Using Topological Data Analysis to Disentangle Complex Data Sets

Maurizio Sanarico
Chief Data Scientist, SDG Group

CERN - Geneve, October 20

sdg group

TOPOLOGY AND TOPOLOGICAL DATA ANALYSIS

- Topology is a branch of mathematics concerned with **spaces** and **maps**

  Formalizes notions of ***proximity*** and ***continuity***
- **Topological data analysis** bringing the algebraic topology paradigm from continuous to data cloud.
- It combines topology, geometry, statistics and computing.
- Topological features are defined to be «real» if they persist over a range of values of a control parameter
- TDA allows learning global representations from *local information*
- Provides a trade-off between complexity and feasibility

TOPOLOGICAL DATA ANALYSIS (TDA)

- TDA comprises “a collection of powerful tools that can quantify shape and structure in data in order to answer questions from the data’s domain.” [Elizabeth Munch]
- Motivation for TDA:
  - Data is huge, often high-dimensional, and complex
  - Traditional techniques have not “kept up”
    - i.e. Rely on *overly-simplistic* assumptions or are opaque.
- Basic Idea:

  Data has shape and shape carries important information

  This shape can be rigorously quantified via topological signatures

WHY TOPOLOGICAL DATA ANALYSIS

- It is powerful, with strong mathematical background, very general and complementary with respect to machine learning mainstream
- It is interpretable
- Can be used for
  - Exploratory / hypothesis generating analysis
  - Space partitioning for apply more focused local models and limit the curse of dimensionality
  - Generate new variables with somewhat untouched content (e.g. Metrization of persistent landscapes with L^p-norms)
  - Characterize the performance of a predictive model (Fiber of failure method).

TDA: THE MAIN STREAMS

- Persistent homology: extends homology (invariants: connected components and «holes» of various dimension) to data ➔ discern «true» from «artifacts».
  - Multimensional persistent homology, Local persistent homology, Zig-zag persistent homology
- Mapper: build a topological network
  - Multiscale mapper, Multinerve mapper, Stable paths in Mapper
    - All of them try to solve stability problems
- Morse-Smale regression and complex analysis: explore and characterize extrema in gradient fields

ANOTHER CLASSIFICATION OF TDA

1. **Topological simplification**
    1. Interpretation of unstructured data
    2. Dimensionality reduction
2. **Quantification of topological information**
    1. Persistent homology
    2. Distances between homological structures
3. **Topology of Gradient Fields**
    1. Morse-Smale complex analysis

THE GOAL OF TDA

- Basic Idea: **Data has shape**
- This shape can be rigorously quantified with topology:

  Using topological signatures
- Such signatures act as **summaries of the data**
- The Goal of TDA:

  Use tools from topology to make meaningful signatures of the data

  Topological signatures lead to topological **invariants**, and such invariants ***enable greater understanding of the relationships in—and transformations of—real data***

WHATS A TOPOLOGICAL SIGNATURE?

- Informally, a topology on a set is a *description of* ***how elements in the set are spatially related***
- Can be seen as a formalization of clustering of arbitrary shape
- A topological signature is a **simplified representation** of the topology of a given space
- Often, a [discrete] representation used as a topological signature is a **simplicial complex**
- ✴ 0-simplex == vertex
- ✴ 1-simplex == edge
- ✴ 2-simplex == triangle
- ✴ 3-simplex == tetrahedron
- ✴ … *k*-simplex == …

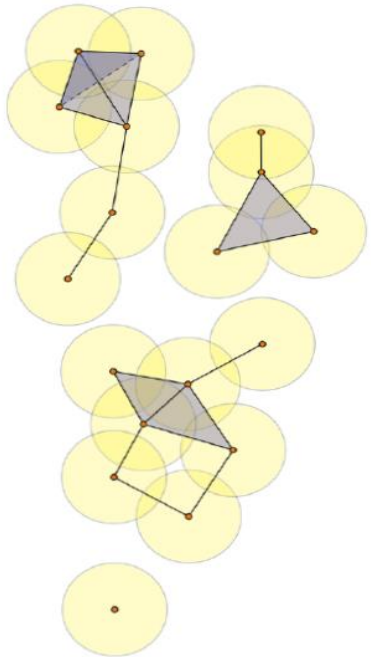
EXAMPLE OF A TOPOLOGICAL SIGNATURE

- Perhaps the most common topological signature is the so-called **Rips Complex** $VR_\varepsilon(X)$ obtained by building a simplex between all points which have pairwise distances less than ε.
- The simplicial complex formed by non-empty intersections of

$$B(X_1, \varepsilon) \cap B(X_2, \varepsilon) \cap \cdots \cap B(X_n, \varepsilon)$$

where

$$B(X_i, \varepsilon) = \{x \in X \mid d(X_i, x) \leq \varepsilon\}$$



TDA: PERSISTENT HOMOLOGY

Just a couple of sentences

- Characterize the topological content of data
- Find topological invariants (connected components, holes, voids,...) and discern them from artifact (topological noise)
- Main tools: persistence diagrams (PDs), barcodes, persistent landscapes and persistent images

    - Vectorization of PDs generates features that can be used in statistical models to add information that standard variables don’t contain

MAPPER: A TOPOLOGICAL SIGNATURE

- **Mapper**: *Perhaps* the most used signature in modern TDA *applications*
- Created by Singh, Mémoli, and Carlsson [Mapper]
- Simplest interpretation:
- Interprets any set of data in as “point cloud data”, turns data into a *simplified topological graph:*
  - *A set of connected nodes each one being a cluster.*

- “ Mapper takes as input both a possibly high-dimensional dataset and a map defined on the data and produces a **summary of the data** by using a cover of the codomain of the map.”

