

From Complexity to Comprehensibility

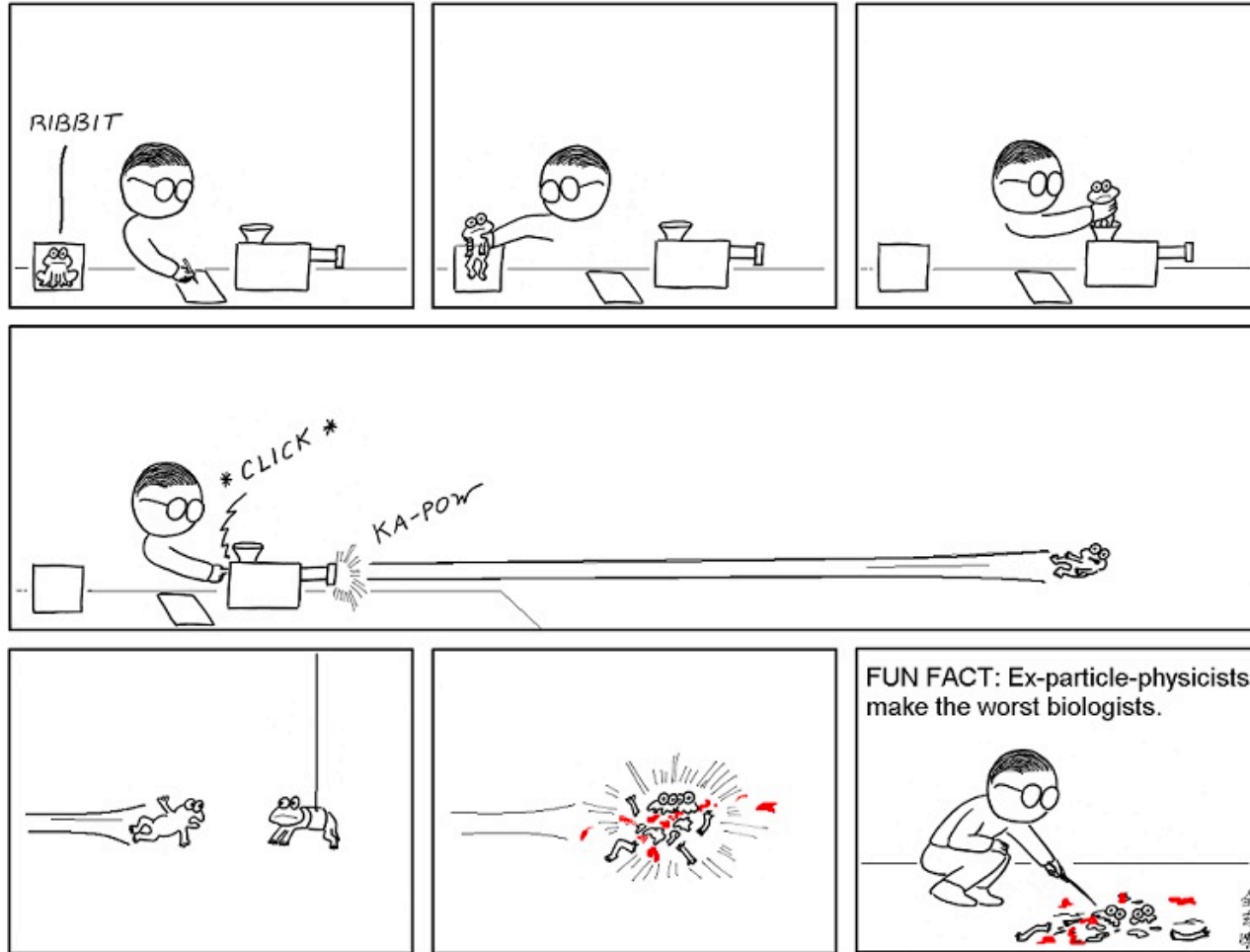
An Integrative View on Biological and Agricultural Systems

Prof Jan Aerts

Inaugural lecture - 23/11/2023



(c) Jan Aerts, 2023



<https://abstrusegoose.com/156>

Reductionist



Holistic

Wicked ("real") problems

How do we prevent 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%?

System

Question

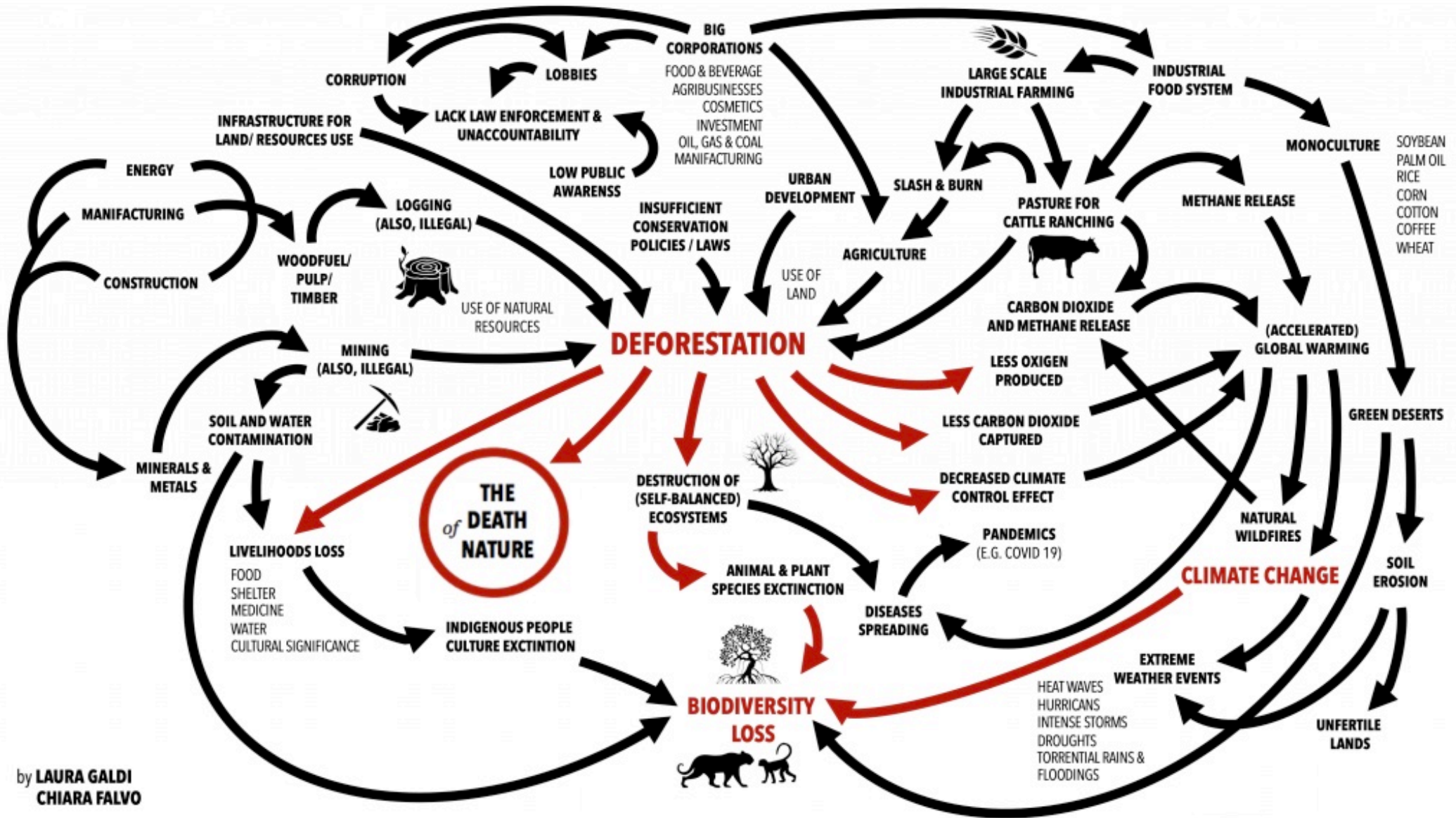
Open



Component

Hypothesis

Closed



by **LAURA GALDI**
CHIARA FALVO

What if we don't embrace complexity and uncertainty?

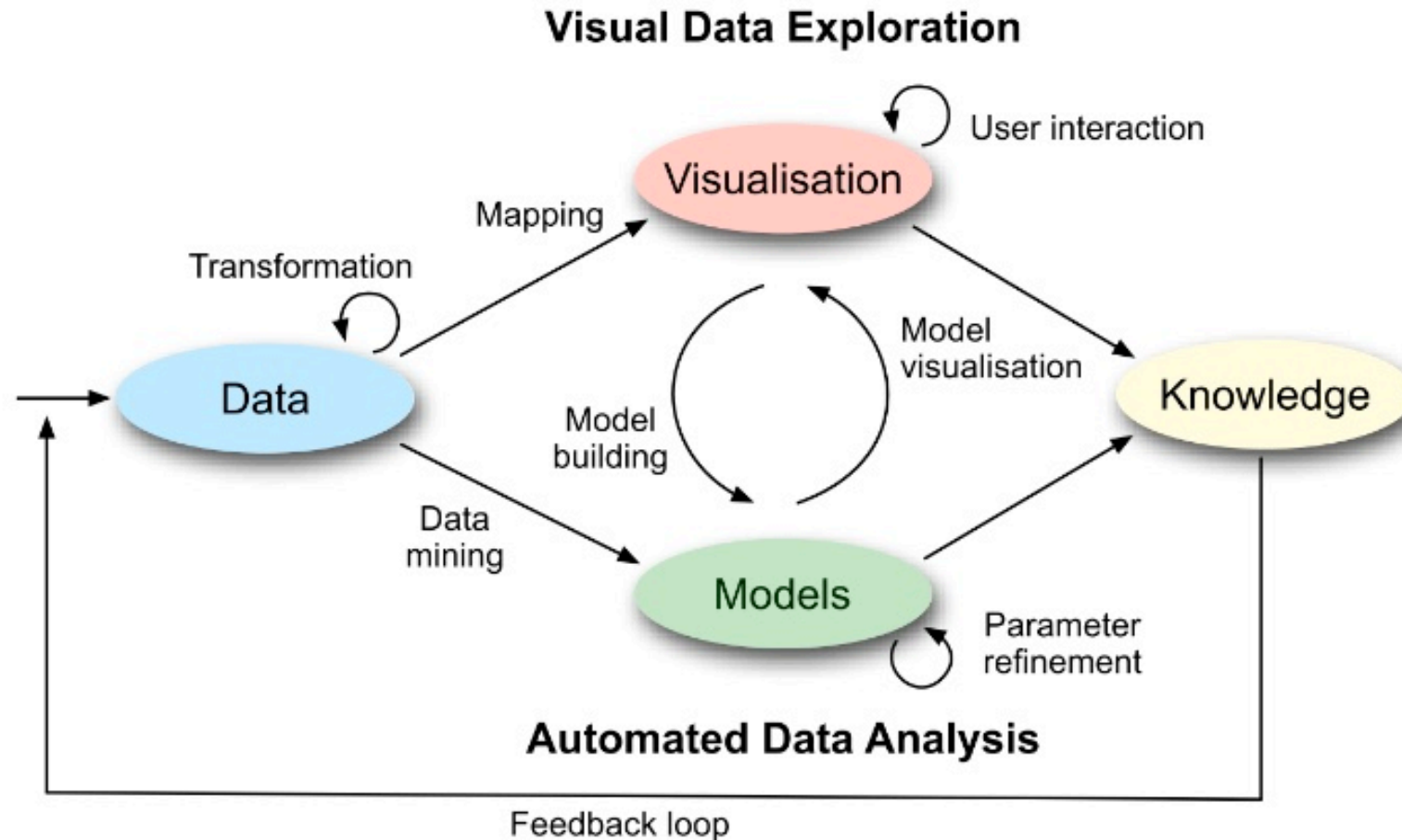
- misrepresentation of systems
- models with poor predictive power
- suboptimal interventions and decisions from limited insight
- reduced resilience to disturbances
- ...

