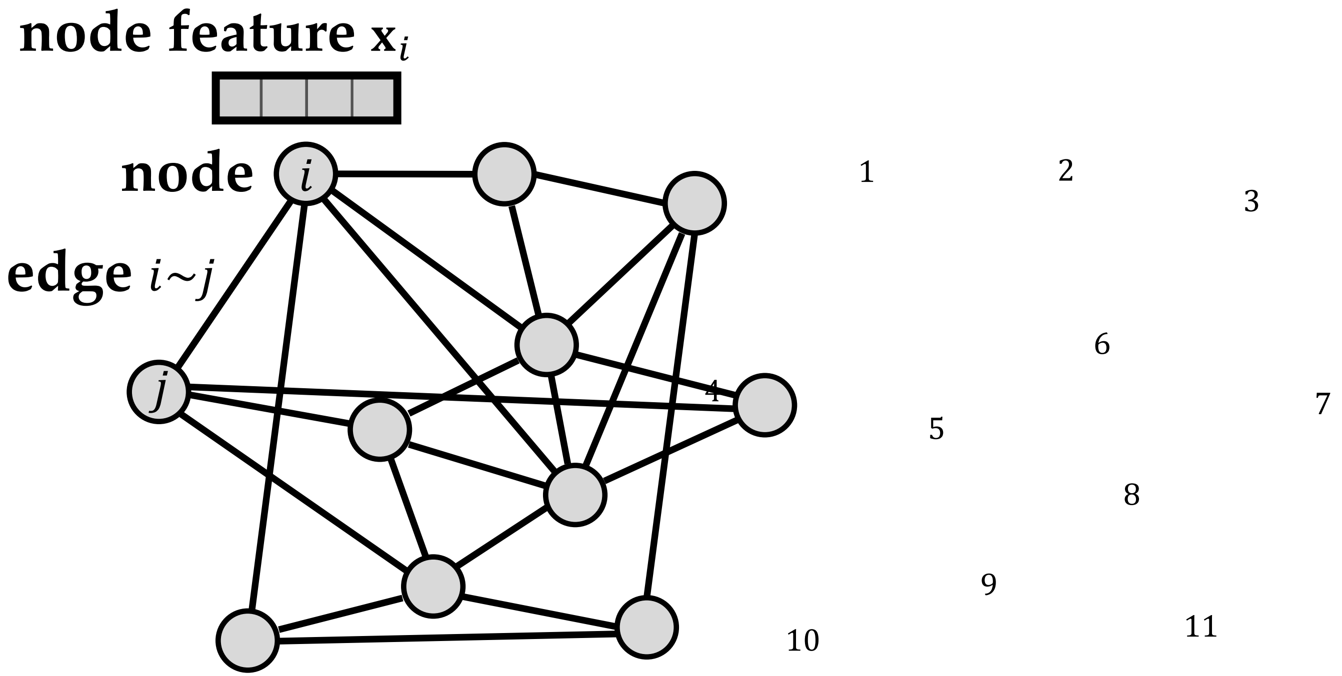
graph

Graphs



graph

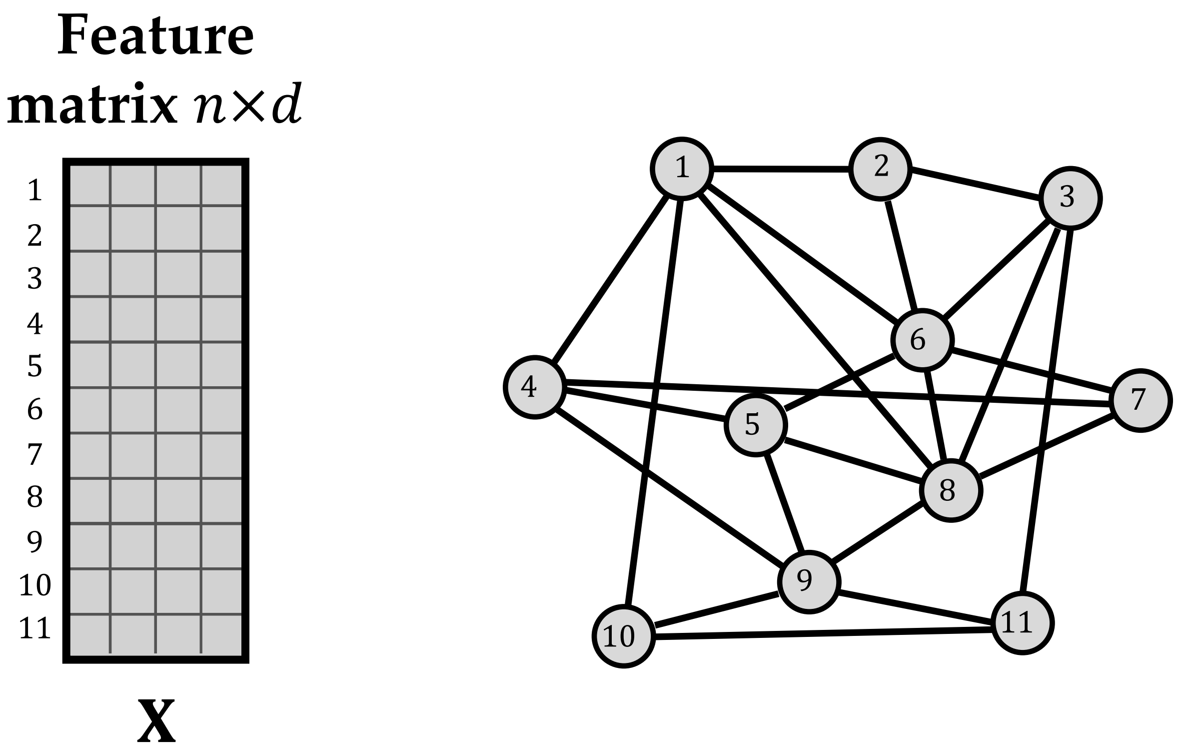
Key Structural Properties of Graphs



graph

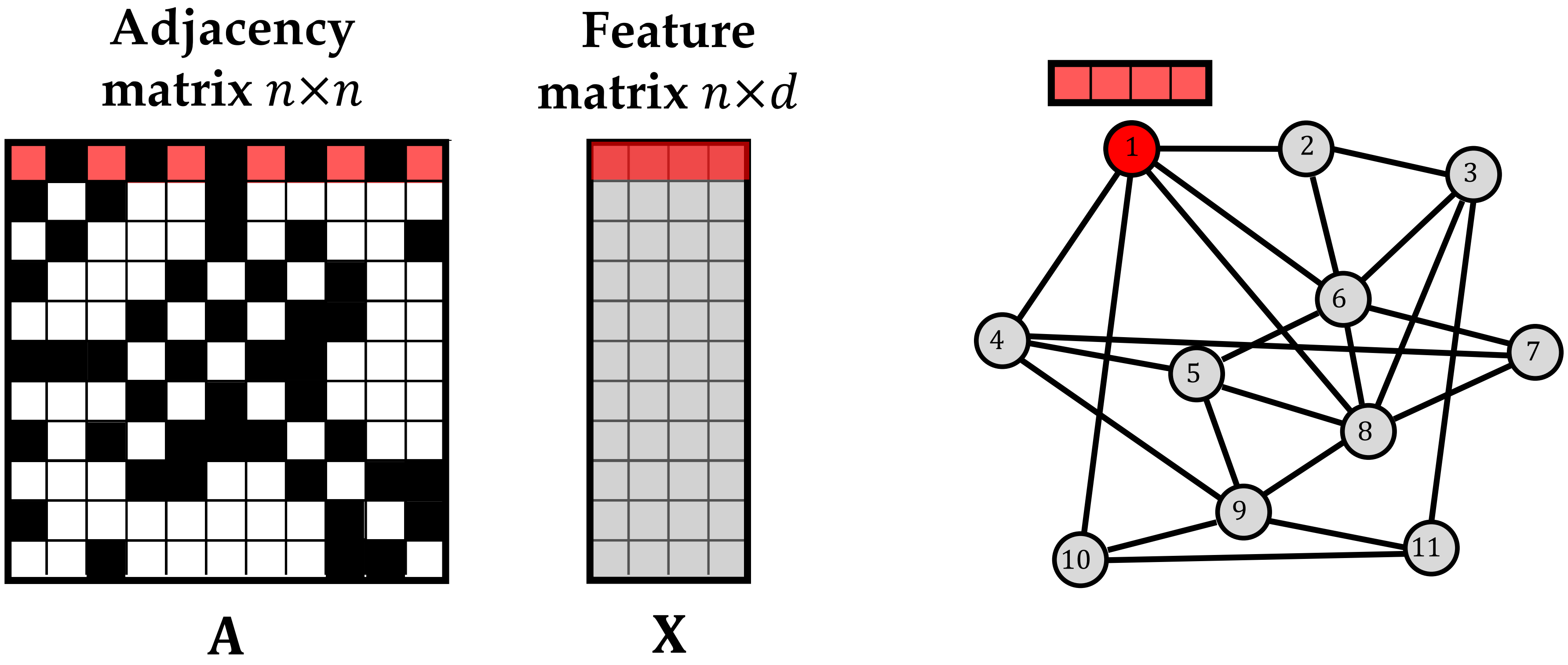
arbitrary ordering of nodes

Key Structural Properties of Graphs



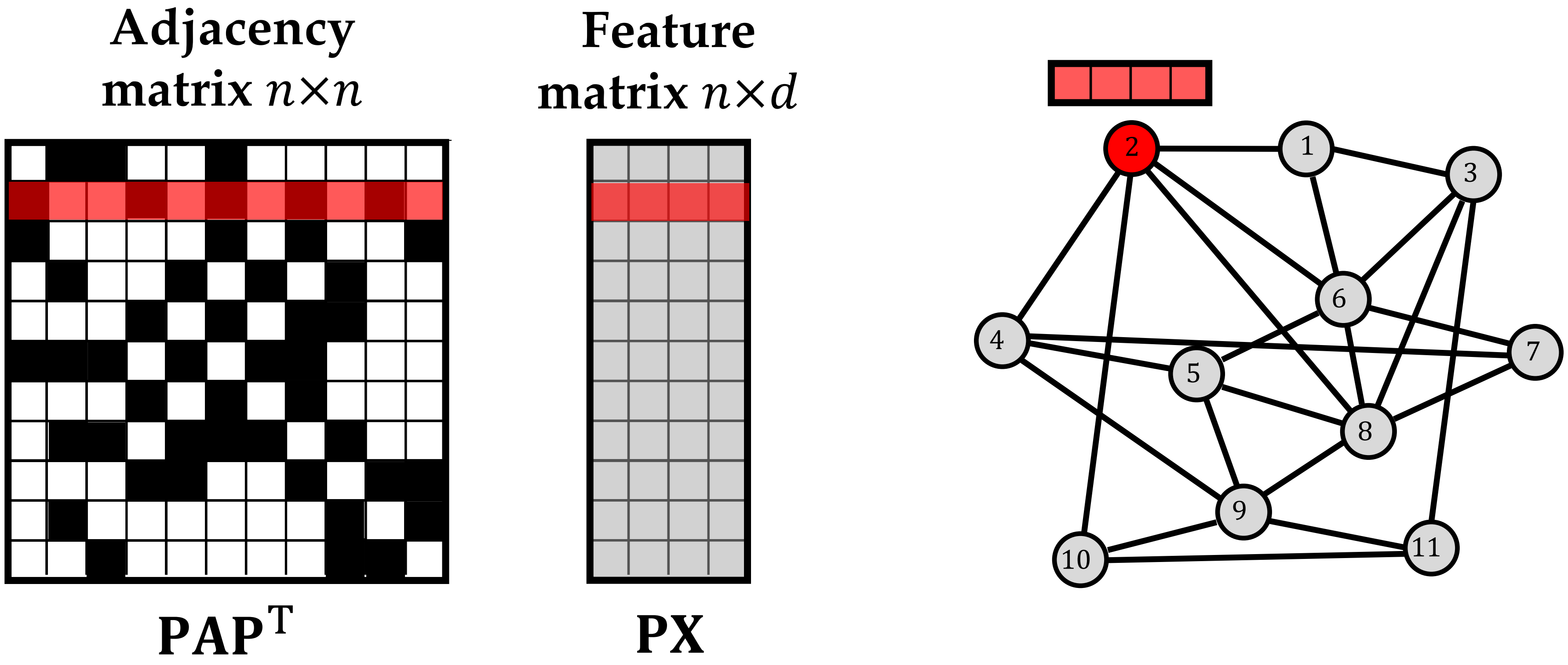
arbitrary ordering of nodes

Key Structural Properties of Graphs



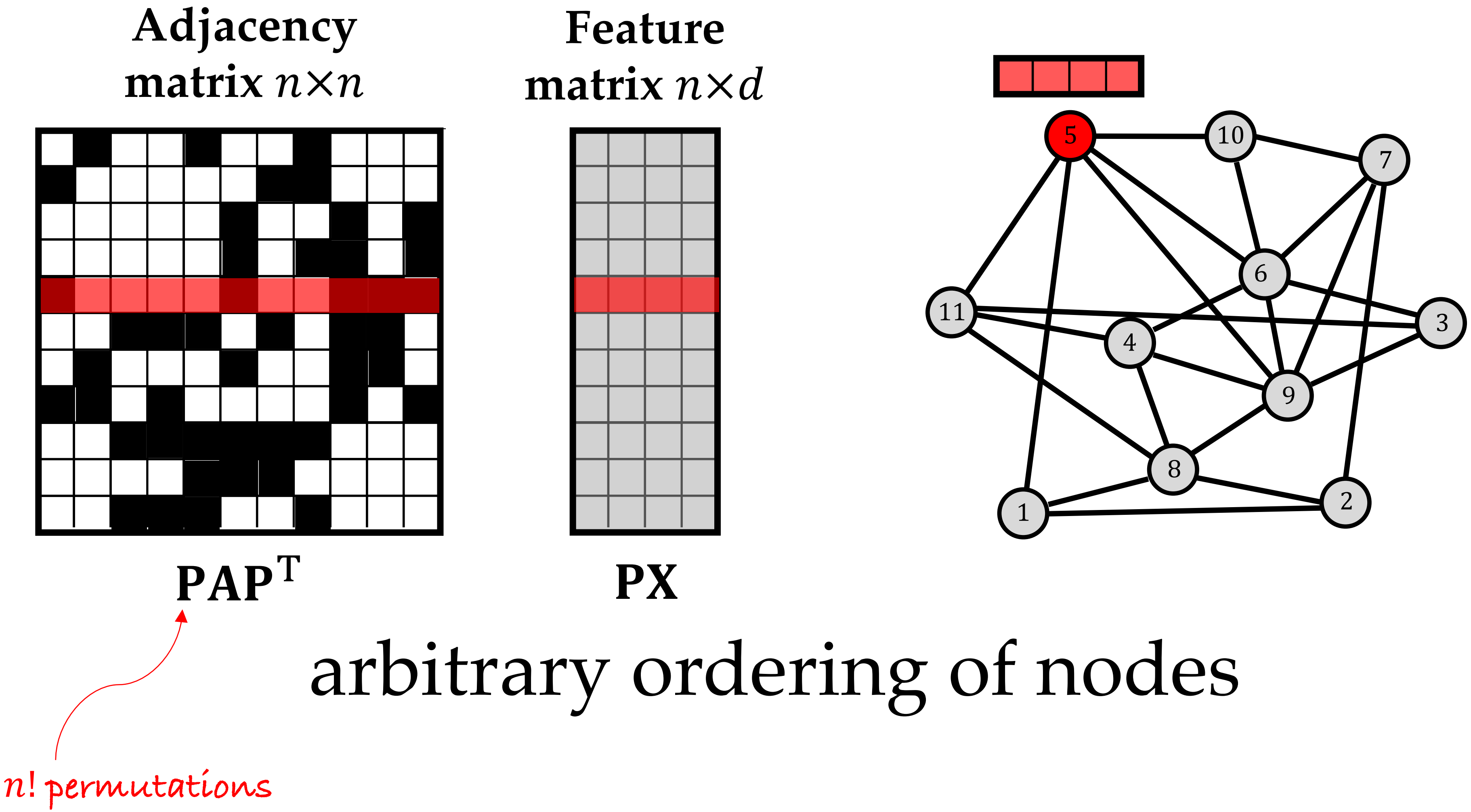
arbitrary ordering of nodes

Key Structural Properties of Graphs



arbitrary ordering of nodes

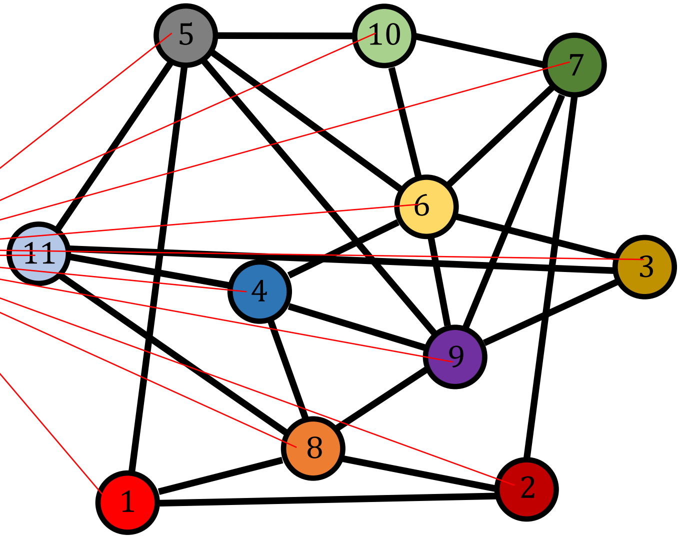
Key Structural Properties of Graphs



arbitrary ordering of nodes

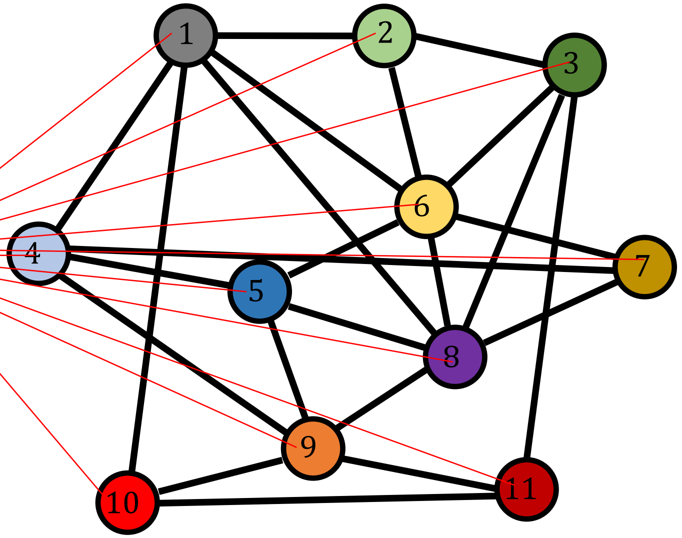
Invariant Graph Functions

graph function $f(\mathbf{X}, \mathbf{A})$



Invariant Graph Functions

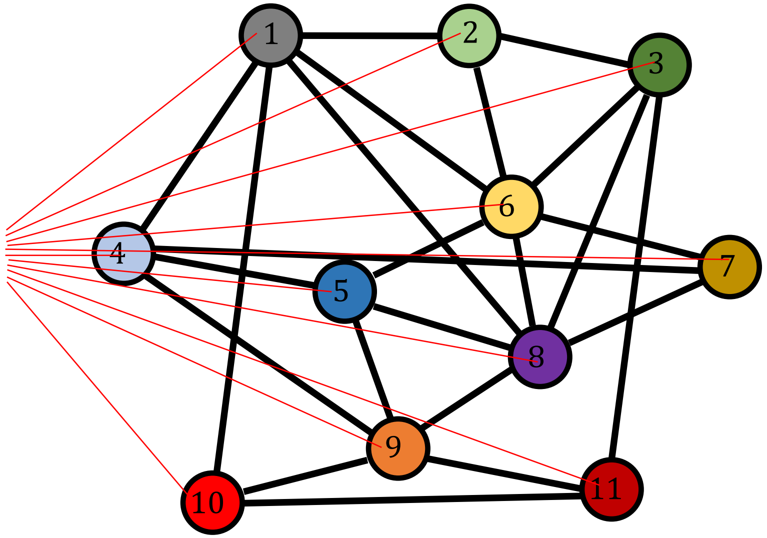
graph function $f(\mathbf{X}, \mathbf{A})$



Invariant Graph Functions

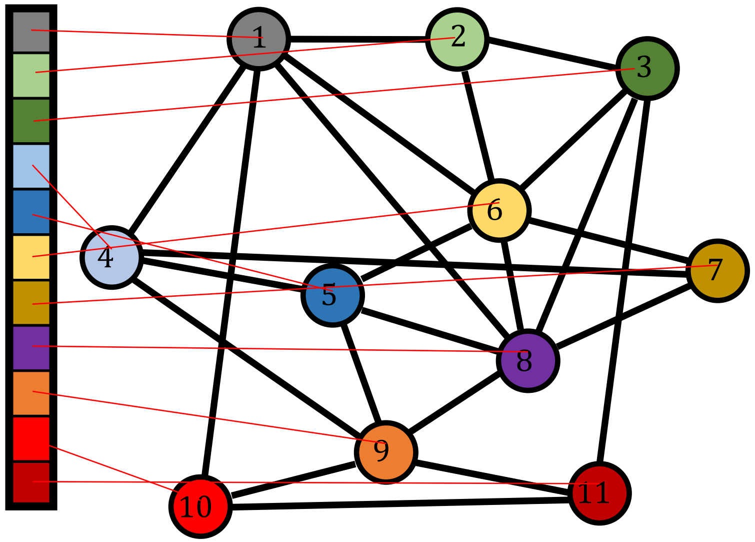
permutation-invariant

$$f\left(\mathbf{PX}, \mathbf{PAP}^\top\right) = f(\mathbf{X}, \mathbf{A})$$



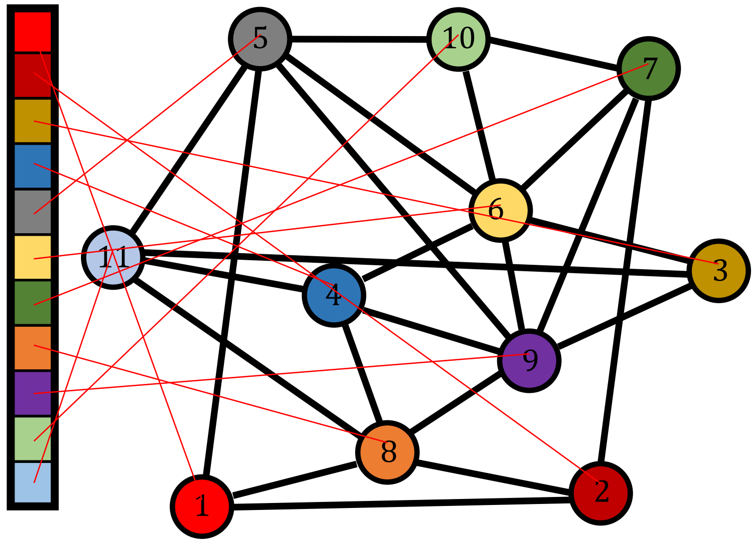
Equivariant Graph Functions

node function $\mathbf{F}(\mathbf{X}, \mathbf{A})$



Equivariant Graph Functions

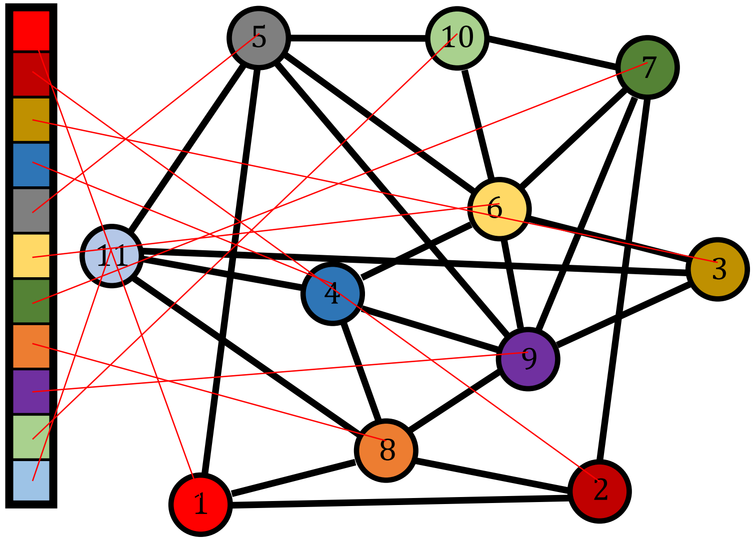
node function $\mathbf{F}(\mathbf{X}, \mathbf{A})$



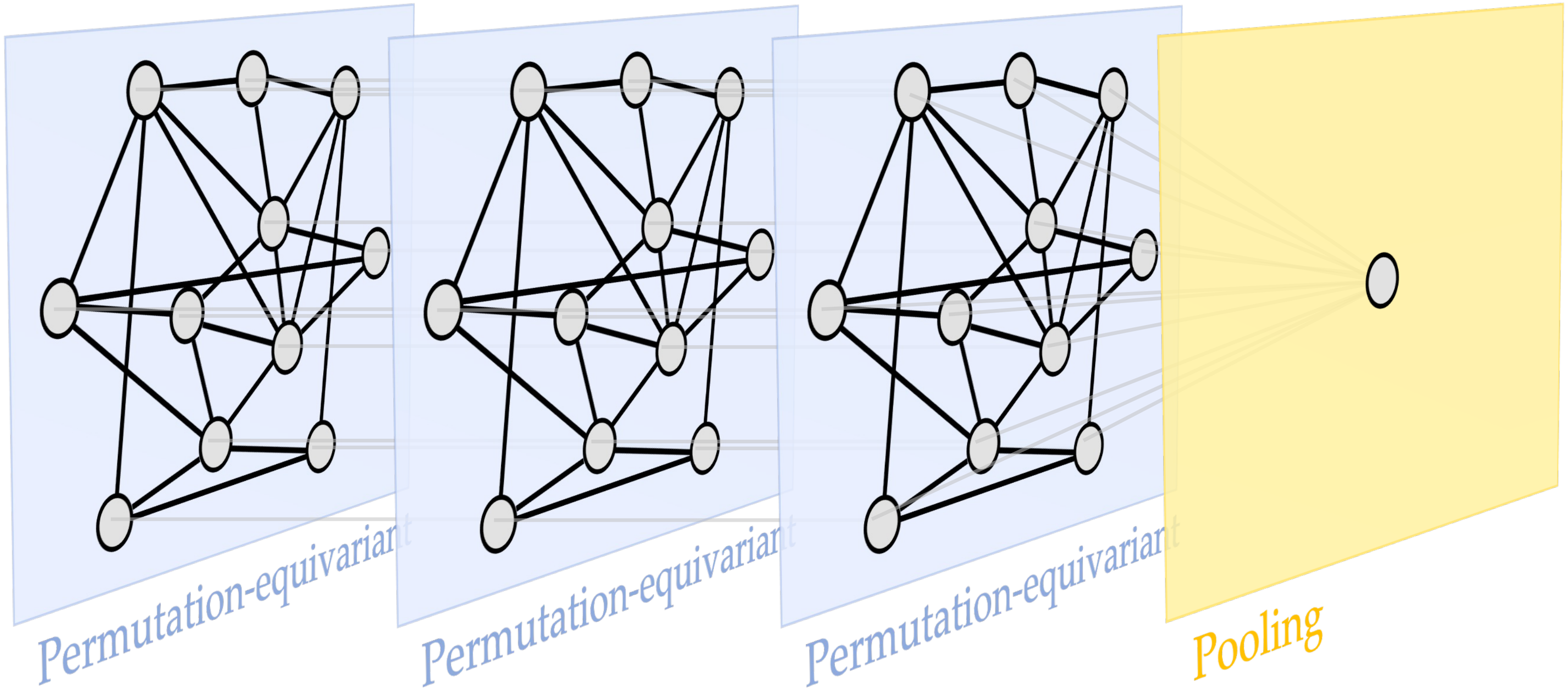
Equivariant Graph Functions

permutation-equivariant

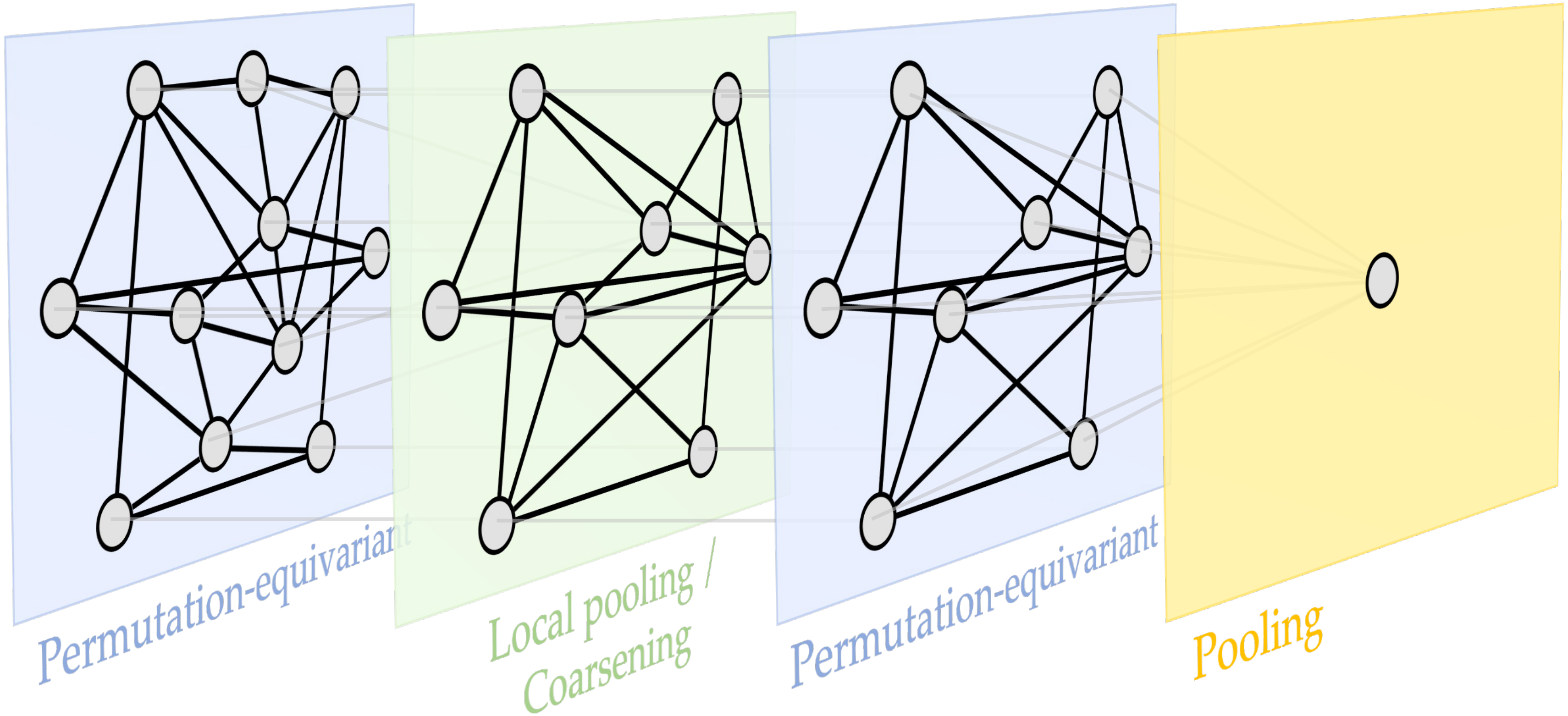
$$\mathbf{F}\left(\mathbf{P}\mathbf{X}, \mathbf{P}\mathbf{A}\mathbf{P}^\top\right) = \mathbf{P}\mathbf{F}(\mathbf{X}, \mathbf{A})$$



Graph Neural Networks



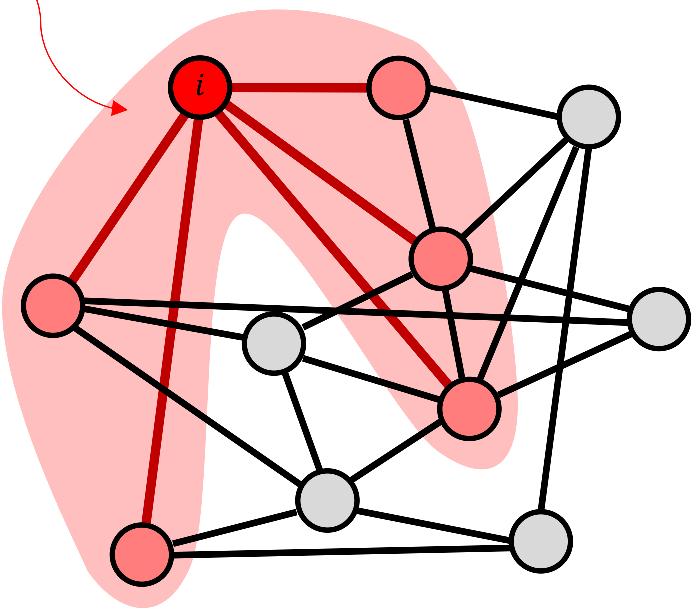
Graph Neural Networks



A General Blueprint for Constructing Graph Functions

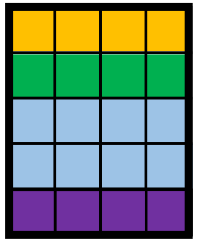
neighbourhood

$\mathcal{N}_i = \{j : i \sim j \text{ or } j \sim i\}$



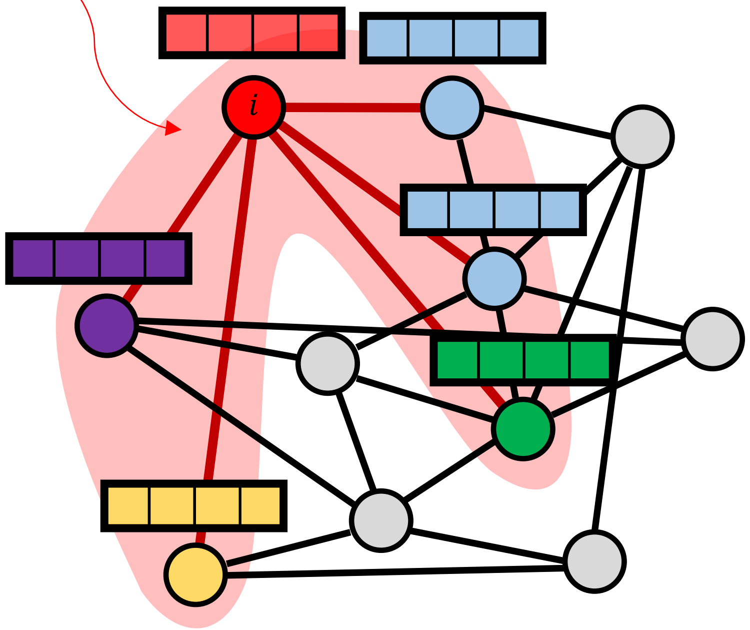
A General Blueprint for Constructing Graph Functions

multiset of
neighbour features



$$\mathbf{X}_{\mathcal{N}_i} = \{\!\{\mathbf{x}_{j\in\mathcal{N}_i}\}\!\}$$

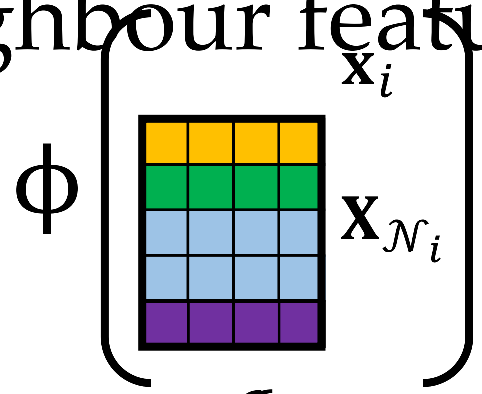
neighbourhood
$\mathcal{N}_i = \{j : i{\sim}j \text{ or } j{\sim}i\}$



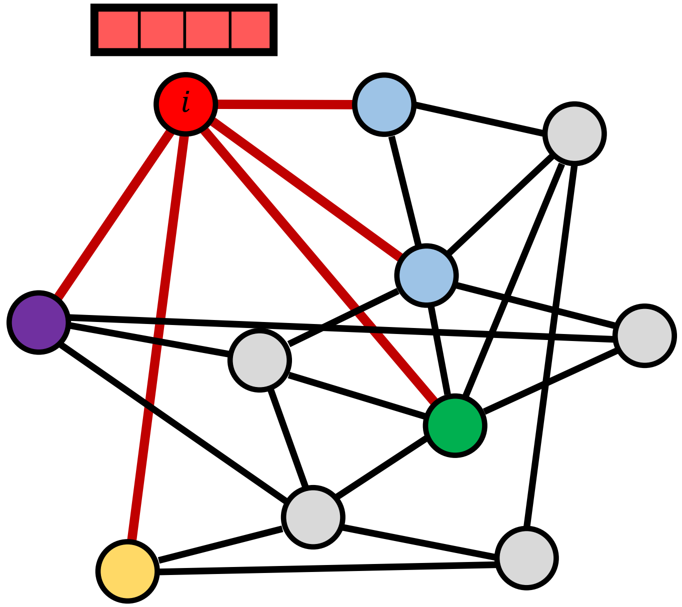
A General Blueprint for Constructing Graph Functions

local function

multiset of
neighbour features

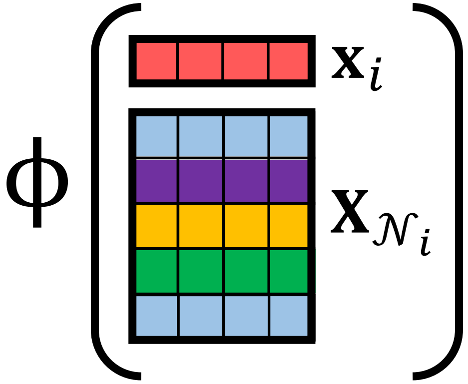


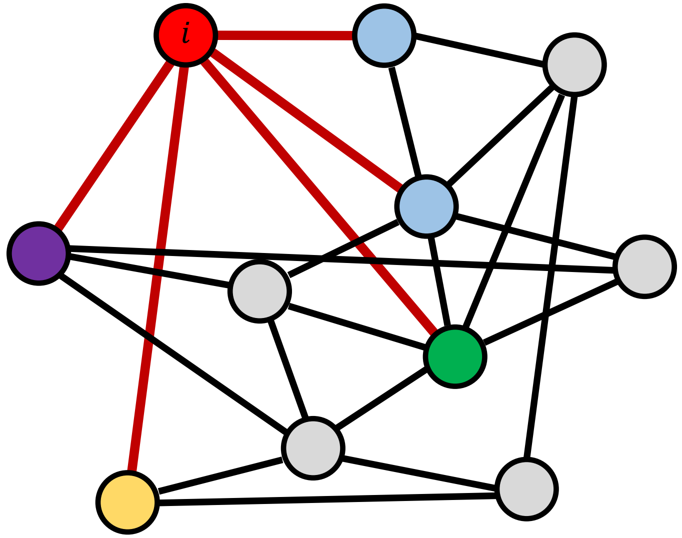
$$\mathbf{X}_{\mathcal{N}_i} = \{\!\{ \mathbf{x}_{j \in \mathcal{N}_i} \}\!\}$$



A General Blueprint for Constructing Graph Functions

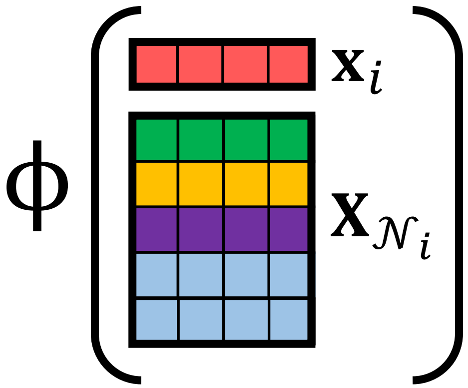
local function



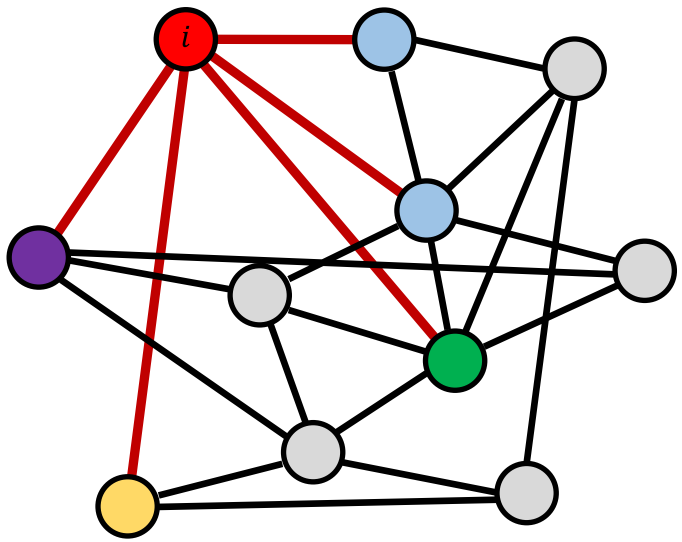


A General Blueprint for Constructing Graph Functions

local function



permutation invariant

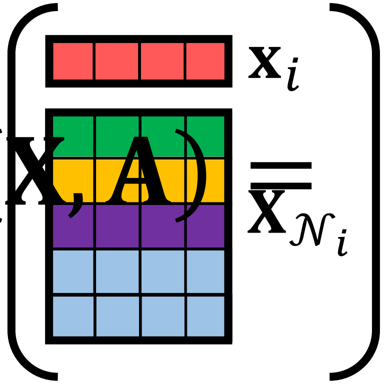


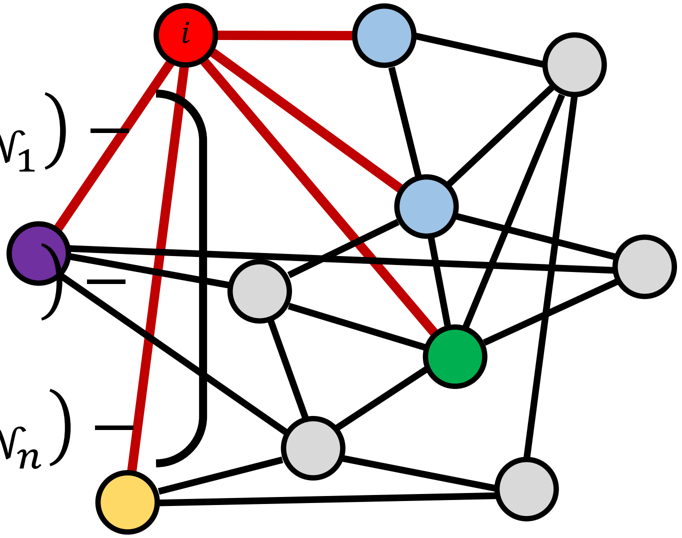
A General Blueprint for Constructing Graph Functions

local function

$$\mathbf{F}(\mathbf{X},\mathbf{A}) = \begin{bmatrix} - \phi(\mathbf{x}_1, \mathbf{X}_{\mathcal{N}_1}) - \\ \vdots \\ - \phi(\quad , \quad) - \\ \vdots \\ - \phi(\mathbf{x}_n, \mathbf{X}_{\mathcal{N}_n}) - \end{bmatrix}$$

$$\phi\left(\quad\right)$$





permutation invariant

A General Blueprint for Constructing Graph Functions

$$\mathbf{F}(\mathbf{X}, \mathbf{A}) = \begin{bmatrix} - \phi(\mathbf{x}_1, \mathbf{X}_{\mathcal{N}_1}) - \\ \vdots \\ - \phi(\mathbf{x}_i, \mathbf{X}_{\mathcal{N}_i}) - \\ \vdots \\ - \phi(\mathbf{x}_n, \mathbf{X}_{\mathcal{N}_n}) - \end{bmatrix}$$

permutation equivariant

"Flavours" of Graph Neural Networks

permutation-invariant
aggregation operator, e.g. sum

$$f(\mathbf{x}_i) = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} \psi(\mathbf{x}_j)\right)$$

new feature of
node i

learnable
functions

"Flavours" of Graph Neural Networks

importance of node j to
the representation of i

$$f(\mathbf{x}_i) = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} c_{ij}\psi(\mathbf{x}_j)\right)$$

"convolutional"

Defferard et al. 2016; Kipf, Welling 2016 (GCN)

"Flavours" of Graph Neural Networks

$$f(\mathbf{x}_i) = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} a(\mathbf{x}_i, \mathbf{x}_j)\psi(\mathbf{x}_j)\right)$$

"attentional"

Monti et al. 2017; Veličković et al. 2018 (GAT)

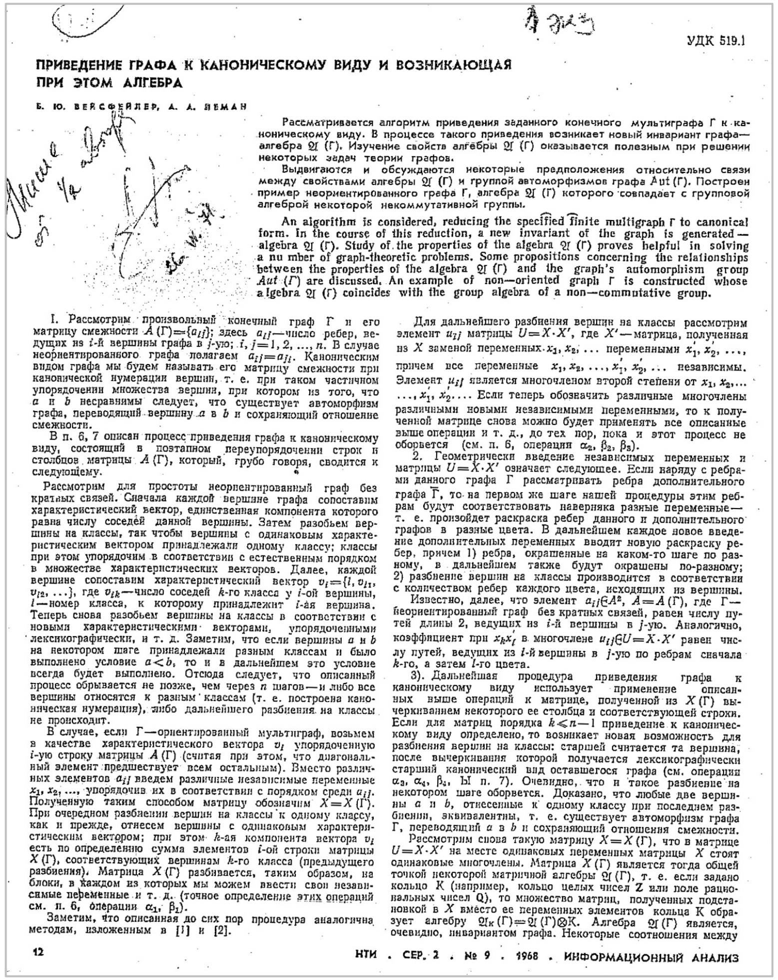
Message Passing

$$f(\mathbf{x}_i) = \phi\left(\mathbf{x}_i, \bigsqcup_{j \in \mathcal{N}_i} \psi(\mathbf{x}_i, \mathbf{x}_j)\right)$$

"message passing"

Gilmer et al. 2017 (MPNN); Battaglia et al 2018; Wang et al. 2018

Weisfeiler-Lehman Test



УДК 519.1

ПРИВЕДЕНИЕ ГРАФА К КАНОНИЧЕСКОМУ ВИДУ И ВОЗНИКАЮЩАЯ ПРИ ЭТОМ АЛГЕБРА

Б. Ю. ВЕЙСФЕЙЛЕР, А. А. ЛЕМАН

Рассматривается алгоритм приведения заданного конечного мультиграфа Г к каноническому виду. В процессе такого приведения возникает новый инвариант графа — алгебра $\mathfrak{A}$ (Г). Изучение свойств алгебры $\mathfrak{A}$ (Г) оказывается полезным при решении некоторых задач теории графов.

Выдвигаются и обсуждаются некоторые предположения относительно связи между свойствами алгебры $\mathfrak{A}$ (Г) и группой автоморфизмов графа Aut (Г). Построен пример неориентированного графа Г, алгебра $\mathfrak{A}$ (Г) которого совпадает с групповой алгеброй некоторой некоммутативной группы.

An algorithm is considered, reducing the specified finite multigraph Г to canonical form. In the course of this reduction, a new invariant of the graph is generated — algebra $\mathfrak{A}$ (Г). Study of the properties of the algebra $\mathfrak{A}$ (Г) proves helpful in solving a nu mber of graph-theoretic problems. Some propositions concerning the relationships between the properties of the algebra $\mathfrak{A}$ (Г) and the graph's automorphism group *Aut* (Г) are discussed. An example of non—oriented graph Г is constructed whose algebra $\mathfrak{A}$ (Г) coincides with the group algebra of a non—commutative group.

1. Рассмотрим произвольный конечный граф Г и его матрицу смежности $A(\Gamma)=\{a_{ij}\}$; здесь a_{ij} — число ребер, ведущих из i-й вершины графа в j-ую; $i, j=1, 2, \ldots, n$. В случае неориентированного графа полагаем $a_{ij}=a_{ji}$. Каноническим видом графа мы будем называть его матрицу смежности при канонической нумерации вершин, т. е. при таком частичном упорядочении множества вершин, при котором из того, что a и b несравнимы следует, что существует автоморфизм графа, переводящий вершину a в b и сохраняющий отношение смежности.

В п. 6, 7 описан процесс приведения графа к каноническому виду, состоящий в поэтапном переупорядочении строк и столбцов матрицы $A(\Gamma)$, который, грубо говоря, сводится к следующему.

Рассмотрим для простоты неориентированный граф без кратных связей. Сначала каждой вершине графа сопоставим характеристический вектор, единственная компонента которого равна числу соседей данной вершины. Затем разобьем вершины на классы, так чтобы вершины с одинаковым характеристическим вектором принадлежали одному классу; классы при этом упорядочим в соответствии с естественным порядком в множестве характеристических векторов. Далее, каждой вершине сопоставим характеристический вектор $v_i=\{l, v_{i1}, v_{i2}, \ldots\}$, где v_{ik} — число соседей k-го класса у i-ой вершины, l — номер класса, к которому принадлежит i-ая вершина. Теперь снова разобьем вершины на классы в соответствии с новыми характеристическими векторами, упорядоченными лексикографически, и т. д. Заметим, что если вершины a и b на некотором шаге принадлежали разным классам и было выполнено условие $a<b$, то и в дальнейшем это условие всегда будет выполнено. Отсюда следует, что описанный процесс обрывается не позже, чем через n шагов — и либо все вершины относятся к разным классам (т. е. построена каноническая нумерация), либо дальнейшего разбиения на классы не происходит.

В случае, если Г — ориентированный мультиграф, возьмем в качестве характеристического вектора v_i упорядоченную i-ую строку матрицы $A(\Gamma)$ (считая при этом, что диагональный элемент предшествует всем остальным). Вместо различных элементов a_{ij} введем различные независимые переменные $x_1, x_2, \ldots$, упорядочив их в соответствии с порядком среди a_{ij}. Полученную таким способом матрицу обозначим $X=X(\Gamma)$. При очередном разбиении вершин на классы к одному классу, как и прежде, отнесем вершины с одинаковым характеристическим вектором; при этом k-ая компонента вектора v_i есть по определению сумма элементов i-ой строки матрицы $X(\Gamma)$, соответствующих вершинам k-го класса (предыдущего разбиения). Матрица $X(\Gamma)$ разбивается, таким образом, на блоки, в каждом из которых мы можем ввести свои независимые переменные и т. д. (точное определение этих операций см. п. 6, операции α_1, β_1).

Заметим, что описанная до сих пор процедура аналогична методам, изложенным в [1] и [2].

Для дальнейшего разбиения вершин на классы рассмотрим элемент u_{ij} матрицы $U=X\cdot X'$, где X' — матрица, полученная из X заменой переменных $x_1, x_2, \ldots$ переменными $x_1', x_2', \ldots$, причем все переменные $x_1, x_2, \ldots, x_1', x_2', \ldots$ независимы. Элемент u_{ij} является многочленом второй степени от $x_1, x_2, \ldots, x_1', x_2', \ldots$ Если теперь обозначить различные многочлены различными новыми независимыми переменными, то к полученной матрице снова можно будет применять все описанные выше операции и т. д., до тех пор, пока и этот процесс не оборвется (см. п. 6, операции α_2, β_2, β_3).

2. Геометрически введение независимых переменных и матрицы $U=X\cdot X'$ означает следующее. Если наряду с ребрами данного графа Г рассматривать ребра дополнительного графа $\overline{\Gamma}$, то на первом же шаге нашей процедуры этим ребрам будут соответствовать наверняка разные переменные — т. е. произойдет раскраска ребер данного и дополнительного графов в разные цвета. В дальнейшем каждое новое введение дополнительных переменных вводит новую раскраску ребер, причем 1) ребра, окрашенные на каком-то шаге по разному, в дальнейшем также будут окрашены по-разному; 2) разбиение вершин на классы производится в соответствии с количеством ребер каждого цвета, исходящих из вершины.

Известно, далее, что элемент $a_{ij}\in A^2$, $A=A(\Gamma)$, где Г — неориентированный граф без кратных связей, равен числу путей длины 2, ведущих из i-й вершины в j-ую. Аналогично, коэффициент при $x_k x_l'$ в многочлене $u_{ij}\in U=X\cdot X'$ равен числу путей, ведущих из i-й вершины в j-ую по ребрам сначала k-го, а затем l-го цвета.

3). Дальнейшая процедура приведения графа к каноническому виду использует применение описанных выше операций к матрице, полученной из $X(\Gamma)$ вычеркиванием некоторого ее столбца и соответствующей строки. Если для матриц порядка $k\leqslant n-1$ приведение к каноническому виду определено, то возникает новая возможность для разбиения вершин на классы: старшей считается та вершина, после вычеркивания которой получается лексикографически старший канонический вид оставшегося графа (см. операции α_3, α_4, β_4, Ы п. 7). Очевидно, что и такое разбиение на некотором шаге оборвется. Доказано, что любые две вершины a и b, отнесенные к одному классу при последнем разбиении, эквивалентны, т. е. существует автоморфизм графа Г, переводящий a в b и сохраняющий отношения смежности.

Рассмотрим снова такую матрицу $X=X(\Gamma)$, что в матрице $U=X\cdot X'$ на месте одинаковых переменных матрицы X стоят одинаковые многочлены. Матрица $X(\Gamma)$ является тогда общей точкой некоторой матричной алгебры $\mathfrak{A}(\Gamma)$, т. е. если задано кольцо К (например, кольцо целых чисел $\mathbb{Z}$ или поле рациональных чисел Q), то множество матриц, полученных подстановкой в X вместо ее переменных элементов кольца К образует алгебру $\mathfrak{A}_K(\Gamma)=\mathfrak{A}(\Gamma)\otimes K$. Алгебра $\mathfrak{A}(\Gamma)$ является, очевидно, инвариантом графа. Некоторые соотношения между

12 НТИ . СЕР. 2 . № 9 . 1968 . ИНФОРМАЦИОННЫЙ АНАЛИЗ

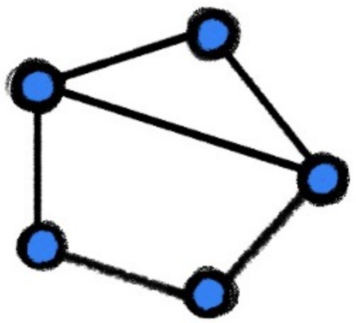
A. Lehman

B. Weisfeiler

1968

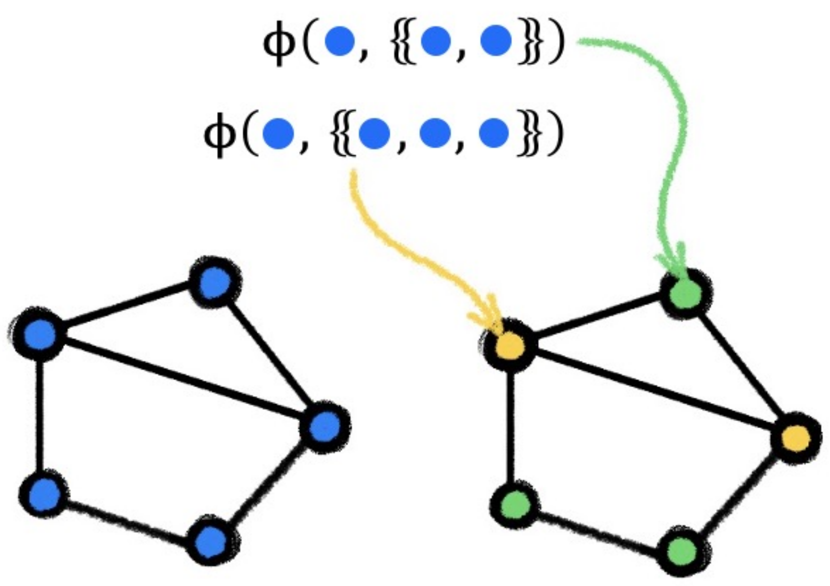
Weisfeiler, Lehman 1968; Portraits: Ihor Gorskiy

Weisfeiler-Lehman Test



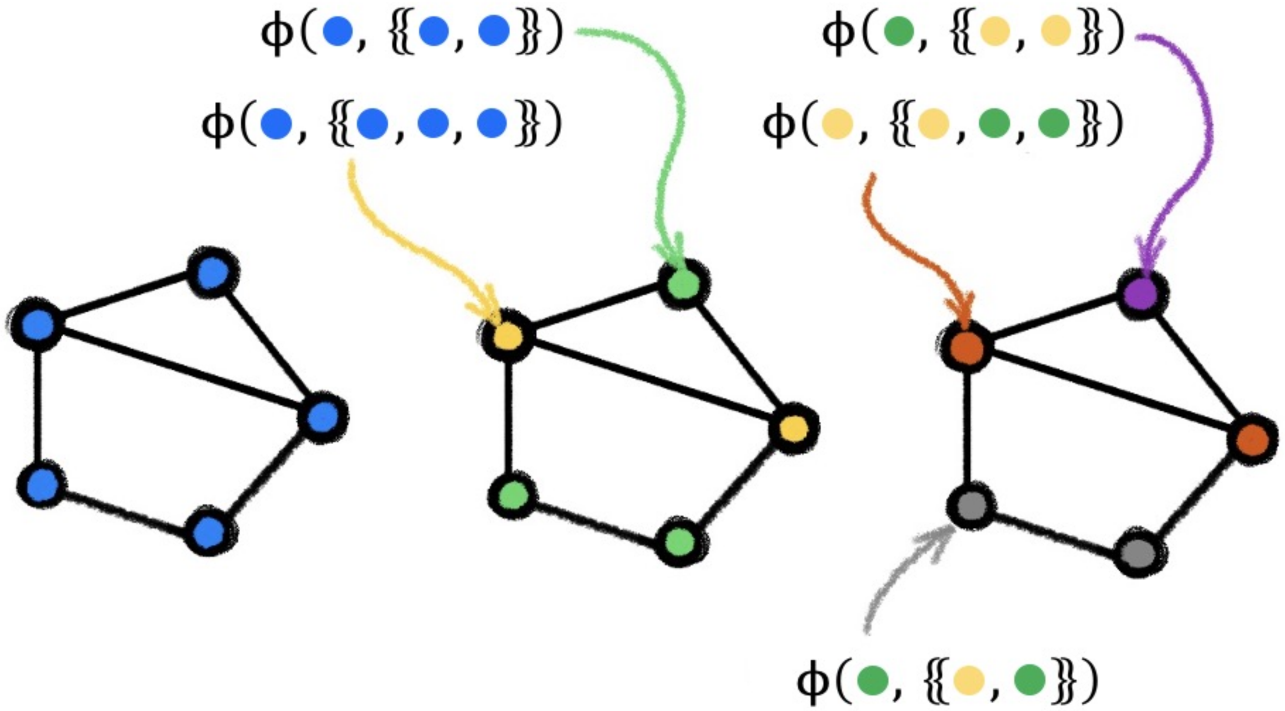
Weisfeiler, Lehman 1968

Weisfeiler-Lehman Test



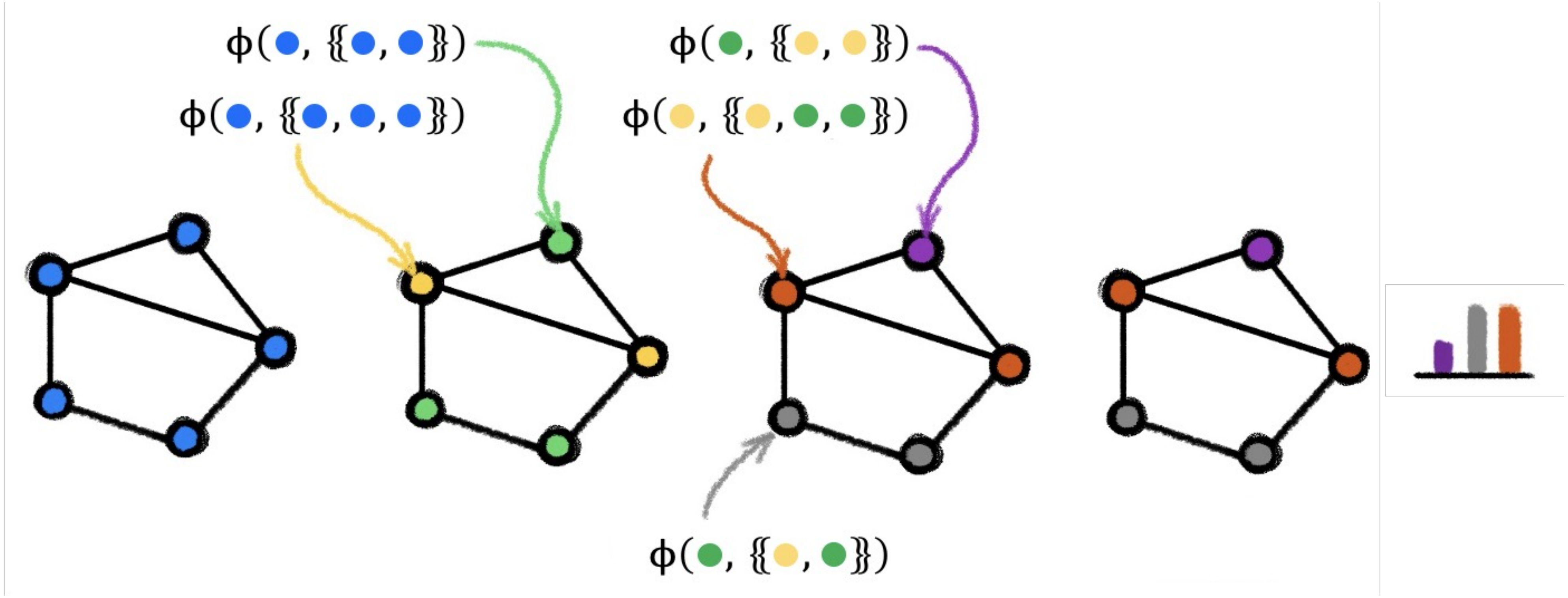
Weisfeiler, Lehman 1968

Weisfeiler-Lehman Test



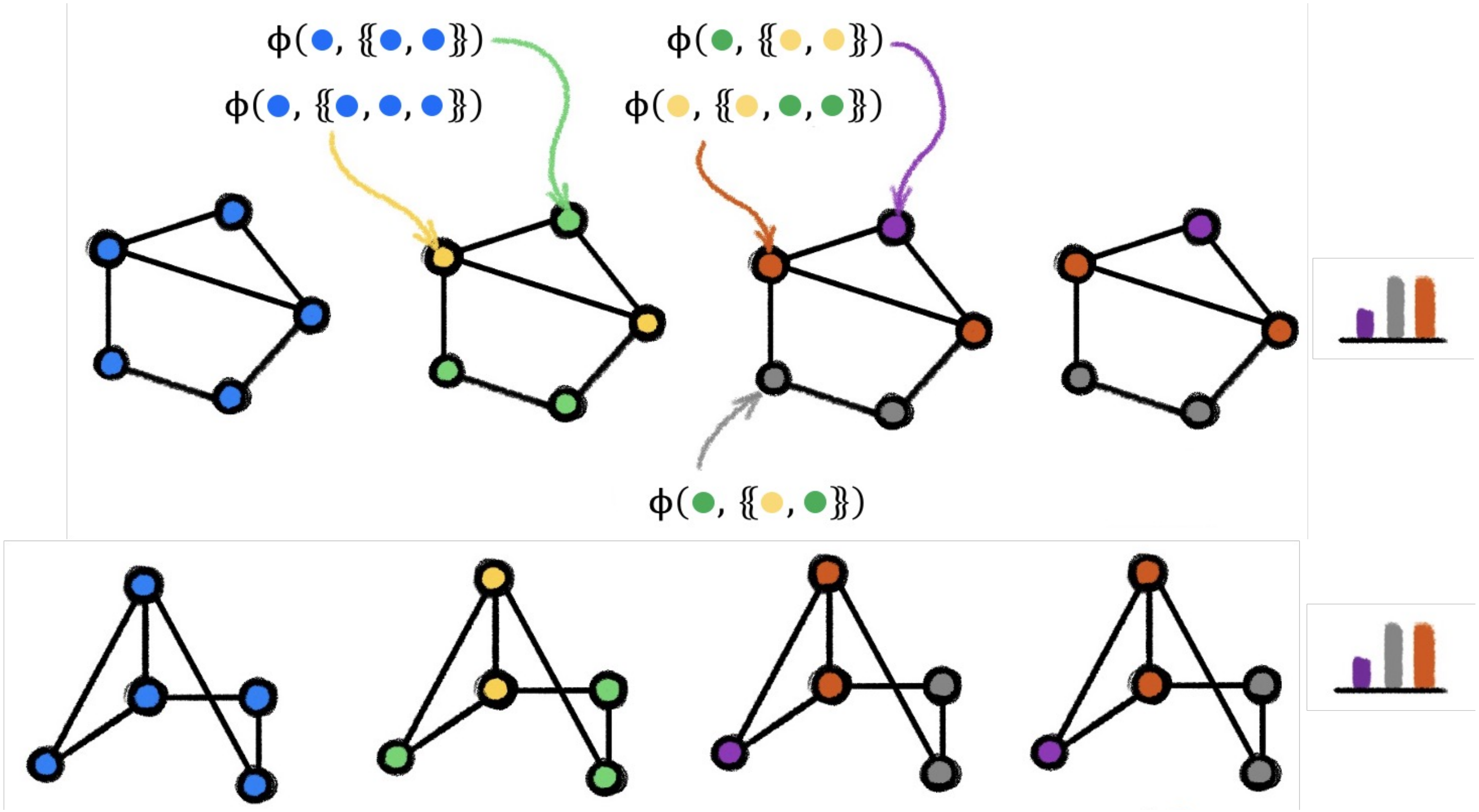
Weisfeiler, Lehman 1968

Weisfeiler-Lehman Test

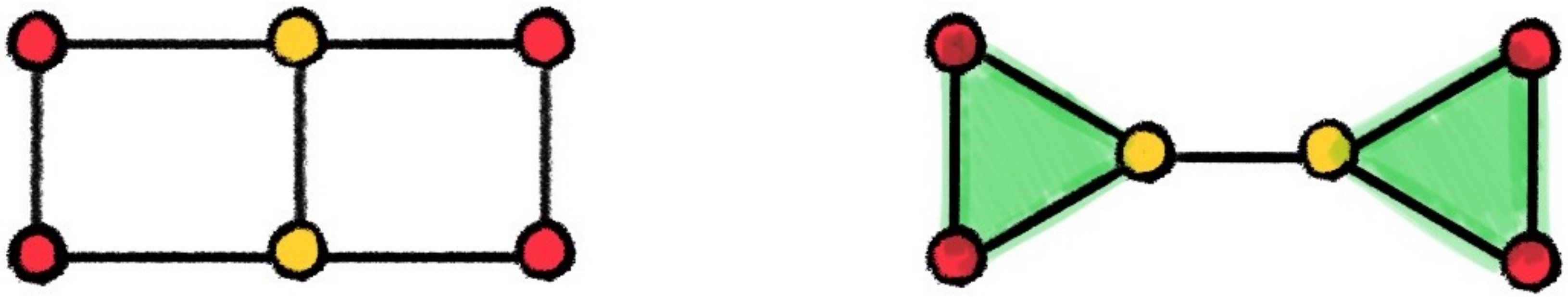


Weisfeiler, Lehman 1968

Weisfeiler-Lehman Test

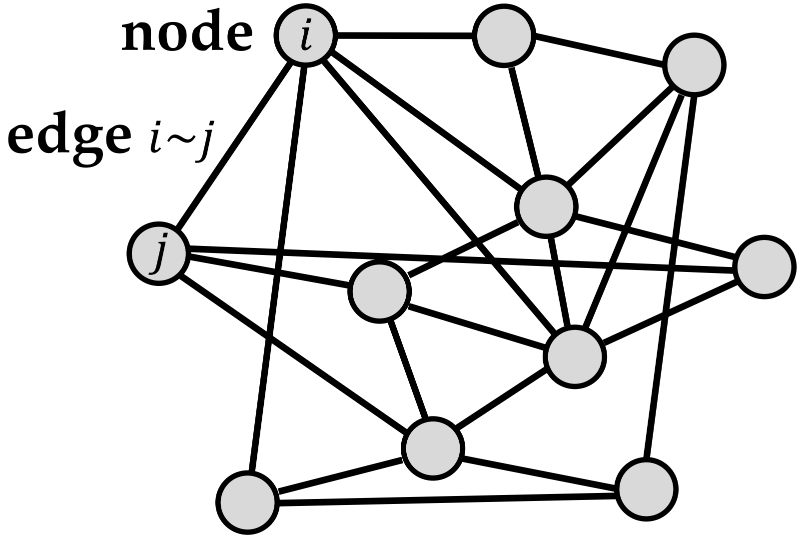


necessary but insufficient condition!