MAPPER ON TIME-RELATED EVENTS: IoTs AT HOME

1) TDA uses rectangular matrix of events (rows) by 3 time and 24 sensor sums variables (columns)

ID	activity	sensor	TIME: Current event time (seconds)	TIME: Window start time (seconds)	TIME: Window duration (seconds)
1	cook	I07	166	133	33
1	cook	M17	167	135	32
1	cook	M17	170	137	33
1	cook	AD1-A	173	137	36
1	cook	M17	173	140	33
1	cook	AD1-A	176	151	25
1	cook	AD1-B	176	153	23
1	cook	M17	178	153	25
1	cook	AD1-A	180	154	26
1	cook	AD1-B	180	154	26
1	cook	AD1-A	182	156	26
1	cook	AD1-B	185	156	29
1	cook	AD1-A	188	158	30
1	cook	AD1-B	188	160	28
1	cook	M17	190	161	29
1	cook	M17	194	161	33
1	cook	I05	199	162	37
1	cook	AD1-A	200	163	37
1	cook	M17	207	164	43
1	cook	I05	207	166	41
1	cook	M17	213	167	46
1	cook	AD1-A	215	170	45
1	cook	M17	219	173	46
1	cook	M17	221	173	48

SENSOR FOR CURRENT EVENT

AD1_A	AD1_B	AD1_C	D01	E01	I01	I02	I03	I04	I05	I06	I07	I08	M01	M07	M08	M09	M13	M14	M15	M16	M17	M18	M23	asterisk
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

SENSOR SUMS IN SLIDING WINDOW

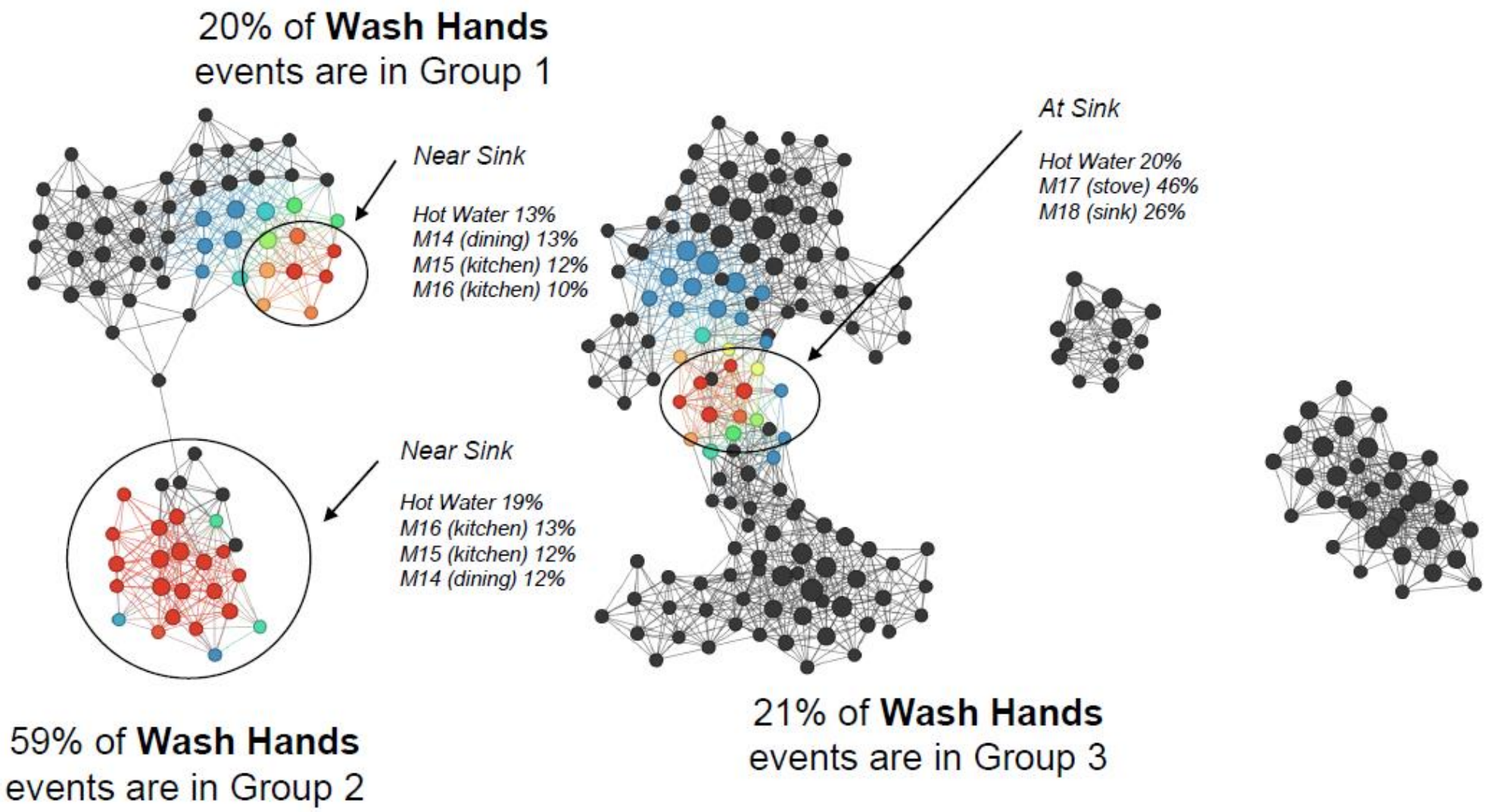
AD1_As	AD1_Bs	AD1_Cs	D01s	E01s	I01s	I02s	I03s	I04s	I05s	I06s	I07s	I08s	M01s	M07s	M08s	M09s	M13s	M14s	M15s	M16s	M17s	M18s	M23s	asterisk_s
0	0	0	2	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	9	1	0	0
0	0	0	2	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	9	1	0	0
0	0	0	2	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	10	0	0	0
1	0	0	1	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	10	0	0	0
1	0	0	1	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	10	0	0	0
2	0	0	1	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	9	0	0	0
2	1	0	0	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	9	0	0	0
2	1	0	0	0	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	9	0	0	0
3	1	0	0	0	1	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	9	0	0	0
3	2	0	0	0	0	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	9	0	0	0
4	2	0	0	0	0	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
4	3	0	0	0	0	0	2	1	1	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
5	3	0	0	0	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
5	4	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
5	4	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
5	4	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
5	4	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
6	4	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	7	0	0	0
6	4	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	8	0	0	0
6	4	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	7	0	0	0
6	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0
7	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0
7	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0
6	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0

2) Actual activity is "ground truth" (i.e., what we know and want the TDA to be able to detect).

