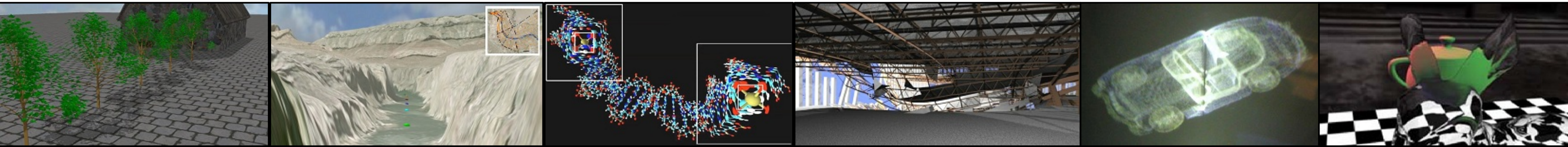
Using Topological Data Analysis to Improve Data Visualization

Paul Rosen

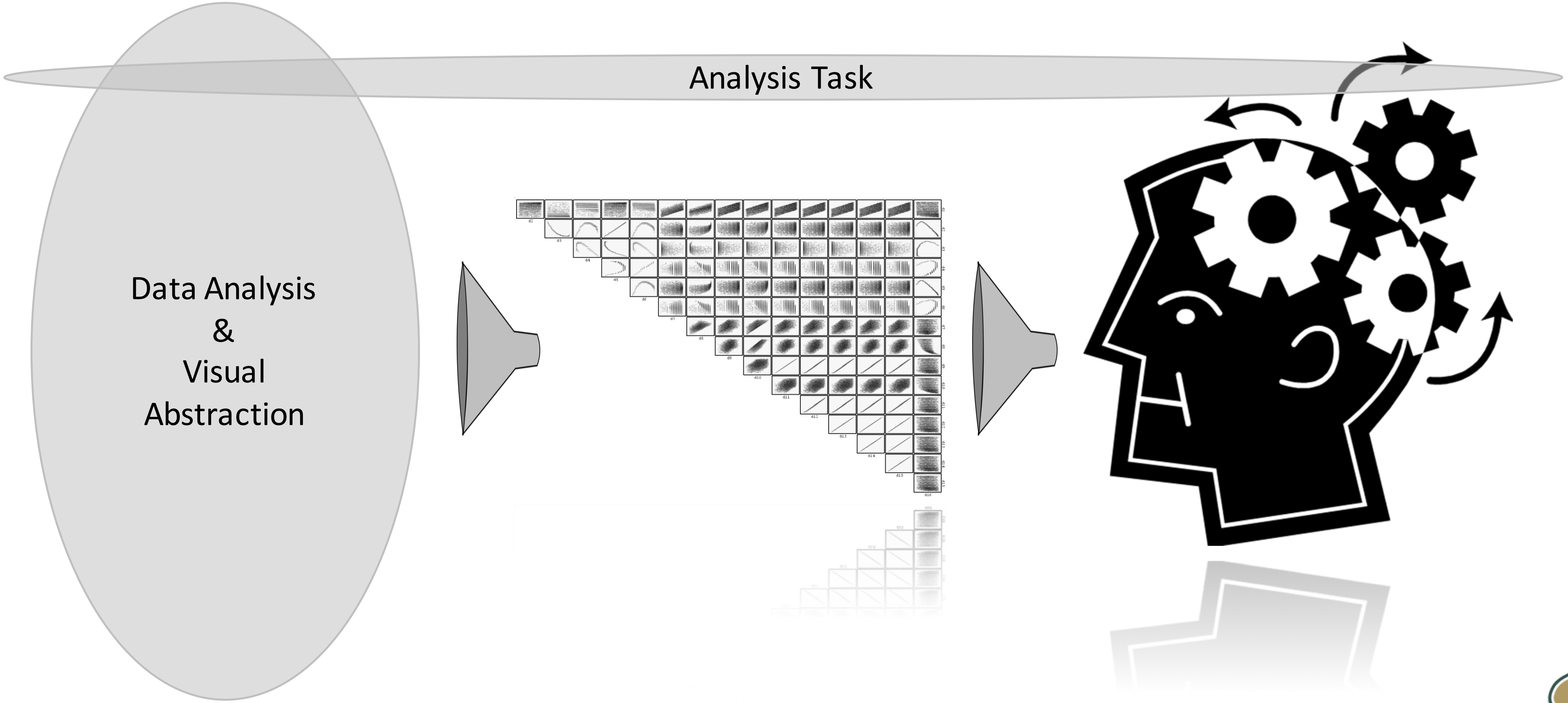
Assistant Professor

Computer Science and Engineering

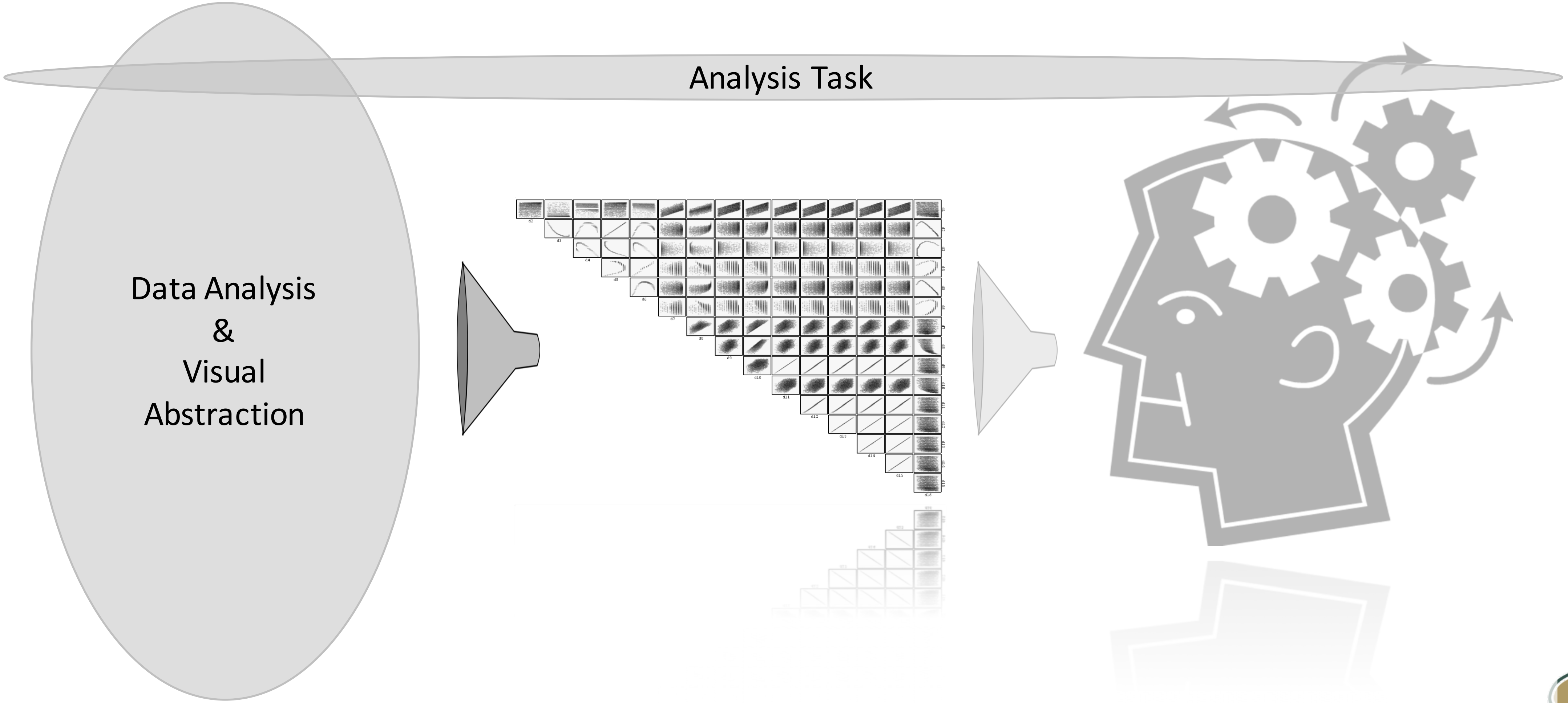
University of South Florida

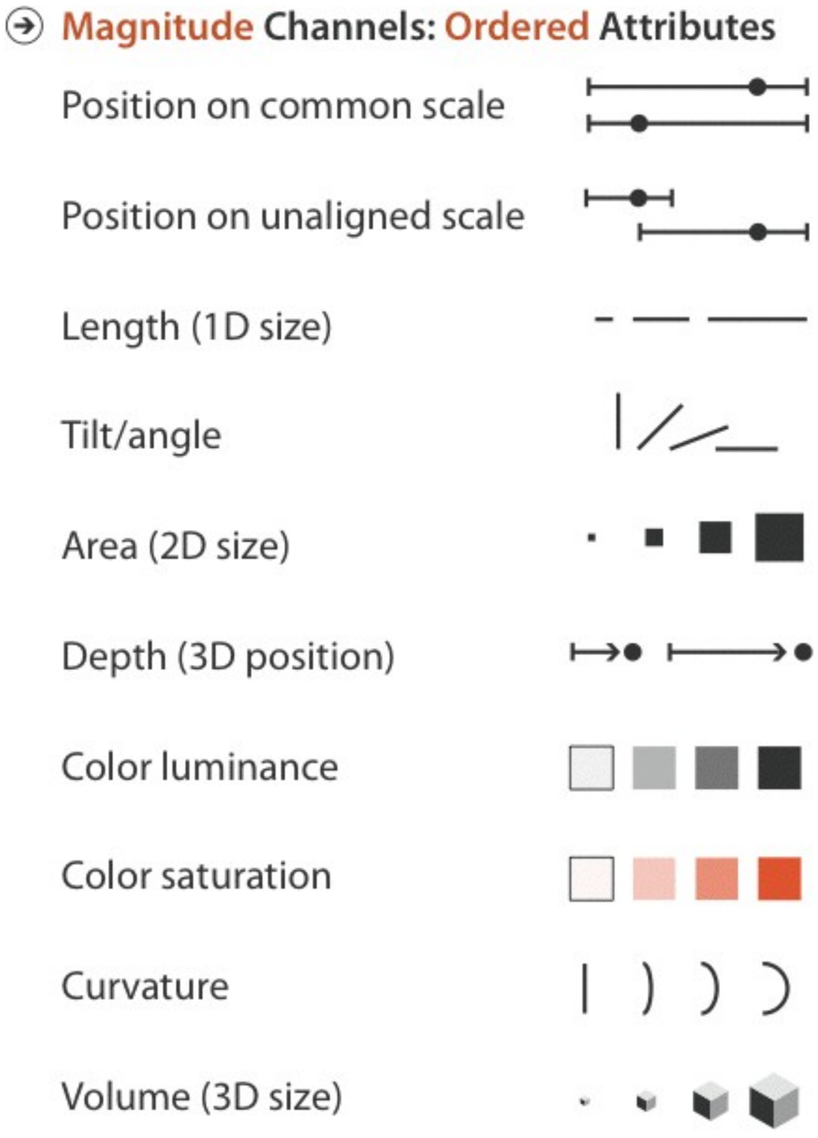


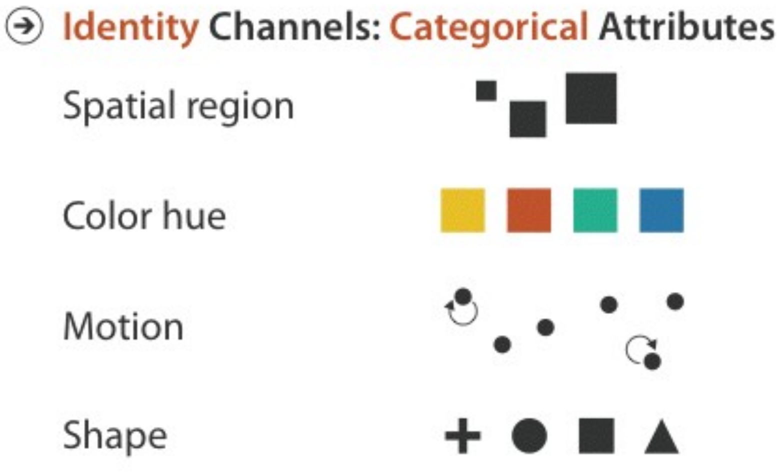
WHAT WE DO IN VISUALIZATION



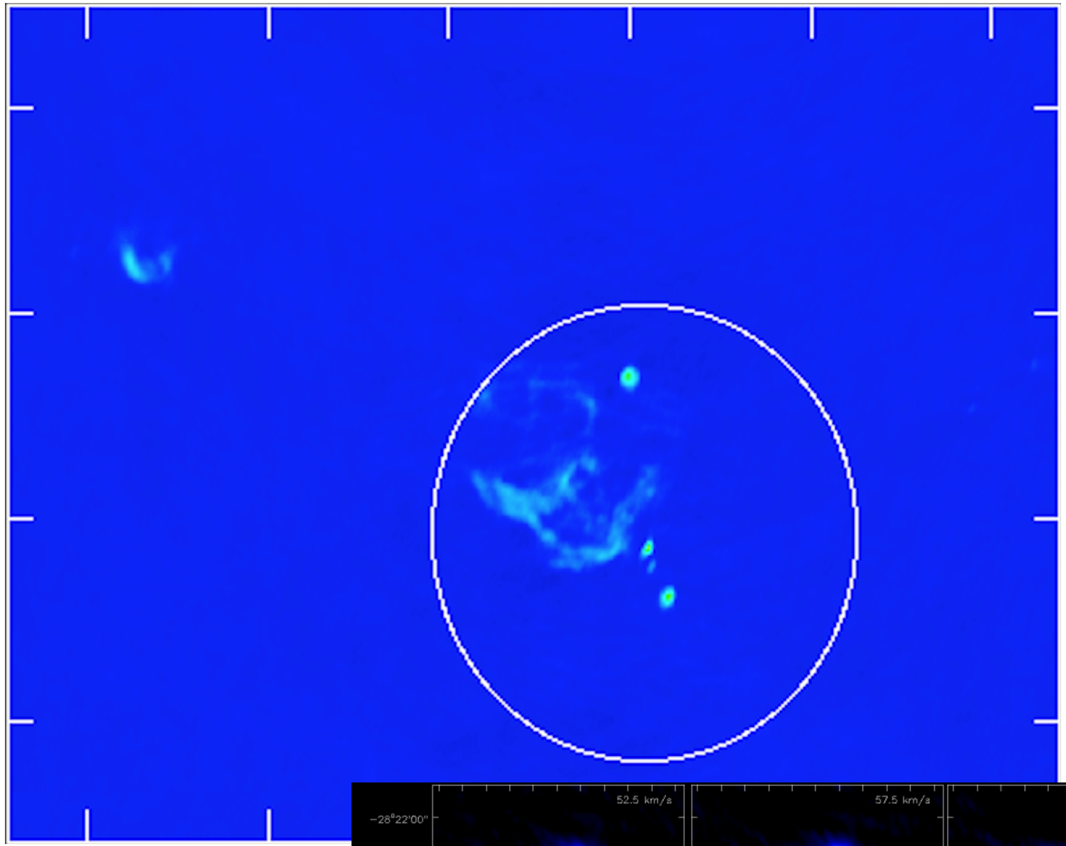
WHAT WE DO IN VISUALIZATION

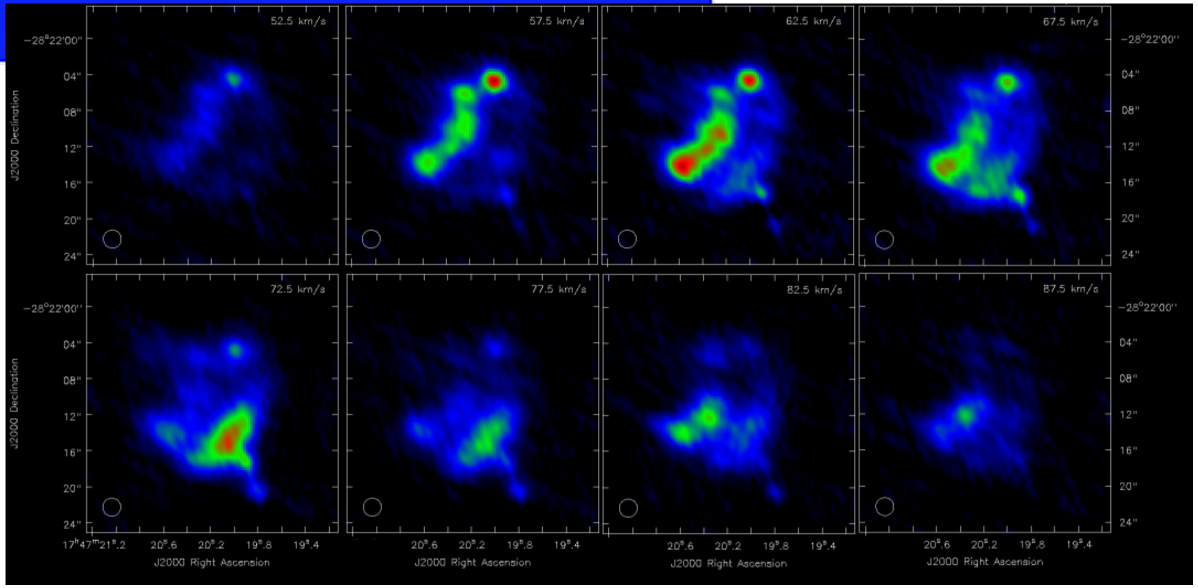


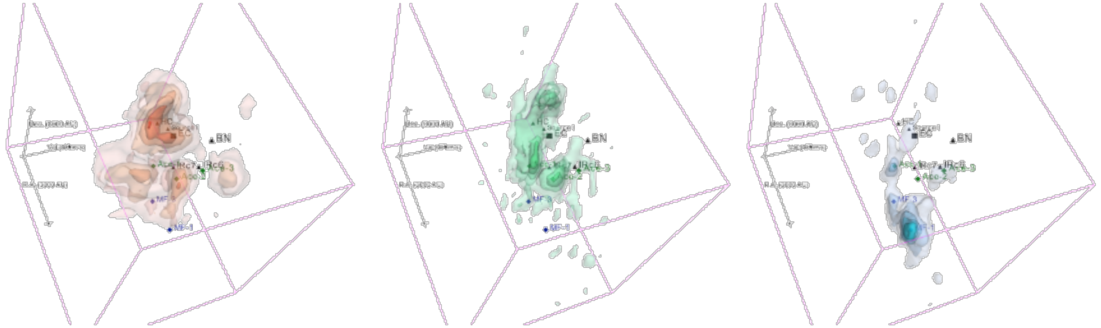


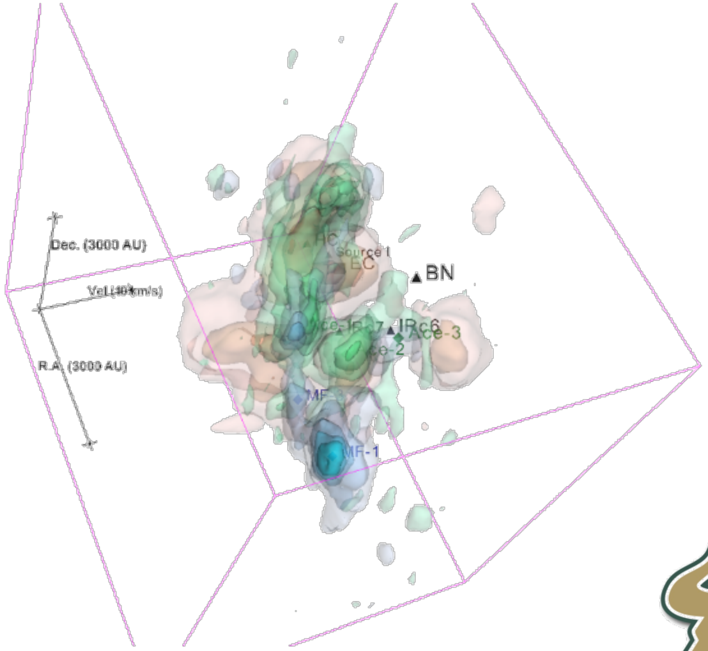


Visual Encoding Channels

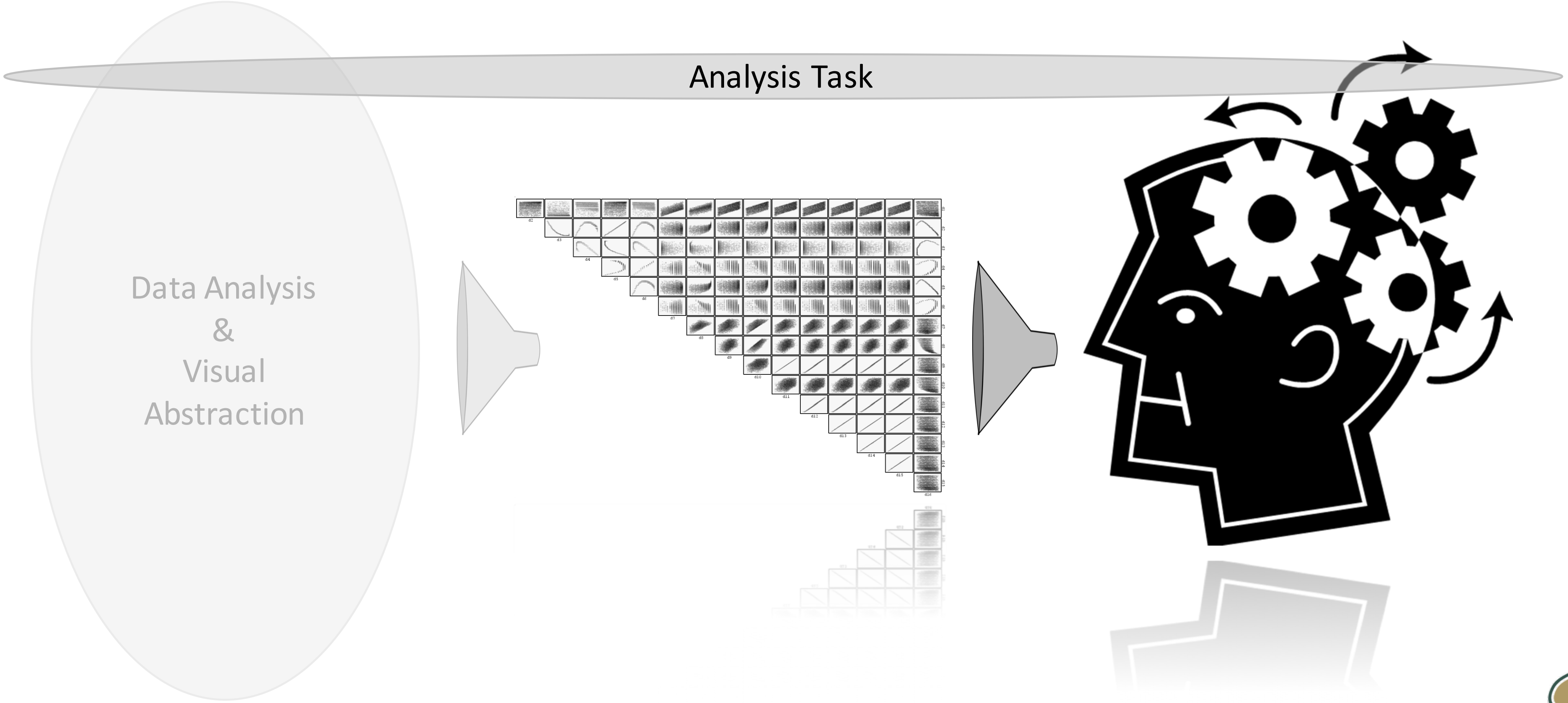



52.5 km/s
57.5 km/s
62.5 km/s
67.5 km/s
72.5 km/s
77.5 km/s
82.5 km/s
87.5 km/s
J2000 Declination
J2000 Right Ascension

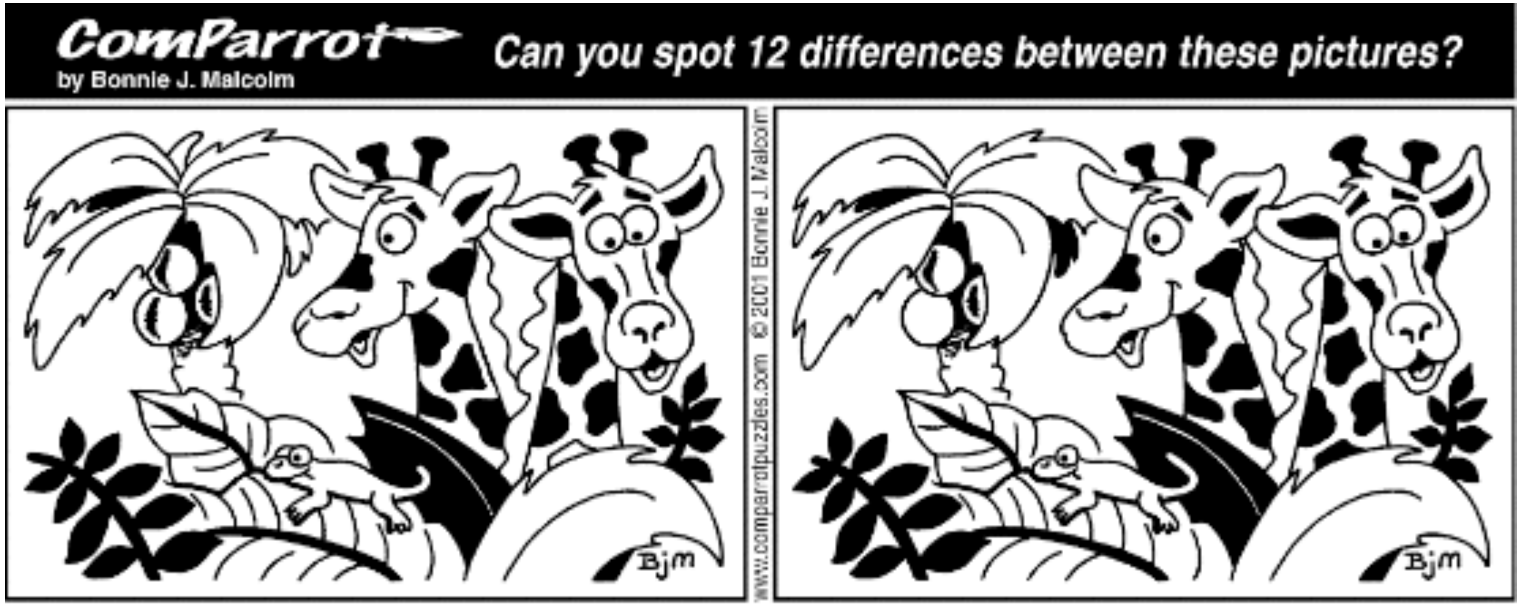



Dec. (3000 AU)
R.A. (3000 AU)
BN
Source I
IRc6
MF-1

WHAT WE DO IN VISUALIZATION

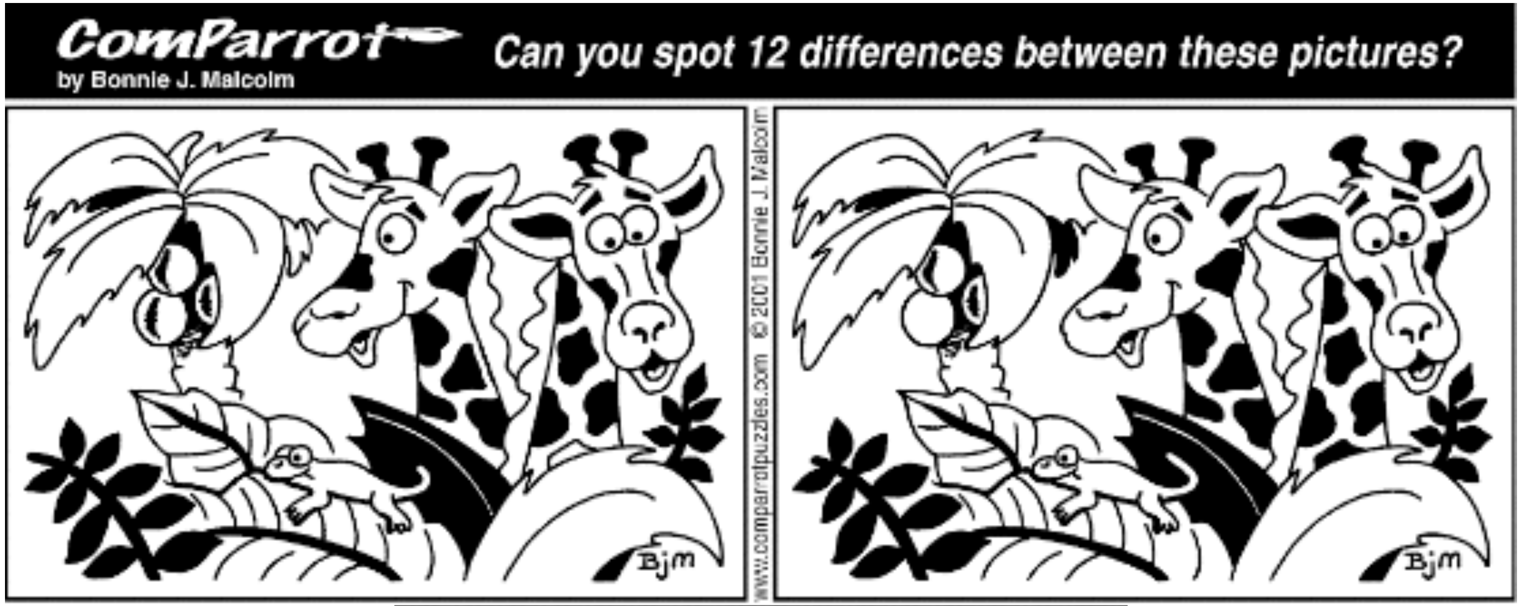


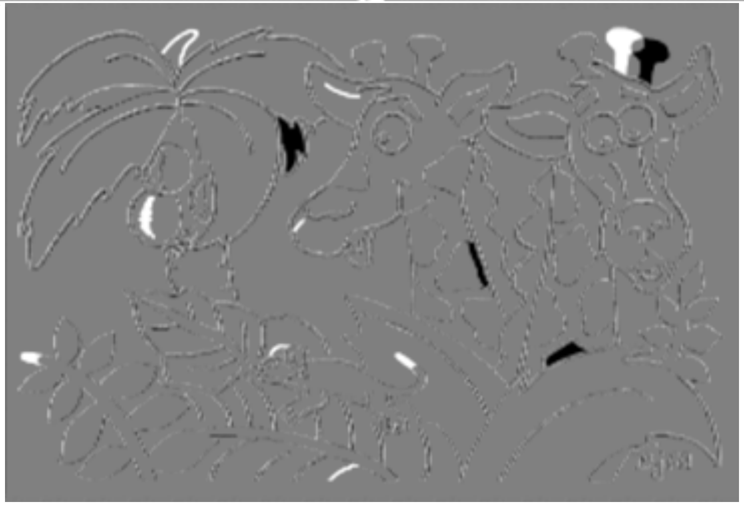
A LIMITED VISUAL SYSTEM

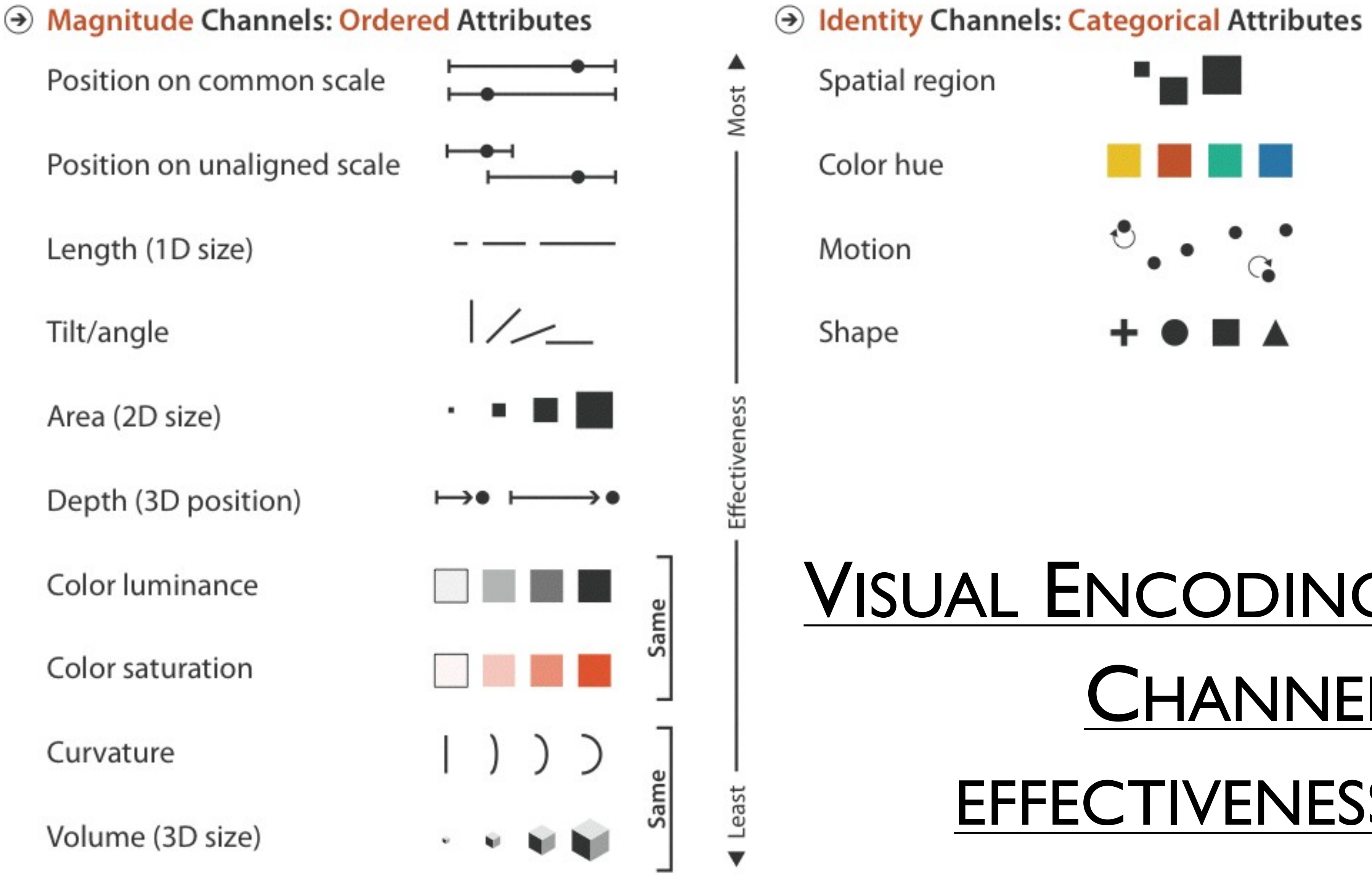


MALCOLM 2001

A LIMITED VISUAL SYSTEM







VISUAL ENCODING CHANNEL EFFECTIVENESS

Perceptual Issues

Physical construction of perceptual machinery results in many limitations and considerations, including:

