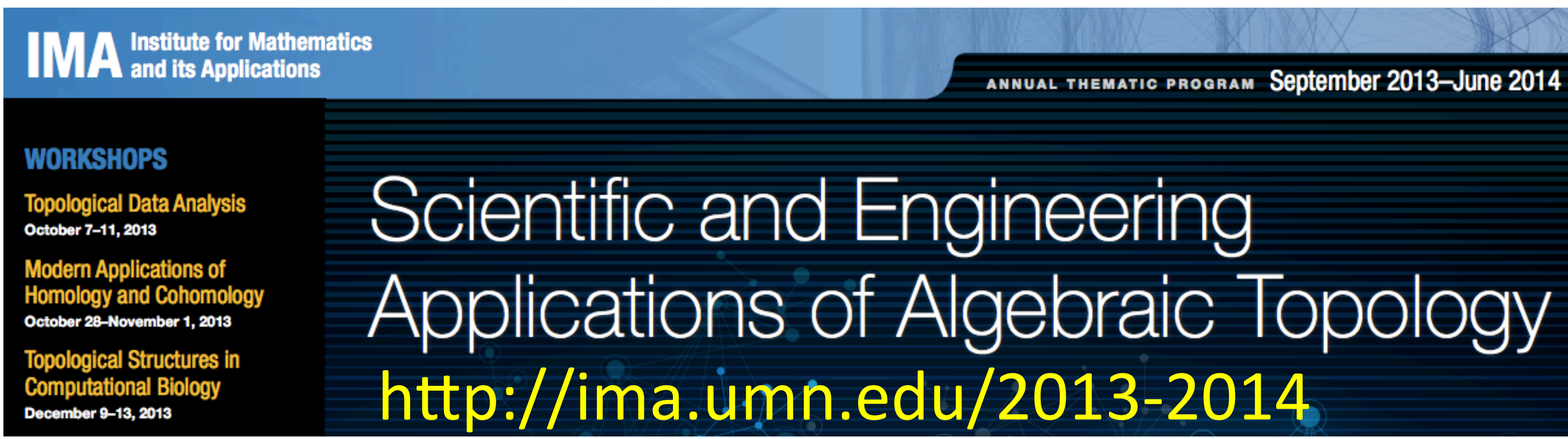
Welcome to

MATH:7450 (22M:305) Topics in Topology: Scientific and Engineering Applications of Algebraic Topology

Week 1: Introduction to Topological Data Analysis via Mapper Software

Sept: Persistent Homology plus topics from student and speaker input.



Topological Data Analysis

October 7-11, 2013

PROGRAM APPLICATION

Organizers

Robert Adler, Technion, Israel
Gunnar Carlsson, Stanford University
John Harer, Duke University

Modern Applications of Homology and Cohomology

October 28 to November 1, 2013

PROGRAM APPLICATION

Organizers

Andrew Blumberg, University of Texas at Austin
Lek-Heng Lim, University of Chicago
Yuan Yao, Peking University

Topological Structures in Computational Biology

December 9-13, 2013

PROGRAM APPLICATION

Organizers

Gunnar Carlsson, Stanford University
Christine Heitsch, Georgia Institute of Technology
Susan Holmes, Stanford University
Konstantin Mischaikow, Rutgers, The State University of New Jersey

Are you interested in analyzing data?

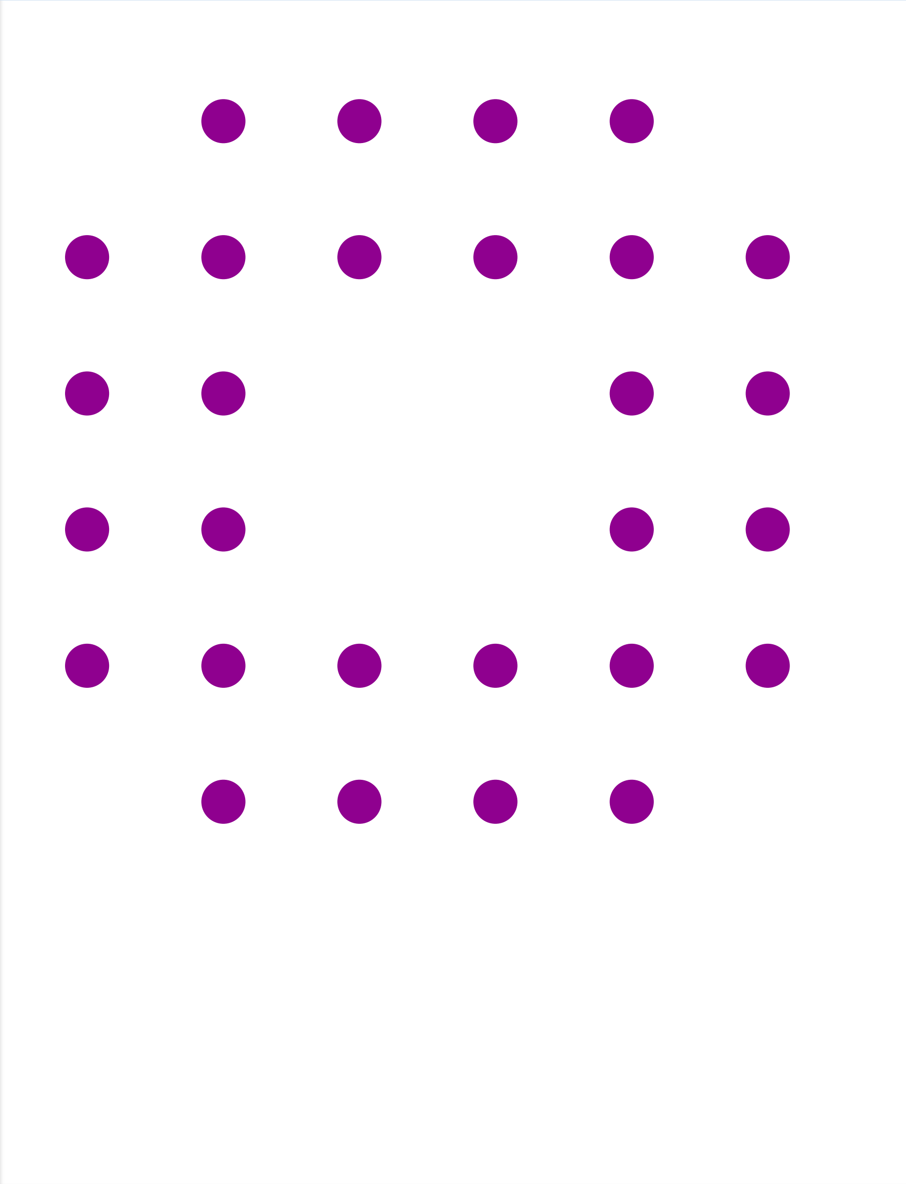
Do you have data to analyze?