3) The specific sensor triggered and ID are explanatory variables

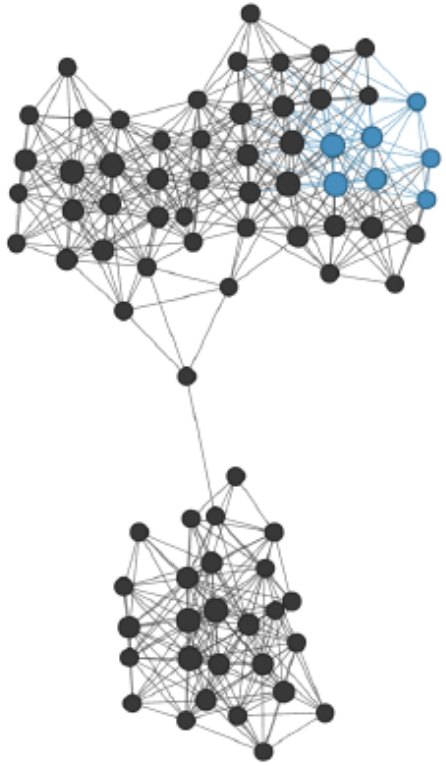
APARTMENT ACTIVITIES

Colored by # of washed hands events



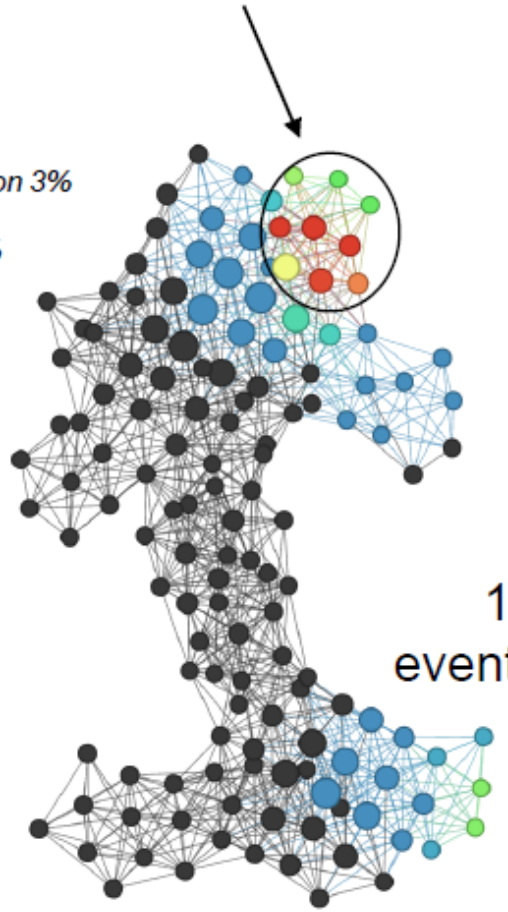
APARTMENT ACTIVITIES

Colored by # of clean events



Put items away and start cleaning

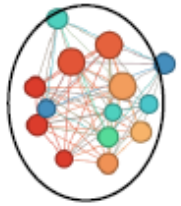
Hot Water 24%
Cabinet 6%
Oatmeal 5%
Bowl 4%
Measuring spoon 3%
Raisins 3%
M18 (sink) 24%



17% of **Clean** events are in Group 3

Cleaning

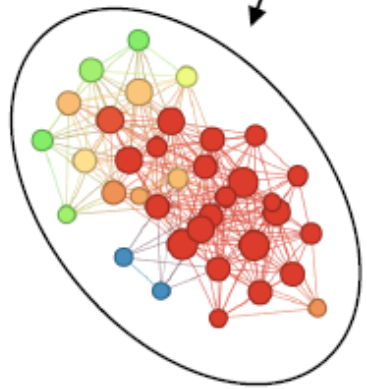
Hot Water 32%
Cold Water 8%
Medicine container 6%
M18 (sink) 11%



15% of **Clean** events are in Group 4

Cleaning

Hot Water 29%
Cold Water 15%
Medicine container 6%
M18 (sink) 11%
M15 (kitchen) 4%



68% of **Clean** events are in Group 5

TDA: MAPPER AT WORK

- Explore the topological and geometrical (local) structure of data
- Identify outliers, groups with different shapes (also shapes difficult to discover with other metrods, like flares, loops, local branches,...)
- Prepare data for further analysis
- Characterize the performance of a model (Fibers of Failure method) giving the opportunity to understand where to focus improvement actions (e.g., oversampling regions of the data space or define a special layer in a deep neural network as a correction layer).
- Can also incorporate supervised learning methods (e.g. Supervised UMAP, semi-supervised UMAP).

