Advanced Data Visualization

CS 6965

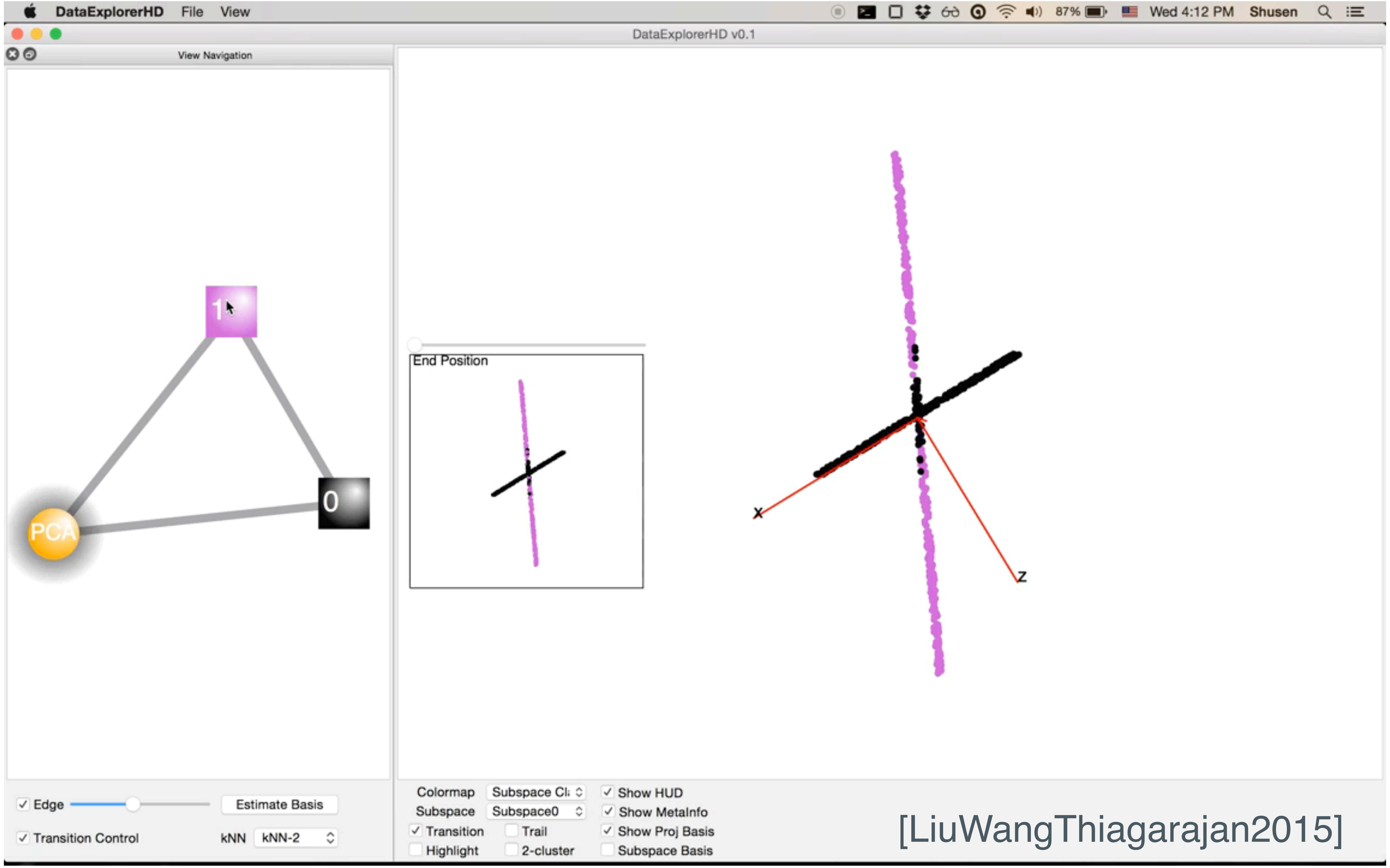
Spring 2018

Prof. Bei Wang Phillips

University of Utah

Lecture 01

Introduction



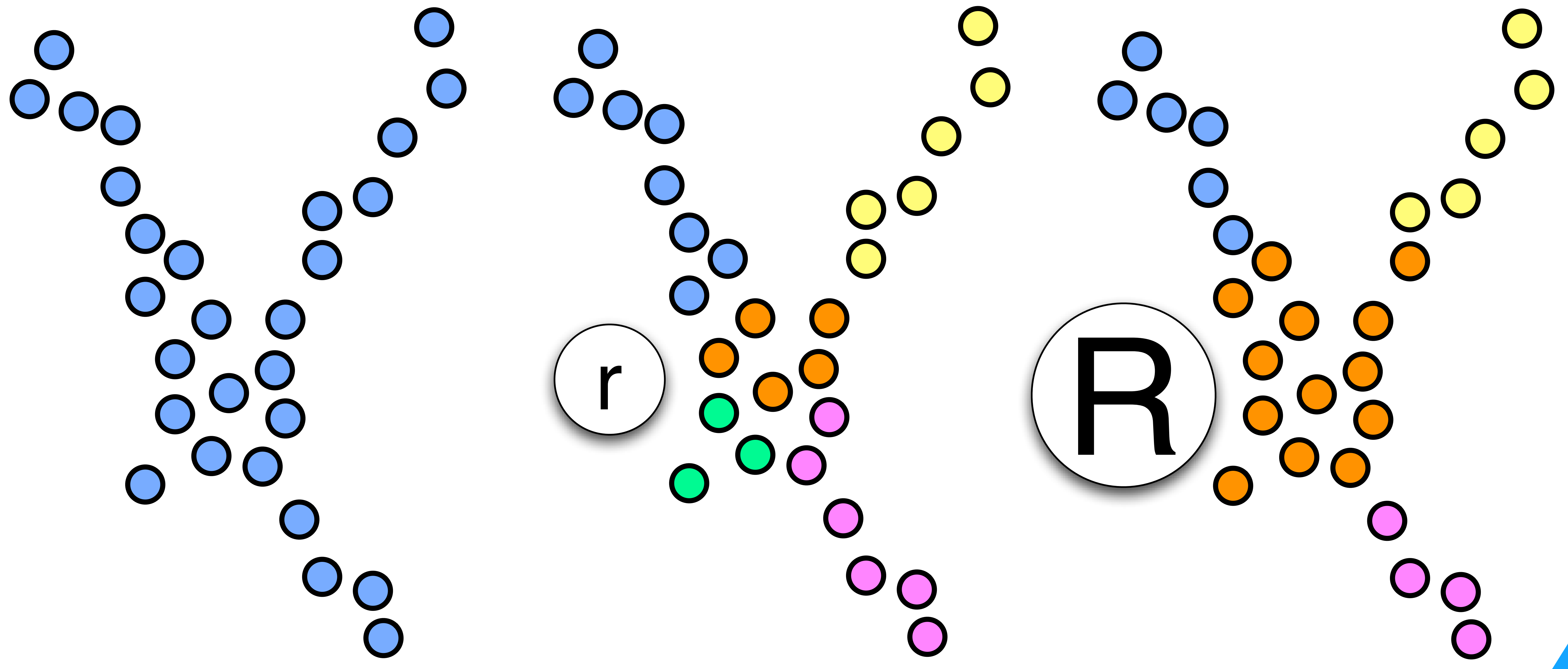
Visualization is an integral part of data analysis

Who am I?

Prof. Bei Wang Phillips

Data Analysis and Data Visualization

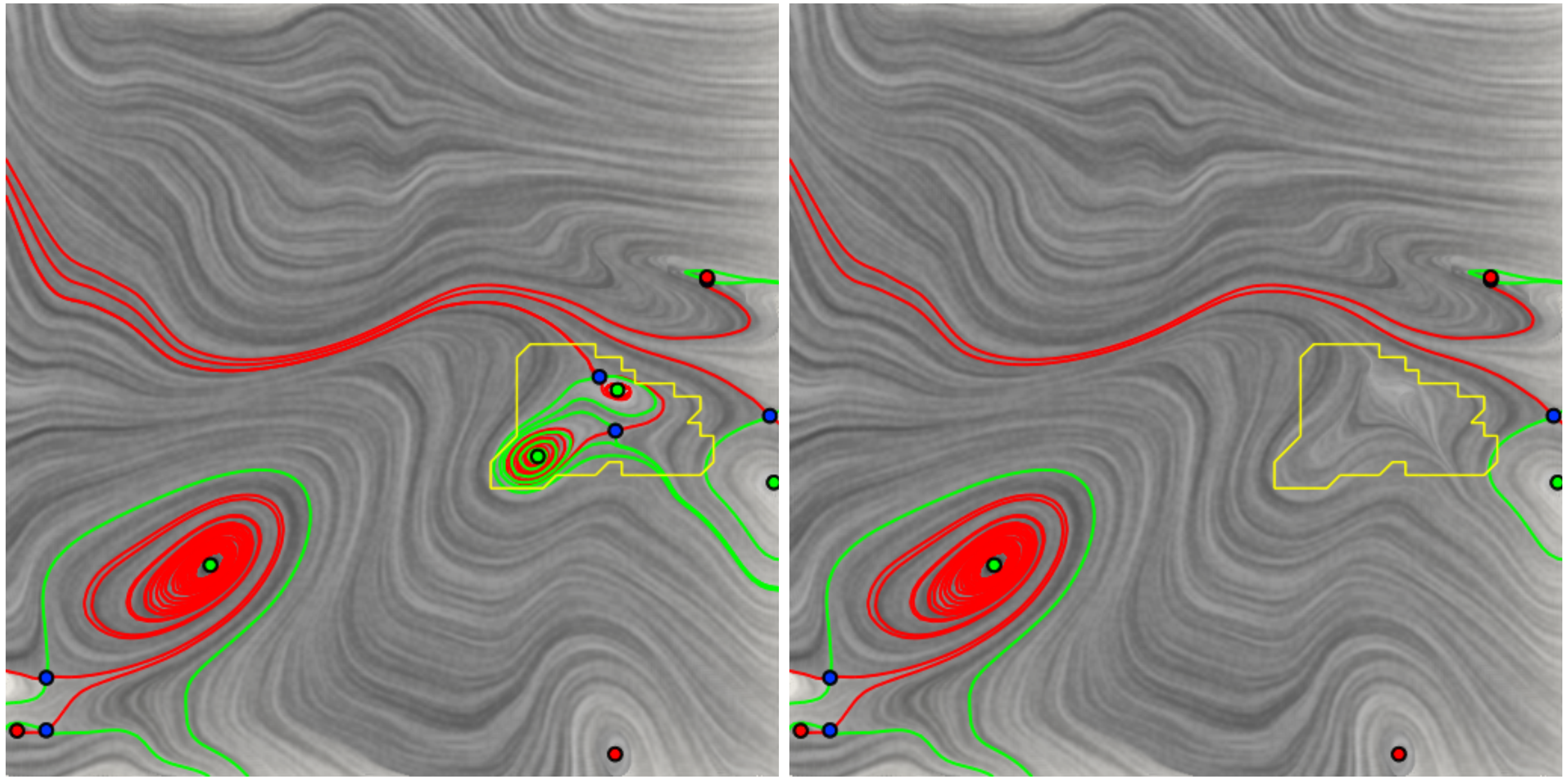
beiwang@sci.utah.edu
http://www.sci.utah.edu/~beiwang/