- Foviation
- Blind spot
- Mach Banding
- Simultaneous contrast
- Contrast Sensitivity
- Popout, Conjunction, & Gestalt Principles
- Colorblindness
- Color Relativity
- Color Distinguishability
- Separability in Color Blending
- Limited 3D capacity
- Change blindness

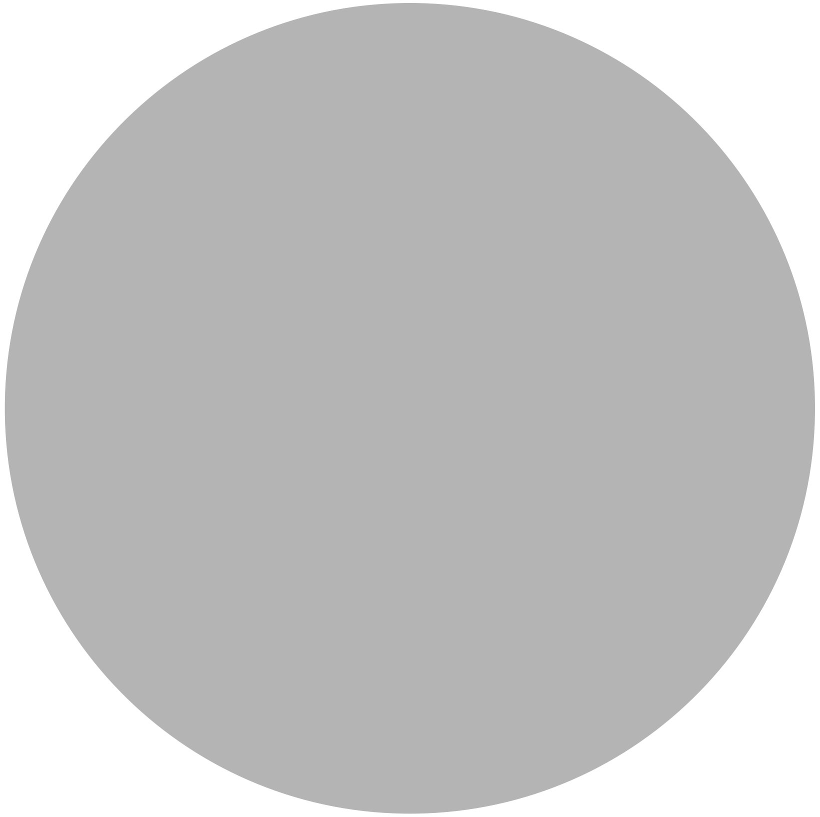
Perceptual Issues

Physical construction of perceptual machinery results in many limitations and considerations, including:

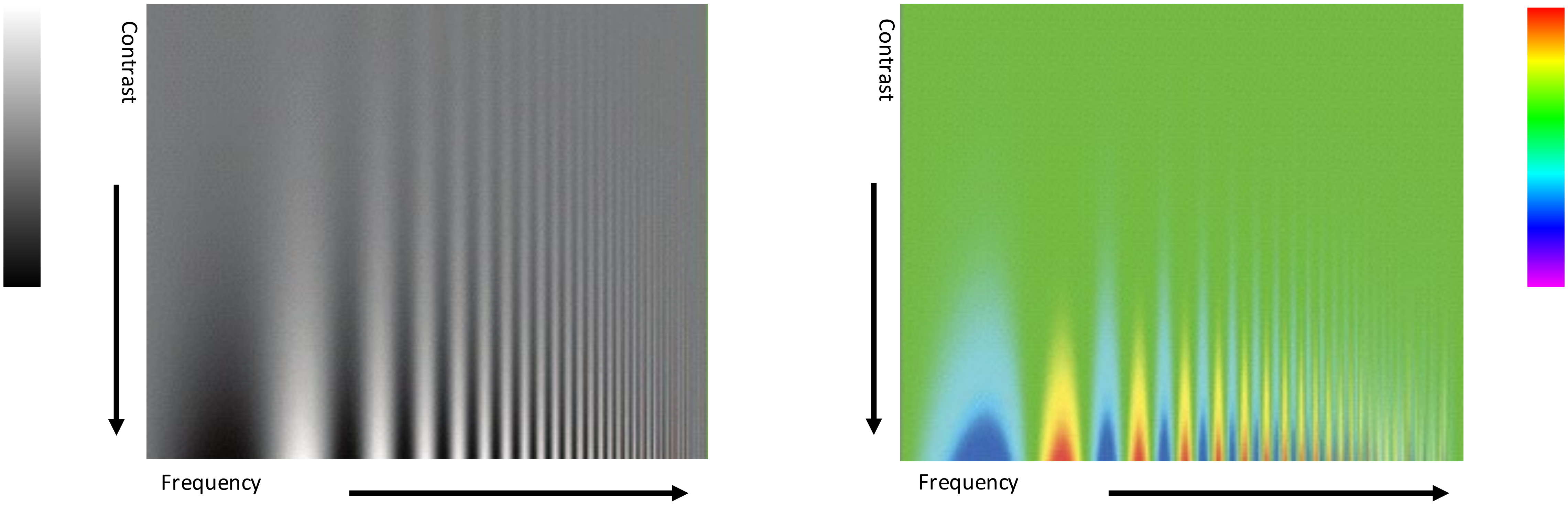
Foviation
Blind spot
Mach Banding
Simultaneous contrast
Contrast Sensitivity
Popout, Conjunction, & Gestalt Principles
Colorblindness
Color Relativity
Color Distinguishability
Separability in Color Blending
Limited 3D capacity
Change blindness

SIMULTANEOUS CONTRAST

SIMULTANEOUS CONTRAST

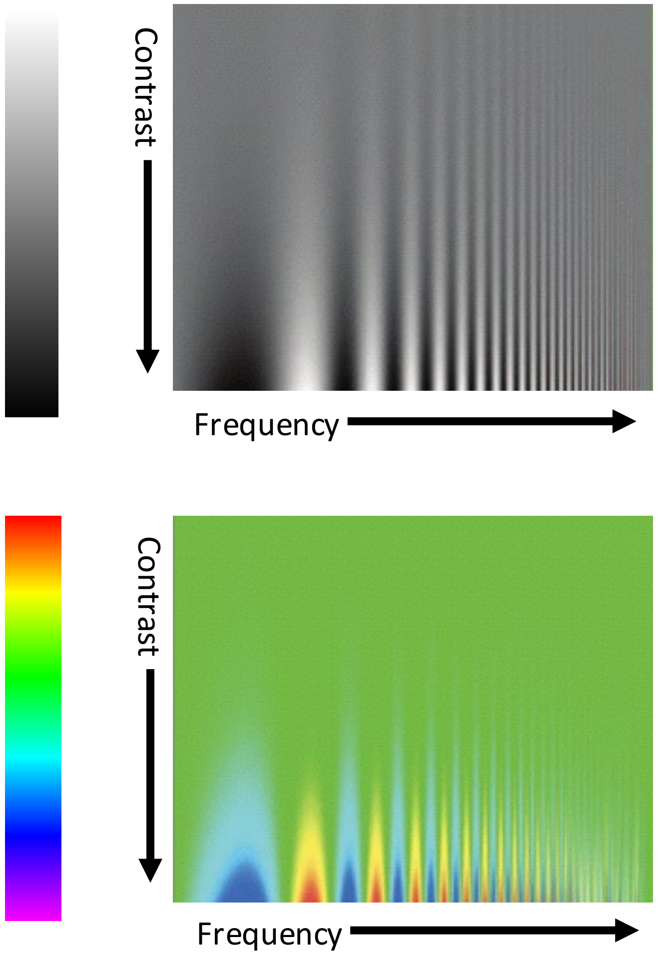


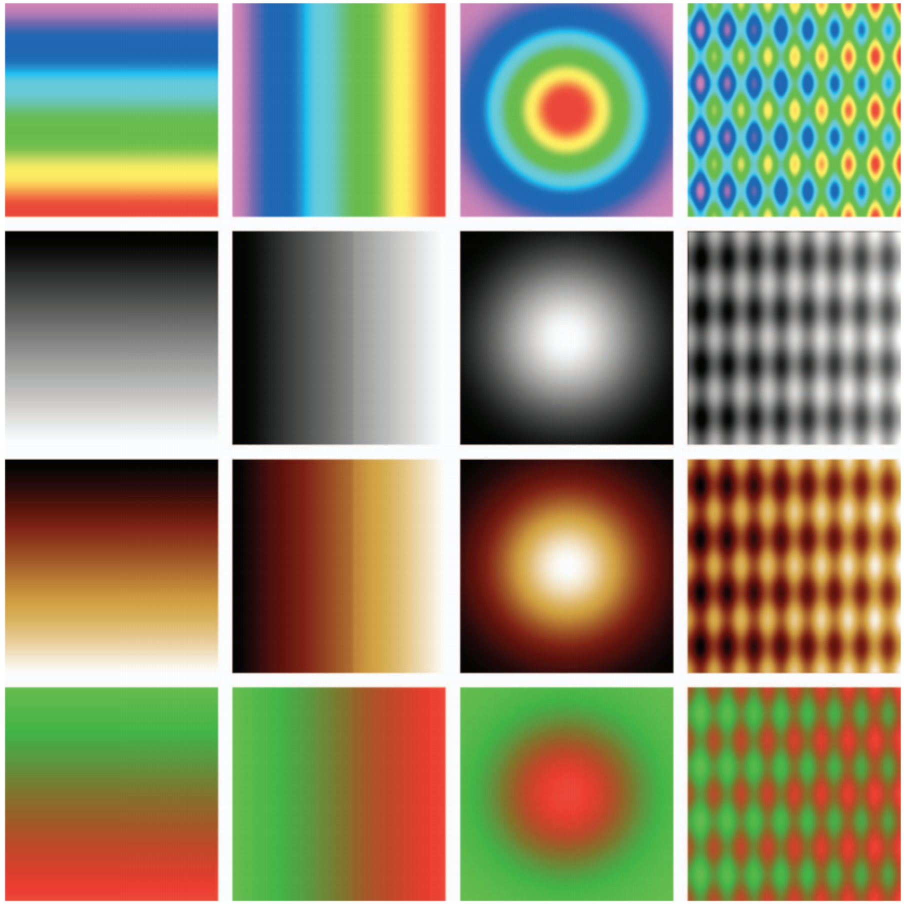
CONTRAST SENSITIVITY



BORLAND 2007

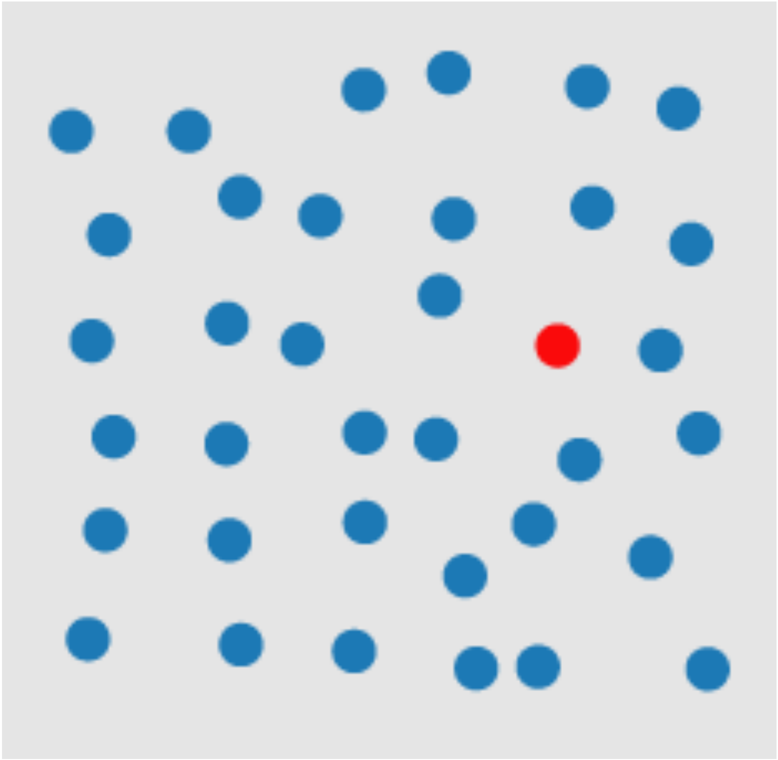
Contrast Sensitivity





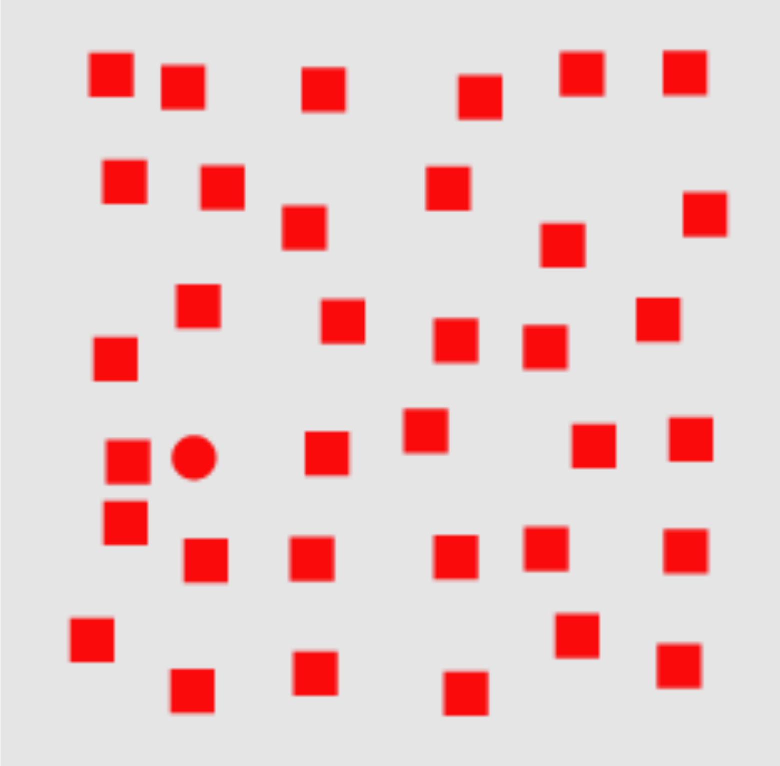
Borland 2007

POPOUT—PICK THE OUTLIER

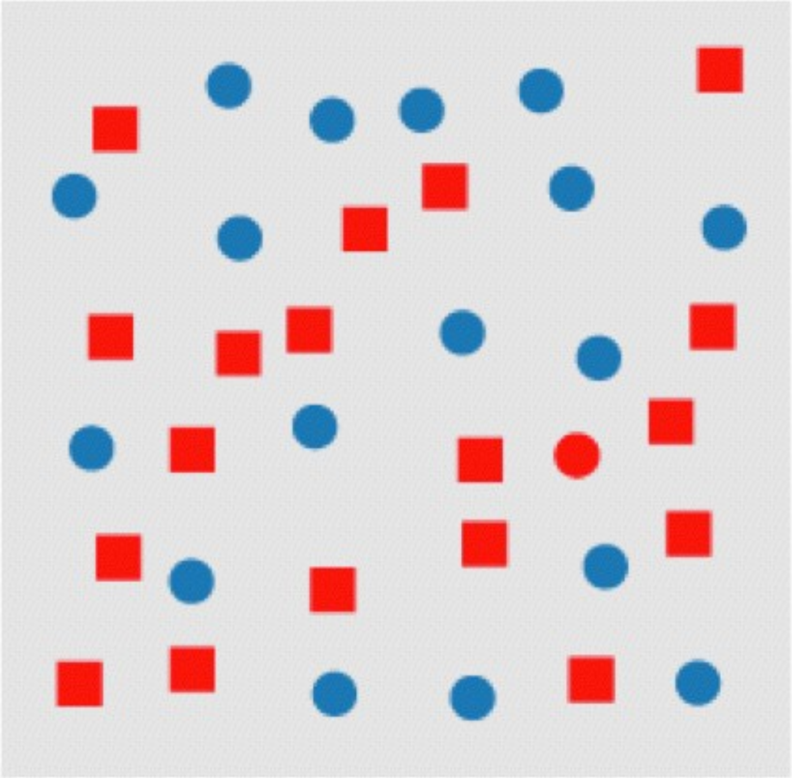


HEALEY 2007

Popout—Pick the outlier



CONJUNCTION



HEALEY 2007

LIMITED VISUAL COMPREHENSION

Overestimation of capacity

Wide variance in capacity

Undefinable "optimal" visualization

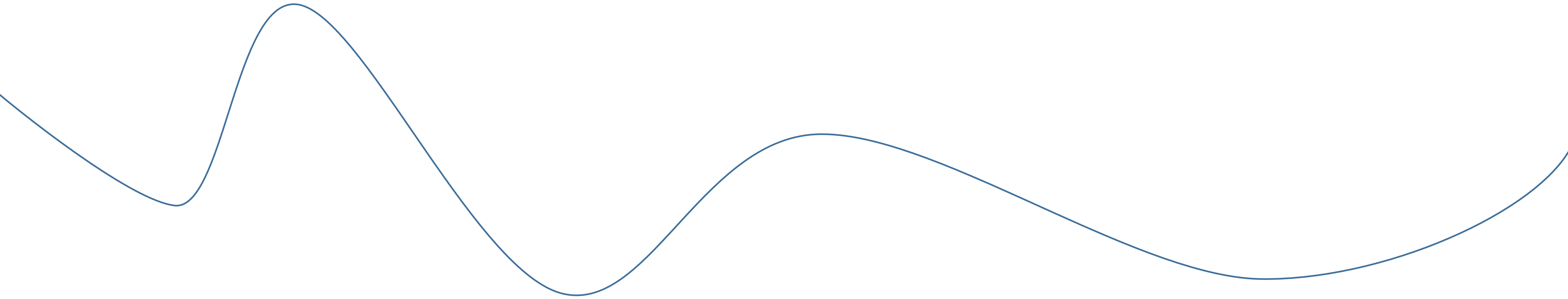
Visualization Challenge

with very bandwidth limited—we need to compress the data while maintaining features salient to analysis

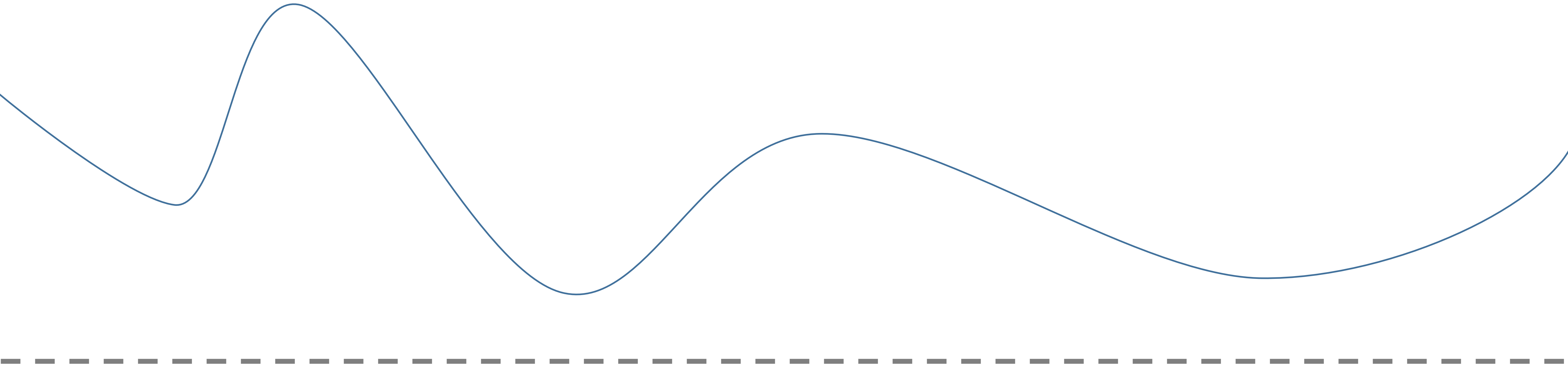
TOPOLOGICAL DATA ANALYSIS AND VISUALIZATION

study of approaches to EXTRACT structure from NOISY or COMPLEX data and REPRESENT that data in an actionable form

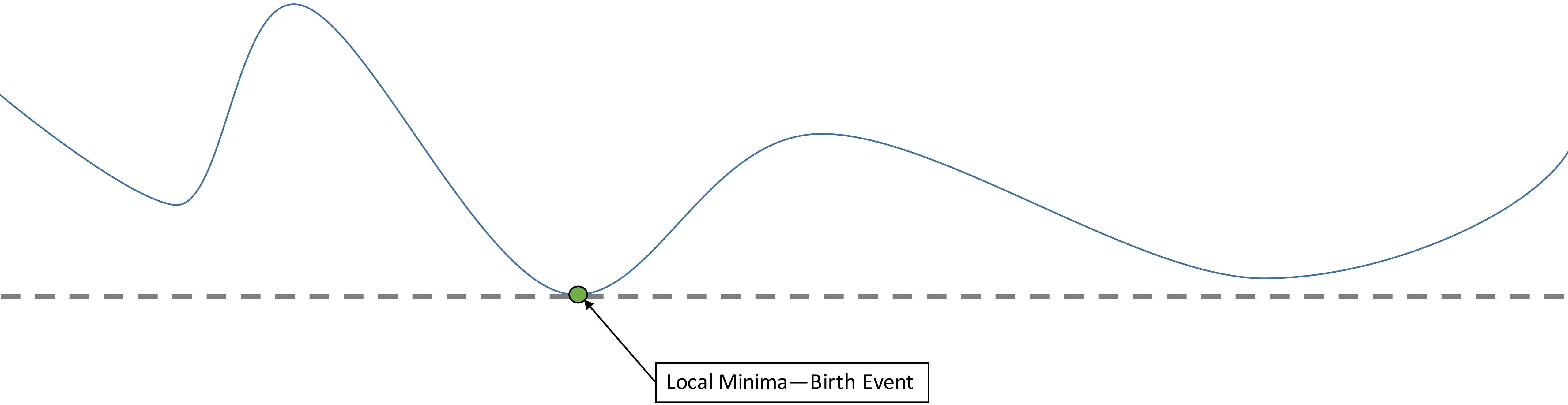
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



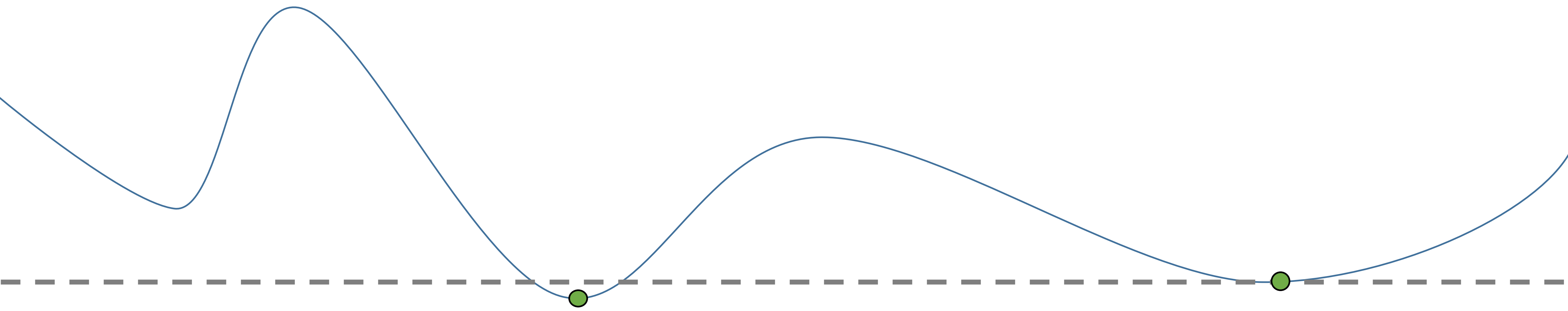
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



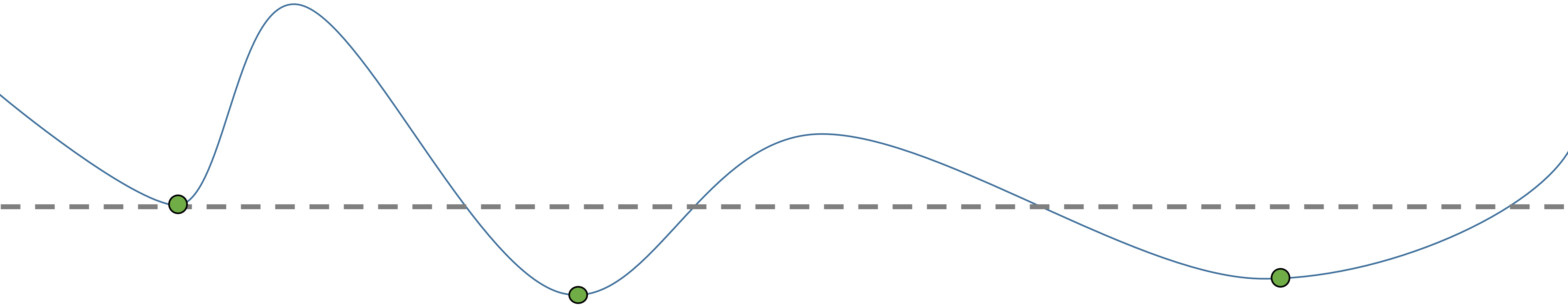
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



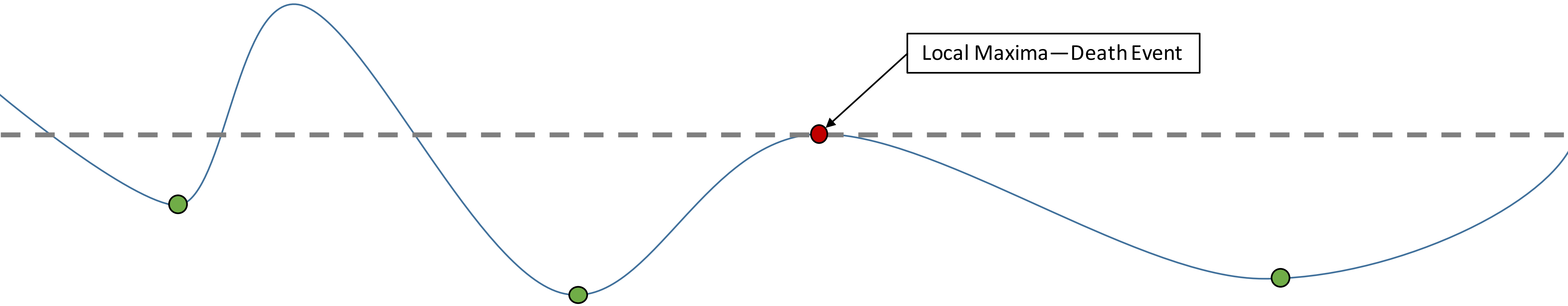
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



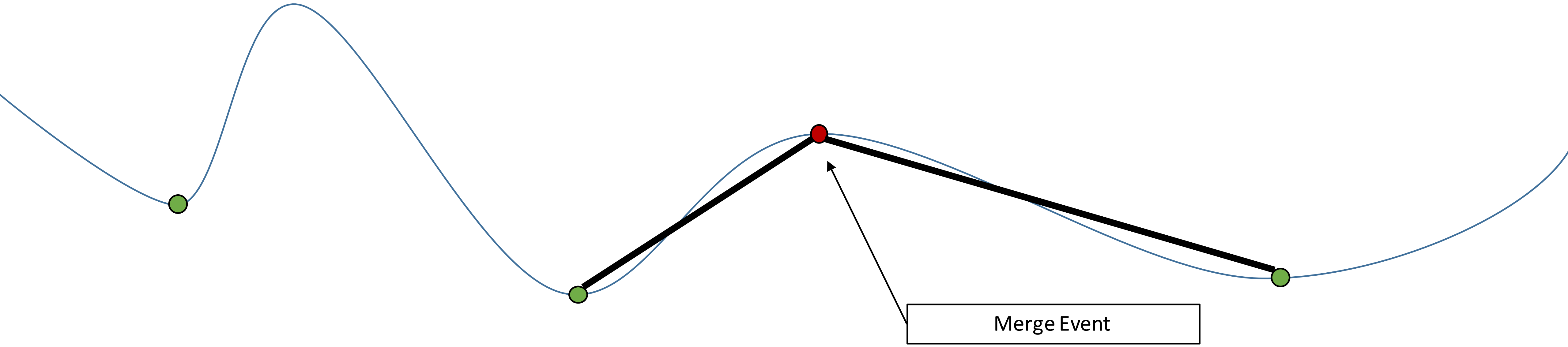
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



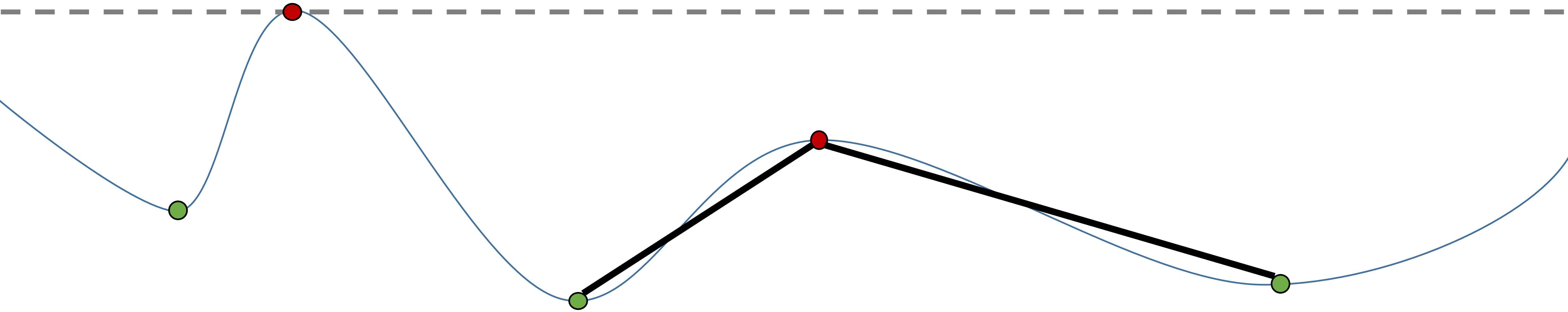
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



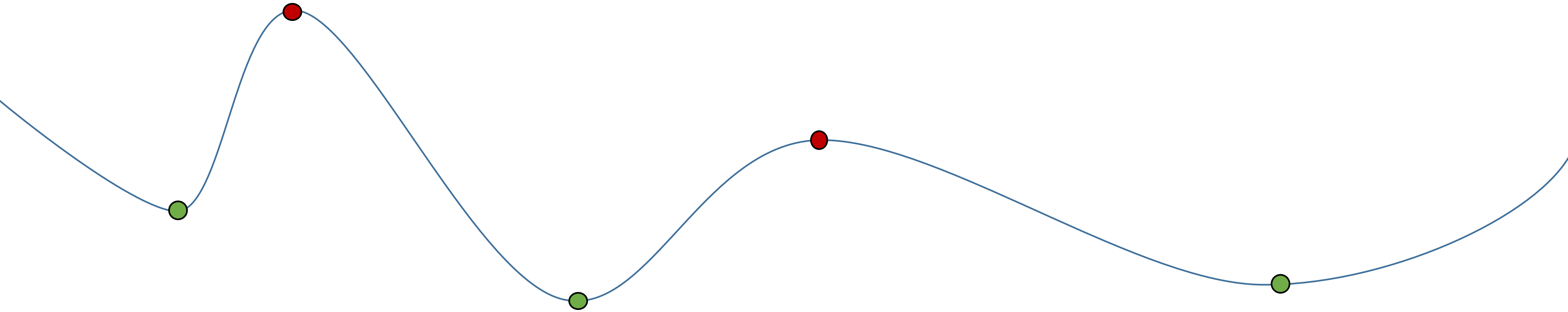
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



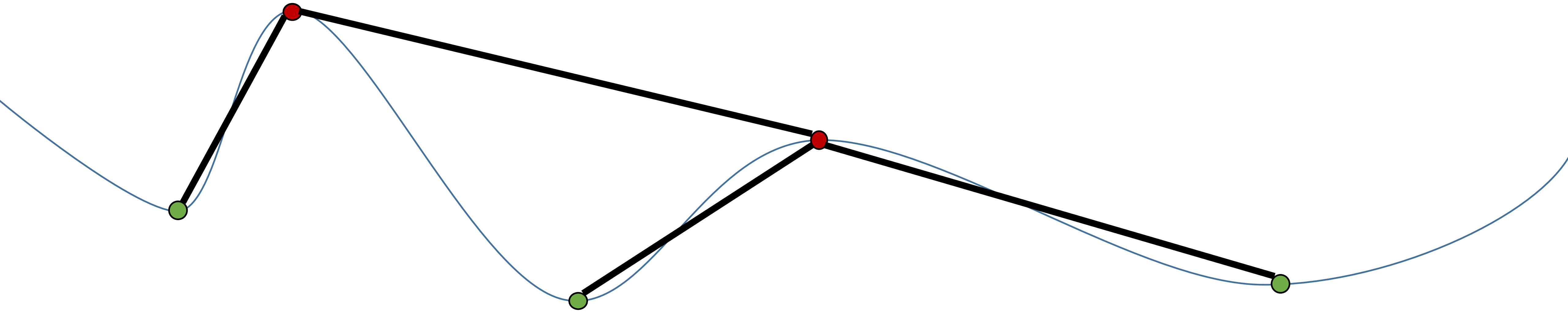
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