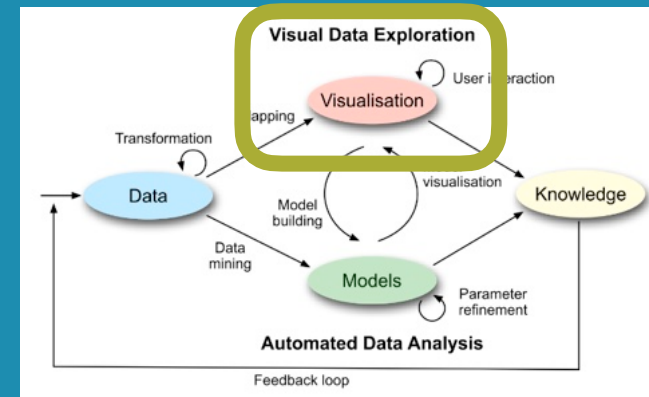
What do we need to do this?

- continual curiosity
- input from different stakeholders/experts and tacit knowledge
- collective sense-making
- contextualisation and externalisation of assumptions

=> combine computational strengths of automated analysis with human analytical skills + create shared mental model

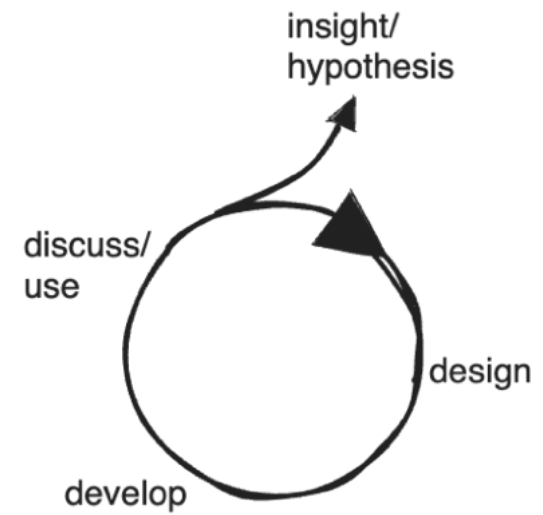
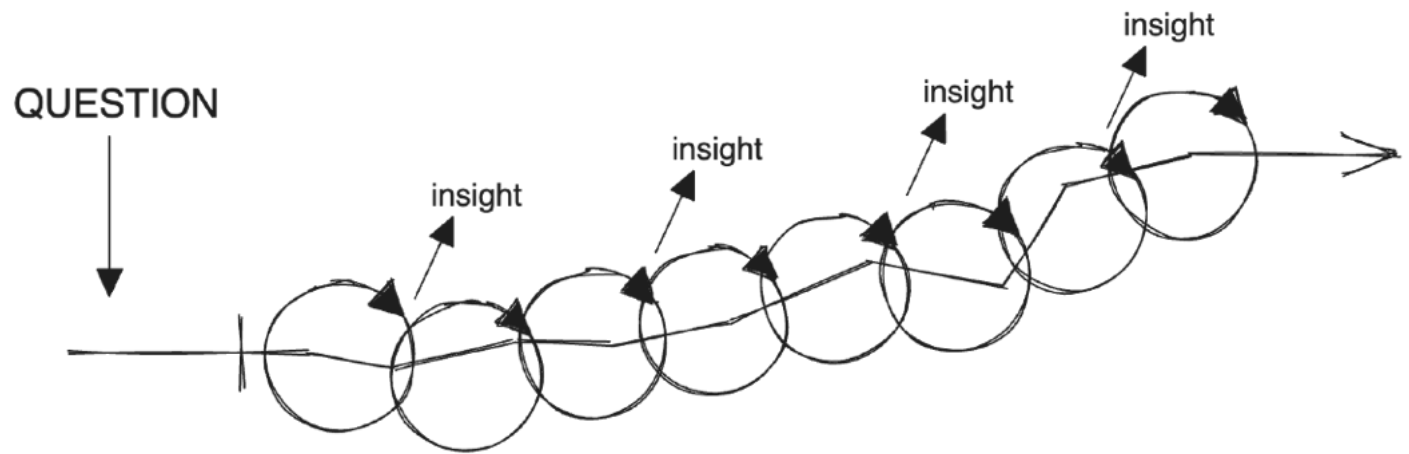
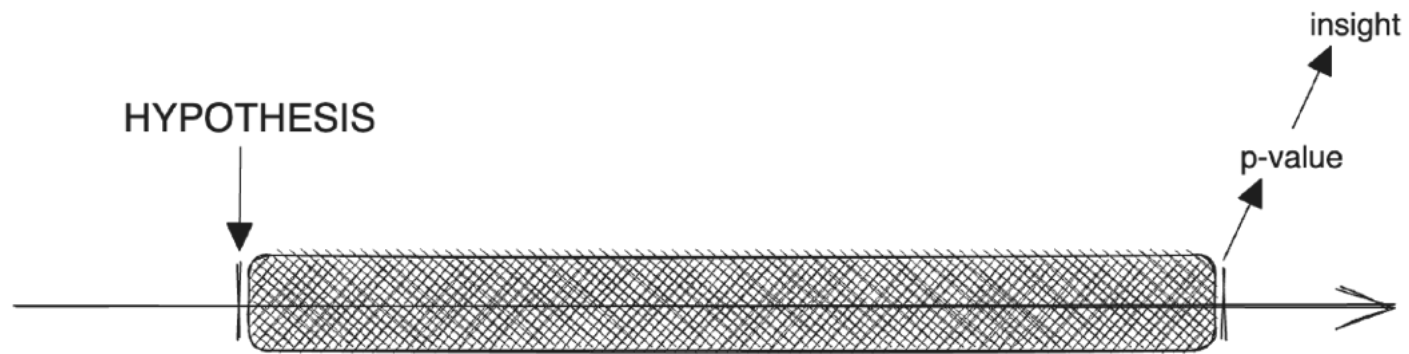
Visual Analytics





:: A :: (Interactive) data visualisation

A tool for thought





Discuss/use

what they want != what they need (underlying assumptions)

game storming methodology and co-creation

insights => new questions

Game storming



card sorting

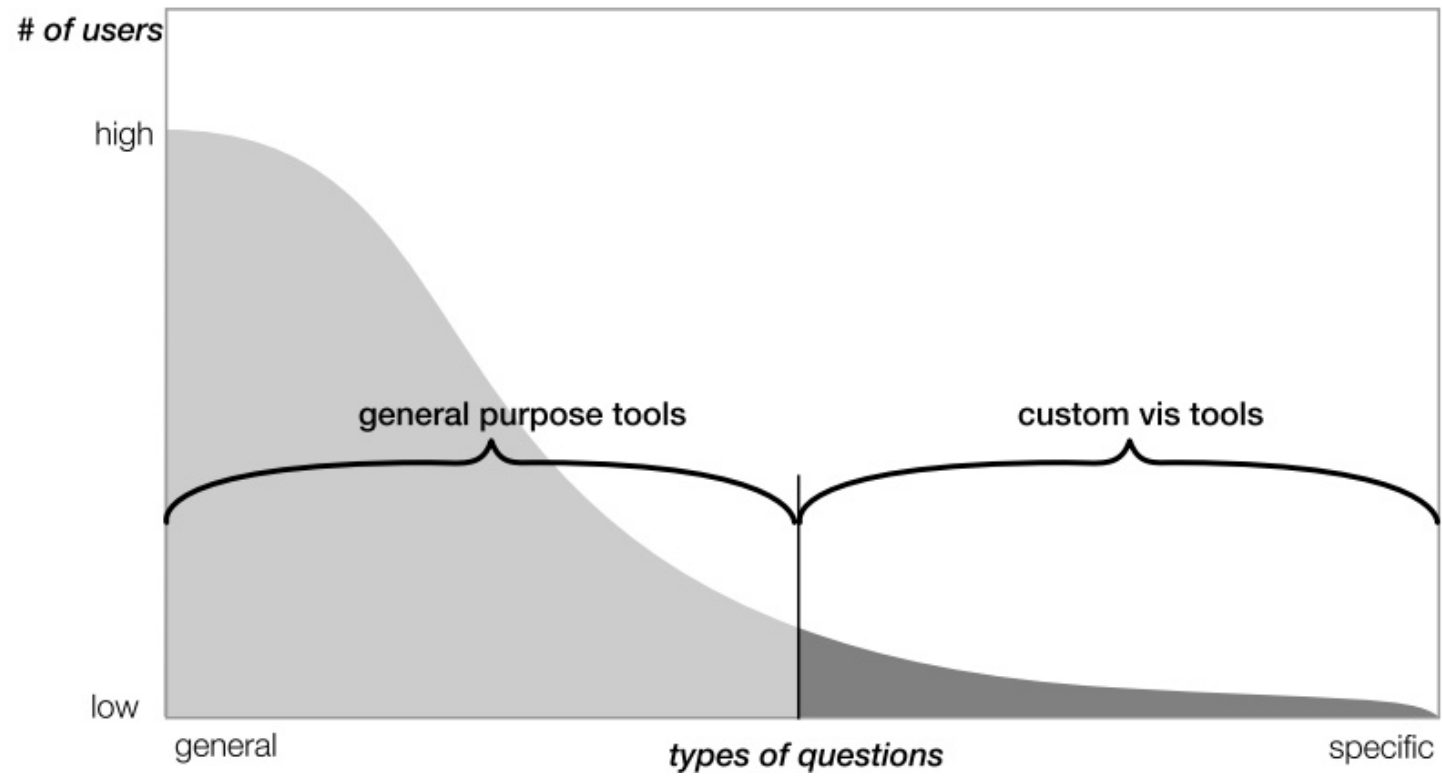
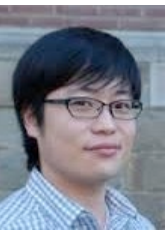


Design

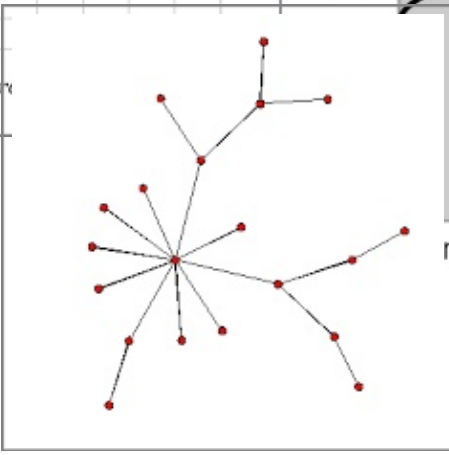
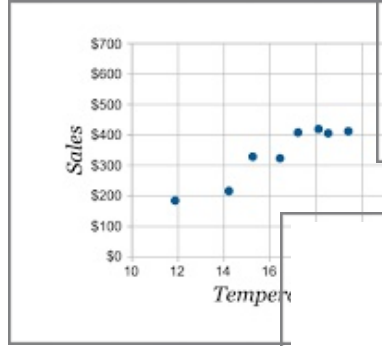
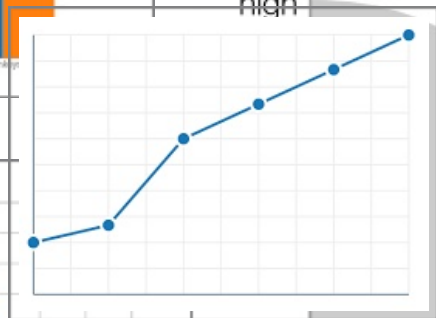
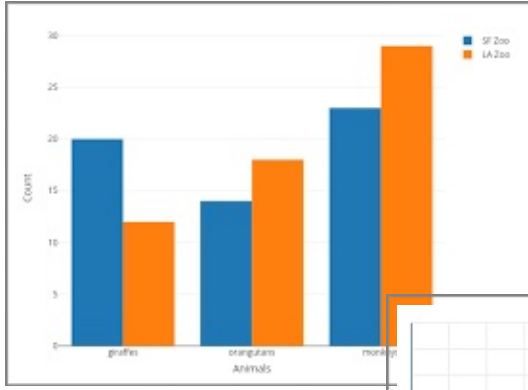
Multimodality in agriculture: challenging/impossible to integrate algorithmically

- fruit cultivar selection: multiomics + colour + yield + tasting panel + chemical analysis + ...
- plant infection: multiomics of plant + multiomics of pathogen + environment (water, temperature, ...) + management practices + chemical analysis + ...; time-series
- crop yield: satellite imagery + chemical analysis + management practices + cultivar + ...; time-series
- soil health: multiomics of crop + soil microbiome + environment (weather, soil type, ...) + management practices + ...; time-series
- ...

