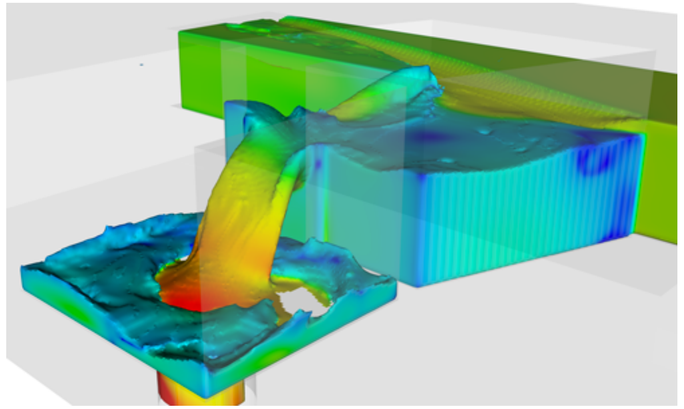
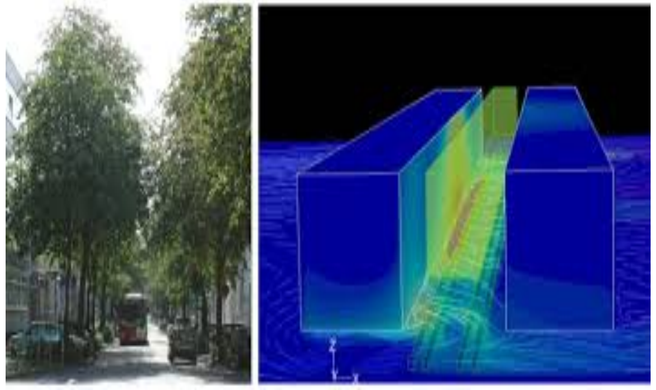
# Data Driven Modeling

Vahid Moosavi
Senior Researcher
ETH Zurich
www.vahidmoosavi.com

**22 April, 2019**

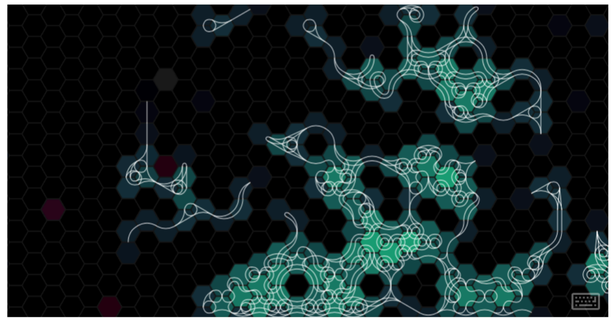
# Computational Modeling Across Engineering Domains, from 1960s…



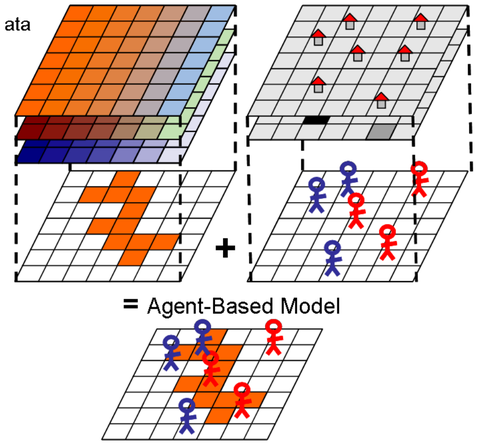


Computational Fluid Dynamics

Steady State

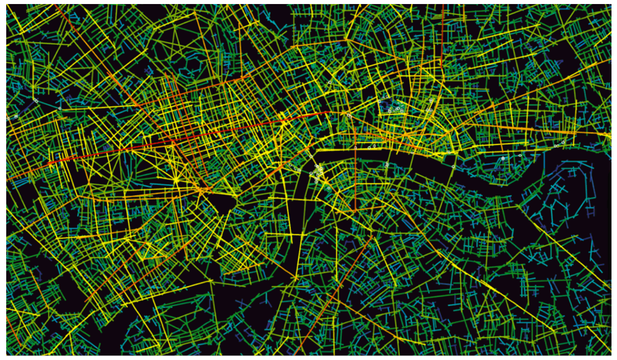


Cellular Automata




Agent Based Modeling

Emergence

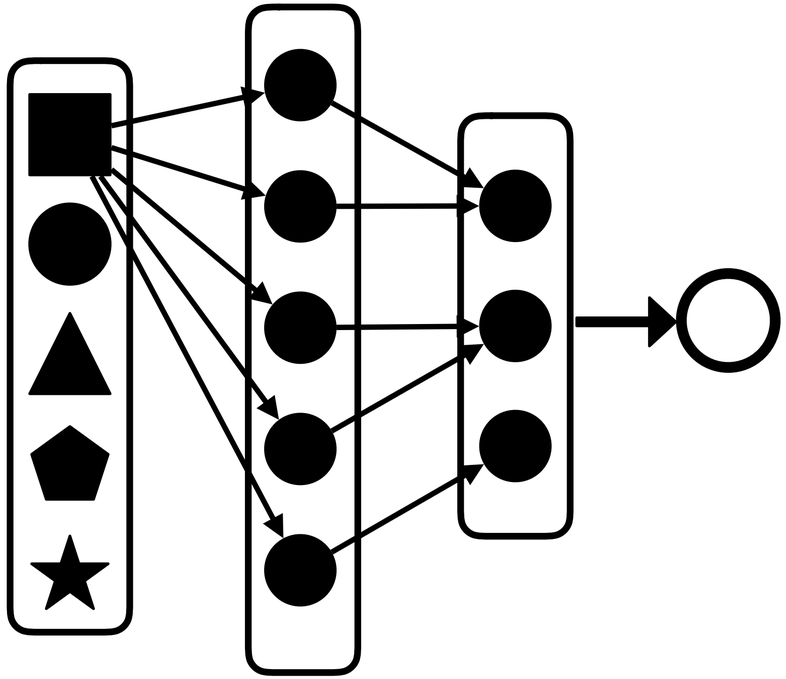


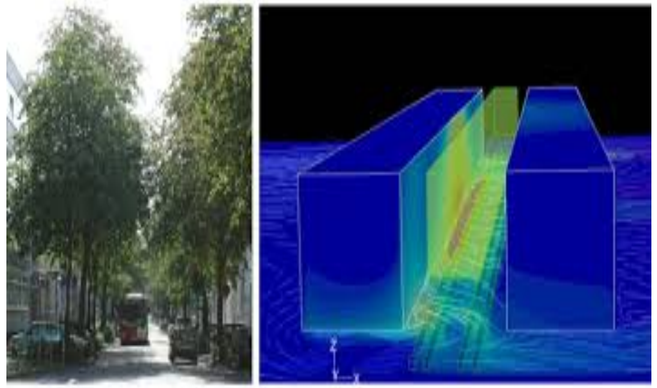
Graph/Network Theory

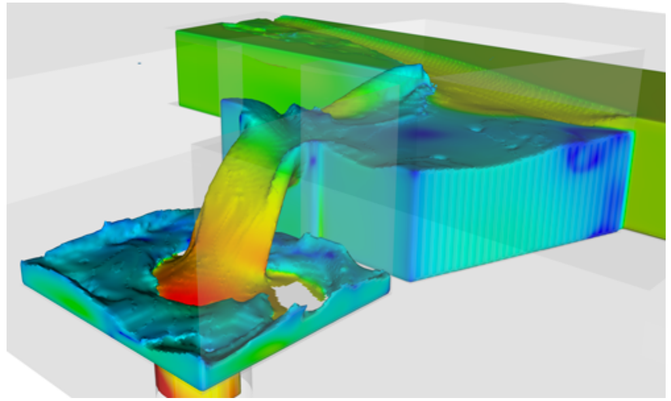
Equilibrium

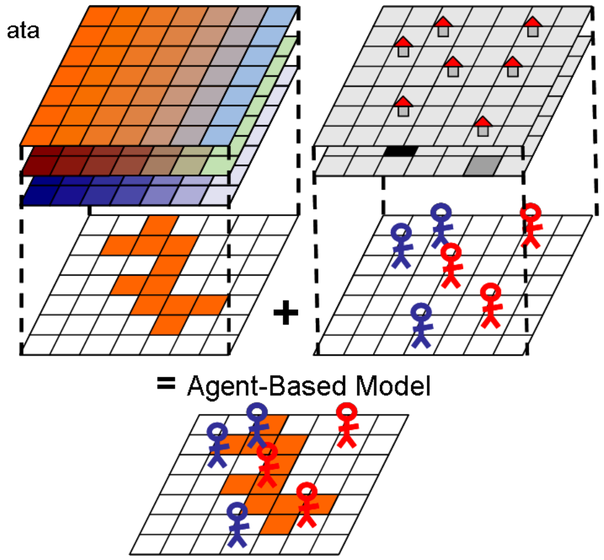
**Systems of (Known Differential) Equations**

# Computation as Automation/Mapping of Knowledge/Logic







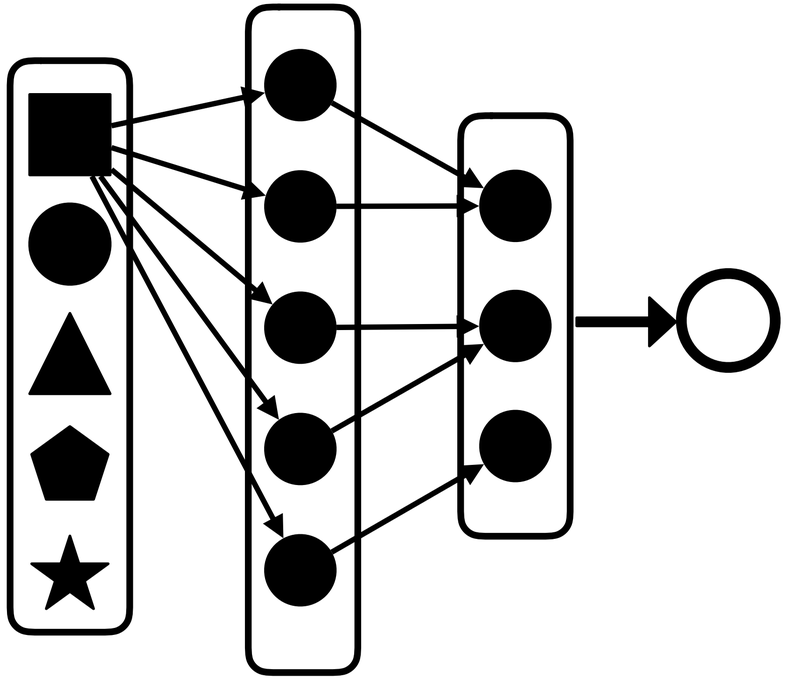


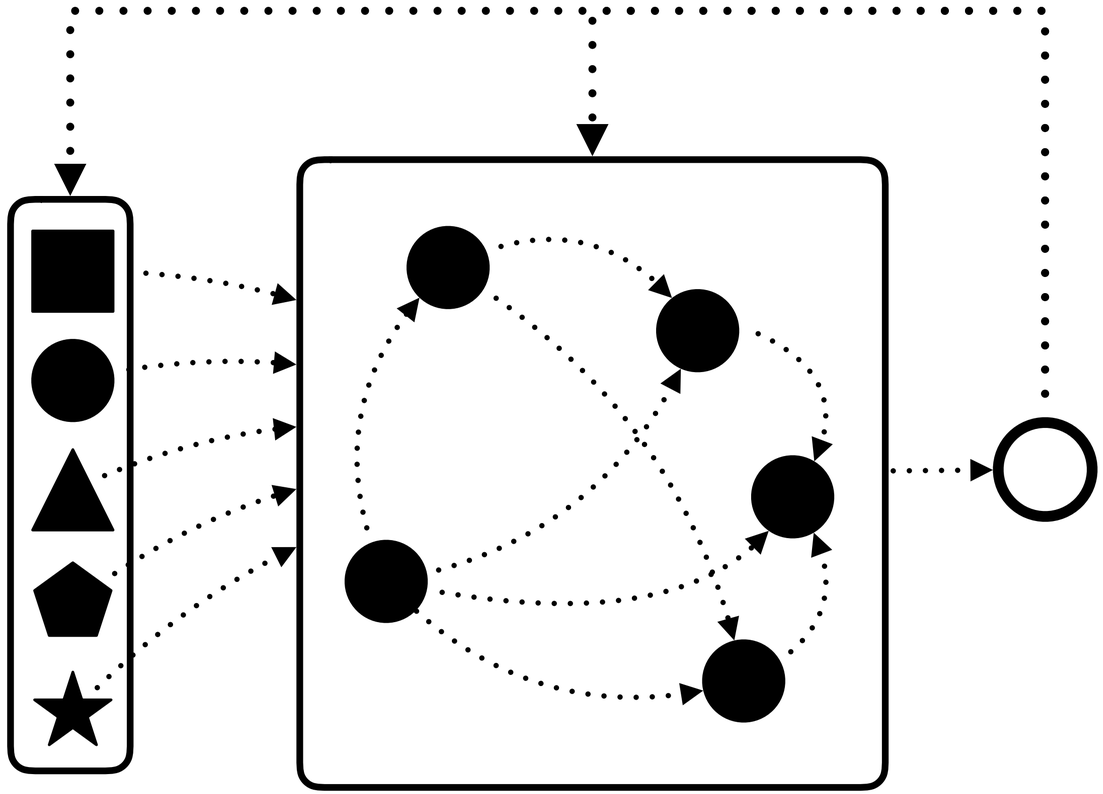

**Automation,**

**Knowledge Mapping**

# Machine Learning 1950s…

**Another way of thinking about computation**



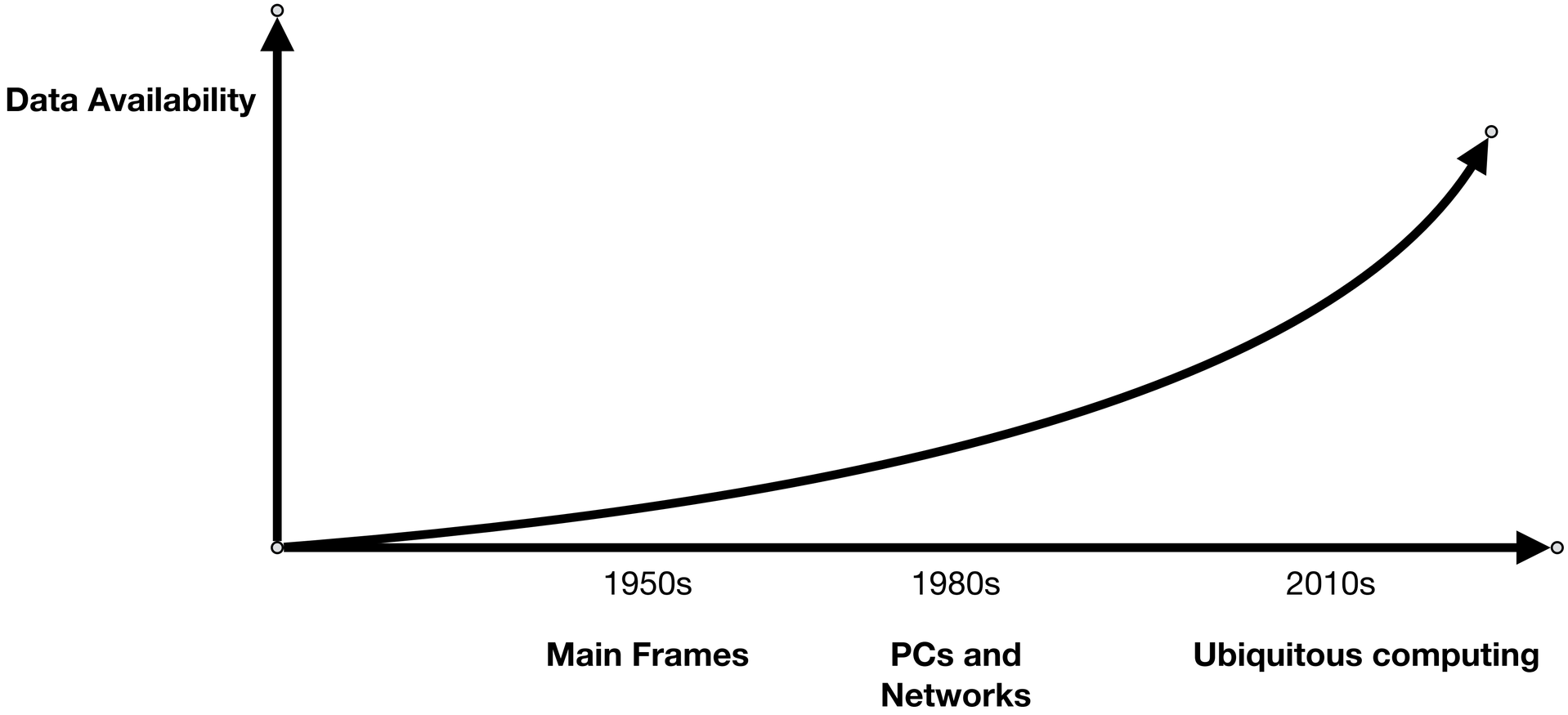


**Knowing**
**Logic**
**Symbol**

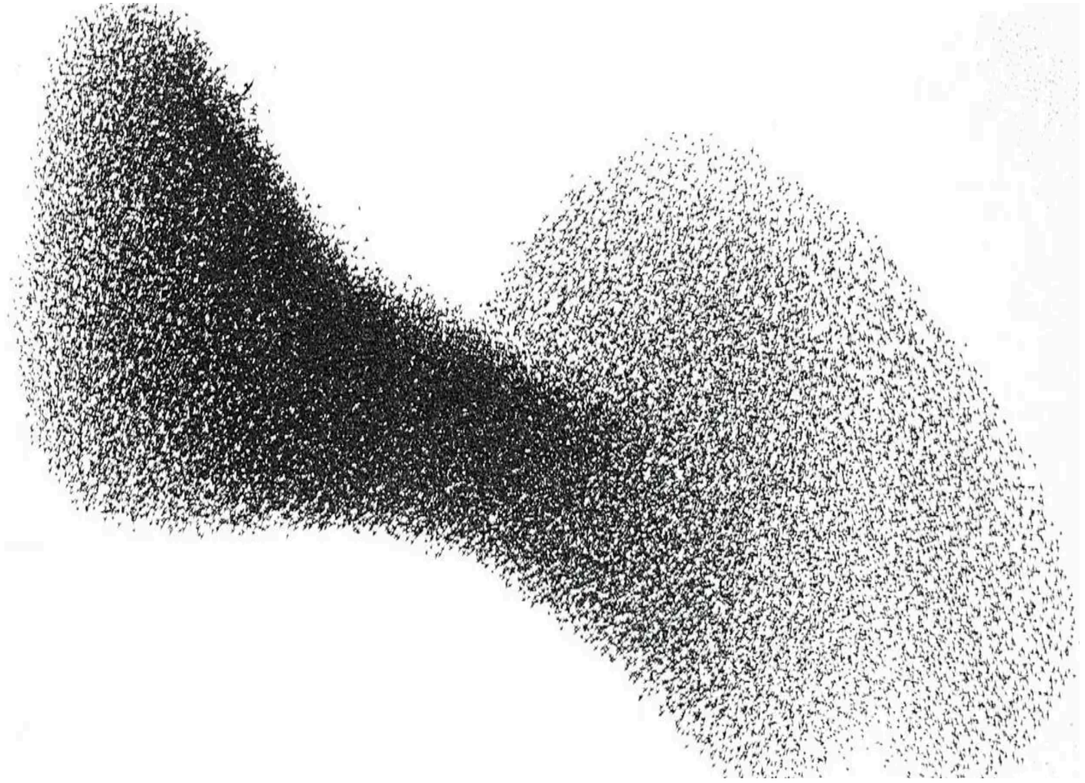
**Learning**
**Algebra**
**Data Vector**

