Topology in Biomedicine

Bala Krishnamoorthy
Washington State University

Search "Bala WSU"

Centers for Disease Control and Prevention

Data Science: Man's Game?

(image: www)

Prof. Susan Holmes
Stanford U.

Data Science: Man's Game?

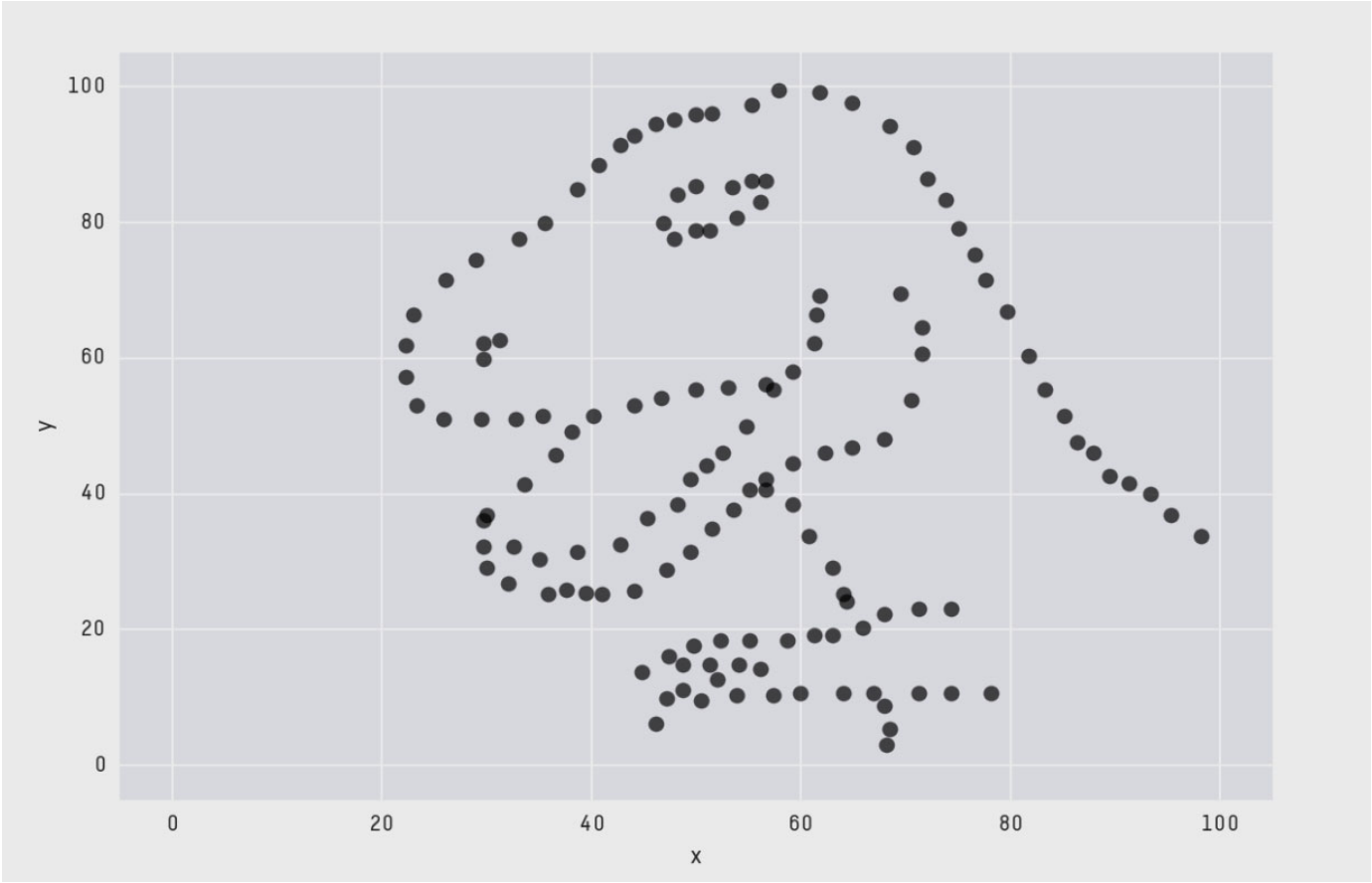
(image: www)

Prof. Susan Holmes
Stanford U.

"... my data is bigger than yours!"

The Datasaurus

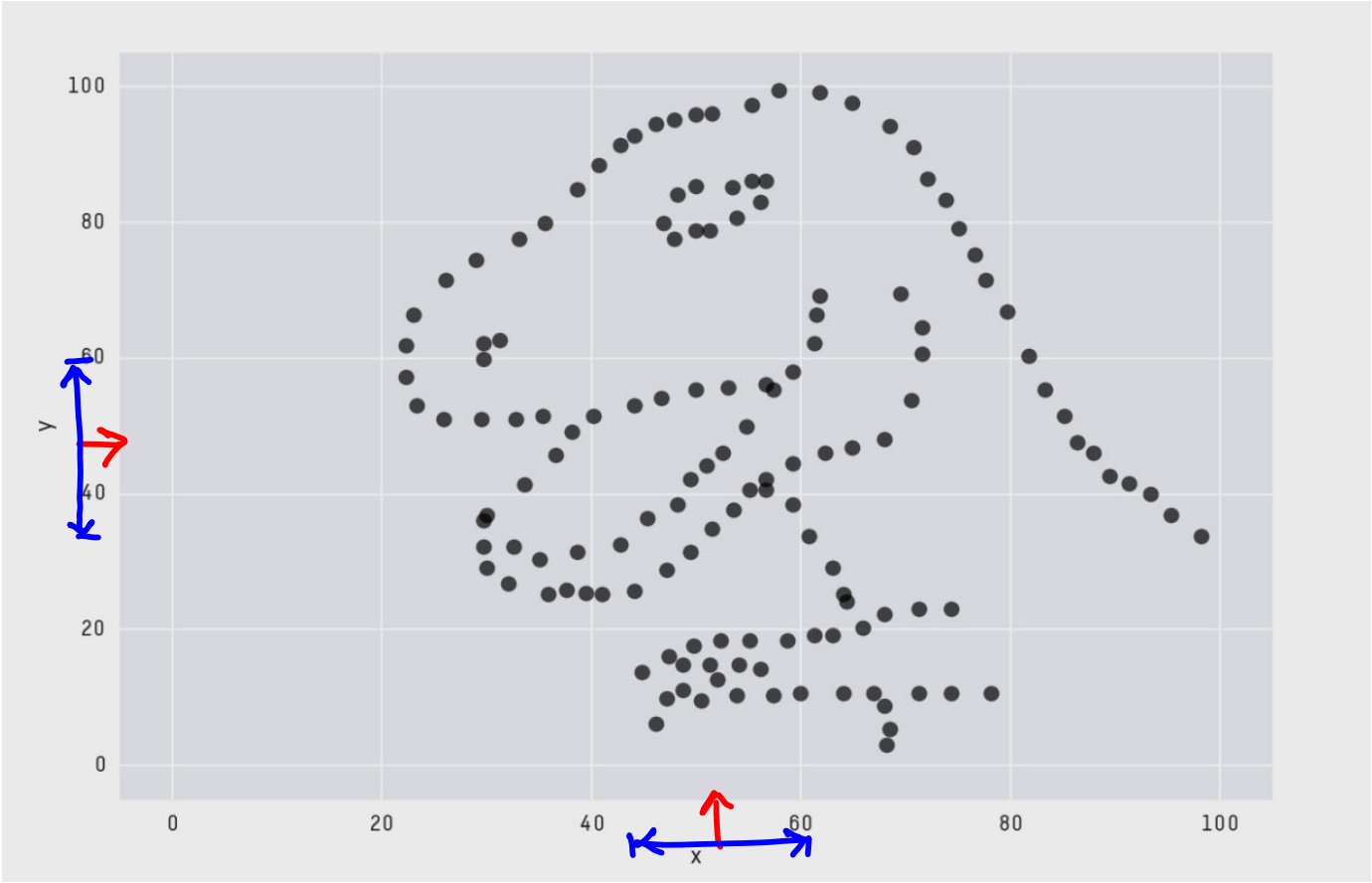
(Matejka & Fitzmaurice, 2017)



The Datasaurus

(Matejka & Fitzmaurice, 2017)

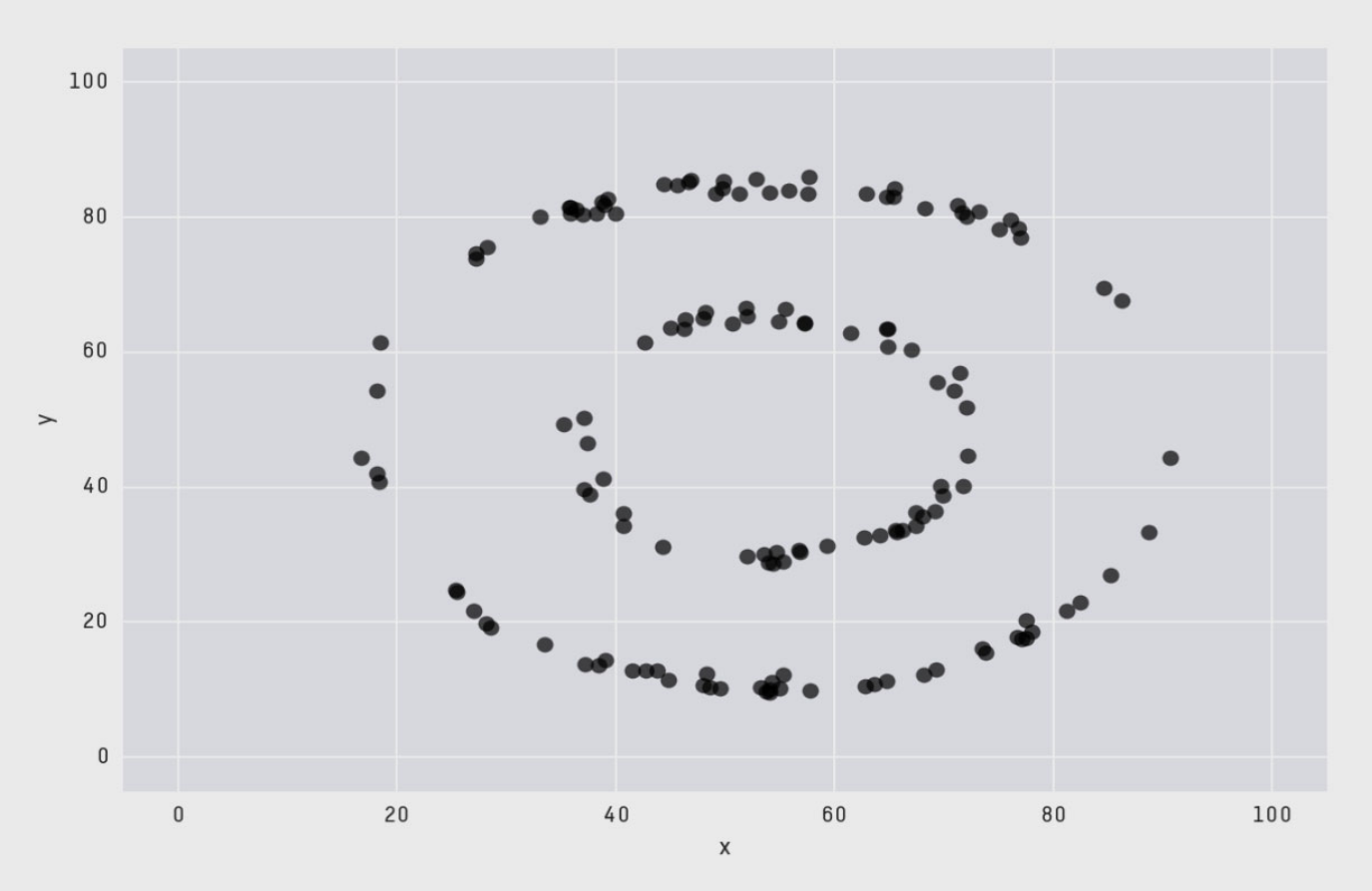
y-stats (average, std. dev)



x-stats (average, std. dev)

The Datasaurus

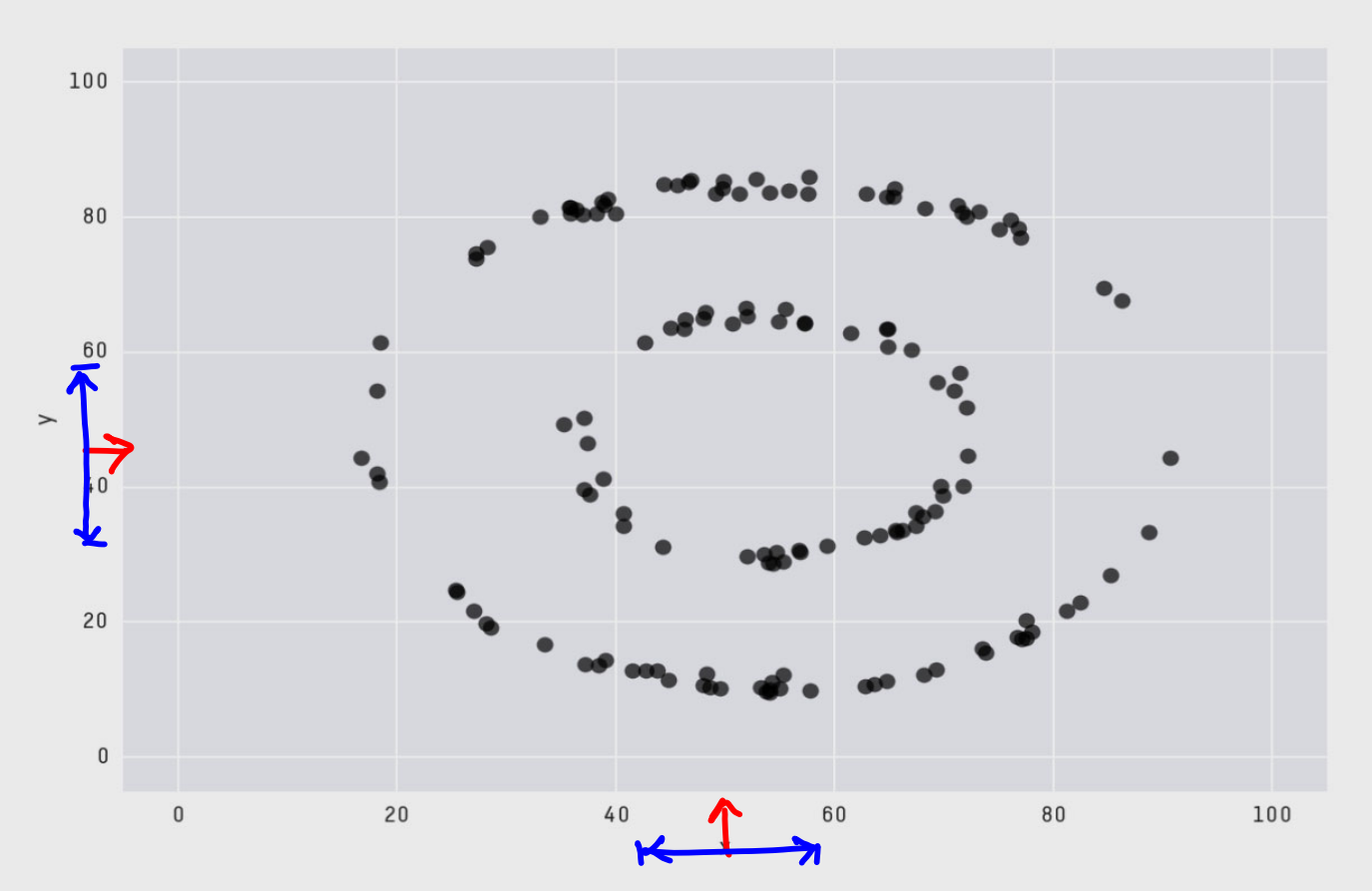
(Matejka & Fitzmaurice, 2017)



The Datasaurus

(Matejka & Fitzmaurice, 2017)