Design

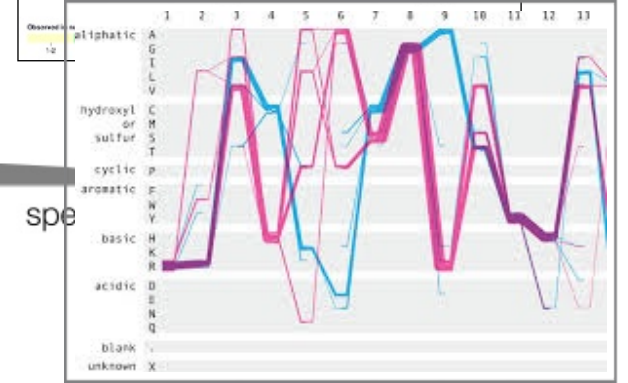
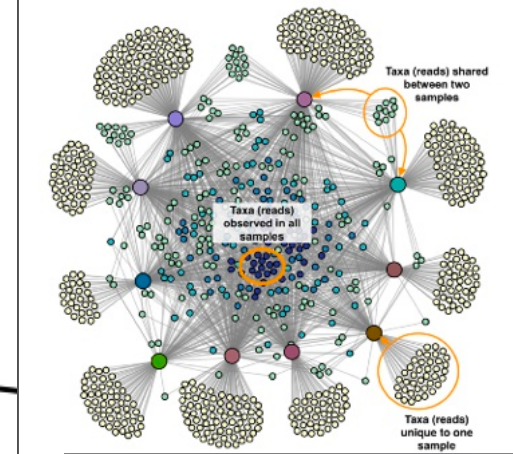


Design



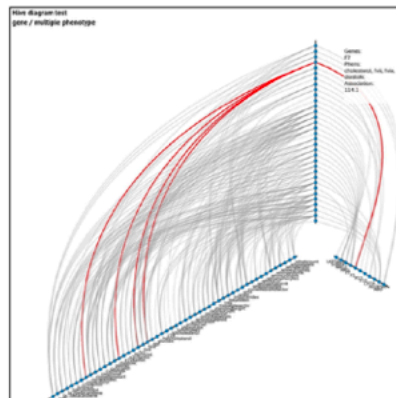
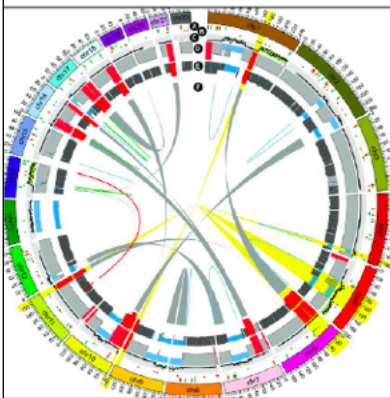
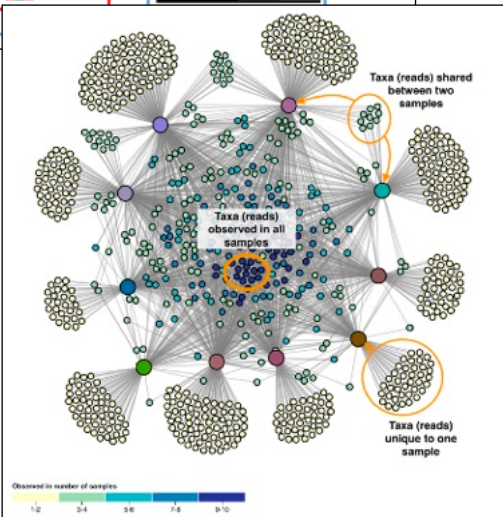
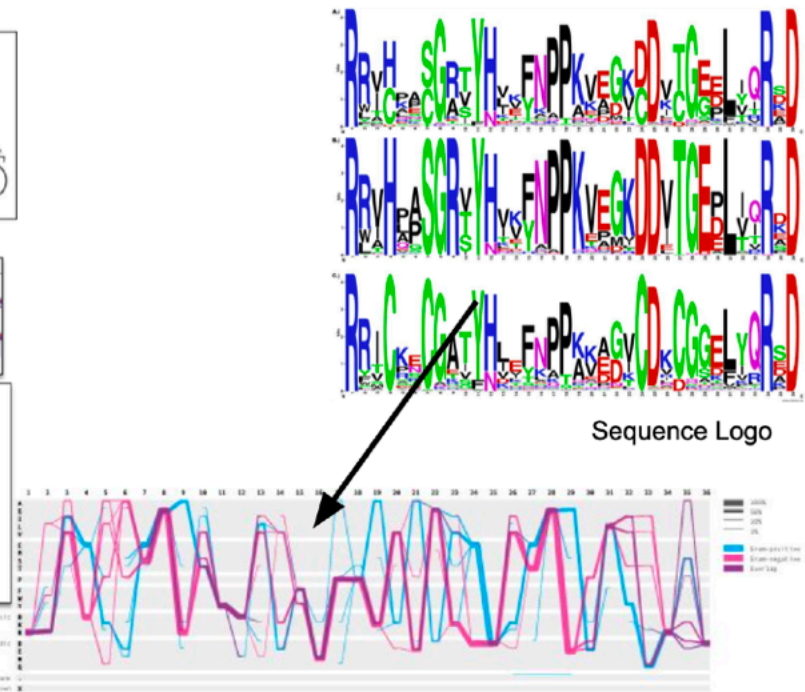
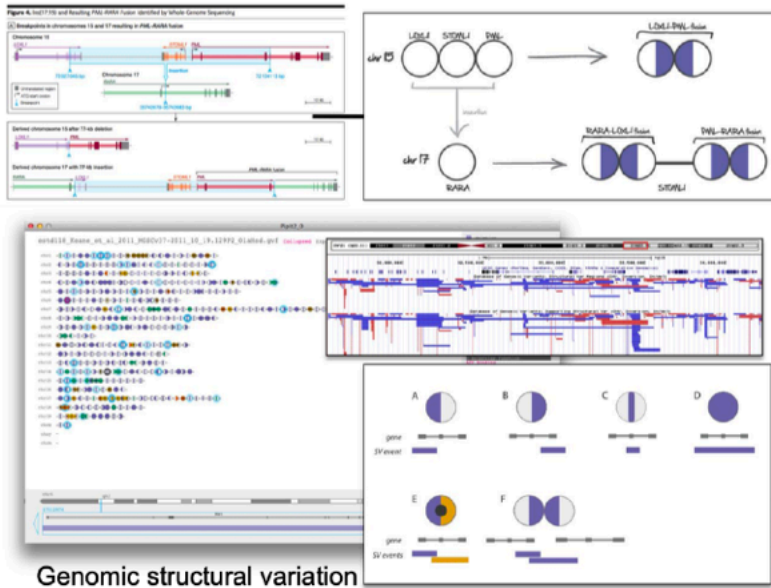
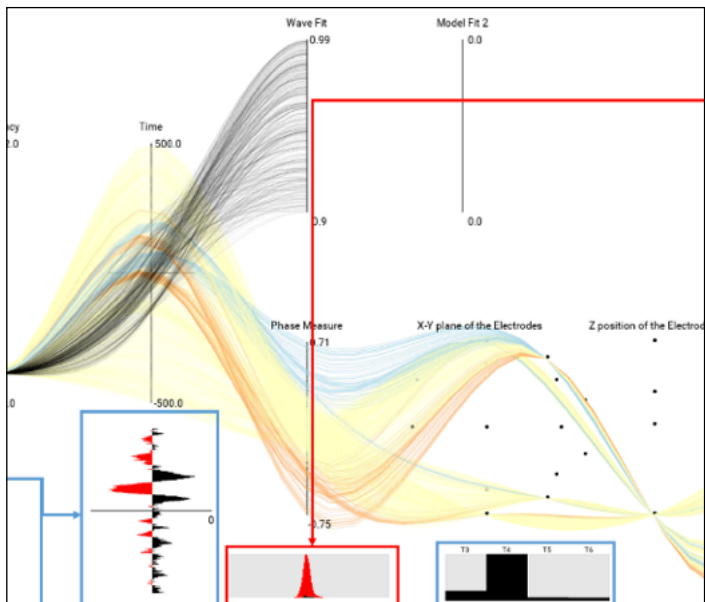
general purpose tools