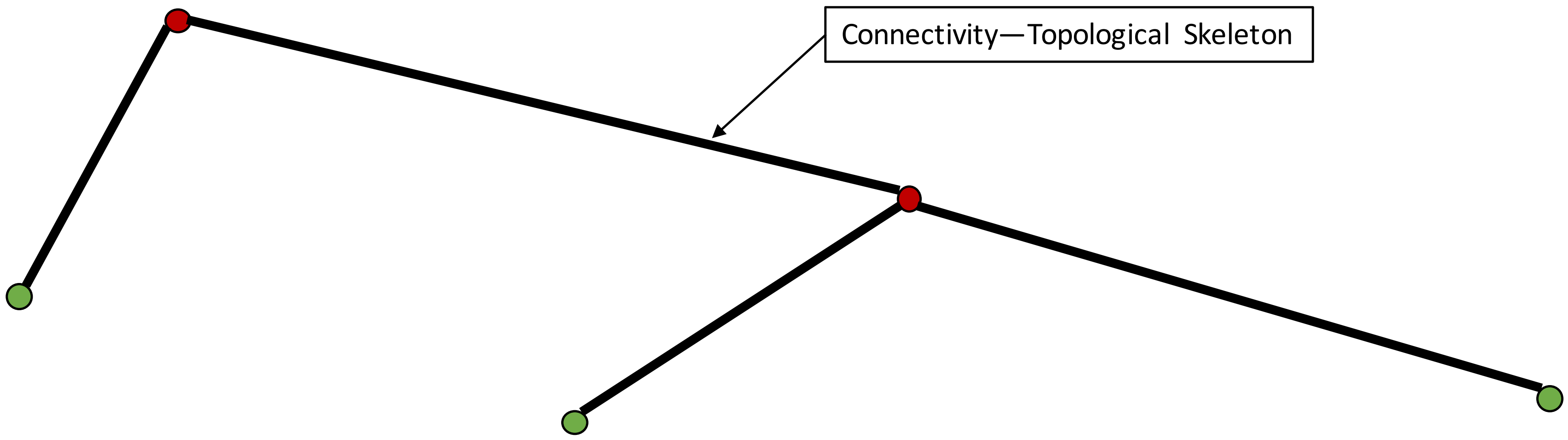
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



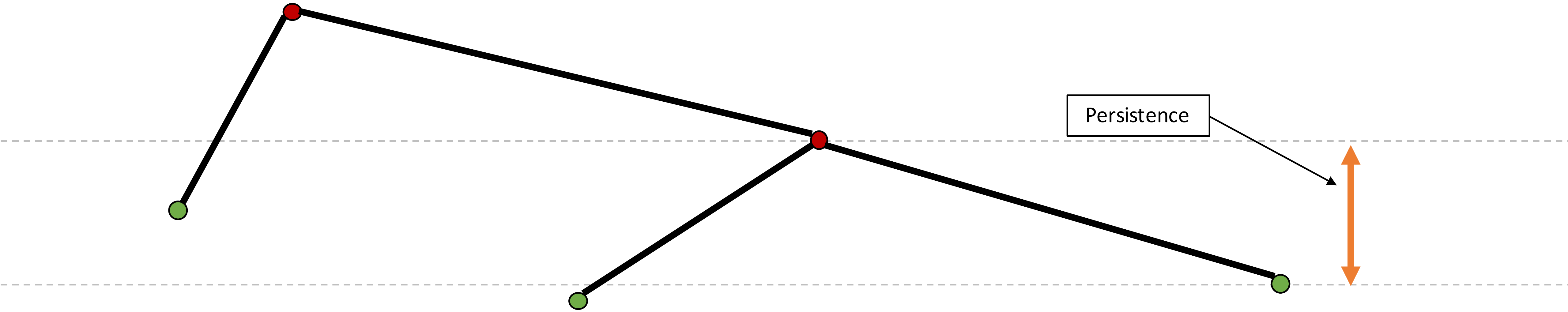
A standard example of TDA—the merge tree



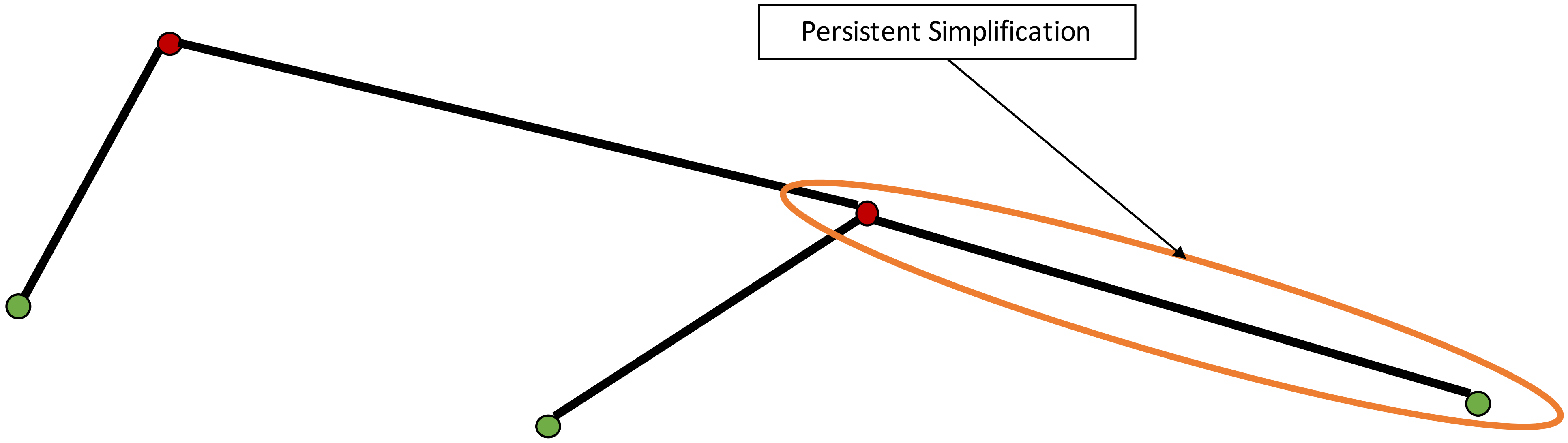
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



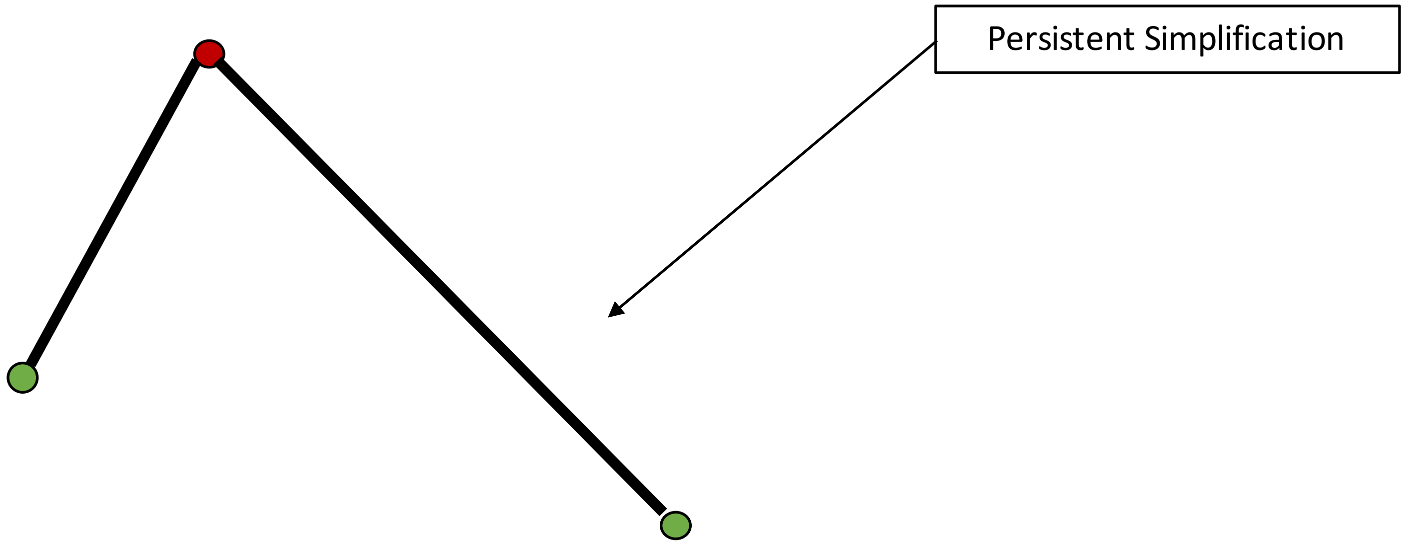
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



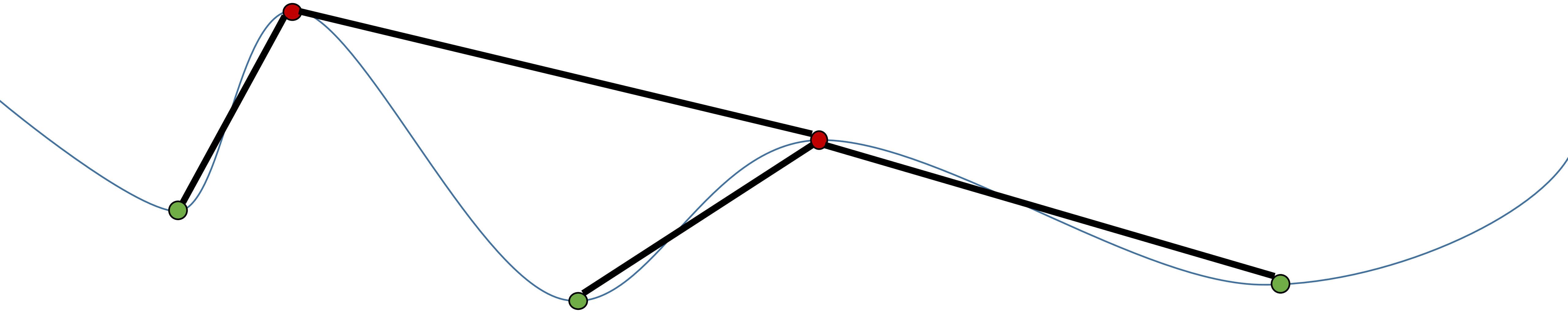
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



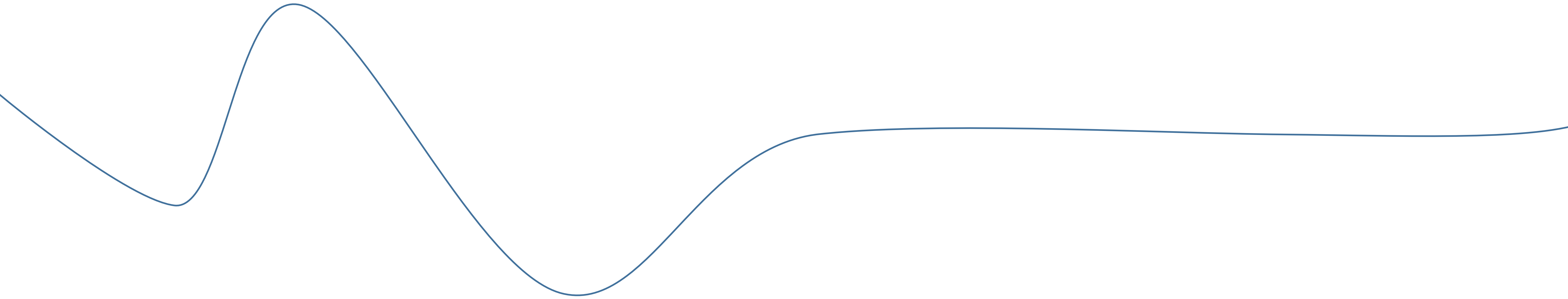
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



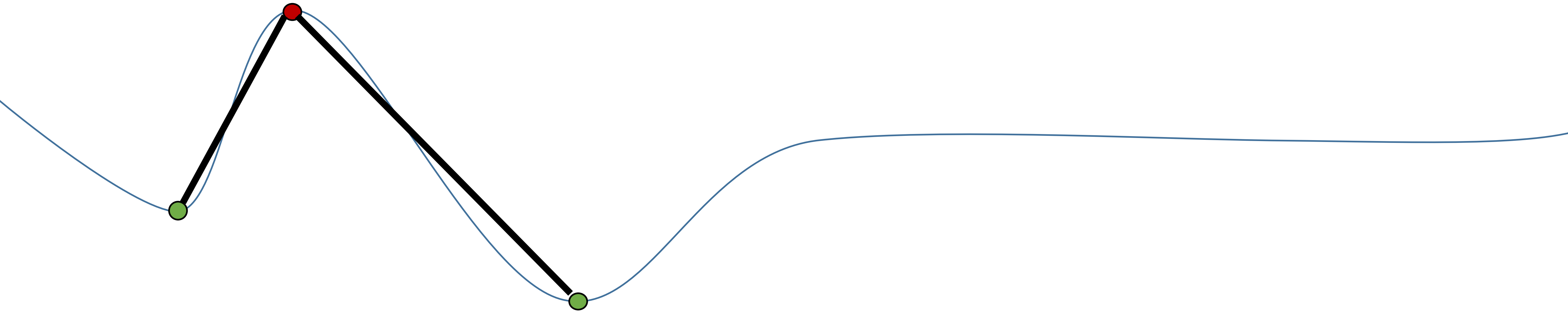
A STANDARD EXAMPLE OF TDA—THE MERGE TREE



A standard example of TDA—the merge tree



A STANDARD EXAMPLE OF TDA—THE MERGE TREE



So, what does TDA give us?

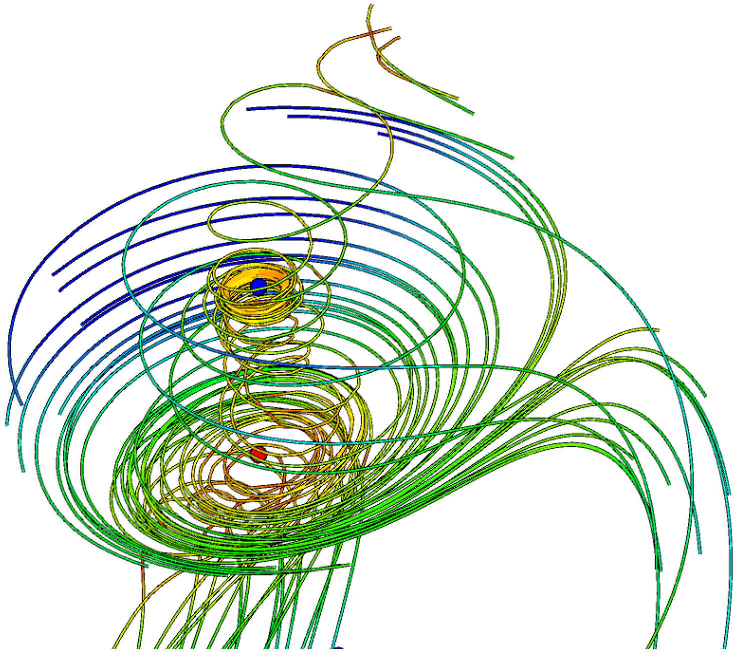
Multiscale skeleton of features in the data

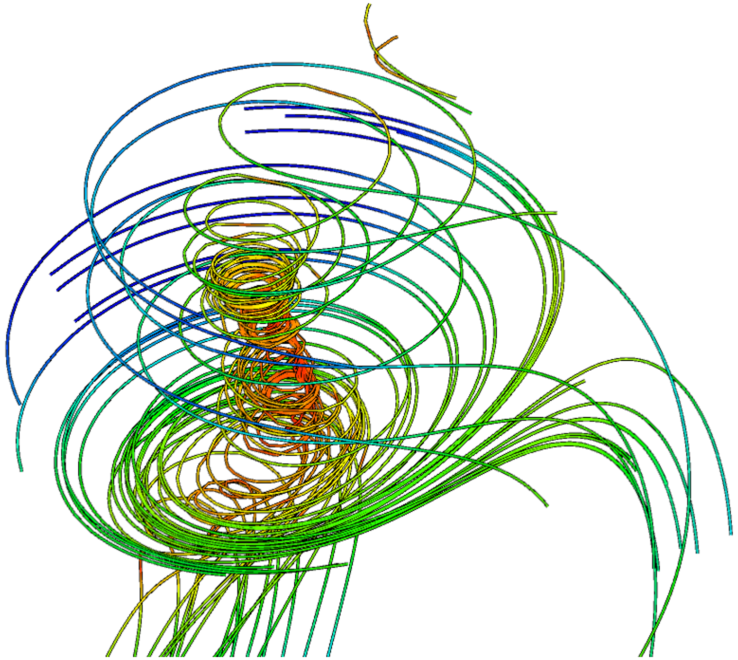
Works in any number of dimensions

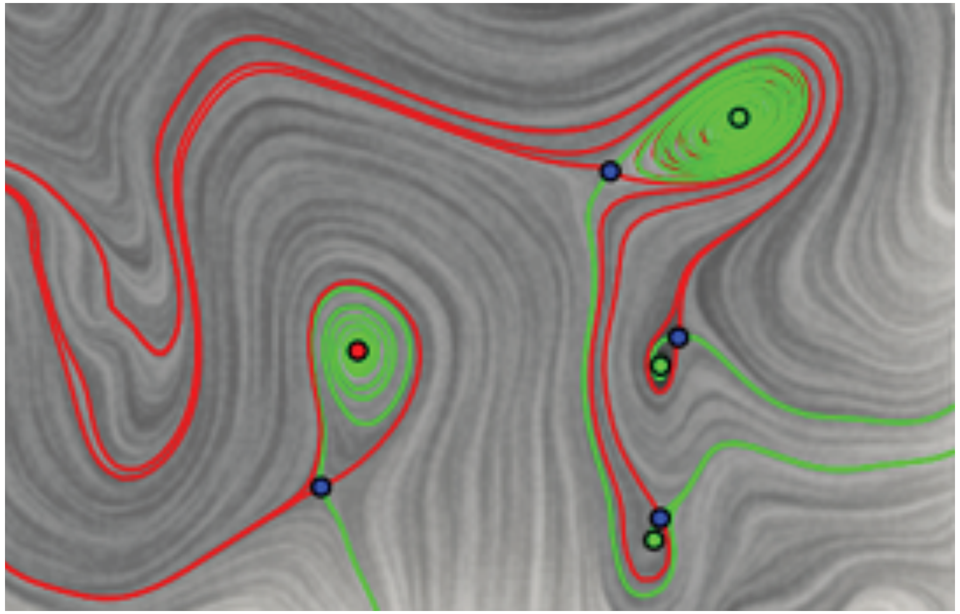
Handles discrete data by building filtrations on data

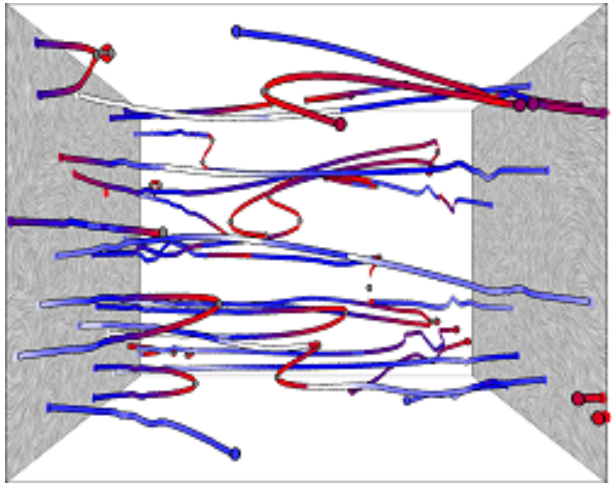
The visualization challenge—applying this approach to meaningfully reduce visual complexity

Successful TDA Applications
Vector Field Simplification





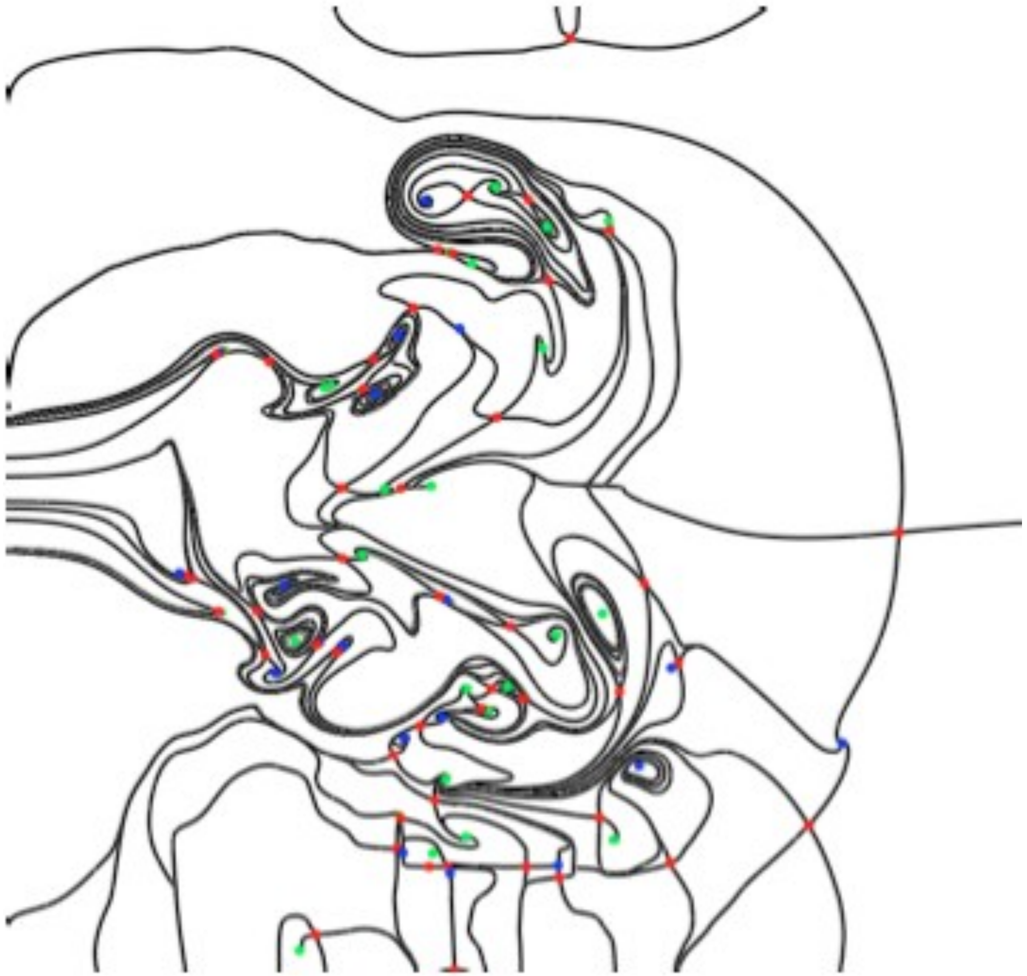




- P. Skraba, P. Rosen, B. Wang, G. Chen, H. Bhatia, V. Pascucci, Simplifying critical points in 3d vector fields with robustness, to appear IEEE Transactions on Visualization and Computer Graphics, 2016.
- P. Skraba, B. Wang, G. Chen, P. Rosen, Robustness-Based Simplification for 2D Steady and Unsteady Flows, IEEE Transactions on Visualization and Computer Graphics, 21(8): 930-944, 2015.
- P. Skraba, B. Wang, G. Chen, P. Rosen. 2d vector field simplification based on robustness. In IEEE Pacific Visualization Symposium, PacificVis, 2014.
- B. Wang, P. Rosen, P. Skraba, H. Bhatia, V. Pascucci. Visualizing robustness of critical points for 2d time-varying vector fields. In Computer Graphics Forum (EuroVis), 2013.

MOTIVATION

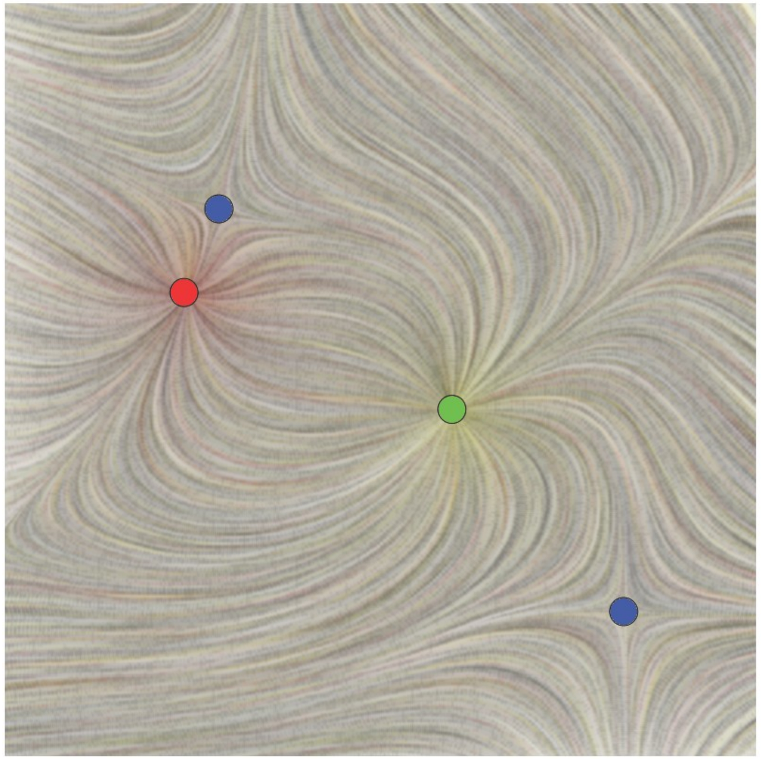
Increasing gap between ***increasing size and complexity of VF data*** and ***limited bandwidth of our visual perception channel***

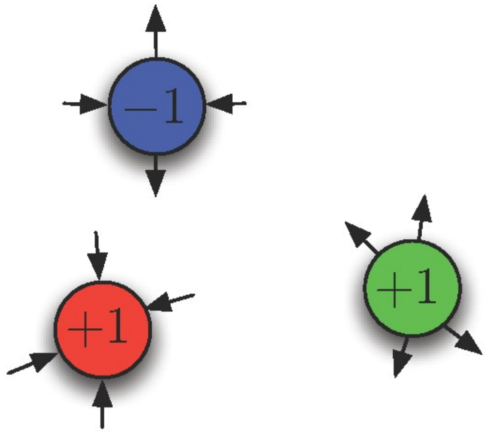


Swirling jet simulation
[Tricoche, Scheuermann and Hagen 2001]

Degrees

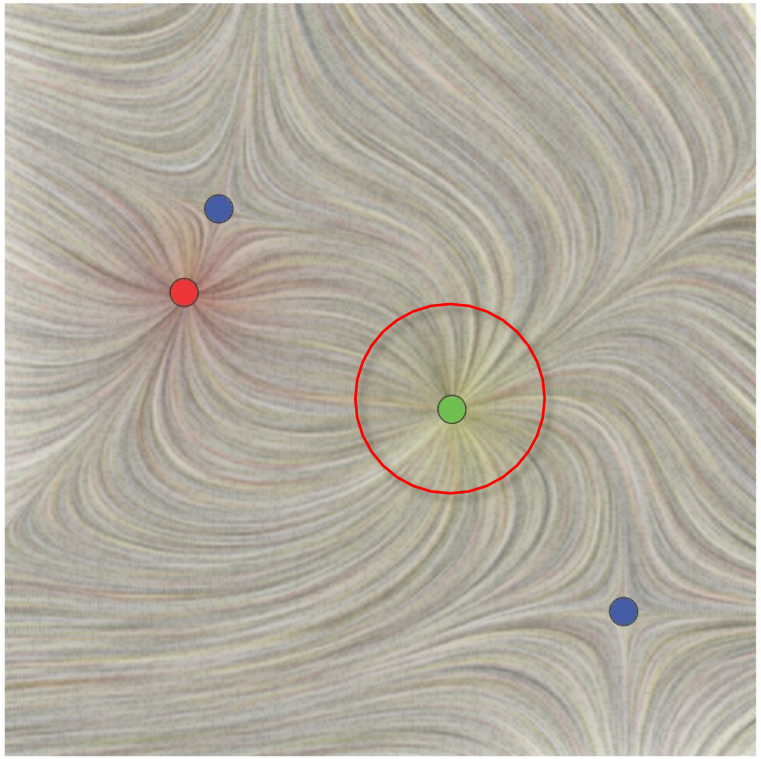
In 2D, deg(x) of a critical point x equals its Poincare index

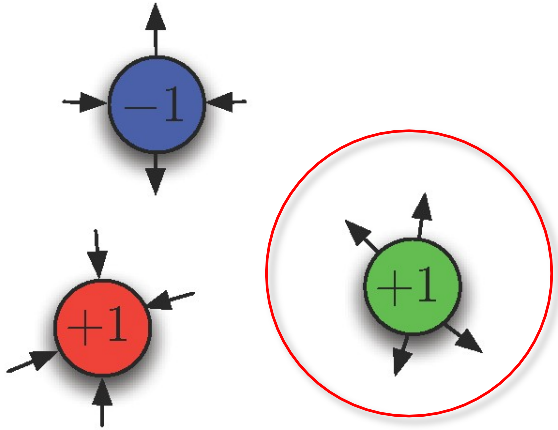




DEGREES

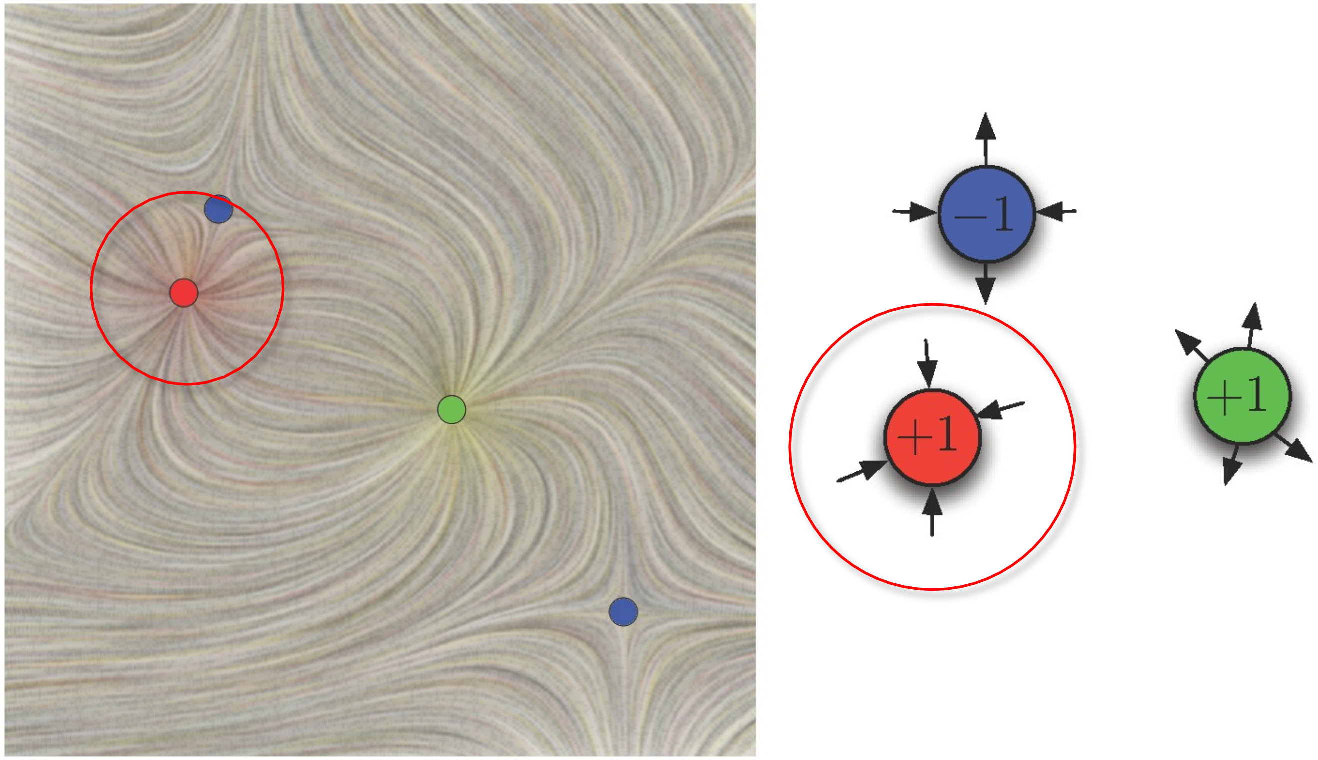
In 2D, deg(x) of a critical point x equals its Poincare index





