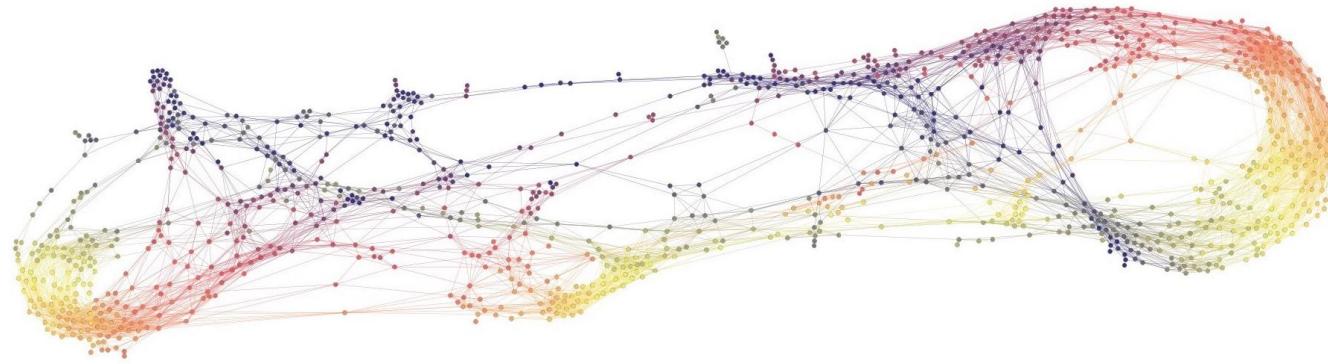


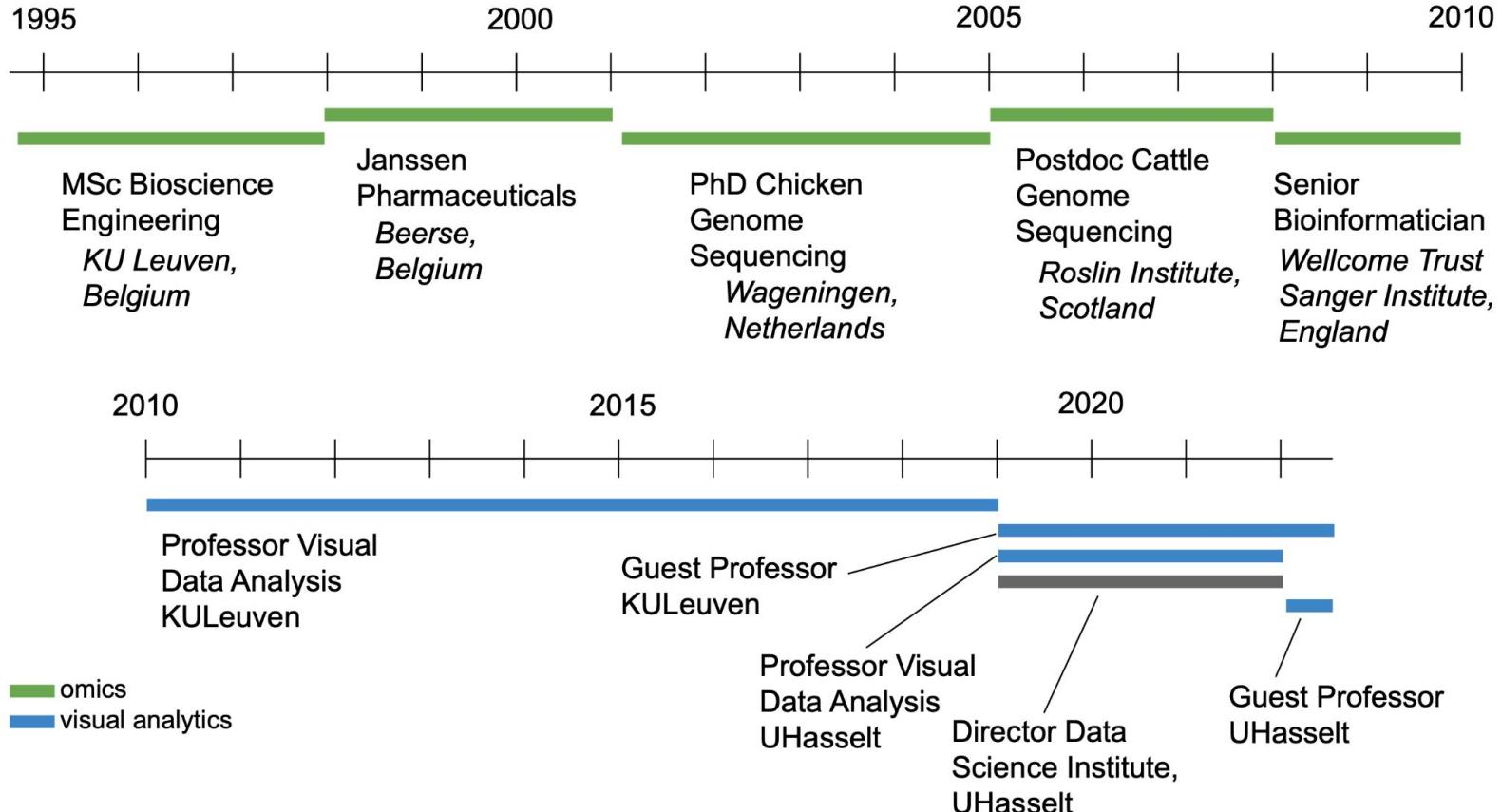
:: Augmented Intelligence ::

Visual Analytics as Hypothesis-Generating Engine



Dr Ir Jan Aerts
Visual Data Analysis Lab
<http://vda-lab.io>
jan.aerts@vda-lab.io

Professional background



Why Augmented Intelligence / Visual Analytics?

Real problems = wicked problems

= problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognise

- finitude: Knowledge or resources are finite
- complexity: Complex interactions between partially nested hierarchies of complex systems with multiple feedback and feedforward loops
- normativity: Human values and norms can be inextricably intertwined with how the problem is formulated or resolved

Wicked ("real") problems (How do we solve climate change?)



Complex problems (How do we reduce CO₂ release in the air?)



Tame/engineering problems (How do we improve efficiency of this engine with 25%?)

Other issues

- thick data: small but rich
- tacit knowledge dismissed
- data too complex to comprehend => embrace complexity?
- algorithmic black boxes => xAI
- blurring responsibility: algorithm vs expert
- hidden assumptions
 - unforeseen consequences of ML
 - reliance on statistics for data exploration

=> decision "support" (?) systems

Humans are better at:

- detecting noise
- perceiving a wide variety of stimuli
- pattern perception
- relevant recall of information
- flexible problem solving
- inductive reasoning
- hypothesising

Machine are better at:

- repetitive and precise operations
- fast response to signals
- performing with high accuracy
- multi-tasking
- deductive reasoning
- logic

data → information → knowledge → understanding



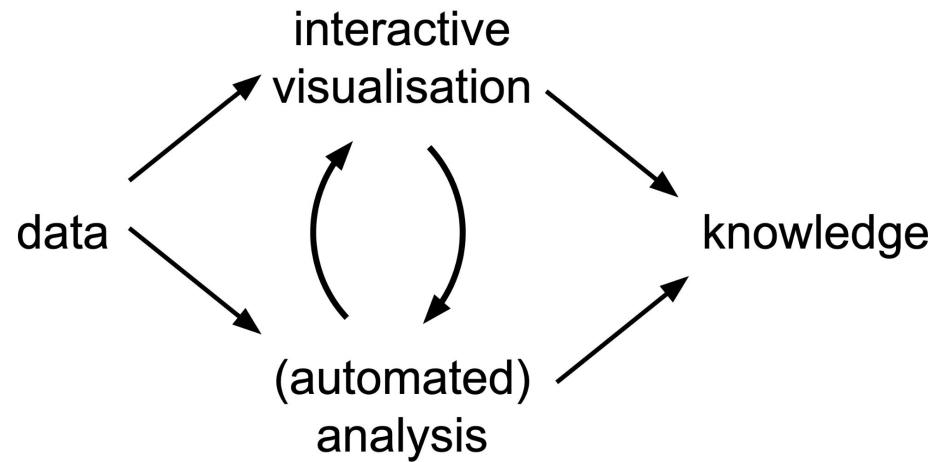
who, what,
where, when

how

why

role of human

Visual Analytics



towards contextualisation: empowering the expert to make better-informed decisions

4 main aspects

1. Data management
2. Novel visual and interaction design
3. Topological Data Analysis
4. Multi-Layer Networks

1. Data management

data modeling of complex data

relational databases

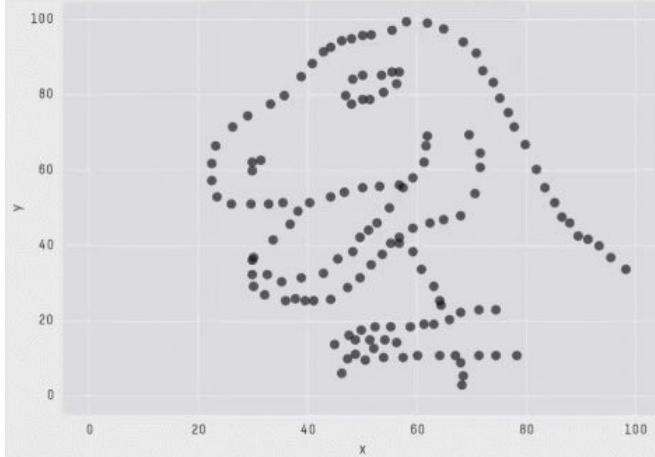
document-oriented databases

graph databases

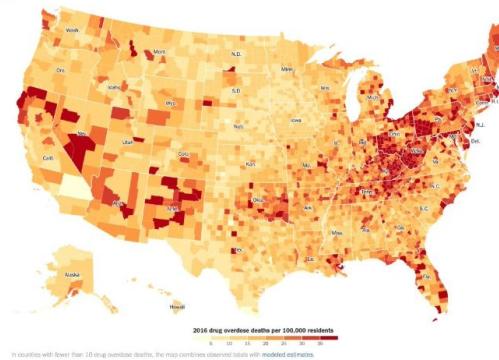
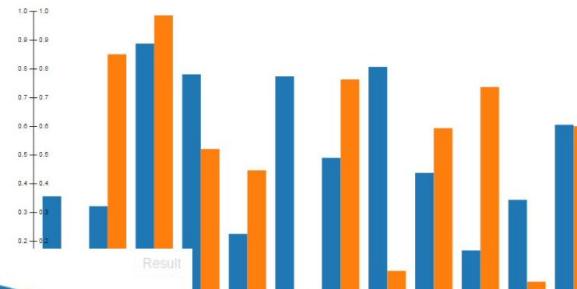
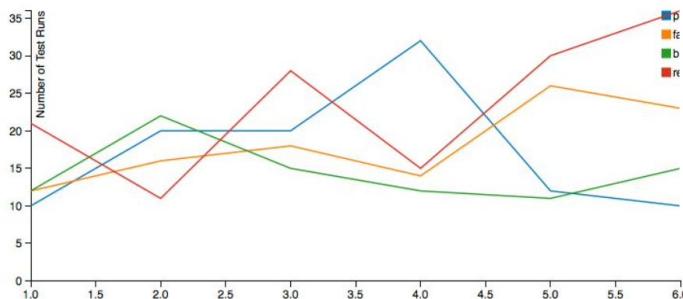
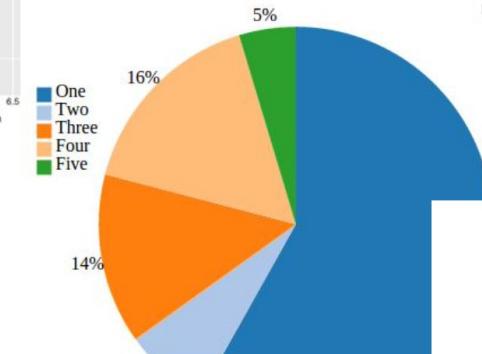
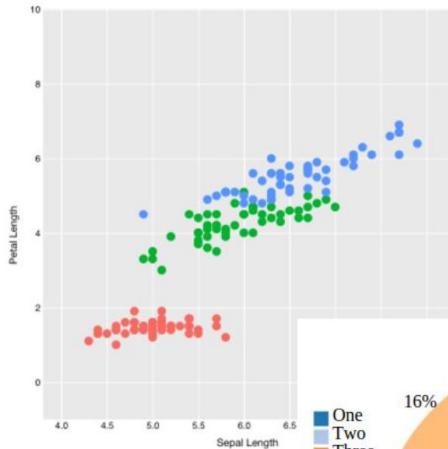