custom vis tools

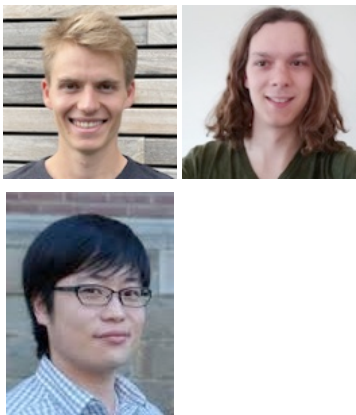


types of questions

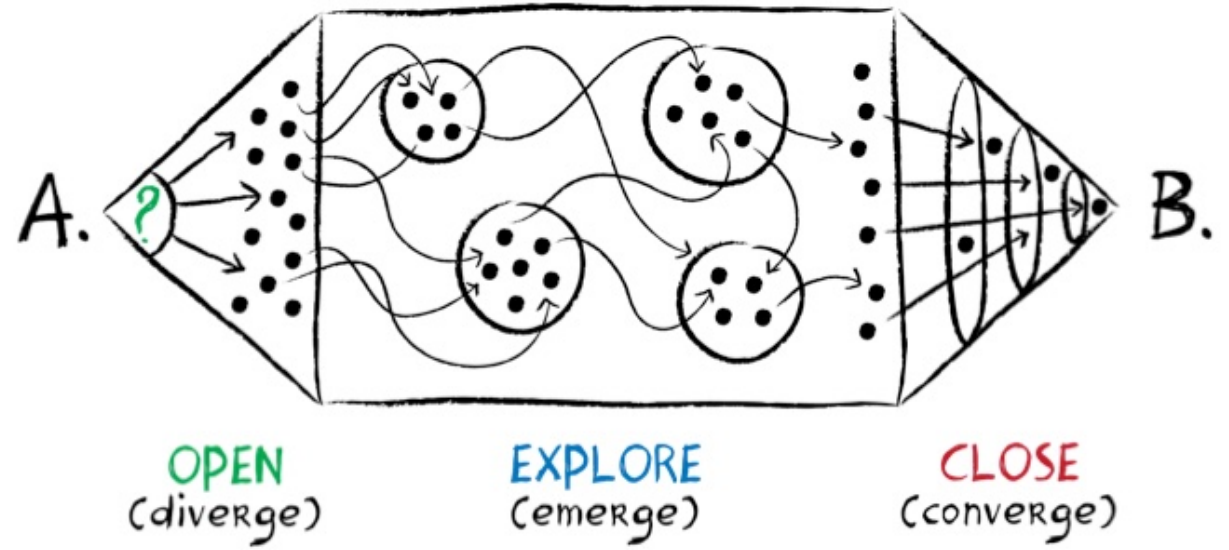
spe

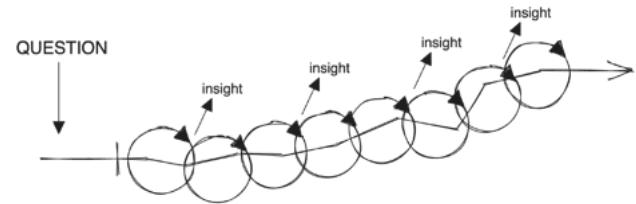
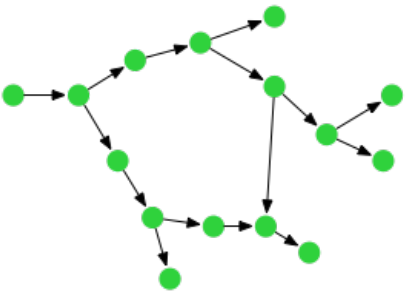
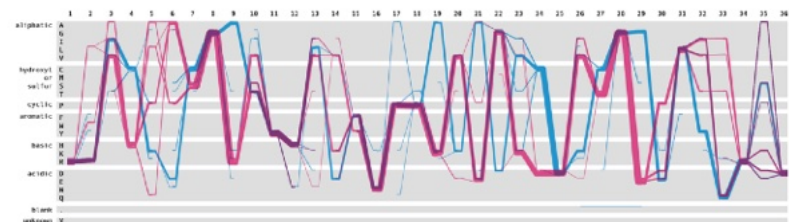
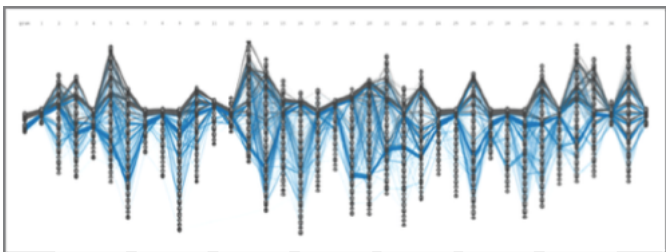
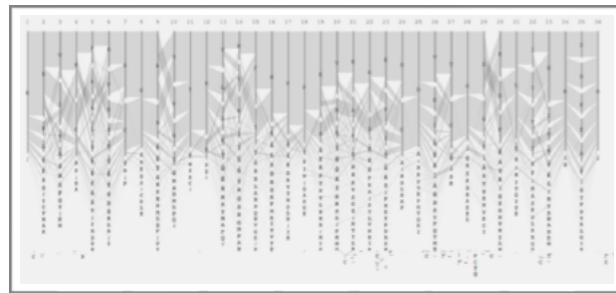
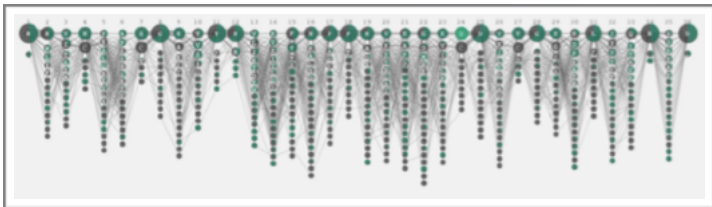
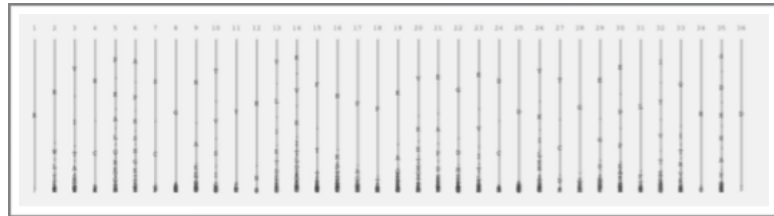
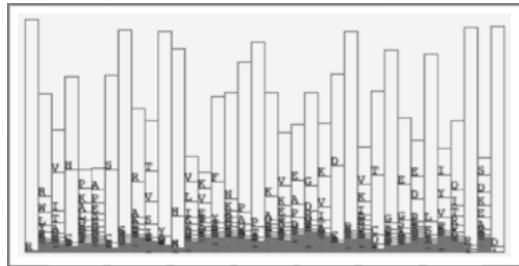
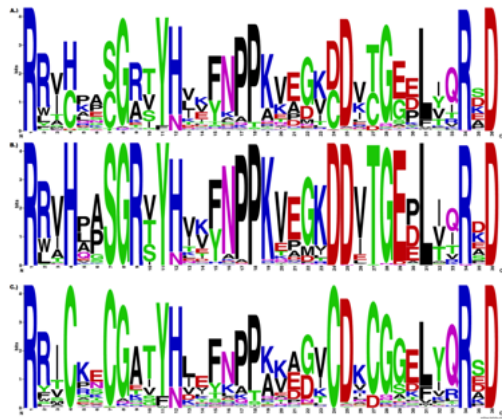
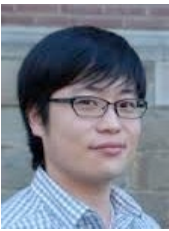


TIEVIs
Temporal Information from Patient Records

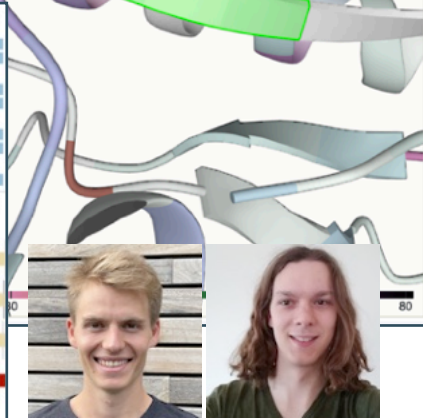
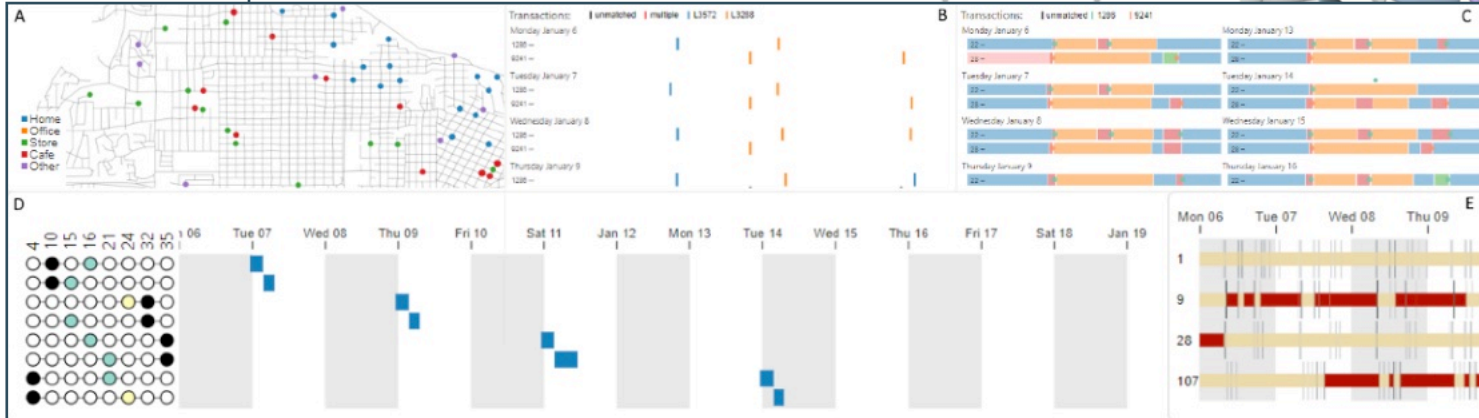
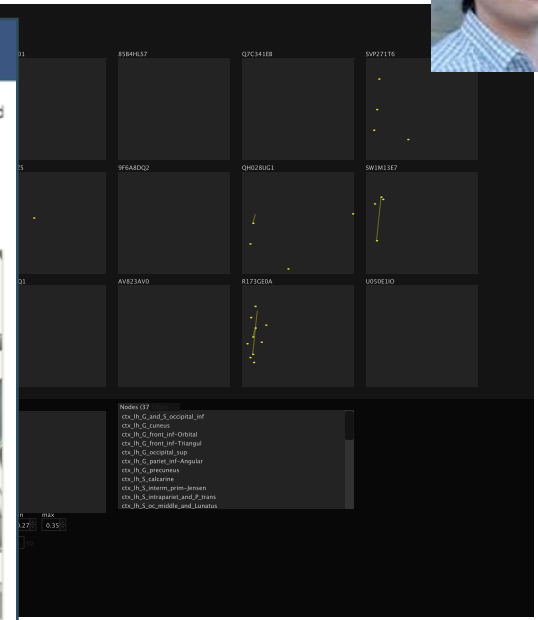
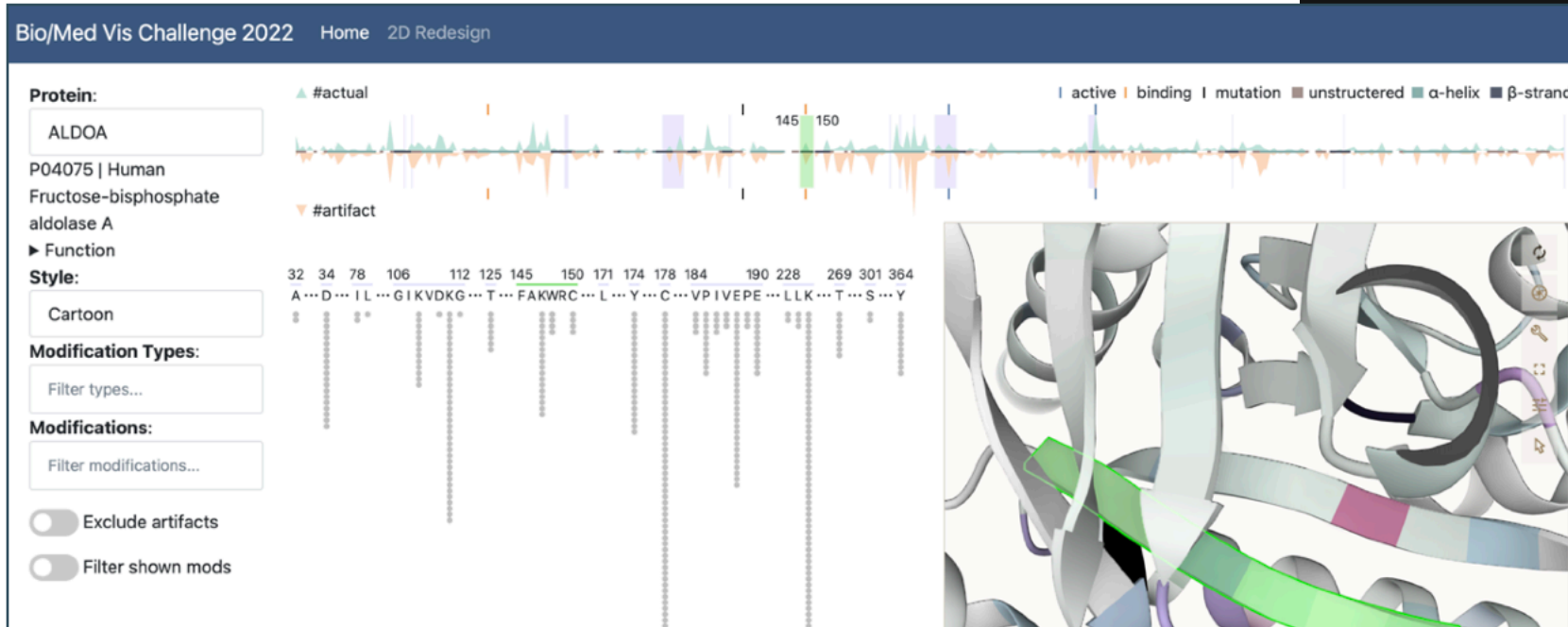