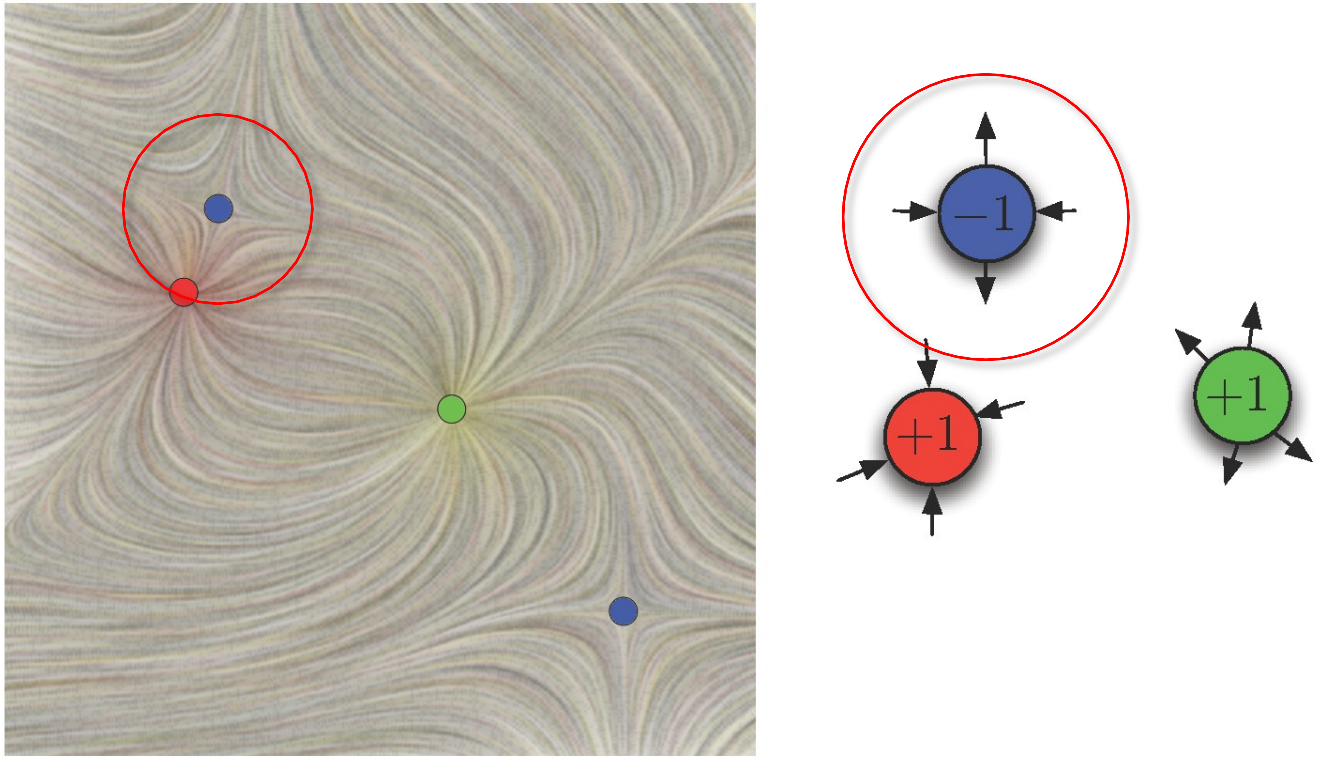
DEGREES

In 2D, deg(x) of a critical point x equals its Poincare index



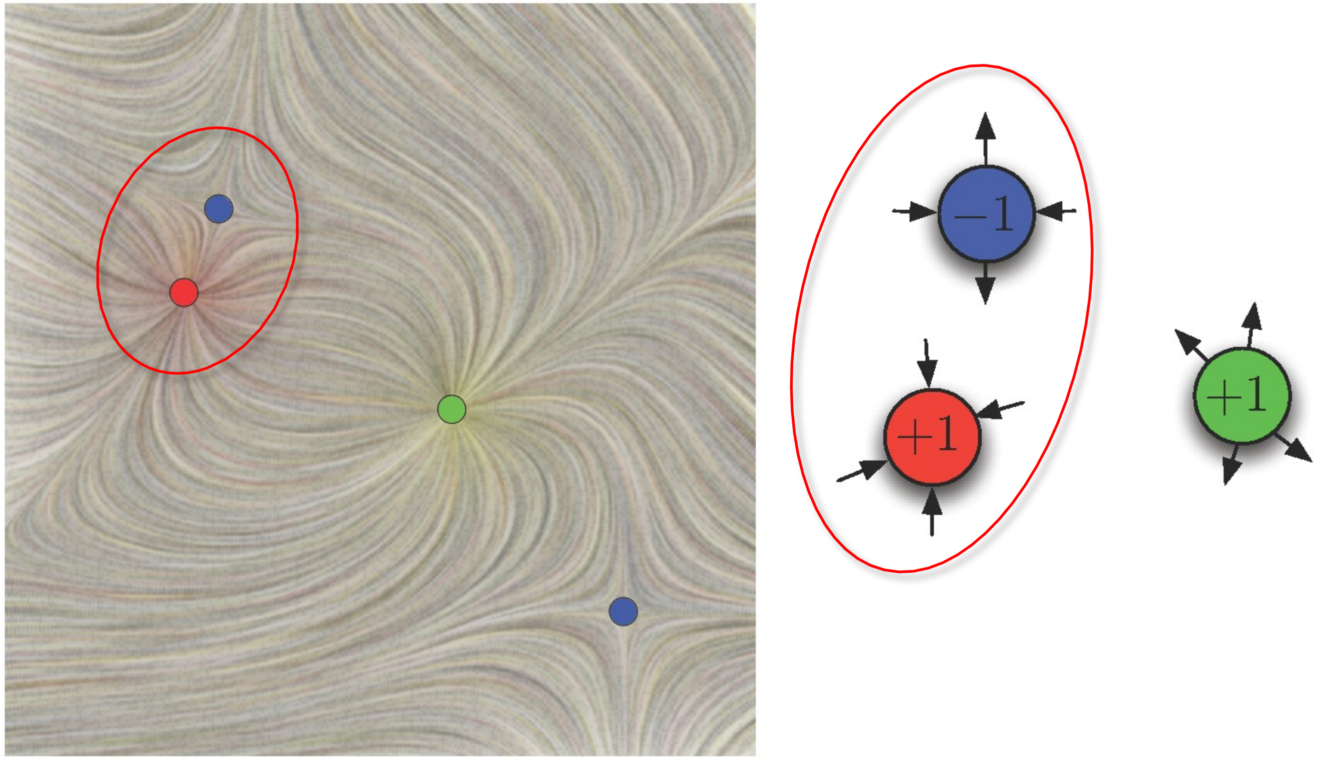
DEGREES

In 2D, deg(x) of a critical point x equals its Poincare index



DEGREES

If a component has degree zero, it is possible to replace the VF inside C with a VF free of critical points



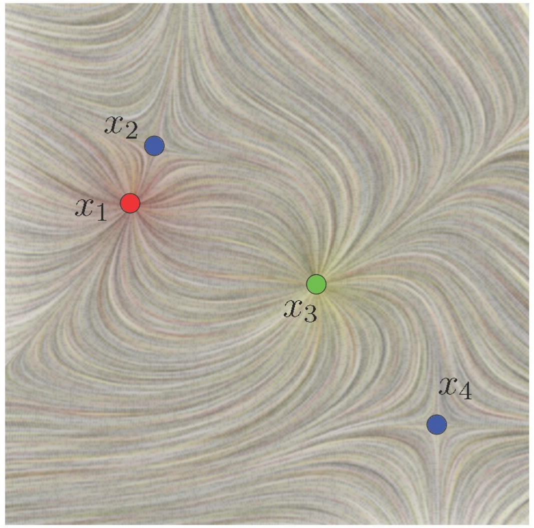
COMPUTATION OF ROBUSTNESS: MERGE TREE

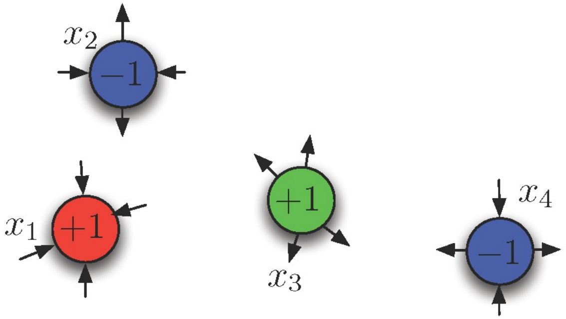
Consider the gradient magnitude function

Construct a merge tree around the critical points

MERGE TREE OF GRADIENT MAGNITUDE FUNCTION

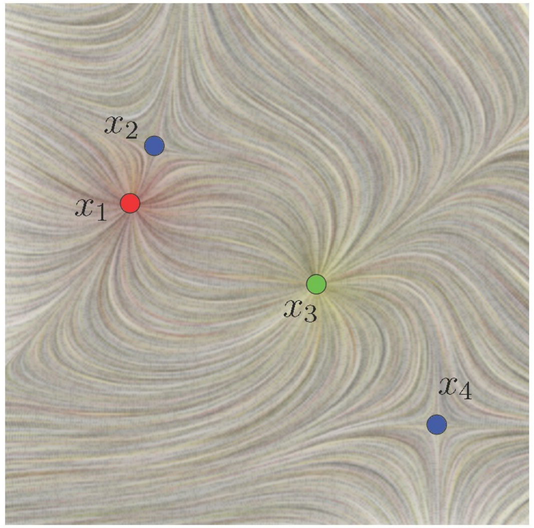
Track connected components of F_r as they appear and merge, as *r* increases from 0

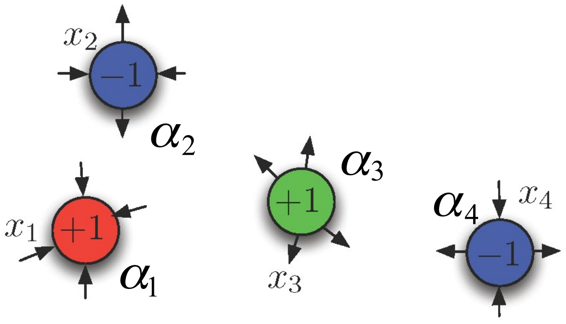


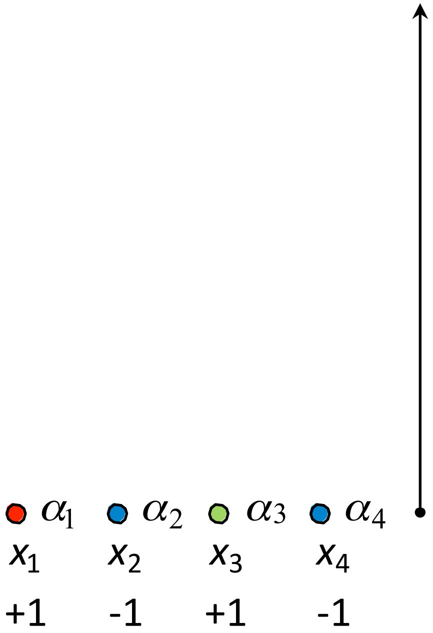


MERGE TREE OF GRADIENT MAGNITUDE FUNCTION

Track connected components of F_r as they appear and merge, as *r* increases from 0

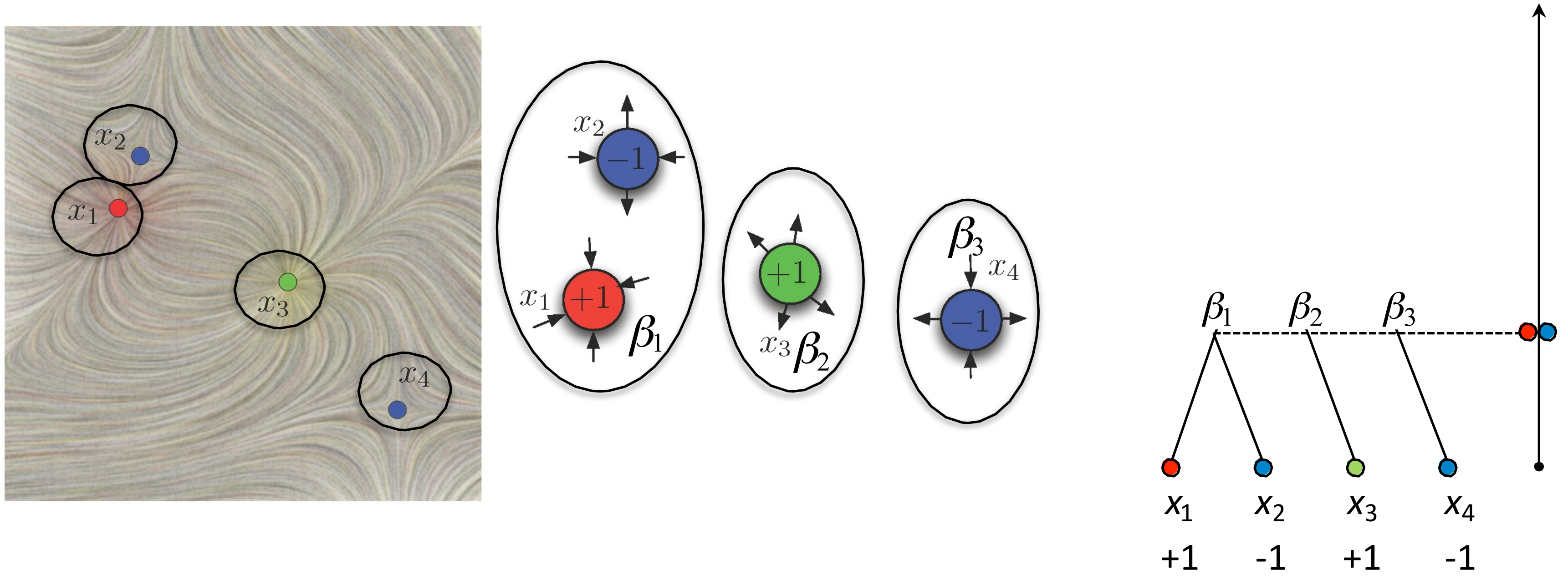






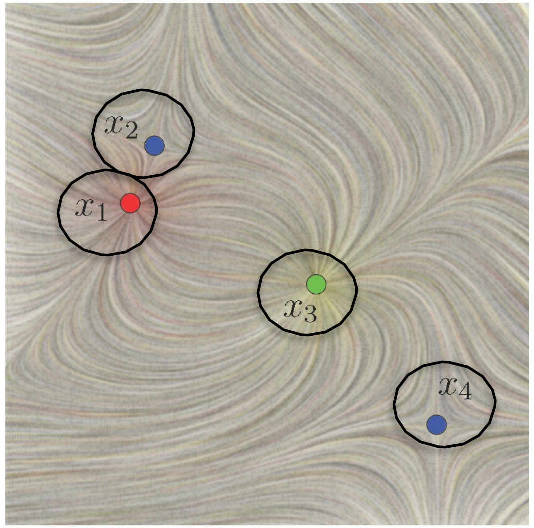
Merge tree of gradient magnitude function

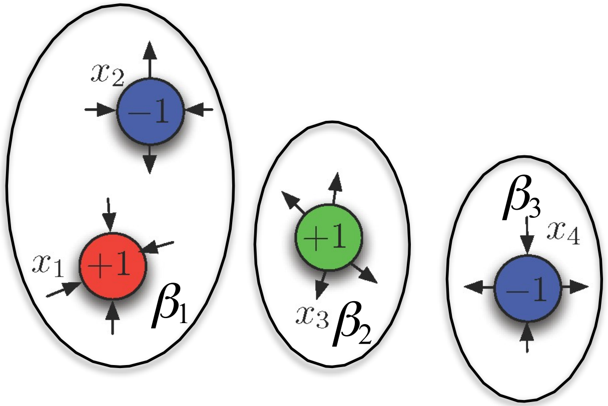
Track connected components of F_r as they appear and merge, as *r* increases from 0

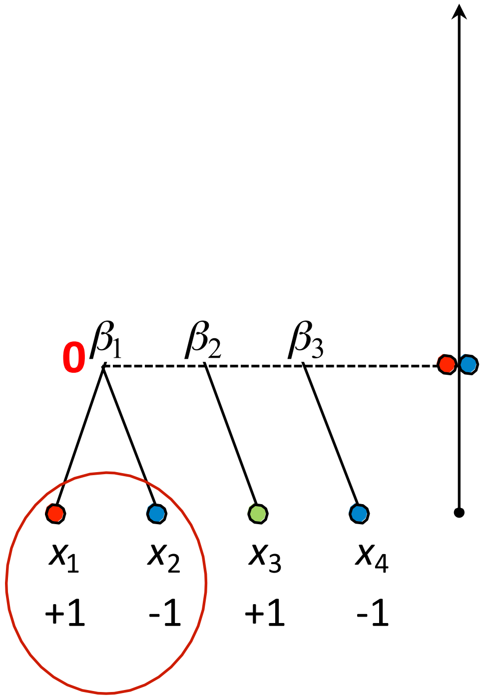


Merge tree of gradient magnitude function

Track connected components of F_r as they appear and merge, as r increases from 0

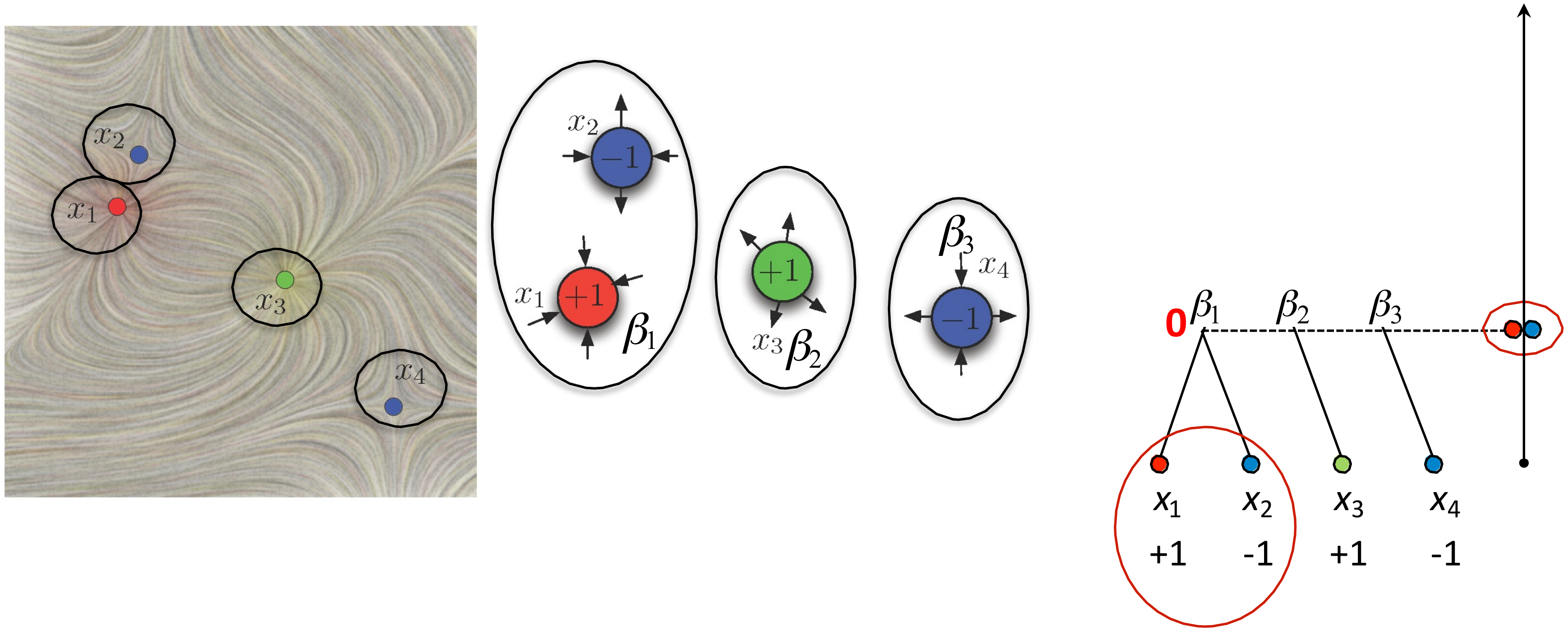






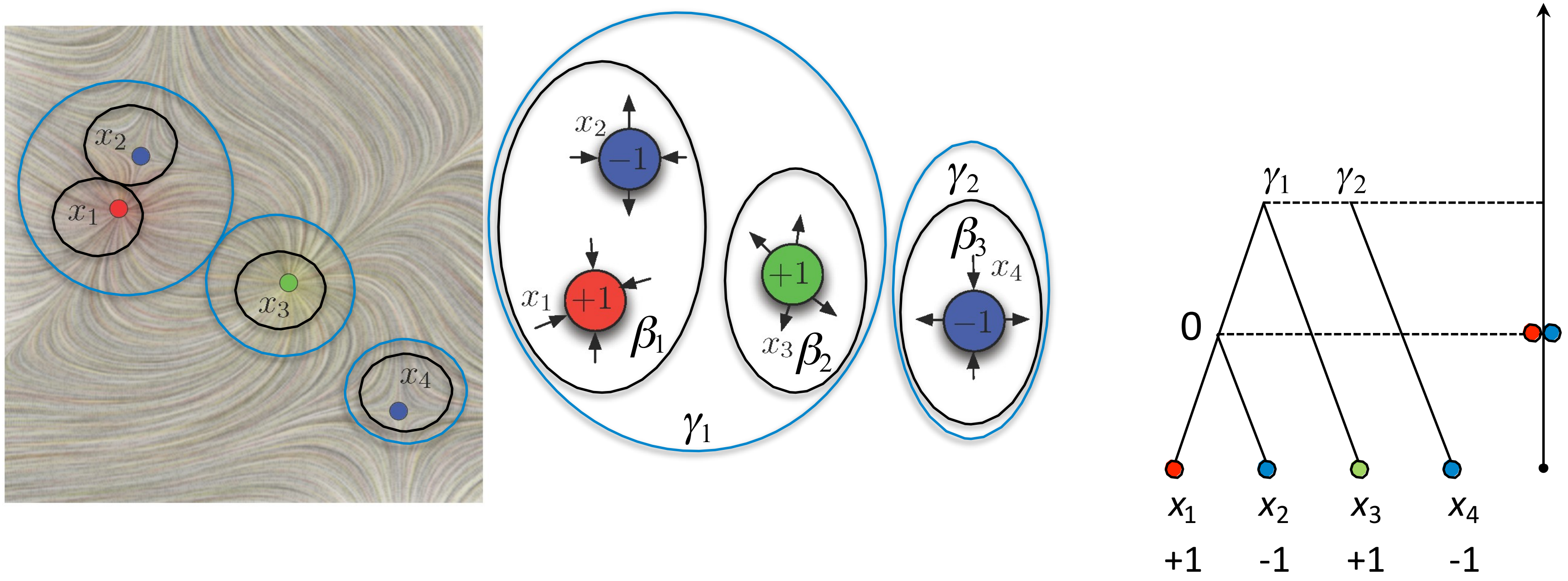
Merge tree of gradient magnitude function

Track connected components of F_r as they appear and merge, as r increases from 0



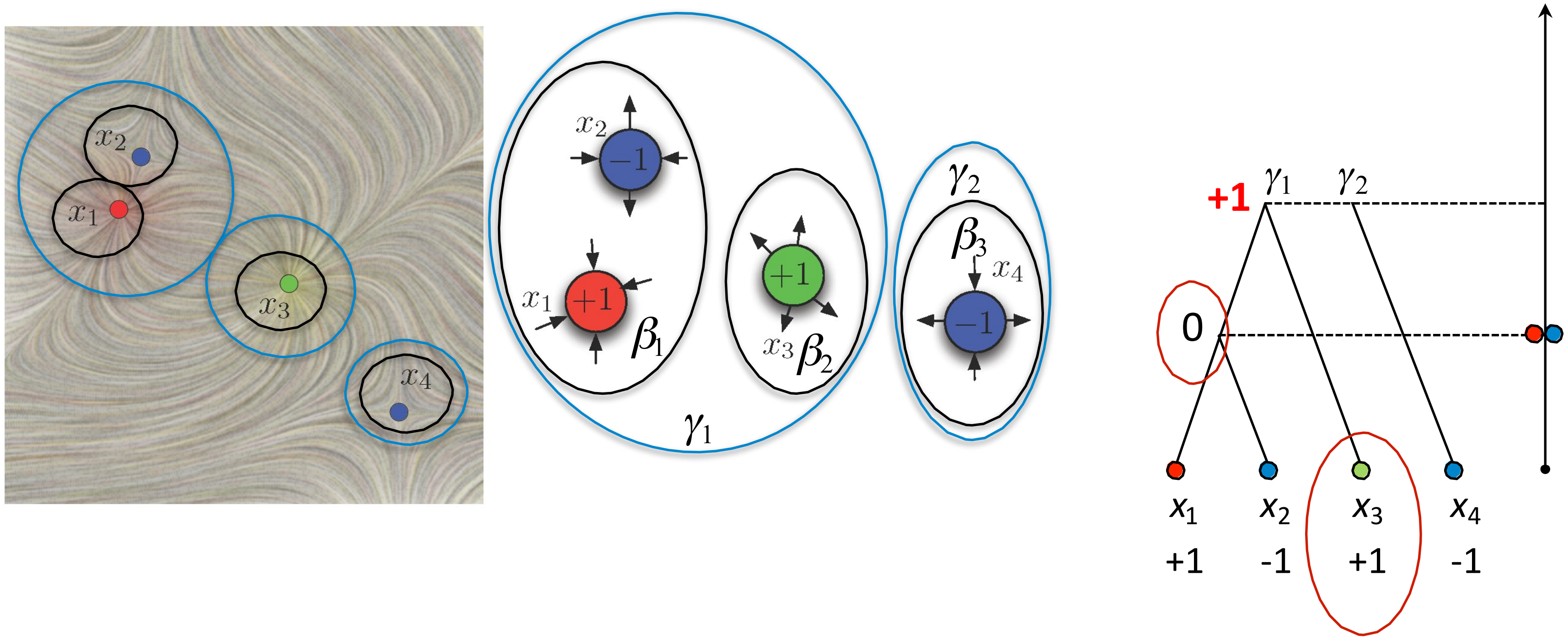
Merge tree of gradient magnitude function

Track connected components of F_r as they appear and merge, as *r* increases from 0



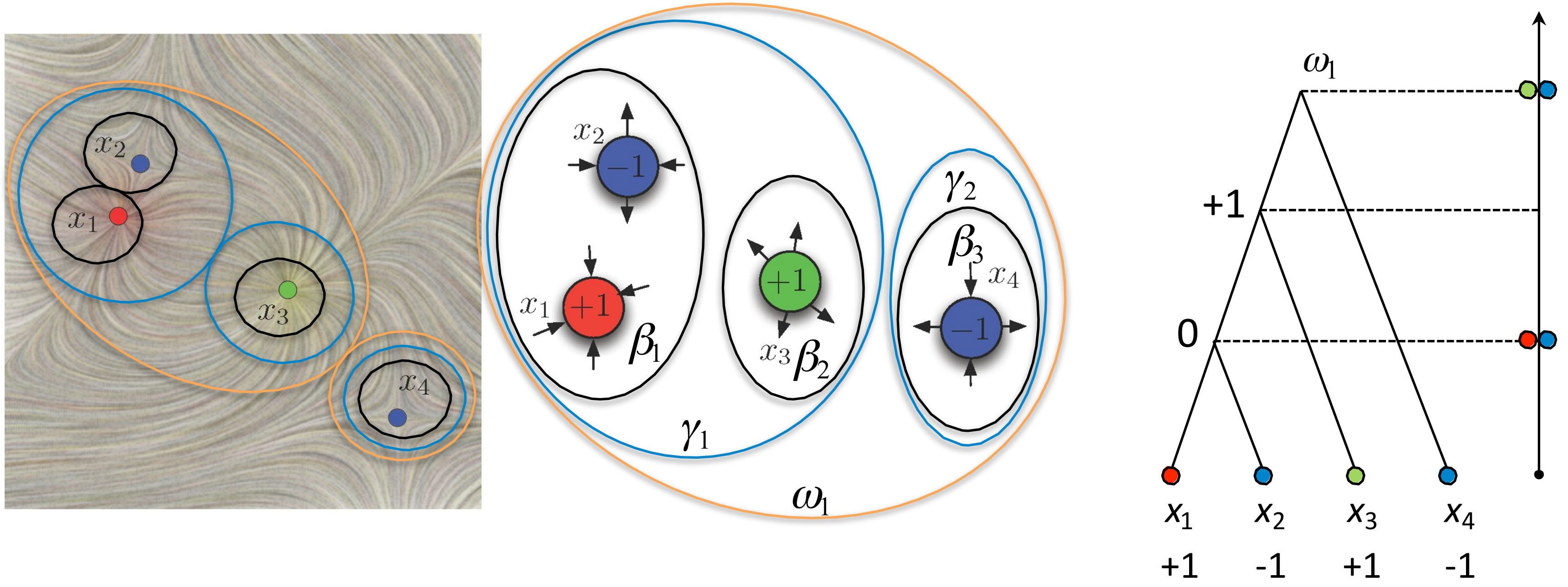
Merge tree of gradient magnitude function

Track connected components of F_r as they appear and merge, as *r* increases from 0



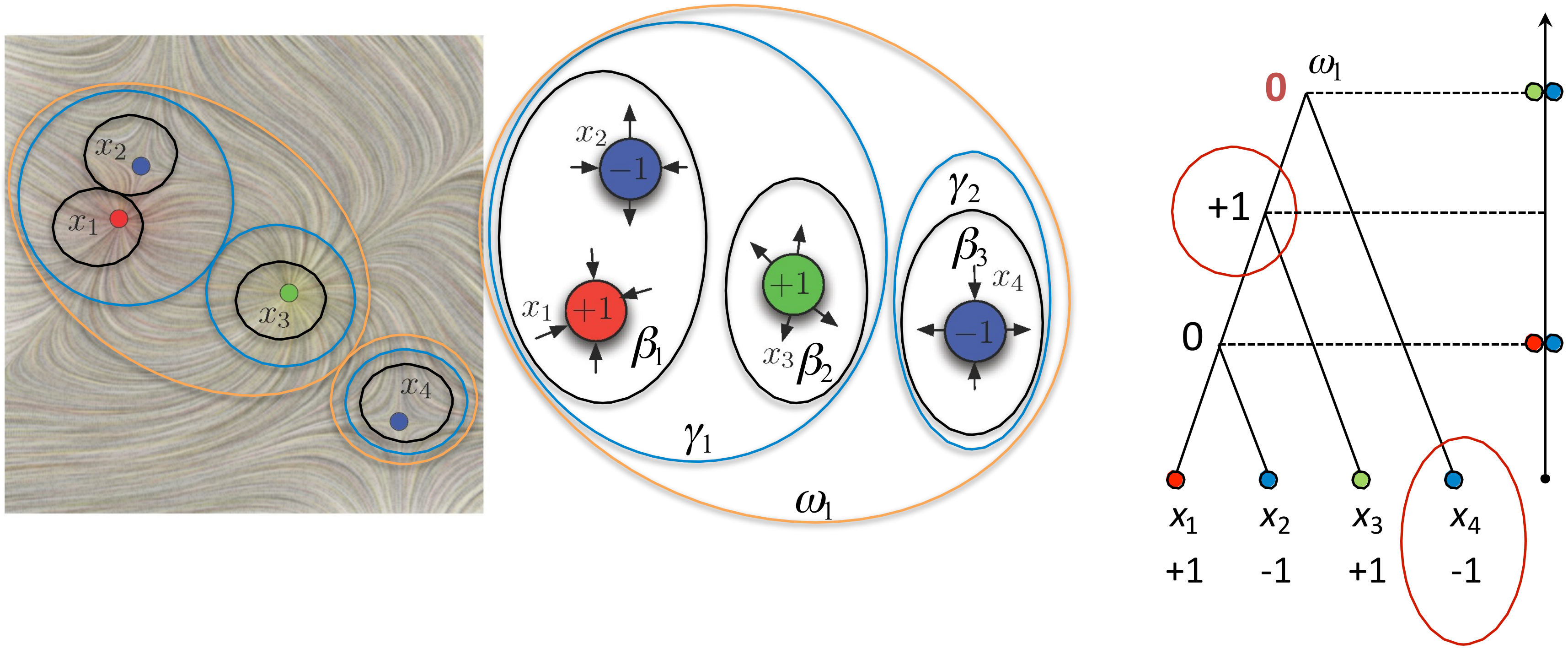
Merge tree of gradient magnitude function

Track connected components of F_r as they appear and merge, as r increases from 0



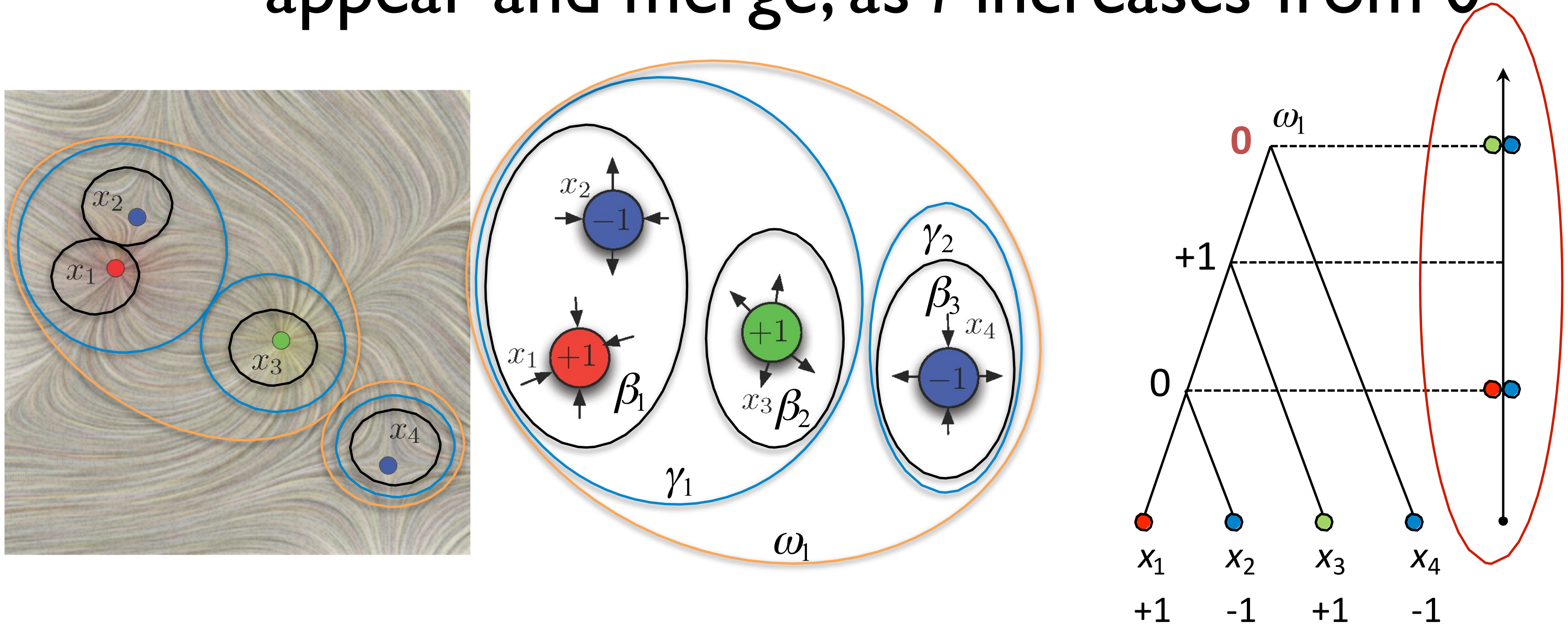
MERGE TREE OF GRADIENT MAGNITUDE FUNCTION

Track connected components of F_r as they appear and merge, as *r* increases from 0