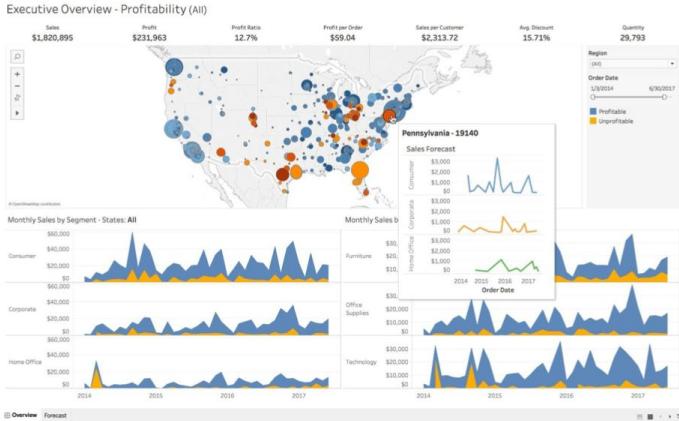
2. Visual and interaction design

X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

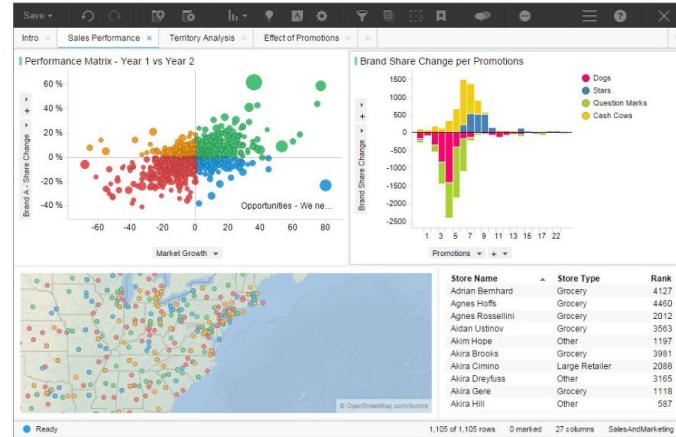


X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526





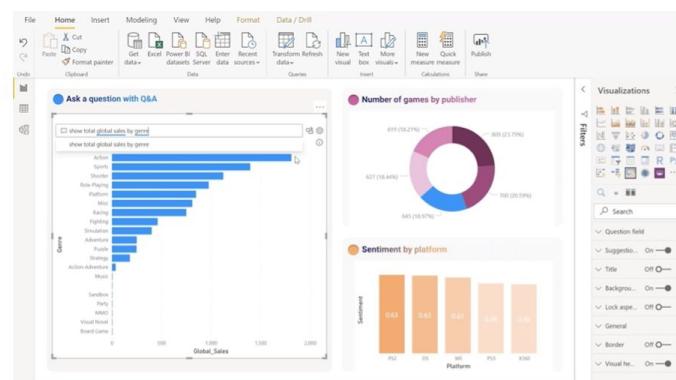
Tableau



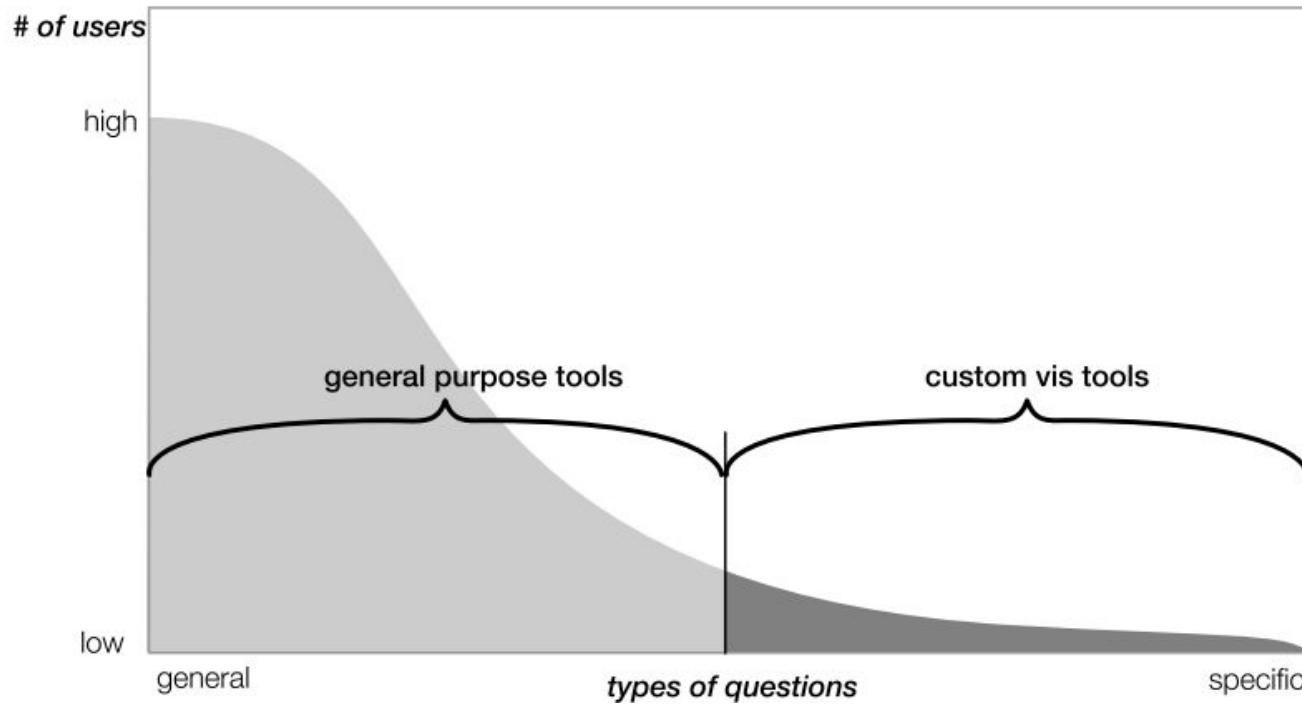
Spotfire



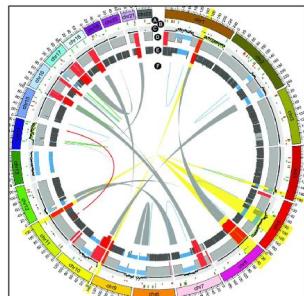
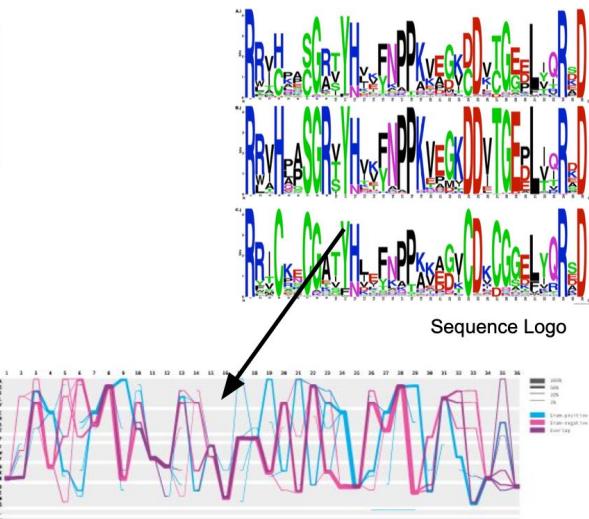
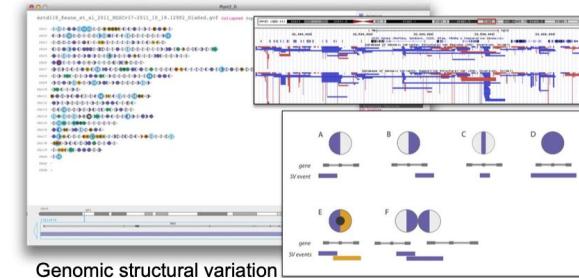
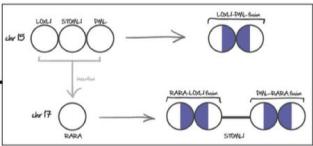
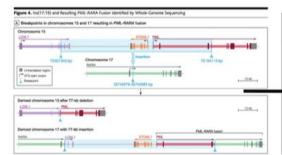
Qlikview



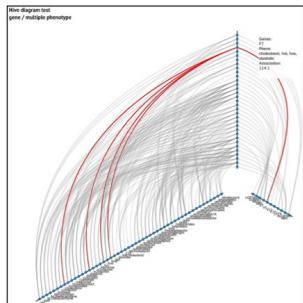
Power BI



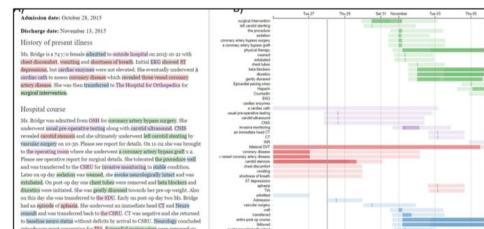
Contextualisation for domain experts



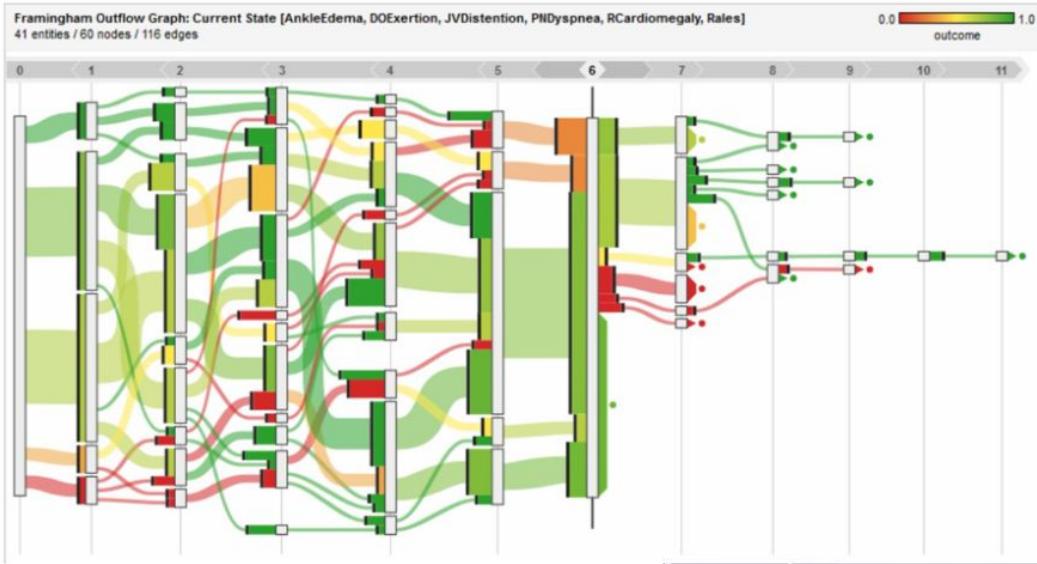
Circos plot
Genomic variation in mouse



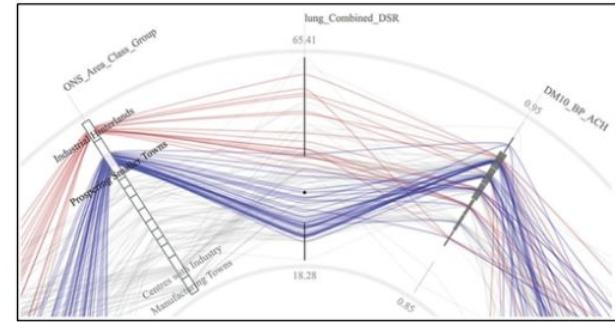
Hive plot
Association single gene - multiple phenotypes



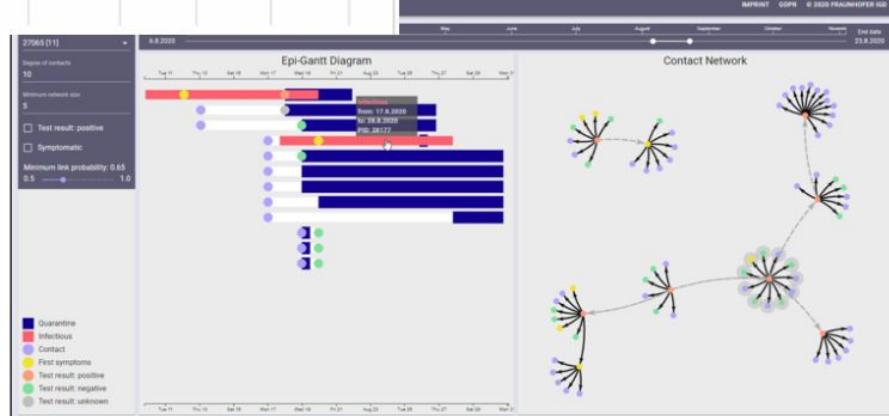
TIEVis
Temporal Information from Patient Records



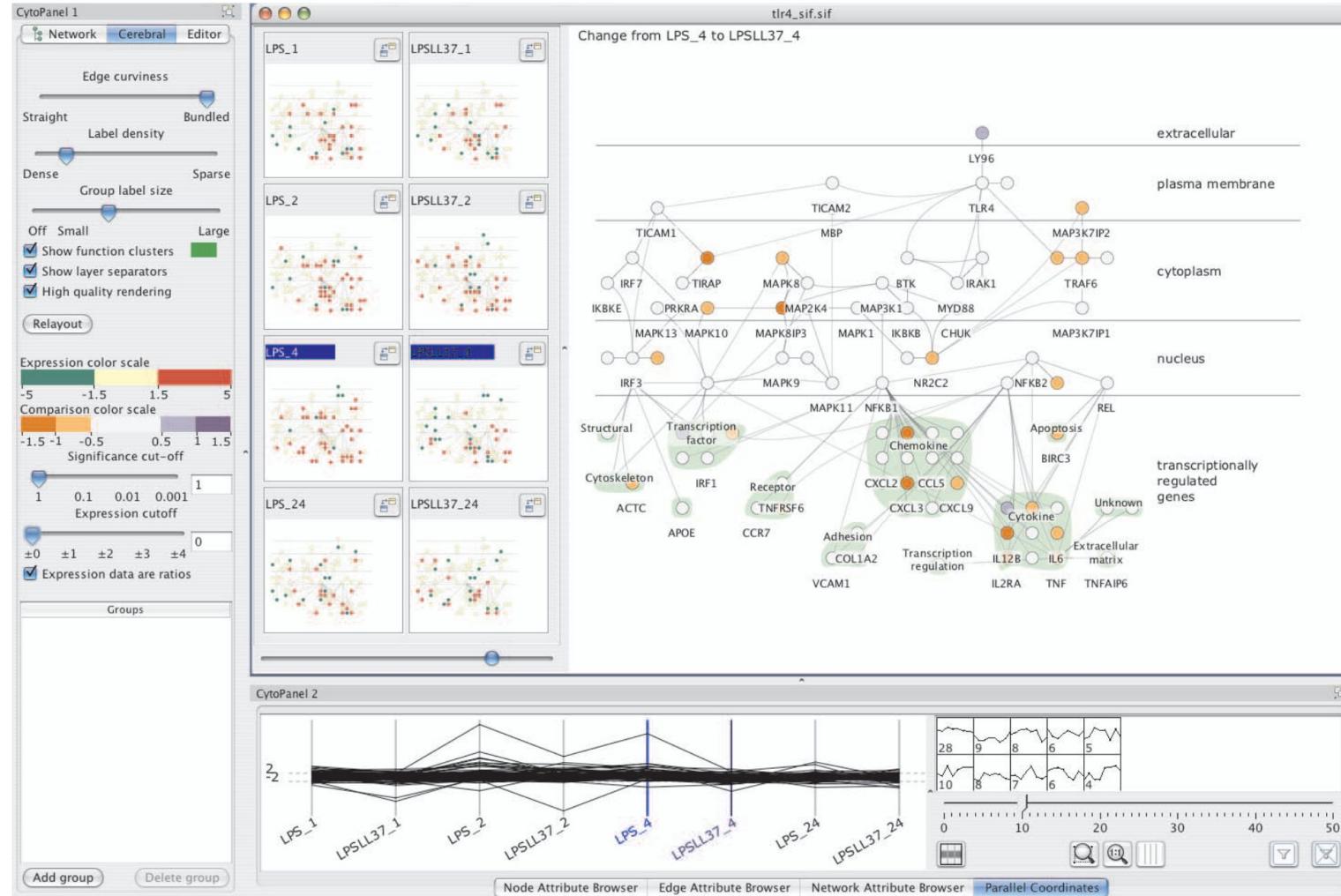
Outflow (Wongsuphasawat & Gotz)
variations in the order of symptom
onset and outcome for heart failure
patients