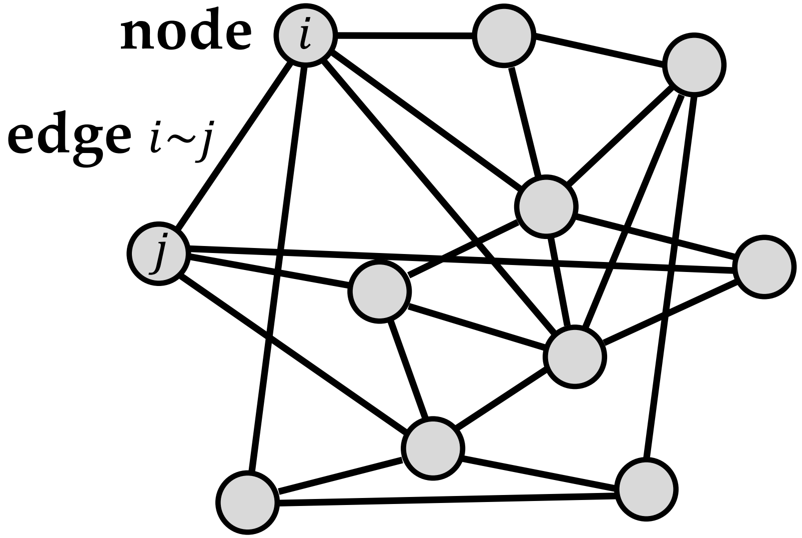
non-isomorphic graphs that are WL-equivalent

Special Cases of GNNs



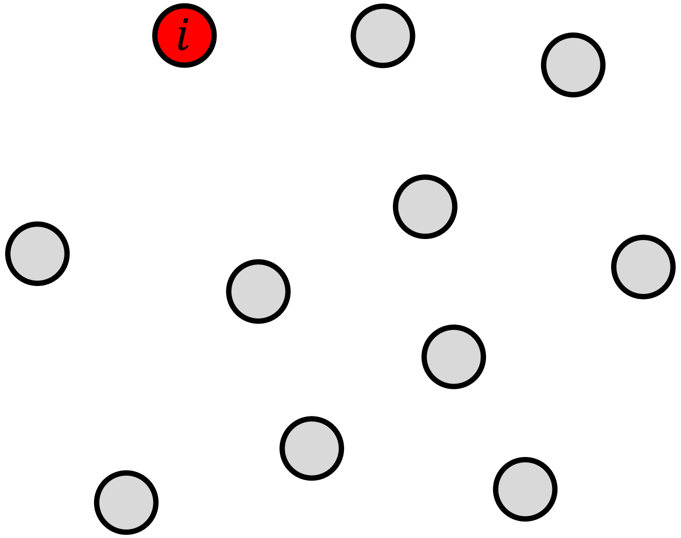
graph

Special Cases of GNNs



graph set

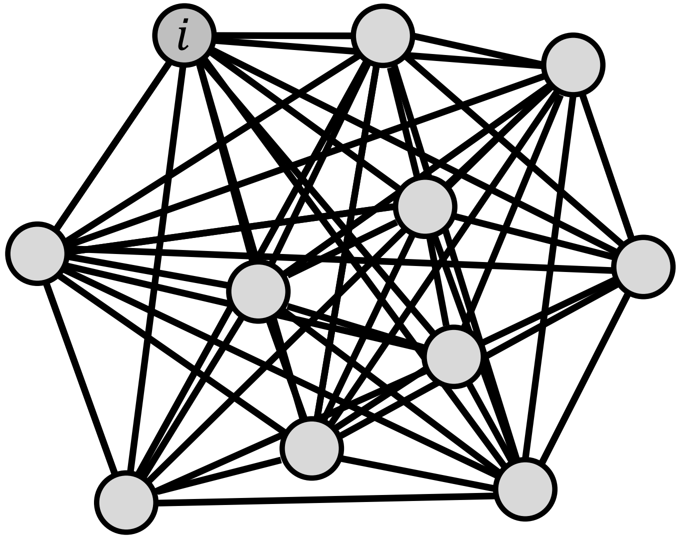
DeepSets



$$\phi(\mathbf{x}_i)$$

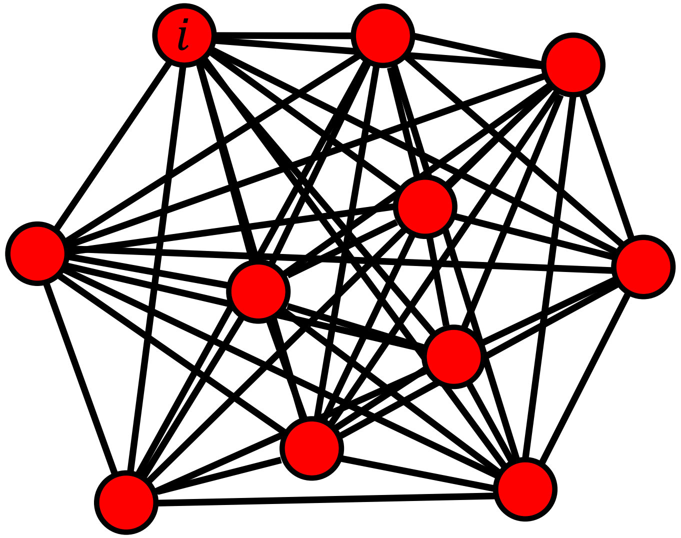
Zaheer et al 2017; Qi et al 2017

Transformers



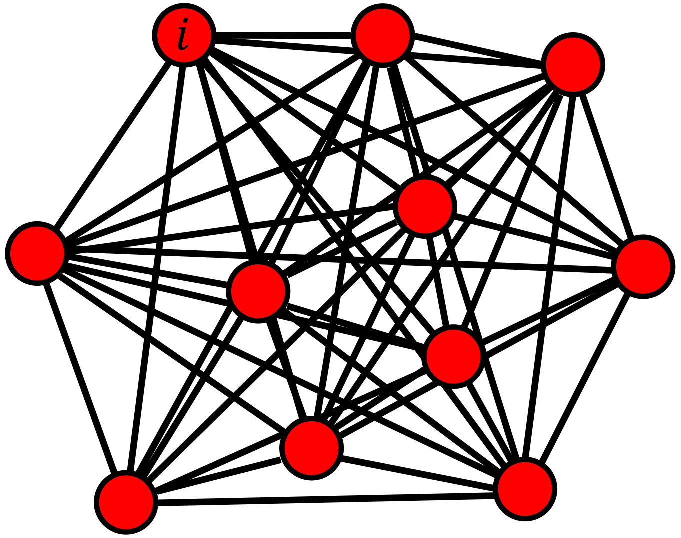
complete graph

Transformers



$$\phi\left(\mathbf{x}_i,\ \mathop{\boxed{}}_{j=1}^{n} \cancel{c_{ij}\,\psi(\mathbf{x}_j)}\right)$$

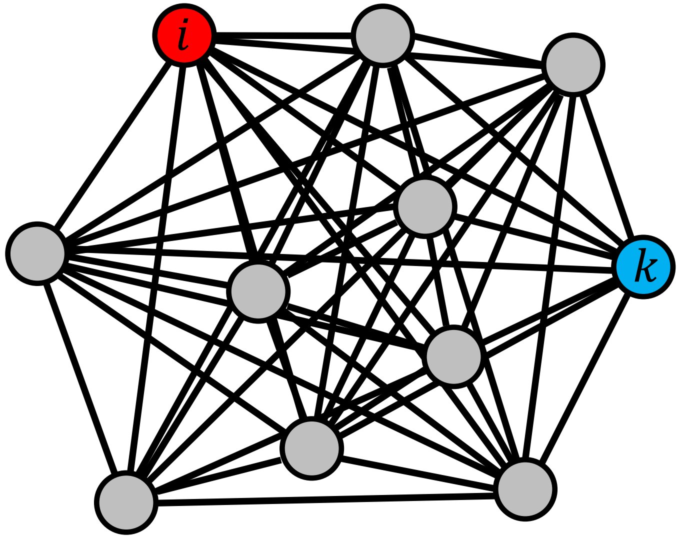
Transformers



$$\phi\left(\mathbf{x}_i, \mathop{\square}_{j=1}^{n} a(\mathbf{x}_i, \mathbf{x}_j)\psi(\mathbf{x}_j)\right)$$

Vaswani et al. 2017

Transformers

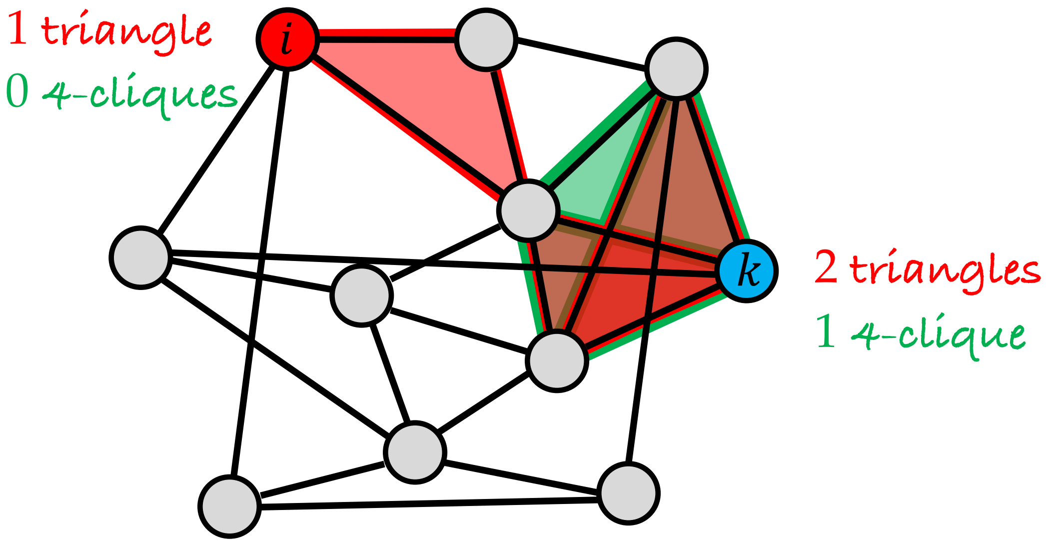


$$\phi\left(\mathbf{x}_i, \underset{j=1}{\overset{n}{\square}} a(\mathbf{x}_i, \mathbf{x}_j, \mathbf{p}_i, \mathbf{p}_j)\psi(\mathbf{x}_j)\right)$$

positional encoding

Vaswani et al. 2017

Graph Substructure Networks

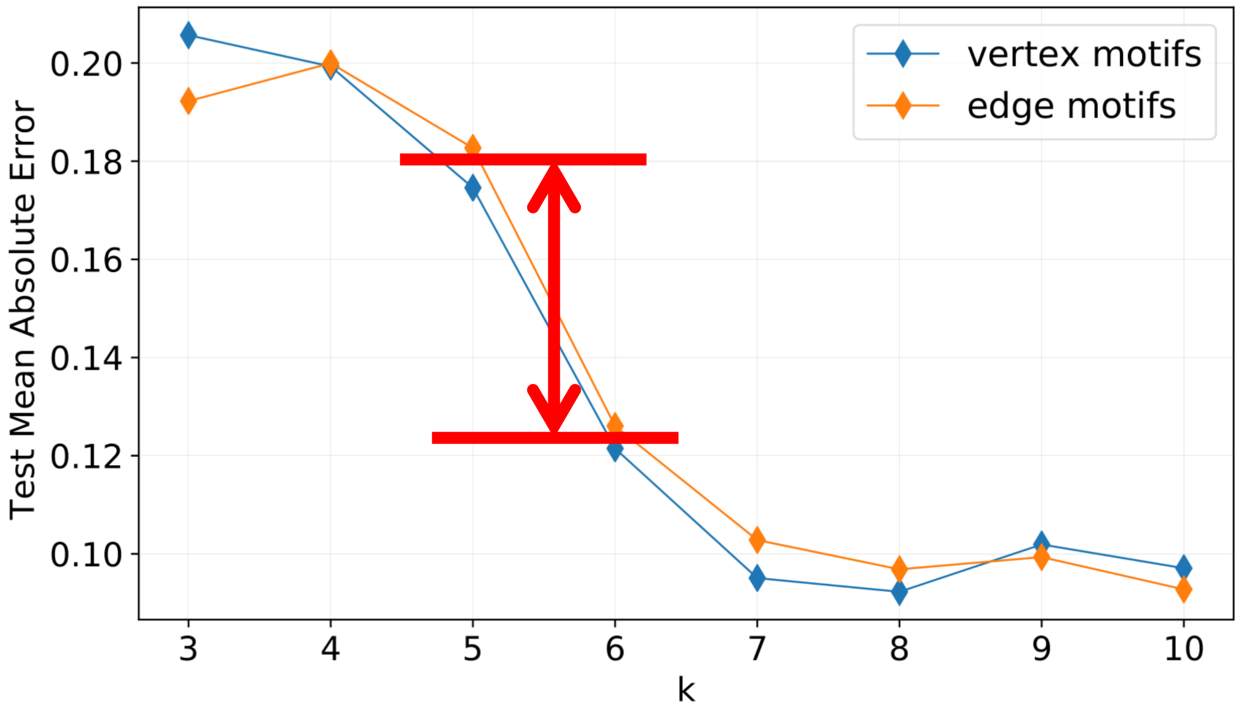


$$\phi\left(\mathbf{x}_i, \underset{j \in \mathcal{N}_i}{\square} \psi(\mathbf{x}_i, \mathbf{x}_j, \mathbf{p}_i)\right)$$

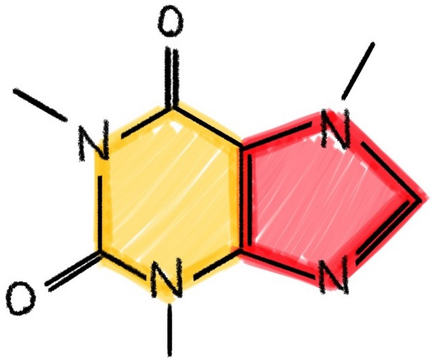
structural encoding

Bouritsas, Frasca et B 2020

Graph Substructure Networks



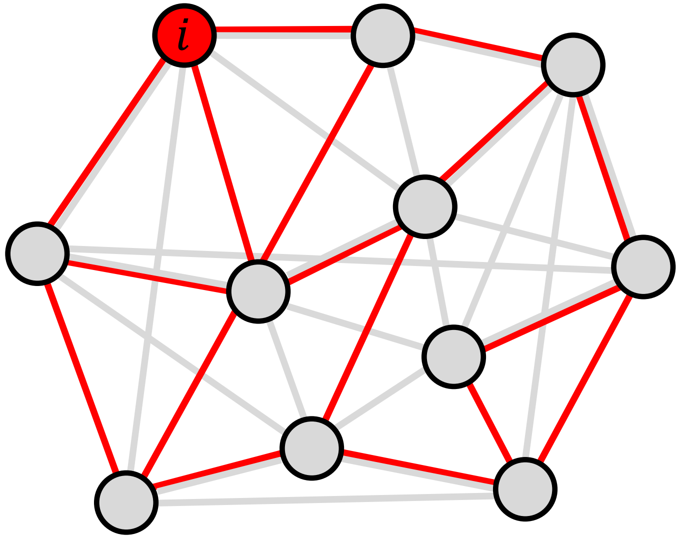
Molecule property prediction on ZINC
using GSN with *k*-cycles



Molecule of caffeine

Bouritsas, Frasca et B 2020

decouple computational graph from the input graph



sampling

rewiring

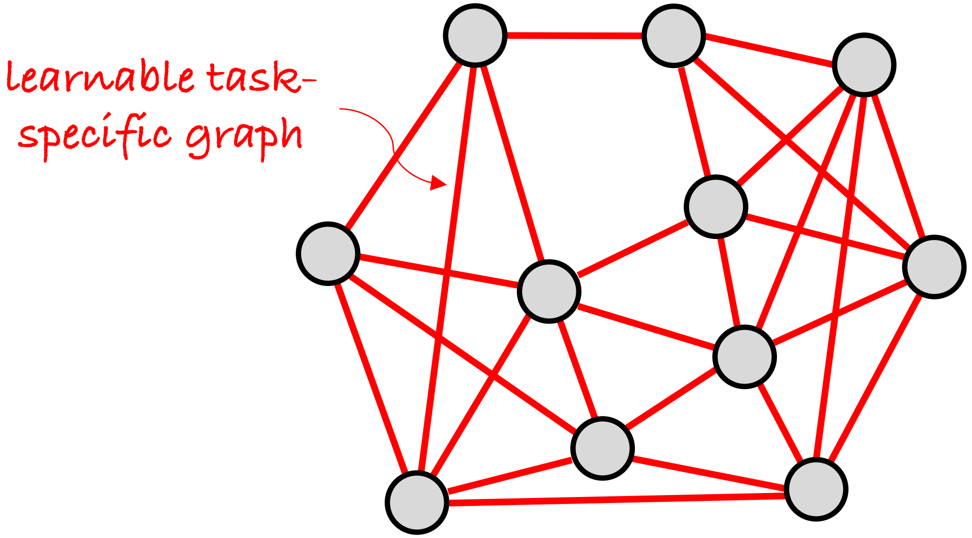
multi-hop filters

$$\phi\left(\mathbf{x}_i, \bigsqcup_{j \in \mathcal{N}_i'} \psi(\mathbf{x}_i, \mathbf{x}_j)\right)$$

new graph

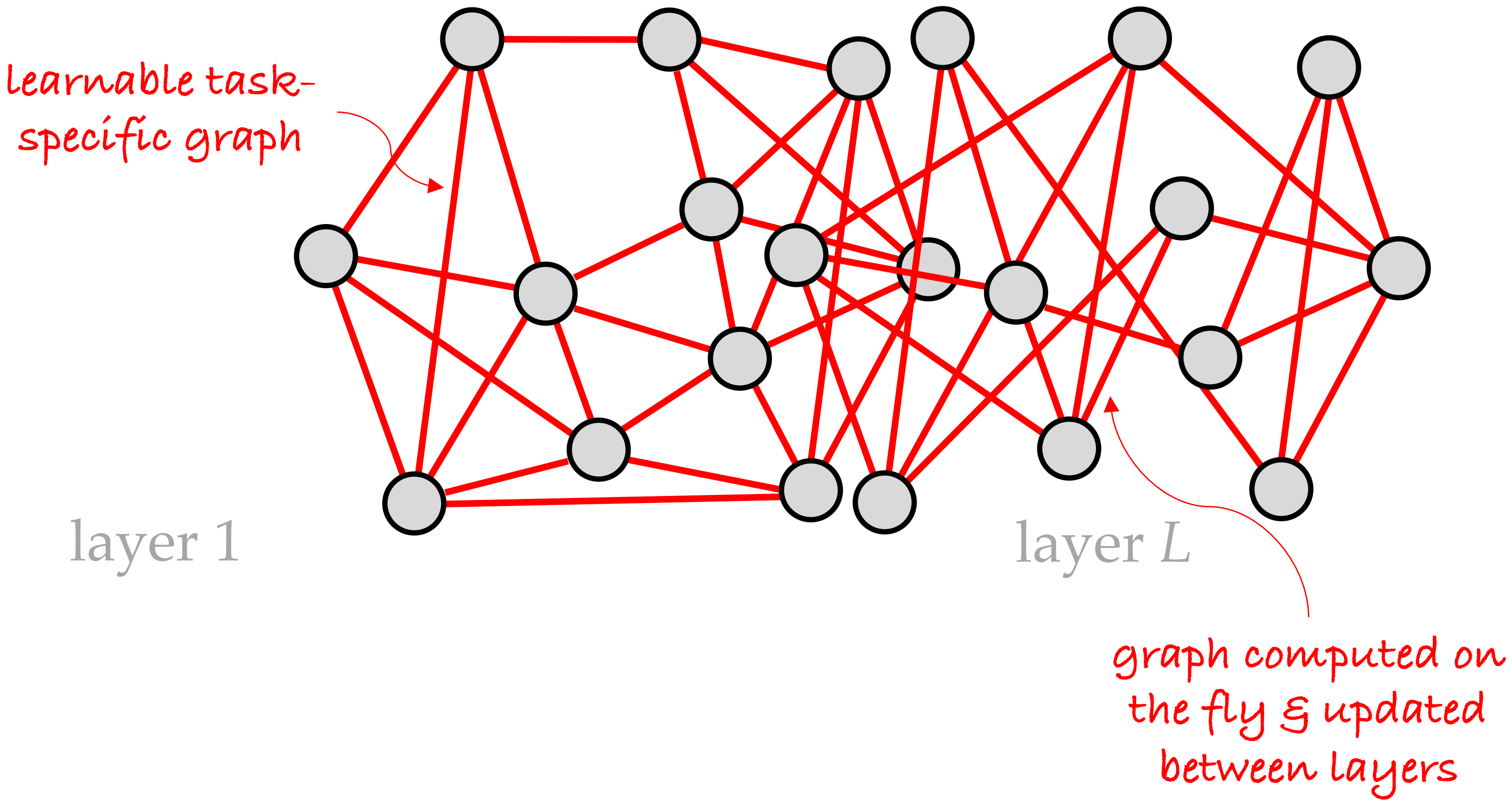
Hamilton et al 2017; Klicpera et al. 2019; Alon & Yahav 2020; Frasca, Rossi et B 2020

Latent Graph Learning

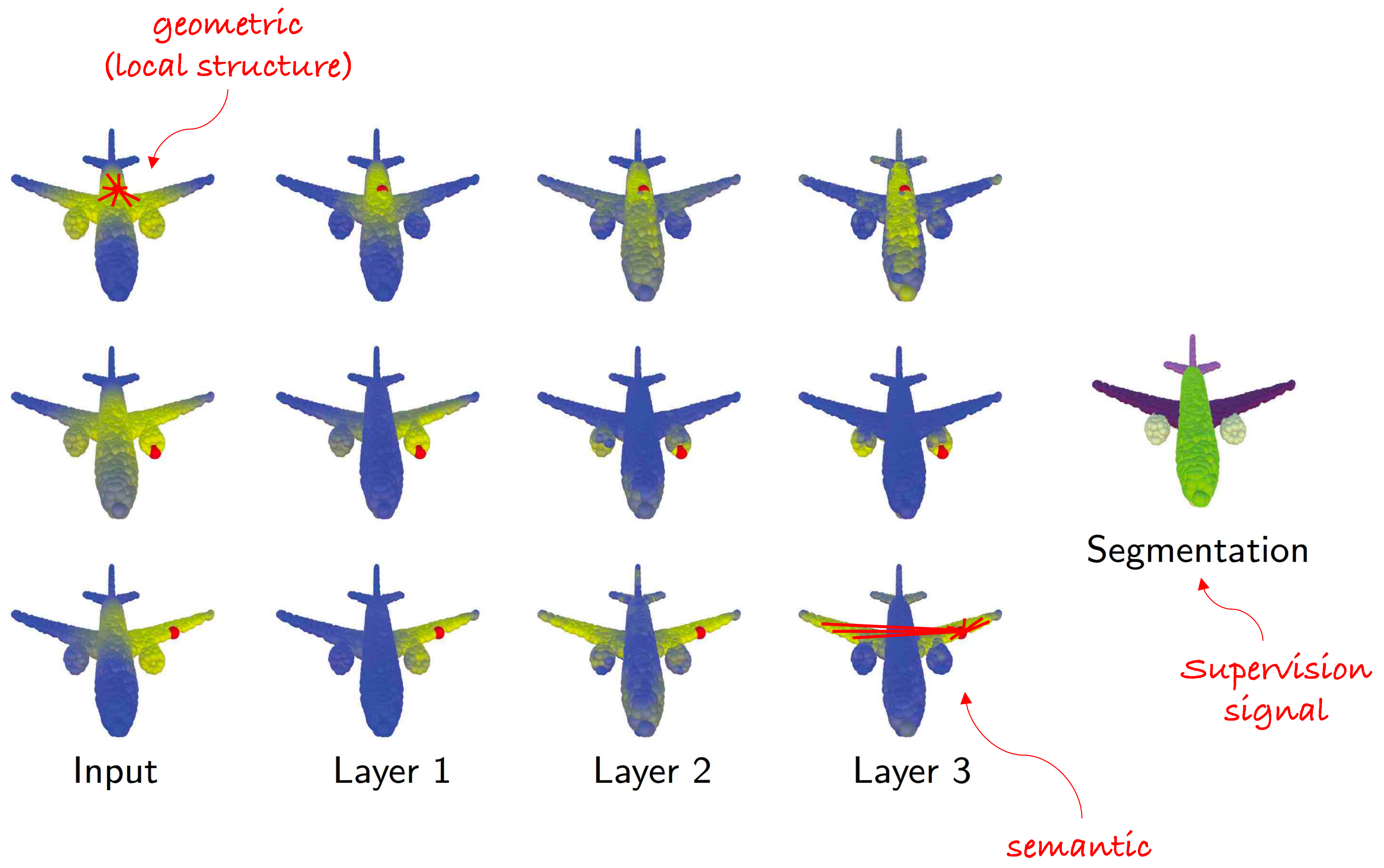


Wang et B 2018; Franceschi et al. 2019; Kipf et al. 2020; Kazi, Cosmo et B 2020; Cranmer et al. 2020

Dynamic Graph CNN

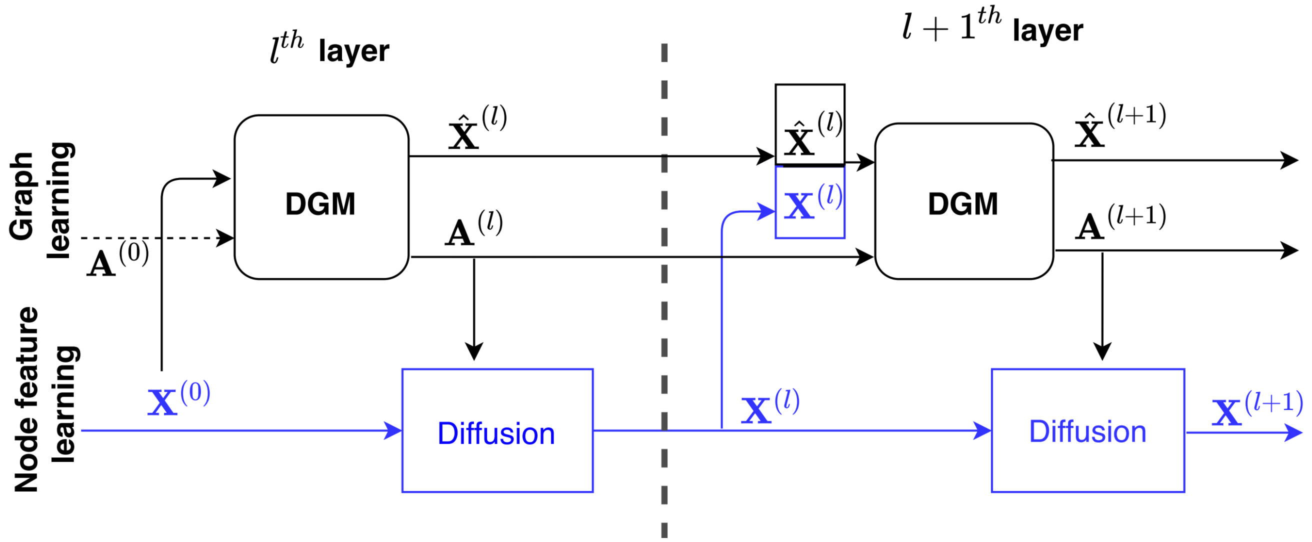


Wang et B 2018; Franceschi et al. 2019; Kipf et al. 2020; Kazi, Cosmo et B 2020; Cranmer et al. 2020



Wang et B 2018

Differentiable Graph Module



Differentiable Graph Module (DGM) allowing to construct the graph from the data and use it for feature learning

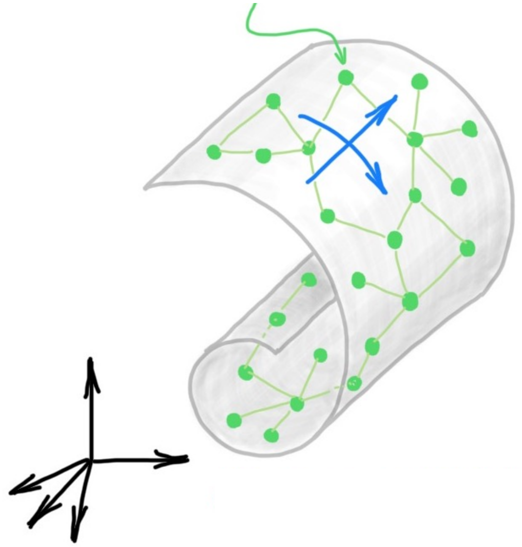
Method	TADPOLE		UK Biobank	
	Transductive	Inductive	Transductive	Inductive
DGCNN	84.59±4.33	**82.99±4.91**	58.35±0.91	**51.84±8.16**
LDS	**87.06±3.67**	†	OOM	†
cDGM	**92.91±2.50**	**91.85±2.62**	**61.32±1.51**	**55.91±3.49**
dDGM	**94.10±2.12**	**92.17±3.65**	**63.22±1.12**	**57.34±5.32**

Disease classification accuracy

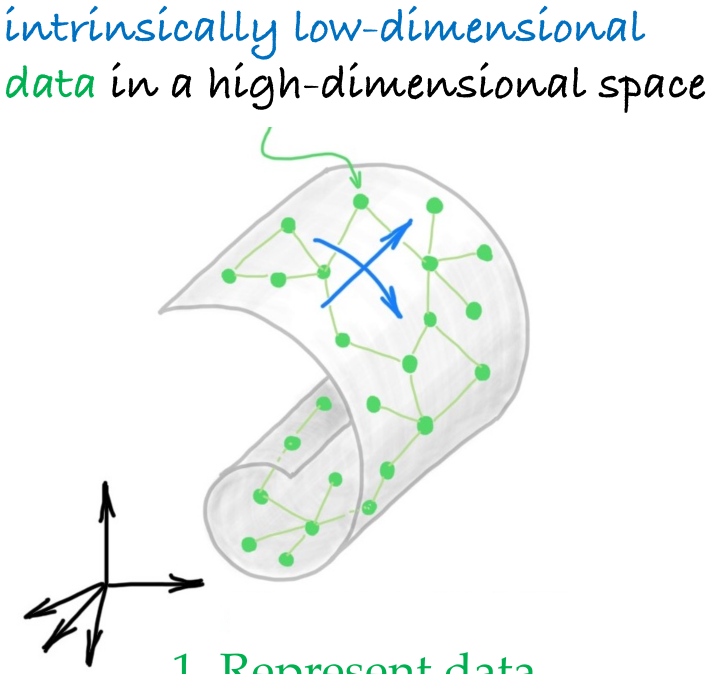
Kazi, Cosmo et B 2020

Manifold Learning

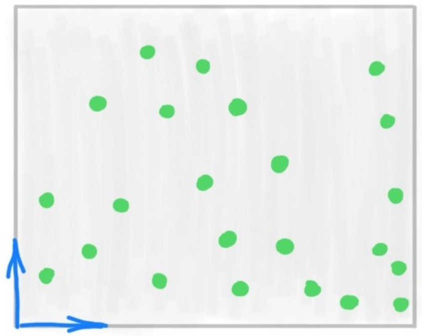
intrinsically low-dimensional
data in a high-dimensional space



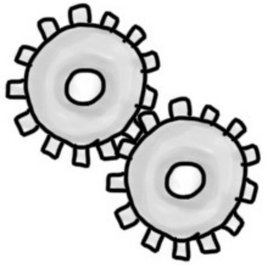
Manifold Learning



1. Represent data structure as a graph

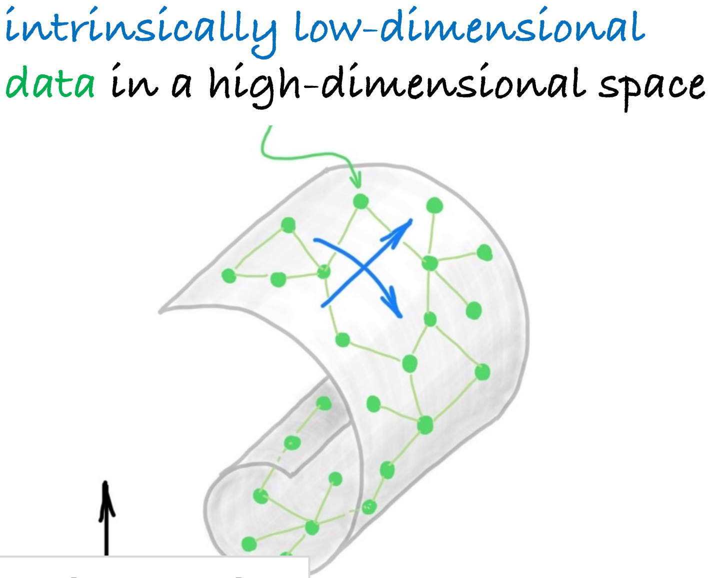


2. Compute low-dimensional embedding



3. Apply ML

Manifold Learning 2.0

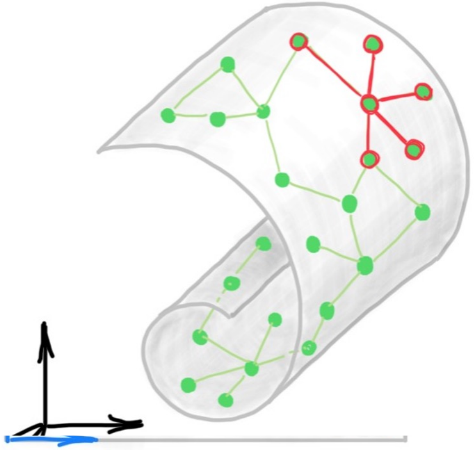


Latent graph neural networks: Manifold learning 2.0?

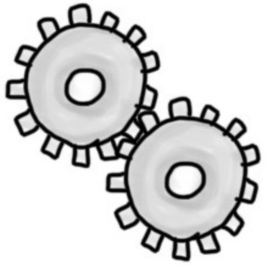
Can we use graph neural networks when the graph is unknown?

Michael Bronstein
Sep 10, 2020 · 12 min read ★

resent data
ure as a graph



1. Represent data structure as a graph apply ML directly on the graph



3. Apply ML

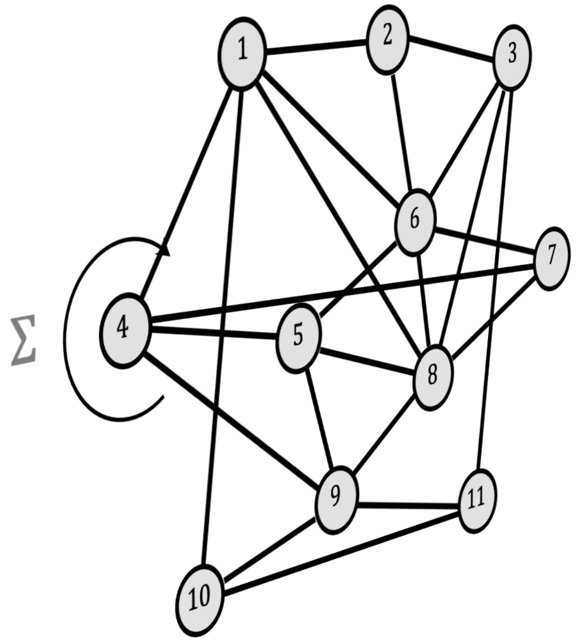
GNNs BEYOND NODES & EDGES

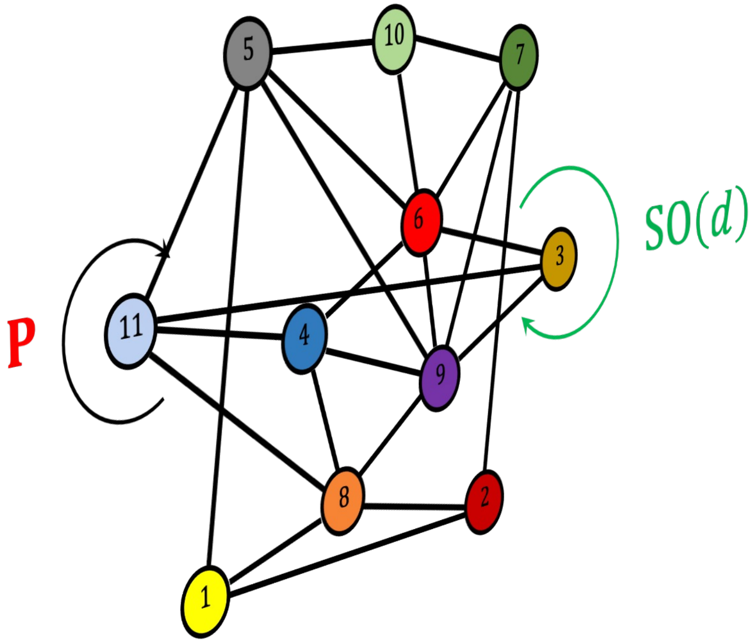
Data Symmetry in Geometric Graphs

Graph $G = (V, E)$

Node features $\mathcal{X}(G)$

functions $\mathcal{F}(\mathcal{X}(\Omega))$





Permutation group Σ_n

Permutation matrix P

Rotation R

Equivariant message passing

$$\mathbf{F}\left(\mathbf{PXR}, \mathbf{PAP}^\top\right) = \mathbf{PF}(\mathbf{X}, \mathbf{A})\mathbf{R}$$

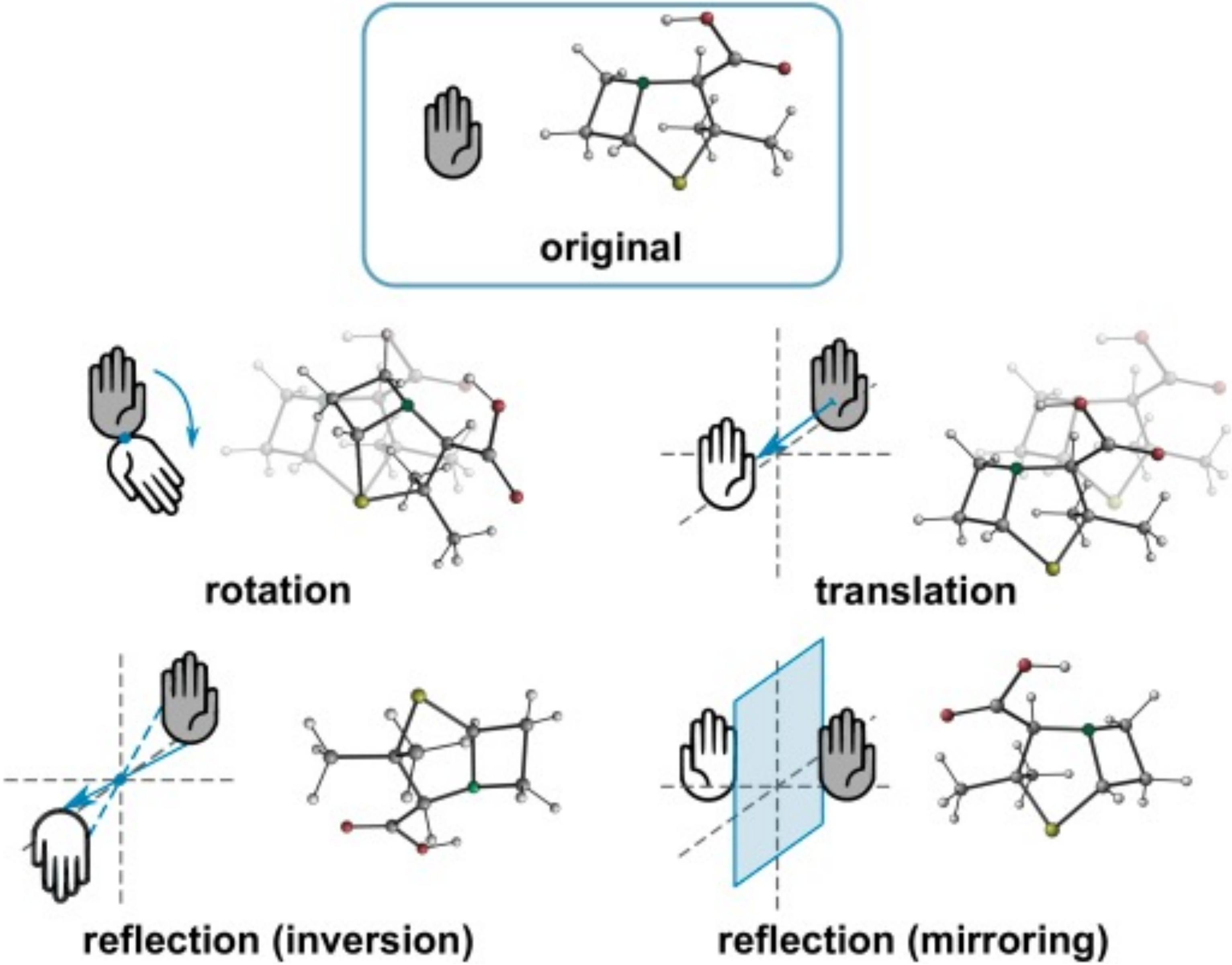
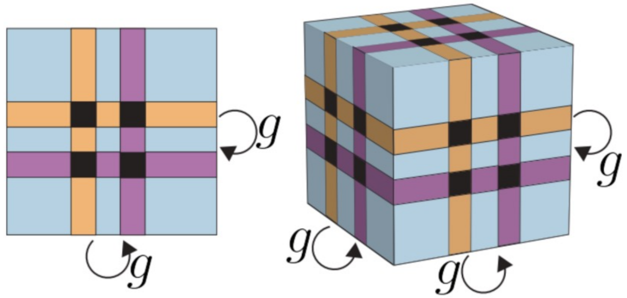


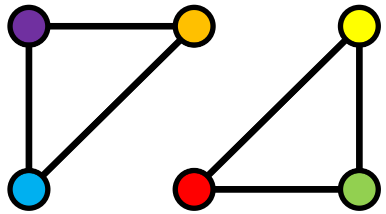
Figure: Atz et al. 2021

Towards More Expressive GNNs



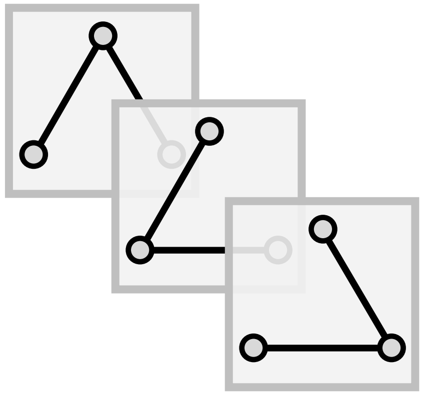
k-WL tests

Maron et al. 2019
Morris et al. 2019



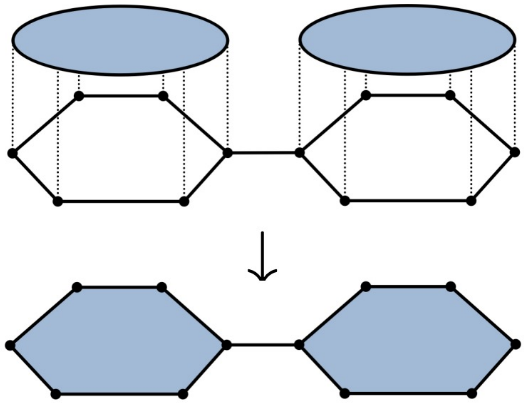
Positional & Structural encoding

Monti, Otness et B 2018
Sato 2020
Dwivedi et al. 2020
Bouritsas, Frasca et B 2020



Subgraph GNNs

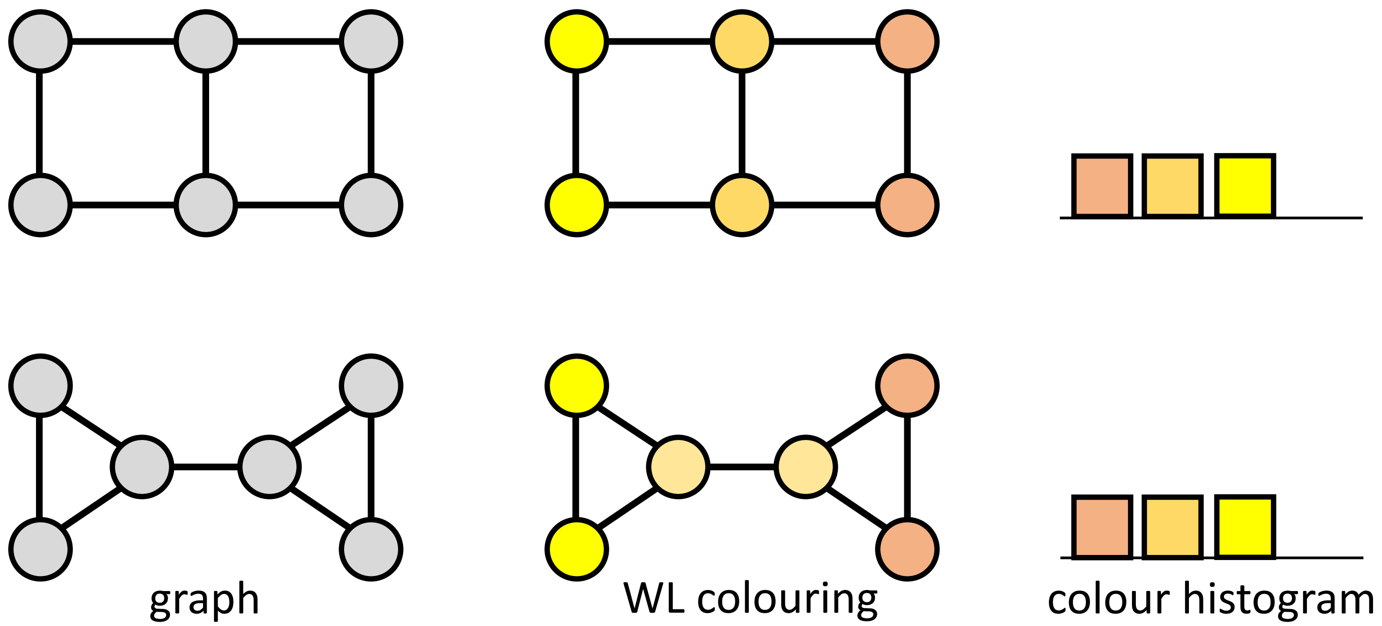
Papp et al. 2021
Cotta et al. 2021
Zhao et al. 2021
Bevilacqua, Frasca et B 2021



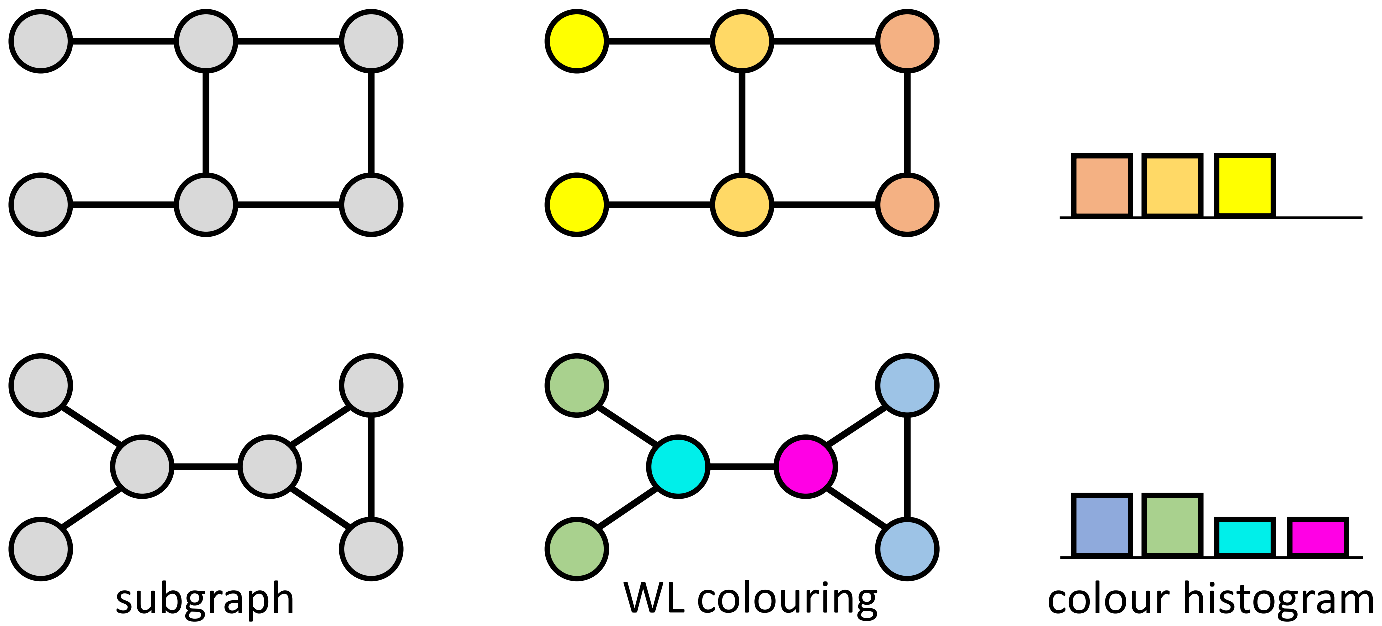
Topological message passing

Bodnar, Frasca et B 2021

Subgraph GNNs



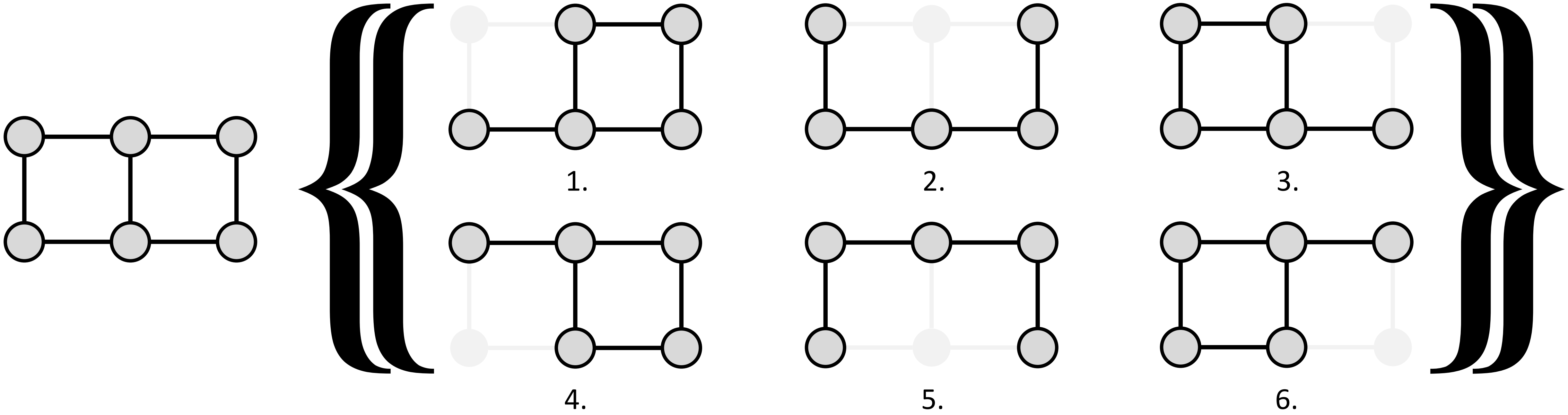
Subgraph GNNs



Graph perturbation allows to distinguish between structures otherwise indistinguishable by Weisfeiler-Lehman

Papp et al. 2021; Cotta et al. 2021; Zhao et al. 2021 Bevilacqua, Frasca et B. 2022