# Machine Learning 1950s…

**However, machine learning is data demanding**



# Machine learning research is evolving like a swarm

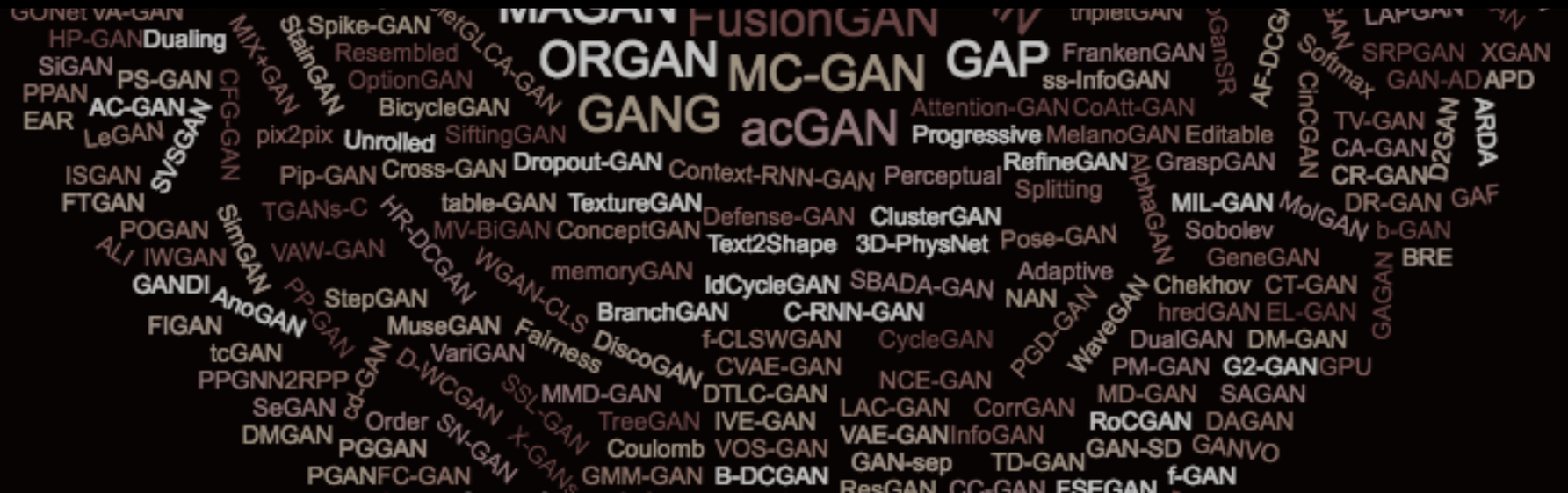


Reinforcement learning
Supervised learning
Unsupervised learning
Feature selection
Prediction
Clustering
Classification
Feature extraction
Decision trees
K-means
Naive Bayes classifier
Nearest Neighborhood Method
Kernel trick
Ensemble models
Support Vector Machines
Meta-parameter optimization
Random forests
Neural Network
Representation learning
Random fields
Deep learning
Topological data analysis
Dimensionality reduction
Energy based models
Manifold learning
Cross validation
Self Organizing Maps
Dictionary learning
Neural encoding
Sparse encoding
Over-fitting and generalization
PCA
Similarity measures
Gradient descent
Function approximation
Space transformation
Curse of dimensionality
Fourier transformation
Precision/Recall
Matrix operations
Model complexity
Markov chains
Meta-heuristics
Vector spaces
Linear systems
Regression models
Accuracy measures
Hill climbing methods
System of equations
Nonparametric probability distributions
Objective minimization
Random walk
Data reduction
Invariance features
Resampling
Least square method
Probabilistic distributions
Data Visualization
Maximum likelihood
Law of large numbers
Causality/correlations
Data normalization
Hypothesis testing
Central limit theorem
Descriptive statistics
Histograms
Bayes' rule
Mathematical programing
Random variables
Outliers
Kolmogrov axioms

# GAN Zoo!

Generative Adversarial Networks


**As a non-computer scientist how to navigate in this space?**



# Towards a Systemic View to Machine Learning

## Function

- Questions/Targets
- Measure of fitness/success
  Measure of similarity
- **Least Squares**
  **(Legendre 1805, Gauss 1809)**
  **Loss Function Design (2018)**

## Structure

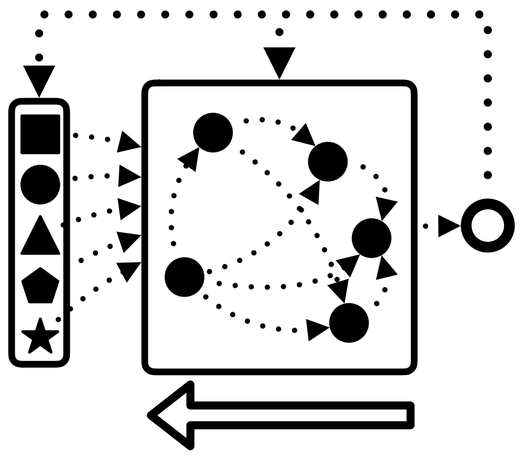
- Data Representation:
  How to deal with Invariances?
- Linear Algebra
- **Computational Graph**

## Process

- How to get better fitness/scores?
- Optimization
- **Back-propagation (1960s)**
  **Gradient Descent (Cauchy 1870s)**
  **Chain Rule (Leibniz, 1676)**
  Evolutionary strategies

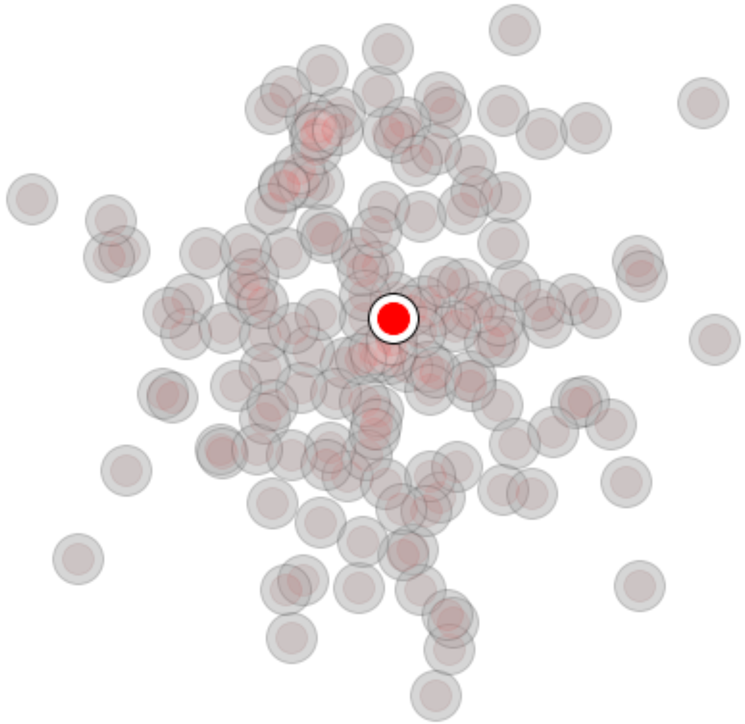
## Context

- Different Data
  Modalities
- Homogenous (Heterogenous) data
  Euclidean and Non-euclidean space

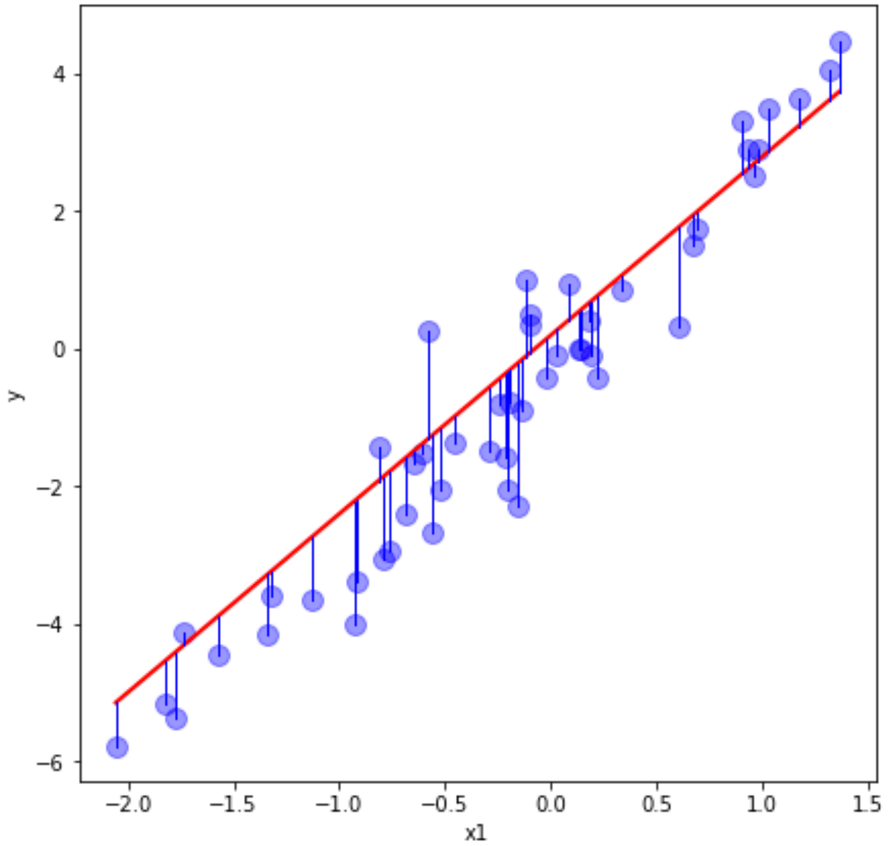


**Linear Regression**

$$S = \sum_{i=1}^{n} (y_i - f(x_i, \beta))^2 = \sum_{i=1}^{n} r^2$$



space of potential functions and the optimum function



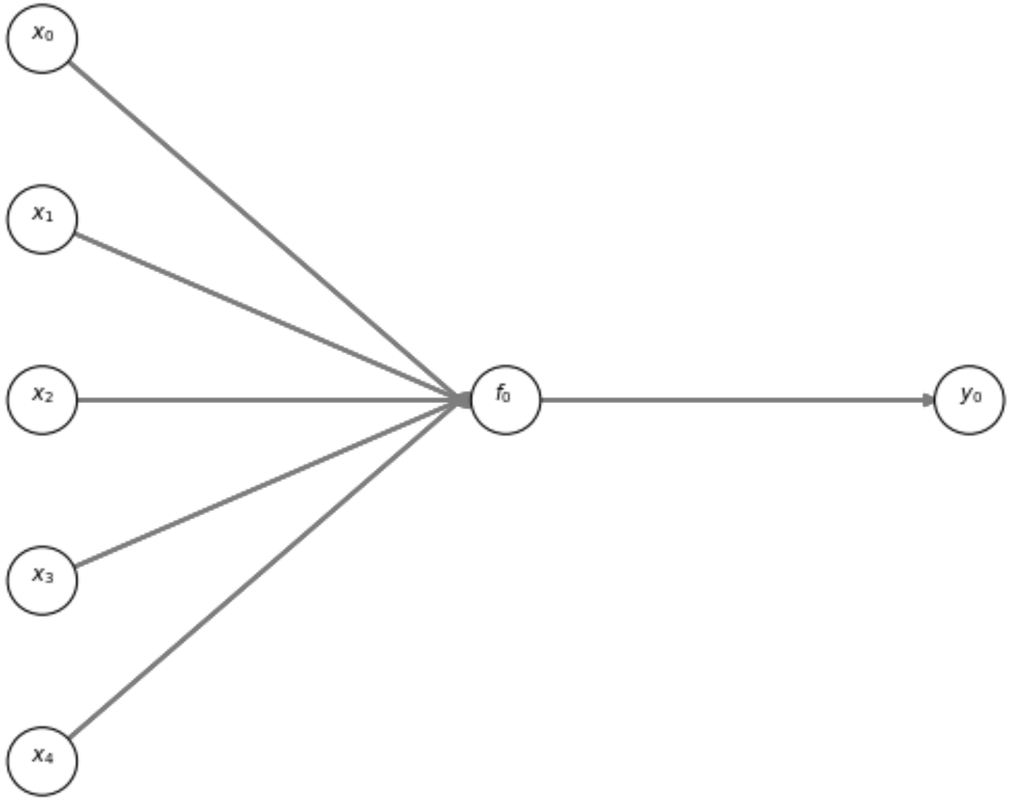
# Three Stages of Machine Learning From the Aspect of “Computational Graph”

# 1. A "central memory" and a "global prototype"

**Linear Regression**

**Before 1980s**

**Computational Graph**



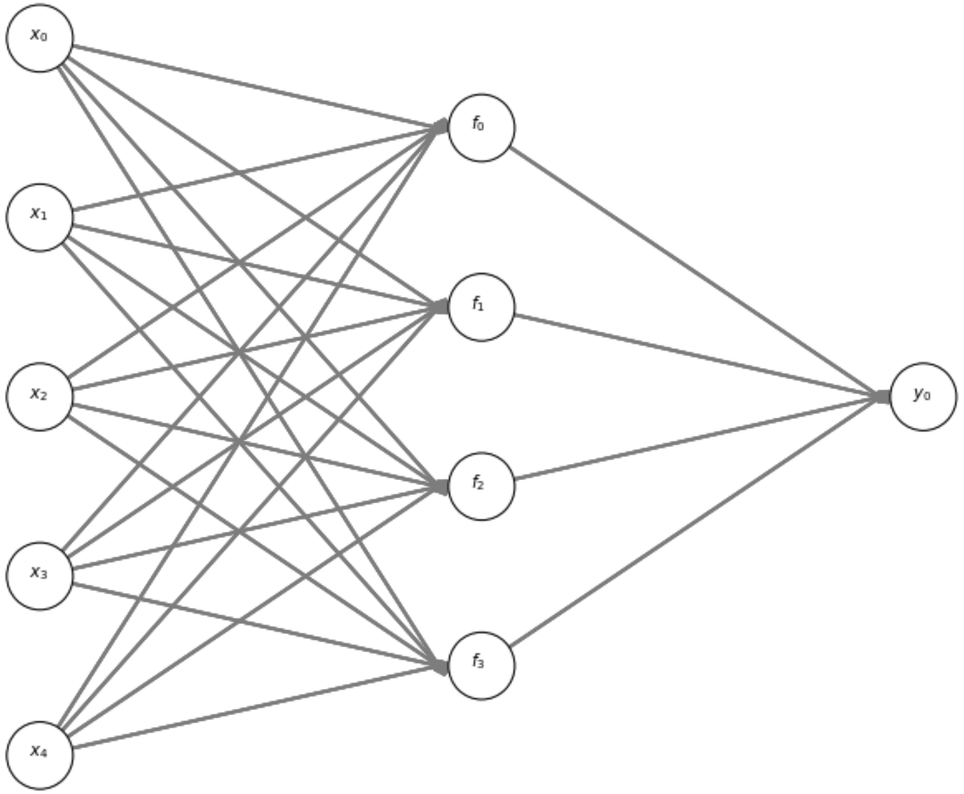

**Central memory means that this abstract point can generate complete instances of the objects!**

# 2. Centralized Memory, Distributed Prototypes with Multiple Global Views

**Manifold Learning**

**From 1980s to 2010**

**Computational Graph**



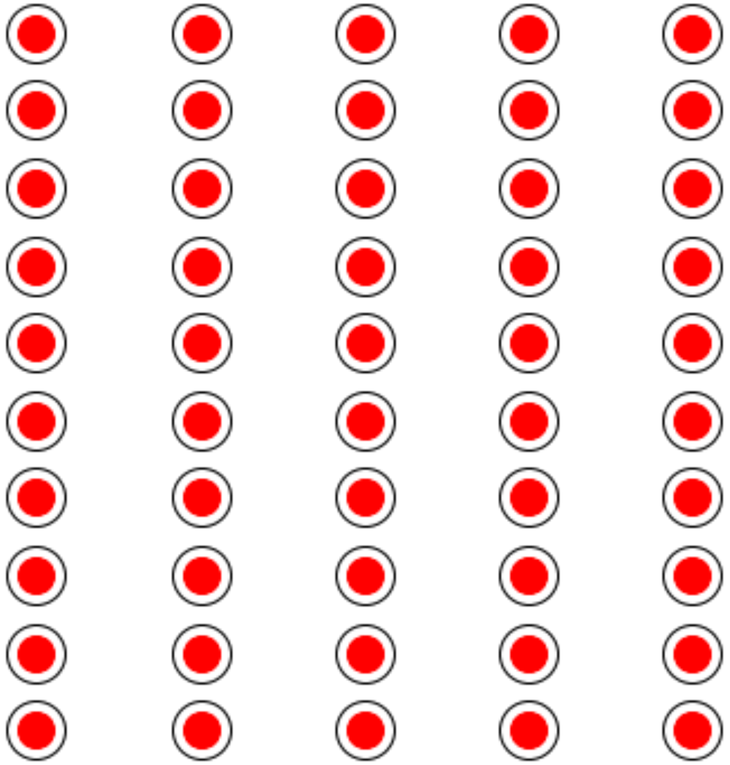

**Key issue:** How to orchestrate these abstract points?

**The main problem:** The capacity of the model grows linearly as the function of number of points!

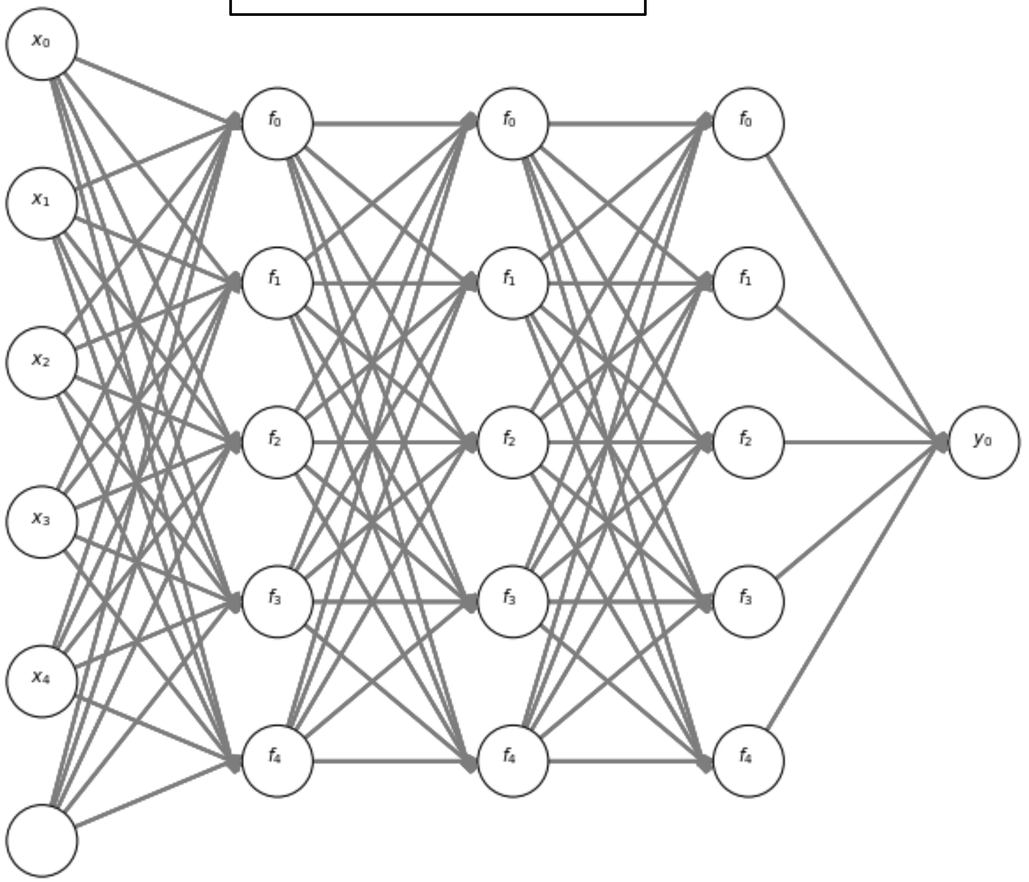
# 3. Distributed Memory and No Explicit Prototypes and No Overall View