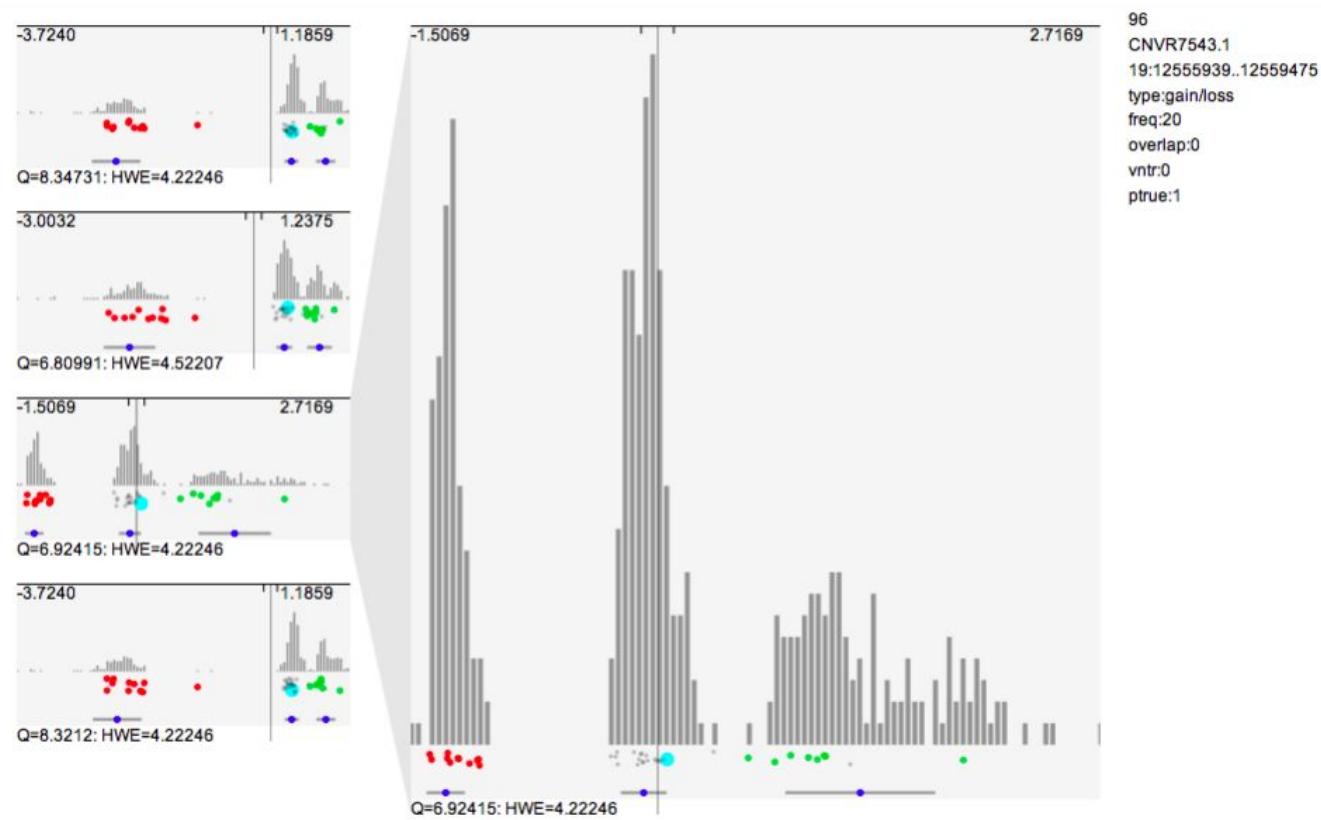


NHS data (Borland et al)

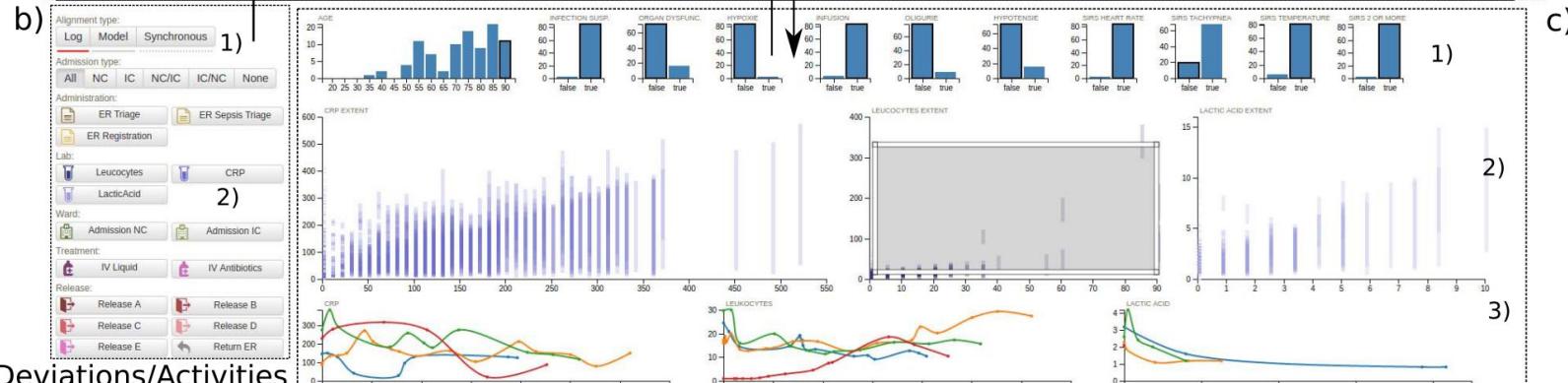
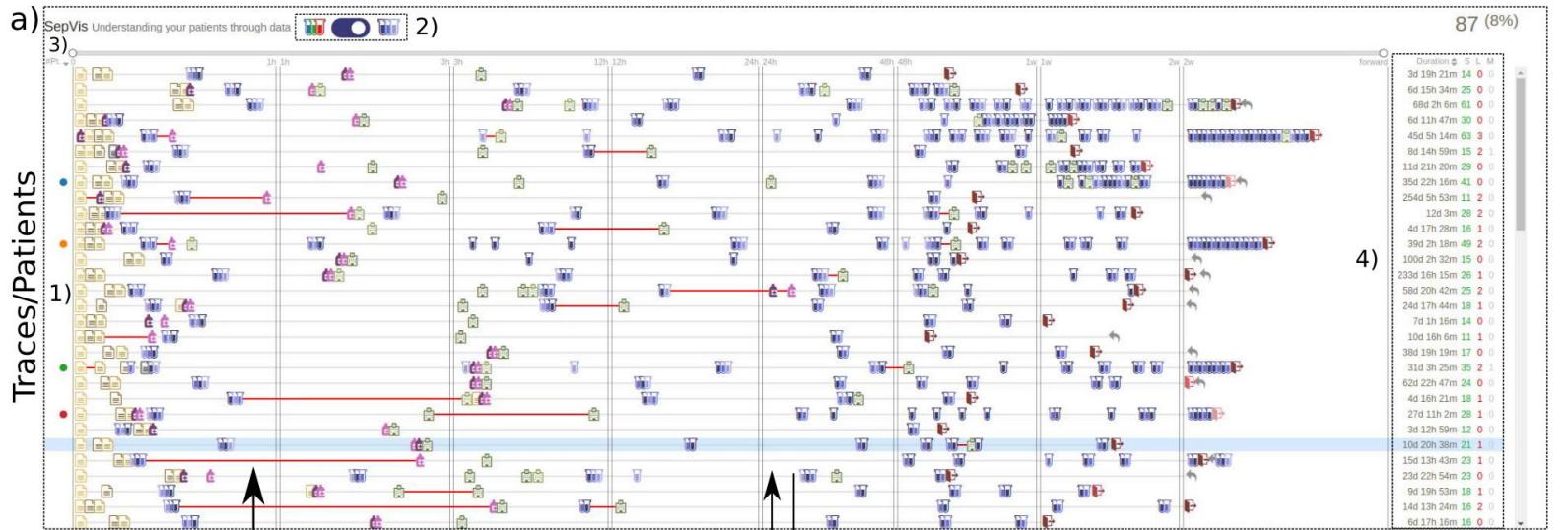


Detection & VA of COVID-19 clusters (Antweiler et al)





classification of copy number variations in 1000 Genomes project (Aerts, 2009)



Data distributions

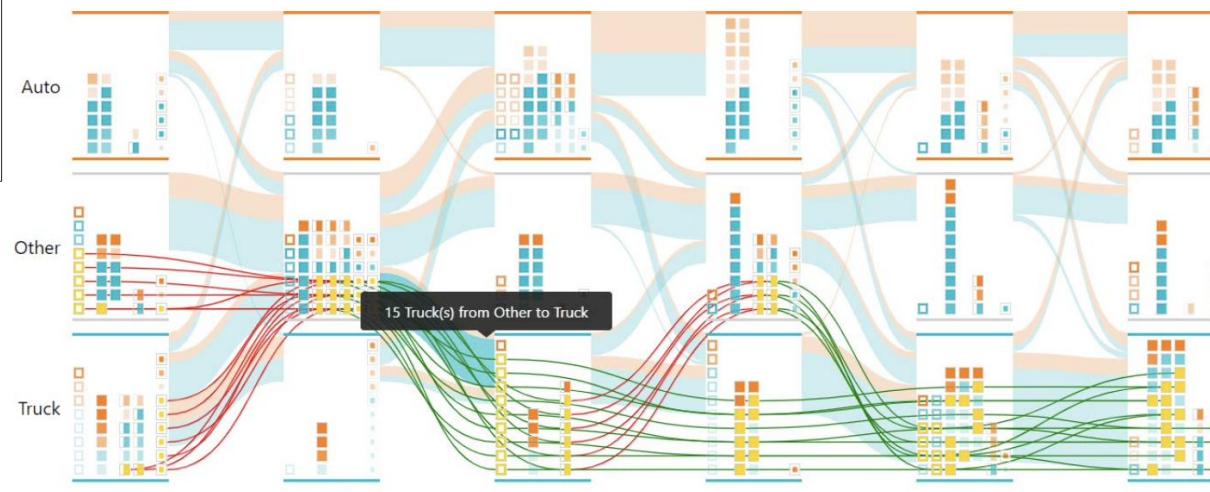
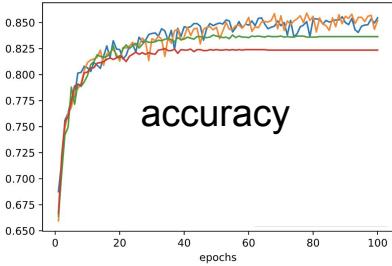
Own involvement in biological data visualisation community

- IEEE/ISMB Biovis symposium: OC, GC and SC (since 2012)
- European Bioinformatics Course on Biological Data Visualisation
- 3-week course Programming for Evolutionary Biology Leipzig/Berlin (Germany)
- CODATA/RDA Research Data Science Summer School, Trieste (Italy)
- ...