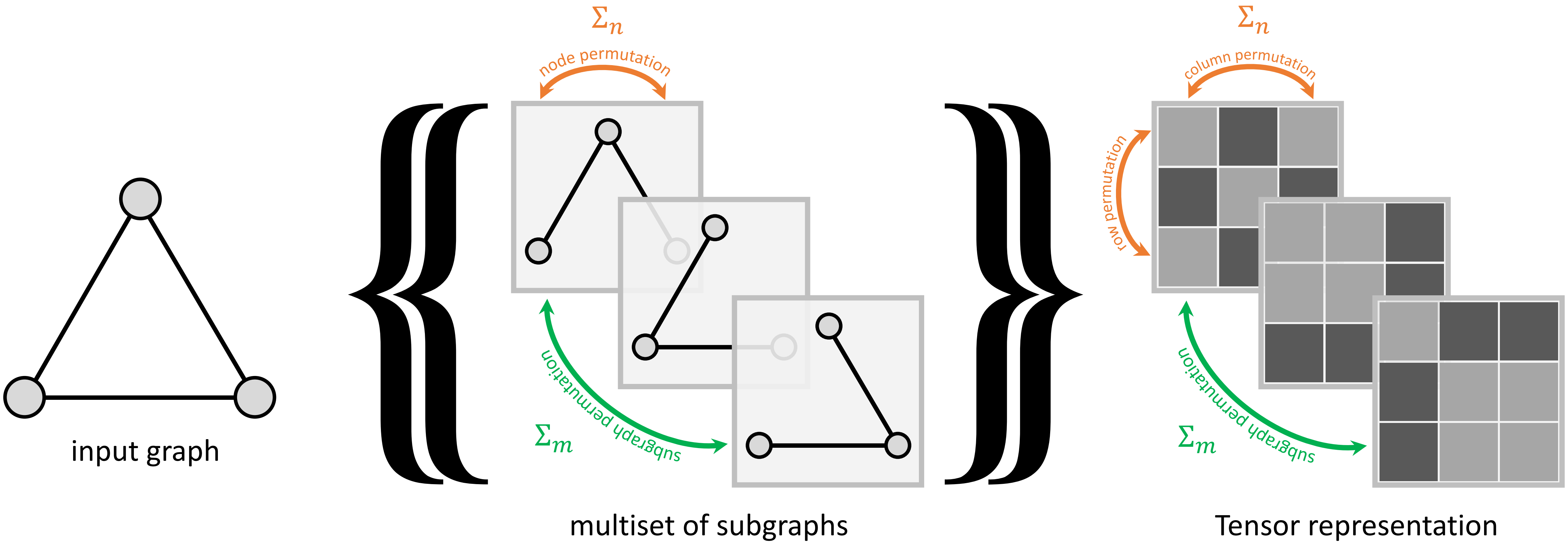
Collection of Subgraphs



A multiset of subgraphs obtained by node deletion

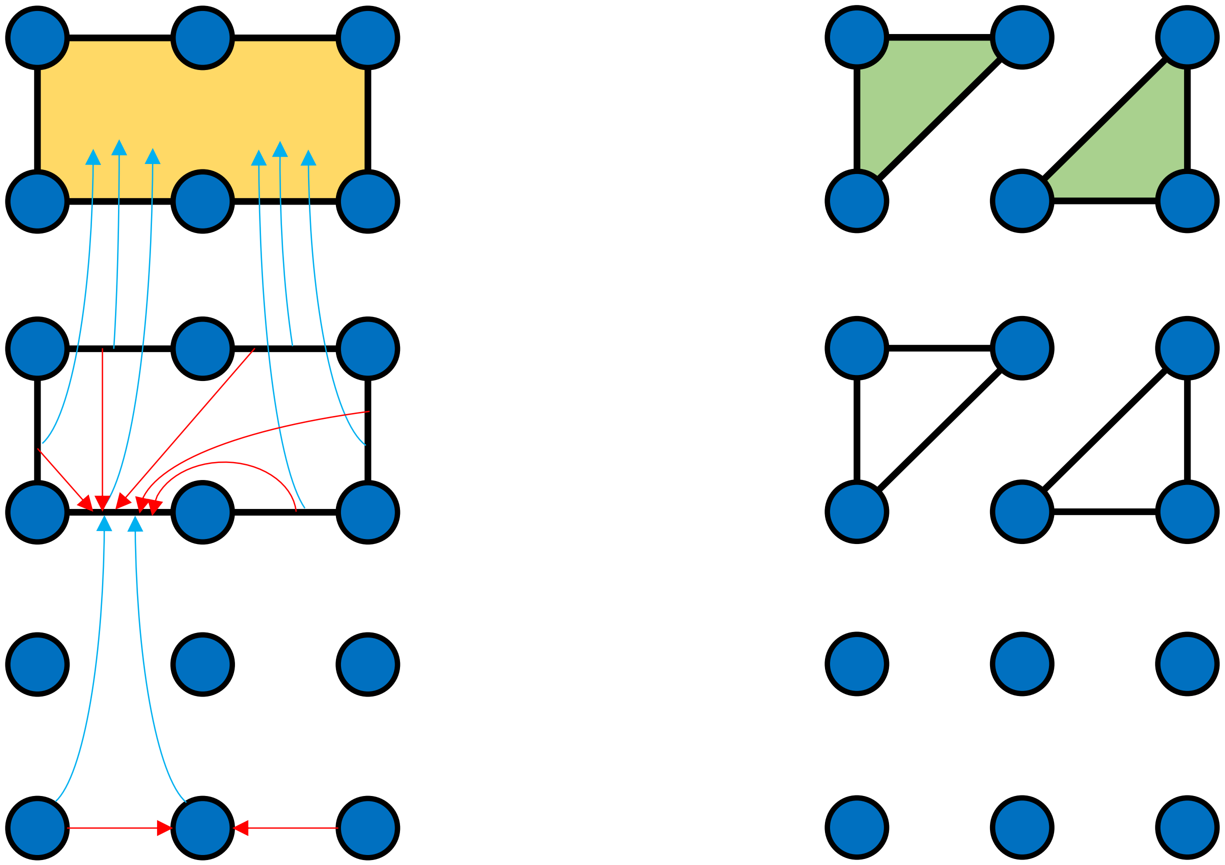
Bevilacqua, Frasca et B. 2022

Equivariant Subgraph Aggregation Networks



Bevilacqua, Frasca et B. 2022

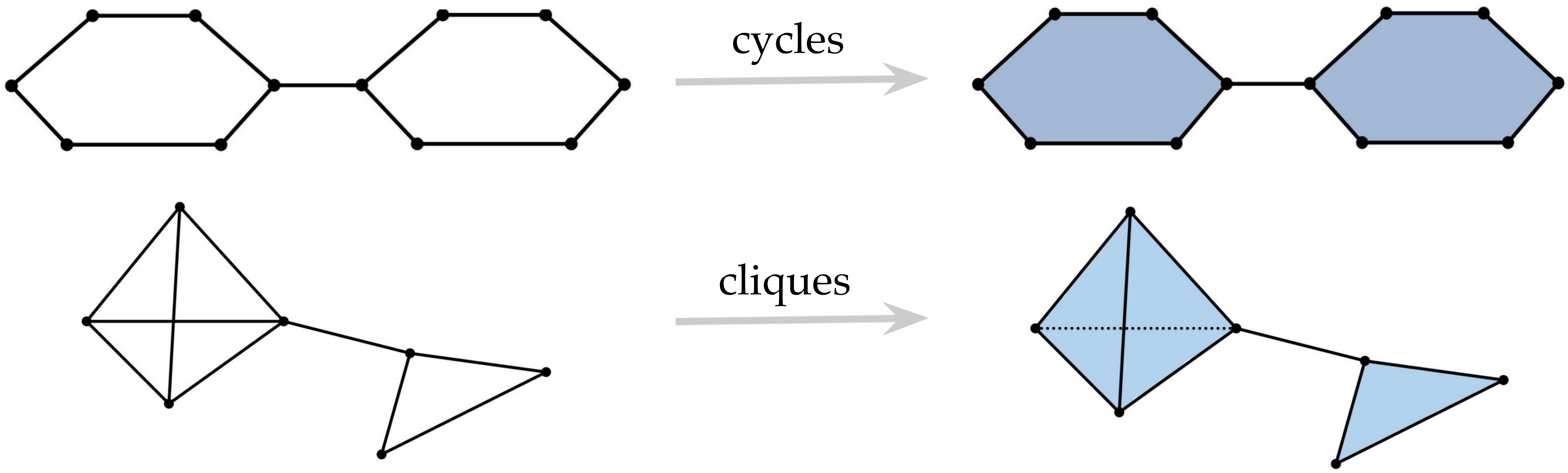
Topological Message Passing



Lift the graph into a cell complex

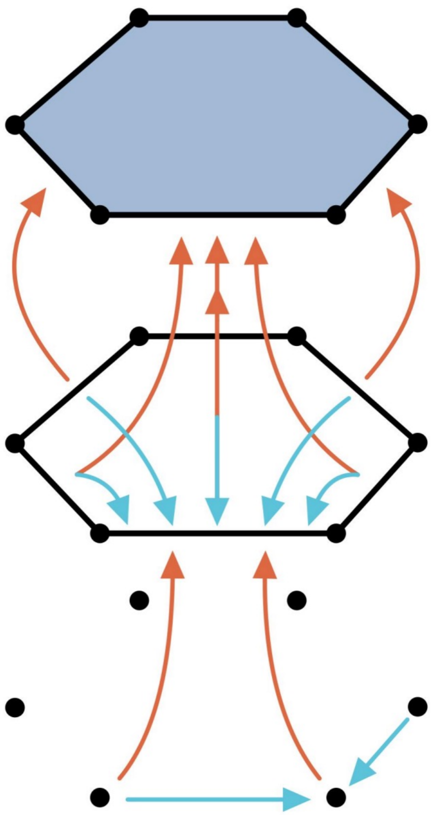
Bodnar, Frasca et B. 2021

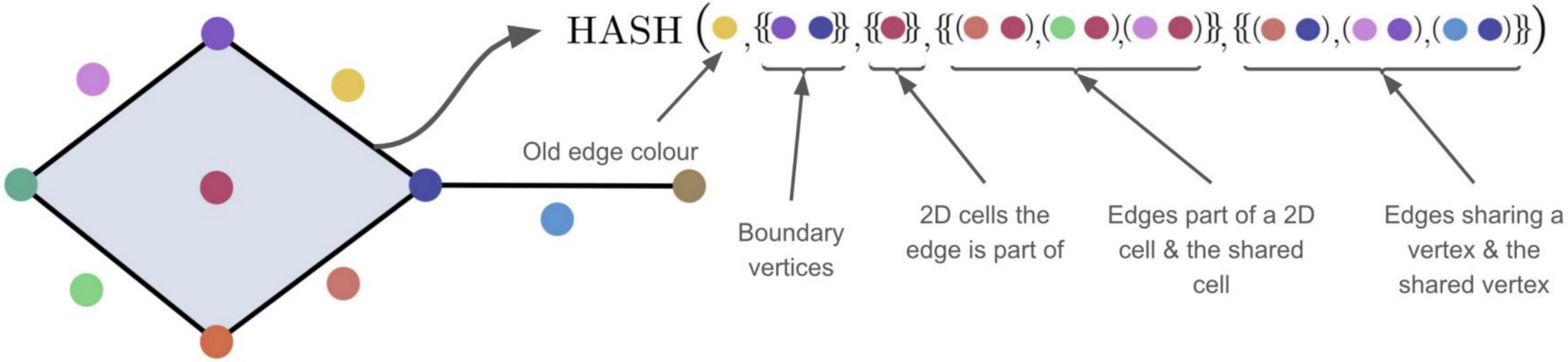
Cellular Lifting Maps



Bodnar, Frasca et B. 2021

Cellular Weisfeiler-Lehman





Theorem 15. *For some finite k, there exists a pair of graphs indistinguishable by 3-WL but distinguishable by CWL with the lifting maps from Corollary 14. For the clique complex and induced cycle liftings, the statement holds for $k \geq 4$. For the simple cycle based lifting, it holds for $k \geq 8$.*

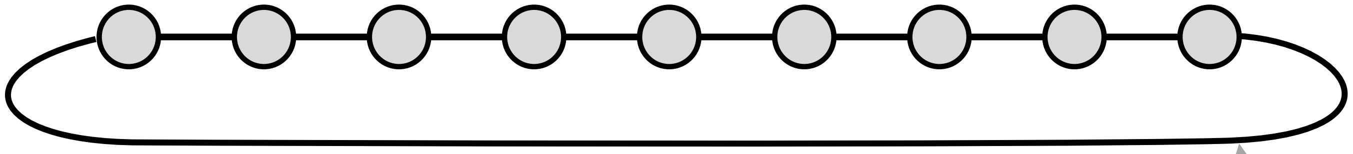
Corollary 14. *For all $k \geq 3$, the following lifting transformations make CWL strictly more powerful than the WL test. (1) The clique complex lifting considering cliques of size at most k. (2) The map that attaches 2-cells to all the simple cycles of size at most k. (3) The map that attaches 2-cells to all the induced cycles of size at most k. (4) The union of all the transformations above.*

Cellular WL is strictly more powerful than WL with appropriate lifting transformation

Bodnar, Frasca et B. 2021

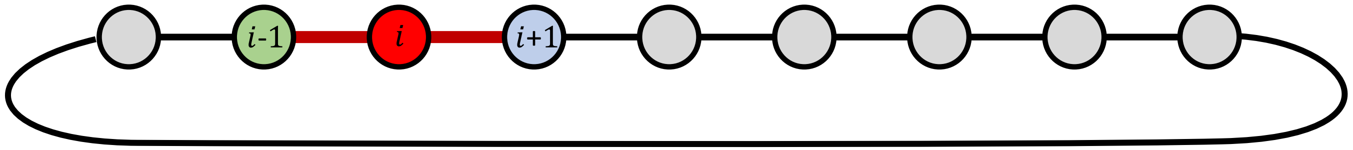
GRIDS

Grids vs Graphs



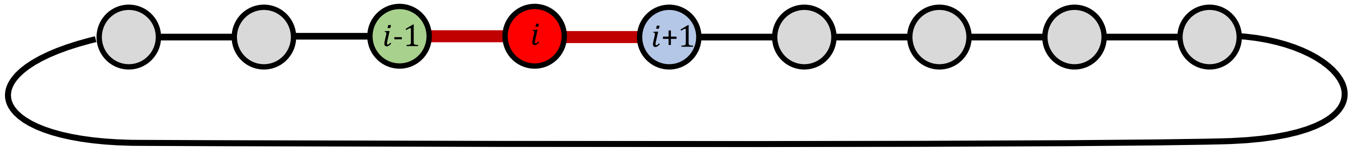
grid = ring graph

Grids vs Graphs



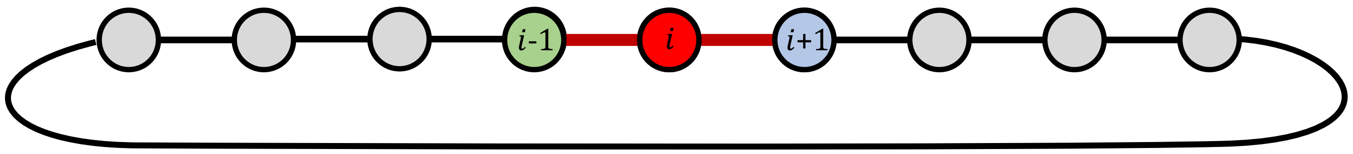
fixed neighbourhood structure

Grids vs Graphs



fixed neighbourhood structure

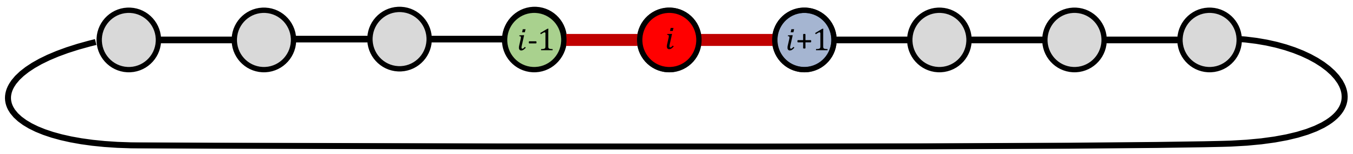
Grids vs Graphs



local aggregation function

$$f(\mathbf{x}_i) = \phi(\mathbf{x}_i, \{\mathbf{x}_{i-1}, \mathbf{x}_{i+1}\})$$

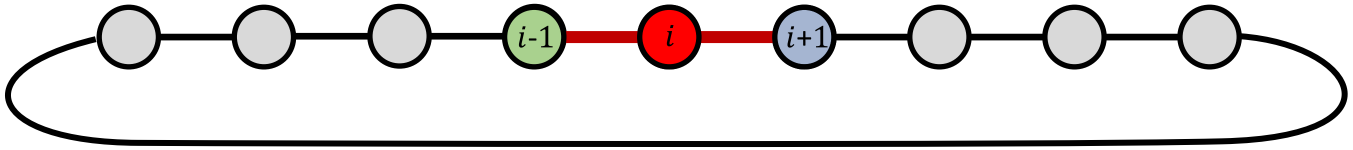
Grids vs Graphs



local aggregation function

$$f(\mathbf{x}_i) = \phi(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})$$

Grids vs Graphs



linear local aggregation function

$$f(\mathbf{x}_i) = a\mathbf{x}_{i-1} + b\mathbf{x}_i + c\mathbf{x}_{i+1}$$

Convolution

$$f(\mathbf{X}) = \begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & & & & \\ & & a & b & c \\ c & & & a & b \end{bmatrix} \begin{bmatrix} \mathbf{X} \end{bmatrix}$$

circulant matrix = convolution

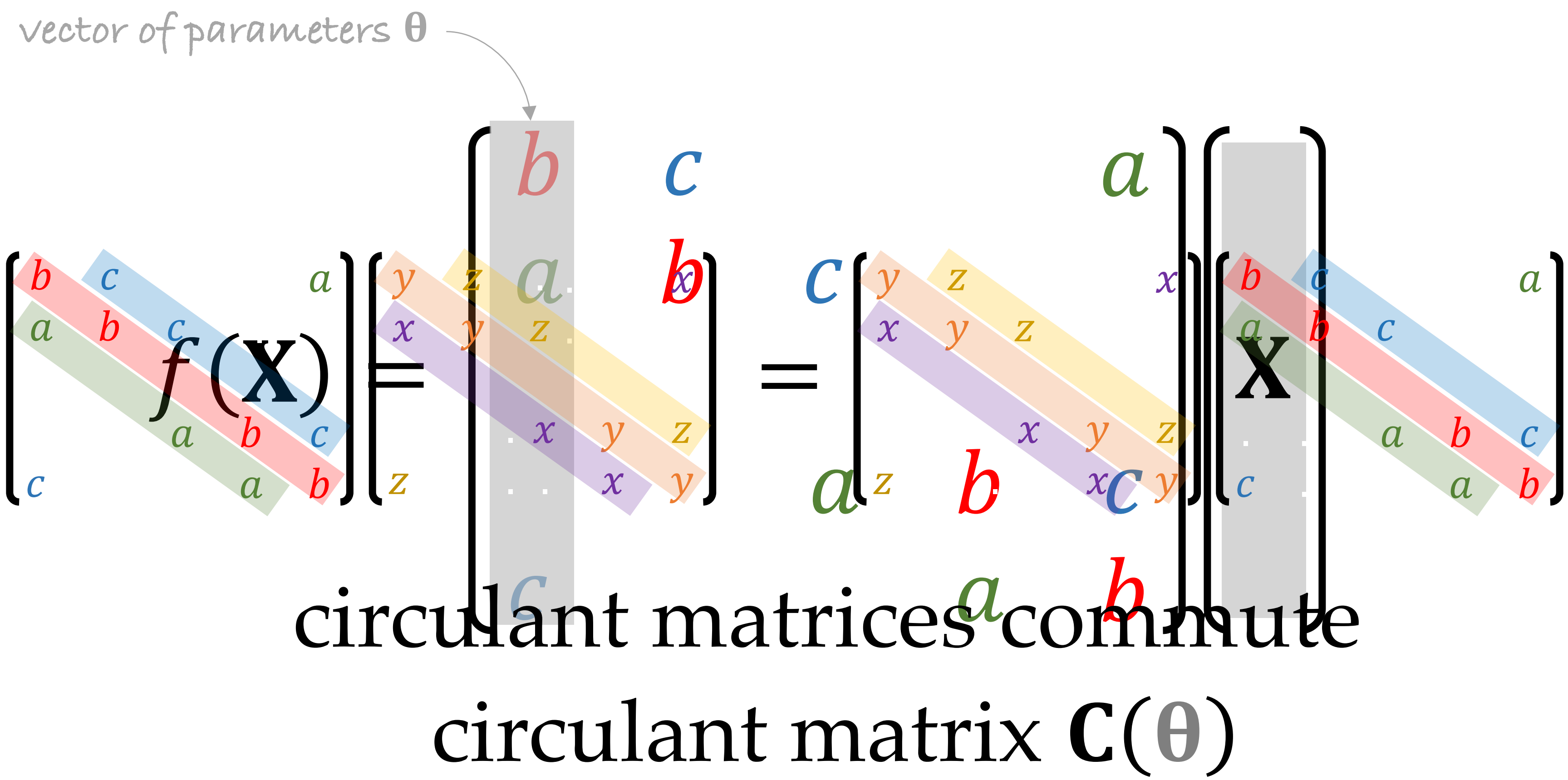
Convolution

vector of parameters $\boldsymbol{\theta}$

$$f(\mathbf{X}) = \begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & & & & \\ & & a & b & c \\ c & & & a & b \end{bmatrix} \begin{bmatrix} \\ \mathbf{X} \\ \\ \end{bmatrix}$$

circulant matrix $\mathbf{C}(\boldsymbol{\theta})$

Deriving Convolution from Symmetry



Deriving Convolution from Symmetry

$$
\begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & \ddots & \ddots & \ddots & \\ & & a & b & c \\ c & & & a & b \end{bmatrix}
\overset{\textbf{shift S}}{\begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 1 & & & & \end{bmatrix}}
=
\overset{\textbf{shift S}}{\begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 1 & & & & \end{bmatrix}}
\begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & \ddots & \ddots & \ddots & \\ & & a & b & c \\ c & & & a & b \end{bmatrix}
$$

circulant matrix $\Longrightarrow$ commutes with shift

Deriving Convolution from Symmetry

$$\begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & \ddots & \ddots & \ddots & \\ & & a & b & c \\ c & & & a & b \end{bmatrix} \overset{\textbf{shift S}}{\begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 1 & & & & \end{bmatrix}} = \overset{\textbf{shift S}}{\begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 1 & & & & \end{bmatrix}} \begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & \ddots & \ddots & \ddots & \\ & & a & b & c \\ c & & & a & b \end{bmatrix}$$

$$\text{convolution} \Longrightarrow \text{shift-equivariant}$$

$$\mathbf{CS} = \mathbf{SC}$$

Deriving Convolution from Symmetry

$$\begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & & \ddots & & \\ & & a & b & c \\ c & & & a & b \end{bmatrix} \overset{\textbf{shift S}}{\begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 1 & & & & \end{bmatrix}} = \overset{\textbf{shift S}}{\begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 1 & & & & \end{bmatrix}} \begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & & \ddots & & \\ & & a & b & c \\ c & & & a & b \end{bmatrix}$$

convolution $\Leftrightarrow$ shift-equivariant

convolution **emerges** from translation symmetry

Deriving Fourier Transform from Symmetry

same eigenbasis for all convolutions

different eigenvalues for each convolution

$$\begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & & & & \\ & & a & b & c \\ c & & & a & b \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} — & \mathbf{u}_1^* & — \\ & \vdots & \\ — & \mathbf{u}_n^* & — \end{bmatrix}$$

commuting matrices are jointly diagonalisable

by eigenvectors of **S**

Deriving Fourier Transform from Symmetry

Fourier basis

$$\mathbf{u}_k = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 \\ e^{\mathrm{i}\frac{2\pi}{n}k} \\ \vdots \\ e^{\mathrm{i}\frac{2\pi}{n}(n-1)k} \end{bmatrix}$$

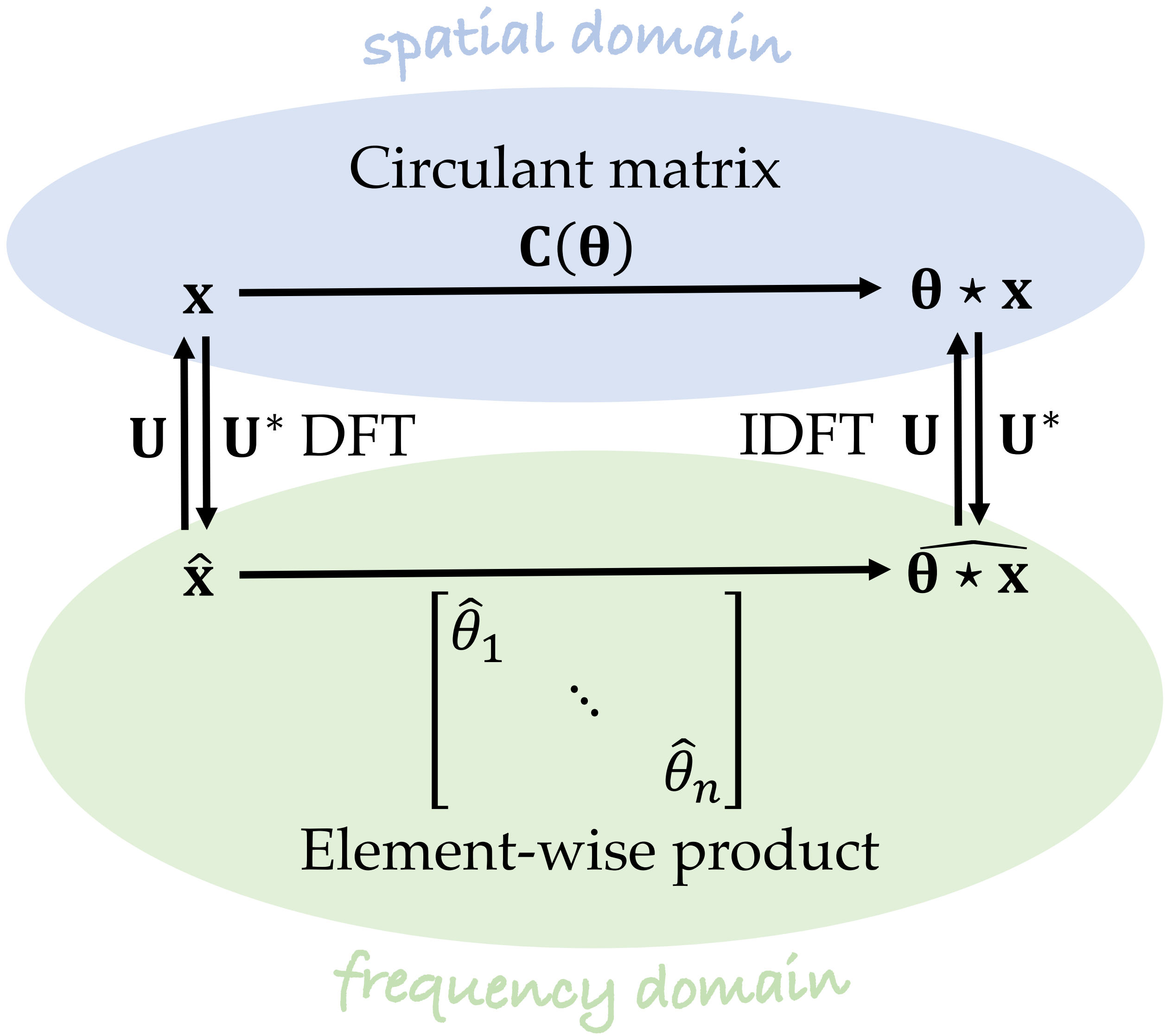
Fourier transform

$$\hat{\boldsymbol{\theta}} = \mathbf{U}^* \boldsymbol{\theta}$$

$$\begin{bmatrix} b & c & & & a \\ a & b & c & & \\ & & \ddots & & \\ & & a & b & c \\ c & & & a & b \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 & & & \\ & \hat{\theta}_2 & & \\ & & \ddots & \\ & & & \hat{\theta}_n \end{bmatrix} \begin{bmatrix} — & \mathbf{u}_1^* & — \\ & \vdots & \\ — & \mathbf{u}_n^* & — \end{bmatrix}$$

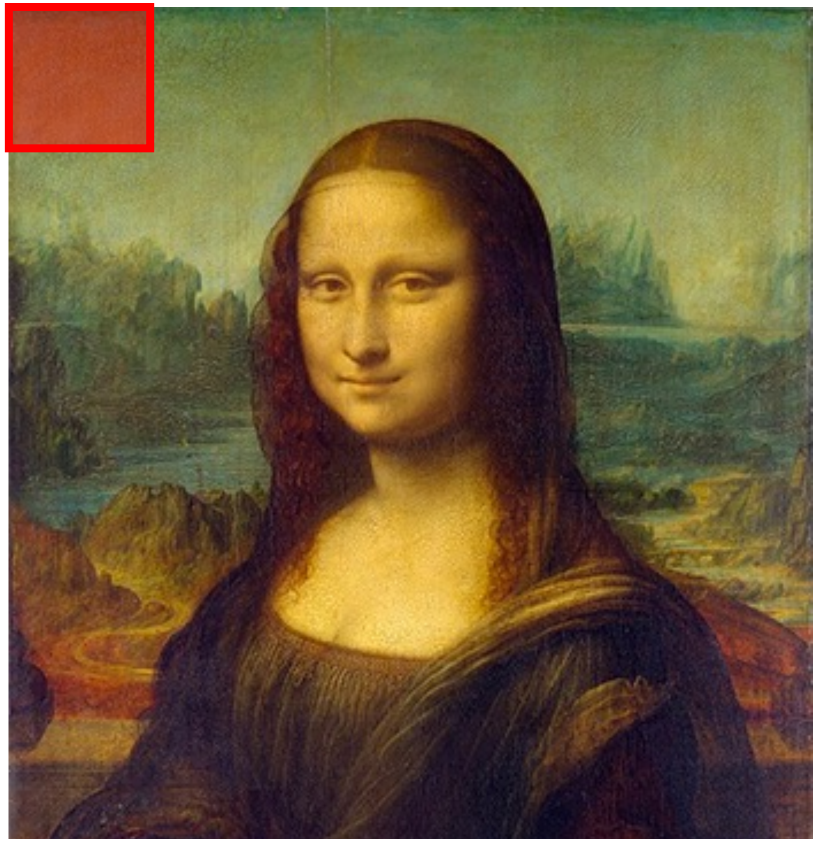
$\boldsymbol{\theta}$

commuting matrices are jointly diagonalisable by Fourier Transform


spatial domain
Circulant matrix
$\mathbf{C}(\boldsymbol{\theta})$
$\mathbf{x}$
$\boldsymbol{\theta} \star \mathbf{x}$
$\mathbf{U}$
$\mathbf{U}^*$ DFT
IDFT $\mathbf{U}$
$\mathbf{U}^*$
$\hat{\mathbf{x}}$
$\widehat{\boldsymbol{\theta} \star \mathbf{x}}$
$\begin{bmatrix} \hat{\theta}_1 & & \\ & \ddots & \\ & & \hat{\theta}_n \end{bmatrix}$
Element-wise product
frequency domain

GROUPS

Convolution, revisited



Convolution, revisited

convolution = matching shifted filter

$$(x \star \psi)(u) = \langle x, T_u \psi \rangle = \int_{-\infty}^{+\infty} x(v)\psi(u-v)\mathrm{d}v$$

shift vector

shift operator

domain Ω = symmetry group $\mathfrak{G}$

Group Convolution

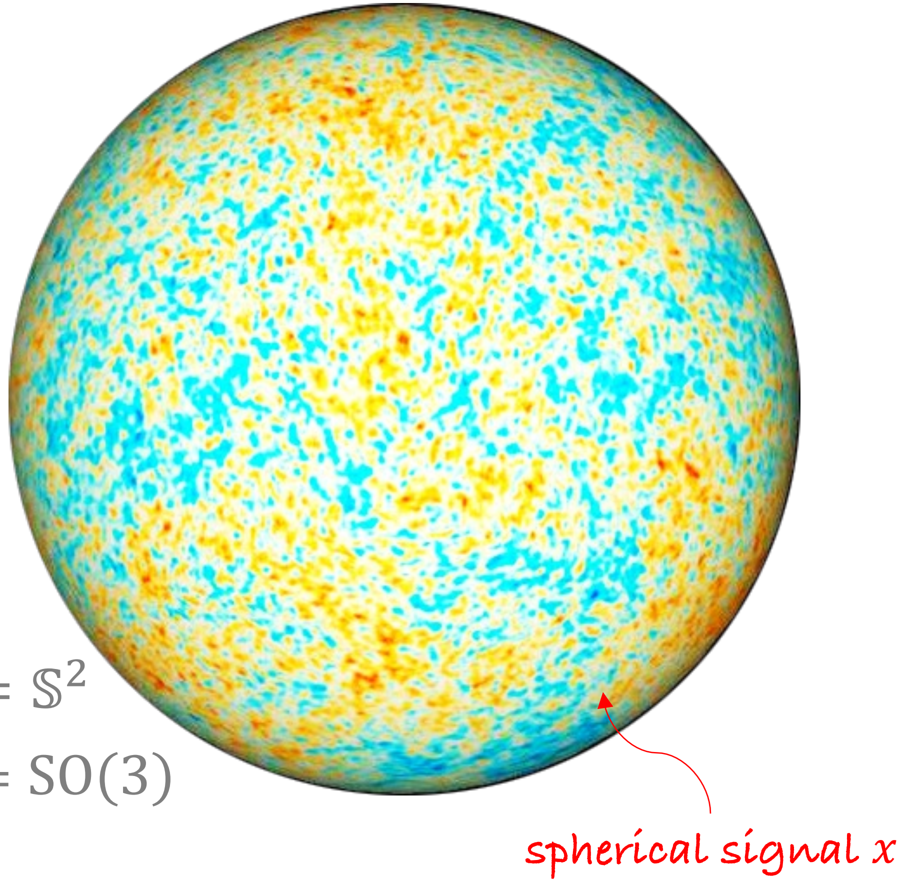
convolution = matching transformed filter

$$(x \star \psi)(\mathfrak{g}) = \langle x, \rho(\mathfrak{g})\psi \rangle = \int_{\Omega} x(v)\psi(\mathfrak{g}^{-1}v)\mathrm{d}v$$

group element

group representation

Convolution on the Sphere



sphere $\Omega = \mathbb{S}^2$

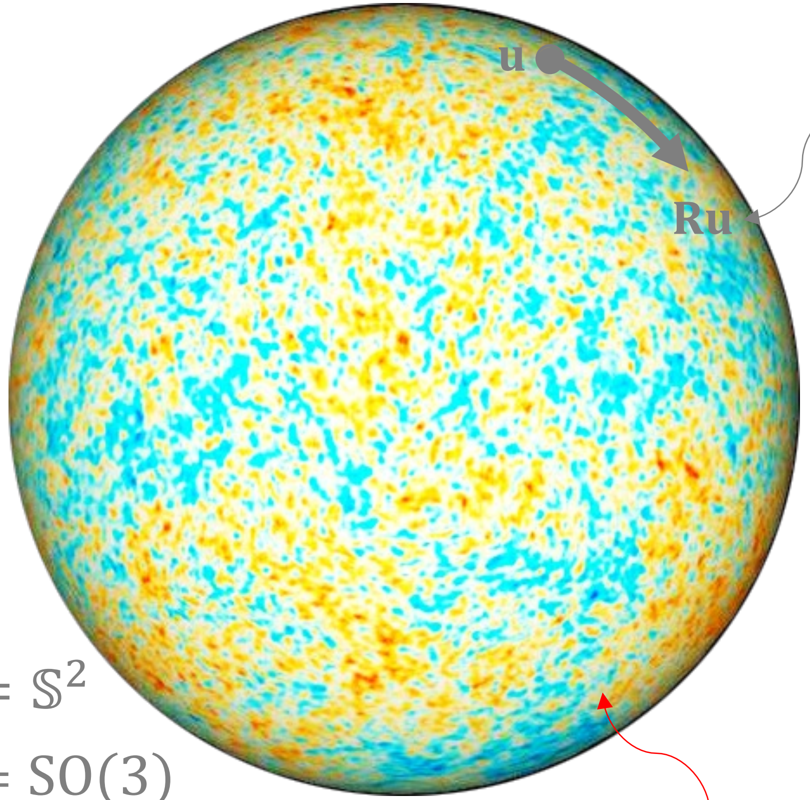
rotation group $\mathfrak{G} = \mathrm{SO}(3)$

Cohen, Welling 2016

Convolution on the Sphere

$$(x \star \psi)(\mathbf{R}) = \int_{\mathbb{S}^2} x(\mathbf{u})\psi(\mathbf{R}^{-1}\mathbf{u})d\mathbf{u}$$

signal on SO(3)

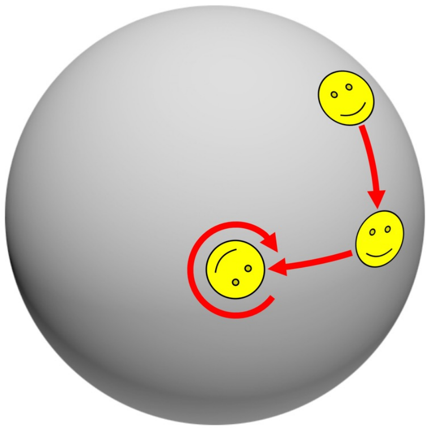


rotation transformation

sphere $\Omega = \mathbb{S}^2$

rotation group $\mathfrak{G} = \mathrm{SO}(3)$

spherical signal x



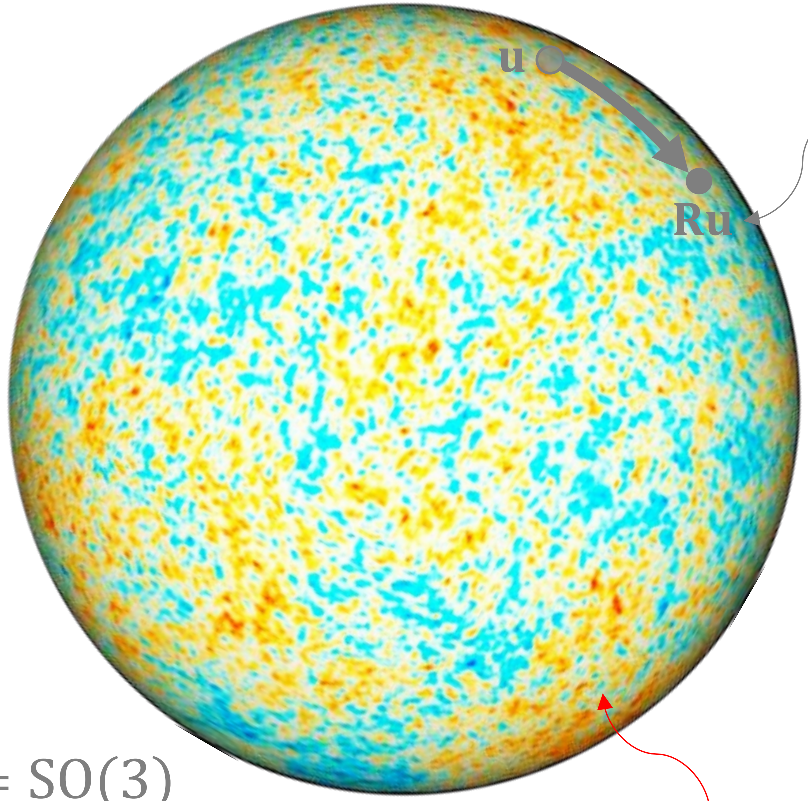
Cohen, Welling 2016

Convolution on the Sphere

$$((\color{red}{x} \star \color{red}{\psi}) \star \color{red}{\phi})(\mathbf{R}) = \int_{\mathrm{SO}(3)} (\color{red}{x} \star \color{red}{\psi})(\mathbf{Q})\color{red}{\phi}(\mathbf{R}^{-1}\mathbf{Q})\mathrm{d}\mathbf{Q}$$

Convolution on SO(3)

rotation transformation

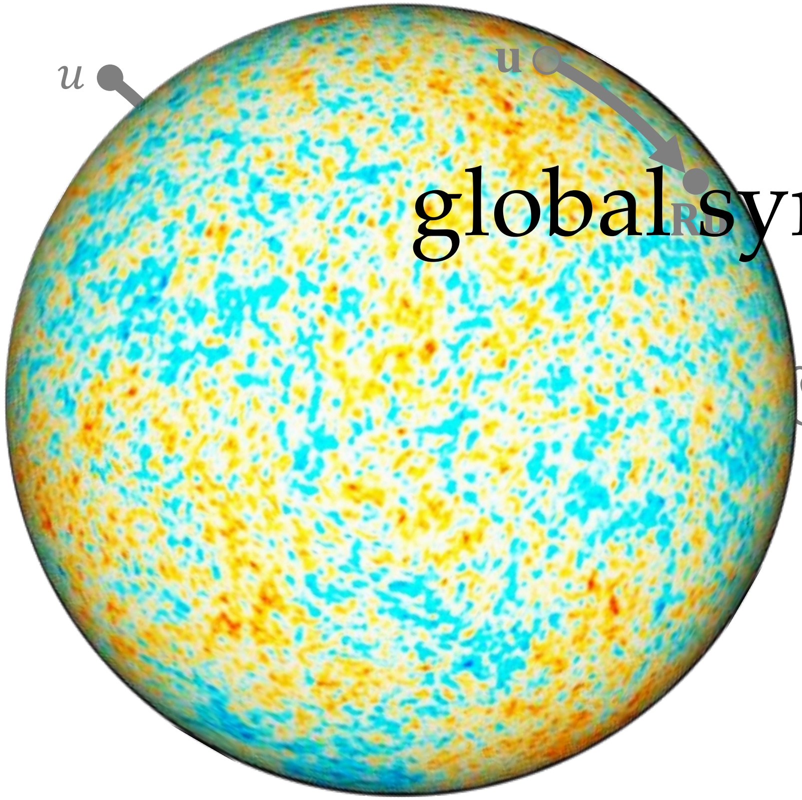


spherical signal x

$\Omega = \mathfrak{G} = \mathrm{SO}(3)$

Cohen, Welling 2016

Homogeneous Spaces

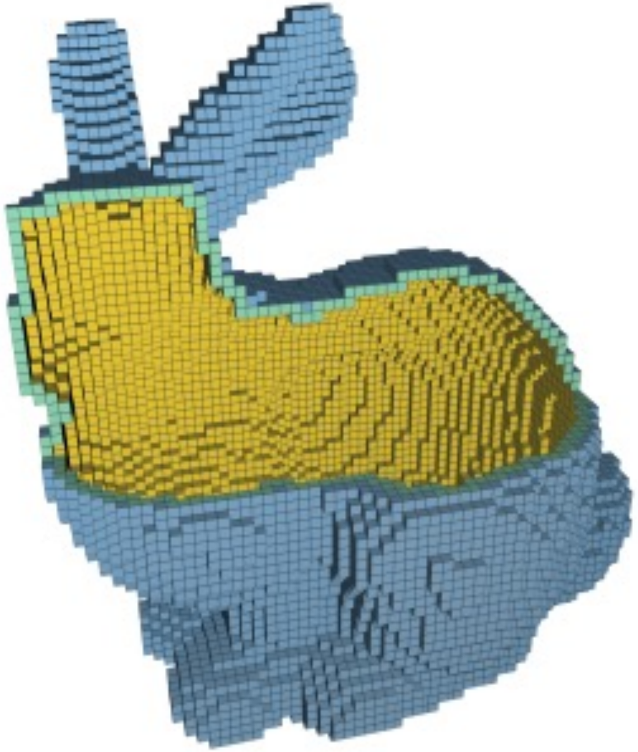


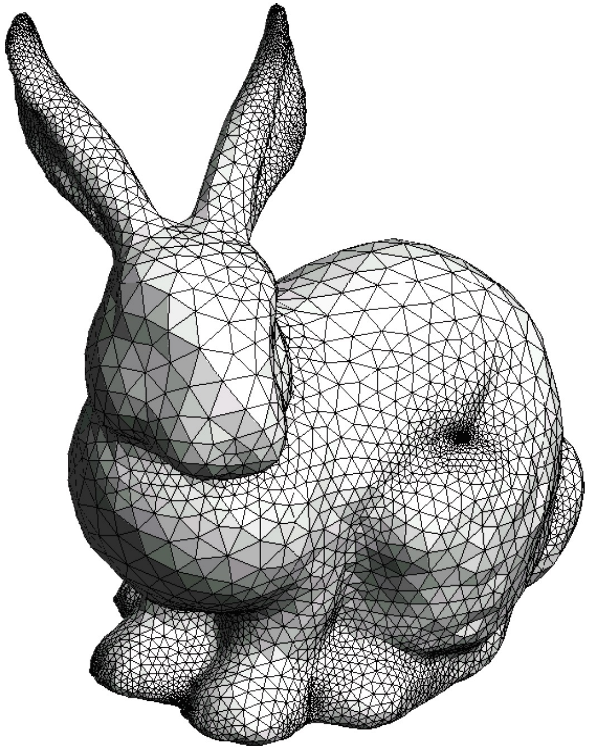
global symmetry group

$\mathfrak{G}$ s.t. $gu = v$

MANIFOLDS & MESHES

Why Manifolds?





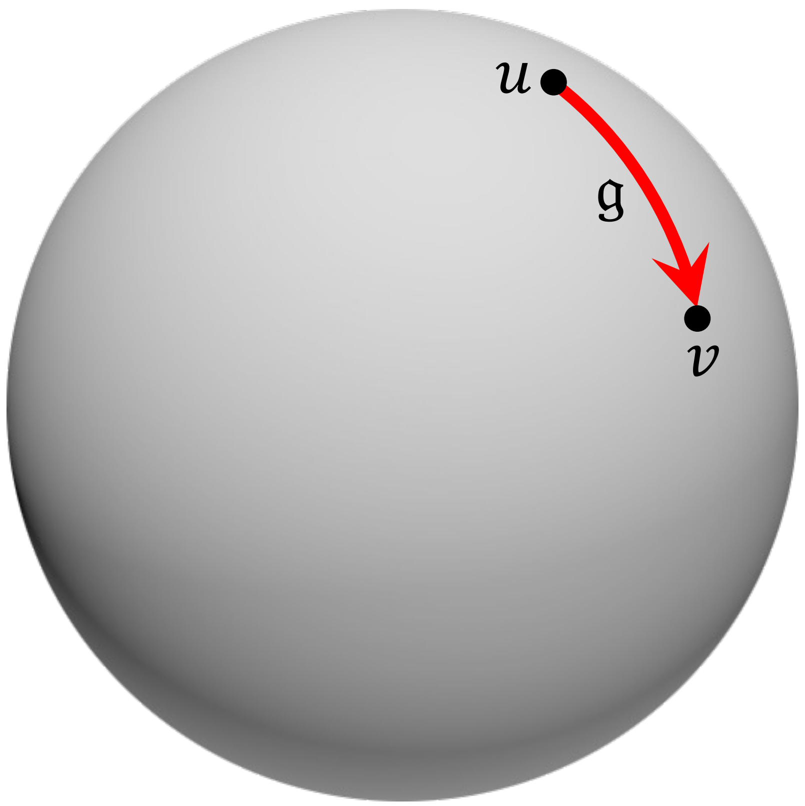
More efficient representation: no "waste" for internal structures

Natural model for deformable shapes

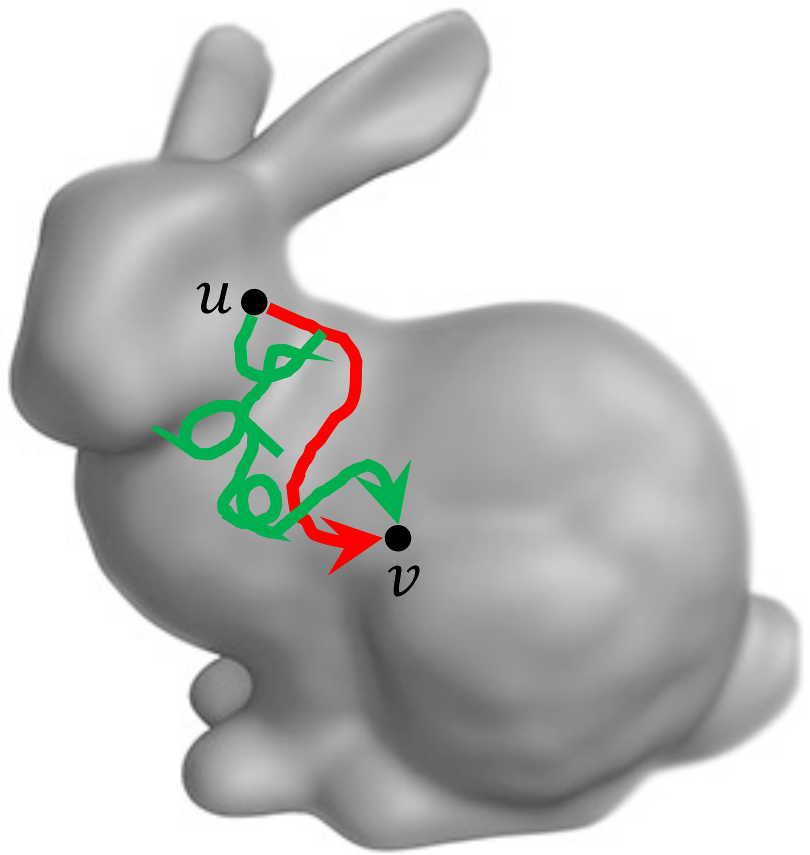
Why Manifolds?

In protein modeling, abstract out internal structure that is irrelevant for interactions + allow some conformation changes

Homogeneous Spaces

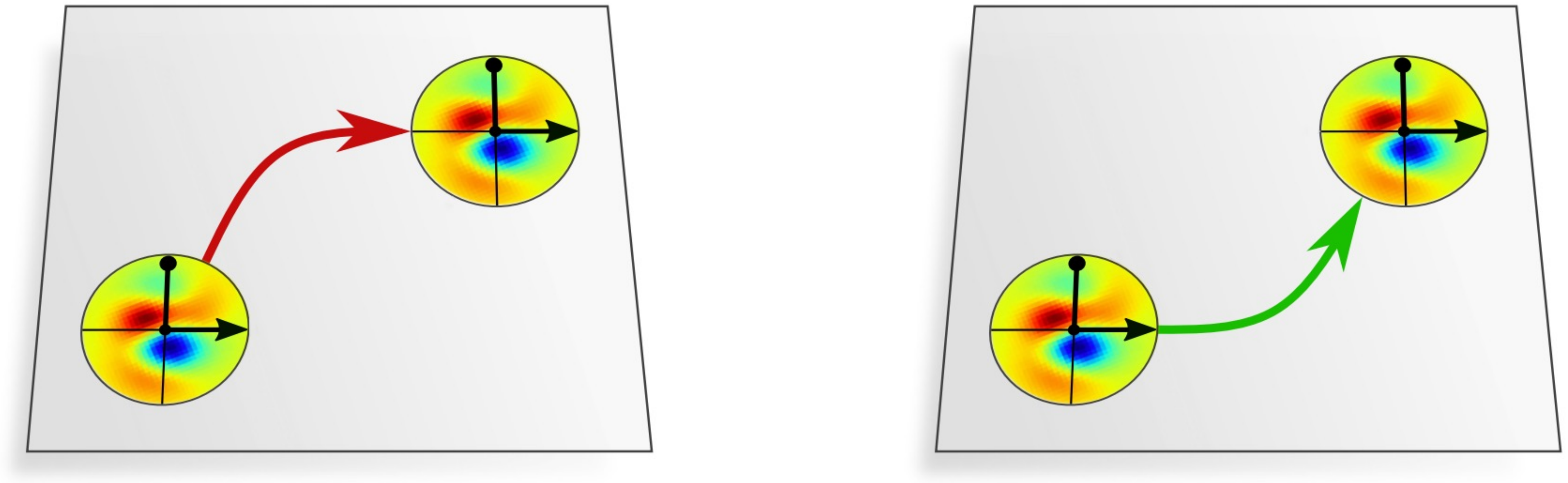


global symmetry group
$\exists \mathfrak{g} \in \mathfrak{G}$ s.t. $\mathfrak{g}u = v$



no useful global
symmetry group

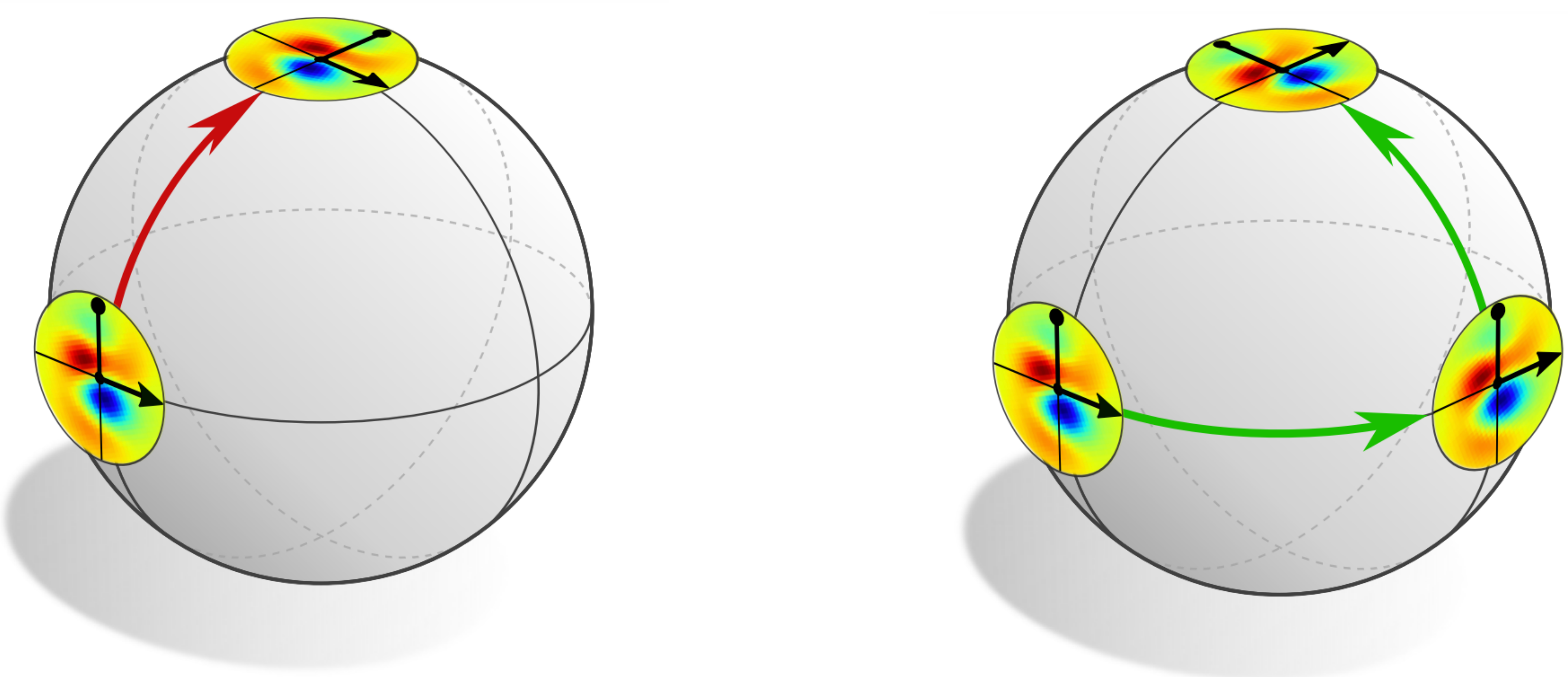
Euclidean Convolution



Euclidean space: Transport the filter around the domain

Figure: Weiler et al. 2021

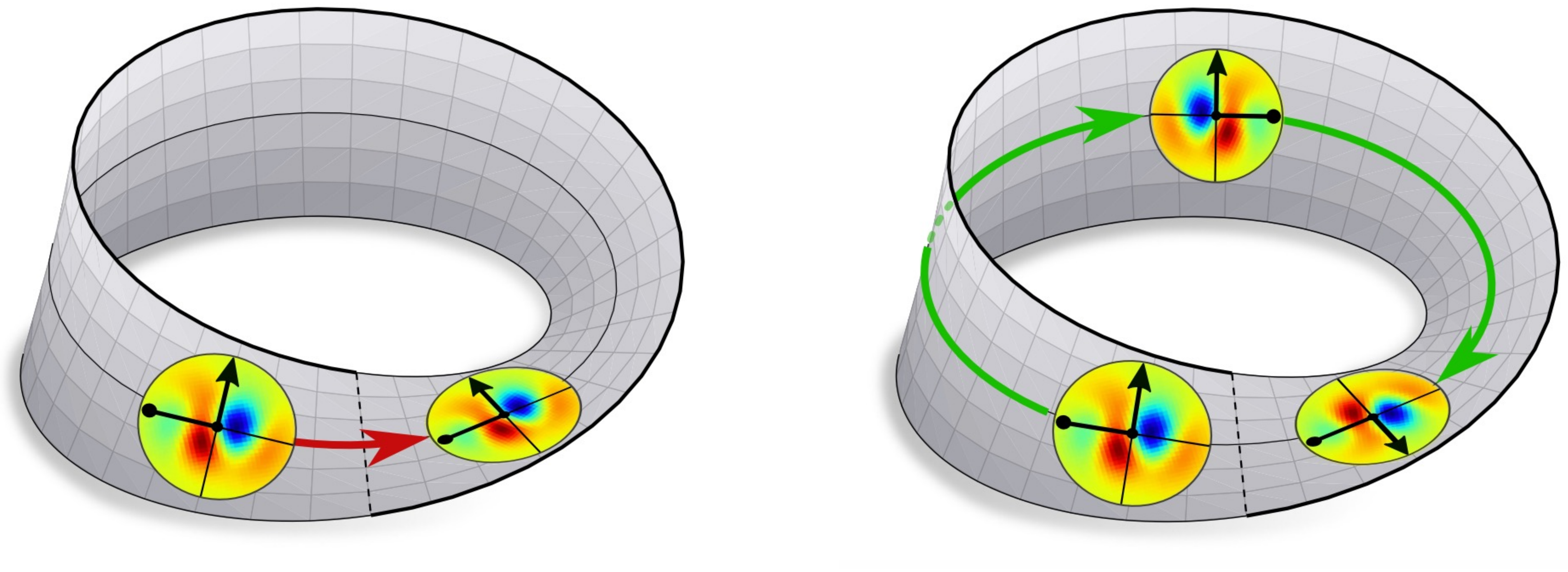
Non-Euclidean Convolution



Manifold: Result of transport is *path dependent*

Figure: Weiler et al. 2021

Non-Euclidean Convolution

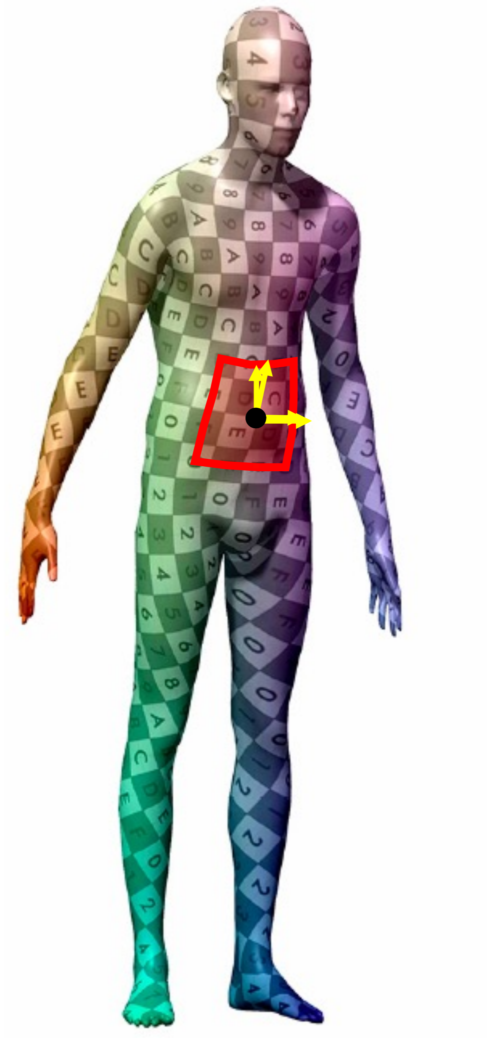


Manifold: Result of transport is *path dependent*

Figure: Weiler et al. 2021

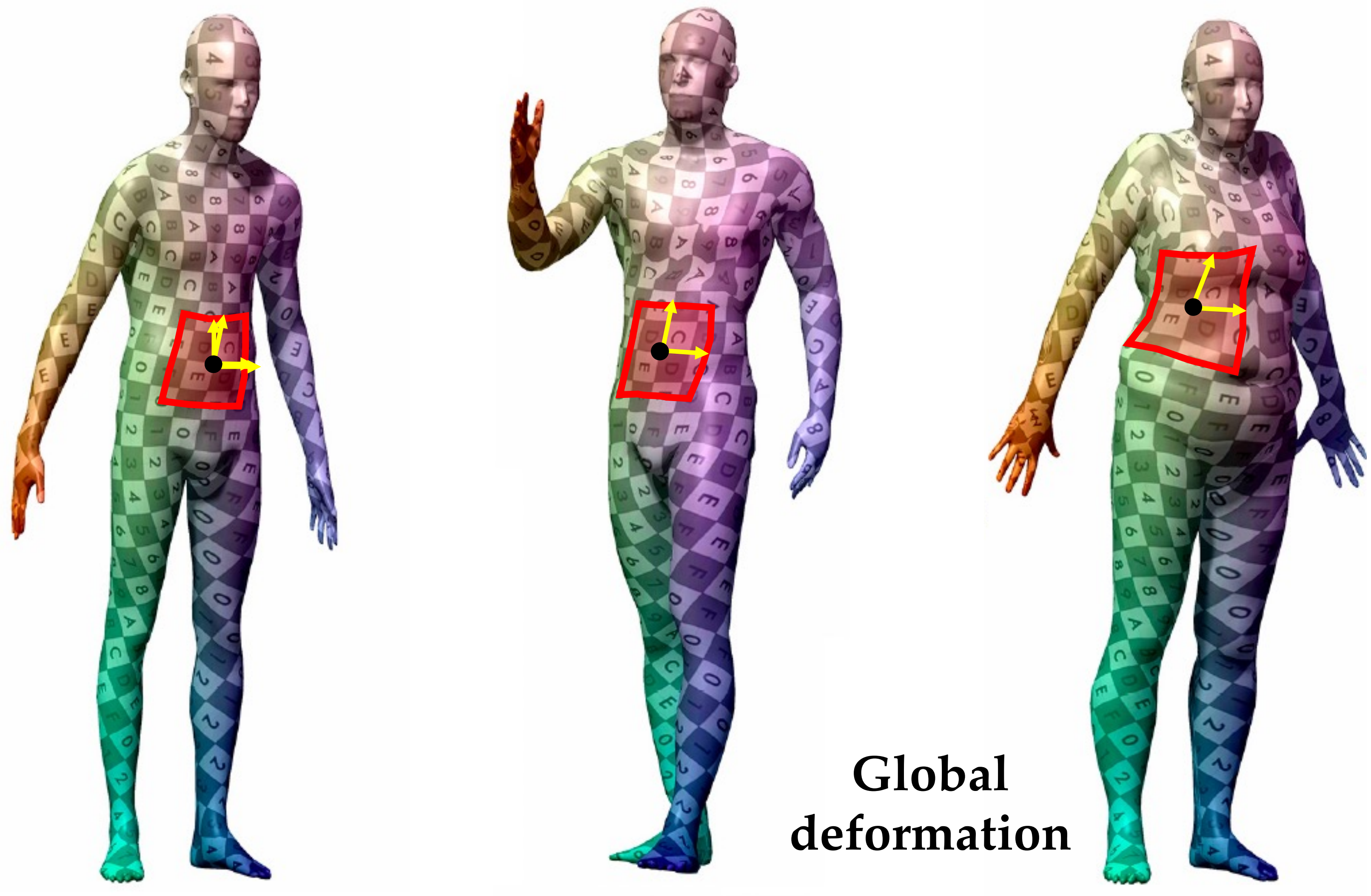
Two Types of Invariance

Local gauge transformation



Two Types of Invariance

Local gauge transformation



Global deformation

Non-Euclidean Convolution Recipes

Fixed path

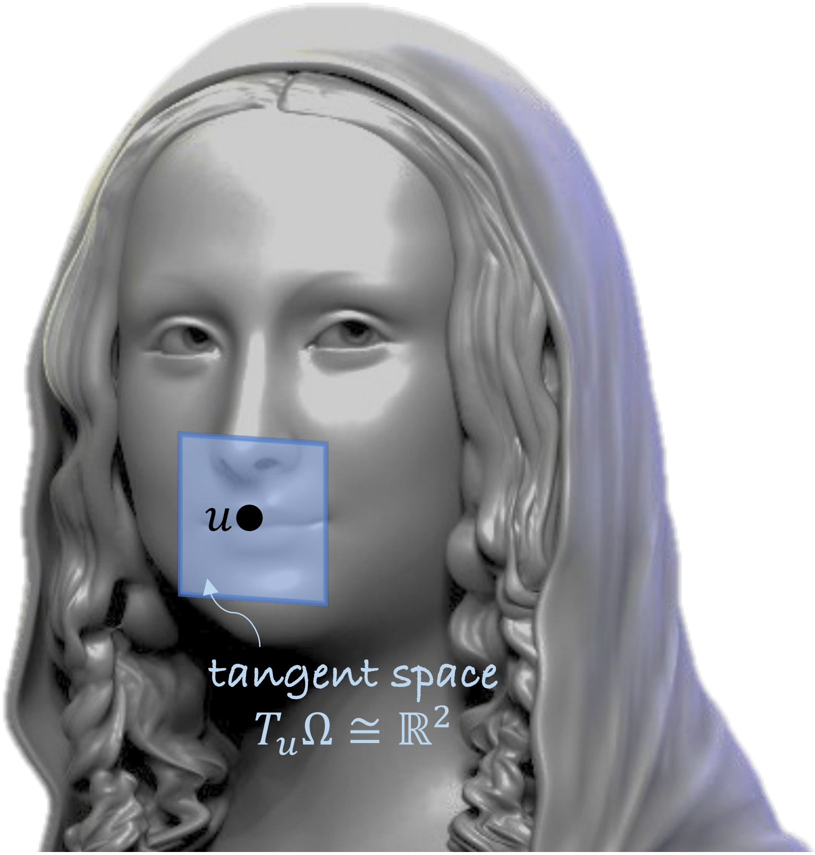
Fixed gauge

"Group pooling"

Isotropic filter (spectral)

Gauge-equivariant filter

Manifolds



manifold Ω

manifold = locally Euclidean space

Riemannian metric = local length/direction

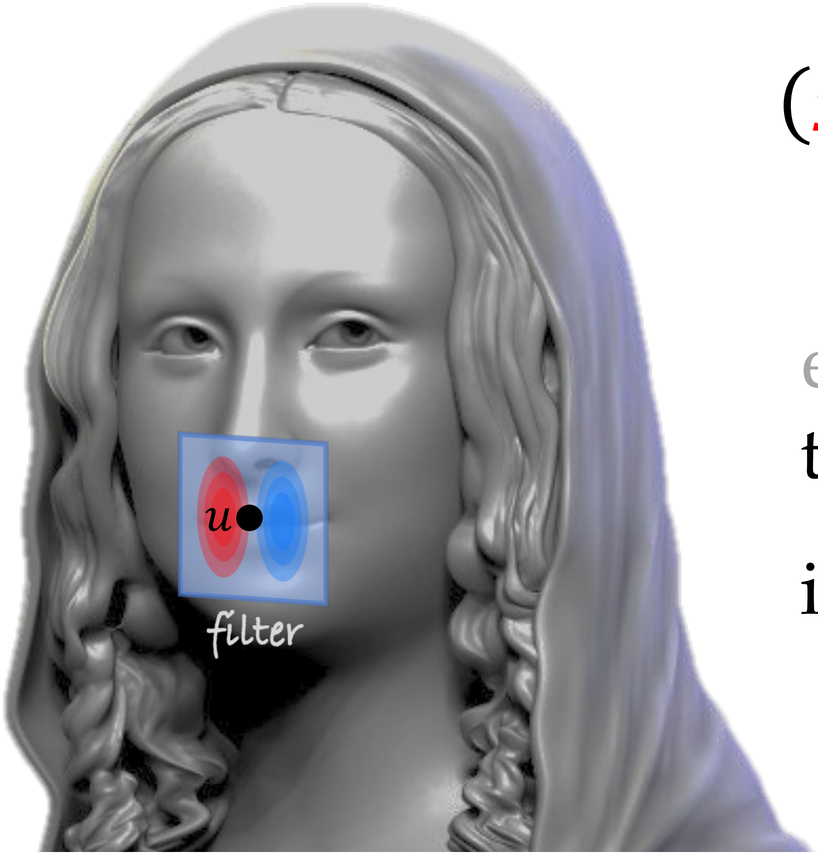
Intrinsic quantity = expressed solely in terms of the Riemannian metric

Isometry = metric-preserving deformation

Geodesic CNNs

Exponential map
$\exp_u : T_u\Omega \to \Omega$

$$(x \star \psi)(u) = \int_{T_u\Omega} \psi(v)\, x(\exp_u v)\, \mathrm{d}v$$



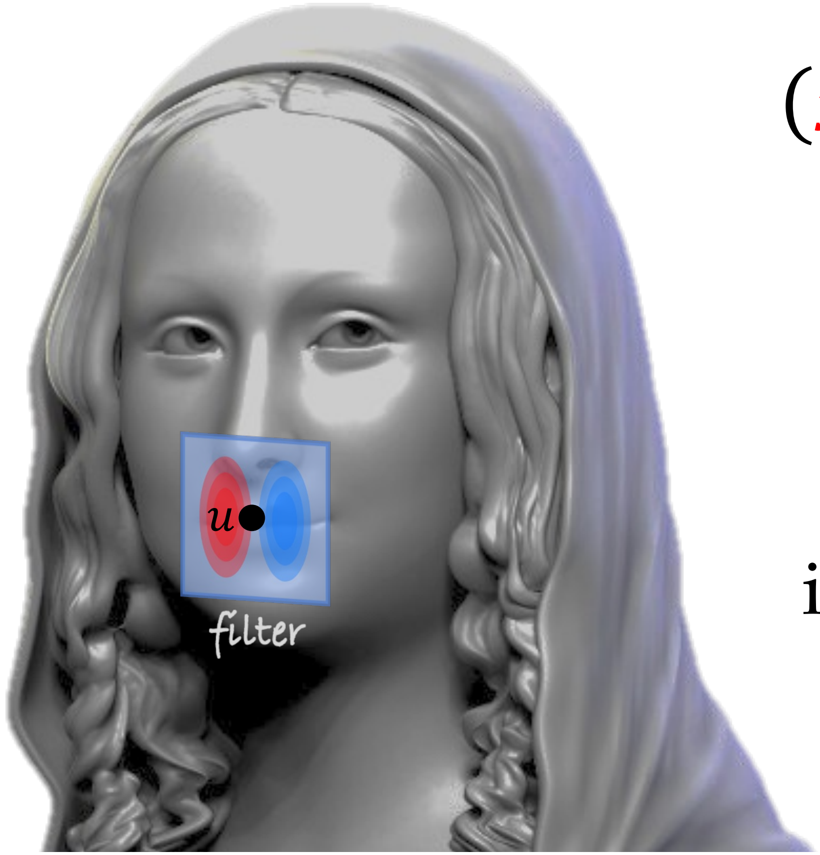
manifold Ω
isometry group $\mathrm{Iso}(\Omega)$

$\exp_u$ is an intrinsic map allowing to express the signal x locally in the tangent space $T_u\Omega$

intrinsic filter = invariant to isometries

Masci et al. 2015; Boscaini et al. 2016; Monti et al. 2017

Geodesic CNNs



manifold Ω
isometry group $\mathrm{Iso}(\Omega)$

$$(x \star \psi)(u) = \int_{T_u\Omega} \psi(v)\,x(\exp_u v)\,\mathrm{d}v$$

Problem: these are abstract vectors!