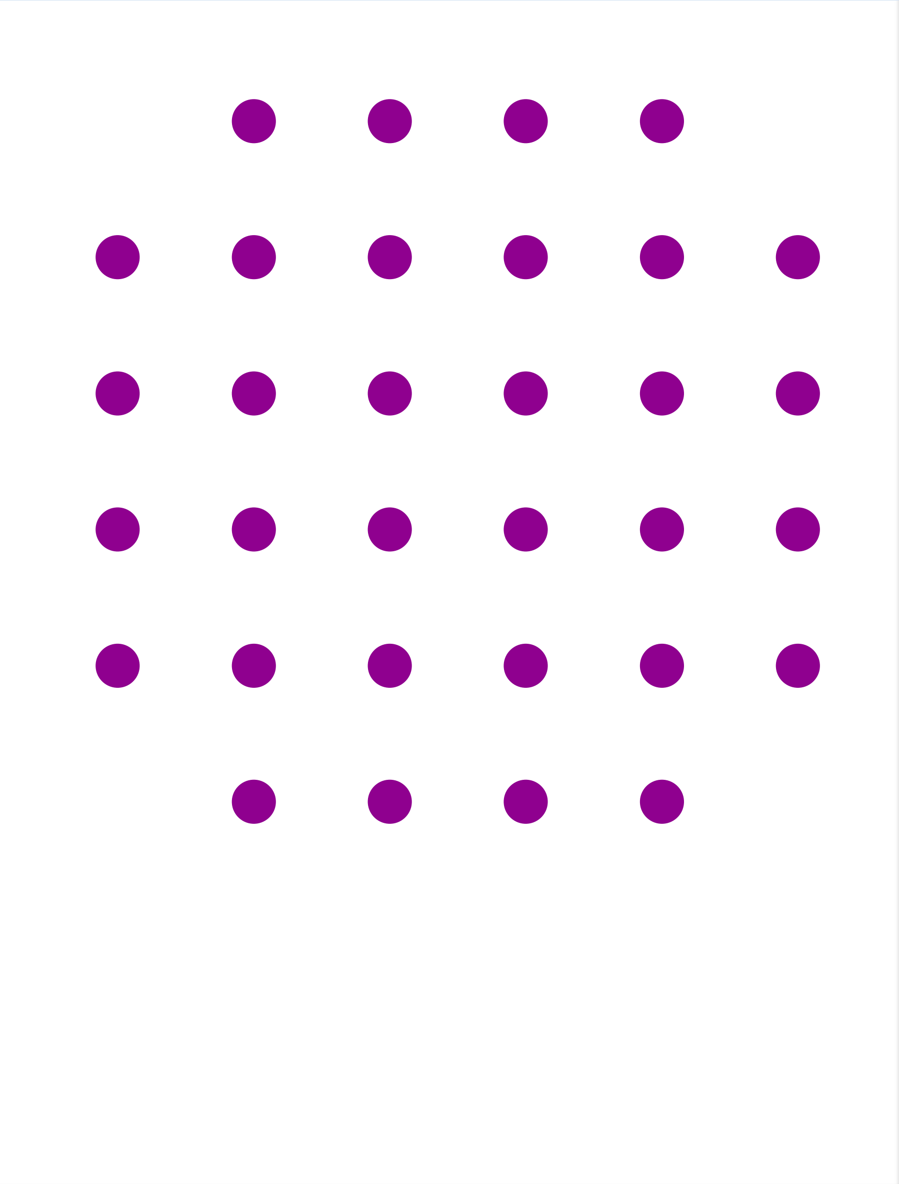
Would you like collaborators?

If so, let me know by mid-September.

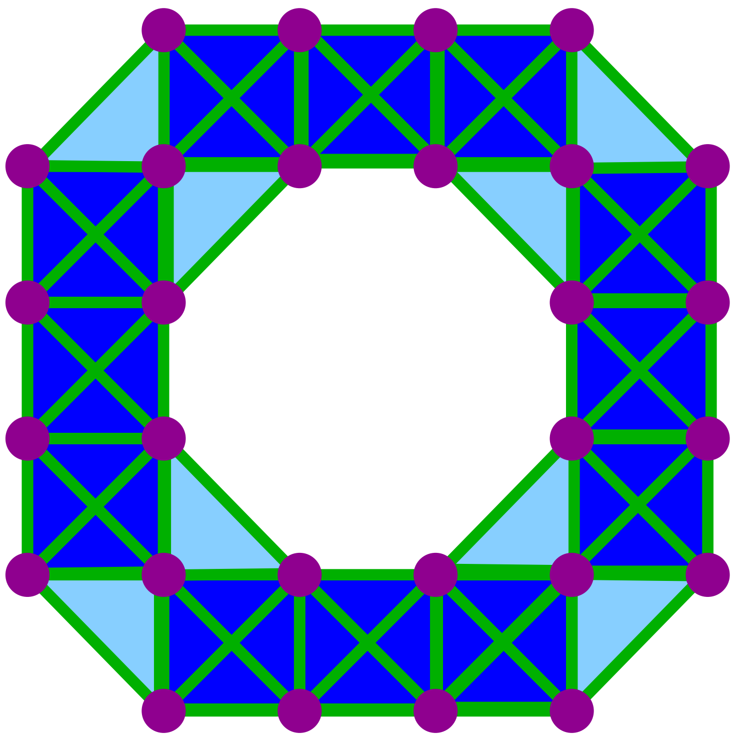
Do you have any recommendations regarding online (or offline) material?

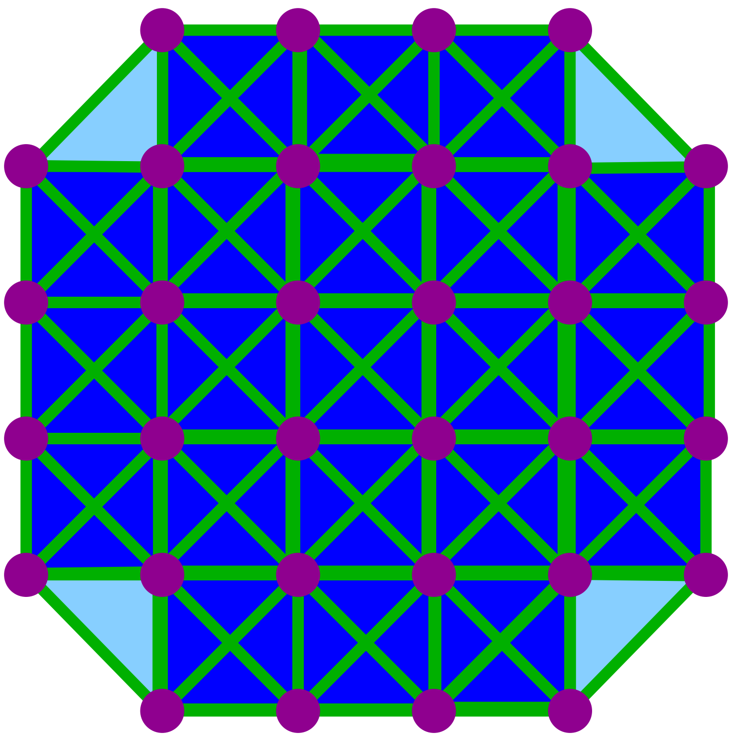
From Preparatory Lecture 6
Creating a simplicial complex from data





From Preparatory Lecture 6
Creating a simplicial complex from data



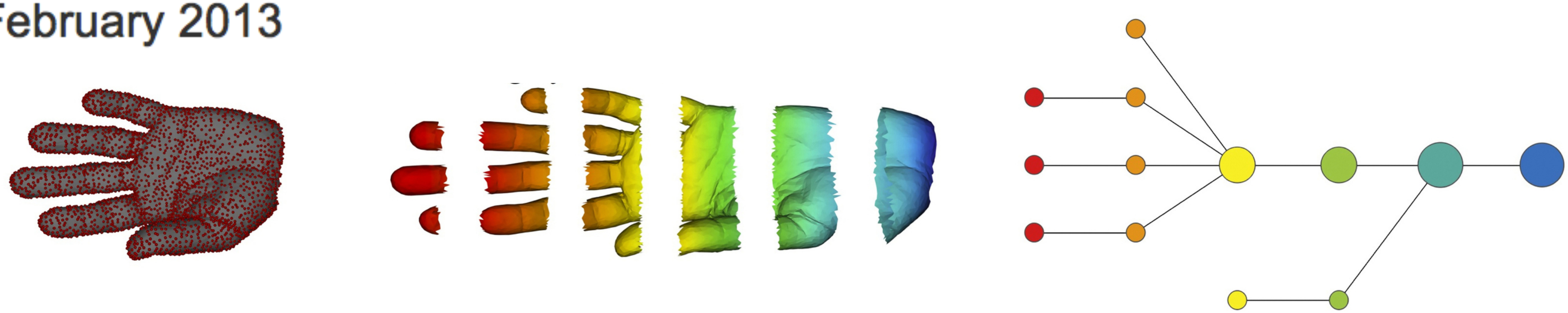


Extracting insights from the shape of complex data using topology

P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson & G. Carlsson

Affiliations | Contributions | Corresponding authors

Scientific Reports **3**, Article number: 1236 | doi:10.1038/srep01236
Received 13 September 2012 | Accepted 06 December 2012 | Published 07 February 2013



http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

A Original Point Cloud

A) Data Set

Example: Point cloud data representing a hand.