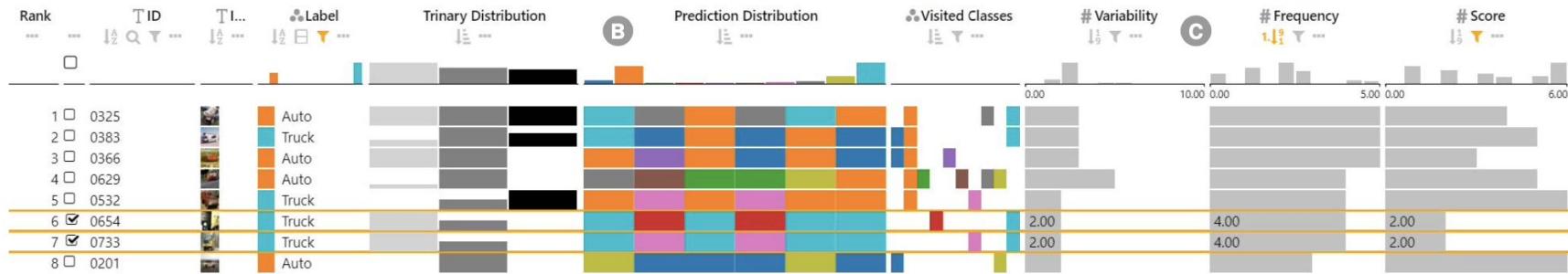
Explainable AI

- global vs local explanations
- intrinsic vs post-hoc explanations
- model-specific vs model-agnostic explanations

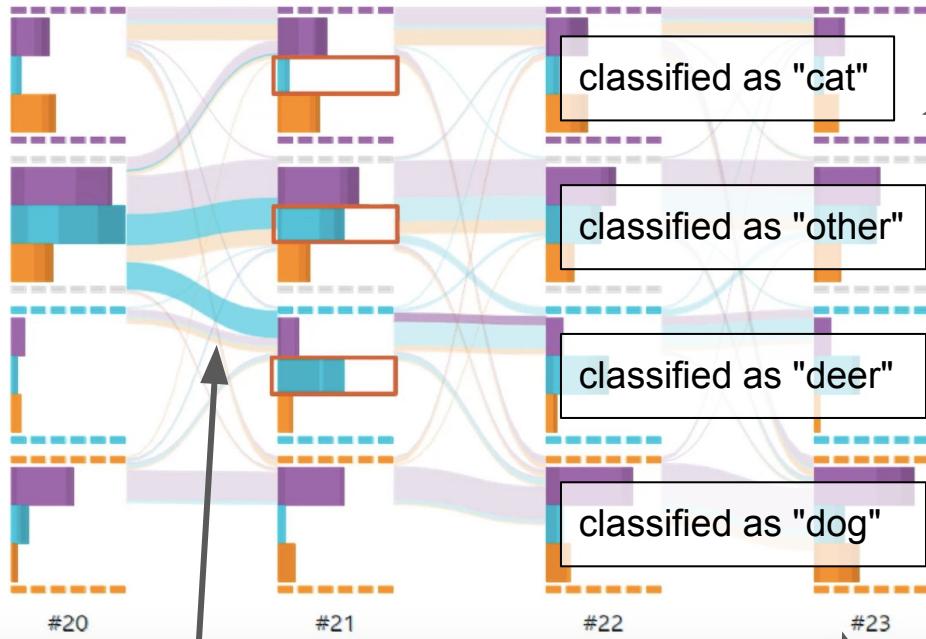


Plane
Auto
Bird
Cat
Deer
Dog
Frog
Horse
Ship
Truck
Other

Stable
Incoming
Outgoing
In & Out



InstanceFlow (Puehringer, Hinterreiter & Streit, 2020)

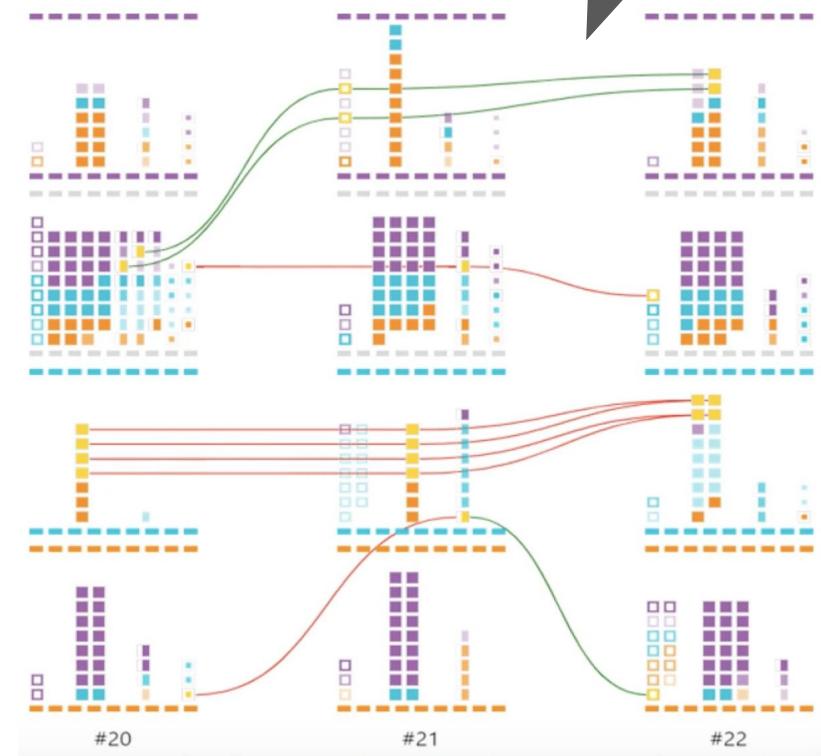


movement of instances
between classes

epoch

aggregated view

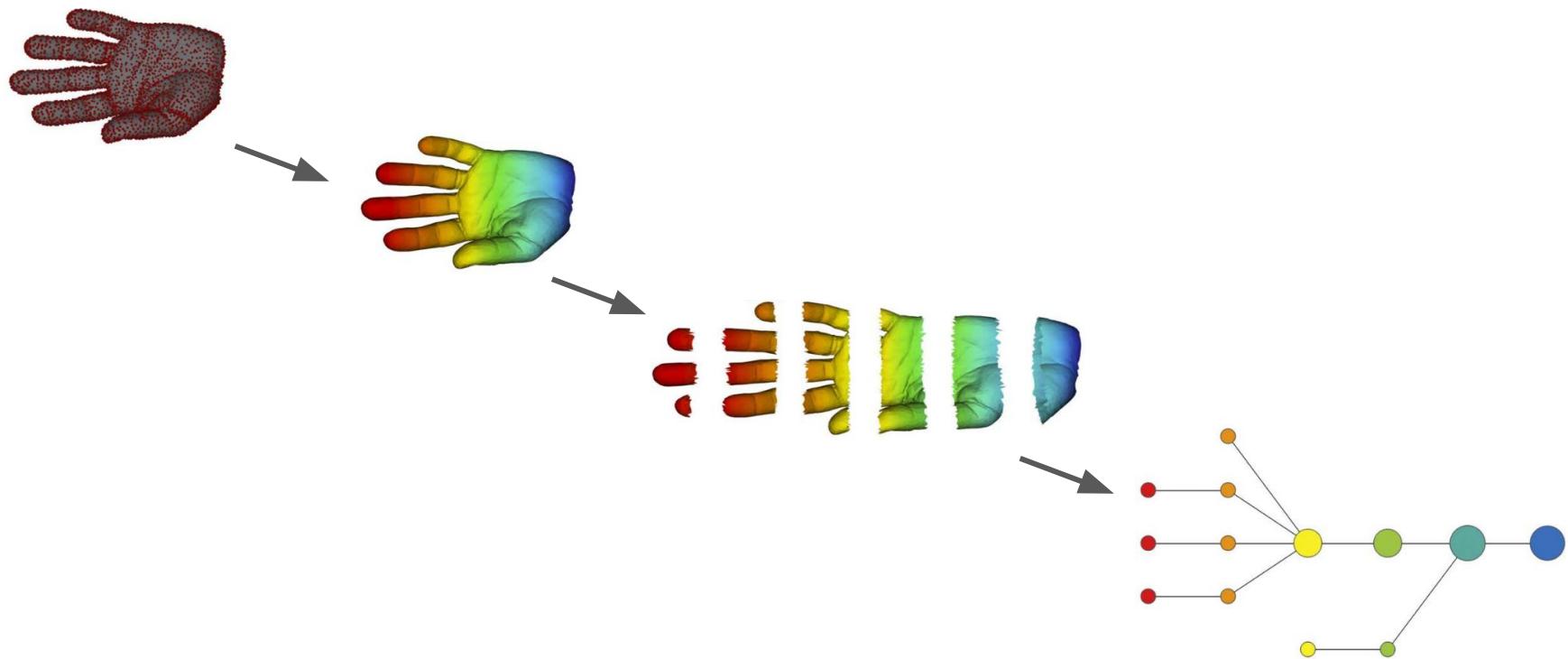
instance view



=> Supercharging interactive visualisation using machine learning =>

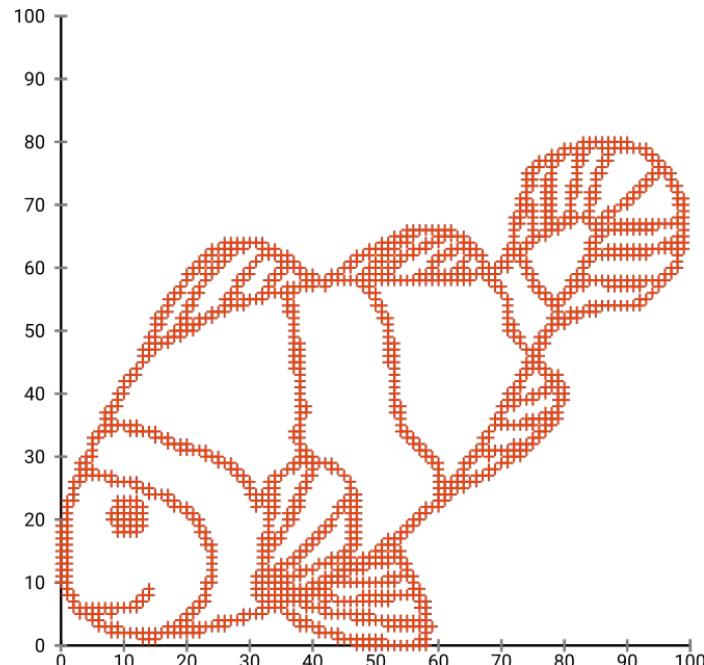
3. Topological Data Analysis (TDA)

Data has shape, shape has meaning, meaning has value

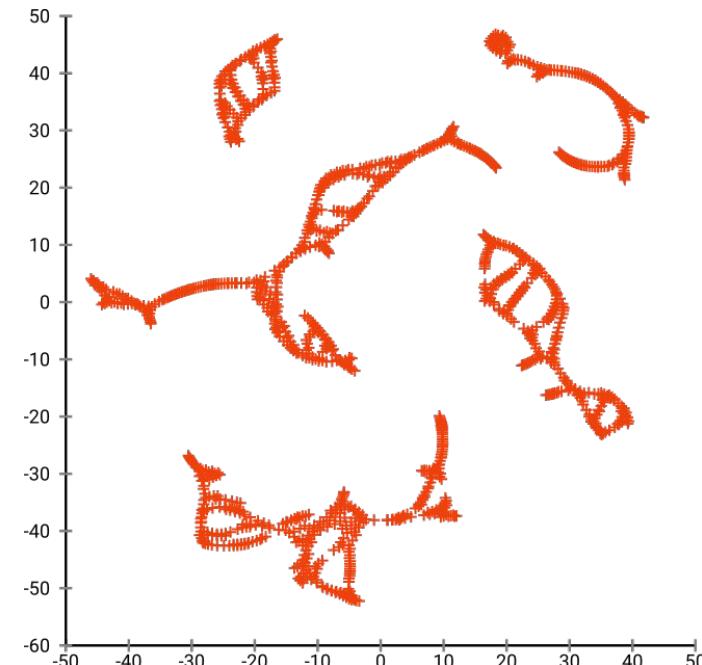


TDA =~ local clustering in a global context

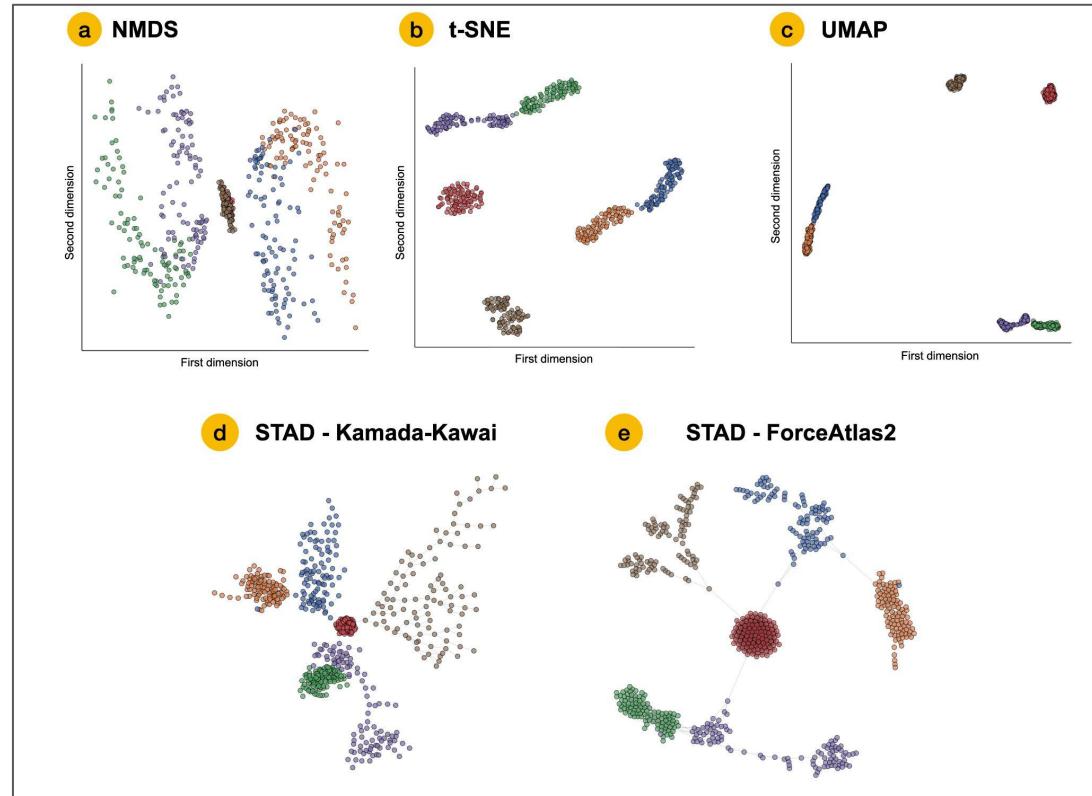
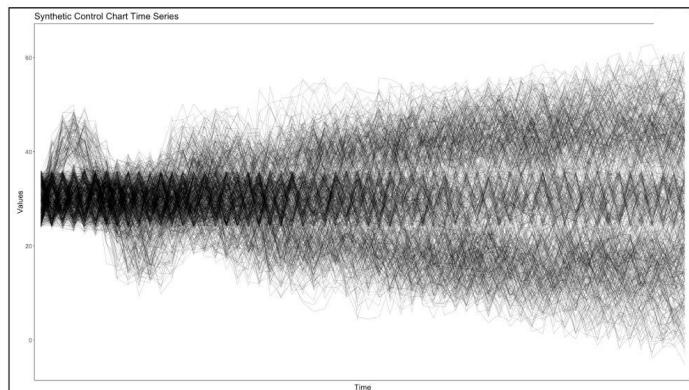
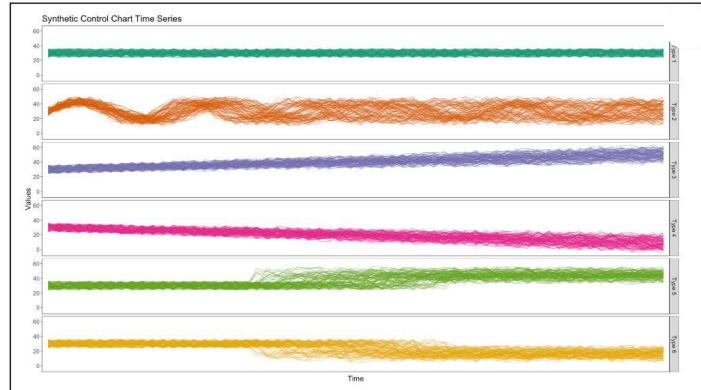
original data