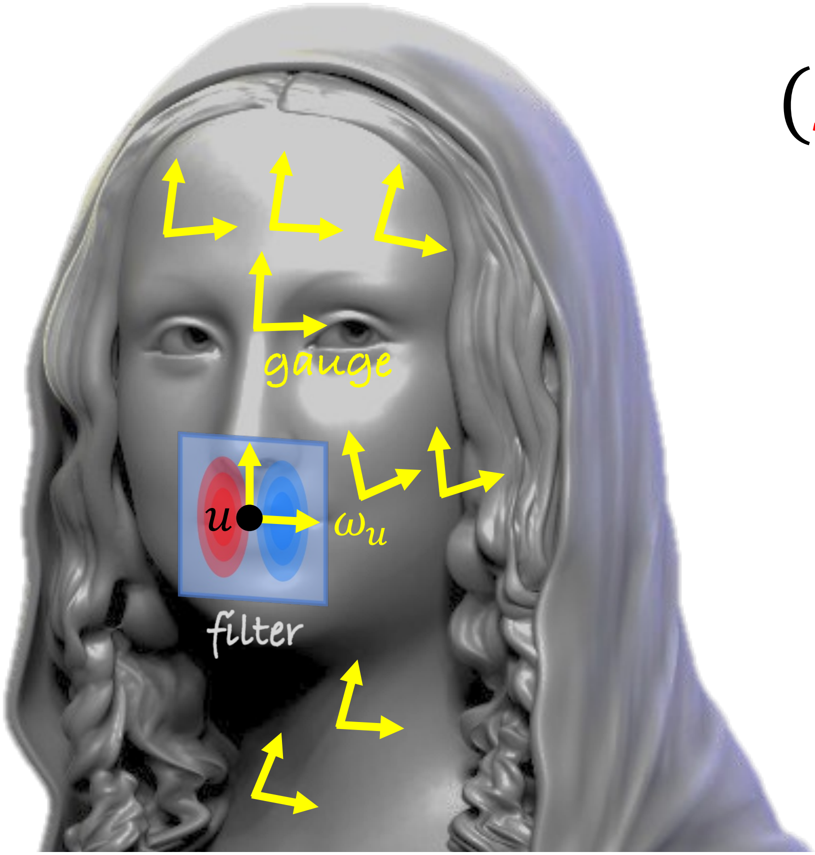
intrinsic filter = invariant to isometries

Masci et al. 2015; Boscaini et al. 2016; Monti et al. 2017

Geodesic CNNs

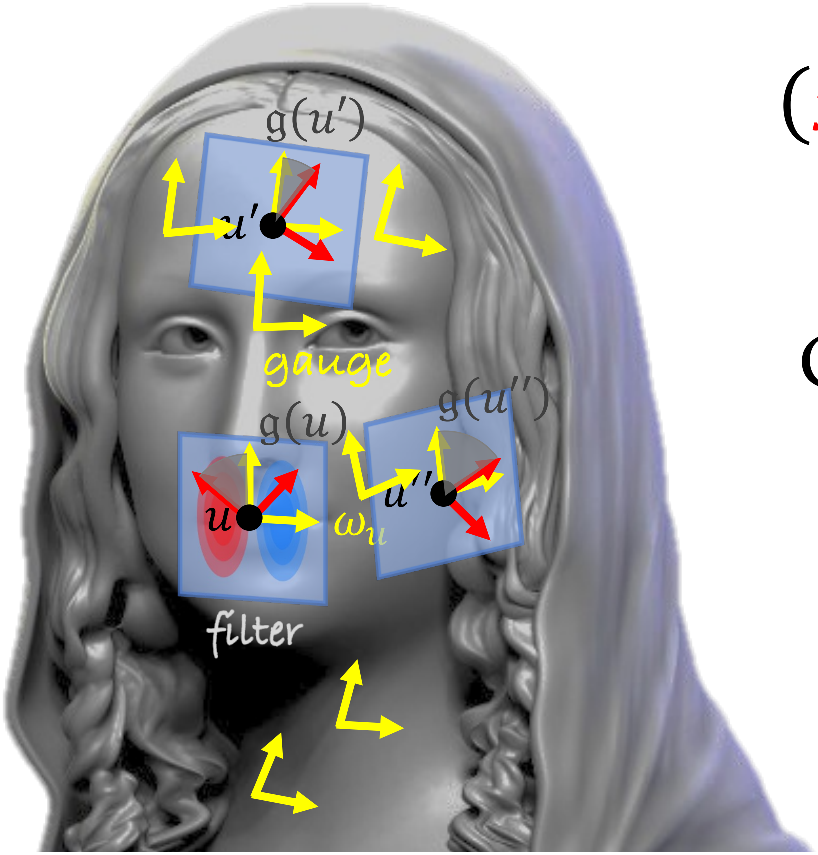
local reference frame
$\omega_u : \mathbb{R}^2 \to T_u\Omega$

$$(x \star \psi)(u) = \int_{\mathbb{R}^2} \psi(\mathbf{v})\, x(\exp_u \omega_u \mathbf{v})\, d\mathbf{v}$$



manifold Ω
isometry group $\mathrm{Iso}(\Omega)$

Gauge Transformations



manifold Ω
structure group 𝔊

$$(x \star \psi)(u) = \int_{\mathbb{R}^2} \psi(\mathbf{v})\, x(\exp_u \omega_u \mathbf{v})\, d\mathbf{v}$$

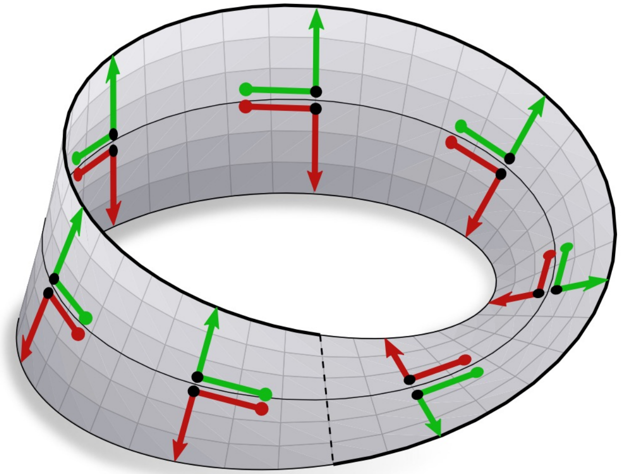
Gauge defined up to gauge transformation

$$\mathfrak{g}: \Omega \rightarrow \mathfrak{G}$$

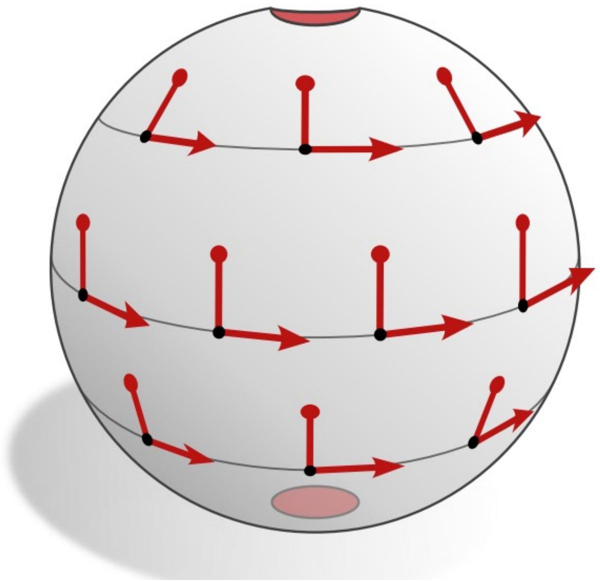
Cohen et al. 2019; Weiler et al. 2021

Structure Group

rotation
SO(2)



reflection
R



fixed gauge
{id}

Weiler et al. 2021

Structure Group

A gauge is defined up to a *gauge transformation* $\mathfrak{g}: \Omega \to \mathfrak{G}$

• "Naked" manifold	$\mathrm{GL}(s)$	invertible matrices
• Manifold+orientation	$\mathrm{GL}^{+}(s)$	invertible matrices with det>0
• Manifold+volume	$\mathrm{SL}(s)$	matrices with det=1
• Manifold+metric	$\mathrm{O}(s)$	orthogonal matrices
• Manifold+metric+orientation	$\mathrm{SO}(s)$	orthogonal matrices with det=1
• Manifold+frame field	$\{\mathrm{id}\}$	identity (no ambiguity)

Weiler et al. 2021

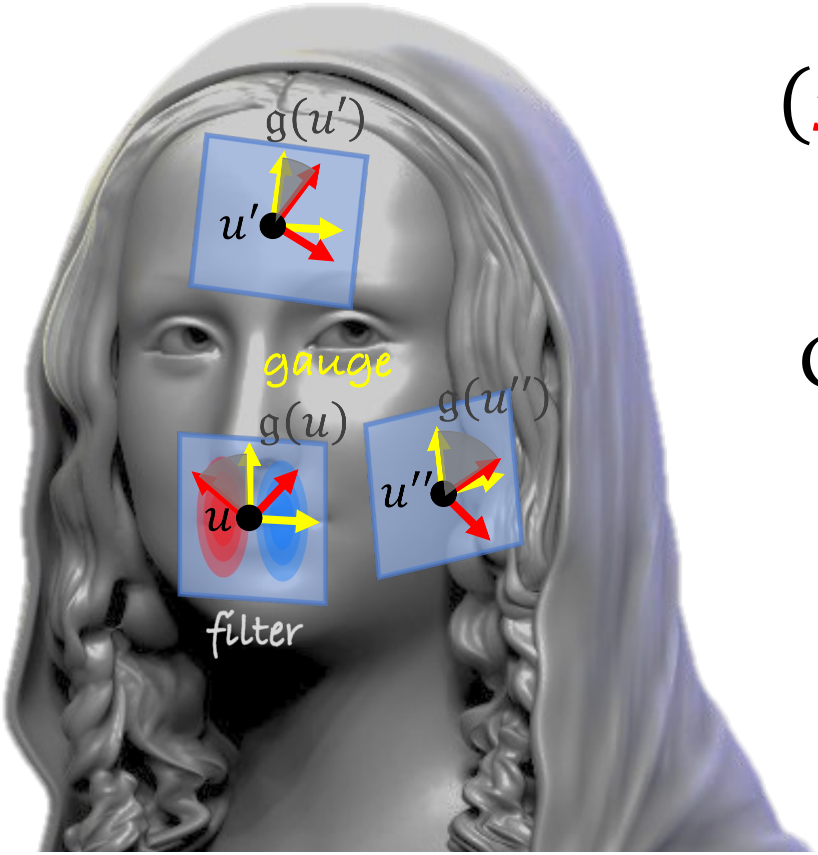
Structure Group

A gauge is defined up to a *gauge transformation* $\mathfrak{g}: \Omega \to \mathfrak{G}$

• "Naked" manifold	$GL(s)$	invertible matrices
• Manifold+orientation	$GL^+(s)$	invertible matrices with det>0
• Manifold+volume	$SL(s)$	matrices with det=1
• Manifold+metric	$O(s)$	orthogonal matrices
• Manifold+metric+orientation	$SO(s)$	orthogonal matrices with det=1
• Manifold+frame field	$\{id\}$	identity (no ambiguity)

Weiler et al. 2021

Gauge-equivariant CNNs



manifold Ω

structure group $\mathfrak{G} = \mathrm{SO}(2)$

$$(x \star \psi)(u) = \int_{\mathbb{R}^2} \psi(\mathbf{v})\, x(\exp_u \omega_u \mathbf{v})\, d\mathbf{v}$$

Gauge defined up to gauge transformation

$$\mathfrak{g}: \Omega \rightarrow \mathrm{SO}(2)$$

gauge-equivariant filter

$$\psi(\mathfrak{g}^{-1}\mathbf{v}) = \rho(\mathfrak{g}^{-1})\psi(\mathbf{v})\rho(\mathfrak{g})$$

Cohen et al. 2019

"Hairy Ball" (a.k.a. Poincaré-Hopf) Theorem

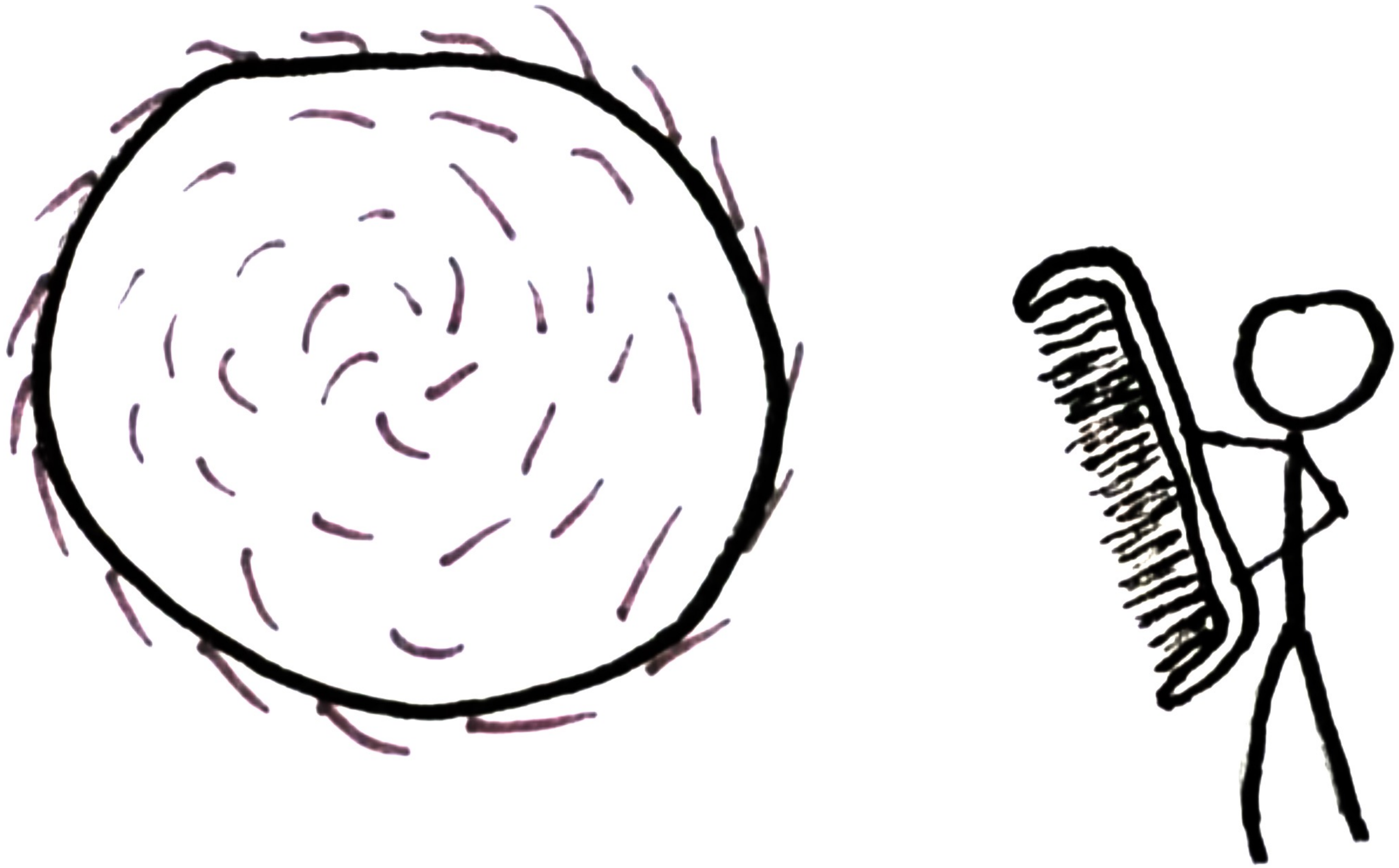
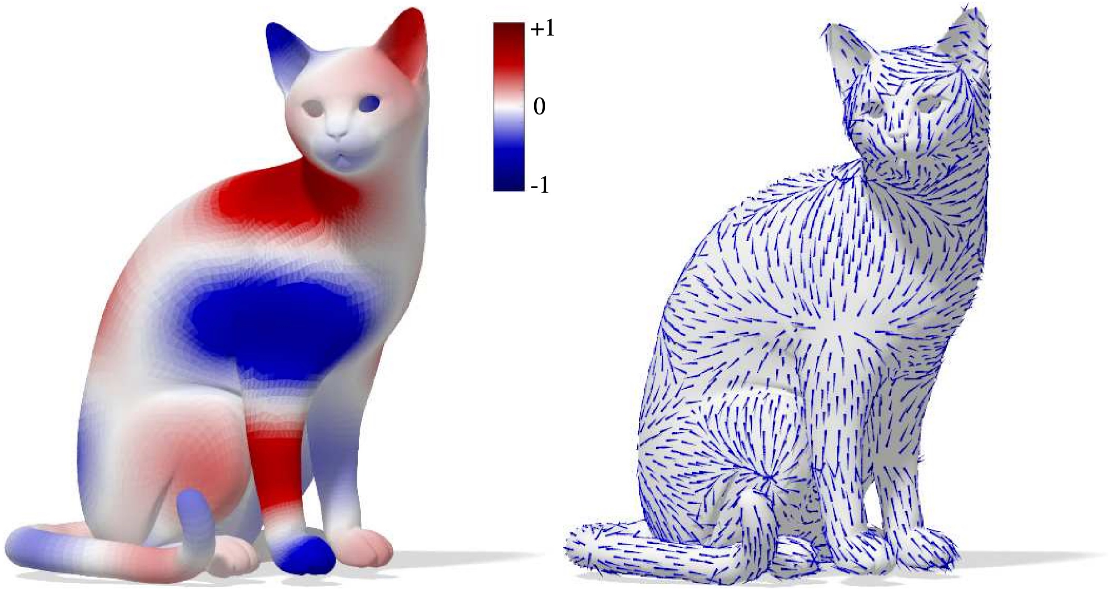
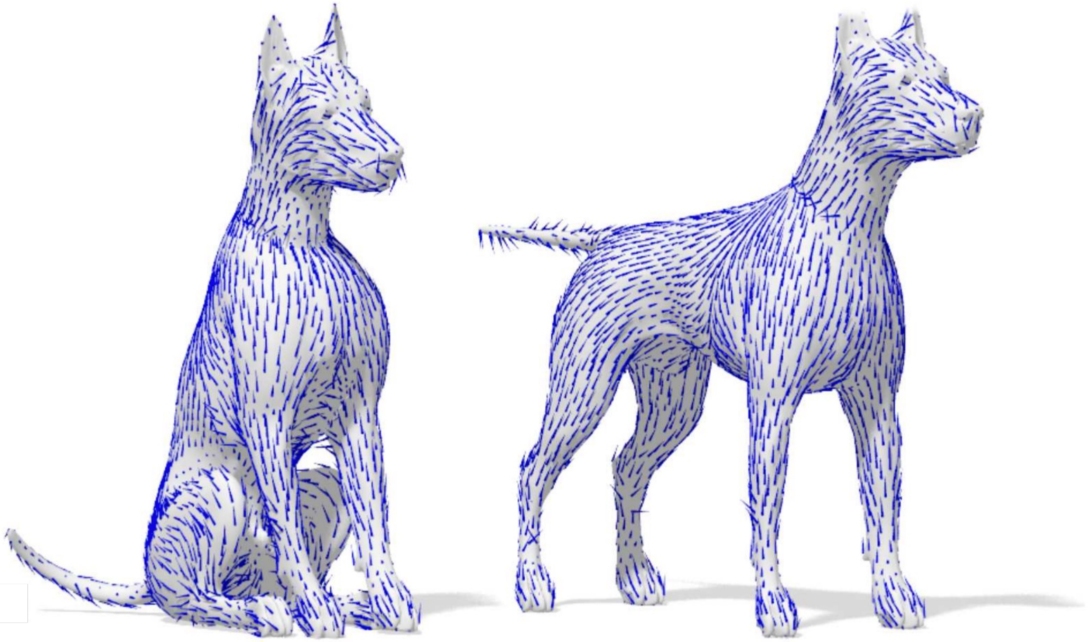


Image: Minutephysics

Theory vs Practice: Stable Gauges



Gradient of intrinsic function



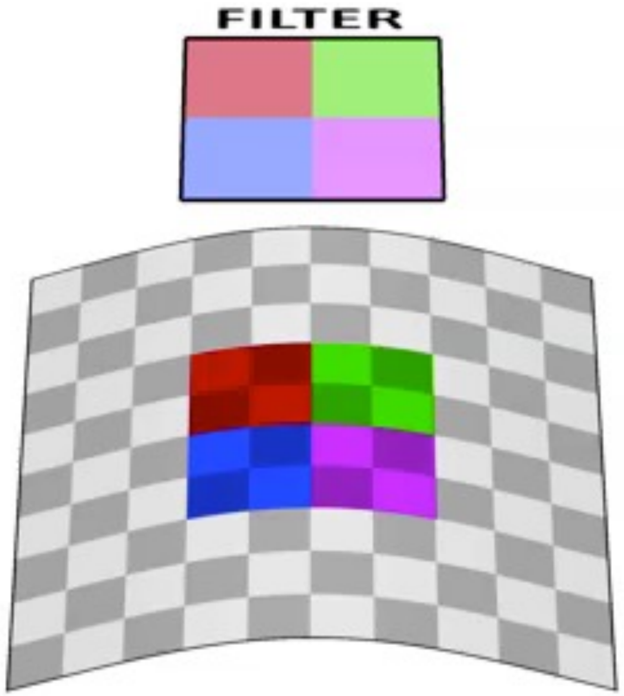
Deformation-invariant stable gauge

Melzi et al. 2019; Monti et al. 2017

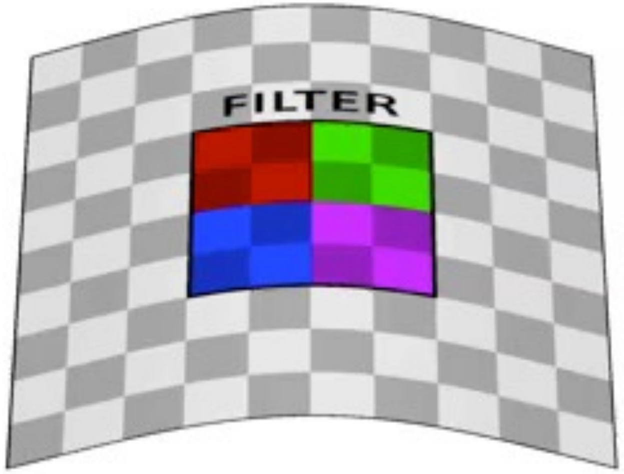
Geodesic CNNs

Anisotropic intrinsic filters on a manifold

Masci et al. 2015; Boscaini et al. 2016; Monti et al. 2017



Euclidean (extrinsic) convolution

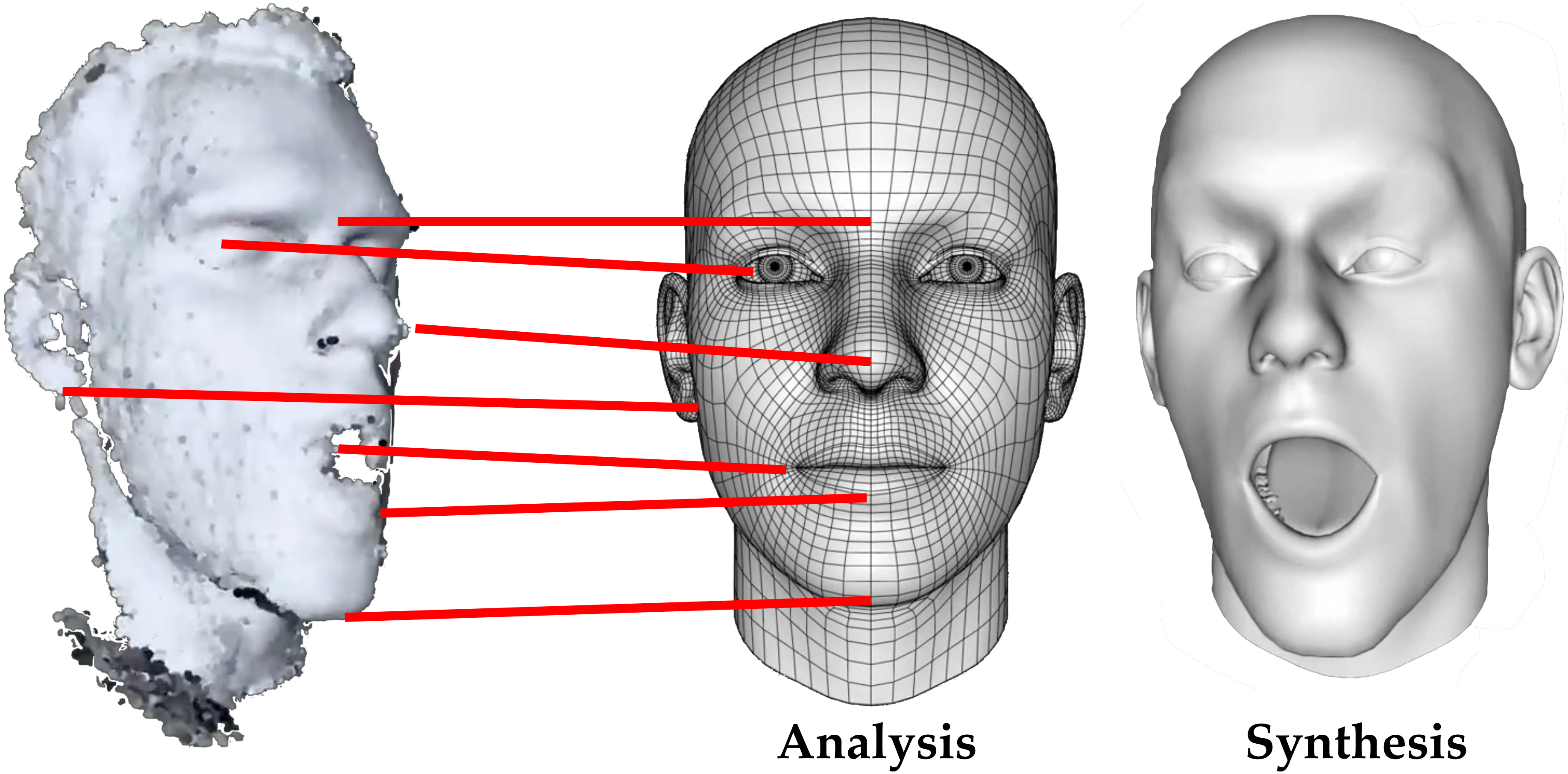


Geometric (intrinsic) convolution

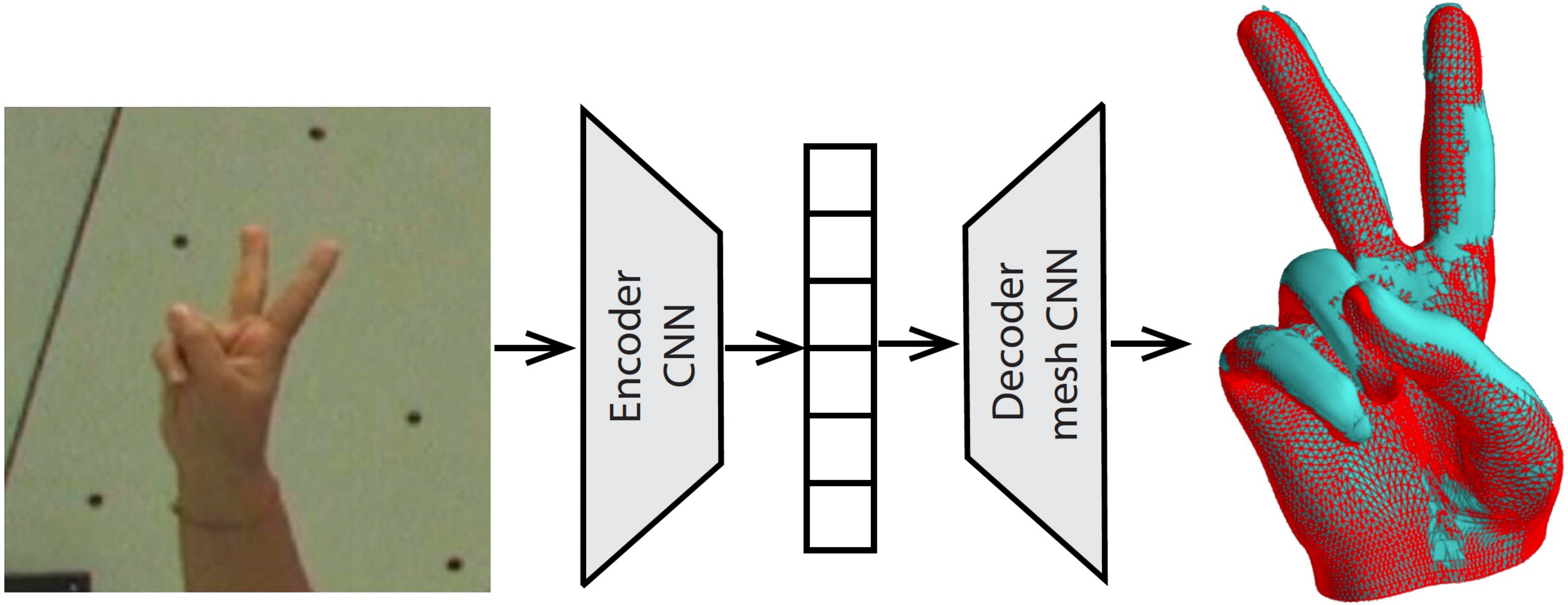
Masci et al. 2015

FaceShift 2015

Shape Analysis & Synthesis



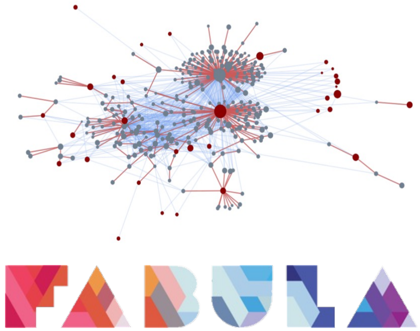
3D Hand Reconstruction



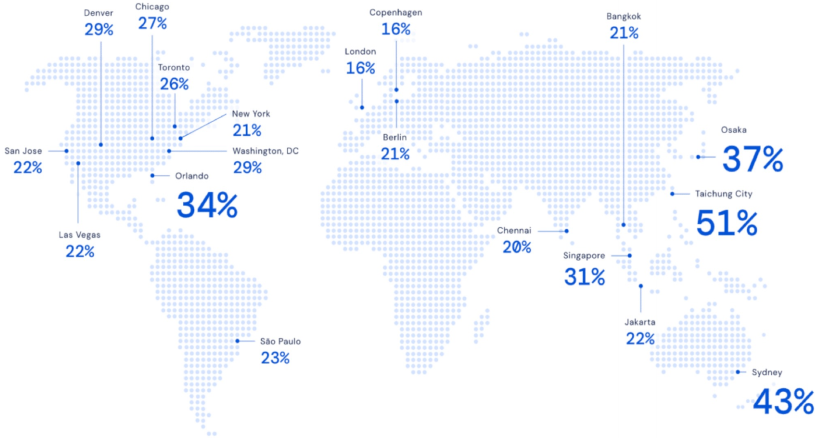
Kulon et B 2020

Kulon et B 2020

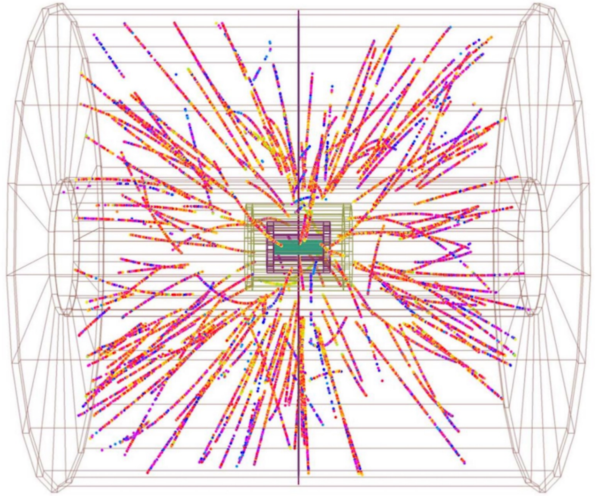
Snap Acquires Ariel AI To Enhance AR Features



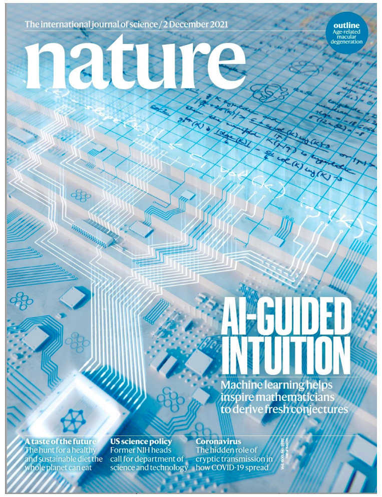
Fake news detection



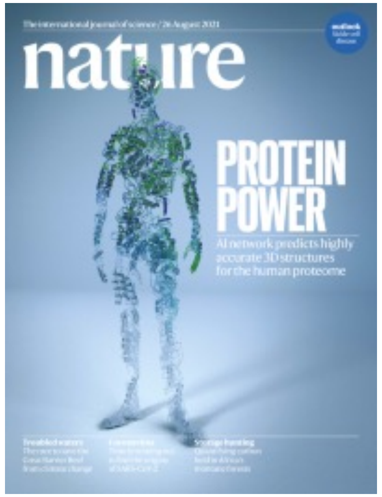
Navigation

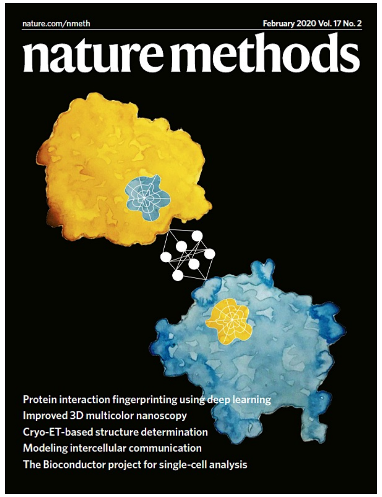


Particle physics



Pure math



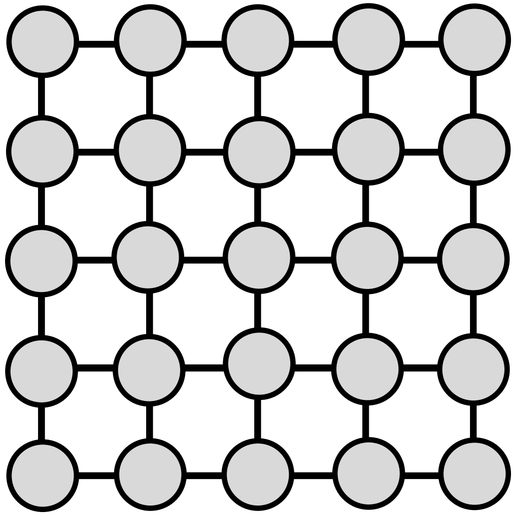


Structural biology

Drug discovery

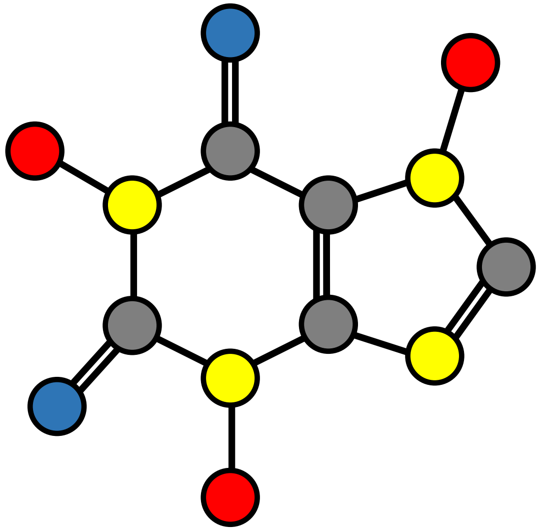
WRAP UP

The "5G" of Geometric Deep Learning



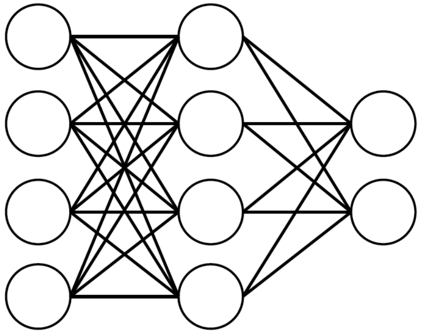
Grids

Groups

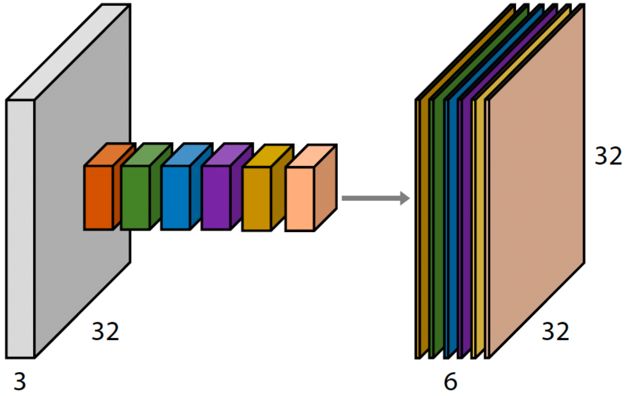


Graphs

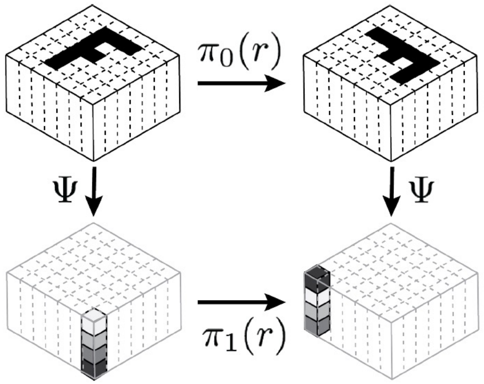
Geodesics & Gauges



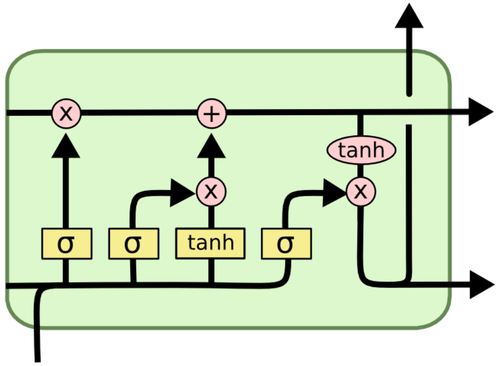
Perceptrons
Function regularity



CNNs
Translation

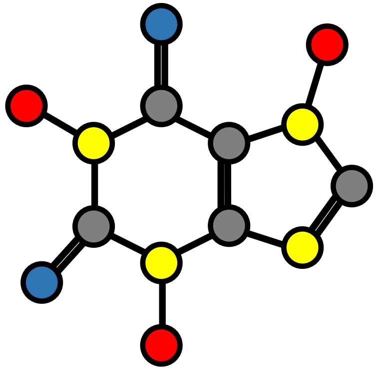


Group-CNNs
Translation+Rotation,
Global groups



LSTMs
Time warping

DeepSets / Transformers
Permutation



GNNs
Permutation

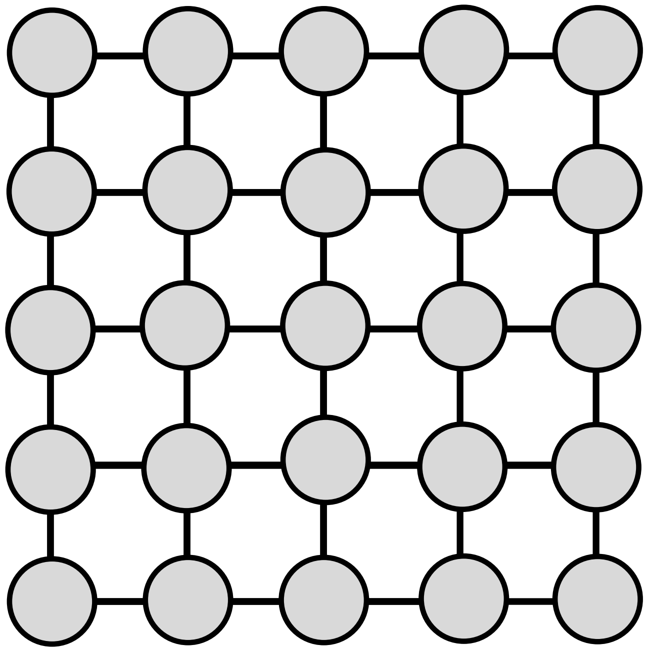
Intrinsic CNNs
Isometry / Gauge choice

“The knowledge of certain principles easily compensates the lack of knowledge of certain facts”

—Claude Adrien Helvétius

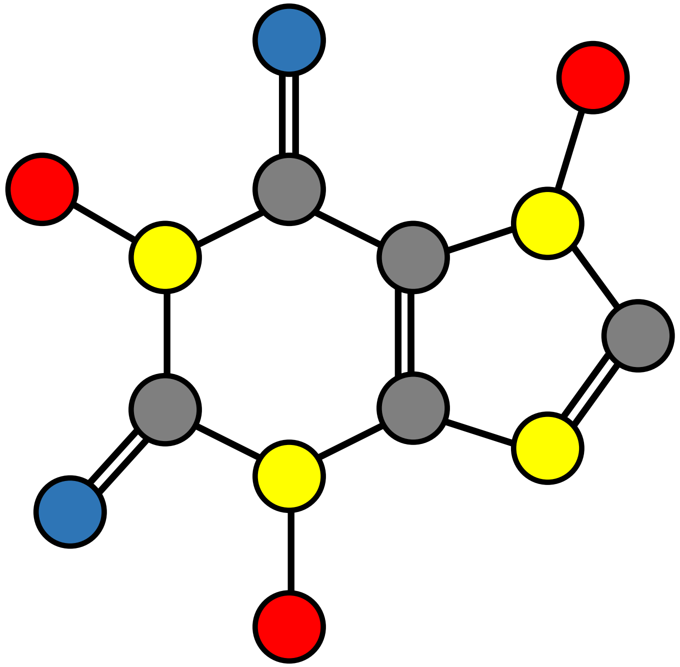
PHYSICS-INSPIRED GDL

Instances of GDL Blueprint: Different domains



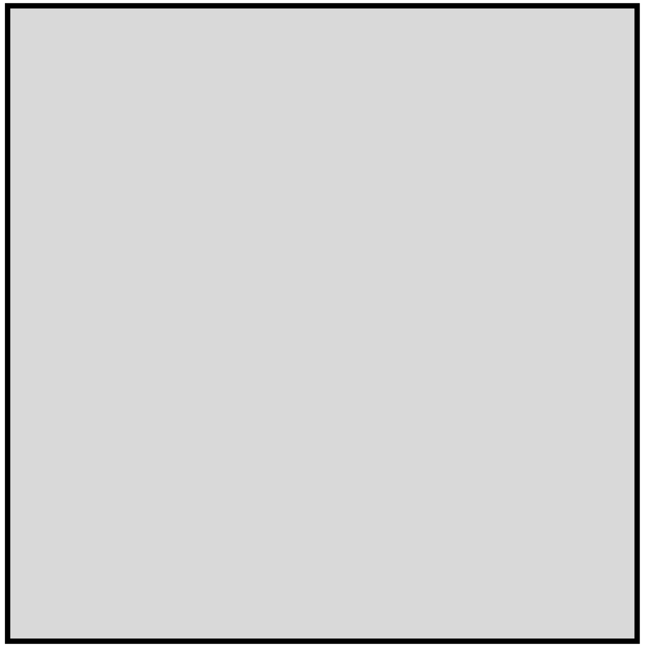
Grid

Mesh



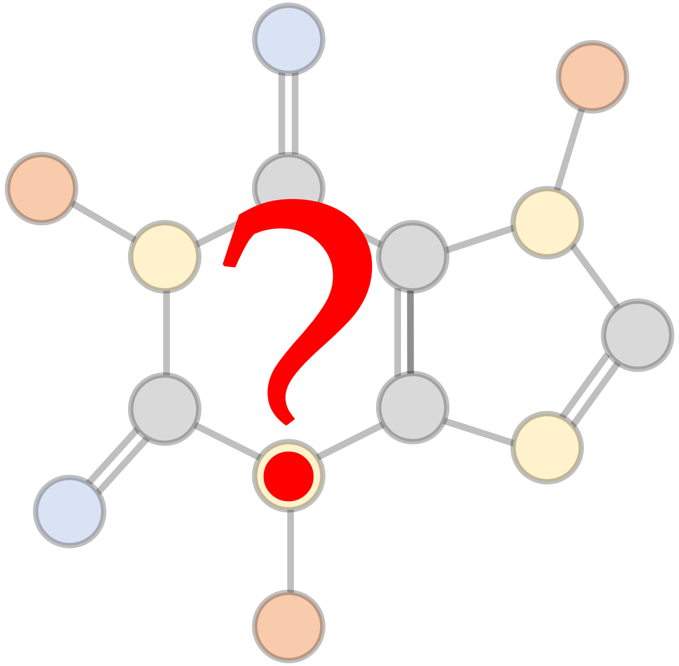
Graph

Instances of GDL Blueprint: Different domains



Plane (Homogeneous space)

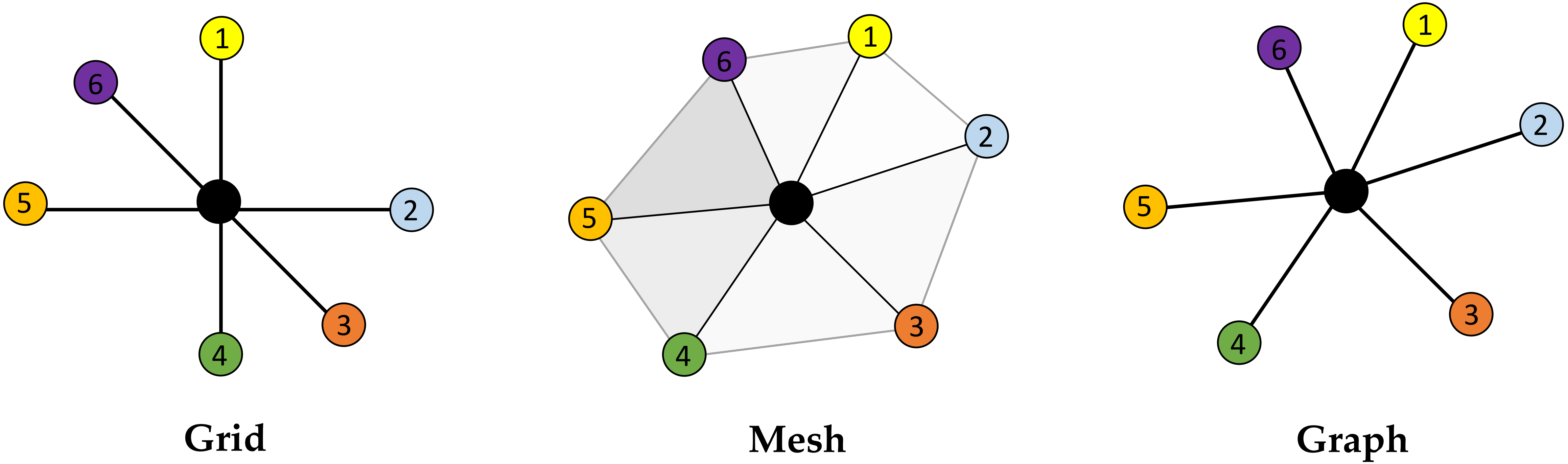
Manifold



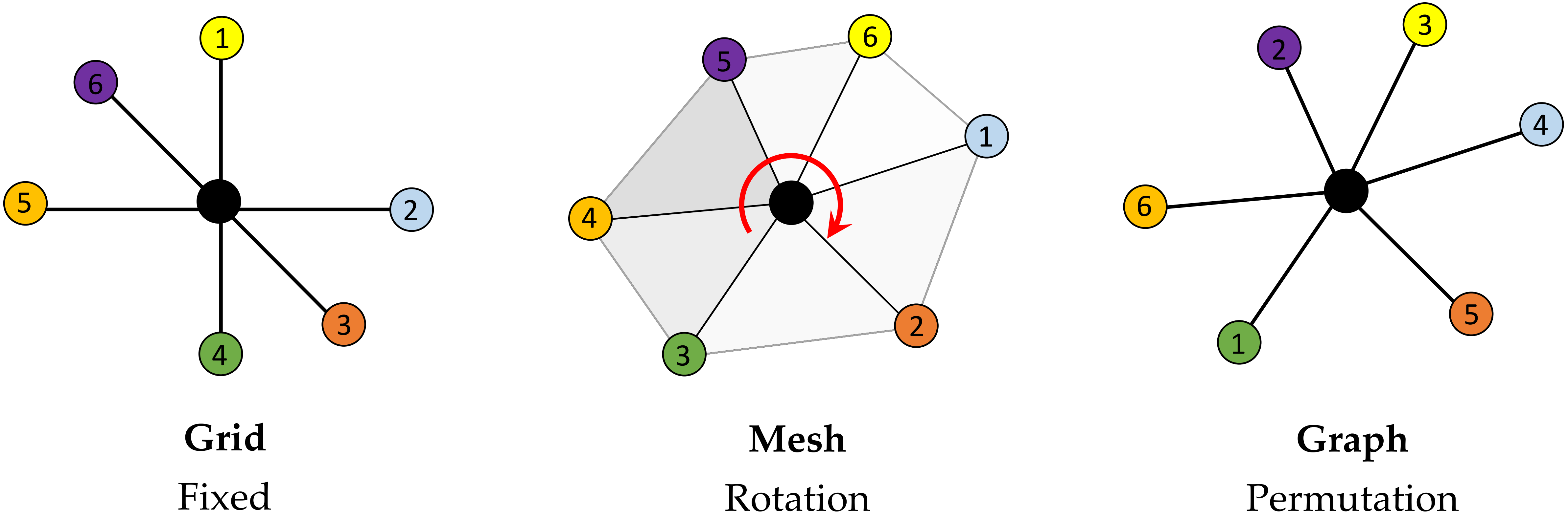
Graph

1. Continuous model for graphs?

Graphs vs Meshes vs Grids



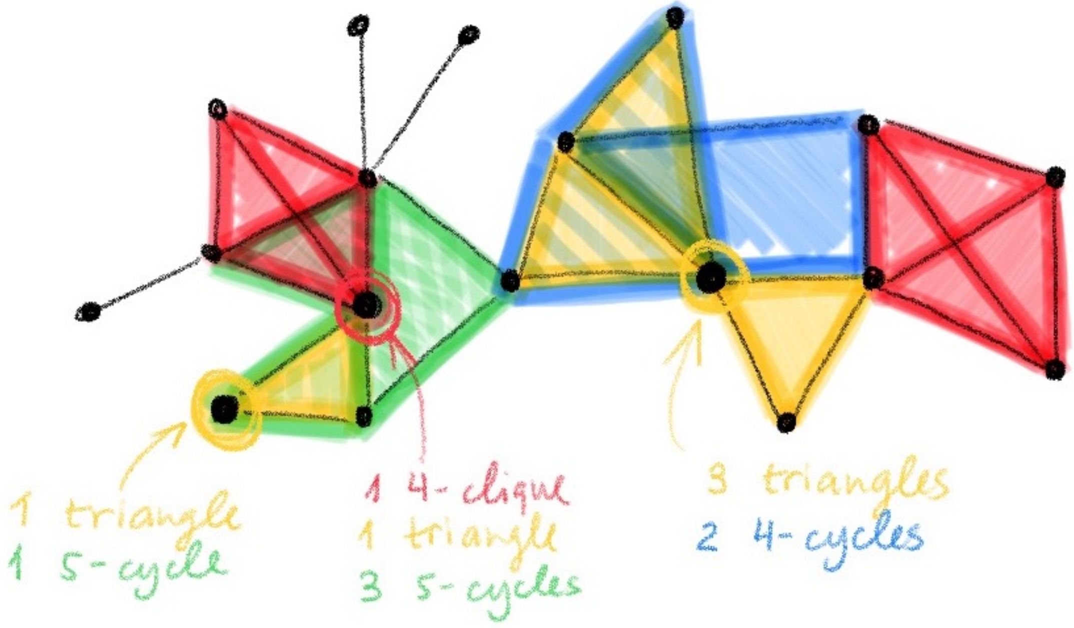
Graphs vs Meshes vs Grids



Graphs have the least structure

Positional Encoding Approaches

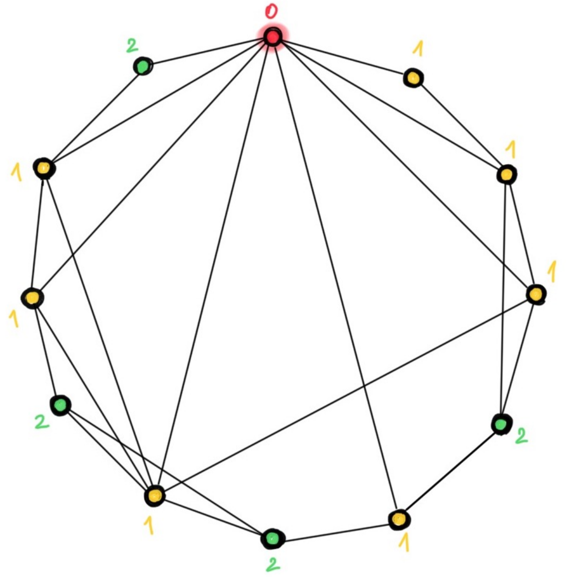
- Random node features[1]
- Graph Laplacian eigenvectors[2]
- Graph substructure counts[3]
- Bags of subgraphs[4]

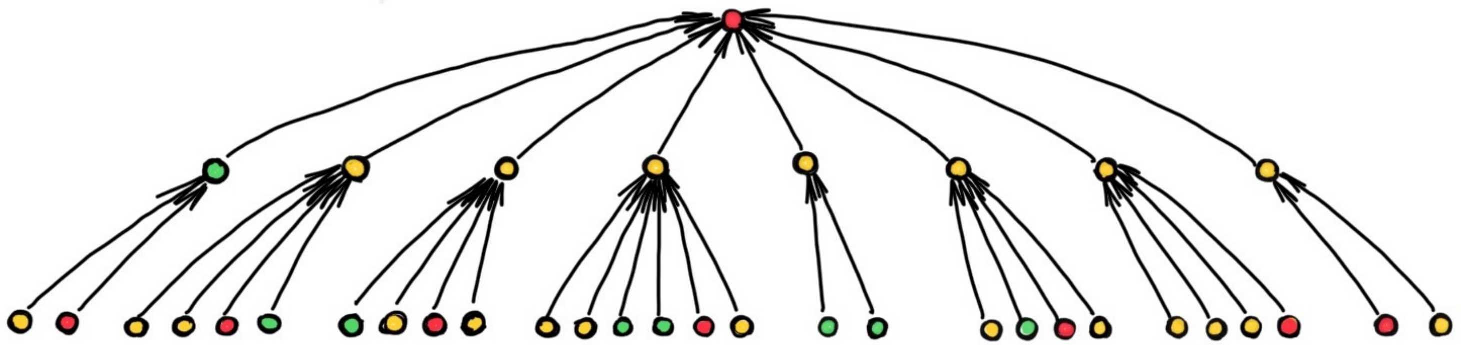


2. How to choose positional encoding?

[1]Sato et al. 2020; [2]Vaswani et al. 2017; Qiu et al. 2020; Dwivedi et al. 2020; [3]Bouritsas, Frasca, et B. 2020; [4]Bevilacqua, Frasca, Lim, et B., Maron 2021