TDA: FOCUS ON MAPPER

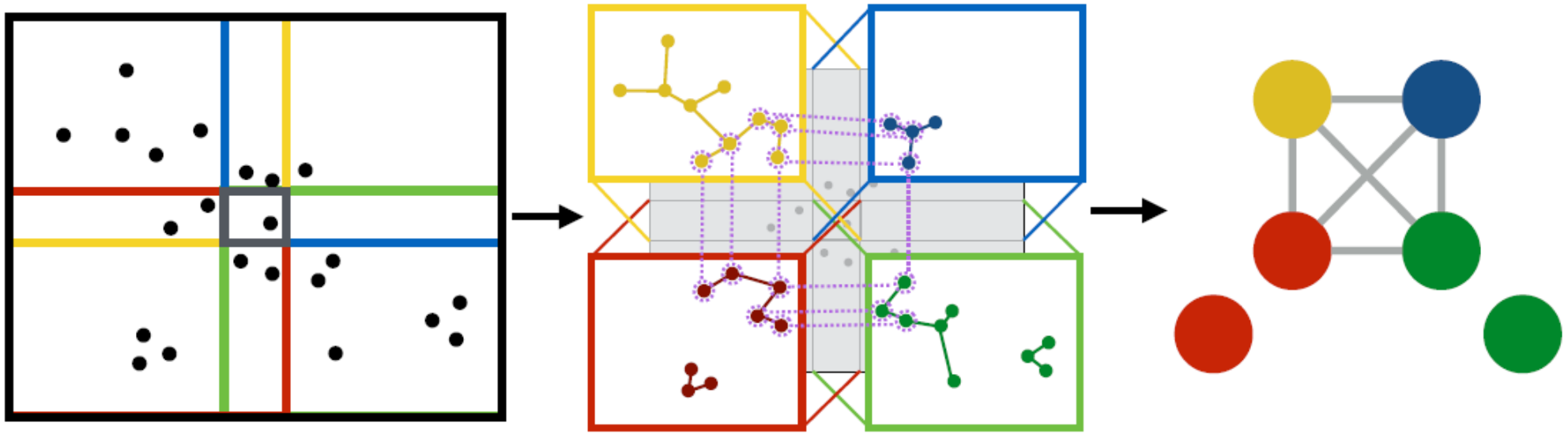
- Relatively easy to program
- Conceptually not so simple
- From modeling point of view requires various not obvious choices
  - A recent development try to overcome this situation by using learnable filter functions (Deep Graph Mapper: Seeing Graphs through the Neural Lens – Bodnar et al).
- Very flexible
- Stability may be an issue (theoretical solution through multiscale version of Mapper or through some parameter selection backed by theorems)
- Somewhat loosely related methods (Mapper can embed all of them):
  - Projection pursuit
  - Isomap
  - Locally linear embedding
  - MDS

TDA MAPPER: A BIRDVIEW (FROM M. PIEKENBROCK, 2018)

$f : X \to Z$

$X_\alpha = f^{-1}U_\alpha$

$f^*(\mathcal{U}) \to M(\mathcal{U}, f)$



$\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$

$C(X_\alpha) \to f^*(\mathcal{U})$

(1-skeleton example)

MAPPER: THE PARAMETERIZATION

1. Define a reference map / filter function: f: $X \rightarrow Z$
2. Construct a covering $\{U_\alpha\}_{\alpha \in A}$ of Z (A is called the index set)
3. Construct the subsets X_α via $f^{-1}(U_\alpha)$
4. Applying a clustering algorithm C to the sets X_α
5. Obtain a cover $f^*(U)$ of X_α by considering the path-connected components of X
    1. Clusters from nodes / 0-simplices
    2. Non-empty intersections from edges / 1-simplices
6. The Mapper is the nerve of this cover: $M(U,f) = N(f^*(U))$

SOME FURTHER CONCEPTS

Mapper provides a compressed **description** of **the shape of a data set** (expressed via the codomain of the mapping)

Mappers is quite general:

- *Any* mapping function can be used (or a composition of different functions)
- Covers may be constructed *arbitrarily*
- *Any* clustering algorithm may be used
- The resulting graph is often much easier to interpret than, e.g. individual scatter plots of pairwise relationships
- Mapper is often paired with **high-dimensional data**, and is generally used to see the 'true' shape or structure of the data

Mapper on Amazon Reviews

CLUSTERING AMAZON REVIEWS

Working with 11000 Amazon reviews of a coffee machine for domestic use, the goal was to create clusters with similar content of the test message.

The analysis was performed using a purely text-based filter function, i.e., in a totally unsupervised way.

The latter analysis creates a force towards building clusters that are homogeneous in rating and heterogenous in text.

It was also assessed whether content was related to customer ratings.

THE AMAZON REVIEWS CASES

Parameterization of Mapper

Input data: Text of Amazon reviews

Projection: TFIDF vectorization ➔ ngrams up to order 6 ➔ TruncatedSVD with 110 components ➔ ISOMAP / UMAP with 2 components

Filter function: ➔ (1) UMAP / ISOMAP with two components

Resolution: 4 - 5 – 7 - 10 ➔ selected 5 for final representation

Overlap: 20%-30%

Local clustering algorithm: Agglomerative Clustering with (n_clusters=3,
linkage="complete",
affinity="cosine")