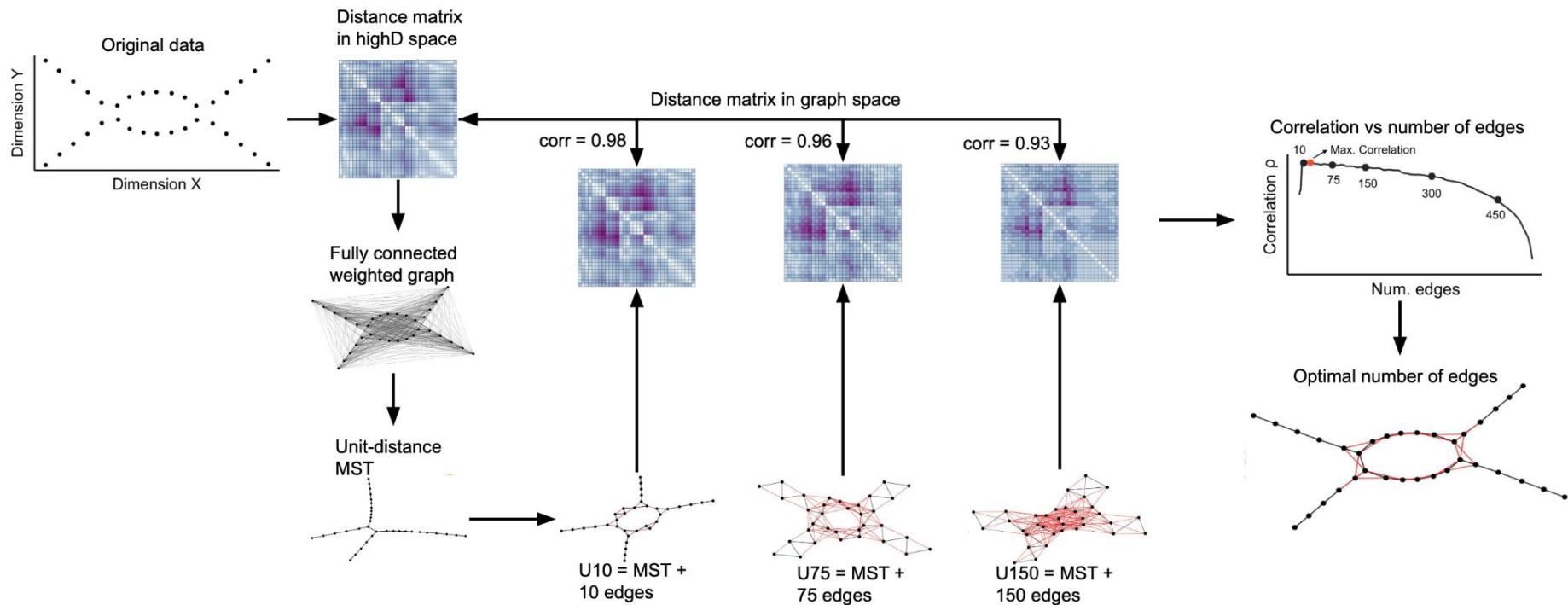
tSNE projection



TDA =~ local clustering in a global context



STAD - Simplified Topological Approximation of Data

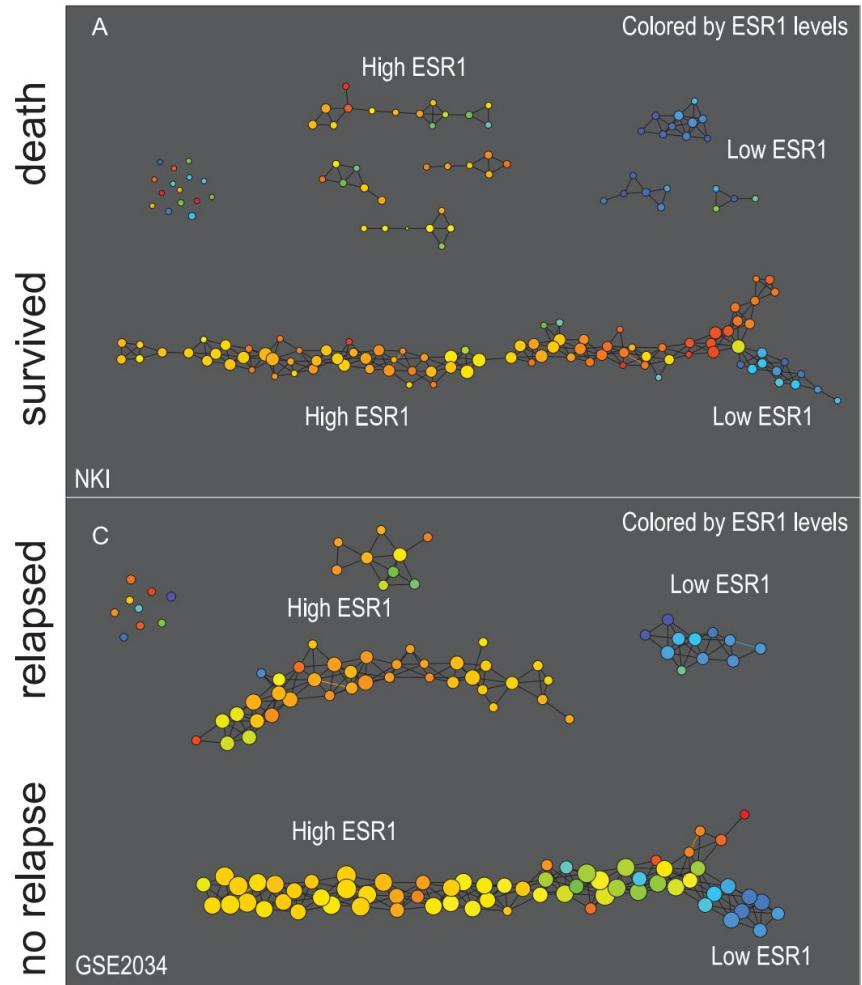


TDA to:

- understand data
- understand ML models

TDA to understand data

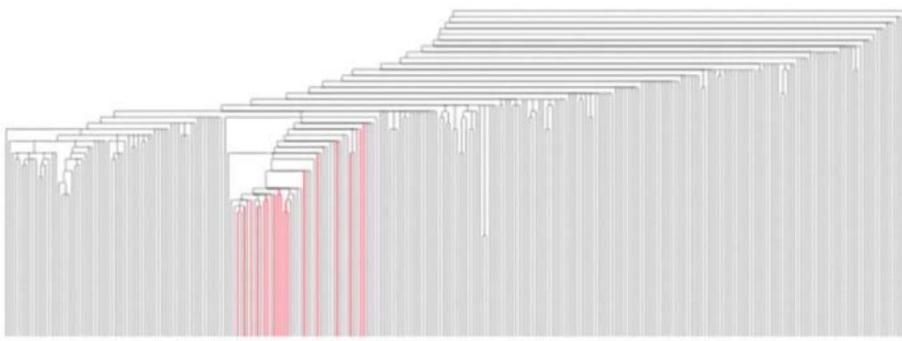
Example: breast cancer patient stratification



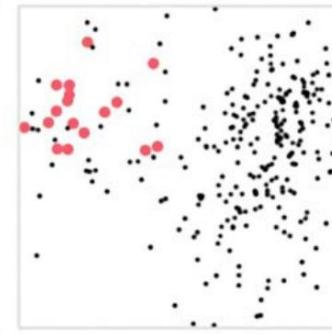
Lum et al, 2012; Ayasdi

ER- did not survive

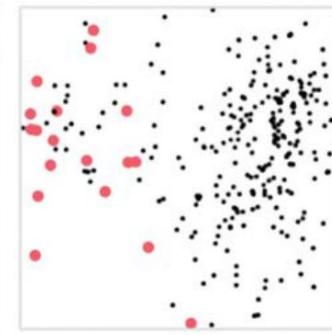
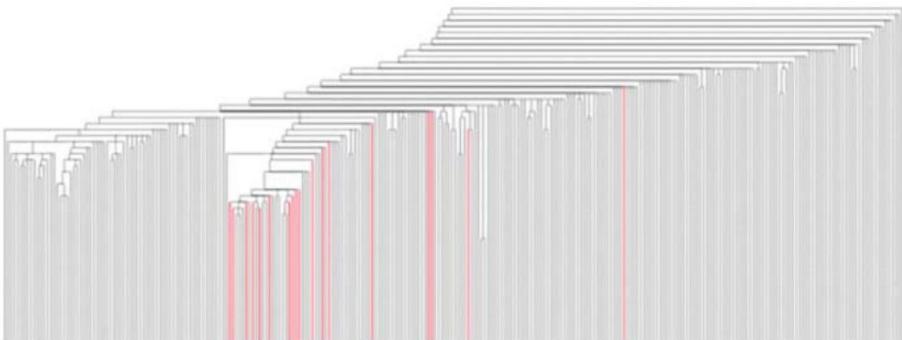
Clustering