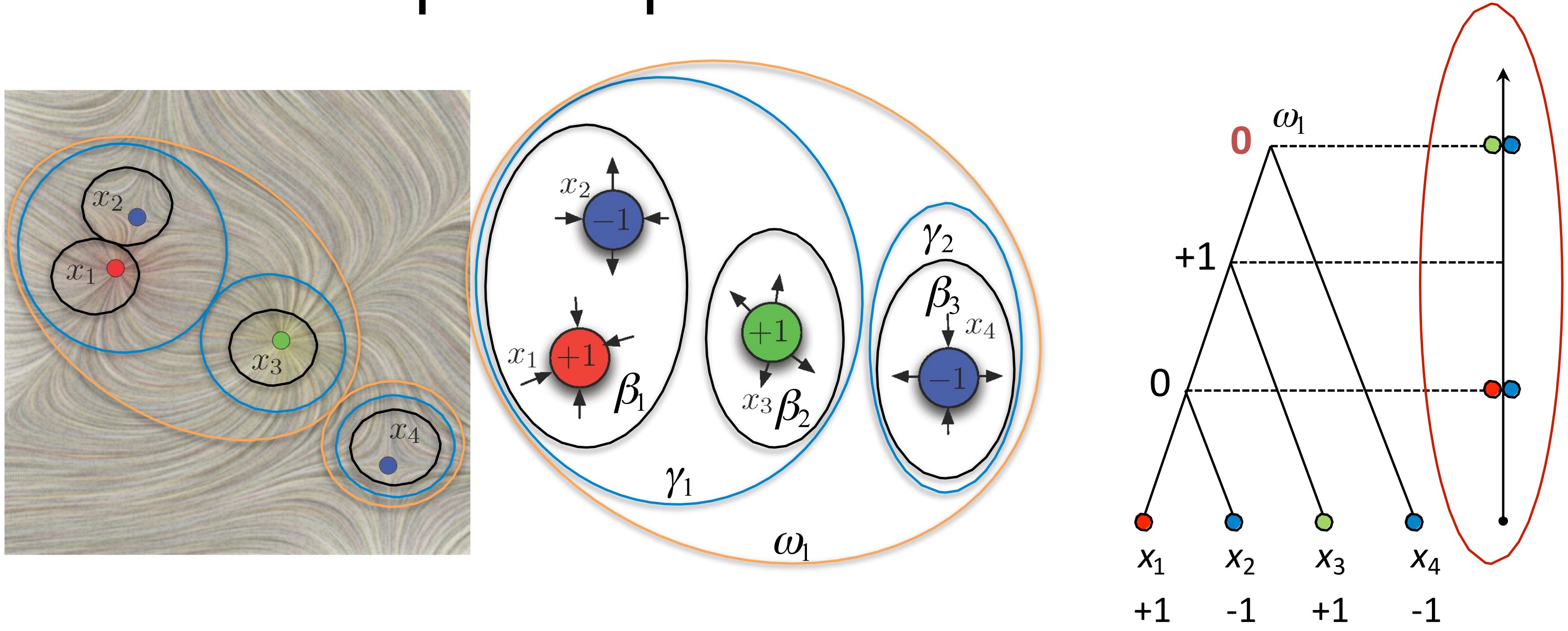
MERGE TREE OF GRADIENT MAGNITUDE FUNCTION

Track connected components of F_r as they appear and merge, as *r* increases from 0



ROBUSTNESS

Robustness quantifies the stability of a critical point with respect to perturbations of the vector fields



Robustness: $\mathrm{rb}(x_1) = \mathrm{rb}(x_2) = r_1$, $\mathrm{rb}(x_3) = \mathrm{rb}(x_4) = r_2$

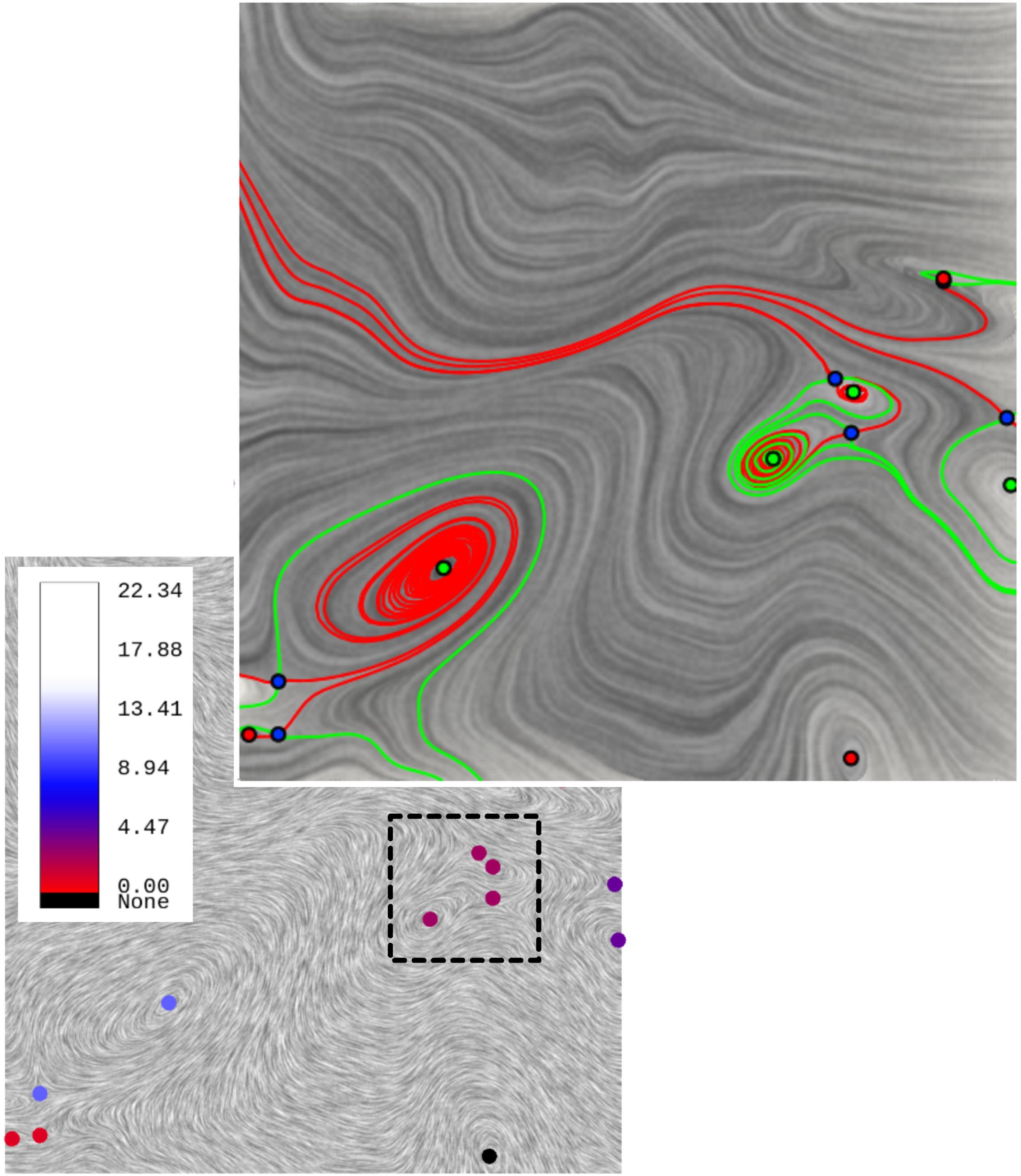
CHAZAL, PATEL AND SKRABA 2011

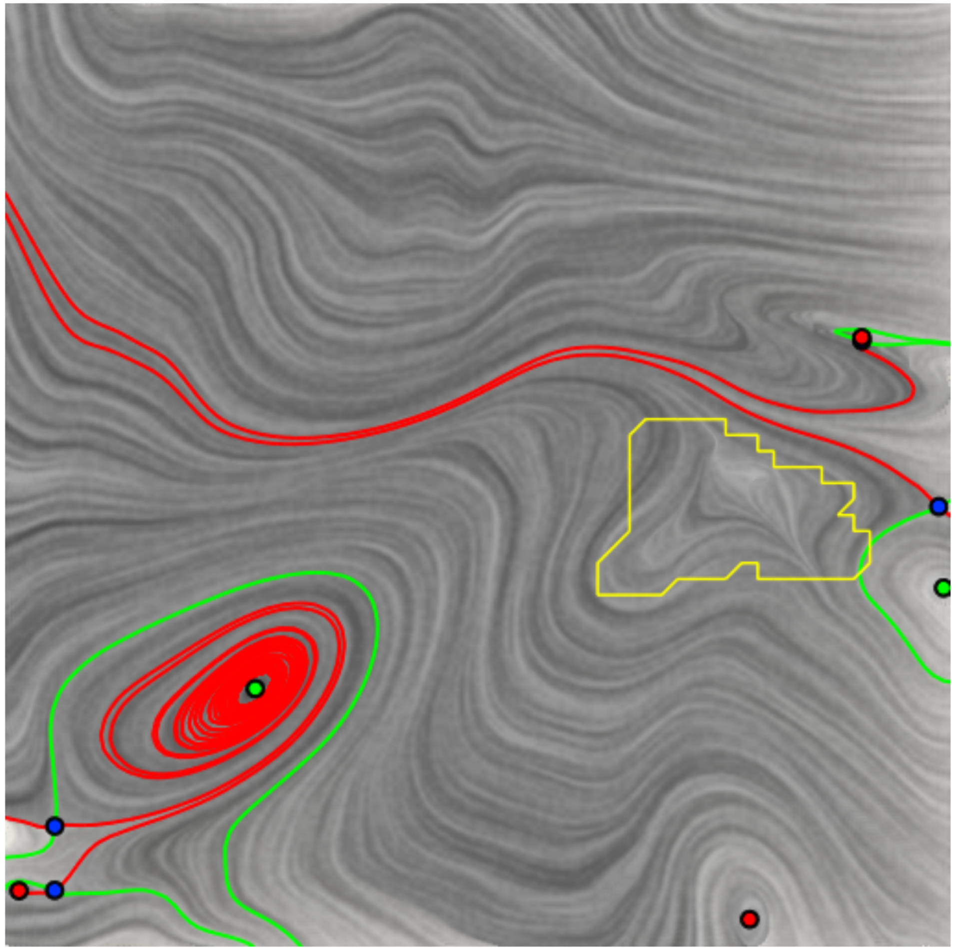
SIMPLIFICATION OF THE VECTOR FEILD

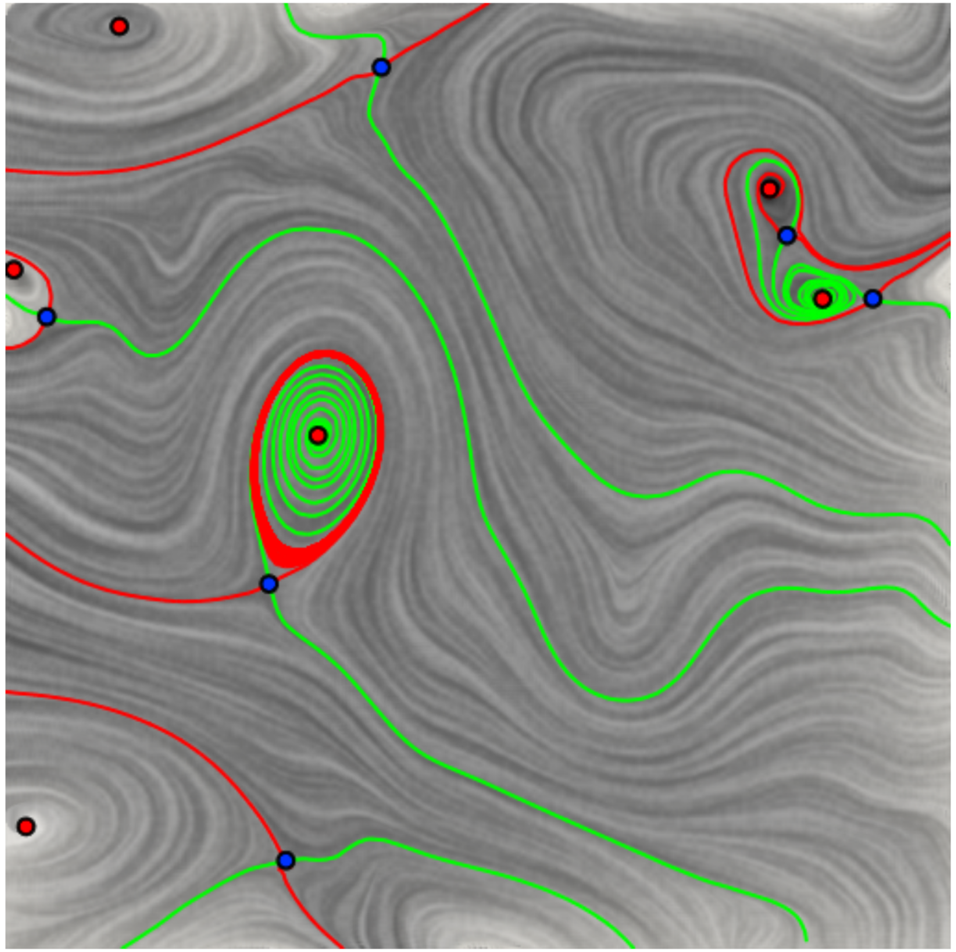
Operates only on sublevelset

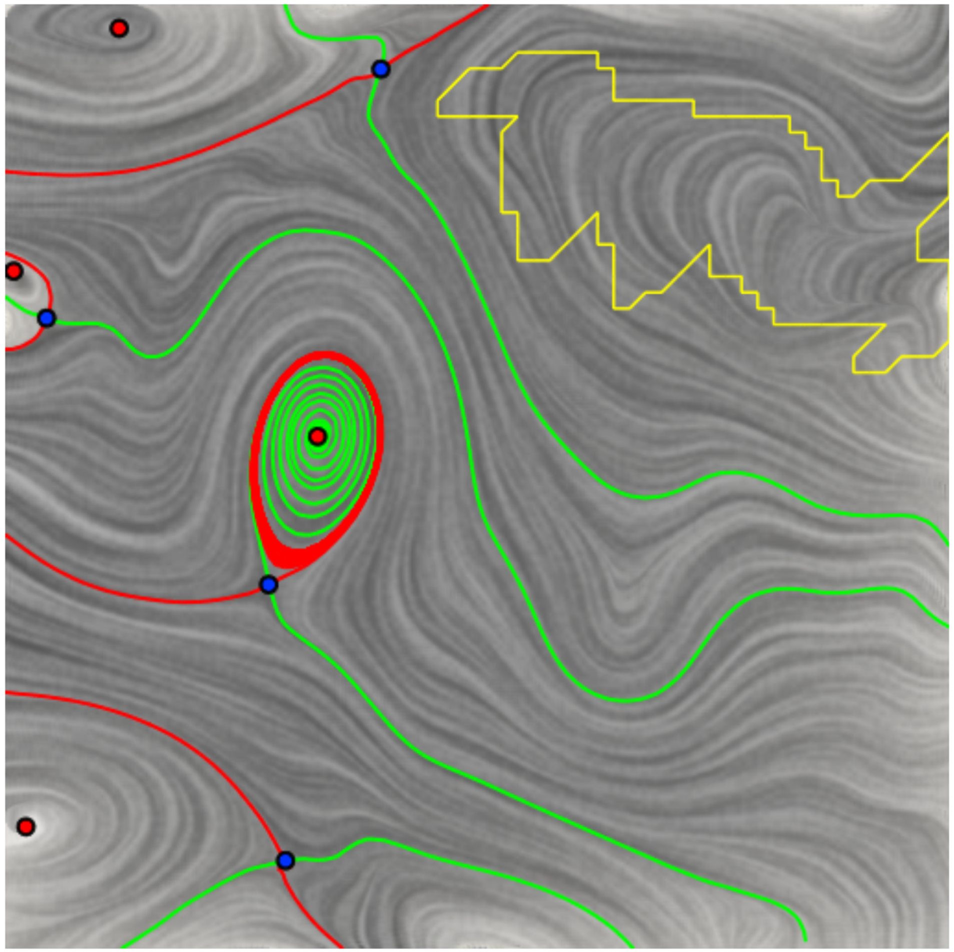
Maintains boundary, provides a minimum perturbation to simplify

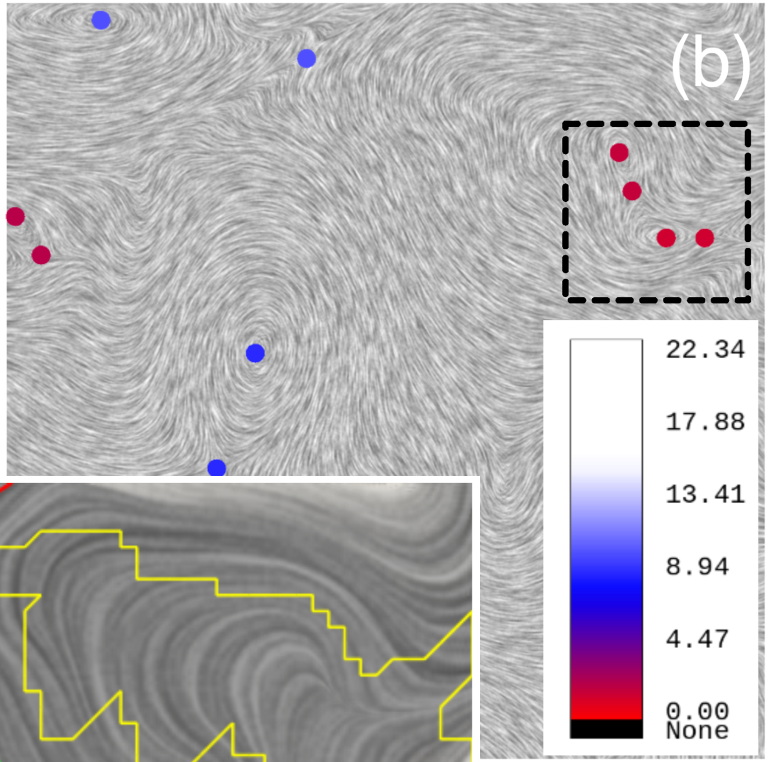
For details, see P. Skraba, B. Wang, G. Chen, P. Rosen. 2d vector field simplification based on robustness. In IEEE Pacific Visualization Symposium, PacificVis, 2014.


22.34
17.88
13.41
8.94
4.47
0.00
None








(b)
22.34
17.88
13.41
8.94
4.47
0.00
None

VF Conclusion

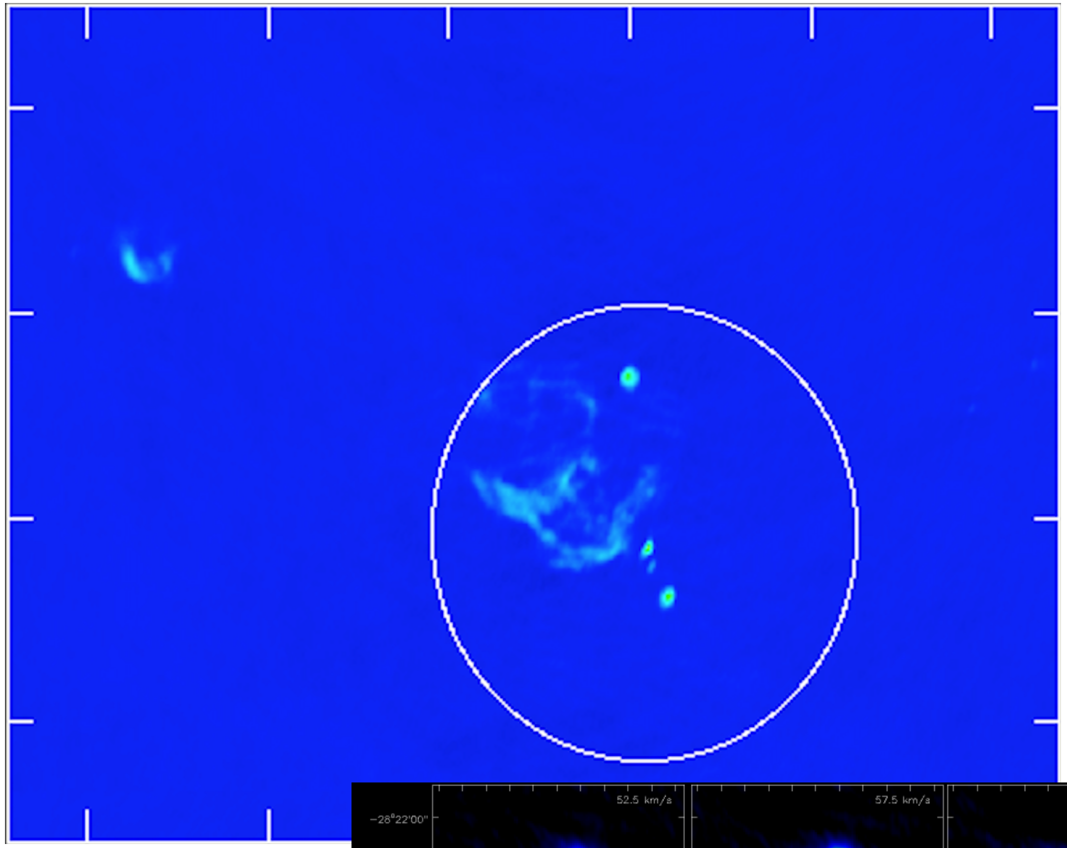
Robustness identifies critical points considered turbulence or noise

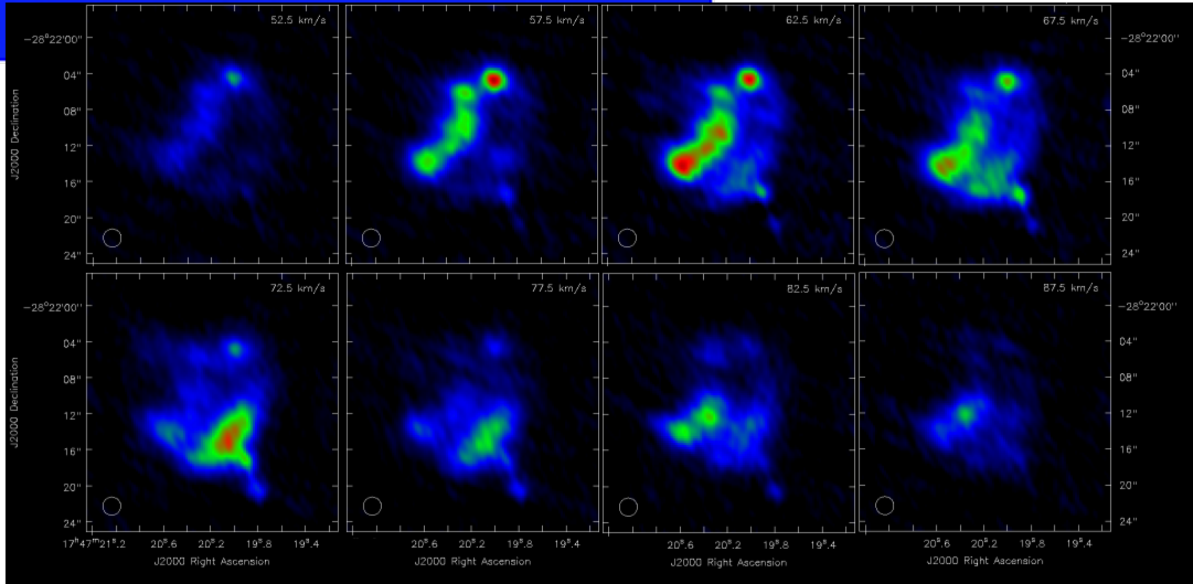
Simplification removes those critical points to focus on global flow structure

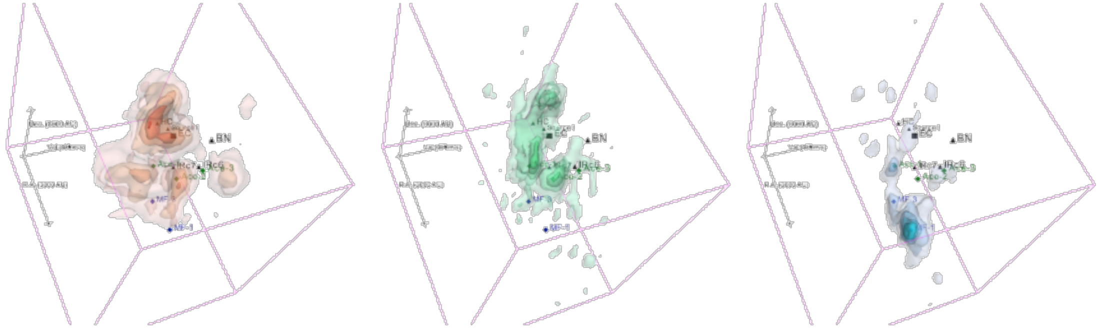
We have also published extensions to time-varying and 3d vector fields

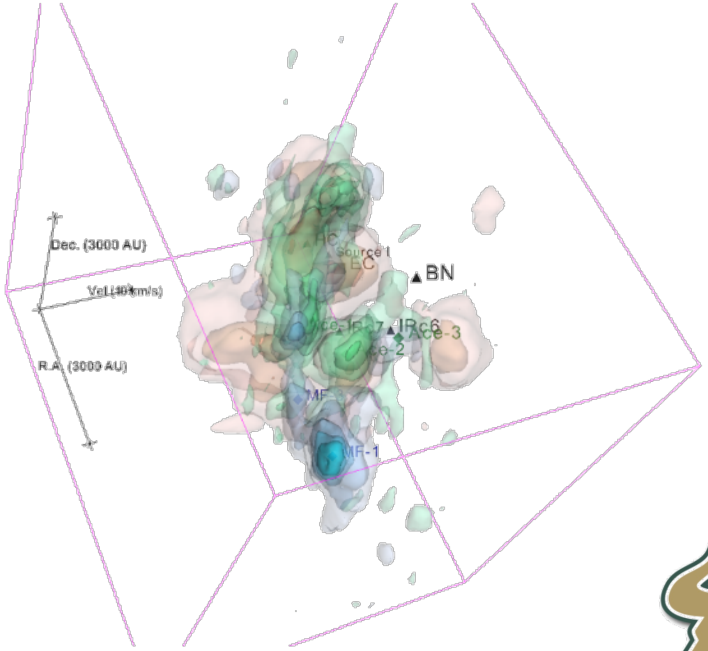
New TDA Applications

Radio Astronomy

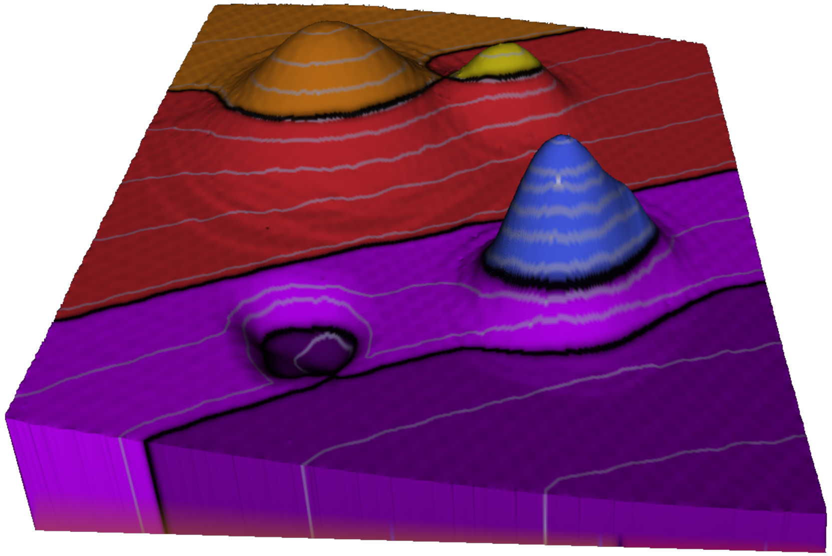


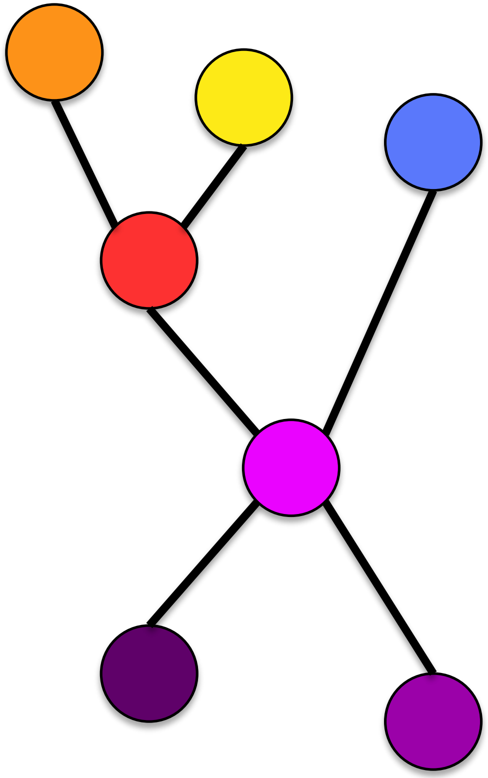

52.5 km/s
57.5 km/s
62.5 km/s
67.5 km/s
72.5 km/s
77.5 km/s
82.5 km/s
87.5 km/s
J2000 Declination
J2000 Right Ascension
−28°22'00"
04"
08"
12"
16"
20"
24"
17h47m21s.2
20s.6
20s.2
19s.8
19s.4