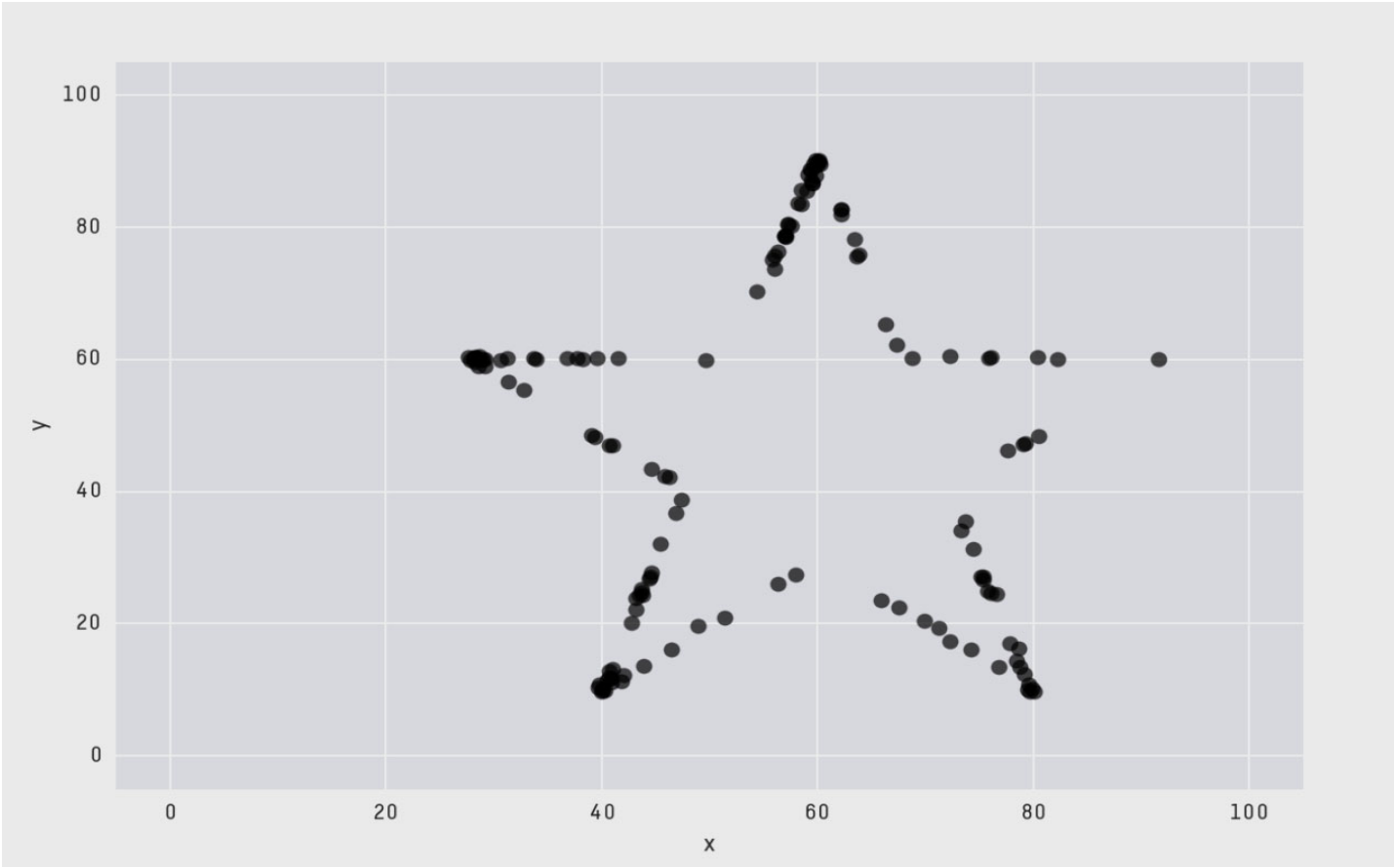
y-stats (average, std. dev)



x-stats (average, std. dev)

The Datasaurus

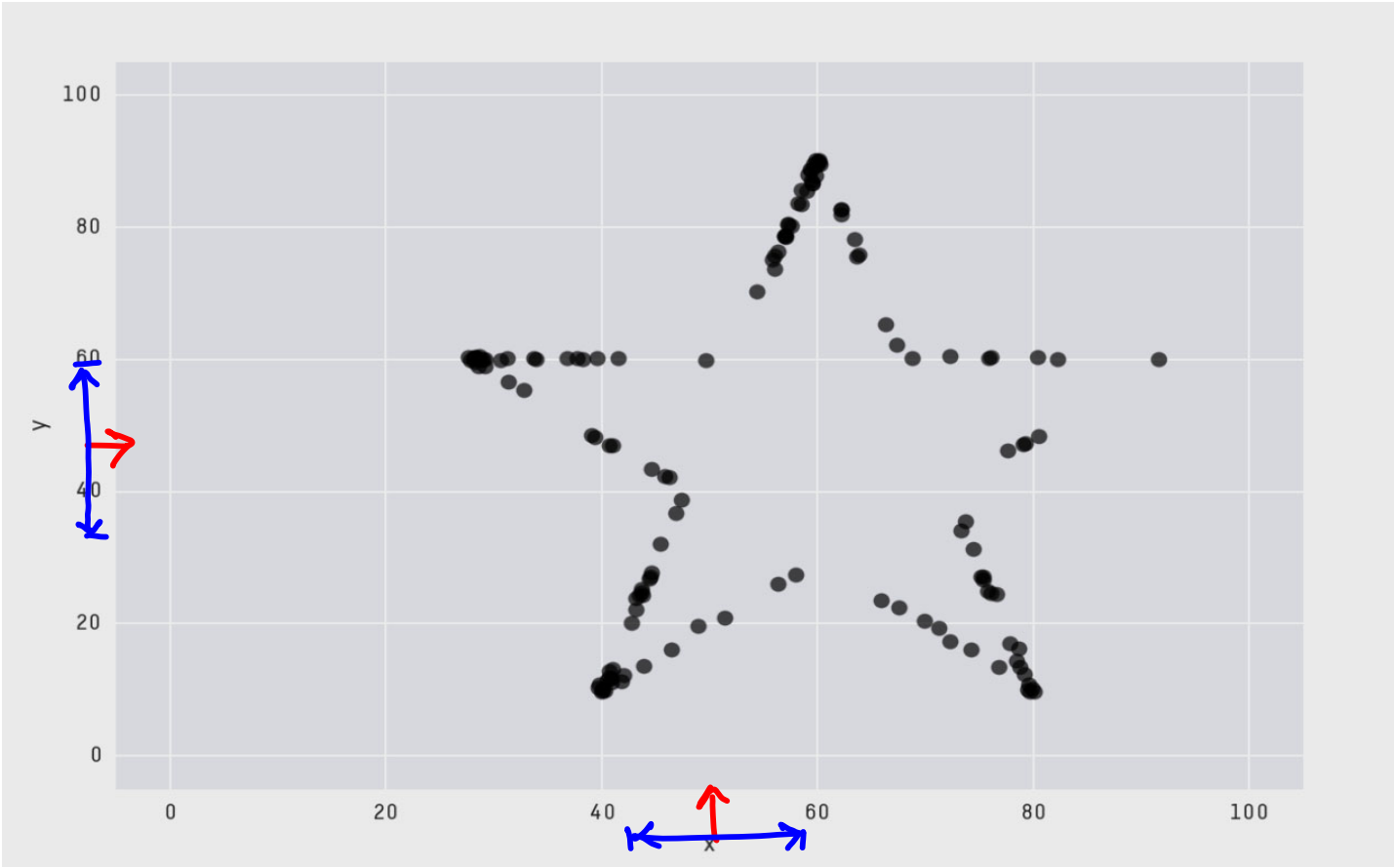
(Matejka & Fitzmaurice, 2017)



The Datasaurus

(Matejka & Fitzmaurice, 2017)

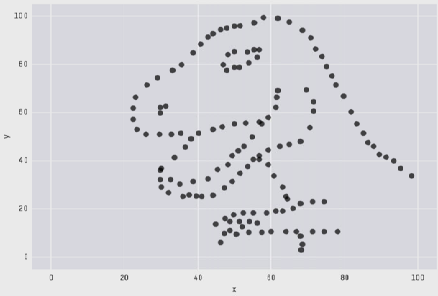
y-stats (average, std. dev)



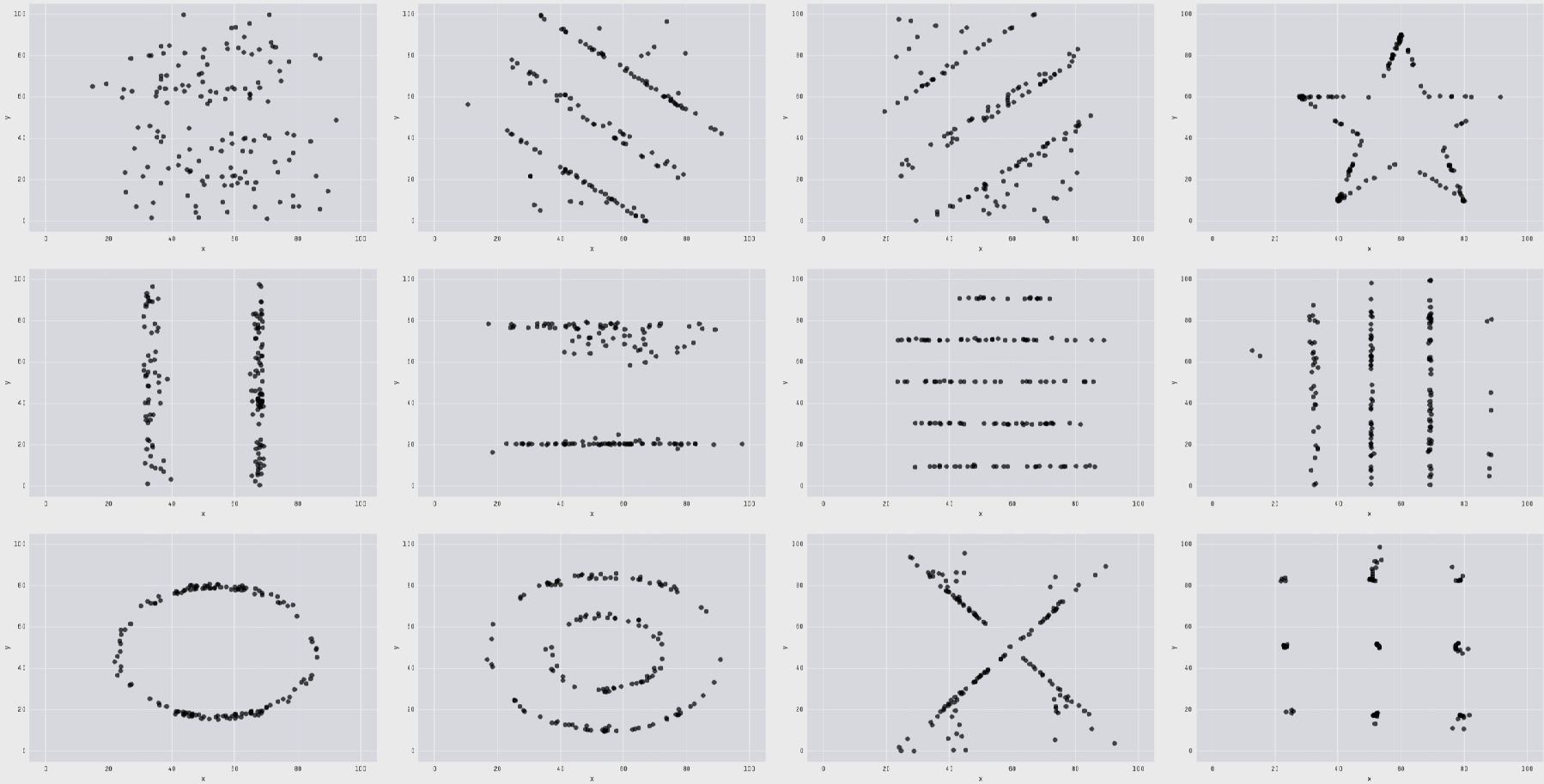
x-stats (average, std. dev)

The Datasaurus Dozen!

(Matejka & Fitzmaurice, 2017)

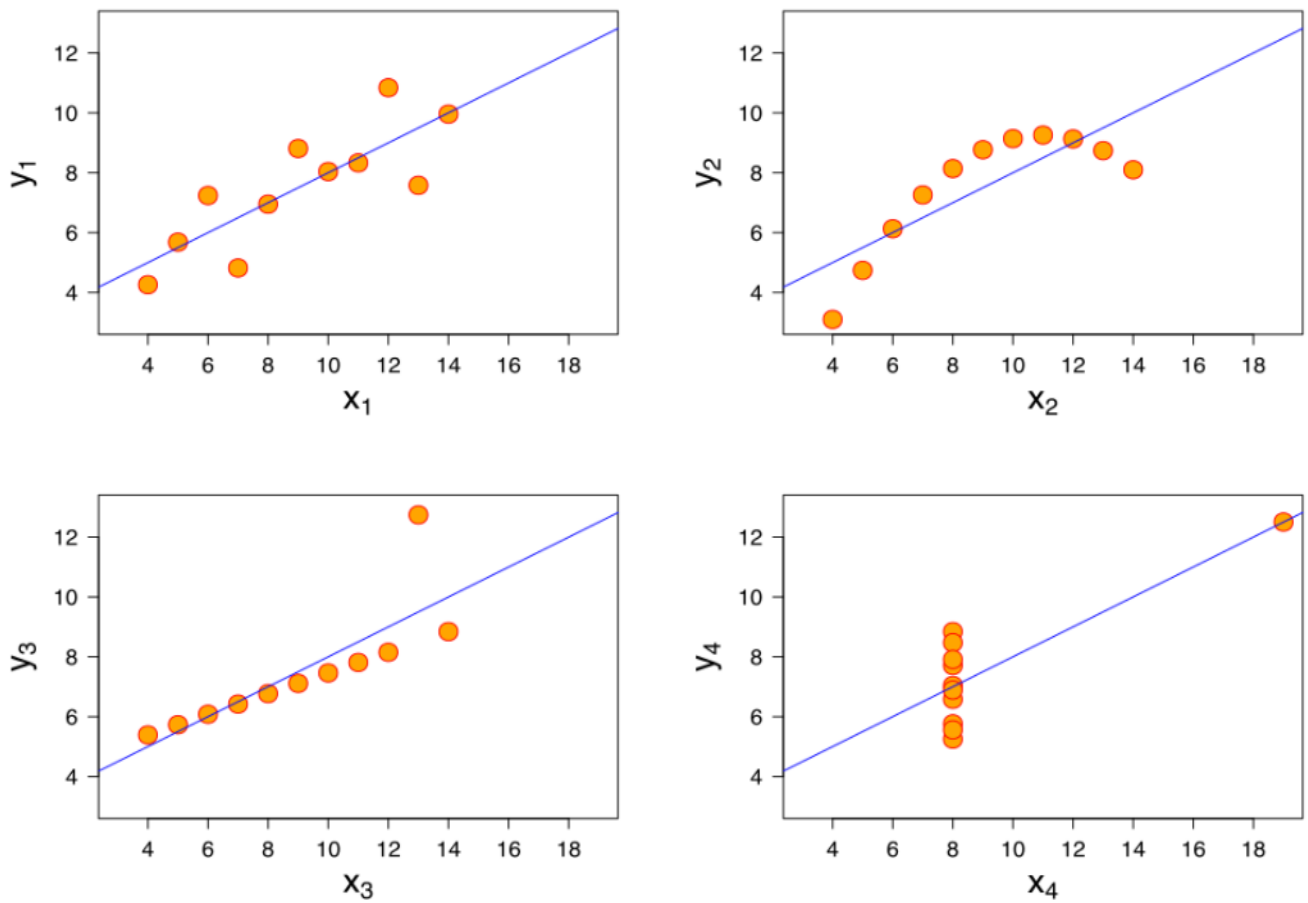


X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06



Anscombe's Quartet

(Anscombe, 1973)

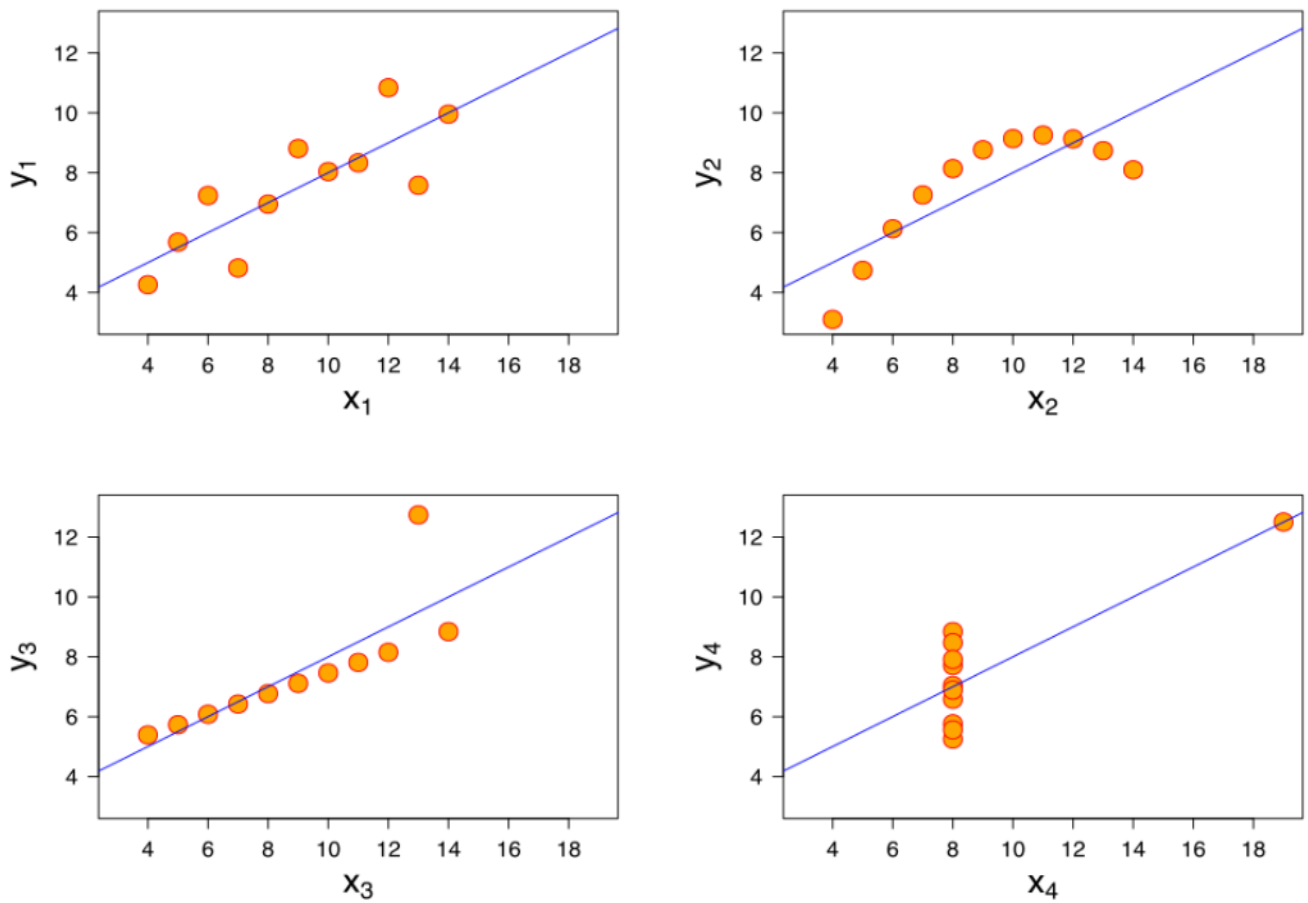


identical summary statistics!