Over-squashing & Bottlenecks





In small-world graphs metric ball volume $\mathrm{vol}(B_k) = \sum_{j \in B_k} d_j$ grows exponentially with ball radius k

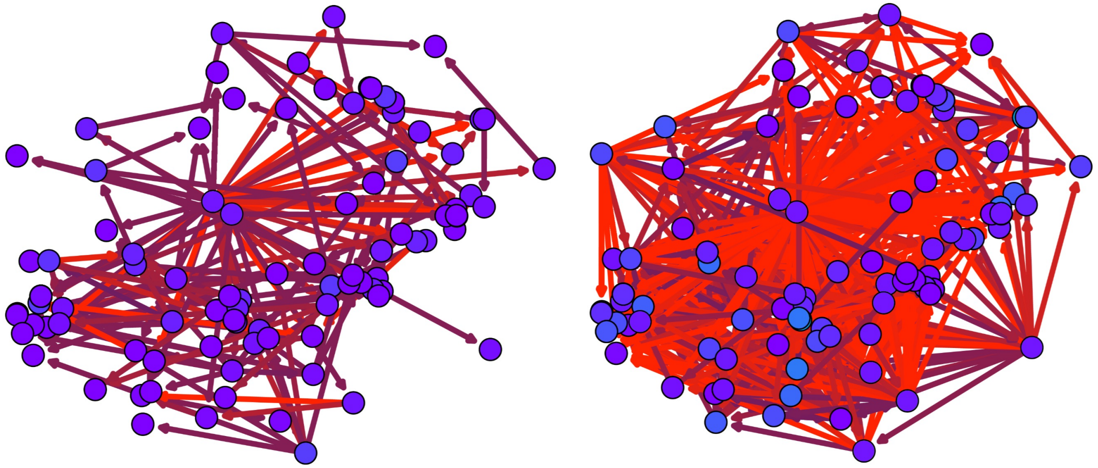
Long-distance dependency + Fast volume growth
= Over-squashing

Alon, Yahav 2020

Graph Rewiring

Decouple **input graph** from **information propagation graph** (at the expense of link to WL)

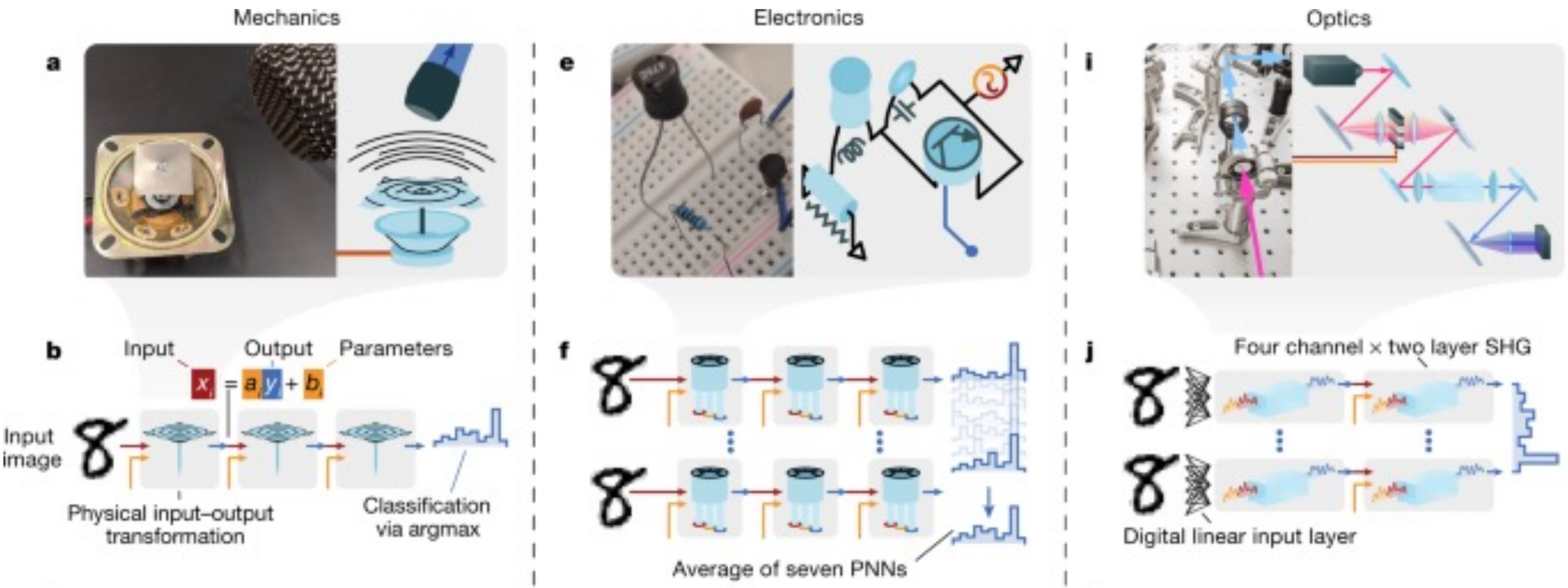
- Neighbourhood sampling (GraphSAGE)[1]
- Multi-hop filters (SIGN)[2]
- Complete graph[3]
- Topology diffusion (DIGL)[4]
- Learnable graph (Dynamic Graph CNN)[5]



3. How to rewire the graph?

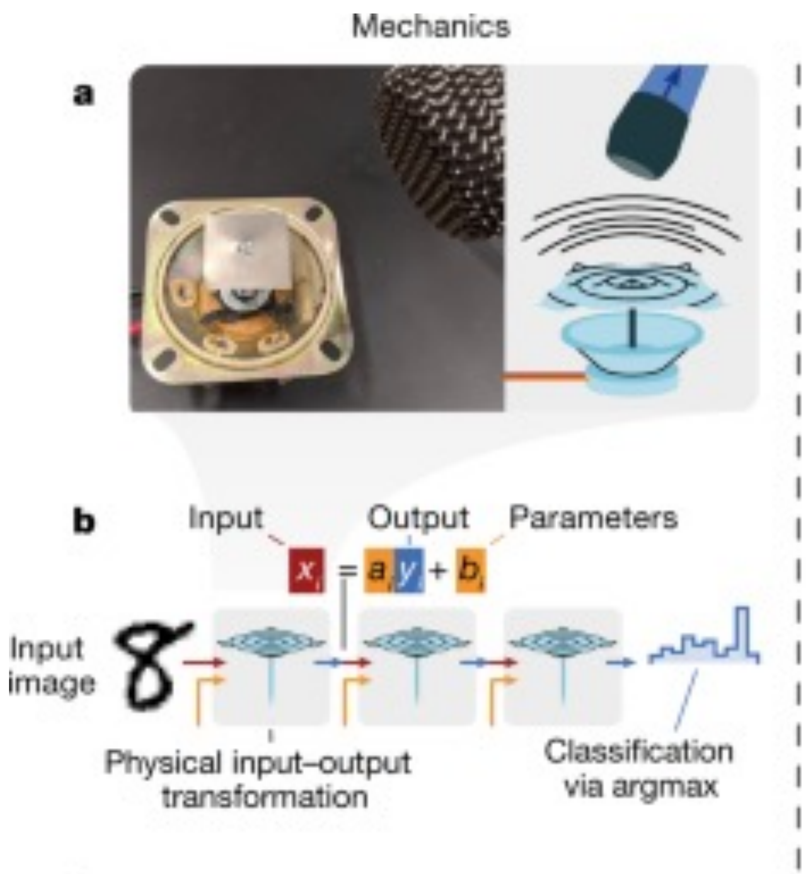
[1]Hamilton et al. 2017; [2]Rossi, Frasca, et B. 2020; [3]Alon, Yahav 2020; [4]Klicpera et al. 2019; [5]Wang et B 2018; Kazi, Cosmo, et B. 2020

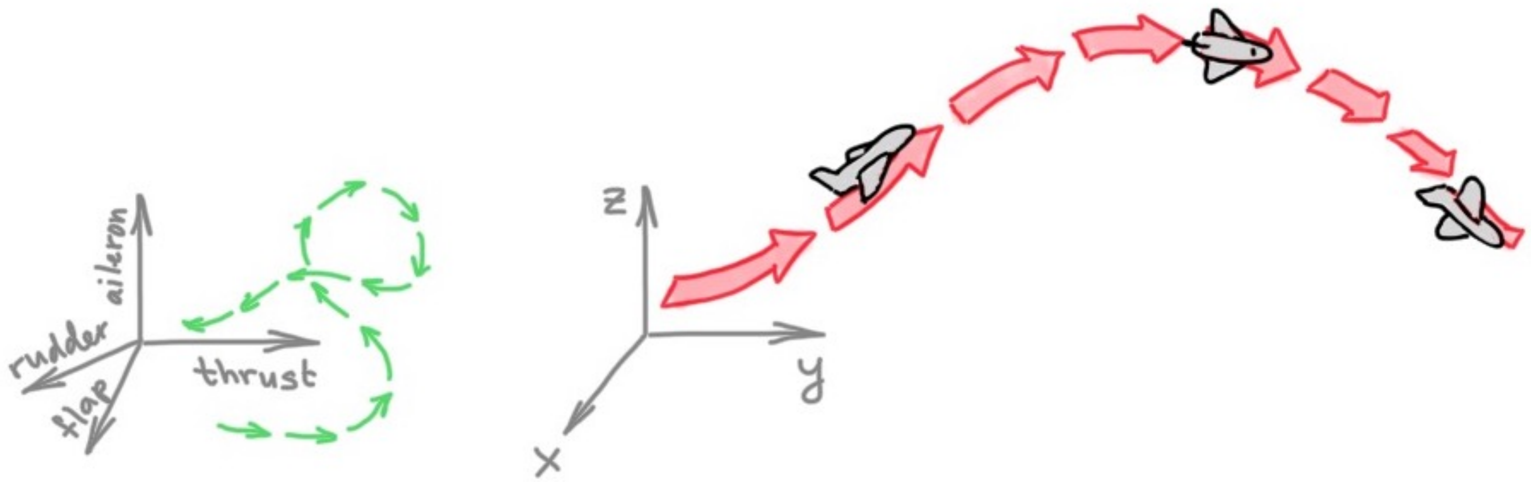
Physical systems as learning metaphor



Wright et al. 2022

Physical systems as learning metaphor





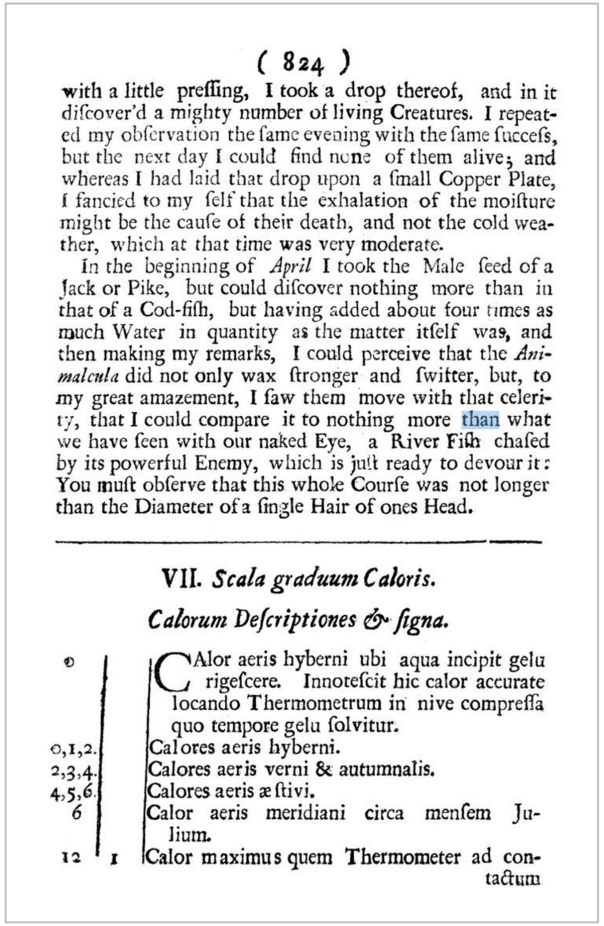
parameter space

system state space

Wright et al. 2022

Diffusion

Newton Law of Cooling: "the [temperature] a hot body loses in a given time is proportional to the temperature difference between the object and the environment"



(824)

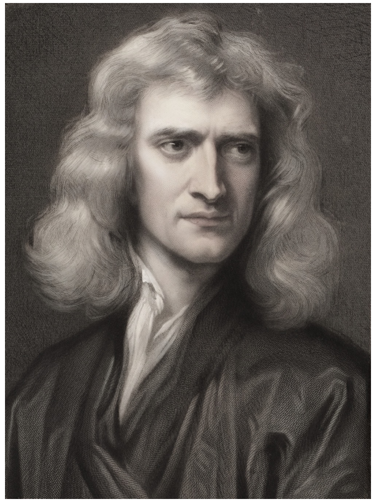
with a little preſſing, I took a drop thereof, and in it diſcover'd a mighty number of living Creatures. I repeated my obſervation the ſame evening with the ſame ſucceſs, but the next day I could find none of them alive; and whereas I had laid that drop upon a ſmall Copper Plate, I fancied to my ſelf that the exhalation of the moiſture might be the cauſe of their death, and not the cold weather, which at that time was very moderate.

In the beginning of *April* I took the Male ſeed of a Jack or Pike, but could diſcover nothing more than in that of a Cod-fiſh, but having added about four times as much Water in quantity as the matter itſelf was, and then making my remarks, I could perceive that the *Animalcula* did not only wax ſtronger and ſwifter, but, to my great amazement, I ſaw them move with that celerity, that I could compare it to nothing more than what we have ſeen with our naked Eye, a River Fiſh chaſed by its powerful Enemy, which is juſt ready to devour it: You muſt obſerve that this whole Courſe was not longer than the Diameter of a ſingle Hair of ones Head.

VII. *Scala graduum Caloris.*

Calorum Deſcriptiones & ſigna.

0 | Calor aeris hyberni ubi aqua incipit gelu rigeſcere. Innoteſcit hic calor accurate locando Thermometrum in nive compreſſa quo tempore gelu ſolvitur.
0,1,2. | Calores aeris hyberni.
2,3,4. | Calores aeris verni & autumnalis.
4,5,6. | Calores aeris æſtivi.
6 | Calor aeris meridiani circa menſem Julium.
12 | 1 | Calor maximus quem Thermometer ad contactum



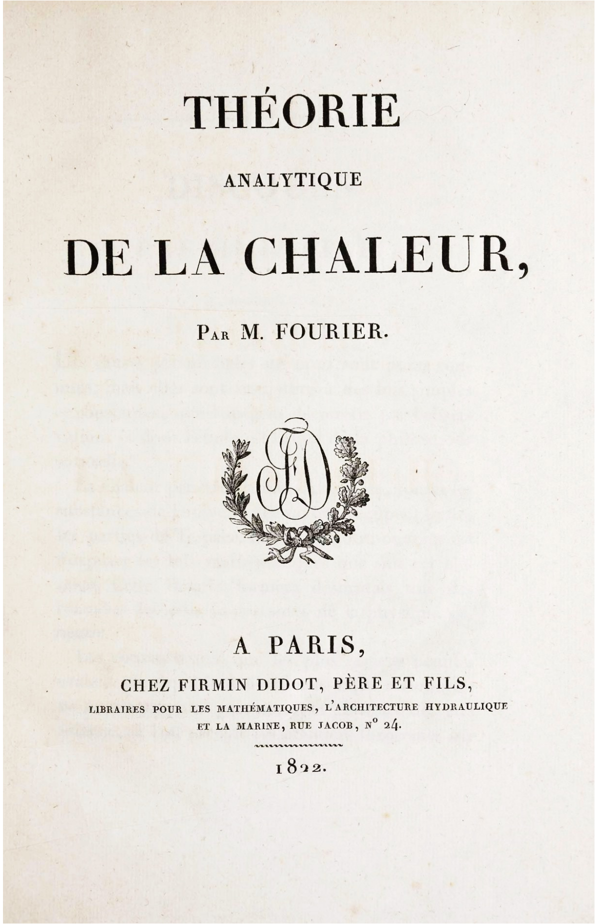
Isaac Newton

Anonymous 1701

Diffusion

Fourier Heat Conduction Law:
"heat flux resulting from thermal conduction is proportional to the magnitude of the temperature gradient and opposite to it in sign"

$$h \propto -\nabla x$$


THÉORIE
ANALYTIQUE
DE LA CHALEUR,
PAR M. FOURIER.

A PARIS,
CHEZ FIRMIN DIDOT, PÈRE ET FILS,
LIBRAIRES POUR LES MATHÉMATIQUES, L'ARCHITECTURE HYDRAULIQUE
ET LA MARINE, RUE JACOB, N° 24.
1822.

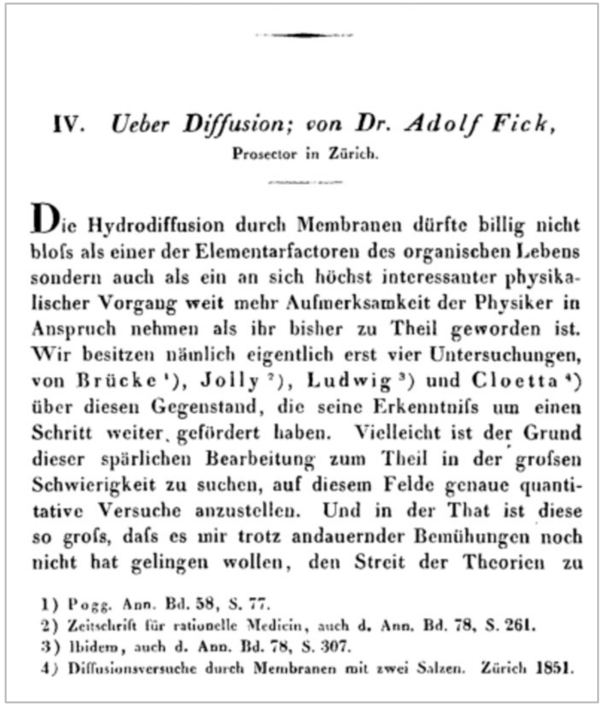

Joseph Fourier

Fourier 1822

Diffusion

Fick (Second) Law of Diffusion:
"temperature change in time equals the heat flux through volume"

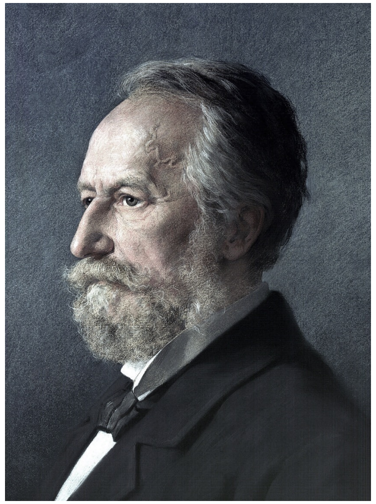
$$\frac{\partial}{\partial t}x = -\mathrm{div}(h)$$



IV. *Ueber Diffusion; von Dr. Adolf Fick,*
Prosector in Zürich.

Die Hydrodiffusion durch Membranen dürfte billig nicht blofs als einer der Elementarfactoren des organischen Lebens sondern auch als ein an sich höchst interessanter physikalischer Vorgang weit mehr Aufmerksamkeit der Physiker in Anspruch nehmen als ihr bisher zu Theil geworden ist. Wir besitzen nämlich eigentlich erst vier Untersuchungen, von Brücke[1]), Jolly[2]), Ludwig[3]) und Cloetta[4]) über diesen Gegenstand, die seine Erkenntnifs um einen Schritt weiter gefördert haben. Vielleicht ist der Grund dieser spärlichen Bearbeitung zum Theil in der grofsen Schwierigkeit zu suchen, auf diesem Felde genaue quantitative Versuche anzustellen. Und in der That ist diese so grofs, dafs es mir trotz andauernder Bemühungen noch nicht hat gelingen wollen, den Streit der Theorien zu

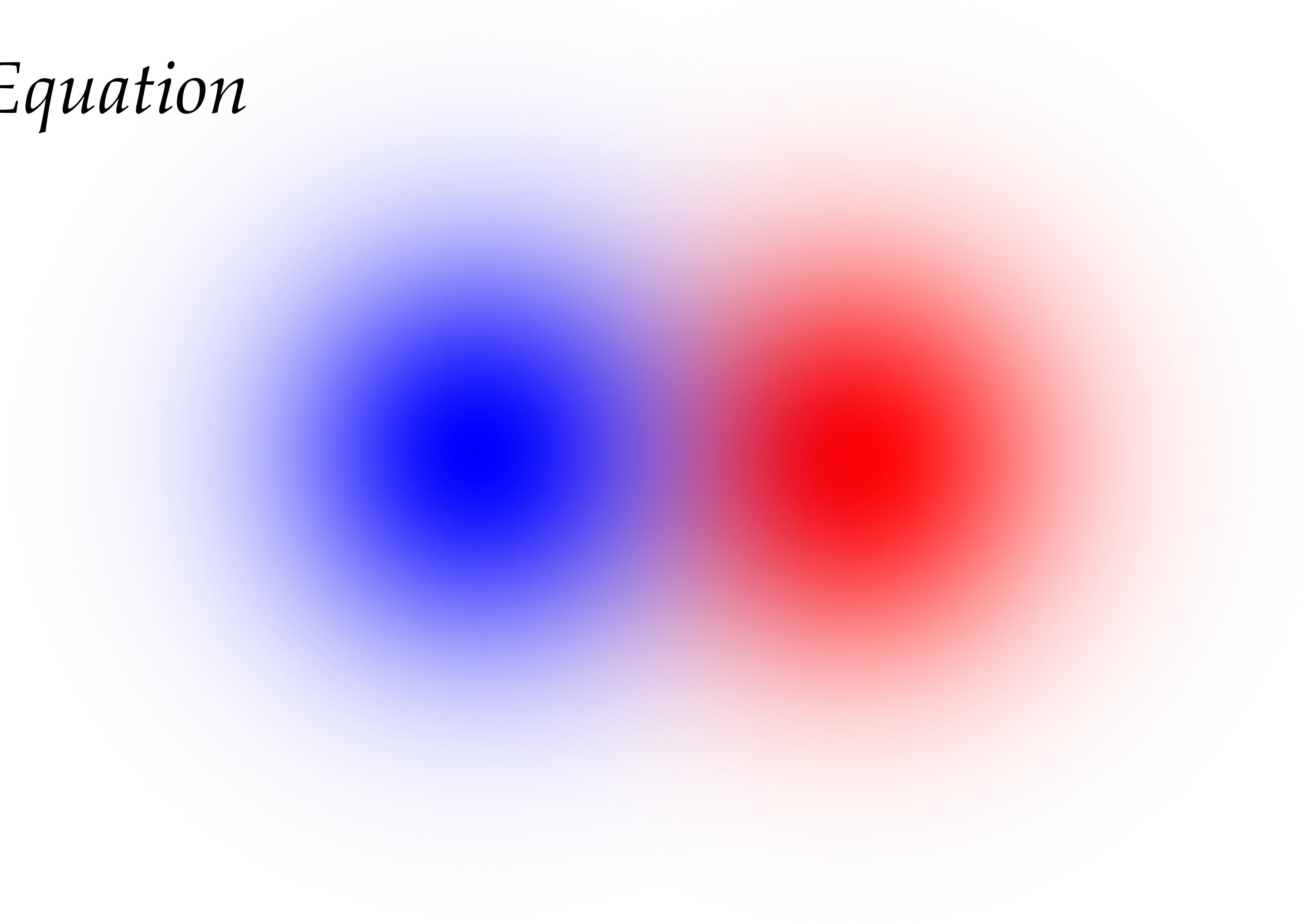
1) Pogg. Ann. Bd. 58, S. 77.
2) Zeitschrift für rationelle Medicin, auch d. Ann. Bd. 78, S. 261.
3) Ibidem, auch d. Ann. Bd. 78, S. 307.
4) Diffusionsversuche durch Membranen mit zwei Salzen. Zürich 1851.



Adolf Eugen Fick

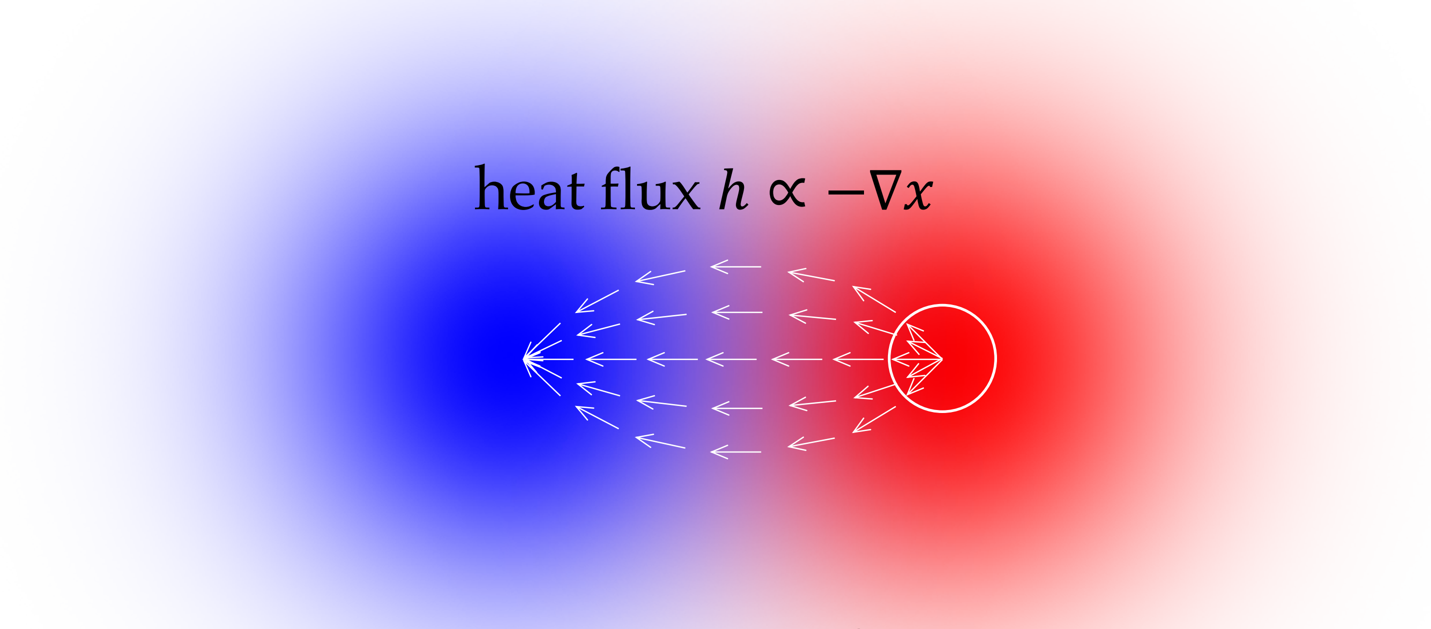
Fick 1855

Diffusion Equation



Fourier 1822; Fick 1855

Diffusion Equation



conservation condition: $\frac{\partial}{\partial t}x = -\mathrm{div}(h)$
("no heat created or disappears")

$$\frac{\partial}{\partial t}x(\mathbf{u},t) = \mathrm{div}\big(a\nabla x(\mathbf{u},t)\big)$$

Fourier 1822; Fick 1855

Diffusion Equation

constant diffusivity

$$\frac{\partial}{\partial t}x = \operatorname{div}(c\nabla x)$$

Diffusion Equation

$$\frac{\partial}{\partial t}x = c\Delta x$$

Homogeneous
Isotropic

1. Gradient flow of the *Dirichlet energy*

$$\mathcal{E}[x] = \frac{1}{2}\int_{\Omega} \|\nabla x(\mathbf{u})\|^2 \mathrm{d}\mathbf{u}$$

Diffusion Equation

$$\frac{\partial}{\partial t}x = c\Delta x$$

Homogeneous
Isotropic

1. Gradient flow of the *Dirichlet energy*

$$\mathcal{E}[x] = \frac{1}{2}\int_{\Omega} \|\nabla x(\mathbf{u})\|^2 \mathrm{d}\mathbf{u}$$

2. Closed form solution: Gaussian filter

$$x(\mathbf{u}, t) = x(\mathbf{u}, 0) \star \frac{1}{(4\pi t)^{d/2}} e^{-\|\mathbf{u}\|^2/4t}$$

Diffusion Equation

Position-dependent diffusivity $a(u)$

Position & direction dependent diffusivity $\mathbf{A}(u)$

$$\frac{\partial}{\partial t}x = c\Delta x$$

Homogeneous
Isotropic

$$\frac{\partial}{\partial t}x = \operatorname{div}(a\nabla x)$$

Non-homogeneous
Isotropic

$$\frac{\partial}{\partial t}x = \operatorname{div}(\mathbf{A}\nabla x)$$

Non-homogeneous
Anisotropic

Diffusion Equation in Image Processing

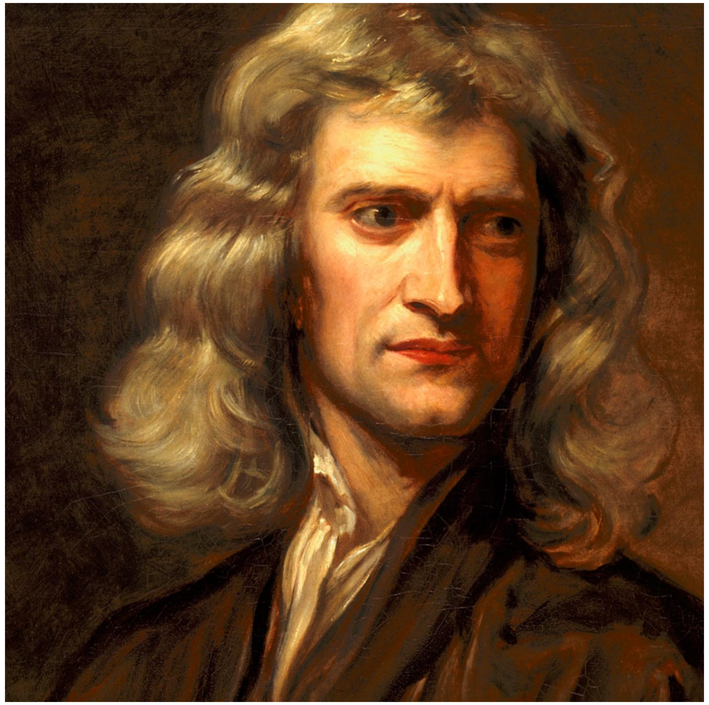
Edge indicator $a(u) \propto \|\nabla x(u)\|^{-1}$

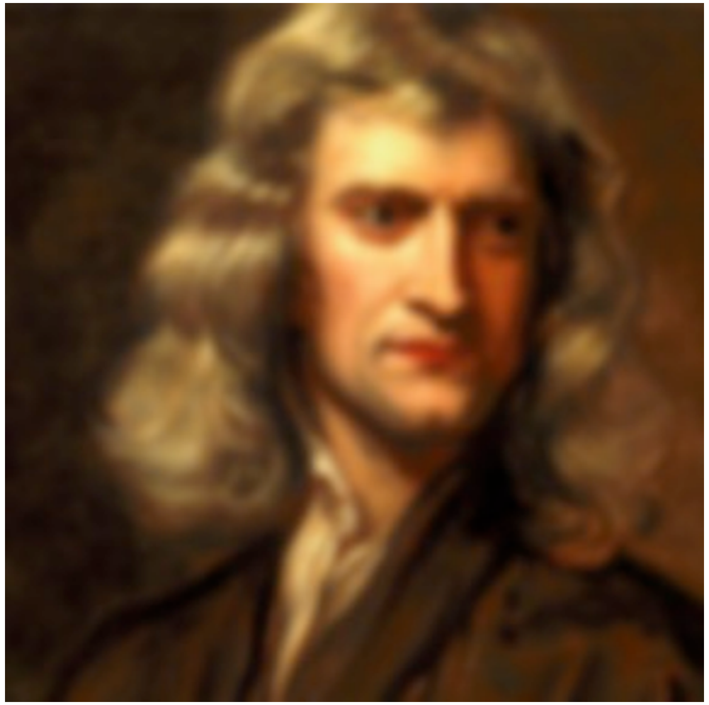
$$\frac{\partial}{\partial t}x = \operatorname{div}(a(x)\nabla x)$$

"Do not diffuse across edges"

Perona, Malik 1990

Diffusion in Image Processing



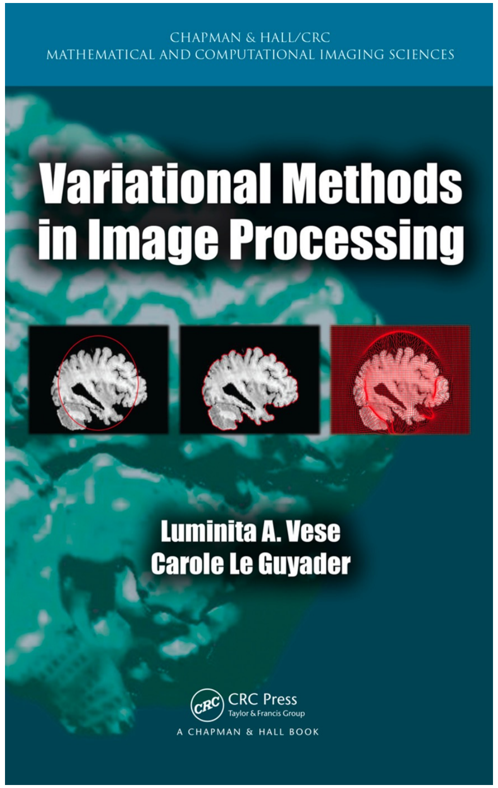


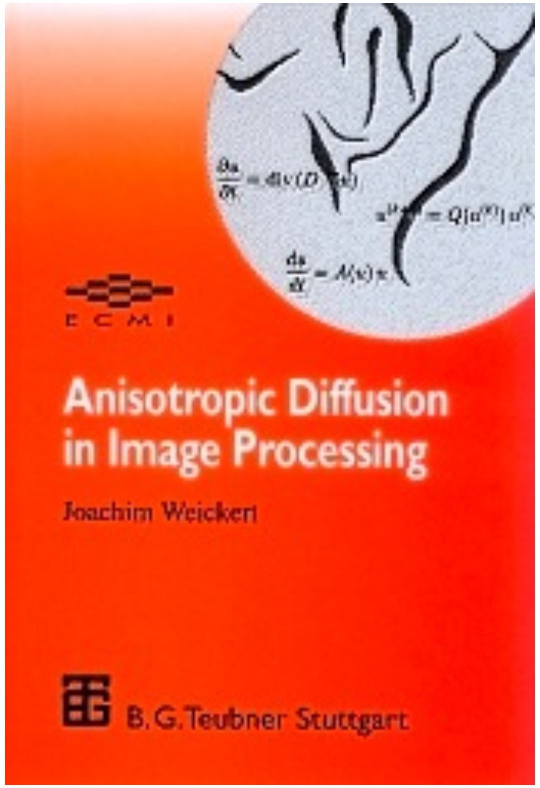
Homogeneous diffusion

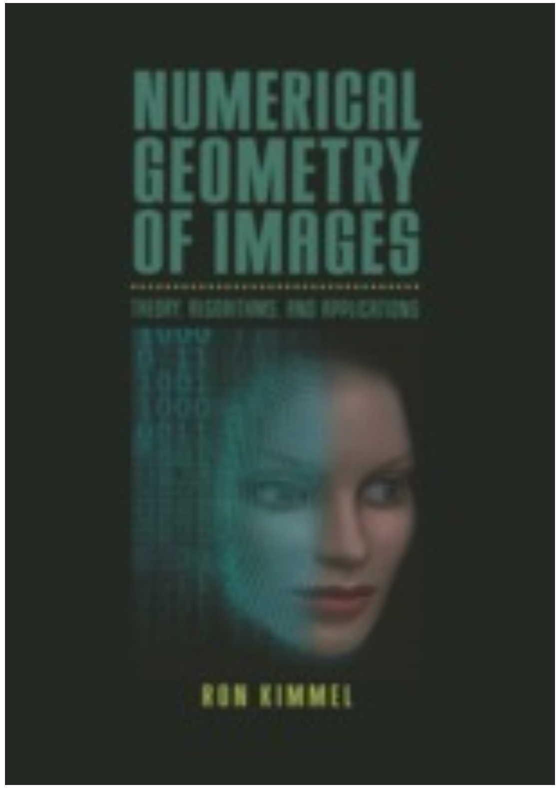
Non-homogeneous diffusion

Perona, Malik 1990; Kimmel et al. 1997; Sochen et al. 1998; Tomasi, Manduchi 1998; Weickert 1998; Buades et al. 2005

Diffusion in Image Processing







Perona, Malik 1990; Kimmel et al. 1997; Sochen et al. 1998; Tomasi, Manduchi 1998; Weickert 1998; Buades et al. 2005

Diffusion Equation on Graphs

$$\frac{\partial}{\partial t}\mathbf{x}_i(t) = \underbrace{\sum_{j:(i,j)\in E}}_{\text{div}} \underbrace{a\left(\mathbf{x}_i(t), \mathbf{x}_j(t)\right)}_{\substack{\text{diffusivity} \\ \mathbf{A}(\mathbf{X})}} \underbrace{\left(\mathbf{x}_j(t) - \mathbf{x}_i(t)\right)}_{\substack{\text{gradient} \\ \nabla\mathbf{X}}}$$

Chamberlain, Rowbottom, et B. 2021

Diffusion Equation on Graphs

$$\frac{\partial}{\partial t}\mathbf{x}_i(t) = \sum_{j:(i,j)\in E} a\left(\mathbf{x}_i(t), \mathbf{x}_j(t)\right)\left(\mathbf{x}_j(t) - \mathbf{x}_i(t)\right)$$

Explicit (Forward Euler) discretization: $t = k\tau$

$$\underbrace{\frac{\mathbf{x}_i^{(k+1)} - \mathbf{x}_i^{(k)}}{\tau}}_{\text{forward difference}} = \sum_{j:(i,j)\in E} a\left(\mathbf{x}_i^{(k)}, \mathbf{x}_j^{(k)}\right)\left(\mathbf{x}_j^{(k)} - \mathbf{x}_i^{(k)}\right)$$

Diffusion Equation on Graphs

$$\frac{\partial}{\partial t}\mathbf{x}_i(t) = \sum_{j:(i,j)\in E} a\left(\mathbf{x}_i(t), \mathbf{x}_j(t)\right)\left(\mathbf{x}_j(t) - \mathbf{x}_i(t)\right)$$

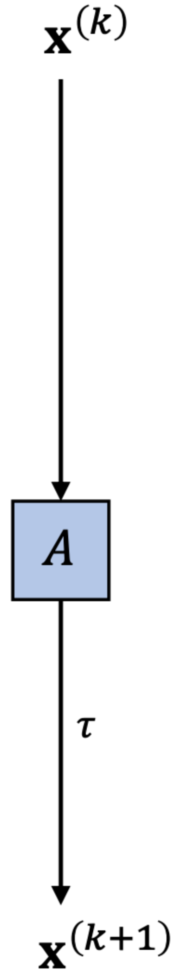
Explicit (Forward Euler) discretization: $t = k\tau$

$$\mathbf{x}_i^{(k+1)} = \sum_{j:(i,j)\in E} a\left(\mathbf{x}_i^{(k)}, \mathbf{x}_j^{(k)}\right)\mathbf{x}_j^{(k)}$$

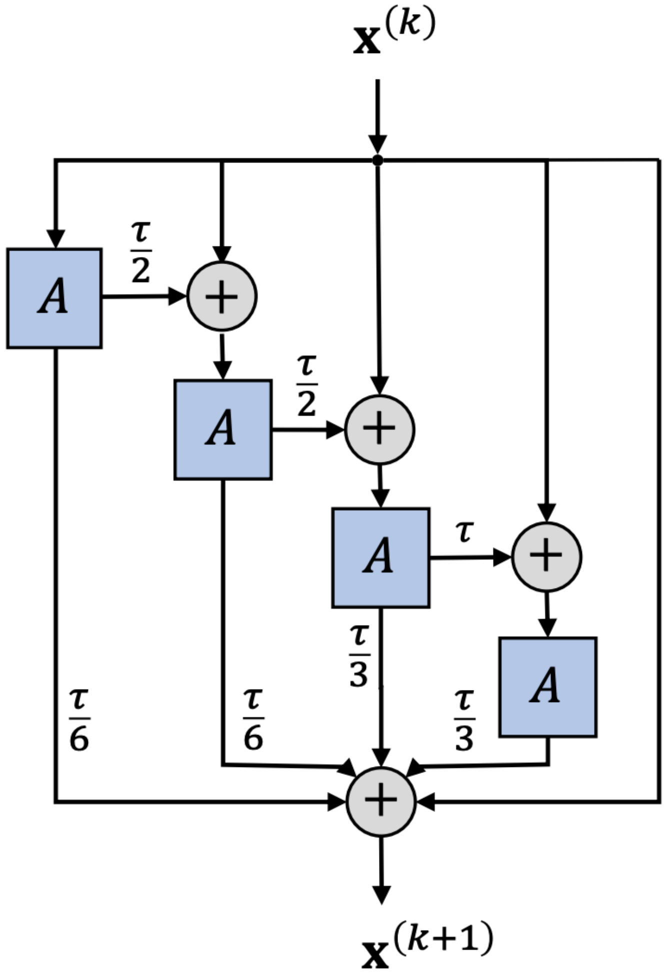
normalised $\sum_j a_{ij} = 1$
unit step $\tau = 1$

GAT is a particular discretisation of graph diffusion

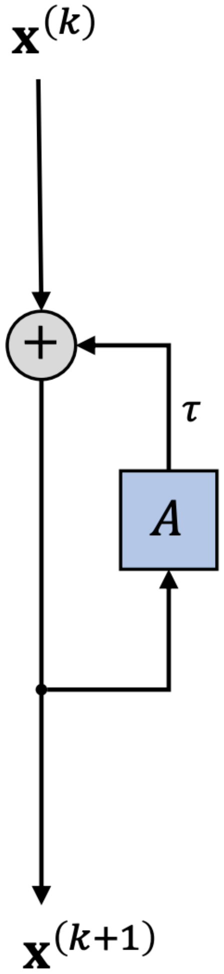
Chamberlain, Rowbottom, et B. 2021



Explicit
Fixed step



Explicit
Multi-step (Runge-Kutta)



Implicit

Graph Neural Diffusion (GRAND)

Given graph $G = (V, E)$ with input node features $\mathbf{X}_{\text{in}}$

- Set initial condition: $\mathbf{X}(0) = \phi(\mathbf{X}_{\text{in}})$
- Solve graph diffusion eqn: $\mathbf{X}(T) = \mathbf{X}(0) + \int_0^T \operatorname{div}\Big(\mathbf{A}\big(\mathbf{X}(t)\big)\boldsymbol{\nabla}\mathbf{X}(t)\Big)\,\mathrm{d}t$
  using an iterative solver
- Output: $\mathbf{Y} = \psi\big(\mathbf{X}(T)\big)$

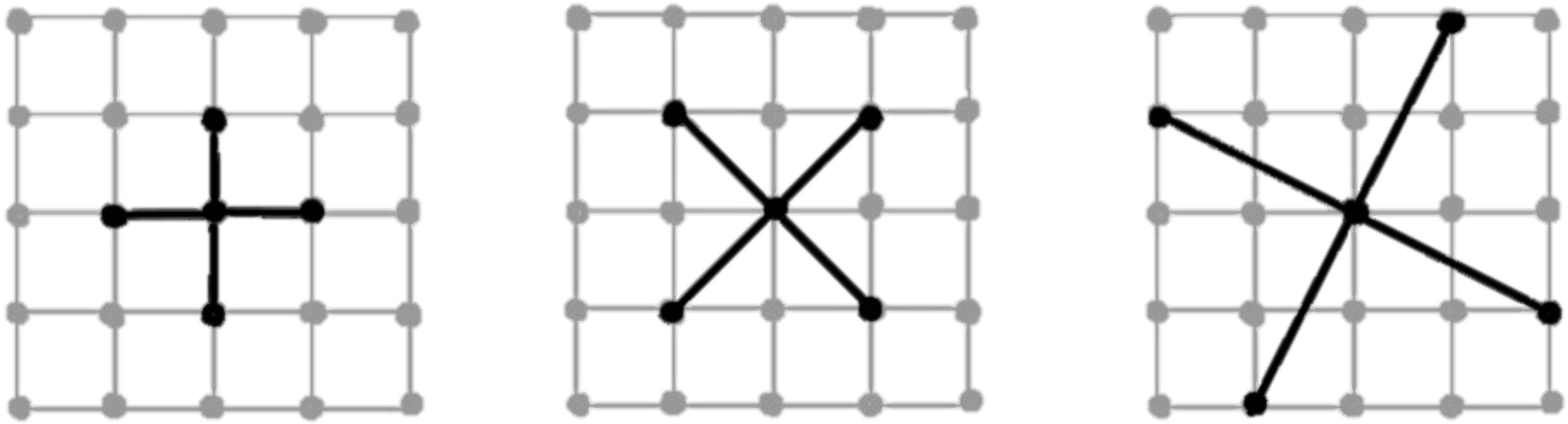
where ϕ, ψ and the diffusivity A are learnable functions

Chamberlain, Rowbottom, et B. 2021

What do we gain?

- New perspectives on old problems (e.g. oversmoothing, bottlenecks, etc)
- New architectures
  - Many GNNs can be formalised as a discretised Graph Diffusion equation
  - More efficient solvers (multistep, adaptive, implicit, multigrid, etc.)
  - Implicit schemes = multi-hop filters
- Theoretical guarantees (e.g. stability, convergence, etc.)
- Deep links to other fields less known in GNN literature (e.g. differential geometry and algebraic topology)

Spatial Derivative: Graph Rewiring?

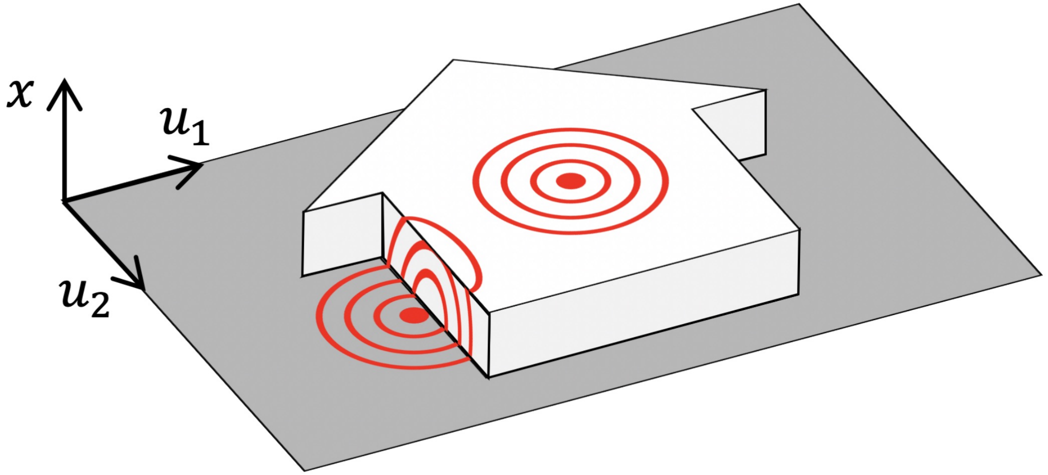


Different discretisations of 2D Laplacian

Images as embedded manifolds

$$\frac{\partial}{\partial t}\mathbf{x} = -\mathrm{div}(a(\mathbf{x})\nabla\mathbf{x})$$

Non-linear diffusion



$$\frac{\partial}{\partial t}\mathbf{z} = \Delta_{\mathbf{G}}\mathbf{z}$$

Non-Euclidean diffusion

Kimmel et al. 1997; Sochen et al. 1998

Beltrami flow

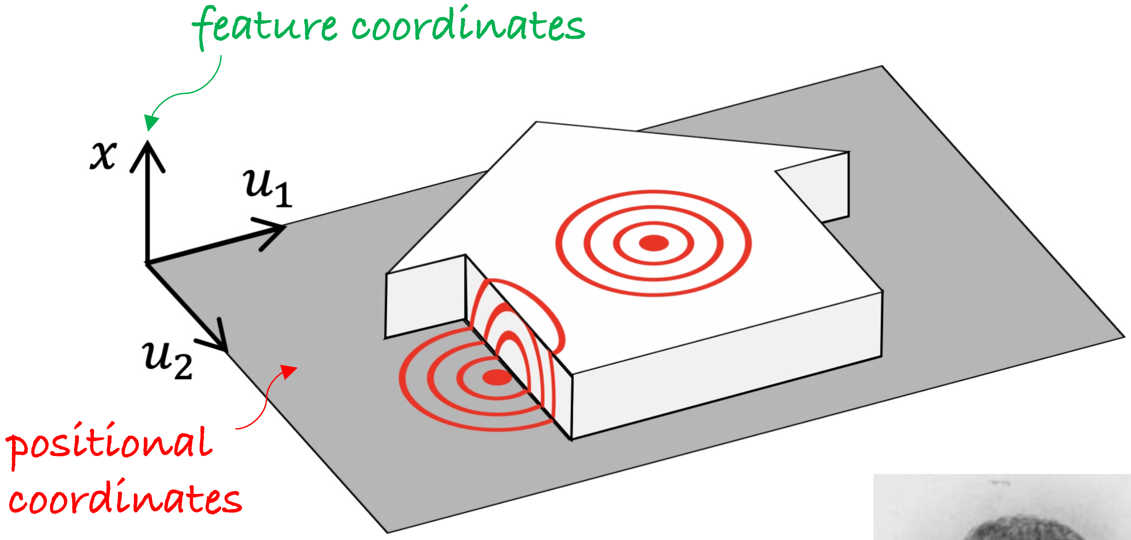
- Consider image as embedded 2-*manifold*

$$\mathbf{z}(\mathbf{u}) = \big(\mathbf{u}, \alpha\mathbf{x}(\mathbf{u})\big)$$

- *Pullback metric:* 2×2 matrix

$$\mathbf{G} = \mathbf{I} + \alpha^2 \big(\nabla_{\mathbf{u}}\mathbf{x}(\mathbf{u})\big)^{\top} \nabla_{\mathbf{u}}\mathbf{x}(\mathbf{u})$$

- *Beltrami flow* = gradient flow of the *Polyakov energy* (harmonic energy of the embedding used in string theory)



$$\frac{\partial}{\partial t}\mathbf{z} = \Delta_{\mathbf{G}}\mathbf{z}$$

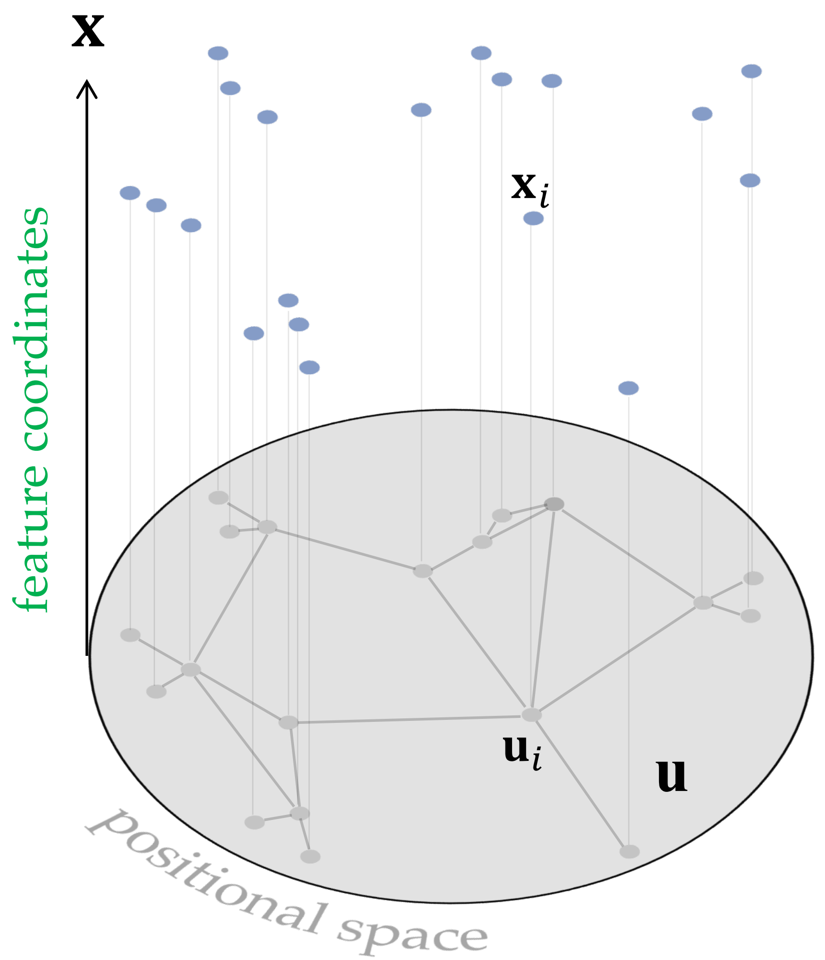
Eugenio Beltrami

Kimmel et al. 1997; Sochen et al. 1998

Graph Beltrami flow

- Graph with positional and feature node coordinates $\mathbf{z}_i = (\mathbf{u}_i, \mathbf{x}_i)$
- **Graph Beltrami flow**

$$\frac{\partial}{\partial t}\mathbf{z}_i = \sum_{j:(i,j)\in E} a(\mathbf{z}_i, \mathbf{z}_j)(\mathbf{z}_j - \mathbf{z}_i)$$



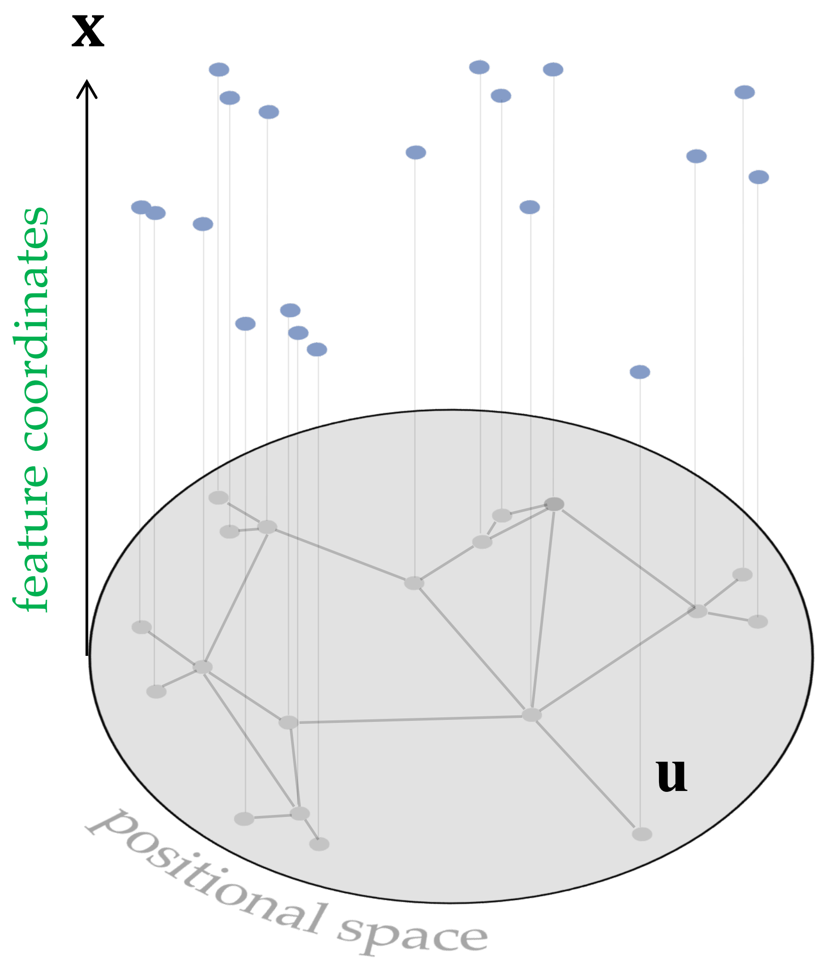
Chamberlain, Rowbottom, et B. 2021

Graph Beltrami flow

- Graph with positional and feature node coordinates $\mathbf{z}_i = (\mathbf{u}_i, \mathbf{x}_i)$
- **Graph Beltrami flow**

$$\frac{\partial}{\partial t}\mathbf{z}_i = \sum_{j:(i,j)\in E} a(\mathbf{z}_i, \mathbf{z}_j)(\mathbf{z}_j - \mathbf{z}_i)$$

  - Evolution of $\mathbf{x}$ = feature diffusion



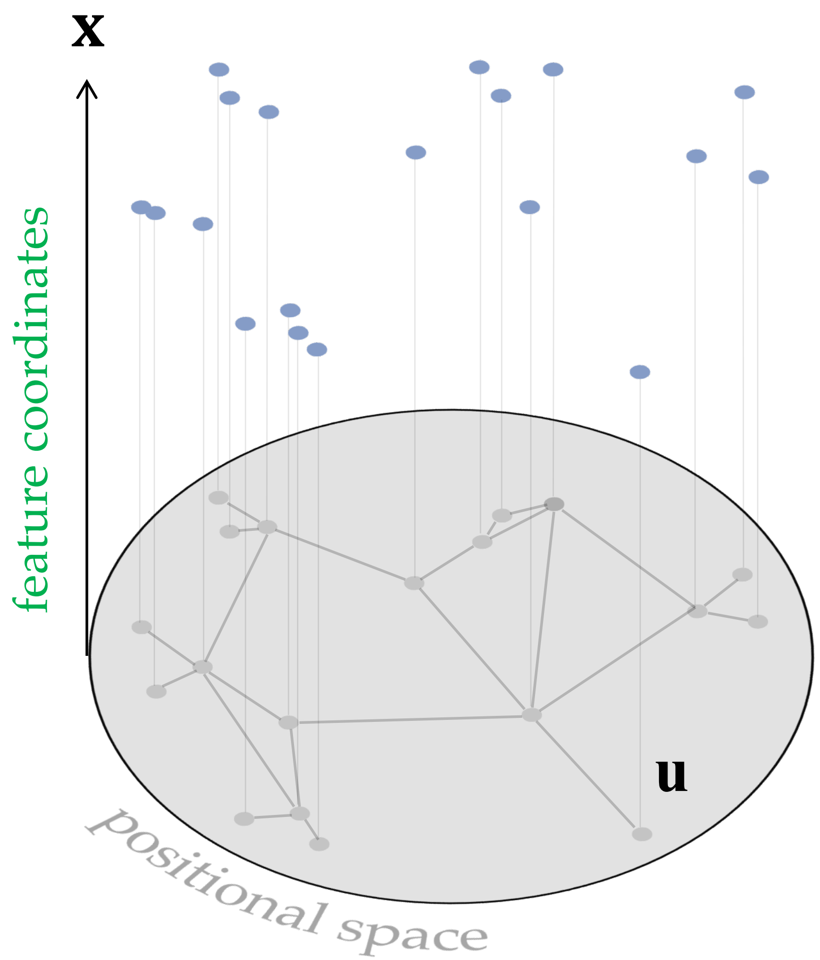
Chamberlain, Rowbottom, et B. 2021

Graph Beltrami flow

- Graph with positional and feature node coordinates $\mathbf{z}_i = (\mathbf{u}_i, \mathbf{x}_i)$
- **Graph Beltrami flow**

$$\frac{\partial}{\partial t}\mathbf{z}_i = \sum_{j:(i,j)\in E} a(\mathbf{z}_i, \mathbf{z}_j)(\mathbf{z}_j - \mathbf{z}_i)$$

  - Evolution of $\mathbf{x}$ = feature diffusion
  - Evolution of $\mathbf{u}$ = graph rewiring



Chamberlain, Rowbottom, et B. 2021

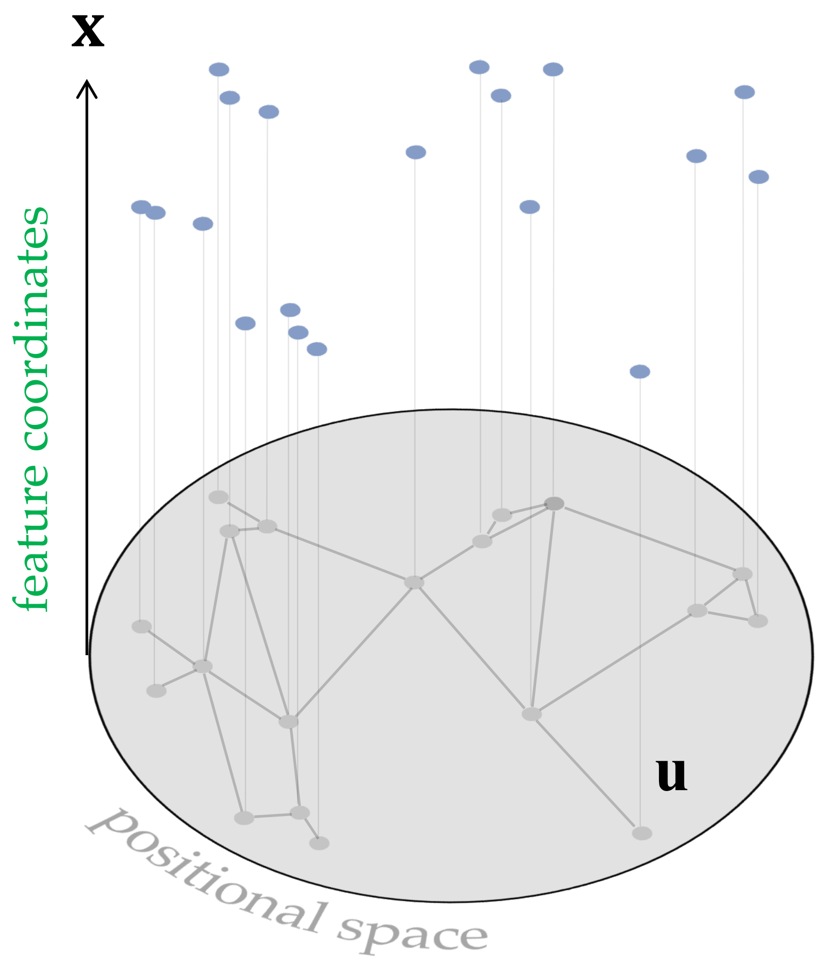
Graph Beltrami flow

- Graph with positional and feature node coordinates $\mathbf{z}_i = (\mathbf{u}_i, \mathbf{x}_i)$
- **Graph Beltrami flow**

$$\frac{\partial}{\partial t}\mathbf{z}_i = \sum_{j:(i,j)\in E'} a(\mathbf{z}_i, \mathbf{z}_j)(\mathbf{z}_j - \mathbf{z}_i)$$