BN

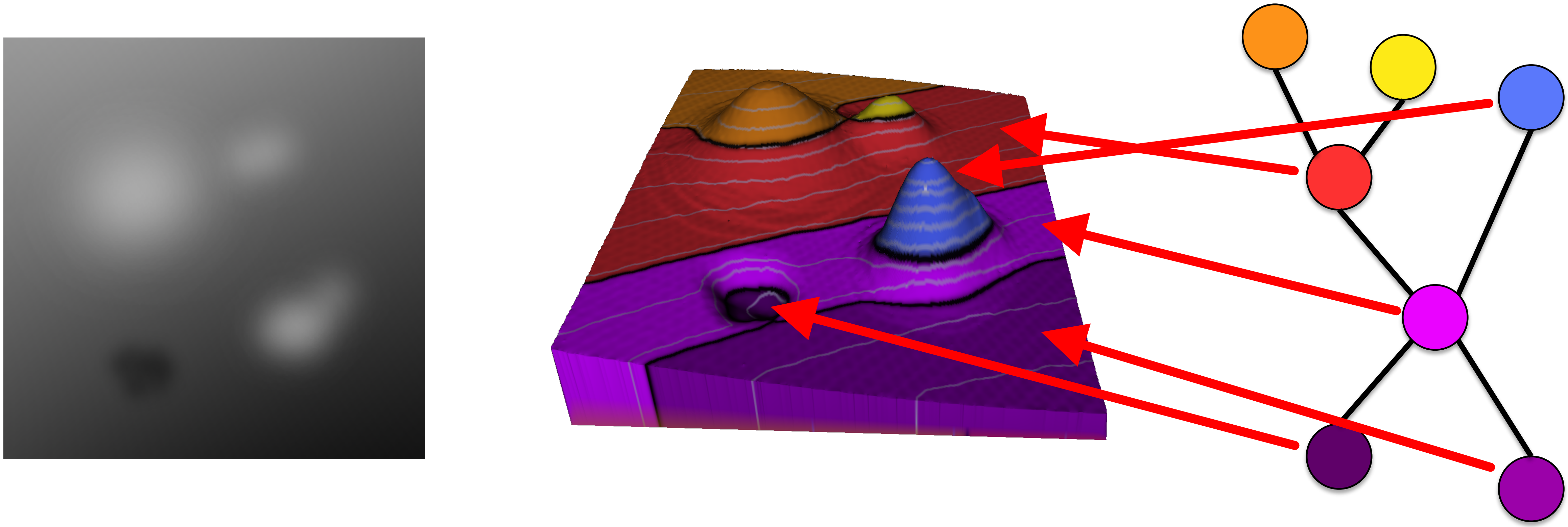

Dec. (3000 AU)
R.A. (3000 AU)
BN
Source I
IRc6
MF-1

Contour tree

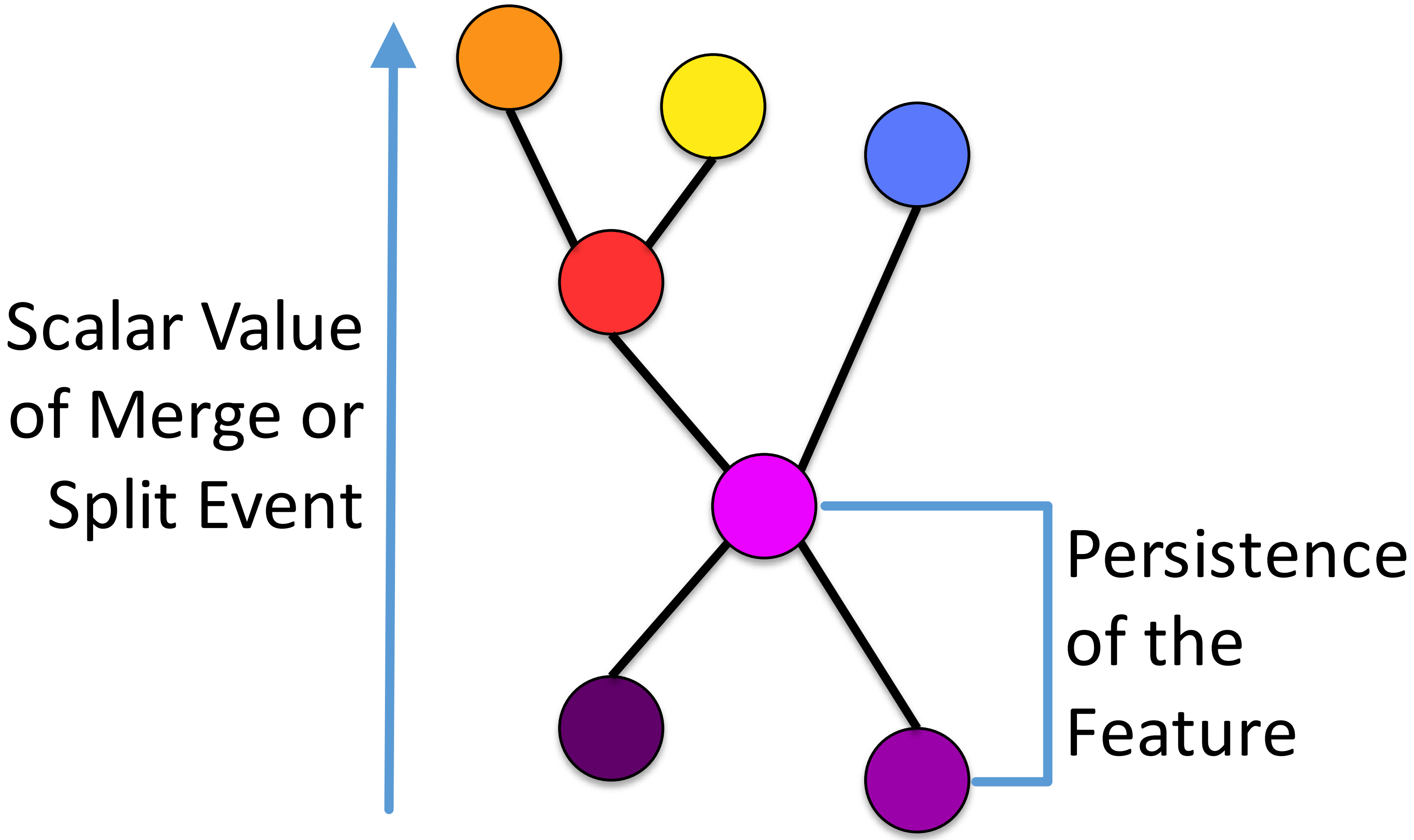




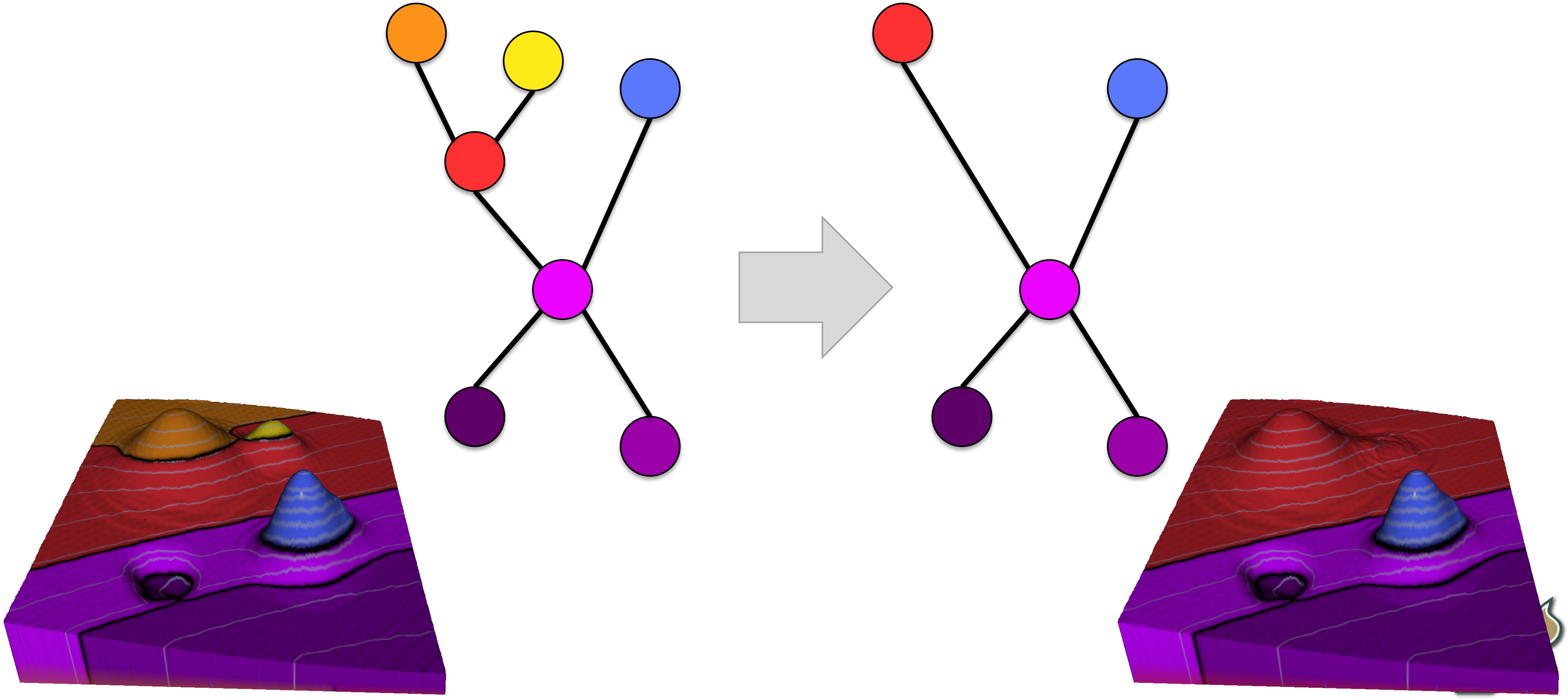
Contour tree



A Closer Look at the Contour tree



Feature Removal



Brand new project

No results to report yet

Radio astronomy data is very noisy (low SNR)—our hope is that contour tree simplification helps in pattern detection

RA/Postdoc positions available to work on the problem

Successful TDA Applications
Nuclear Engineering

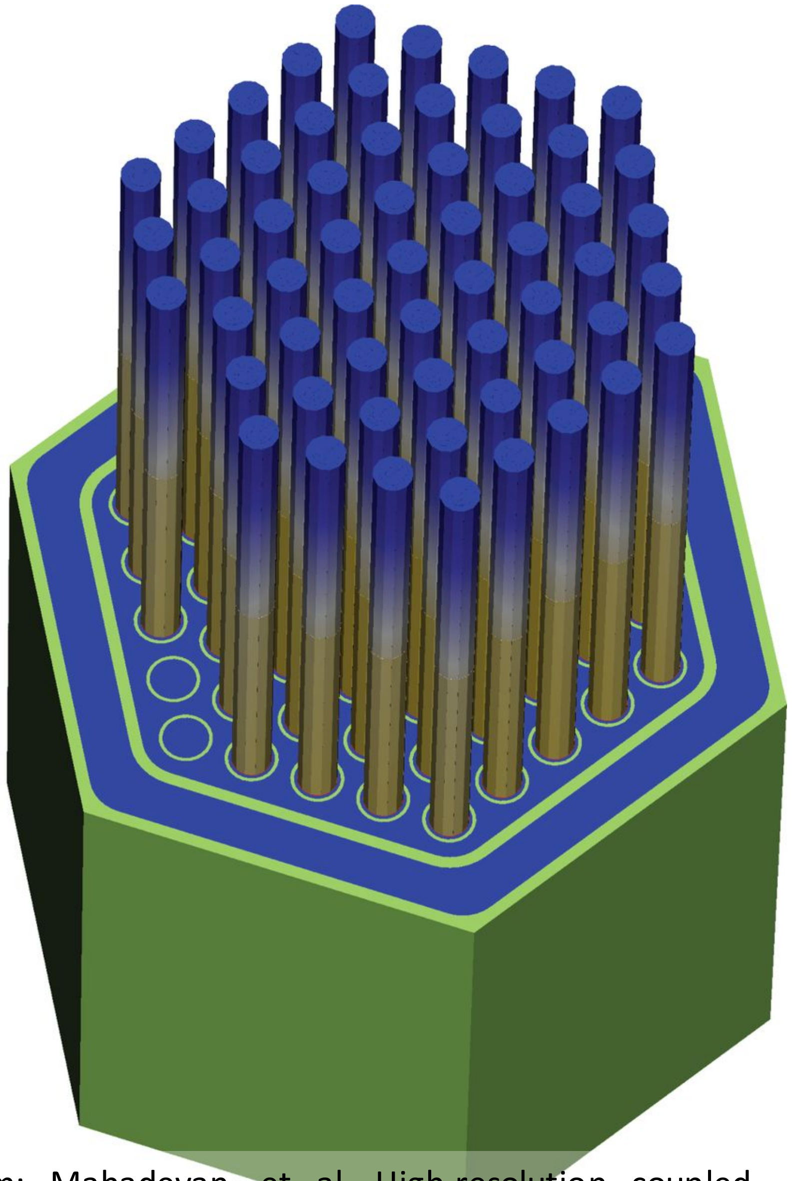
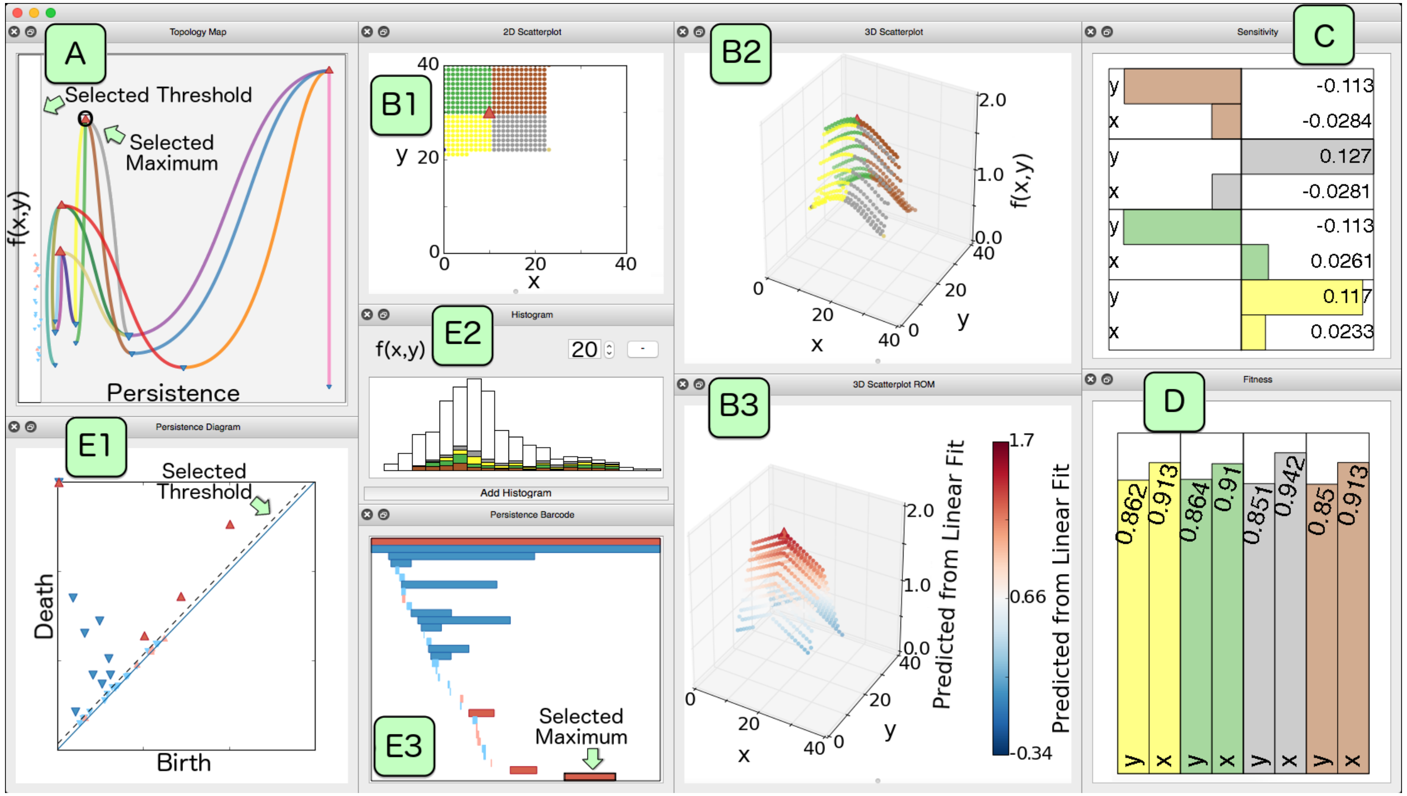
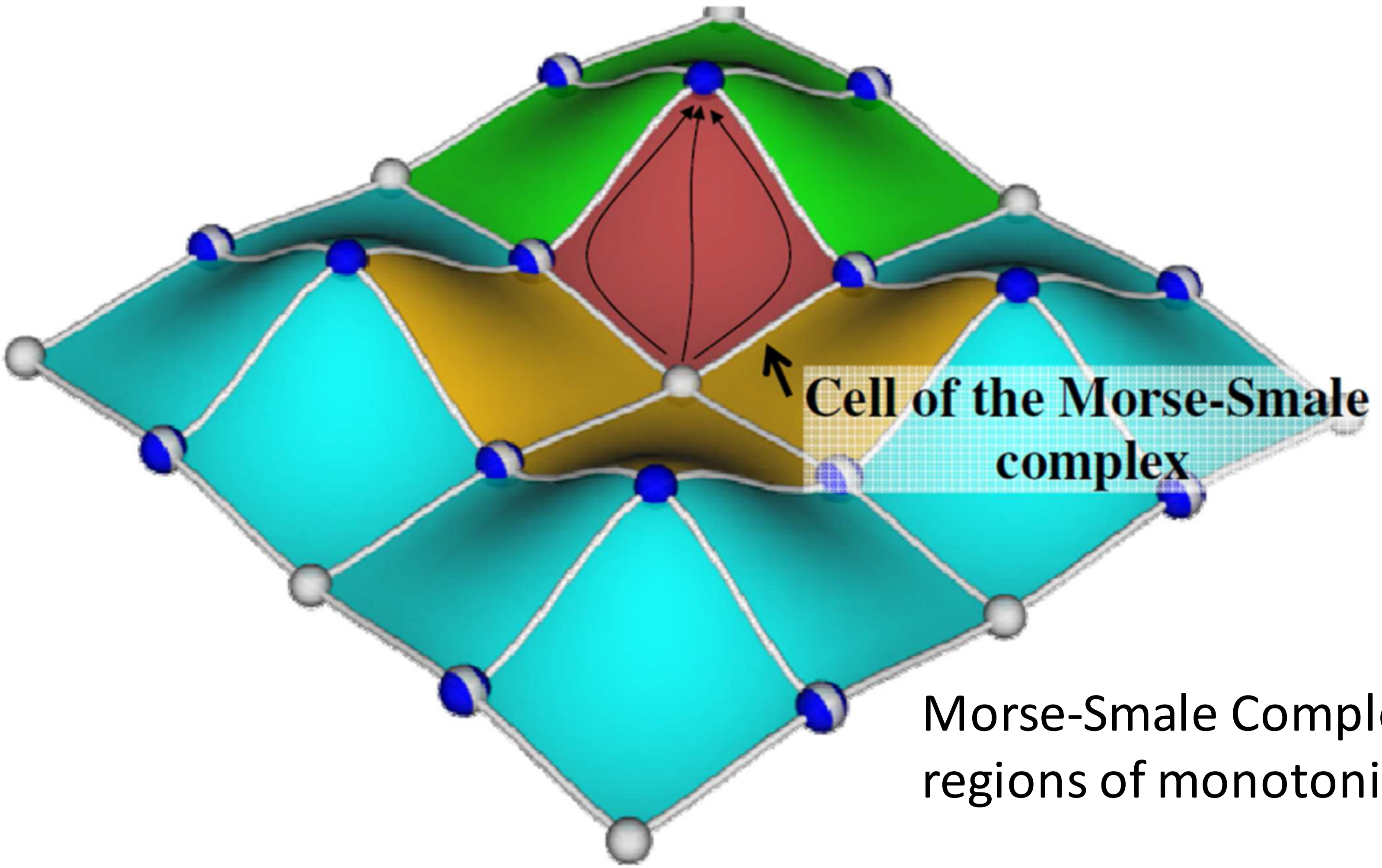


Image from: Mahadevan, et al. High-resolution coupled physics solvers for analysing fine-scale nuclear reactor design problems. Phil. Trans. of Royal Society of London, 2014



D. Maljovec, B. Wang, P. Rosen, A. Alfonsi, G. Pastore, C. Rabiti, V. Pascucci, "Topology-Inspired Partition-Based Sensitivity Analysis and Visualization of Nuclear Simulations", to appear IEEE Pacific Visualization, 2016.

MORSE-SMALE COMPLEX



Morse-Smale Complex defines regions of monotonic behavior

IMAGE COURTESY GUONING CHEN

Morse-Smale Regression

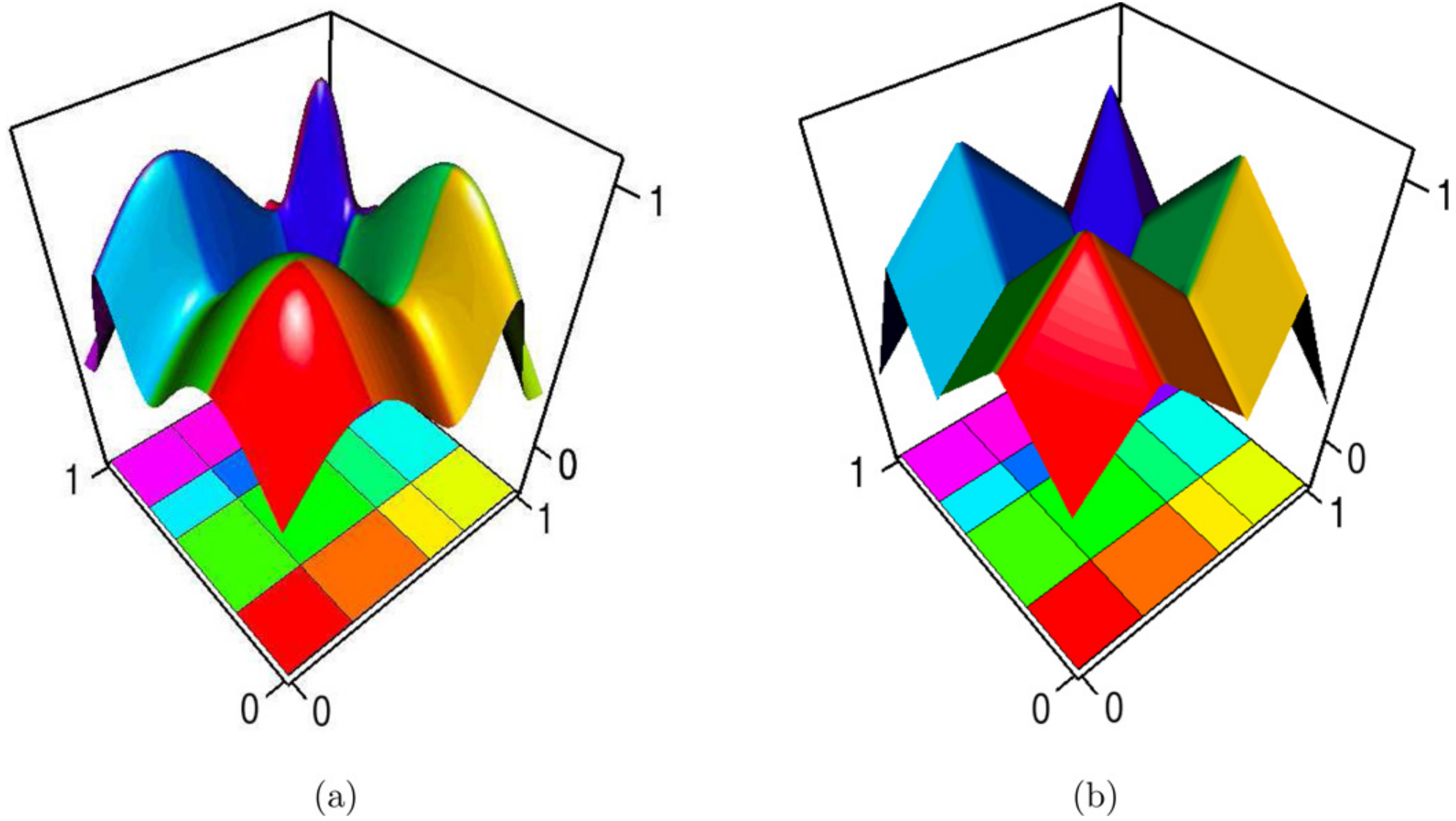
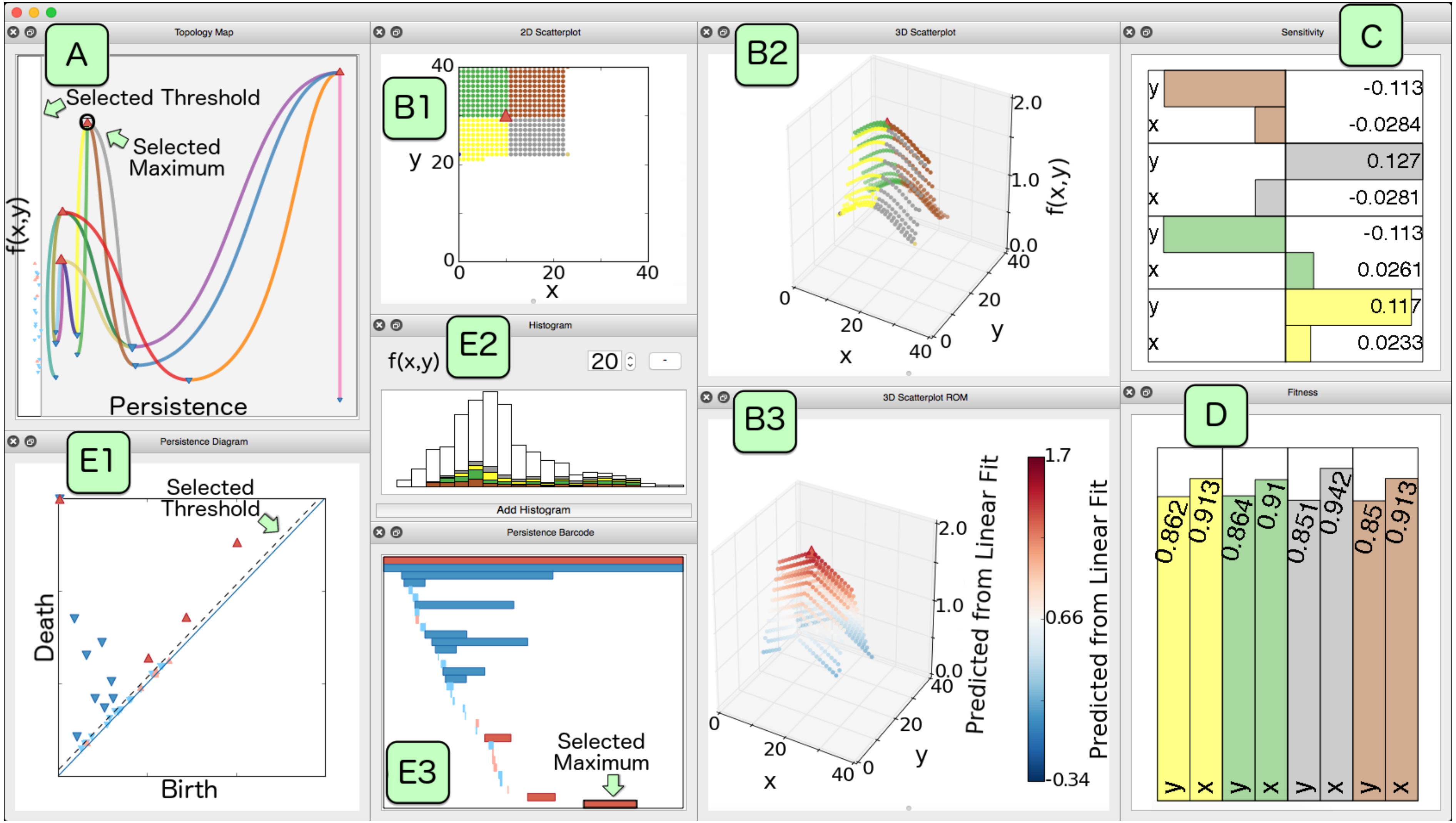
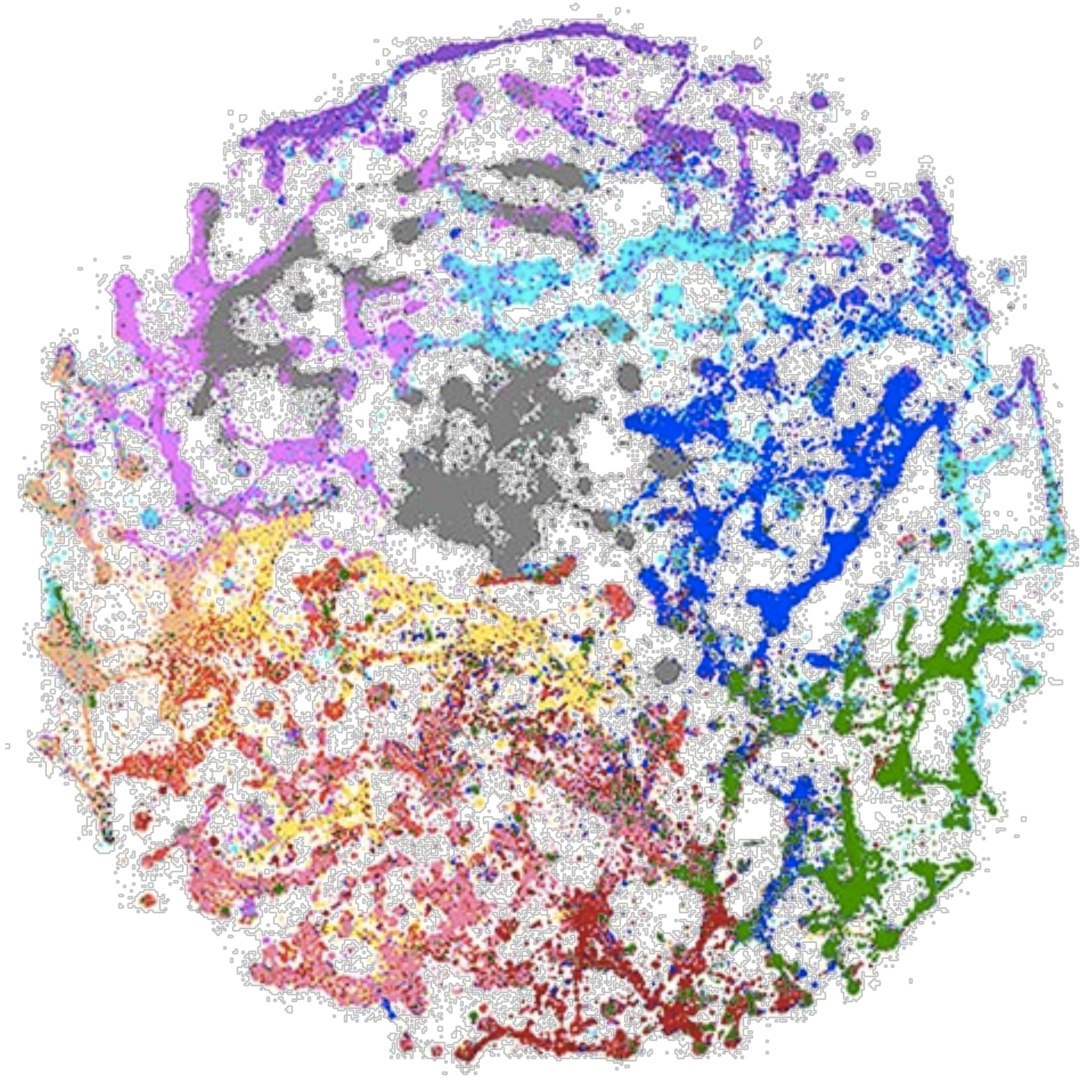


Image from: Gerber, et al. "Morse-Smale Regression." Journal of computational and graphical statistics, 2013.



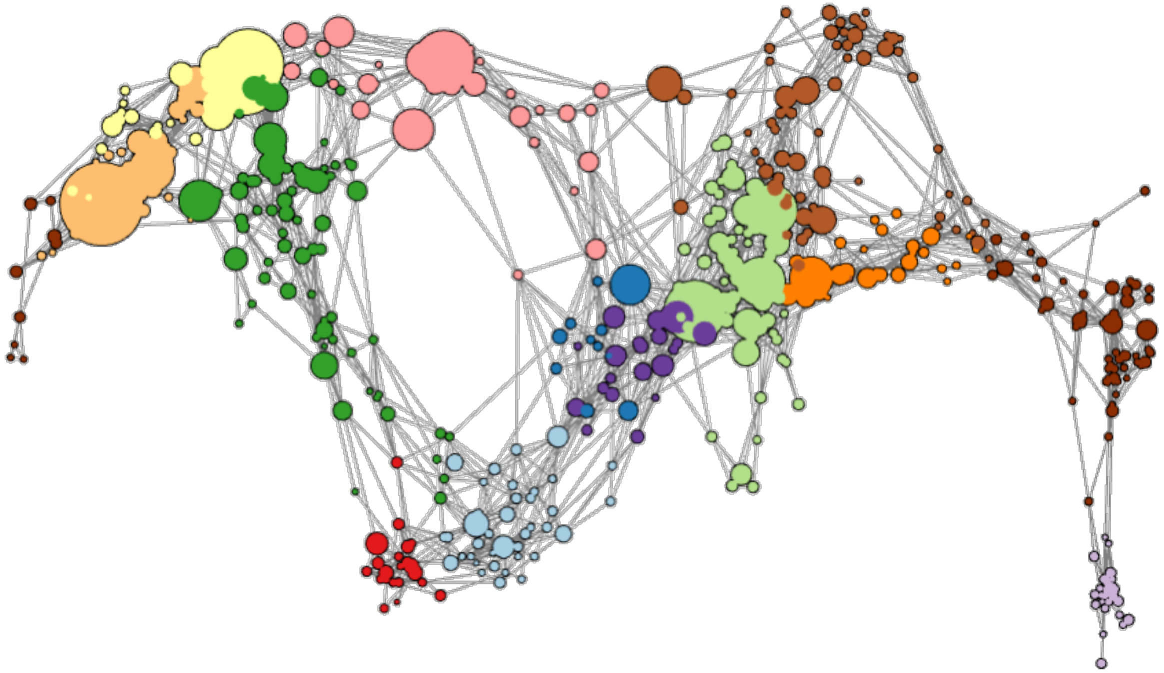
Topology Map
A
Selected Threshold
Selected Maximum
f(x,y)
Persistence
2D Scatterplot
B1
40
y 20
0
0 20 40
x
3D Scatterplot
B2
2.0
1.0
0.0
f(x,y)
0 20 40
x
40 20 0
y
Sensitivity
C
y -0.113
x -0.0284
y 0.127
x -0.0281
y -0.113
x 0.0261
y 0.117
x 0.0233
Histogram
E2
f(x,y)
20
Add Histogram
Persistence Diagram
E1
Selected Threshold
Death
Birth
Persistence Barcode
E3
Selected Maximum
3D Scatterplot ROM
B3
2.0
1.0
0.0
Predicted from Linear Fit
1.7
0.66
-0.34
Predicted from Linear Fit
Fitness
D
0.862
0.913
0.864
0.91
0.851
0.942
0.85
0.913
y x y x y x y x