(image: Wikipedia)

Anscombe's Quartet

(Anscombe, 1973)



identical summary statistics!

(image: Wikipedia)

* Need to visualize data!

DATA → INSIGHT

Challenges

* not just large sizes

DATA → INSIGHT

Challenges

* not just large sizes
* complexity!

DATA → INSIGHT

Challenges

* not just large sizes
* complexity!
* need to visualize hi-dimensional data

DATA → INSIGHT

Challenges

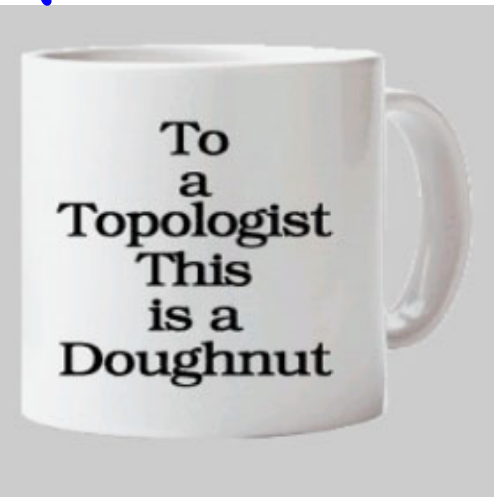
* not just large sizes
* complexity!
* need to visualize hi-dimensional data
* need to extract "hidden" structure

TOPOLOGY ?

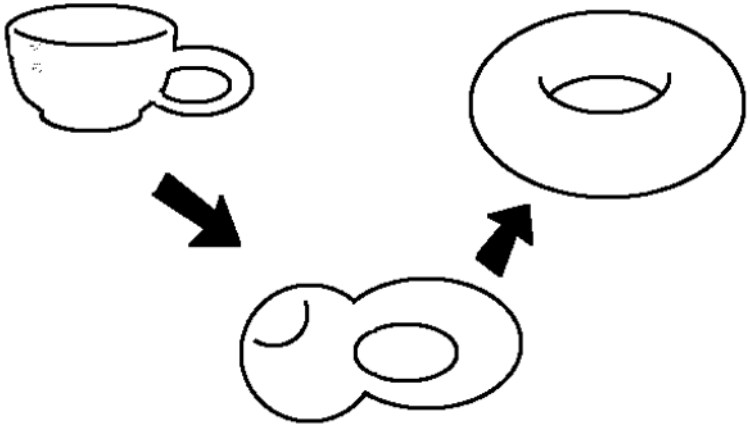
* branch of mathematics that studies how space is "connected"

TOPOLOGY ?

* branch of mathematics that studies how space is "connected"



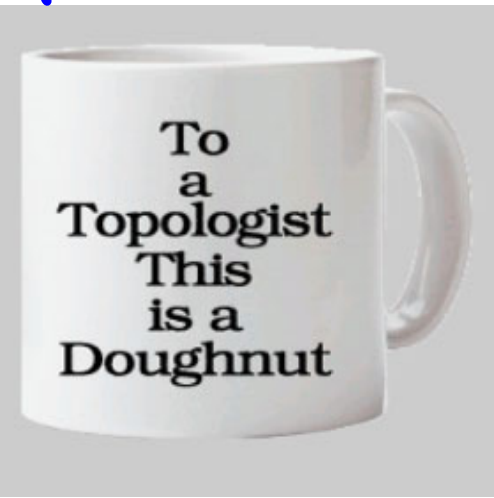
(images: www)

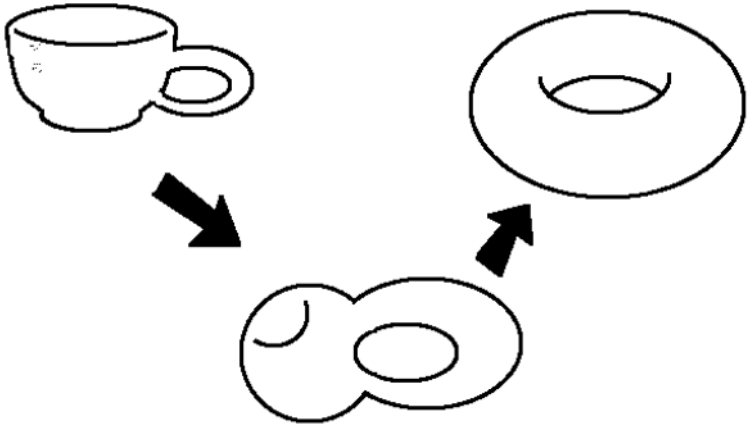


TOPOLOGY ?

* branch of mathematics that studies how space is "connected"

(images: www)



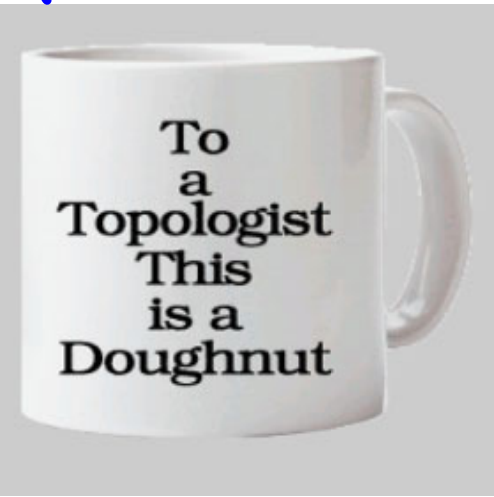


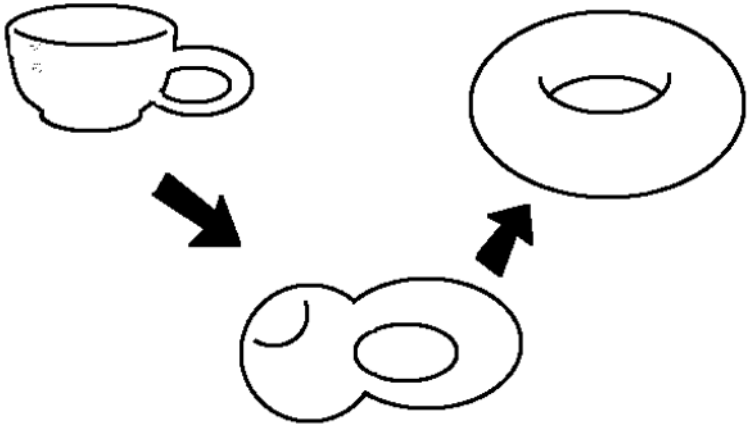
? Is topology of any use at all ?

TOPOLOGY ?

* branch of mathematics that studies how space is "connected"

(images: www)

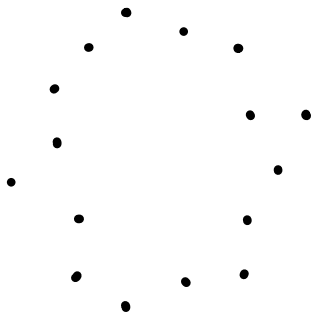




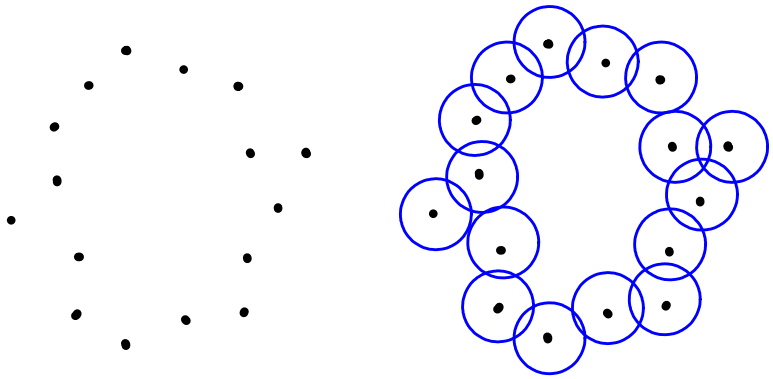
? Is topology of any use at all ?