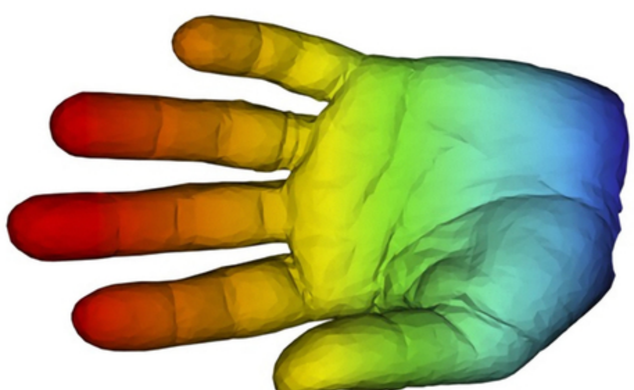
http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

A Original Point Cloud

Data Set:

Example: Point cloud data representing a hand.

B Coloring by filter value



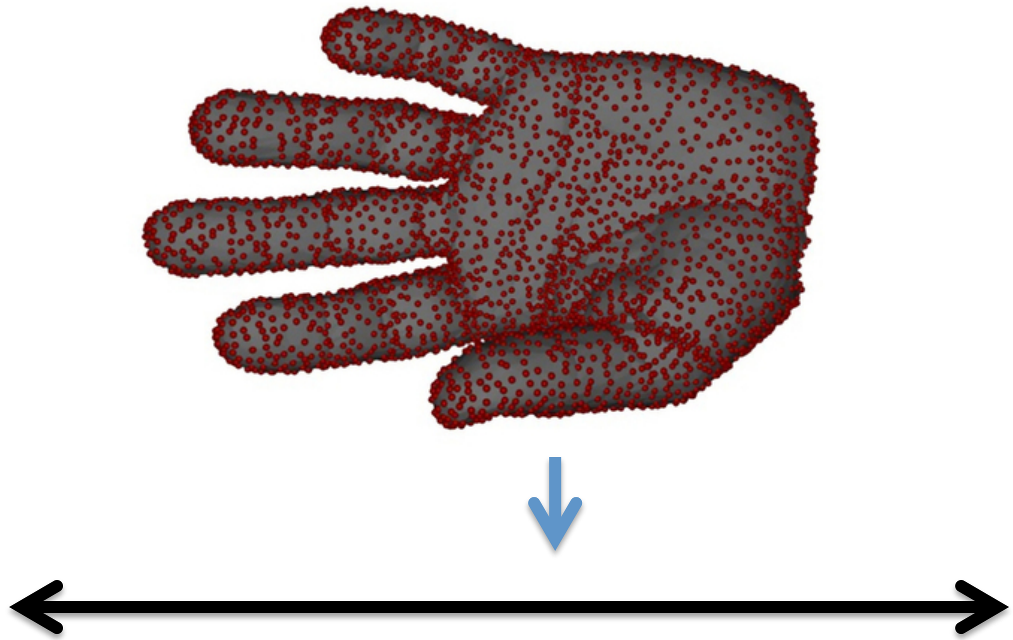
Function f : Data Set → **R**

Example: x-coordinate

f : (x, y, z) → x

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value



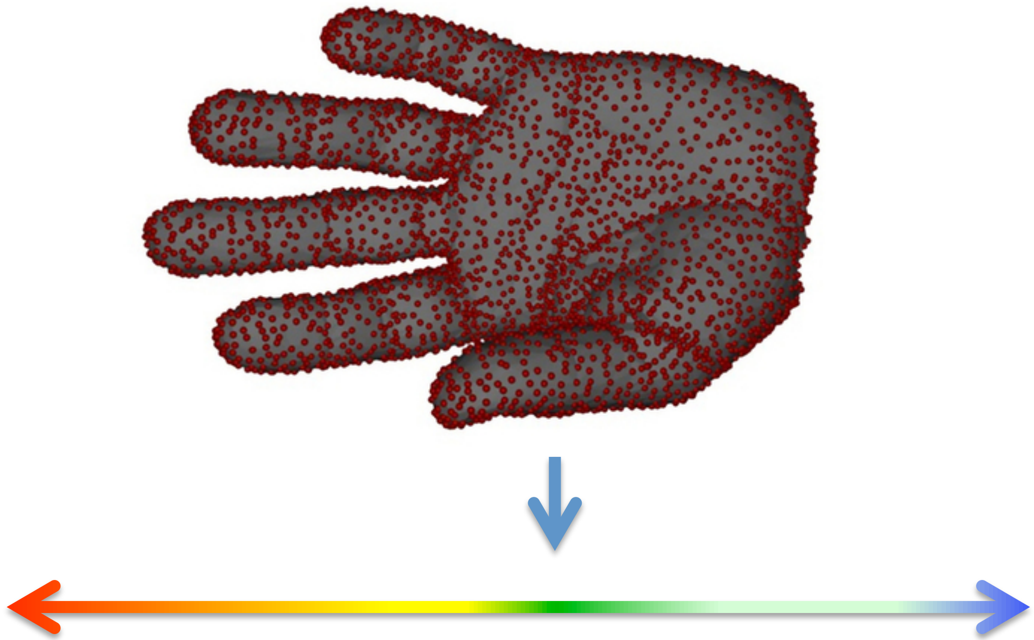
Function f : Data Set → **R**

Ex 1: x-coordinate

$f : (x, y, z) \rightarrow x$

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value



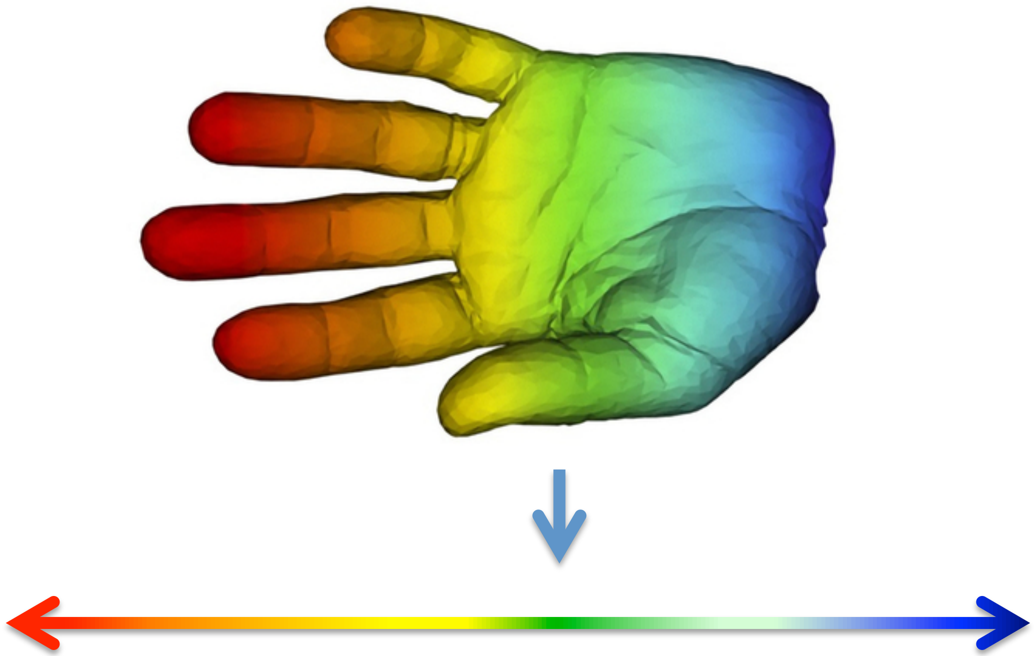