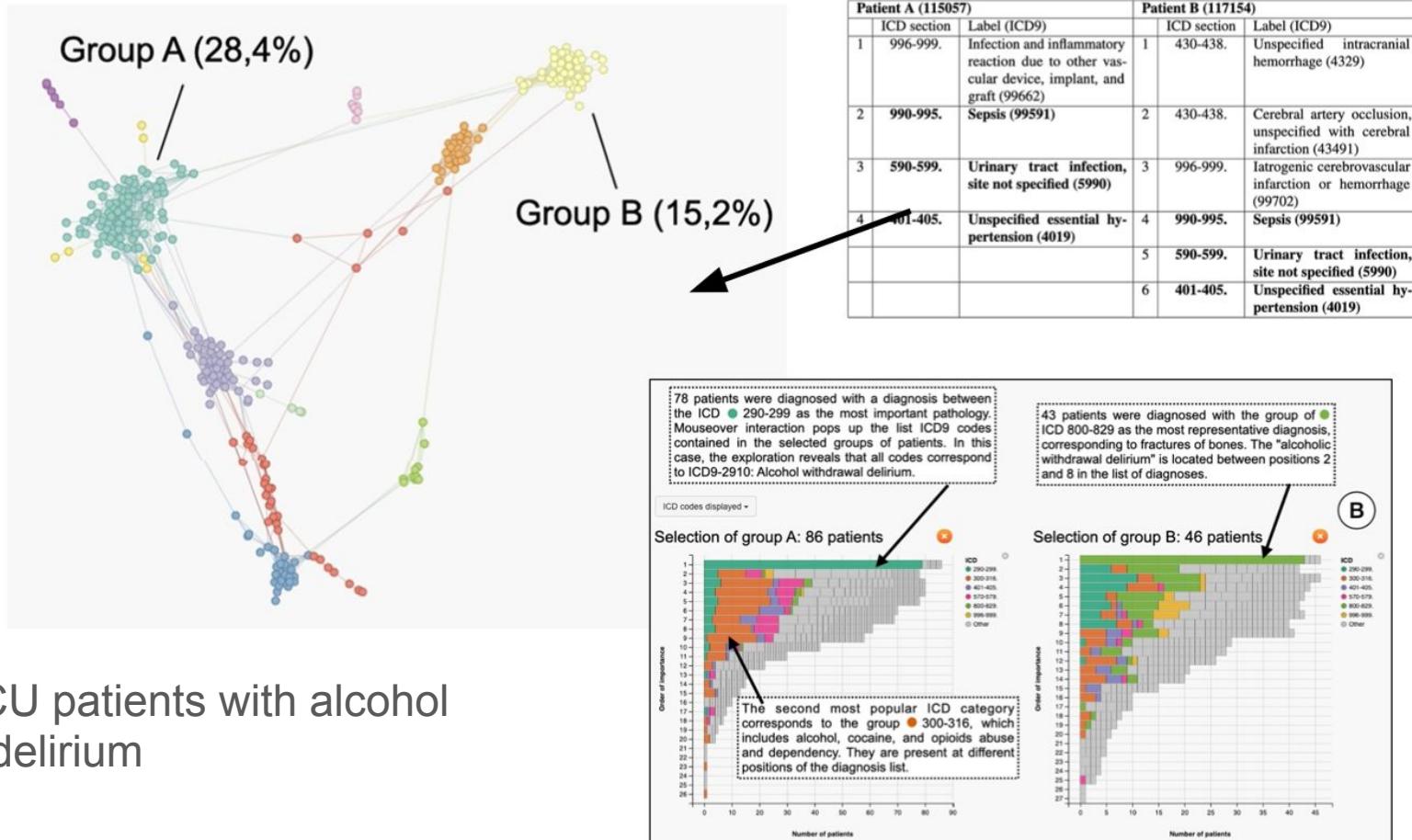


PCA



ER- survived

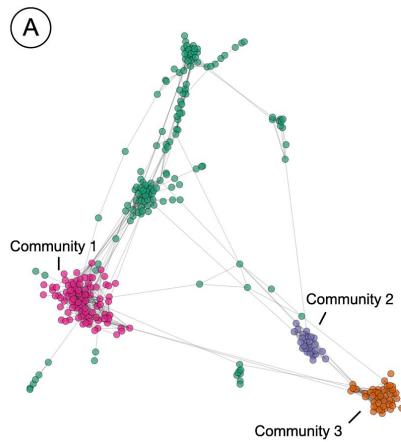




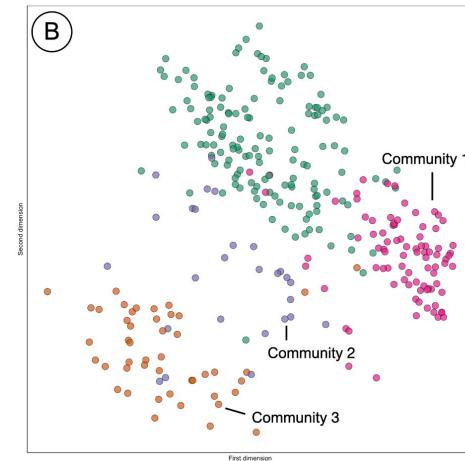
Example: ICU patients with alcohol withdrawal delirium

TDA vs tSNE/MDS/UMAP

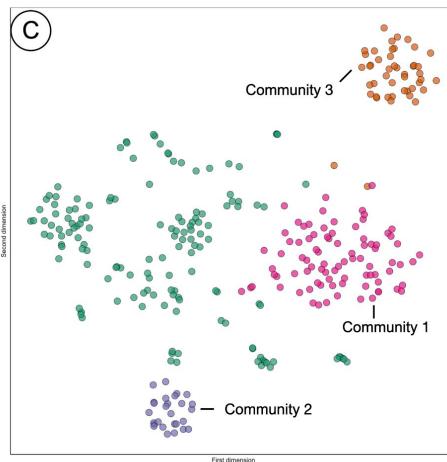
STAD-R



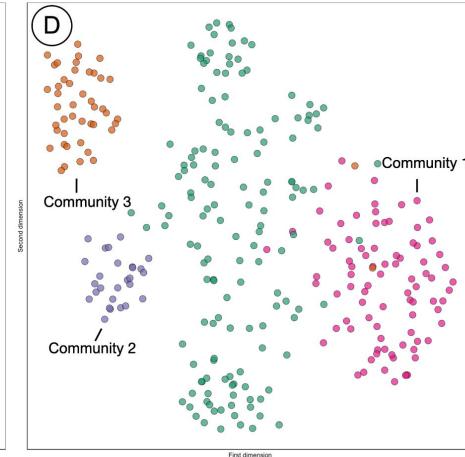
MDS



t-SNE

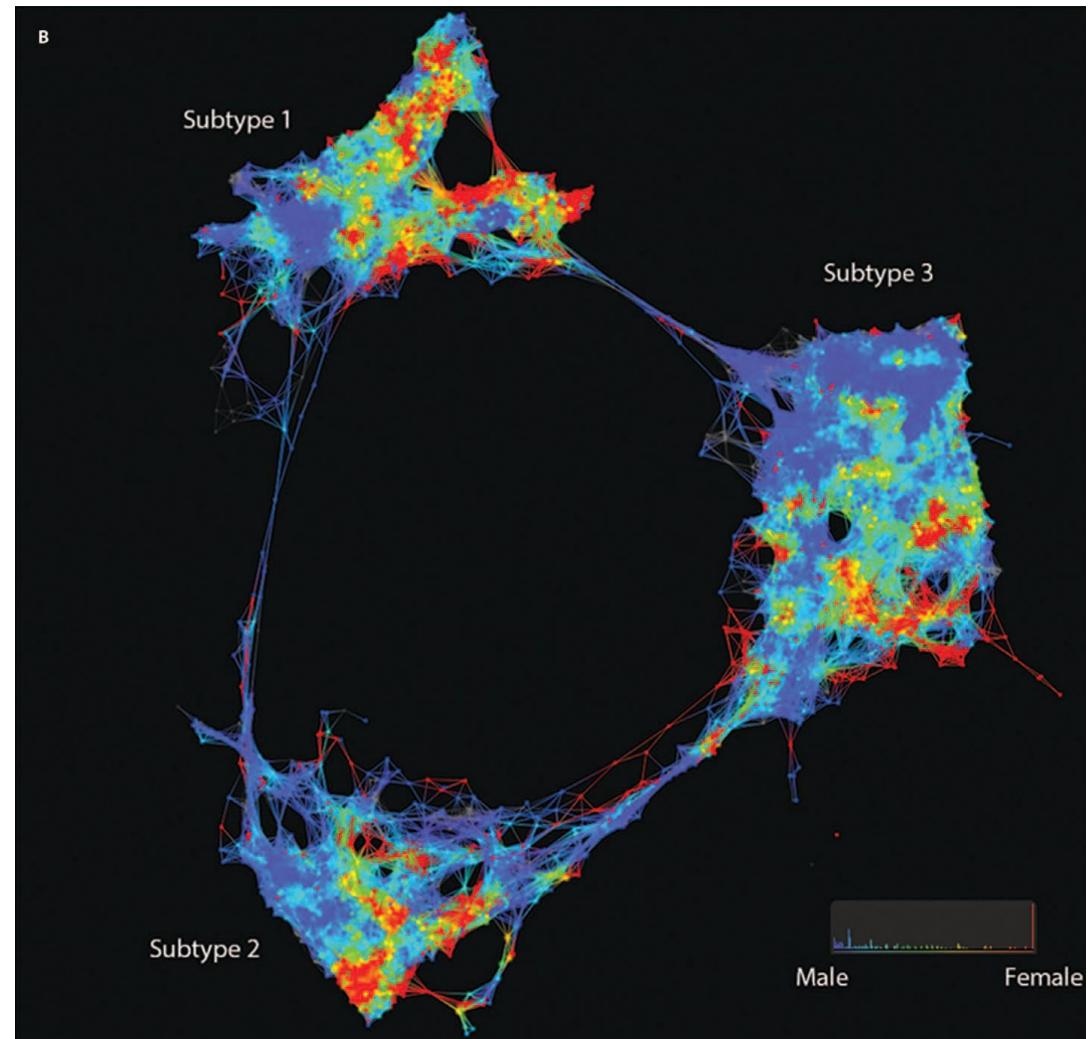


UMAP

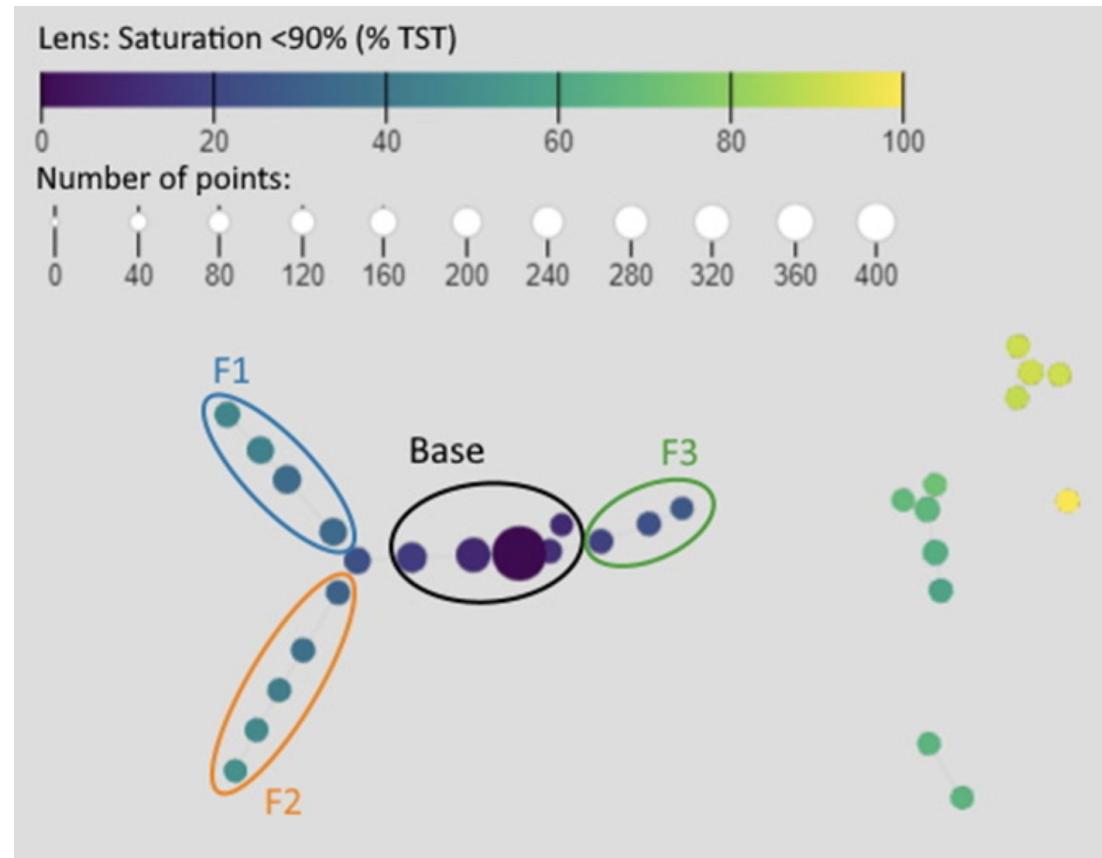


Community 1 Community 2 Community 3

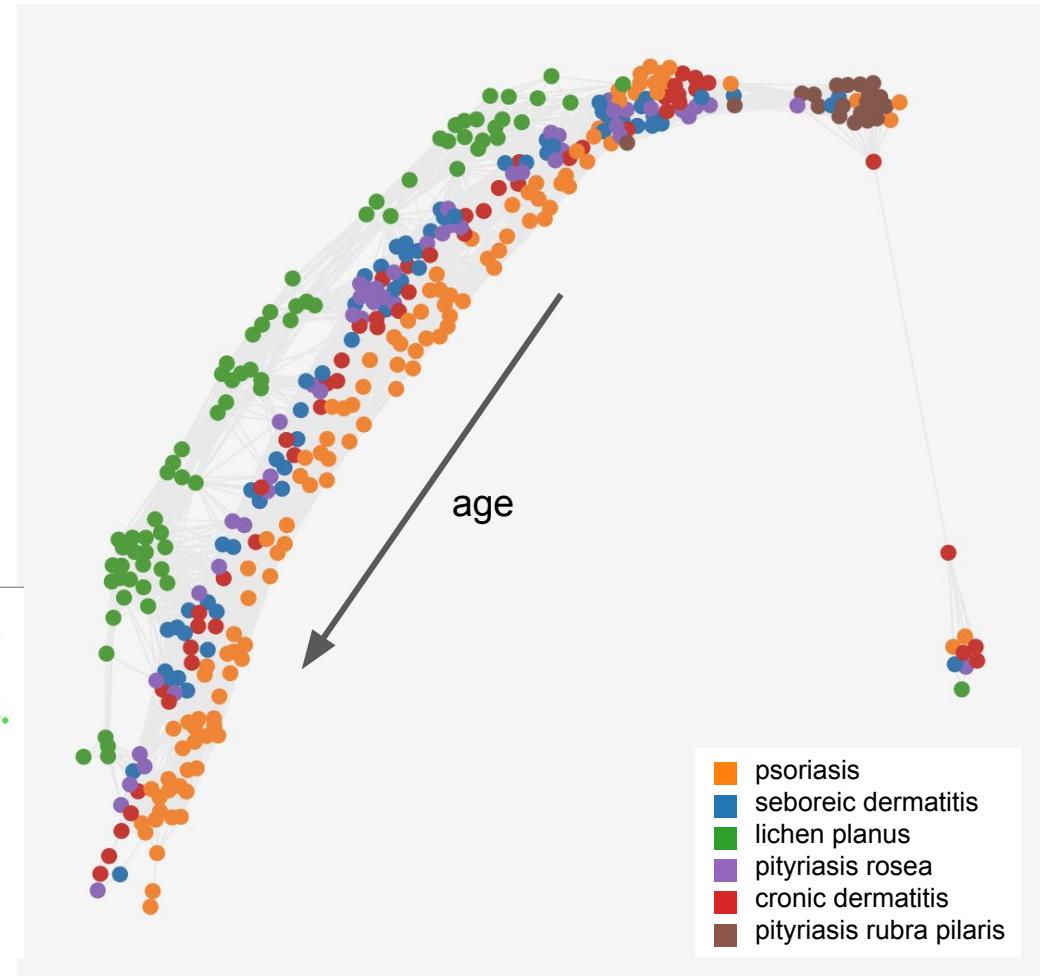
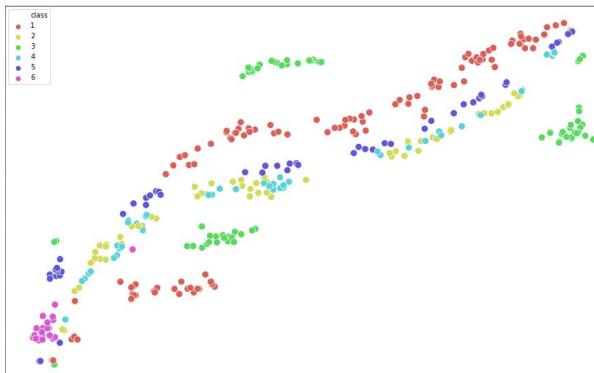
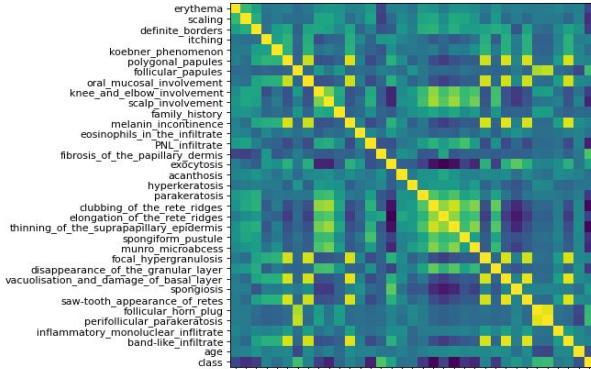
Example: towards precision medicine in Type2 Diabetes