A A A B B C C

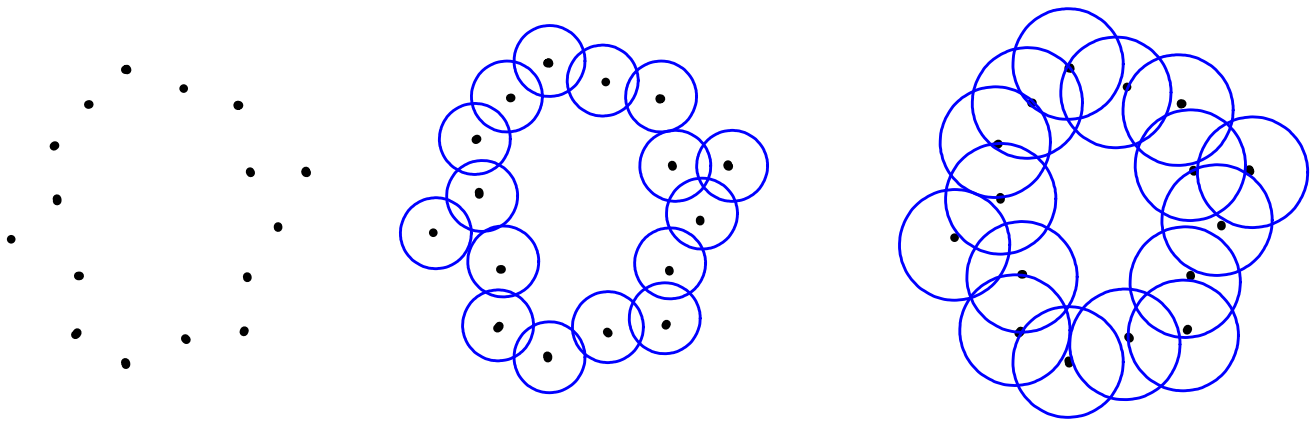
Persistent Homology



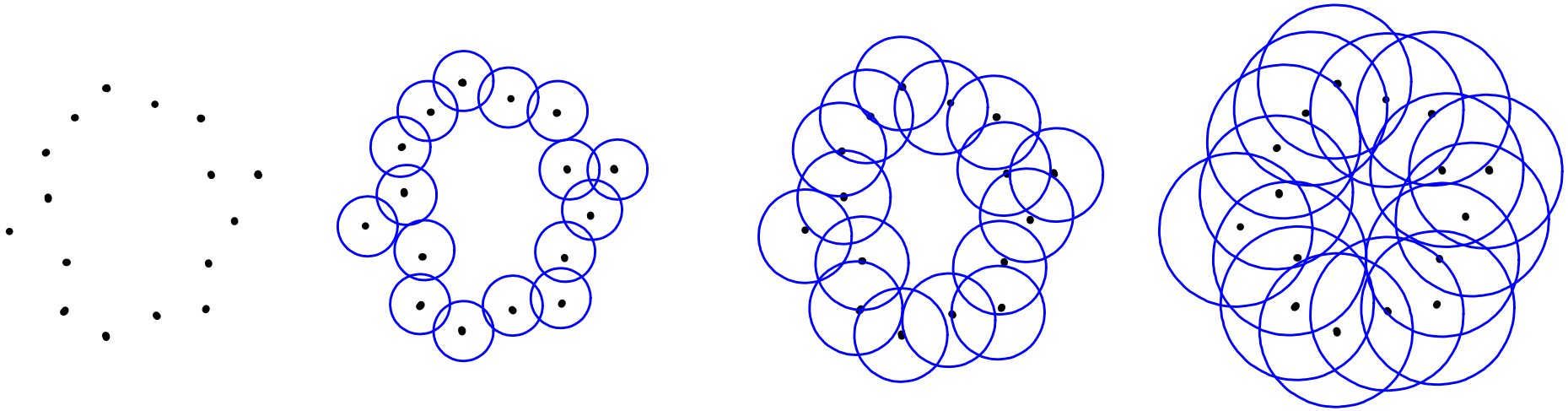
Persistent Homology



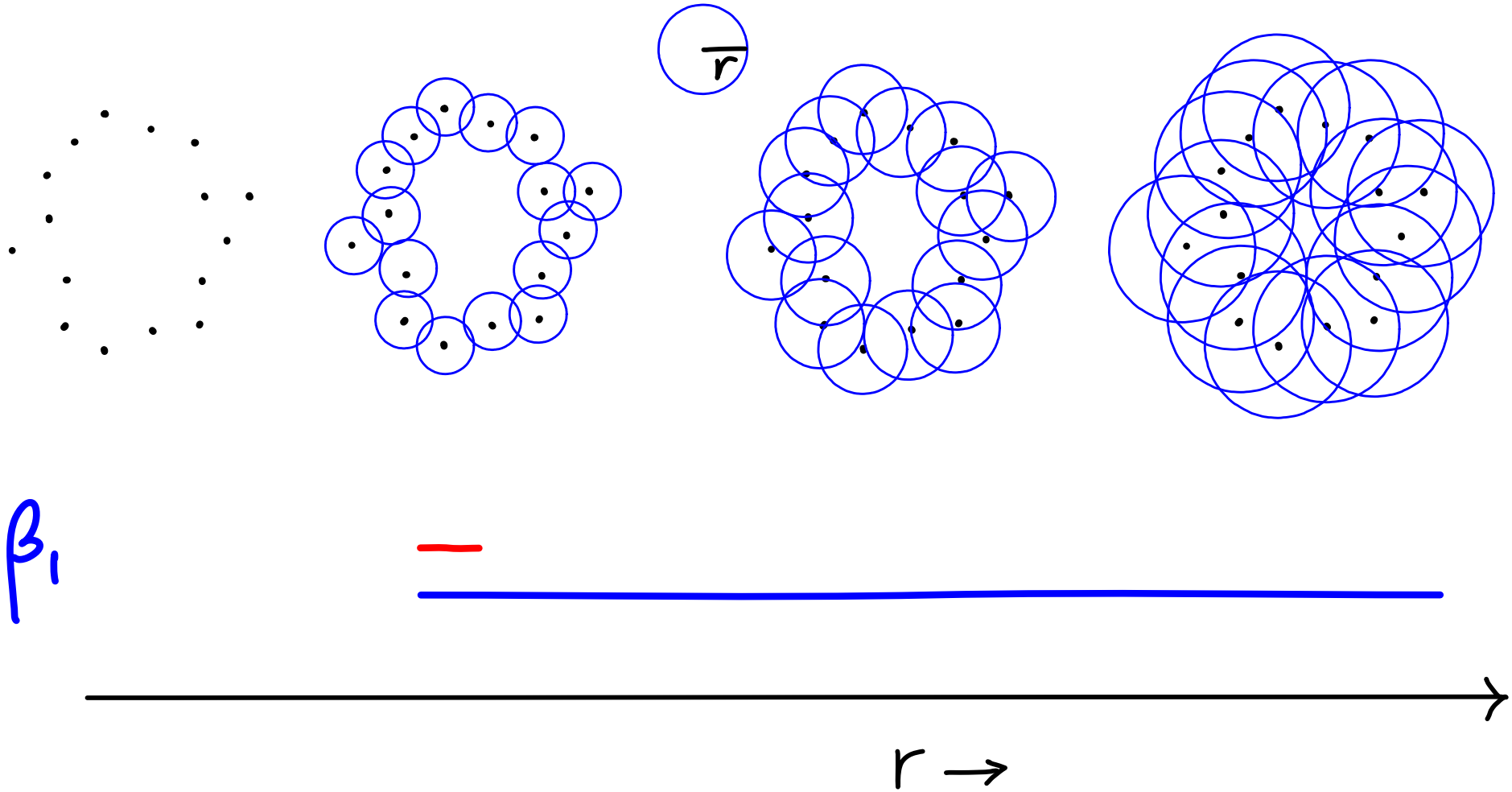
Persistent Homology



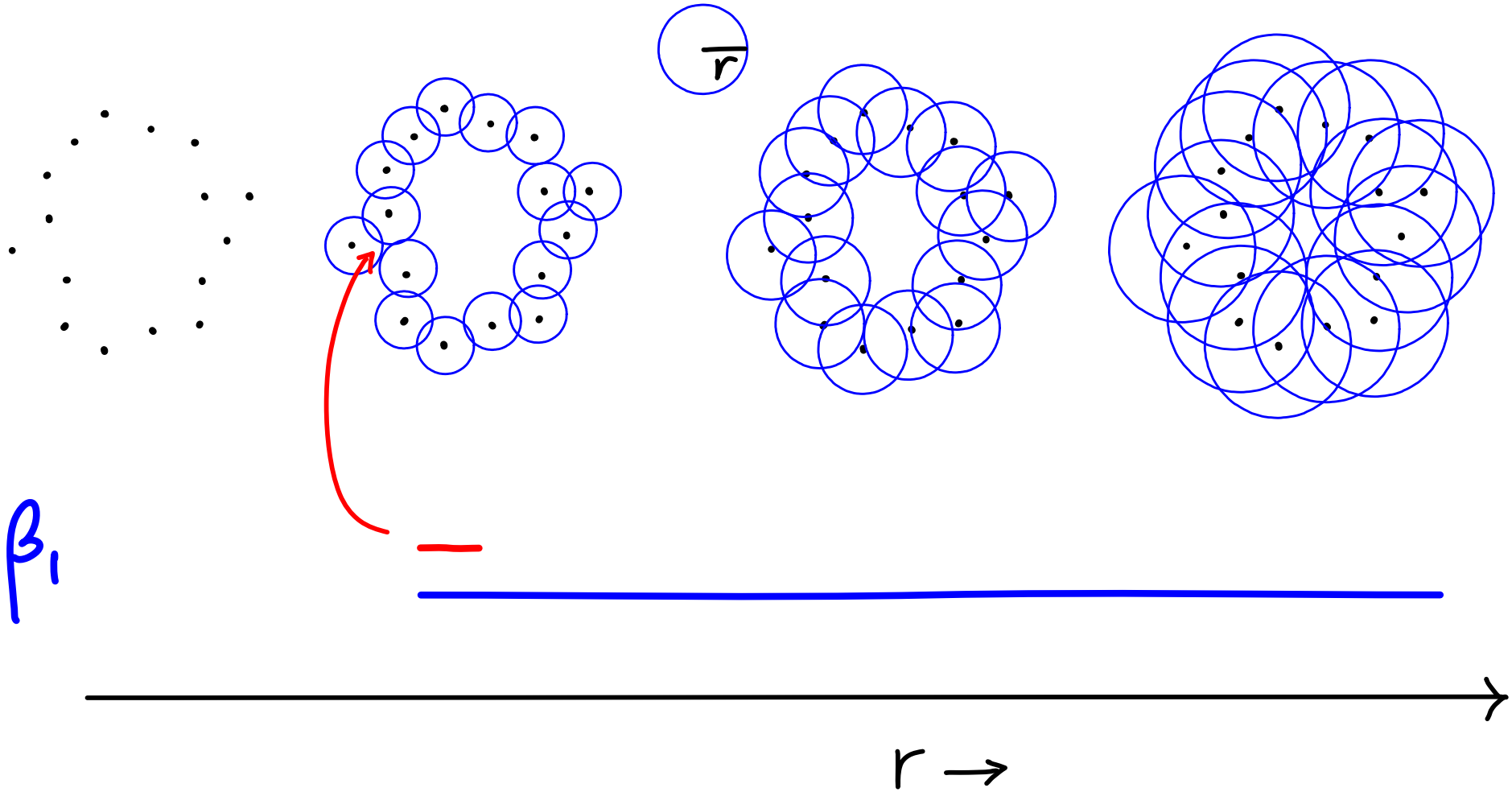
Persistent Homology



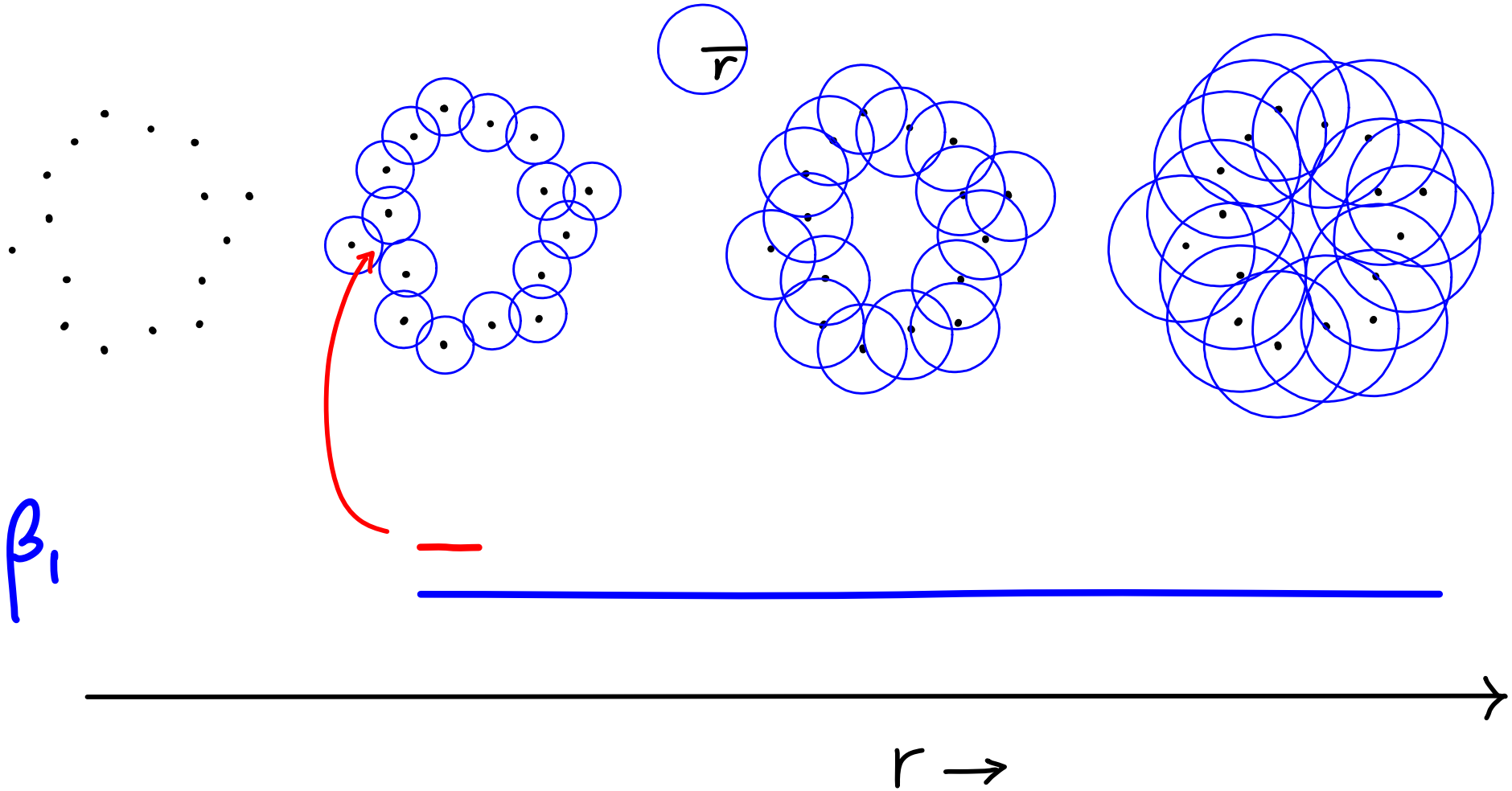
Persistent Homology



Persistent Homology



Persistent Homology



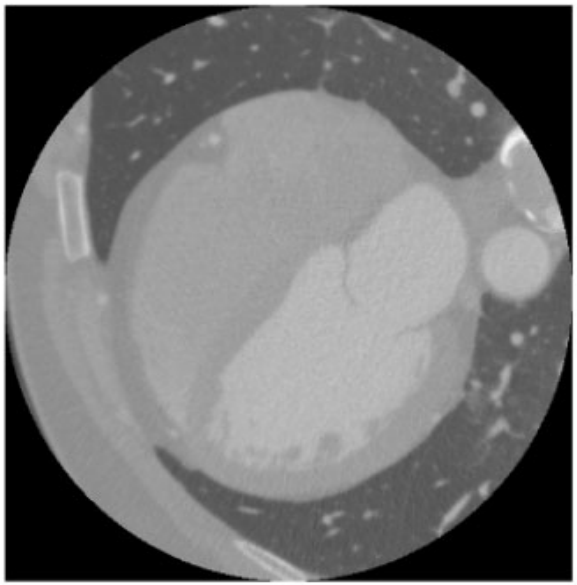
* "barcodes" for shapes!

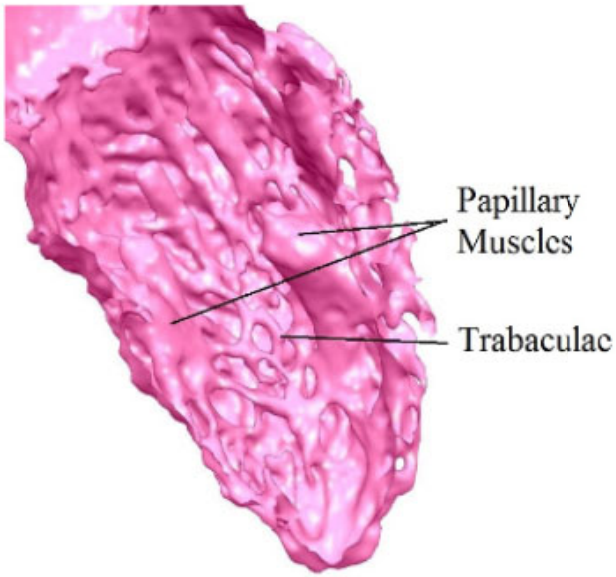
TOPOLOGY FINDS HOLES!

* 3D structure of proteins (Edelsbrunner,...)

TOPOLOGY FINDS HOLES!

* 3D structure of proteins (Edelsbrunner,...)
* Cardiac trabeculae (Chen et al., 2017)





Holes in Cancer Gene Expr. Data

(Lockwood & K., 2015)

* 10s-100s patients, ~50,000 gene expressions

Holes in Cancer Gene Expr. Data

(Lockwood & K., 2015)

* 10s-100s patients, ~50,000 gene expressions

huge dim! distances not useful

Holes in Cancer Gene Expr. Data

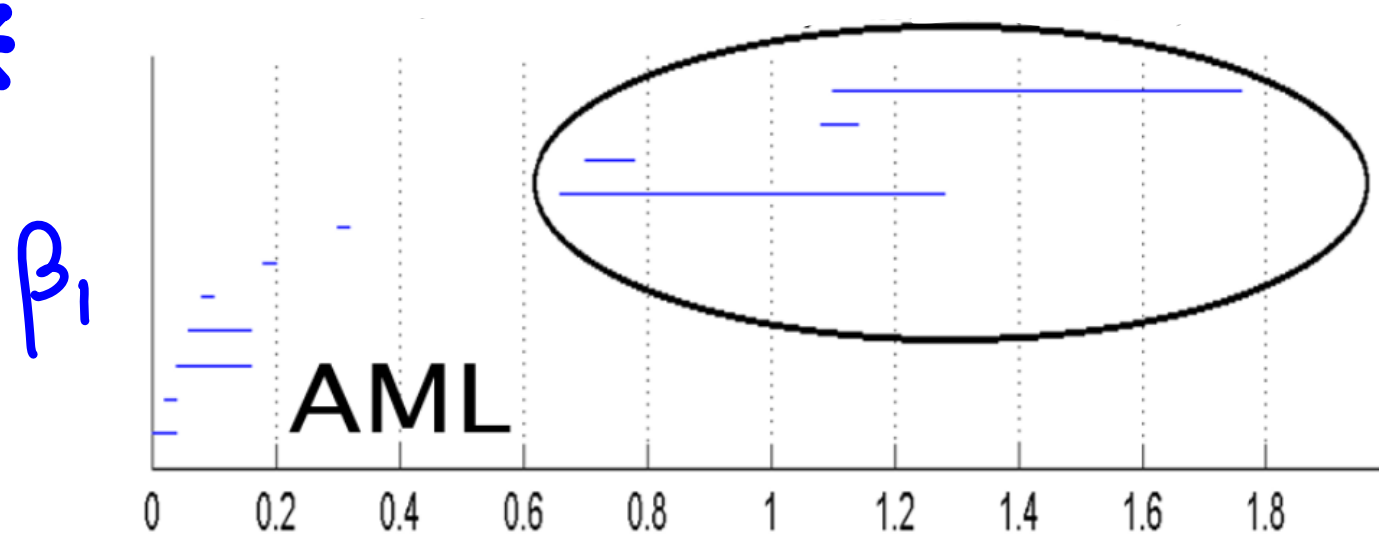
(Lockwood & K., 2015)

* 10s-100s patients, ~50,000 gene expressions
  huge dim! distances not useful
* "dualize" → genes in patients space

Holes in Cancer Gene Expr. Data

(Lockwood & K., 2015)

* 10s-100s patients, ~50,000 gene expressions
  huge dim! distances not useful
* "dualize" → genes in patients space
*

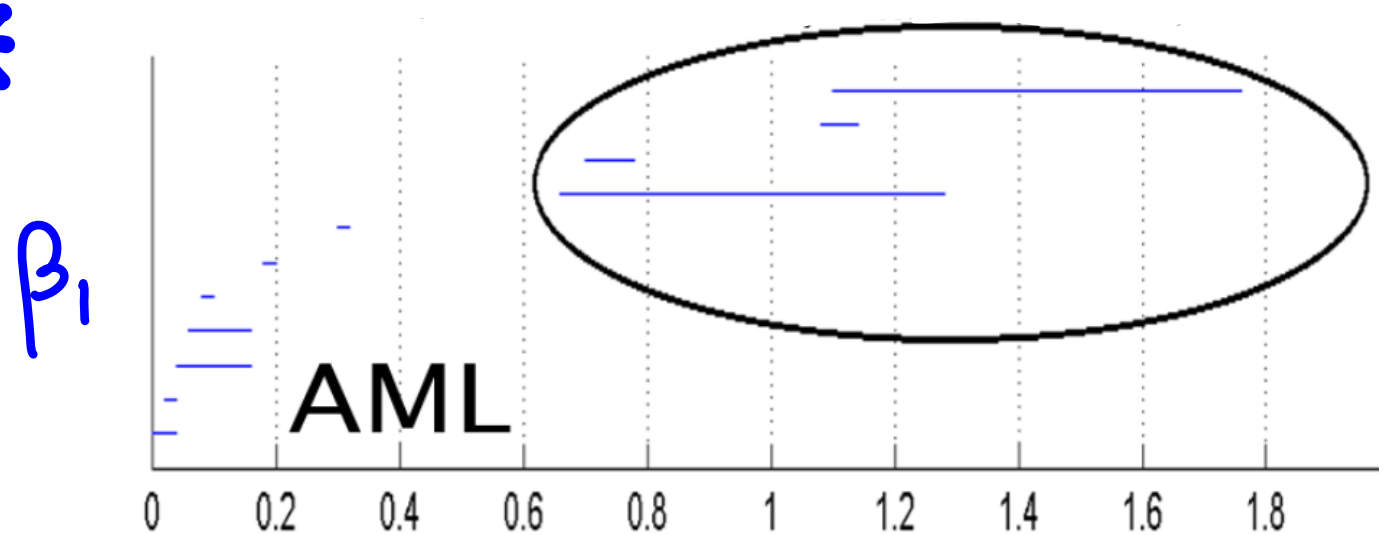


→ loops

Holes in Cancer Gene Expr. Data

(Lockwood & K., 2015)