Function f : Data Set → **R**

Ex 1: x-coordinate

$f : (x, y, z) \rightarrow x$

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value



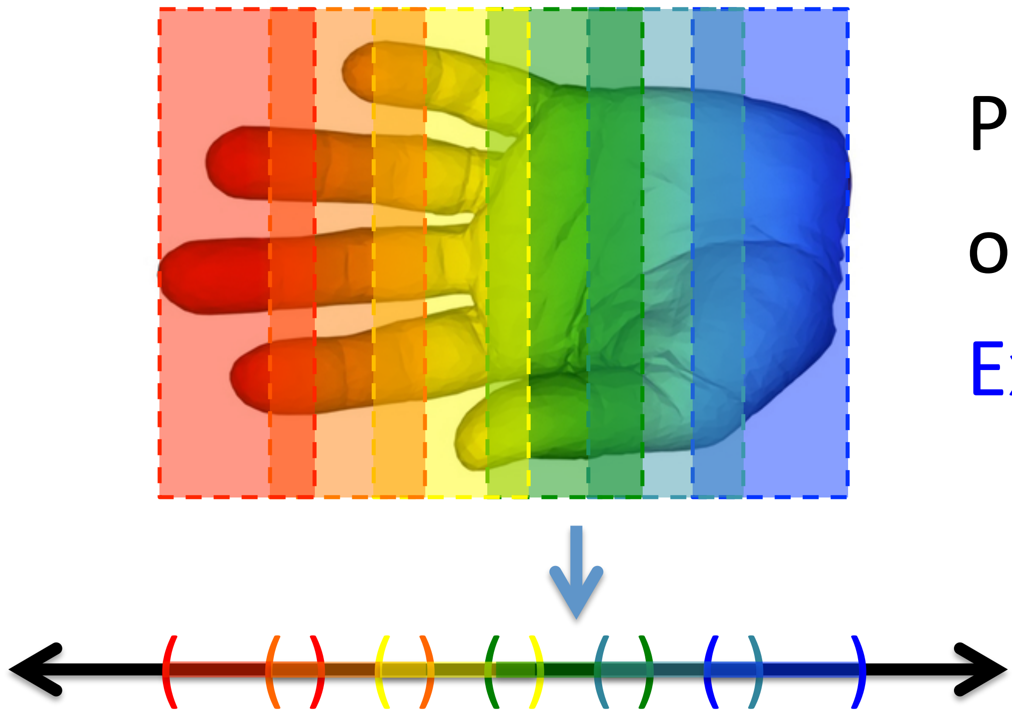
Function f : Data Set → **R**

Ex 1: x-coordinate

$f : (x, y, z) \rightarrow x$

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value



Put data into overlapping bins.

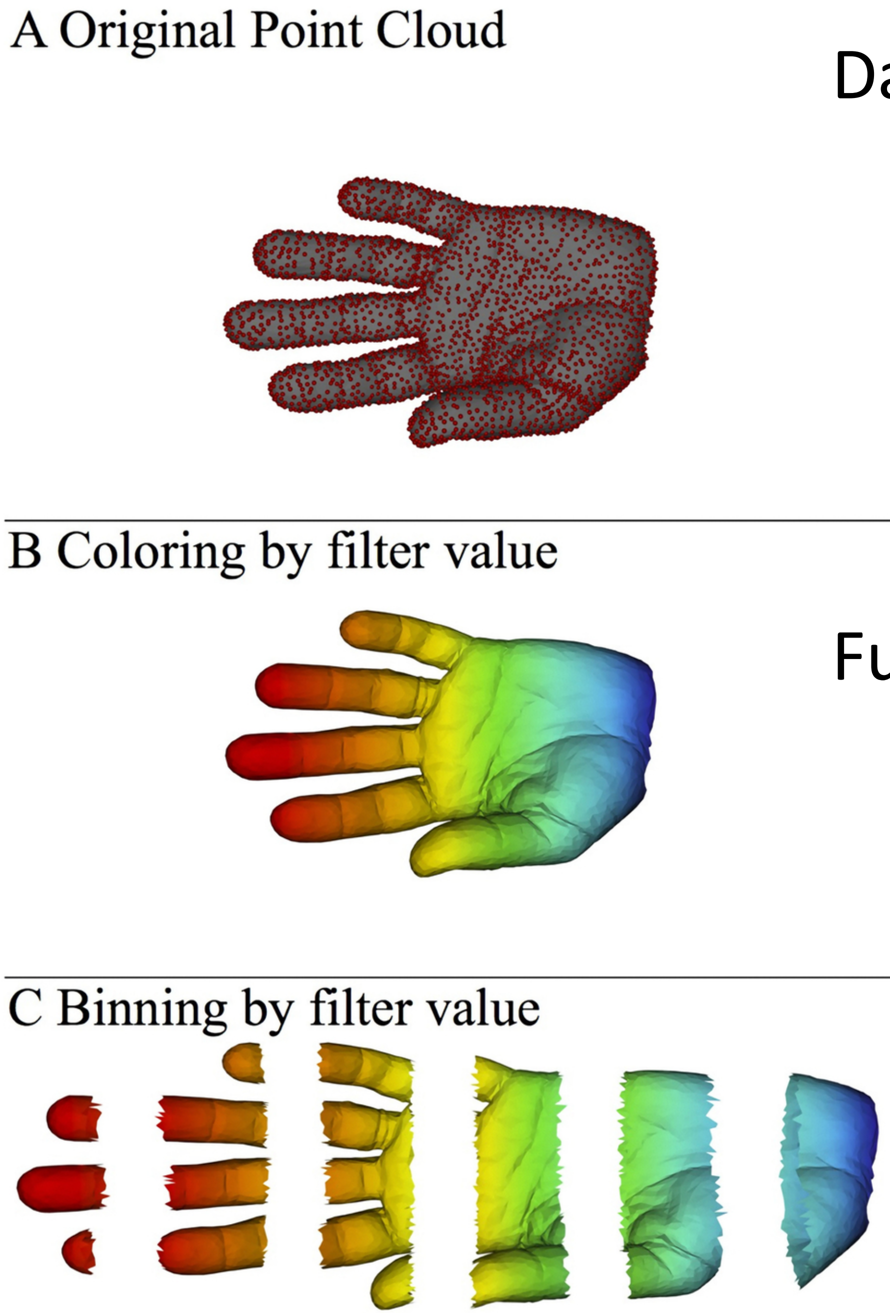
Example: $f^{-1}(a_i, b_i)$

Function f : Data Set → **R**

Ex 1: x-coordinate

$f : (x, y, z) \rightarrow x$

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html



Data Set

Example: Point cloud data representing a hand.