[-] CLUSTER DETAILS [+] MAPPER SUMMARY [+] HELP

Latent Semantic Char-gram Analysis with Isometric Embedding

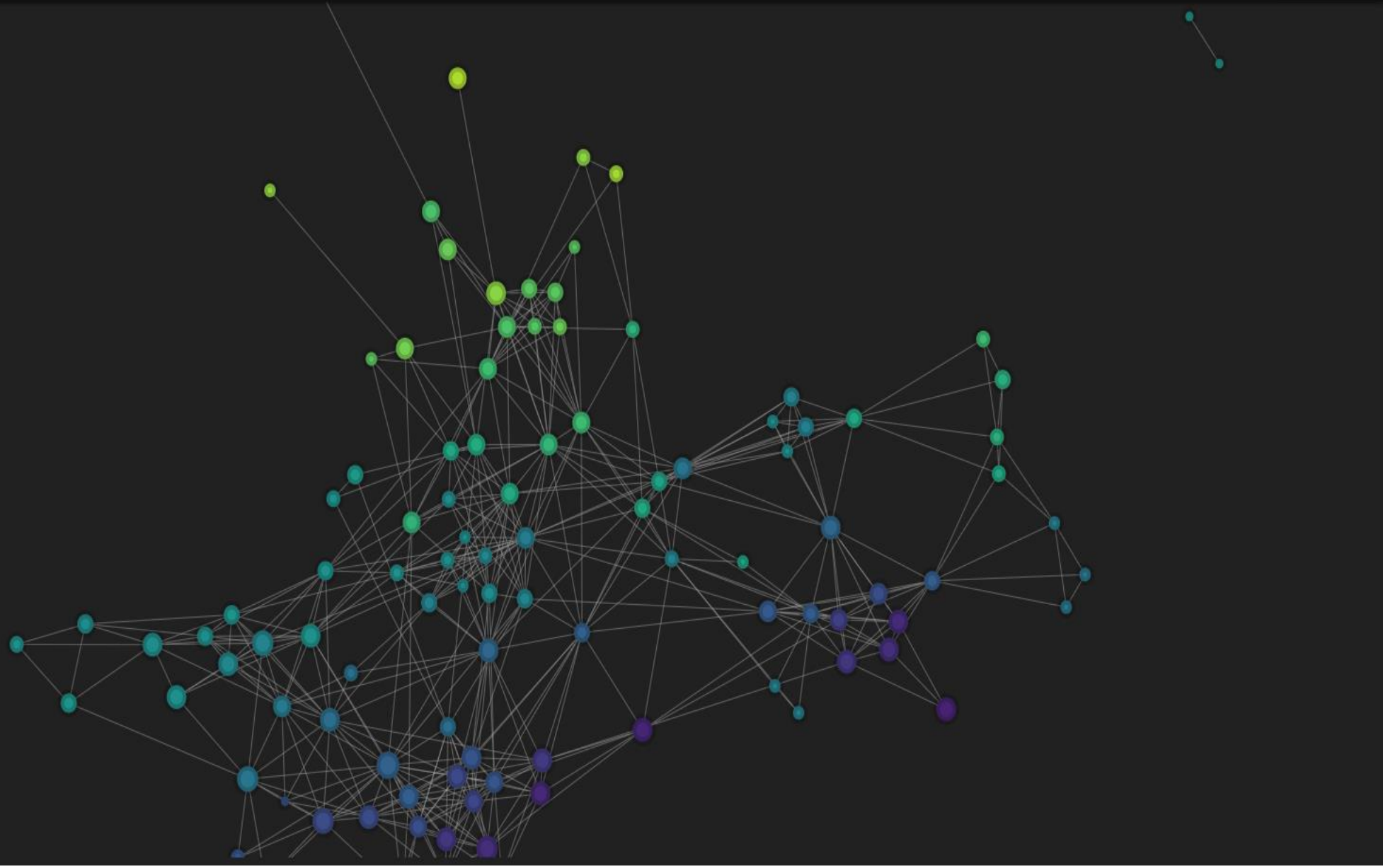
Lens	Mean	Max	Min
ISO1	0.877	0.947	0.822
ISO2	0.783	0.918	0.721

ABOVE AVERAGE

Feature	Mean	STD
stars	0.602	2.7x
love	0.231	1.3x
excellent	0.108	0.6x
great machine	0.034	0.2x
price	0.038	0.1x
best	0.037	0.1x
product	0.046	0.1x
great coffee	0.032	0.0x
worth	0.018	0.0x

BELOW AVERAGE

Feature	Mean	STD
coffee	0.054	0.6x
espresso	0.003	0.5x



[-] CLUSTER DETAILS [+] MAPPER SUMMARY [+] HELP

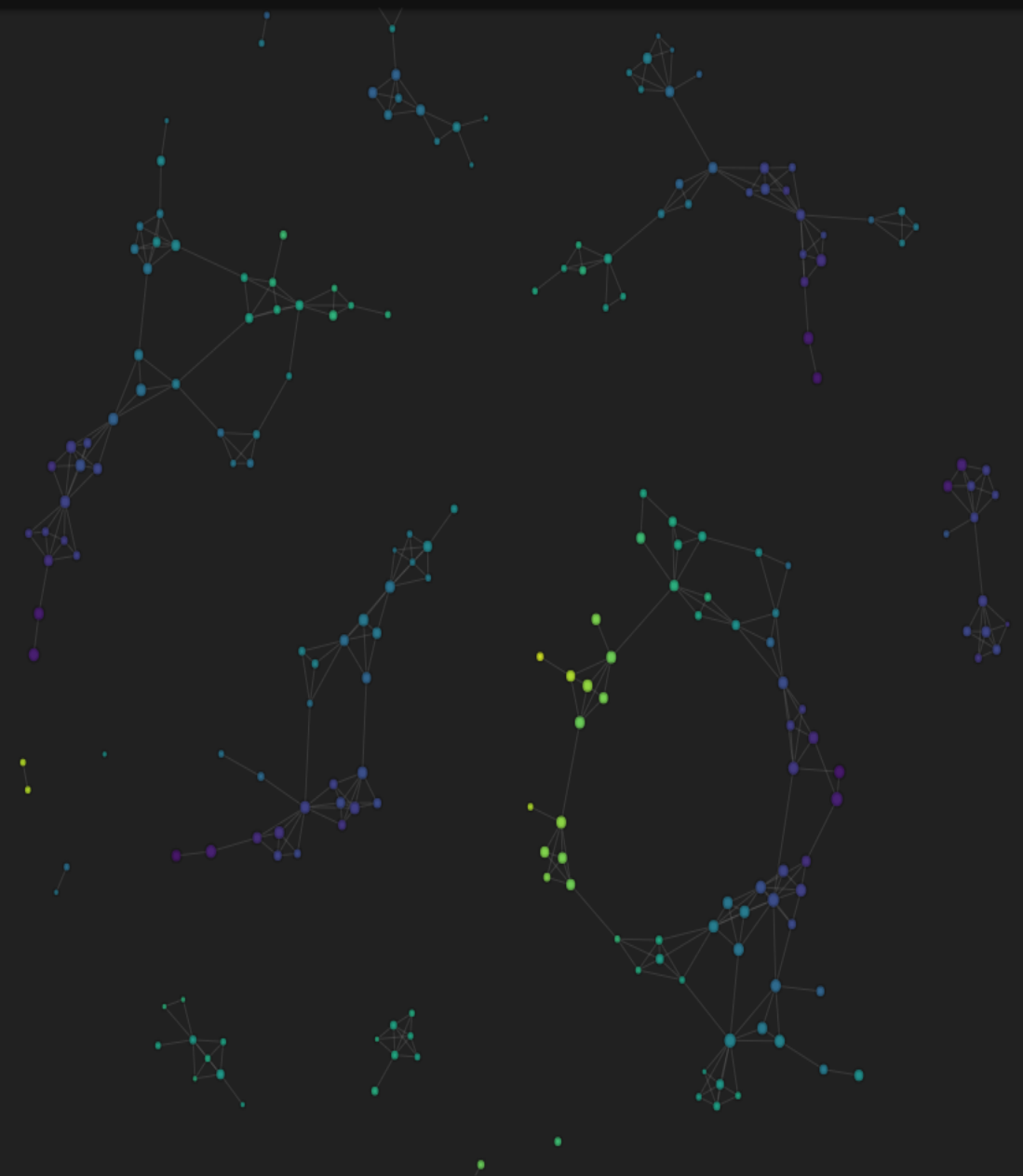
Latent Semantic Char-gram Analysis with Isometric Embedding