* 10s-100s patients, ~50,000 gene expressions
  huge dim! distances not useful
* "dualize" → genes in patients space
*

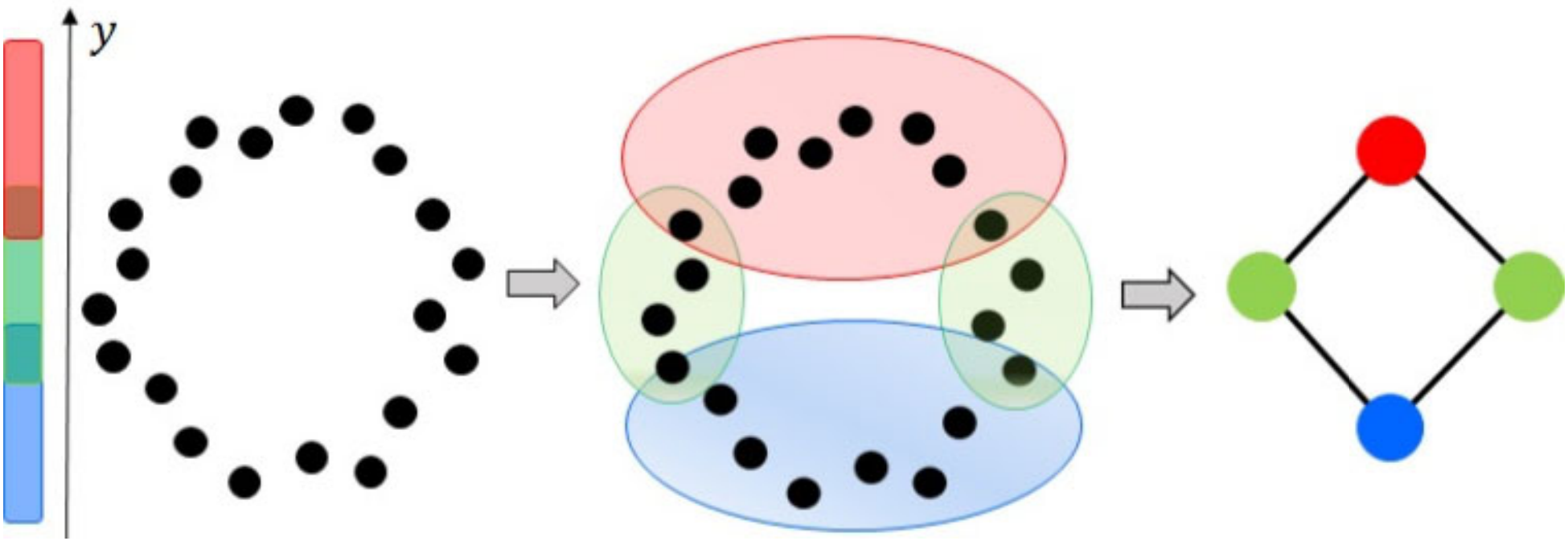


→ loops

✓ loop genes implicated in cancer(s)

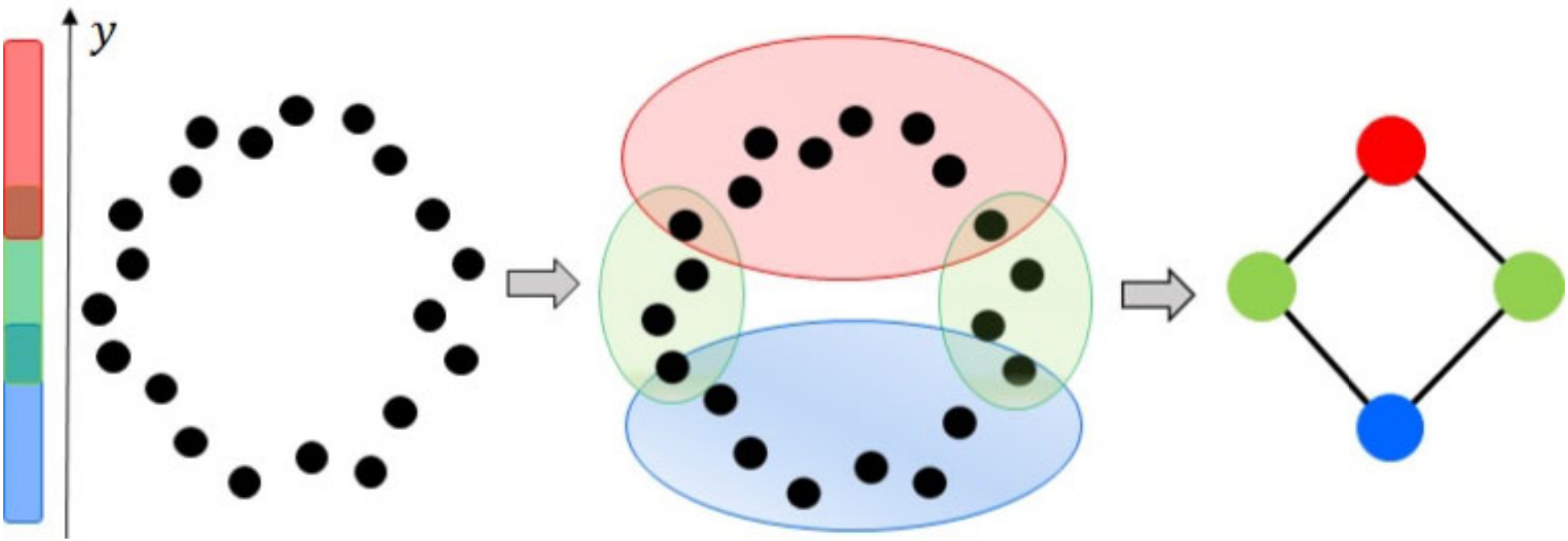
Mapper

(Singh, Memóli, Carlsson, 2007)



MAPPER

(Singh, Memóli, Carlsson, 2007)



* increasingly used in diverse applications!

TOPOLOGICAL DATA ANALYSIS (TDA)

* highly compact representation of data that "reveals structure"

Topological Data Analysis (TDA)

* highly compact representation of data that "reveals structure"
  - ✓ coordinate-free representation

Topological Data Analysis (TDA)

* highly compact representation of data that "reveals structure"
  - ✓ coordinate-free representation
  - ✓ insensitive to small changes in data

Topological Data Analysis (TDA)

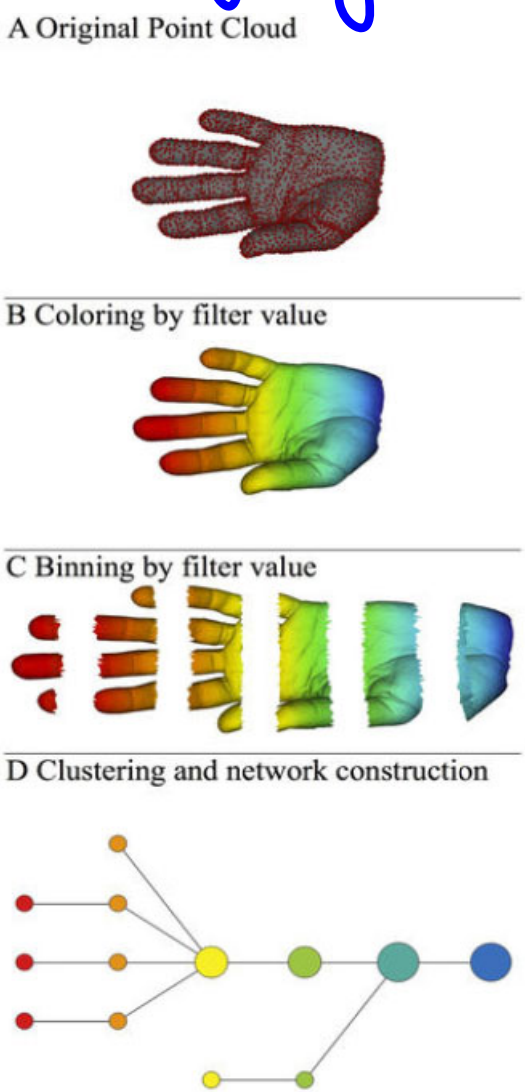
* highly compact representation of data that "reveals structure"
  - ✓ coordinate-free representation
  - ✓ insensitive to small changes in data
  - ✓ highly compressed representation

Topological Data Analysis (TDA)

* highly compact representation of data that "reveals structure"
  - ✓ coordinate-free representation
  - ✓ insensitive to small changes in data
  - ✓ highly compressed representation
  - ✓ excellent visualization(s)

Topological Data Analysis (TDA)

* highly compact representation of data that "reveals structure"

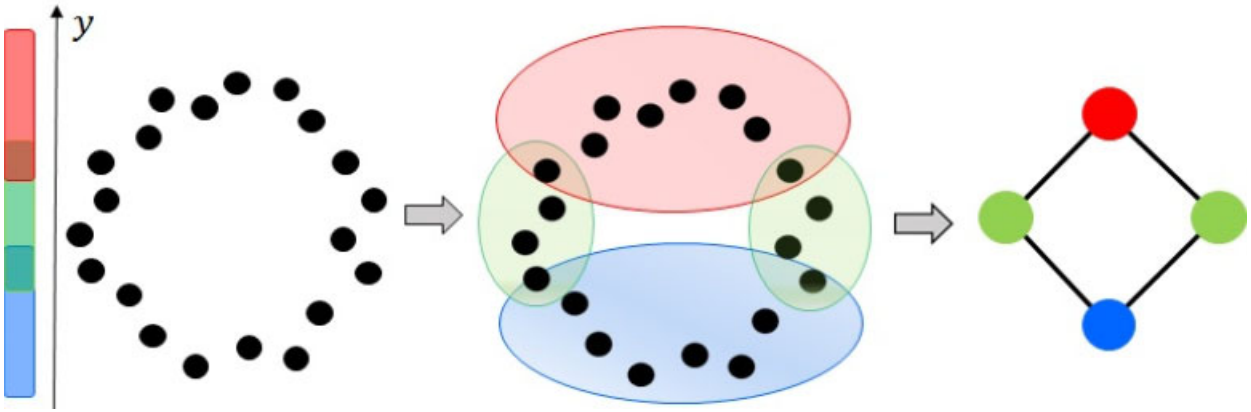


- ✓ coordinate-free representation
- ✓ insensitive to small changes in data
- ✓ highly compressed representation
- ✓ excellent visualization(s)

(Lum et al. 2013, Nature)

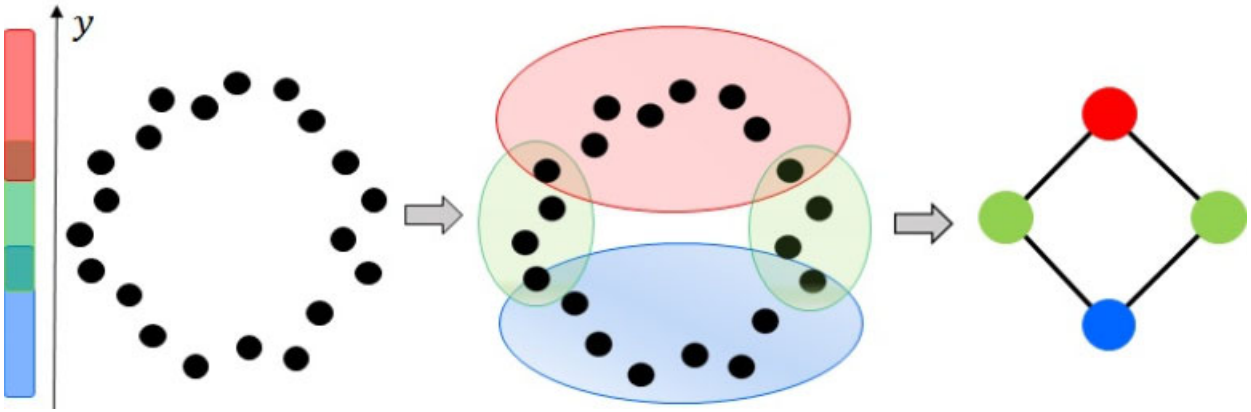
Mapper: Details

✓ "stability" over interval length, overlap %



Mapper: Details

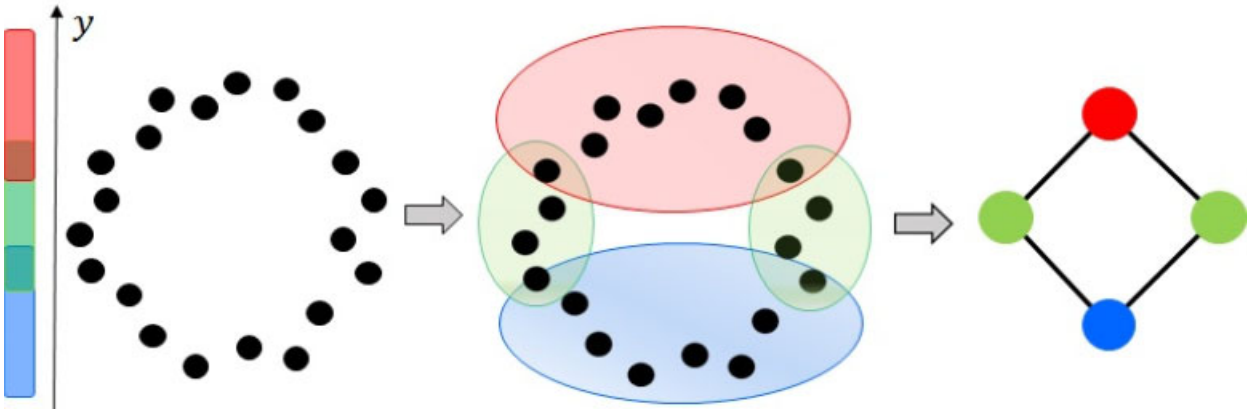
✓ "stability" over interval length, overlap %



✓ multiple functions — can be parallelized

MAPPER: DETAILS

✓ "stability" over interval length, overlap %

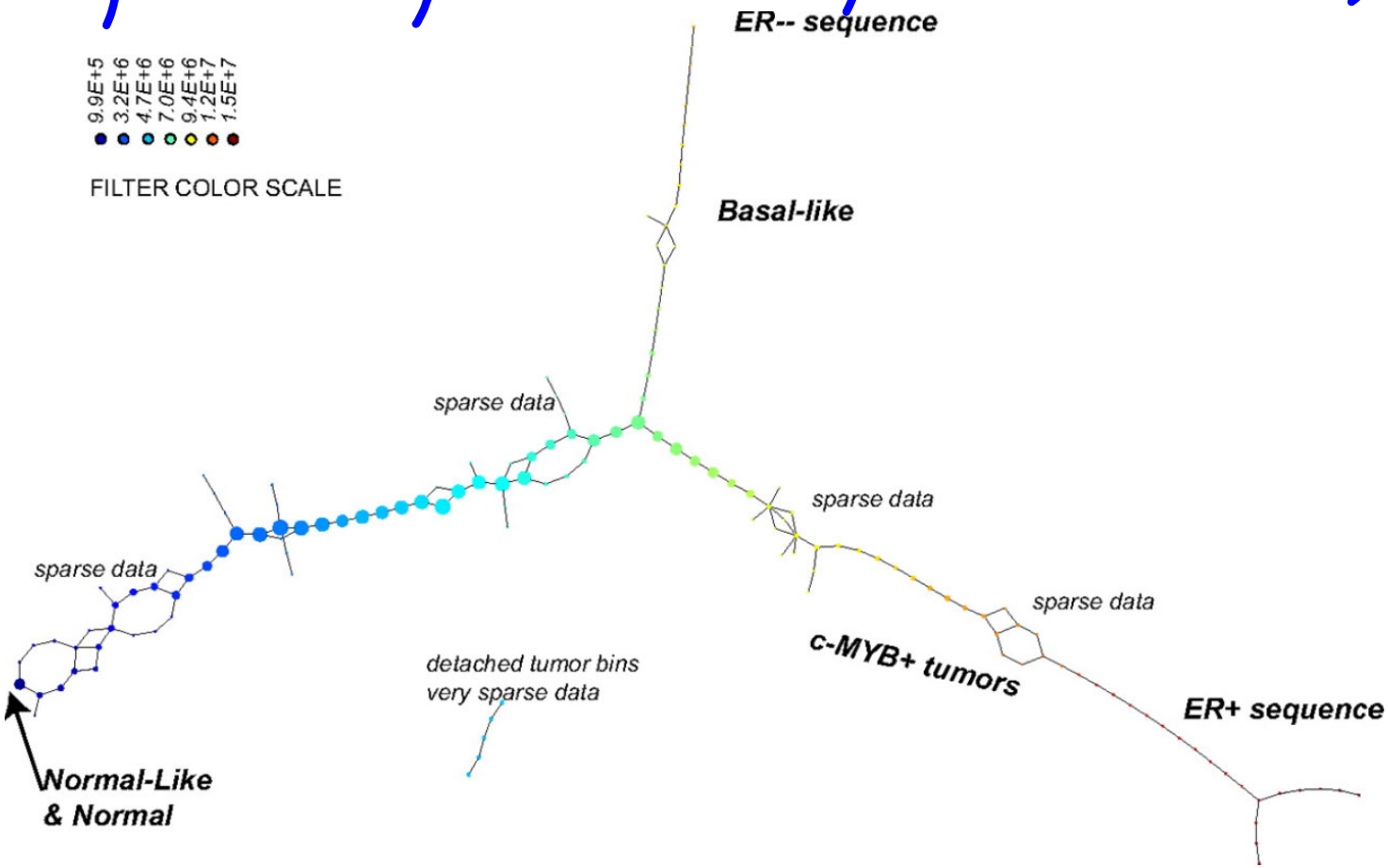


✓ multiple functions — can be parallelized

✓ variable selection using a "topological stepwise regression"