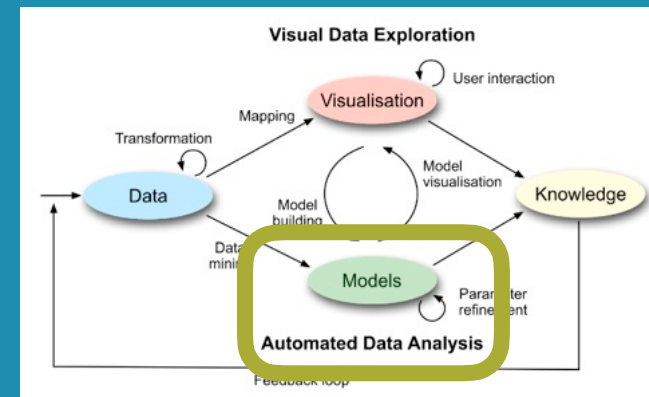
Snowflake
Microbiome visualisation





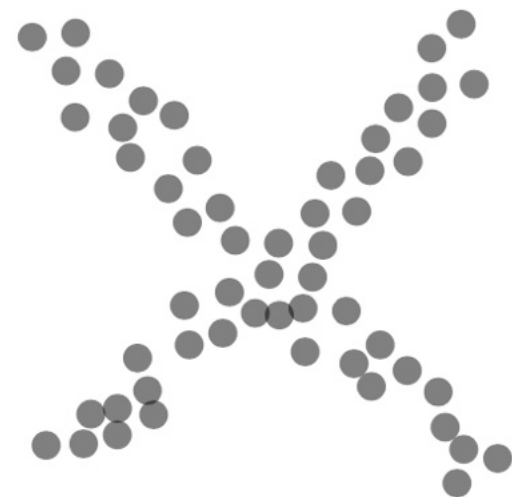
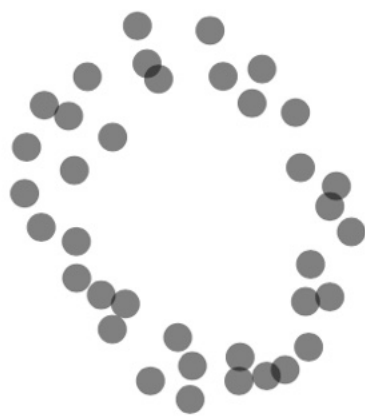
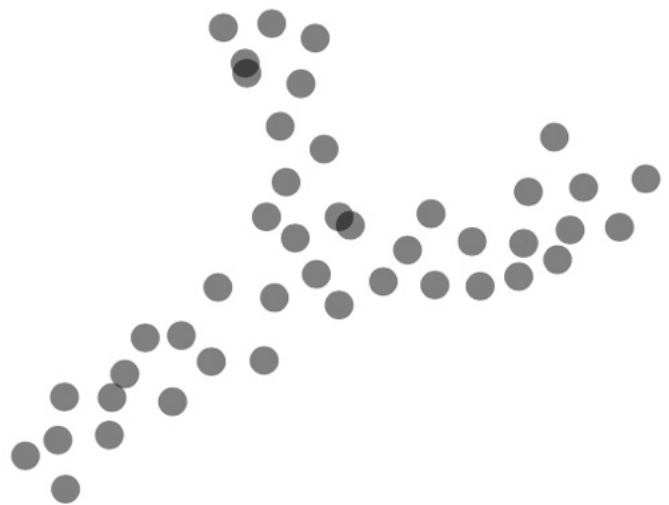
Develop



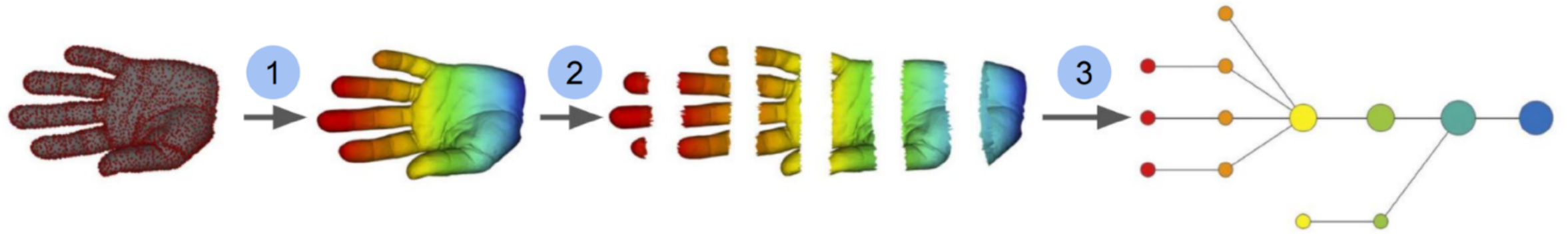


::B-1:: Topological Data Analysis

Data has shape, shape has meaning

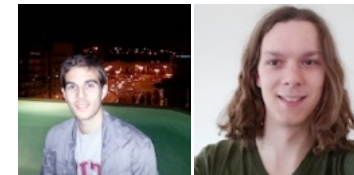


high-dimensional data => 0-dimensional network

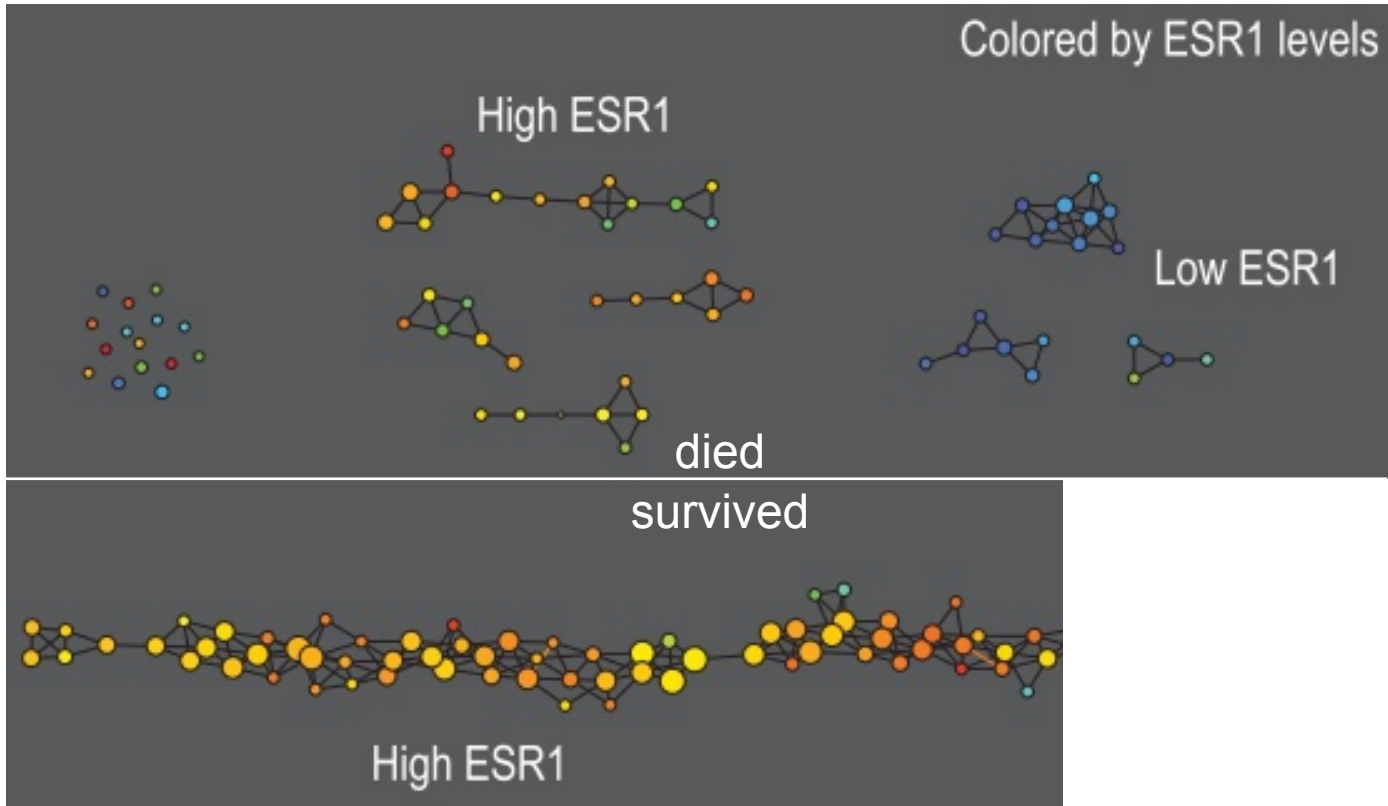


Adapted from Lum et al, 2014

alternative: STAD - Simplified Topological Approximation of Data



Breast cancer patients (NKI dataset)



Extracting insights from the shape of complex data using topology

P. Y. Lum¹, G. Singh¹, A. Lehman¹, T. Ishkanov¹, M. Vejdemo-Johansson², M. Alagappan¹, J. Carlsson³ & G. Carlsson^{1,4}

Scientific Reports volume 3, Article number: 1236 (2013)

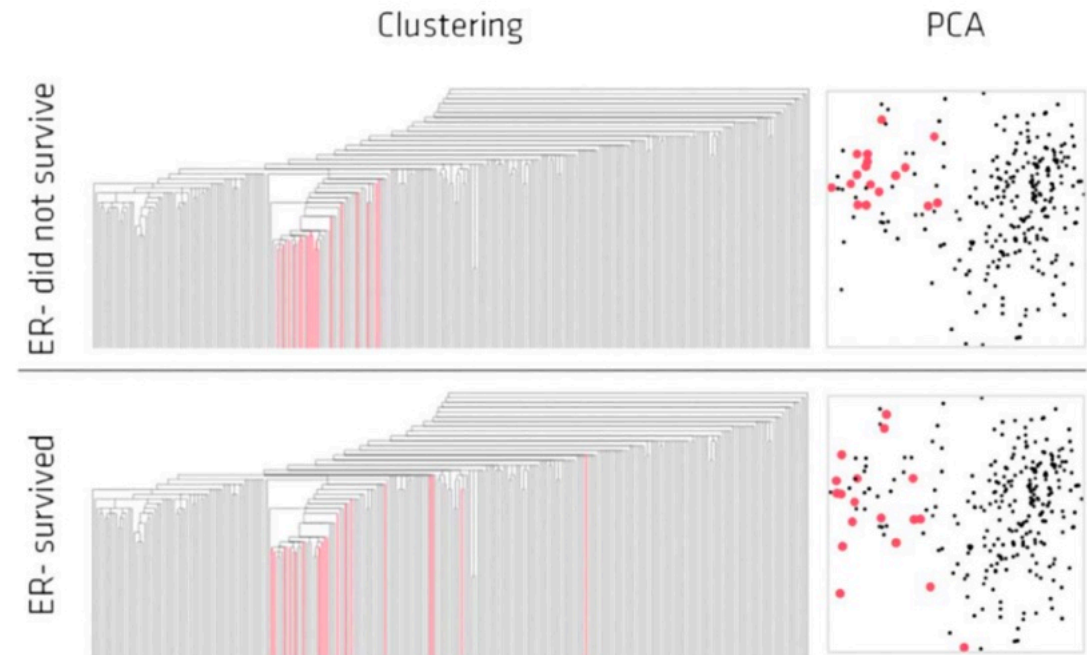
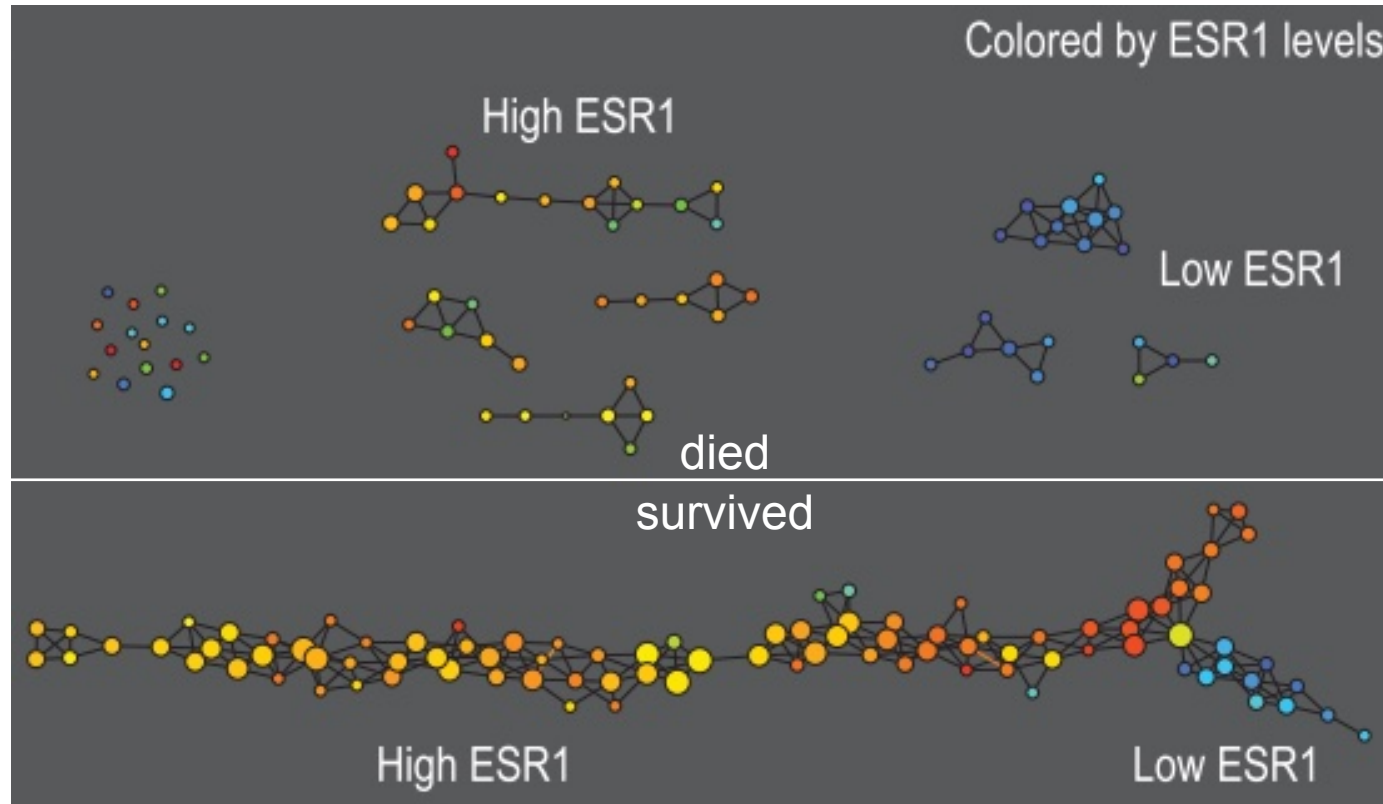