**Deep Learning**

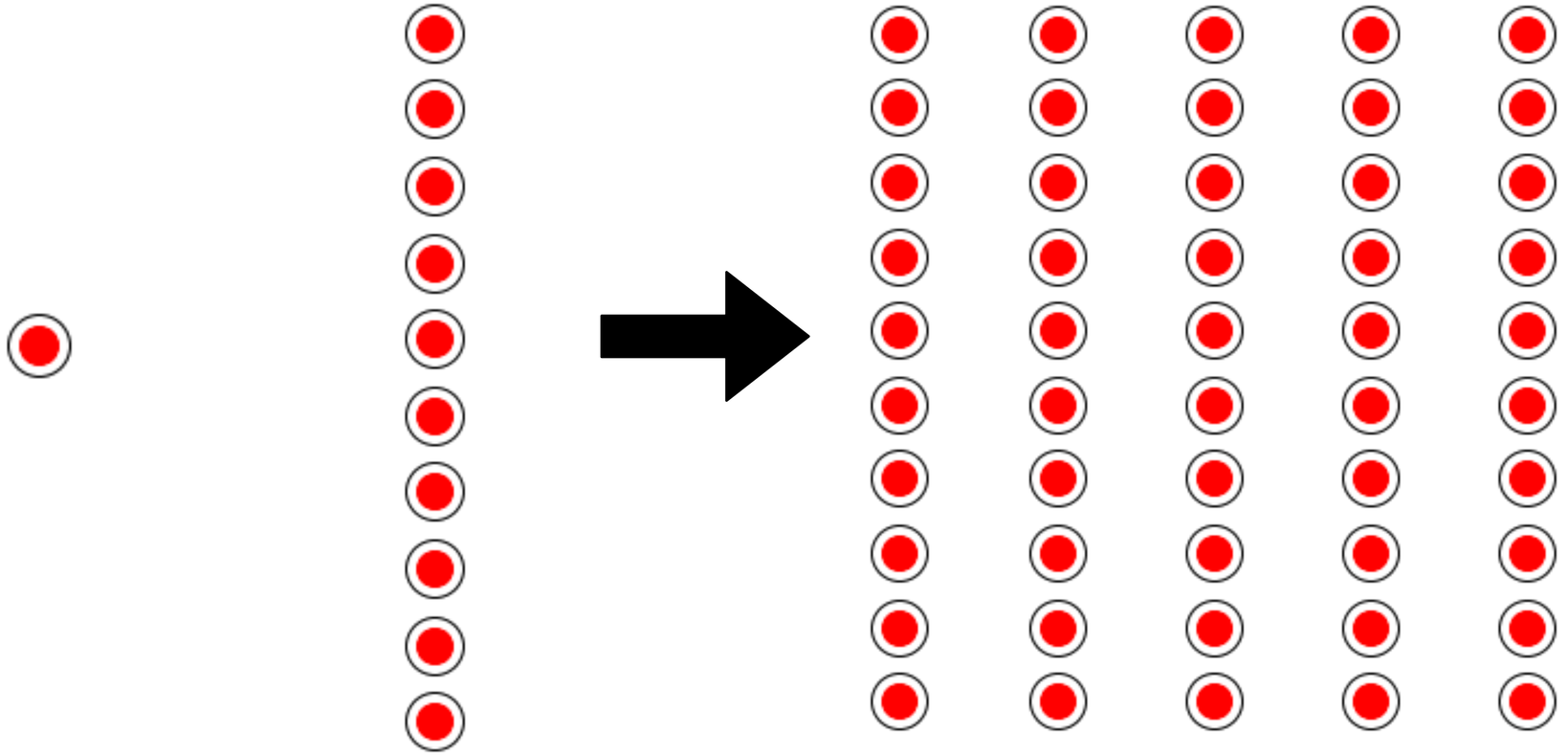
**From 2010s**

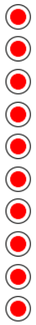


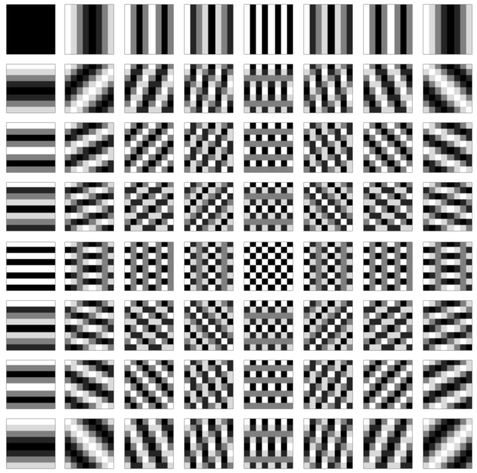
**Computational Graph**

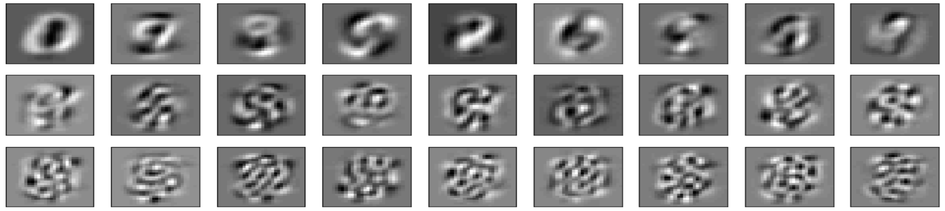


# I think a major shift is happening here!

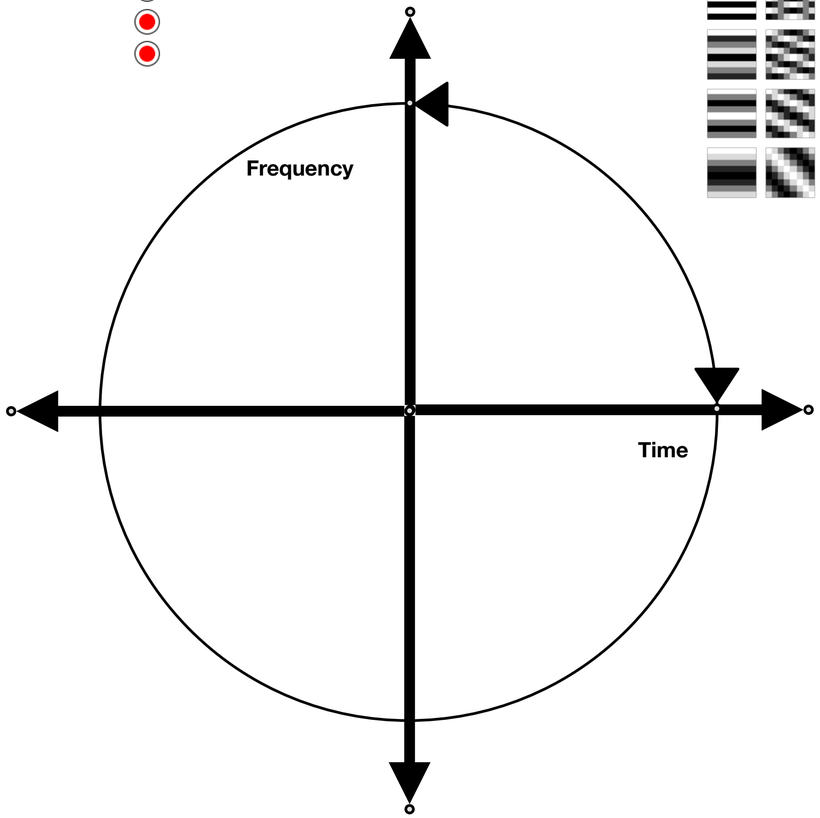








## Fourier Transform



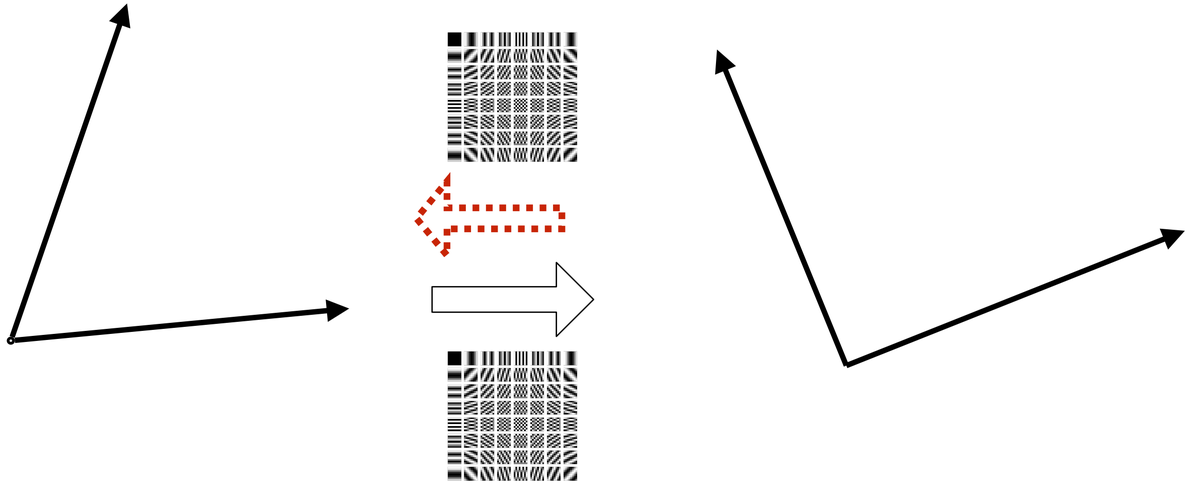

**Vector Space**

**Linear transform**

**Meaningful transform**

**Orthogonal basis system (e.g. PCA)**

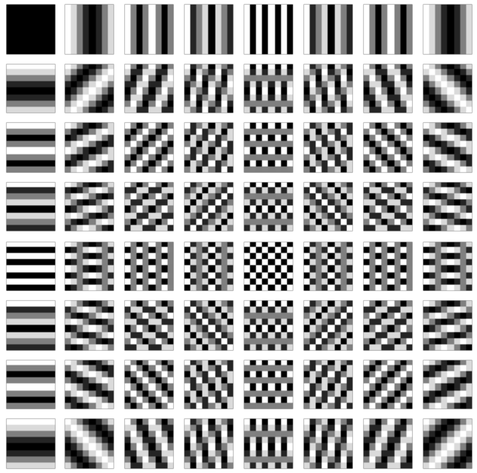
**Idealization is hidden here!**

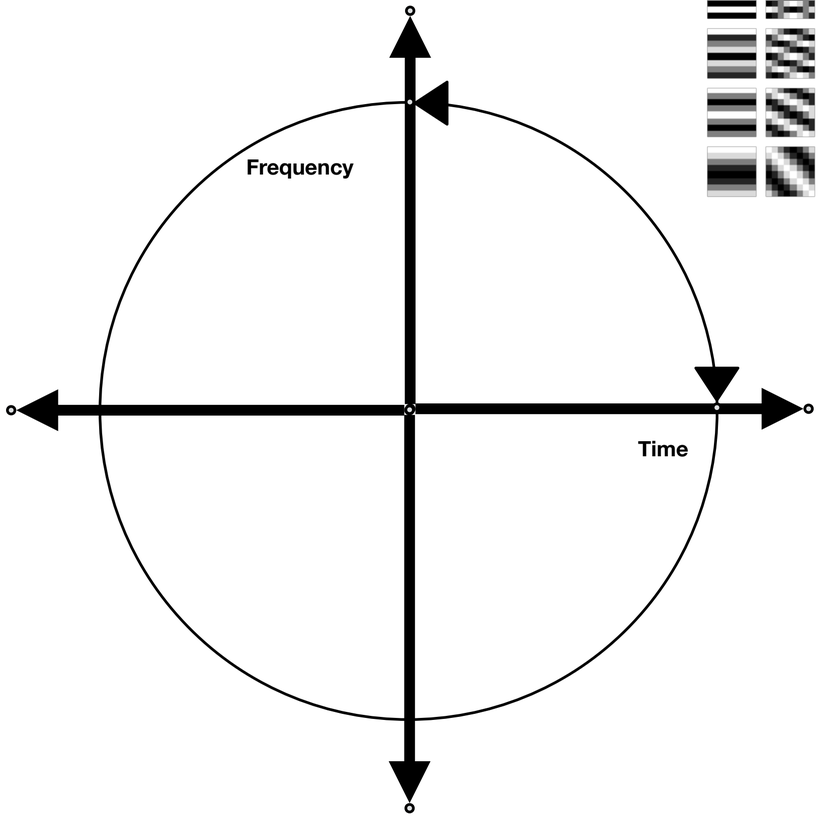


**By linear transformation we can jump in either space of time or frequency**

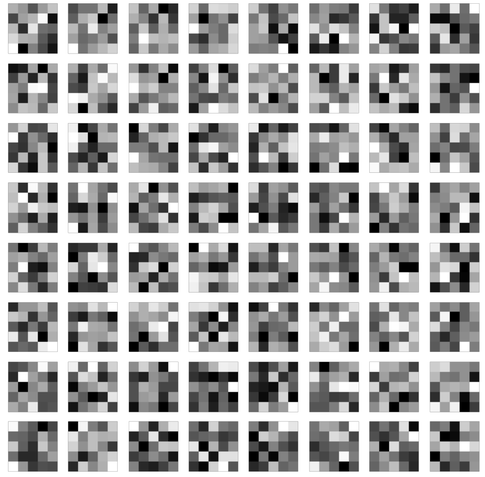
Credit to: Elias Zafiris

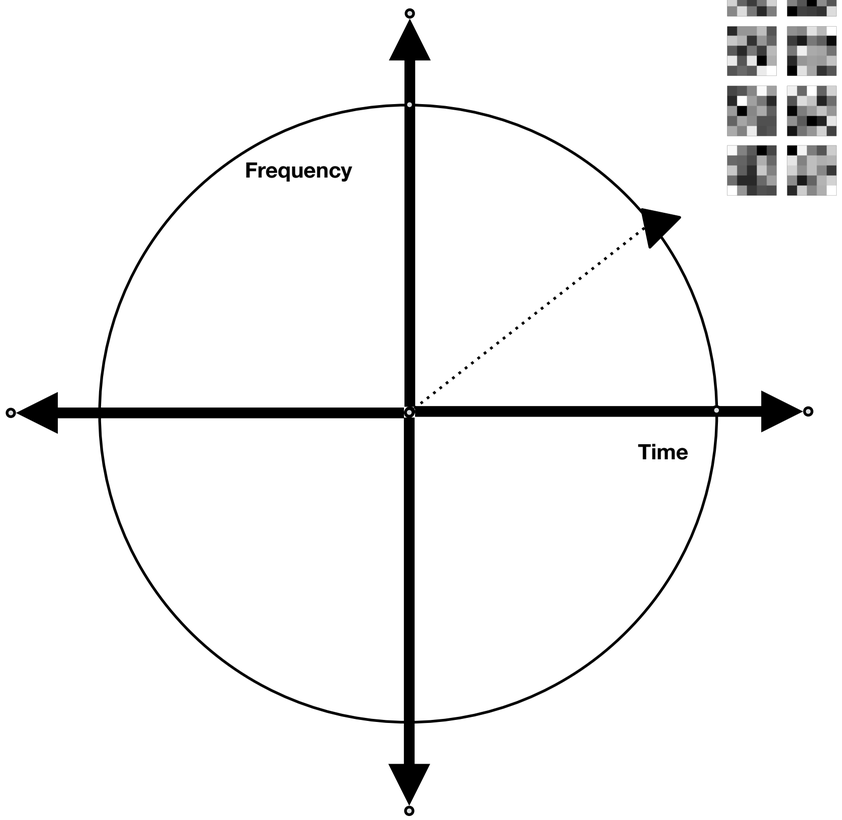
## Fourier Transform



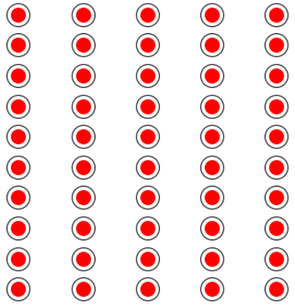


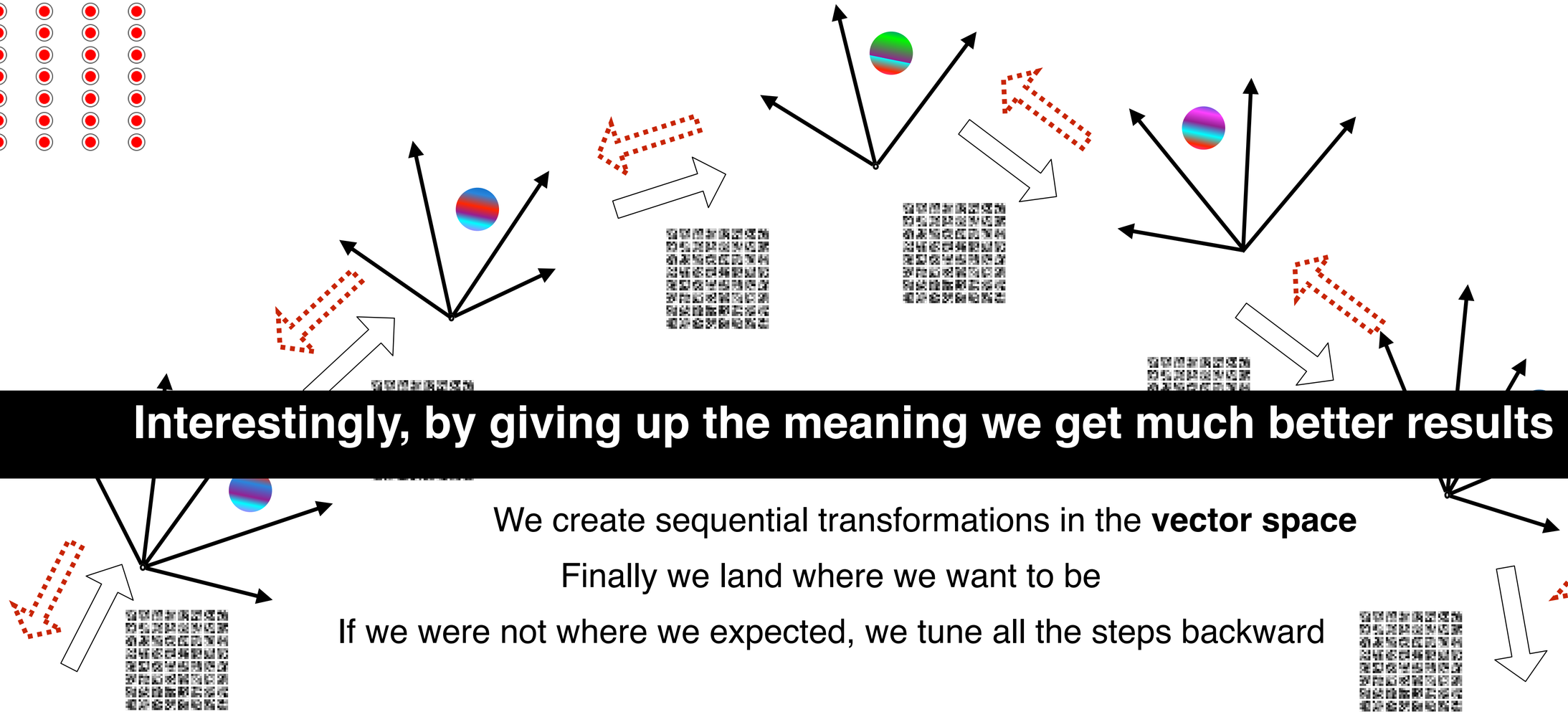

## Deep Learning (CNNs)






**Is there any meaning in the space between time and frequency?**



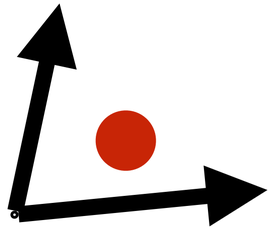


**Interestingly, by giving up the meaning we get much better results**

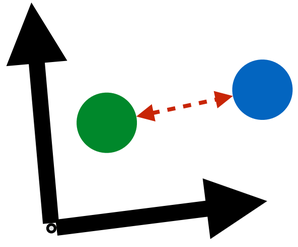
We create sequential transformations in the **vector space**

Finally we land where we want to be

If we were not where we expected, we tune all the steps backward



**A meaningful context**



# Three Stages of Machine Learning From the Aspect of “Computational Graph”

Linear Regression

Addition

Memory as a Particle or Wave

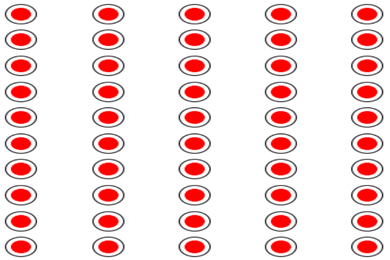
King/God

Manifold Learning

Multiplication

Memory as a Particle or Wave

Political Parties/Brands



Deep Learning

Power

Memory as a Particle and Wave

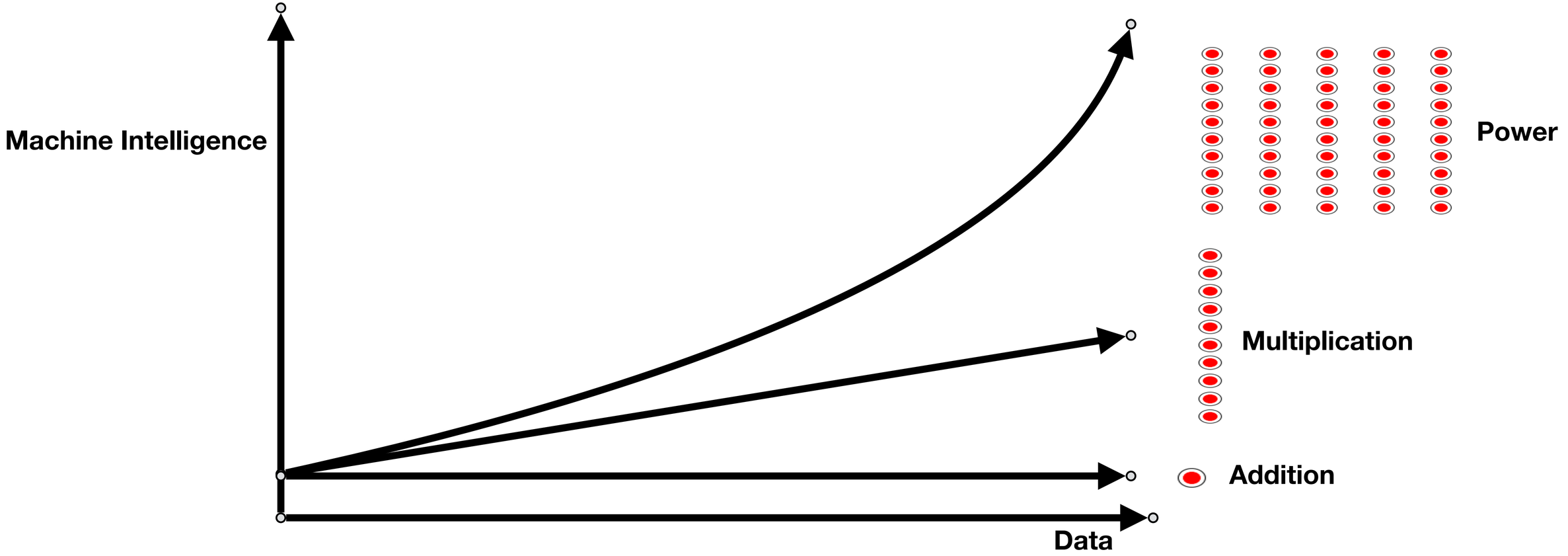
Individuals?