Structural inference

of point cloud data & stratification learning

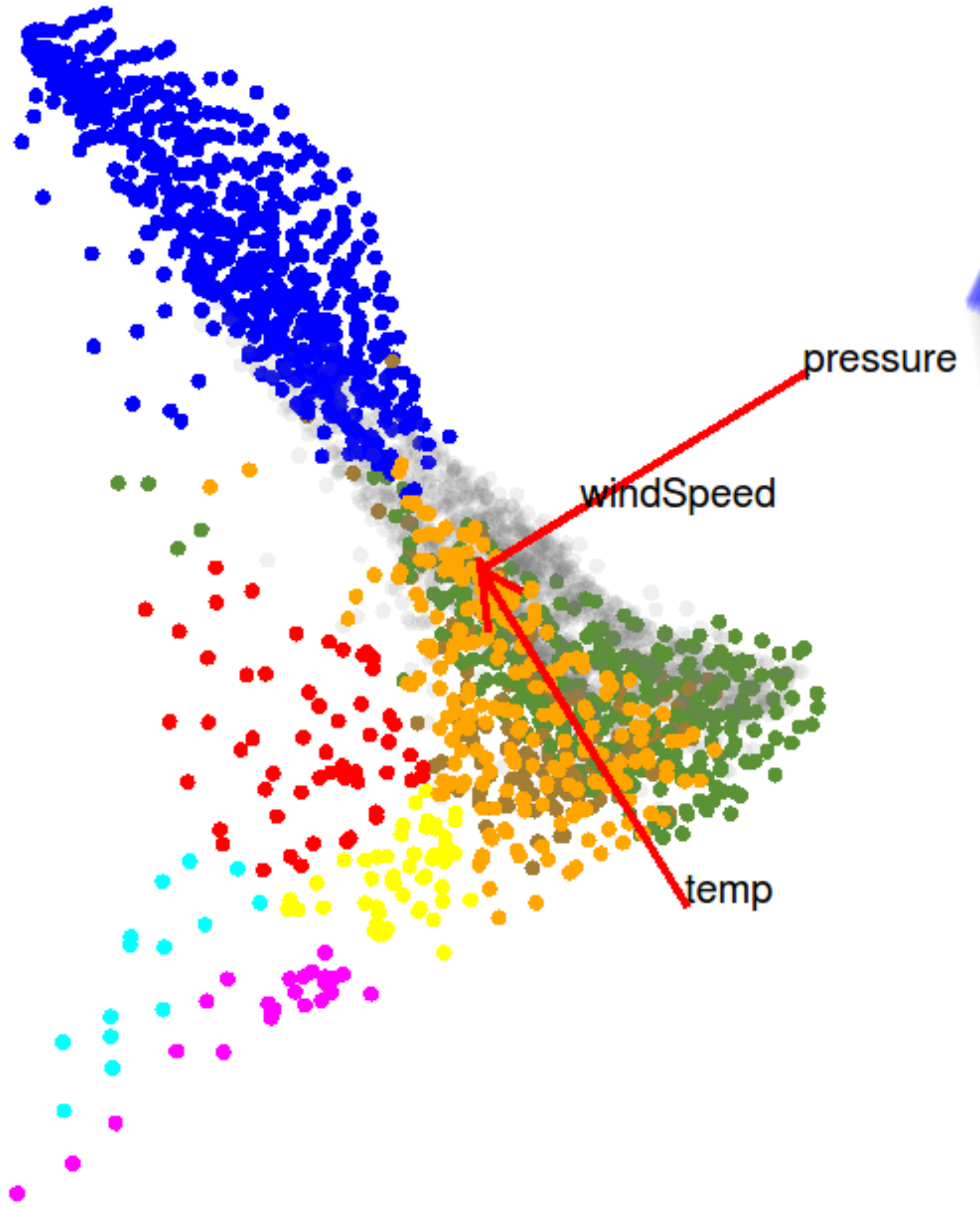
Source: Bei Wang

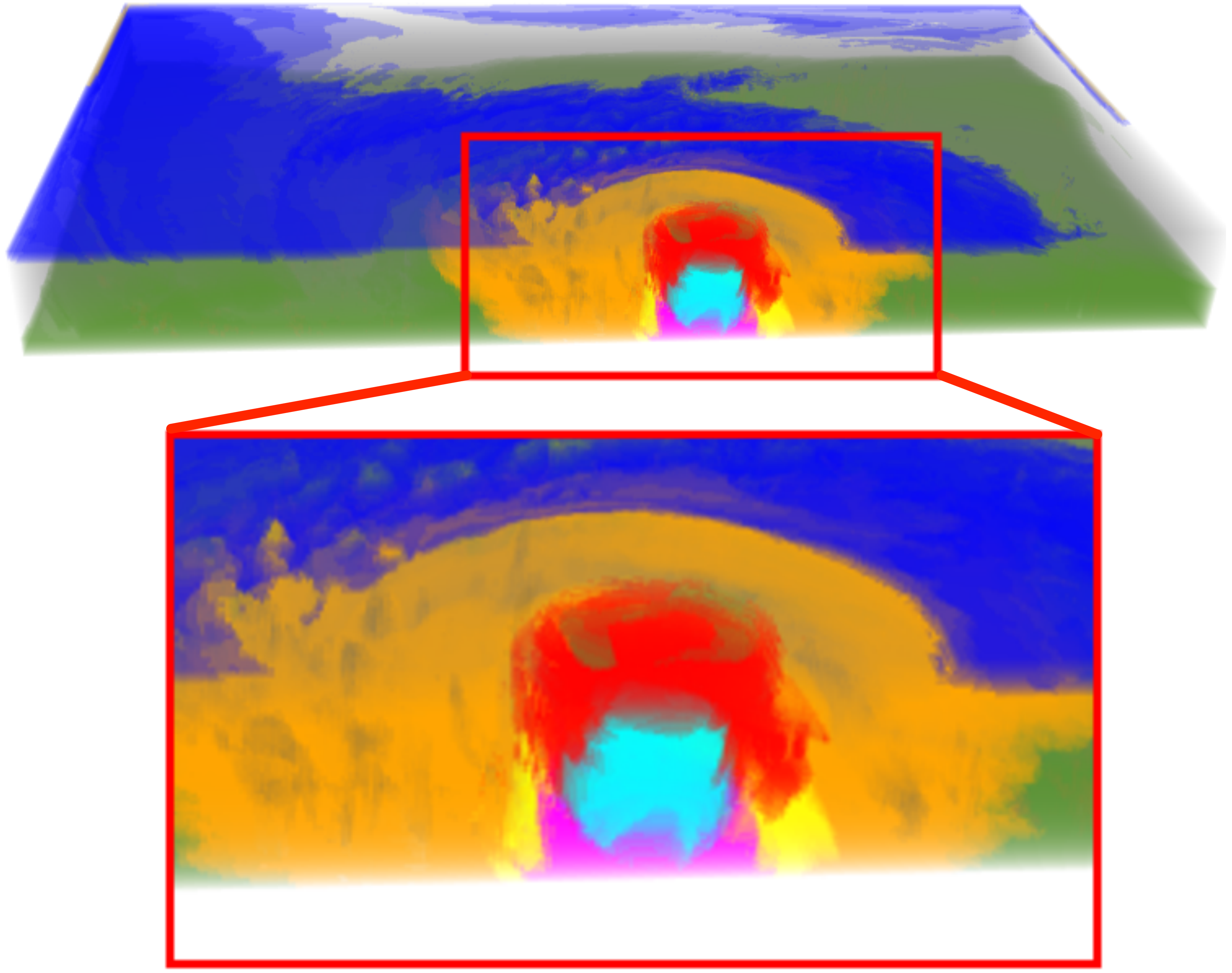


Robust Feature Extraction

of vector field data

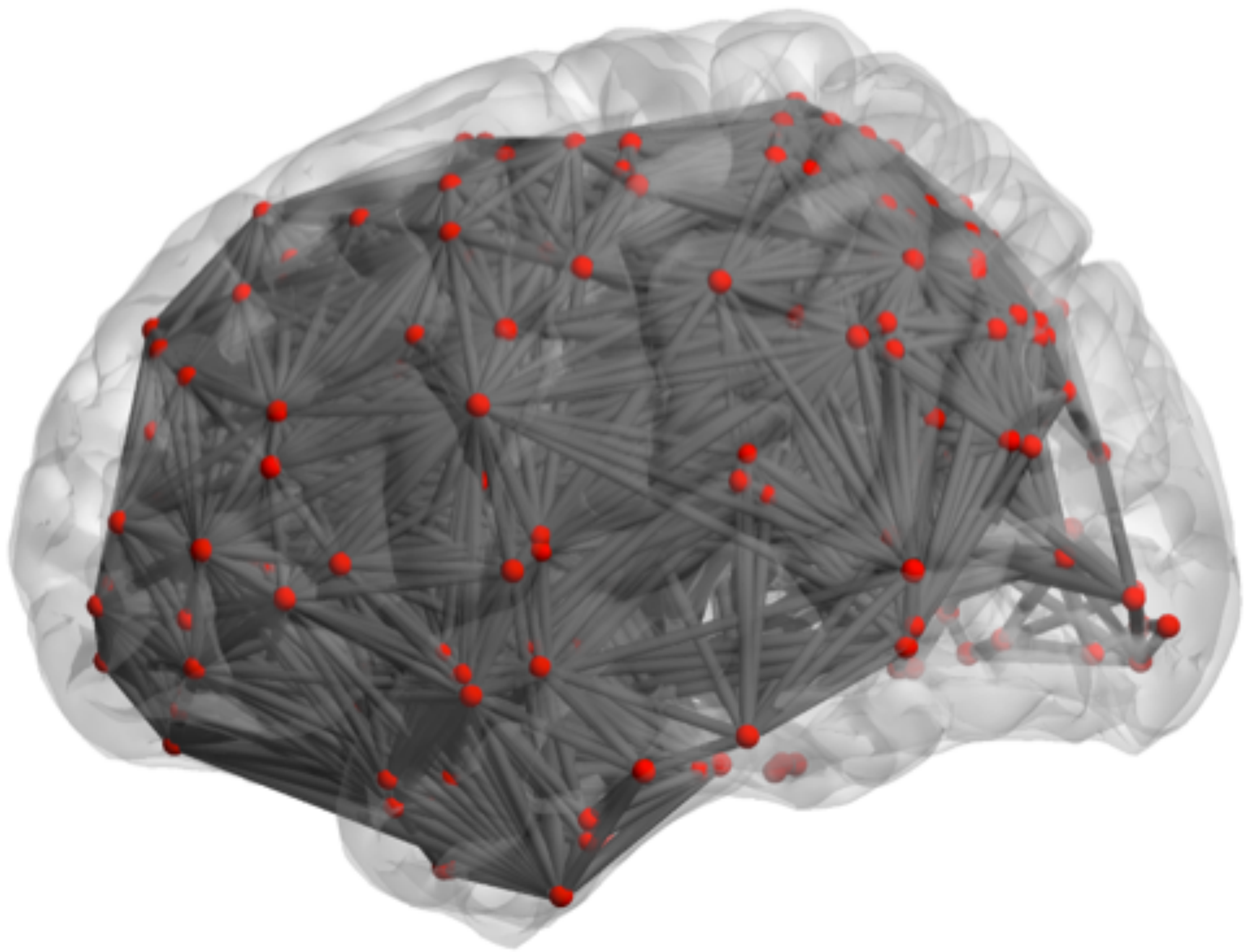
[SkrabaWangChen2015]

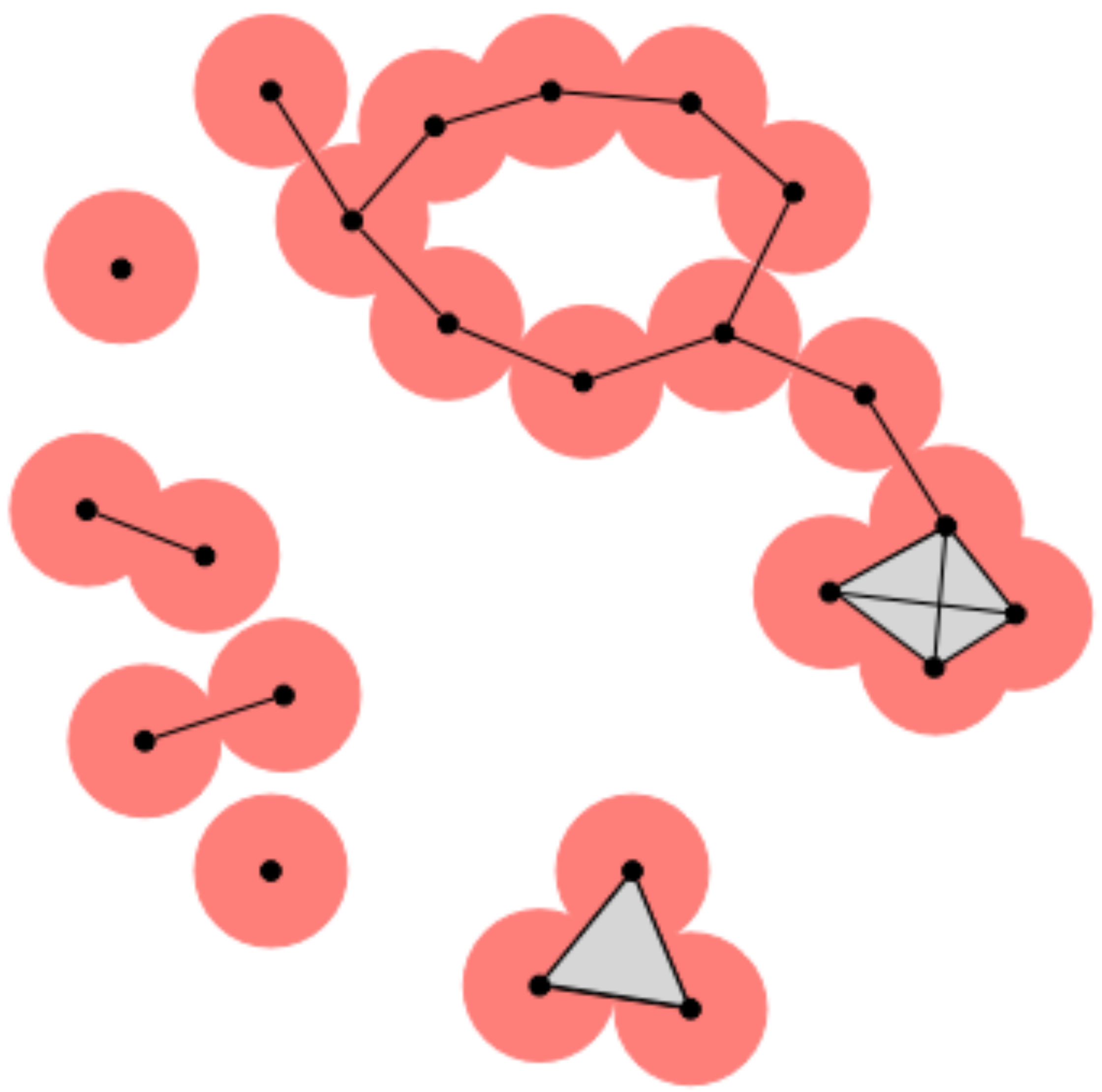




Visual Analytics
of high-dim data

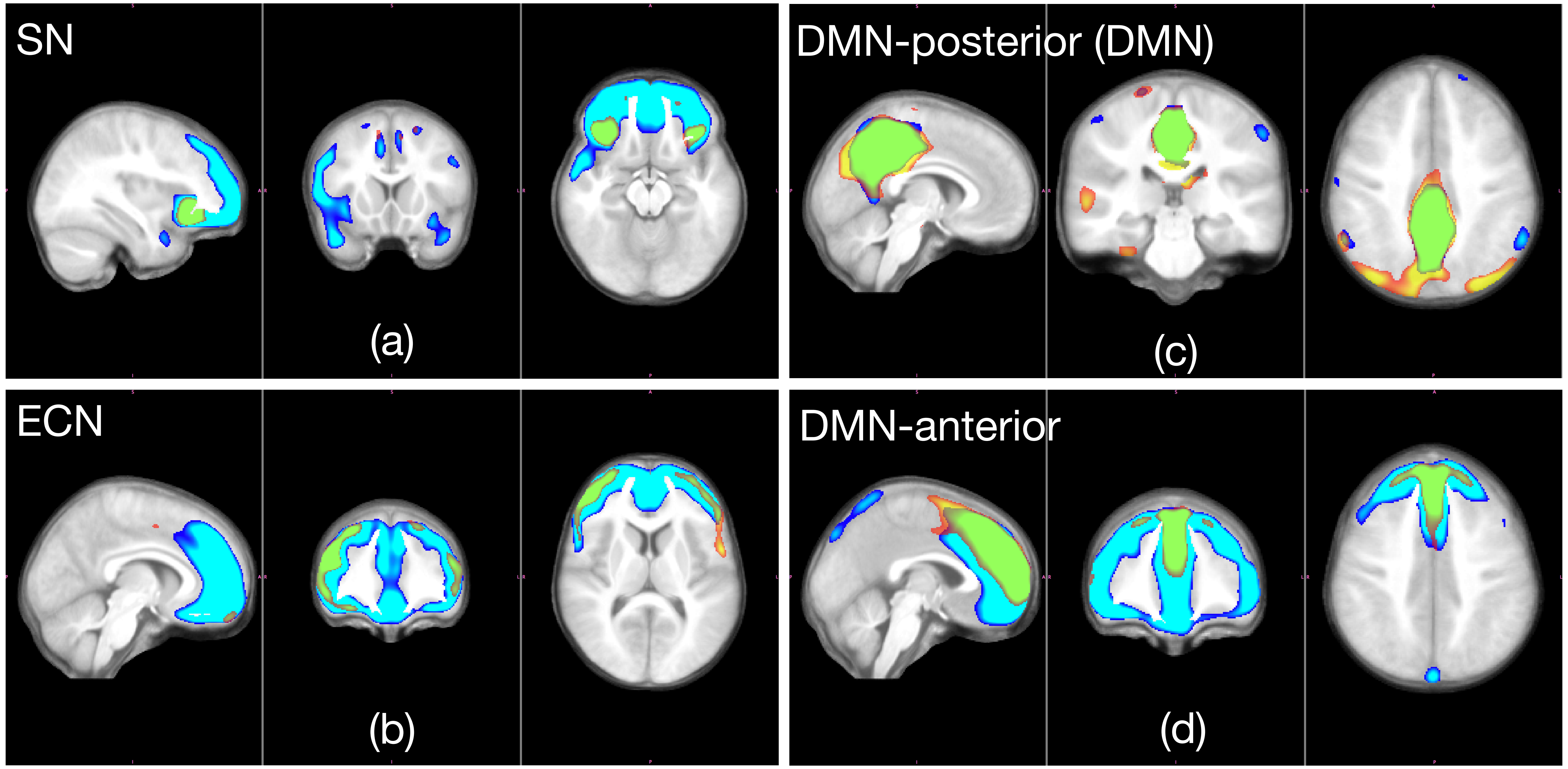
[LiuWangThiagarajan2014]





Visualizing
brain networks & social networks

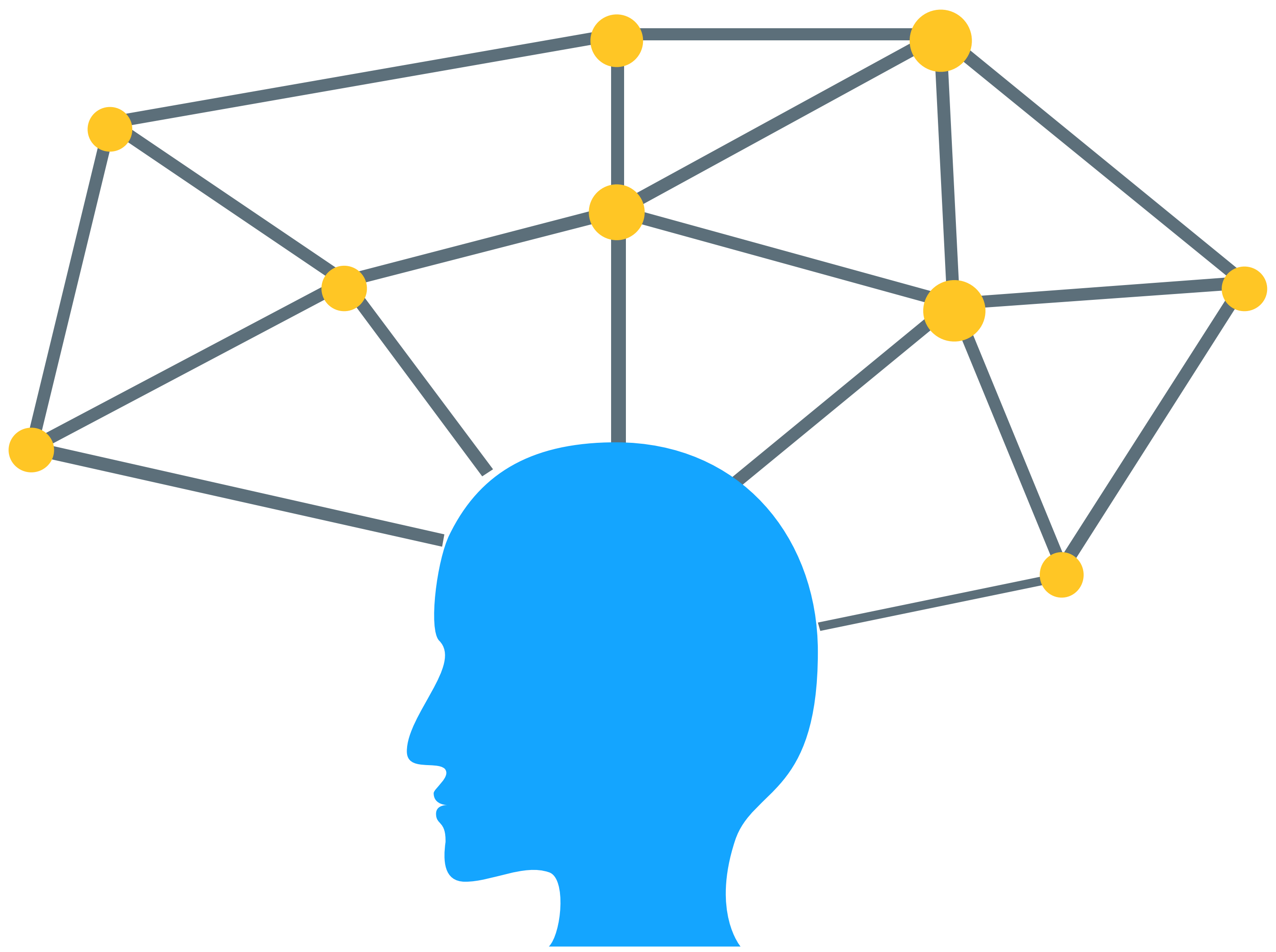
[WongPlandeWang2016]



[PalandeJoseZielinski2017]

Topology Inspired

machine learning & statistics



Visualization

is the secret weapon for

Machine learning

Course Objective

- Enable the students to become familiar with new and innovative techniques that combine data analysis with data visualization, from algorithmic and implementation perspectives.