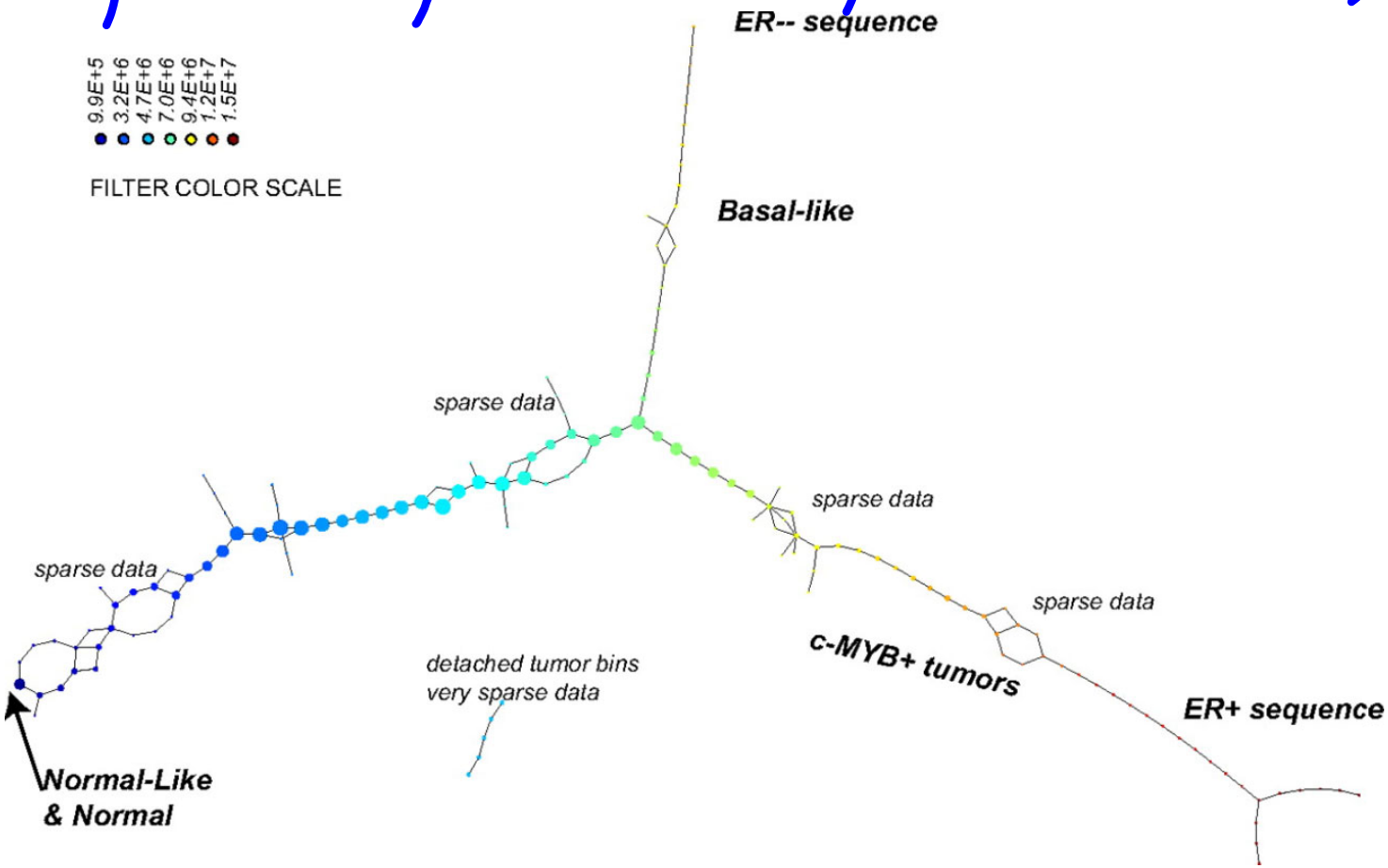
Breast Cancer Survival

Nicolau, Levine, Carlsson, 2011 (PNAS)



Breast Cancer Survival

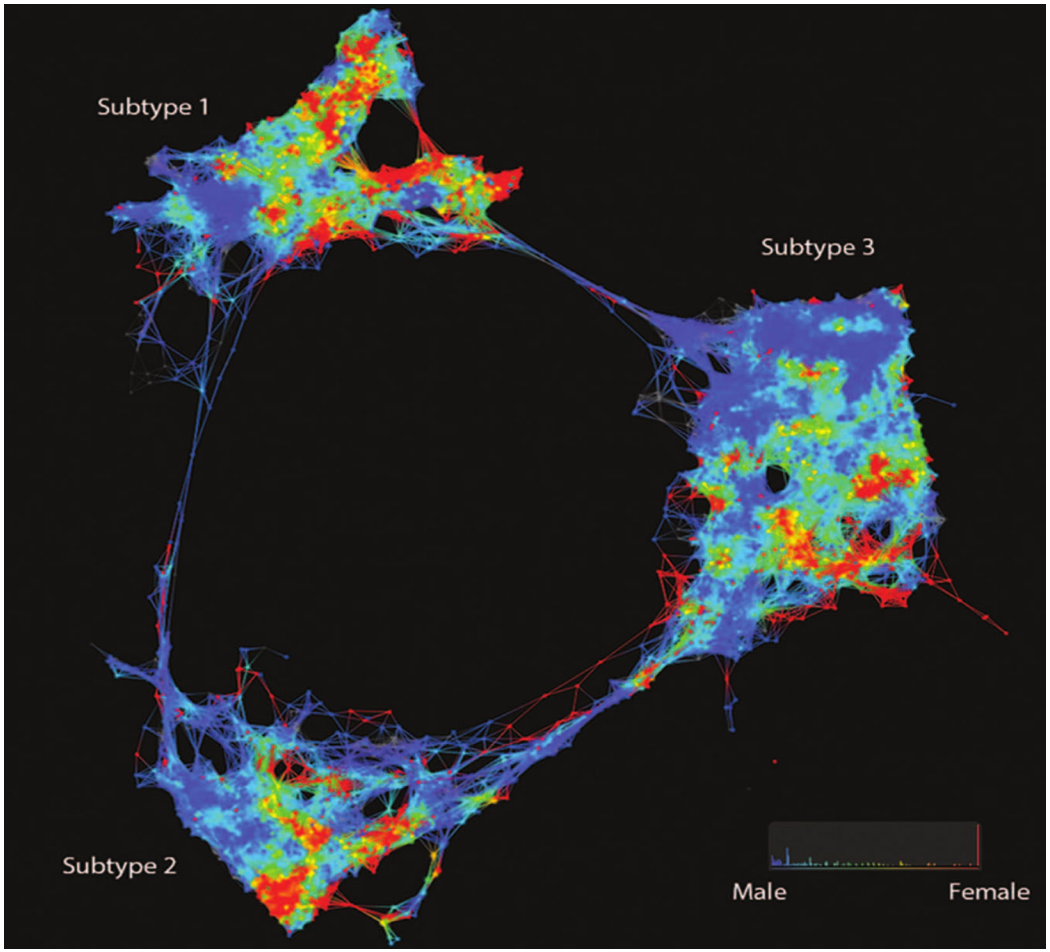
Nicolau, Levine, Carlsson, 2011 (PNAS)



* identified unique group of ER^+ breast cancers with 100% survival and no metastasis

Type 2 Diabetes Subtypes

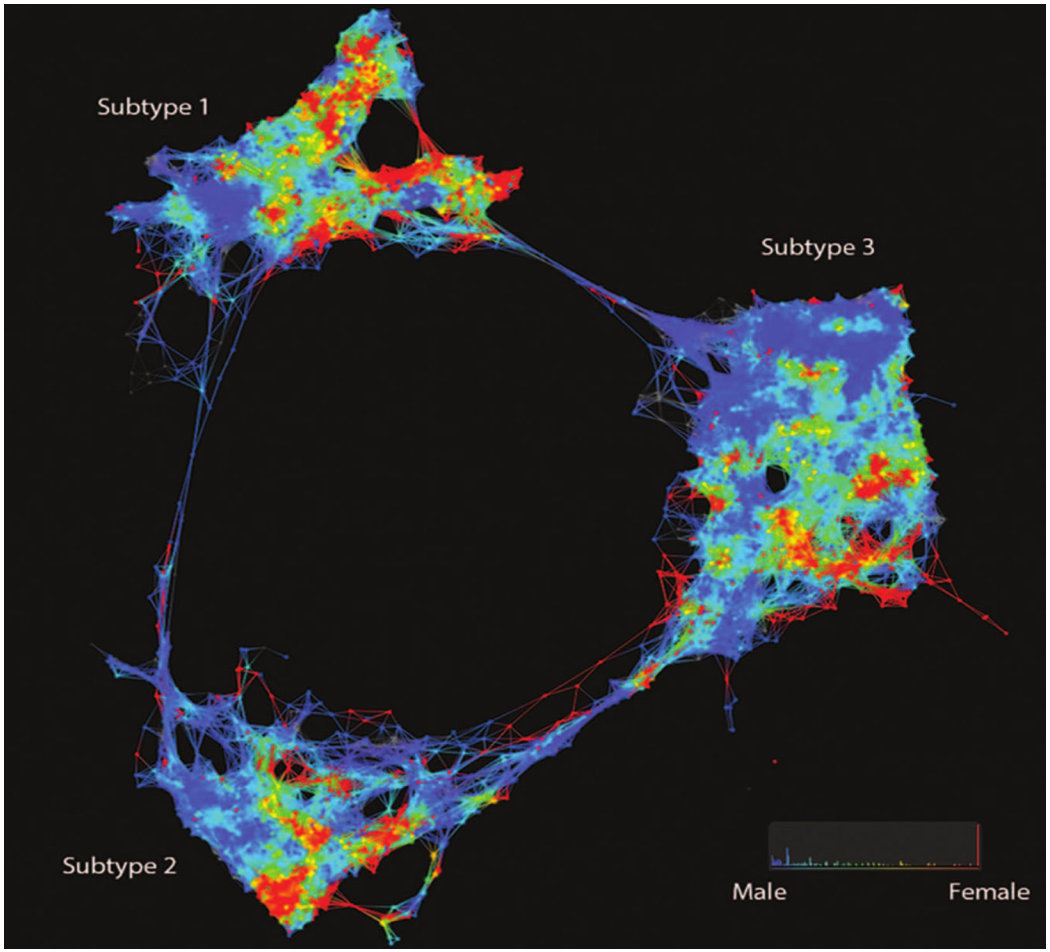
Li et al., 2015 (Science)



* med. records + genotype data

Type 2 Diabetes Subtypes

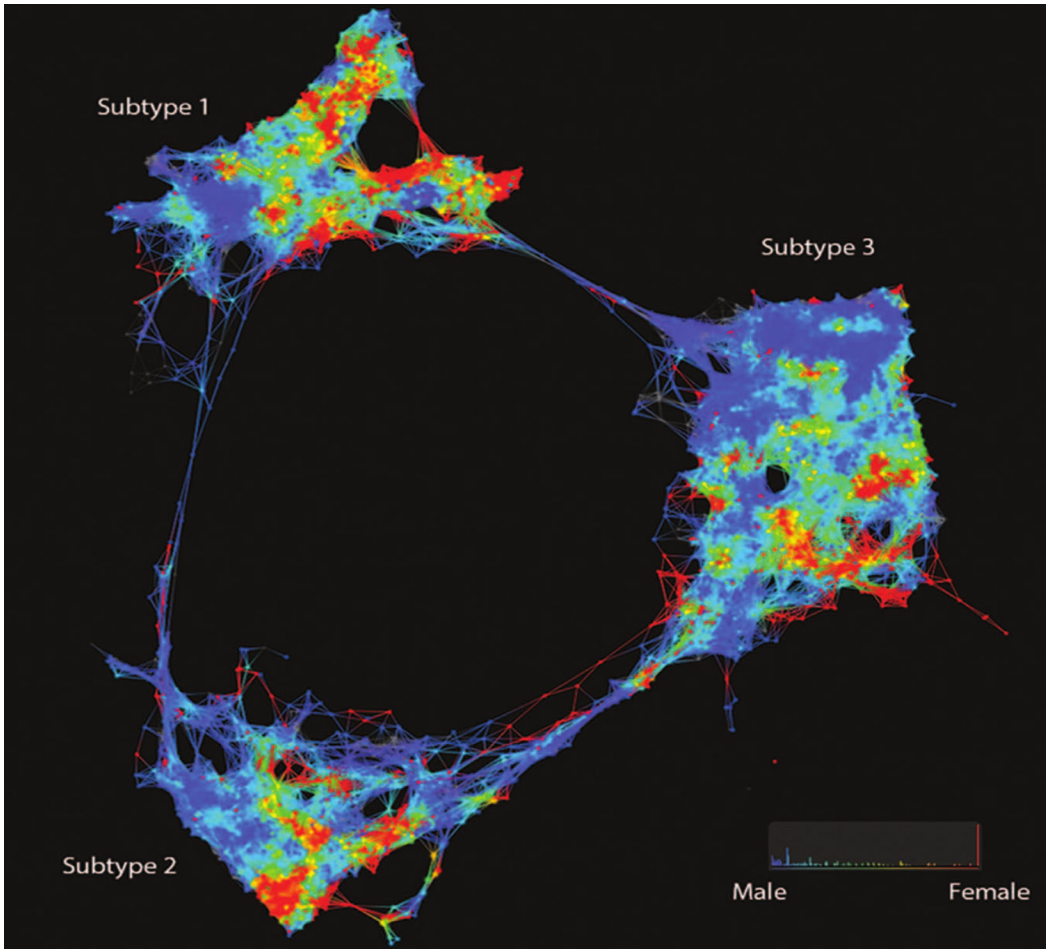
Li et al., 2015 (Science)



* med. records + genotype data
* 3 subgroups w/ specific genetic markers

Type 2 Diabetes Subtypes

Li et al., 2015 (Science)