**Do we need new ways of discussion and communications?**

**There is a risk to shortcut to stage one!**

# Conclusions: Concerns and Possibilities

**This exponentially growing capacity might result to Tyrannic setups if it is centralized.**




**But also with literacy it could be seen as a new capacity for super individuals**

# My Focus On Techniques and Typology of Problems

**Data Driven Modeling Across Domains**

An Orthogonal View to Classical Scientific Modeling

Vahid Moosavi

**Forthcoming Book**

## Teaching to graduate students from 2016

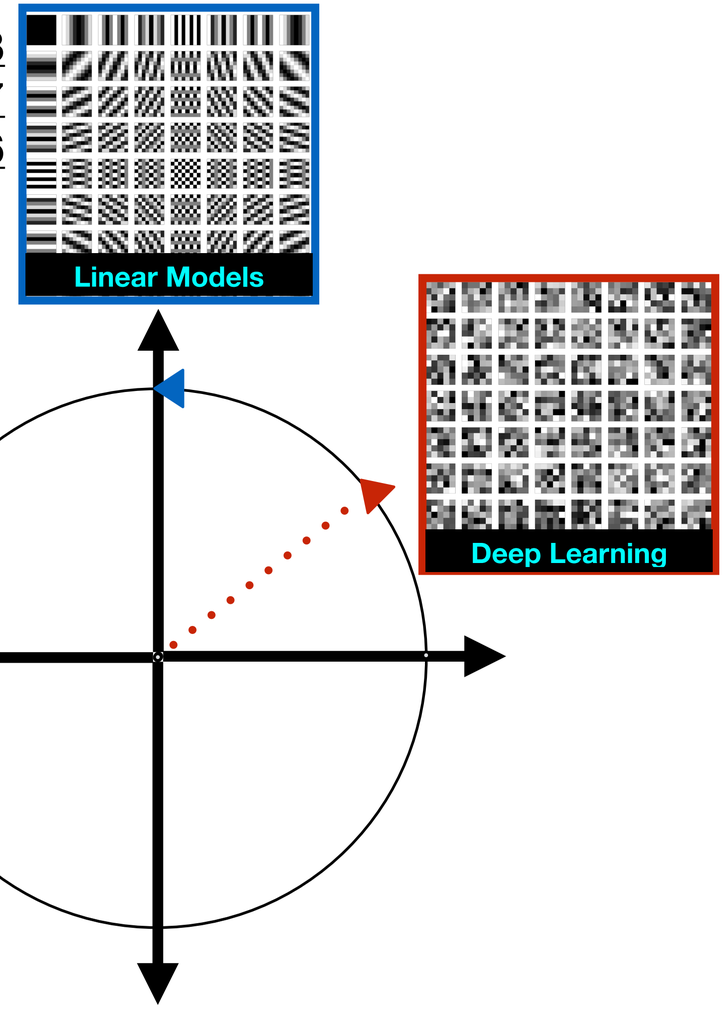
GitHub

**Lecture Notes, Codes and Videos**

https://github.com/sevamoo/data_driven_modeling_2018

https://github.com/sevamoo/data_driven_modeling_2017

https://github.com/sevamoo/data_driven_modeling_2016




**Statistics and Probability**

**Linear Algebra**

**Optimization**

**Machine Learning**

# Conclusions: Concerns and Possibilities

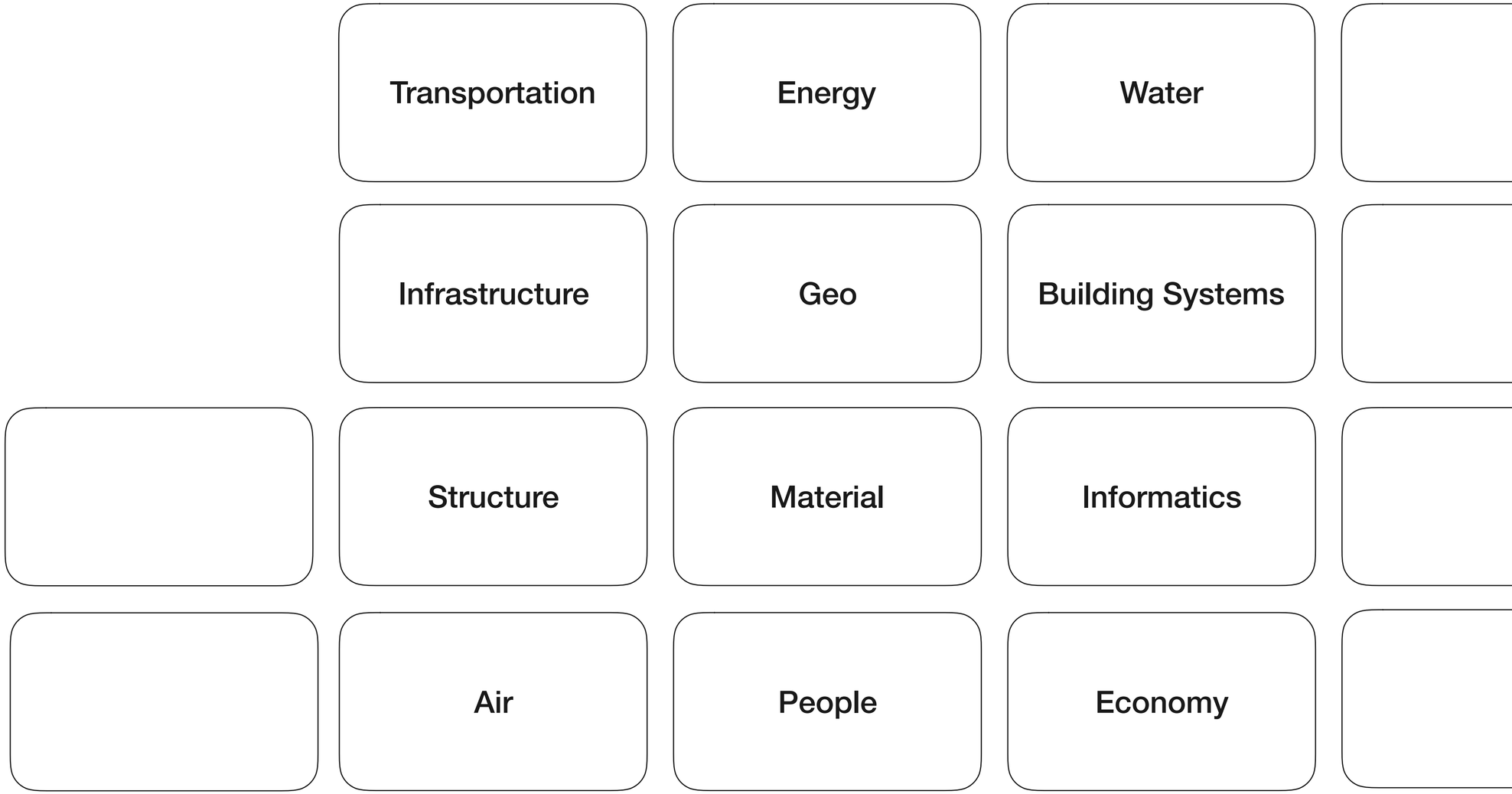
Machine learning is good at finding answers:
**An alarm for domain experts!**

Machine learning is not able to find good questions:
**A good combination for humans with coding literacy**

Classical disciplines (Centered around expertise) need to redefine themselves.

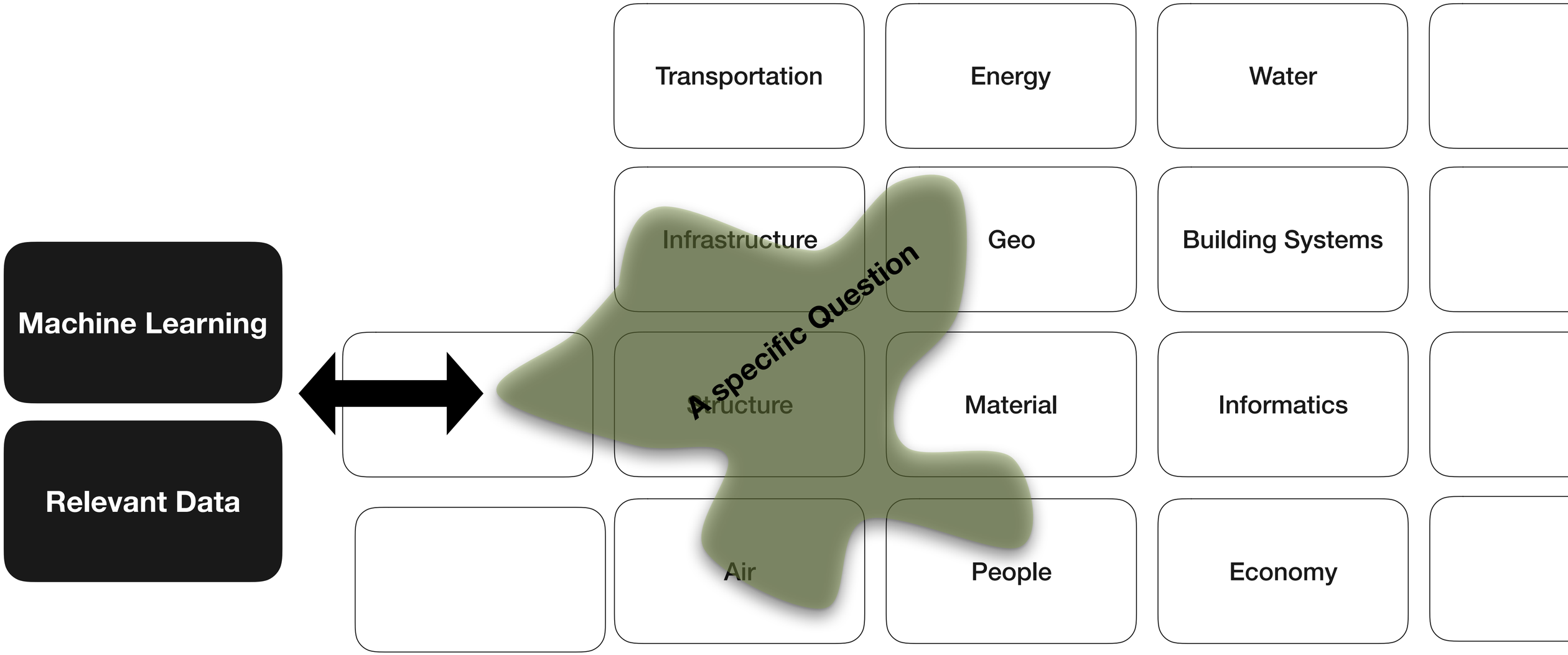
# Disciplinary Research (Working within given frames)

… Is focused on **a set of fixed questions** and knowing their answers

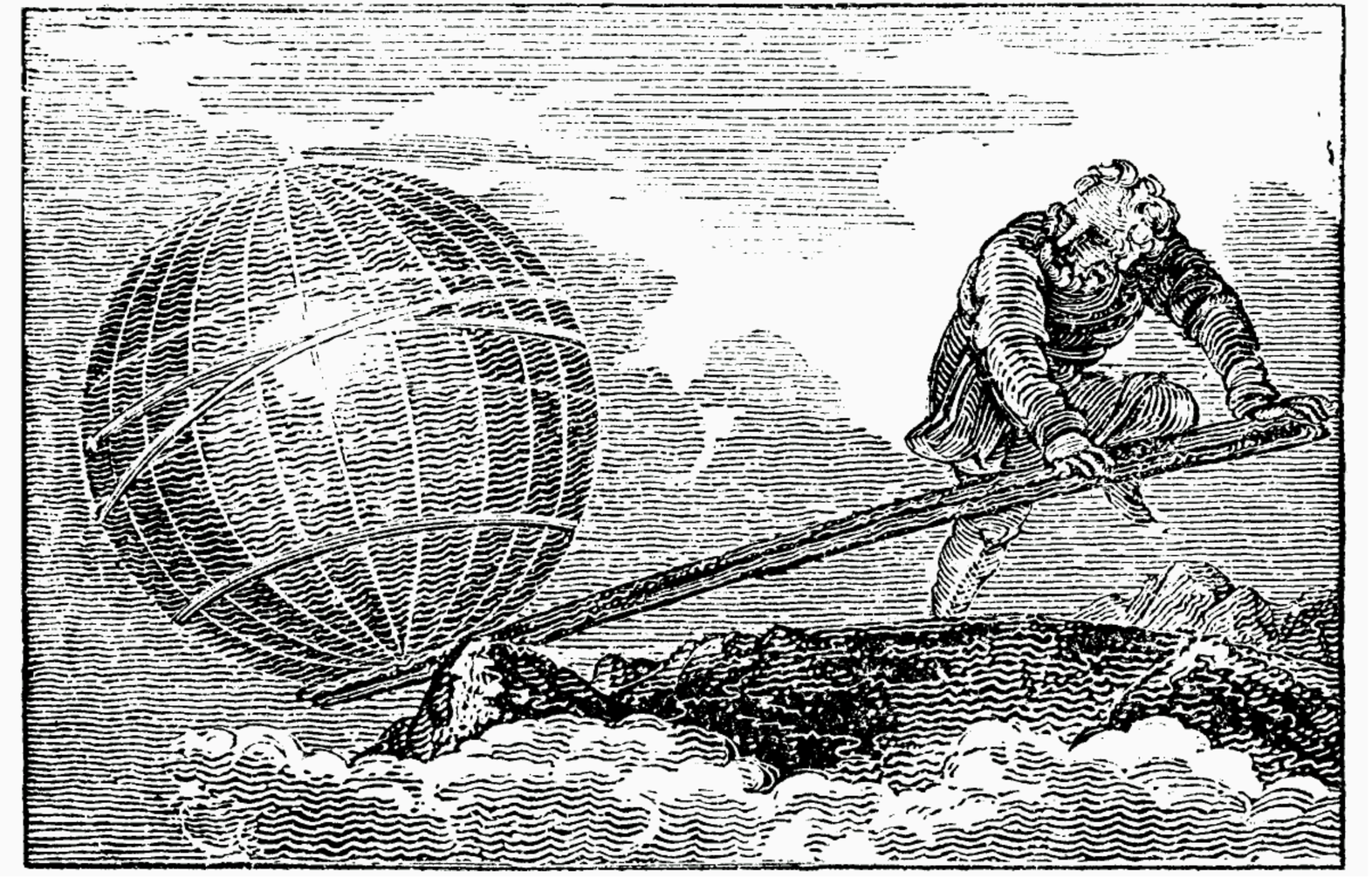


# Data Driven Research (Machine Learning + Data)

… Is focused on **asking** good questions and learning their answers



# Applications



**Give me a lever long enough and a fulcrum on which to place it, and I shall move the world… ,Archimedes**

**Technology is the answer but what was the QUESTION? (Cedric Price)**

# Applications

## Urban and Spatial Applications

- Fast and Scalable Urban Flood Risk Estimation
- Urban Air Quality at the Local and Global Scales
- Urban Forms and Urban Morphology
- Remote Sensing and Slum Detection
- Exploratory City Mining
- Urban Traffic Simulation

## Economic and Financial Problems

- Urban Economy and Real Estate Market Dynamics
- Systemic Risk in World Economic Networks

## Other Application Domains

- Structural Design and Design Space Exploration
- Atmospheric Science
- Natural Language Processing
- Manufacturing System Modeling
- Supply Chain Management

# Network

## Main Collaborators

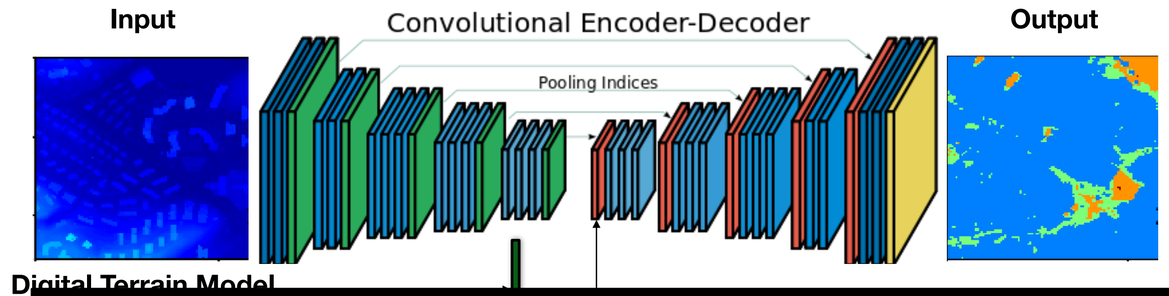
- Future Cities Lab, Singapore
- Center for Environmental Sensing and Modeling (CENSAM), MIT, SMART, Singapore
- Department of Urban Water Management, EAWAG, ETH Domain, Switzerland
- German Remote Sensing Data Center, Land Surface, German Aerospace Center (DLR)
- Implenia AG. (Construction Company), Switzerland
- Cambridge Centre for Climate Science, UK
- Chair for Structural Design, , ETH Zurich
- Block Research Group, ETH Zurich
- Polyhedral Structures Lab, University of Pennsylvania

## Format of Data Driven Applications

- Academic research and education
- Startup and consulting to industry

# (1) Learning Physics

**Slow Physics Based Simulations + Machine (Deep) Learning : Fast Emulators (Data)**




**In this use case:** The model learns the relation between the rain patterns, terrain model, and the expected max water levels for a specific region. The data comes from a slow physics based simulator.

## Next Possible Steps:

Generic Data Driven Urban Flow Engines? **Water, Wind, Heat, …**

2. Real time weather forecasts + Real time flood maps

Collaborators: Dr. Joao Paulo Leitao, Department of Urban Water Management, EAWAG, Switzerland, Guo Zifeng, CAAD ETH Zurich

**Early results**: João P. Leitão, Mohamed Zaghloul and ***Vahid Moosavi***, Modelling overland flow from local inflows in “almost no-time” using Self-Organizing Maps, Proceedings of 11th International Conference on Urban Drainage Modelling, Palermo Italy, 2018.