Hot Market, Cool Startups

- AtScale: multidimensional analysis, calculation engine to run against any BI visualization tool, prediction-defined aggregates.
- Noodle.ai: maths, algorithms that learn, glass box (not black box) algorithms that allow executives to understand risks (probabilities) and causality, and make informed decisions.
- Periscope: helps data scientists quickly build customized, highly detailed visualizations of their data.
- Ayasdi: analyze and build predictive models using big data or high-dimensional data sets; hypothesis-free, automated analytics at scale; topological data analysis.
- Gluent: data virtualization technology that makes possible what it calls "hybrid data" computing.

Source: http://www.crn.com/slide-shows/applications-os/300088266/the-10-coolest-big-data-startups-of-2017-so-far.htm

New, Cutting-Edge & Emerging

- Visualization research venues: recent publications, conferences
- Emerging research topics
- Known and recent techniques employed by industry

Goal

Successful completion of the course will enable the students to:

- Obtain a deeper understanding of visualization as a powerful tool for data analysis, in particular, machine learning [User]
- Apply emerging and innovative techniques to data in various application domains [Expert User]
- Pursue new research directions in data analysis and data visualization [Developer, Researcher]

Prerequisite

- Students are expected to have basic knowledge of data structures and algorithmic techniques, bachelor-level knowledge in mathematics or computer science, and working knowledge of programming, ideally with Python and/or C++.

- Targeted audience: PhD students, master students and very-motivated upper level undergraduate students.

- The students are not required to be majoring in CS, but it is preferable that the students have some background in algorithms and/or other data science related courses, and have working knowledge of programming, ideally with Python and/or C++.

Assignments and Grading

- 4 assignments in the form of mini projects (60 points, 60%)
- Final project (40 points, 40%)
  - Final project proposal (10 points, 10%)
  - Final project report (25 points, 20%; 5 points are for progress report; 20 points for final report)
  - Final project presentation (5 points, 5%)
- Additional 10 bonus points may be available in the form of bonus assignment questions.
- Grading:
  - A 100-93 A- 93-90
  - B+ 90-87 B 87-83 B- 83-80
  - C+ 80-77 C 77-73 C- 73-70
  - D+ 70-67 D 67-63 D- 63-60
  - E 60-0

Course Communications

Website:

- http://www.sci.utah.edu/~beiwang/teaching/cs6965-spring-2018.html
- Primary source for course information, schedule, etc.

- Canvas:
  - Communication from instructor via course announcement
  - Secondary source for course information
  - Homework submission portal
  - Check to make sure you receive class announcement **daily**

- Email: beiwang@sci.utah.edu for questions on the course

Study large and complex data

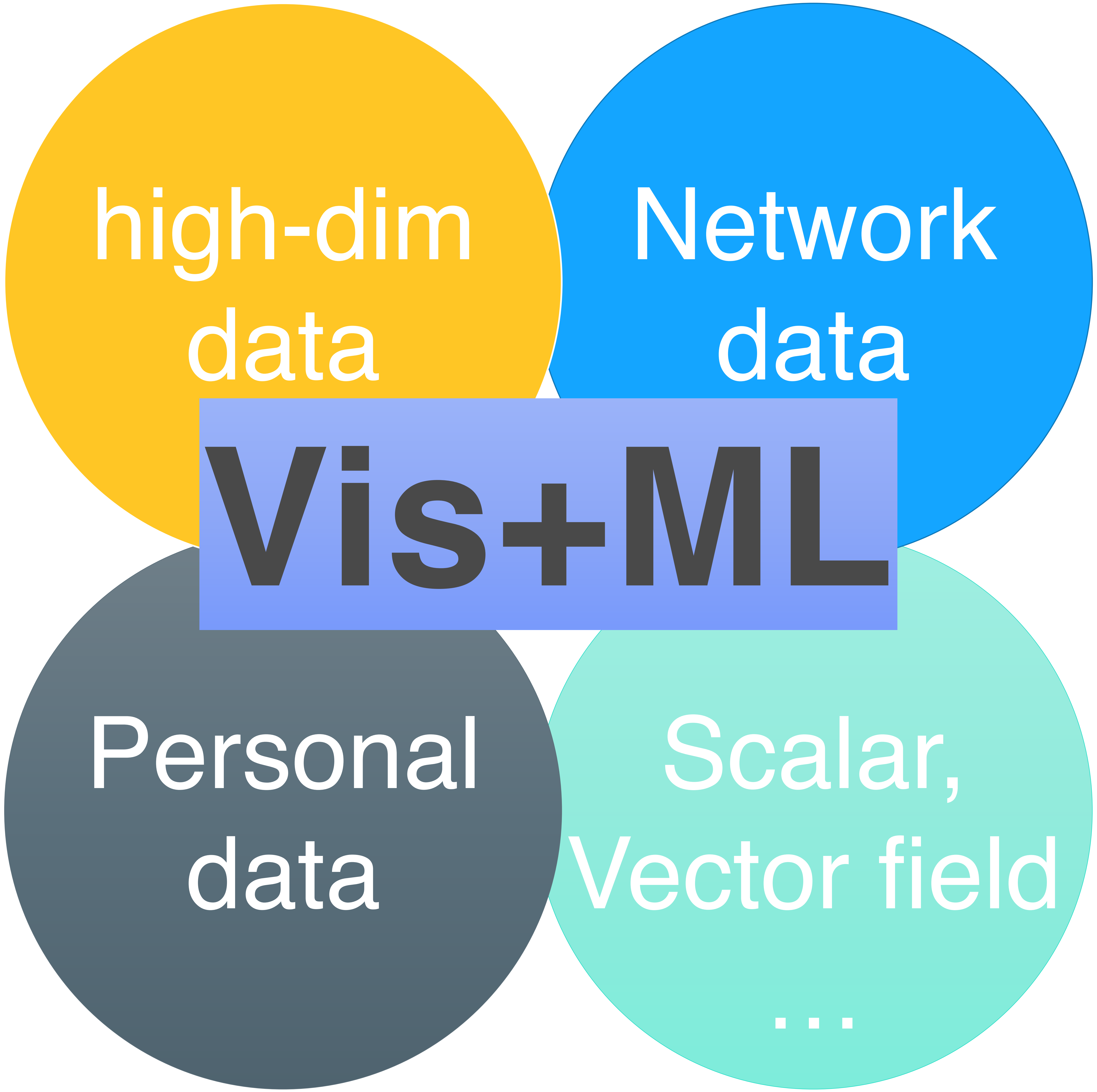
high-dim data

Network data

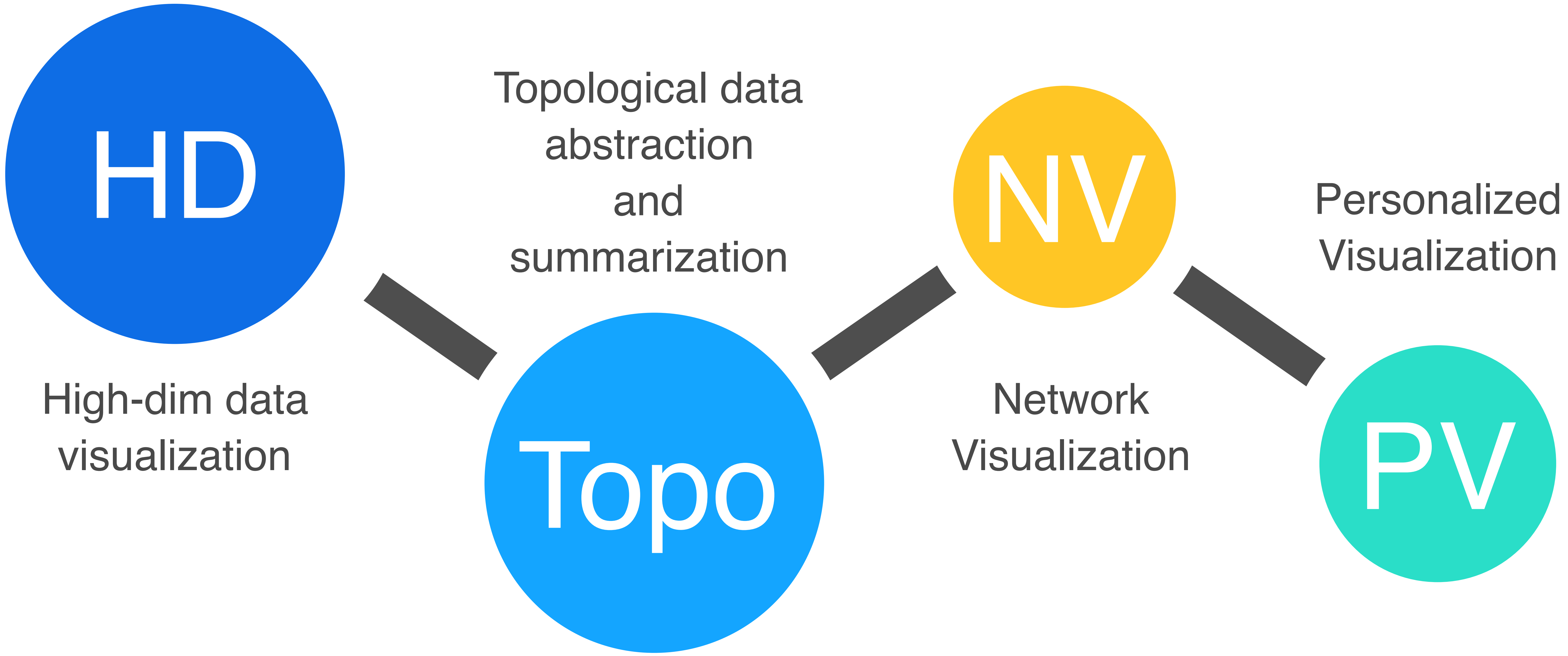
Personal data

Scalar, Vector field

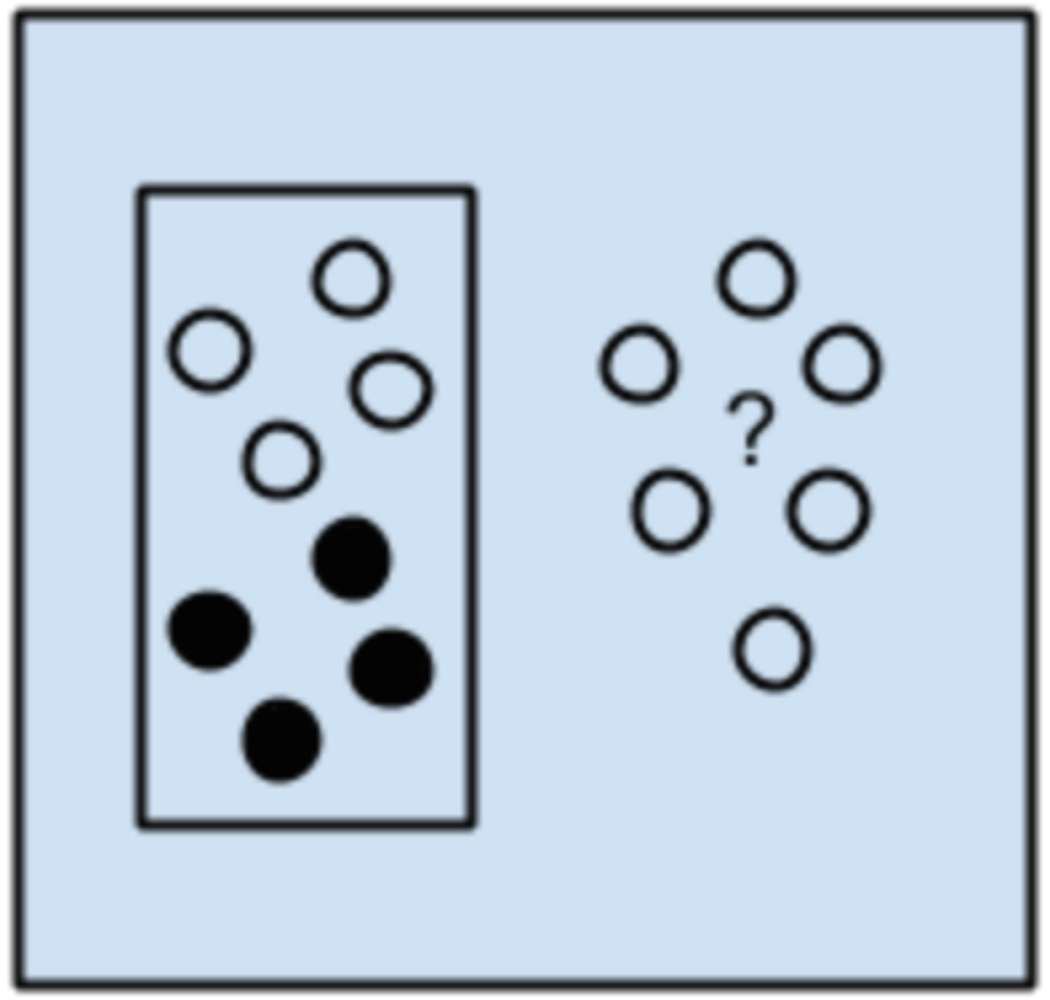
...


high-dim data
Network data
Vis+ML
Personal data
Scalar, Vector field
...