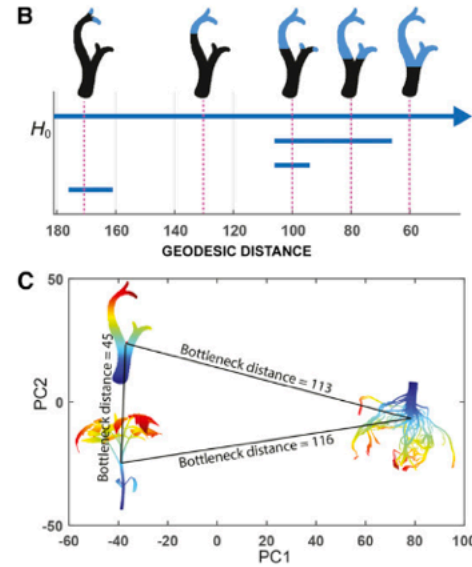
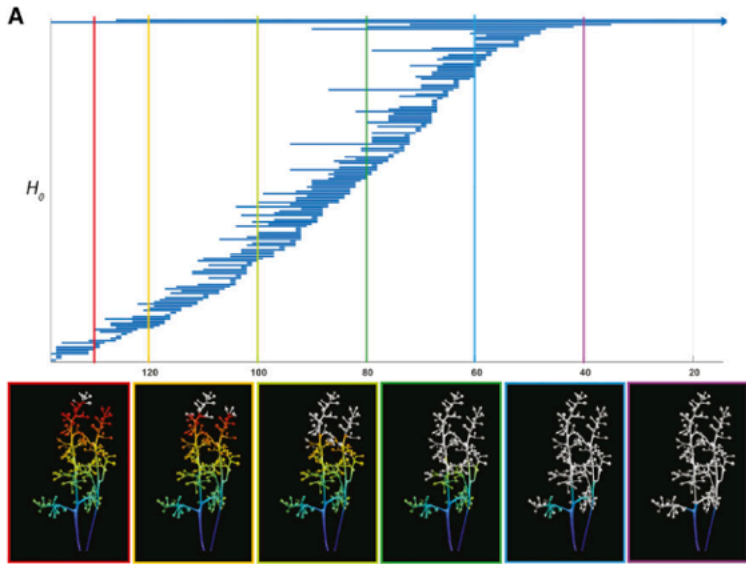
Breast cancer patients (NKI dataset)

Extracting insights from the shape of complex data using topology

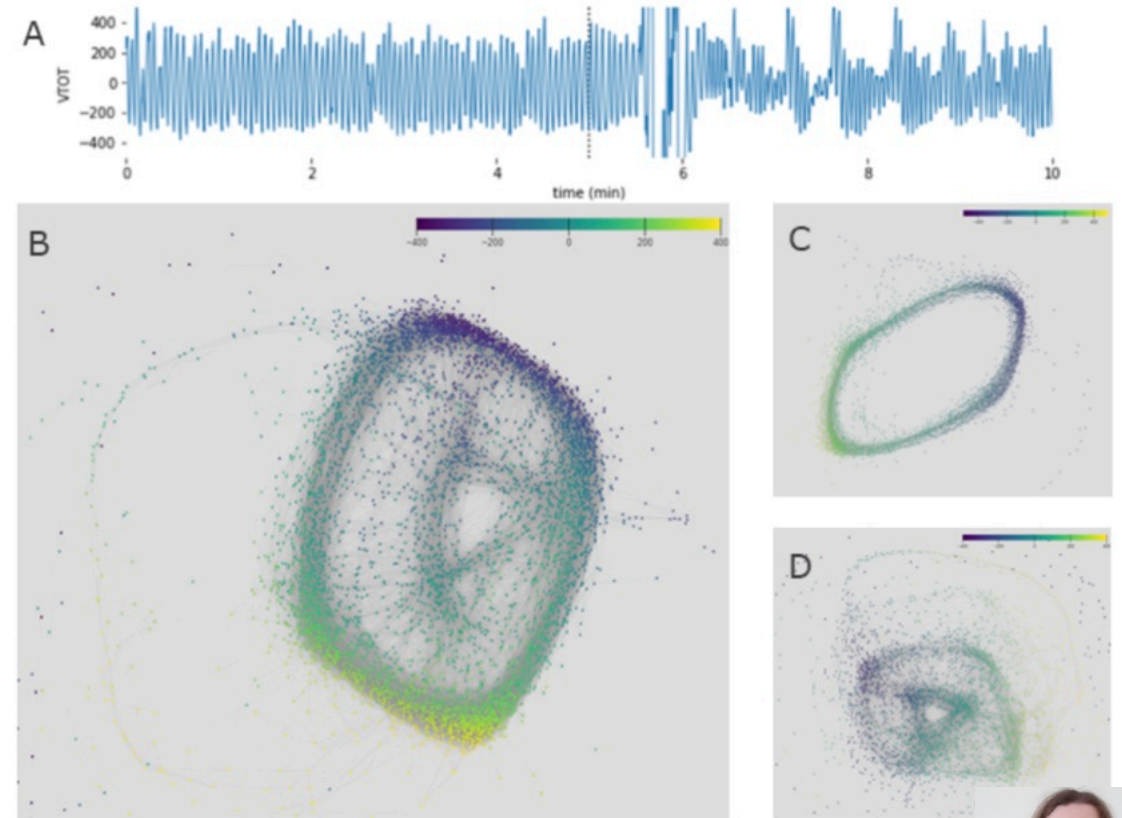
P. Y. Lum¹, G. Singh¹, A. Lehman¹, T. Ishkanov¹, M. Vejdemo-Johansson², M. Alagappan¹, J. Carlsson³ & G. Carlsson^{1,4}

Scientific Reports volume 3, Article number: 1236 (2013)





persistence barcode to describe shape of a grape cluster (Li et al, 2017)



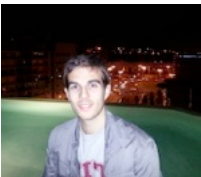
chest volume while breathing: topology of sleep apnea



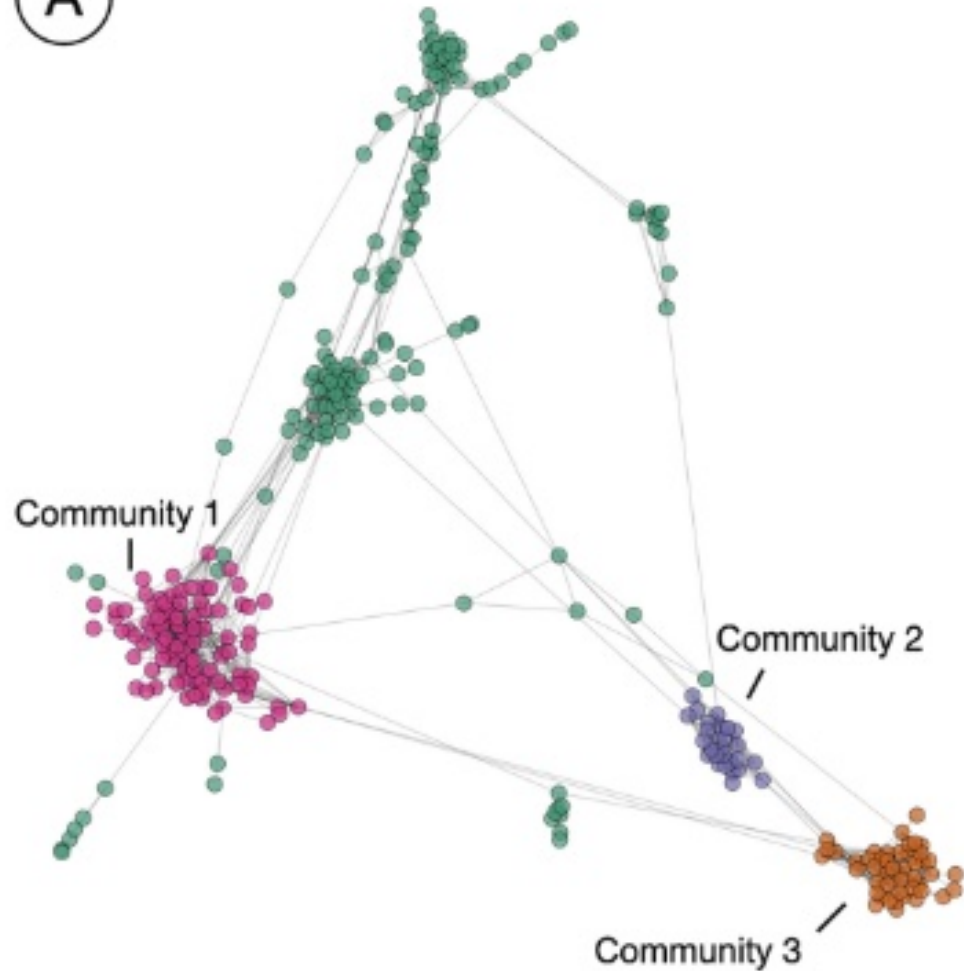
Inserting domain knowledge: 1. choice of distance metric

Patient A (115057)			Patient B (117154)		
ICD section	Label (ICD9)		ICD section	Label (ICD9)	
1	996-999.	Infection and inflammatory reaction due to other vascular device, implant, and graft (99662)	1	430-438.	Unspecified intracranial hemorrhage (4329)
2	990-995.	Sepsis (99591)	2	430-438.	Cerebral artery occlusion, unspecified with cerebral infarction (43491)
3	590-599.	Urinary tract infection, site not specified (5990)	3	996-999.	Iatrogenic cerebrovascular infarction or hemorrhage (99702)
4	401-405.	Unspecified essential hypertension (4019)	4	990-995.	Sepsis (99591)
			5	590-599.	Urinary tract infection, site not specified (5990)
			6	401-405.	Unspecified essential hypertension (4019)

$$M(c_A, c_B) = \ln\left(1 + \frac{1}{\max(\text{position}(c_A), \text{position}(c_B))}\right) \longrightarrow S(X, Y) = \sum_n^{i=1} M(X \cap Y)$$



(A)

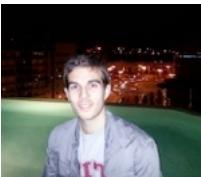


MIMIC-III dataset: ICU patients who have a diagnosis of “alcohol withdrawal delirium”