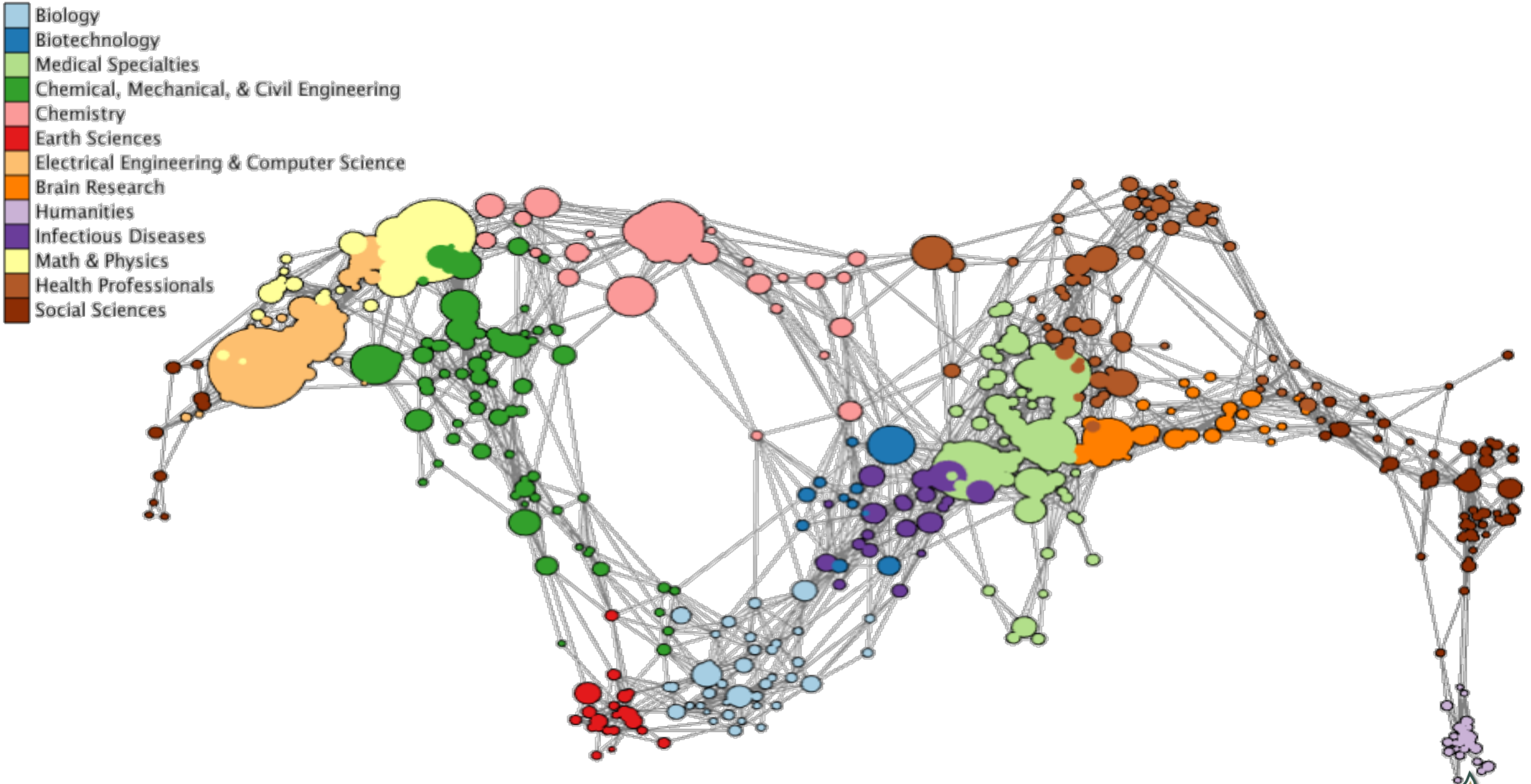
New TDA Applications
Network/Graph Visualization



The Map of Science

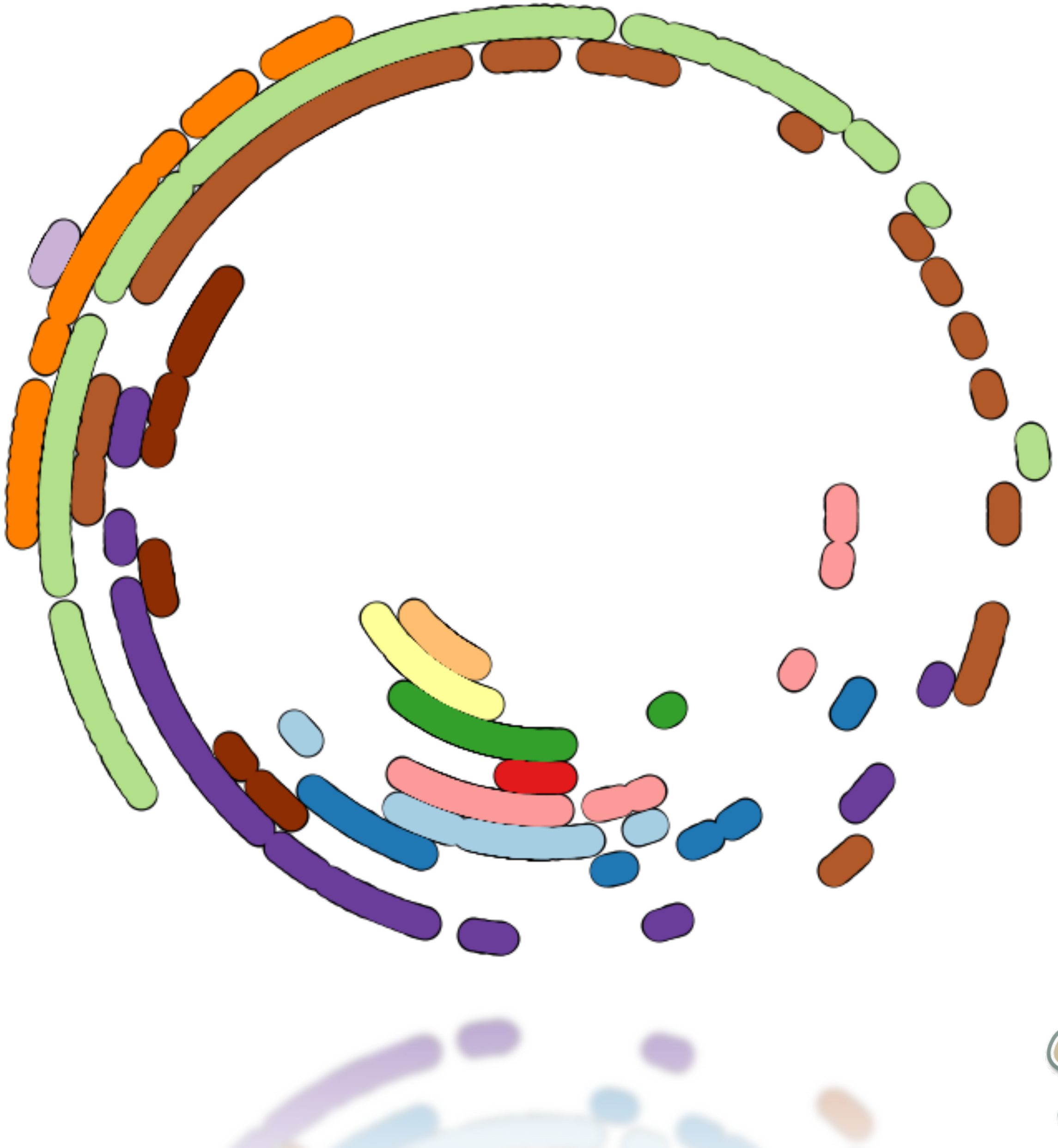
understanding
cross-disciplinary
relationships in science




Biology
Biotechnology
Medical Specialties
Chemical, Mechanical, & Civil Engineering
Chemistry
Earth Sciences
Electrical Engineering & Computer Science
Brain Research
Humanities
Infectious Diseases
Math & Physics
Health Professionals
Social Sciences

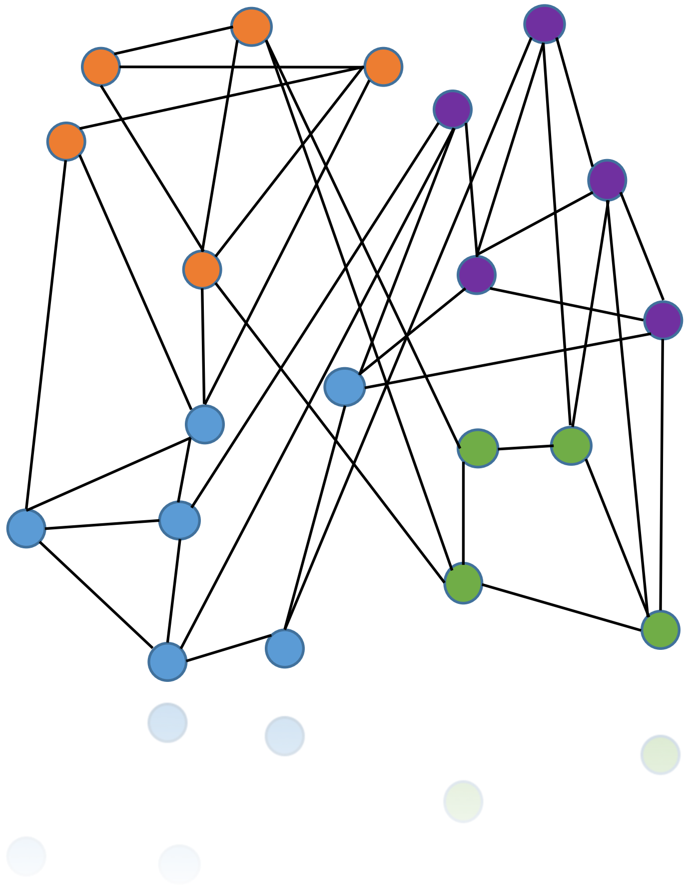
Circular Parameterization

topological analysis tool
that finds circles in
high-dimensional data, projects
data onto those circles,
and visualizes results



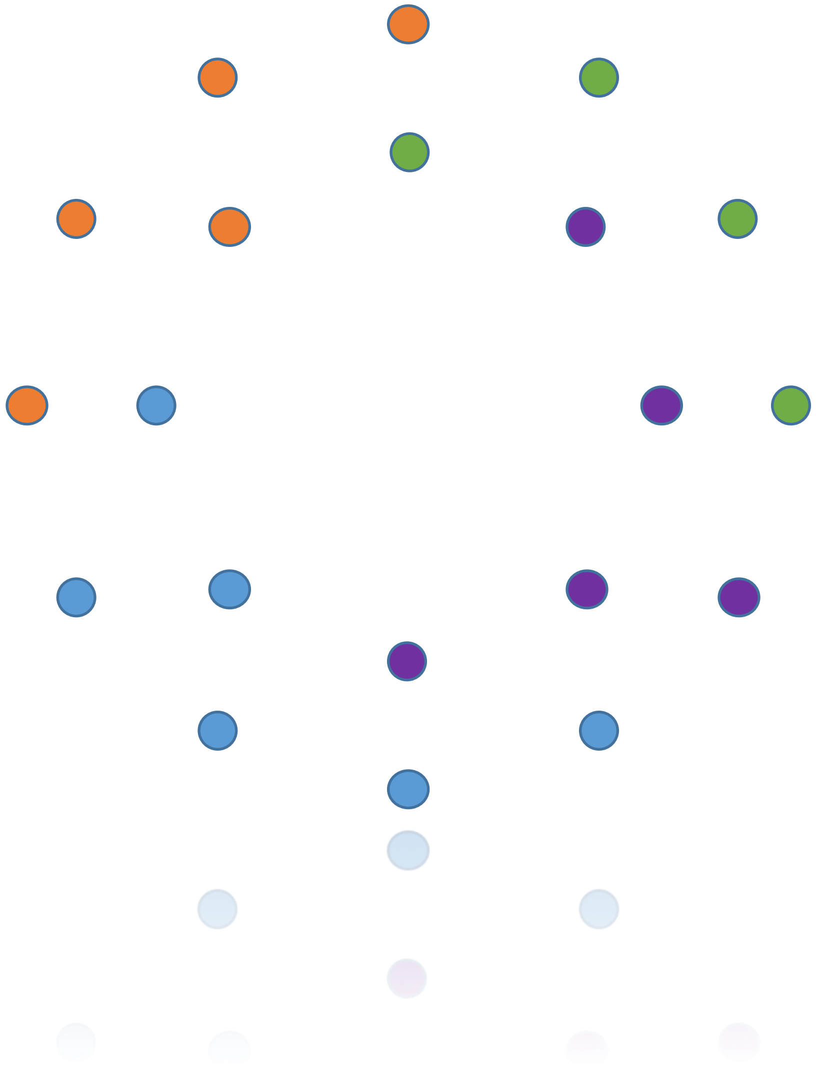
EMBED THE GRAPH

Form a distance matrix to embed in some high dimension



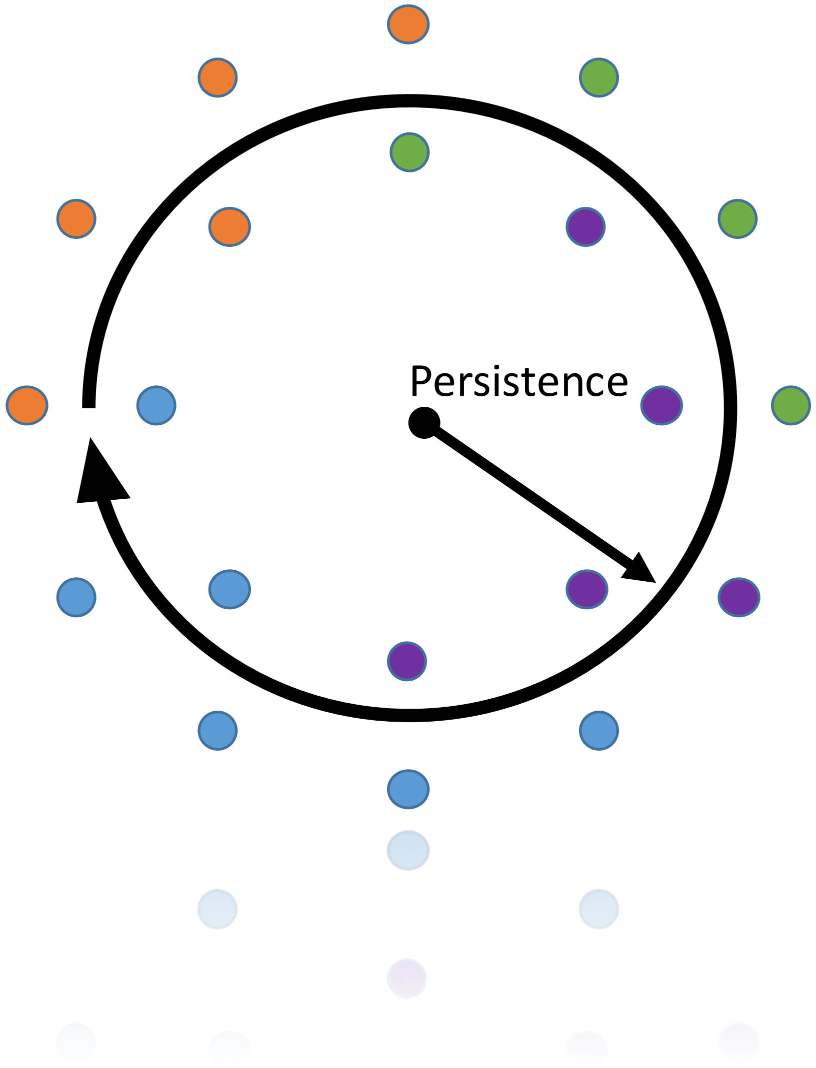
EMBED THE GRAPH

Form a distance matrix to embed in some high dimension



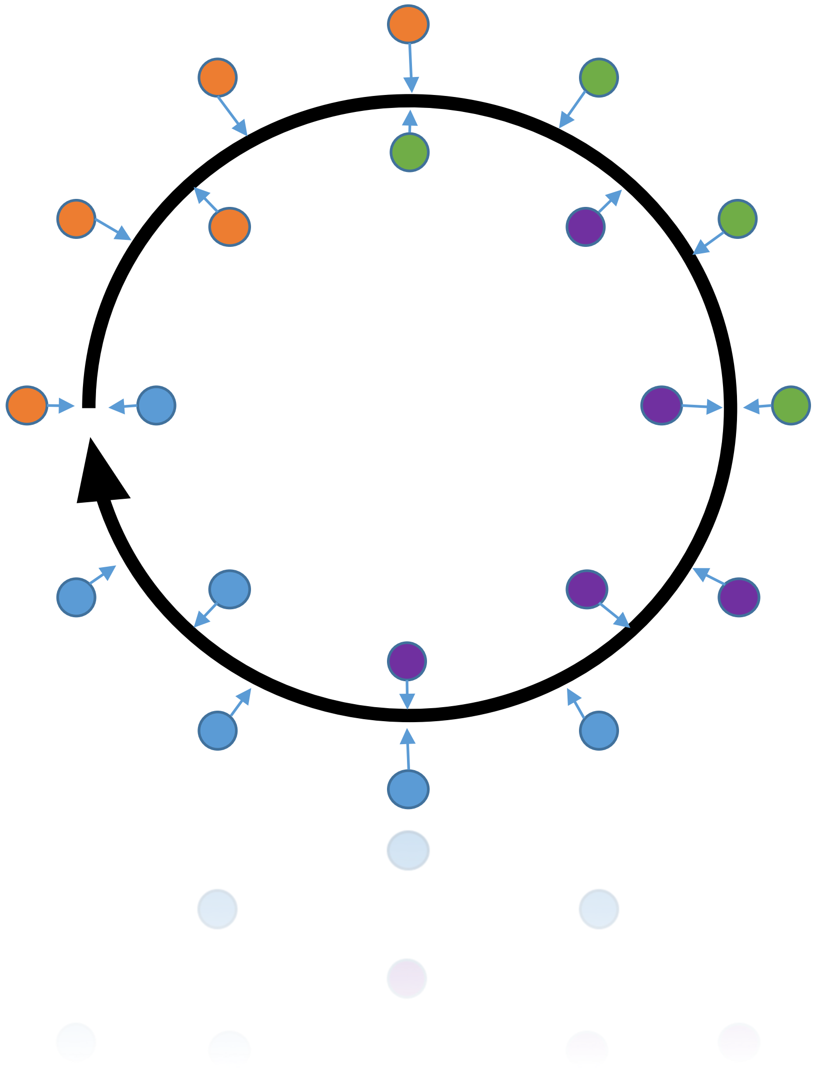
Extract cycles

Find high persistence cylces



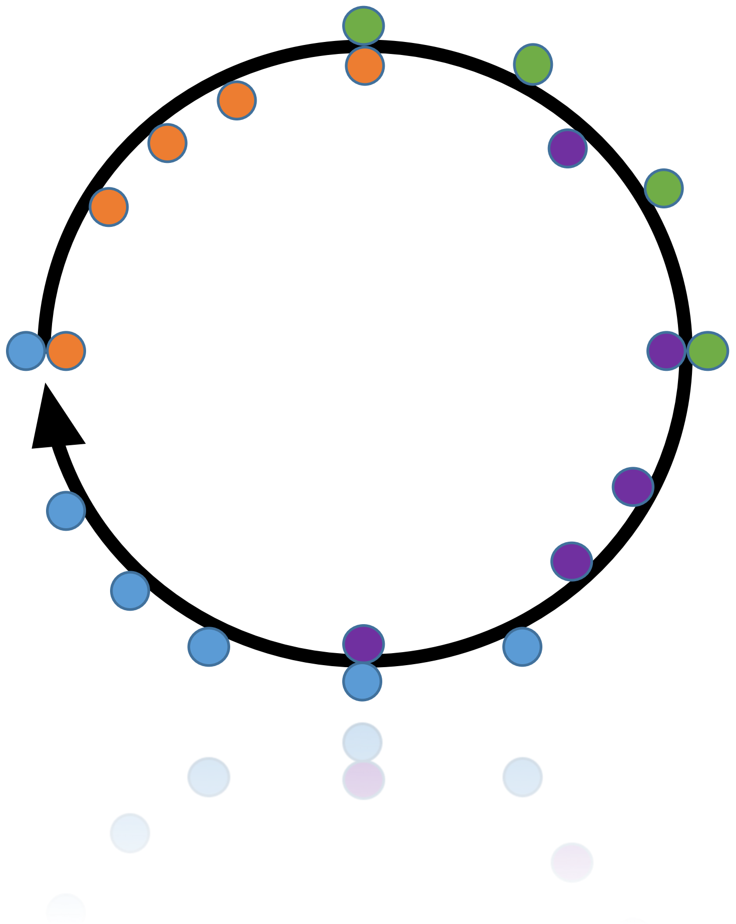
PROJECT THE DATA

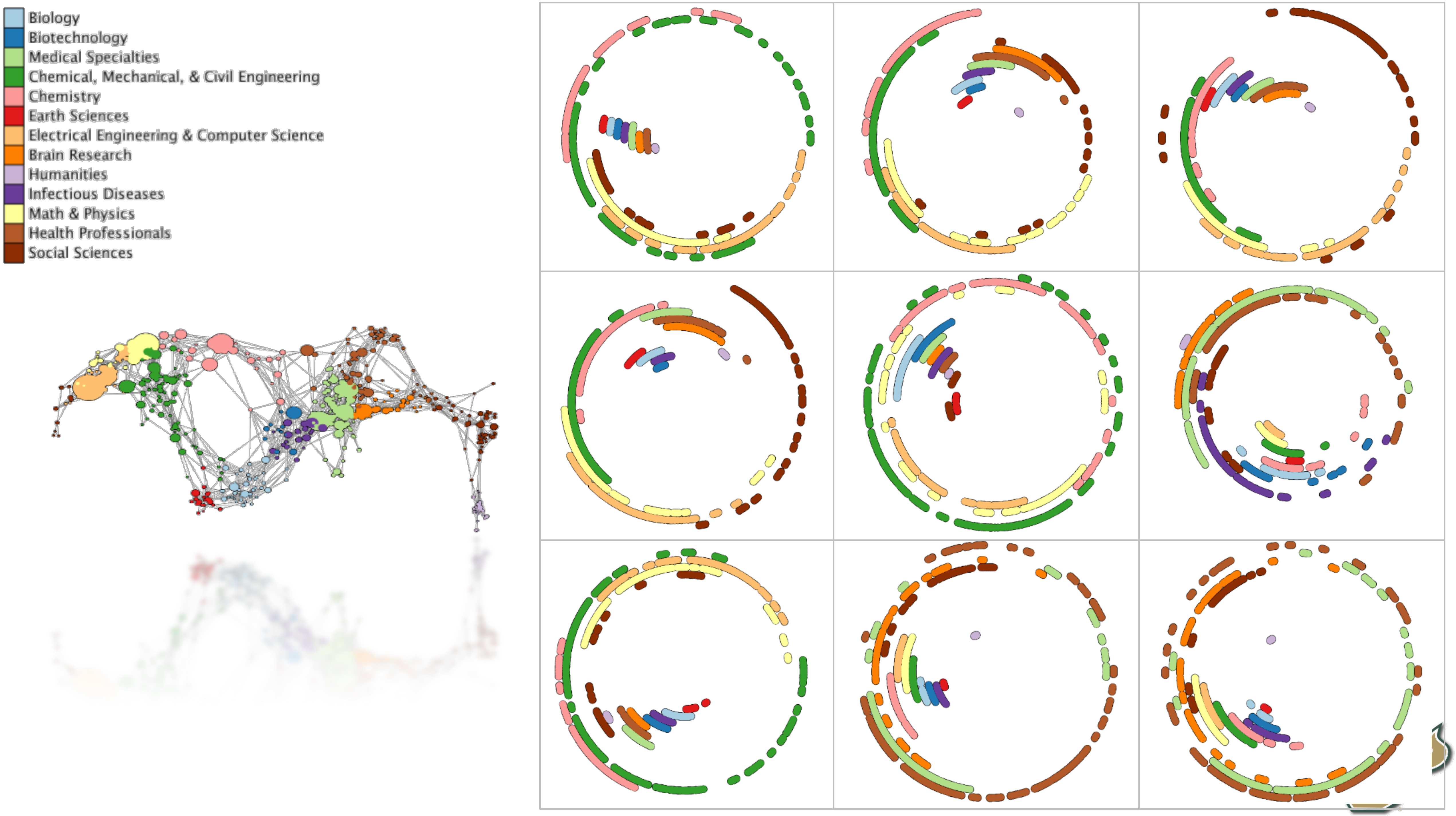
Using a circular parameterization based upon a cohomological reduction method

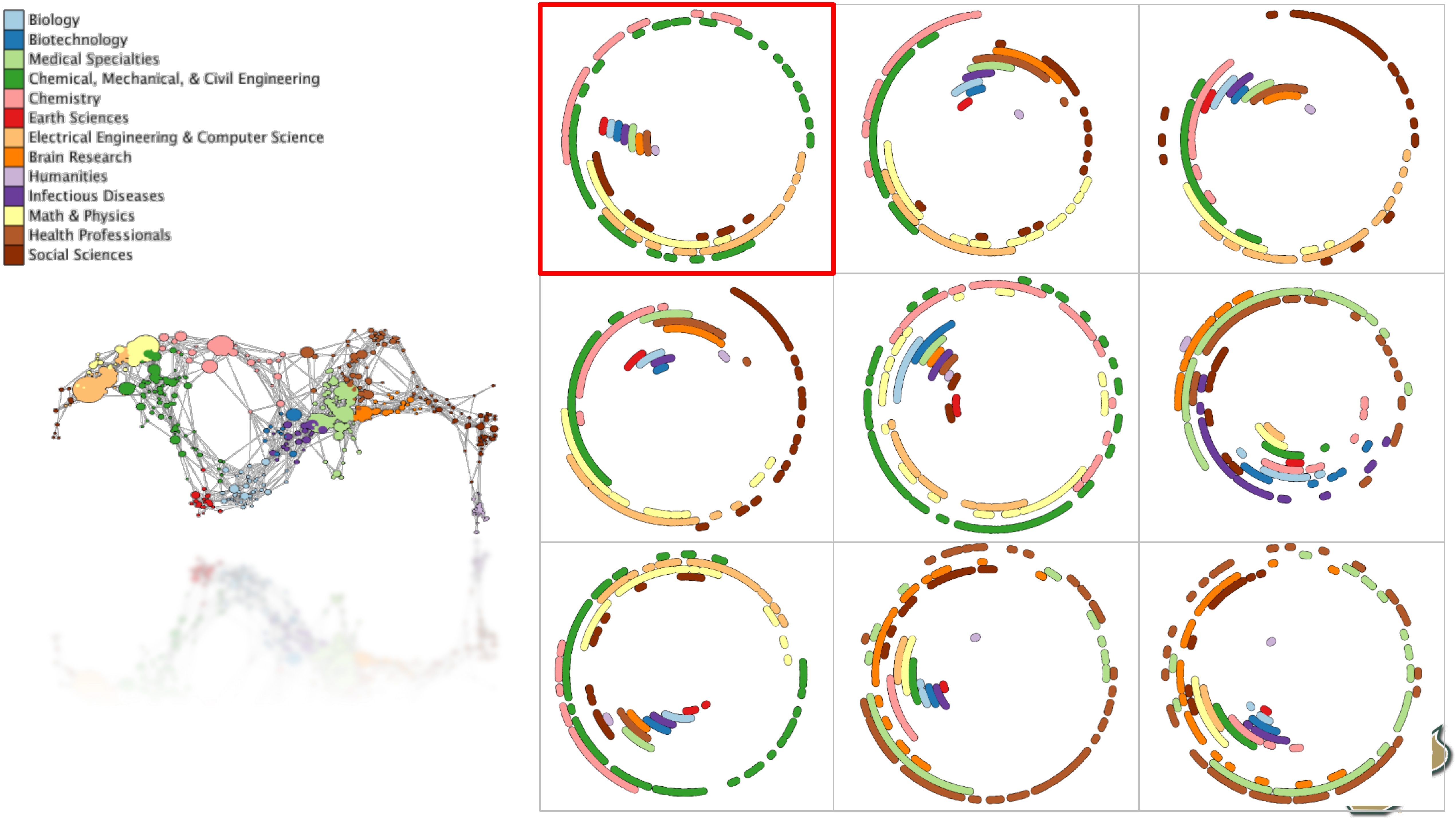


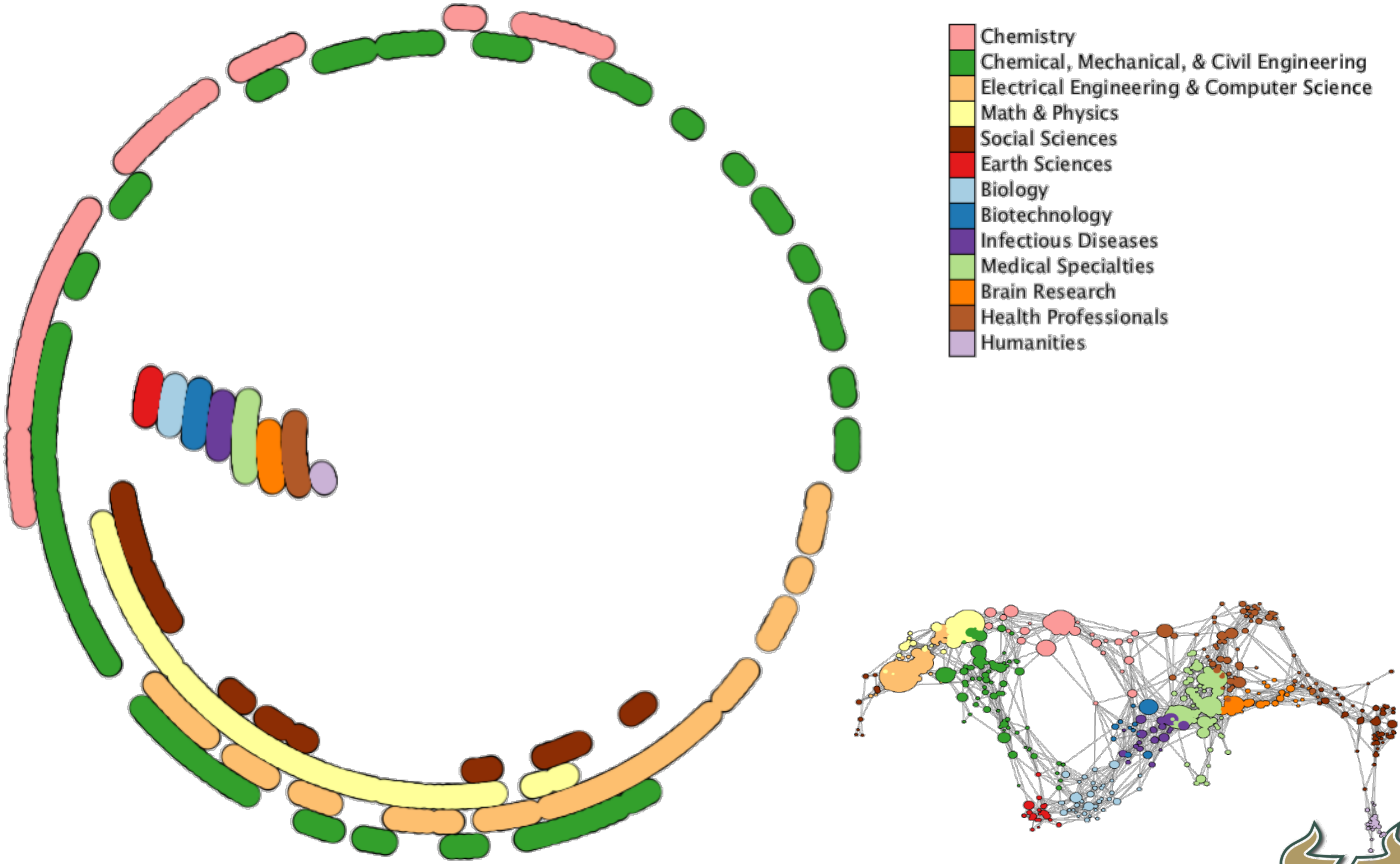
PREMISE

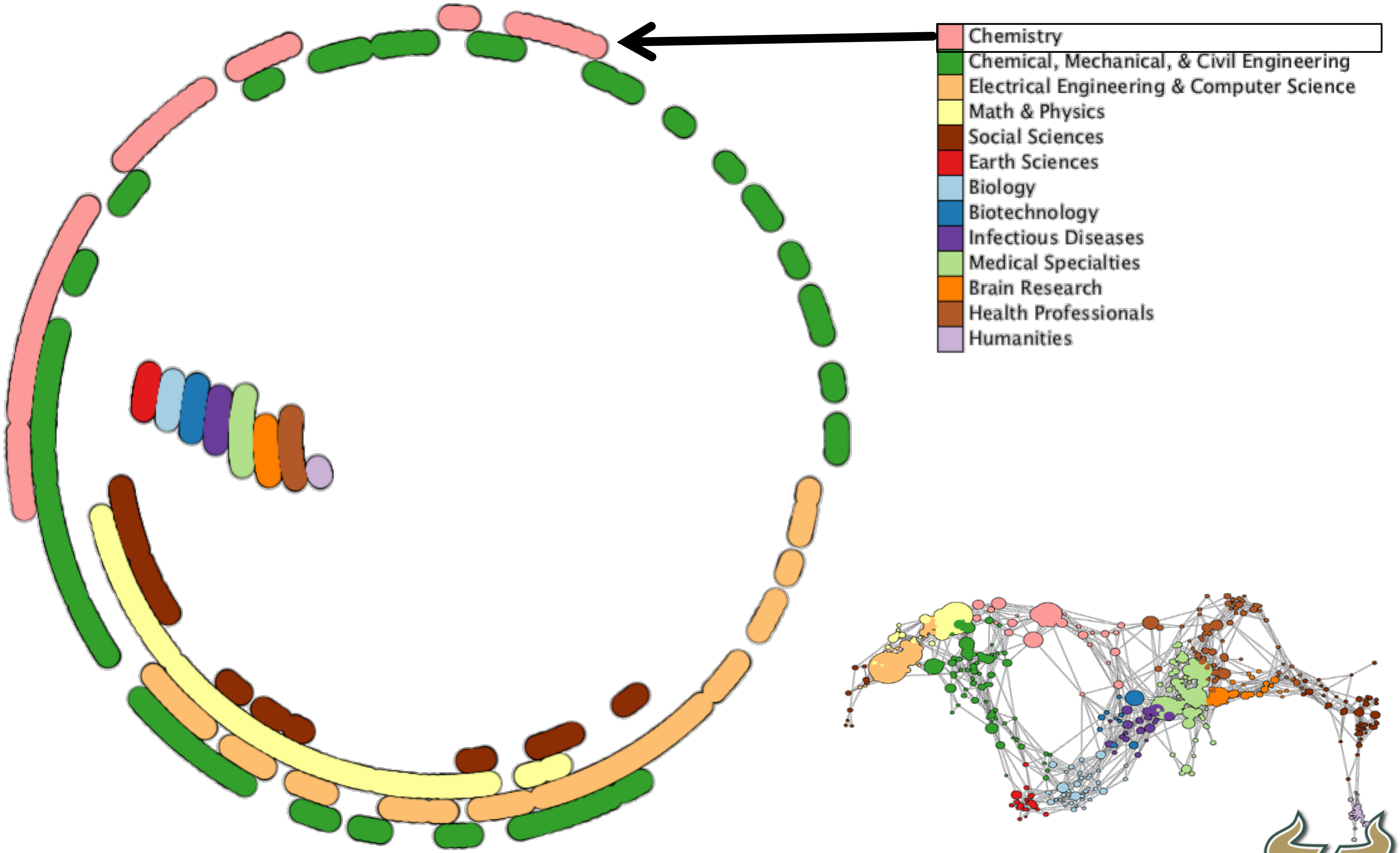
Properly embedded in high dimensional space—items at similar locations along the circle are more likely to interact.