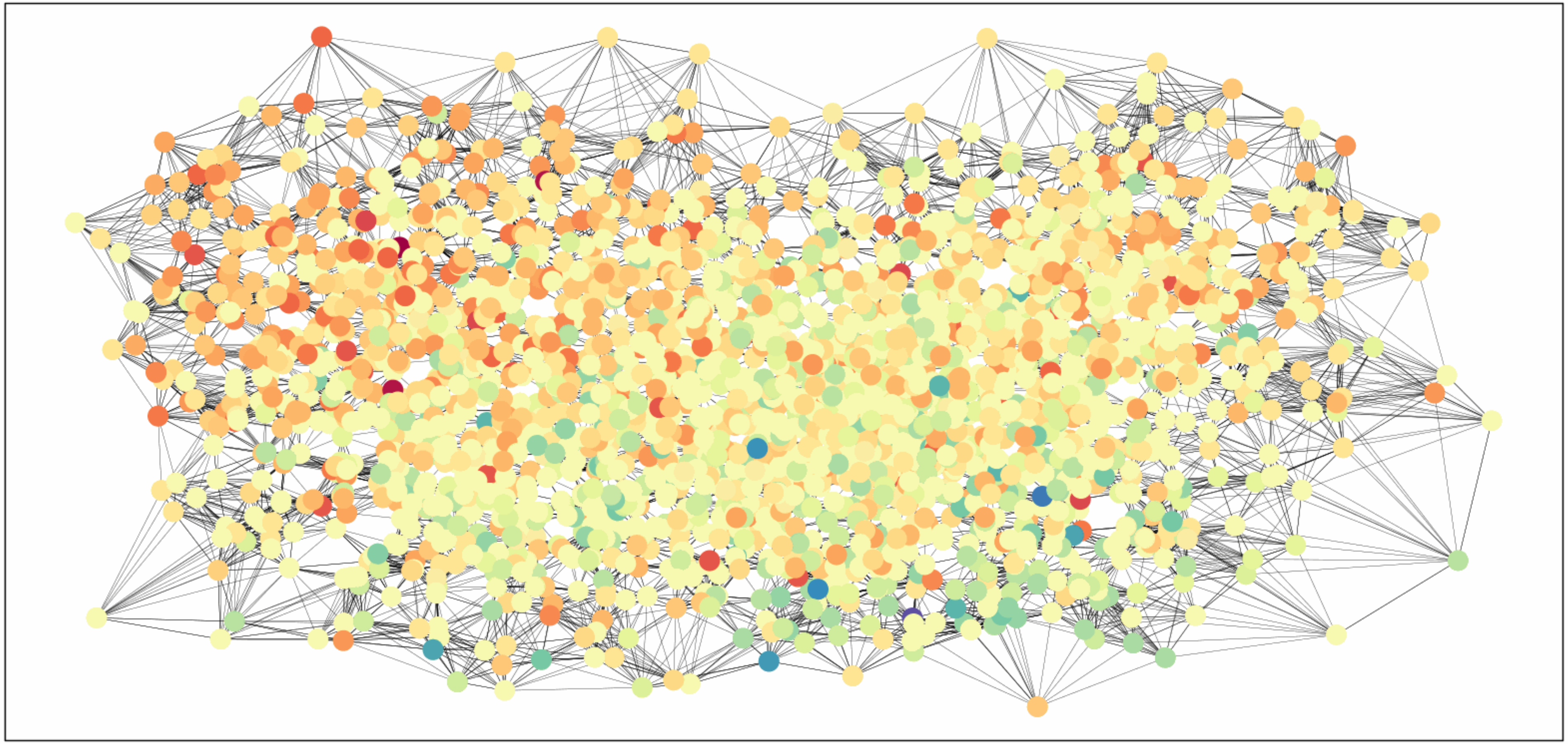
rewired graph

  - Evolution of $\mathbf{x}$ = feature diffusion
  - Evolution of $\mathbf{u}$ = graph rewiring



Chamberlain, Rowbottom, et B. 2021

Graph Beltrami flow



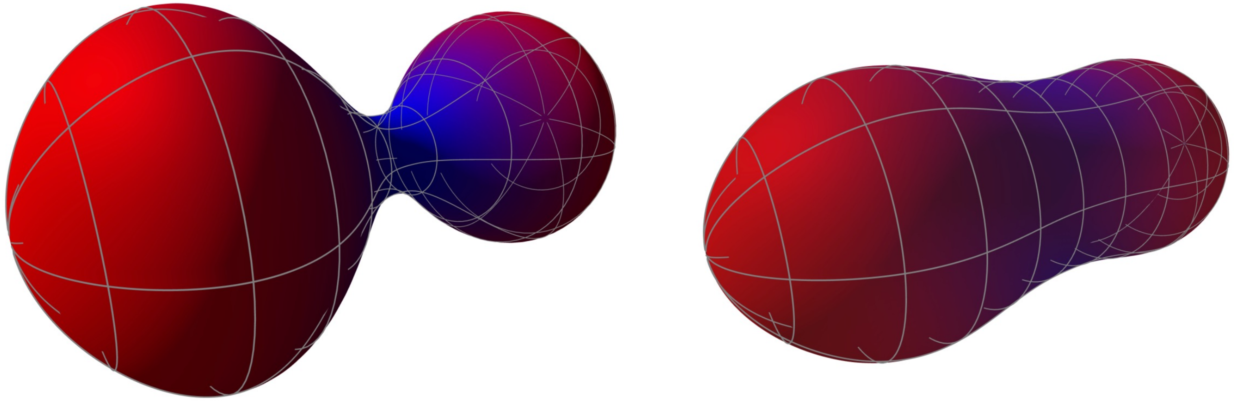
Evolution of positional/feature components + rewiring of the Cora graph

Chamberlain, Rowbottom, et B. 2021

Ricci flow

- **Ricci flow:** "diffusion of the Riemannian metric"

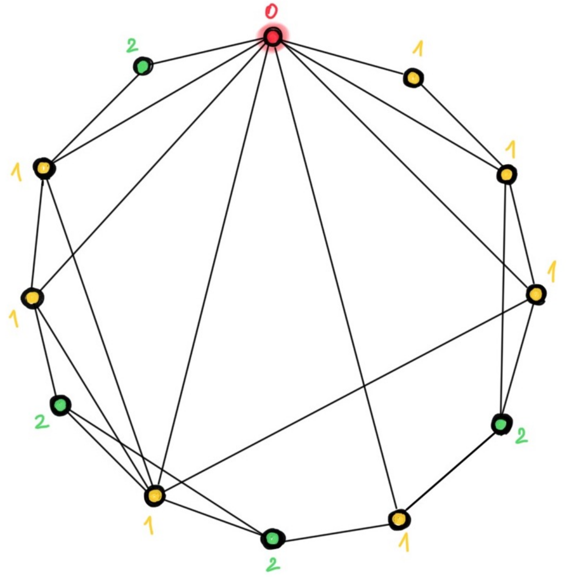
$$\frac{\partial g_{ij}}{\partial t} = R_{ij}$$

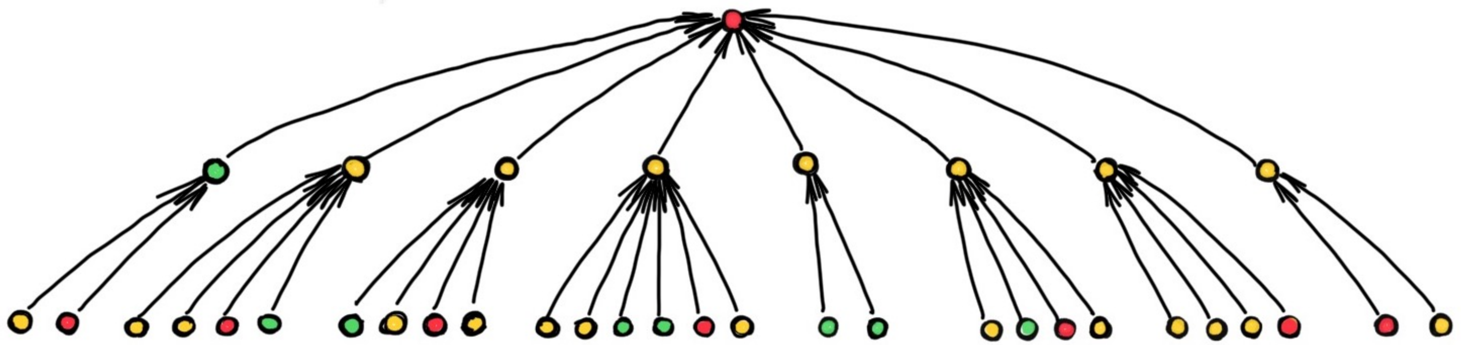


Evolution of a manifold under Ricci flow

Ricci 1903; Hamilton 1988; Perelman 2003

Over-squashing & Bottlenecks





In small-world graphs metric ball volume $\mathrm{vol}(B_k) = \sum_{j \in B_k} d_j$ grows exponentially with ball radius k

Long-distance dependency + Fast volume growth
= Over-squashing

Alon, Yahav 2020

Characterisation of Over-squashing in GNNs

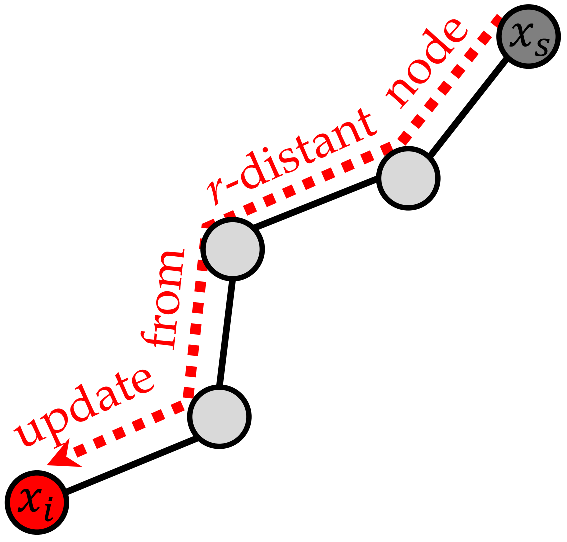
- Multilayer MPNN-type GNN of the form

$$x_i^{(\ell+1)} = \phi_\ell\left(x_i^{(\ell)}, \sum_{j=1}^{n} \hat{a}_{ij}\psi_\ell\left(x_i^{(\ell)}, x_j^{(\ell)}\right)\right)$$

- $|\nabla\phi_\ell| \leq \alpha$ and $|\nabla\psi_\ell| \leq \beta$ for $\ell = 0,1,\ldots,L$.

Lemma 1 (sensitivity): Let node s be geodesically $d_G(i,s) = r + 1$ away from node i. Then

$$\left|\frac{\partial x_i^{(r+1)}}{\partial x_s}\right| \leq (\alpha\beta)^{r+1}\left(\widehat{\mathbf{A}}^{r+1}\right)_{is}$$



Over-squashing: small Jacobian $\left|\frac{\partial x_i^{(r+1)}}{\partial x_s}\right|$ leads to poor information propagation

Topping, di Giovanni, et B. 2021

Characterisation of Over-squashing in GNNs

- Multilayer MPNN-type GNN of the form

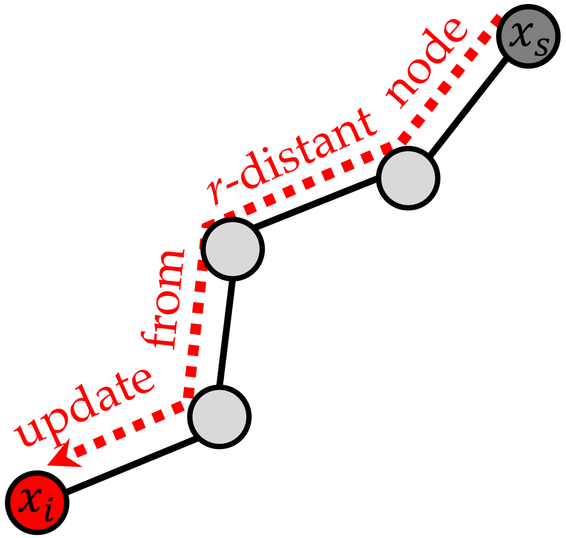
$$x_i^{(\ell+1)} = \phi_\ell\left(x_i^{(\ell)}, \sum_{j=1}^{n} \hat{a}_{ij}\psi_\ell\left(x_i^{(\ell)}, x_j^{(\ell)}\right)\right)$$

- $|\nabla\phi_\ell| \leq \alpha$ and $|\nabla\psi_\ell| \leq \beta$ for $\ell = 0,1,\ldots,L$.

Lemma 1 (sensitivity): Let node s be geodesically $d_G(i,s) = r+1$ away from node i. Then

$$\left|\frac{\partial x_i^{(r+1)}}{\partial x_s}\right| \leq (\alpha\beta)^{r+1}\left(\widehat{\mathbf{A}}^{r+1}\right)_{is}$$

It's the graph structure ("bottleneck") to blame!



Over-squashing: small Jacobian $\left|\frac{\partial x_i^{(r+1)}}{\partial x_s}\right|$ leads to poor information propagation

Topping, di Giovanni, et B. 2021

Characterisation of Over-squashing in GNNs

- Multilayer MPNN-type GNN of the form

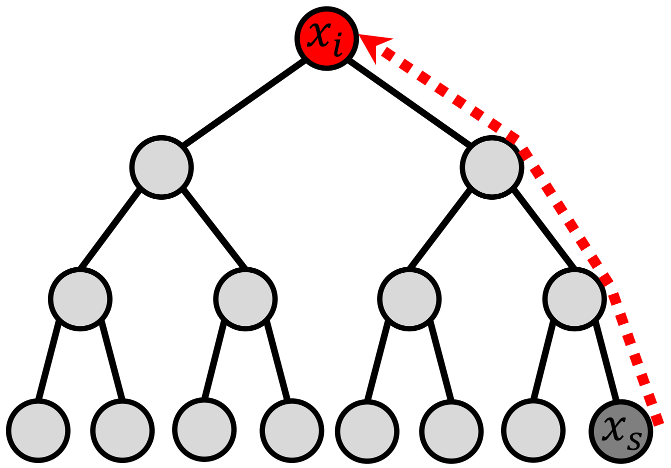
$$x_i^{(\ell+1)} = \phi_\ell\left(x_i^{(\ell)}, \sum_{j=1}^{n} \hat{a}_{ij}\psi_\ell\left(x_i^{(\ell)}, x_j^{(\ell)}\right)\right)$$

- $|\nabla\phi_\ell| \leq \alpha$ and $|\nabla\psi_\ell| \leq \beta$ for $\ell = 0,1,\dots,L$.

Lemma 1 (sensitivity): Let node s be geodesically $d_G(i,s) = r + 1$ away from node i. Then

$$\left|\frac{\partial x_i^{(r+1)}}{\partial x_s}\right| \leq (\alpha\beta)^{r+1}\left(\hat{\mathbf{A}}^{r+1}\right)_{is}$$

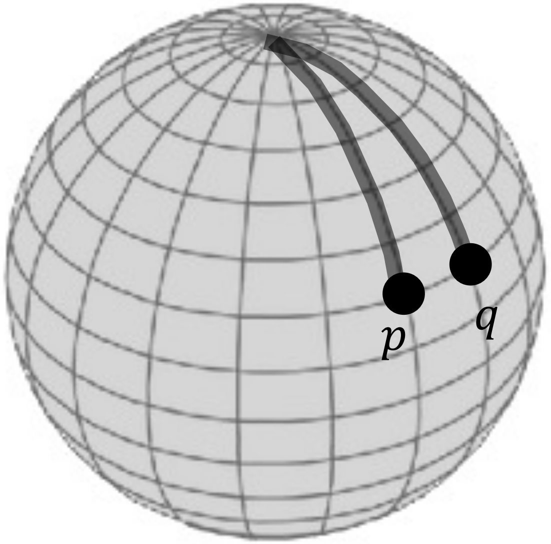
It's the graph structure ("bottleneck") to blame!



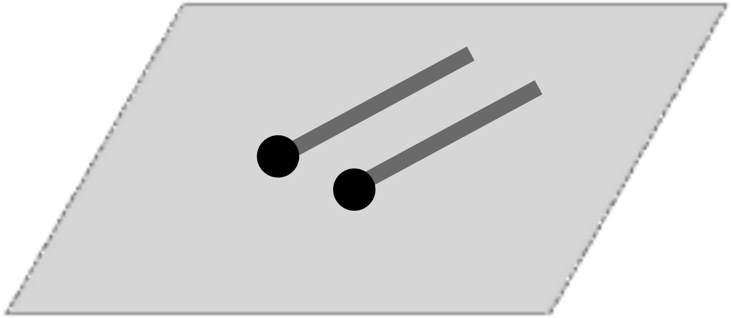
Pathological example: binary tree
$\left(\hat{\mathbf{A}}^{r+1}\right)_{is} = \frac{1}{2} \cdot 3^{-r}$

Topping, di Giovanni, et B. 2021

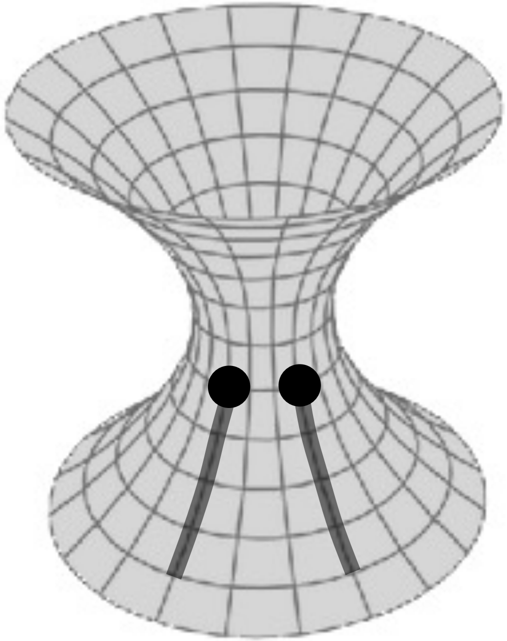
Ricci Curvature on Manifolds



Spherical (>0)



Euclidean (=0)

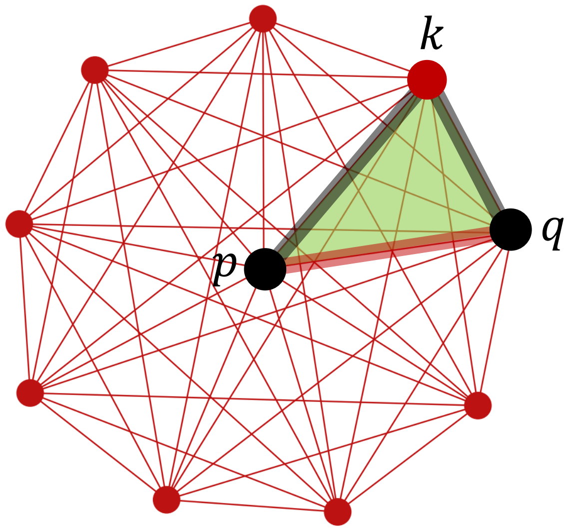


Hyperbolic (<0)

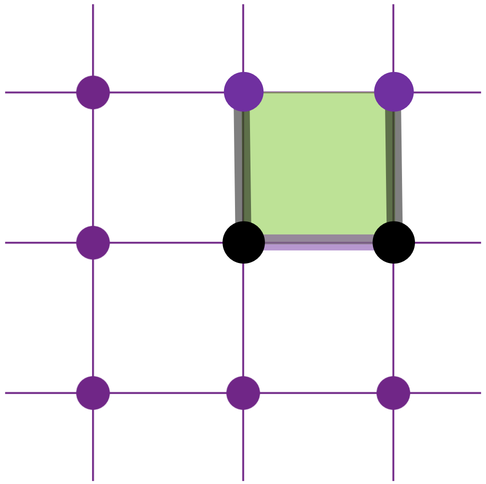
"geodesic dispersion"

Ricci 1903

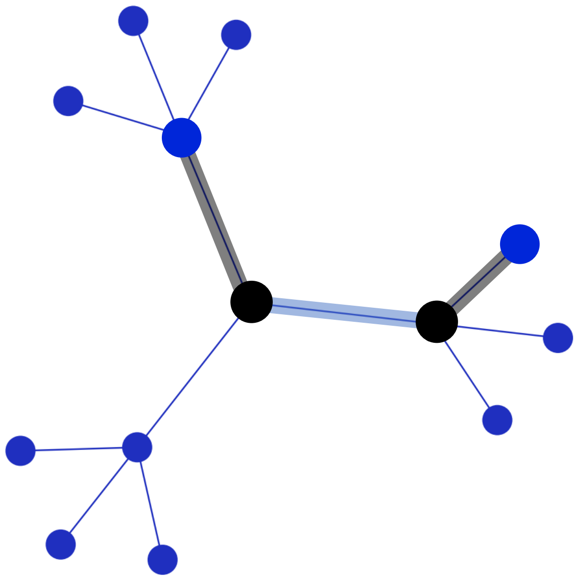
Ricci Curvature on Graphs



Clique (>0)



Grid (=0)



Tree (<0)

Forman 2003; Ollivier 2007; Topping, di Giovanni, et B. 2021

Balanced Forman Curvature

Balanced Forman Curvature of edge $i\sim j$ in simple unweighted graph $\mathrm{Ric}(i,j) = 0$ if $\min\{d_i, d_j\} = 1$ and otherwise

Triangles based at i~j

Max number of 4-cycle based at i~j traversing the same node

$$\mathrm{Ric}(i,j) = \frac{2}{d_i} + \frac{2}{d_j} + 2\frac{|\sharp_\Delta(i,j)|}{\max\{d_i,d_j\}} + \frac{|\sharp_\Delta(i,j)|}{\min\{d_i,d_j\}} + \frac{\gamma_{\max}^{-1}}{\max\{d_i,d_j\}}\left(|\sharp_\square^i(i,j)| + |\sharp_\square^j(i,j)|\right) - 2$$

Degree of i

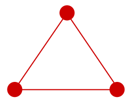
Neighbours of i forming a 4-cycle based at i~j (w/o diagonals)

Forman 2003; Topping, di Giovanni, et B. 2021

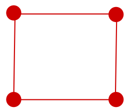
Balanced Forman Curvature

Balanced Forman Curvature of edge $i{\sim}j$ in simple unweighted graph $\mathrm{Ric}(i,j)$ =0 if $\min\{d_i, d_j\} = 1$ and otherwise

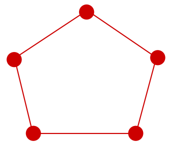
$$\mathrm{Ric}(i,j) = \frac{2}{d_i} + \frac{2}{d_j} + 2\,\frac{|\sharp_\Delta(i,j)|}{\max\{d_i,d_j\}} + \frac{|\sharp_\Delta(i,j)|}{\min\{d_i,d_j\}} + \frac{\gamma_{\max}^{-1}}{\max\{d_i,d_j\}}\left(\left|\sharp_\square^i(i,j)\right| + \left|\sharp_\square^j(i,j)\right|\right) - 2$$



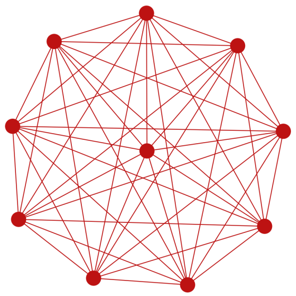
Cycle C_3: $\frac{3}{2}$



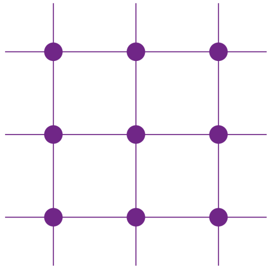
C_4: 1



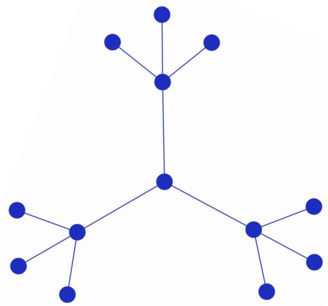
$C_{n\geq 5}$: 0



Clique K_n: $\frac{n}{n-1}$



Grid G_n: 0



Tree T_r: $\frac{4}{r+1} - 2$

Forman 2003; Topping, di Giovanni, et B. 2021

Over-squashing & Bottleneck via Curvature

Theorem 1 (main result): Consider an MPNN with $L \geq 2$ layers and $|\nabla\phi_\ell| \leq \alpha$ and $|\nabla\psi_\ell| \leq \beta$. Let $i \sim j$ with $d_i \leq d_j$ and assume $\exists\delta$ s.t. $0 < \delta < \max\{d_i, d_j\}^{1/2}$, $\delta < \gamma_{\max}^{-1}$ and $\mathrm{Ric}(i,j) \leq -2 + \delta$. Then, there exist nodes $Q \subset \{s\colon d_G(i,s) = 2\}$ of size $|Q| > 1/\delta$ s.t.

small δ = negative curvature

more nodes

$$\frac{1}{|Q|}\sum_{k\in Q}\left|\frac{\partial x_k^{(\ell+2)}}{\partial x_i^{(\ell)}}\right| < (\alpha\beta)^2\delta^{1/4}$$

stronger over-squashing

Over-squashing is caused by negatively-curved edges!

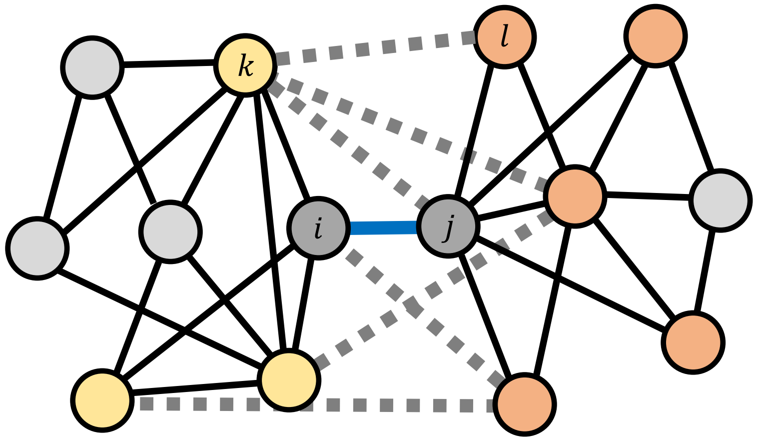
Topping, di Giovanni, et B. 2021

Stochastic Discrete Ricci Flow (SDRF)

Input: graph $G = (V, E)$, temperature $\tau > 0$, (optional C)

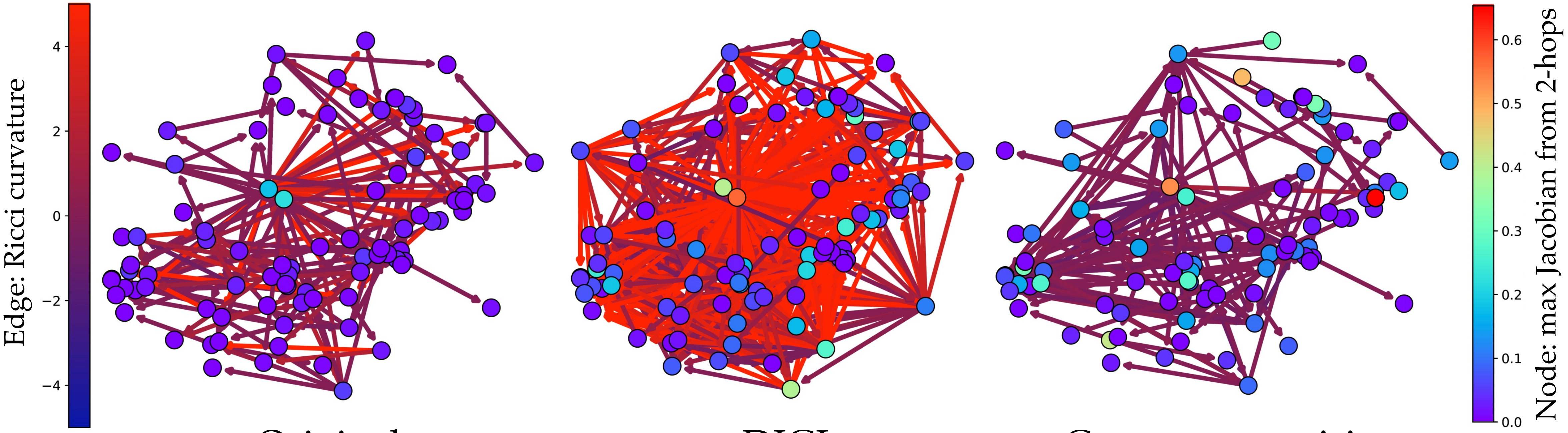
- For edge $i{\sim}j$ with smallest $\mathrm{Ric}(i, j)$
  - Calculate the improvement $\delta_{kl} = \mathrm{Ric}_{G'}(i, j) - \mathrm{Ric}(i, j)$ from adding edge $k{\sim}l$ with $k \in B_1(i)$ and $l \in B_1(j)$
  - Sample index k, l with probability $\mathrm{Softmax}(\tau\delta_{kl})$ and add edge $k{\sim}l$ to E'
- (optional) Remove edge $i{\sim}j$ with largest $\mathrm{Ric}(i, j) > C$

Output: new graph $G' = (V, E')$



Topping, di Giovanni, et B. 2021

Curvature- vs Diffusion-based Rewiring



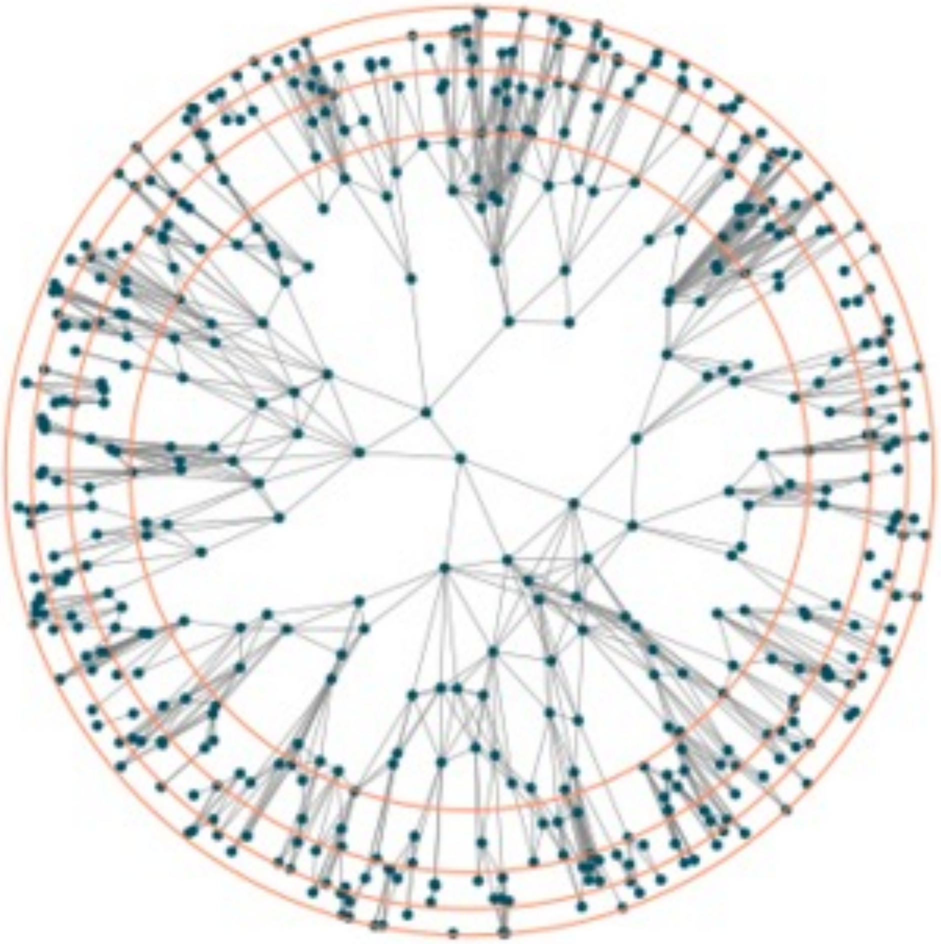
Original
Cornell graph

DIGL
+308% edges

Curvature rewiring
+36% / -36% edges

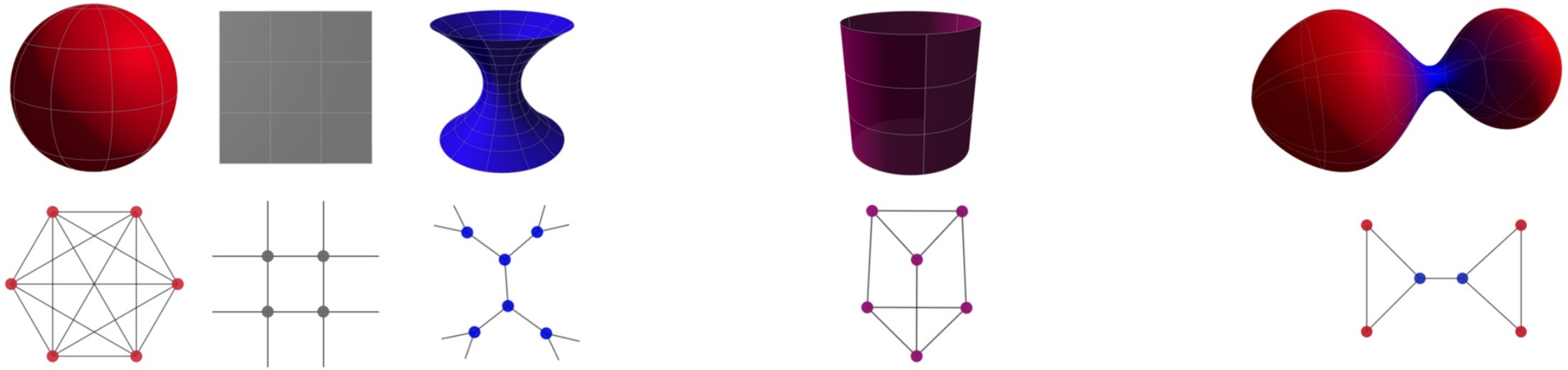
Topping, di Giovanni, et B. 2021; Klicpera et al. 2019 (DIGL)

"Network Geometry"



Boguna et al. 2020

Heterogeneous Embeddings



Space-forms
(sphere, plane, hyperboloid)

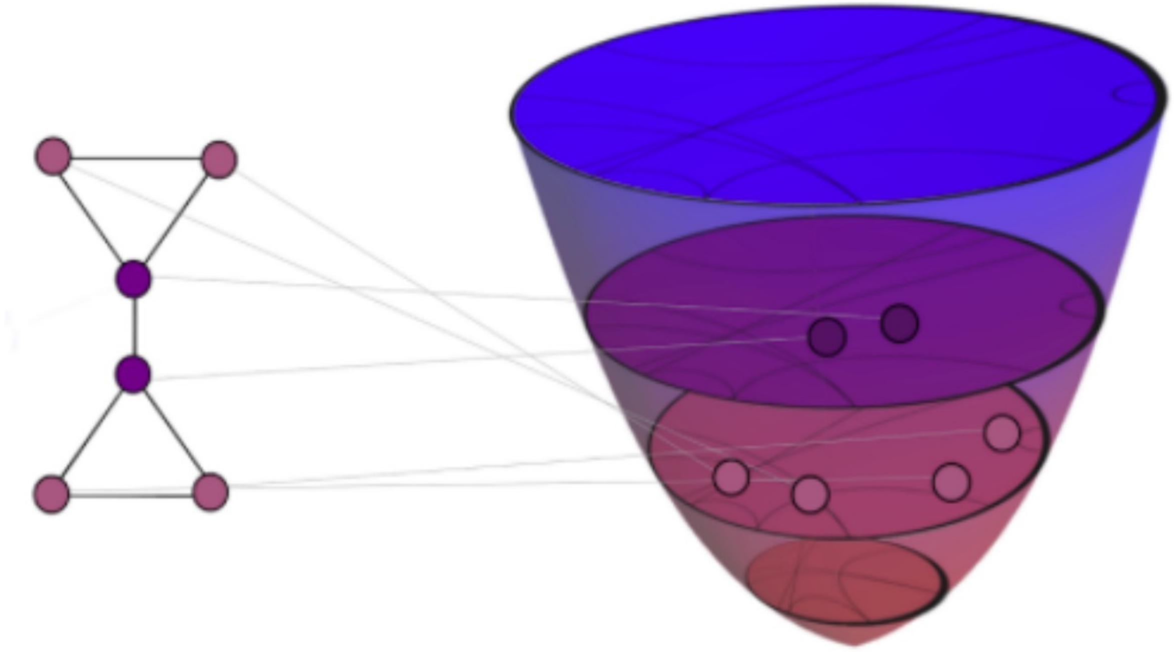
Homogeneous
(product manifolds)

Heterogeneous

Embedding space must have geometry compatible with that of the graph!

Di Giovanni, Luise, B. 2022

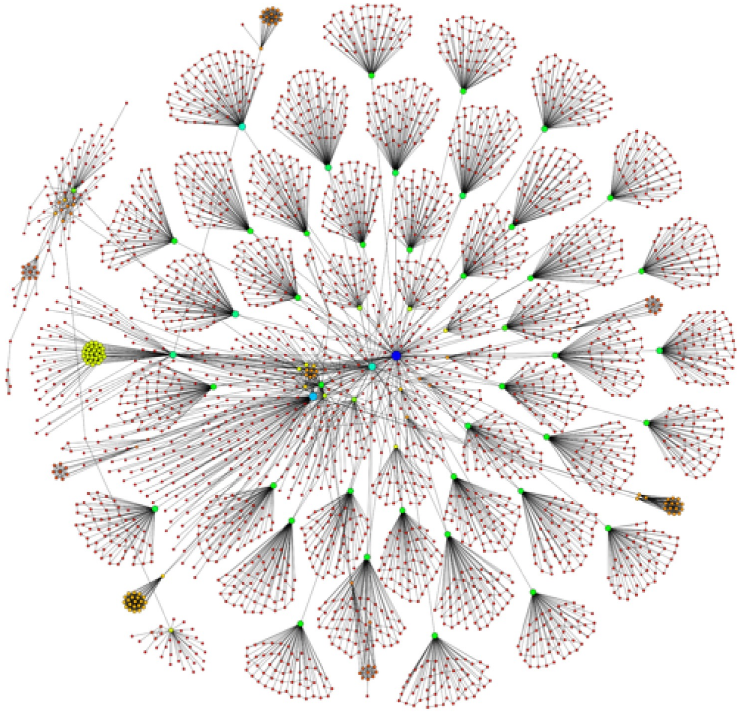
Heterogeneous Embeddings



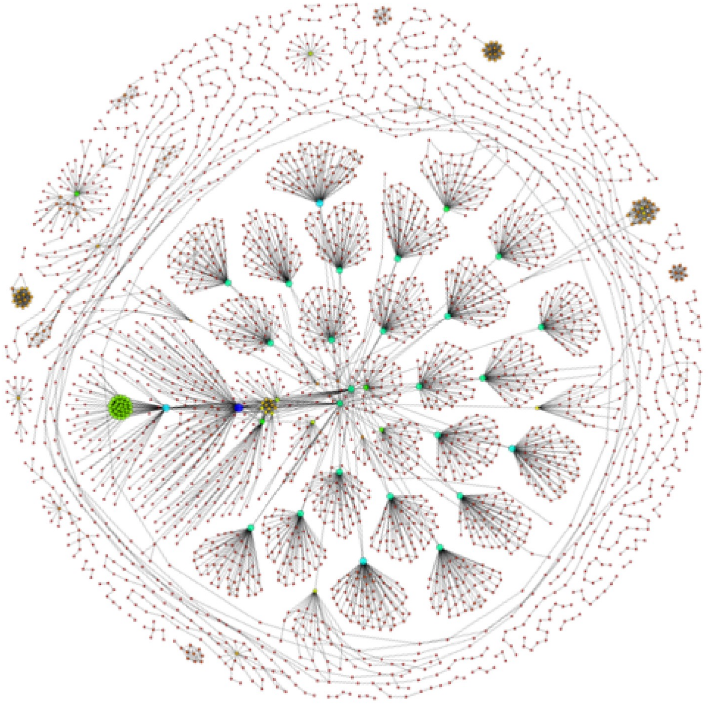
Graph embedding into a heterogeneous manifold constructed as a product with rotationally-symmetric factor with controllable Ricci curvature

Di Giovanni, Luise, B. 2022

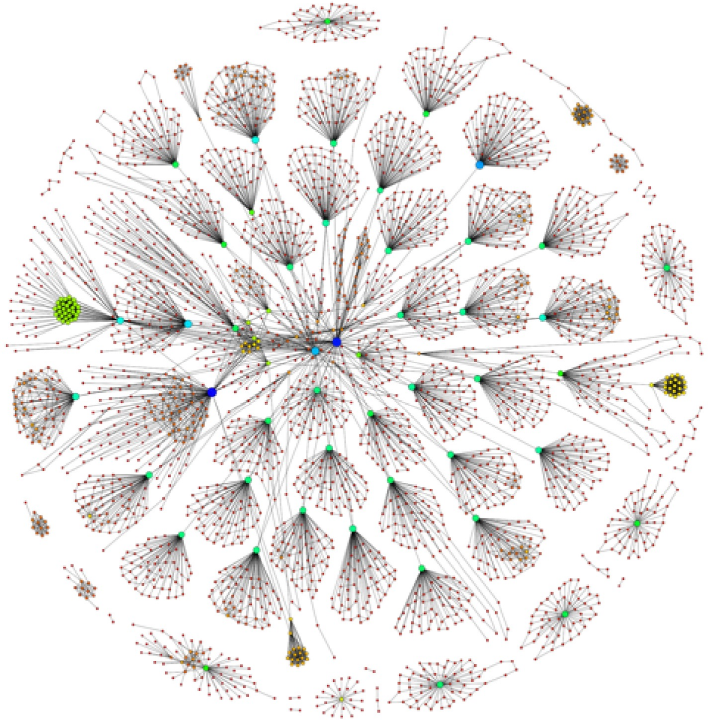
Heterogeneous Embeddings



Groundtruth graph



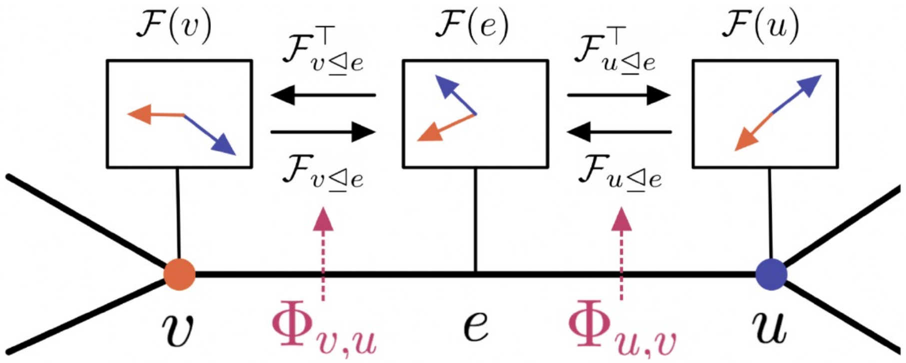
Reconstruction from homogeneous embedding



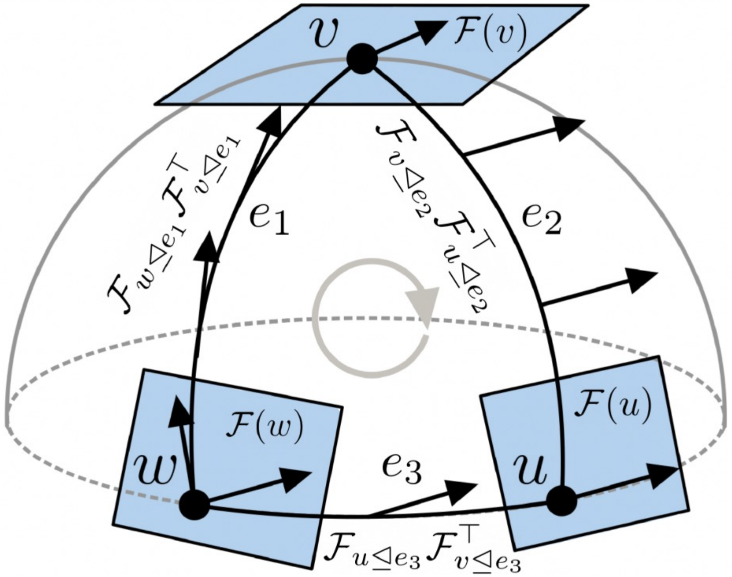
Reconstruction from heterogeneous embedding

Di Giovanni, Luise, B. 2022

Cellular Sheaves



Cellular sheaf



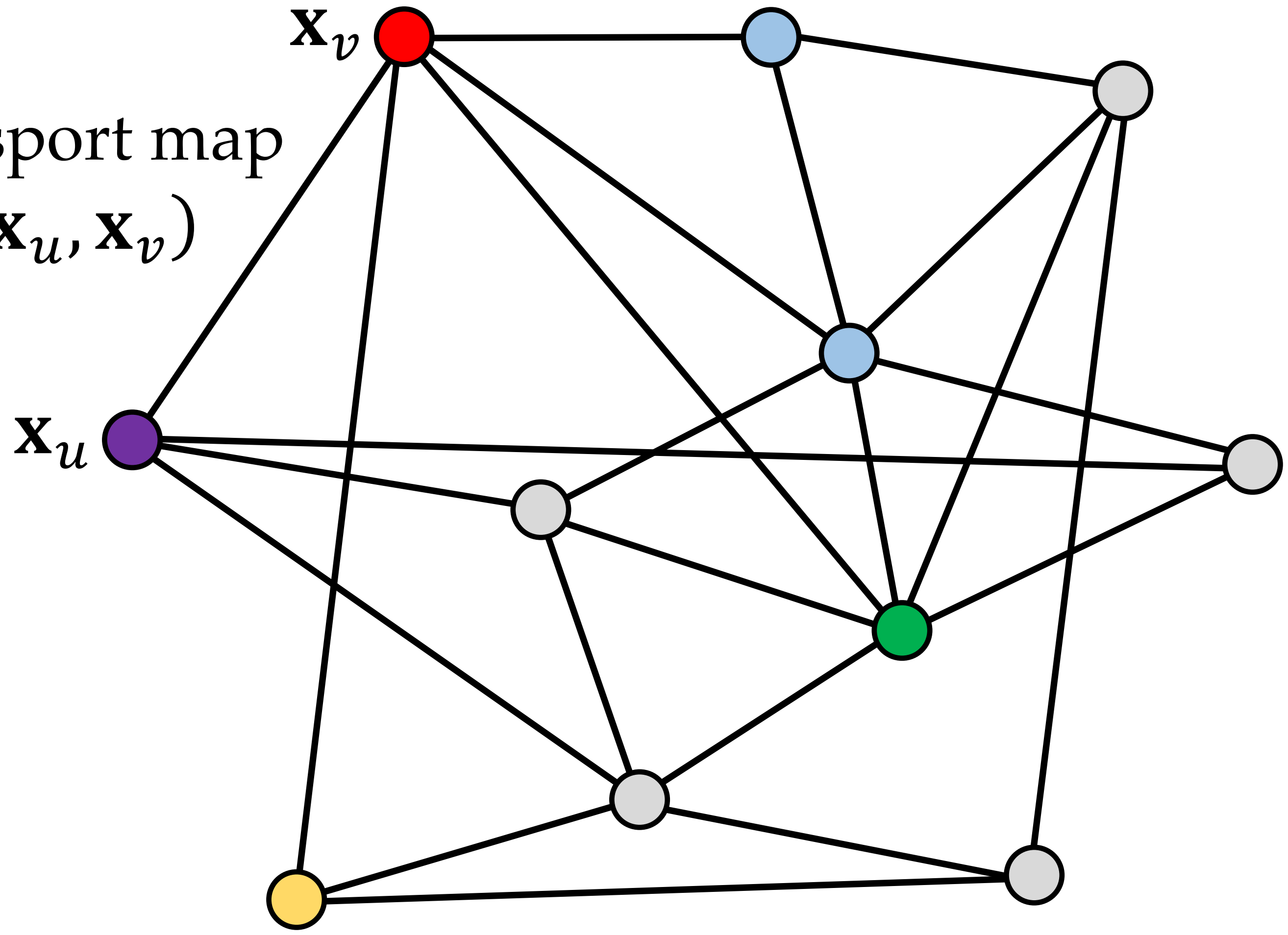
Analogy to parallel transport
on manifolds

Endow graph with "geometry" leading to richer diffusion with better separation, ability to cope with heterophily, and no oversmoothing

Bodnar, di Giovanni, et B. 2022

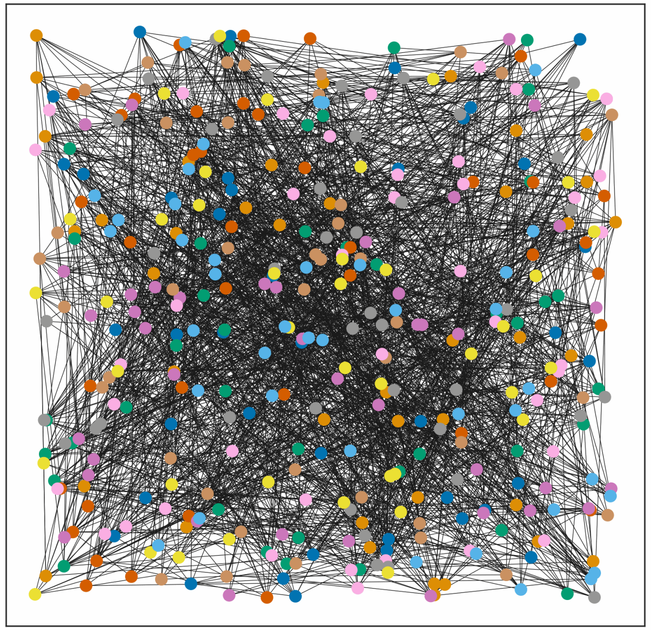
Cellular Sheaves

learnable transport map
$\mathcal{F}_{u \trianglelefteq v} = \mathbf{\Phi}(\mathbf{x}_u, \mathbf{x}_v)$



Sheaf Laplacian: $(\Delta_{\mathcal{F}}\mathbf{x})_v = \sum_{v,u \trianglelefteq e} \mathcal{F}_{v \trianglelefteq e}^{\mathrm{T}} (\mathcal{F}_{v \trianglelefteq e}\mathbf{x}_v - \mathcal{F}_{u \trianglelefteq e}\mathbf{x}_u)$

Diffusion on Cellular Sheaves



$$\dot{\mathbf{X}}(t) = -\Delta_{\mathcal{F}}\mathbf{X}(t) \qquad \text{with i.c. } \mathbf{X}(0) = \mathbf{X}$$

Node classification = limit of sheaf diffusion equation with an appropriate sheaf, alternative to WL

Bodnar, di Giovanni, et B. 2022

Alternative to Weisfeiler-Lehman?

Proposition 7. *Let $\mathcal{G}$ be the set of connected graphs $G = (V, E)$ with two classes $A, B \subset V$ such that for each $v \in A$, there exists $u \in A$ and an edge $(v, u) \in E$. Then*

$$\mathcal{H}^d_{\text{sym}} := \{(\mathcal{F}, G) \mid \mathcal{F}_{v \trianglelefteq e} = \mathcal{F}_{u \trianglelefteq e},\ \det(\mathcal{F}_{v \trianglelefteq e}) \neq 0\}$$

has linear separation power over $\mathcal{G}$.

GCN-type architectures (symmetric scalar transport maps) can achieve separation in **homophilic** setting

Bodnar, di Giovanni, et B. 2022

Alternative to Weisfeiler-Lehman?

Proposition 7. *Let $\mathcal{G}$ be the set of connected graphs $G = (V, E)$ with two classes $A, B \subset V$ such that for each $v \in A$, there exists $u \in A$ and an edge $(v, u) \in E$. Then*

$$\mathcal{H}^d_{\text{sym}} := \{(\mathcal{F}, G) \mid \mathcal{F}_{v \trianglelefteq e} = \mathcal{F}_{u \trianglelefteq e},\ \det(\mathcal{F}_{v \trianglelefteq e}) \neq 0\}$$

has linear separation power over $\mathcal{G}$.

Proposition 8. *Let $\mathcal{G}$ be the set of connected bipartite graphs $G = (A, B, E)$, with partitions A, B forming two classes and $|A| = |B|$. Then $\mathcal{H}^1_{\text{sym}}$ cannot linearly separate any graph in $\mathcal{G}$ for any initial conditions $\mathbf{X}(0) \in \mathbb{R}^{n \times f}$.*

GCN-type architectures (symmetric scalar transport maps) are **not powerful enough** in the **heterophilic** setting

Bodnar, di Giovanni, et B. 2022

Alternative to Weisfeiler-Lehman?

Proposition 10. *Let $\mathcal{G}$ contain all the connected graphs with two classes. Then,*

$$\mathcal{H}^d := \{(\mathcal{F}, G) \mid \det(\mathcal{F}_{v \trianglelefteq e}) \neq 0\}$$

has linear separation power over $\mathcal{G}$.

Heterophilic settings require **asymmetric** transport maps (between neighbour nodes from different classes)

Bodnar, di Giovanni, et B. 2022

Alternative to Weisfeiler-Lehman?

Proposition 10. *Let $\mathcal{G}$ contain all the connected graphs with two classes. Then,*

$$\mathcal{H}^d := \{(\mathcal{F}, G) \mid \det(\mathcal{F}_{v \trianglelefteq e}) \neq 0\}$$

has linear separation power over $\mathcal{G}$.

Heterophilic settings require **asymmetric** transport maps (between neighbour nodes from different classes)

Explains recent works using negatively-weighted edges

Bodnar, di Giovanni, et B. 2022

Alternative to Weisfeiler-Lehman?

Proposition 10. *Let $\mathcal{G}$ contain all the connected graphs with two classes. Then,*

$$\mathcal{H}^d := \{(\mathcal{F}, G) \mid \det(\mathcal{F}_{v \trianglelefteq e}) \neq 0\}$$

has linear separation power over $\mathcal{G}$.

Proposition 11. *Let G be a connected graph with $C \geq 3$ classes. Then $\mathcal{H}^1$ cannot linearly separate any $\mathbf{X} \in \mathbb{R}^{n \times f}$.*

Sheaves of higher dimension are necessary for problems with multiple classes

Bodnar, di Giovanni, et B. 2022

Alternative to Weisfeiler-Lehman?

Theorem 15. *Let $\mathcal{G}$ be the class of connected graphs with $C \leq 2\mathrm{d}$ classes. Then, for all $\mathrm{d} \in \{2, 4\}$,*

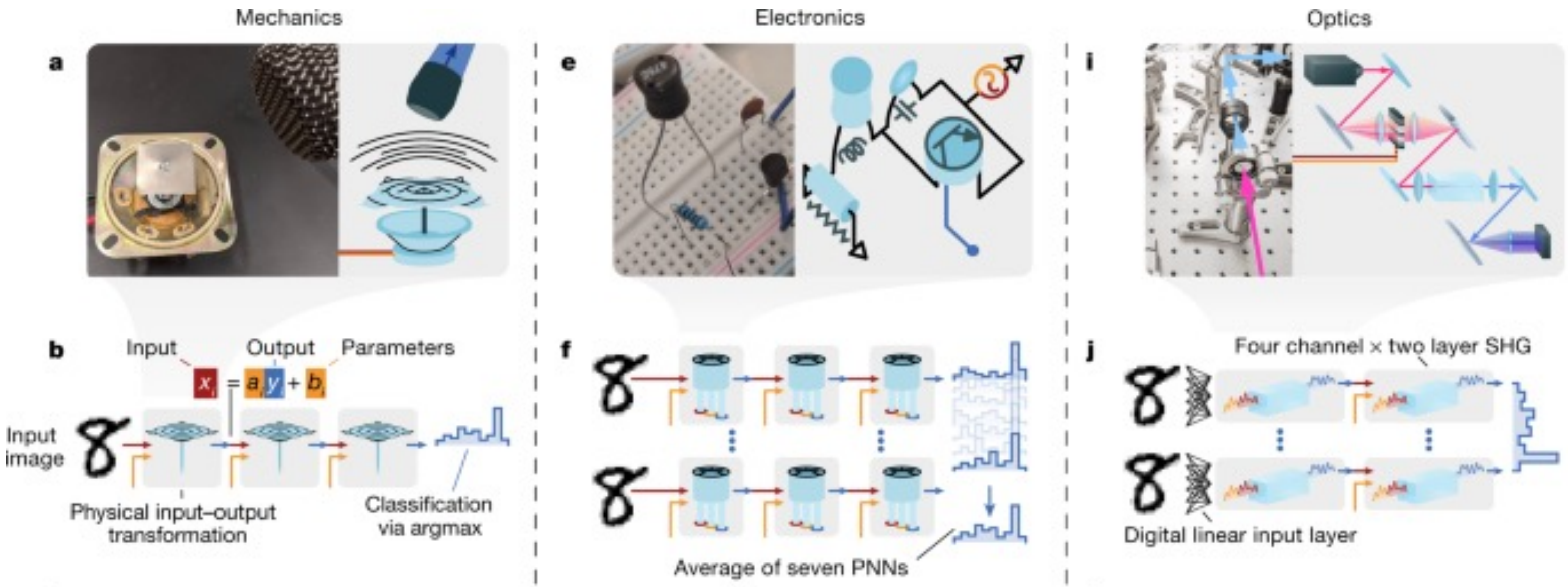
$$\mathcal{H}^{\mathrm{d}}_{\mathrm{orth}} := \{(\mathcal{F}, G) \mid \mathcal{F}_{v \trianglelefteq e} \in O(d)\}$$

has linear separation power over $\mathcal{G}$.