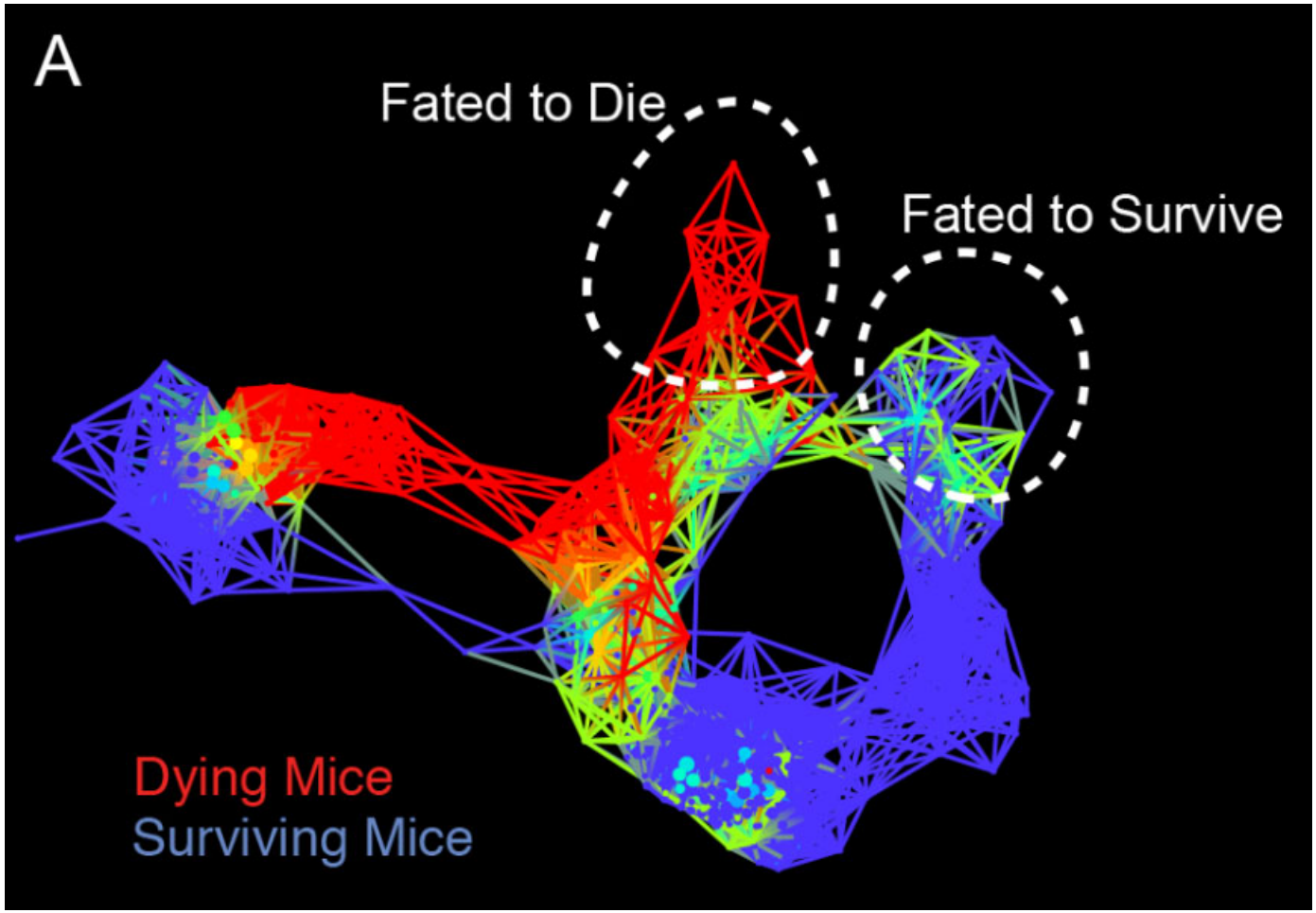


* med. records + genotype data
* 3 subgroups w/ specific genetic markers

* Rostami et al. (2025) — new subtype of luminal B breast cancer

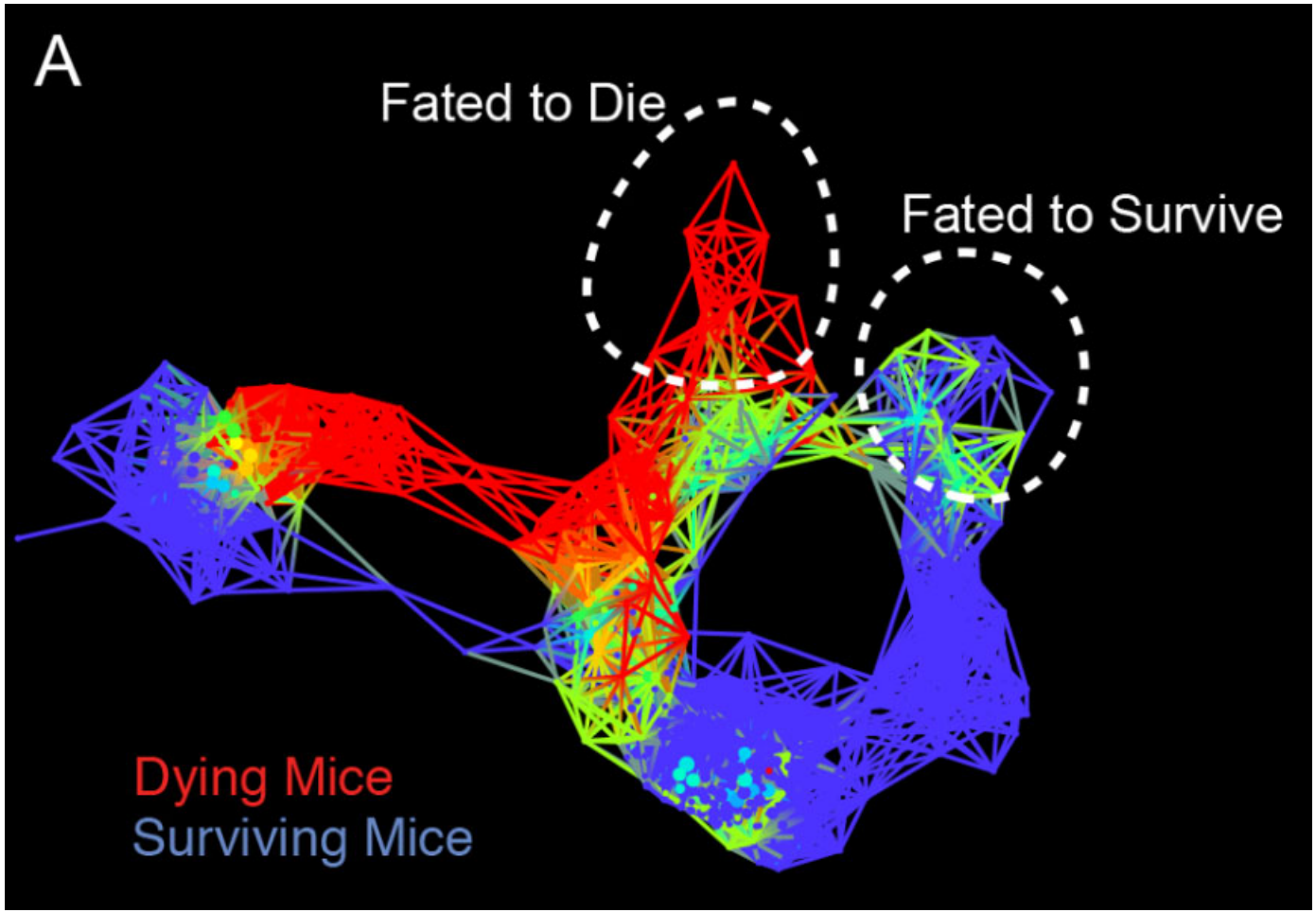
Infection Resilience

Torres et al. 2016 (PLoS Biology)



Infection Resilience

Torres et al. 2016 (PLoS Biology)



* resilient hosts take small loop thru disease space

Patient Trajectories

(Mahdobhi, ..., K, 2019)

* How does the agents and doses affect length of stay?

Patient Trajectories

(Mahdobhi, ..., K, 2019)

* How does the agents and doses affect length of stay?
* Duke Antimicrobial Stewardship (DASON)

Number of hospitals	25
Number of hospital unit-categories	9
Number of distinct patient-admission records	349, 610
Number of adult patients	334, 207
Number of male patients	148, 540
Number of female patients	201, 052
Average LOS per admission	7 days
Longest LOS	405 days
Number of antibacterials used	66
Most used antibacterial	Vancomycin
Average DOT per admission	6
Number of agent ranks	4
Most used agent rank	rank 3

Patient Trajectories

(Mahdobhi, ..., K, 2019)

* How does the agents and doses affect length of stay?
* Duke Antimicrobial Stewardship (DASON)

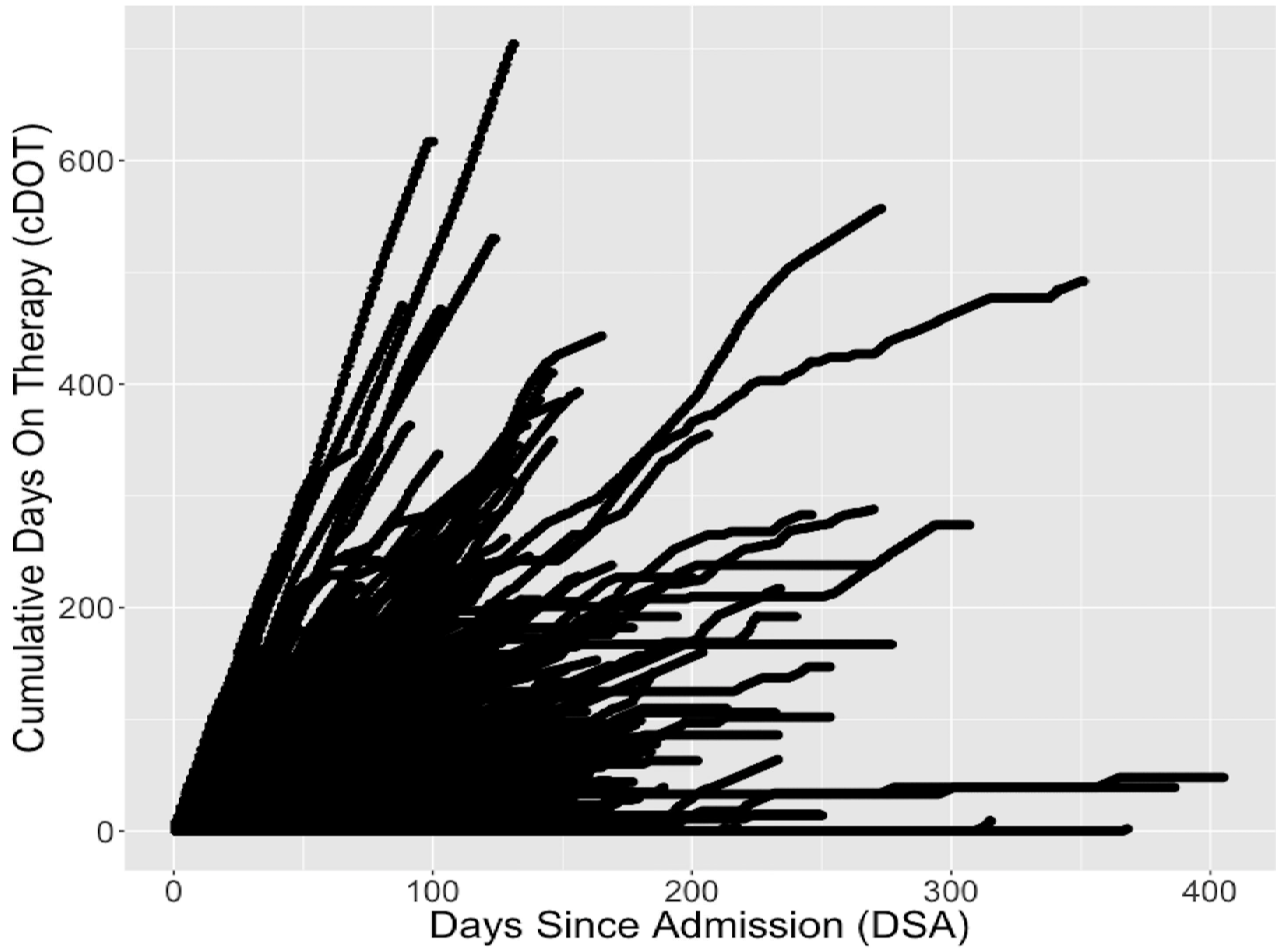
Number of hospitals	25
Number of hospital unit-categories	9
Number of distinct patient-admission records	349, 610
Number of adult patients	334, 207
Number of male patients	148, 540
Number of female patients	201, 052
Average LOS per admission	7 days
Longest LOS	405 days
Number of antibacterials used	66
Most used antibacterial	Vancomycin
Average DOT per admission	6
Number of agent ranks	4
Most used agent rank	rank 3

cDOT vs. DSA

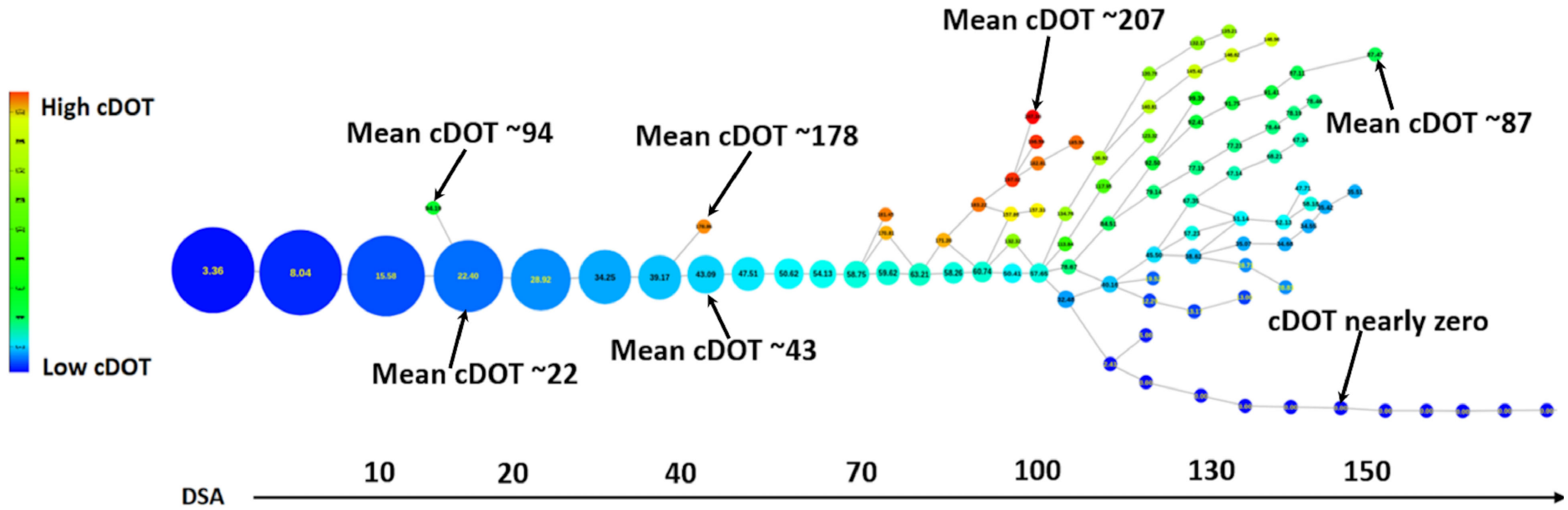
* cDOT: cumulative Days On Therapy
  DSA: Days Since Admission

cDOT vs. DSA

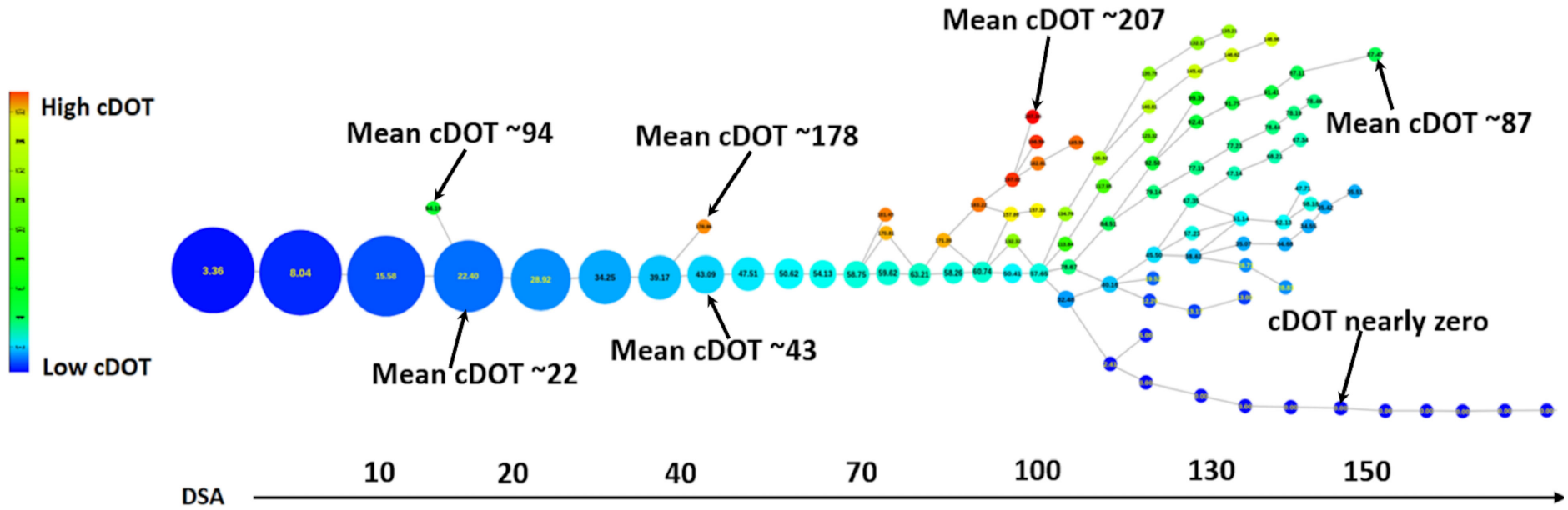
* cDOT : cumulative Days On Therapy
  DSA : Days Since Admission