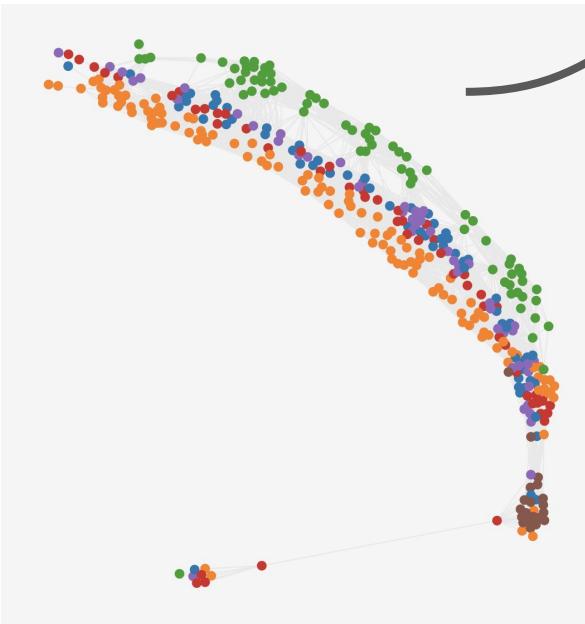
...applied to sleep apnea patients



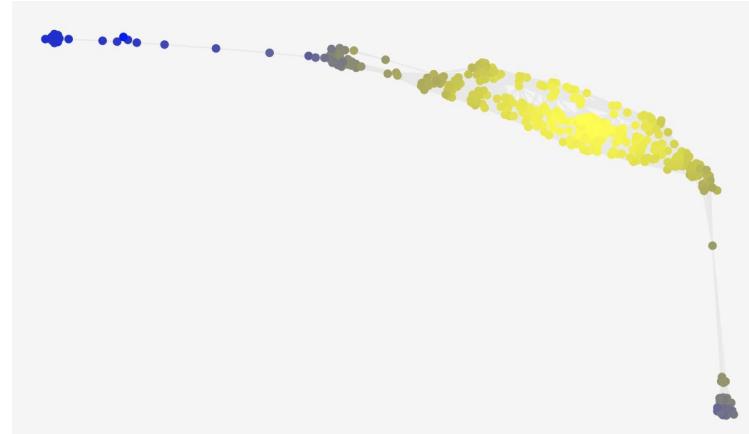
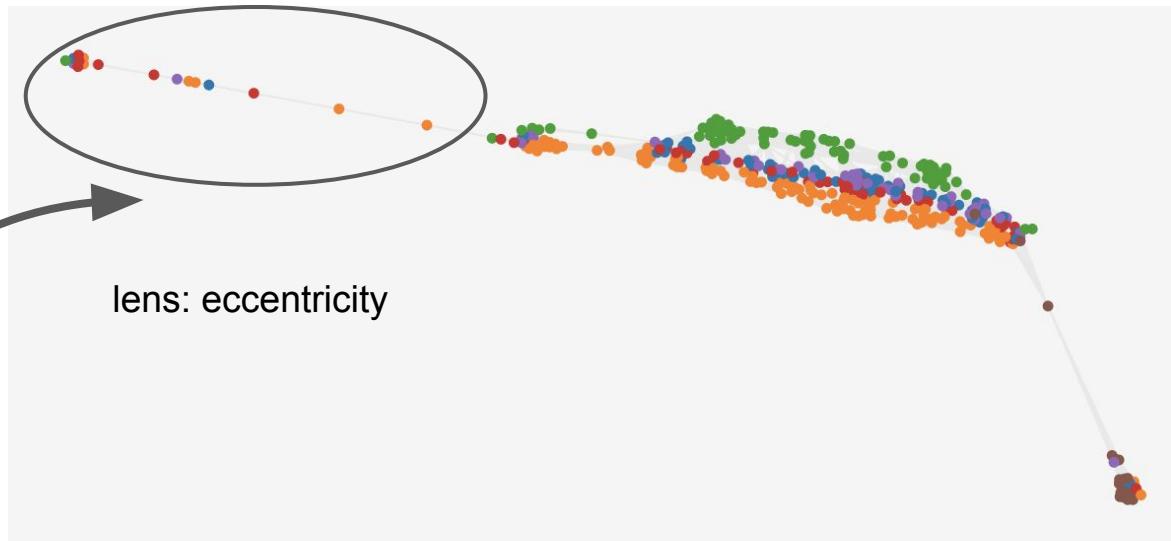
...applied to dermatology classification



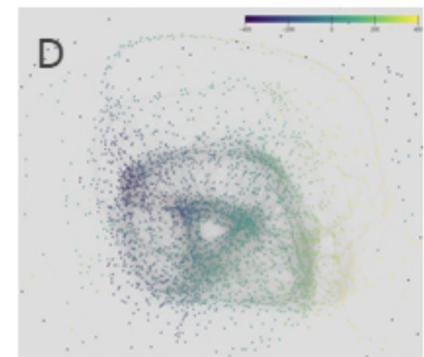
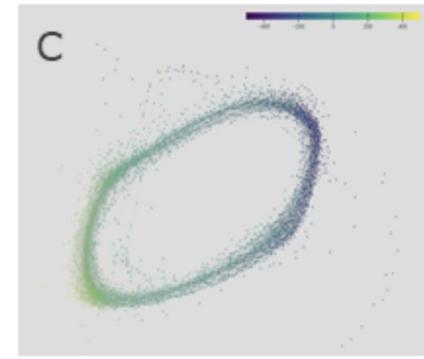
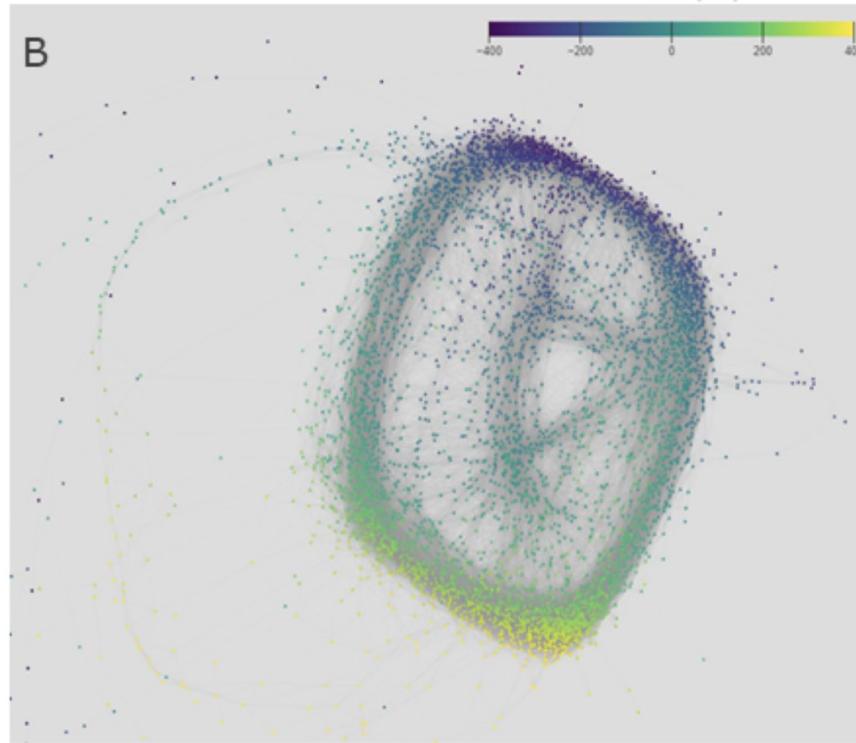
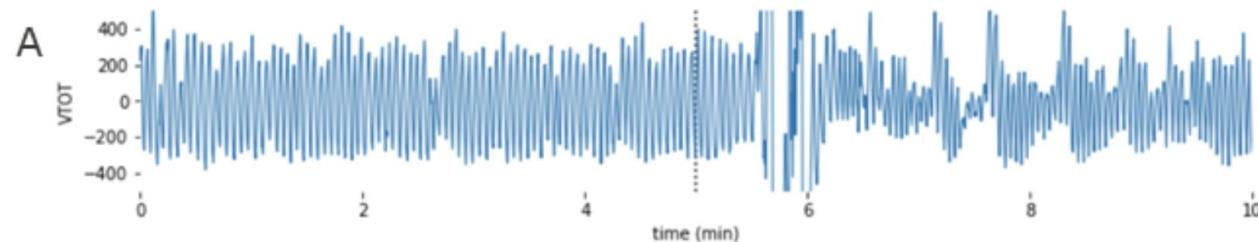
no lens



lens: eccentricity



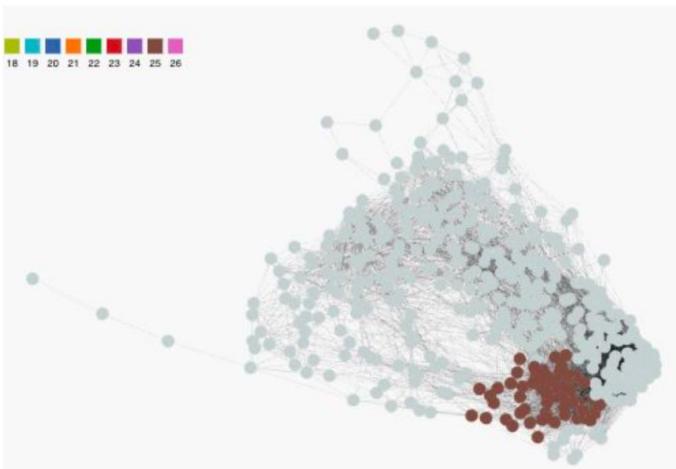
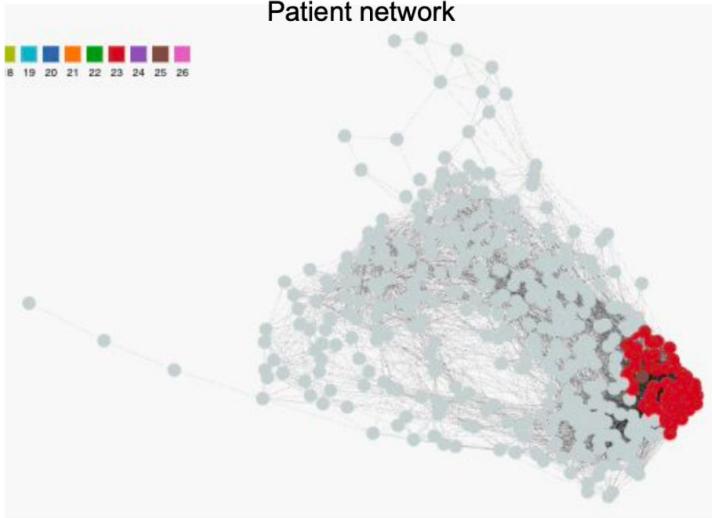
...applied to chest
volume in sleep
apnea patients



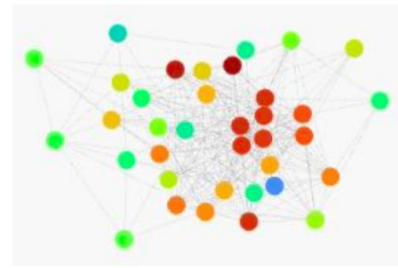
TDA to understand/tweak models

Feature modeling: TDA on transposed matrix

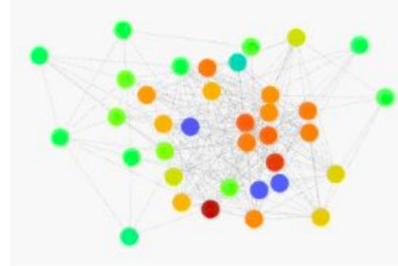
Use case: sleep apnea patients => select features that best discriminate data