Bodnar, di Giovanni, et B. 2022

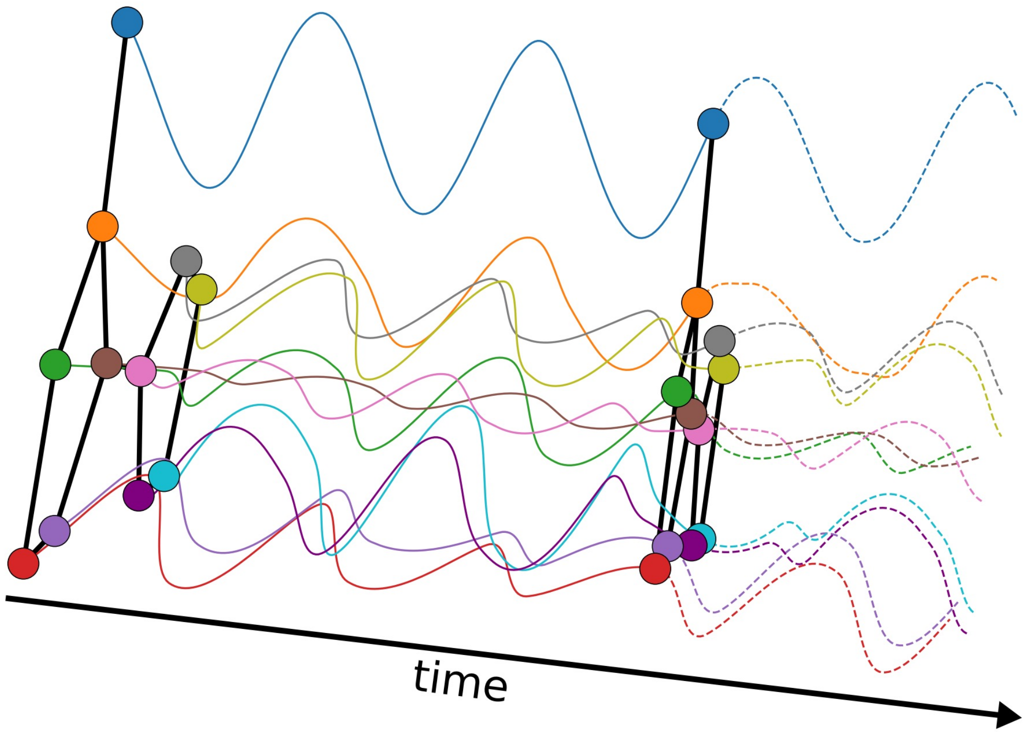
Image: Michael Galkin

Physical systems as learning metaphor



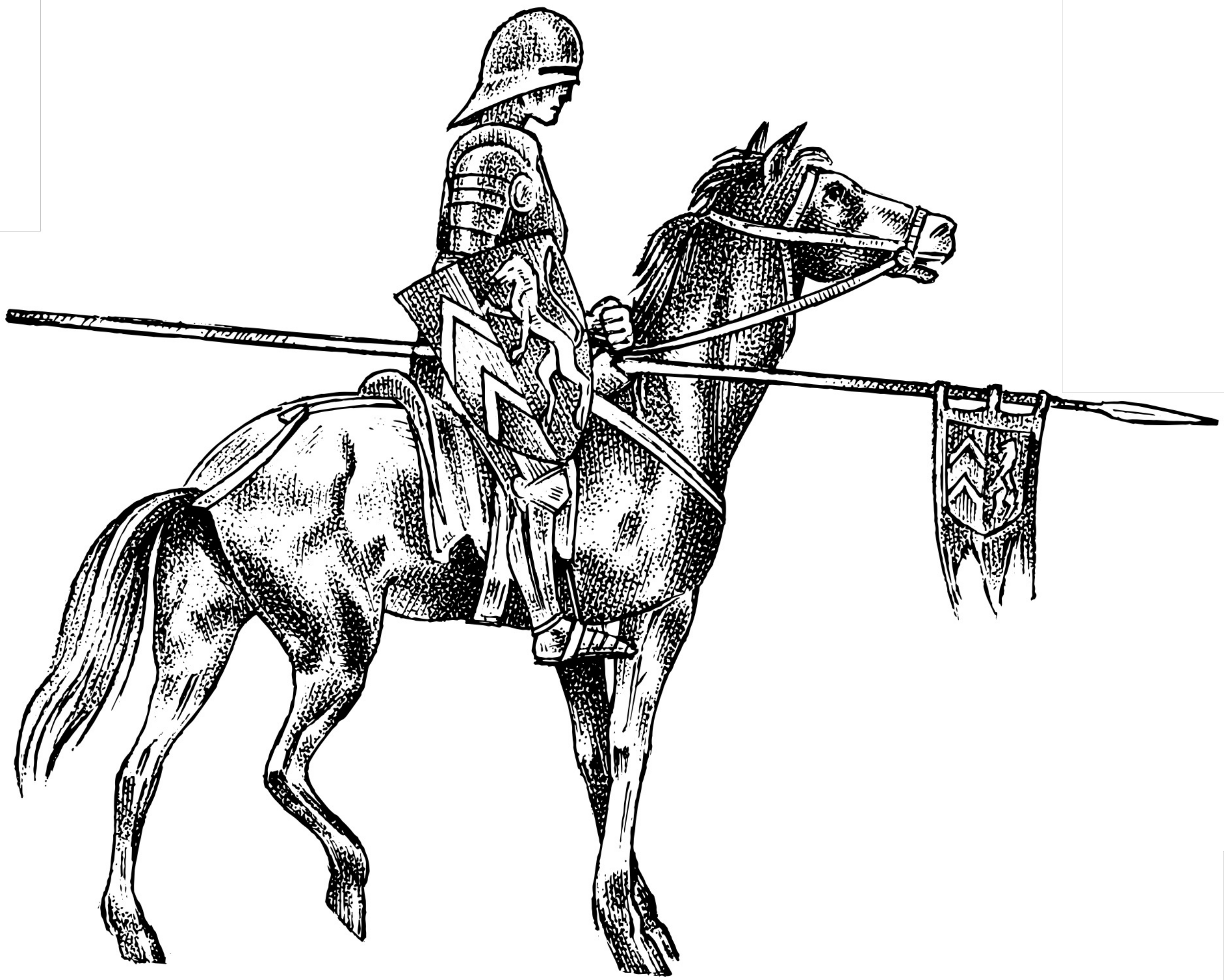
Wright et al. 2022

Graph-Coupled Oscillators



Dynamics of a system of coupled oscillators on a molecular graph

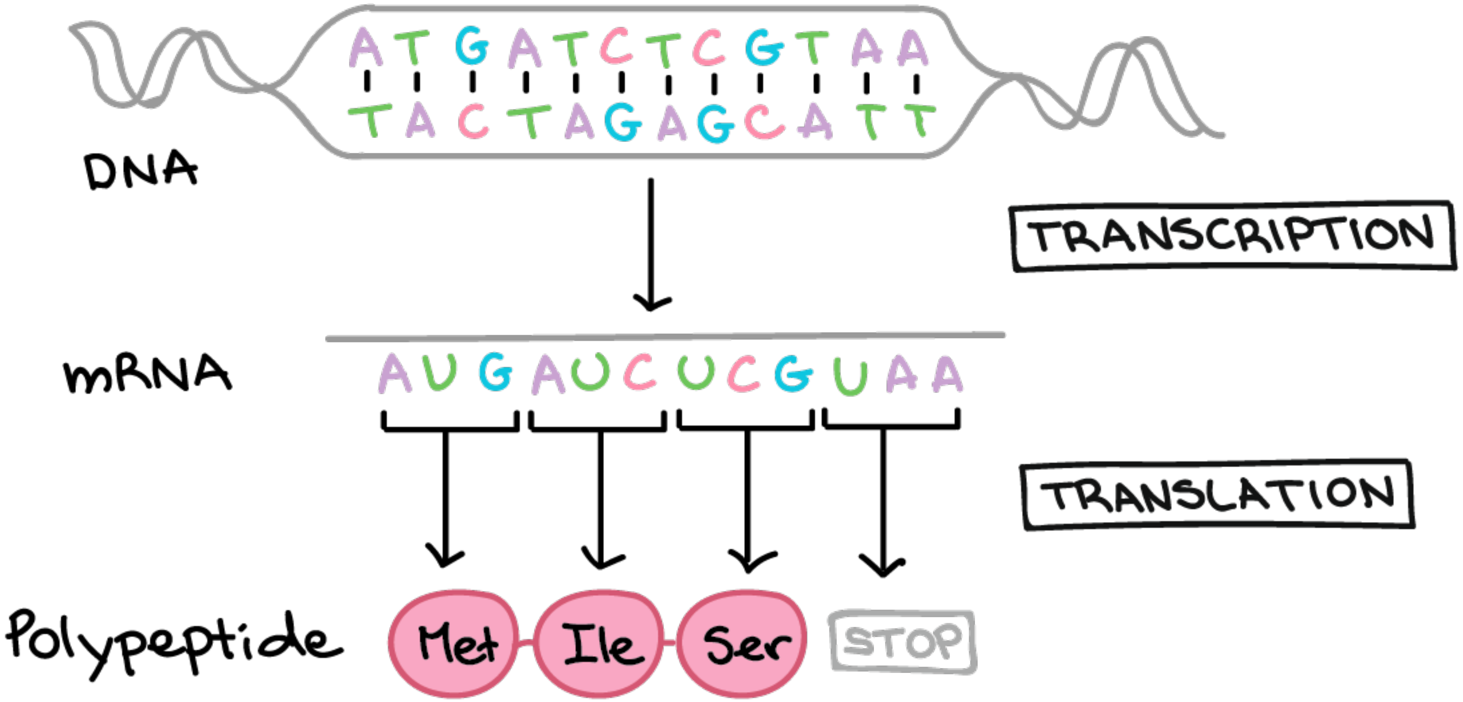
Rusch et B. 2022



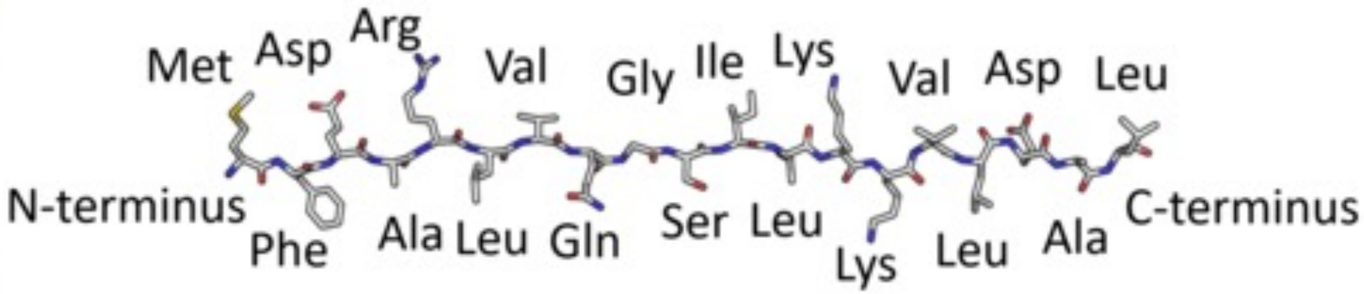
Are we done with Message Passing?

APPLICATIONS IN STRUCTURAL BIOLOGY

“Central Dogma”

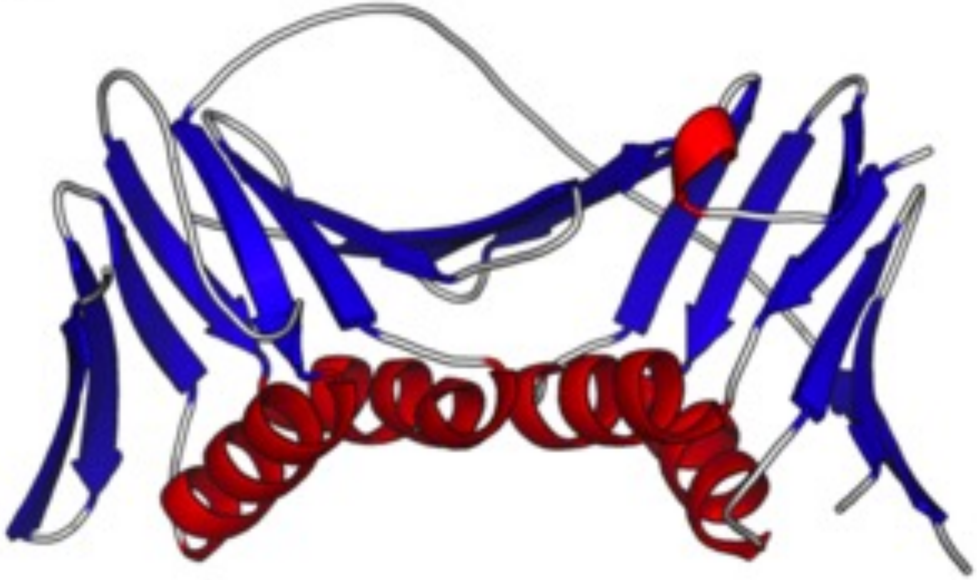


TTT Phe	TCT Ser	TAT Tyr	TGT Cys
TTC Phe	TCC Ser	TAC Tyr	TGC Cys
TTA Leu	TCA Ser	TAA stop	TGA stop
TTG Leu	TCG Ser	TAG stop	TGG Trp
CTT Leu	CCT Pro	CAT His	CGT Arg
CTC Leu	CCC Pro	CAC His	CGC Arg
CTA Leu	CCA Pro	CAA Gln	CGA Arg
CTG Leu	CCG Pro	CAG Gln	CGG Arg
ATT Ile	ACT Thr	AAT Asn	AGT Ser
ATC Ile	ACC Thr	AAC Asn	AGC Ser
ATA Ile	ACA Thr	AAA Lys	AGA Arg
ATG Met	ACG Thr	AAG Lys	AGG Arg
GTT Val	GCT Ala	GAT Asp	GGT Gly
GTC Val	GCC Ala	GAC Asp	GGC Gly
GTA Val	GCA Ala	GAA Glu	GGA Gly
GTG Val	GCG Ala	GAG Glu	GGG Gly



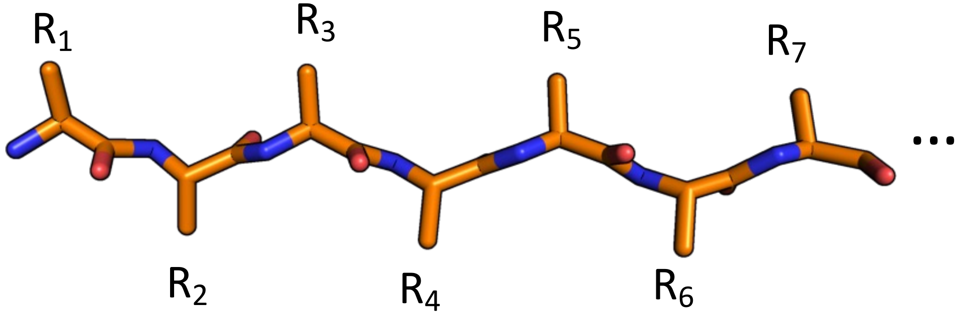
β-Sheet (3 strands)

α-helix

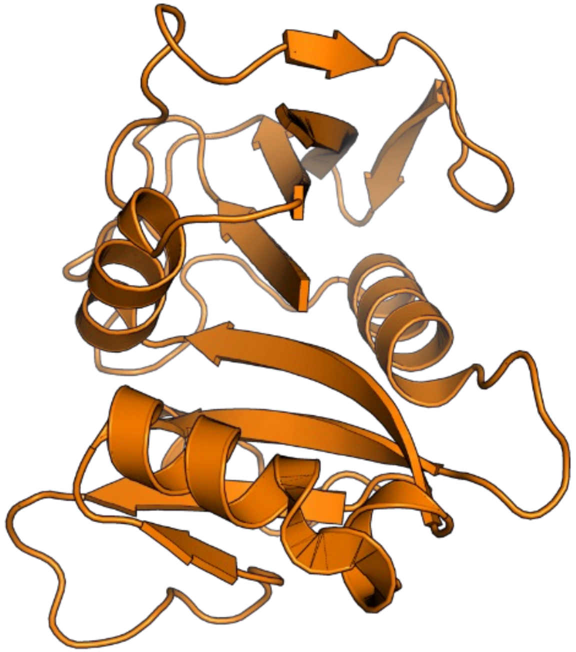


Protein Folding

Protein folding



Primary protein structure

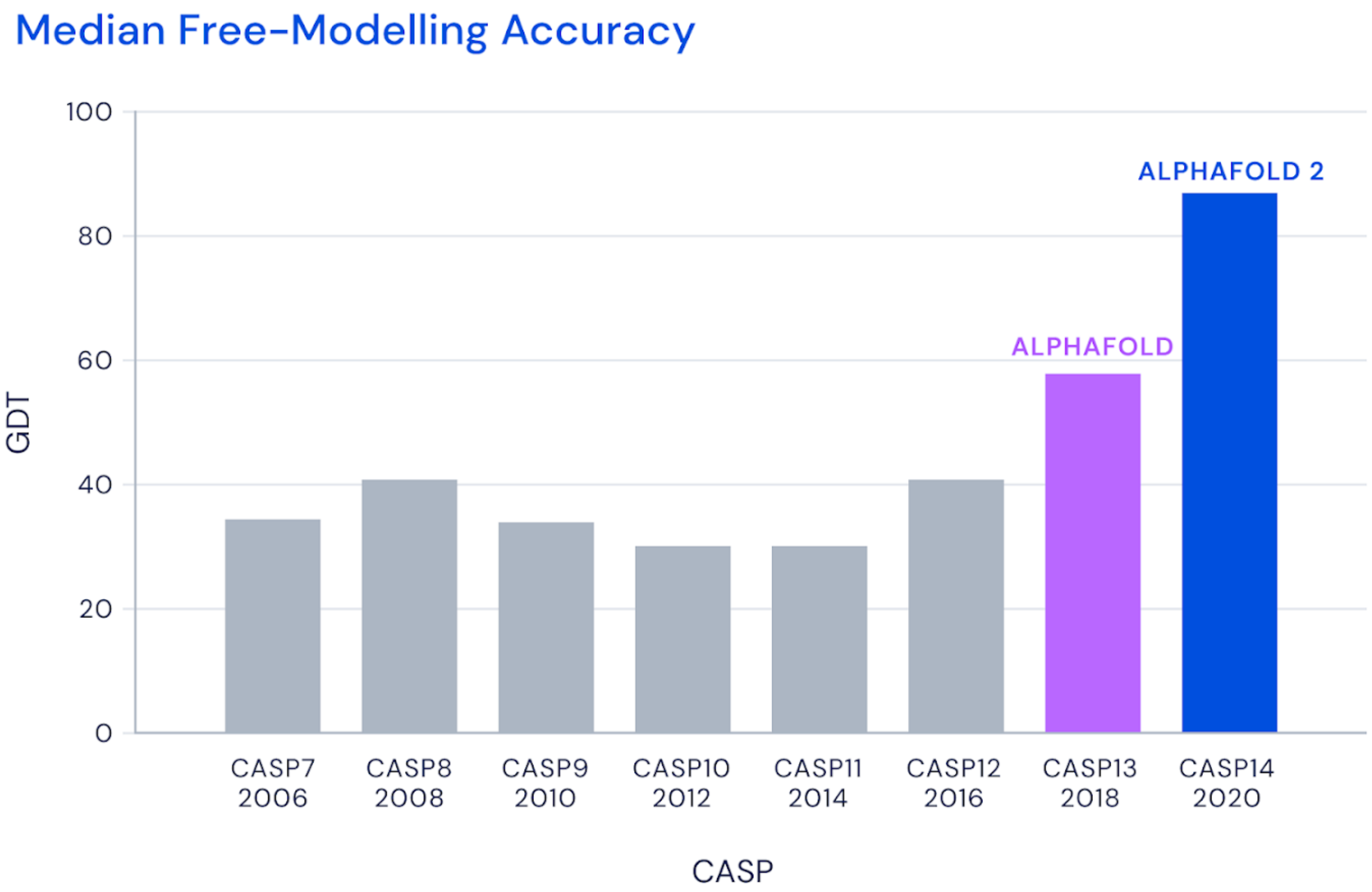


Tertiary protein structure

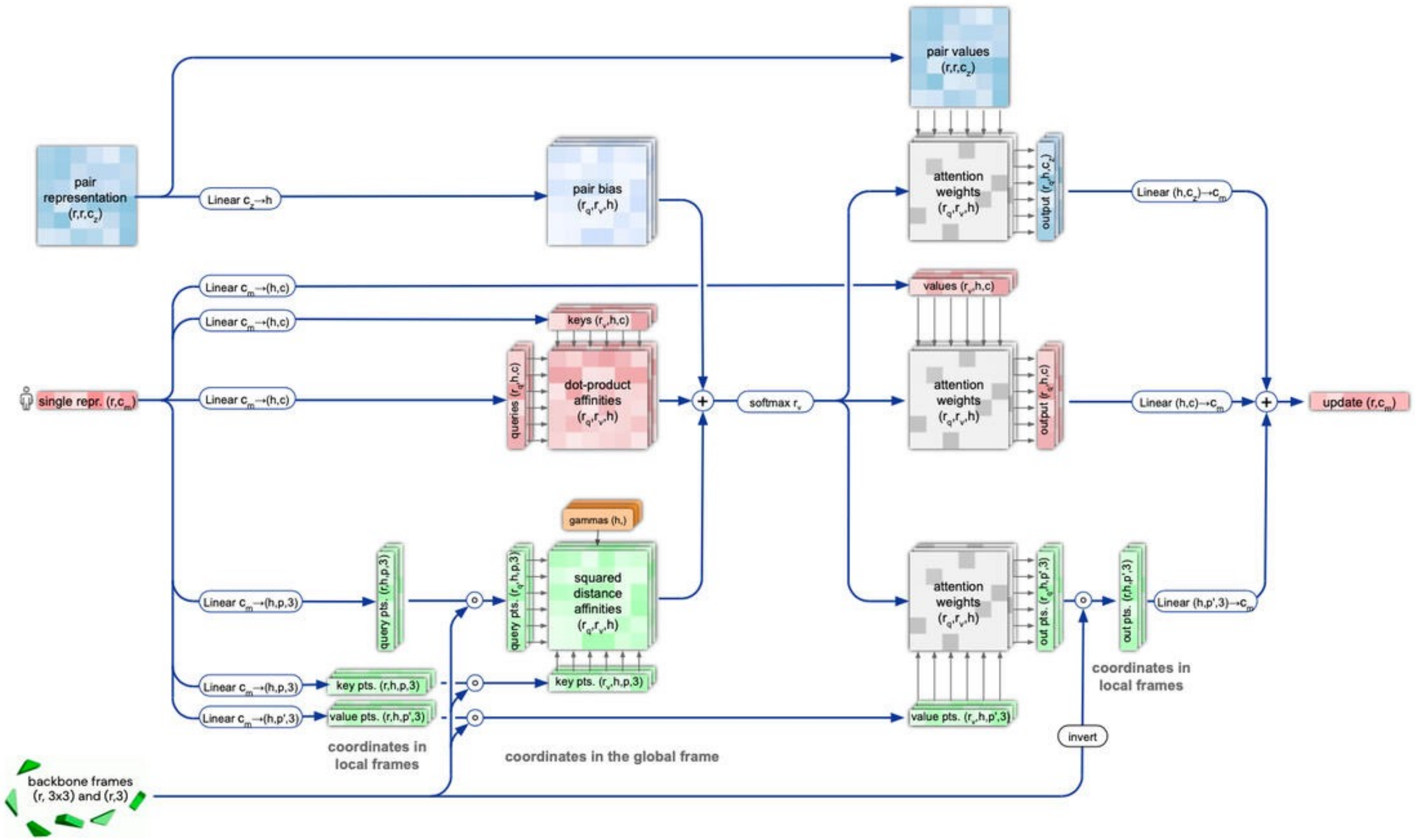
"[protein folding] is determined by […] the aminoacid sequence in a given environment"

—Christian Anfinsen (1972 Nobel Laureate in Chemistry)

“ImageNet Moment” of Structural Biology

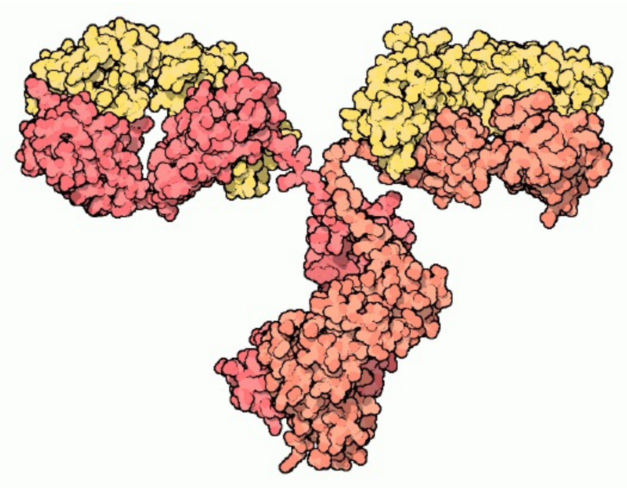


AlphaFold 2

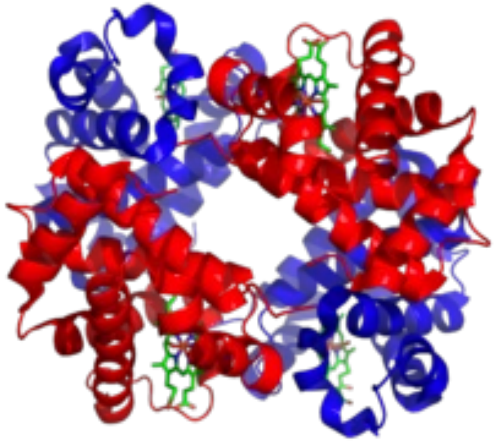


Invariant Point Attention

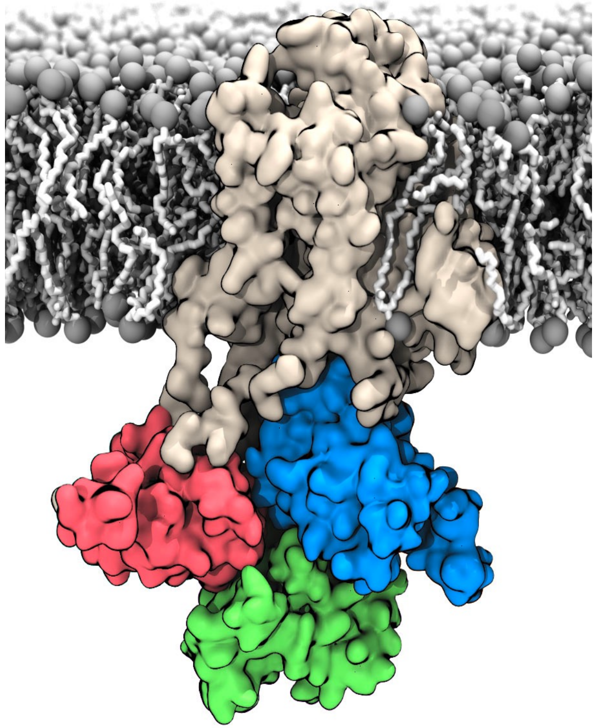
Jumper et al. 2021



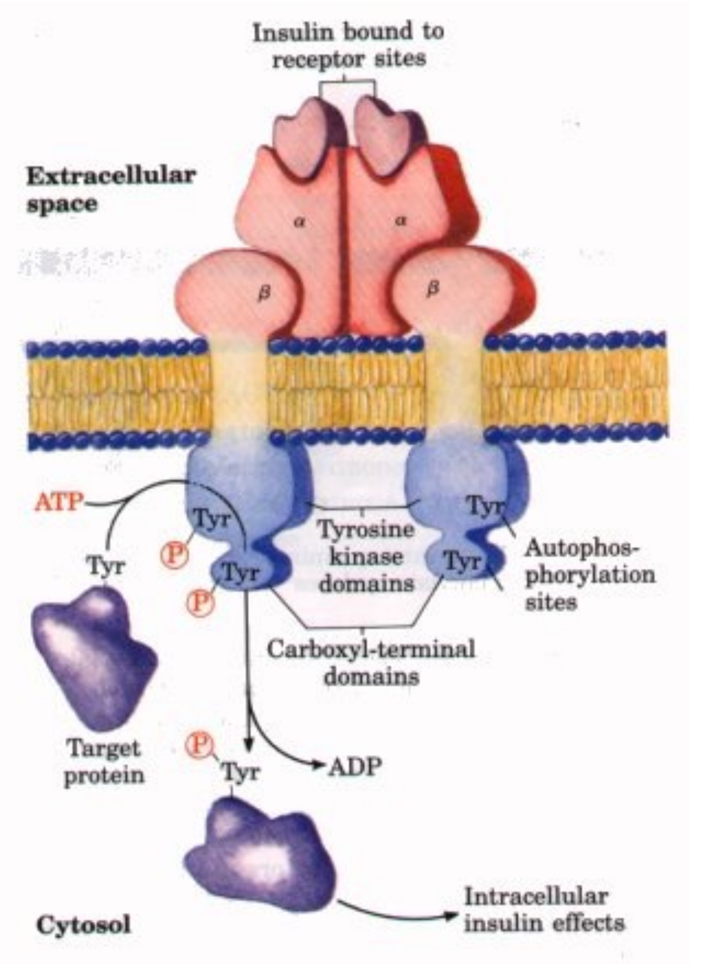
Defense (antibody)



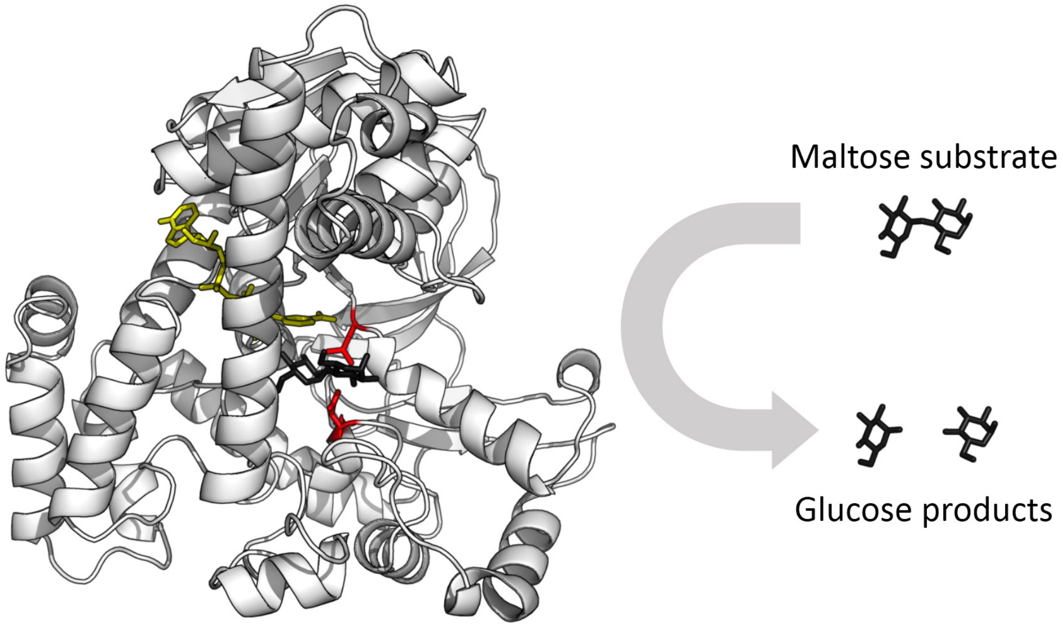
Storage (haemoglobin)



Transport (calcium pump)



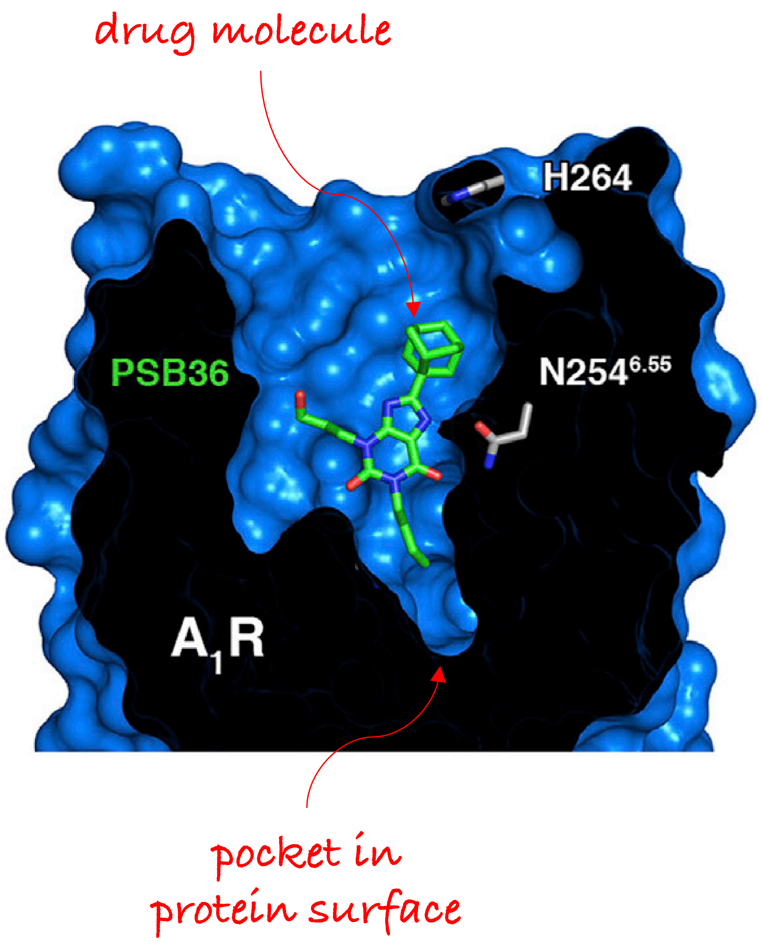
Communication (insulin)



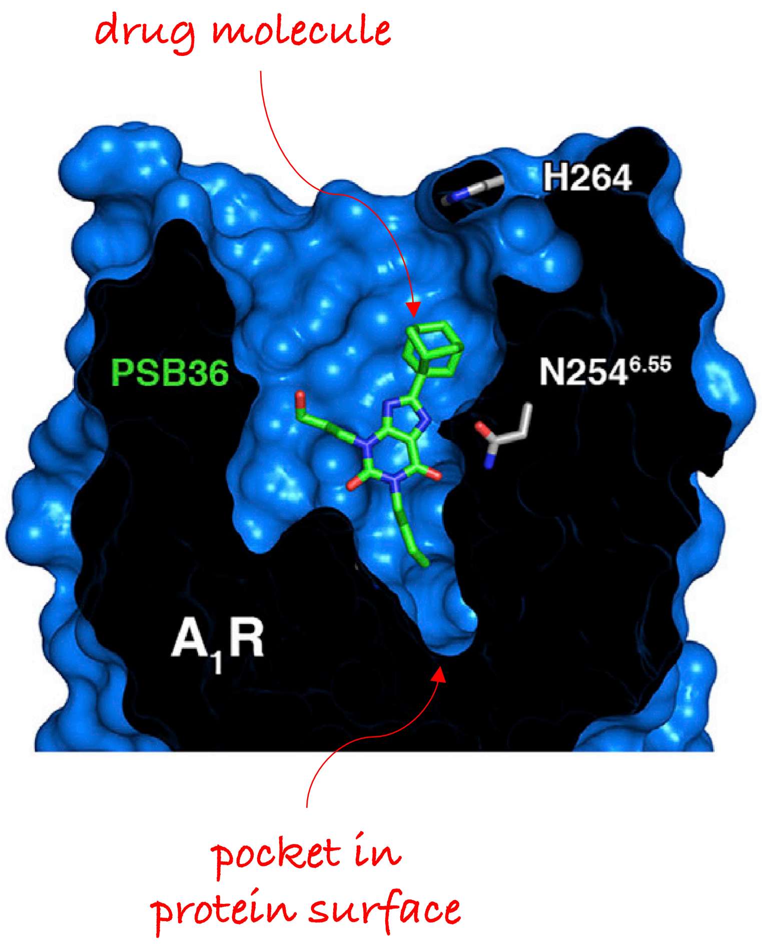
Catalysis (enzyme)

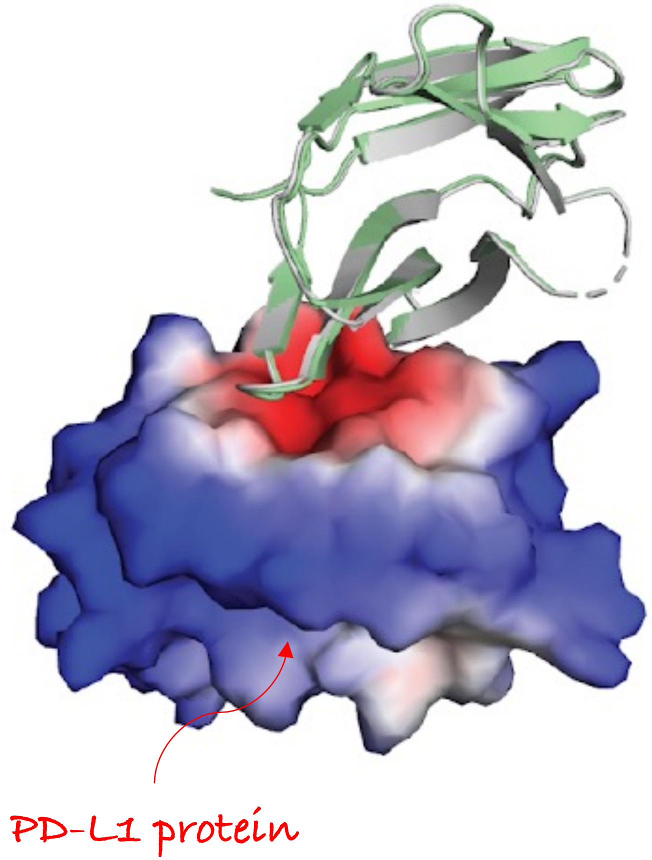
Structure (collagen)

Small Molecule Drugs

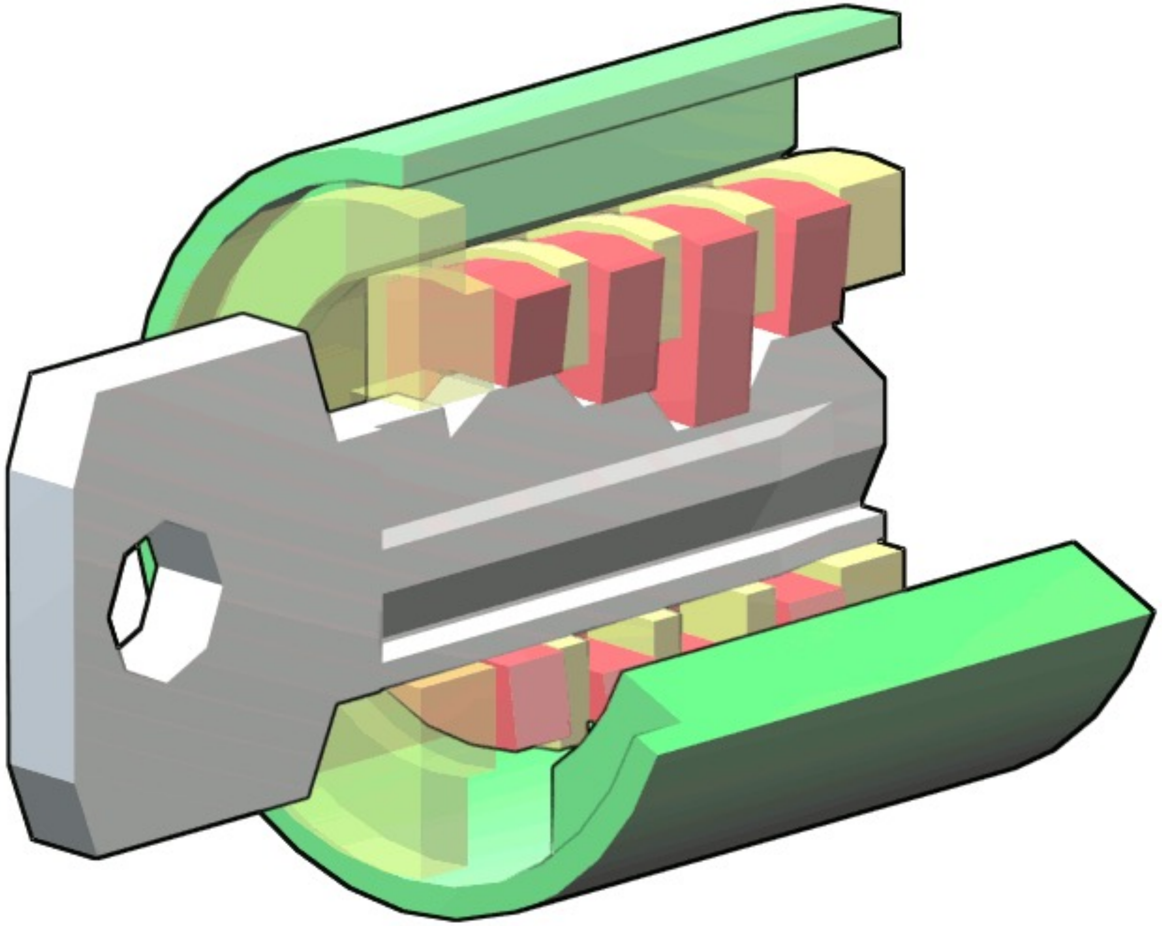


Protein-Protein Interactions





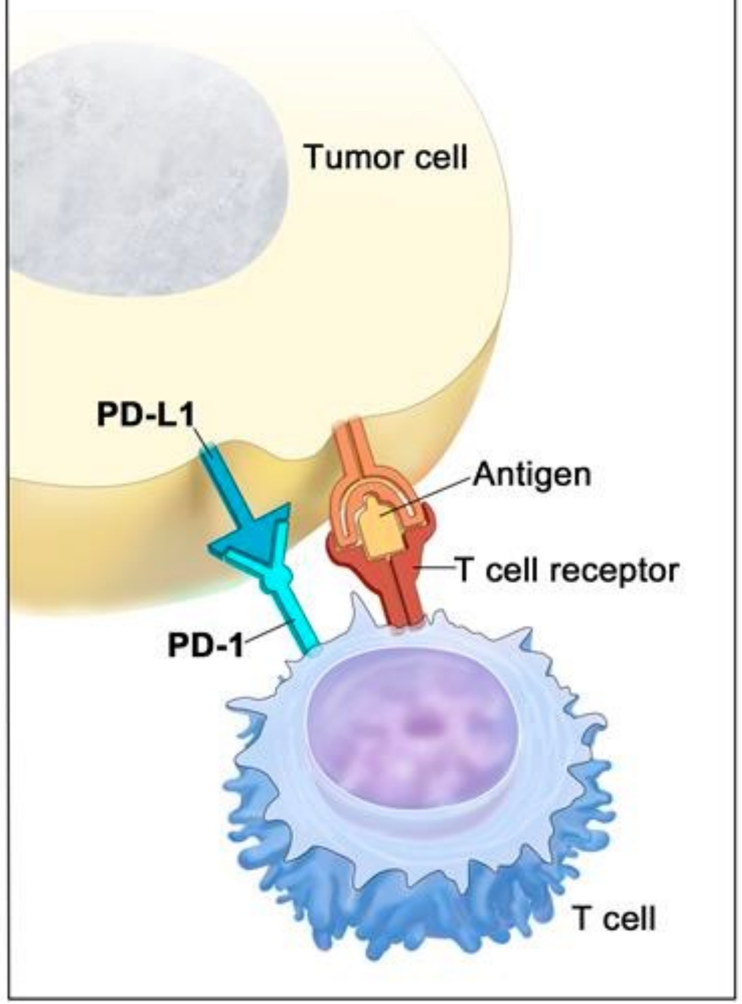
Lock-Key Metaphor



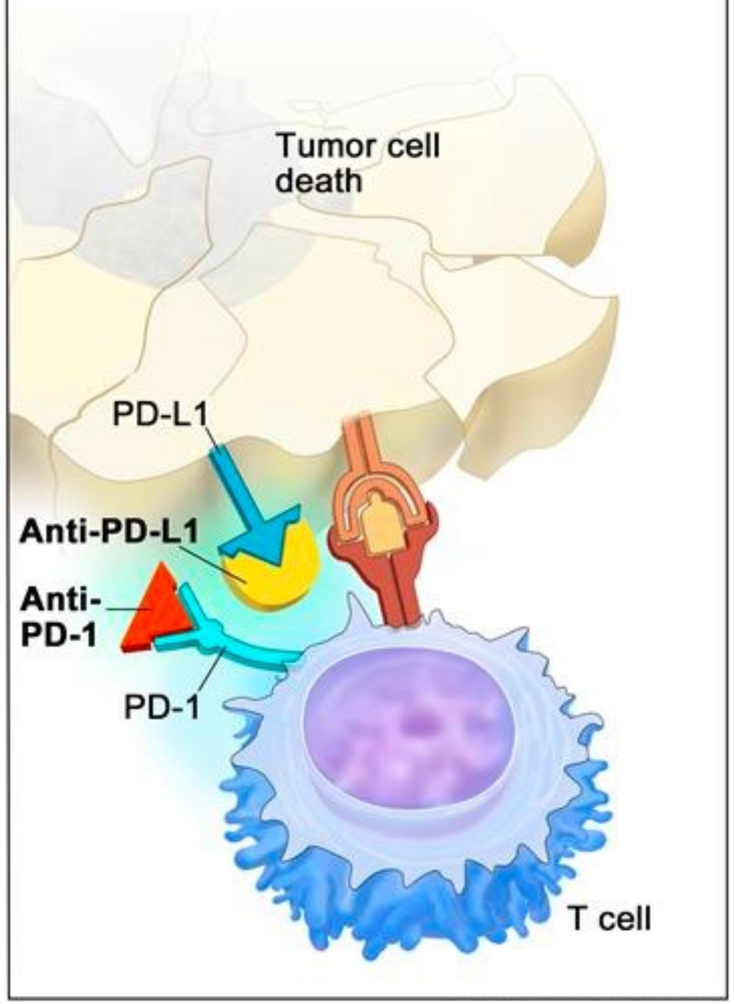
Emil Fischer “Schlüssel-Schloss-Prinzip” 1894

2018 Nobel Prize
PD-proteins role in
immunotherapy

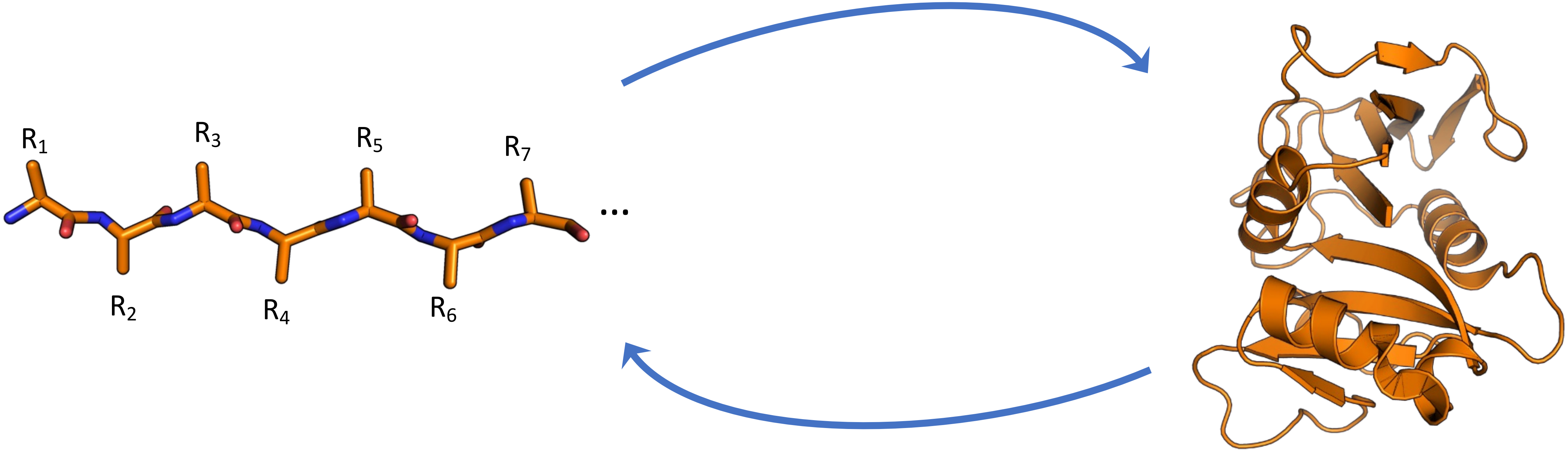
PD-L1 binds to PD-1 and inhibits
T-cell killing of tumor cell



Blocking PD-L1 or PD-1 allows
T-cell killing of tumor cell

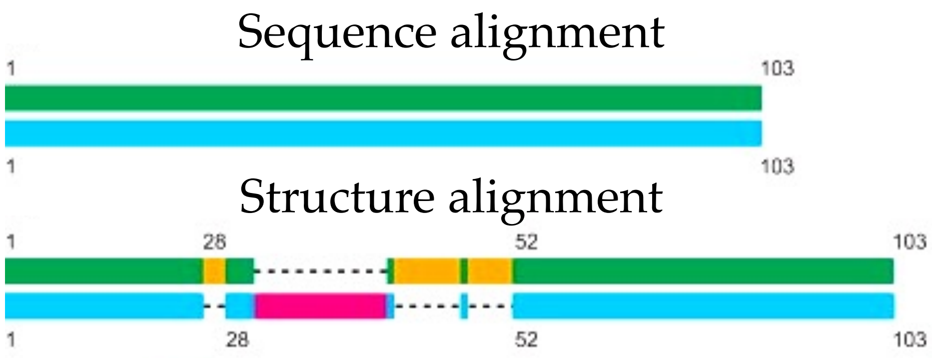


Protein Design = "Inverse Folding"



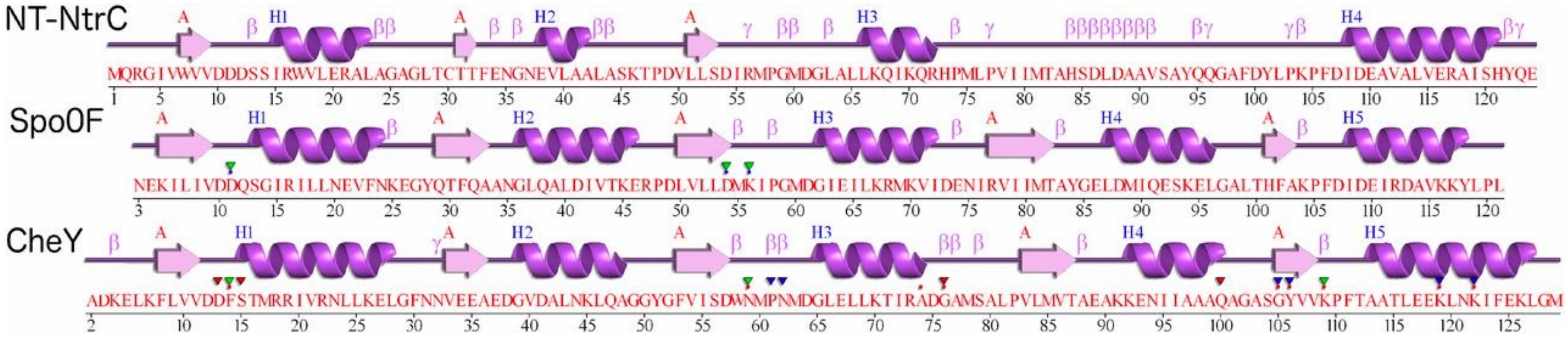
Sequence → Structure → Function

Sequence vs Structure



Similar sequence, dissimilar structure

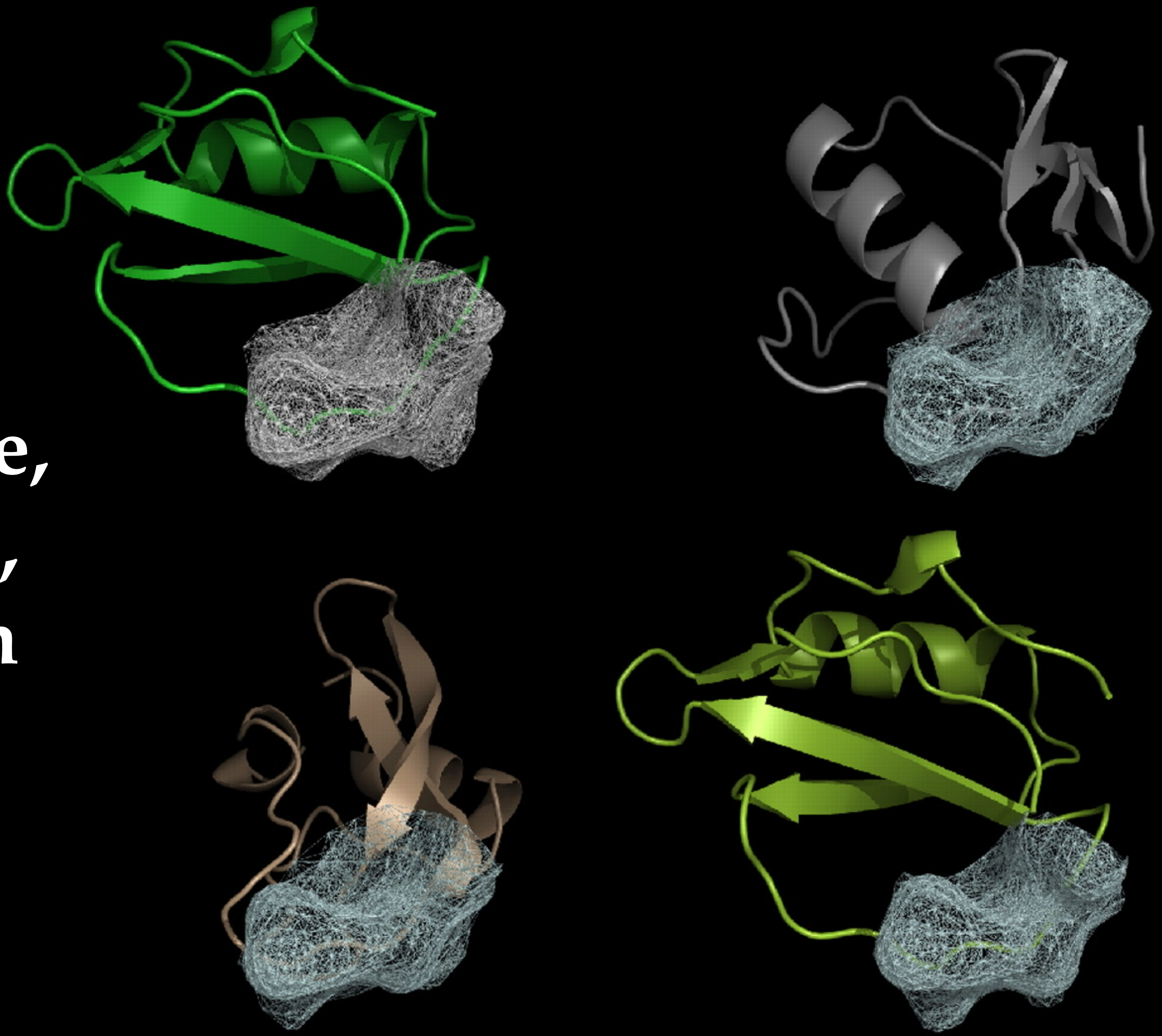
Kosloff, Kolodny 2008



He et al. 2017

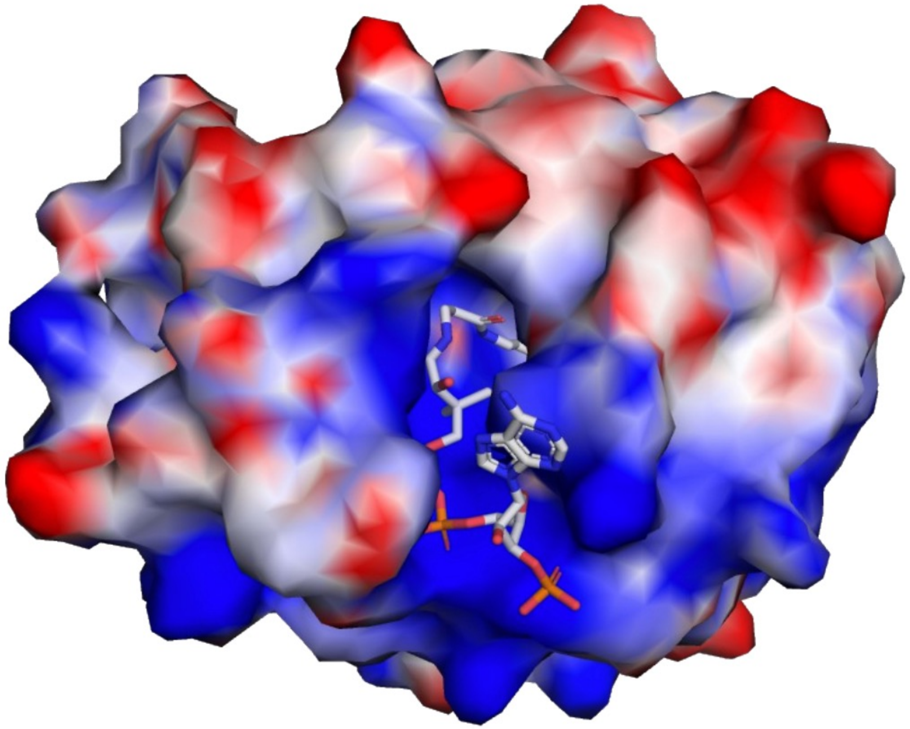
Dissimilar sequence, similar structure

Dissimilar sequence, dissimilar structure, but similar function



Yin et al. 2009

Why surface?

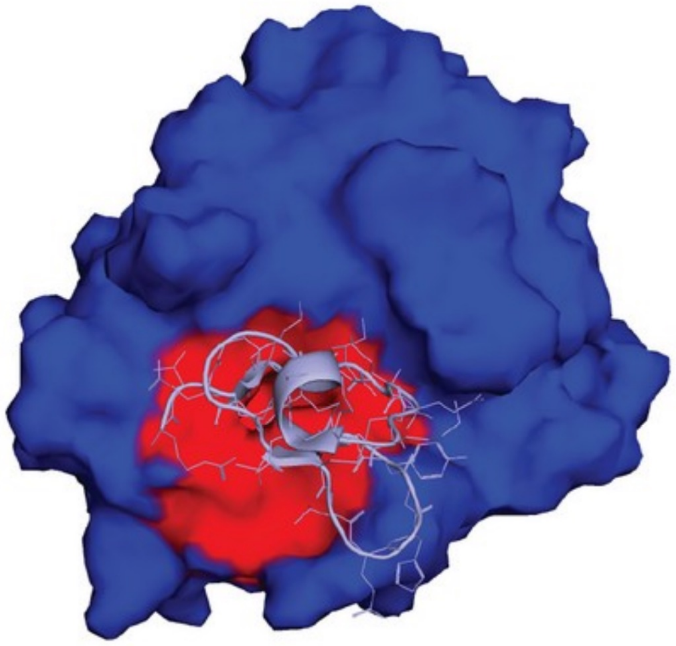


Contains interaction patterns
(geometric/chemical)

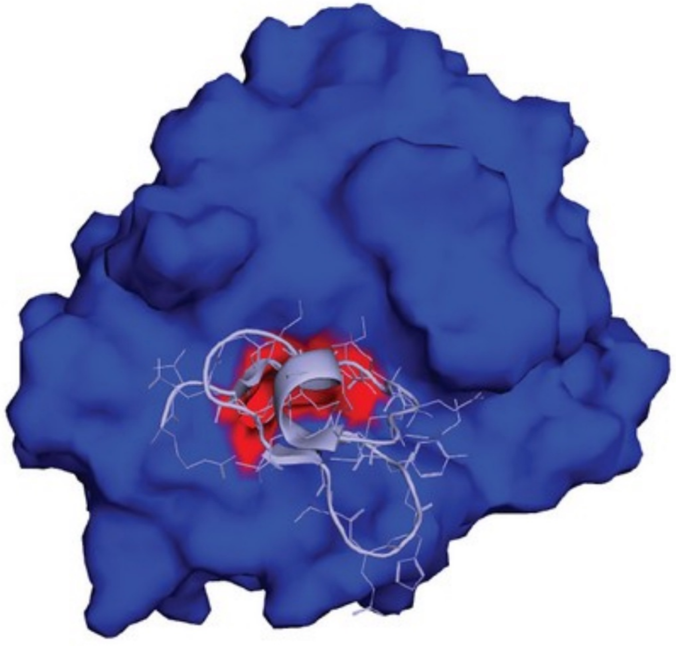
Abstracts out internal structure

Why surface?