Function f : Data Set → **R**

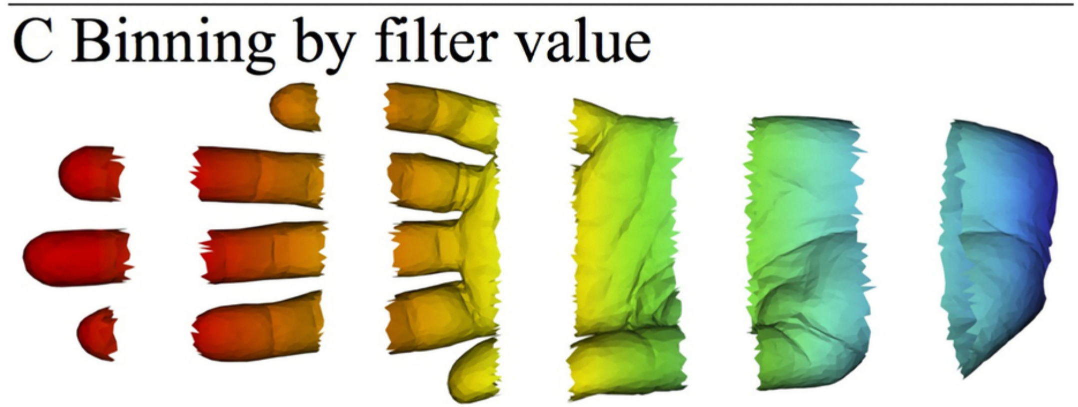
Example: x-coordinate

$f : (x, y, z) \rightarrow x$

Put data into overlapping bins.

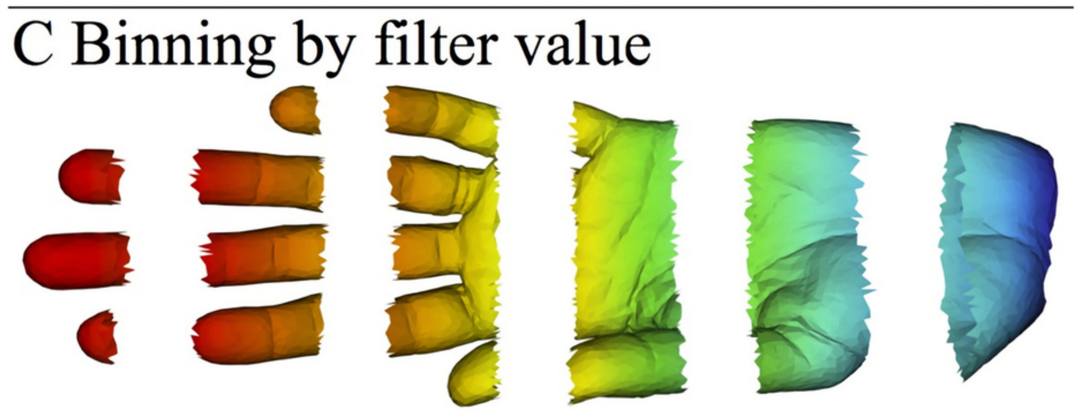
Example: $f^{-1}(a_i, b_i)$

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html



D) Cluster each bin

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html



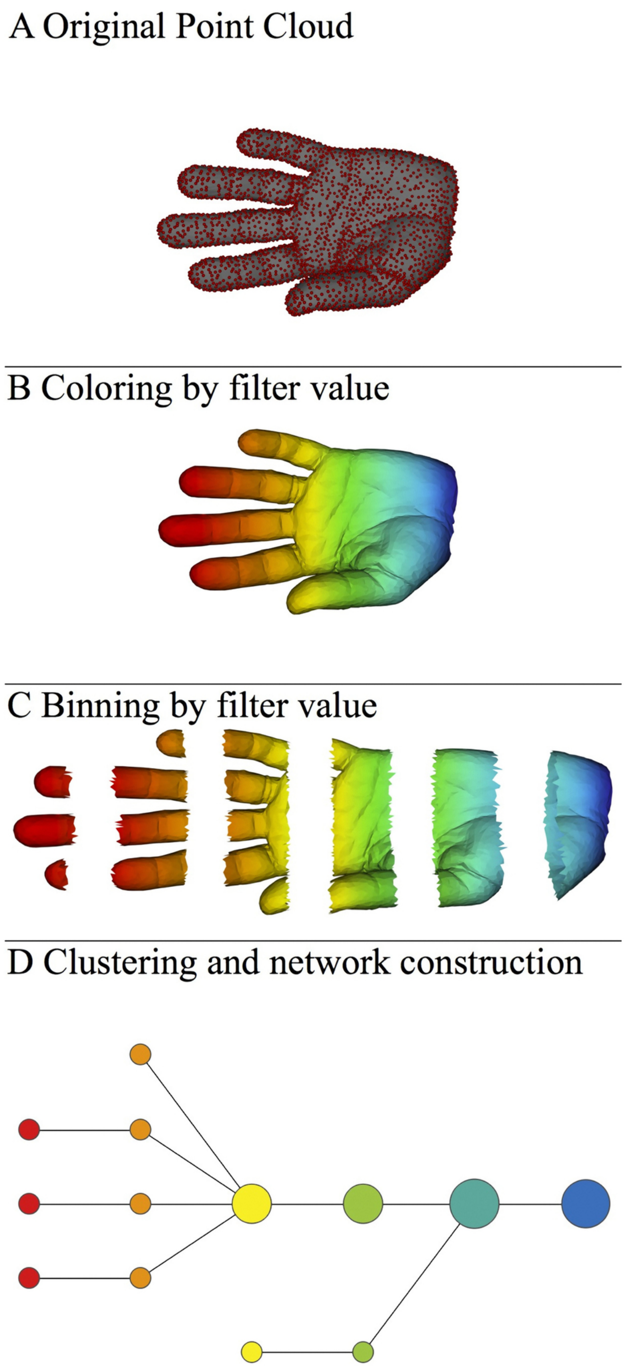
D) Cluster each bin

& create network.

Vertex = a cluster of a bin.

Edge = nonempty intersection between clusters

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html



A) Data Set

Example: Point cloud data representing a hand.

B) Function f : Data Set → **R**

Example: x-coordinate

$f : (x, y, z) \rightarrow x$

C) Put data into overlapping bins.

Example: $f^{-1}(a_i, b_i)$

D) Cluster each bin & create network.

Vertex = a cluster of a bin.

Edge = nonempty intersection between clusters

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Note: we made many, many choices

It helps to know what you are doing when you make choices, so collaborating with others is highly recommended.

A Original Point Cloud

We chose how to model the data set

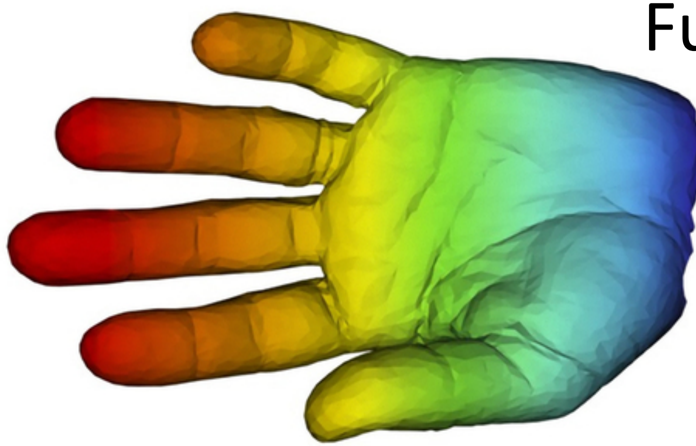
A) Data Set

Example: Point cloud data representing a hand.