# Optimization in Statically Indeterminate Structural Systems

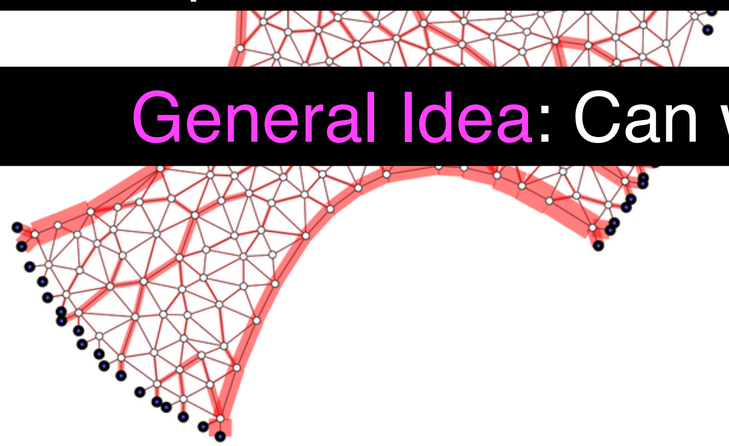
## Minimizing Volume = Minimizing Load path

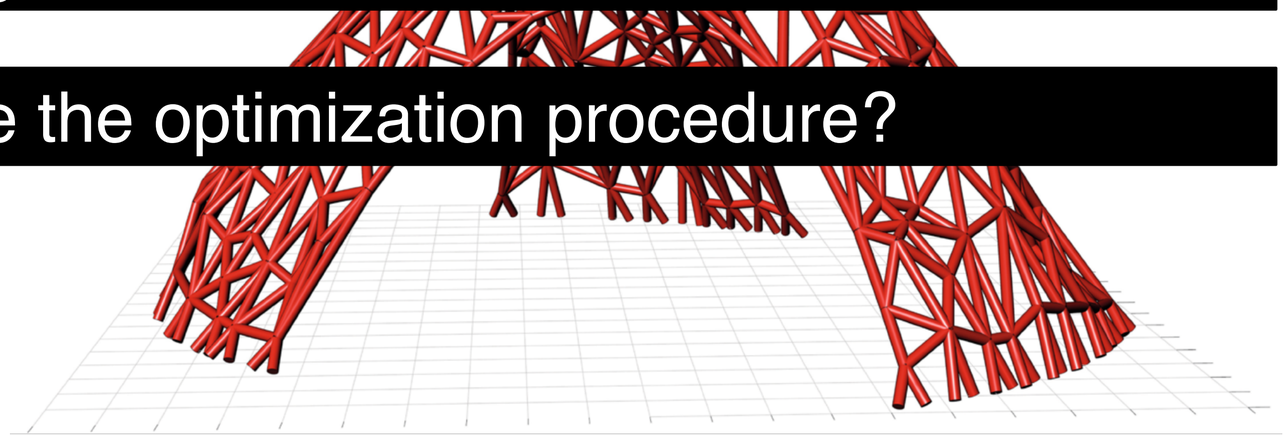
$$\min \sum_i V_i = \min \sum_i A_i \, l_i = \min \frac{1}{\sigma} \sum_i |f_i| \, l_i$$

**Challenges**

1. The Optimum load path depends on the chosen independent force densities!
2. Number of independent sets is exponentially increasing!
3. Not all the optimum results are interesting forms!

**General Idea**: Can we guide the optimization procedure?

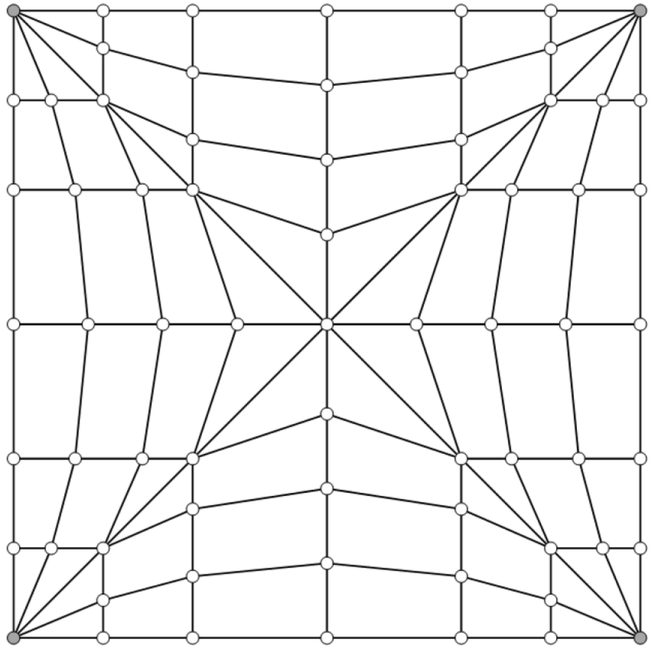




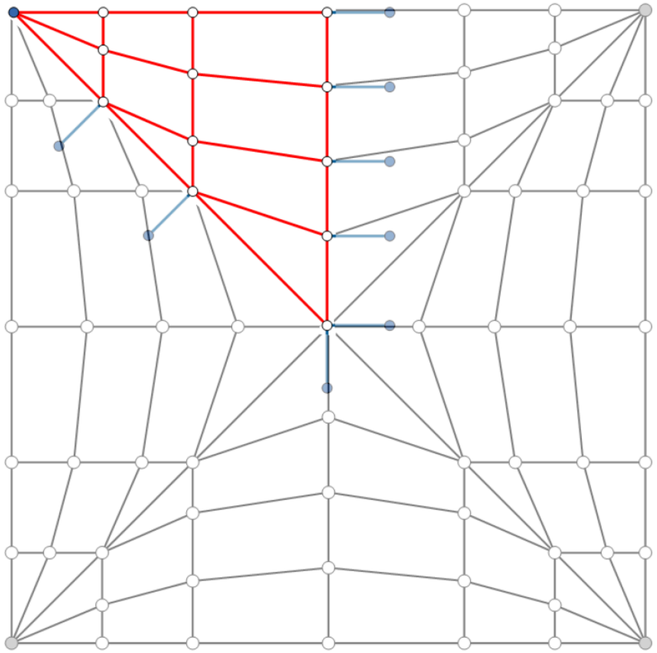
Liew, A., Avelino, R., ***Vahid Moosavi***, Van Mele, T. and Block, P., Optimising the load-path of compression-only thrust networks through independent sets, (Accepted at Structural and Multidisciplinary Optimization) Chair for Structural Design and Block Research Group

# Machine Learning Guided Optimization

Large space to analyze



Original Network
M = 140 edges
K = 20 independent edges

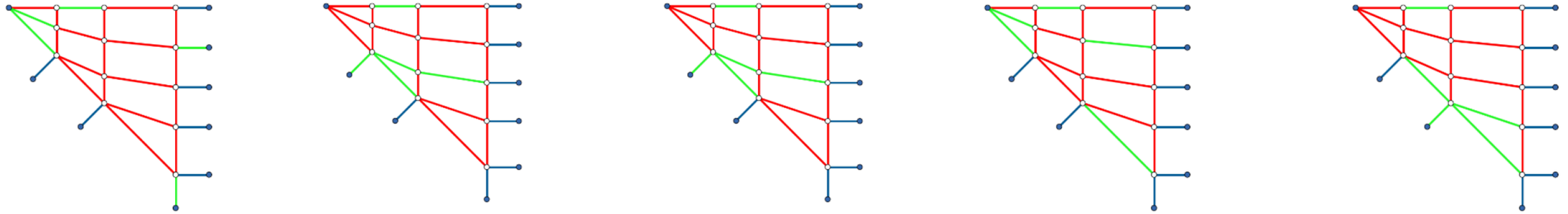


Symmetrical
M = 29 edges
K = 5 independent edges

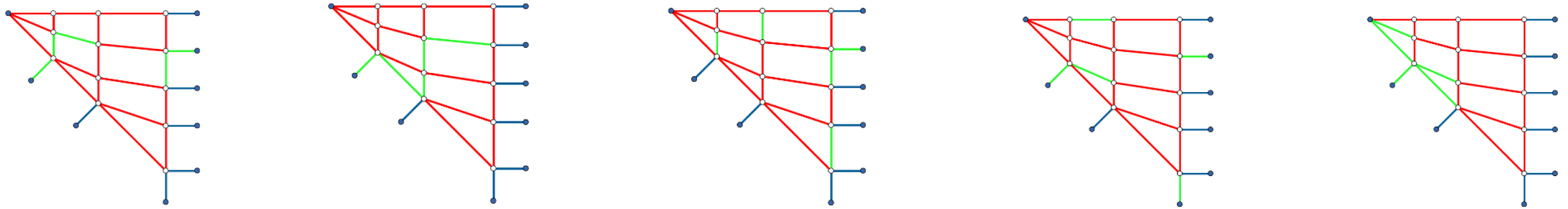
After the symmetry the problem presents:
118755 combinations of 5 different edges

# Machine Learning Guided Optimization

5 “best” set of independents – led to good results

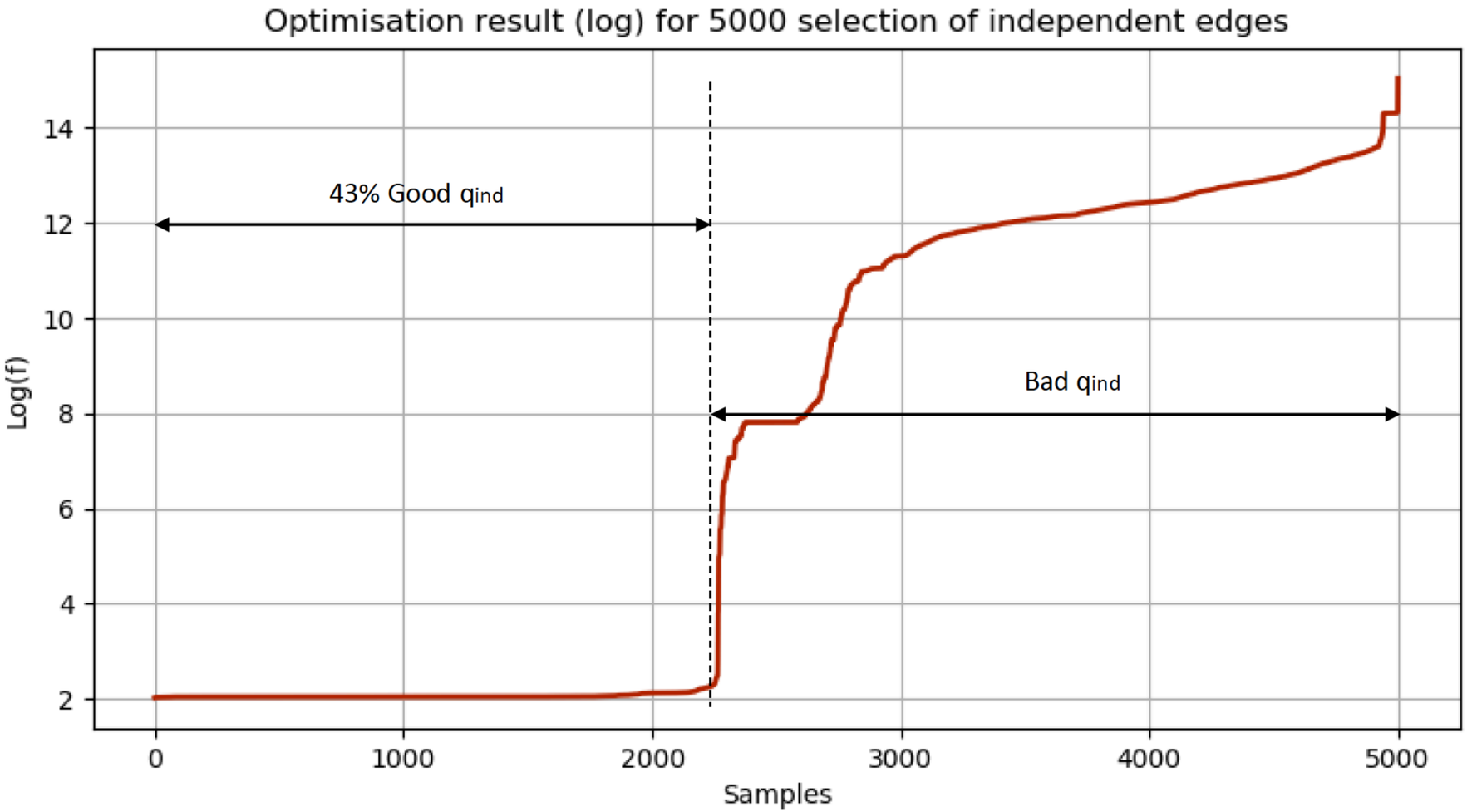


5 “worst” set of independents – led to bad results



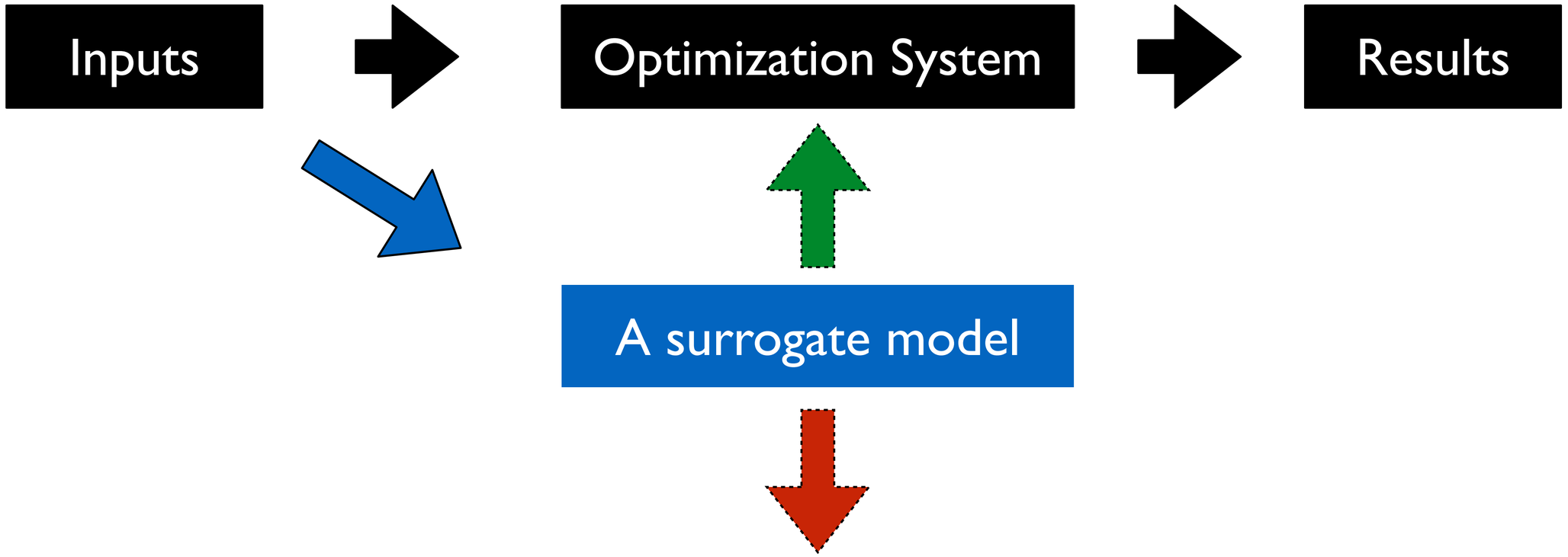
Independent edges  Normal edges  Symmetric edges

# Machine Learning Guided Optimization

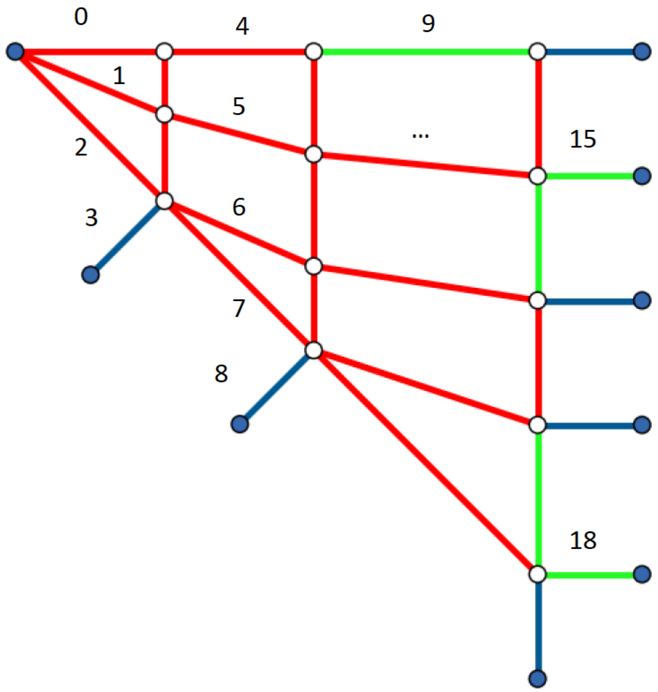


# Machine Learning Guided Optimization

General Idea: What if we learn the behavior of optimization procedure?



# Machine Learning Guided Optimization

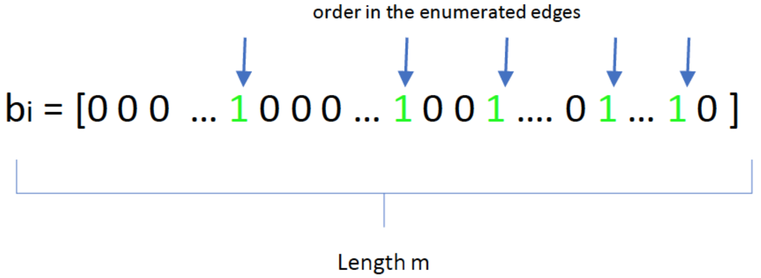



Independent edges

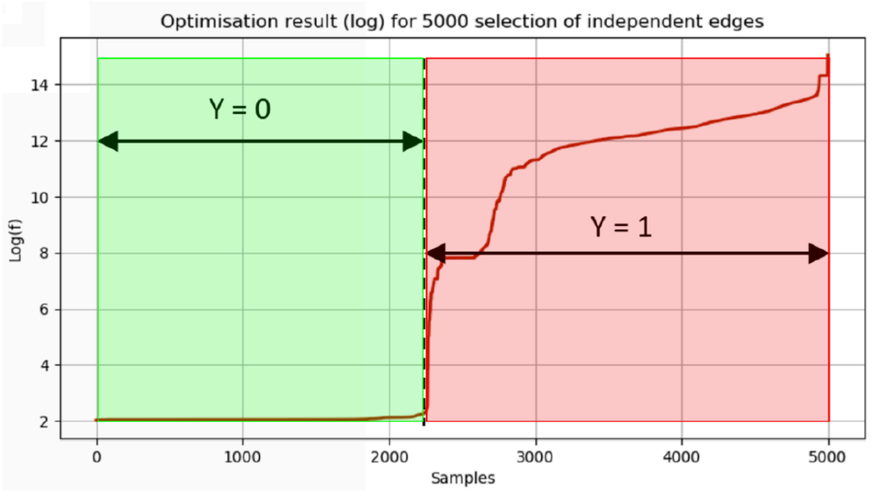
Normal edges

Symmetric edges

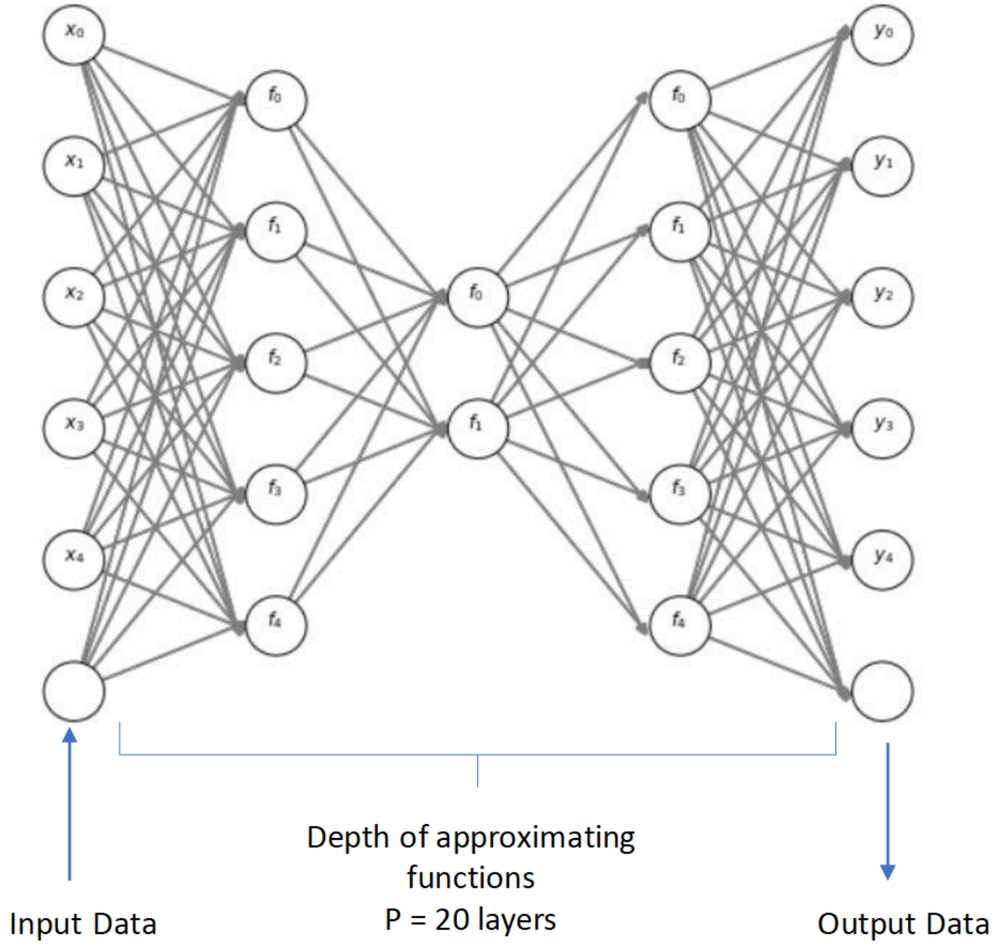
Feature:
Independent edges in a binary format