Mutually Inclusive Modules

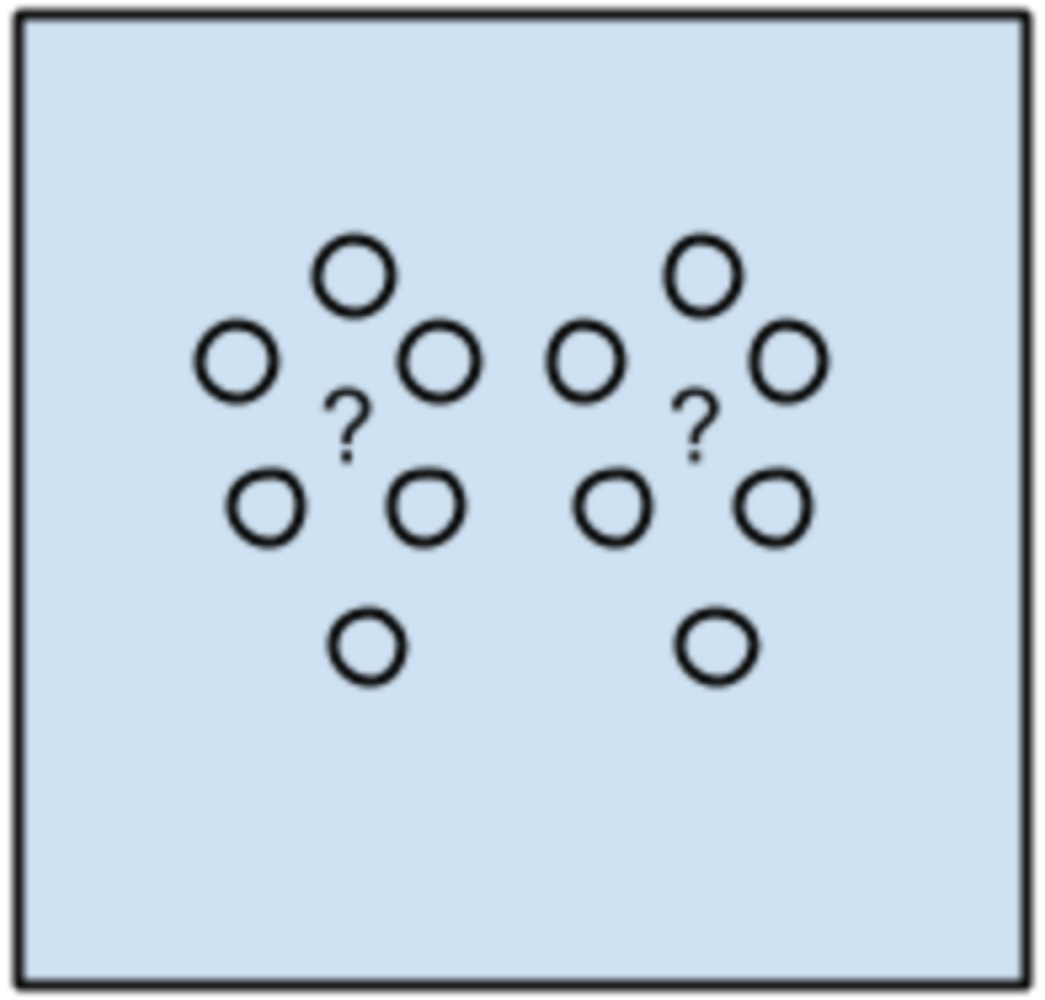


Machine Learning At a Glance



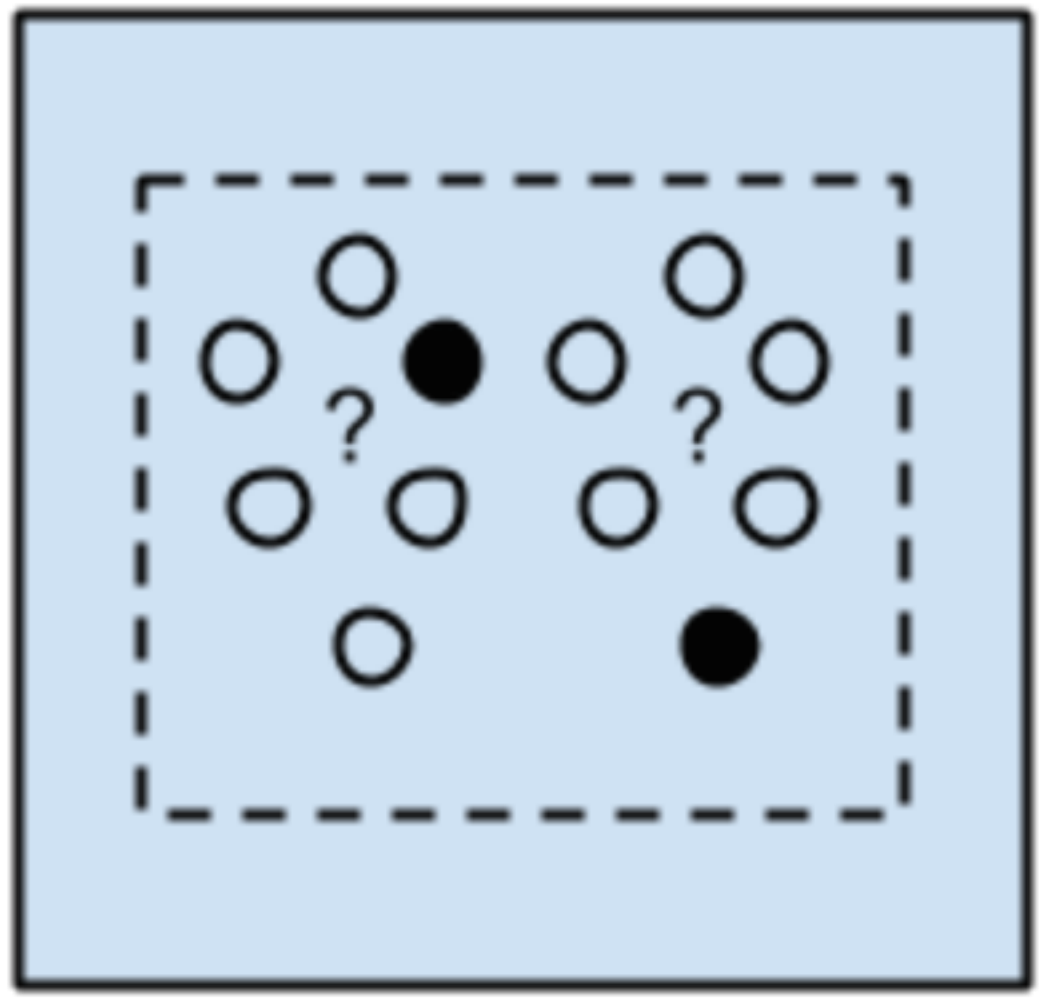
Supervised Learning

Problems: Classification
Regression
Algorithms: Logistic Regression
Back Propagation Neural Network



Unsupervised Learning

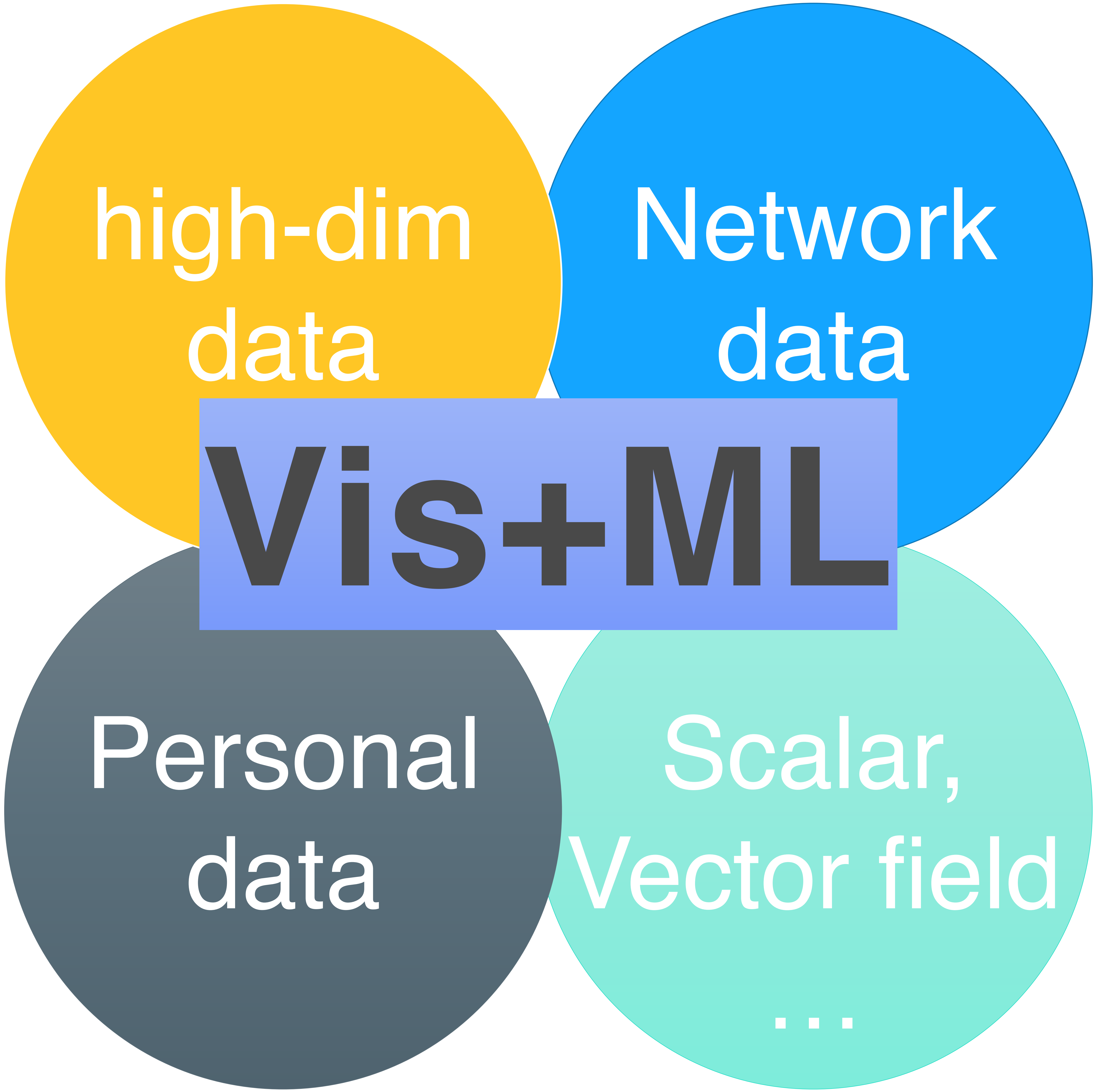
Problems: Clustering
Dimensionality Reduction
Algorithms: k-means, Data Mining,
Topological Data Analysis



Semi-supervised Learning

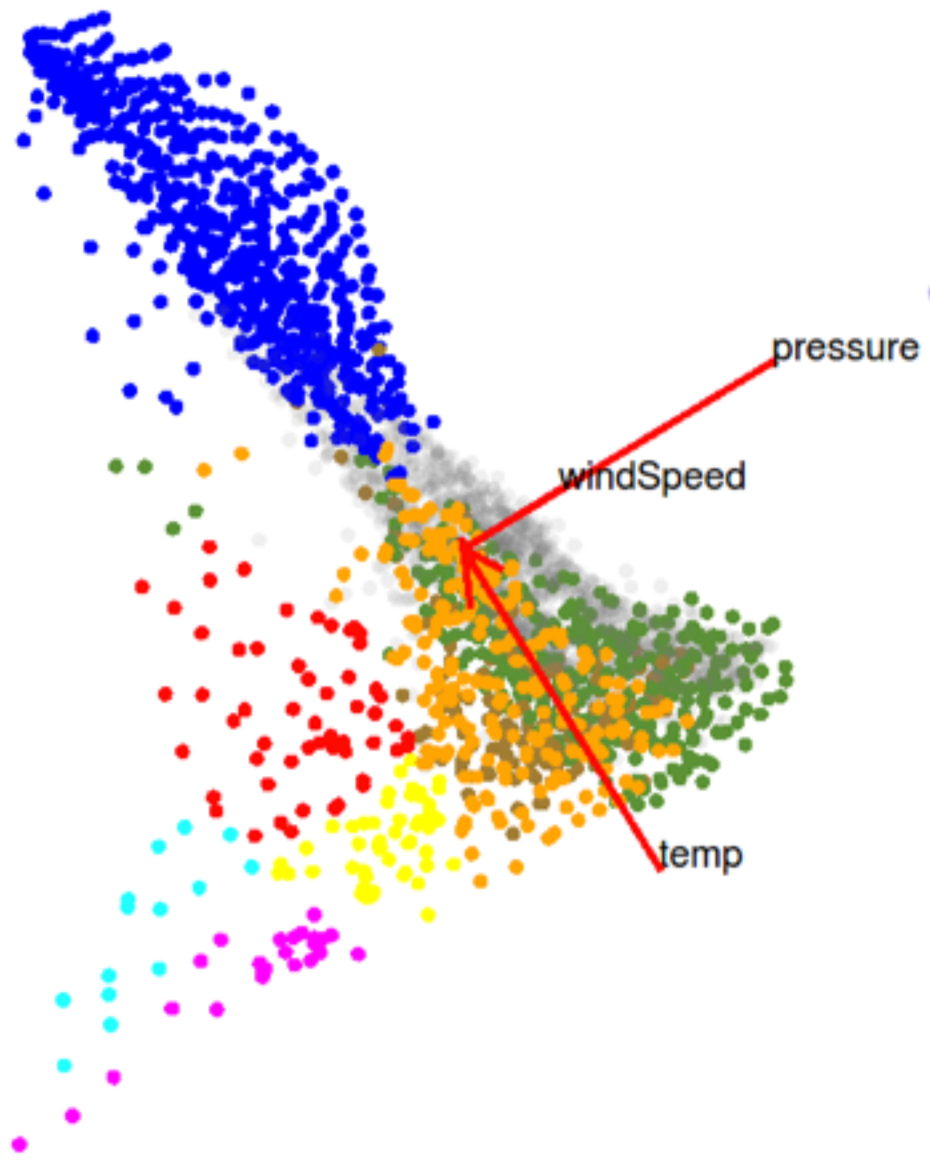
Problems: Classification
Regression
Algorithms: extensions to flexible
algorithms, model unlabelled data

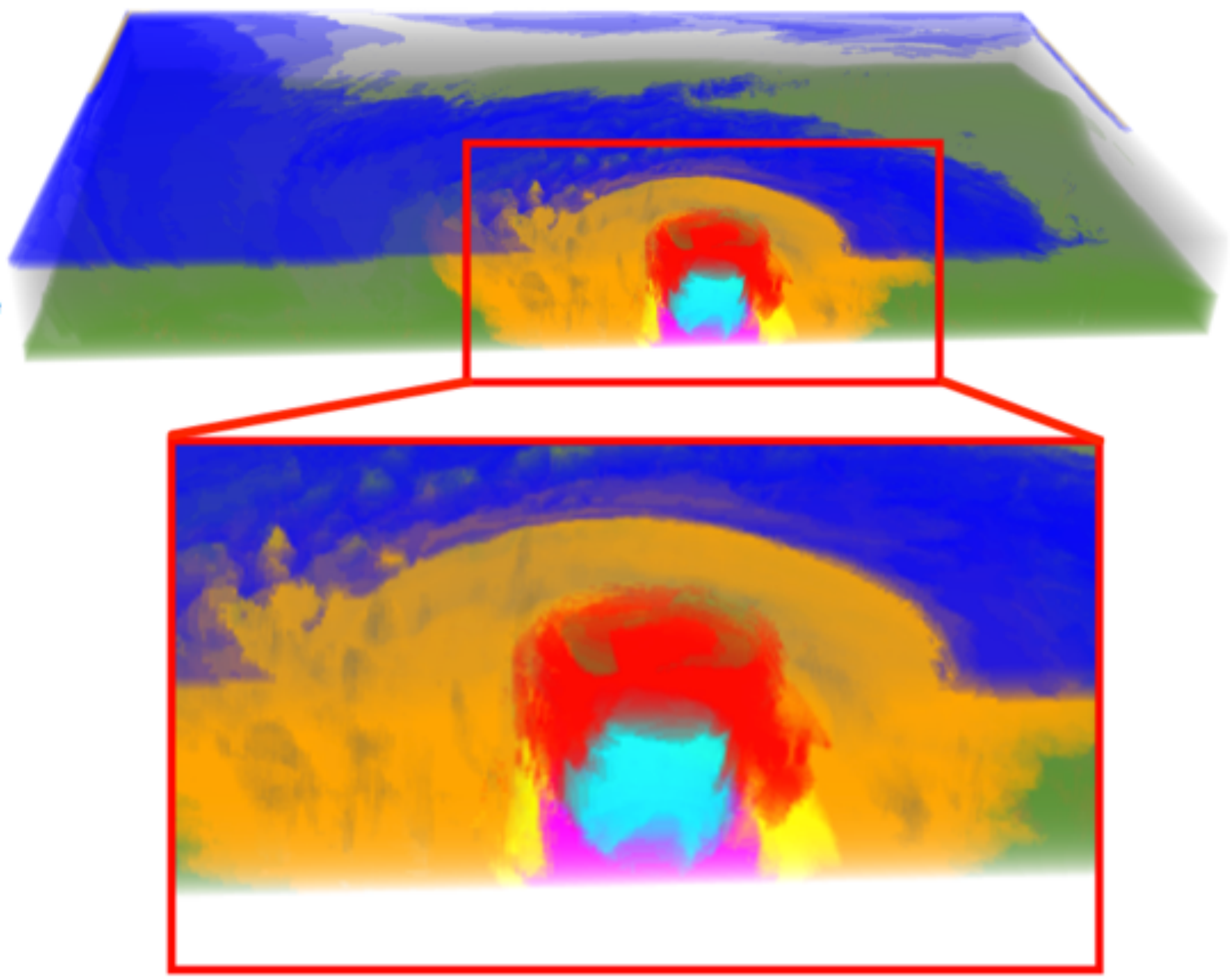
Source: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/


high-dim data
Network data
Vis+ML
Personal data
Scalar, Vector field
...