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value

Chose filter function



Function f : Data Set → **R**

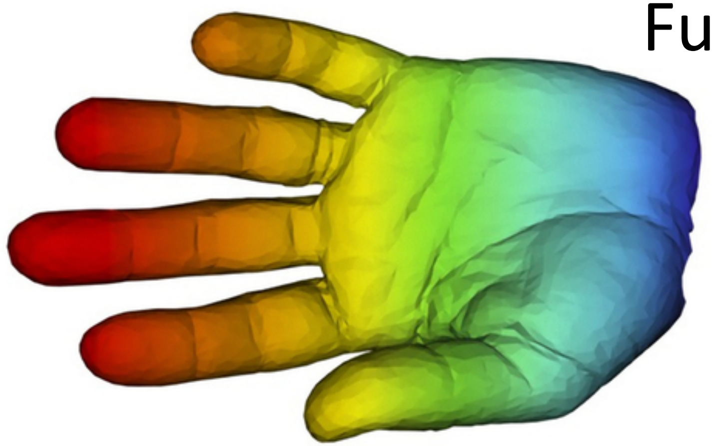
Ex 1: x-coordinate

$f : (x, y, z) \rightarrow x$

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value

Chose filter function



Function f : Data Set → **R**

Ex 1: x-coordinate

f : (x, y, z) → x

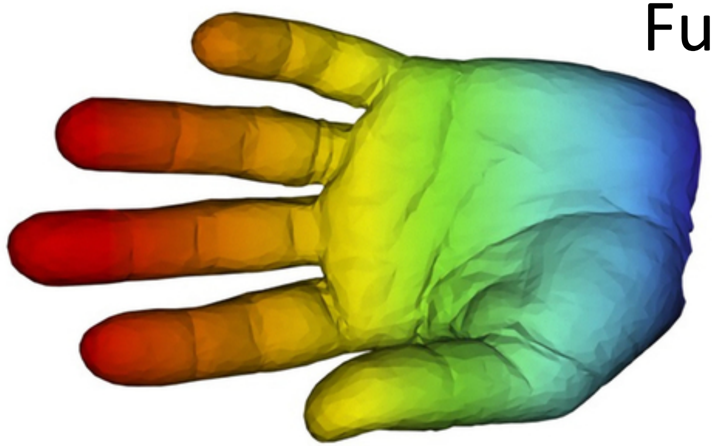
Ex 2: y-coordinate

g : (x, y, z) → y

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value

Chose filter function



Function f : Data Set → **R**

Ex 1: x-coordinate

f : (x, y, z) → x

Possible filter functions:

Principle component analysis

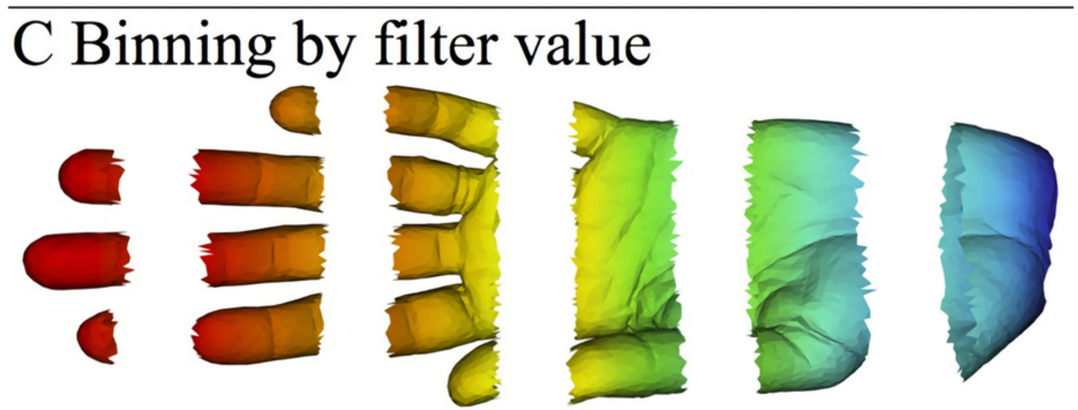
L-infinity centrality:

f(x) = max{d(x, p) : p in data set}

Norm: f(x) = | |x | | = length of x

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Chose bins



Put data into overlapping bins.

Example: $f^{-1}(a_i, b_i)$

If equal length intervals:

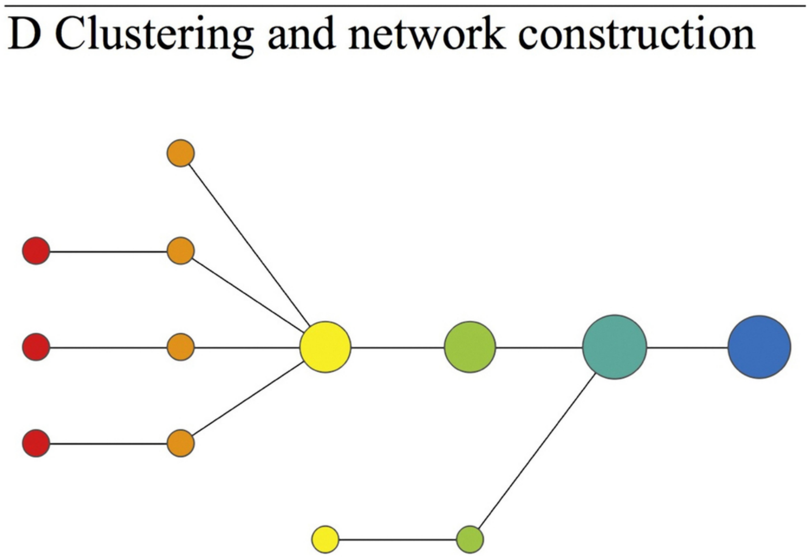
Choose length.

Choose % overlap.

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Chose how to cluster.

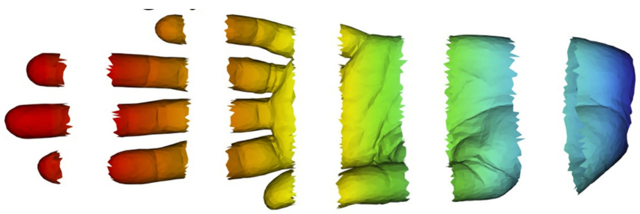
Normally need a definition of distance between data points



Cluster each bin & create network.

Vertex = a cluster of a bin.

Edge = nonempty intersection between clusters



http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Note: we made many, many choices

It helps to know what you are doing when you make choices, so collaborating with others is highly recommended.

Note: we made many, many choices

"It is useful to think of it as a camera, with lens adjustments and other settings. A different filter function may generate a network with a different shape, thus allowing one to explore the data from a different mathematical perspective."

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Note: we made many, many choices

"It is useful to think of it as a camera, with lens adjustments and other settings. A different filter function may generate a network with a different shape, thus allowing one to explore the data from a different mathematical perspective."

False positives???