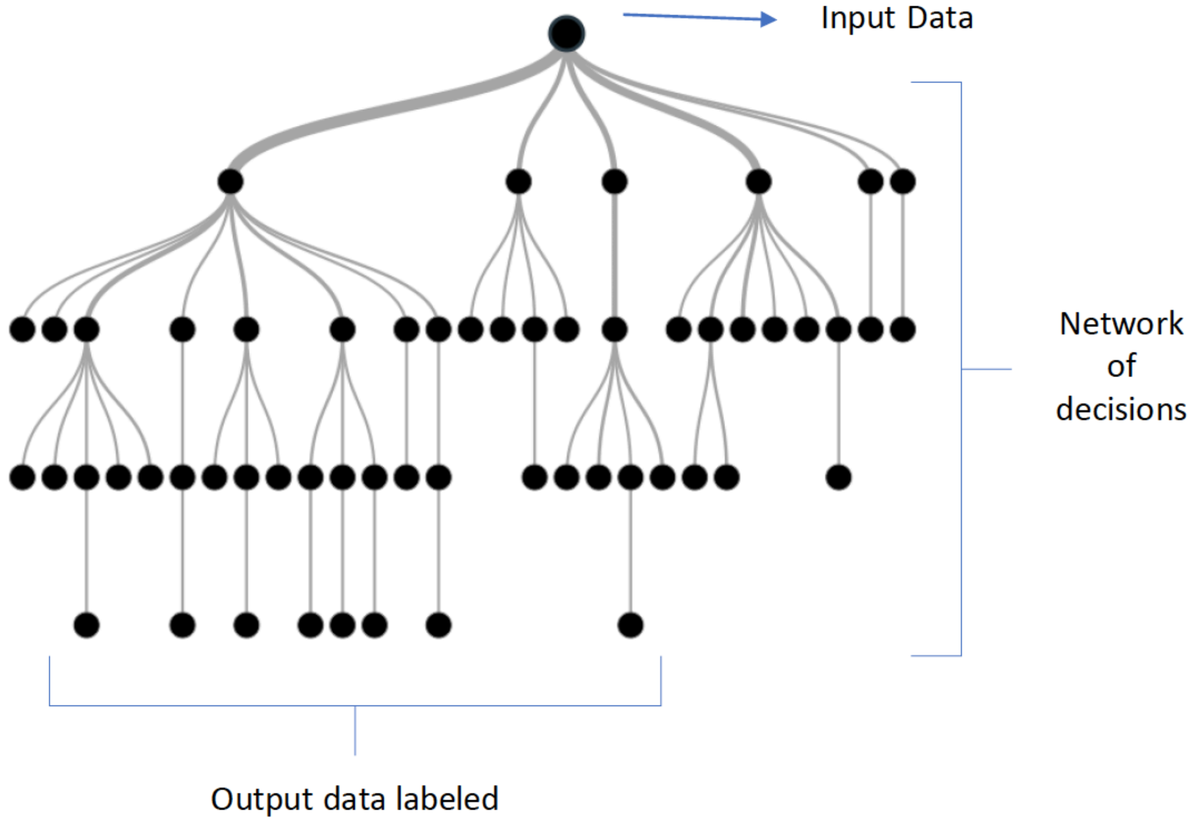



Label:
optimisation result: Load-path value simplified to 0/1 information meaning good or bad pattern



# Machine Learning Guided Optimization



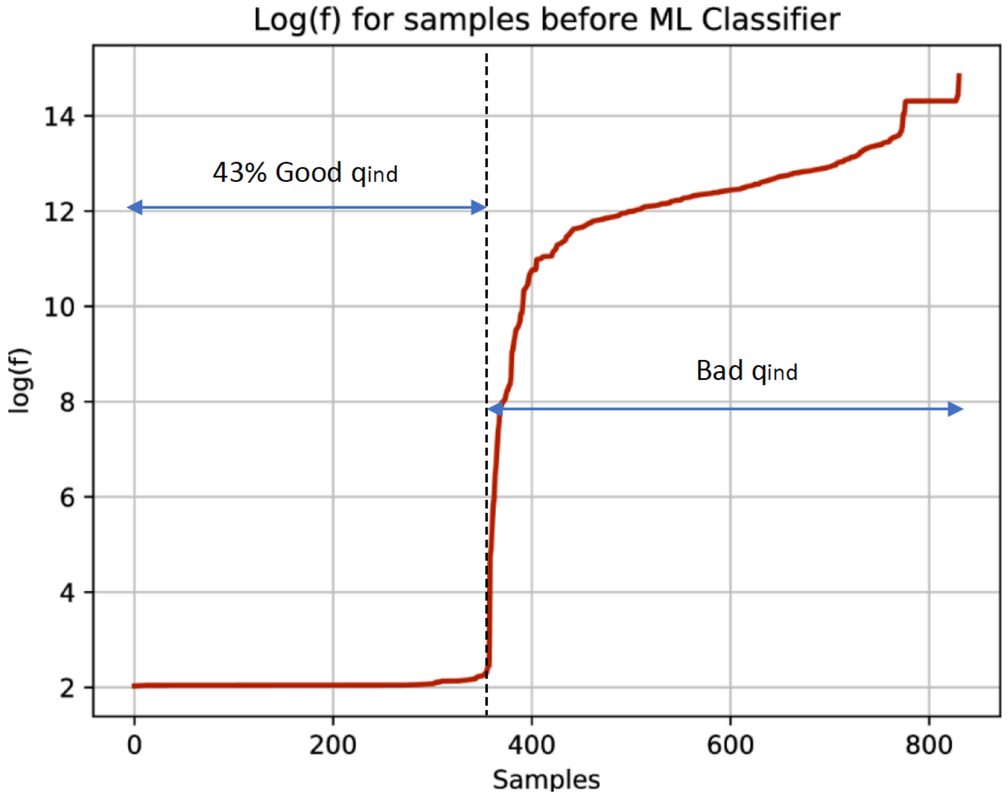




## Neural Networks

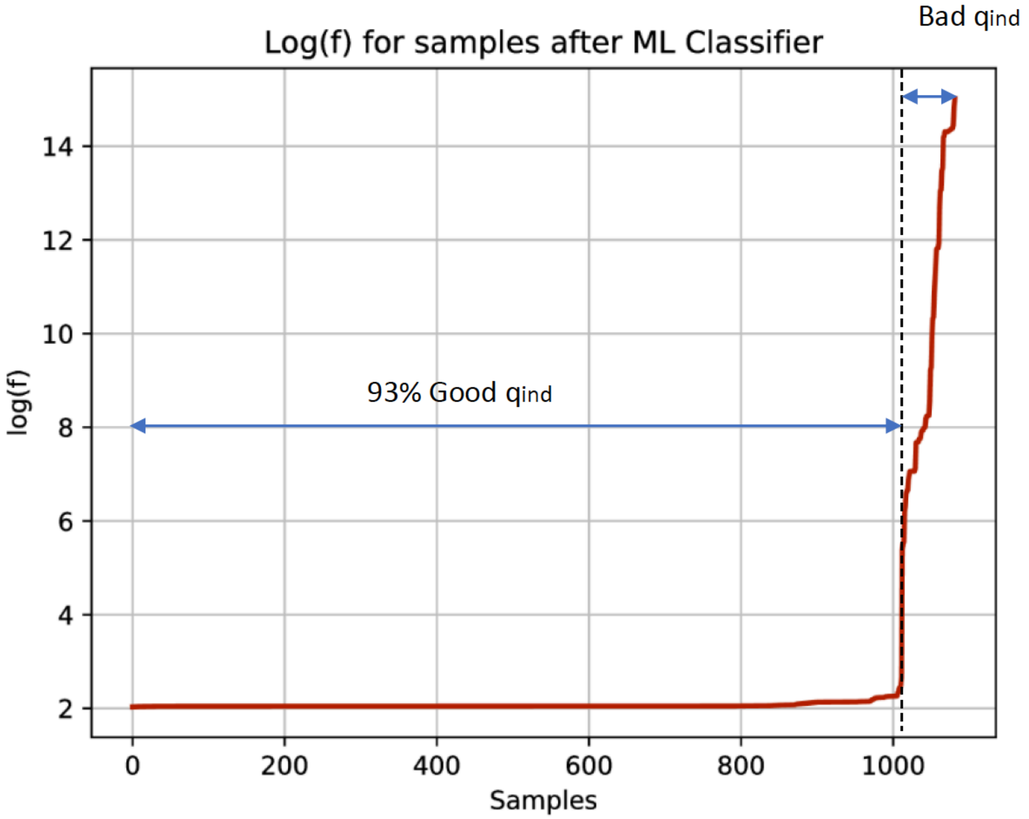
## Random Forest

# Machine Learning Guided Optimization

**Learning process made with 831 randomized samples** → **Learning** → **Calculation proceeded with samples accepted by the machine**

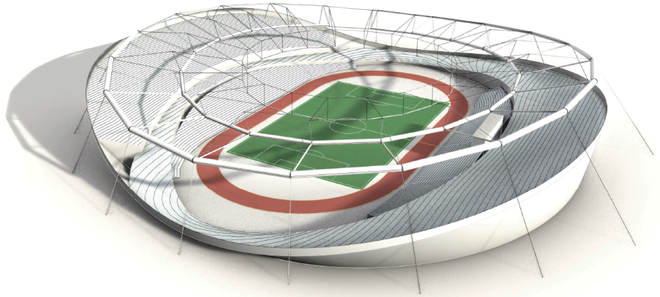


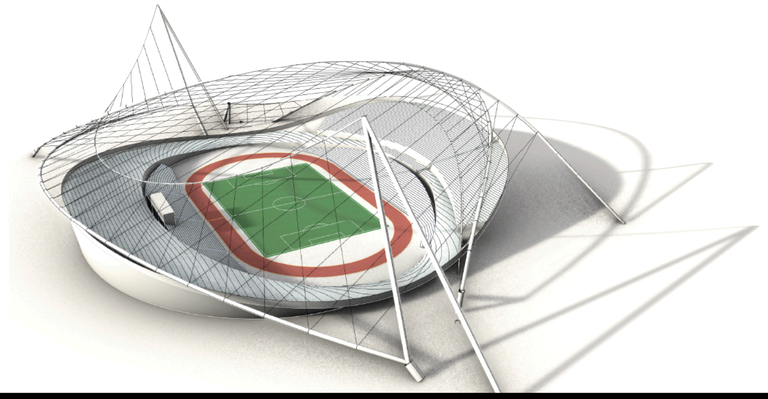

Randomised $q_{ind}$ presented typically 43% of success "Good Samples"

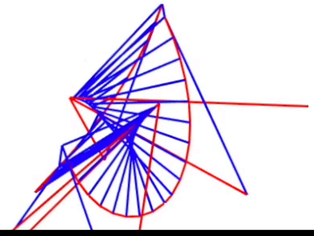



The machine identifies "Good loking" $q_{ind}$ and presents 97% of success

# (2) Fast Parametric Models, But Hard to Control the Outcome!







How to **learn the nonlinear relation** between force and form?

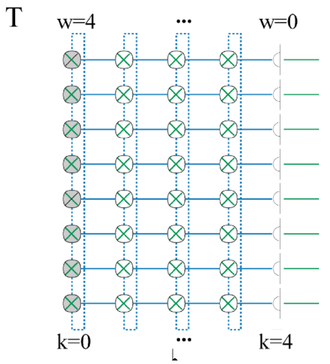
How to **explore the design space based on user preferences**?

Form generation in the context of graphic static

Collaborators: 1-Chair for Structural Design ETH Zurich, 2- Block Research Group, ETH Zurich, and 3- PSL, University of Pennsylvania: 3 PhD Students + 2 Master Students

# Data-Driven Structural Design Beyond Optimization

**… Rather than trying to understand the one to one relations.**




(a)

**Generating thousands of forms via initial random inputs**

(c)

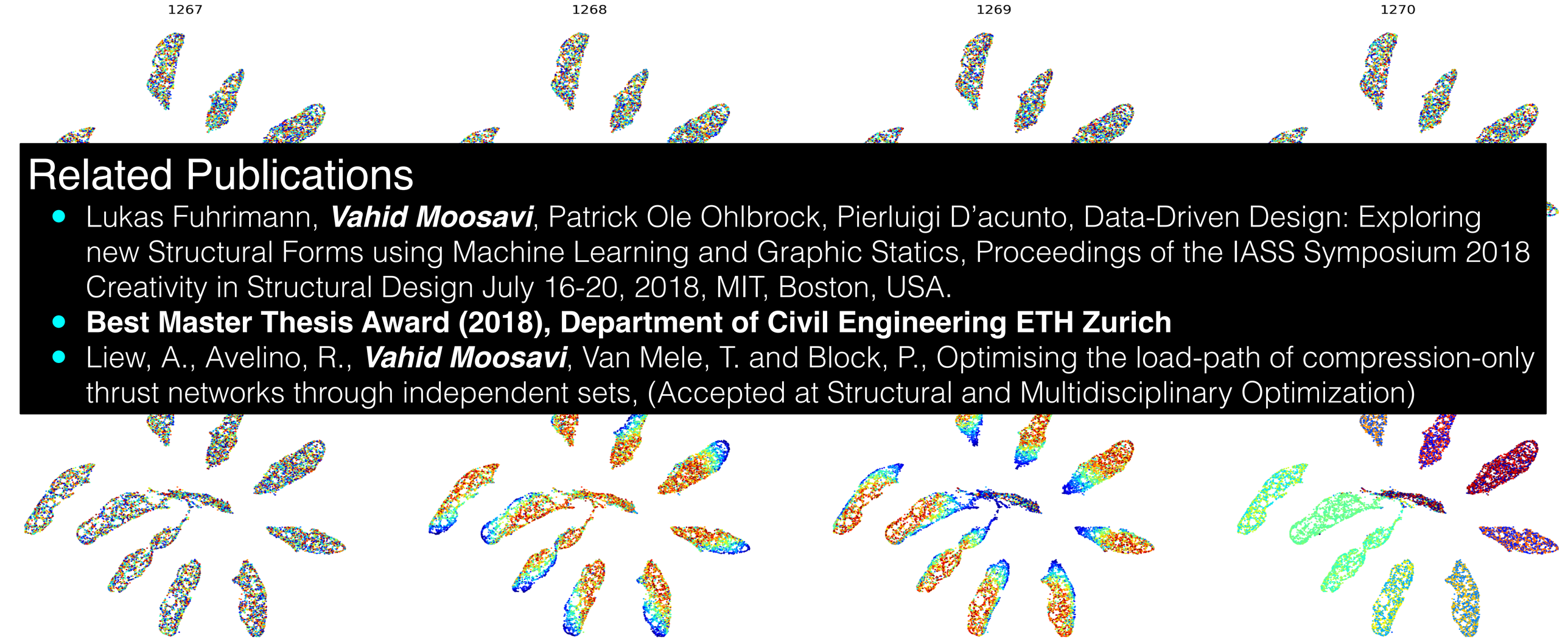
**Learning to generate what user prefers**
**(Dynamic prediction/classification)**

(b)

**Exploring the high-dimensional form space**
**(Dynamic dimensionality reduction and clustering)**

# (2) Data-Driven Structural Design Beyond Optimization

**Dimensionality Reduction: The effect of design variables on the final geometric forms**




## Related Publications

- Lukas Fuhrimann, ***Vahid Moosavi***, Patrick Ole Ohlbrock, Pierluigi D'acunto, Data-Driven Design: Exploring new Structural Forms using Machine Learning and Graphic Statics, Proceedings of the IASS Symposium 2018 Creativity in Structural Design July 16-20, 2018, MIT, Boston, USA.
- **Best Master Thesis Award (2018), Department of Civil Engineering ETH Zurich**
- Liew, A., Avelino, R., ***Vahid Moosavi***, Van Mele, T. and Block, P., Optimising the load-path of compression-only thrust networks through independent sets, (Accepted at Structural and Multidisciplinary Optimization)

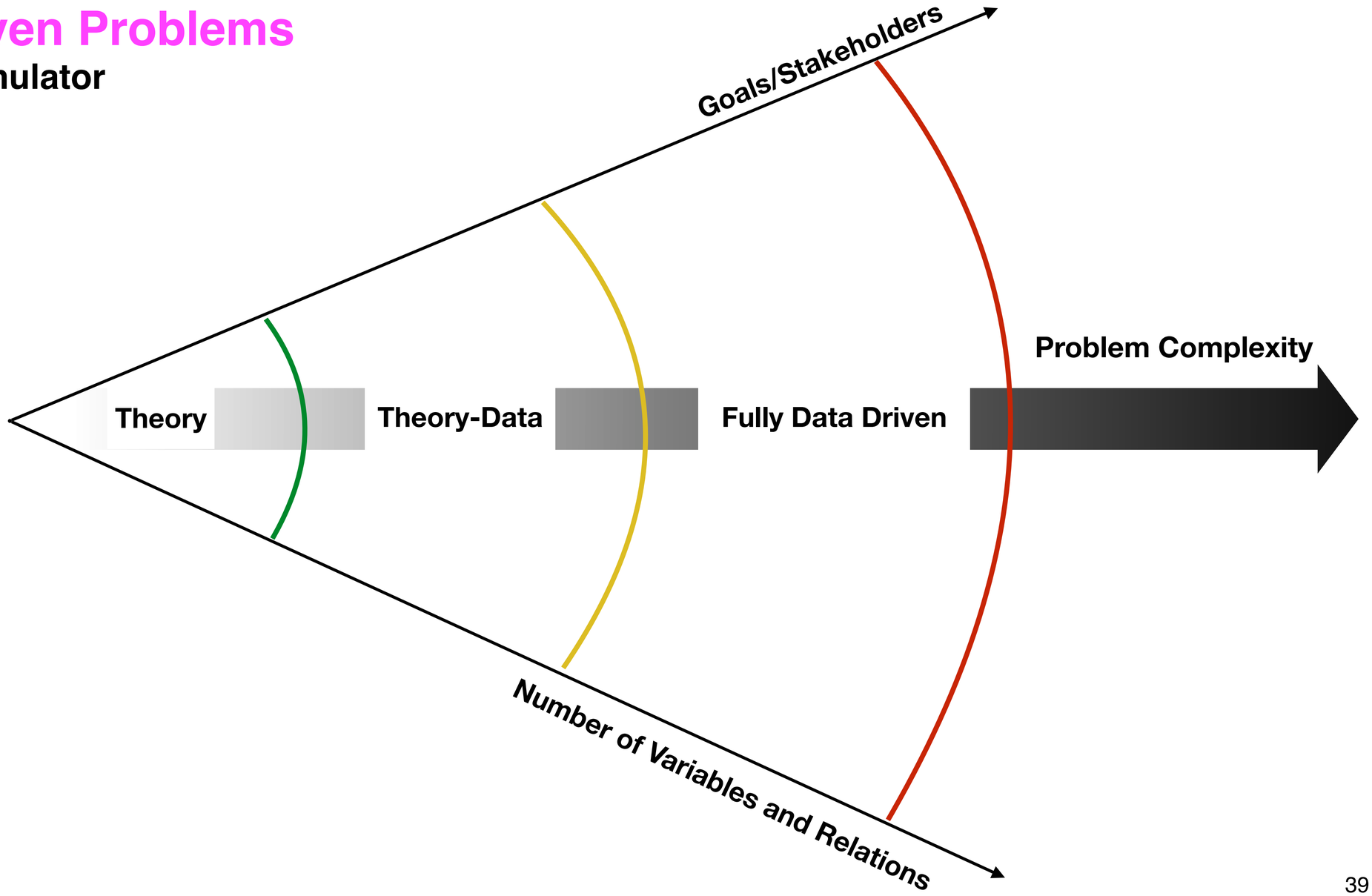
Collaborators: 1-Chair for Structural Design ETH Zurich, 2- Block Research Group, ETH Zurich, and 3- PSL, University of Pennsylvania: 3 PhD Students + 2 Master Students