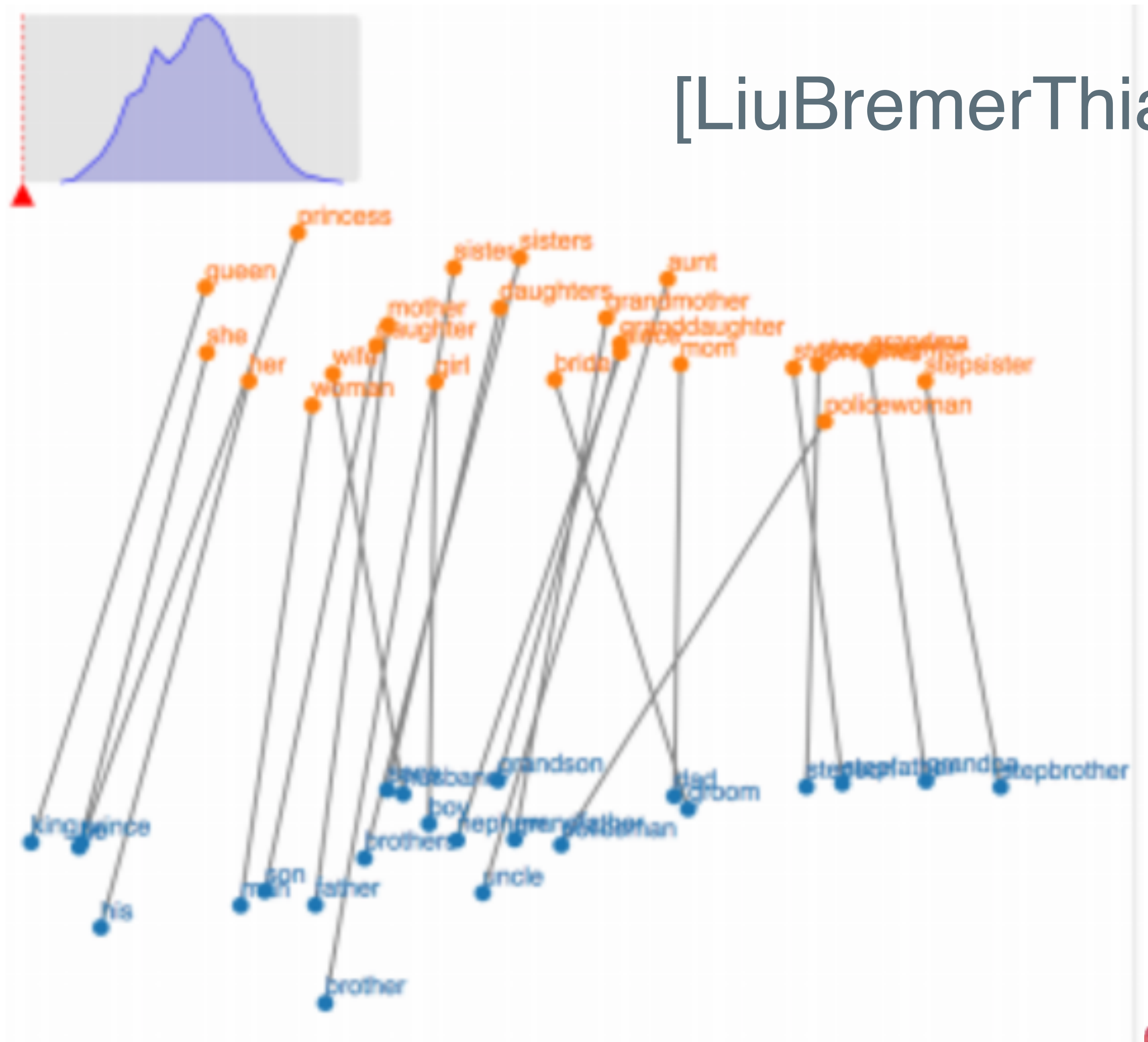
1:HD
High-dim VIS

Obtain insight from high-dimensional data through ML and interactive VIS

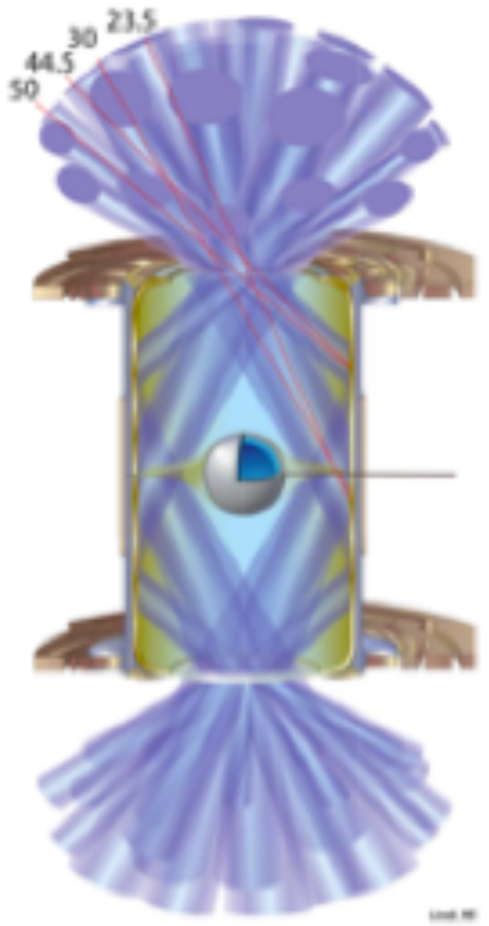


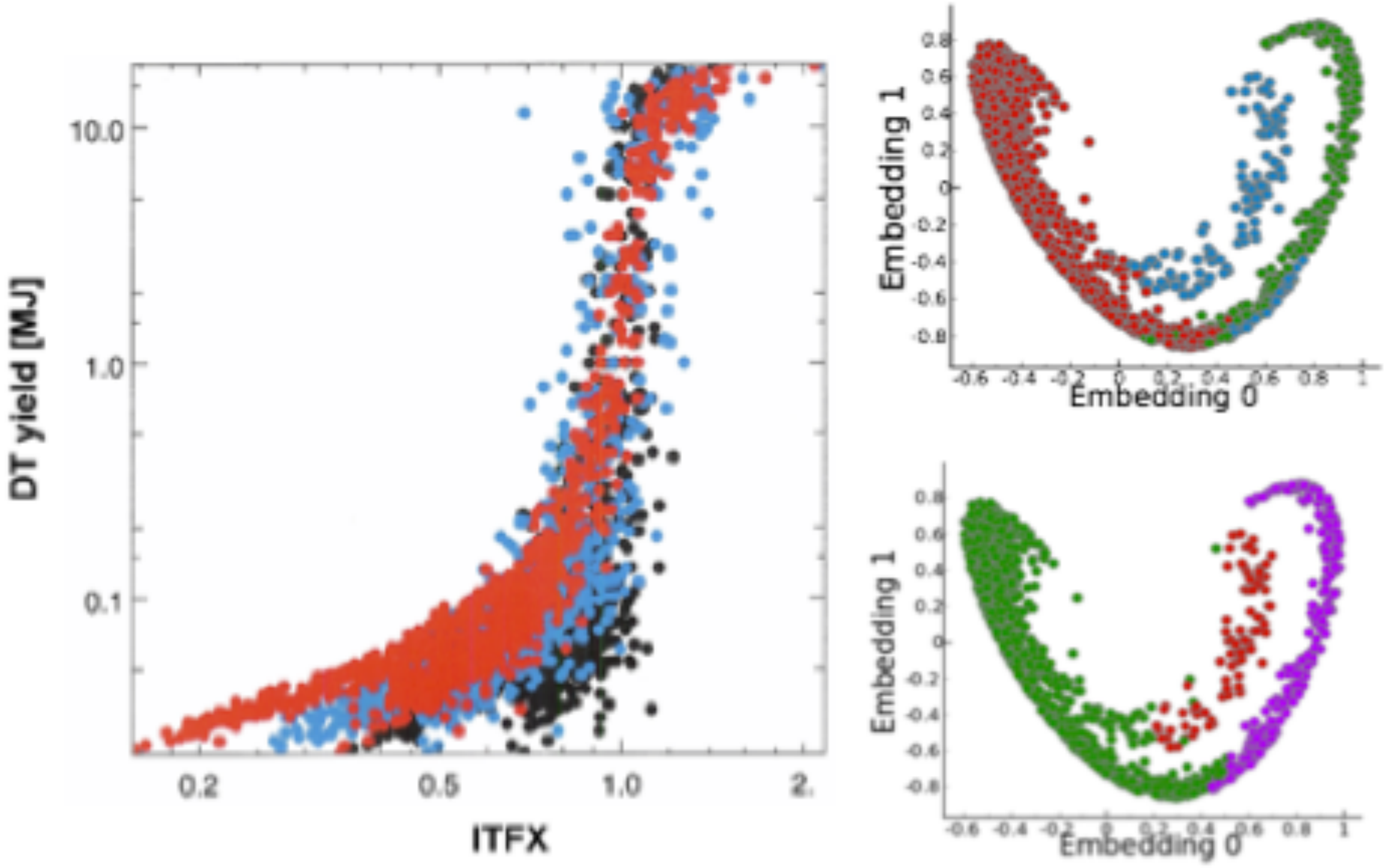


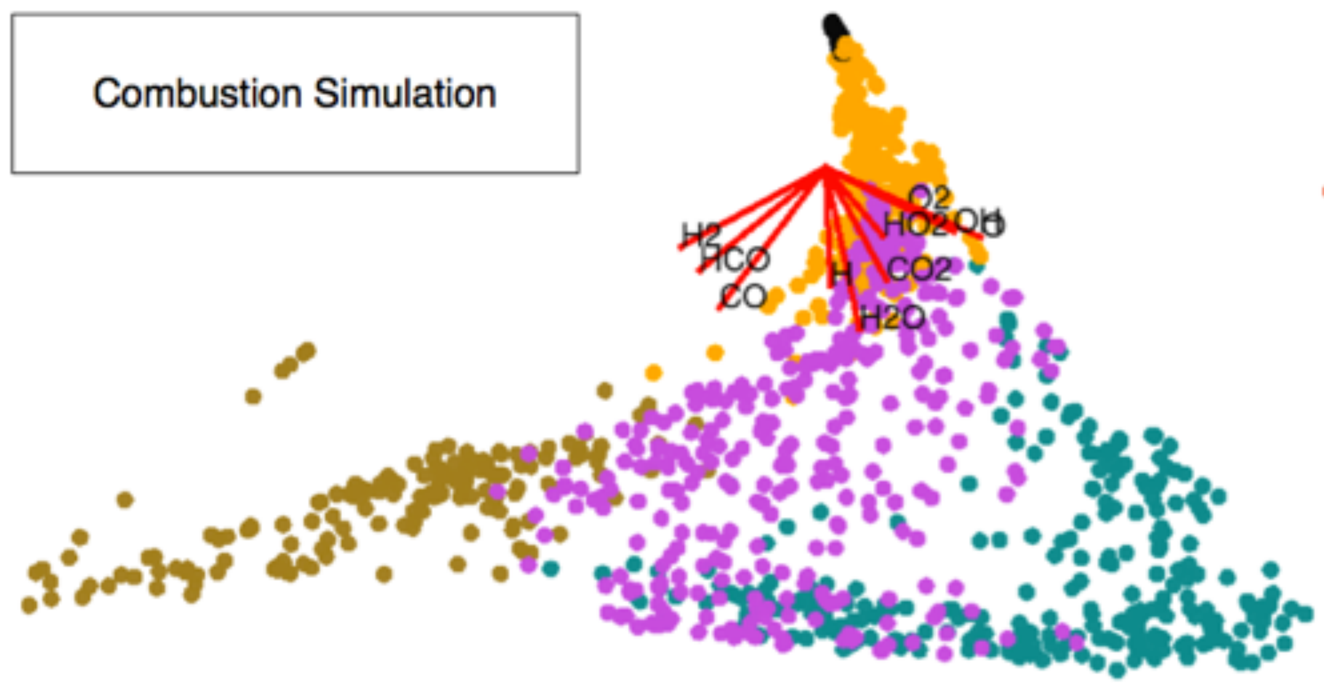
[LiuBremerThiagarajan2017]

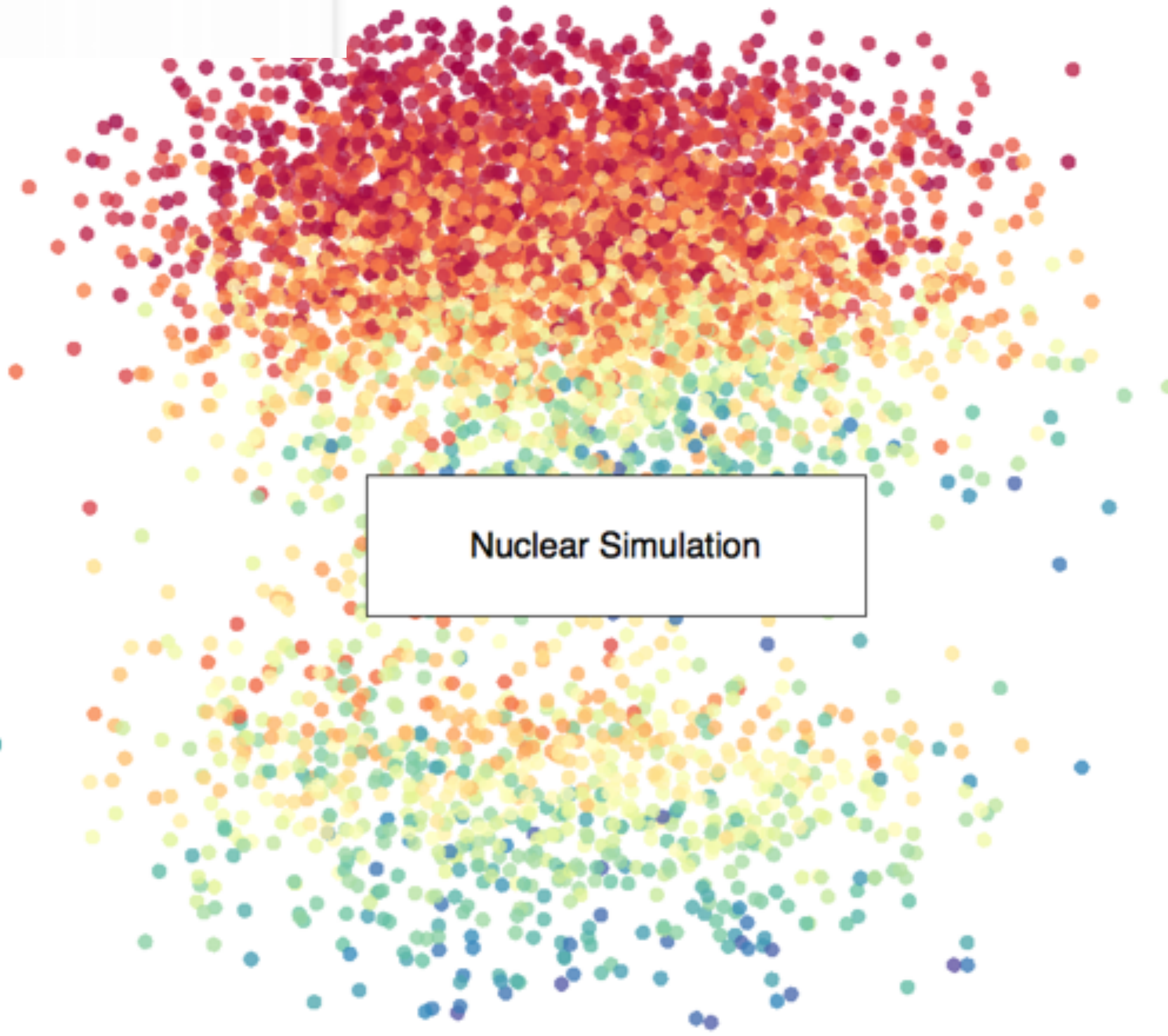


[LiuWangThiagarajan2014]