Feature network



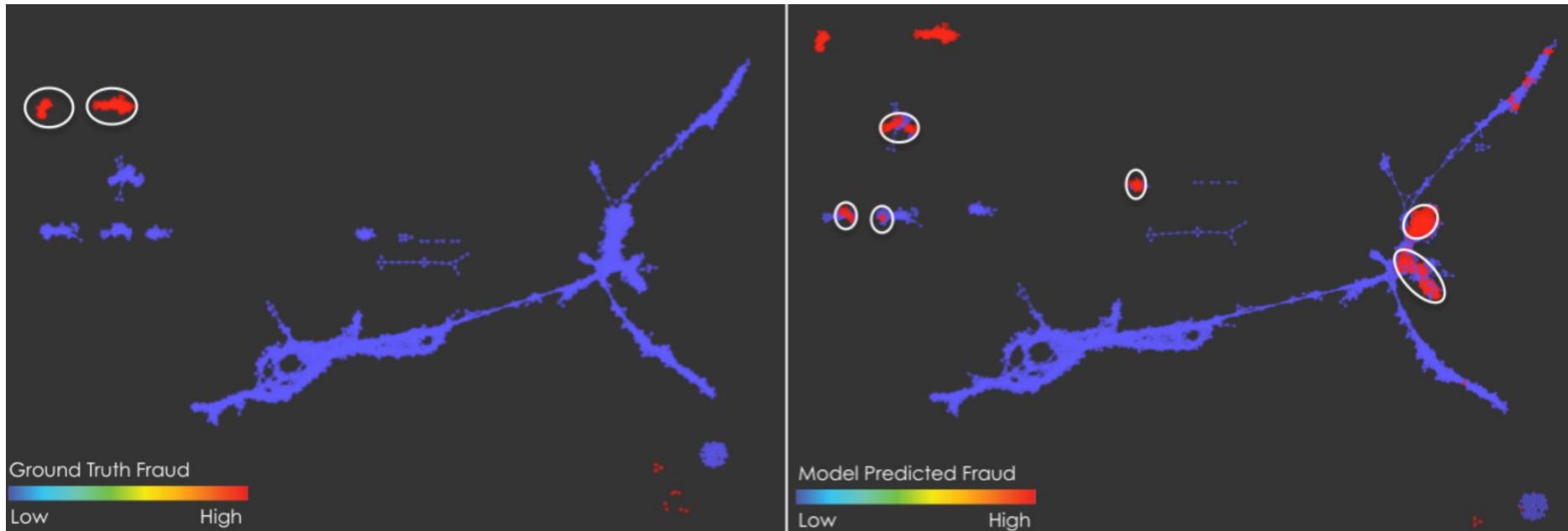
- high CSA index
- high MSA index



- low average saturation during sleep
- low % TIB with saturation 90-100%

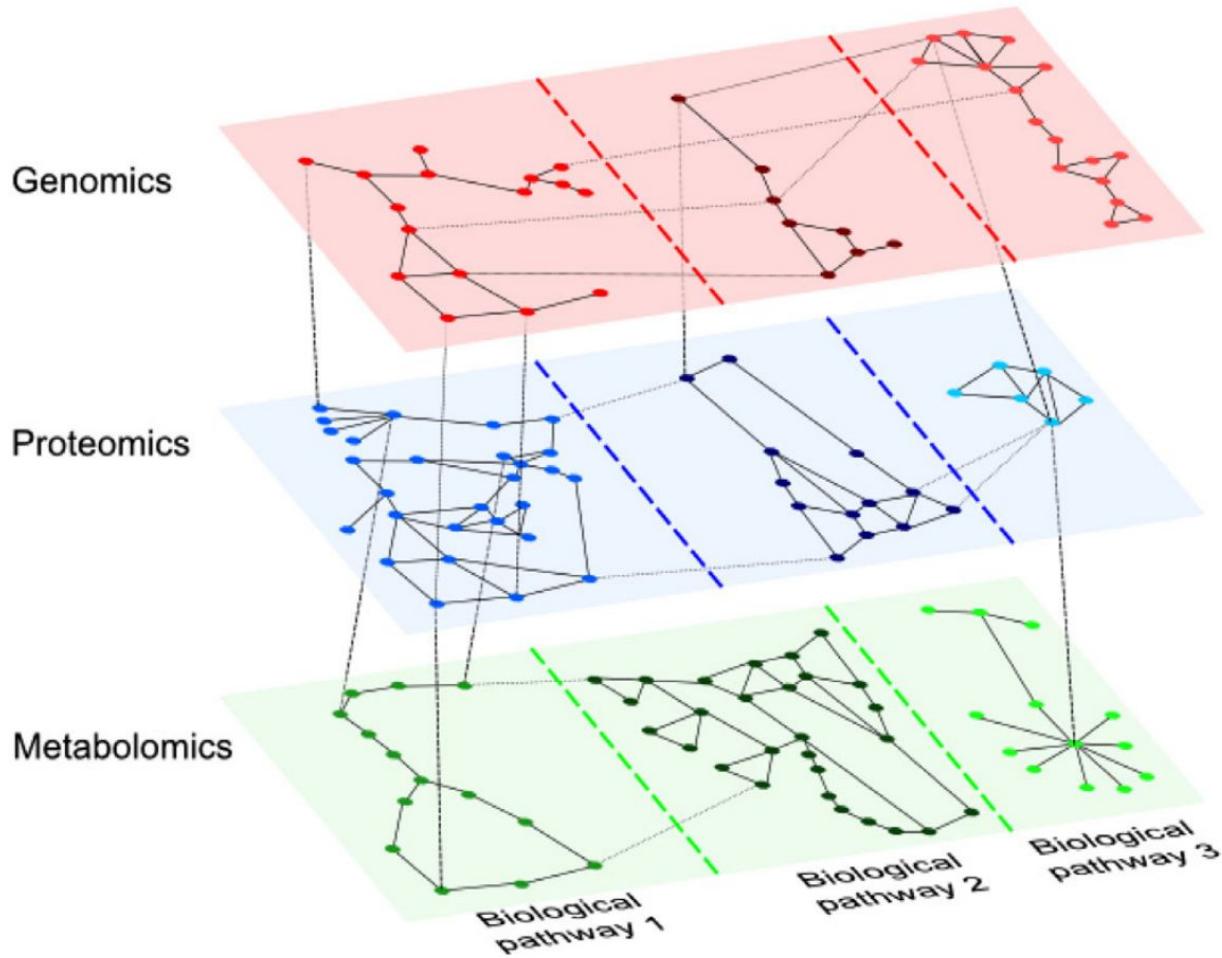
Validate & improve models

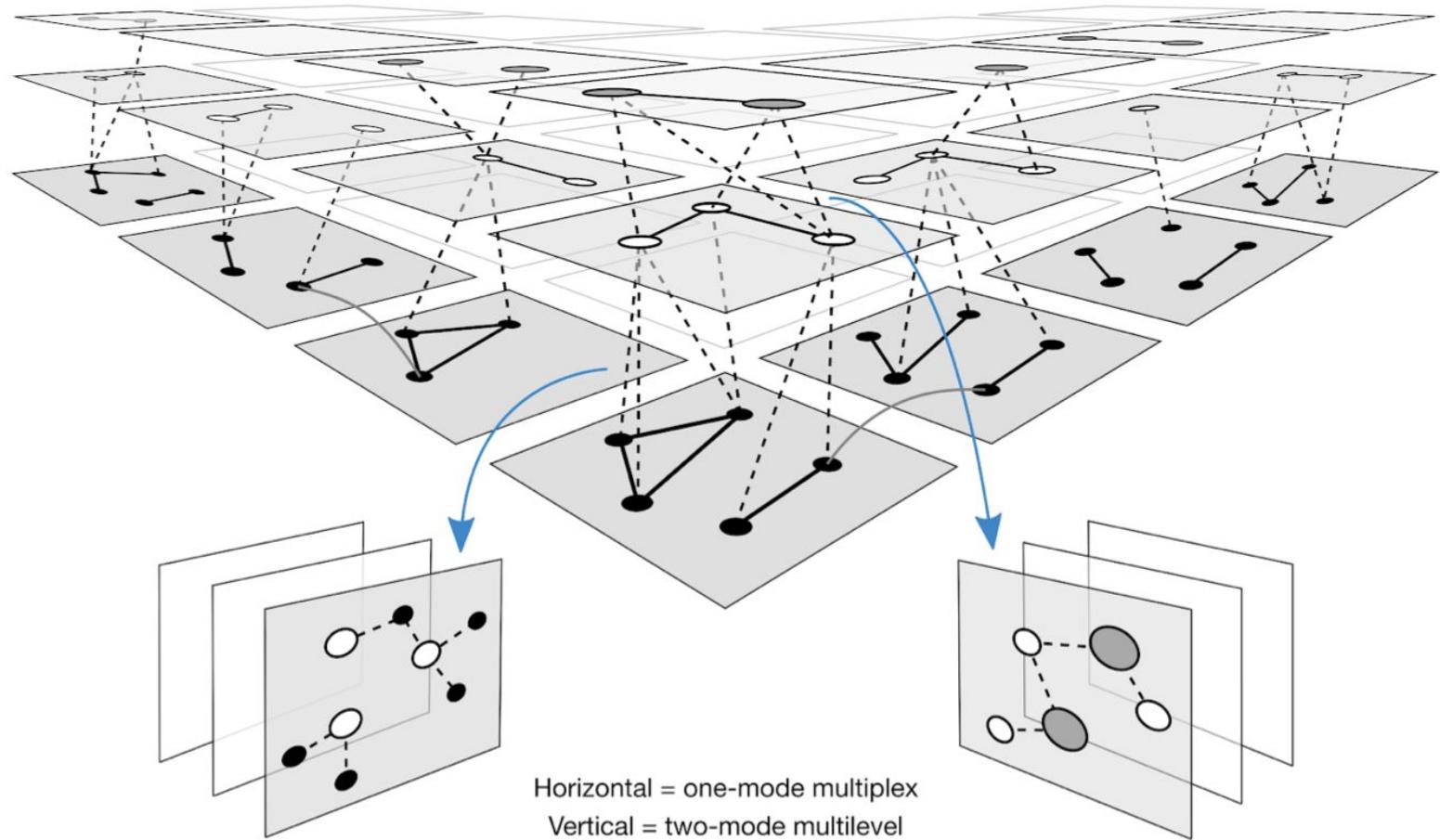
...by comparing prediction with ground truth



=> Looking at data from different perspectives =>

4. Multi-layer Networks (MLN)





WP1

Original data

	param1	param2	param3	param4	...
Individual 1	8	15	1.2	184	...
Individual 2	7	14	1.4	160	...
Individual 3	12	19	3.9	156	...
Individual 4	5	96	12.6	57	...
...



Distance matrix

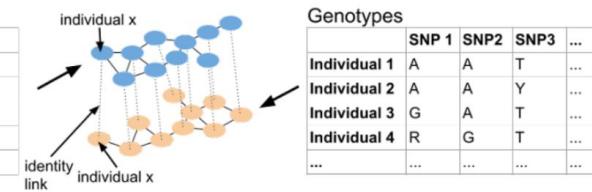
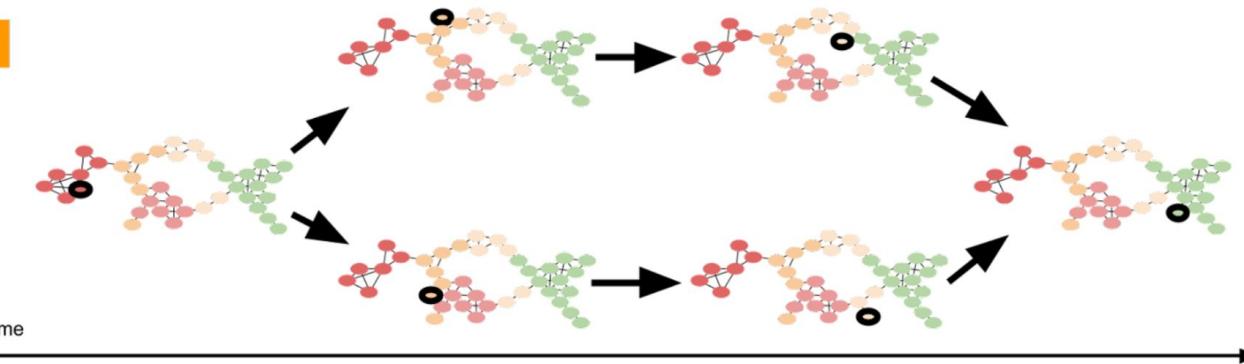
	Individual 1	Individual 2	Individual 3	Individual 4	...
Individual 1					...
Individual 2					...
Individual 3					...
Individual 4					...
...					...

TDA network

**WP2**

Blood values

	WBC	RBC	lymphs	...
Individual 1	2.1	3.0	42	...
Individual 2	4.0	4.7	44	...
Individual 3	0.8	4.5	87	...
Individual 4	1.1	96	72	...
...

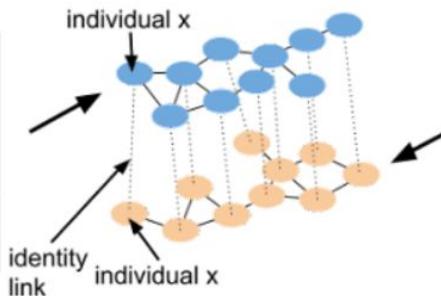
**WP3**

WP1



Blood values

	WBC	RBC	lymphs	...
Individual 1	2.1	3.0	42	...
Individual 2	4.0	4.7	44	...
Individual 3	0.8	4.5	87	...
Individual 4	1.1	96	72	...
...



Genotypes

	SNP 1	SNP2	SNP3	...
Individual 1	A	A	T	...
Individual 2	A	A	Y	...
Individual 3	G	A	T	...
Individual 4	R	G	T	...
...



Questions / Discussion