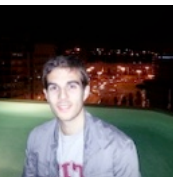
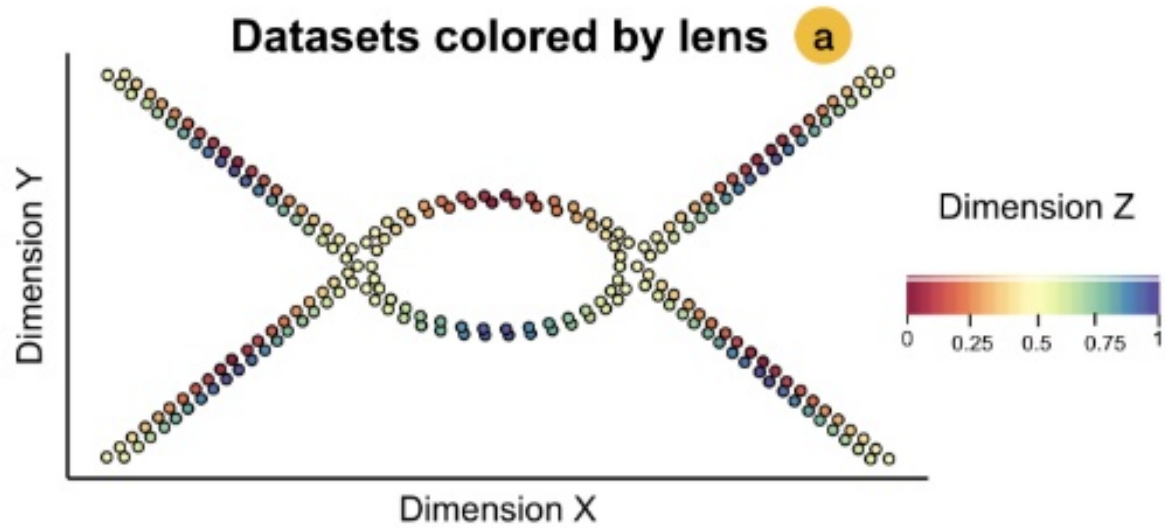
Community 1: main diagnosis = alcohol withdrawal delirium

Community 2: main diagnosis = intracranial injuries (e.g. concussion)

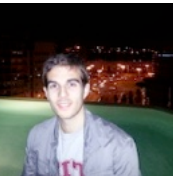
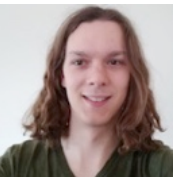
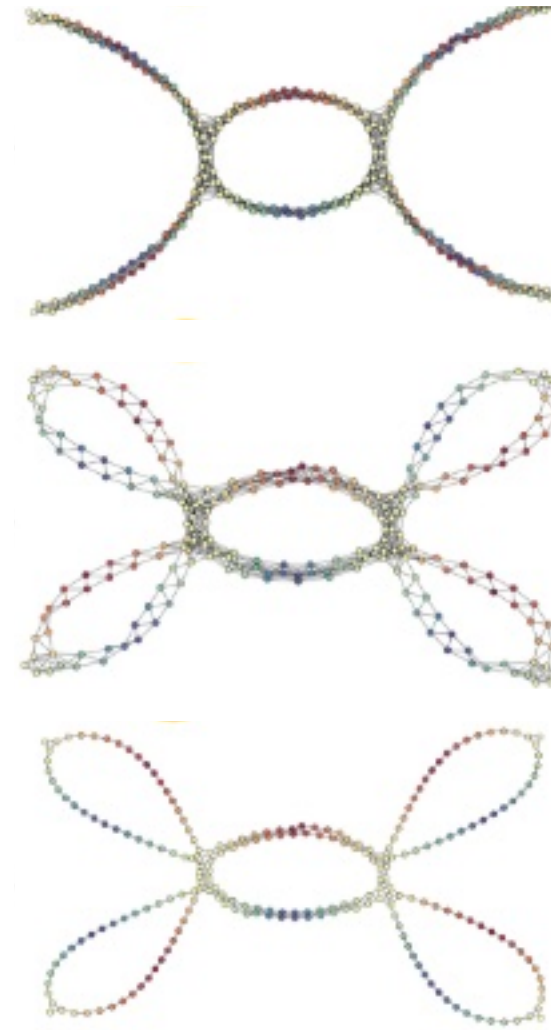
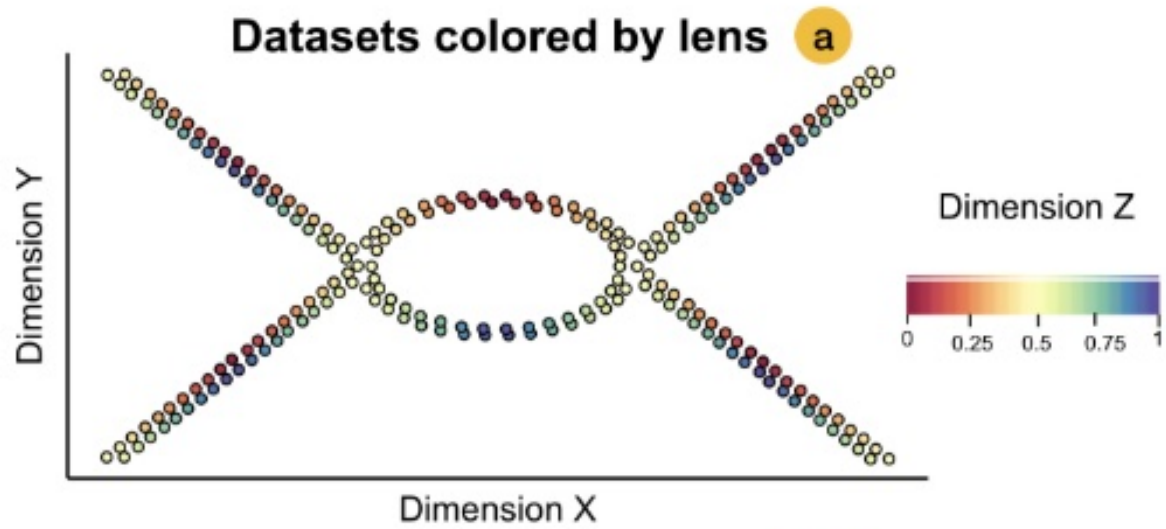
Community 3: main diagnosis = head trauma/fractures



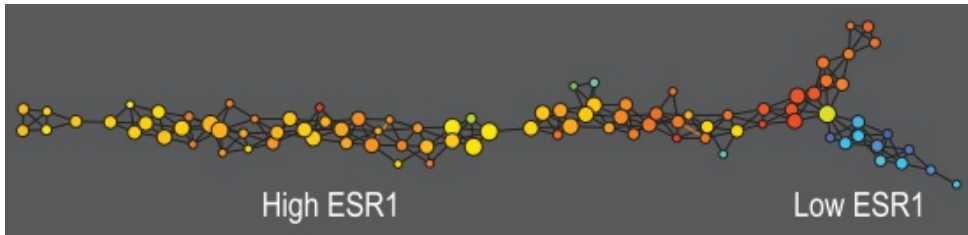
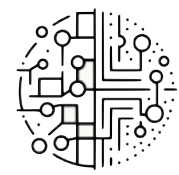
Inserting domain knowledge: 2. lens



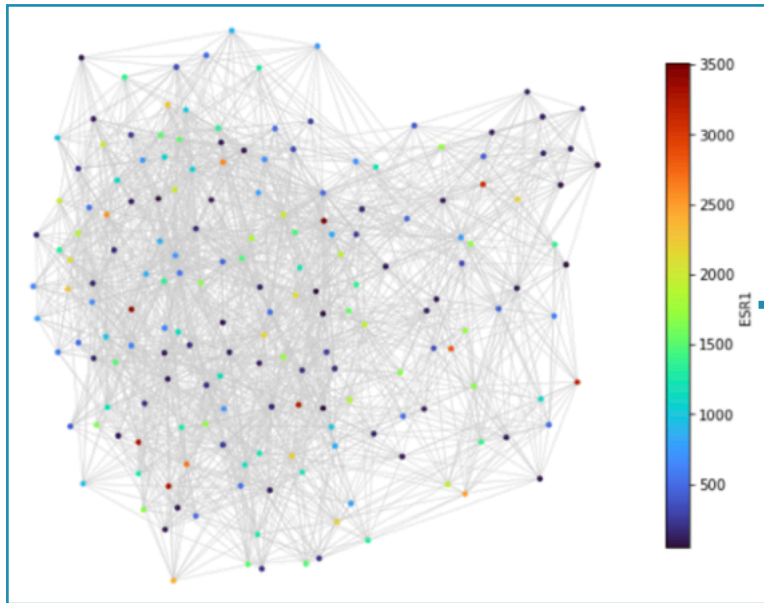
Inserting domain knowledge: 2. lens



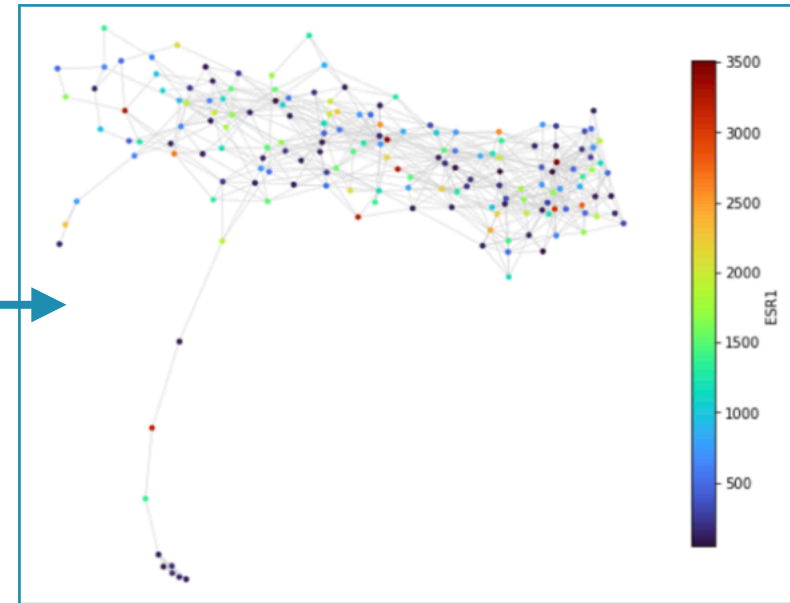
"Lensed" UMAP: dimensionality reduction with domain knowledge



UMAP



UMAP with lens (L_∞ -centrality)