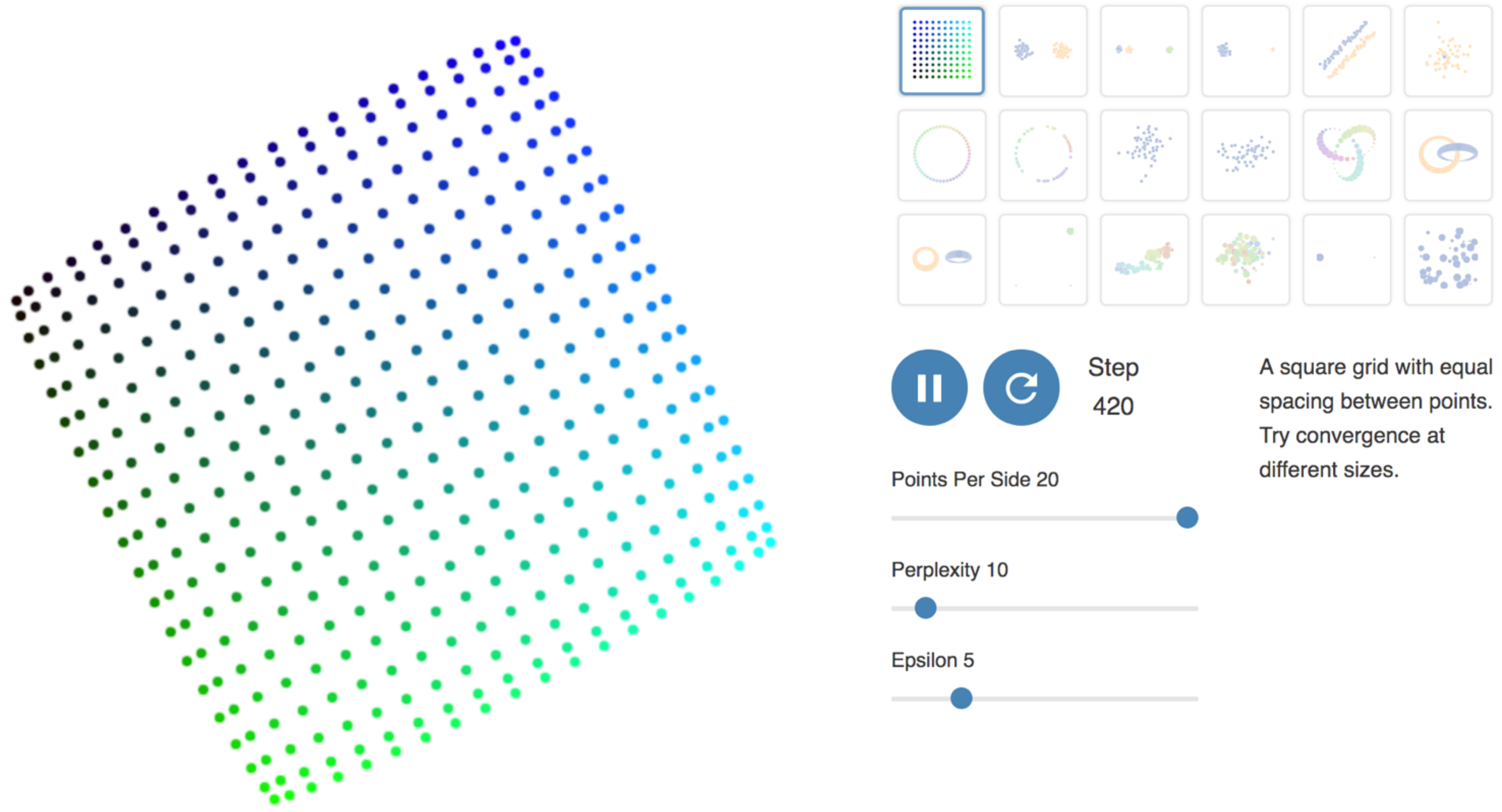






[BremerMaljovecSaha2015]

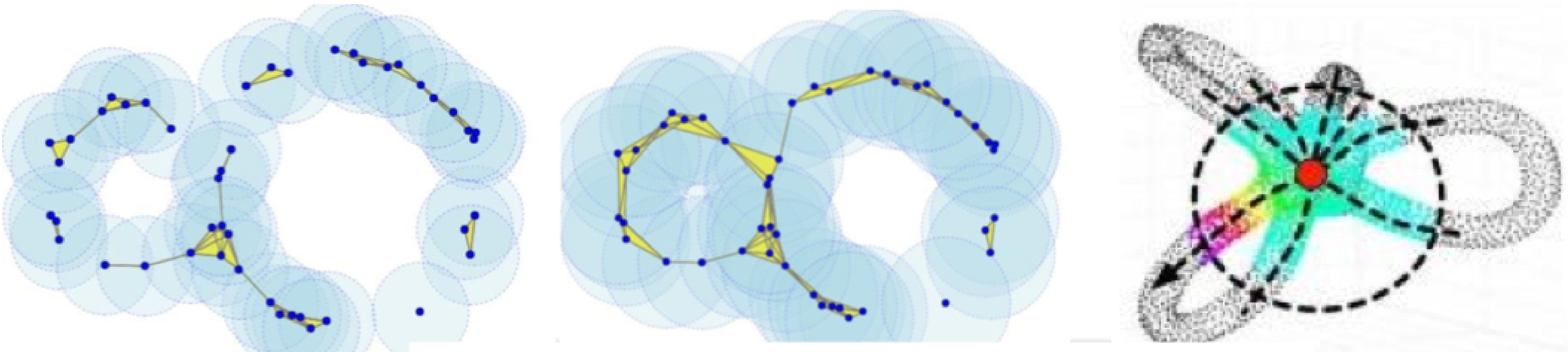
[LiuWangBremer2014]



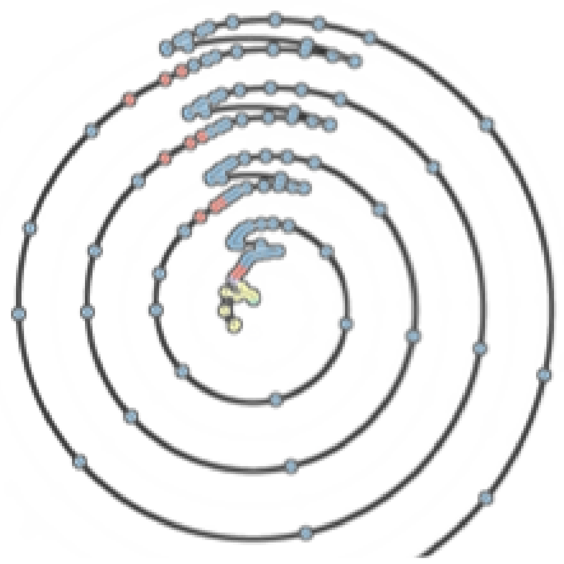
Source: https://distill.pub/2016/misread-tsne/

2:TOPO

Topological abstraction & summarization

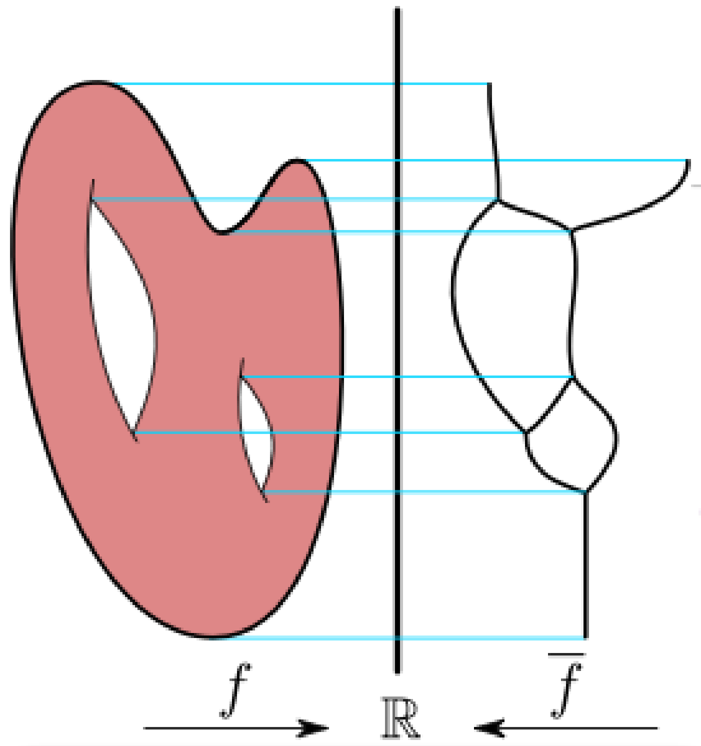


Persistent Homology, Cohomology, Local Homology

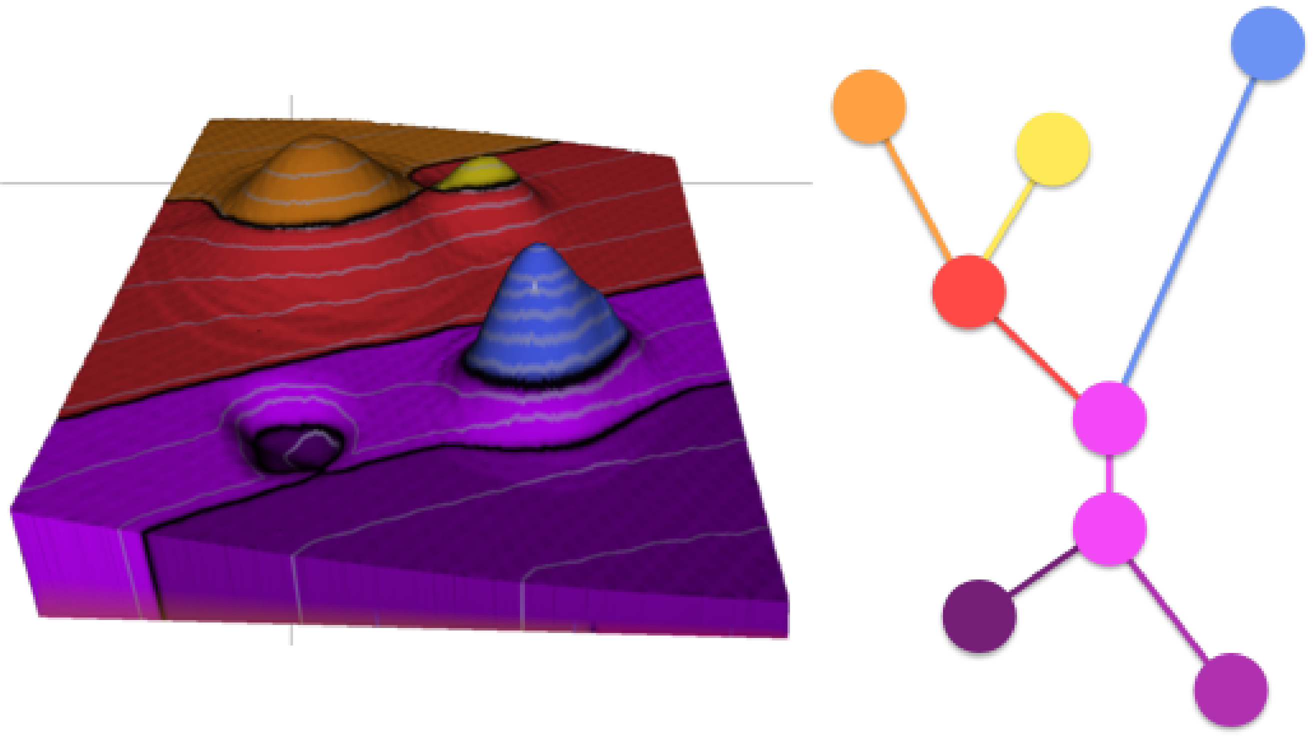


Cyclic Structures

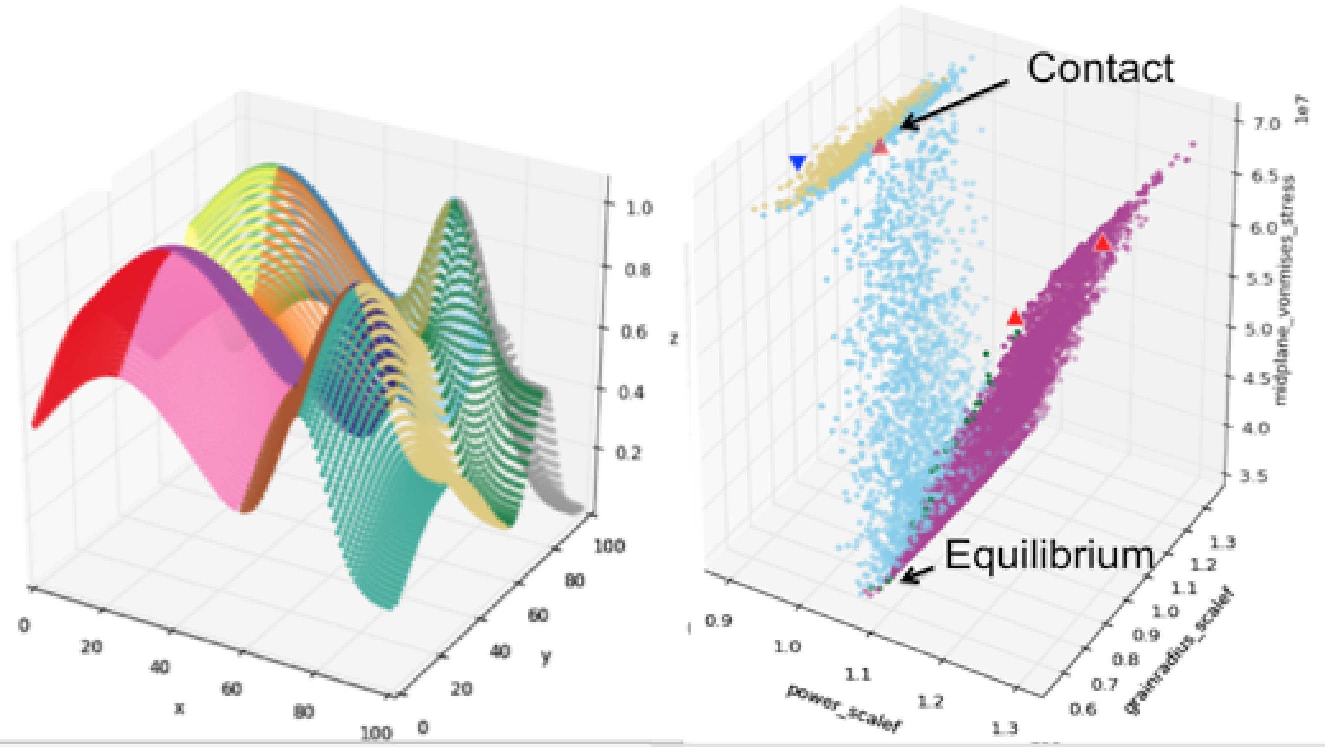
Topological data analysis and visualization capture the shape of complex data