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Note: we made many, many choices

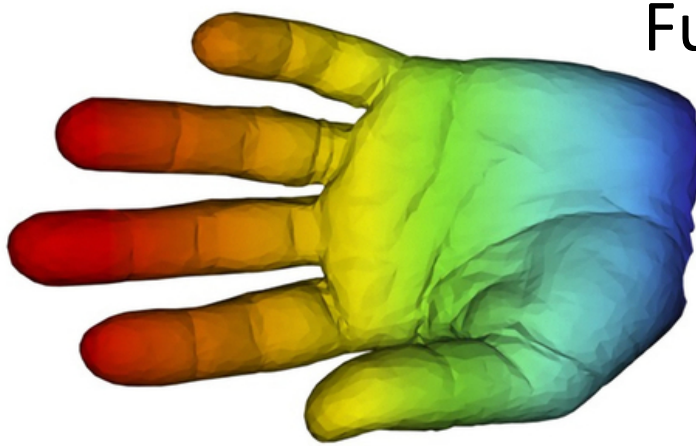
"It is useful to think of it as a camera, with lens adjustments and other settings. A different filter function may generate a network with a different shape, thus allowing one to explore the data from a different mathematical perspective."

False positives vs Persistence

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

B Coloring by filter value

Chose
filter
function



Function f : Data Set → **R**

Ex 1: x-coordinate

$f : (x, y, z) \rightarrow x$

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

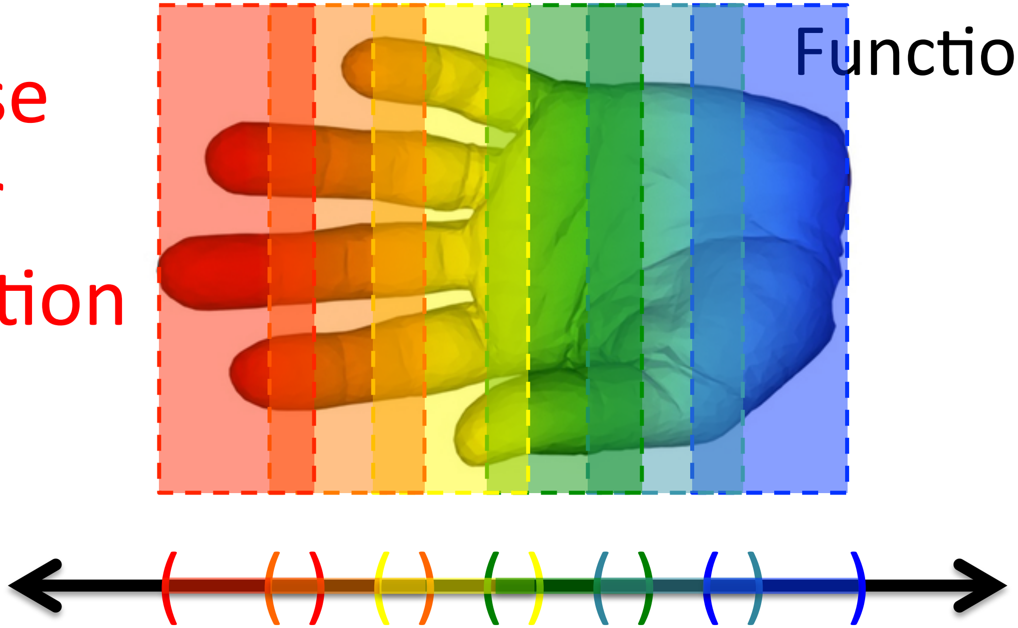
B Coloring by filter value

Chose filter function

Function f : Data Set → **R**

Ex 1: x-coordinate

$f : (x, y, z) \rightarrow x$



http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Chose filter

Only need to cover the data points.

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Chose
filter

Only need to cover the data points.

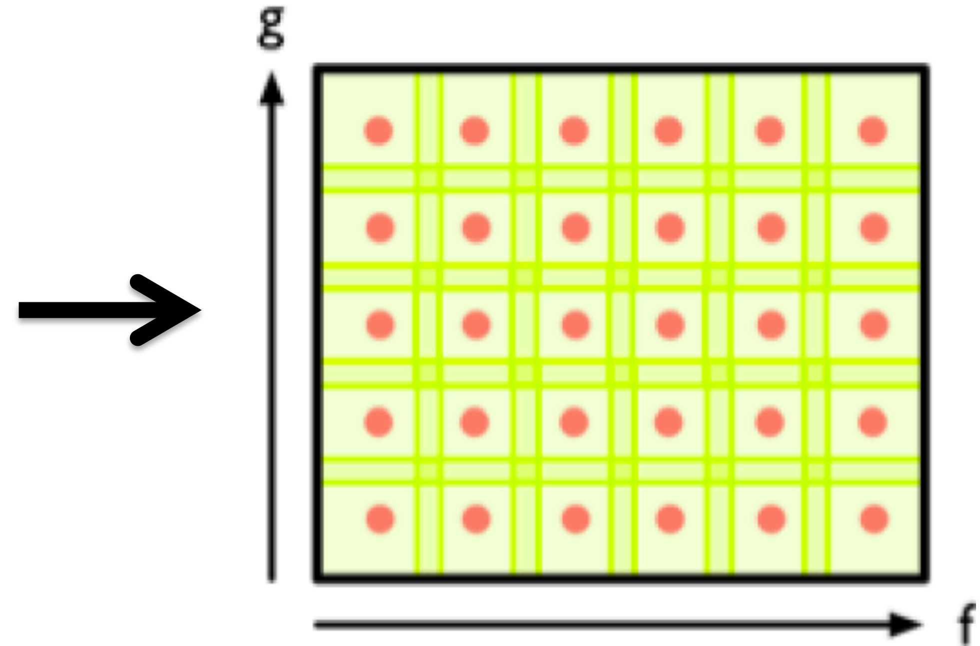
http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Chose filter

Only need to cover the data points.

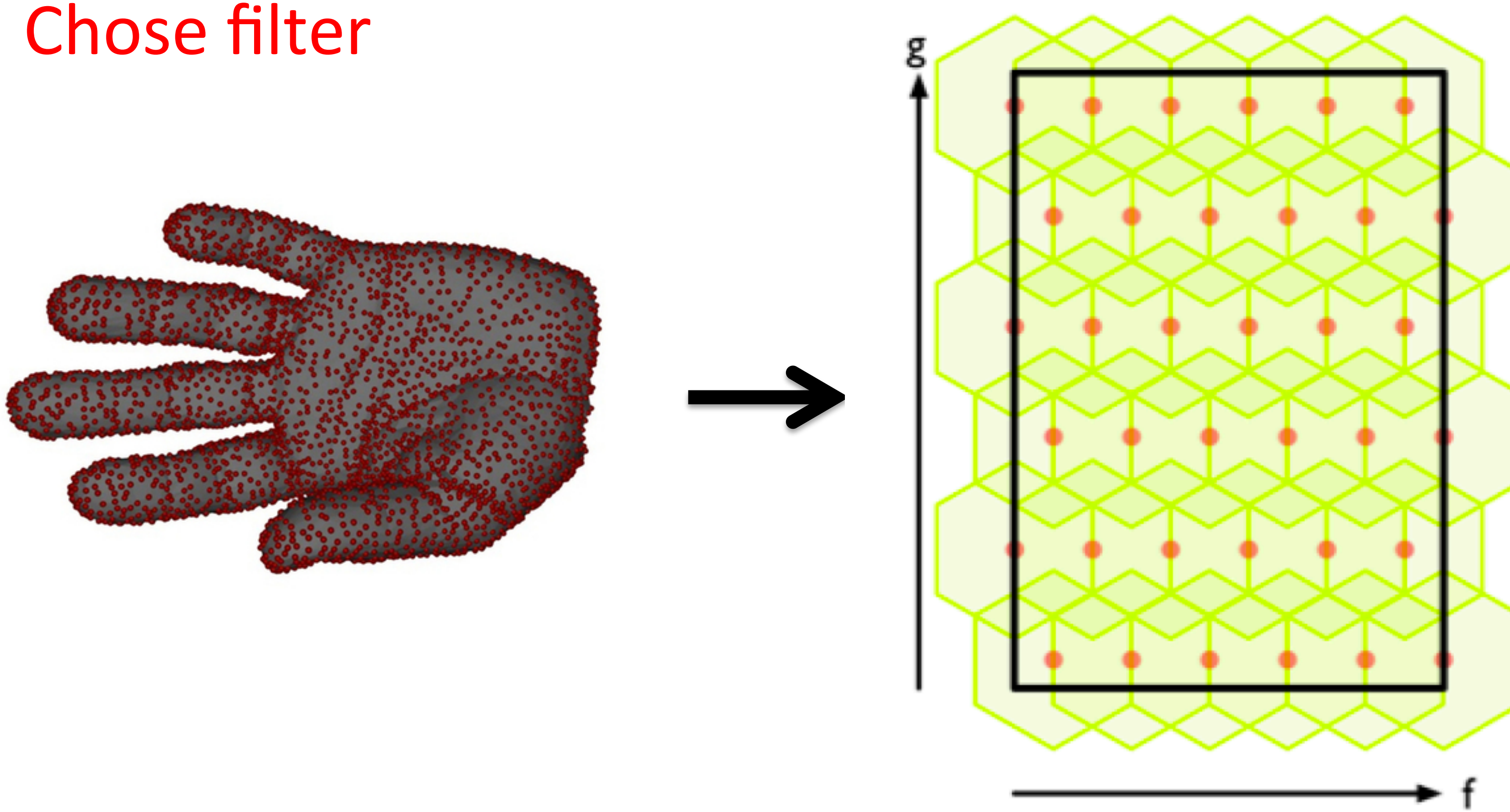
http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Chose filter