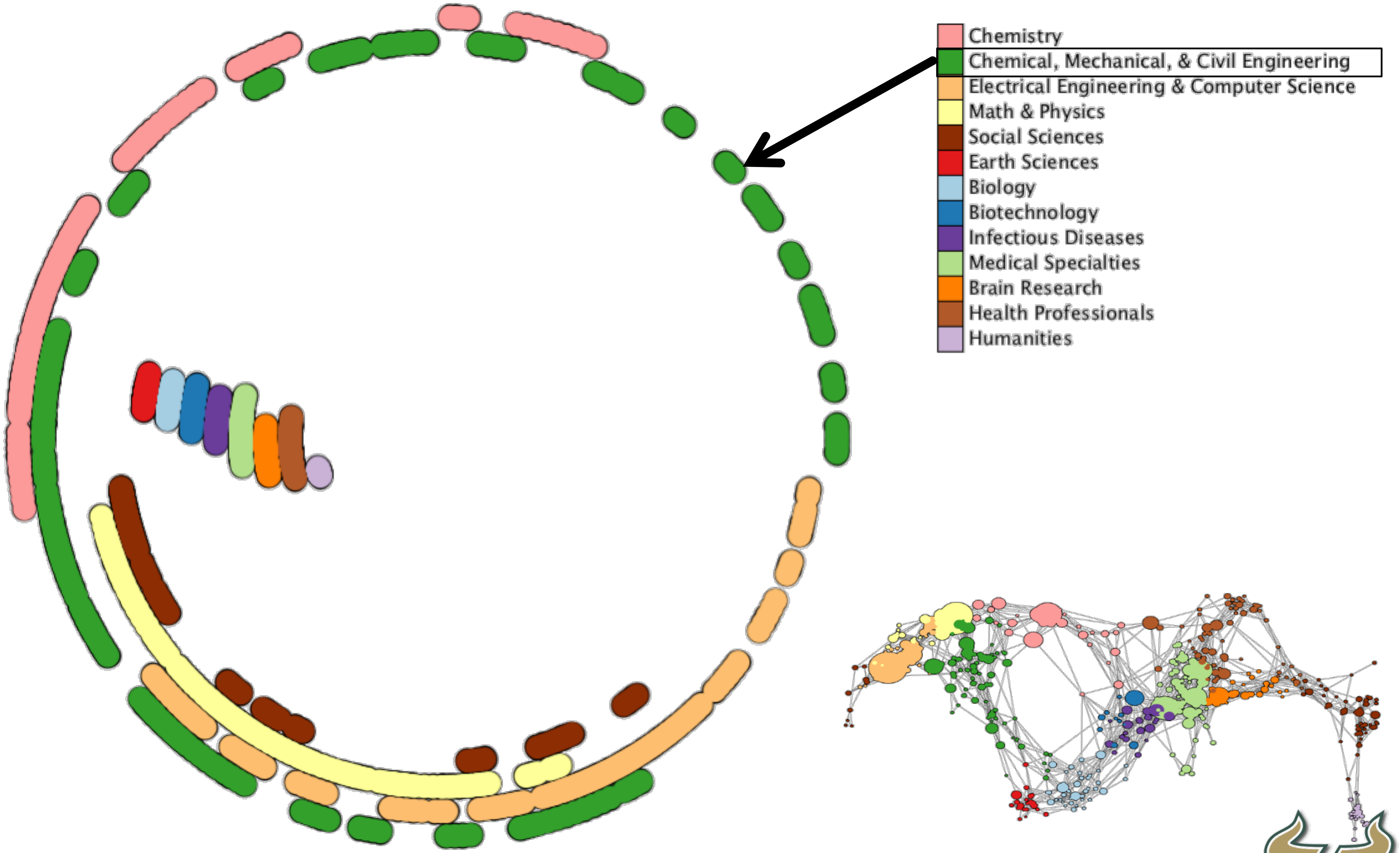



Biology
Biotechnology
Medical Specialties
Chemical, Mechanical, & Civil Engineering
Chemistry
Earth Sciences
Electrical Engineering & Computer Science
Brain Research
Humanities
Infectious Diseases
Math & Physics
Health Professionals
Social Sciences

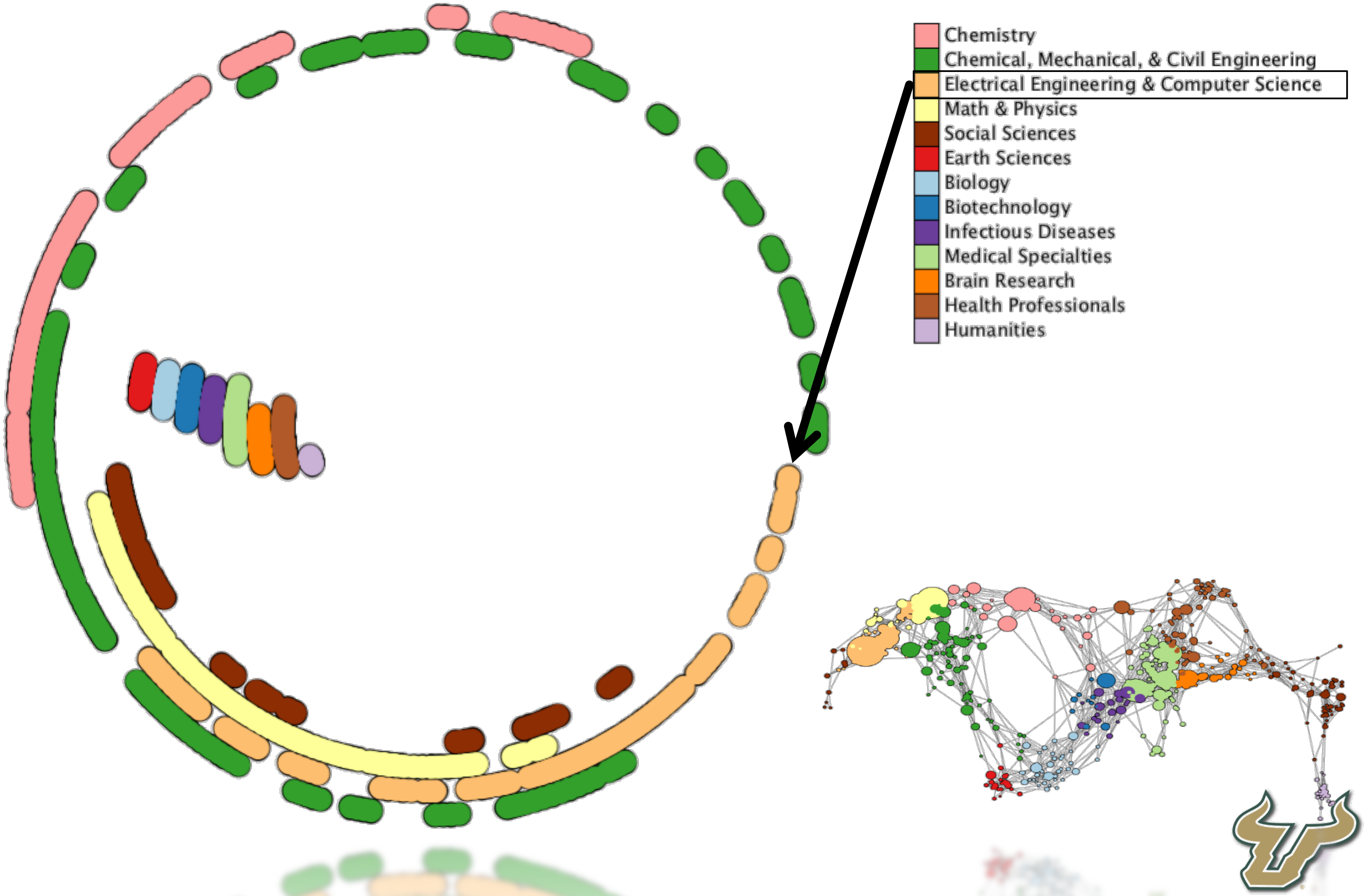

Biology
Biotechnology
Medical Specialties
Chemical, Mechanical, & Civil Engineering
Chemistry
Earth Sciences
Electrical Engineering & Computer Science
Brain Research
Humanities
Infectious Diseases
Math & Physics
Health Professionals
Social Sciences

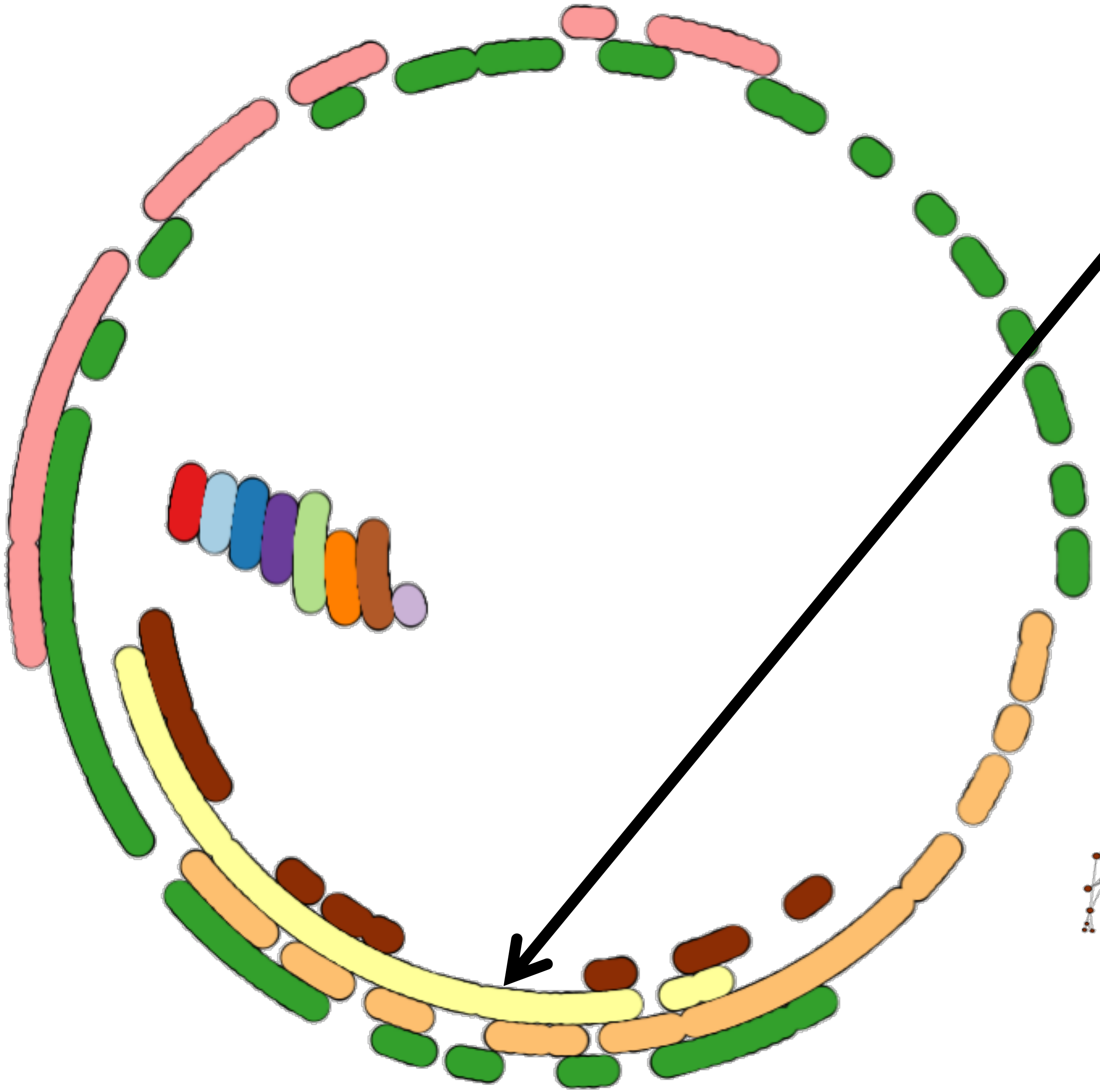

Chemistry
Chemical, Mechanical, & Civil Engineering
Electrical Engineering & Computer Science
Math & Physics
Social Sciences
Earth Sciences
Biology
Biotechnology
Infectious Diseases
Medical Specialties
Brain Research
Health Professionals
Humanities

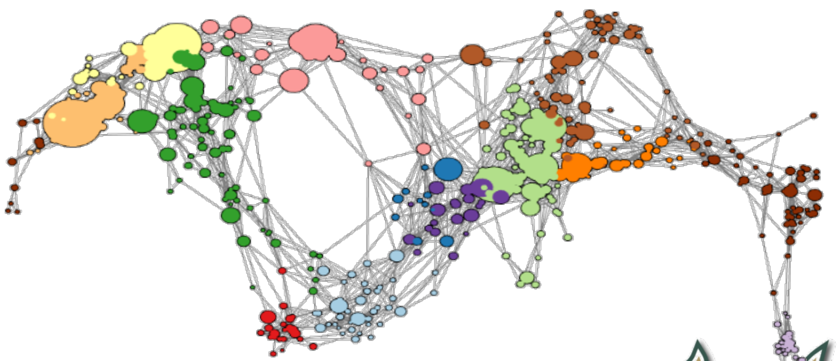

Chemistry
Chemical, Mechanical, & Civil Engineering
Electrical Engineering & Computer Science
Math & Physics
Social Sciences
Earth Sciences
Biology
Biotechnology
Infectious Diseases
Medical Specialties
Brain Research
Health Professionals
Humanities

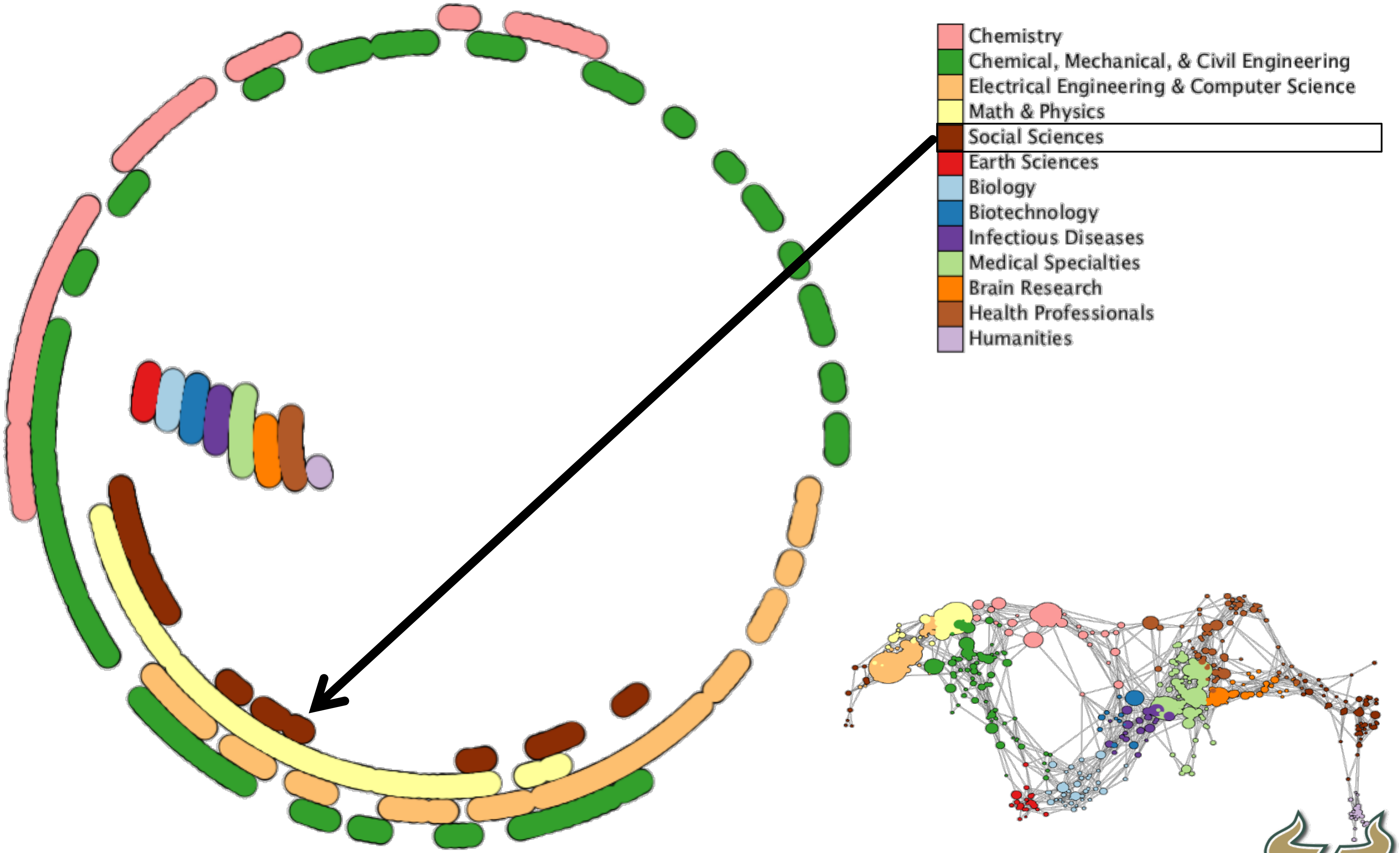


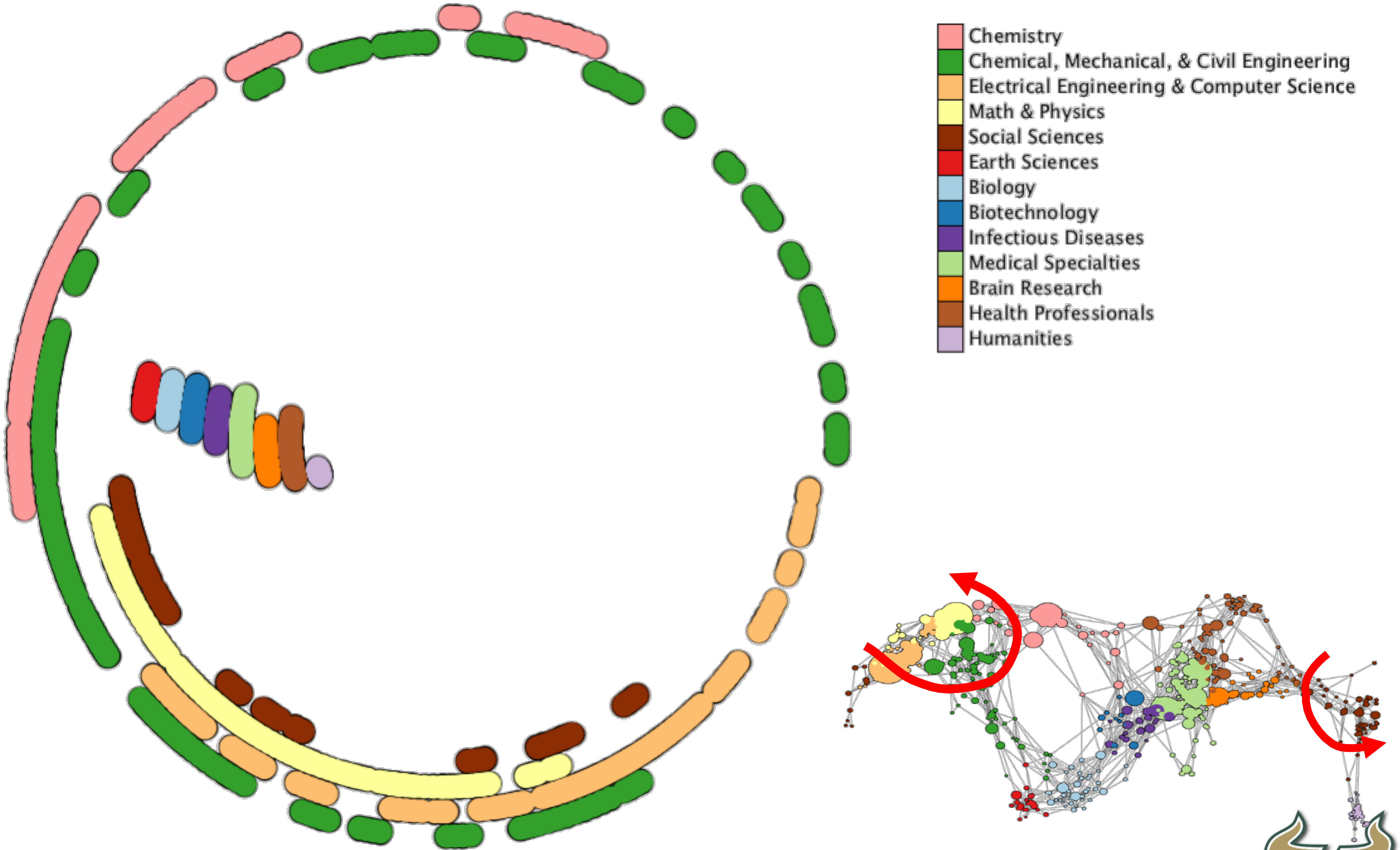
Chemistry
Chemical, Mechanical, & Civil Engineering
Electrical Engineering & Computer Science
Math & Physics
Social Sciences
Earth Sciences
Biology
Biotechnology
Infectious Diseases
Medical Specialties
Brain Research
Health Professionals
Humanities


Chemistry
Chemical, Mechanical, & Civil Engineering
Electrical Engineering & Computer Science
Math & Physics
Social Sciences
Earth Sciences
Biology
Biotechnology
Infectious Diseases
Medical Specialties
Brain Research
Health Professionals
Humanities


Chemistry
Chemical, Mechanical, & Civil Engineering
Electrical Engineering & Computer Science
Math & Physics
Social Sciences
Earth Sciences
Biology
Biotechnology
Infectious Diseases
Medical Specialties
Brain Research
Health Professionals
Humanities



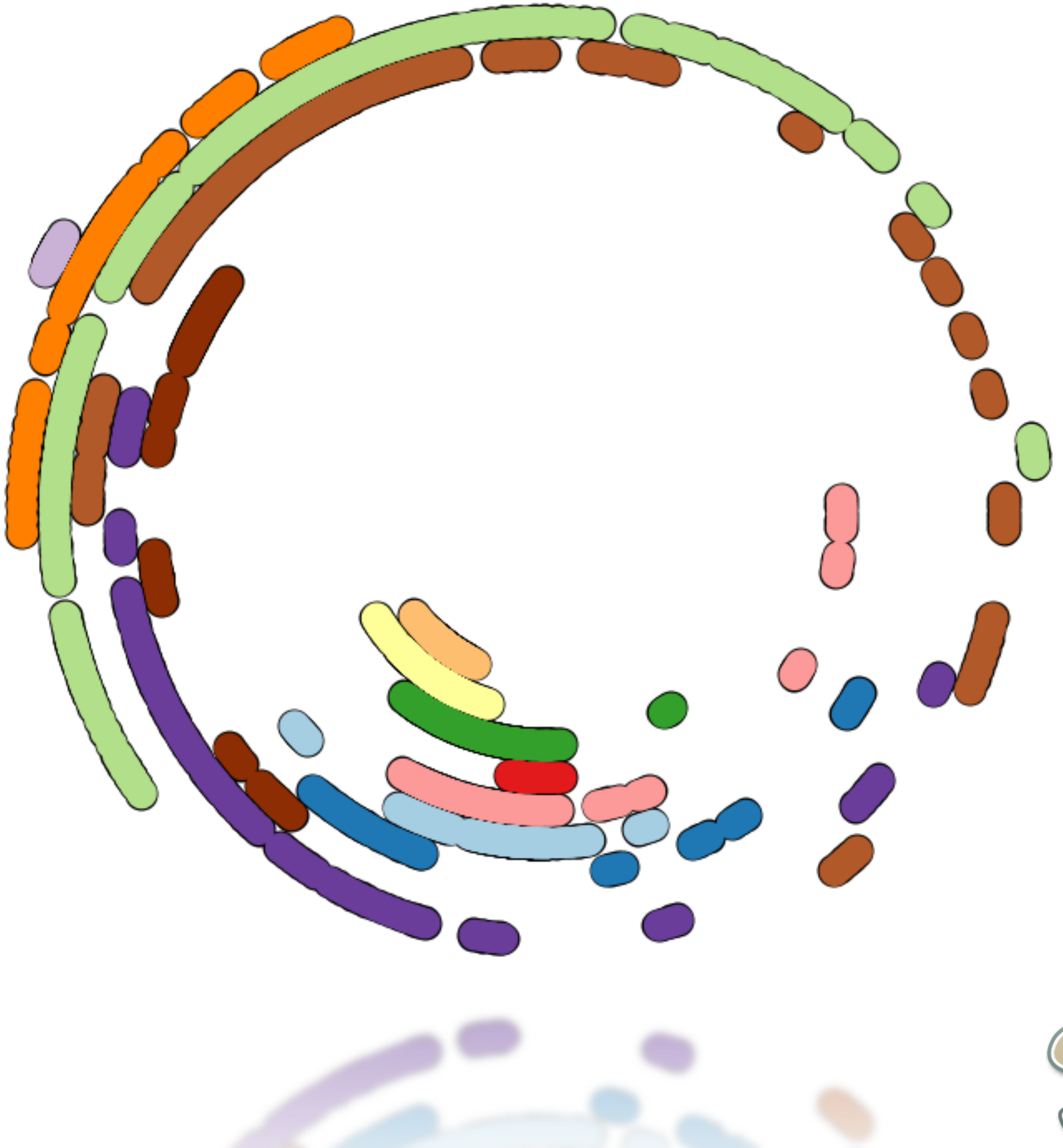

Chemistry
Chemical, Mechanical, & Civil Engineering
Electrical Engineering & Computer Science
Math & Physics
Social Sciences
Earth Sciences
Biology
Biotechnology
Infectious Diseases
Medical Specialties
Brain Research
Health Professionals
Humanities


Chemistry
Chemical, Mechanical, & Civil Engineering
Electrical Engineering & Computer Science
Math & Physics
Social Sciences
Earth Sciences
Biology
Biotechnology
Infectious Diseases
Medical Specialties
Brain Research
Health Professionals
Humanities

WHY TOPOLOGICAL DATA ANALYSIS

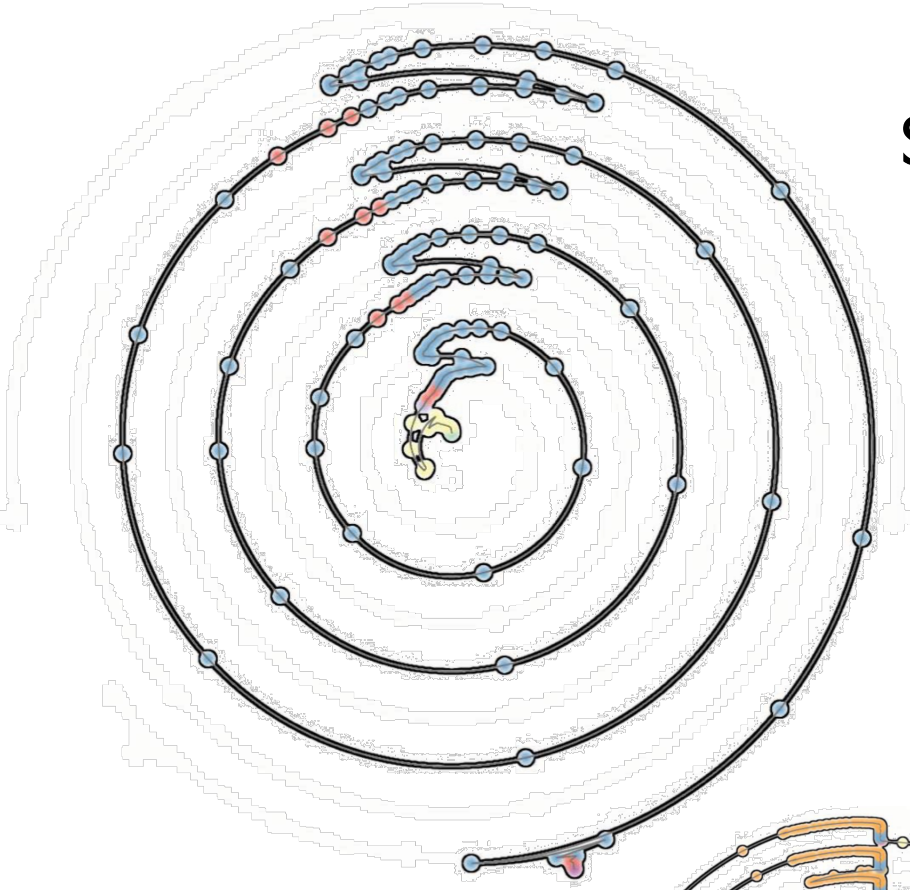
reduced visual complexity to aid exploration that may otherwise be impossible

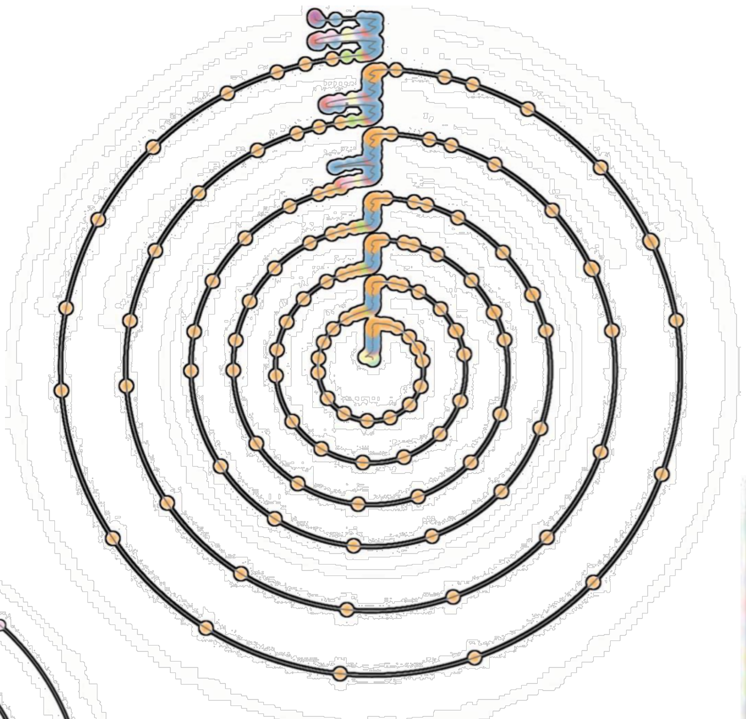
RA/Postdoc positions available to work on the problem

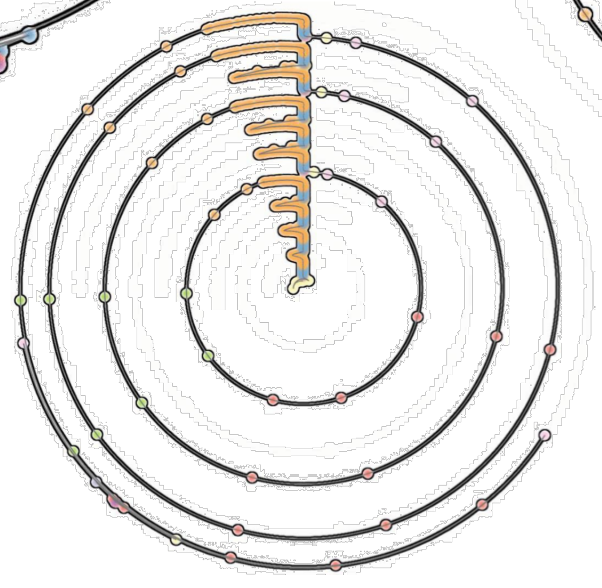


Successful TDA Applications

Software Performance Analysis: Cyclical Behavior in Memory Reference Traces







```
File: sort.cpp
 1:void bubblesort(std::vector<double>& v){
 2:  for(unsigned end=v.size()-1; end >= 0; end--){
 3:    bool swapped = false;
 4:    for(unsigned i=0; i<end; i++){
 5:      if(v[i] > v[i+1]){
 6:        std::swap(v[i], v[i+1]);
 7:        swapped = true;
 8:      }
 9:    }
10:    if(!swapped) break;
11:  }
12:}
```

A.N.M. I. Choudhury, B. Wang, P. Rosen, and V. Pascucci. Topological analysis and visualization of cyclical behavior in memory reference traces. In IEEE Pacific Visualization Symposium, PacificVis, pages 9 -16, 2012.

Conclusions

Application of TDA to visualization is a relatively new approach for taming big data problems

Highly applicable to *certain classes of data* (essentially data embeddable in an N-dimensional space) and *certain classes of visual analysis task* (analyzing certain types of multiscale structure)

Coauthors, Collaborators, & Students:

On Everything
Bei Wang, University of Utah

Radio Astronomy
Chris Johnson, University of Utah
Jeff Kern, NRAO
Betsy Mills, NRAO

Network Graph
Mustafa Hajij – University of S. Florida
Carlos Scheidegger – University of Arizona

Vector Field
Harsh Bhatia – LLNL
Roxana Bujack, UC Davis
Guoning Chen – University of Houston
Valerio Pascucci – University of Utah
Primoz Skraba – Jozef Stefan Institute

Software Visualization
Roni Choudhury – Kitware, Inc.

Nuclear Simulation
Andrea Alfonsi – Idaho National Lab
Dan Maljovec – University of Utah
Valerio Pascucci – University of Utah
Giovanni Pastore – Idaho National Lab
Cristian Rabiti – Idaho National Lab

Funding Acknowledgements

Lawrence Livermore National Laboratory

Pacific Northwest National Laboratory Analysis In Motion (AIM) Initiative

NSF Information Integration & Informatics, Algorithmic Foundations (IIS-1513616)

National Radio Astronomy Observatory (ALMA Development Study)

QUESTIONS?

CONTACT:

Paul Rosen

prosen@usf.edu

http://www.cspaul.com