Reeb Graph



Contour and Contour Trees



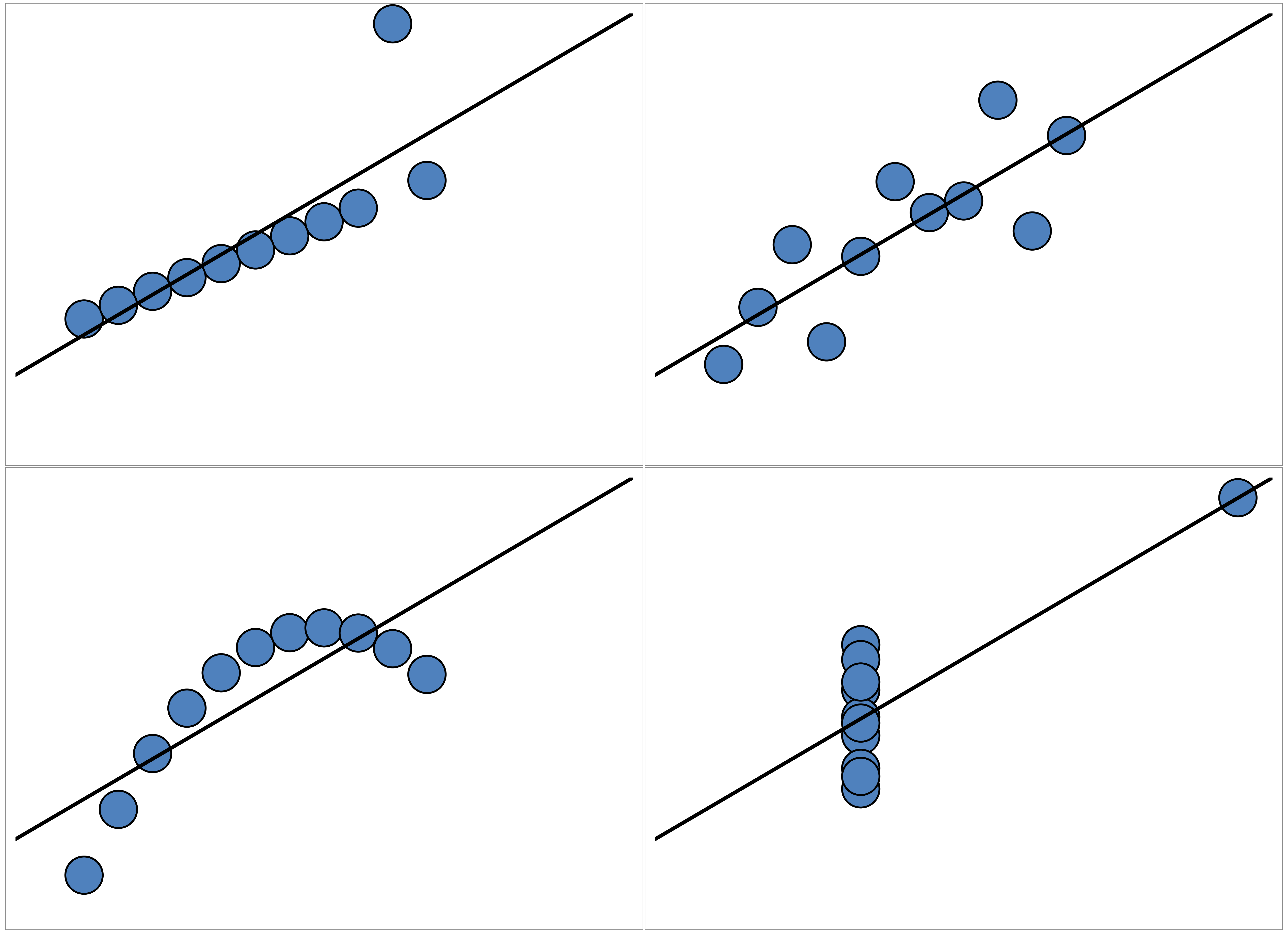
Morse-Smale Complexes

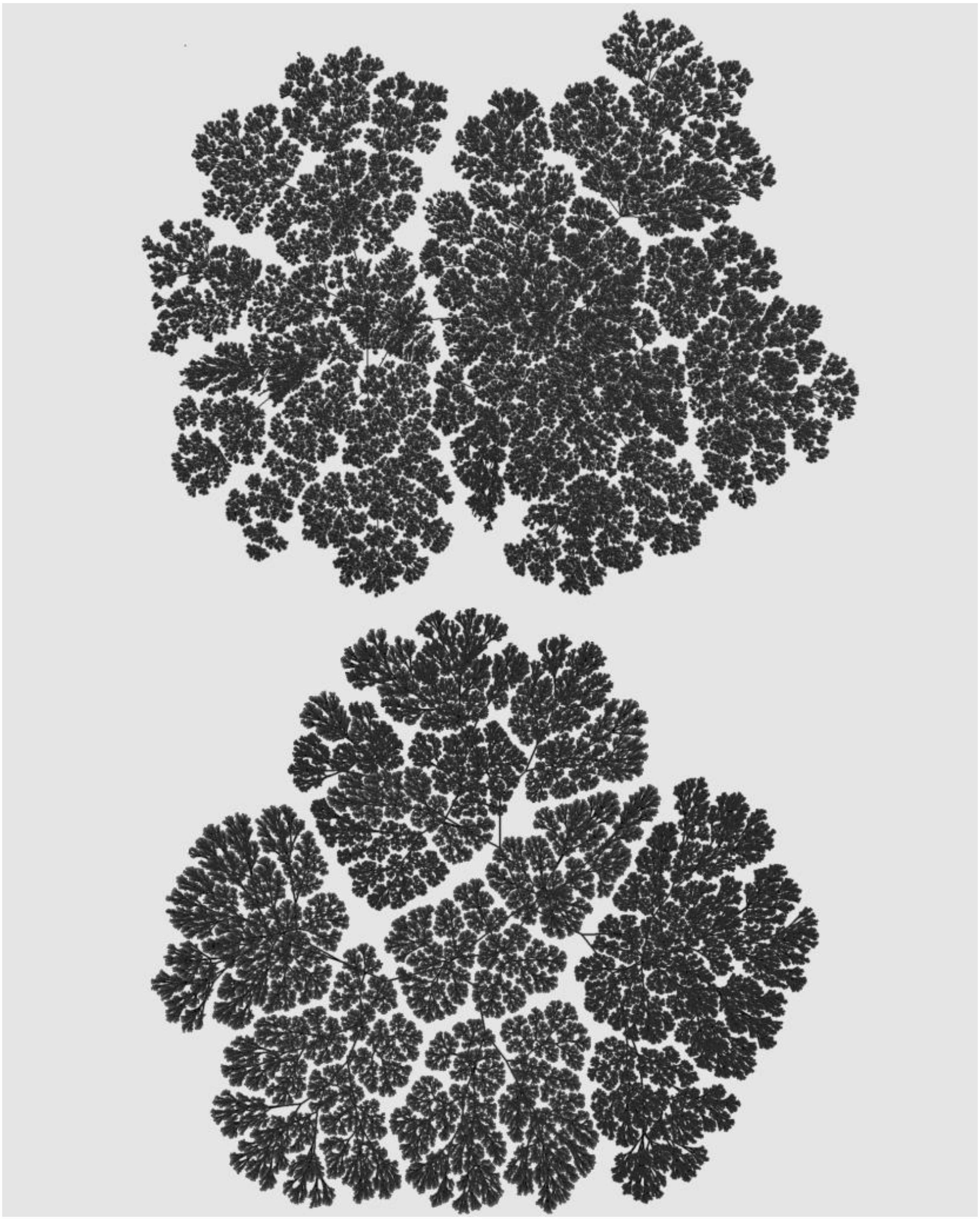
Scalar & vector field data

3:NV

Network Vis

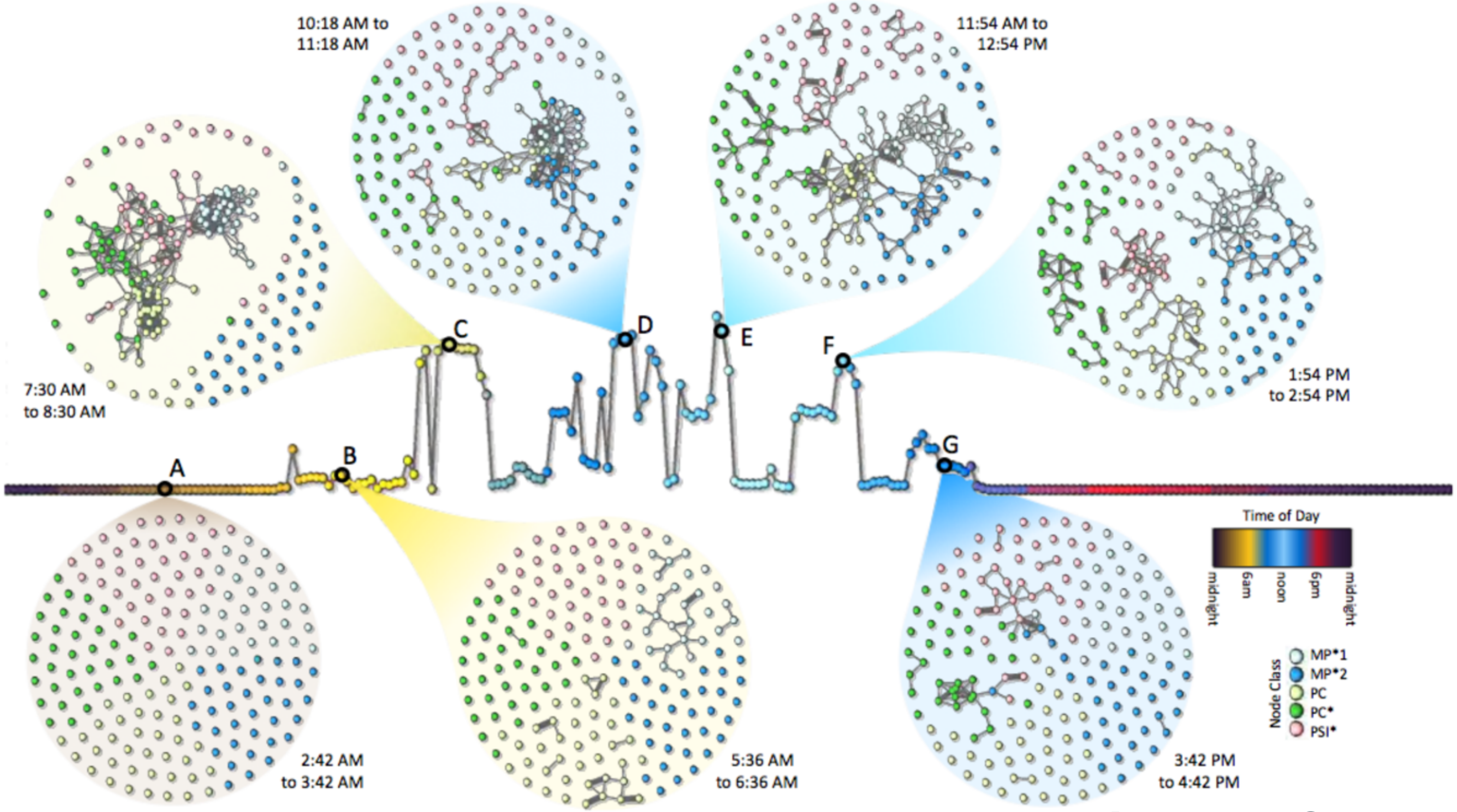
A picture is worth a 1000 words, but





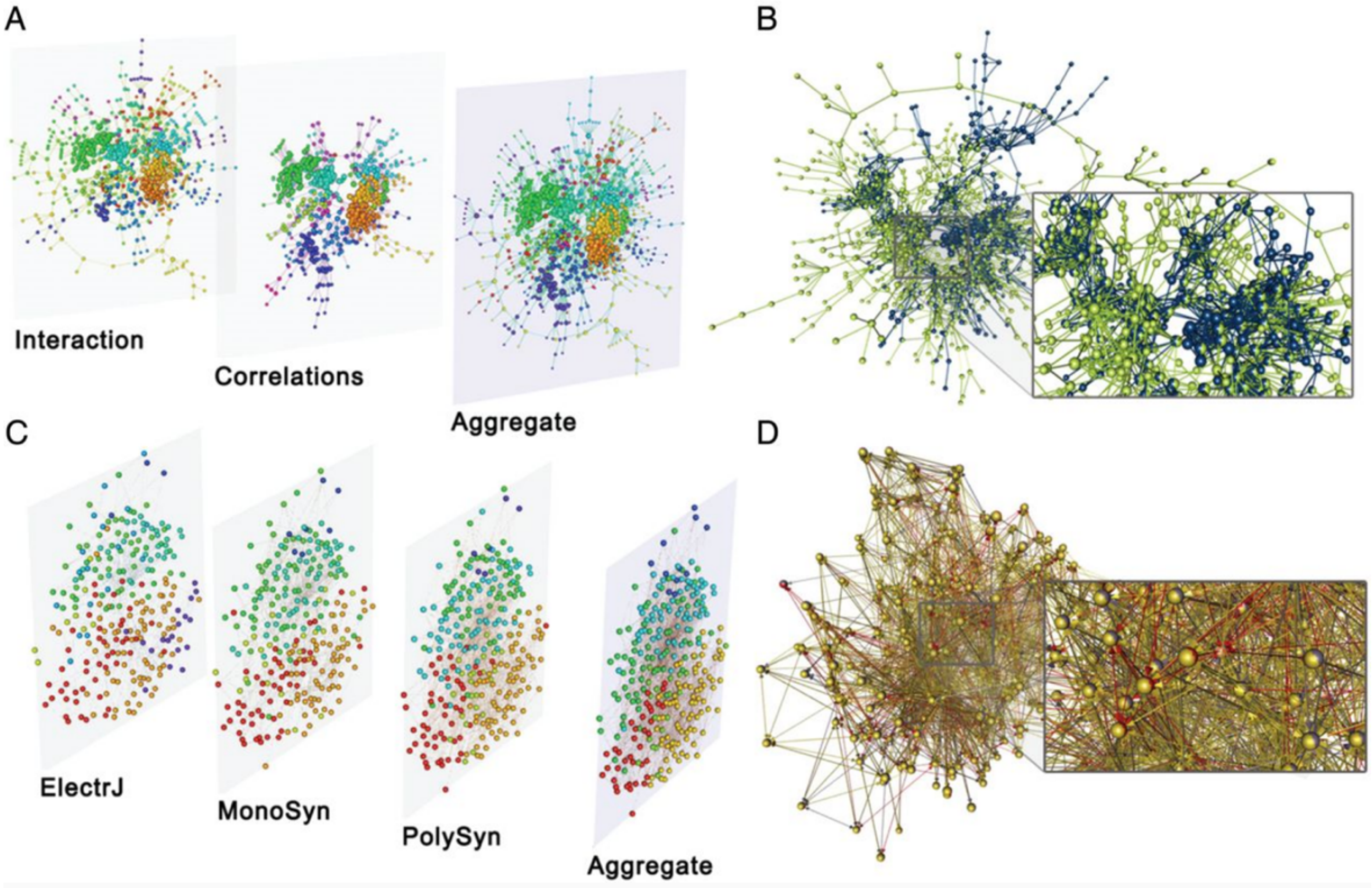
Source: Carlos Scheidegger

Static vs time-varying networks



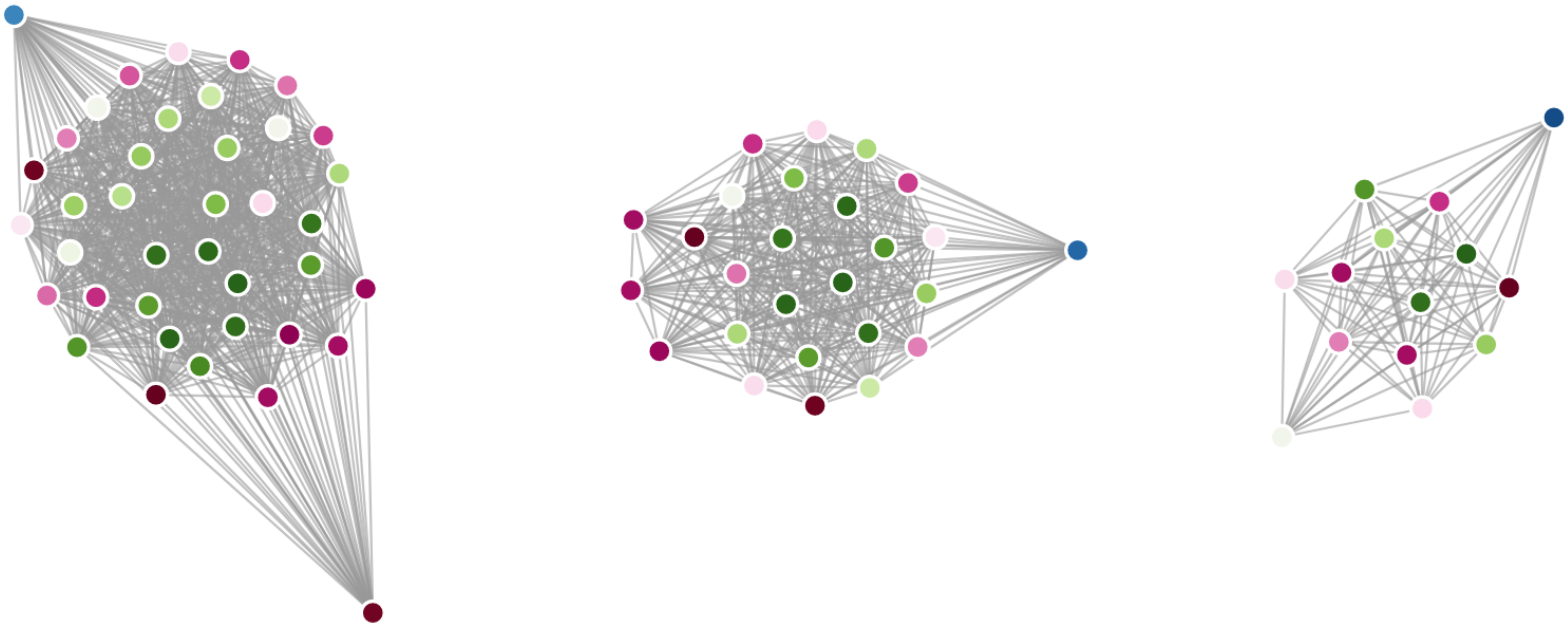
[HajijWangScheideggerRosen2018]