http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Chose filter



http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Topological Data Analysis (TDA): Three key ideas of topology that make extracting of patterns via shape possible.

1.) coordinate free.

- No dependence on the coordinate system chosen.
- Can compare data derived from different platforms
- vital when one is studying data collected with different technologies, or from different labs when the methodologies cannot be standardized.

http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Topological Data Analysis (TDA): Three key ideas of topology that make extracting of patterns via shape possible.

2.) invariant under "small" deformations.

- less sensitive to noise

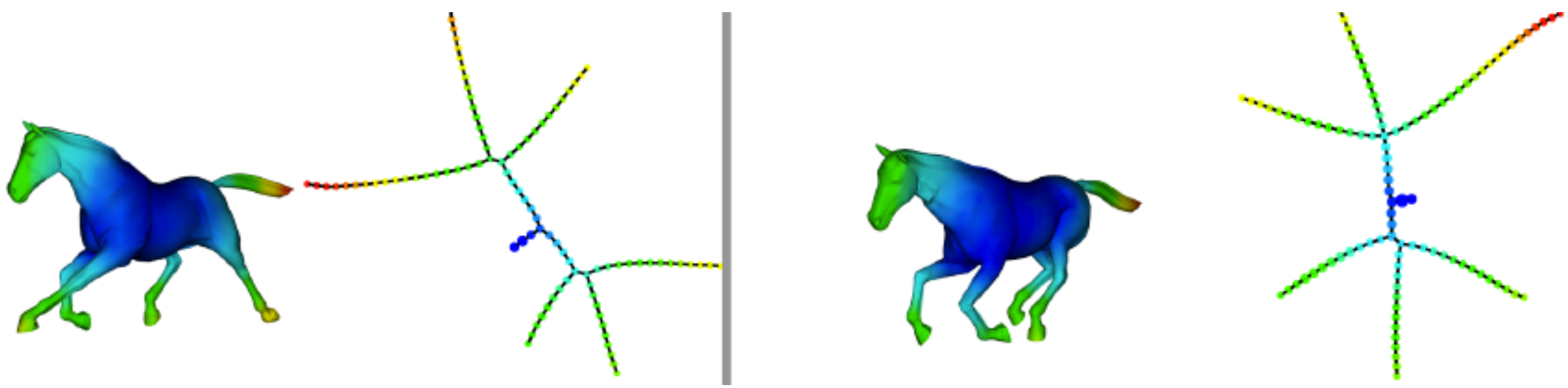


Figure from http://comptop.stanford.edu/u/preprints/mapperPBG.pdf

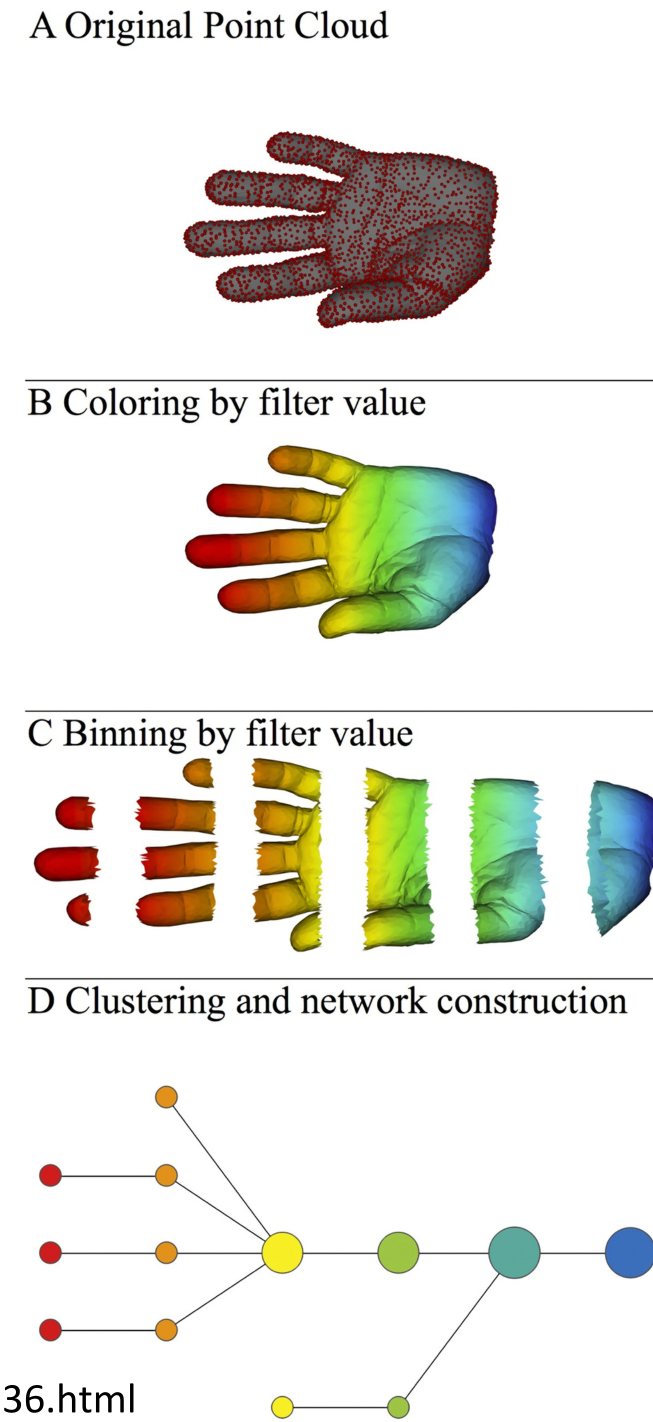
http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition, Singh, Mémoli, Carlsson

Topological Data Analysis (TDA): Three key ideas of topology that make extracting of patterns via shape possible.

3.) compressed representations of shapes.

- Input: dataset with thousands of points
- Output: network with 13 vertices and 12 edges.



http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html