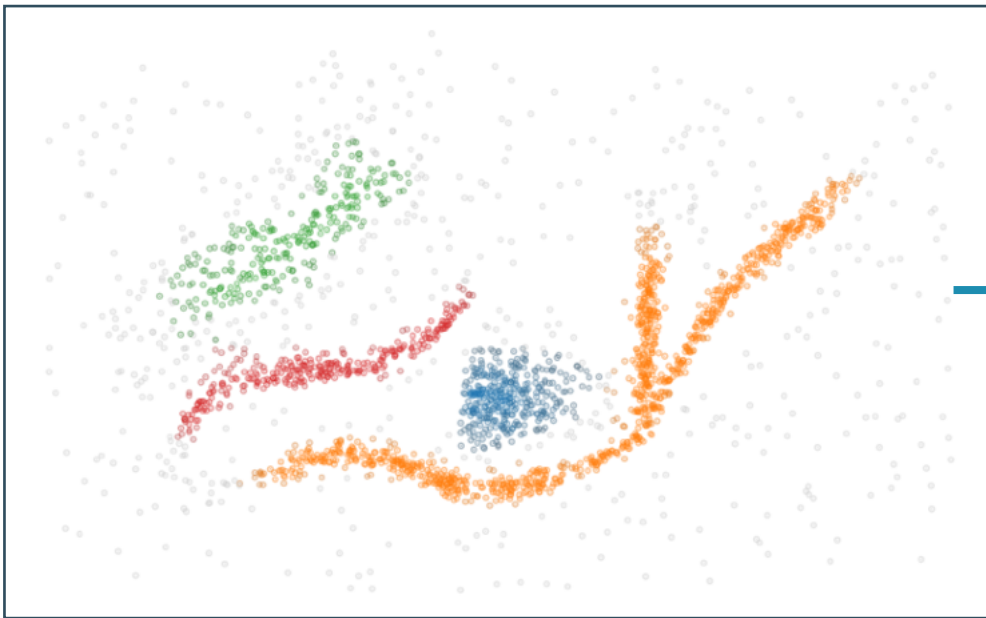


FLASC - FLAre-Sensitive Clustering

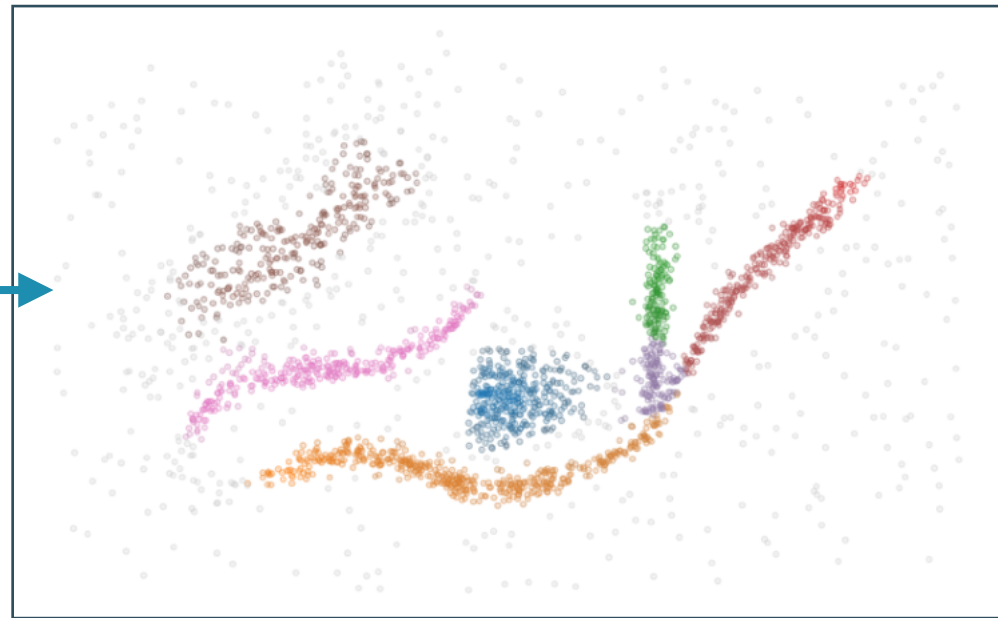
extension to HDBSCAN clustering algorithm

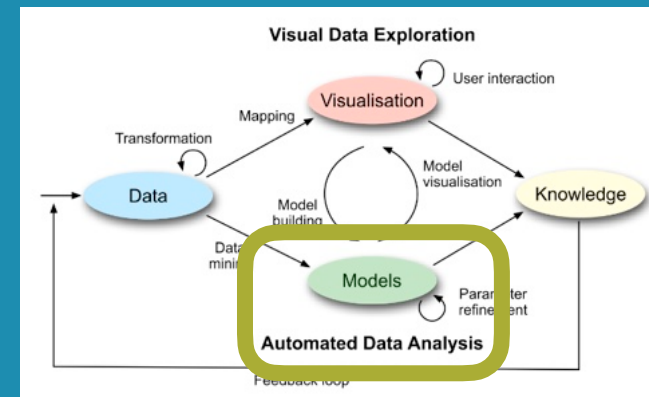


HDBSCAN

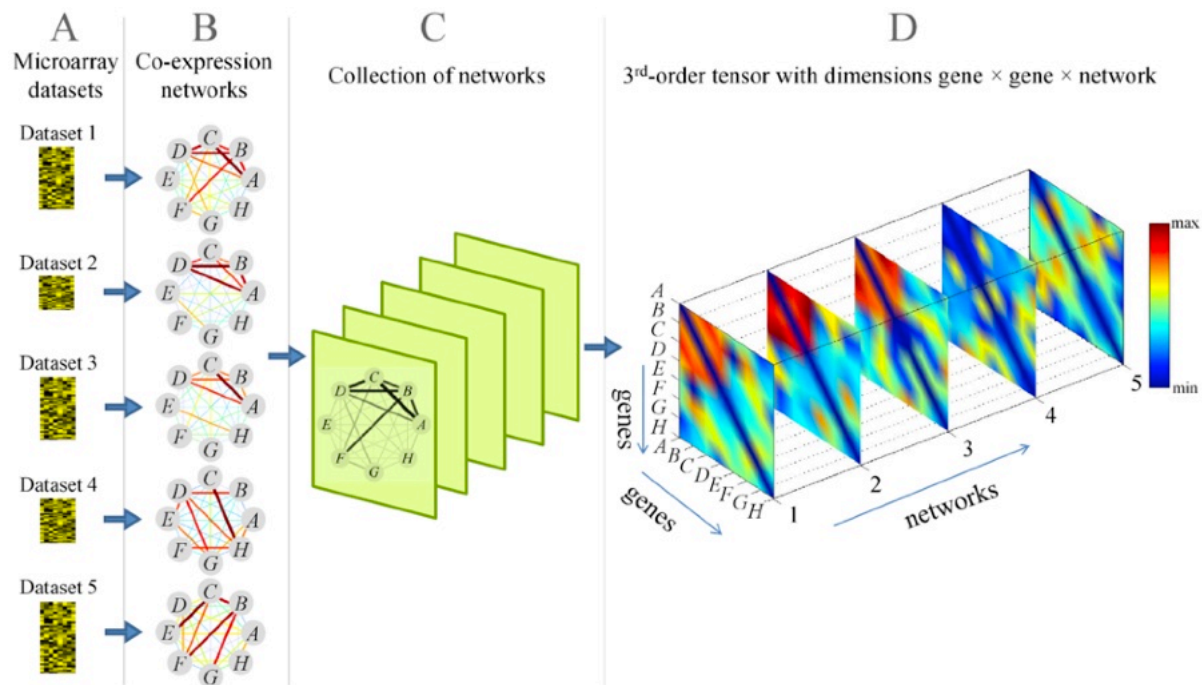


FLASC





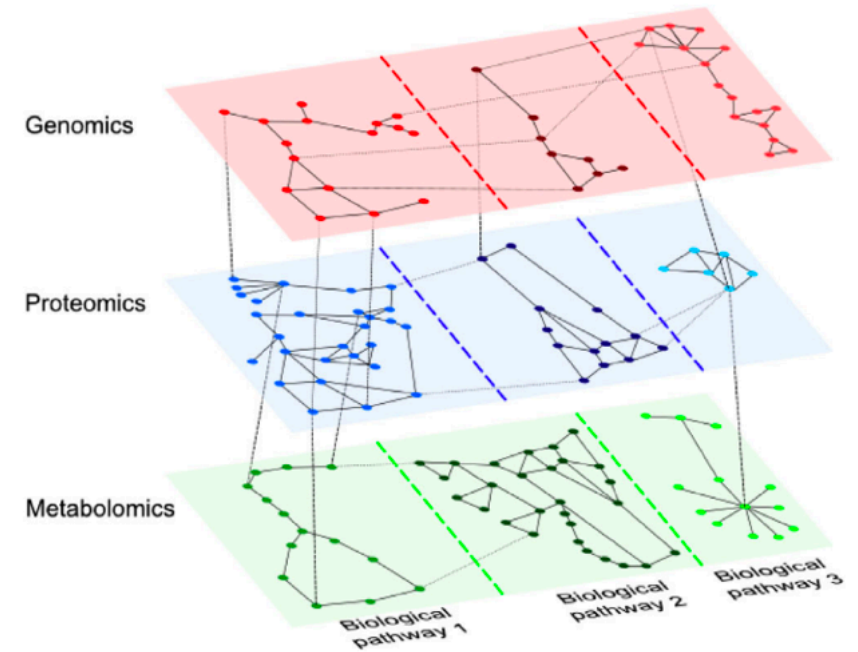
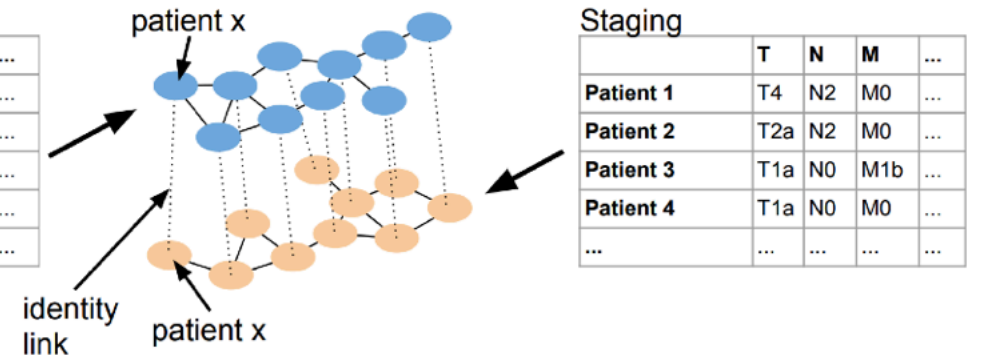
::B-2:: Multi-layer networks



Li et al. PLoS Comp Biol, 2011

Patient blood values

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



McGee et al. Computer Graphics Forum, 2019

Future research

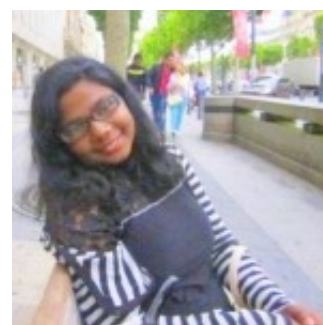
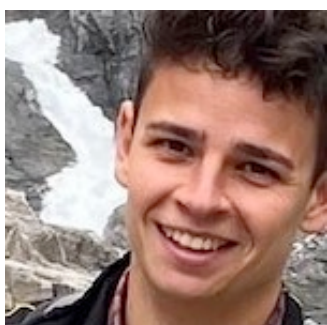
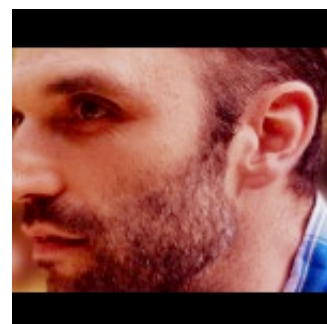
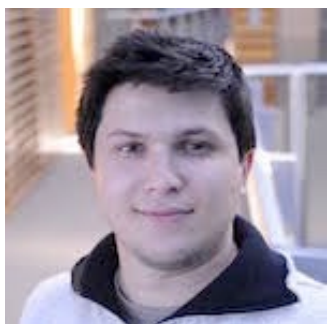
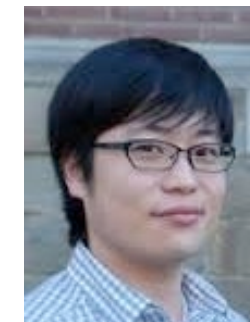
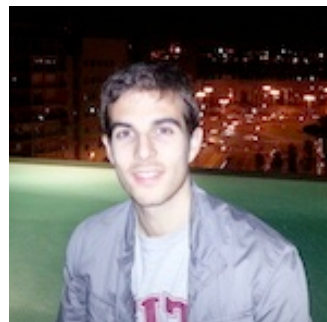
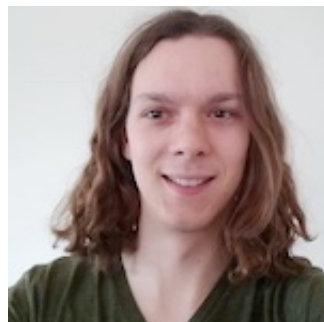
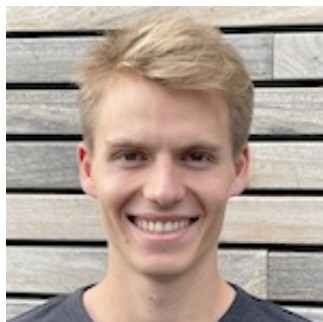
- Integration of multi-modal bio/agricultural data
- Interactive exploration of multi-level geolocation data
- Topological data analysis for parameter space exploration
- Multi-resolution representations and integration of time-series data
- ...

Final aim

- ... to help experts and lay individuals embrace complexity
- ... to help experts and lay individuals consider uncertainty
- ... to support interactions between expert and lay individual

All to get a better understanding of our natural world and work towards tackling loss of biodiversity, increasing biosystem resilience , and ensuring agricultural sustainability.

Thanks to...





jan.aerts@kuleuven.be
<http://vda-lab.io>