PROJECTION STATISTICS

Lens	Mean	Max	Min
ISO1	0.104	0.201	0.027
ISO2	0.542	0.633	0.456

ABOVE AVERAGE

Feature	Mean	STD
problem	0.047	0.5x
delonghi	0.079	0.5x
did	0.05	0.4x
machine	0.199	0.4x
service	0.055	0.4x
months	0.068	0.4x
water	0.092	0.4x
espresso	0.12	0.4x



THE AMAZON REVIEWS CASE

Reviews were recoded as a binary variable with rating 1,2,3 = 0 and 4,5 = 1.

Predictive ability of clusters was assessed by splitting data with unsupervised clusters (based only on review texts) into train and test samples and then apply random forest.

THE AMAZON REVIEWS CASE

We used the cluster IDs as the unique feature in a predictive model (Random Forest) ➔ Cluster ID actually represents groups of connected clusters (sub-networks).

We ran the RF algorithm 10 times on a set of train/test samples with 70/30 split, 15 times on a set of train/test samples with 80/20 split and 30 times on a set of train/test samples with 90/10 split.

The results were then pooled to account for sampling variability.

It is interesting noting that individual cluster are interesting also from the point of view of trying to understand missclassified cases and to pick patterns of text that are associated to conflicting ratings

On next page the average results for each random split

THE AMAZON REVIEWS CASE

70/30 split

	Predicted Rating	
Observed Rating	0	1
0	5230	999
1	1090	5781

Sensitivity = 84.1%
Specificity = 83.9%
Precision = 85.3%

80/20 split

	Predicted rating	
Observe rating	0	1
0	4992	1158
1	1101	5909

Sensitivity = 84.3%
Specificity = 81.1%
Precision = 83.6%

THE AMAZON REVIEWS CASE

90/10 split

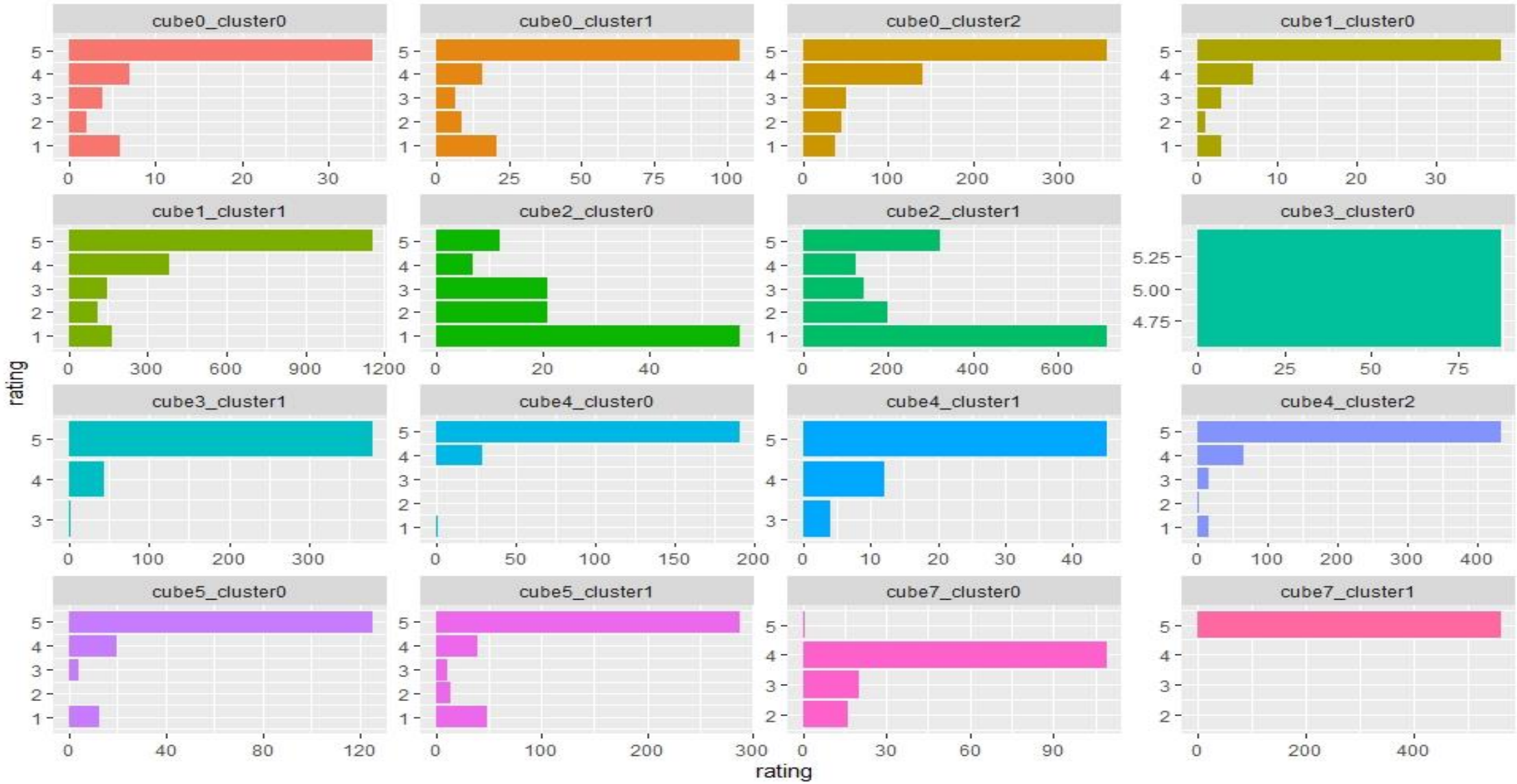
	Predicted rating	
Observed rating	0	1
0	5256	997
1	1048	5879