Multilayer & multivariate networks



[DomenicoPorterArenas2015]

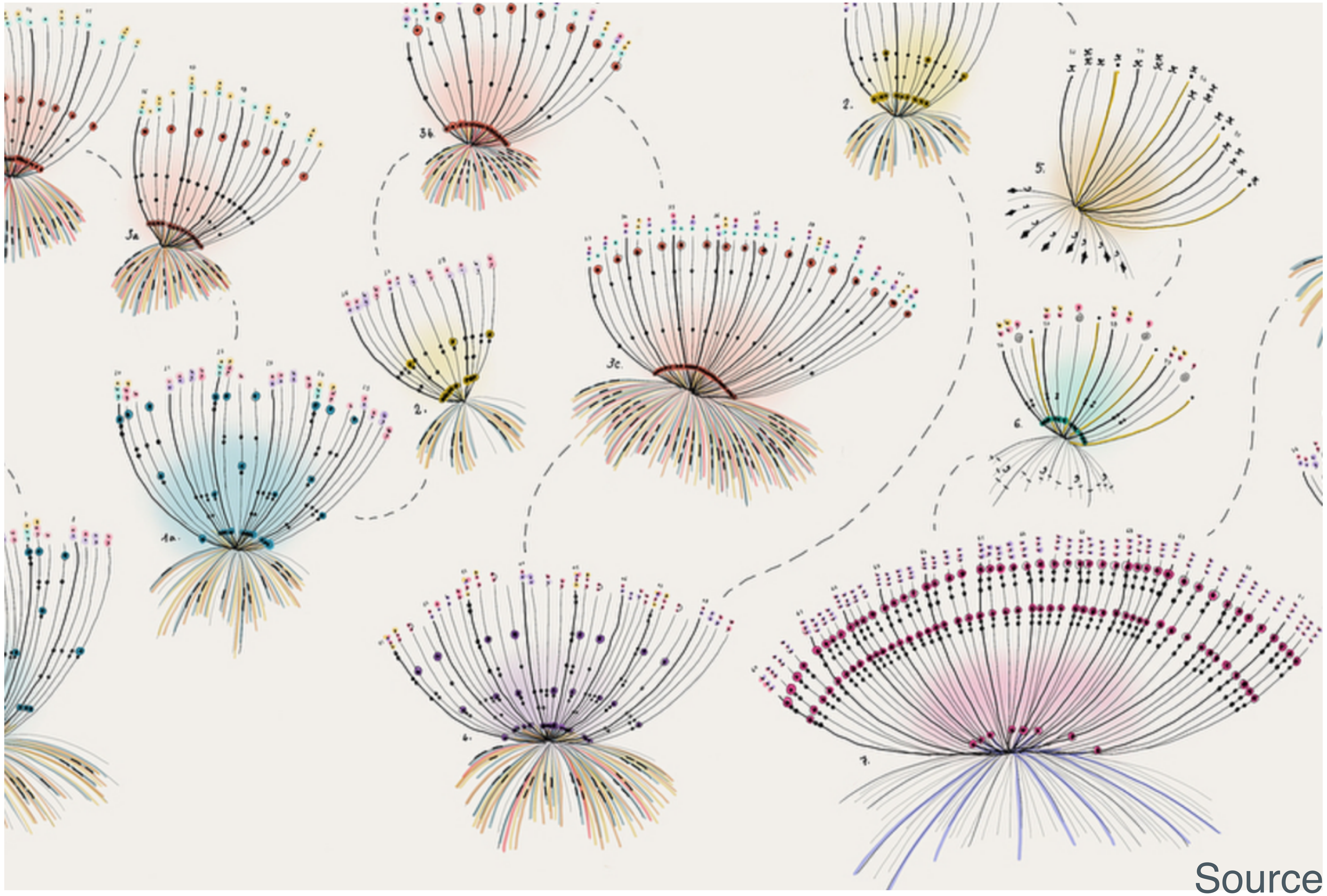
Scalability



Source: Bei Wang and Sravankuma Neerati

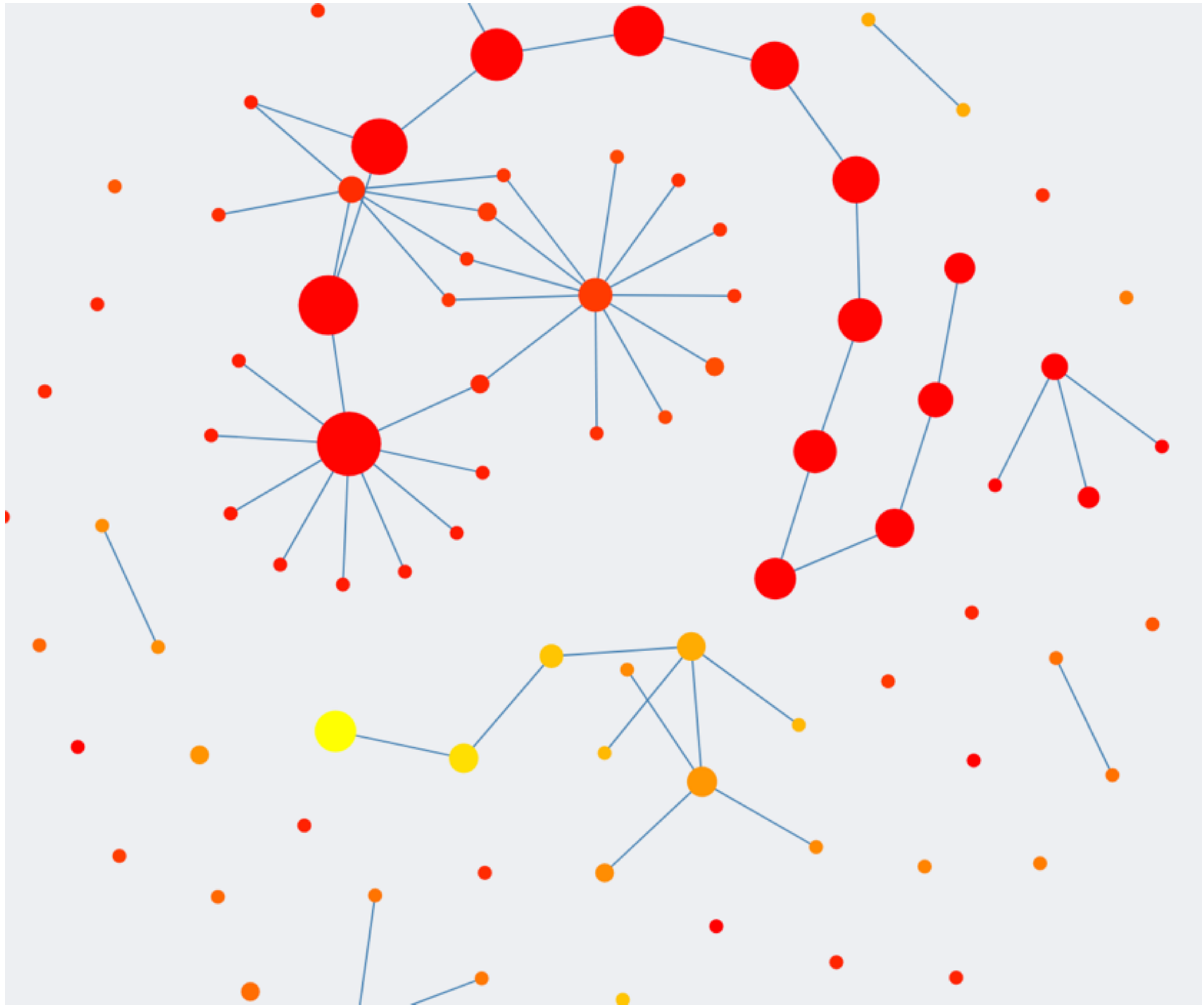
4:PV
Personalized Vis

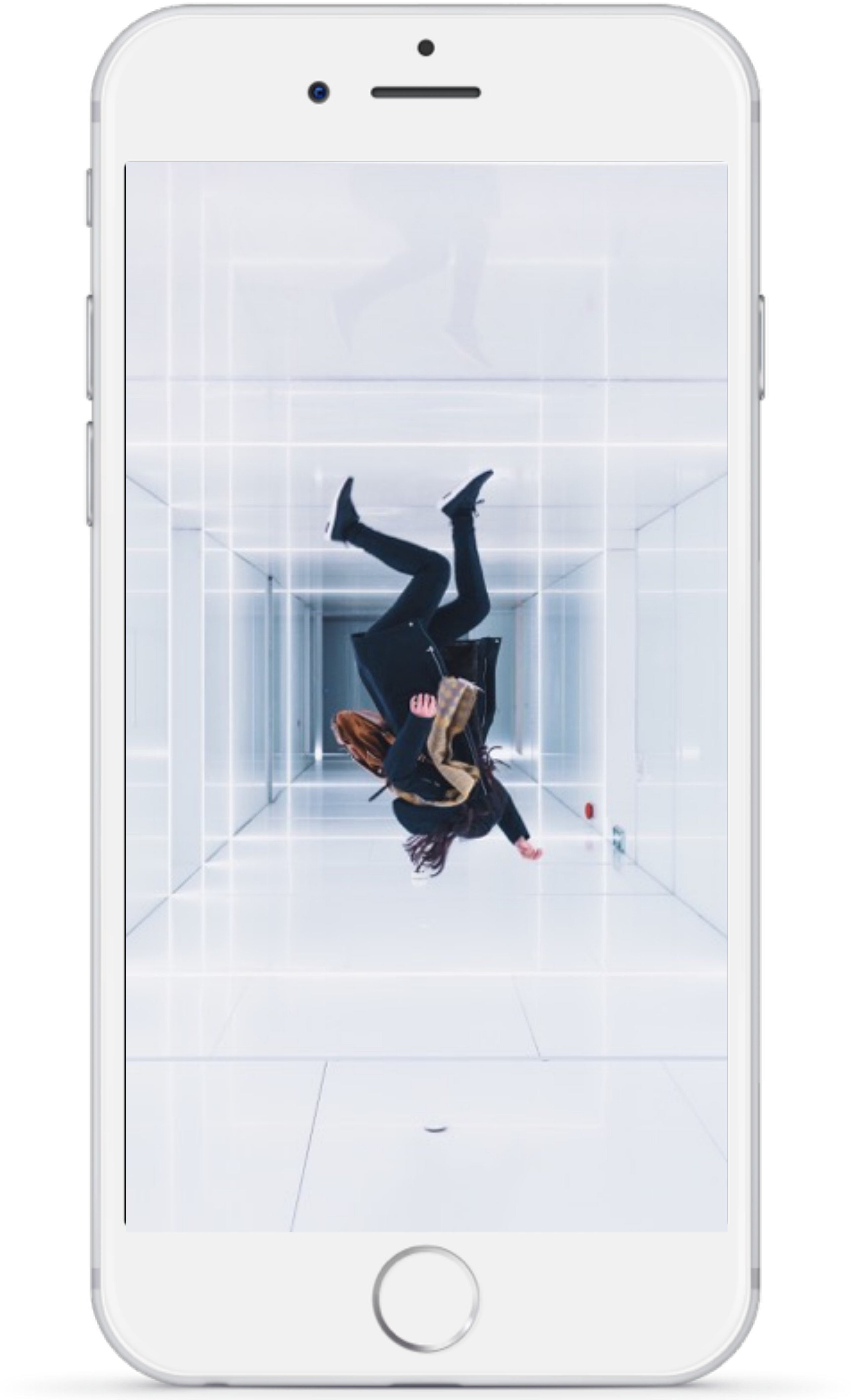
Visualizing for individuals



Source: Giorgia Lupi, Accurat

Visualizing personal data





Visualization on mobile devices

Class Syllabus and Final Project

- Final project key dates:
  - Project team creation: due February 8.
  - Project proposal description: due March 6.
  - Project progress report: due March 27.
  - Project final report: due April 24.
  - Project presentations: on April 24 (9:10 - 10:30 a.m.) and April 27th (8:00 - 10:00 a.m.)
- **http://www.sci.utah.edu/~beiwang/teaching/cs6965-spring-2018/syllabus-spring-2018.pdf**
- **http://www.sci.utah.edu/~beiwang/teaching/cs6965-spring-2018/schedule.html**

How to succeed in class

- Attend lectures
- Start thinking about final project early
- Ask questions in class
- Getting help: office hour, Tuesday 10:30 to 11:30 a.m. or by appointment, by email with title “CS 6965”
- **Learning programming along the way: D3.js, TTK, Python, etc.**

Mandatory readings

- Scikit-learn tutorial:
  - http://scikit-learn.org/stable/tutorial/basic/tutorial.html

Getting ready for mini-project 1
Python, D3.js, etc.

- Install and read the documentation of kepler-mapper:
  - https://github.com/MLWave/kepler-mapper
- Interactive Data Visualization for the Web, 2nd Ed.
  - http://alignedleft.com/work/d3-book-2e

Slide Deck References

- [LiuWangThiagarajan2015]: Visual Exploration of High-Dimensional Data through Subspace Analysis and Dynamic Projections. Shusen Liu, Bei Wang, Jayaraman J. Thiagarajan, Peer-Timo Bremer and Valerio Pascucci. Computer Graphics Forum (CGF), 34(3), pages 271-280, 2015.
- [SkrabaWangChen2015]: Robustness-Based Simplification of 2D Steady and Unsteady Vector Fields. Primoz Skraba, Bei Wang, Guoning Chen and Paul Rosen. IEEE Transactions on Visualization and Computer Graphics (TVCG), 21(8), pages 930 - 944, 2015.
- [LiuWangThiagarajan2014]: Multivariate Volume Visualization through Dynamic Projections. Shusen Liu, Bei Wang, Jayaraman J. Thiagarajan, Peer-Timo Bremer and Valerio Pascucci. IEEE Symposium on Large Data Analysis and Visualization (LDAV), 2014.
- [PalandeJoseZielinski2017]: Revisiting Abnormalities in Brain Network Architecture Underlying Autism Using Topology-Inspired Statistical Inference. Sourabh Palande, Vipin Jose, Brandon Zielinski, Jeffrey Anderson, P. Thomas Fletcher and Bei Wang. International Workshop on Connectomics in NeuroImaging (CNI) at MICCAI, 2017.
- [WongPlandeWang2016]: Kernel Partial Least Squares Regression for Relating Functional Brain Network Topology to Clinical Measures of Behavior. Eleanor Wong, Sourabh Palande, Bei Wang, Brandon Zielinski, Jeffrey Anderson and P. Thomas Fletcher. International Symposium on Biomedical Imaging (ISBI), 2016.
- [BremerMaljovecSaha2015]: ND2AV: N-Dimensional Data Analysis and Visualization -- Analysis for the National Ignition Campaign. Peer-Timo Bremer, Dan Maljovec, Avishek Saha, Bei Wang, Jim Gaffney, Brian K. Spears and Valerio Pascucci. Computing and Visualization in Science (CVS), 17(1), Pages 1-18, 2015.

Slide Deck References

- [LiuBremerThiagarajan2017]: Visual Exploration of Semantic Relationships in Neural Word Embeddings. Shusen Liu, Peer-Timo Bremer, Jayaraman J. Thiagarajan, Vivek Srikumar, Bei Wang, Yarden Livnat and Valerio Pascucci. IEEE Transactions on Visualization and Computer Graphics, 2017.
- [LiuWangBremer2014]: Distortion-Guided Structure-Driven Interactive Exploration of High-Dimensional Data. Shusen Liu, Bei Wang, Peer-Timo Bremer and Valerio Pascucci. Computer Graphics Forum (CGF), 33(3), pages 101-110, 2014.
- [HajijWangScheideggerRosen2018]: Visual Detection of Structural Changes in Time-Varying Graphs Using Persistent Homology. IEEE Pacific Visualization Symposium (conditionally accepted), 2018.
- [DomenicoPorterArenas2015]: MuxViz: a tool for multilayer analysis and visualization of networks. Manlio De Domenico, Mason A. Porter and Alex Arenas. Journal of Complex Networks, 2015.

Thanks!

Any questions?

You can find me at: **beiwang@sci.utah.edu**

CREDITS

Special thanks to all people who made and share these awesome resources for free:

- Presentation template designed by Slidesmash
- Photographs by unsplash.com and pexels.com
- Vector Icons by Matthew Skiles

Presentation Design

This presentation uses the following typographies and colors:

Free Fonts used:

http://www.1001fonts.com/oswald-font.html

https://www.fontsquirrel.com/fonts/open-sans

Colors used