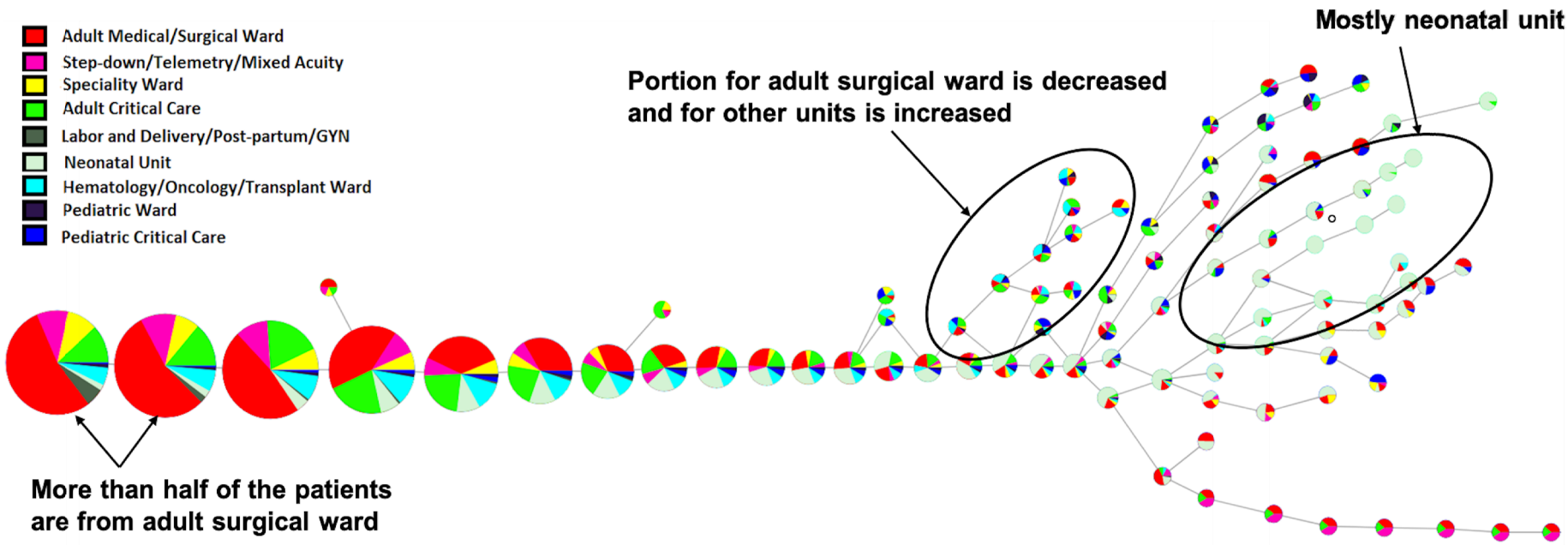


TDA: cDOT, DSA

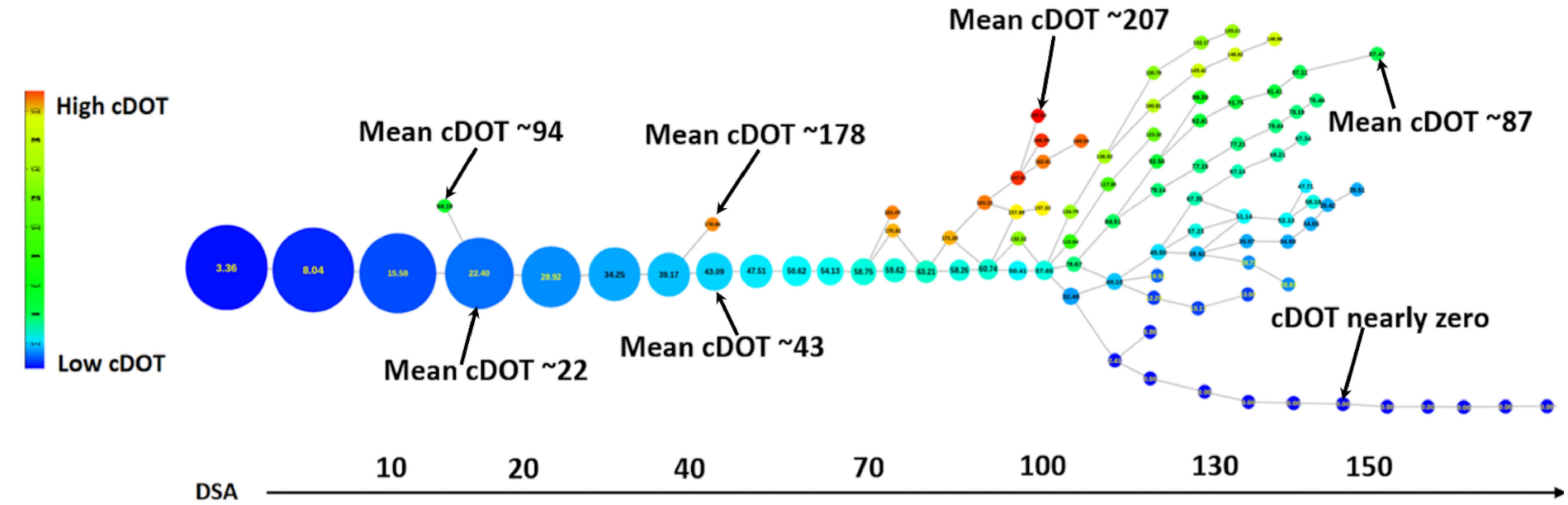

High cDOT
Low cDOT
Mean cDOT ~94
Mean cDOT ~178
Mean cDOT ~207
Mean cDOT ~87
Mean cDOT ~22
Mean cDOT ~43
cDOT nearly zero
3.36
8.04
15.58
22.40
28.92
34.25
39.17
43.09
47.51
50.62
54.13
58.75
59.62
63.21
58.26
60.74
60.41
47.49
52.49
10
20
40
70
100
130
150
DSA

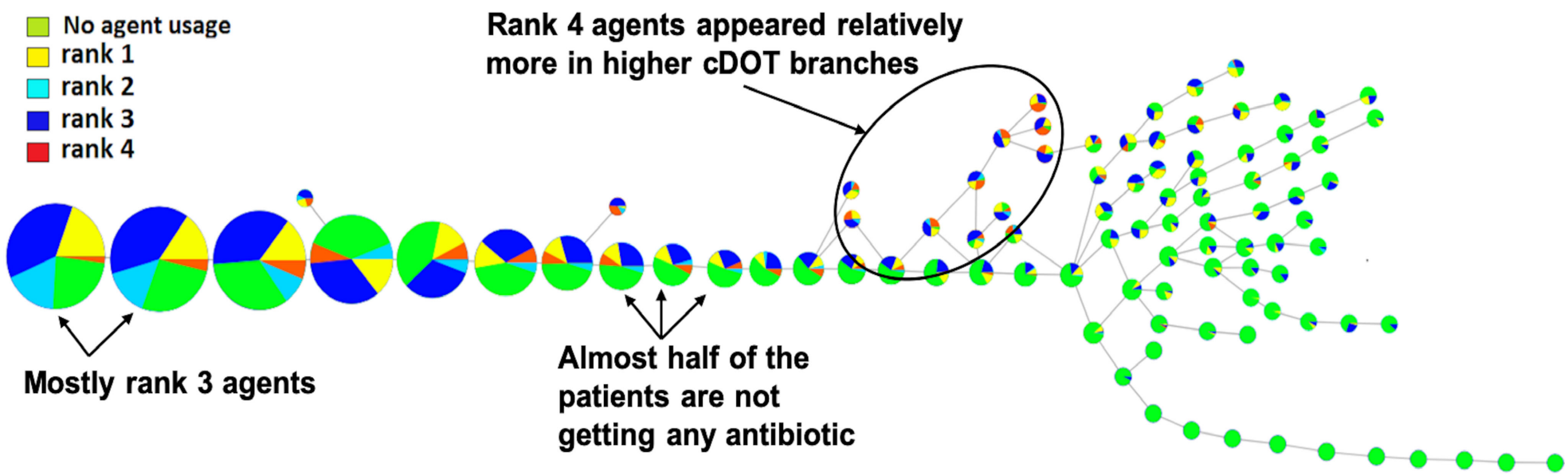
TDA: cDOT, DSA, Units



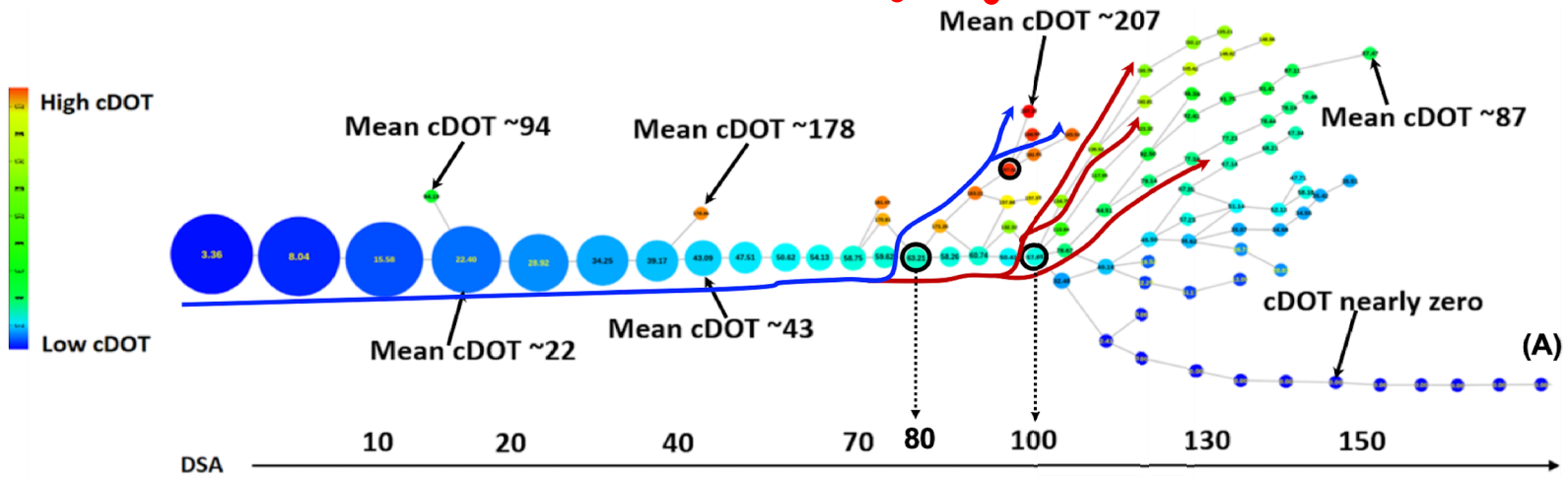


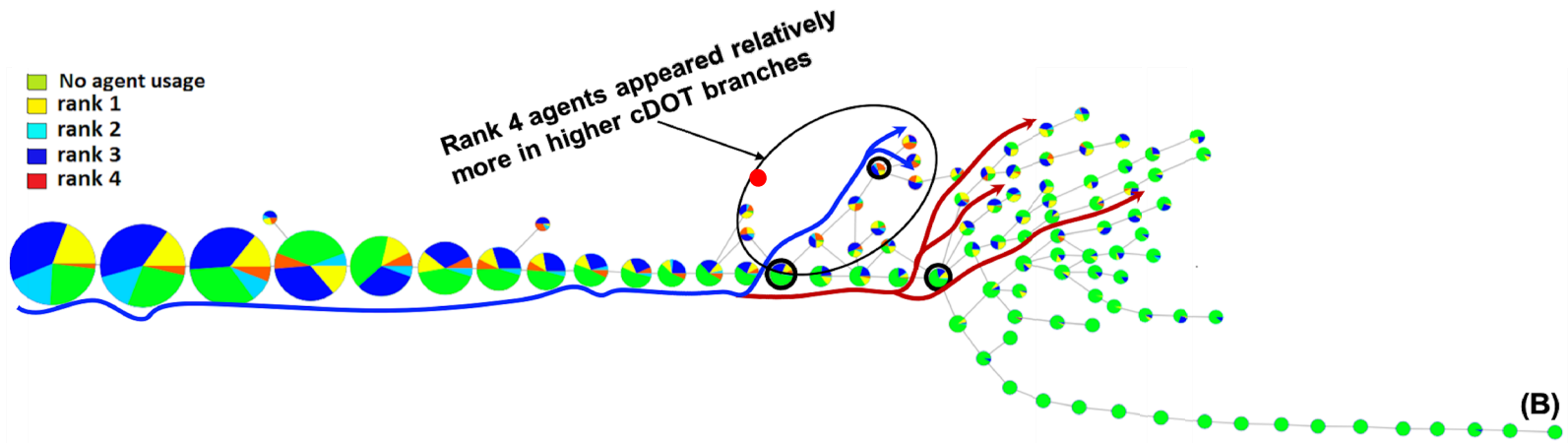
TDA: cDOT, DSA, AGENTS

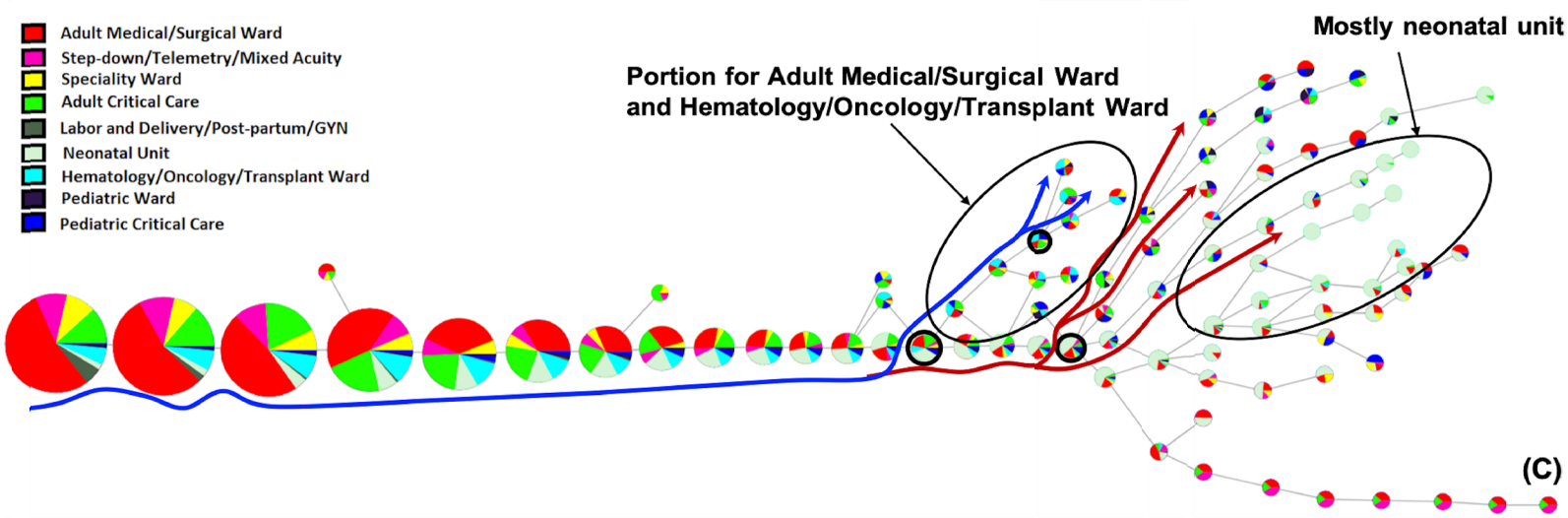




TDA: FLARES


High cDOT
Low cDOT
Mean cDOT ~94
Mean cDOT ~178
Mean cDOT ~207
Mean cDOT ~87
Mean cDOT ~22
Mean cDOT ~43
cDOT nearly zero
(A)
DSA
10
20
40
70
80
100
130
150


No agent usage
rank 1
rank 2
rank 3
rank 4
Rank 4 agents appeared relatively more in higher cDOT branches
(B)


Adult Medical/Surgical Ward
Step-down/Telemetry/Mixed Acuity
Speciality Ward
Adult Critical Care
Labor and Delivery/Post-partum/GYN
Neonatal Unit
Hematology/Oncology/Transplant Ward
Pediatric Ward
Pediatric Critical Care
Portion for Adult Medical/Surgical Ward and Hematology/Oncology/Transplant Ward
Mostly neonatal unit
(C)

TDA: The Takeaway

✓ versatile tool for "partial clustering"

TDA: The Takeaway

- ✓ versatile tool for "partial clustering"
- ✓ insightful visualizations of data

TDA: The Takeaway

- ✓ versatile tool for "partial clustering"
- ✓ insightful visualizations of data
- ✓ software available (Scikit-TDA, GUDHI, ...)

open source!

TDA: The Takeaway

- ✓ versatile tool for "partial clustering"
- ✓ insightful visualizations of data
- ✓ software available (Scikit-TDA, GUDHI, ...)

  open source!

- ● may not help in every case...

TDA: The Takeaway

- ✓ versatile tool for "partial clustering"
- ✓ insightful visualizations of data
- ✓ software available (Scikit-TDA, GUDHI, ...)
  open source!
- ● may not help in every case...
- ✓ add to your data analysis toolbox

TDA: The Takeaway

- ✓ versatile tool for "partial clustering"
- ✓ insightful visualizations of data
- ✓ software available (Scikit-TDA, GUDHI, ...)
  open source!
- ● may not help in every case...
- ✓ add to your data analysis toolbox

Thank You!