Sensitivity = 84.8%
Specificity = 84.0%
Precision = 85.4%

Therefore, clusters achieve a good predictive performance and being unsupervised can be applied to new reviews without overfitting problems.
The Fiber of Failure method and the characterization of individual clusters trying to find what distinguishes good predictions from incorrect predictions.

THE AMAZON REVIEWS CASE

Cluster Examples (showing some of the largest) with rating



THE OTHER APPLICATIONS

- Discover sequences of alarms/warnings, identifying the first one and characterize the impact on performance on a packaging line with 1800 types of alarms and warnings (Alarm flood problem).
- Cluster WES data (Whole Exome Sequence) to characterize the genetic contribution to Covid19 severity (still to be submitted for publication).
- Use in a NLP pipeline to classify CRM messages into categories based on the content.

SOME NEW INTERESTING DEVELOPMENTS

1. Fibers of Failure: Classifying Errors in Predictive Models (Carlsson et al.).
2. Towards a topological-geometrical theory of group equivariant non-expansive operators for data analysis and machine learning – GENEOS (Bergomi et al.)
3. Deep Graph Mapper: Seeing Graphs through the Neural Lens (Bodnar et al.)
4. Parametric machines: a fresh approach to architecture search (Vertechi et al.)
5. Partition-Based Graph Abstraction (Wolf et al.)

THE FIBER-OF-FAILURE (FOF) METHOD

- The next slide illustrates the FoF method applied to predicted failures of a optical fiber network:
- The color of the dot is related to accuracy of prediction going from perfect prediction of positive cases (deep blue) to perfect prediction of negative cases (deep red).
- Intermediate colors point out to varying degrees of accuracy, the size of the dot are proportional to number of cases.
- In the picture the yellow dots represents worst cases.
- One of the goals is to tell what distinguishes good from bad predictions

Test

p_3	0.63	0.671	0.586
p_4	9.124	9.698	8.519

ABOVE AVERAGE

Feature	Mean	STD
default	12.212	7.6x
dsdmm	0.219	7.1x
equipment	4.798	7.1x
config	1.162	7.0x
diagnose	0.03	6.9x

BELOW AVERAGE

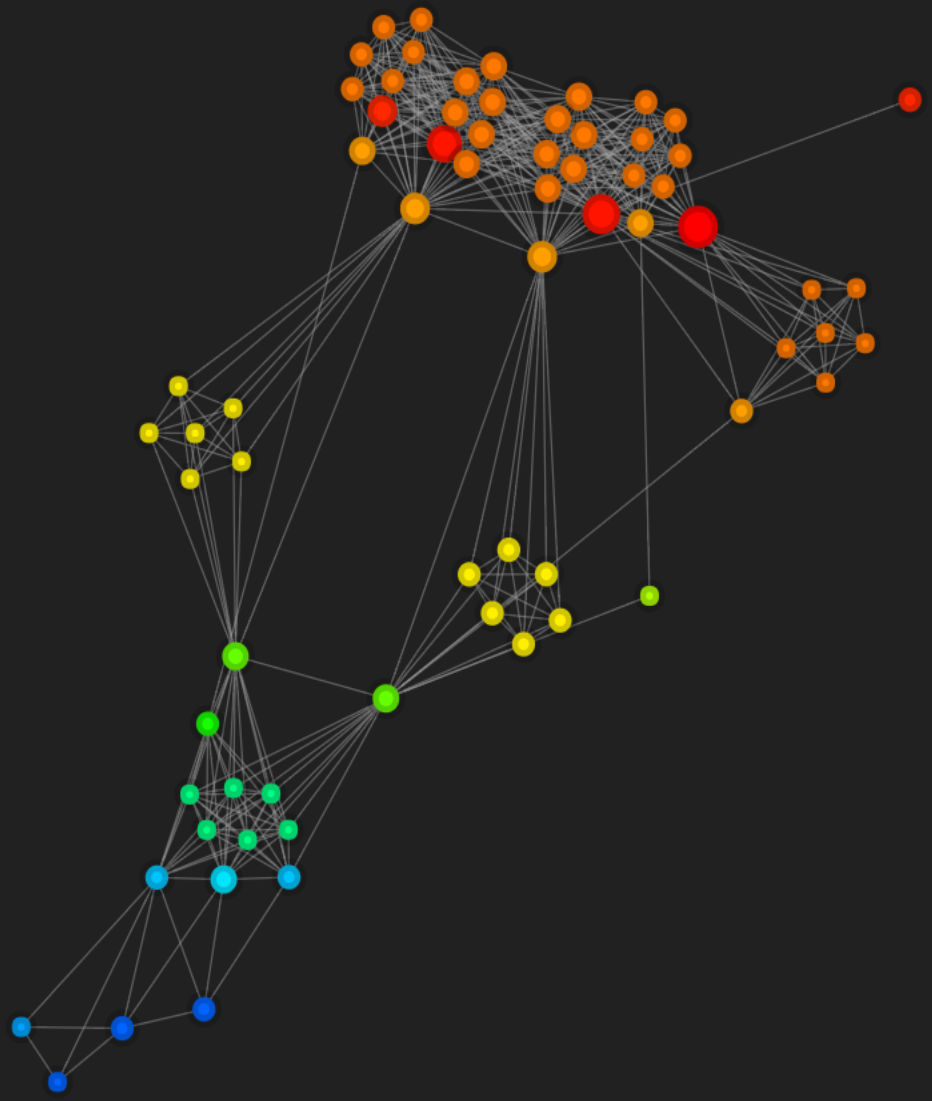
Feature	Mean	STD
hwpstnctprightcw	0.0	0.3x
hwpstnctprightcallhold	0.0	0.3x
hwpstnctprightmwi	0.0	0.3x
hwpstnctprightct	0.0	0.3x
hwpstnctpright threeparty	0.0	0.3x