Euclidean ball (R=12Å)

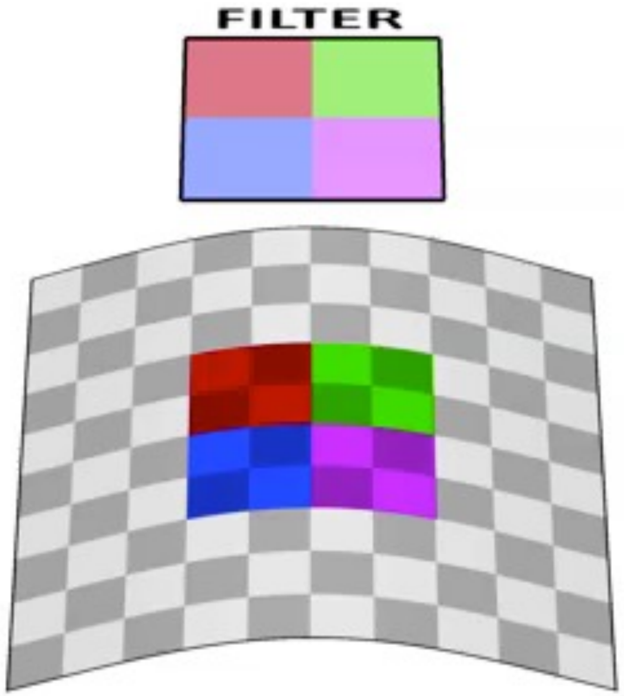


Geodesic ball (R=12Å)

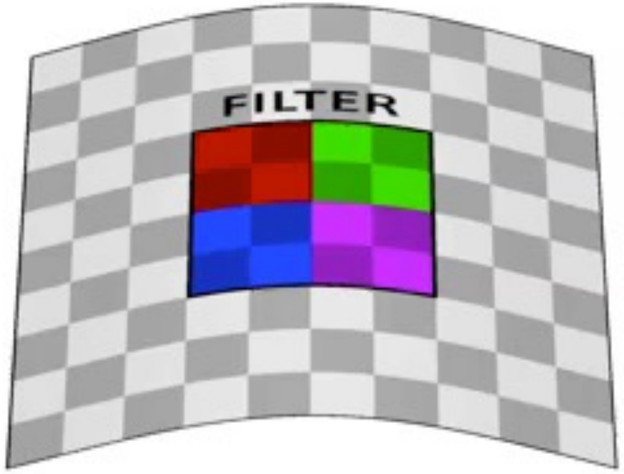


Better model of interaction interfaces

Protein surfaces are deformable



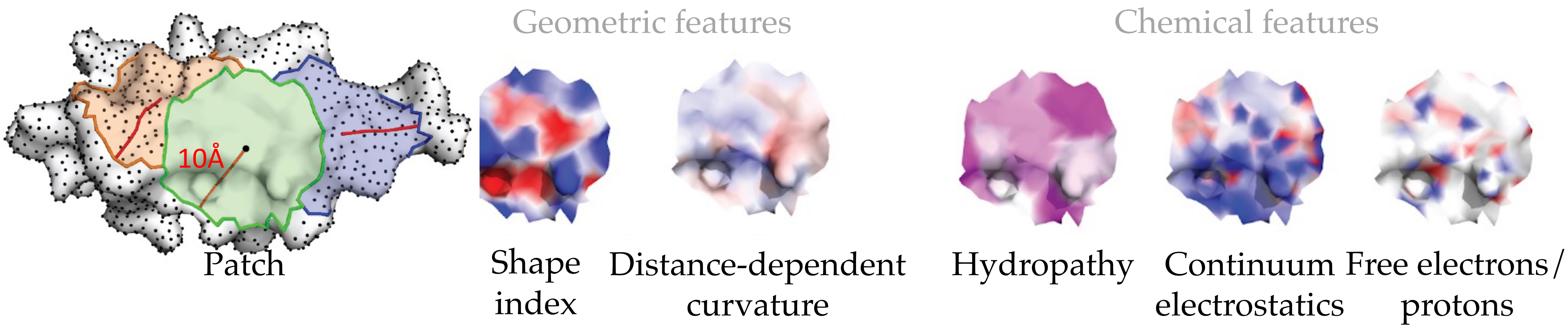
Euclidean (extrinsic) convolution



Geometric (intrinsic) convolution

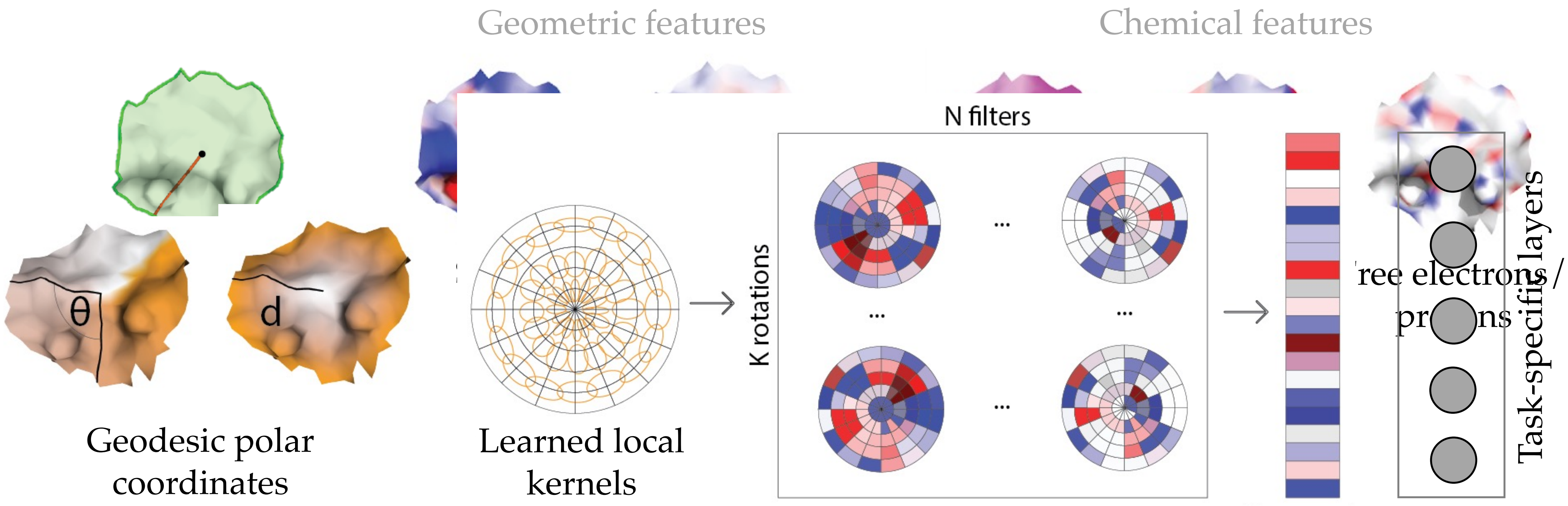
Masci et al. 2015

Molecular Surface Interaction Fingerprints (MaSIF)



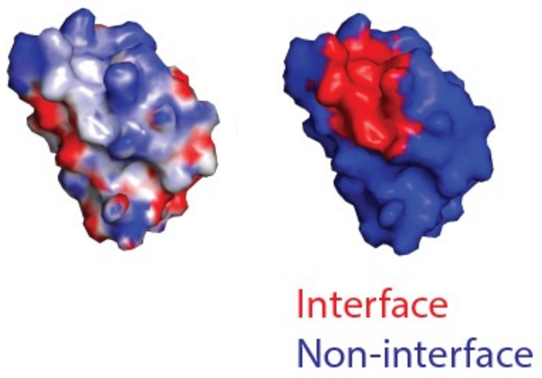
Gainza, Sverrisson et B 2020

Molecular Surface Interaction Fingerprints (MaSIF)

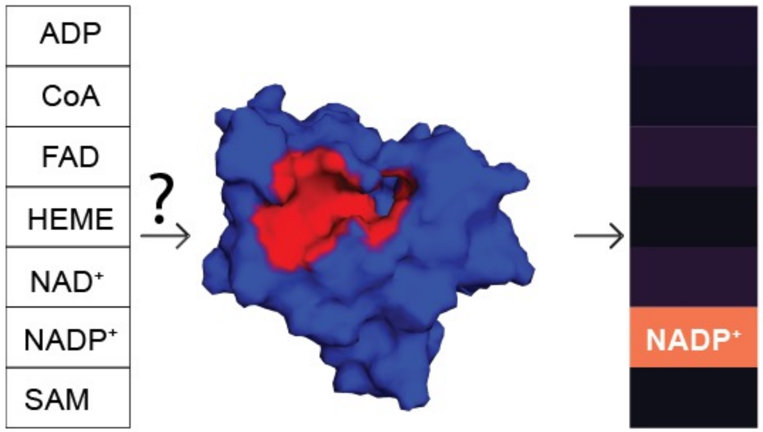


Gainza, Sverrisson et B 2020

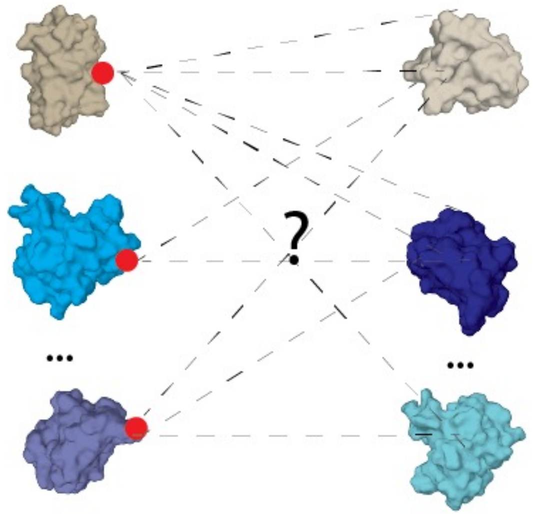
MaSIF Applications



Interface site prediction
MaSIF-site



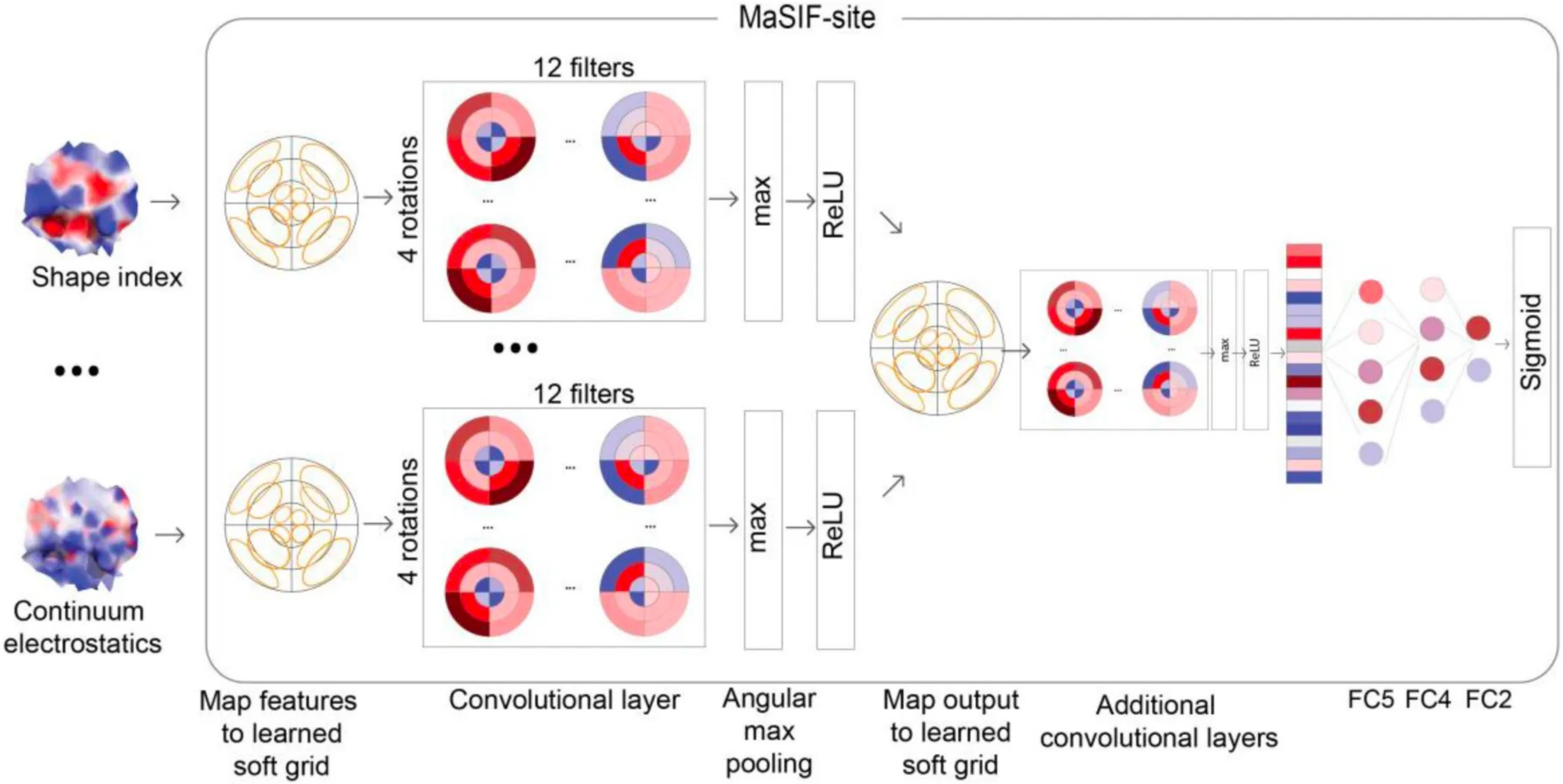
Pocket classification
MaSIF-ligand



Fast PPI search
MaSIF-search

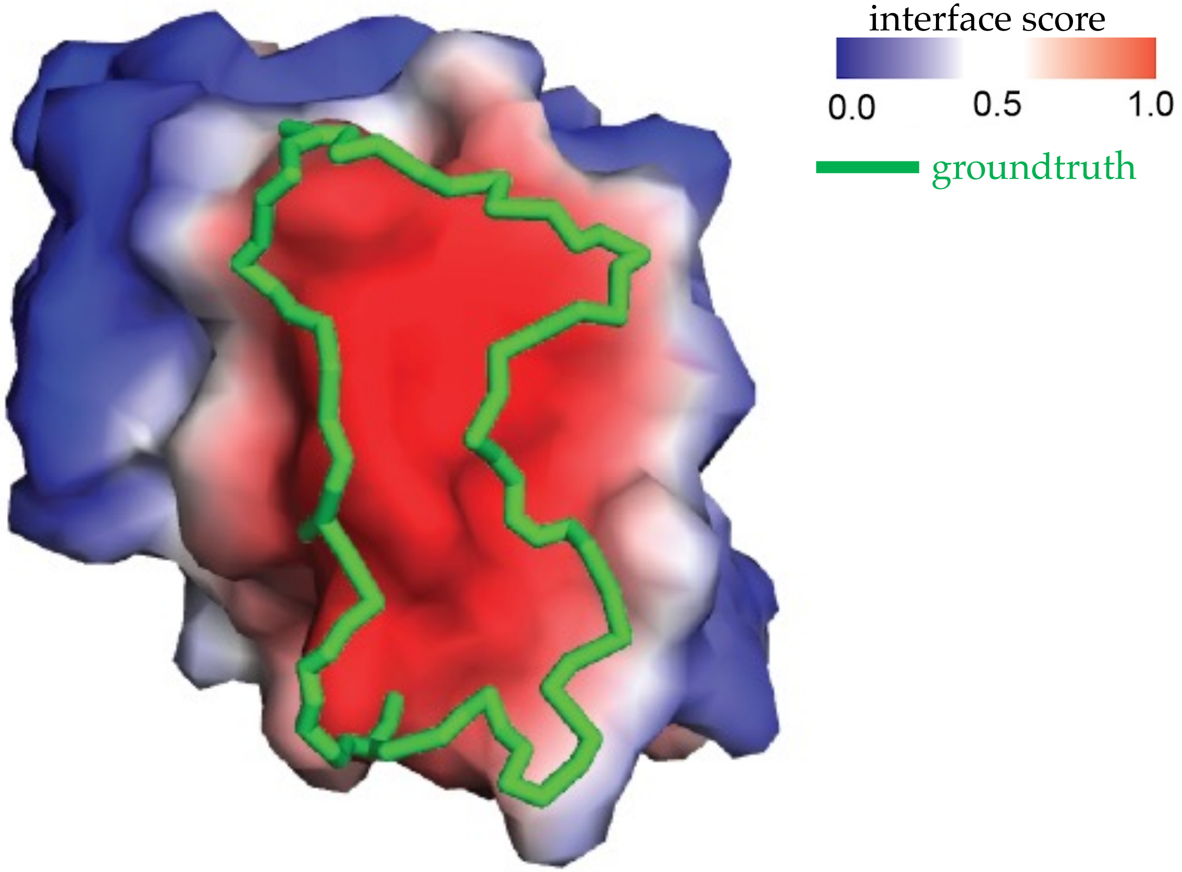
Gainza, Sverrisson et B 2020

MaSIF: Binding Site Prediction



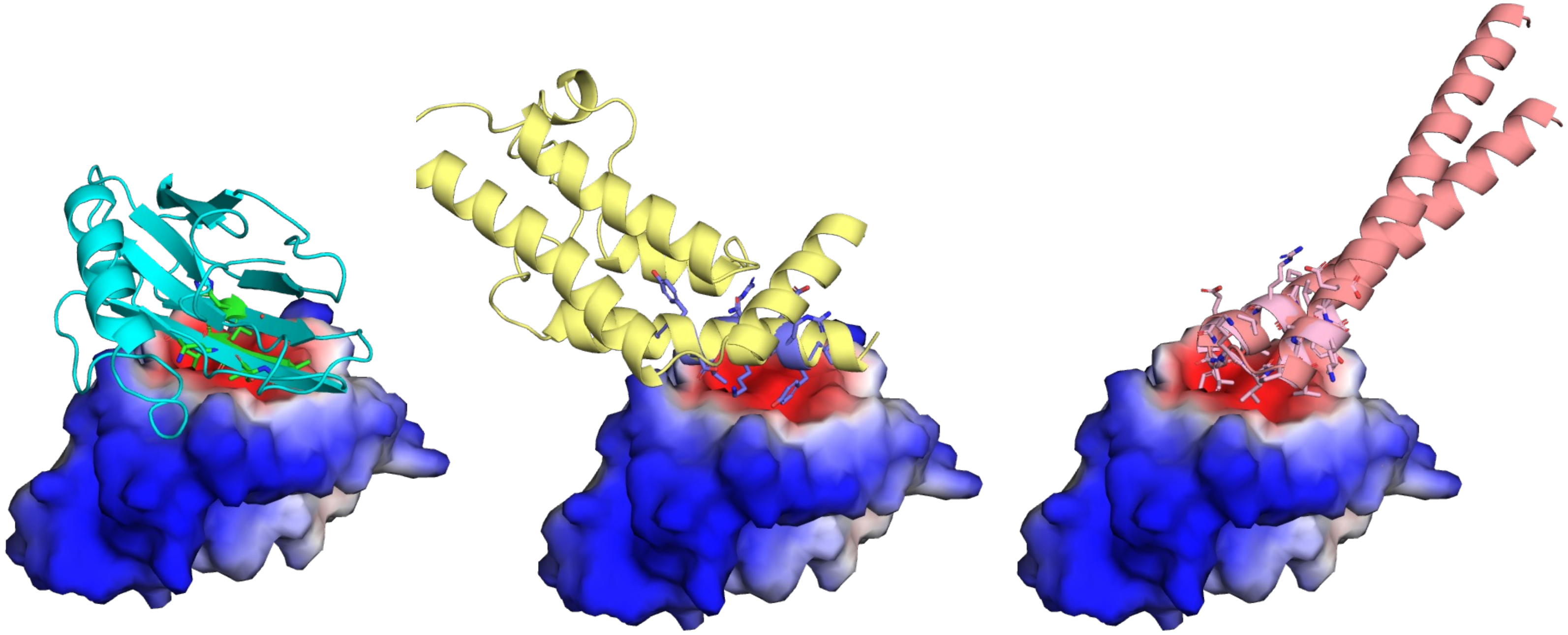
Gainza, Sverrisson et B 2020

MaSIF: Binding Site Prediction

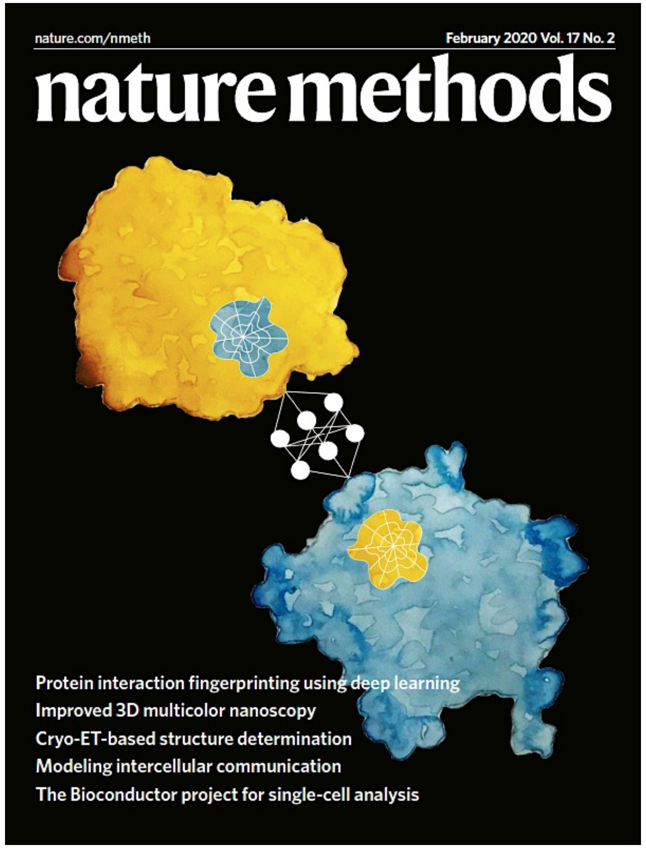


Gainza, Sverrisson et B 2020

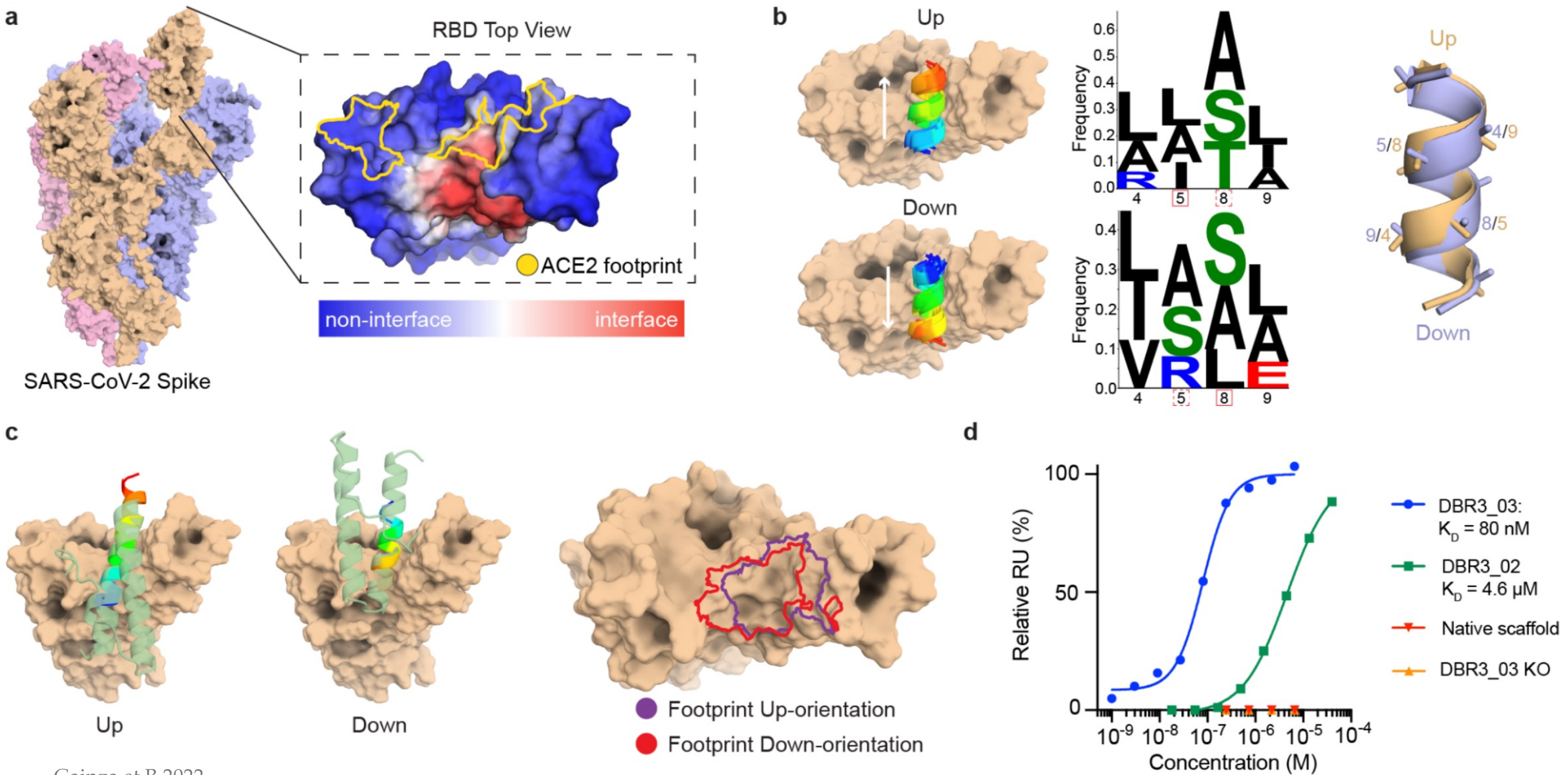
De novo *Protein Design*



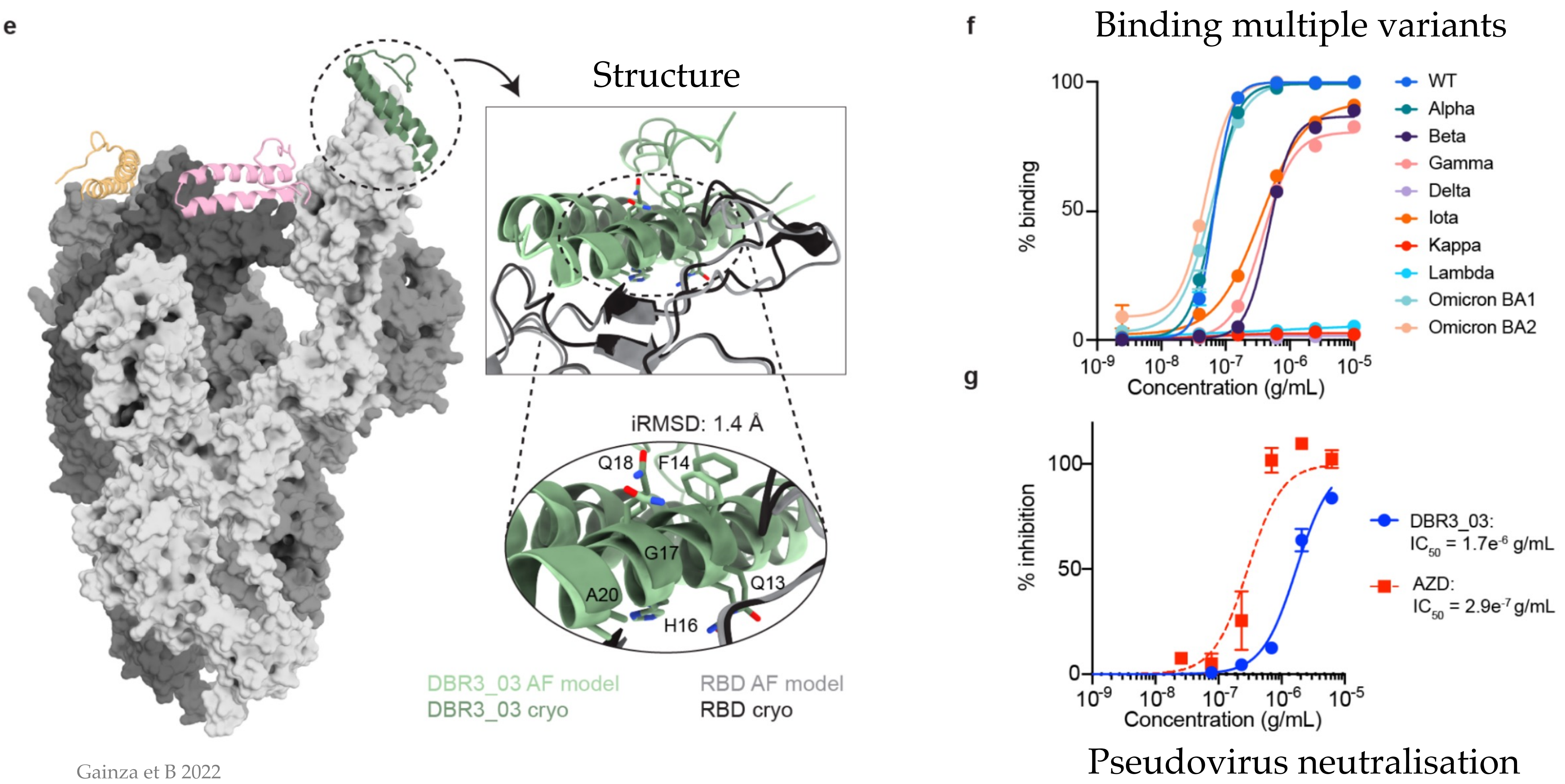
Binding site identification + designed binders
for the oncological target PD-L1



Gainza, Sverrisson et B 2020; Gainza et B 2022

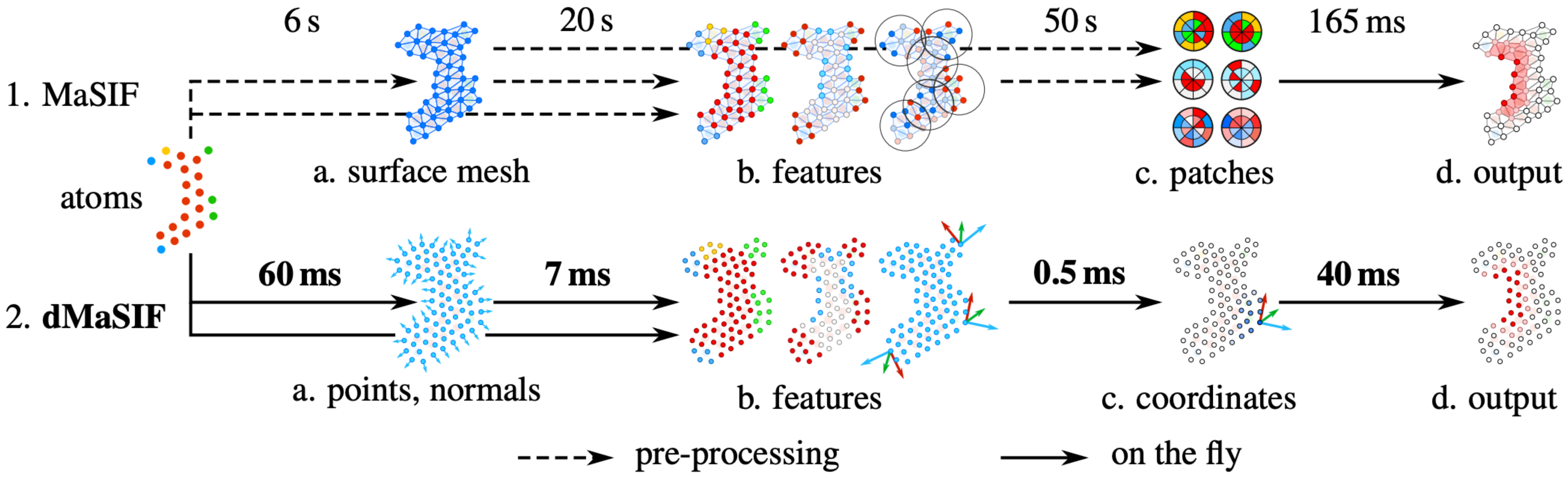


Gainza et B 2022



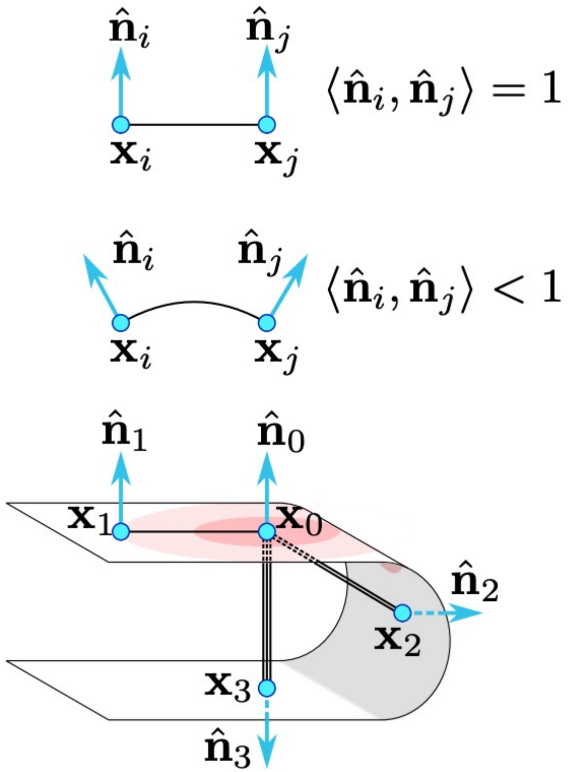
Gainza et B 2022

dMaSIF



Sverrisson, Feydy et B 2021; Feydy et B 2020

dMaSIF

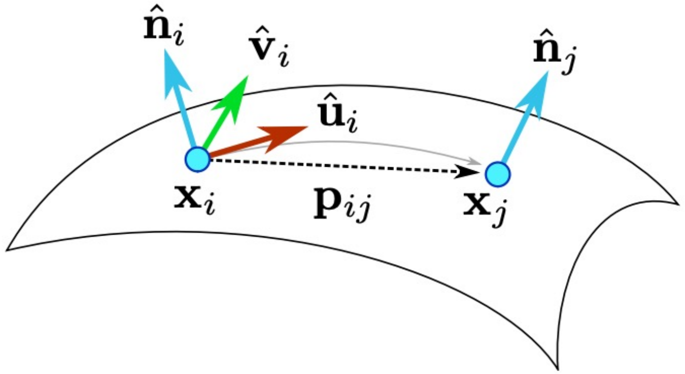


(a) Quasi-geodesic distance d_{ij}

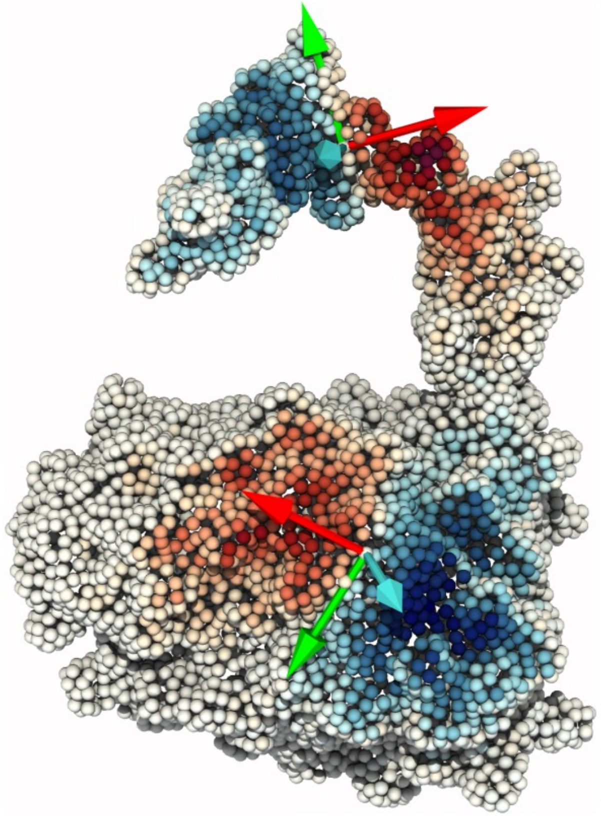
$$\mathbf{p}_{ij} = \left[\ (\mathbf{x}_j - \mathbf{x}_i)^\top\ \right] \cdot \left[\ \hat{\mathbf{n}}_i \ \middle|\ \hat{\mathbf{u}}_i \ \middle|\ \hat{\mathbf{v}}_i\ \right]$$

$$\mathbf{q}_{ij} = \left[\ (\hat{\mathbf{n}}_j - \hat{\mathbf{n}}_i)^\top\ \right] \cdot \left[\ \hat{\mathbf{n}}_i \ \middle|\ \hat{\mathbf{u}}_i \ \middle|\ \hat{\mathbf{v}}_i\ \right]$$

$$\mathrm{Conv}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{f}_j)$$
$$= \mathrm{Window}(\mathrm{d}_{ij}) \cdot \mathrm{Filter}(\mathbf{p}_{ij}\ ,\ \mathbf{q}_{ij}) \cdot \mathbf{f}_j$$



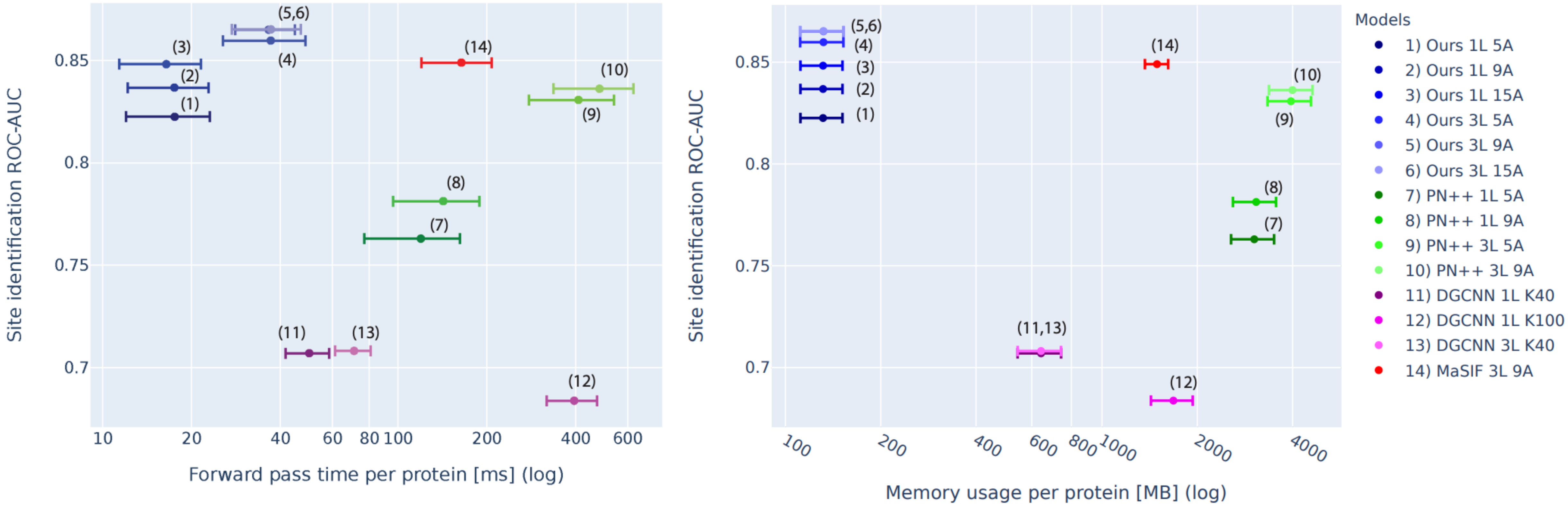
(b) Quasi-geodesic convolution



(c) $\mathrm{Filter}(\mathbf{p}_{ij}) = \mathbf{p}_{ij}^{\hat{\mathbf{u}}}$.

Sverrisson, Feydy et B 2021

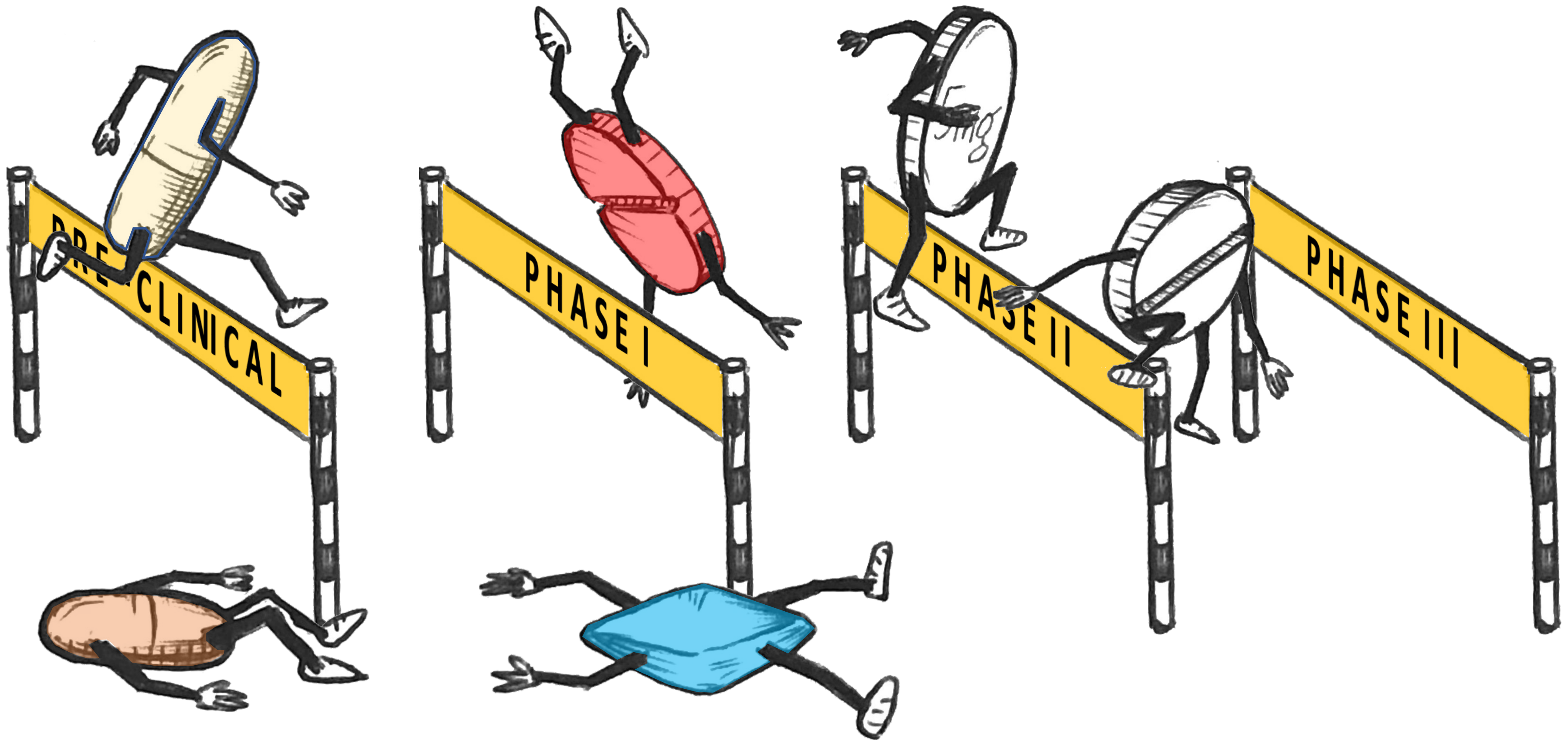
dMaSIF



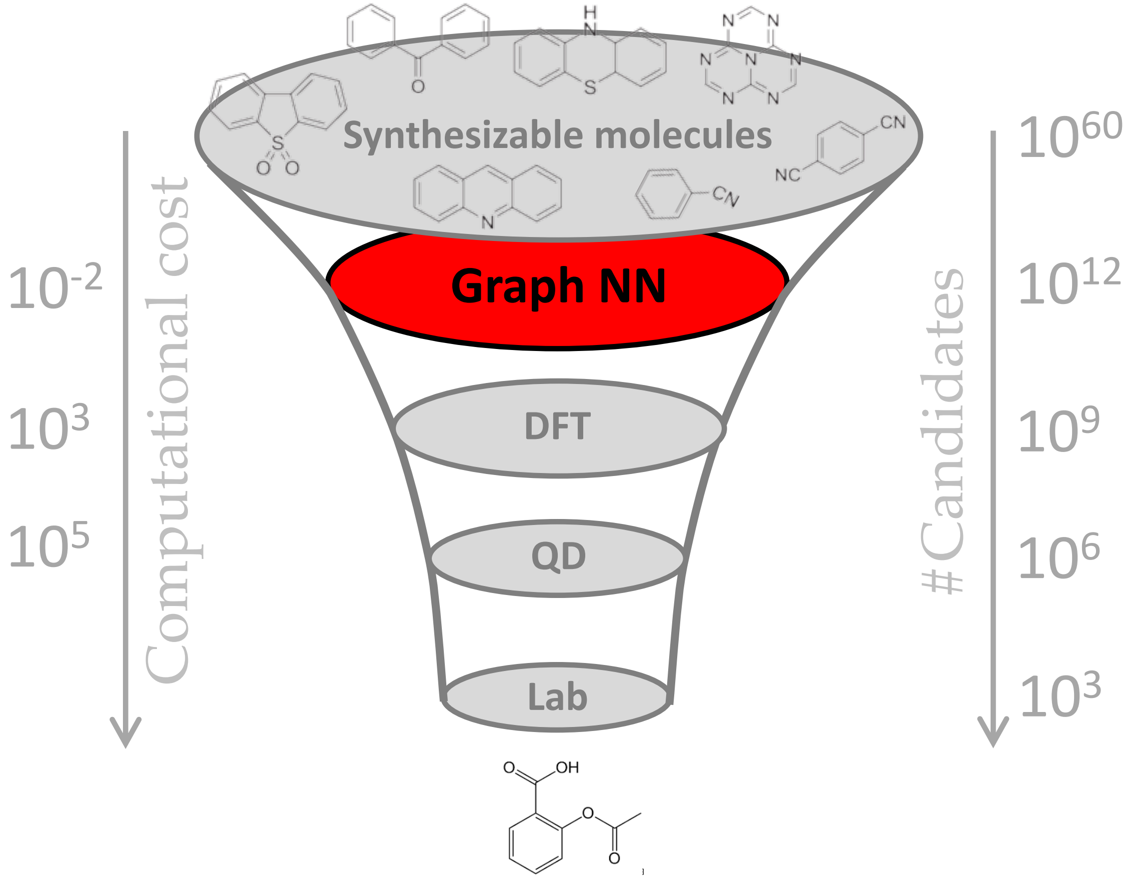
Sverrisson, Feydy et B 2021; Feydy et B 2020

CHEMISTRY & DRUG DESIGN

Drug Discovery & Design



Virtual Drug Screening



"Computational funnel"

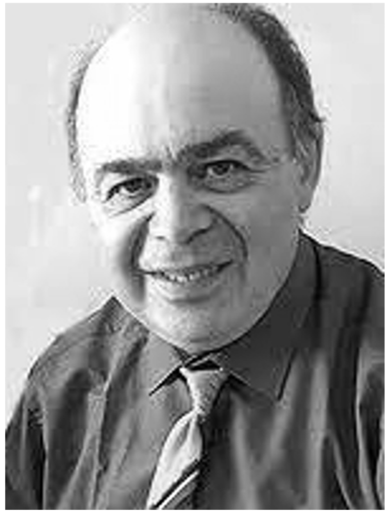
Duvenaud et al. 2015; Gilmer et al. 2017; Jin et al. 2020

Chemical Precursors of Graph Neural Networks

D. Kireev

ChemNet

1995



I. Baskin

Neural descriptors

1997

C. Merkwirth

Molecular graph net

2005

D. Duvenaud

Molecular fingerprints

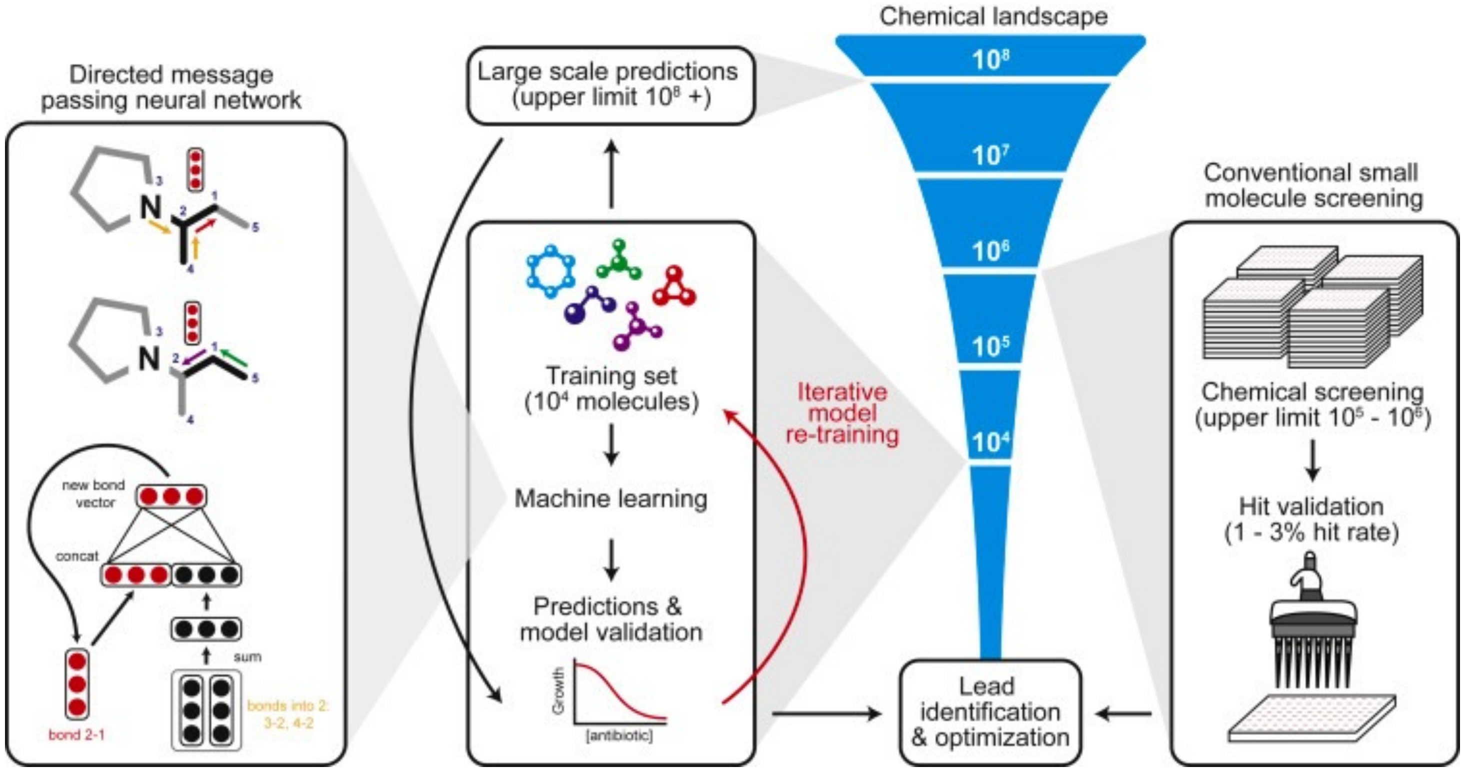
2015

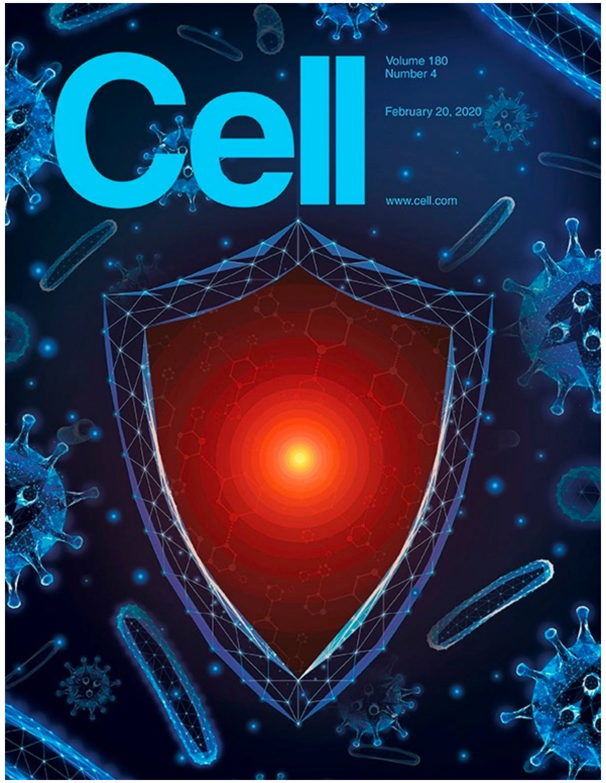
J. Gilmer

MPNNs

2017

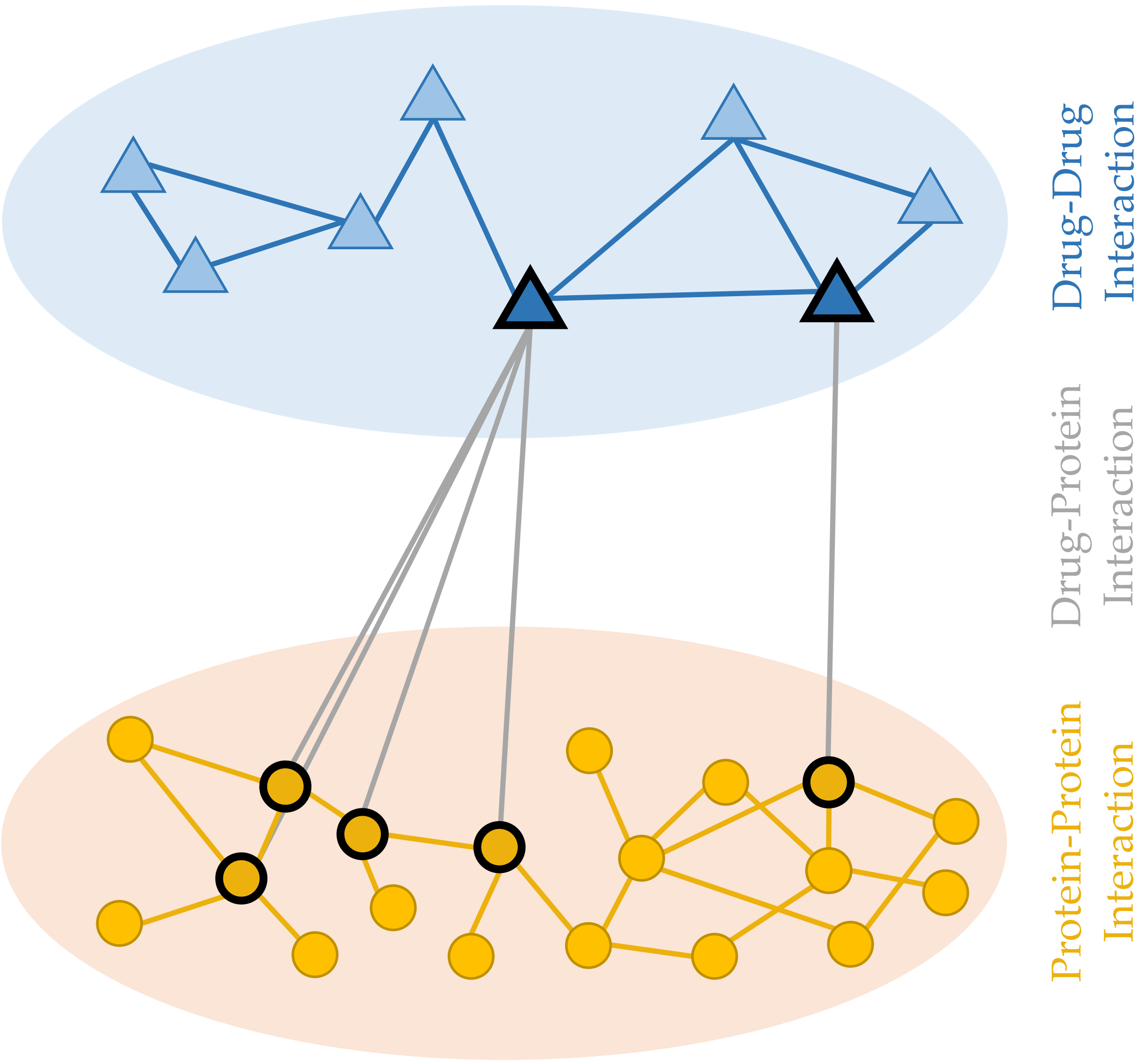
New Antibiotic Discovery





Stokes et al. 2020

Drug repositioning & Combinatorial therapy

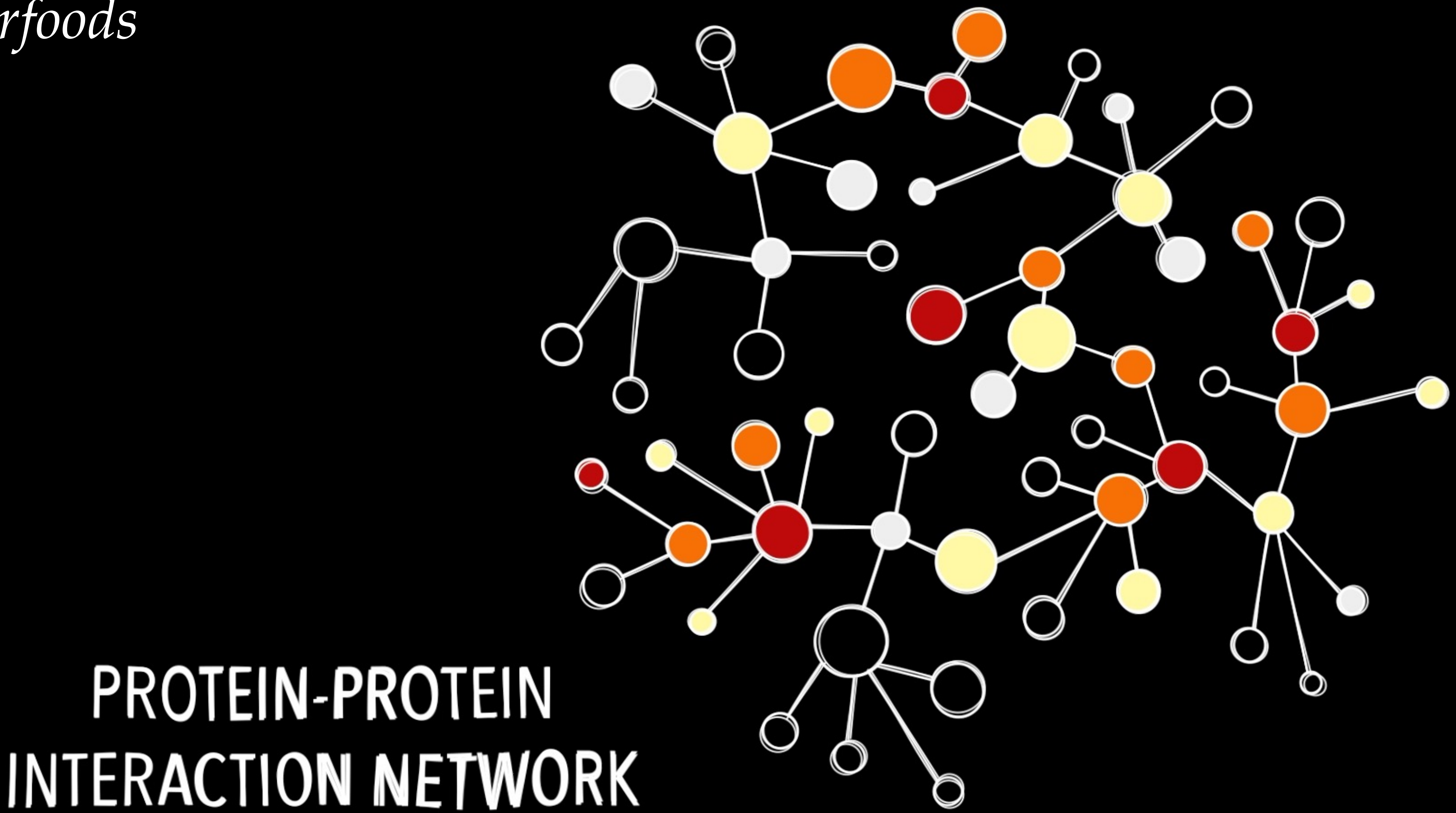


Zitnik et al. 2018

Hyperfoods

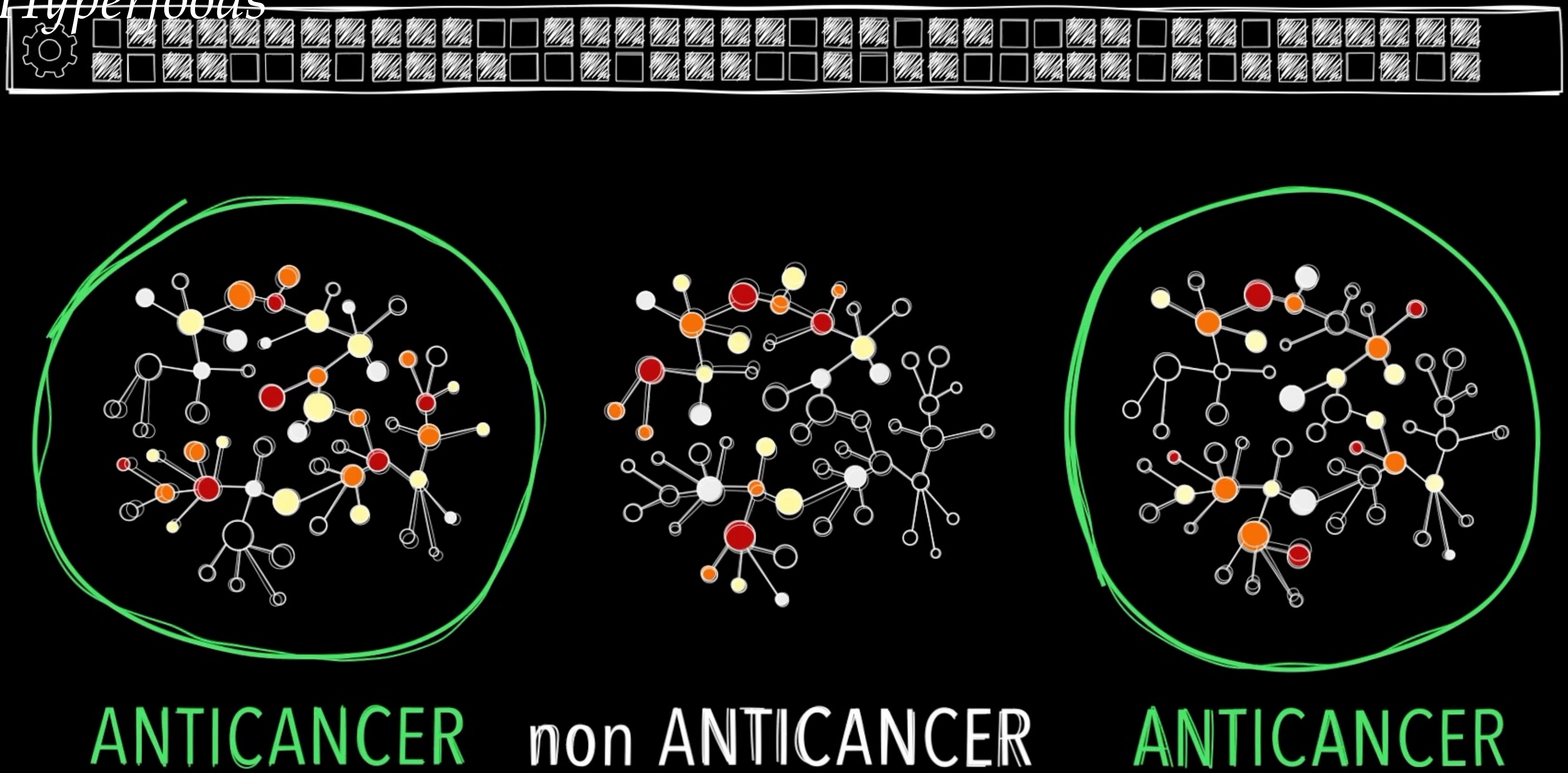
Veselkov et B 2019

Hyperfoods



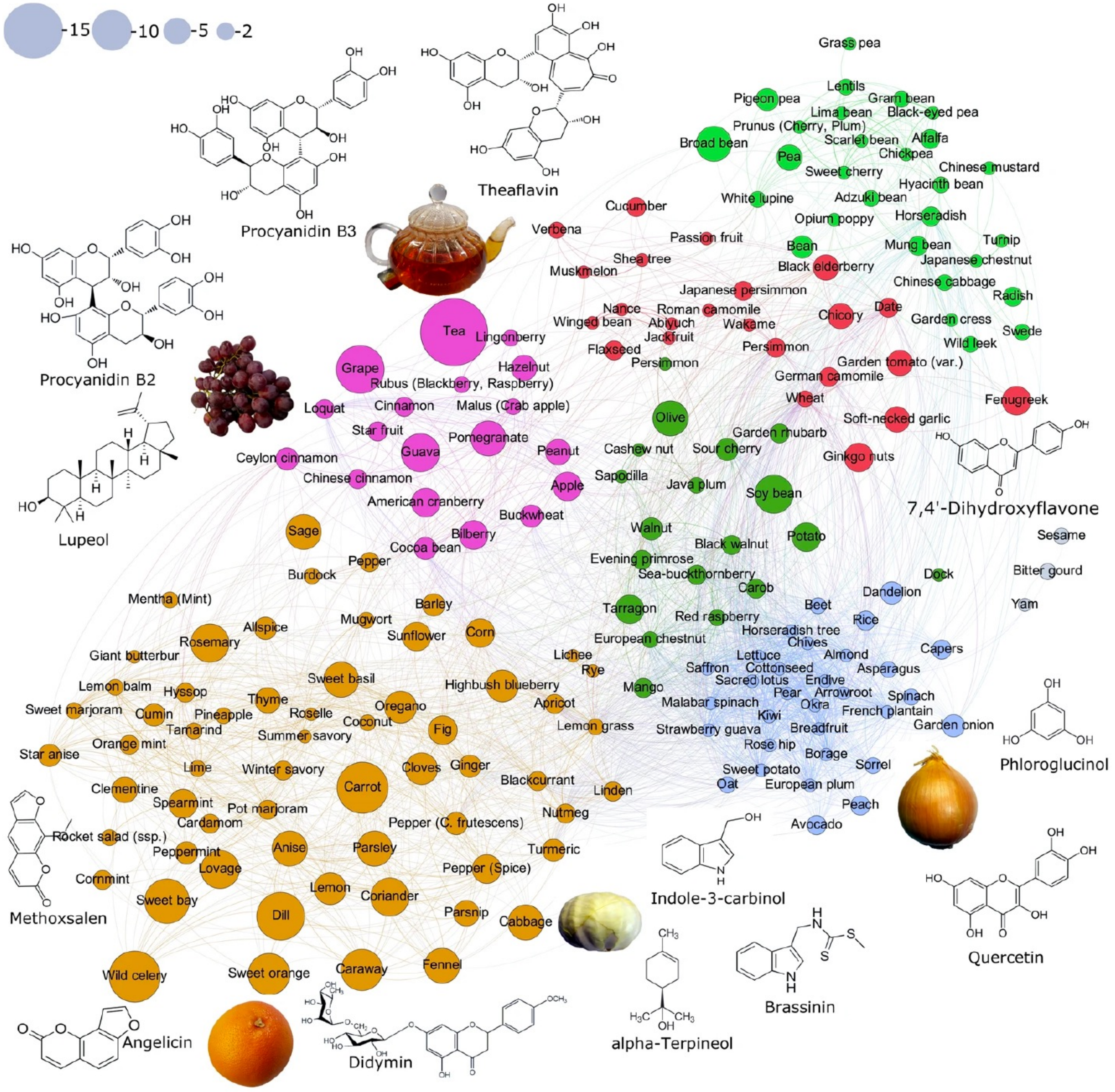
Veselkov et B 2019

Hyperfoods



Veselkov et B 2019

Hyperfoods



Veselkov et B 2019

Video: Vodafone Foundation / Recipes: Bruno Barbieri
Based on Laponogov et B 2020