# Majority Of Data Driven Problems

**Out of Scope of a Single Simulator**

**+**

**Between Several Domains!**



# (4) Planetary Urban Modeling (Urbanized Planet as Internet)

Lightweight Software Applications

Question (Data) Driven

Challenges/ Goals

Energy
Density
Water
Spatial form
People
Air
Economy
Heat
Real estate
Planning
Spatial distribution of resources

Technology

Simulations
Data Fusion
Machine Learning
Crawlers
APIs
Data dumps

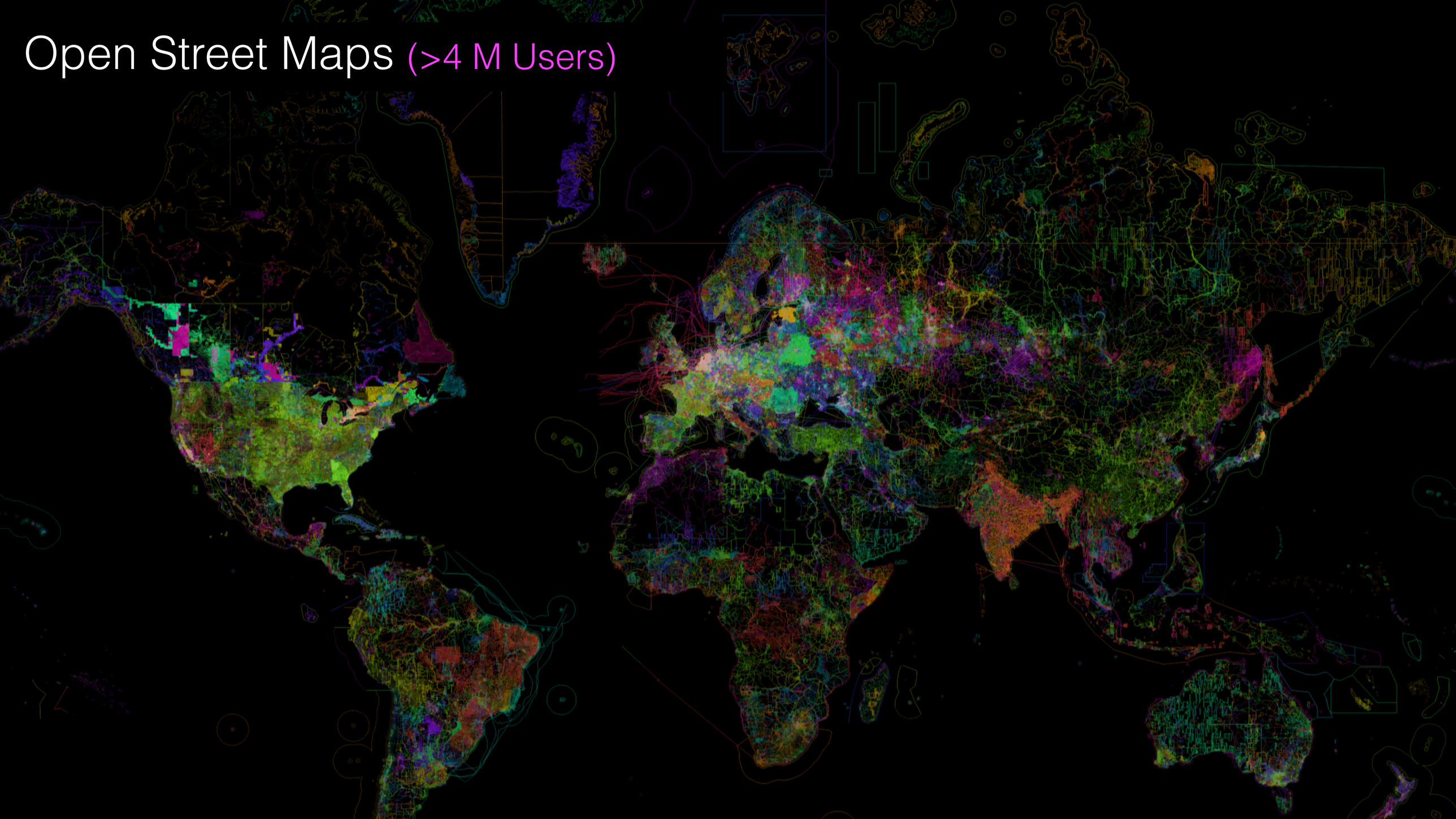
Data Streams

planet
Descartes Labs
NOAA
UN HABITAT FOR A BETTER URBAN FUTURE
NASA
USGS science for a changing world
earth observatory
opernicus Europe's eyes on Earth
UNDP
UNITED NATIONS
esa
DigitalGlobe
OPEN DATA CUBE
OpenStreetMap
mapbox
Open Governments
National Census
IOT
Local sensors
tripadvisor
Real estate portals
Hotels
airbnb
UBER

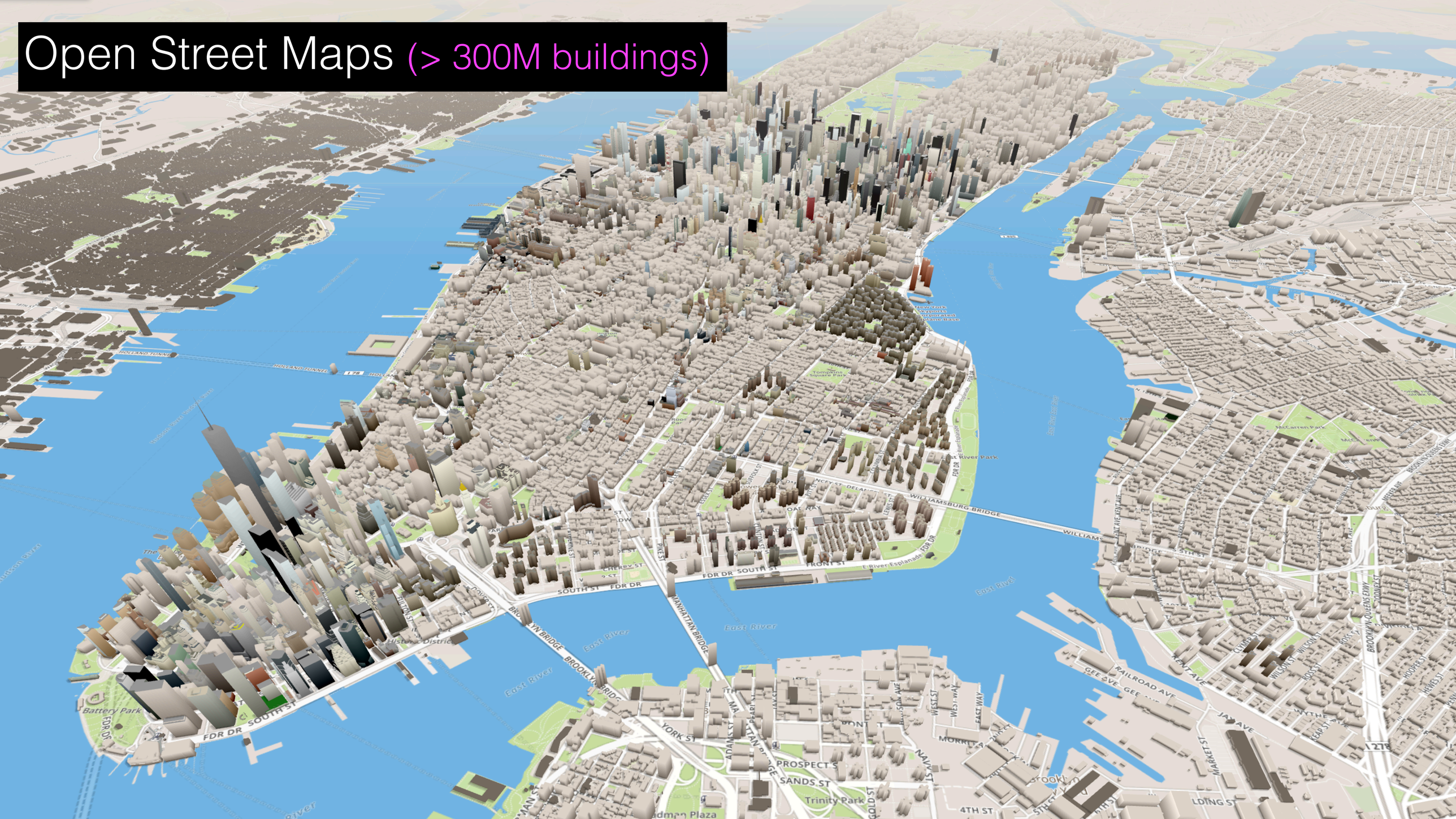
Complicated Software Applications

Model Driven

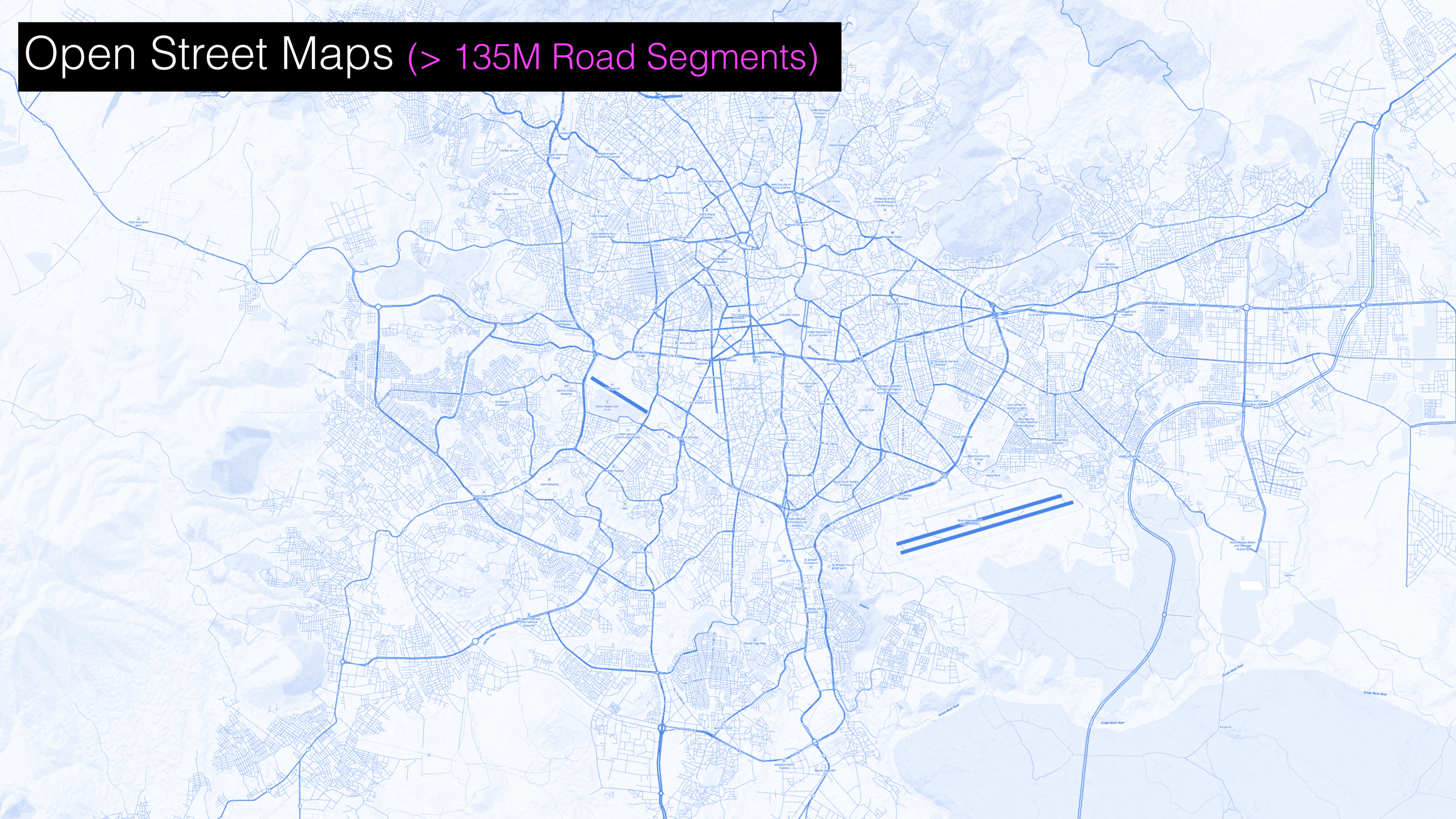
# Open Street Maps (>4 M Users)

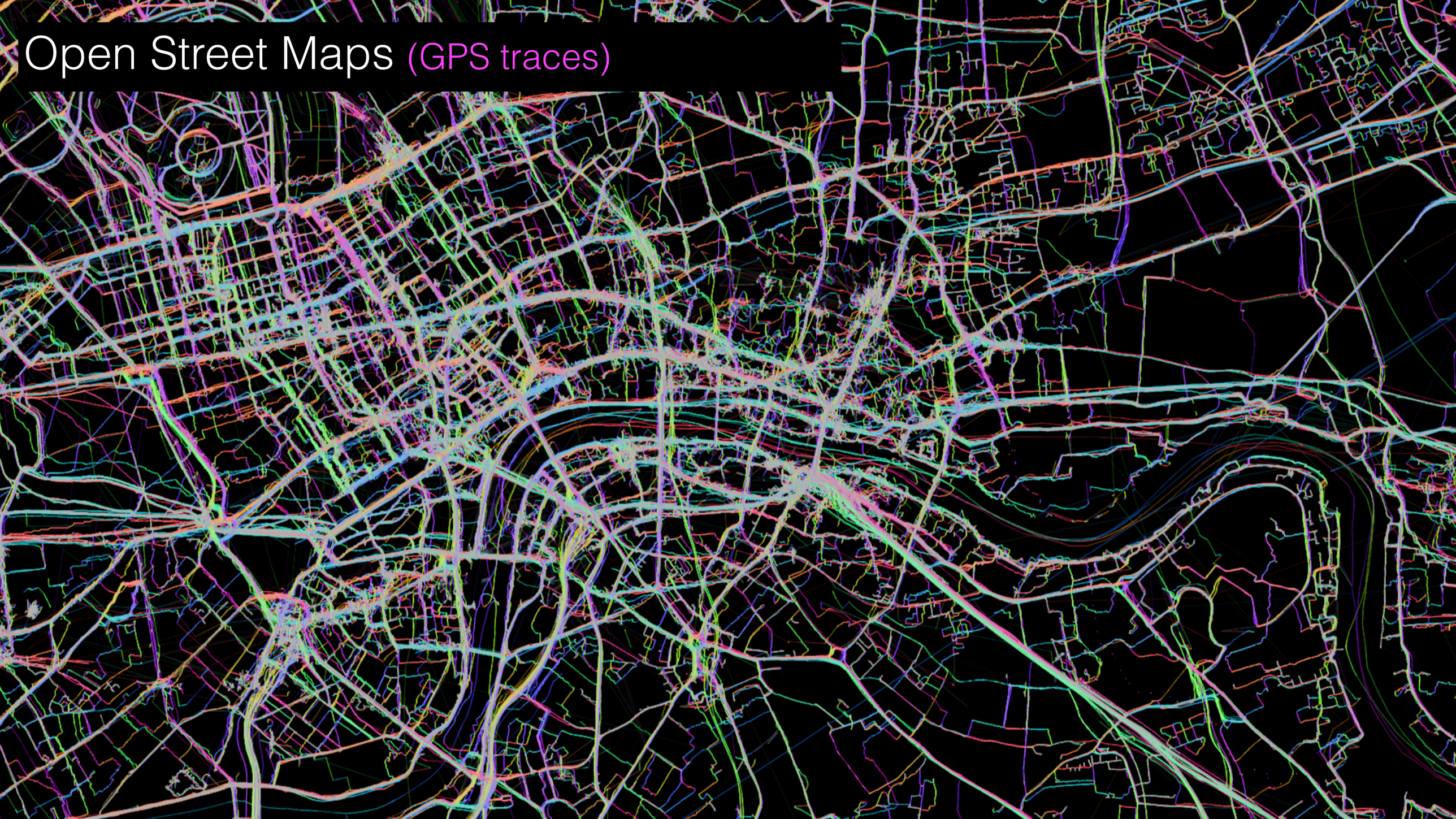


# Open Street Maps (> 300M buildings)

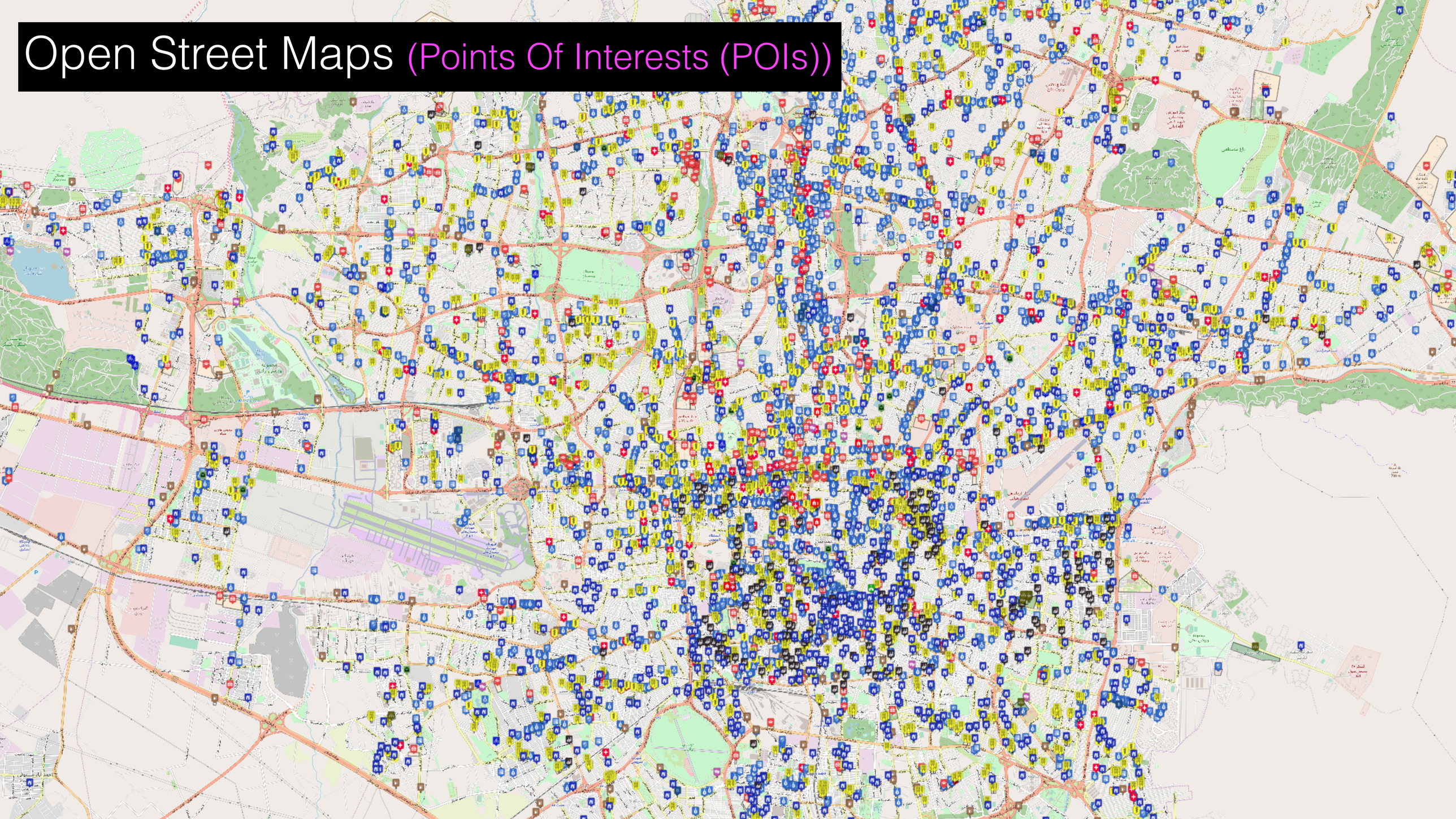


# Open Street Maps (> 135M Road Segments)




Open Street Maps (GPS traces)