SIZE

11

MEMBERS

1 1 1 1 1 1 1 1 1 1 1



Graph Meta [-]

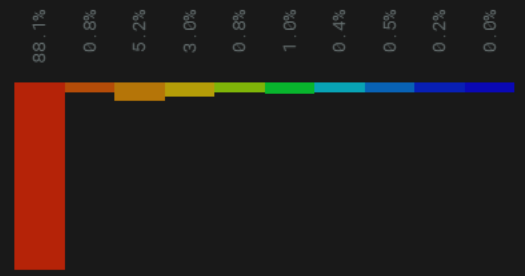
NODES 69

EDGES 473

TOTAL SAMPLES 9914

UNIQUE SAMPLES 8280

DISTRIBUTION



THE TOOLS

Two among the many

- Kepler Mapper: Nice HTML interactive output, flexible and powerful, TDA limited to Mapper.
- Giotto TDA: A general framework for TDA, open source but also available to commercial enterprises. Actively developed and Maintained.
- Giotto implementation of Mapper offers some interesting features,one particularly interesting is the interactive graph with the possibility to change parameters and see immediately the result.

Two Mapper implementations: Kepler vs. Giotto-tda

To date, several implementations of Mapper are available. I present two Python open source alternatives.

	Kepler Mapper	Giotto-tda Mapper
First release in	2017	2019
Stars on GitHub*	507	308**
Visualization aspects	Highly developed	Basic, incl. 3D
Hyperparameter interactivity	✗	✓
Compatibility with scikit-learn	Partial	Complete

*Updated on 18/10/2020 **Stars for the entire giotto-tda library

Focus on Giotto-tda

Mapper is one of the many components available in Giotto-tda.

Using memory caching, the user can interactively change the parameters of the Mapper Algorithm and visualise the results.

Giotto-tda is a high-performance topological machine learning toolbox in Python built on top of scikit-learn. (more on https://giotto-ai.github.io/gtda-docs/0.3.0/index.html)

Principles:

- Seamless integration with scikit-learn
- Code modularity
- Standardisation
- Innovation
- Performance
- Data structures

Main tools available:

- Mapper
- Persistence homology
- Persistence diagrams
- Topological features extraction from graphs
- Topological features extraction from time series
- Topological features from images

What is Giotto?

Giotto represents an ecosystem of AI tools and services controlled by the Swiss company L2F.

Giotto by L2F

In partnership with: EPFL, Blue Brain Project, heig-vd, ReDS, Schweizerische Eidgenossenschaft, Confédération suisse, Confederazione Svizzera, Confederaziun svizra, Swiss Confederation, Innosuisse – Swiss Innovation Agency

Python Open source development

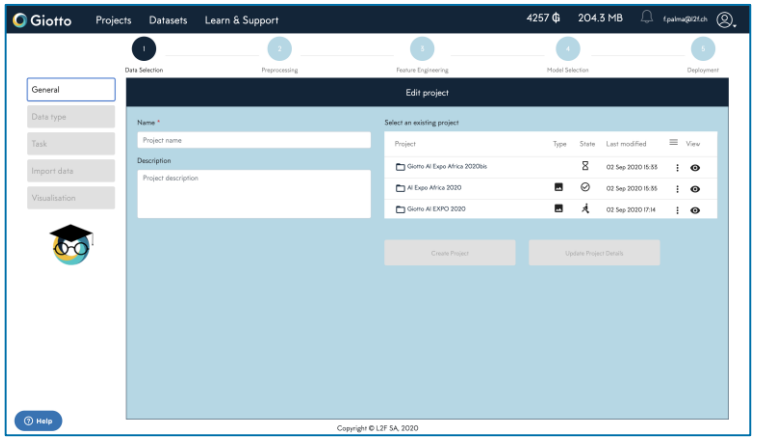
Giotto-TDA

Giotto-Time

www.giotto.ai/libraries

https://github.com/giotto-ai

No-code AI, visual programming platform



www.giotto.ai

Training and enterprise solutions

www.giotto.ai/automl

business@giotto.ai

SPEAKER INFORMATION

Maurizio Sanarico

Chief Data Scientist, SDG Group

+ 39 3485165701

Maurizio.sanarico@sdggroup.com

sdggroup.com

SDG GROUP

Strategy.

GLOBAL

CORPORATE **PROJECTS**

+ 1.000

SPECIALIZED **CONSULTANTS**

100 M €

GROSS **SALES**

+ 500

ACTIVE **CLIENTS**

+ 20

INTERNATIONAL **OFFICES**

Decision.

One of the world's fastest growing Data Analytics consulting firms. Thanks to our proven expertise in Data Analytics, we support organisations in designing a Data Driven strategy that enables them to be increasingly competitive in their target market.

Governance.

"We are specialized niche players, offering an in-depth analytics expertise and working with clients to empower their business strategy to become successful data-driven enterprises."

Luca Quagini, Founder & CEO SDG Group

SDG GROUP

Alcune aziende che hanno già scelto SDG Group