# Open Street Maps (Points Of Interests (POIs))



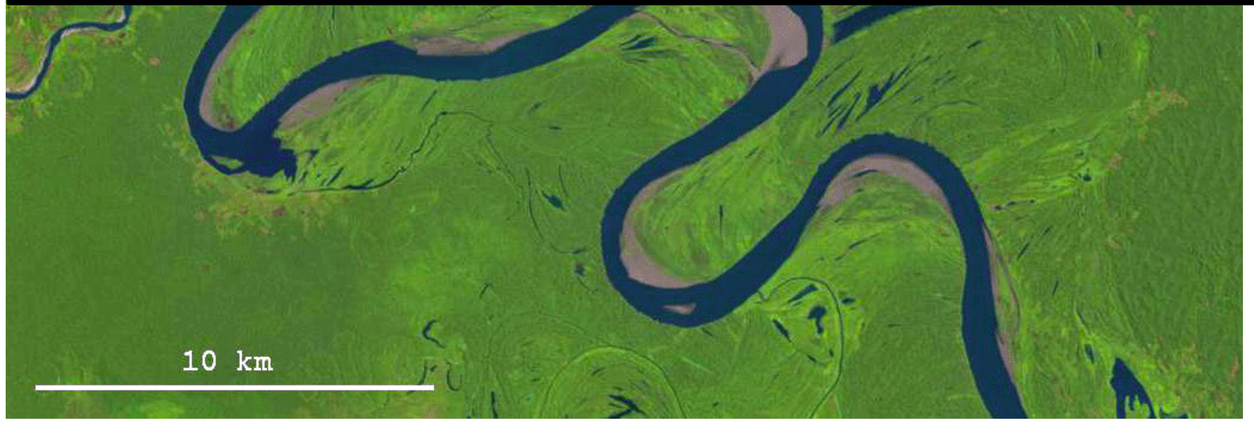
# Open Street Maps (Waterways=river>1.1M)

https://taginfo.openstreetmap.org/keys/waterway#values

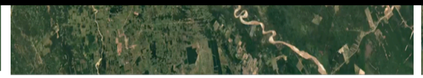

## Monitoring at the scale of planet?

## Dynamics of Rivers and Lakes

## Urbanization speed







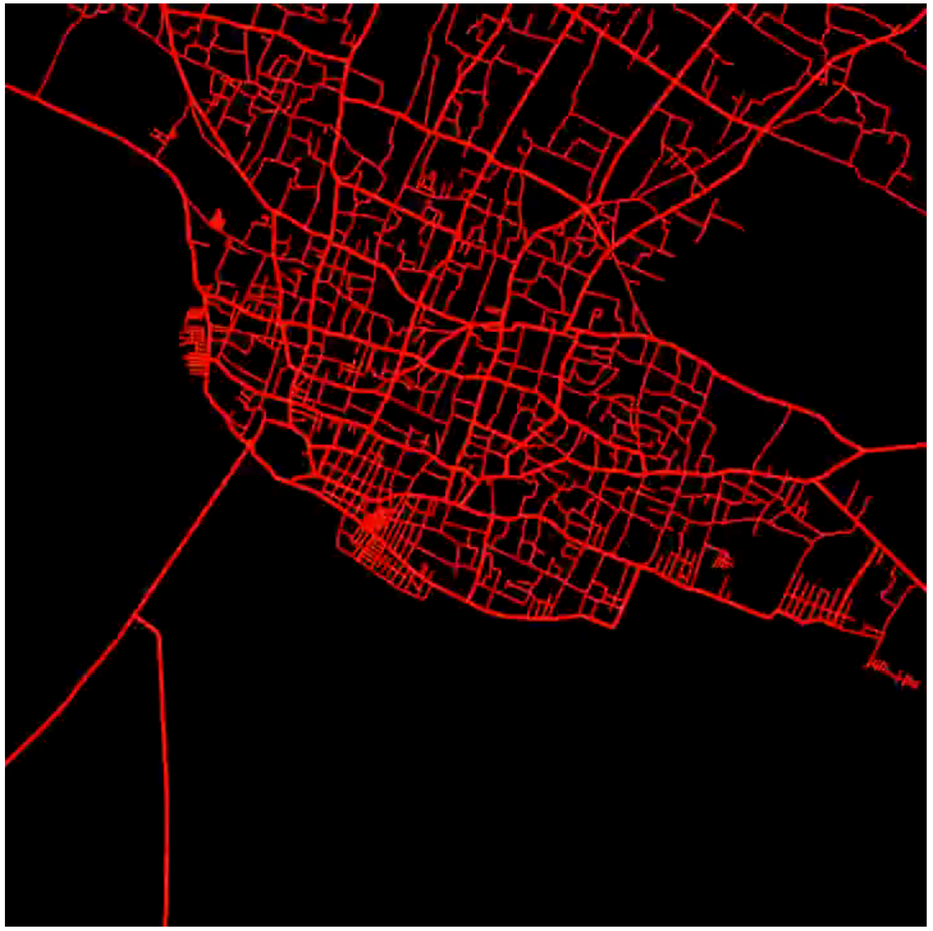


# (4) Planetary Urban Modeling: Urban Form Across the Planet

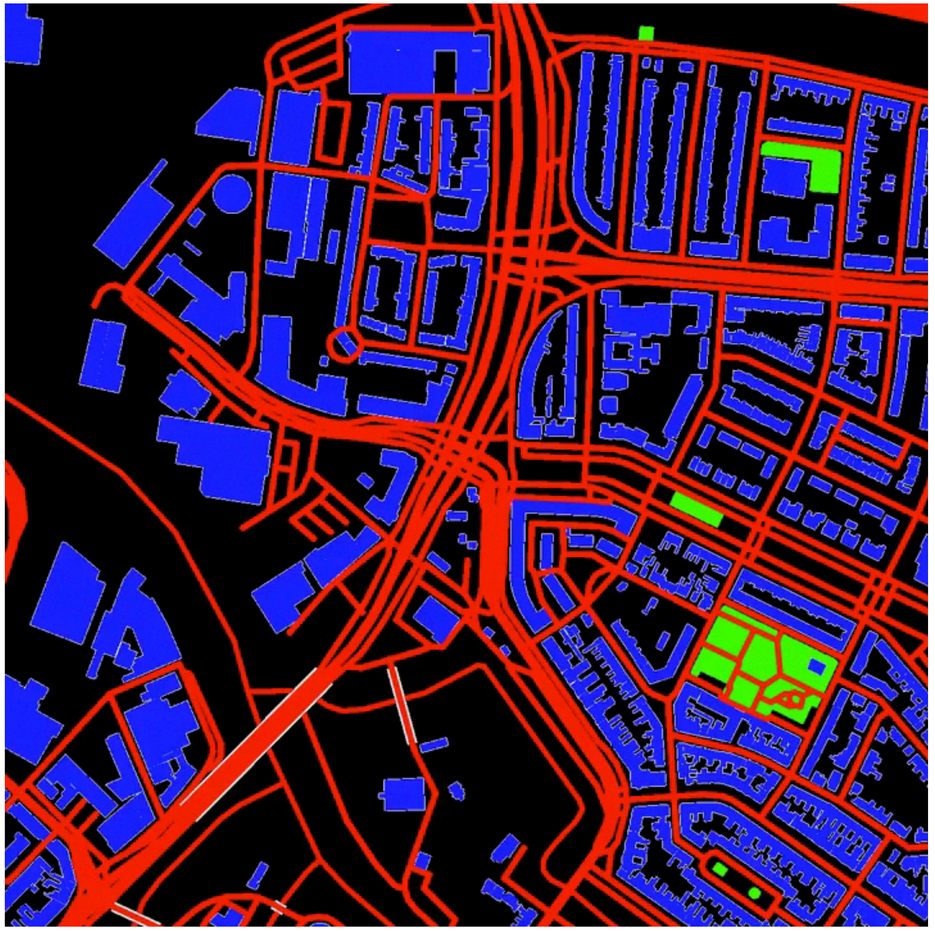
**Urban Forms of 1.1 Million Locations, Freely Available…**

# (4) Planetary Urban Modeling: Urban Form Across the Planet

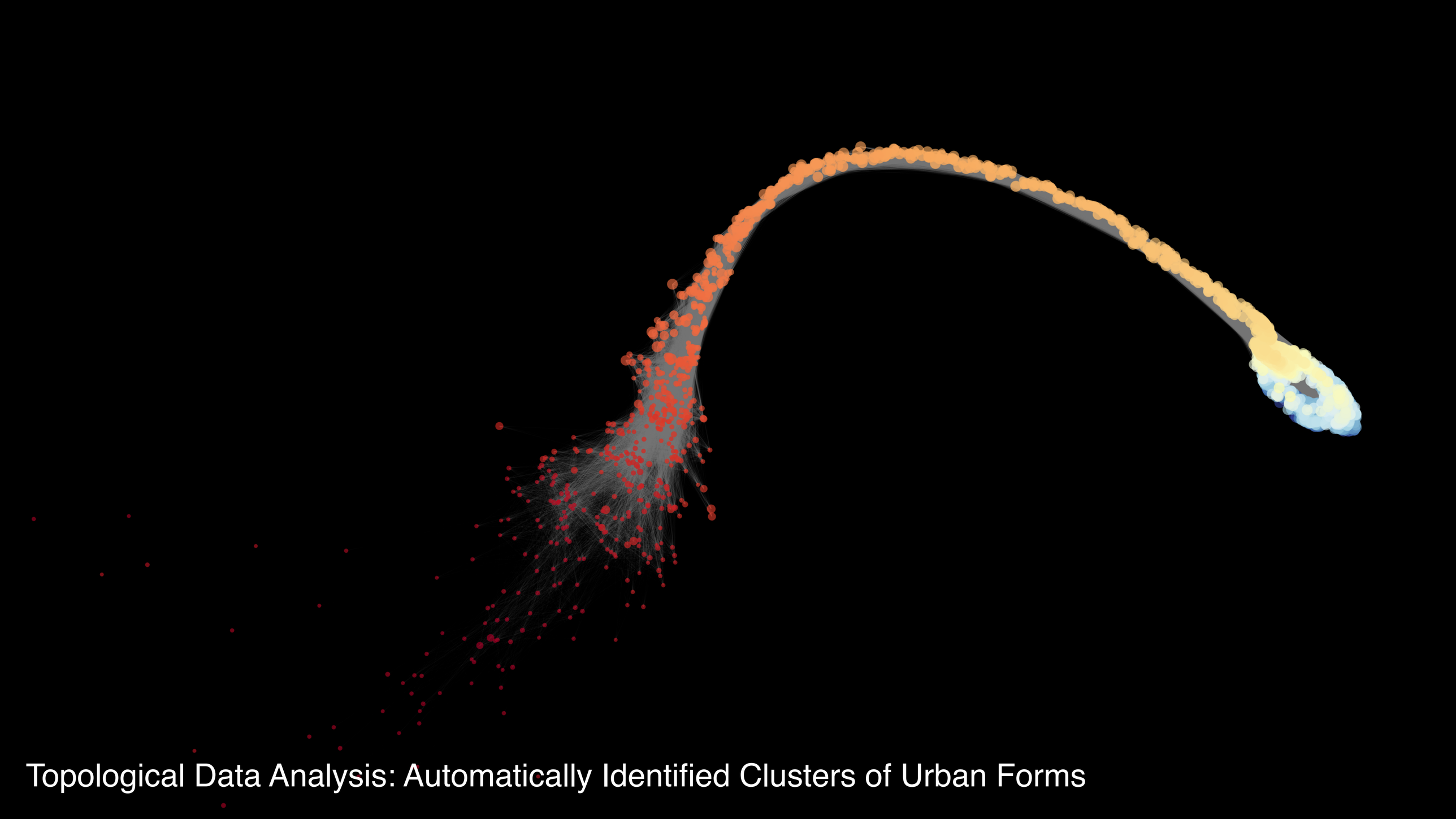
**Urban Forms in 1.1 Million Locations + Deep Learning from Computer Vision**



**Each Frame**



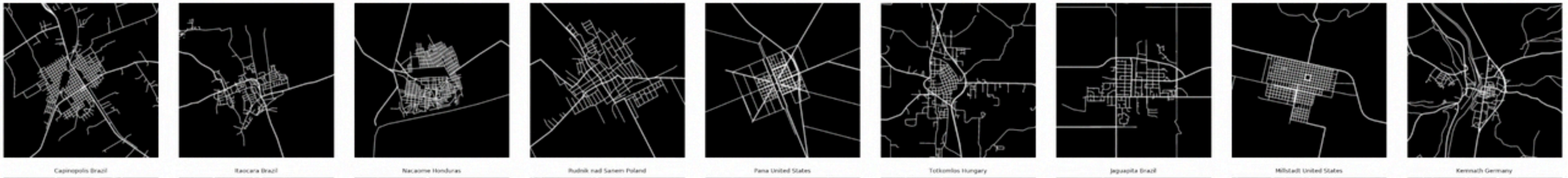
**One Location**



Topological Data Analysis: Automatically Identified Clusters of Urban Forms

# A search engine of complex urban situations

**Can we find Twin Cities in the informational space?**



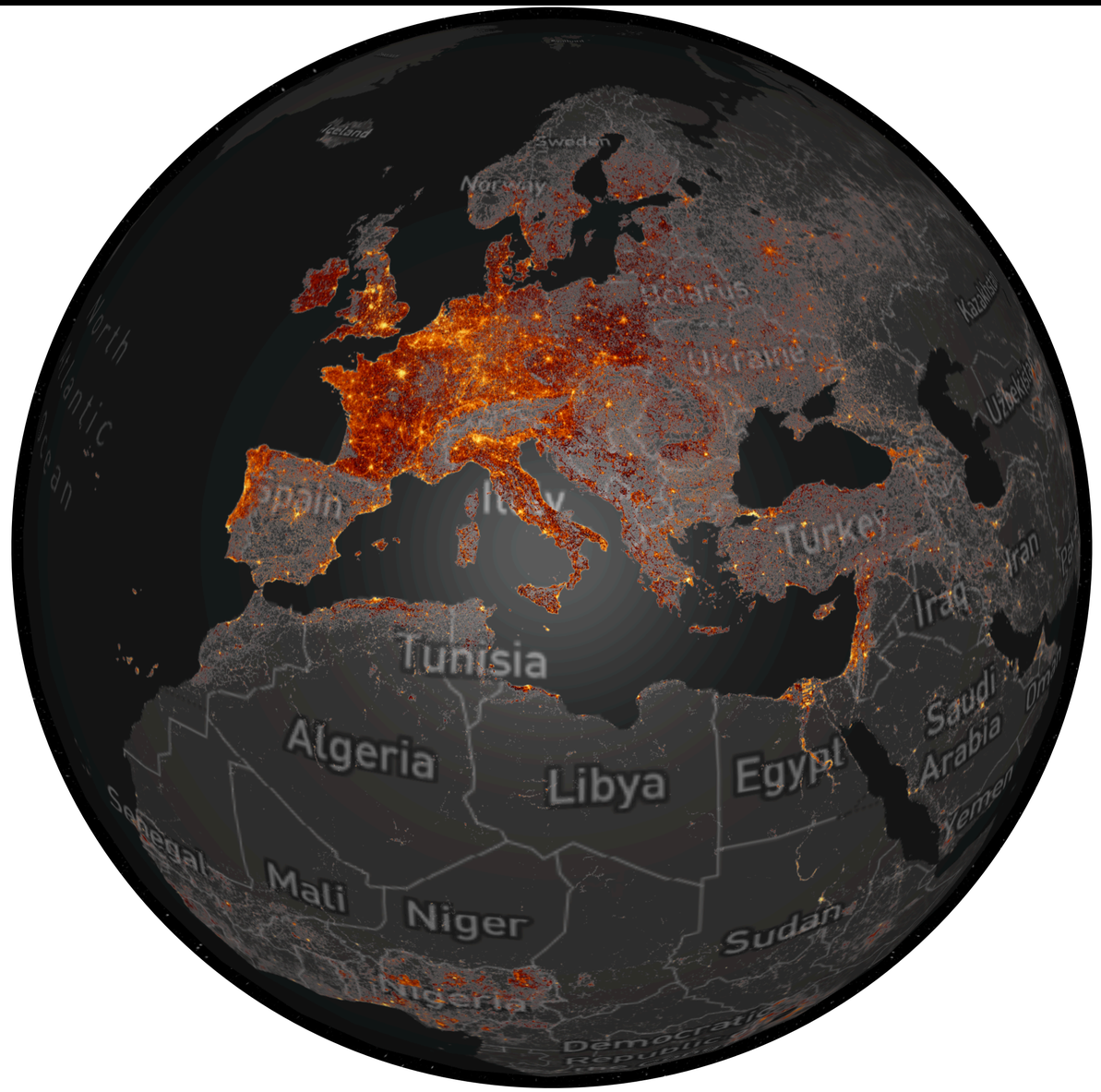

## …As a Comparative Approach: To Train the Models Against a Target: Traffic Quality, Air Quality, Modes of Transport, Poverty, …


Project website: https://sevamoo.github.io/cityastext/

Related Publications

- **Vahid Moosavi**, Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet, https://arxiv.org/abs/1709.02939.

# Urban Density/Diversity/Accessibility: A Global Assessment

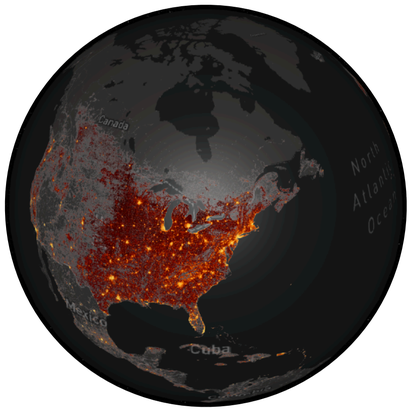


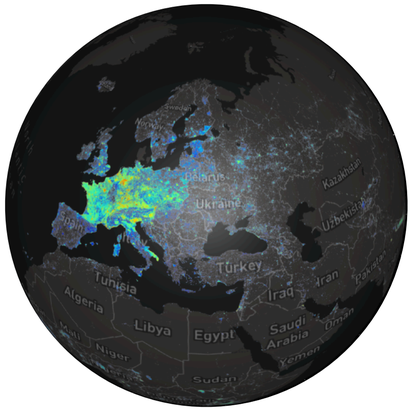

Road Networks, Buildings, Resources,…

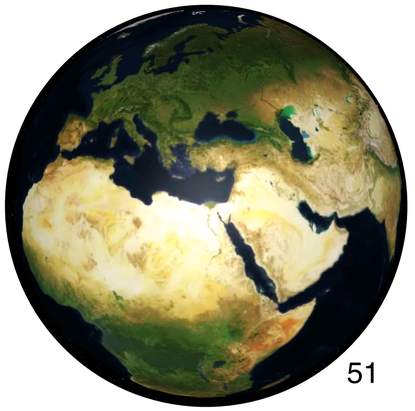
**>150 M** **>320 M** **>50 M**

**Open Source Data**

**Open Source Code**



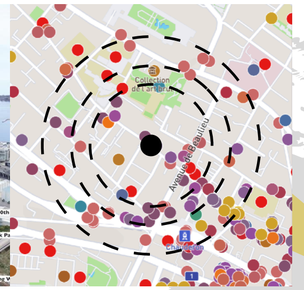



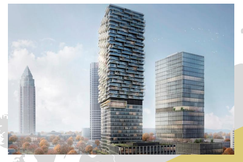



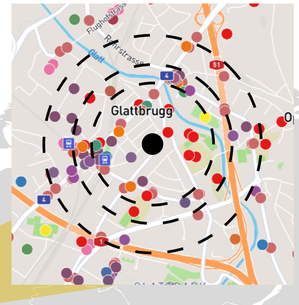
**Project website**: https://sevamoo.github.io/urban-modeling-at-scale/

# A Google Like Search Engine of Spatial Search (Nearly real time)

**Commercial Real Estate Development**: Multidimensional Analysis of Millions of Locations






Distribution of different activities around each POI

- Similar sites to a chosen location out of 20M points

Project website: https://sevamoo.github.io/reference_project/mapboxgl_cluster_INSIDE.html

Collaborators: Swiss Prime Site, and Dr. Christian Kraft, Institute of real estate and financial services, Lucerne University of Applied Sciences, Switzerland

www.vahidmoosavi.com
https://github.com/sevamoo